Universidade de Aveiro Departamento de Biologia
2010

Tatiana Barros Reis Cereija

Caracterização da inibição da tripsina pela boofilina

Universidade de Aveiro Departamento de Biologia
2010

Tatiana Barros Reis Cereija

Caracterização da inibição da tripsina pela boofilina

Characterization of trypsin inhibition by boophilin

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular, realizada sob a orientação científica do Doutor Pedro José Barbosa Pereira, Professor auxiliar convidado do Departamento de Biologia da Universidade de Aveiro.

Apoio financeiro da FCT e do FSE no âmbito do III Quadro Comunitário de Apoio.

o júri

presidente

Professora Doutora Adelaide Almeida
Professora Auxiliar do Departamento de Biologia da Universidade de Aveiro

Professor Doutor Pedro José Barbosa Pereira
Professor auxiliar convidado do Departamento de Biologia da Universidade de Aveiro
Investigador principal no Instituto de Biologia Molecular e Celular, Porto

Doutora Isaura Simões
Investigadora auxiliar no Centro de Neurociências e Biologia Molecular da Universidade de Coimbra

palavras-chave

Boofilina, trombina, tripsina, Kunitz, inibidor, cascata de coagulação, anticoagulante, trombose.

resumo

No decurso da evolução, enquanto os vertebrados desenvolviam um complexo sistema para prevenir as perdas de sangue, os hematófagos criavam mecanismos anticoagulantes capazes de contra-atacar a reposta hemostática do hospedeiro, por forma a facilitar a extração, o armazenamento e a digestão do sangue. Muitos destes mecanismos assentam na inibição específica da trombina, proteína que tem um papel central na hemostasia. Apesar do elevado número de inibidores naturais de trombina, apenas uma pequena parte foi caracterizada bioquimicamente e, dentro desta, só alguns inibidores foram estudados de um ponto de vista estrutural. De todos aqueles com estrutura conhecida, a boofilina é o único que possui dois domínios canónicos do tipo Kunitz. No entanto, apesar de possuir domínios canónicos, a boofilina inibe a trombina através de um mecanismo não-canónico, no qual o seu N-terminal é inserido no centro catalítico da trombina enquanto que o domínio C-terminal interage com o exosite I da proteinase. Além disso, ao contrário de todos os outros inibidores caracterizados, a boofilina, em complexo com a trombina, é capaz de interagir com uma proteinase serínica adicional do tipo tripsina. Esta segunda interação envolve unicamente o domínio N-terminal, presumivelmente através do *loop* reactivo. *In vivo*, é possível que para além de inibir a trombina, a boofilina seja também capaz de inibir outra proteinase serínica da cascata de coagulação como o fXa, resultando não só na inibição da trombina mas também na diminuição da sua produção. Este aspecto faz da boofilina um bom modelo para a desenvolvimento de novos anticoagulantes terapêuticos. Pretende-se, por isso, clarificar este segundo mecanismo de inibição através do estudo da interação entre a boofilina e a tripsina. A boofilina (D1D2) e o seu domínio N-terminal isolado (D1) foram expressos em *Pichia pastoris*. Ambas as proteínas foram purificadas por cromatografia de afinidade numa coluna de tripsina seguida, no caso da D1D2, por cromatografia de troca iónica. Uma vez purificadas, a constante de inibição de cada uma das proteínas foi determinada e a D1 foi cristalizada por difusão vapor. Os cristais obtidos eram ortorrômbicos (grupo espacial $P2_12_12_1$), tinham uma percentagem invulgarmente baixa de solvente (23%) e difrataram com uma resolução mínima de 1.80 Å. A estrutura cristalográfica da D1 releva uma conformação similar à observada no complexo trombina-boofilina. O complexo tripsina-D1 foi também preparado *in vitro* e a sua cristalização está em curso.

keywords Boophilin, thrombin, trypsin, Kunitz, inhibitor, coagulation cascade, anticoagulant, thrombosis.

abstract During evolution, while vertebrates developed a complex system to prevent blood loss, blood-feeding animals evolved anticoagulant mechanisms to counteract the haemostatic response of their hosts in order to facilitate blood drawing, storage and digestion. Many of these anticoagulant mechanisms rely on the specific inhibition of thrombin, a central enzyme in haemostasis. Although a number of a natural thrombin inhibitors from haematophagous animals have been described only a few have been fully characterised biochemically, and even fewer have been studied from a structural viewpoint. From those structurally characterised to date, only boophilin displays two tandem canonical Kunitz fold domains. However, despite its regular BPTI-like domains, boophilin inhibits thrombin by a non-canonical mechanism, in which its N-terminus inserts into the active site cleft of thrombin, while the C-terminus Kunitz domain interacts with the exosite I of the proteinase. Moreover, in contrast to all the other natural thrombin inhibitors characterised, when in complex with thrombin boophilin retains the capability to interact with an additional (non-thrombin) trypsin-like serine proteinase molecule. In this second interaction only the N-terminal domain is involved, presumably through its conserved reactive-loop. *In vivo*, it is likely that besides inhibiting thrombin, boophilin can target another serine proteinase of the coagulation cascade, such as fXa. This would result not only in thrombin inhibition, but also in the impairment of thrombin generation, making boophilin a good model for the design of new therapeutic anticoagulants. For this reason, this work aimed at clarifying this second inhibition mechanism by studying the trypsin-boophilin complex. Boophilin (D1D2) and its N-terminal domain (D1) were expressed in *Pichia pastoris*. Both proteins were purified by immobilized-trypsin affinity chromatography followed, in the case of D1D2, by ion-exchange chromatography. Once purified, the inhibition constant of both proteins was determined and D1 was crystallized by vapour diffusion. The crystals obtained were orthorhombic (space group $P2_12_12_1$), had an unusually low solvent content of 23% and diffracted X-rays to beyond 1.80 Á. The crystallographic structure of the first domain of boophilin reveals a conformation of the reactive site loop similar to that observed in the thrombin-boophilin complex. The D1-trypsin complex was also prepared *in vitro* and its crystallization is underway.

Contents

Abbreviations

ADP	Adenosine diphosphate
Ala	Alanine
APS	Ammonium persulfate
Arg	Arginine
Asp	Aspartic acid
AU	Absorbance units
BIS-TRIS	Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
BMGY	Buffered glycerol-complex medium
BMMY	Buffered methanol-complex medium
BSA	Bovine serum albumine
Bz	Benzoyl
cDNA	Complementary deoxyribonucleic acid
gDNA	Genomic deoxyribonucleic acid
Cys	Cysteine
C-terminus	Carboxyl-terminus
D1	Isolated first domain of boophilin
D1D2	Full-length boophilin
DIC	Disseminated intravascular coagulation
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
dTTP	2'-deoxythymidine 5'-triphosphate
EDTA	Ethylenediaminetetraacetic acid
gDNA	Genomic deoxyribonucleic acid
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine

His	Histidine
K_i	Inhibition constant
K_m	Michaelis-Menten constant
LB	Luria-Bertani
Lys	Lysine
MES	2-Morpholinoethanesulfonic acid monohydrate
MOPS	3-(N-Morpholino)propanesulfonic acid
MPD	2-Methyl-1,3-propanediol
N-terminus	Amino-terminus
NPGB	4-nitrophenyl 4-guanidinobenzoate hydrochloride
OD	Optical density
PAI	Plasminogen activator inhibitor
PAS	Periodic acid-Schiff
PCR	Polymerase chain reaction
PEG	Poly(ethylene) glycol
Phe	Phenylalanine
pI	Isoelectric point
pNA	4-nitranilide
rpm	Rotation per minute
Ser	Serine
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
SOC	Super optimal broth
TAE	Tris, acetic acid, EDTA
TAFI	Thrombin activatable fibrinolysis inhibitor
TAFIa	Activated thrombin activatable fibrinolysis inhibitor
TE	Tris-HCl, EDTA
TEMED	N,N,N',N'-Tetramethylethylenediamine
TF	Tissue factor
TFB I	Standard transformation buffer I
TFB II	Standard transformation buffer II

TFPI	Tissue factor pathway inhibitor
tPA	Tissue-type plasminogen activator
Tris	Tris(hydroxymethyl)aminomethane
Val	Valine
V_{max}	Maximum reaction rate
YPD	Yeast peptone dextrose

1. Introduction

1.1. Haemostasis

Haemostasis is a group of well-regulated processes that are responsible for keeping the blood fluid and for avoiding blood loss after a vessel injury by the formation of a fast and local haemostatic plug. Normal haemostasis depends on the closely integrated and coordinated action of vascular wall, platelets and coagulation factors (Cotran *et al.*, 2000).

After a vessel injury, the smooth muscle cells of the vessel wall contract in order to minimize blood loss. The platelets adhere to the collagen of the subendothelial matrix that is exposed by the vessel injury. Platelets become activated suffering morphological changes and releasing the content of large vesicles named granules. These granules contain adhesion proteins (fibrinogen and thrombospondin), serotonin (that promotes retention of procoagulant proteins on the platelet surface), ADP and other platelet-activating factors. Besides collagen, the cell membrane tissue factor (TF) is also exposed, and both initiate the coagulation cascade (Cotran *et al.*, 2000).

1.2. Coagulation cascade

The coagulation cascade (Figure 1) has been traditionally presented as two relatively independent pathways: the intrinsic and the extrinsic pathways that converge into a common pathway (Ajjan and Grant, 2006; Cotran *et al.*, 2000; Furie and Furie, 1988).

The intrinsic pathway is initiated by the activation of factor (f) XI, which in turn activates fIX to fIXa. FIXa in complex with its cofactor fVIIIa activates fX, initiating the common pathway (Cotran *et al.*, 2000). The physiological initiator of the intrinsic pathway is not clear (Thiagarajan and Narayanan, 2001). The release of collagen from the subendothelial layer, after vessel injury, has been suggested as the initiator (Cotran *et al.*, 2000). *In vitro*, this process can be induced by

exposing vessels to negatively charged non-physiological surfaces such as glass and other silicates (Jesty, 2001).

The extrinsic pathway is initiated by TF. TF activates fVII, forming the TF/fVIIa complex that induces the conversion of limited amounts of fIX to fIXa and catalyses the conversion of fX to fXa. The activation of fX marks the beginning of the common pathway. fXa forms a complex with fVa, phospholipids and calcium named prothrombinase complex, which activates thrombin by prothrombin cleavage (Mann *et al.*, 2003). Thrombin is involved in a large number of reactions within and outside the coagulation cascade. The major role of thrombin in blood coagulation is the conversion of fibrinogen into fibrin. It also contributes for fibrin polymerization and stabilization by activating fXIII. Thrombin further amplifies thrombin generation by proteolytic activation of fV, fVIII and fXI. Besides its action in the coagulation cascade, thrombin is also involved in inflammation and tissue repair (Thiagarajan and Narayanan, 2001). The fibrin network, obtained at the end of the coagulation cascade, together with red and white blood cells and platelets, forms the blood clot that avoids blood loss through the site of injury (Cotran *et al.*, 2000).

Nowadays, this rigid division is less often used to represent the coagulation mechanism (Jesty, 2001; Mann *et al.*, 2003). The contribution of the intrinsic pathway to haemostasis seems to be small. People who lack fXII are clinically normal, even to the extent of being able to undergo major surgery. In contrast, a deficiency in TF has never been observed in humans, and in studies with transgenic mice, more than 80% of TF knock out animals die from haemorrhage (Jesty, 2001). Thus, the extrinsic pathway seems to have an important role in coagulation cascade initiation, while the intrinsic pathway contributes to the generation of additional thrombin molecules by a positive feedback mechanism. These two pathways are interdependent, in contrast to what their rigid division suggests. (Thiagarajan and Narayanan, 2001).

1.2.1. Clotting regulation

The coagulation cascade is tightly controlled by negative feedback mechanisms as well as by circulating and locally produced inhibitors (Figure 1). In this way, clot formation is restricted to the site of injury and the coagulation cascade is turned off once the blood clot is formed (Berkner, 2001).

When thrombin molecules diffuse away from the site of injury they are caught by thrombomodulin, present on the surface of uninjured endothelial cells. Thrombomodulin and thrombin form a complex that activates protein C, the major natural anticoagulant. Activated protein C binds to its cofactor protein S, produced by endothelial cells, and cleaves fVIIIa and fVa, shutting down the coagulation cascade. (Cotran *et al.*, 2000) This negative feedback mechanism ensures a pulse of thrombin formation over the period during which cofactors are in their active state (Jesty, 2001).

There are also circulating inhibitors that shut off the activated clotting enzymes. Once the inhibitor-protease complex is formed it is rapidly removed from circulation and degraded (Berkner, 2001). Tissue factor pathway inhibitor (TFPI) blocks TF-fVII initiation of coagulation in a two-step mechanism: first, TFPI combines with fXa and then this complex inhibits TF-fVIIa complex. The need of fXa for inhibition of TF-fVIIa complex by TFPI allows the activation of significant amounts of fXa before inactivation of the TF-fVIIa complex (Jesty, 2001). Another circulating inhibitor is antithrombin III, a member of the serpin family of serine endopeptidase inhibitors (Rawlings *et al.*, 2010). This inhibitor is present in plasma at levels largely exceeding those of its targets or their precursors (Jesty, 2001). Physiologically, antithrombin III inactivates thrombin and fXa and, to a lesser extent, factors IXa, XIa and XIIa (Bafunno and Margaglione, 2010). The inhibition mechanism used against thrombin is the best documented. Antithrombin III binds to thrombin, is cleaved but then traps thrombin in a complex (Berkner, 2001). The inhibition of clotting enzymes by antithrombin III is greatly increased by heparin, an highly sulfated polysaccharide (Jesty, 2001).

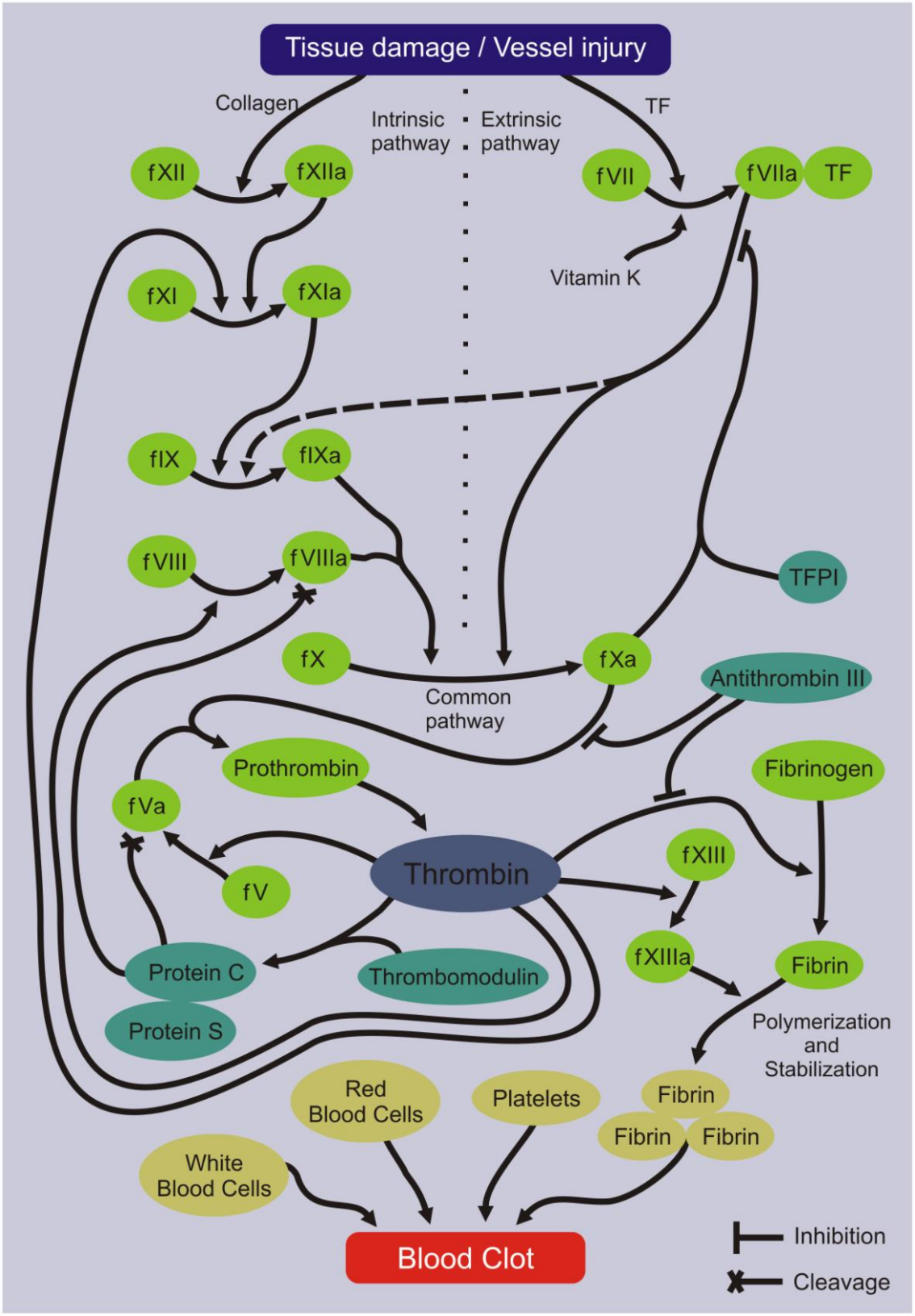


Figure 1 – Schematic representation of the coagulation cascade and its regulation. The coagulation cascade is initiated after a vessel injury when collagen or TF are exposed. The

successive activation of coagulation zymogens leads to the formation of thrombin, a proteinase that has a central role in coagulation and haemostasis. Thrombin cleaves fibrinogen into fibrin that, together with blood cells and platelets, forms the blood clot at the site of the injury. The coagulation cascade is tightly regulated in order to restrict clot formation to the site of injury and to turn off the cascade once the blood clot is formed.

1.2.2. Clot stabilization and lysis

The stability of the fibrin clot is determined by a delicate balance between coagulation and fibrinolysis. The fibrinolytic system is initiated after the formation of fibrin when both plasminogen and tissue-type plasminogen activator (tPA) bind to the fibrin surface to generate plasmin by proteolytic cleavage of plasminogen. The interactions between the lysine-binding sites of plasminogen and tPA and the C-terminal lysine residues of partially degraded fibrin result in the formation of a ternary complex and increase the catalytic efficiency of plasmin formation. Plasmin stimulates its own formation by generating new C-terminal lysine residues after fibrin cleavage (Figure 2). This positive feedback loop is stopped by plasminogen activator inhibitor (PAI), α_2-antiplasmin and thrombin activatable fibrinolysis inhibitor (TAFI), thus protecting the fibrin clot against lysis (Bouma and Mosnier, 2003; Cotran *et al.*, 2000).

PAI inhibits tPA interfering with plasminogen activation while α_2-antiplasmin binds to plasmin forming an irreversible complex that is unable to lyse fibrin. Activated TAFI (TAFIa) inhibits fibrinolysis by removing C-terminal lysine and arginine residues from partially degraded fibrin, thereby lowering its binding capacity for plasminogen and inhibiting tissue plasminogen activator-mediated fibrin degradation. TAFI is activated by thrombin, but for this activation to occur higher concentrations of thrombin are required when compared to the amount necessary for fibrin formation. This fact combined with the short half-life of TAFI (about 10 minutes at 37ºC) suggest that regulation of its functional activity occurs most likely at the level of activation (Bouma and Mosnier, 2003).

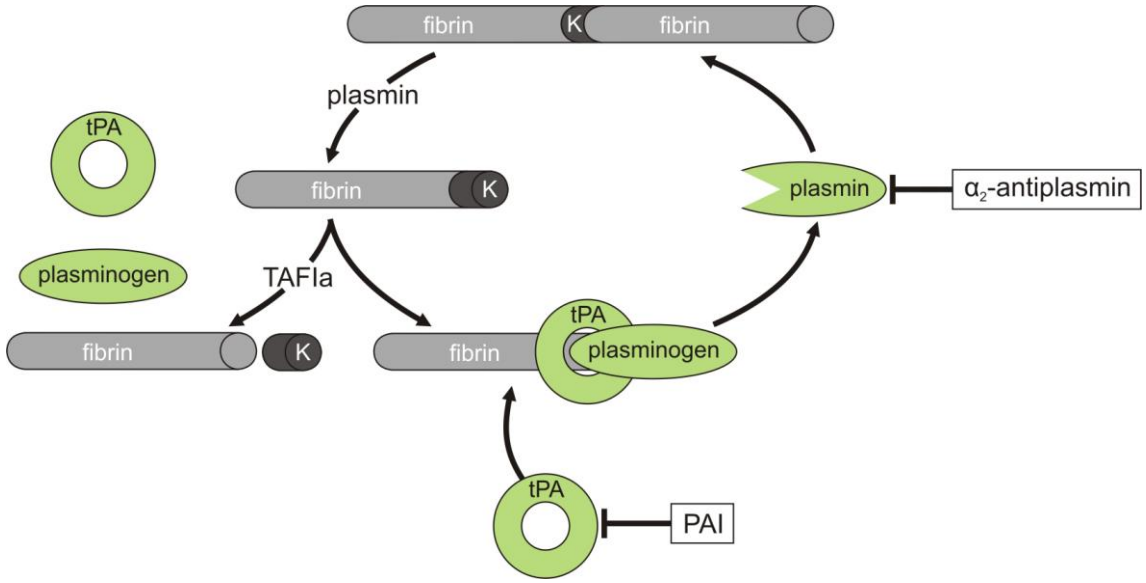


Figure 2 – Clot stabilization and lysis. Fibrinolysis is initiated when plasminogen is cleaved to plasmin by tPA. Plasmin cleaves fibrin into soluble fibrin degradation products that contribute to the formation of new molecules of plasmin. This positive feedback loop is halted by PAI, α_2-antiplasmin and TAFI (adapted from (Bouma and Mosnier, 2003)).

1.3. Thrombosis

A deregulation in the processes involved in the maintenance of normal haemostasis can produce two pathologic situations: bleeding, when the deregulation leads to an inability to respond to a vessel injury, and thrombosis, when the haemostatic plug growth is not limited (Damjanov, 2000). There is also a pathology that combines bleeding and thrombosis named disseminated intravascular coagulation (DIC) (Cotran *et al.*, 2000), in which the formation of disseminated small blood clots leads to a large consumption of platelets and coagulation proteins. With no machinery able to respond in case of injury, abnormal bleeding occurs. From this pathologies, thrombosis (in particular thromboembolism) is by far the major cause of morbility and mortality in western societies, largely due to the adopted life style and diet (Ajjan and Grant, 2006; Gross and Weitz, 2008).

1.3.1. Pathogenesis

There are three predisposing conditions for the formation of pathological thrombi: endothelial cell injury, hemodynamic changes and hypercoagulability of the blood. These conditions, also known as the Virchow's triad (Figure 3) can act independently to form a thrombus or can affect each other and have a cumulative effect (Cotran *et al.*, 2000; Damjanov, 2000).

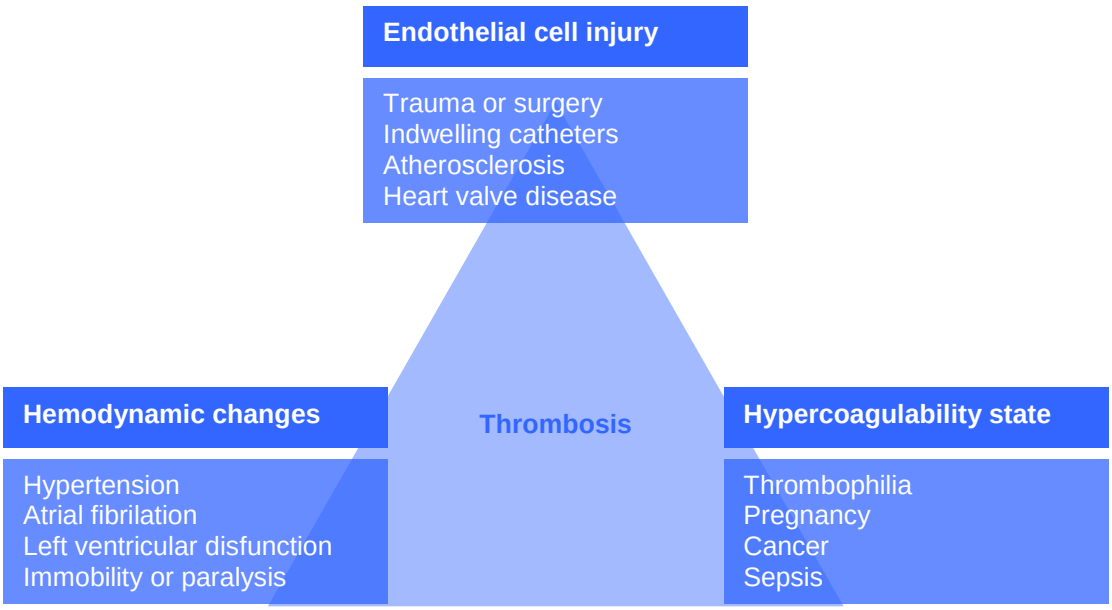


Figure 3 – Representation of the Virchow's triad.

Endothelial injury leads to thrombus formation mainly in heart and arterial vessels. Sites for thrombus formation can include heart chambers (e.g. after an endocardic injury), ulcerated atherosclerotic plaques, inflammation sites or traumatic injury sites. Endothelial injury can also be produced by hemodynamic stress, radiation, bacterial toxins and smoking.

Hemodynamic factors that promote coagulation are of two kinds: those that lead to turbulence and those that slow down the blood flow. The normal blood flow

is laminar meaning that the blood cells flow in the center of vascular lumen separated from the endothelium by acellular plasma that flows slower. Turbulence disturbs this normal flow and allows platelets to be exposed to the vessel wall and to discharge their granules as a result of mechanical stimulation. The turbulence also promotes damage of endothelial cells. On the other hand, slow blood flow promotes sedimentation of the cells and is also less efficient in washing away small thrombi than normal and more vigorous blood flow.

The factors that lead to hypercoagulability of blood can be divided into primary and secondary factors. The primary factors arise from genetic modifications of coagulation cascade factors or of regulators that lead to abnormal or absence of function. The secondary factors are acquired and the most significant are cancer, tissue injury, DIC, chronic cardiac failure, pregnancy, oral contraceptives and smoking.

1.3.2. Pathology

In most situations the formation of small thrombi is not correlated with a pathologic situation because they are lysed. Large thrombi can be incorporated into the vessel wall or recanalized, and thus blood can keep flowing. However, if a thrombus cannot be firmly attached to the vessel wall, recanalized or dissolved two situations can occur: the thrombus can block the vessel in the same place where it was formed or it breaks off from the anchoring surface giving rise to an embolus. Thromboembolus is an embolus of thrombotic origin that flows with blood and blocks small vessels (Cotran *et al.*, 2000; Damjanov, 2000).

The occlusion of vessels leads to an interruption of the blood supply to an organ thus resulting in loss of function. This total occlusion can be observed in the histological section (Figure 4 A), where the whole diameter of a vessel is occupied by a thrombus. The presence of a vessel segment with no content confirms the absence of blood flow (Figure 4 A, thin arrow).

A prolonged insufficiency in blood supply can cause ischemic necrosis, named infarction. Every organ can be affected, resulting in different symptoms and conditions. The most severe thrombosis are those affecting the brain (Figure 4, B), heart or lungs, leading to irreversible damage and/or death (Cotran *et al.*, 2000; Damjanov, 2000).

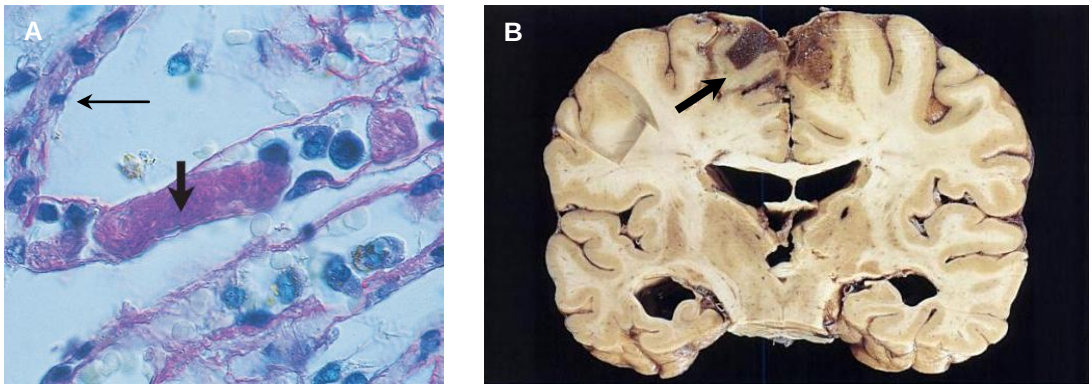


Figure 4 – Thrombosis. A) Microthrombosis in a lung vessel. Histological section PAS stained amplified 150x of pulmonary parenchyma with a capillary containing a thrombus (large arrow). The thrombus is blocking the capillary avoiding the blood flow. Notice the appearance of capillary after the thrombus (thin arrow) (Riede and Werner, 2004) B) Saggital and cortical vein thrombosis. Coronal section of the brain showing a fatal infraction and liquefaction (arrow) of the brain substance drained by the cortical veins (Cooke and Stewart, 2004).

1.4. Searching for therapeutic anticoagulants

Synthetic anticoagulants have been used to prevent and treat thrombosis in the venous and arterial circulation, including heart; however, they have well documented limitations. The increase of bleeding is the most common effect (Eikelboom and Hirsh, 2007) and some anticoagulants do not have a specific reversing agent, such as fondaparinux (Osinbowale *et al.*, 2010). Interaction with other drugs and food were described for warfarin (Becattini *et al.*, 2010). Some

anticoagulants, such as idraparinux and heparin, exert their inhibitory effect on the clotting enzymes indirectly via a potentiation of the inhibitory function of antithrombin, and thus, their inhibitory activity depends on the blood levels of this protein (Alban, 2008; Markwardt, 2003).

The need to overcome all these limitations drives a continual and intense effort to develop new, efficient and safe anticoagulants.

1.4.1. The best targets

Thrombin plays a central role in the process of haemostasis and thrombosis and for this reason it has become the primary target for the development of anticoagulant agents (Markwardt, 2003). However, thrombin has a great number of important activities within and outside the haemostatic system. For this reason, the inhibition of thrombin could affect not only the procoagulant properties of this protease but also interfere with other thrombin activities. As a consequence, the efficacy and the safety of thrombin inhibitors is questionable (Ansell, 2007).

Factor X is a good alternative to thrombin as a therapeutic target. Besides its position at the convergent point of the intrinsic and extrinsic pathways, in contrast to thrombin fXa has limited functions other than being the principal mediator of thrombin generation by prothrombinase activation. Furthermore, as the levels of serine proteinases are amplified at each step of the coagulation cascade, anticoagulants with target coagulation factors higher up than thrombin in the cascade, such fXa, are probably more effective than those directly targeting thrombin (Ansell, 2007).

Thrombin and fXa have evolved from a common ancestral protease related to chymotrypsin. However, in contrast to chymotrypsin that cleaves almost any protein, thrombin and fXa as every other clotting enzyme, cleave only a few specific proteins. The ability of each proteinase to recognize its cofactor and to cleave specific proteins is dependent upon its structure (Berkner, 2001). For this reason, understanding the structure of thrombin and fXa, as well as the

interactions they establish with other proteins is of high practical importance for the design of specific therapeutic anticoagulants. In fact, the initial thrombin publications have produced enormous efforts in the pharmaceutical industry worldwide, aimed at rationally designing and elaborating specific direct thrombin inhibitors, that could help to prevent venous thrombotic events (Bode, 2005).

1.4.1.1. Thrombin

Thrombin (EC 3.4.21.5) is a trypsin-like member of the chymotrypsin family of serine proteases (Rawlings *et al.*, 2010). The human α-thrombin molecule (Stubbs and Bode, 1993) is composed of a light chain (36 residues) and of a heavy chain (259 residues), covalently linked by a disulfide bridge (Cys1-Cys122). The catalytic triad of thrombin, composed of Ser195, His57 and Asp102, occupies the bottom of a rather narrow cleft sandwiched between two large loops “inserted” at positions 60 and 149. In spite of a neutral isoelectric point, the electrostatic surface of thrombin reveals two highly positive areas sandwiching the negative active-site between them. These two positive surface regions are named anion binding exosites I and II, and play major functional roles by regulating thrombin interactions with substrates, cofactors and natural inhibitors (described below). Anion exosite I and II are also known as fibrinogen-binding site and heparin-binding site, respectively, showing their role in the specific interaction with fibrinogen and heparin (Figure 5, A).

The complex structure of thrombin reflects the multifunctional character of this proteinase: thrombin is a procoagulant and anticoagulant and it also plays a role in cellular proliferation and inflammation (Thiagarajan and Narayanan, 2001).

1.4.1.2. Factor Xa

Factor Xa (EC 3.4.21.6) is also a member of the trypsin-like serine proteinase family. As thrombin, fXa is composed of a light chain (16.2 kDa) and a heavy chain (42 kDa) linked together by a disulfide bond (Krupiczojc *et al.*, 2008). The active site, by contrast, is more solvent-exposed, since the 60-loop and 149-loop are 8 and 6-residues shorter respectively, when compared to thrombin. Factor Xa also possesses two charged surface regions that are topologically equivalent to the thrombin exosites (Figure 5, B). Similar to thrombin, the exosite II of fXa is positively charged and is involved in heparin binding. On the other hand, the exosite I of fXa is negatively charged and plays a role in the activation of fX by fIXa (Chen *et al.*, 2004; Manithody *et al.*, 2002).

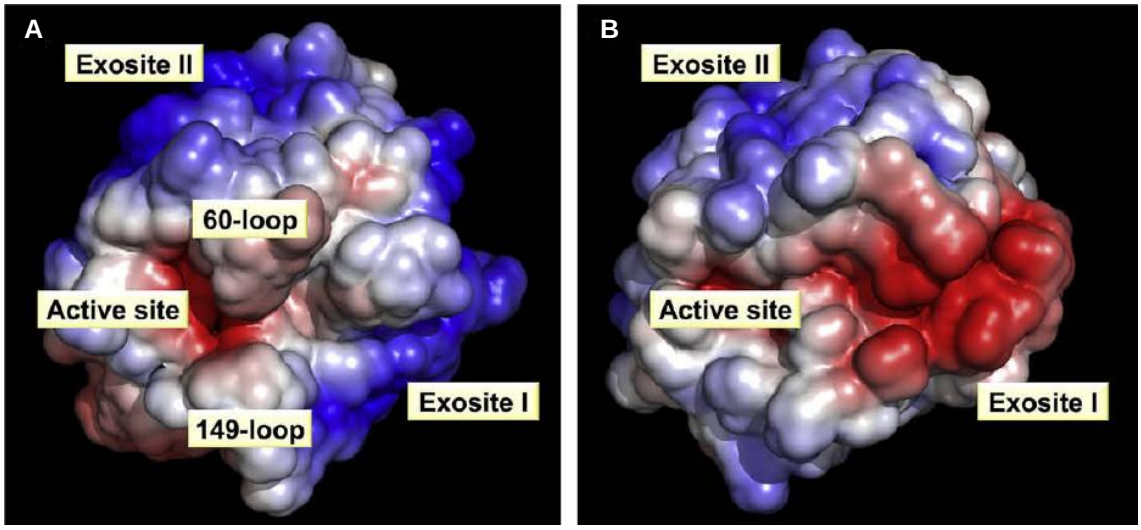


Figure 5 – Electrostatic surface potentials of human thrombin (A) and human fXa (B). The blue and red areas represent the positively and negatively charged surface, respectively. Notice the specific charge distribution of each proteinase and the accessibility of the active-sites (Corral-Rodríguez *et al.*, 2009).

1.5. Coagulation inhibitors from haematophagous animals

During evolution, while vertebrate animals developed a complex haemostatic system to prevent blood loss, blood-feeding animals evolved anticoagulant mechanisms to counteract the haemostatic response of their hosts in order to facilitate blood drawing, storage and digestion. These anticoagulant mechanisms interfere with fibrinolysis, platelet aggregation and blood coagulation (Urata *et al.*, 2003). The blood coagulation inhibitors identified in blood-feeding animals seem to be directed mainly to thrombin and its activator fXa. Unlike physiological inhibitors of blood coagulation proteinases, which are mainly of the serpin and Kunitz families, large molecular diversity is observed in the coagulation inhibitors from haematophagous animals (Koh and Kini, 2009). Although a number of natural inhibitors from haematophagous animals have been described, only a few have been fully characterised, and even fewer have been studied from a structural view-point.

Because of its special role in blood coagulation, thrombin has become the most attractive target for potential therapies against thrombosis. Consequently, special attention was dedicated to the study of haematophagous-derived thrombin inhibitors. To date, only six thrombin inhibitors were structurally characterised revealing six different inhibition strategies in a wide variety of structural scaffolds (Fuentes-Prior *et al.*, 1997; Grütter *et al.*, 1990; Macedo-Ribeiro *et al.*, 2008; Richardson *et al.*, 2000; Rydel *et al.*, 1990; Van de Locht *et al.*, 1995; Van de Locht *et al.*, 1996).

1.5.1. Thrombin inhibitors

The thrombin inhibitor hirudin, isolated from the leech *Hirudo medicinalis*, is the best studied natural thrombin inhibitor (Grütter *et al.*, 1990; Markwardt, 1970; Rydel *et al.*, 1990; Stone and Hofsteenge, 1986). The N-terminal domain of hirudin folds into a globular unit stabilised by three disulfide bridges and the C-terminal

domain shows an extended conformation. During thrombin inhibition, the three N-terminal residues bind to an hydrophobic pocket on the active site of thrombin in a non-canonical manner, while the C-terminal region, rich in acidic residues, interacts with exosite I (Figure 6, A). Hirudin, the most potent thrombin inhibitor known so far (K_i = 22x10^{-15} M), has been used as a model for the development of anticoagulant inhibitors. Nowadays, structural variants of hirudin are produced as recombinant proteins (leupirudin and desirudin) for therapeutic use. However, they have limitations: bleeding, lack of a specific antidote and risk of immunoresponse (Greinacher and Warkentin, 2008).

Another member of the hirudin family is haemadin (Rawlings *et al.*, 2010). Although hirudin and haemadin share low sequence similarity, they display a common three-dimensional structure: a globular N-terminal stabilised by three disulfide bridges with an extended acidic tail. The inhibition of thrombin by haemadin, with an inhibition constant of 2.4x10^{-13} M, involves the binding of the three N-terminal residues of haemadin to the active site of thrombin, similar to hirudin, and the interaction of the C-terminal tail with the exosite II instead of the exosite I (Figure 6, B) (Richardson *et al.*, 2000).

The only thrombin Kazal-type inhibitor structurally characterized is rhodniin, isolated from *Rhodnius prolixus*. Rhodniin is composed by two Kazal-type domains and exhibits an inhibition constant of about 2x10^{-13} M (Friedrich *et al.*, 1993). The first domain canonically binds to the thrombin active site and the second interacts with the exosite I (Figure 6, C) (Van de Locht *et al.*, 1995).

Triabin, isolated from *Triatoma pallidipennis* (Noeske-Jungblut *et al.*, 1995) , shows no similarity with any other inhibitor known, giving name to the triabin family (Rawlings *et al.*, 2010). Its structure shows an eight-stranded β-sheet, which is smoothly rolled and annealed at both edges forming a slightly flattened and conical β-barrel with the overall appearance of a calyx (Fuentes-Prior *et al.*, 1997). The interaction of triabin with the exosite I of thrombin effectively blocks thrombin's cleavage of physiological substrates (Figure 6, D). The inhibition constant found for triabin was 3x10^{-12} M (Noeske-Jungblut *et al.*, 1995).

Ornithodorin and boophilin are both from the BPTI-Kunitz family. Ornithodorin (Van de Locht *et al.*, 1996) was isolated from *Ornithodorus moubata*

and is a potent thrombin inhibitor (K_i= 1x10^{-12} M). It is composed of two distorted Kunitz modules: the highly acidic C-terminal domain binds to the exosite I of thrombin and the first N-terminal residues interact with the active site (Figure 6, E). Boophilin (Macedo-Ribeiro *et al.*, 2008) was isolated from *Boophilus microplus*, recently reclassified as *Rhipicephalus microplus* (Barker and Murrell, 2002). This inhibitor is composed of two canonical Kunitz-type domains. During thrombin inhibition, the N-terminal residues of boophilin bind across thrombin's active site cleft in a parallel, hirudin-like manner while the C-terminal domain interacts with thrombin's exosite I (Figure 6, F). Boophilin exhibits an inhibition constant of 1.8x10^{-9} M against thrombin.

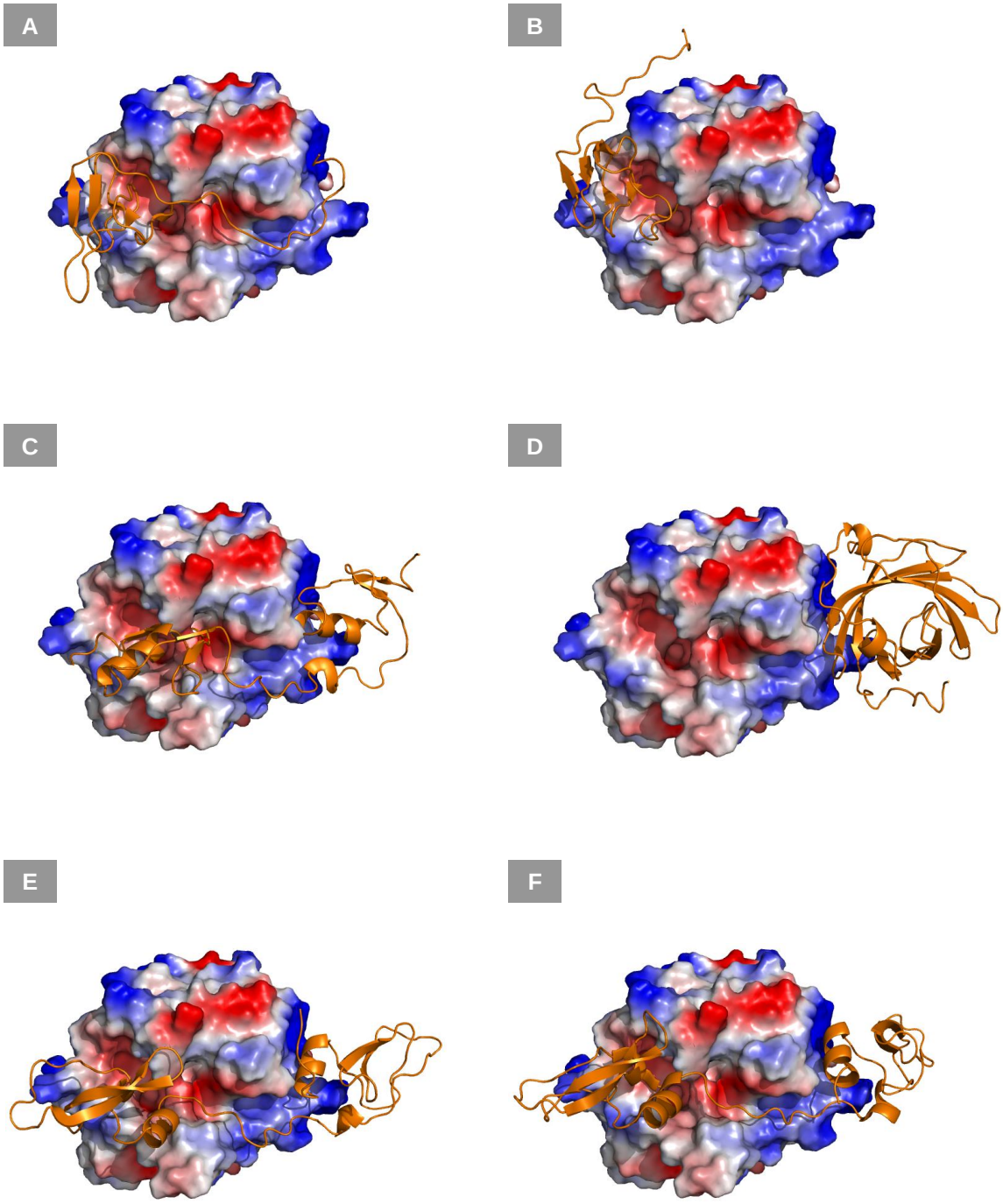


Figure 6 – Interactions between thrombin and six of its inhibitors. Thrombin is represented by its electrostatic surface potential while inhibitors are shown as orange ribbons. A) Hirudin-thrombin complex (PDB: 1HRT); B) Haemadin-thrombin complex (PDB: 1E0F); C) Rhodniin-thrombin complex (PDB: 1TBQ); D) Triabin-thrombin complex (PDB: 1AVG); E) Ornithodorin-thrombin complex (PDB: 1TOC); F) Boophilin-thrombin complex (PDB: 2ODY).

1.5.1.1. Boophilin

There are two described variants of boophilin: isoform H2 and isoform G2. The mRNA encoding for the mature protein of each isoform has eight nucleotide differences, four of which result in amino acids exchanges. However, these substitutions are conservative: in the N-terminal domain, the aspartic acid present at position 66 of isoform G2 is replaced by a glutamic acid in isoform H2; in the linker region Ser78 and Gly82 of isoform G2 are substituted in isoform H2 by an asparagine and a serine, respectively; and in the C-terminal domain, Arg110 of isoform G2 is replaced by a glutamine in isoform H2 (Figure 7) (Macedo-Ribeiro *et al.*, 2008). Both isoforms have a molecular weight of about 13.9 kDa and an isoelectric point of 4.3.

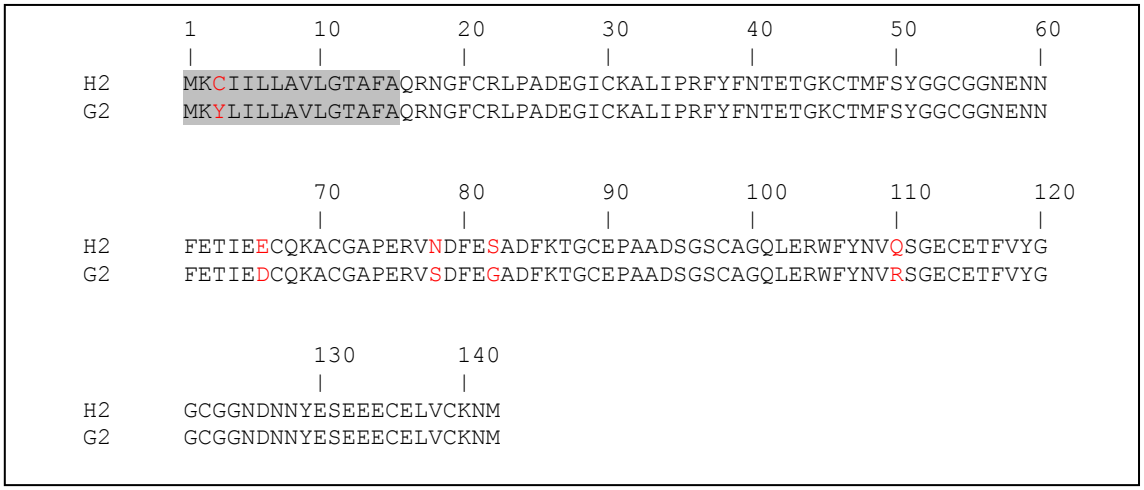


Figure 7 – Amino acid alignment between the two isoforms of boophilin. The signal peptide is shadowed in grey. The amino acid differences between both isoforms are highlighted in red.

Boophilin inhibits thrombin, the central proteinase of the coagulation cascade (Macedo-Ribeiro *et al.*, 2008). From a therapeutic point of view, boophilin could be used not only as an anticoagulant but also as a model for the design of

synthetic anticoagulant drugs. On the other hand, thrombin inhibitors are critical for haematophagous animals during feeding and digestion. For this reason it has been proposed that inhibitors targeting thrombin, such as boophilin, could be used as targets for the control of pathogen-transmitting parasites (Wikel and Bergman, 1997).

1.5.1.1.1. Inhibitor of clot formation

In vitro, boophilin is able to inhibit several serine proteases, including thrombin, trypsin, plasmin and urokinase-type plasminogen activator. Despite the ability to inhibit at least these proteases, boophilin is specially active against thrombin and trypsin. The activity of boophilin against these two proteases is significantly reduced in absence of the N-terminal domain (Macedo-Ribeiro *et al.*, 2008).

The crystallographic structure of the thrombin-boophilin complex shows that this inhibitor is composed by two canonical BPTI-Kunitz domains connected by a 10-residue linker. In spite of the presence of a reactive site loop in the N-terminal domain, boophilin inhibits thrombin according to a non-canonical mechanism, in which the N-terminal Arg17 blocks the active site of thrombin while the negatively charged surface of the C-terminal domain establishes electrostatic and hydrophobic interactions with the exosite I of thrombin (Figure 8) (Macedo-Ribeiro *et al.*, 2008).

The most interesting feature about boophilin is that thrombin-bound boophilin retains the capability to interact with a trypsin-like serine proteinase. The presence of an alanine at the P1 position of the C-terminal domain of boophilin strongly disfavours it canonical binding to a trypsin-like protease and the ability of the isolated N-terminal domain to inhibit trypsin suggests that boophilin inhibits a subset of trypsin-like serine proteinases according to a canonical mechanism with the Lys31 occupying the S1 pocket of the bound enzyme. In fact, the structure of the thrombin-boophilin complex shows that Lys31 remains completely accessible

for interaction with an additional proteinase (Figure 8). (Macedo-Ribeiro *et al.*, 2008).

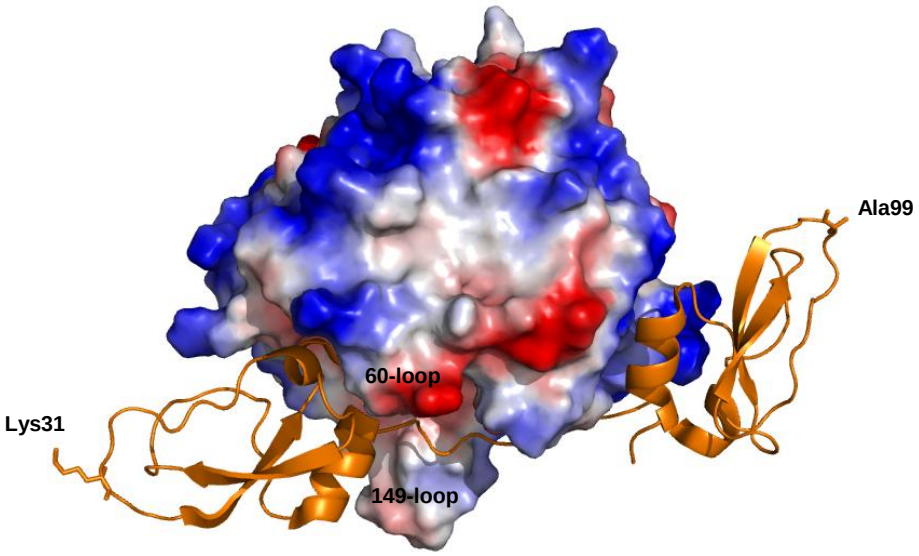


Figure 8 – Boophilin in complex with thrombin (PDB: 2ODY). Thrombin is represented by its electrostatic surface potential while boophilin is shown as an orange ribbon. Boophilin is composed of two canonical BPTI-Kunitz domains. The presence of a lysine at the P1 position of the N-terminal domain of boophilin favours it canonical binding to a trypsin-like proteinase.

In vivo, it is possible that besides inhibiting thrombin, boophilin targets another trypsin-like serine proteinase of the coagulation cascade. Meizothrombin is an intermediate during the prothrombin activation by the prothrombinase complex (fXa-fVa). If boophilin is also able to inhibit meizothrombin, that already has the active site and exosite I formed, fXa seems to be the most attractive candidate because of its physical proximity. This would result not only in thrombin inhibition, but also in the impairment of thrombin generation, making boophilin a good model for the design of new therapeutic anticoagulants. However, *in vitro* fXa is only marginally inhibited by free boophilin. It is possible that the formation of the ternary complex depends on the proper orientation of both membrane-bound serine proteinases, and probably requires additional boophilin-fVa interactions (Macedo-Ribeiro *et al.*, 2008). This strategy is already used by some inhibitors,

such as TFPI. The TFPI inhibits fVIIa using its first Kunitz domain, but only after its second Kunitz domain bound fXa (Broze, 1995).

1.5.1.1.2. Target for *Rhipicephalus microplus* control

Boophilin is produced by the cattle tick *Rhipicephalus microplus*, the major ectoparasite affecting livestock in America, Asia, Africa and Oceania. Besides being the vector of *Babesia spp* and *Anaplasma spp*, this blood-sucking parasite causes significant economic impact due to both blood and milk losses and leather damage. The enormous economic losses caused by this parasite could be minimized by the control of tick populations down to acceptable levels. Usually, tick control is performed with acaricides, however, the improper use of these substances has increased the incidence of acaricide-resistant ticks, besides increasing the occurrence of environmental and food contamination. Thus, the development of new control strategies is necessary (Parizi *et al.*, 2009).

It has been proposed that preventing blood feeding of pathogen-transmitting parasites by interfering with specific inhibitors of haemostasis could provide novel approaches for improving human and animal health (Wikel and Bergman, 1997). Boophilin is the best studied and the most potent coagulation inhibitor from *Rhipicephalus microplus* known (Ciprandi *et al.*, 2006; Horn *et al.*, 2000; Macedo-Ribeiro *et al.*, 2008; Ricci *et al.*, 2007) and, for this reason, boophilin seems to be the best target for the development of new anti-*Rhipicephalus microplus* vaccines.

1.6. Objectives

Boophilin is a serine endopeptidase inhibitor able to inhibit *in vitro* thrombin and trypsin at the same time. This ability is unique between the thrombin inhibitors described to date and suggests that boophilin uses two different inhibition mechanisms. Up to now, only the inhibition mechanism used against thrombin was described. *In vivo*, the second inhibition mechanism is probably used against a (non-thrombin) trypsin-like serine proteinase from the coagulation cascade, such as fXa. This ability would result not only in thrombin inhibition but also in the impairment of thrombin generation, making boophilin a good model for the design of new anticoagulant drugs. This thesis proposes to clarify this second inhibition mechanism by studying the trypsin-boophilin complex.

The inhibition of trypsin by the isolated C-terminal domain (second domain) of boophilin is only observed when it is in large excess over the protease. Moreover, the presence of a lysine in the reactive-loop of the N-terminal domain (first domain) of boophilin strongly suggests that trypsin inhibition is achieved by this domain (Macedo-Ribeiro *et al.*, 2008). We propose to clarify the contribution of each isolated domain by comparing the inhibition constants between full-length boophilin and its isolated first domain. However, to understand the inhibition mechanism used by boophilin against trypsin, additional studies are necessary.

The interaction between proteins has been studied by X-ray crystallography, which allows the visualization of the complex at atomic or near atomic level. As the name suggests, for the application of X-ray diffraction, the protein must be first crystallized. The crystallization has become the primary obstacle in X-ray crystallography since protein crystallization is highly affected by their dynamic structure (McPherson, 2004). Assuming that the second domain does not participate in the inhibition of trypsin, the complex composed by full-length boophilin and trypsin would be very flexible and hard to crystallize. If boophilin inhibits trypsin only by its first domain the complex composed by the isolated first domain and trypsin would be a good model to study the inhibition mechanism and would also probably crystallize. However, it is first necessary to confirm that the

isolated first domain has a conformation similar to the first domain of the full-length boophilin. We propose to crystallize and determine the structure of the isolated first domain of boophilin by X-ray crystallography.

If the crystallographic structure of the isolated first domain shows a conformation similar to the first domain of full-length boophilin, we propose to crystallize the complex composed by the isolated first domain and trypsin and to determine it crystallographic structure.

2. Materials and Methods

2.1. Materials

Bovine pancreatic trypsin, *Glycine max* trypsin inhibitor, *Glycine max* trypsin-chymotrypsin inhibitor, BSA, NPGB, primers (3'AOX1 and 5'AOX1), phenol-chloroform-isomyl alcohol, cyanoborohydride solution, 2-mercaptoethanol, benzamidine, TEMED, bromophenol blue, glycine, calcium chloride, BIS-TRIS, potassium phosphate monobasic, potassium phosphate dibasic, potassium acetate, manganese (II) chloride, sodium acetate, sodium cacodylate, sodium oxamate, Triton X-100, ammonium sulfate, MOPS, methanol, tryptone and the glycerol used for crystallography were acquired from Sigma-Aldrich Química S.A. (Sintra, Portugal). PEG 1000, PEG 3350, PEG 8000, MPD, sodium carbonate, myo-inositol, biotin, rubidium chloride and potassium hydroxide were obtained from Fluka, as part of Sigma-Aldrich Química S.A. (Sintra, Portugal). Magnesium sulfate and sodium citrate were acquired from Riedel-de-Haën, distributed by Sigma-Aldrich Química S.A. (Sintra, Portugal). Sodium chloride, yeast extract, Tris, sodium formate, lithium sulfate monohydrate, sodium azide, glacial acetic acid, hydrochloric acid, dextrose, EDTA, potassium chloride, magnesium chloride, MES, sodium phosphate dibasic, sodium phosphate monobasic and sodium hydroxide were purchase from Merck, distributed by VWR (Carnaxide, Portugal). Ethanol was purchased from Aga (Porto, Portugal). Imidazole was acquired from USB Corporation, distributed by Isogen Life Science (Barcelona, Spain). Agar was obtained from Liofilchem and ethidium bromide solution was acquired from AppliChem, both distributed by Frilabo (Maia, Portugal). The glycerol used in culture media was obtained from Panreac (Amadora, Portugal). Bacto™ peptone was purchased from BD Biosciences, distributed by ENZIfarma S.A. (Oeiras, Portugal). Sodium dodecylsulfate (SDS), acrylamide solution, APS and Poly-Prep® Column were obtained from Bio-rad (Amadora, Portugal). PageBlue™ stain, *Pfu* polymerase, *Pfu* polymerase buffer, dATP, dTTP, dCTP, dGTP and 6x Orange Loading Dye were obtained from Fermentas, distributed by Bioportugal (Porto, Portugal). NZYMiniprep Kit was obtained from NZYTech (Lisboa, Portugal). QIAquick® PCR Purification Kit and QIAquick® Gel Extraction Kit were acquired from QIAGEN, distributed by Izasa Portugal (Carnaxide, Portugal). CNBr-activated

Sepharose 4B, Mono Q and Superdex 75 columns and bovine thrombin were obtained from GE Healthcare (Carnaxide, Portugal). AminoLink Plus Coupling Resin was purchased from Thermo Scientific Pierce, distributed by Dagma Lda (Carcavelos, Portugal). Agarose and Zeocin™ were purchased from Invitrogen Life Technologies (Barcelona, Spain). The chromogenic substrate Bz-Phe-Val-Arg-pNA was purchased from Bachem (Weil am Rhein, Germany). Chromozym® TH (Tos-Gly-Pro-Arg-pNA) was purchased from Roche (Amadora, Portugal). Crystal Screen, Crystal Screen 2, PEG-Ion and Index were obtained from Hampton Research (Aliso Viejo, California). Proplex, Morpheus and JCSG-plus were acquired from Molecular Dimensions (Newmarket, United Kingdom).

2.2. Preparation of *Escherichia coli* DH5α competent cells

Escherichia coli DH5α cells were grown in LB agar plates (Appendix I, a, i) at 37 ºC overnight. A single colony was used to inoculate 10 ml of LB medium (Appendix I, a, i). The culture was grown at 37 ºC with shaking (180 rpm) to an optical density (OD) of 0.3 at 550 nm. The biomass was then amplified by adding 4 ml of this culture to 100 ml of LB medium and incubating in the above conditions until an OD of 0.3 at 550 nm was reached. The culture was then incubated on ice for 5 minutes followed by centrifugation at 1000 *g* for 5 minutes at 4 ºC. The supernatant was discarded and the pellet was gently resuspended in 20 ml of ice-cold TFB I (Appendix I, b, i). After centrifugation at 1000 *g* and 4 ºC for 5 minutes, the new supernatant was rejected and the pellet suspended in 2.5 ml of ice-cold TFB II (Appendix I, b, ii). The cell suspension was incubated on ice for 5 minutes, aliquoted into ice-cold microcentrifuge tubes, frozen in liquid nitrogen and stored at -80 ºC.

2.3. Bacterial transformation

Fifty microlitres of *E. coli* DH5α competent cells prepared as described in section 2.2 were thawed on ice and 200 ng of pPICZαB (Appendix III) containing the cDNA for full-length boophilin (D1D2) or for its first domain (D1) was added. The mixture was incubated on ice for 30 minutes and heat-shocked at 42 ºC for 30 seconds. The mixture was incubated on ice for 2 minutes before adding 600 µl of SOC medium (Appendix I, a, ii). Cells were recovered at 37 ºC for 1 hour with shaking (180 rpm). For transformant selection two hundred microlitres of the culture were spread on Low Salt LB agar plates (Appendix I, a, iii) containing 25 µg/ml of Zeocin™ (Invitrogen). A low salt concentration is required for the transformant selection with Zeocin™, since a salt concentration above 90 mM inhibits the drug. The plates were incubated at 37 ºC overnight. A negative control was performed in parallel, without plasmid DNA addition.

2.4. Isolation and purification of plasmid DNA

A single colony of *E. coli* DH5α transformed with pPICZαB-D1D2 or pPICZαB-D1 was grown at 37 ºC overnight with shaking (180 rpm) in 10 ml of Low Salt LB containing 25 µg/ml of Zeocin™. Each vector was purified using the NZYMiniprep Kit (NZYTech). The overnight culture was centrifuged at 4000 *g* for 10 minutes at room temperature and the supernatant was discarded. The cell pellet was resuspended in Buffer A1 and lysed with 250 µl of lysis buffer (Buffer A2). After mixing by gently inverting the tubes 8 times, the mixture was incubated at room temperature for 4 minutes. Three hundred microlitres of neutralization buffer (Buffer A3) were added and mixed by gently inverting the tubes 8 times. The solution was then centrifuged for 10 minutes at 16000 *g* and the supernatant obtained was loaded onto the NZYTech spin column already placed in a 2 ml collecting tube. After centrifugation at 16000 *g* for 1 minute the flow-through was discarded. The column membrane was washed in two steps by centrifugation for 1

minute at 16000 *g* with 500 µl of Buffer AW pre-warmed at 50 ºC followed by 600 µl of Buffer A4. An additional centrifugation at 16000 *g* for 2 minutes was performed in order to remove all ethanol from the column membrane. The NZYTech spin columns were placed into clean 1.5 ml microcentrifuge tubes. Thirty microlitres of sterile water were loaded onto the column and incubated for 1 minute at room temperature. DNA was collected by centrifugation at 16000 *g* for 1 minute. The eluted DNA was quantified by measuring the absorbance at 260 nm and the quality was assessed by the ratio between the absorbance at 260 nm and 280 nm and at 260 nm and 230 nm. This analysis was performed on a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific).

2.5. Genomic DNA extraction

Ten microlitres of each glycerol stock containing *Pichia pastoris* cells transformed with pPICZαB-D1 or pPICZαB-D1D2 were grown in 10 ml of YPD medium (Appendix II, a, i) containing 100 µg/ml of Zeocin^{TM} at 28 ºC and 220 rpm to an OD of 1.5 at 600 nm. After reaching the desired OD, the culture was centrifuged at 1200 *g* for 10 minutes at room temperature. The supernatant obtained was rejected and the pellet was resuspended in 500 µl of sterile water. After transferring to a microcentrifuge tube, the suspension was centrifuged at 16000 *g* for 5 minutes at room temperature. The supernatant obtained was rejected and the cell pellet was resuspended in 200 µl of lysis buffer (Appendix II, b, i). Three hundred milligrams of glass spheres (Ø 0.5 mm) and 200 µl of phenol-chloroform-isoamyl alcohol mixture were added and then all the components were mixed by vortexing at high speed for 3 minutes at room temperature. Two hundred microlitres of TE buffer (10 mM pH 8.0, 1 mM EDTA) were added and the mixture was mixed by vortexing. After centrifugation at 16000 *g* for 5 minutes at room temperature, the aqueous phase (upper phase) was transferred to a clean microcentrifuge tube. Nucleic acids were precipitated by adding 1 ml of 100% ethanol and mixing by inverting the tube. After centrifugation at 16000 *g* for 3

minutes at room temperature, the supernatant was rejected and the pellet resuspended in 400 µl of TE buffer. To remove RNA, 30 µg of RNAse A were added and incubated for 5 minutes at 37 ºC followed by precipitation of genomic DNA with 10 µl of 3 M sodium acetate and 1 ml of 100% ethanol. After centrifugation at 16000 *g* for 3 minutes at room temperature the supernatant was rejected and the pellet was air-dried at room temperature. The genomic DNA was resuspended in 50-100 µl of sterile water until complete dissolution. The DNA was quantified by measuring the absorbance at 260 nm and the quality was analysed by the ratio between the absorbance at 260 nm and 280 nm and at 260 nm and 230 nm. The integrity of genomic DNA was confirmed by DNA electrophoresis as described in section 2.9.

2.6. Polymerase chain reaction

The sequences coding for the D1D2 or the D1 integrated in the genomic DNA of *Pichia pastoris* cells were amplified by PCR. The reaction mixture containing *Pfu* buffer (20 mM Tris-HCl pH 8.8 at 25 ºC, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 0.1 mg/ml BSA, 0.1% (v/v) Triton X-100, 2 mM $MgSO_4$) (Fermentas), 0.2 mM of each dNTP (dATP, dCTP, dGTP, dTTP), 0.5 µM of 3'AOX1 primer (5´-GCAAATGGCATTCTGACATCC-3´), 0.5 µM 5'AOX1 primer (5´-GACTGGTTCCAATTGACAAGC-3´) and 2 U *Pfu* DNA polymerase (Fermentas) were added to about 1 µg of genomic DNA to a final volume of 50 µl. The cDNA was amplified by 30 cycles of denaturation at 95 ºC for 30 seconds, annealing at 57 ºC for 30 seconds and extension at 72 ºC for 2 minutes preceded by a initial denaturation step at 95 ºC for 3 minutes and followed by a final extension step at 72 ºC for 15 minutes. The PCR product obtained after amplification of D1D2 cDNA was re-amplified by a new PCR. The PCR product was first purified as described in section 2.7 and then 36 ng of purified PCR product were added to the reaction mixture to a final volume of 50 µl. The amplification reaction was performed in the same conditions as described previously. The PCR product was analysed by DNA

gel electrophoresis (section 2.9), purified (section 2.7 and 2.8) and sequenced (section 2.10).

2.7. Purification of the PCR product

The PCR product was purified using the QIAquick® PCR Purification Kit (QIAGEN). Two hundred microlitres of binding buffer (Buffer PB) were added to 40 µl of PCR product and mixed by gently pipetting up and down. Once the colour of the mixture had turned violet, 10 µl of 1 M sodium acetate pH 4.1 were added and the sample was applied to a QIAquick spin column already placed into a 2 ml collection tube. After centrifugation at 16000 *g* for 60 seconds at room temperature, the flow-through was discarded and the column was washed with 750 µl of Buffer PE by repeating the previous centrifugation. The flow-through was discarded and the column centrifuged for an additional 1 minute to remove all traces of ethanol. The column was placed in a clean 1.5 ml microcentrifuge tube and 30 µl of sterile water were added to the center of the QIAquick membrane. After incubation for 1 minute at room temperature, the DNA was eluted by centrifugation at 16000 *g* for 1 minute at room temperature. The eluted DNA was quantified by measuring the absorbance at 260 nm and the quality was assessed by the ratio between the absorbance at 260 nm and 280 nm and at 260 nm and 230 nm. This analysis was performed on a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific).

2.8. Purification of DNA from agarose gel

The PCR product obtained after the second amplification of D1D2 cDNA was separated from contaminants by agarose gel electrophoresis. The band containing the D1D2 cDNA was excised from the gel with a sterile scalpel blade under UV light and the cDNA was purified using the QIAquick® Gel Extraction Kit

(QIAGEN). The gel slice was weighed and the agarose dissolved in 3 gel volumes (considering 1 µl = 1 µg) of solubilization buffer (Buffer QC) at 50 ºC for 10 minutes. During this incubation, the sample was mixed by vortexing every 3 minutes. After dissolution, 1 gel volume of isopropanol was added to the sample and mixed by inverting the tube. The sample was then applied to a QIAquick spin column already placed in a 2 ml collection tube. The column was centrifuged at 16000 *g* for 1 minute and the flow-through was discarded. In order to remove all traces of agarose, the column was washed with 0.5 ml of Buffer QG by centrifugation at 16000 *g* for 1 minute. The column was then incubated with 0.75 ml of Buffer PE for 2 minutes at room temperature. After centrifugation at 16000 *g* for 1 minute the flow-through was discarded and the column centrifuged for an additional 1 minute to remove all ethanol from the column membrane. The column was placed into a sterile 1.5 ml microcentrifuge tube and incubated with 30 µl of sterile water for 1 minute at room temperature, and then centrifuged at 16000 *g* for 1 minute at room temperature to collect the purified DNA. The eluted DNA was quantified by measuring its absorbance at 260 nm and its quality was assessed by the ratio between the absorbance at 260 nm and 280 nm and at 260 nm and 230 nm. This analysis was performed on a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific).

2.9. DNA electrophoresis

Genomic DNA and PCR products were analysed on a 2% agarose gel containing 0.2 µg/ml of ethidium bromide. Samples were prepared by adding 1/6 sample volume of 6x Orange DNA Loading Dye (10 mM Tris-HCl pH 7.6, 0.15% orange G, 0.03% xylene cyanol FF, 60% glycerol and 60 mM EDTA) (Fermentas). The agarose gels (50 ml) were run at 100 V in TAE buffer (0.2 M Tris, 0.571% glacial acetic acid, 5 mM EDTA).

2.10. DNA sequence analysis

The cDNA for D1D2 and D1 cloned into pPICZαB or amplified from the genomic DNA of *P. pastoris* were analysed by DNA sequencing using 5'AOX1 (5'-GACTGGTTCCAATTGACAAGC-3') as primer (Eurofins MWG, Germany). The nucleotide sequences obtained for the plasmids were translated and aligned with both boophilin isoforms (GenBank numbers CAC82583.1 and CAC82582.1). The nucleotide sequences obtained for the genomic DNA were translated and aligned with the amino acid sequences obtained for the plasmids. All alignments were performed using the ClustalW tool (Thompson *et al.*, 1994) (http://www.ebi.ac.uk/Tools/clustalw2/index.html).

.

2.11. Expression of D1 or D1D2

Pichia pastoris KM71H cells transformed with pPICZαB coding for D1D2 or D1 fused to a N-terminal signal peptide (α-factor) were grown in YPD agar plates (Appendix II, a, i) containing 100 µg/ml of Zeocin^{TM} for 2 days. Ten millilitres of BMGY (Appendix II, a, ii) were used to grow one colony at 29 ºC and 220 rpm to an OD of 2-6. The cellular mass was amplified by adding the previous culture to 1 l of BMGY and growing in the same conditions to an OD of 2-6. To induce expression, the pellet obtained after centrifugation of the culture was resuspended in 200 ml BMMY (Appendix II, a, iii). The cells were incubated at 29 ºC and 220 rpm for 5 days. Every 24 hours 1 ml of 100% methanol was added to the culture. Since both recombinant proteins were secreted, cells were harvested by centrifugation at 10000 *g* for 5 minutes at room temperature and the supernatant was saved. The expression was confirmed by SDS-PAGE analysis (sections 2.15) and *in vitro* activity assays (2.16).

2.12. Purification of D1D2

D1D2 was first purified by affinity chromatography in an immobilized-trypsin column. This column was prepared as described below.

2.12.1. Preparation of the affinity column

An immobilized-trypsin affinity column was prepared by coupling bovine pancreatic trypsin (Sigma) to CNBr-activated Sepharose 4B (GE Healthcare). This medium was supplied lyophilized in the presence of additives. In order to remove the additives, 5 g of lyophilized powder were resuspended in 100 ml of 1 mM HCl and washed for 20 minutes in 900 ml of the same solution on a sintered glass filter (porosity G3). The swollen medium was washed with 200 ml of 0.1 M $NaHCO_3$ pH 8.3, 0.5 M NaCl (coupling solution). Seven hundred and fifty milligrams of trypsin were dissolved in 12 ml of coupling buffer (at this stage 10 µl of the mixture were saved for protein quantification by measuring the absorbance at 280 nm). Two millilitres of 200 mM benzamidine solution were prepared in coupling buffer and added to the 12 ml of trypsin solution. The trypsin-benzamidine solution was incubated with the medium at 4 ºC overnight. The suspension was filtered on a sintered glass filter (porosity G3) and the flow-through was collected. The benzamidine present in the flow-through was eliminated by 3 cycles of dialysis each against 100 volumes of coupling buffer at 4 ºC. After dialysis the trypsin present in the flow-through was quantified by measuring the absorbance at 280 nm. The medium was washed with 100 ml of coupling buffer to remove the excess of trypsin. The remaining reactive groups of the medium were blocked by incubating with 50 ml of ethanolamine pH 8.0 for 2 hours at room temperature with end-over-end rotation. The medium was washed with 3 cycles of alternating pH. Each cycle included a wash with 50 ml of 0.1 M sodium acetate pH 4.0 containing 0.5 M NaCl followed by 50 ml of 0.1 M Tris-HCl pH 8.0 containing 0.5 M NaCl.

The coupling efficiency was determined by subtracting the amount of trypsin present in the dialysed flow-through to the total protease in the trypsin solution. The percentage of active trypsin coupled to the medium was also quantified in order to determine the effective binding capacity. One millilitre of medium was loaded into a Poly-Prep® Column (Bio-Rad) and equilibrated with 20 ml of 50 mM Tris-HCl pH 8.0, 150 mM NaCl (binding buffer). Twelve milligrams of *Glycine max* trypsin-chymotrypsin inhibitor (Sigma) were dissolved in binding buffer and added to the column. The column was washed with 30 ml of binding buffer and bound material was eluted with 8 ml of 0.5 M KCl pH 2.0. One millilitre fractions were collected and their pH neutralized by adding 50 µl of 1 M Tris (pH not adjusted). The eluted material was quantified by measuring the absorbance at 280 nm. The amount of trypsin inhibitor present in the eluted material was used to quantify the binding capacity of the column.

2.12.2. Purification of D1D2

The supernatant obtained after D1D2 expression was filtered (0.45 µm pore) and applied to the immobilized-trypsin affinity column prepared as described before, previously equilibrated with at least 20 column volumes of 0.1 M Tris-HCl pH 8.0. The column was washed with at least 20 column volumes of 0.1 M Tris-HCl pH 8.0 and bound proteins were eluted by competition with 4 column volumes of 1 M benzamidine in the same buffer. The benzamidine present in the eluted material was eliminated by 3 cycles of dialysis each against 100 sample volumes of 20 mM Tris-HCl pH 8.0, 150 mM NaCl at 4 ºC. Purified material was analysed by SDS-PAGE and by activity assays (see section 2.14 and 2.15) before the second purification step. The dialysed material was diluted 10 times with 20 mM Tris-HCl pH 8.0 and applied to a MonoQ column (5 x 50 mm) (Amersham) pre-equilibrated with at least 5 column volumes of the same buffer. The column was washed with 10 column volumes of 20 mM Tris-HCl pH 8.0, and bound proteins were eluted with a linear gradient from 0 to 1.0 M NaCl in the same buffer. The

absorbance at 214 and 280 nm was monitored on an ÄKTApurifier^{TM} (GE Healthcare) and 1 ml fractions were collected, quantified (section 2.14), analysed by SDS-PAGE (section 2.15) and tested for thrombin and trypsin inhibition (section 2.16).

2.13. Purification of D1

D1 was first purified by affinity chromatography in an immobilized-trypsin column. This column was prepared as described below.

2.13.1. Preparation of the affinity column

An immobilized-trypsin affinity column was prepared by coupling bovine pancreatic trypsin (Sigma) to AminoLink Plus Coupling Resin (Thermo Scientific Pierce). Twenty millilitres of AminoLink Plus Coupling Resin were placed in a chromatography column (2.5 cm x 10 cm) and allowed to settle at room temperature. The bottom cap was opened and the storage solution was drained. The column was equilibrated with 100 ml of 0.1 M sodium phosphate pH 7.2, 0.15 M NaCl (coupling buffer) and incubated with 300 mg of trypsin dissolved in 20 ml of coupling buffer (trypsin solution). Four hundred microlitres of 5 M cyanoborohydride solution were added and the mixture was incubated at 4 ºC overnight with gentle shaking by end-over-end rocking. After releasing the gas formed during the reaction by opening the top cap in a fume hood, the content was drained and collected. The flow-through obtained was used to determine the coupling efficiency by subtracting the trypsin amount present in the flow-through to the amount present in the initial trypsin solution. The column was washed with 80 ml of coupling buffer followed by 100 ml of 1 M Tris-HCl pH 7.4 (quenching buffer). Two millilitres of quenching buffer and 40 µl of 5 M cyanoborohydride solution were added to the column in order to block the remaining active sites of the resin. The

matrix was gently mixed for 30 minutes by end-over-end rocking at room temperature. The gas formed during the reaction was first released in a fume hood and the content was drained. The column was washed with 150 ml of 1M NaCl followed by 200 ml of 20 mM Tris-HCl pH 8.0, 150 mM NaCl and was stored in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 M benzamidine, 0.05% sodium azide at 4 ºC.

Five hundred microlitres of resin were used to quantify the amount of active trypsin bound to the resin. The resin was added to a Poly-Prep® Column (Bio-rad) and equilibrated to room temperature. The storage solution was drained and the column washed with 20 ml of 50 mM Tris-HCl pH 8.0, 150 mM NaCl (binding buffer). Two millilitres of binding buffer with 8.4 mg of trypsin inhibitor from *Glycine max* (Sigma) were added to the column and the mixture was incubated for 1 hour at room temperature. The column was washed with 3 ml of binding buffer and the bound material was eluted with 2 ml of 0.2 M glycine-HCl pH 3.0. Five hundred microlitres fractions were collected and their pH adjusted to 7 by adding 25 µl of 1 M Tris (pH not adjusted). The concentration of the collected fractions was determined by measuring the absorbance at 280 nm. The amount of trypsin inhibitor present in the collected fractions was used to calculate the binding capacity of the column.

2.13.2. Purification of D1

The supernatant obtained after D1 expression was filtered (0.45 µm pore) and applied to the immobilized-trypsin affinity column prepared as described in section 2.13.1 and previously equilibrated with at least 20 column volumes of 20 mM Tris-HCl pH 8.0, 150 mM NaCl. The column was washed with at least 20 column volumes of 20 mM Tris-HCl pH 8.0, 150 mM NaCl and bound proteins were eluted with 4 column volumes of 0.2 M glycine pH 3.0. The collected fractions were neutralized by adding 50 µl of 1 M Tris (pH not adjusted) per 1 ml of collected eluate. The most concentrated fractions were pooled and the buffer was

exchanged by 3 cycles of dialysis each against at least 100 sample volumes of 20 mM Tris-HCl pH 8.0, 150 mM NaCl. The purified material was quantified and analysed by SDS-PAGE and activity assays as described in sections 2.14, 2.15 and 2.16, respectively.

2.14. Protein quantification

Purified D1D2 and D1 were quantified by measuring the absorbance at 280 nm on a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific). This quantification is based on the Lambert-Beer law, which considers a linear relationship between absorbance and protein concentration. This law can be translated by the equation $A = C\varepsilon\ell$, in which *A* corresponds to the absorbance represented in absorbance units (AU), *C* to the protein concentration in molarity (M) , ***ε*** to the extinction coefficient of the protein in units of $M^{-1}cm^{-1}$ and ℓ to the path length of the sample in centimetres (cm). Before quantification, the molar extinction coefficient (ε) for each protein was estimated from the respective amino acid sequence using the ProtParam tool of the ExPaSy Proteomics Server (http://expasy.org/).

2.15. SDS-PAGE

Samples from expression and purification of D1D2 and D1 were analysed by SDS polyacrylamide gel electrophoresis (SDS-PAGE). The SDS-polyacrylamide gels were composed by two layers. The stacking gel, with approximately 1 cm, was composed by 4% of acrylamide, 0.156 M Tris-HCl pH 6.8, 0.2% SDS, 0.1% APS and 0.1% TEMED. The separating gel was composed by 15 or 17% of acrylamide, 0.5 M Tris-HCl pH 8.8, 0.2% SDS, 0.1% APS and 0.1% TEMED. All samples were prepared by adding 1/5 sample volume of loading buffer (10% w/v SDS, 20% v/v glycerol, 0.2 M Tris-HCl pH 6.8, 0.05% Bromophenol Blue, 1.8% 2-

mercaptoethanol) followed by incubation at 100 ºC for 10 minutes. SDS polyacrylamide gels were run at between 80 and 120 V in running buffer (24.8 mM Tris, 191.8 mM glycine and 0.1% SDS). Afterwards the gels were washed 3 times in deionised water in order to remove SDS. Each wash was performed in a microwave on maximum power for 1 minute. Washed gels were stained with PageBlue™ (Fermentas) on a microwave for up to 30 seconds on maximum power followed by gentle agitation at room temperature for 30 minutes. After staining the gels were washed overnight with deionised water.

2.16. Activity assay

The inhibitory activity of D1D2 and D1 recombinant proteins was tested after the expression and purification steps by measuring the residual activity of trypsin and/or thrombin towards Bz-Phe-Val-Arg-pNA (Bachem) and Chromozym® TH (Tos-Gly-Pro-Arg-pNA) (Roche), respectively. Bovine pancreatic trypsin (5 µl, 160 nM) or bovine thrombin (5 µl, 240 nM) was incubated with 1 to 80 µl of protein samples in assay buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% Triton X-100) for 10 minutes at 37 ºC. After this first incubation, 5 µl of synthetic substrate (4 mM Bz-Phe-Val-Arg-pNA or 1.9 mM Chromozym® TH) was added and the reaction mixture (final volume of 100 µl) was incubated at 37 ºC for 10 minutes. The reaction was stopped by adding 50 µl of 50% (v/v) acetic acid after which the absorbance at 405 nm was measured using a Synergy 2 microplate reader (BioTek).

2.17. Titration of trypsin

A stock solution of 10 mg/ml trypsin was prepared in 1 mM HCl. In order to determine the amount of active trypsin present in the solution a titration was performed. Ten microlitres of trypsin solution (diluted 10 times) were incubated

with 10 µl of 10 mM 4-nitrophenyl 4-guanidinobenzoate hydrochloride (NPGB) in 20 mM Tris-HCl pH 8.0, 150 mM NaCl to a final volume of 1 ml. The mixture was performed in a polystyrene cuvette and the reaction was followed by measuring the absorbance at 410 nm for 20 minutes (measurements every 5 seconds) at 37 ºC on a Shimadzu UV-2401 PC (Shimadzu Corporation). Measurement of a blank reaction containing 10 µl of 10 mM NPGB in 20 mM Tris-HCl pH 8.0, 150 mM NaCl (in the absence of proteolytic enzyme) was performed at the same time.

The reaction of trypsin was calculated by subtracting the absorbance values corresponding to NPGB degradation (blank reaction) from the absorbance values obtained for the mixture of NPGB and trypsin. A linear regression analysis was applied to the values obtained (absorbance and time of reaction) and the absorbance for the time 0 determined. At this point the concentration of substrate cleaved was equal to the concentration of active trypsin. The active trypsin concentration was determined by using the equation $A = C\varepsilon\ell$ (Lambert-Beer law), in which *A* corresponds to the absorbance represented in absorbance units (AU), *C* to the concentration of NPGB cleaved in molarity (M), ε to the extinction coefficient of NPGB (= 16.595 l $mmol^{-1}$ cm^{-1}) and ℓ to the path length of the sample in centimetres (cm).

2.18. Determination of the Michaelis-Menten constant of trypsin against Bz-Phe-Val-Arg-pNA

Fifty microlitres of 6 nM active trypsin (in 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mg/ml BSA) were incubated with 50 µl of 1.56, 3.13, 6.25, 12.5, 25, 50, 100, 200, 400 or 800 µM Bz-Phe-Val-Arg-pNA (in 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mg/ml BSA). The synthetic inhibitor cleavage reaction was followed at 37 ºC for 1 hour using a Synergy 2 microplate reader (BioTek). The absorbance of the reaction mixture at 405 nm was measured every minute. Each condition was assayed in triplicate.

The concentration of reaction product (4-nitraniline) for each reading point was calculated by using the equation $A = C\varepsilon\ell$, in which *A* corresponds to the absorbance represented in absorbance units (AU), *C* to the concentration of reaction product in molarity (M), ***ε*** to the extinction coefficient of the reaction product (= 10.4 l $mmol^{-1}$ cm^{-1}) and ℓ to the path length of the sample in centimetres (cm). A linear regression analysis was performed on the linear region of the hydrolysis curve (defined by the product concentration and time of reaction) and only the curves with R square higher than 0.97 were considered for the determination of the reaction rate. The reaction rate for each substrate concentration was given by the slope of the respective equation $Y = mX + b$ (where *Y* corresponds to the concentration of reaction product, *m* to the slope of the straight line defined by the equation, *X* to the time and *b* to the product concentration at time 0) obtained by the linear regression analysis. These values were analysed using GraphPad Prism (Motulsky, 2008). Non-linear regression analysis using the Michaelis-Menten equation

$$v_0 = \frac{V_{max}[S]}{K_m + [S]}$$

(v_0 corresponds to initial reaction rate, V_{max} to the maximum reaction rate, K_m to the Michaelis-Menten constant and [*S*] to the substrate concentration) was performed on substrate concentrations and reaction rates.

2.19. Determination of the inhibition constants of D1 and D1D2 towards bovine trypsin

Fifty microlitres of 12 nM active trypsin (in 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mg/ml BSA) were incubated with 50 µl of 0.75, 1.5, 3, 6, 12, 24, 48, 96 and 192 nM D1D2 or D1 (in 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mg/ml BSA) at 37 ºC for 10 minutes. After adding 100 µl of 400 µM Bz-Phe-Val-Arg-pNA (in 50 mM

Tris-HCl pH 8.0, 100 mM NaCl, 1 mg/ml BSA), the mixture was incubated at 37 ºC for 1 hour. The absorbance of the mixture at 405 nm was measured every 15 seconds. Each condition was assayed in triplicate.

The concentration of reaction product (4-nitraniline) for each reading point was calculated using the equation $A = C\varepsilon\ell$, in which *A* corresponds to the absorbance represented in absorbance units (AU), *C* to the concentration of reaction product in molarity (M), ***ε*** to the extinction coefficient of the reaction product (= 10.4 l $mmol^{-1}$ cm^{-1}) and ℓ to the path length of the sample in centimetre (cm). A linear regression analysis was performed on the linear region of each hydrolysis curve (defined by the product concentration and time of reaction) and only the curves with R square higher than 0.97 were considered for the determination of the reaction rate. The reaction rate for each inhibitor concentration was given by the slope of the respective equation $Y = mX + b$ (where *Y* corresponds to the product concentration, *m* to the slope of the straight line defined by the equation, *X* to the time and *b* to the product concentration at time 0). These values were analysed using GraphPad Prism (Motulsky, 2008). Non-linear regression analysis using the Morrison equation

$$v = v_0\left(1-\left(\left([E]_0+[I]+K_{I-app}\right)-\frac{\sqrt{\left([E]_0+[I]+K_{I-app}\right)^2-4[E]_0[I]}}{2[E]_0}\right)\right)$$

(*v* is the measured reaction rate, v_0 is the reaction rate in the absence of inhibitor, $[E]_0$ is the total enzyme concentration, [*I*] is the added inhibitor concentration and $K_{I\text{-}app}$ is the apparent equilibrium inhibitor constant) was performed on inhibitor concentrations and reaction rates.

2.20. Crystallization of D1

Initial crystallization conditions were screened at 293 K using the sitting-drop method with commercial crystallization screens: Crystal Screen, Crystal Screen 2 and PEG-Ion from Hampton Research, and Morpheus from Molecular Dimensions. The drops contained equal volumes (2 µl) of protein solution (≈10 mg/ml in 20 mM Tris-HCl pH 8.0, 150 mM NaCl) and precipitant, equilibrated against a 300 µl reservoir. D1 crystals were obtained using 1.0 M lithium sulfate monohydrate and 2% (w/v) PEG 8000 as precipitant. Crystals were sequentially transferred to solutions of precipitant containing 5, 10, 15, 20 and 25% of glycerol as cryoprotectant and then cryo-cooled in liquid nitrogen.

2.21. Data collection and processing

Diffraction data was collected on an Onyx CCD Detector in an Oxford Diffraction Gemini PX Ultra diffractometer using CuKα radiation. A single cryo-cooled crystal of D1 was used and data were measured in 1º oscillation steps with an 80 mm sample-to-detector distance and 6 minutes exposure per frame. Two hundred and ninety one frames were collected.

Diffraction data were integrated with CrysAlis^{Pro} Software (Oxford Diffraction) and scaled using SCALA (Evans, 2006) from the CCP4 program suite (Collaborative Computational Project No. 4, 1994).

2.22. D1 structure solution

The structure of D1 was solved by molecular replacement with PHASER (McCoy *et al*., 2007) from the CCP4 program suite (Collaborative Computational Project No. 4, 1994) using the coordinates for the N-terminal domain of boophilin

(residues Gln16 to Ala58) from PDB entry 2ODY (Macedo-Ribeiro *et al*., 2008) as search model. The model was subjected to alternating cycles of manual building with *Coot* (Emsley *et al.*, 2010) and crystallographic refinement with PHENIX (Adams *et al.*, 2010).

2.23. Trypsin-D1 complex preparation

The macromolecular complex was prepared by mixing bovine pancreatic trypsin (Sigma) with a 10% molar excess of purified D1. The mixture was incubated on ice for 1 hour, and the trypsin-D1 complex was separated from the excess of inhibitor by size exclusion chromatography on a Superdex 75 column (Amersham Pharmacia), using 20 mM Tris-HCl pH 8.0, 150 Mm NaCl as the mobile phase. The absorbance at 280 nm was monitored on an ÄKTApurifier^{TM} (GE Healthcare). Fractions of 1 ml were collected and analyzed by SDS-PAGE (see section 2.15) and those containing the trypsin·D1 complex were pooled and concentrated on a 3 kDa cut-off centrifugal concentrator (Vivaspin 2, GE Healthcare).

2.24. Crystallization of trypsin-D1 complex

Initial crystallization conditions were screened at 293 K using sitting drops with commercial crystallization screens: Crystal Screen, Crystal Screen 2, PEG-Ion and Index from Hampton Research, and Proplex, Morpheus and JCSG-plus from Molecular Dimensions. The drops contained equal volumes (2 µl) of complex solution (≈10 mg/ml in 20 mM Tris-HCl pH 8.0, 150 mM NaCl) and precipitant solution, equilibrated against 300 µl reservoir. Crystals were obtained using 0.10 M carboxylic acids (0.033 M sodium formate, 0.033 M sodium citrate, 0.033 M sodium oxamate), 0.1 M buffer pH 6.5 (0.025 M imidazole, 0.025 M sodium

cacodylate, 0.025 M MES, 0.025 M BIS-TRIS), 12.5% (v/v) MPD, 12.5% (w/v) PEG 1000, 12.5% (w/v) PEG 3350 as precipitant solution. Some crystals were cryo-cooled in liquid nitrogen.

In order to find a condition able to produce better crystals, new solutions were prepared by varying the pH (pH 6.0, 6.5 and 7.0) and/or the percentage of MPD, PEG 1000 and PEG 3350 present in the initial condition (see Table 1 for details). The complex solution was incubated with these new solutions at 293 K in sitting-drop plates.

Table 1 – Scheme of new precipitant solutions composition. Twenty four precipitant solutions (PS) were prepared by varying the buffer pH and the percentage of MPD, PEG 1000 and PEG 3350 used in the initial precipitant solution.

		0.1 M Buffer (pH)			
		6.0	6.5	7.0	
MPD; PEG 1000; PEG 3350 (% of each)	2.5	PS 1	PS 2	PS 3	Carboxylic acids (0.1 M)
	5	PS 4	PS 5	PS 6	
	7.5	PS 7	PS 8	PS 9	
	10	PS 10	PS 11	PS 12	
	12.5	PS 13	PS 14	PS 15	
	15	PS 16	PS 17	PS 18	
	17.5	PS 19	PS 20	PS 21	
	20	PS 22	PS 23	PS 24	

3. Results and discussion

3.1. DNA sequence analysis of pPICZαB-D1 and pPICZαB-D1D2

The expression vectors pPICZαB-D1 and pPICZαB-D1D2 required for the expression of D1 or D1D2, respectively, in yeast were initially sequenced (MWG Eurofins) in order to verify that the nucleotide sequence was correct, using primer 5'AOX1. Figure 9 represents the amino acid sequence of the D1 and D1D2 recombinant proteins encoded by these vectors. Both proteins are fused to an N-terminal secretion signal termed α-factor. This peptide corresponds to the sequence of the *S. cerevisiae* α-factor prepro peptide and allows the secretion of the D1 and D1D2 proteins into the expression medium.

After the translation of the recombinant protein α-factor-D1 or α-factor-D1 D2 in *P. Pastoris* the N-terminal signal peptide is processed. The processing involves the removal of the pre signal (represented in dark grey in Figure 9, A) by a signal peptidase in the endoplasmic reticulum followed by cleavage by Kex2 endopeptidase after Lys-Arg (Kex2 endopeptidase processing site) present at the end of the signal peptide sequence (Cereghino and Cregg, 2000).

There are two described variants of boophilin: isoform H2 and isoform G2 with GenBank access numbers CAC82583.1 and CAC82582.1, respectively. The D1 and D1D2 recombinant proteins encoded by the plasmids pPICZαB-D1 and pPICZαB-D1D2, respectively were aligned with these two isoforms (Figure 9, B). The D1 sequence corresponds to the first BPTI-Kunitz domain of boophilin G2 isoform. In its sequence a mutation was found in residue position 27, where the glutamic acid was substituted by a glutamine (Figure 9, B). The D1D2 sequence encodes for the full-length boophilin G2 isoform. Its sequence had two mutations when compared to the sequence of boophilin G2 isoform. The first mutation was similar to the one found in D1 sequence. The second mutation was localized in residue position 92, where the alanine was substituted by a threonine (Figure 9, B). Both mutations are conservative and consequently are not predicted to severely affect the inhibitor-proteinase interaction. The crystallographic structure of D1D2 in complex with thrombin (PBD: 2ODY) also supports this idea since it shows that these mutations are localized in regions that are not involved in

thrombin-D1D2 interaction (Figure 10). Because boophilin used two different inhibition mechanisms against trypsin and thrombin, the effect of these mutations on trypsin-boophilin interaction should be clarified.

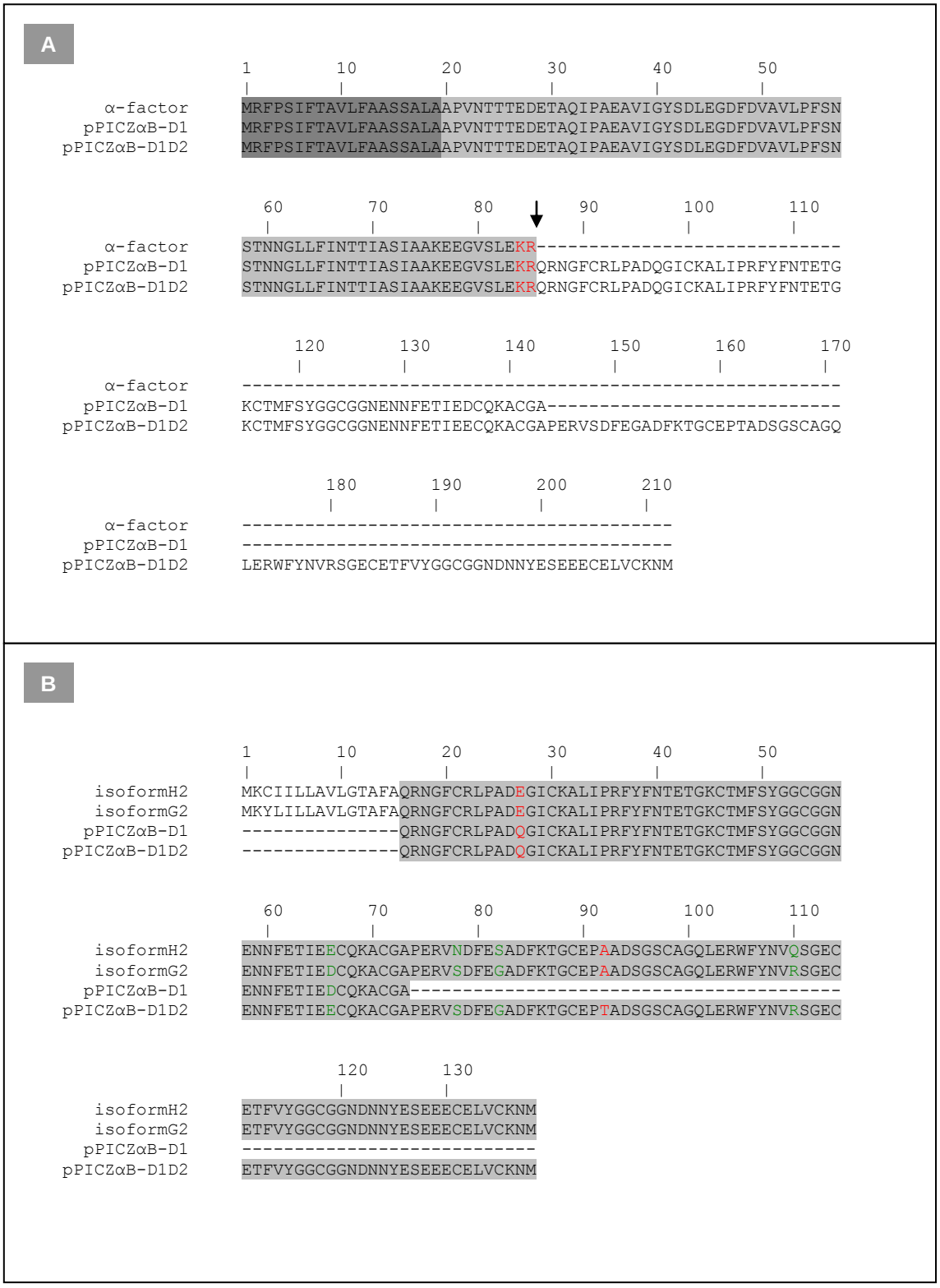


Figure 9 – Amino acid sequence alignment of recombinant D1 and D1D2 proteins encoded by pPICZαB-D1 and pPICZαB-D1D2, respectively, with the α-factor (A) and with boophilin isoforms H1 and G2 (B). A – The sequence represented in grey corresponds to the α-factor signal sequence. The pre signal sequence of the α-factor is shadowed in dark grey. α-factor is separated from D1 or

D1D2 proteins after cleavage by Kex2 endopeptidase. The Kex2 recognition site is represented highlighted in red and the cleavage site is indicated by an arrow. B – The sequences for mature D1 and D1D2 are shown in grey. The amino acid differences between the two isoforms of boophilin are shown in green. Two mutations (red) were found in the D1D2 sequence. One mutation is also present in the D1 sequence.

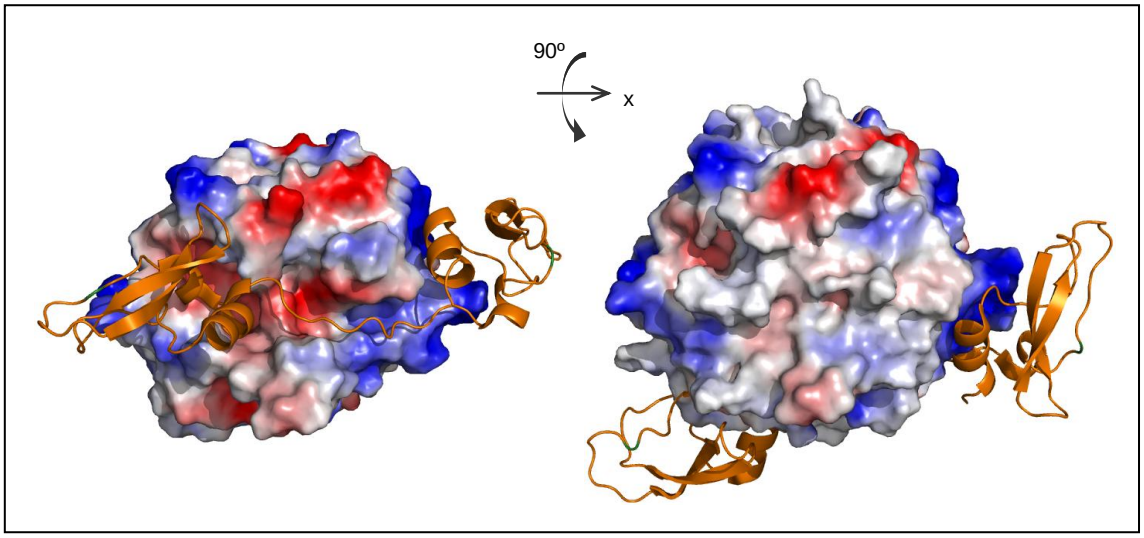


Figure 10 – Crystallographic structure of boophilin in complex with thrombin showing the localization of mutations. The two images show the same complex in different positions. The right image was rotated 90º compared to the left image. Thrombin is represented by its electrostatic surface potential, while boophilin is shown as an orange ribbon. The localization of the mutations is in green.

3.2. Genomic DNA analysis

The expression vectors pPICZαB-D1 and pPICZαB-D1D2 were integrated into the *P. pastoris* genome in order to maximize the stability of transformants. To confirm that no mutations had occurred during the transformation and integration processes, the gDNA of the transformed cells was extracted for subsequent sequence analysis of the integrated sequences.

After gDNA extraction, the quality of DNA obtained was assessed by spectrophotometric analysis. (see Table 2). The ratio between the absorbance at 260 nm and 280 nm obtained for each sample was between 1.80 and 2.00, revealing that the samples were not contaminated with proteins. On the other hand, the ratio between the absorbance at 260 nm and 230 nm was smaller than the common values of 2.0-2.2. This indicated that all samples were contaminated with phenol. The concentration of this contaminant was higher in the DNA purified from cultures of pPICZαB-D1D2 transformants (D1D2 (A) and D1D2 (B)).

The quality of gDNA extracted was also assessed by electrophoretic analysis (Figure 11). This analysis revealed that the concentration of the gDNA extracted form each culture of pPICZαB-D1D2 transformants was significantly higher compared to the gDNA obtained from each culture of pPICZαB-D1 transformants (D1 (A) and D1 (B)). This concentration difference was not in agreement with the concentration values determined by spectrophotometry. Probably due to the high DNA concentration, it was also possible to observe some DNA degradation in D1D2 (A) and D1D2 (B) samples when examined by electrophoresis (Figure 11). The gDNA extracted from each culture of pPICZαB-D1 transformants showed a high degree of integrity (Figure 11), however the amount of DNA applied to the gel was significantly lower compared to the D1D2 (A) and D1D2 (B) samples.

Table 2 – Spectrophotometric analysis of genomic DNA.

Sample	Concentration (ng/µl)	$\frac{OD_{260nm}}{OD_{280nm}}$	$\frac{OD_{260nm}}{OD_{230nm}}$
D1 (A)	567.6	1.91	1.75
D1 (B)	907.0	1.91	1.74
D1D2 (A)	823.0	1.88	1.49
D1D2 (B)	1071.0	1.90	1.47

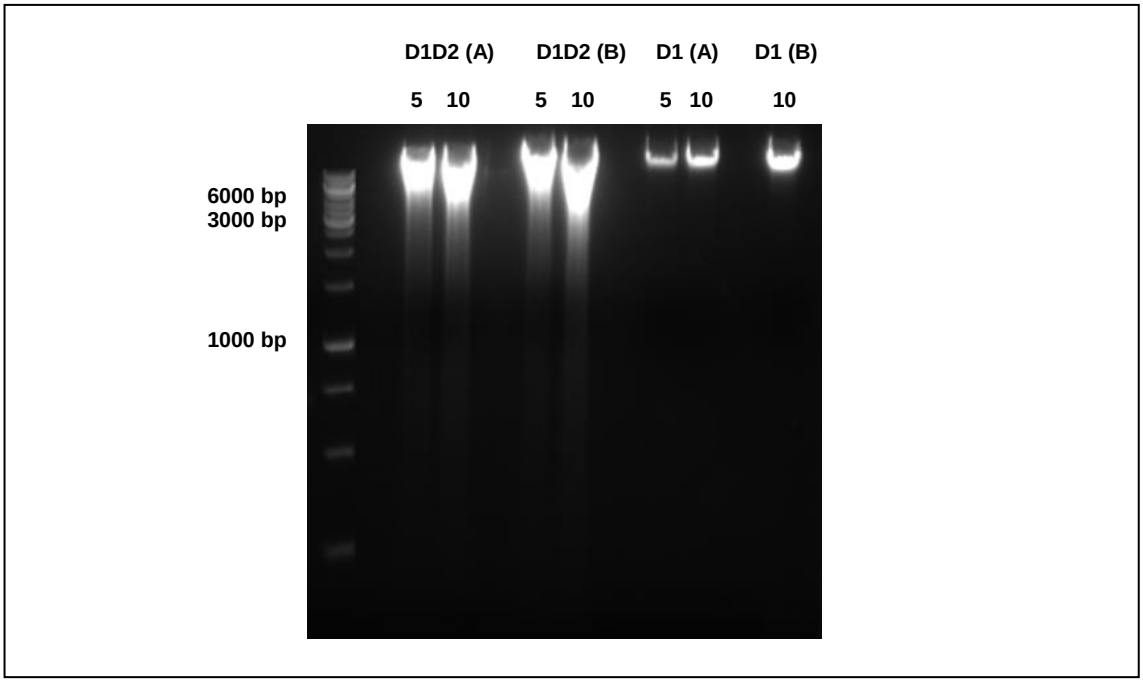


Figure 11 – Electrophoretic analysis of the isolated genomic DNA from pPICZαB-D1D2 (D1D2 A and B) and pPICZαB-D1 (D1 A and B) transformants. Five (5) and ten (10) microlitres of each sample were separated in a 2% agarose gel.

3.3. Amplification of α-factor-D1 and α-factor-D1D2 cDNA

The gDNA extracted from pPICZαB-D1D2 (A and B) and pPICZαB-D1 (A and B) transformants were used as templates for the amplification of α-factor-D1D2 and α-factor-D1 cDNA by PCR using the primers 3'AOX1 (5'-GCAAATGGCATTCTGACATCC-3') and 5'AOX1 (5'-GACTGGTTCCAATTGACAAGC-3'). For the reaction conditions tested, the amount of PCR product obtained when D1D2 DNA was used as a template was lower than when the template was D1 DNA (Figure 12). The low amount of PCR product obtained when D1D2 DNA was used as a template was probably due to Pfu DNA polymerase inhibition. The inhibition of the enzyme was probably caused by the presence of significant amounts of phenol in the D1D2 DNA samples and/or by high DNA concentration in the PCR mixture due to an incorrect quantification.

A small band of contaminant with about 700 bp was observed in both D1D2 (A) and negative control PCR reactions (Figure 12, Lane B and E). This contaminant had a size similar to that of the α-factor-D1 PCR product (approximately 700 bp). As a possible α-factor-D1 cDNA contaminant, it was ignored and the D1 (A) PCR product (700 bp) was purified and analysed by DNA sequencing.

In order to increase the amount of α-factor-D1D2 cDNA a new PCR was performed using the D1D2 (B) PCR product (Figure 12, Lane A) as template. The amount of cDNA obtained was significantly higher; however the same band of contaminant was observed (Figure 13). The cDNA of α-factor-D1D2 was isolated from the gel and the DNA purified and analysed by DNA sequencing.

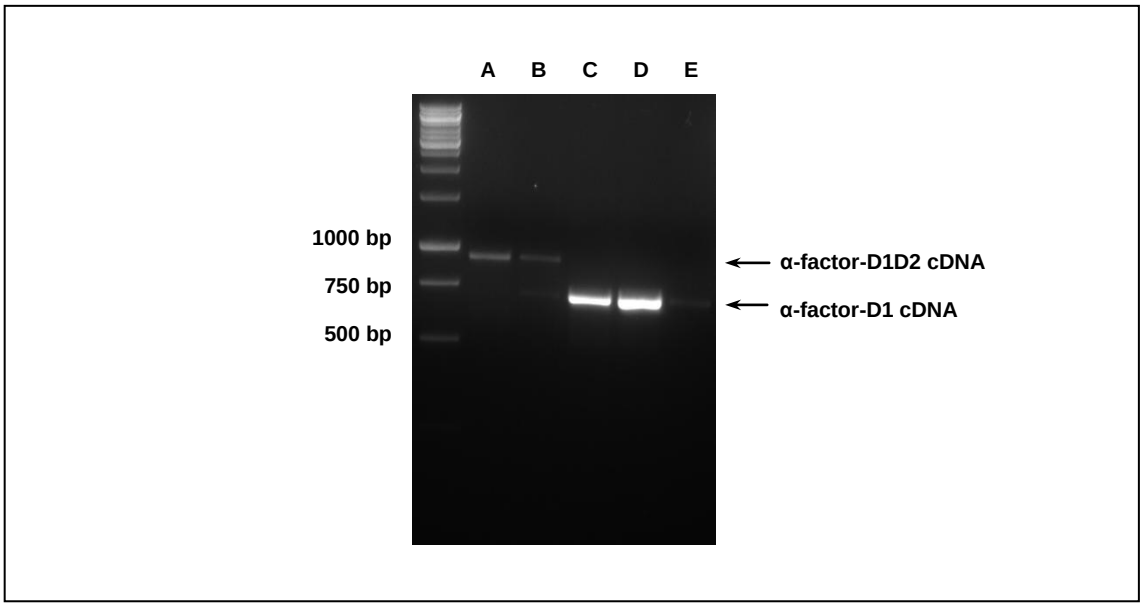


Figure 12 – Electrophoretic analysis of the products obtained by PCR reaction performed with genomic DNAs. Ten microlitres of each sample were resolved in a 2% agarose gel. Templates in each lane: A – D1D2 (B); B – D1D2 (A); C – D1 (B); D – D1 (A); E – Negative control.

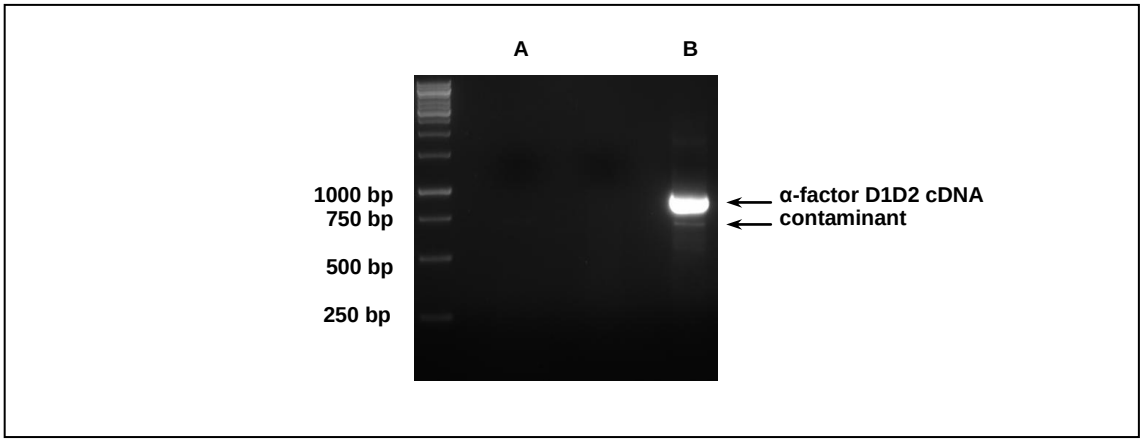


Figure 13 – PCR product obtained after amplification of α-factor-D1D2 (B) cDNA. Ten microlitres of PCR reaction mixture were separated in a 2% agarose gel. A band of contaminant was observed in the PCR reaction mixture. A – Negative control. B – PCR reaction mixture.

3.4. DNA sequence analysis of α-factor-D1 and α-factor-D1D2 cDNA integrated into *Pichia pastoris* genome

The purified α-factor-D1 and α-factor-D1D2 cDNA amplified by PCR were analysed by DNA sequencing. The DNA sequences obtained were translated to amino acid sequence and aligned with the α-factor-D1 or α-factor-D1D2 amino acid sequences encoded by the plasmids. No differences were observed between the sequences encoded by the cDNA of the plasmids or the cDNA integrated into the yeast cells genome (Figure 14). Thus, no mutations that could lead to an amino acid exchange had occurred during transformation and integration of the plasmids into the yeast cells.

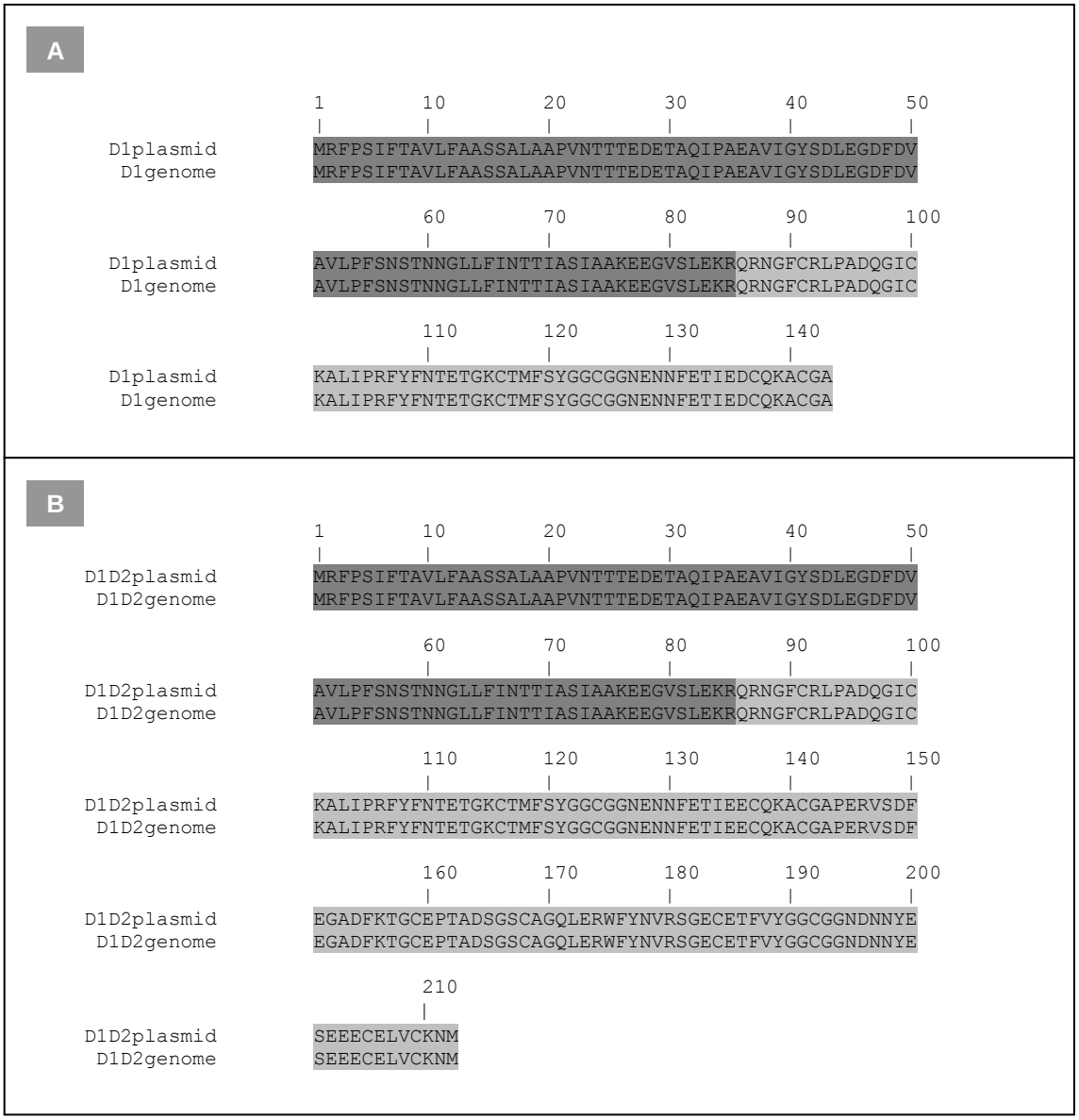


Figure 14 – Amino acid sequence alignment between the sequence encoded by the plasmids pPICZαB-D1 and pPICZαB-D1D2 or by the α-factor-D1 or α-factor-D1D2 cDNAs integrated into the genome. The sequence of α-factor is represented in dark grey and the sequences of D1 and D1D2 are represented in light grey. No differences were observed between the sequences encoded by the plasmids or by the genome.

3.5. Expression of D1D2

After obtaining the desired biomass, the expression of recombinant D1D2 was induced by substituting the growth medium (containing glycerol) for the induction medium (containing methanol). The induction was performed for 5 days, by adding methanol every 24 hours.

The expression and secretion of D1D2 was confirmed by SDS-PAGE and by activity assays at the end of the fifth day of induction. The supernatant obtained after D1D2 expression originated two bands in a SDS polyacrylamide gel (Figure 15, Lane 3). The higher molecular band is the more intense and corresponds to D1D2. This protein revealed an apparent molecular weight of about 18 kDa, which is higher than the value determined by mass spectrometry (13.9 kDa). This anomalous migration in SDS-polyacrylamide gels appear to the related to the acidic character of D1D2 (theorical pI: 4.41), which could interfere with SDS binding. The lower molecular weight band showed an apparent molecular weight of about 14 kDa and it is a putative degradation product of the full-length protein. It was also possible to observe the presence of a smear in the D1D2 supernatant lane (Figure 15, Lane 3), which corresponds to the medium components and to the other proteins secreted by *P. pastoris*.

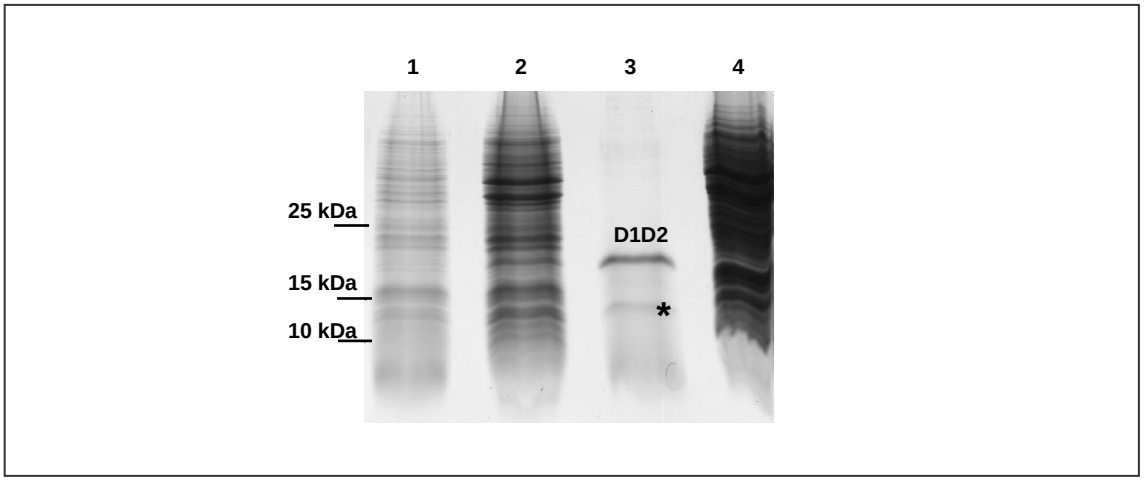


Figure 15 – Expression and secretion of D1D2 protein in *P. pastoris*. Ten microlitres of each sample were separated in a 15% polyacrylamide gel. The gel was stained with PageBlue^{TM}. Lane 1 – Culture before induction. Lane 2 – Culture after induction. Lane 3 – Culture medium collected after the fifth day of induction. The band with higher molecular weight corresponds to D1D2 and the one with lower molecular weight (*) to degradation products. Lane 4 – Whole cell extract.

The presence of D1D2 in the culture medium was also confirmed by testing its ability to inhibit trypsin and thrombin in an *in vitro* activity assay. Both proteinases were inhibited by the supernatant obtained after D1D2 expression (Figure 16). The inhibition was dose-dependent.

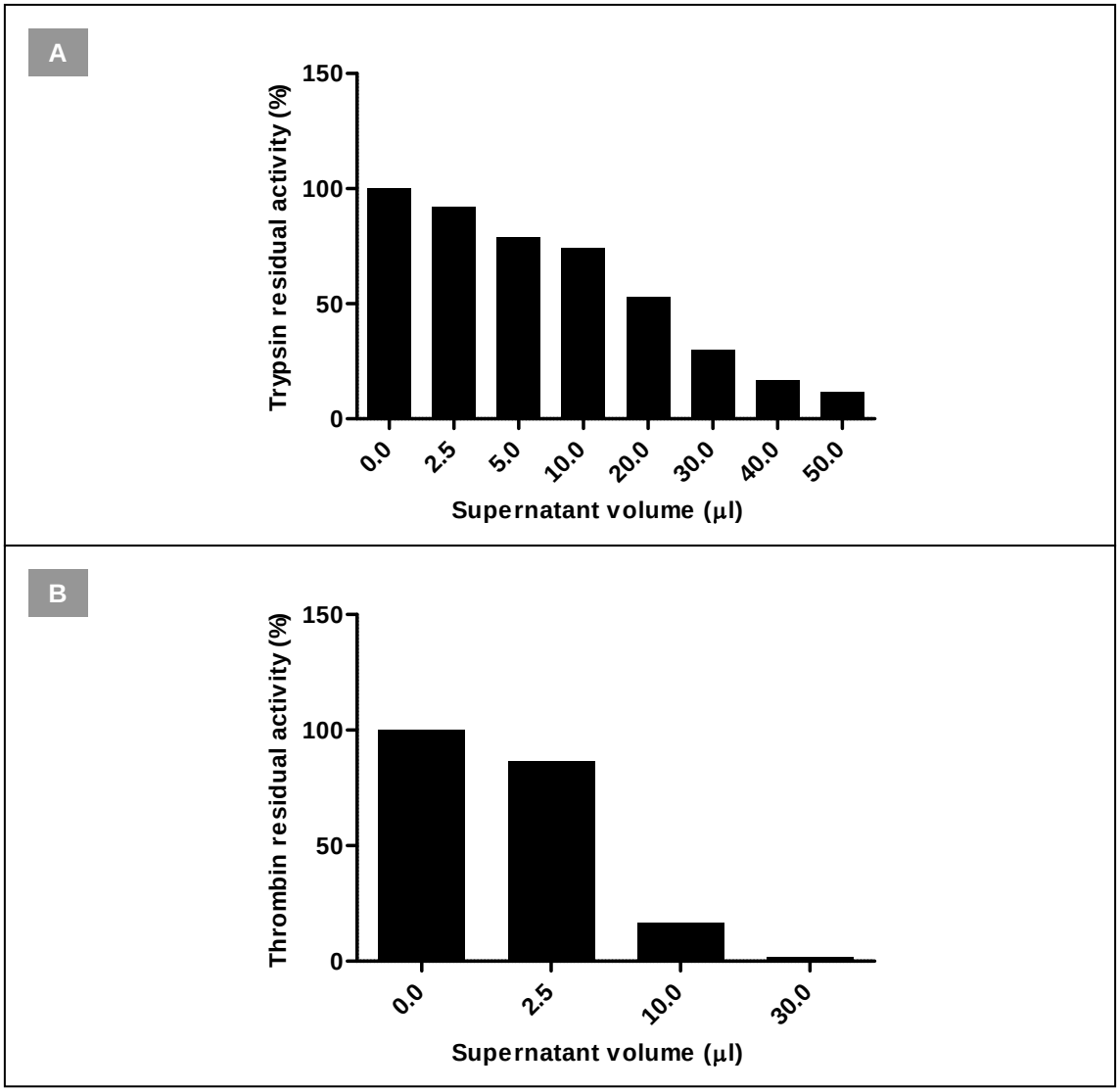


Figure 16 – The supernatant obtained after D1D2 expression showed anti-trypsin and anti-thrombin activity. Different volumes of supernatant (diluted 300x) were incubated with trypsin (A) and thrombin (B) and the residual activity was measured at 405 nm after addition of a synthetic substrate.

3.6. Purification of D1D2

The secretion of D1D2 to the culture medium helped the purification of the recombinant protein, since the yeast secretes only low levels of its own proteins. D1D2 was purified in two chromatographic steps: affinity chromatography followed

by an ion-exchange chromatography. For the first step of purification an immobilized-trypsin column was prepared. The efficiency of trypsin coupling was confirmed and the binding capacity of the column was quantified.

3.6.1. Preparation of the affinity column

The amount of trypsin coupled to the CNBr-activated Sepharose 4B was determined by subtracting the trypsin amount present in the flow-through obtained after incubating the medium with a trypsin solution of known concentration from the trypsin amount present in this last one. The concentration of trypsin solution was determined by measuring the absorbance at 280 nm and the value obtained was 62,322 mg/ml. In the 12 ml of trypsin solution there were 748 mg of trypsin. After coupling, the flow-through obtained was dialyzed and then quantified by measuring the absorbance at 280 nm. From the 748 mg of trypsin used only 16.5 mg was present in flow-through. In conclusion, ninety eight percent (731.5 mg) of the initial trypsin was bound to the medium.

Once prepared, the binding capacity of the column was quantified. Twelve milligrams of *Glycine max* trypsin-chymotrypsin inhibitor (7.5 kDa) was applied to a 1 ml column. The eluted fractions were quantified by measuring the absorbance at 280 nm. Only 1.691 mg of trypsin inhibitor was eluted from the column. Since the interaction of trypsin-trypsin inhibitor is of 1:1 molar ratio then the number of moles of active trypsin bound to the column is equal to the number of moles of eluted trypsin inhibitor. Thus, 2.25×10^{-7} mol of trypsin was active in the 1 ml column (binding capacity of the column). This means that only 12.8% (93.7 mg) of coupled trypsin (731.5 mg) could be inhibited.

3.6.2. Purification of D1D2

Recombinant D1D2 was first purified by affinity chromatography on an immobilized-trypsin affinity column. This first purification step allowed to separate the D1D2 protein and the putative degradation product that was able to interact with trypsin from other medium components. The fractions collected during elution with 1 M benzamidine showed two bands after separation on a SDS polyacrylamide gel (Figure 17, Lane 3-6). The higher molecular weight band with about 18 kDa is the more intense and corresponds to D1D2. If the lower molecular weight band (approximately 13 kDa) corresponds to the degradation product it probably contains the first domain of boophilin; otherwise it would not interact with trypsin. After elution the samples were pooled together and the benzamidine removed by dialysis. In spite of the high concentration of benzamidine used a significant amount of D1D2 remained attached to the column after elution (Figure 17, Lane 2). The elution by competition with 1 M benzamidine was not effective. Higher concentration of benzamidine would increase the cost associated to the elution, and would require extensive dialysis in order to eliminate the benzamidine from the purified fractions. Thus, the elution with benzamidine should be replaced by a more efficient and affordable elution method.

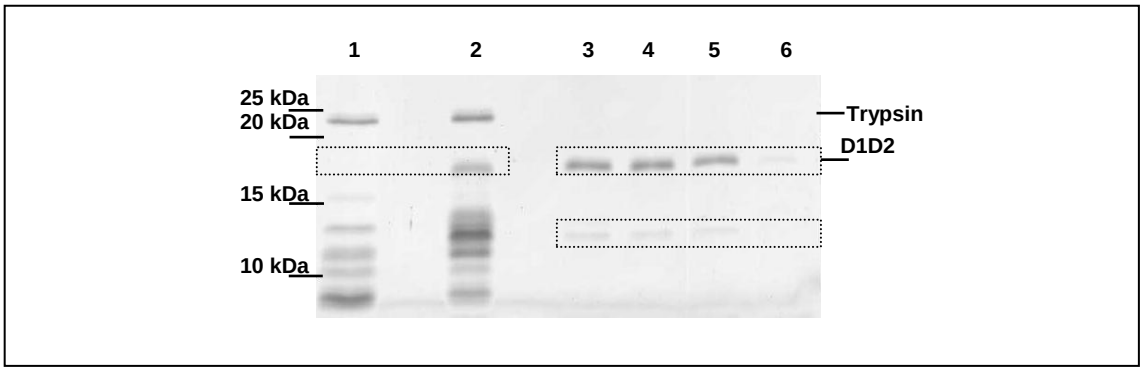


Figure 17 – Immobilized-trypsin affinity chromatography as first step on D1D2 purification. Ten microlitres of each sample were separated in a 15% polyacrylamide gel. The gel was stained with PageBlue^{TM}. Lane 1 – Column material before injection of supernatant. The higher molecular weight corresponds to trypsin and the lower molecular weight bands to trypsin degradation

products. Lane 2 – Column material after injection of supernatant and elution of bound proteins. D1D2 band is present in this lane (indicated by a rectangle). Lane 3 to 6 – Fractions collected during elution of bound proteins.

If the lower molecular band observed on the purified fraction after separation by SDS-PAGE corresponds to a degradation product of D1D2 it certainly contains D1. On the other hand, the difference on the molecular weight observed between the two bands suggests that the D2 is absent in the degradation product.

The theorical isoelectric point of D1 is 6.29, which should be similar to the isoelectric point of the degradation product found after the first step of D1D2 purification. The D1D2 has a lower theorical isoelectric point compared to D1 (4.41). Thus, the difference in pI could be used to separate these two proteins. The charge difference observed between D1D2 and D1 was significantly high to promote their separation by an ion-exchange chromatography in a positively charged column (Mono Q) at pH 8.0.

The injection of the sample and elution of bound protein were monitored at 214 nm and 280 nm on an ÄKTApurifier^{TM} (GE Healthcare). The most sensitive setting is at 214 nm, which detects absorbance of the peptide bond. The absorbance at 280 nm is due to the presence of tyrosine and tryptophan (aromatic residues). The tyrosine and tryptophan content of various proteins varies with narrow limits and thus, the absorbance at this wavelength as been used as a rapid and fairly sensitive measure of protein concentration. In the chromatogram obtained during elution of bound proteins with a linear gradient from 0 to 1.0 M NaCl two peaks were observed. The first peak was very small and almost undetectable at 280 nm (Figure 18, A, peak *). This peak eluted at a concentration of about 0-0.2 M of NaCl. The second peak was bigger and presented a small shoulder (Figure 18, A, peak **). This higher peak was obtained at approximately 0.4 M of NaCl. The ionic strength needed to produce this peak was higher compared to the first one, and thus it must correspond to D1D2 because this protein is more negatively charged than D1 at pH 8.0.

The flow-through collected during sample injection and the fractions collected during elution of bound proteins were resolved in a SDS-polyacrylamide gel. The concentration of the first peak fractions was too low and no proteins were detected by SDS-PAGE. The fractions of the second peak showed a single band with a molecular weight of about 18 kDa similar to the D1D2 protein (Figure 18, B). Despite the presence of a shoulder in the second peak, no additional bands were detected. Due to the fact that the sample injected into the column was fairly diluted (in order to decrease the concentration of salt) the protein concentration of the flow through was also low. Therefore, in order to analyse the content of the flow-through by SDS-PAGE the fraction was first lyophilized and resuspended in the smaller volume of water possible. On a SDS polyacrylamide gel, the concentrated flow-through showed a single band with a molecular weight similar (approximately 13 kDa) to the degradation products found in the affinity chromatography fractions (Figure 18, C).

The activity assay with trypsin showed anti-trypsin activity for all fractions of both peaks (Figure 19, A) and for flow-through (Figure 19, B). This assay, much more sensitive than SDS-PAGE, showed the complete separation of the two proteins obtained after affinity chromatography (Figure 19). A second activity assay with thrombin was also performed, confirming the presence of D1D2 only in the fractions corresponding to the second peak. This assay was not quantified and the inhibitory activity was only confirmed by the absence of yellow colour after incubation with substrate, in contrast with the positive control. The yellow colour results from substrate cleavage by the enzyme, being absence in the negative control, in which no enzyme or inhibitor was added.

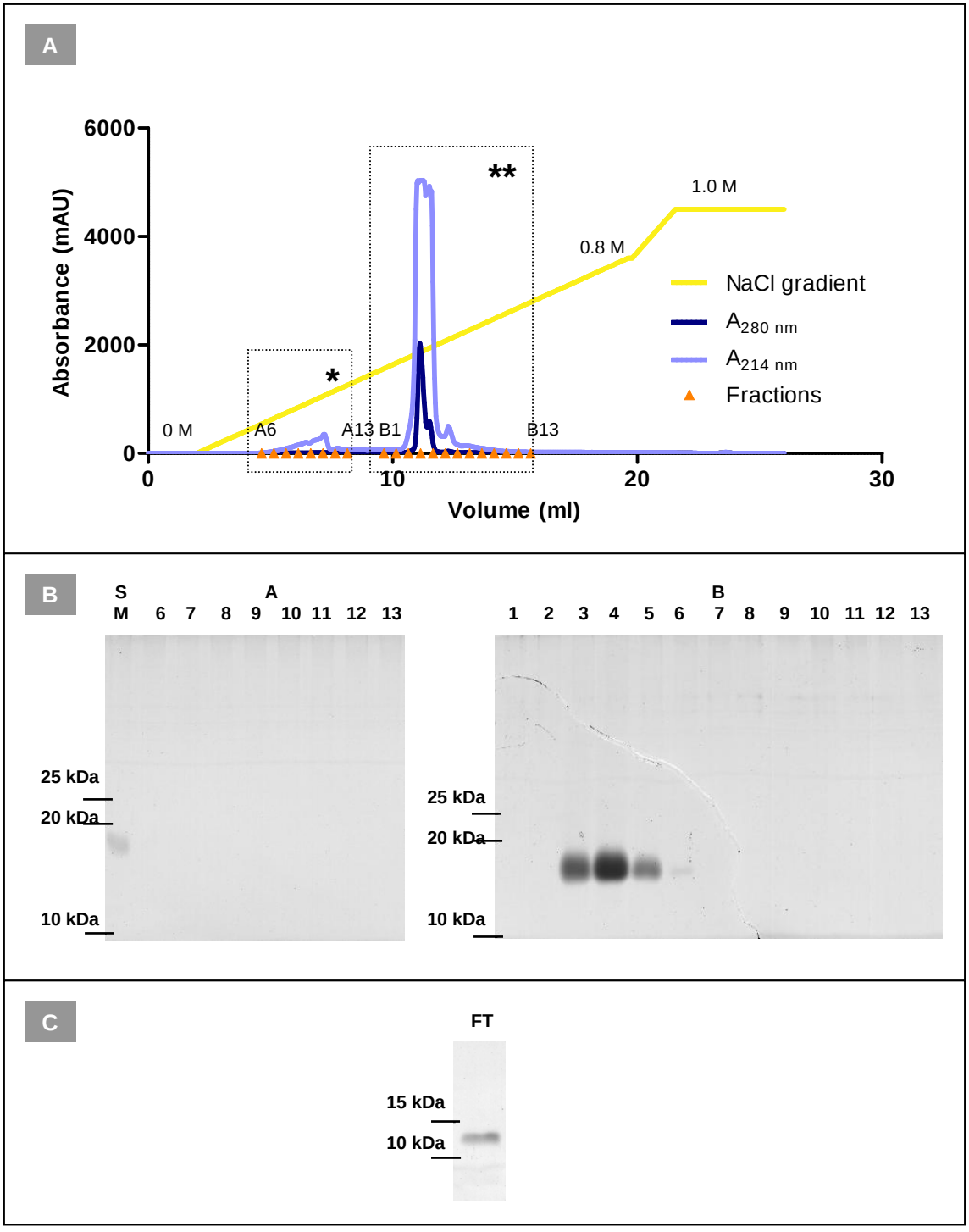


Figure 18 – Purification of D1D2 by ion exchange chromatography. A – Chromatogram obtained during elution of bound proteins. First peak (*) corresponds to the degradation products containing the D1 and the second peak (**) to D1D2. B – Ten microlitres of start material (SM) and of each collected fraction (orange in the chromatogram) were separated in 15% polyacrylamide gels. The gel was stained with PageBlue™. Only fractions B3 to B6 had detectable amounts of protein. C –

Ten microlitres of flow through material (FT), concentrated by liofilization, was resolved in a 15% polyacrylamide gel stained with PageBlue™.

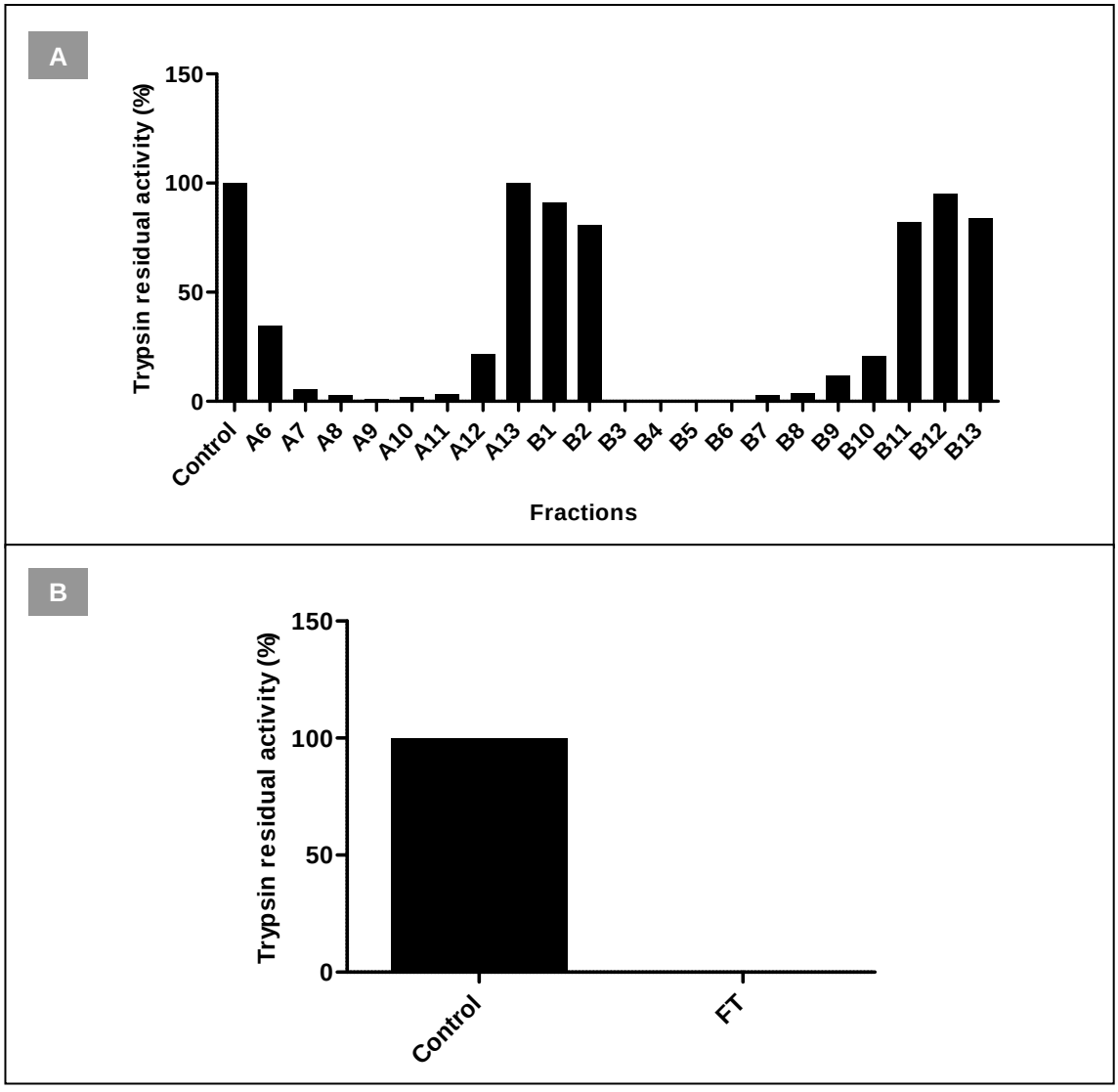


Figure 19 – Anti-trypsin activity was observed for fractions of both peaks and flow-through. Five microlitres of each collected fraction (A) and 50 µl of flow-through (B) were incubated with trypsin and its residual activity was determined by measuring the absorbance at 405 nm after adding the substrate.

3.7. Expression of D1

The expression of D1 was confirmed at the end of the fifth day of induction by SDS-PAGE and activity assays. The supernatant obtained after D1 expression showed a single band on the SDS polyacrylamide gel with an apparent molecular weigh of about 10 kDa (Figure 20, Lane 3). This band corresponds to D1. It was also possible to observe the presence of a light smear in D1 supernatant lane (Figure 20, Lane 3), which corresponds to the medium components and to the other proteins secreted by *P. pastoris*.

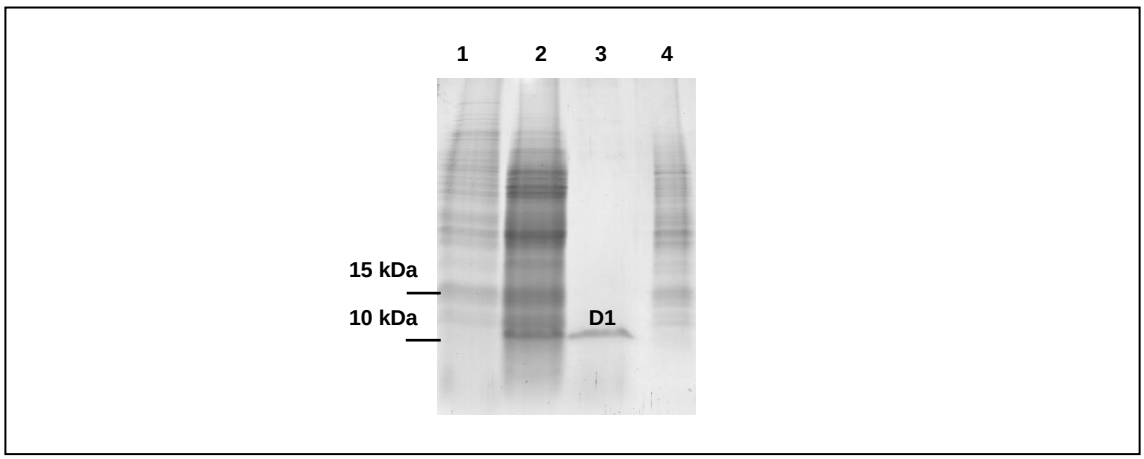


Figure 20 – Expression and secretion of D1. Ten microlitres of each sample were resolved in a 15% polyacrylamide gel stained with PageBlue™. Lane 1 – Culture before induction Lane 2 – Culture after induction; Lane 3 – Supernatant. Only the D1 band was detected. Lane 4 – Whole cell extract.

The presence of D1 in the supernatant was also confirmed by performing activity assays with trypsin and thrombin. Only trypsin was inhibited by the D1 supernatant (Figure 21). D1 was unable to inhibit thrombin, suggesting that the second domain of boophilin is required for the interaction of the inhibitor with the thrombin proteinase.

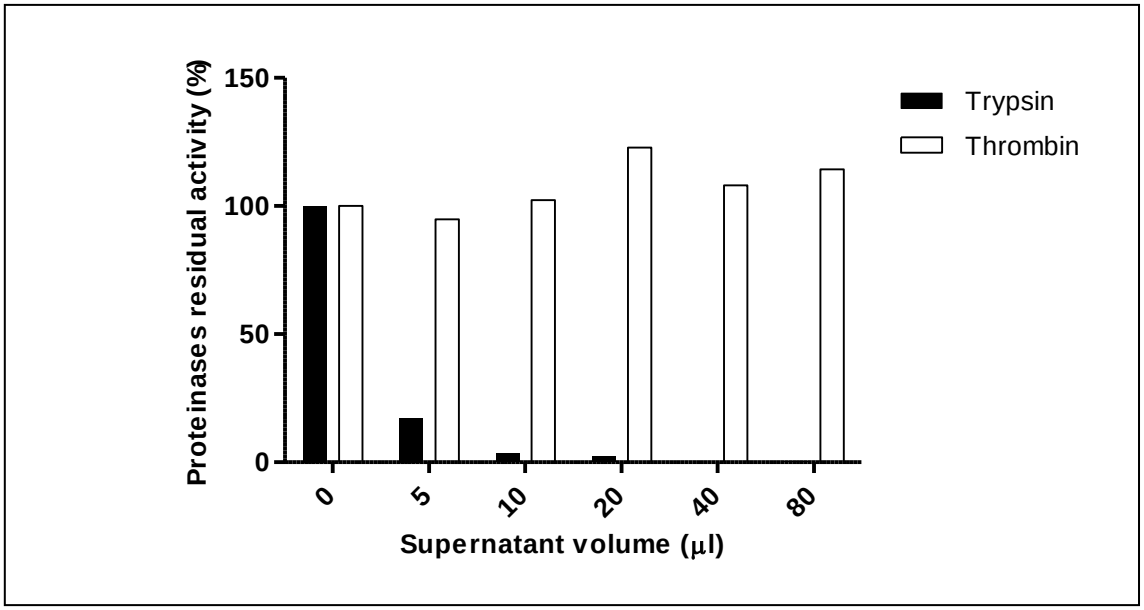


Figure 21 – Supernatant obtained after D1 expression showing anti-trypsin activity. No inhibition was observed for thrombin. Different volumes of supernatant (diluted 50x) were incubated with trypsin and thrombin and their residual activity was measured at 405 nm after adding the respective substrate.

3.8. Purification of D1

The secretion of D1 by *P. pastoris* was a great first step of purification, because *P. pastoris* only secretes low levels of its own proteins. D1 was purified by a single affinity chromatography in an immobilized-trypsin column. The column used to purify D1 was prepared and its binding capacity determined.

3.8.1. Preparation of the affinity column

The trypsin amount coupled to the AminoLink Plus Coupling Resin (Thermo Scientific Pierce) was determined by subtracting the trypsin amount present in flow-through obtained after incubating the resin with a trypsin solution of known

concentration from the trypsin amount present in the initial trypsin solution. The concentration of trypsin solution and flow-through was determined by measuring the absorbance at 280 nm. The total amount of trypsin present in trypsin solution was 246.5 mg and from those, only 1.41 mg were found in flow-through. In conclusion, 99.43% (245.09 mg) of the initial trypsin was coupled to the resin.

Once prepared, the binding capacity of the column was quantified by applying trypsin inhibitor from *Glycine max* (20.1 kDa) to a 500 µl column. The concentration of eluted material was determined by measuring the absorbance at 280 nm. The total mass of eluted inhibitor was 0.523 mg. Since the interaction of trypsin-trypsin inhibitor is of 1:1 molar ration then the number of moles of active trypsin bound to the column is equal to the number of moles of eluted trypsin inhibitor. Thus, 2.66×10^{-8} mol of active trypsin was present in the 500 µl column. The binding capacity of the column prepared was 5.32×10^{-8} mol per millilitre of column. In conclusion, from the 245.09 mg of coupled trypsin, 10.3% (25.3 mg) was active.

3.8.2. Purification of D1

The purification of D1 was achieved by a single chromatography on an immobilized-trypsin affinity column. The absorbance at 280 nm of eluted fractions was determined in order to identify the fractions containing protein (Figure 22, A). The fractions containing a detectable amount of protein at 280 nm were analysed by SDS-PAGE and only a band with a molecular weight similar to that of D1 was detected. An activity assay with trypsin was also performed (Figure 23) and all fractions tested showed significant inhibitory activity against trypsin. The fractions 4-11 were polled together since they contained significant amounts of D1.

In order to avoid protein loss, the column flow-through was collected in fractions (FT1, FT2, and FT3) and analysed by SDS-PAGE (Figure 22, B) and by activity assays (Figure 23). Those flow-through fractions containing inhibitory activity against trypsin were reinjected into the column and bound proteins eluted.

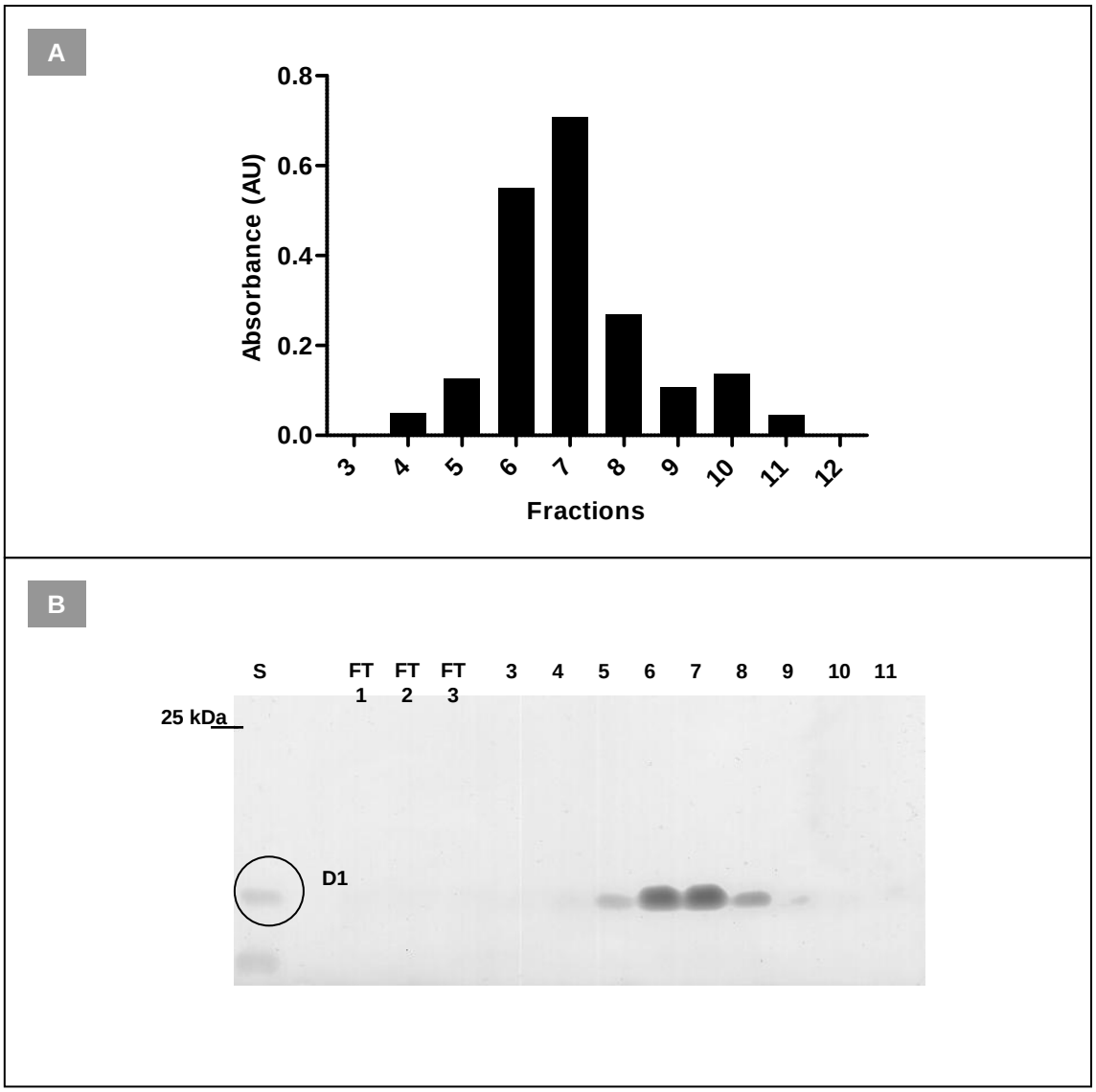


Figure 22 – Purification of D1 by affinity chromatography. A –The absorbance at 280 nm of the purified fractions was determined and only those fractions corresponding to the peak of elution (3-12) are shown in the graphic. No significant absorbance values were observed for the other fractions. B – Ten microlitres of supernatant (S) and of each flow through fraction (FT1, FT2, FT3) and 5 microlitres of each purified fraction (3-11) were separated in a 15% polyacrylamide gel stained with PageBlue^{TM}.

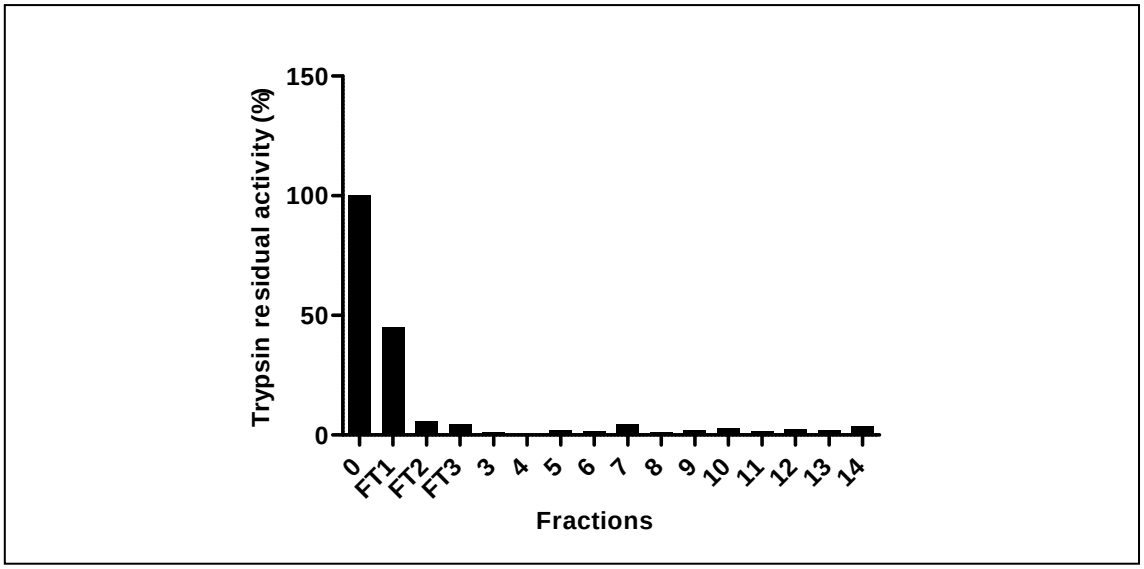


Figure 23 – Anti-trypsin activity was observed for all purified fractions of D1. One microlitre of each purified fraction and 5 µl of each flow through fraction were incubated with trypsin and its residual activity was measured at 405 nm after adding the substrate. Anti-trypsin activity was observed for all purified fractions and for all flow-through fractions (FT1, FT2, and FT3).

3.9. Trypsin titration

In order to study the inhibition of trypsin by D1 or by D1D2 it was necessary to first titrate trypsin. Trypsin is a very active enzyme that is prone to autolysis. Trypsin self-cleavage leads to the inactivation of its enzymatic activity. Therefore the amount of active trypsin in stock solutions subsequently used in the enzymatic assays was calculated. This was done by incubating trypsin with NPGB and by following the reaction at 410 nm for 20 minutes (measurements every 5 seconds) at 37ºC (Figure 24, A, dark blue line). The cleavage of NPGB produces a yellow colour detectable at the wavelength used.

NPGB is a suicide trypsin substrate meaning that after being cleaved by the protease the cleavage product is not released. Since the cleavage product remains attached to trypsin the cleavage of additional substrate molecules is consequently avoided. This feature allows the quantification of active trypsin in a

certain solution by determination of the amount of substrate (from a known substrate concentration solution) that is cleaved by the enzyme. However, NPGB by itself is slightly unstable in aqueous solutions and starts to degrade (Figure 24, A, light blue line). For this reason, the degradation of NPGB in the absence of protease was also followed in the same experimental conditions. This was used as a blank control. The absorbance values corresponding to the cleavage of NPGB by trypsin were later calculated by subtracting the absorbance values of the NPGB degradation from the absorbance values of the mixture (Figure 24, B). Three independent experiments were performed.

A linear regression analysis was performed on absorbance values (dependent variable) and time of reaction (independent variable) obtained for each experiment. The slope ($\Delta y/\Delta x$, where Δy is the difference between two absorbance values and Δx is the interval of time in which Δy is observed) given by the equation $Y = mX + b$ (where *Y* corresponds to the absorbance represented in AU, *m* to the reaction rate in AU sec^{-1}, *X* to the time in seconds, and *b* to the absorbance in AU at time 0) obtained after linear regression analysis is near zero revealing the inhibition of trypsin by NPGB (see Table 3). The number of active trypsin molecules is given by the number of substrate molecules cleaved at time 0. The absorbance at time 0 was determined and the concentration of active trypsin calculated by using the equation $A = C\varepsilon\ell$, in which *A* corresponds to the absorbance represented in (AU), *C* to the concentration of cleaved NPGB in molarity (M), ε to the extinction coefficient of NPGB (= 16.595 l $mmol^{-1}$ cm^{-1}) and ℓ to the path length of the sample in centimetre. The concentration of active trypsin was given by the mean of the concentration values determined in each experiment (Table 3).

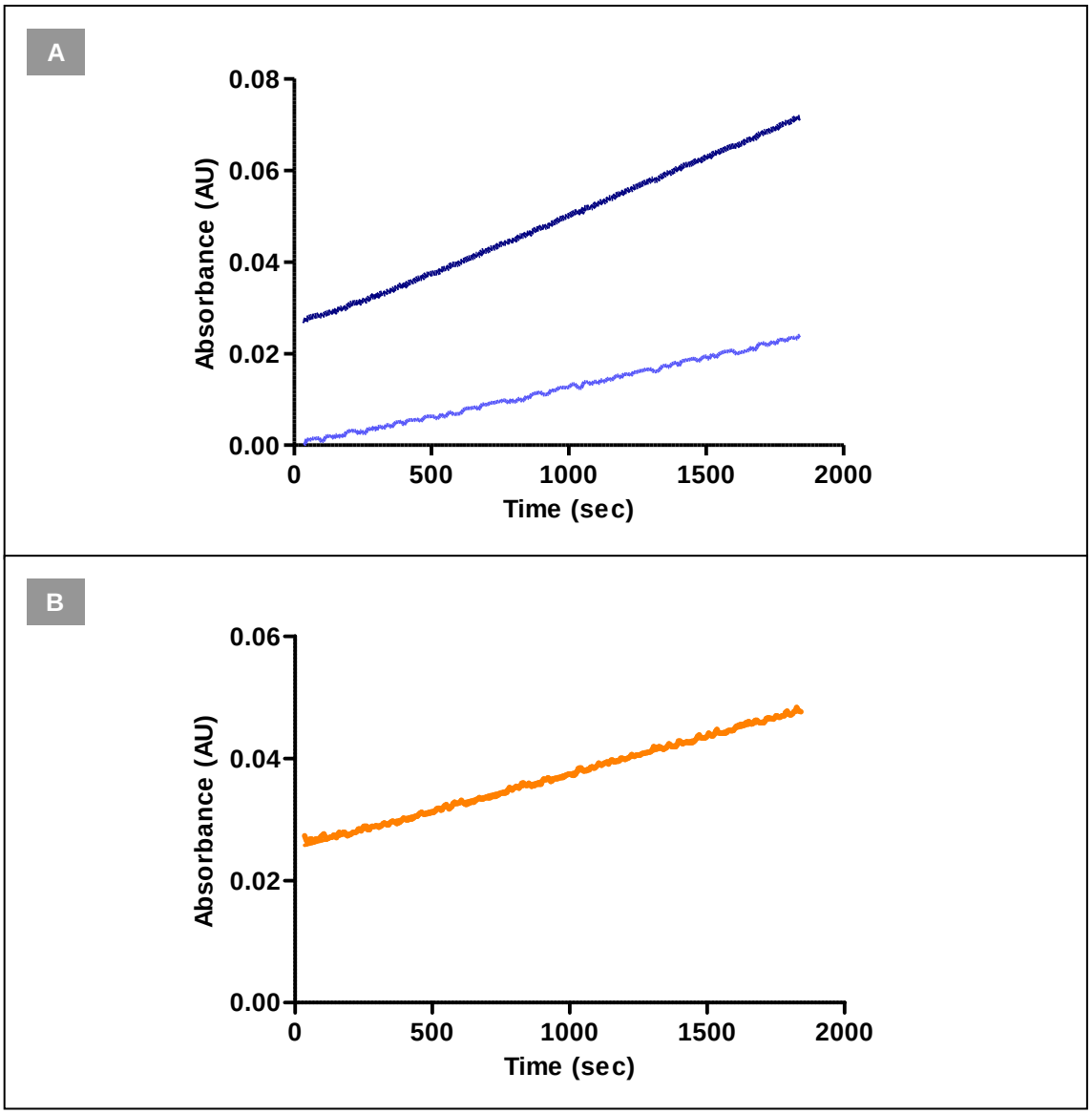


Figure 24 – Titration of trypsin with NPGB. A – A mixture containing trypsin and NPGB was incubated at 37 ºC and the reaction was followed by measuring the absorbance at 410 nm (dark blue). In parallel, the degradation of NPGB in the assay buffer was also followed (light blue). B – The cleavage of NPGB by trypsin was obtained by subtracting the absorbance values produced by NPGB degradation from the absorbance values of the mixture containing trypsin and NPGB.

Table 3 – Active trypsin concentration. Three independent experiments were performed. The active trypsin concentration was given by the mean of the concentration values obtained for each experiment.

Number of experiments	Linear regression	Active trypsin concentration (mg/ml)
1	$Y = 9.0x10^{-6}X + 2.73x10^{-2}$ $R^2 = 0.990$	7.83
2	$Y = 8.0x10^{-6}X + 2.53x10^{-2}$ $R^2 = 0.988$	7.26
3	$Y = 5.0x10^{-5}X + 2.53x10^{-2}$ $R^2 = 0.998$	7.26
		7.45 ± 0.33

3.10. Determination of the maximum reaction rate and Michaelis-Menten constant of trypsin against Bz-Phe-Val-Arg-pNA

In order to determine the inhibition constants of D1 and D1D2 against trypsin it was necessary to calculate not only the concentration of active trypsin in the stock solutions used (explained above) but also to determine the Michaelis-Menten constant of trypsin against the synthetic substrate Bz-Phe-Val-Arg-pNA (explained below).

The maximum reaction rate (V_{max}) and the Michaelis-Menten constant (K_m) are two parameters which define the kinetic behaviour of an enzyme. The V_{max} is a rate of reaction and it is the limiting value of the initial rate when all the active sites are occupied with substrate. This parameter is related to the amount of enzyme or substrate used in the reaction. By contrast, K_m is independent on the enzyme or substrate concentration and it is always the substrate concentration required to achieve half of V_{max}. This parameter is inherent to the enzyme and is a measure of its activity. K_m is related to the dissociation constant and gives an indication of the enzyme affinity to a specific substrate: high K_m values indicate high affinity.

In order to determine these two parameters, trypsin was incubated with different concentrations of the synthetic substrate Bz-Phe-Val-Arg-pNA and the reaction was followed at 405 nm for 1 hour (measurements every minute) at 37 ºC.

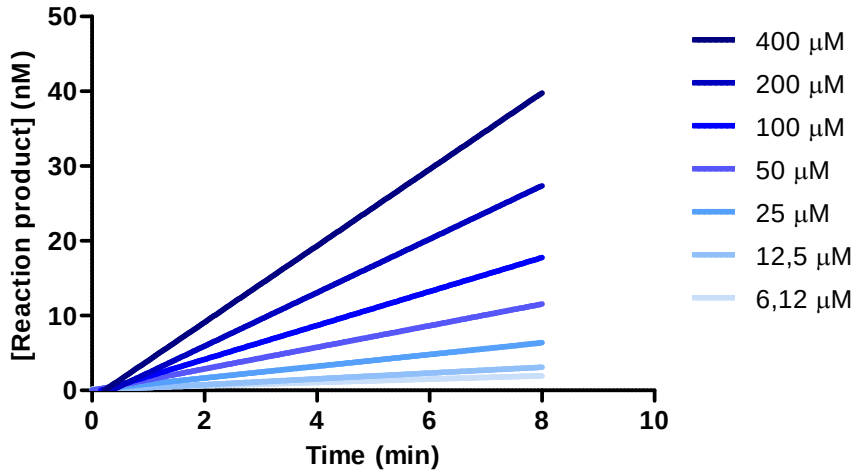
The concentration of product for each reading point was calculated using the equation $A = C\varepsilon\ell$ where *A* corresponds to the absorbance represented in absorbance units (AU), *C* to the concentration of reaction product in molarity (M), ***ε*** to the extinction coefficient of the reaction product (=10.4 l $mmol^{-1}$ cm^{-1}) and ℓ (=0.2598 cm) to the path length of the sample. A linear regression analysis was performed on the linear region of each hydrolysis curve (where the concentration of product was the dependent variable and the time of reaction the independent variable) and only the curves with R square above 0.97 were considered for the determination of V_{max} and K_m. The reaction rate for each substrate concentration was given by the slope of the respective equation $Y = mX + b$ (where *Y* corresponds to the concentration of reaction product represented in molarity, *m* to the slope of the straight line defined by the equation, *X* to the time in minutes, and *b* to the product concentration in molarity at time 0) obtained by linear regression analysis (Figure 25).

In order to obtain V_{max} and K_m a non-linear regression analysis using the Michaelis-Menten equation

$$v_0 = \frac{V_{max}[S]}{K_m + [S]}$$

(v_0 corresponds to initial reaction rate, V_{max} to the maximum reaction rate, K_m to the Michaelis-Menten constant and [*S*] to the substrate concentration) was performed using reaction rate values (dependent variable) and substrate concentration (independent variable). The V_{max} and K_m values obtained were of 10.62 µM/min and 367.7 µM, respectively (see details on Figure 26).

A

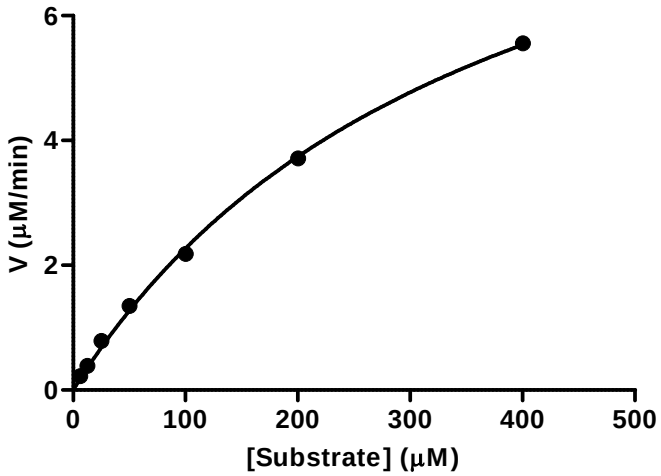


B

Substrate concentration (μM)	Equation	R^2
400	Y = **5.5580**X - 2.3527	0.9994
200	Y = **3.7127**X - 1.4575	0.9991
100	Y = **2.1816**X + 0.040	0.9990
50.0	Y = **1.3462**X + 0.3172	0.9970
25.0	Y = **0.7854**X + 0.0932	0.9986
12.5	Y = **0.3858**X $2.0x10^{-15}$	0.9938
6.12	Y = **0.2221**X + 0.1810	0.9720

Figure 25 – Kinetic reaction of trypsin in presence of different substrate concentrations. A – Linear region of the hydrolysis curves obtained by incubating trypsin with different substrate concentrations. B – Linear regression analysis performed on the data present in A. The reaction rate of each reaction was given by the slope of the equation obtained by linear regression analysis (bold).

A



B

Best-fit values	
V_{max}	10,62 µM/min
K_m	367,4 µM
Standard Error	
V_{max}	0,5373 µM/min
K_m	31,71 µM
95% Confidence Intervals	
V_{max}	9.237 to 12.00 µM/min
K_m	285.9 to 448.9 µM
Goodness of Fit	
Degrees of Freedom	5
R square	0,9987
Absolute Sum of Squares	0,03060 μM^2
$S_{y.x}$	0,07822 µM
Number of points	
Analyzed	7

Figure 26 – Determination of maximum reaction rate and Michaelis-Menten constant of trypsin against Bz-Phe-Val-Arg-pNA. A – Graphic representation showing the reaction rate of trypsin in presence of different substrate concentrations. B – Statistical data obtained by non-linear analysis using the Michaelis-Menten equation.

3.11. Determination of the inhibition constant of D1D2 and D1 towards bovine trypsin

The inhibition constant (K_i) is a measure of the affinity between an inhibitor and an enzyme. This parameter is a dissociation constant that gives the inhibitor concentration required to produce half of the maximal inhibition. As a dissociation constant, a lower K_i indicates that the inhibitor does not dissociate easily from an enzyme, which is translated by a more effective inhibition.

One way of understanding the inhibition mechanism of a specific inhibitor is to measure the enzyme velocity at a variety of inhibitor concentrations. Thus, trypsin was incubated with different concentrations of D1D2 or D1, and after adding the substrate the kinetic reaction was monitored by measuring the absorbance at 405 nm for 1 hour (measurements every 15 seconds) at 37 ºC.

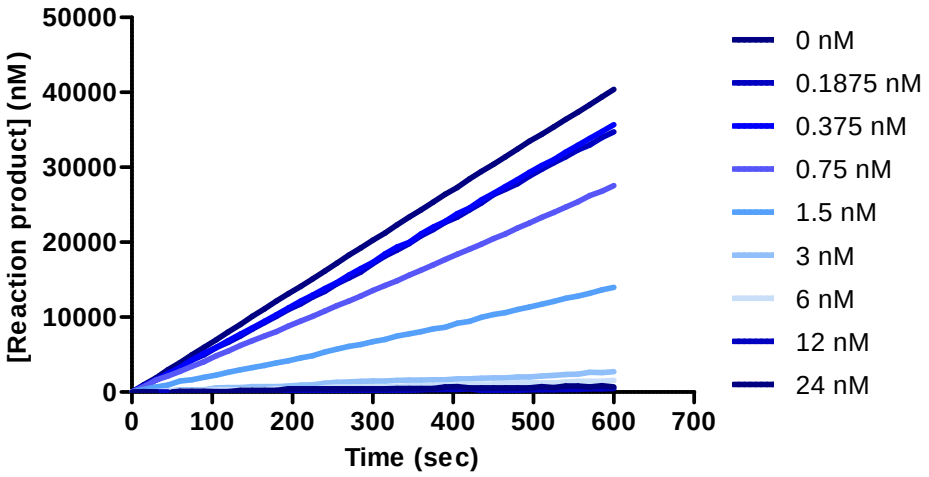
The concentration product at each reading point was calculated using the equation $A = C\varepsilon\ell$ where *A* corresponds to the absorbance represented in AU, *C* to the concentration of reaction product in molarity, ***ε*** to the extinction coefficient of the reaction product (=10.4 l mmol^{-1} cm^{-1}) and ℓ (=0.5196 cm) to the path length of the sample. A linear regression analysis was performed on the linear region of each hydrolysis curve (where the concentration of reaction product and the reaction time were the dependent and the independent variables, respectively) and only the curves with R square above 0.97 were considered for the determination of the reaction rate. The reaction rate was given by the slope of the equation $Y = mX + b$ (*Y* corresponds to the concentration of reaction product represented in molarity, *m* to the slope of the straight line defined by the equation, *X* to the time in seconds, and *b* to the concentration of reaction product in molarity at time 0) obtained by linear regression analysis (Figures 27 and 28).

The inhibition constants were obtained by non-linear regression analysis using the Morrison equation

$$v = v_0\left(1-\left(\left([E]_0+[I]+K_{I-app}\right)-\frac{\sqrt{\left([E]_0+[I]+K_{I-app}\right)^2-4[E]_0[I]}}{2[E]_0}\right)\right)$$

(v is the measured reaction rate, v_0 is the reaction rate in the absence of inhibitor, $[E]_0$ is the total enzyme concentration, $[I]$ is the added inhibitor concentration and $K_{I\text{-}app}$ is the apparent equilibrium inhibitor constant) on the reaction rate values (dependent variable) and inhibitor concentrations (independent variable) (Figures 29 and 30). The inhibition constant determined for D1D2 was $2.366 \times 10^{-2} \pm 2.976 \times 10^{-2}$ nM and for D1 was $0.9438 \times 10^{-2} \pm 2.873 \times 10^{-2}$ nM. The inhibition constant determined for D1D2 was approximately two times higher than the value obtained for D1. This can be explained by the flexibility of the D1D2-trypsin complex. Assuming that the inhibition of trypsin by D1D2 is achieved by the Lys31 of the first Kunitz domain, the second domain would remain free and the D1D2 would be very flexible. This flexibility could lead to an increase of the dissociation of D1D2-trypsin complex. On the other hand, the interval in which the K_i values of D1 and D1D2 diverge were overlapped and thus the differences observed were not significant. The results obtained support the hypothesis that the second domain of boophilin is not involved in trypsin inhibition. However, the elucidation of the inhibition mechanism requires additional data from other techniques such as X-ray crystallography.

A

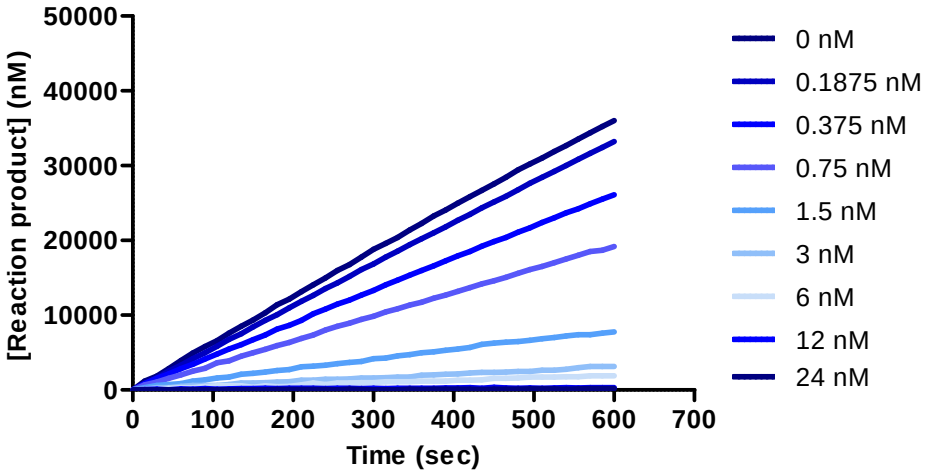


B

D1D2concentration (nM)	Equation	R^2
0	Y = **66.6000**X + 183.6924	0.9998
0.1875	Y = **58.0294**X – 176.7243	0.9998
0.375	Y = **60.1282**X – 429.0560	0.9999
0.75	Y = **46.2644**X – 178.5264	0.9998
1.5	Y = **23.9578**X – 333.3454	0.9990
3	Y = **7.8549**X – 2583.5890	0.9988
6	Y = **3.3554**X – 692.1107	0.9992
12	Y = **1.5589**X – 520.6742	0.9992
24	Y = **0.8367**X + 176.9457	0.9932

Figure 27 – Determination of trypsin reaction rate in presence of different D1D2 concentrations. A – Linear region of the hydrolysis curves obtained by incubating trypsin with different D1D2 concentration. B – Linear regression analysis performed on the data present in A. The reaction rate of trypsin for each inhibitor concentration was given by the slope of the equation obtained by linear regression analysis (bold).

A

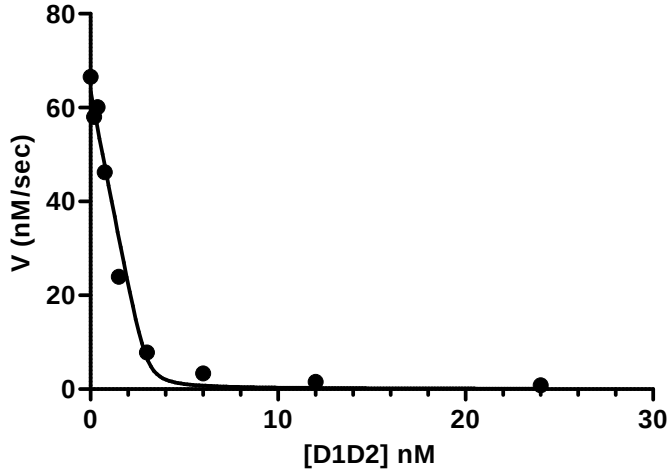


B

D1D2concentration (nM)	Equation	R^2
0	Y = **60.7232**X + 48.6963	0.9999
0.1875	Y = **60.7440**X – 217.6116	0.9999
0.375	Y = **67.5390**X – 18254.4971	0.9998
0.75	Y = **39.0818**X – 4782.5355	0.9994
1.5	Y = **18.8004**X – 3620.9887	0.9999
3	Y = **4.8767**X – 1482.1912	0.9998
6	Y = **1.3955**X – 209.6986	0.9992
12	Y = **0.5087**X – 3.7704	0.9950
24	Y = **3.3181**X – 8.1759	0.9832

Figure 28 – Determination of trypsin reaction rate in presence of different concentrations of D1. A – Linear region of the hydrolysis curves obtained by incubating trypsin with different concentrations of D1. B – Linear regression analysis performed on the data present in A. The reaction rate of trypsin for each D1 concentration was given by the slope of the respective equation obtained by linear regression analysis (bold).

A

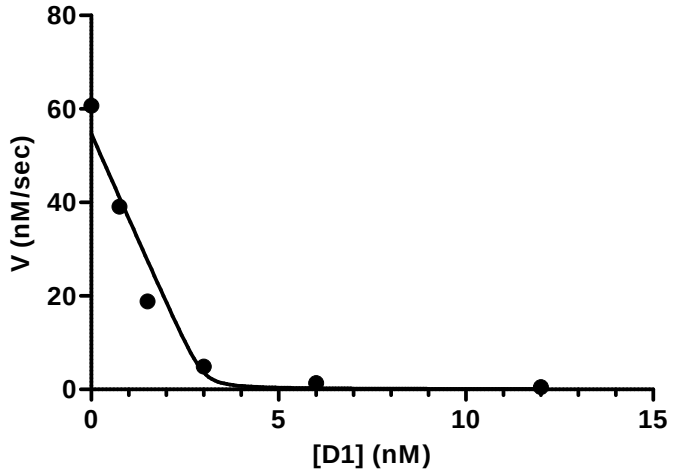


B

Best-fit values	
V_o	63.54 nM/sec
[T]	3000 nM
K_i	$2.386x10^{-2}$ nM
[S]	200000 nM
K_m	367400 nM
Standard Error	
V_o	2.226 nM/sec
K_i	$2.976x10^{-2}$ nM
95% Confidence Intervals	
V_o	58.27 to 68.80 nM/sec
K_i	-0.04653 to 0.09425 nM
Goodness of Fit	
Degrees of Freedom	7
R square	0.9810
Absolute Sum of Squares	117.7 nM^2
$S_{y.x}$	4.101 nM
Number of points	
Analyzed	9

Figure 29 – Determination of the inhibition constant of D1D2 towards trypsin. A – Graphic obtained after non-linear regression analysis using the Morrison equation showing the decrease of the reaction rate of trypsin with the increase of D1D2 concentration. B – Statistic data obtained by non-linear analysis using the Morrison equation, where V_o is the enzyme velocity in the absence of inhibitor, [T] is the trypsin concentration used, K_i is the inhibition constant obtained, [S] is the substrate concentration used, and K_m is the Michaelis-Menten constant of trypsin against the substrate used.

A



B

Best-fit values	
V_o	54.77 nM/sec
[T]	3000 nM
K_i	$0.9438x10^{-2}$ nM
[S]	200000 nM
K_m	367400 nM
Standard Error	
V_o	4.104 nM/sec
K_i	$2.873x10^{-2}$ nM
95% Confidence Intervals	
V_o	43.38 to 66.17 nM/sec
K_i	-0.07033 to 0.08921 nM
Goodness of Fit	
Degrees of Freedom	4
R square	0.9593
Absolute Sum of Squares	120.9 nM^2
$S_{y.x}$	5.498 nM
Number of points	
Analyzed	6

Figure 30 – Determination of the inhibition constant of D1 towards trypsin. A – Graphic obtained after non-linear regression analysis with the Morrison equation showing the decrease of the reaction rate of trypsin with the increase of D1 concentration. B – Statistic data obtained by non-linear regression analysis using the Morrison equation, where V_o is the enzyme velocity in the absence of inhibitor, [T] is the trypsin concentration used, K_i is the inhibition constant obtained, [S]

is the substrate concentration used, and K_m is the Michaelis-Menten constant of trypsin against the substrate used.

3.12. D1 crystallization

X-ray crystallography has been used to study the interaction between proteins. As the name suggests, for application of X-ray diffraction to a macromolecule it must be first crystallized. The crystallization step has emerged as a primary obstacle in macromolecular crystallography since the highly dynamic structure of proteins interferes with the crystallization process. Because there is strong data supporting the hypothesis that the second domain of boophilin is not necessary for the inhibition of trypsin, the crystallization of the D1D2 in complex with trypsin could be very difficult. Once the second domain of boophilin does not interact with trypsin, it would remain free and boophilin would be very flexible. This flexibility produces a very heterogeneous sample, which frequently affects crystallization. Because D1 seems to be the only domain of boophilin that interacts with trypsin, the problem of the complex flexibility could be solved by using the trypsin-D1 complex instead of trypsin-D1D2 complex to study the inhibition mechanism of boophilin against trypsin. However, it is first necessary to confirm that the isolated first domain of boophilin exhibits a conformation similar to the first domain of D1D2. In order to determine the crystallographic structure of D1, this protein was first crystallized.

Crystallization assays were setup with purified recombinant D1 using the sitting drop method. In sitting drop vapour diffusion the drop containing the protein solution combined with precipitant solution is dispensed onto a small platform, which is contiguous through the vapour phase with a reservoir of much larger volume. A group of very small needle crystals was observed 13 days after the incubation of D1 solution with the precipitant (Figure 31, A). Cuboidal crystals started to grow from this group being observed for the first time on the 34th day (Figure 31, B). Their growth was followed for 4 months (Figure 31) after which the size of the crystals did not seem to increase. Afterward some crystals were cryo-

cooled in liquid nitrogen. One of these cryo-cooled crystals was subjected to X-ray diffraction (section 3.13) and the collected data was used to determine the crystallographic structure of D1 (section 3.14).

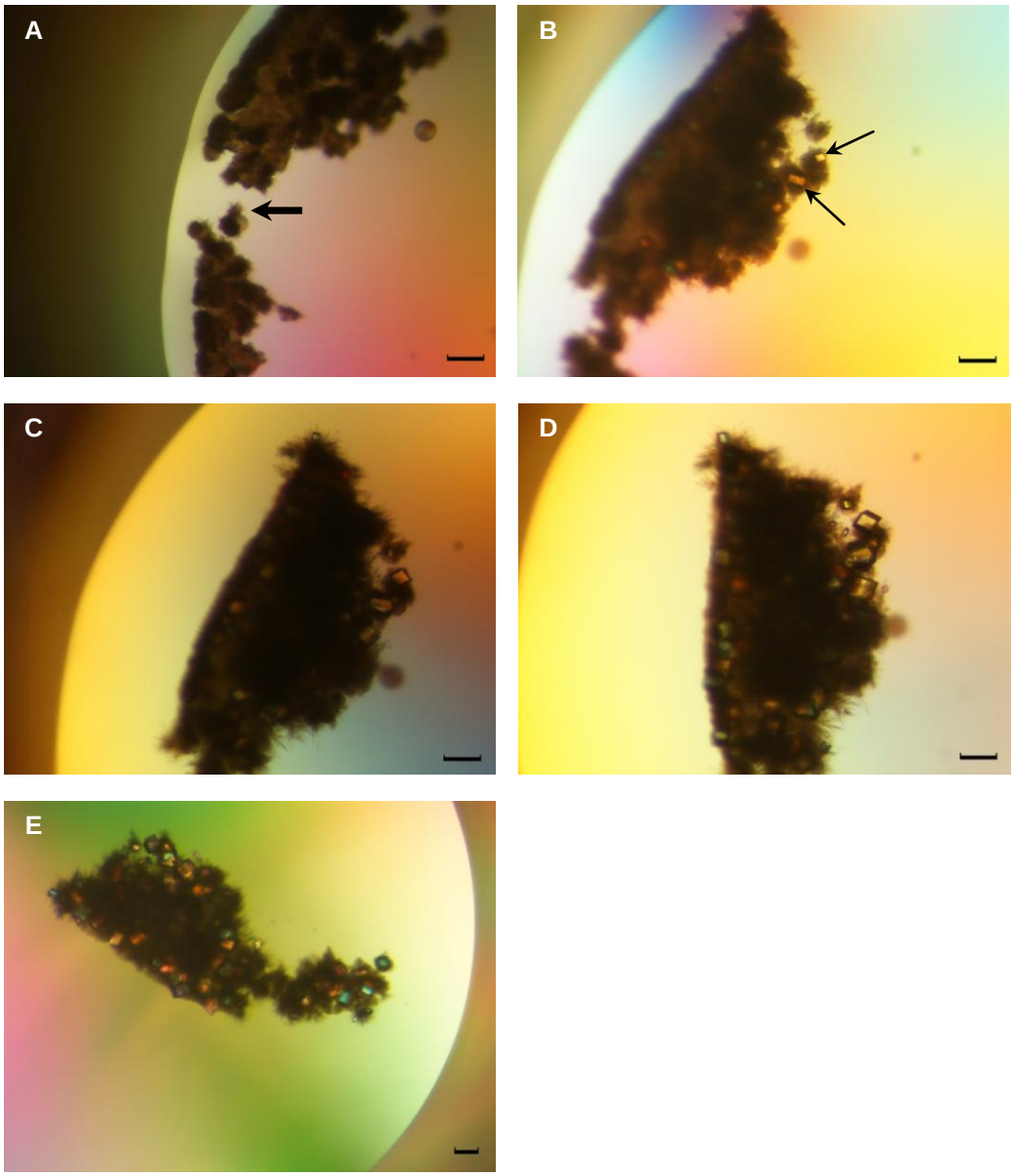


Figure 31 – Appearance and growth of D1 crystals. The drop contained equal volumes of D1 solution and precipitant. Growth was followed for 4 months. The needle crystals group (large arrow) and cuboidal crystals (thin arrows) were first observed on the 13rd day (A) and 34th day (B), respectively. The crystal growth was followed (C – 57th day; D – 85th day) until the day of their collection (E – 102th day) (Scale bars, 0.2 mm).

3.13. Data collection, processing and refinement

Diffraction data to 1.80 Å was collected from a single cryo-cooled (100 K) D1 crystal. The crystal measured belonged to the orthorhombic space group $P2_12_12_1$, with unit cell constant of a = 24.81 Å, b = 30.74 Å, c = 54.97 Å, and had an unusual low solvent content of about 23%. The molecular coordinates of the first domain of boophilin (residues Gln16 to Ala58) in complex with thrombin from PDB entry 2ODY (Macedo-Ribeiro *et al*., 2008) were used as a search model to solve the structure of D1. The model was subjected to alternating cycles of manual building with *Coot* (Emsley *et al.*, 2010) and refinement with PHENIX (Adams *et al.*, 2010). The data-collection, processing and refinement statistics are summarized in Table 4.

Table 4 – Statistics of data collection and refinement.

Crystallographic analysis	
Resolution range (Å)	14.74 – 1.80
Space group	$P2_12_12_1$
Unit cell dimensions (Å)	a = 24.8 b = 30.7 c = 55.0
Multiplicity (overall/outer shell)	5.1/2.5
R_{merge} [a] (overall)	0.058
Completeness (%) (overall/outer shell)	97.19/87.6
I/σ (I) (overal/outer shell)	7.1/5.1
Mathews coefficient ($Å^3$ Da^{-1})	1.60
Solvent content (%)	23.3
Structure Refinement	
R_{factor} [b]/Free R_{factor} [c] (%)	0.1400/0.1689
Nº of unique reflections (working/test set)	3894/182
Water molecules	68
Total number of non-hydrogen atoms	511
Number of protein non-hydrogen atoms	443
Average overall B-factor ($Å^2$)	9.79
Average protein B-factor ($Å^2$)	8.17

Average water B-factor (Å^2)	20.31
r.m.s.d. bond lengths (Å)	0.014
r.m.s.d. bond angles (º)	1.423
Ramachandran plot statistics	
Residues in allowed regions	55 (98.21%)
Residues in disallowed regions	0 (0%)
Estimated coordinate error	
E.s.d. from Luzzati plot (Å)	0.2574
DPI[d] (Å)	0.1067

[a]$R_{merge} = \Sigma_h\Sigma_i|I_{hi}-\langle I_h\rangle|/\Sigma_h\Sigma_i\langle I_h\rangle$, where I_{hi} is the observed intensity of the i-th measurement of reflection (h), including symmetry-related ones, and $\langle I_h\rangle$ is the mean intensity of the i observations of reflection h over all measurements of I_{hi}.

[b]$R_{factor} = \Sigma||F_o|-|F_c||/\Sigma|F_o|$, where $|F_o|$ and $|F_c|$ are observed and calculated structure factor amplitudes, respectively.

[c]Free R_{factor} is the cross-validation R-factor computed for a randomly chosen subset of 5% of the total number of reflections, which were not used during refinement.

[d]Diffraction-data precision indicator

3.14. D1 structure

The crystallographic structure of D1 presents a canonical Kunitz fold domain composed by two anti-parallel β-sheets followed by an α-helix. The structure is maintained by the presence of three disulfide bridges established between Cys21-Cys71, Cys30-Cys54 and Cys46-Cys67 (Figure 32).

The structure of D1 by itself reveals a conformation that is similar to the conformation of the first domain (residues Gln16 to Ala58) in the context of the D1D2 in complex with thrombin (PDB entry 2ODY) (Macedo-Ribeiro *et al.*, 2008) (see superposition of the two structures in Figure 33). The reactive site loop of the first domain of D1D2 presumably involved in trypsin inhibition is preserved in D1. The mutation found in D1 (Glu27→Gln27) by DNA sequencing did not introduce any significant structure modifications (Figure 33). Since D1 protein by itself seems

to fold correctly, we hypothesize that this might be a suitable protein to study the interaction between trypsin and boophilin. Besides the disulfide bonds, no additional post-translational modifications were observed. This fact validates the use of *Pichia pastoris* as an expression system for the production of D1. Additional post-translational modifications, such as glycosylation, could affect not only the activity/affinity of D1, but also its crystallization.

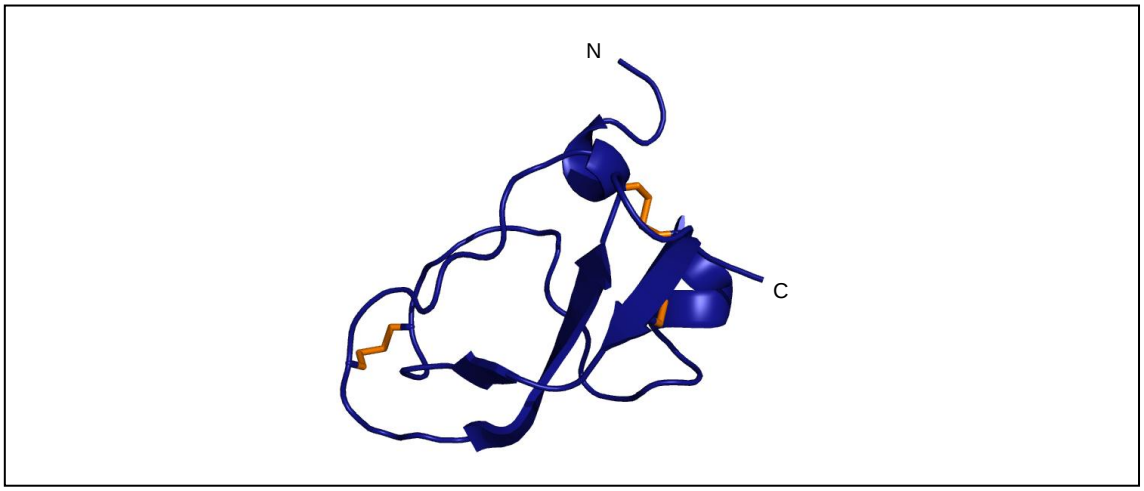


Figure 32 – Crystallographic structure of D1. The structure of D1 shows two anti-parallel β-sheets followed by an α-helix. The structure is maintained by the presence of three disulfide bridges (orange).

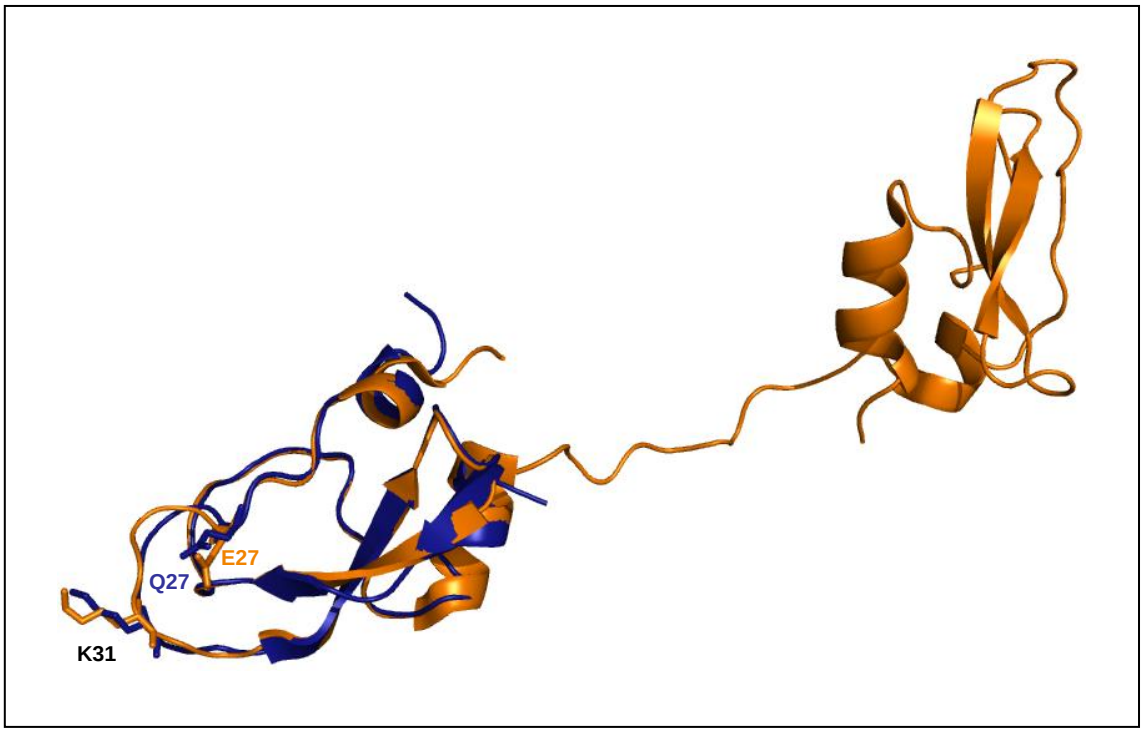


Figure 33 – Crystallographic structure of D1 (blue) aligned with the first domain of D1D2 (orange). The structure of D1 reveals a conformation similar to the structure of the first domain of full-length boophilin. The mutation found (residue position 27) by DNA sequencing does not affect the structure of the D1. It is also possible to observe that the reactive-loop (containing K31) is preserved.

3.15. Trypsin-D1 complex purification

With the aim of crystallizing D1 in complex with trypsin for better understanding the mechanism of inhibition of trypsin by boophilin, thrombin-D1 complexes were prepared. The trypsin-D1 complex was formed by incubating trypsin with a molar excess of D1 (mixture). This excess of D1 was used to ensure that all trypsin molecules were in complex with a D1 molecule. After incubation, excess D1 was removed by a size-exclusion chromatography (Figure 34, A), which promotes the separation of the complex from free D1 according to their size differences. The chromatogram of the mixture shows a higher peak followed by two smaller peaks. In comparison with the chromatogram obtained after performing

a size exclusion chromatography under the same conditions of trypsin (23.8 kDa) alone, the higher peak produced by the mixture was eluted before trypsin. This result indicates that the peak corresponds to the higher molecular weight trypsin-D1 complex (30.2 kDa). The two smaller peaks, eluted after trypsin, correspond to the free D1 and degradation products.

The fractions containing the peaks of elution were analysed by SDS-PAGE. It is not possible to confirm the presence of a complex by this technique; however, it is able to show the presence or absence of trypsin and boophilin in the same fraction and also the degradation degree or contamination. On SDS polyacrylamide gel the fractions corresponding to the higher peak showed the presence of trypsin, D1 and degradation products of trypsin (Figure 34, B, lanes 25-27), while the smaller peaks showed a light band of trypsin and two close small molecular weight bands. From these two bands, the one with higher molecular weight corresponds to D1 and the other to degradation products. A closer observation of the most concentrated fractions of the higher peak reveals also the presence of this small molecular weight band. It is possible that this small molecular weight band corresponds to cleaved D1. During trypsin-D1 interaction trypsin probably cleaves the reactive-loop of the D1 without releasing the reaction product and thus becoming inhibited. After treating the complex with β-mercaptoethanol and incubating at 95 ºC for 10 minutes the complex dissociates and the disulfide bonds (between Cys21-Cys71 and Cys30-Cys54) that connect the two segments of cleaved D1 are broken. The two fragments produced should be Gln16-Lys31 (1806.0 Da) and Ala32-Ala58 (4589.0 Da). The first fragment is too small and should be undetectable in a SDS polyacrylamide gel. The second fragment has almost the molecular weight of D1 (6377.1 Da) and could correspond to the smaller molecular weight band observed.

Only fractions 25-27 were pooled together in order to prevent the contamination with non-complexed proteins. The pool of these fractions was concentrated and used in crystallization assays (see section below).

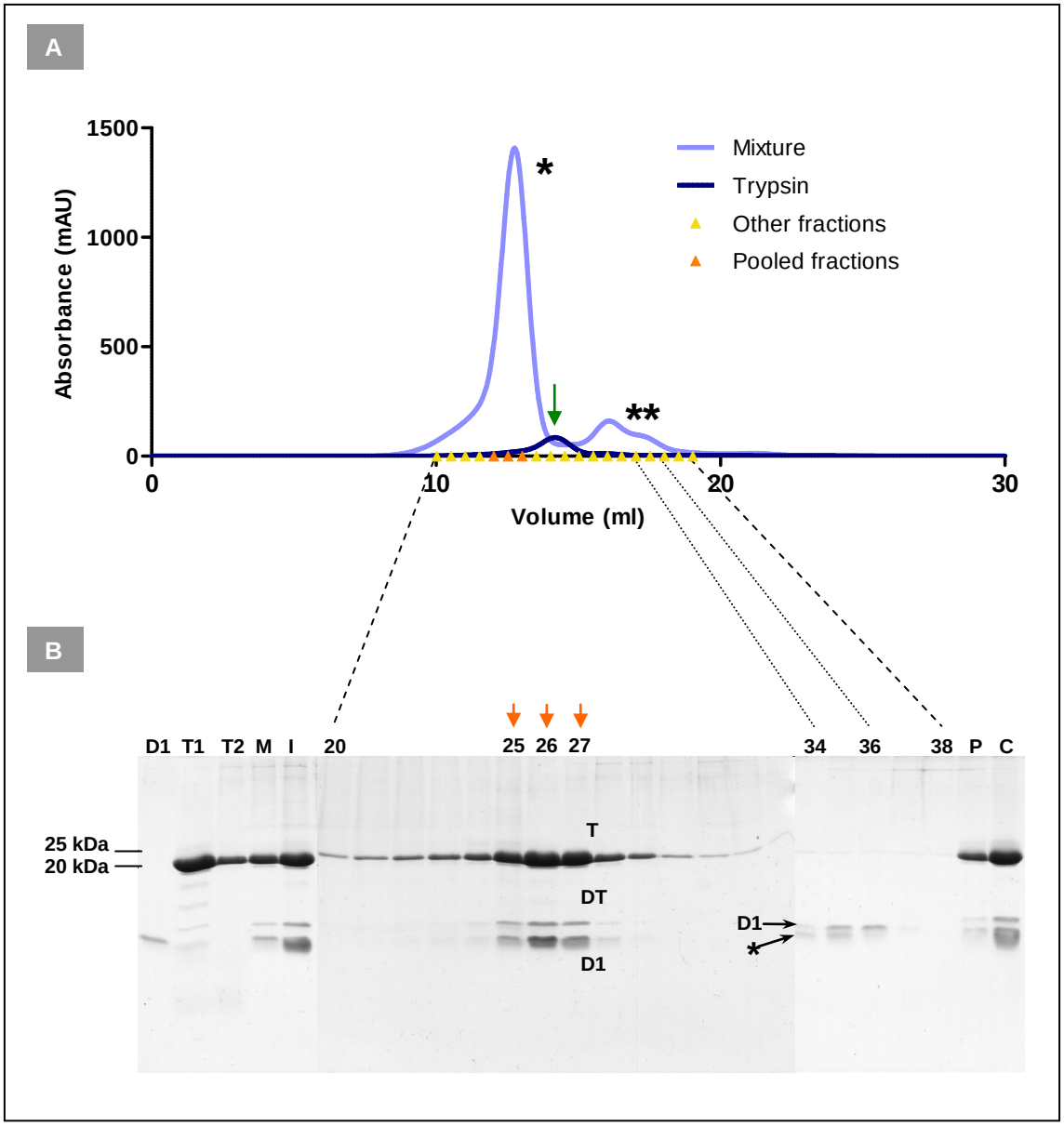


Figure 34 – Purification of 1:1 trypsin-D1 complex by size exclusion chromatography. A – Chromatogram obtained during the purification of the complex from the mixture (Light blue curve). The complex is eluted first (*) compared to trypsin (arrow) followed by the excess of inhibitor (**). The pooled fractions are represented by orange arrows and the other analyzed fractions by yellow arrows. B – Ten microlitres of D1 (D1), 1µl of trypsin (T1), 1µl of trypsin incubated on ice during 12 hours (T2) , 10 µl of the mixture (M), 1 µl of concentrated mixture injected into the column (I), 5 µl of fractions 20-28, 10 µl of fractions 29-38, 3 µl of the pool of fractions 25-27 (P) and 1 µl of concentrated pool (C) were separated in a 17% SDS polyacrylamide gel. Fractions 25-27 show the bands correspondent to trypsin (T), D1 (D1) and products of trypsin degradation (DT). Fractions 34-36 shows two bands, in which one (*) of them has a smaller apparent molecular weight than D1.

3.16. Trypsin-D1 complex crystallization

Crystallization assays were setup with an *in vitro* complex composed by commercial bovine trypsin and purified recombinant D1. Fragile crystal plates were obtained by incubating the complex solution with an equal volume of precipitant. These crystals were observed for the first time 7 days after the incubation with the precipitant solution (Figure 32, A) and their growth was more significant until the 12nd day. Crystals were cryo-cooled in liquid nitrogen at the 80th day (Figure 32, E). These crystals were measured at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). However, they were too fragile, being destroyed by X-ray radiation before a complete data set was obtained.

In order to find a condition in which better crystals could grow new precipitant solutions were prepared by varying the pH and the concentration of precipitant. However, no crystals were obtained in these new precipitant solutions. It was possible to reproduce the crystals by incubating the trypsin-D1 complex solution with the precipitant solution of the commercial kit. However, no crystals were obtained in a solution with a similar composition prepared in our laboratory, even using the same batch of complex solution. The water used to prepare the precipitant solution could be the reason why we are not able to obtain crystals with our precipitant solution. In order to confirm this possibility, new solutions should be prepared using the same water used in the commercial precipitant.

Besides the optimization of the condition where the crystals were found, new conditions should be tested in order to find other that also promote crystallization of the complex. These new conditions involve not only new precipitant solutions (new crystallization screens and additive screens), but also different crystallization methods (e.g. hanging drop) and even different temperatures (e. g. 4 ºC and room temperature).

Moreover, the complex solution used in the crystallization assays had contaminants (see Figure 34, B) that could interfere with the crystallization of the complex. Thus, an additional purification step could be required.

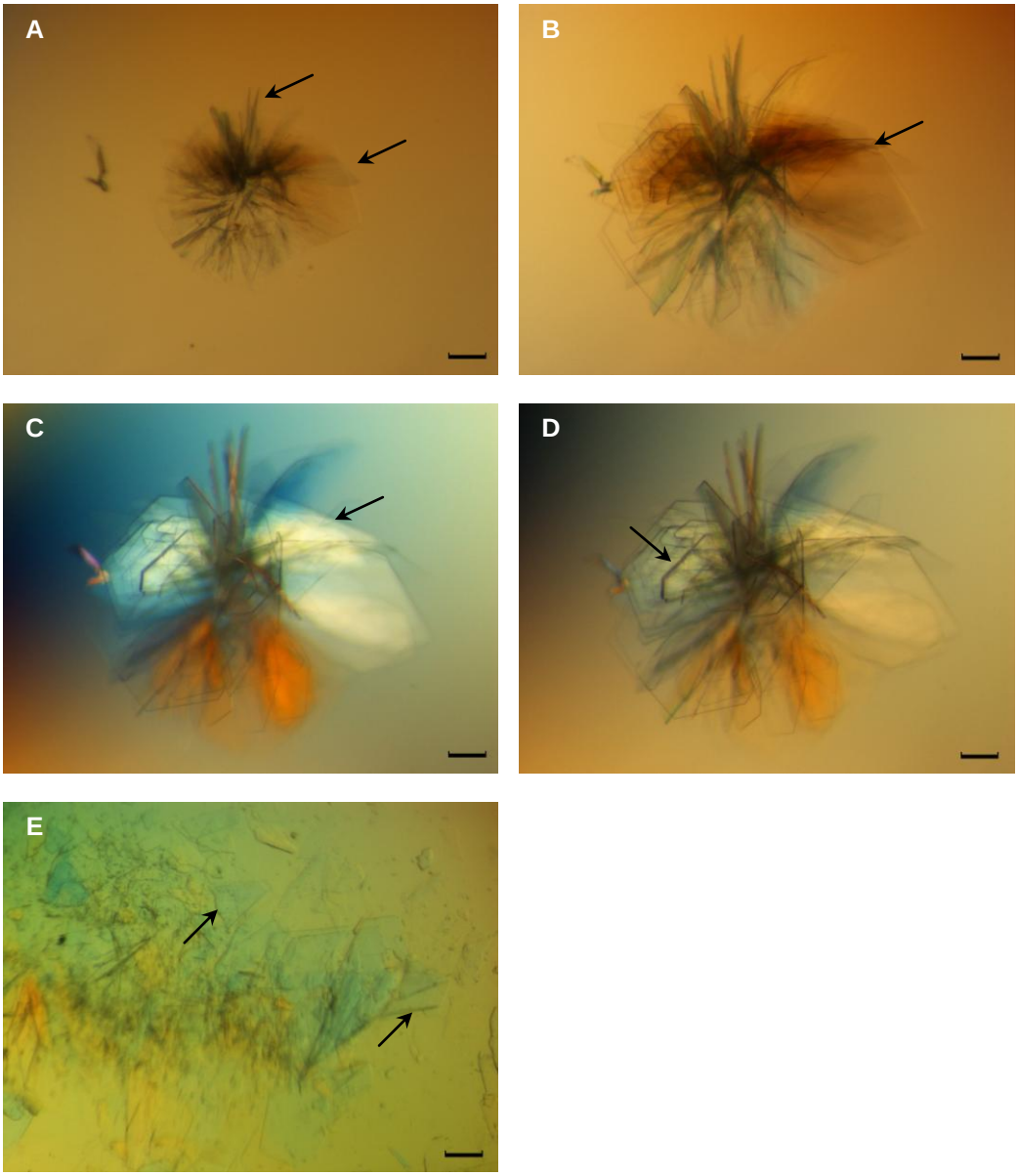


Figure 35 – Crystals obtained by incubating complex solution with the precipitant. Crystal plates (arrows) were first observed on the 7th day (A). The crystals growth was followed (B – 12nd day; C – 30th day; D – 60th day) until the day of their collection (80th day). The crystals were fragile and crush very easily (E - 80th day after collecting some crystals fragments) (Scale bars, 0.2 mm).

4. Conclusion

Pichia pastoris has many of the advantages of higher eukaryotic expression systems such as protein processing, folding, and posttranslational modification. Cell manipulation is almost as easy as in *E. coli* and generally gives higher expression levels. However, the growth and expression times are much longer (Cereghino and Cregg, 2000).

The *Pichia pastoris* expression system proved to be an adequate system for expression of D1D2 and D1. Both recombinant proteins were expressed at high levels and correctly folded, as confirmed by the ability to inhibit trypsin and/or thrombin.

This expression system has also an additional advantage. *Pichia pastoris* secrets only low levels of its own proteins (Cereghino and Cregg, 2000). This feature associated to the secretion of recombinant proteins, as a consequence of their fusion with the α-factor secretion signal from *S. cerevisiae*, is a great first step of purification. The secretion of the recombinant proteins not only avoids all the work associated to cell lysis but also promotes the easy separation of the recombinant proteins from all cellular components. The advantage of this feature was confirmed specially in D1 purification, in which a single affinity chromatography in an immobilized-trypsin column produced highly pure protein. In the case of D1D2, an additional chromatography was required. The D1D2 seems to be unstable and the degradation occurred during protein expression produced a band on a SDS polyacrylamide gel with an apparent molecular weight similar to D1 (approximately 10 kDa). This result suggests that the degradation occurred in the inter-domain linker, producing probably D1 and D2. The presence of D2 was not confirmed but since it has low affinity for trypsin it is probably eliminated during the affinity chromatography step. On the other hand, D1 shows a high affinity for trypsin, being eluted with D1D2 during the affinity chromatography. In order to separate D1D2 from D1 an additional chromatography was performed based on their charge difference. At pH 8.0, D1 and D1D2 showed a significant charge difference that allowed their separation by ion-exchange chromatography using an anion exchanger. At the end of this additional purification step D1D2 was highly pure.

The high purity of D1 and D1D2 was required for the determination of their inhibition constants and crystallization. Despite the inhibition constant determined for D1D2 being two times higher than the inhibition constant determined for D1, the intervals in which they diverge were overlapped. Thus, no significant differences were observed between the inhibitory activity of D1 and D1D2 against trypsin. This finding supports the hypothesis that the second domain of boophilin is not required for trypsin inhibition. However, additional data are required clarify the inhibition mechanism used by boophilin against trypsin, such as X-ray crystallography study of boophilin in complex with trypsin.

The only technique that allows direct visualization of protein structure at the atomic or near atomic level is X-ray crystallography. As the name suggests, for the application of X-ray diffraction, the macromolecule must be first crystallized. However, not only must crystals be grown, but they must be reasonably large, high-quality crystals that are suitable for a high-resolution X-ray diffraction analysis. The crystallization step has emerged as the primary obstacle in X-ray crystallography. There is no comprehensive theory that could guide our efforts to crystallize a particular protein and thus the approaches to protein crystallization remain largely empirical and demand patience and perseverance (McPherson, 2004). It is a matter of searching, as systematically as possible, and wait for success to smile. There are many factors that affect the crystallization of proteins and some of them are summarized in Table 5.

Table 5 – Factors affecting crystallization (McPherson, 2004).

Physical	Chemical	Biochemical
Temperature	pH	Purity
Surface	Precipitant type	Aggregation state of macromolecules
Methodology	Precipitant concentration	Post-translational modifications
Gravity	Ionic strength	Source of macromolecules
Pressure	Supersaturation degree	Proteolysis/hydrolysis
Vibrations	Reductive/oxidative environment	Chemical modifications

Electrostatic/magnetic fields	Macromolecules concentration	Inherent symmetry of the macromolecule
Dielectric properties	Metal ions	Macromolecule stability
Viscosity	Crosslinkers/polyions	Isoelectric point
Equilibration rate	Detergents, surfactants, amphophiles	History of the sample
Homogeneous/heterogeneous macromolecules	Non-macromolecular impurities	Ligands/inhibitors/effectors

One of the factors that affect the crystallization is the heterogeneity of the sample. This heterogeneity can be produced not only by different molecular species but also by different conformations of the same molecule. This means that a flexible molecule is less likely to crystallize than a static one. The trypsin-D1D2 complex would certainly have this problem. During inhibition of trypsin, the second domain of boophilin would be able to move freely, producing complex molecules with different symmetries. Because the first domain seems to be the only domain of boophilin to interact with trypsin, the flexibility issue could be avoided by using the trypsin-D1 instead of the trypsin-D1D2 complex to study the inhibition of trypsin by boophilin. However, for this to be possible D1 must have a conformation similar to the first domain of D1D2. The crystallographic structure obtained for D1 shows no significant differences compared to the first domain of D1D2, and thus, the use of trypsin-D1 complex to study the inhibition of trypsin by boophilin is validated.

The trypsin-D1 complex was incubated with about 720 precipitant solutions, including commercial crystallizations screens and optimization solutions. Some crystals were obtained but most of them were of very bad quality, totally unsuitable for X-ray analysis. The best crystals obtained were those presented on this thesis, but even those were too sensitive to the X-ray radiation.

As showed above, the number of factors that affect crystallization is extremely high and a very small variation in a parameter (e.g. temperature, protein concentration, precipitant pH) could be the difference between have or does not have crystals. Therefore, the number of conditions that can be used in order to find one able to produce high-quality crystals is huge.

In the case of trypsin-D1 complex, beside the use of new precipitant solutions that combine different salts, crystallizing agents and pH (and also different concentrations of each), other conditions such as temperature (4ºC) and crystallization method (hanging drop) could also be used. However, because so many precipitant solutions were used with no success, the main problem could be the protein complex itself. The trypsin-D1 complex could also be flexible or have symmetry that disfavour the perfect fit between the molecules, giving rise to the fragile crystals obtained. This problem could force the use of a different approach to study the inhibition mechanism of trypsin and boophilin. A ternary complex composed by trypsin-boophilin-thrombin could be more stable compared to trypsin-D1 complex and may be easier to crystallize. Another problem that could be affecting the crystallization of trypsin-D1 complex is the presence of contaminants in the complex solution. As confirmed by SDS-PAGE, the complex purified by size-exclusion chromatography showed bands of trypsin degradation. In spite of some proteins being able to crystallize from very heterogeneous mixtures, in general, however, the likelihood of success in crystal growth is greatly advanced by increased homogeneity. Therefore, another approach to produce better crystals is to use a more homogeneous complex solution that could be obtained by an additional purification step.

.

5. Bibliography

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Ajjan, R., and Grant, P.J. (2006). Coagulation and atherothrombotic disease. Atherosclerosis *186*, 240-259.

Alban, S. (2008). Pharmacological strategies for inhibition of thrombin activity. Curr Pharm Des *14*, 1152-1175.

Ansell, J. (2007). Factor Xa or thrombin: is factor Xa a better target? J Thromb Haemost *5*, 60-64.

Bafunno, V., and Margaglione, M. (2010). Genetic basis of thrombosis. Clin Chem Lab Med *48*, S1-S11.

Barker, S.C., and Murrell, A. (2002). Phylogeny, evolution and historical zoogeography of ticks: a review of recent progress. Exp Appl Acarol *28*, 55-68.

Becattini, C., Lignani, A., and Agnelli, G. (2010). New anticoagulants for the prevention of venous thromboembolism. Drug Des Devel Ther *4*, 49-60.

Berkner, K.L. (2001). Blood clotting: general pathway (John Wiley & Sons, Ltd).

Bode, W. (2005). The structure of thrombin, a chameleon-like proteinase. J Thromb Haemost *3*, 2379-2388.

Bouma, B.N., and Mosnier, L.O. (2003). Thrombin activatable fibrinolysis inhibitor (TAFI) at the interface between coagulation and fibrinolysis. Pathophysiol Haemost Thromb *33*, 375-381.

Broze, G.J.J. (1995). Tissue factor pathway inhibitor and the revised theory of coagulation. Annu Rev Med *46*, 103-112.

Cereghino, J.L., and Cregg, J.M. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol Rev *24*, 45-66.

Chen, L., Manithody, C., Yang, L., and Rezaie, A.R. (2004). Zymogenic and enzymatic properties of the 70-80 loop mutants of factor X/Xa. Protein Sci *13*, 431-442.

Ciprandi, A., de Oliveira, S.K., Masuda, A., Horn, F., and Termignoni, C. (2006). *Boophilus microplus*: Its saliva contains microphilin, a small thrombin inhibitor. Exp Parasitol *114*, 40-46.

Collaborative Computational Project No. 4 (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr *50*, 760-763.

Cooke, R.A., and Stewart, B. (2004). Nerves system. In Colour atlas of anatomical pathology, C. Livingstone, ed. (United Kingdom).

Corral-Rodríguez, M.A., Macedo-Ribeiro, S., Pereira, P.J.B., and Fuentes-Prior, P. (2009). Tick-derived Kunitz-type inhibitors as antihemostatic factors. Insect Biochem Mol Biol *39*, 579-595.

Cotran, R.S., Kumar, V., and Collins, T. (2000). Distúrbios hemodinâmicos, trombose e choque. In Robbins Patologia Estrutural e Funcional, G. Koogan, ed. (Rio de Janeiro).

Damjanov, I. (2000). Fluid and hemodynamic disorders. In Pathology for the Health-Related Professions, W.B.S. Company, ed. (Philadelphia).

Eikelboom, J.W., and Hirsh, J. (2007). Combined antiplatelet and anticoagulant therapy: clinical benefits and risks. J Thromb Haemost *5*, 255-263.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.

Evans, P. (2006). Scalling and assessment of data quality. Acta Crystallogr D Biol Crystallogr *62*, 72-82.

Friedrich, T., Kröger, B., Bialojan, S., Lemaire, H.G., Höffken, H.W., Reuschenbach, P., Otte, M., and Dodt, J. (1993). A Kazal-type inhibitor with thrombin specificity from *Rhodnius prolixus*. J Biol Chem *268*, 16216-16222.

Fuentes-Prior, P., Noeske-Jungblut, C., Donner, P., Schleuning, W.-D., Huber, R., and Bode, W. (1997). Structure of the thrombin complex with triabin, a lipocalin-like exosite-binding inhibitor derived from a triatominebug. Proc Natl Acad Sci USA *94*, 11845-11850.

Furie, B., and Furie, B.C. (1988). The molecular basis of blood coagulation. Cell *53*, 505-518.

Greinacher, A., and Warkentin, T.E. (2008). The direct thrombin inhibitor hirudin. Thromb Haemost *99*, 819-829.

Gross, P.L., and Weitz, J.I. (2008). New anticoagulants for treatment of venous thromboembolism. Arterioscler Thromb Vasc Biol *28*, 380-386.

Grütter, M.G., Priestle, J.P., Rahuel, J., Grossenbacher, H., Bode, W., Hofsteenge, J., and Stone, S.R. (1990). Crystal structure of the thrombin-hirudin complex: a novel mode of serine protease inhibition. EMBO J *9*, 2361-2365.

Horn, F., Coutinho dos Santos, P., and Termignoni, C. (2000). Boophilus microplus anticoagulant protein: An antithrombin inhibitor isolated from the cattle tick saliva. Arch Biochem Biophys *384*, 68-73.

Jesty, J. (2001). Blood coagulation (John Wiley & Sons, Ltd).

Koh, C.Y., and Kini, R.M. (2009). Molecular diversity of anticoagulants from haematophagous animals. J Thromb Haemost *102*, 437-453.

Krupiczojc, M.A., Scotton, C.J., and Chambers, R.C. (2008). Coagulation signalling following tissue injury: Focus on the role of factor Xa. Intern J Biochem Cell Biol *40*, 1228-1237.

Macedo-Ribeiro, S., Almeida, C., Calisto, B.M., Friedrich, T., Mentele, R., Stürzebecher, J., Fuentes-Prior, P., and Pereira, P.J.B. (2008). Isolation, cloning and structural characterisation of boophilin, a multifunctional kunitz-type proteinase inhibitor from the cattle tick. PLoS ONE *3*, e1624.

Manithody, C., Yang, L., and Rezaie, A.R. (2002). Role of basic residues of the autolysis loop in the catalytic function of factor Xa. Biochemistry *41*, 6780-6788.

Mann, K.G., Butenas, S., and Brummel, K. (2003). The dynamics of thrombin formation. Arterioscler Thromb Vasc Biol *23*, 17-25.

Markwardt, F. (1970). Hirudin as an inhibitor of thrombin. Methods in Enzymology *19*, 924-932.

Markwardt, F. (2003). Historical perspective of the development of thrombin inhibitors. Pathophysiol Haemost Thromb *32(suppl 3)*, 15-22.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

McPherson, A. (2004). Introduction to protein crystallization. Methods *34*, 254-265.

Motulsky, H. (2008). GraphPad Prism, GraphPad Software Inc, ed. (San Diego, CA, USA).

Noeske-Jungblut, C., Haendler, B., Donner, P., Alagon, A., Possani, L., and Schleuning, W.-D. (1995). Triabin, a highly potent exosite inhibitor of thrombin. J Biol Chem *270*, 28629-28634.

Osinbowale, O., Ali, L., and Chi, Y. (2010). Venous thromboembolism: a clinical review. Postgrad Med *122*, 54-65.

Parizi, L.F., Pohl, C., Masuda, A., and Junior, I. (2009). New approaches toward anti-*Rhipicephalus (Boophilus) microplus* tick vaccine. Rev Bras Parasitol Vet *18*, 1-7.

Rawlings, N.D., Barret, A.J., and Bateman, A. (2010). MEROPS: the peptidase database. Nucleic Acids Res *38*, D227-D233.

Ricci, C., Pinto, A., Berger, M., and Termignoni, C. (2007). A thrombin inhibitor from the gut of *Boophilus microplus* ticks. Exp Appl Acarol *42*, 291-300.

Richardson, J.L., Kroger, B., Hoeffken, W., Sadler, J.E., Pereira, P., Huber, R., Bode, W., and Fuentes-Prior, P. (2000). Crystal structure of the human [alpha]-thrombin-haemadin complex: an exosite II-binding inhibitor. EMBO J *19*, 5650-5660.

Riede, U.-N., and Werner, M. (2004). Localized circulatory disorders. In Color atlas of pathology: pathologic principles, associated disease and sequelae, Thieme, ed. (New York).

Rydel, T.J., Ravichandran, K.G., Tulinsky, A., Bode, W., Huber, R., Roitsch, C., and Fenton, J.W.n. (1990). The structure of a complex of recombinant hirudin and human alpha-thrombin. Science *249*, 277-280.

Stone, S.R., and Hofsteenge, J. (1986). Kinetics of the inhibition of thrombin by hirudin. Biochemistry *25*, 4622-4628.

Stubbs, M.T., and Bode, W. (1993). A player of many parts: the spotlight falls on thrombin's structure. Thromb Research *69*, 1-58.

Thiagarajan, P., and Narayanan, A.S. (2001). Thrombin (John Wiley & Sons, Ltd).

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4673-4680.

Urata, J., Shojo, H., and Kaneko, Y. (2003). Inhibition mechanisms of hematophagous invertebrate compounds acting on the host blood coagulation and platelet aggregation pathways. Biochimie *85*, 493-500.

Van de Locht, A., Lamba, D., Bauer, M., Huber, R., Friedrich, T., Kröger, B., Höffken, W., and Bode, W. (1995). Two heads are better than one: crystal structure of the insect derived double domain Kazal inhibitor rhodniin in complex with thrombin. EMBO J *14*, 5149-5157.

Van de Locht, A., Stubbs, M.T., Bode, W., Friedrich, T., Bollschweiler, C., Höffken, W., and Huber, R. (1996). The ornithodorin-thrombin crystal structure, a key to the TAP enigma? EMBO J *15*, 6011-6017.

Wikel, S.K., and Bergman, D. (1997). Tick-host immunology: significant advances and challenging opportunities. Parasitol Today *13*, 383-389.

Appendix

I. Bacterial culture media and solutions

a. Bacterial culture media

i. Luria-Bertani (LB) medium

1% tryptone
1% NaCl
0.5% yeast extract

Dissolve tryptone, NaCl and yeast extract in deionised water. For LB agar plates, add 15 g of agar per litre of medium. Sterilize by autoclaving.

ii. Super optimal broth medium (SOC medium)

0.5% yeast extract
2% tryptone
10 mM NaCl
2,5 mM KCl
10 mM $MgCl_2$
10 mM $MgSO_4$
20 mM glucose

Dissolve yeast extract, tryptone, NaCl, KCl, $MgCl_2$ and $MgSO_4$ in deionised water and sterilize by autoclaving. Let the solution cool down and add glucose. Sterilize the final solution by passing it through a 0.2 µm filter.

iii. Low Salt Luria-Bertani (Low Salt LB) medium

1% tryptone
0.5% NaCl
0.5% yeast extract

Dissolve tryptone, yeast extract and NaCl in deionised water. Adjust pH to 7.5 with 1 M NaOH. For Low Salt LB agar plates, add 15 g of agar per litre of medium. Sterilize by autoclaving.

b. Competent cell solutions

i. TFB I

For a final volume of 200 ml, dissolve in sterile water 0.6 g of potassium acetate, 2.4 g of rubidium chloride, 0.294 g of calcium chloride and 2.0 g of manganese (II) chloride and add 30.8 ml of glycerol. Adjust the pH to 5.8 with 0.2 M acetic acid. Sterilize by filtering the solution through a 0.2 µm pore filter. Store the solution at -20ºC.

ii. TFB II

For a final volume of 100 ml, dissolve 0.24 g of MOPS sodium salt, 1.1 g of calcium chloride, 0.12 g rubidium chloride and add 15.4 ml of glycerol. Adjust the pH to 6.5 with 0.5 M sodium hydroxide. Sterilize by filtering the solution through a 0.2 µm pore filter. Store the solution at -20ºC.

II. Yeast culture media and solutions

a. Yeast culture media

i. Yeast extract peptone dextrose medium (YPD)

1% yeast extract
2% Bacto™ peptone
2% dextrose (glucose)

For a final volume of 1000 ml, dissolve 10 g of yeast extract and 20 g of Bacto™ peptone in 900 ml of deionised water. For YPD agar plates add 20 g of agar. Autoclave for 20 minutes on liquid cycle. Add 100 ml of 10x dextrose (Appendix II, b, v).

ii. Buffered glycerol-complex medium (BMGY)

1% yeast extract
2% Bacto™ peptone
100 mM potassium phosphate pH 6.0
1.34% yeast nitrogen base without amino acids (YNB)
$4x10^{-5}$% biotin
1% glycerol

For a final medium volume of 1000 ml, dissolve 10 g of yeast extract and 20 g of Bacto™ peptone in 700 ml of deionised water. After autoclaving during 20 minutes on liquid cycle, allow it to cool down to room temperature, and then add 100 ml of 1 M potassium phosphate buffer pH 6.0 (Appendix II,

b, ii), 100 ml of 10x YNB (Appendix II, b, iii), 2 ml of 500x biotin (Appendix II, b, iv) and 100 ml of 10x glycerol (Appendix II, b, vi).

iii. Buffered methanol-complex medium (BMMY)

1% yeast extract
2% Bacto™ peptone
100 mM potassium phosphate pH 6.0
1.34% yeast nitrogen base without amino acids (YNB)
$4x10^{-5}$% biotin
0.5% methanol

For a final medium volume of 1000 ml, dissolve 10 g of yeast extract and 20 g of Bacto™ peptone in 700 ml of deionised water. After autoclaving during 20 minutes on liquid cycle, allow it to cool down to room temperature, and then add 100 ml of 1 M potassium phosphate pH 6.0 (Appendix II, b, ii), 100 ml of 10x YNB (Appendix II, b, iii), 2 ml of 500x biotin (Appendix II, b, iv) and 100 ml 10x methanol (Appendix II, b, vii).

b. Solutions

i. Lysis buffer

10 mM Tris-HCl pH 8.0
100 mM NaCl
1 mM EDTA
2% Triton X-100
1% SDS

ii. 1 M Potassium phosphate pH 6.0

For a final volume of 1000 ml, combine 132 ml of 1 M of potassium phosphate dibasic with 868 ml of 1 M of potassium phosphate monobasic and confirm that the pH = 6.0 ± 0.1. If the pH needs to the adjusted, use phosphoric acid or potassium hydroxide. Sterilize by autoclaving.

iii. 10x Yeast nitrogen base without amino acids (YNB)

2% potassium phosphate monobasic
10% ammonium sulphate
1% magnesium sulphate
0.2% sodium chloride
0.2% calcium chloride
$4x10^{-3}$% myo-inositol

For a final volume of 1000 ml, add 20 g of potassium phosphate monobasic, 100 g of ammonium sulfate, 10 g of magnesium sulfate, 2 g of sodium chloride, 2 g of calcium chloride and 40 mg of myo-inositol to deionised water. Heat

the solution in order to dissolve. Sterilize by filtering the solution through a 0.2 µm pore filter.

iv. 500x Biotin (0.02% Biotin)

Dissolve 10 mg of biotin in 50 ml of deionised water and sterilize by filtering through a 0.2 µm pore filter.

v. 10x Dextrose (20% dextrose)

Dissolve 200 g of dextrose (D-glucose) in 1000 ml of sterile water and sterilize by autoclaving.

vi. 10x Glycerol (10% Glycerol)

Mix 100 ml of glycerol with 900 ml of deionised water and sterilize by autoclaving.

vii. 10x Methanol (5% methanol)

Mix 5 ml of methanol with 95 ml of deionised water and sterilize by filtering through a 0.2 µm pore filter.

III. Plasmid

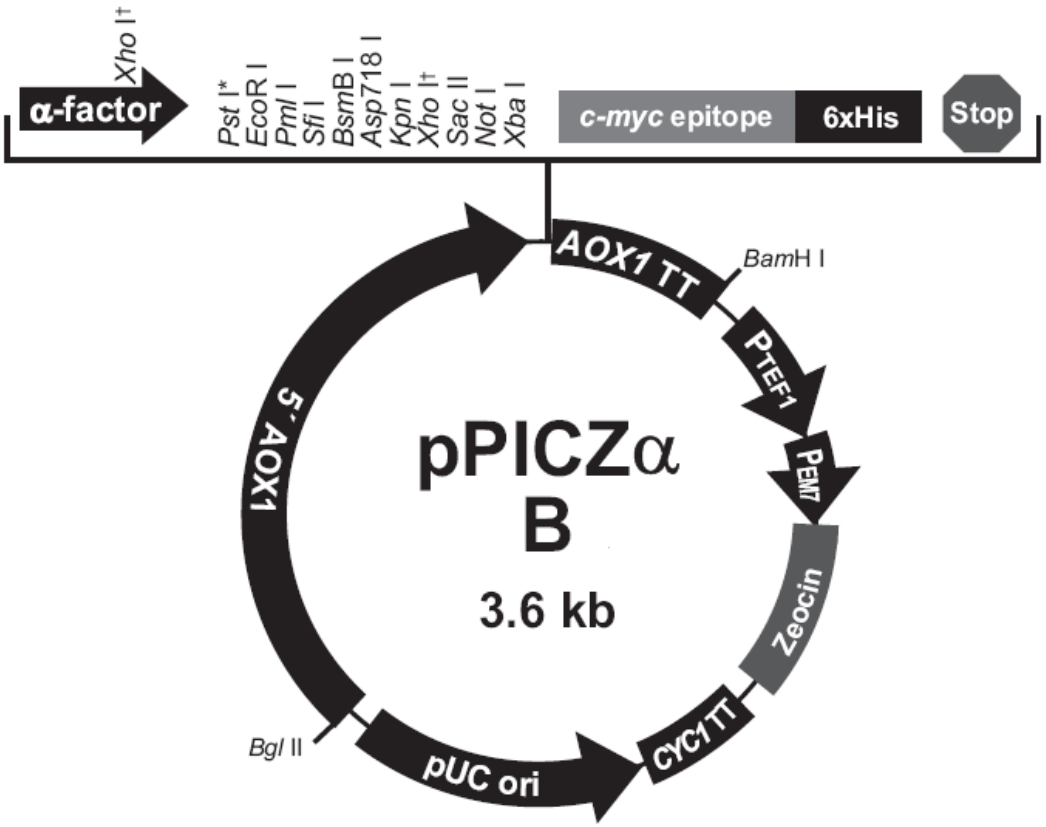


Figure 36 – Map of pPICZαB vector. This vector has the 5' AOX1 promoter, that allows the methanol-inducible, high-level expression of the gene of interest in *P. pastoris*; the α-factor secretion signal (from *S. cerevisiae*), which allows the efficient secretion of most proteins from *P. pastoris* cells; a multiple cloning site with two Xho I sites (†), which allow the cloning of the interest gene in frame with Kex2 cleavage site, resulting in expression of the gene without additional amino acids in the N-terminus; the *c-myc* epitope for the detection of recombinant fusion protein with an anti-*myc* antibody; an C-terminal polyhistidine tag, that allows the purification of the recombinant fusion protein on a metal-chelating resin and also the detection of the recombinant fusion protein with a anti-his antibody; the AOX1 transcription terminator (TT) region that permits efficient 3' mRNA processing for increased mRNA stability; The *TEF1* promoter (from *S. cerevisiae*) which drives the expression of the Zeocin™ resistance gene in *P. pastoris*; the EM7 promoter (synthetic prokaryotic promoter), that drives constitutive expression of the Zeocin™ resistance gene in *E. coli*; the Zeocin™ resistance gene (*Sh ble*) for the selection of transformants in *E.coli* and *P. pastoris*; the *CYC1* transcription termination region (3' end of *S. cerevisiae CYC1* gene), that allows efficient 3' mRNA processing of the Zeocin™ resistance gene for increased stability; and the pUC origin, which allows the replication and maintenance of the plasmid in *E.coli*.