In vitro study of the structure / Function relationship of the proteins GASP-1 and GASP2 : Involvement of the secund kunitz domain in the functional duality of the GASP proteins

Montasir Al Mansi

▶ **To cite this version:**

Montasir Al Mansi. In vitro study of the structure / Function relationship of the proteins GASP-1 and GASP2 : Involvement of the secund kunitz domain in the functional duality of the GASP proteins. Life Sciences [q-bio]. Université de Limoges, 2018. English. NNT : . tel-02786271

HAL Id: tel-02786271
https://hal.inrae.fr/tel-02786271

Submitted on 4 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Thèse de doctorat

UNIVERSITY OF LIMOGES

DOCTORAL SCHOOL 615 - Biological Sciences and Health

PEIRENE EA7500, USC INRA 1061
Génomique AniMale, Amélioration, Adaptation

A Thesis
submitted for the degree of Doctor of Limoges University

Discipline / Specialty: Molecular and Cellular Aspects of Biology

Montasir Al MANSI

14 December 2018

IN VITRO STUDY OF THE STRUCTURE / FUNCTION RELATIONSHIP OF THE PROTEINS GASP-1 AND GASP-2 : Involvement of the secund kunitz domain in the functional duality of the GASP proteins

cosupervised by Pr. **Véronique BLANQUET** and Dr. **Laure BREMAUD**

JURY

President	**A. OULMOUDEN**, Professor, Limoges
Reviewers	**M.P. ELLIES-OURY,** Assistant Professor, INRA, Clermont-Ferrand **J.L. VILOTTE**, Research Director, INRA, Jouy en Josas
Examiners	**F. BRULE-MORABITO**, Assistant Professor, Orléans **L. BREMAUD**, Assistant Professor, Limoges **V. BLANQUET**, Professor, Limoges

Université de Limoges

Dedication

To my father (RIP) and my mother,
To my wife and my daughters,
To my brothers and my sisters,
To my father and my mother in law

ACKNOWLEDGEMENT

This work would not have been possible without the financial support of the French Government through the French Embassy in Jordan. Special thanks to the French Ambassador, ***Mr. David BERTOLOTTI*** *and his dedicated staff:* ***Mrs. Lisbeth NGOUANET****,* ***Dr. Salwa NACOUZI****, and* ***Mrs. Hind MELKAWI****.*

This thesis was carried out at the PEIRENE Laboratory, GAMAA unit, Faculty of Science and Technology, Limoges University, France. I would like to thank the University of Limoges for giving me the opportunity to study and do this research.

I would like to express my sincere gratitude to my advisor ***Professor Véronique BLANQUET*** *for the continuous support of my PhD study and related research, for her patience, motivation, immense knowledge and continuous support. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my PhD study.*

I would like to express my deep gratitude to ***Dr. Laure BRÉMAUD*** *for her patience, encouragement and support during my work in the laboratory through the years that I spent on this work, really I am very grateful for her support.*

My sincere thanks and gratitude go to the members of the jury, especially to ***Dr. Marie-Pierre ELLIES-OURY*** *and* ***Dr. Jean-Luc VILOTTE*** *for accepting to report this thesis. Thanks also to* ***Professor Ahmad OULMOUDEN*** *for agreeing to judge this work and to chair the jury, and* ***Dr. BRULE-MORABITO*** *who has agreed to judge this work as an examiner.*

I would like to express my sincere gratitude and my thanks and appreciation to ***Dr. Didier DELOURME****, for his invaluable support and guidance during all stages of my study.*

I cannot forget ***Dr. Patrick PELISSIER*** *for his invaluable help with all the necessary analyses. He has always done everything possible to help me, I really do appreciate it.*

I would like to express my gratitude for ***Professor Tariq AL-NAJJAR*** *for his continuous support and help during all the time.*

Very special thanks to ***Dr. Khalid ABU KHADRA****, for the countless, valuable and interesting scientific conversations, with my best wishes for him in his career.*

I would like to express my sincere appreciation to my friend ***Dr. Saleem AL-MOGHRABI*** *for his encouragement and support.*

A very special thanks to my father in law ***Dr. Musa AJLOUNI*** *and* ***Eng. Mohammad AJLOUNI*** *for their encouragement and for all the hours they spent in re-reading and fine-tuning of this thesis.*

Many thanks to the staff of the laboratory and the PhD students with whom I shared unforgettable moments: ***Alexis****,* ***Caroline****,* ***Tiffany****,* ***Jessica****,* ***Melanie*** *and* ***Abdel Majeed****. My thanks are extended also to* ***Eric LAPEYRONNIE*** *who just finished his Master's degree and has started his PhD. I wish him a very fruitful and successful future.*

My heartfelt gratitude is extended to my friends in Limoges, ***Saleck SALEM VALL*** *and* ***Abdel Rahman ALKOUNA,*** *for their support during my study.*

I would like to thank ***my mother****,* ***my brothers*** *and* ***my sisters****; whose love and guidance me in whatever I pursue. I wish to thank my loving and supportive wife,* ***Amal****, and my four wonderful children;* ***Raneem****,* ***Rand Noor*** *and* ***Rayyan****, who provide unending motivation and inspiration.*

ABREVIATIONS

ActRIIA/B : activin Receptor type-IIA/B

Akt : Protein kinase B

Alk : Activin-like kinase

BMP : Bone Morphogenic Protein

bp : base pair

BPTI : Bovine Pancreatic Trypsin Inhibitor

Cdk : Cyclin-dependent kinase

Da : Dalton

DNA : DeoxyriboNucleic Acid

ERK : Extracellular signal-Regulated Kinase

GASP : GDF-Associated Serum Protein

GDF : Growth and Differentiation Factor

GST : Glutathion S-Transferase

FLRG : Follistatin related-gene

FoxO : Forkhead box protein O

FST/FS : Follistatin

IgC2 : Immunoglobulin C2-type

IGF : Insulin-like Growth Factor

IPTG : IsoPropyl β-D-1 ThioGalactopyranoside

JNK : Inhibition constant

K_i : Inhibition constant

K_m : Michaelis constant

LTBP : Latent TGF-β-Binding Protein

MAPK : Mitogen-activated protein Kinase

MMP : Matrix MetalloProtease

MRF : Myogenic Regulatory Factor

MSTN : Myostatin

mTOR : mammalian Target Of Rapamycin

MTS : (3-(4,4-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium)

muRF1 : Muscle RING-Finger protein-1

MW : Molecular Weight

Myf : Myogenic Factor

MyoD : Myoblast Determination Protein 1

NTR : Netrin

OD : Optic Density

PA clan : Proteases of mixed nucleophile, superfamily A

PAGE : PolyAcrylamide Gel Electrophoresis

Pax : Paired box protein

PCR : Polymerase Chain Reaction

PDB ID : Protein Data Bank Identifiant

PI3K: PhosphoInositide 3-Kinase

RCL : Reactive Center Loop

R-SMAD : Receptor-regulated SMAD

sFRP : secreted Frizzled-Related Protein

shRNA : short hairpin RNA

SMAD : Sma and Mad related protein

SVF : Fetal bovine serum

TGF : Transforming Growth Factor

TIMP : Tissue Inhibitor of MetalloProtease

WAP : Whey Acidic Protein

WFIKKN : WAP, Follistatin, Immunoglobulin, kazal, Kunitz, Netrin-containing protein

SUMMARY

LIST OF FIGURES

LIST OF TABLES

FOREWORD

Muscle tissue is one of the four basic tissues found in all animals. It can represent up to 40% of the total body weight. It is a heterogeneous and very complex tissue, essential for a large number of functions ranging from motor skills to breathing, maintenance of posture but also the release of heat resulting from muscle activity. It is also the main protein reservoir of the body and a key player in metabolism through its role to maintain homeostasis. Muscle formation or myogenesis is a fundamental biological process and the last twenty years have been marked by the discovery of many molecular factors that play a role in formation, differentiation and muscle mass control. A loss of muscle mass, also called muscular atrophy, can have many physiological or pathological origins. The consequences are also multiple and can, in the most extreme cases, lead to morbidity. All this process is highly regulated at the molecular level (Buckingham, 2006; Bentzinger et al., 2012). Among all the growth factors playing a local role in muscle, particular interests are brought to myostatin.

As a member of the TGF-β superfamily and formerly known as Growth Differentiation Factor-8 (GDF-8), myostatin has many biological functions. In addition to its role in adipogenesis and osteogenesis control, myostatin is also a negative regulator of proliferation and differentiation of muscle cells. Invalidation of myostatin gene in the Mstn $^{-/-}$ mouse model results in a significant increase in muscle mass due to hyperplasia (increase in number) and hypertrophy (increase in size) of muscle fibers (Kambadur et al., 1997; McPherron et al., 1997). Therefore, it is quite interesting to develop new strategies to block the action of myostatin in order to implement new therapeutic treatments against muscle diseases. In another field, these new strategies could also have a significant agronomic and economic impact. Since years, livestock are selected according to criteria considered economically advantageous, such as their ability to produce a larger quantity of meat. For example, in cattle, the muscular character, characterized by a generalized increase in muscle mass (Ménissier, 1982), has been known for almost two centuries. But it was only after the publication of McPherron's work that it was shown by functional and positional cloning that the culard character results from mutations inducing a loss of function of myostatin (Grobet et al., 1998). Numerous studies to inhibit the action of myostatin have been carried out using neutralizing antibodies, direct screening of the gene or overexpression of endogenous or artificial inhibitors capable of complexing with myostatin. The aim is to prevent the

binding of myostatin to its ActRIIB receptor. The most studied endogenous inhibitors are follistatin, FSTL3, propeptide of myostatin, and GASP-1 and GASP-2 proteins. In recent years, many studies have been published on the inhibition of myostatin by follistatin or its own propeptide. Concerning the GASP proteins, the majority of the studies were conducted on the intrinsic anti-proteasic properties of some of their domains. In addition, it has been shown that GASP-1 is able in vivo to increase muscle mass and strength (Haidet et al., 2008). In vitro studies have subsequently confirmed that overexpression of the protein promotes proliferation and differentiation of muscle cells (Bonala et al., 2012). It is in this context that the work of Professor V. Blanquet's team takes place. Its research concerns the functional in vitro and in vivo characterization of Gasp-1 and Gasp-2 using different strategies based on invalidation or overexpression of these genes in mouse. The team showed that mice overexpressing Gasp-1 (Gasp-1-20 line) present an overall increase in muscle mass due to hypertrophy of the myofibers (Monestier et al., 2012a). This increase results of the inhibition of the canonical SMAD2 / 3 pathway and activation of the AKT pathway (Brun et al., 2012; 2014). However, unlike Mstn $^{-/-}$ mice or mice overexpressing follistatin, the Gasp-1-20 line is characterized by an absence of hyperplasia. Molecular analyzes performed on this line indicate an up-regulation of myostatin expression from the embryonic stage, thus explaining the absence of muscle fiber hyperplasia. This work has also shown that muscle hypertrophy can be explained by an increase in postnatal protein synthesis but not by an increase in the number of satellite cells (Brun et al., 2014). Moreover, no significant loss of fat mass was observed in the transgenic mice. It has also been shown that the mutant mice gained weight with age due to an increase in fat mass associated with ectopic fat accumulation. They develop an adipocyte hypertrophy, hyperglycemia, hyperinsulinemia and a hepatic steatosis. All these symptoms are linked to insulin resistance (Périé et al., 2017).

Unlike GASP-1, the role of GASP-2 protein in muscle development has been less studied. Gasp-2 $^{-/-}$ Knockout mice exhibit muscle atrophy and defects in regeneration of fibers after injury. This phenotype results from the inhibition of myostatin inhibition (Lee and Lee, 2013). In vitro studies on murine C2C12 cells, performed in the laboratory, have shown that GASP-2, like GASP-1, also inhibits the proliferation and differentiation of myoblasts (Périé et al., 2016). In vivo GASP-2 overexpression studies are currently underway and should clarify the role of the protein in muscle development.

Within the team, it seemed interesting to complete all the functional analyzes of the GASP proteins by a study of the Structure / Function relationship. GASP-1 and

GASP-2 are classified among the rare heterotypic compound inhibitors. These proteins are, in particular, characterized by several modules, each having anti-protease characteristics. There are numerous studies in the literature relating the involvement of the WAP, Kazal and Kunitz domains in the inhibition of serine proteases. Similarly, the netrin domain, present in GASP proteins and which is also found in many proteins such as TIMPs (Tissue Inhibitors of Metallo-proteinases) or sFRPs (secreted Frizzled-Related Proteins), is associated to an inhibition of metallo-proteinases (Banyai and Patty, 1999; Iyer et al., 2012; Von Maltzahn et al., 2012). Due to their structural and functional diversity, proteases regulate a large number of molecular mechanisms. They are notably involved in the protein degradation, activation of precursors or cell differentiation. Their action is subject to extremely precise regulation mechanisms involving various inhibitors present in the cell, such as serine protease inhibitors, which play an important role in a very large number of biological functions (Potempa et al., 1994). To date, very little data has been published on the anti-protease capabilities of GASP proteins and their involvement in the maturation of myostatin. Only the second Kunitz domain isolated from human GASP-2 protein has been described as capable to inhibit trypsin (Nagy et al., 2003; Liepinsh et al., 2006).

At first, it seemed important to determine whether, in the native conformation of the entire protein, this second Kunitz domain retains its anti-protease properties. In the same way, we have been able to show that GASP-1, like GASP-2, also has this inhibitory capacity. However, although very structurally close, our work has established that there are differences in the specificity of inhibition between the two proteins. The generation of chimeric proteins in which only the second Kunitz domain has been interchanged has shown that this specificity difference is only due to the second Kunitz domain and not to the molecular environment in each protein. Finally, we proposed a structural model to explain this functional duality.

Experimental results including the cloning of several recombinant native and chimeric GASP proteins, their production in a prokaryotic system, their purification, the functionality tests as well as the different kinetic parameters of anti-tryptic activity are presented in the Results section in the form of a publication that is submitted to PLOS Biology. In this paper, we also propose a structural model to explain the functional duality for the anti-trypsin activity of proteins GASP-1 and GASP-2. Additional experimental data are also presented in this section. The presentation of these results follows a bibliographic synthesis divided into three parts that bring together the main scientific data acquired in the recent years on myostatin, GASP proteins and the protease/anti-protease systems.

BIBLIOGRAPHIC REVIEW

CHAPTER I : MYOSTATIN AND ITS SIGNALING PATHWAY

I.1. Transforming Growth Factor beta (TGF-β) superfamily

The transforming Growth Factor β (TGF-β) superfamily consists of 33 members, including TGFs, Bone Morphogenetic Proteins (BMPs), Growth and Differentiation Factors (GDFs, activin and nodal-related proteins, play fundamental roles in the regulation of various biological processes such as growth, development, or regulation of the immune system. These proteins are also involved in homeostasis during development in multiple adult tissues (Derynck and Akhurst, 2007; Derynck and Miyazono, 2008). All proteins found in their signaling pathways are well conserved during evolution and regulate various cellular functions such as adhesion, proliferation, migration, apoptosis and differentiation (Piek *et al.,* 1999). The members of this superfamily can be divided into two groups based on the receptors and the molecules they activate (Hinck, 2012):

- the subfamily containing most TGF-β (TFGβ-1, 2 and 3) as well as GDFs (GDF-8, 9 and 11), BMP-3 and activin/ nodal bind to receptors activating the pathway SMAD2/3.

- BMP proteins (all BMP except BMP-3) and some GDFs (GDF-1, 3, 5, 6 and 7) bind to SMAD1/5/8-coupled receptors.

I.1.1. Biosynthesis and maturation of TGF-β Proteins

All TGF-β proteins have the same organization with a signal sequence, N-terminal propeptide domain (NH_2) and C-terminal domain (COOH) that gives rise to the mature and active form of TGF-β (Kingsley, 1994). The precursors are dimerized stably by a disulfide bridge at the C-terminal portion. Once the secretory signal peptide has been eliminated, the furin family enzymes cleave the precursor at the level of the Arg-Ser-Arg-Arg sequence and release two fragments: the mature C-terminal peptide and the prodomain **(Figure 1)**. The prodomain or LAP (Latency Associated Peptide) remains bound to the TGF-β mature peptide non-covalently and maintains the active dimer as a latent complex, which prevents it from binding to its receptor **(Figure 1)** (Piek *et al.,* 1999; Thies *et al.,* 2001). The latent TGF-β complex is largely more stable than the active dimer and is found predominantly in the blood. Two types of latent

complex are described in the literature (Piek *et al.,* 1999): the LAP or Small Latent Complex and the Large Latent Complex (LLC) which interact with the extracellular matrix. In Golgi, the non-cleaved TGF-β-associated LAP with furin can interact with a non-covalent glycoprotein of the Latent TGF-β-Binding Proteins family (LTBP) to form LLC. LTBPs allow targeting of the latent complex to the extracellular matrix. Activation of the active TGF-β dimer is mediated by proteolytic cleavage induced by members of the bone morphogenetic protein-1/tolloid (BMP-1/TLD) family of metalloproteinases releasing the mature and active form of TGF-β **(Figure 1)** (Wolfman *et al.,* 2003).

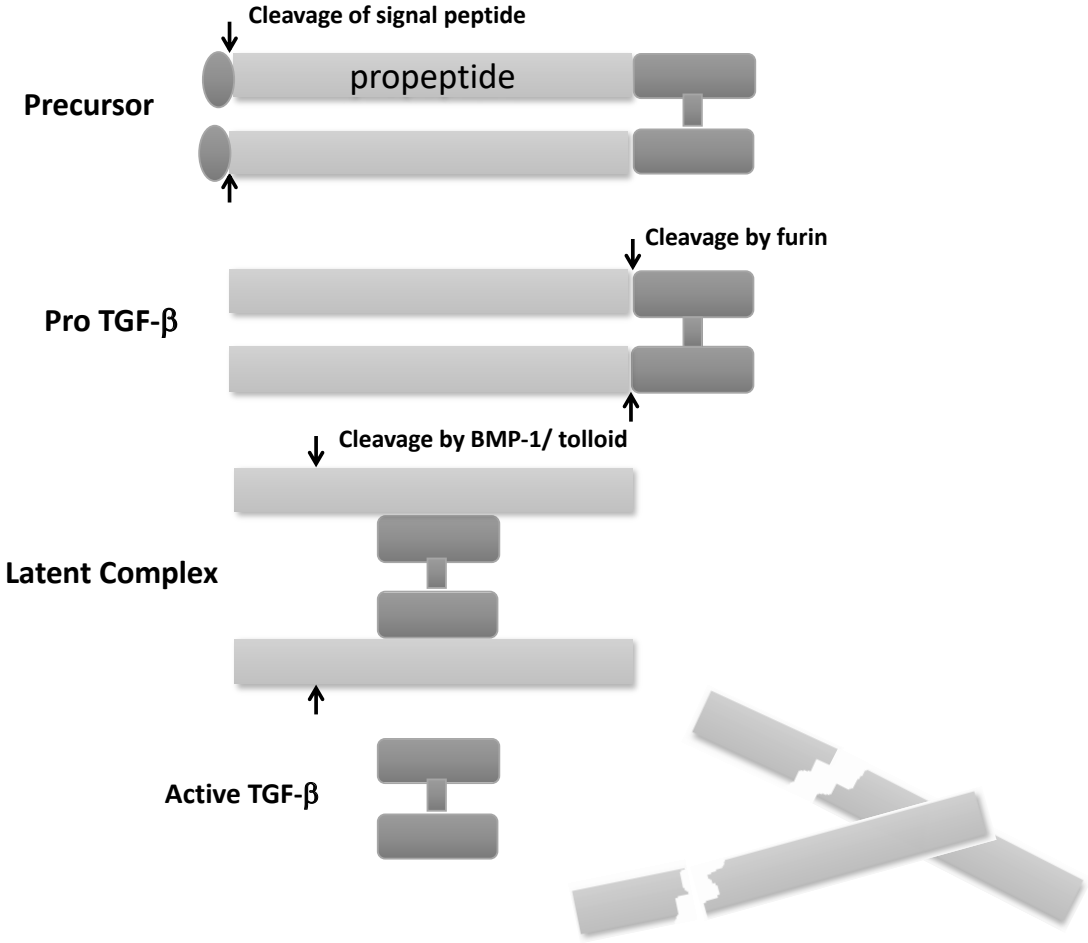


FIGURE 1: BIOSYNTHESIS AND PROCESSING OF MATURE TGF-β

The TGF-β superfamily members are synthesized as a precursor which requires two proteolytic cleavages to generate the active form; the first cleavage allows elimination of the signal peptide and gives the pro-TGF-β. The second cleavage separates the prodomain from the active dimer of TGF-β. The prodomain remains non-covalently bound to the active dimer thus forming the latent complex. A final cleavage performed by members of the BMP-1/tolloid family allows the release of the active form of TGF-β. BMP-1: Bone Morphogenetic Protein-1, TGF-β: Transforming Growth Factor β. (Lee, 2004).

I.1.2. TGF-β signaling pathways

Once activated, TGF-β bind to their receptors. They require two kinds of receptor threonine/serine kinase : type I and type II. There are seven types I (Alk (Activin-like kinase) 1-7) and five type II receptors in the human genome (Schmierer and Hill, 2007). Once TGF-β is bound to its receptor, there is the formation of a heterotetrameric complex in which the type II receptor will phosphorylate and activate the type I receptor (ActRI). This complex once formed will allow the phosphorylation of Receptor-regulated SMAD (R-SMADS): SMAD 1/5/8 for BMPs and GDFs and SMAD2/3 for TGF-βs. **(Figure 2)** (Massague, 1998; Hinck, 2012).

Then SMAD1/5/8 and SMAD2/3 bind to SMAD4 (common mediator-SMAD) to allow translocation to the nucleus and activation or inhibition of the expression of their respective target genes **(Figure 2)** (Heldin *et al.,* 1997; Derynck and Zhang, 2003). Thus, the members of the TGF-β superfamily act along two signaling axes: the SMAD1/5/8 axis and the SMAD2/3 axis in which SMAD4 play a central role (Sartori *et al.,* 2013). These two axes are negatively regulated by SMAD6 and SMAD7 I-SMADs which prevent the phosphorylation of R-SMADs by TGF-β receptors **(Figure 2)** (Massague, 1998).

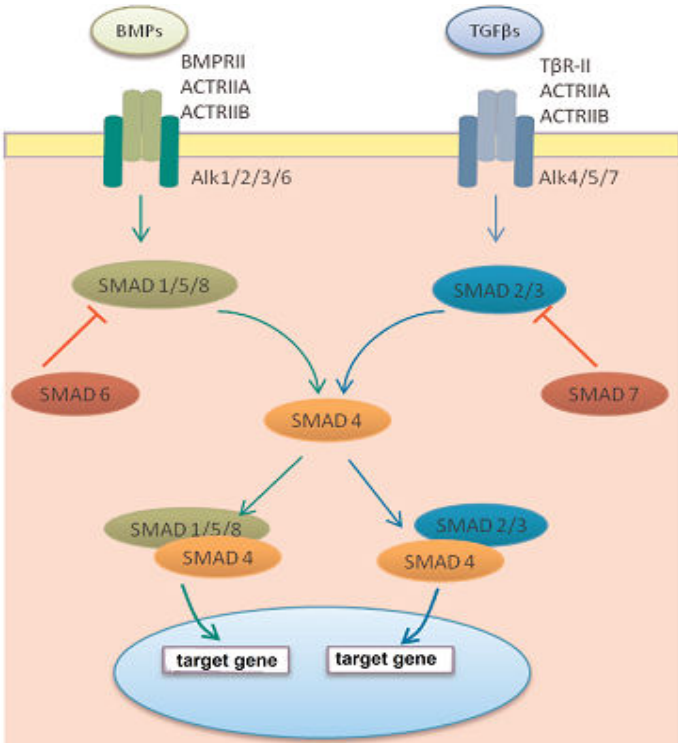


FIGURE 2: CANONICAL SIGNALING PATHWAYS OF BMPs AND TGF-β

Members of the TGF-β superfamily are divided into two groups according to the signaling pathway they activate. The TGF-β subfamily binds to Alk4/5/7 receptors and activates sSMAD2/3. The group of BMPs bind to Alk1/2/3/6 receptors and activate SMAD1/5/8. The co-Smad, SMAD4 can be recruited by the two signaling pathways, playing a central role between the activation of the BMPs and TGF-β pathways. Once SMAD4 is recruited, each signaling pathway can activate or inhibit its target genes. ACTR, Activin Receptor; ALk, Activin-like kinase; BMPR, Bone Morphologic Protein Receptor; SMAD, Sma Mothers Against Decapentaplegic homolog; TβR, TGF-β Receptor; TGF-β, transforming Growth Factor β. (Hink, 2012).

The expression profile and the role of the TGF-β vary during development (Wu and Hill, 2009). For example, TGF-β-1, 2 and 3 inhibit the proliferation of epithelial cells origin but stimulate the growth of mesenchymal cells. BMPs are powerful inducers of bone and cartilage formation and play an important role in development for ventral mesoderm placement, organogenesis and neural tissue differentiation (Chen *et al.*, 2004).

I.2. Myostatin

Myostatin (MSTN), also known as GDF-8, was discovered in 1997 during the search for new members of TGF-β (McPherron *et al.,* 1997). Structurally, myostatin contains all the characteristic features of the TGF-β superfamily: myostatin is

composed of a signal sequence, N-terminal propeptide domain (NH2) and C-terminal domain (COOH) that gives rise to the active ligand. It is one of the rare members of this family whose invalidation is not lethal and which represents the most powerful negative regulators of muscle development. Since then, work on this protein has focused mainly on its effects on skeletal muscle (McPherron *et al.,* 1997).

I.2.1. Structural organization and expression of myostatin

The gene encoding myostatin consists of three exons and is located on chromosome 2 in human and on chromosome 1 in mouse (ENSMUSG00000026100). The gene encodes a transcript of 1131 bases and synthetized a precursor protein of 376 amino acids in mouse **(Figure 3)**. Its expression is detected for the first time in myotoma somites of 9.5 days *post coïtum* mouse embryo. Additionally, the ontogeny of myostatin expression in *pectoralis* muscle coincides roughly with the periods of primary and secondary muscle fiber formation of embryos (Kocamis and Killefer, 2002). In adult, it is mainly expressed in skeletal muscles and heart. A very weak expression can be also detected in adipose tissue, epidermis, mammary glands, neuronal cells of the olfactory cortex (McPherron *et al.,* 1997; Sharma *et al*., 1999; Iwasaki *et al*., 2013). Expression of myostatin was detected in cardiomyocytes and purkinje fibers of cattle heart and tubuloalveolar secretory lobules of lactating mammary glands in pigs (Zhang *et al.,* 2012).

DNA and amino acid sequence of the active region has 100% homology among human, chicken, murine, and porcine species, which suggests a highly conserved function (Miar *et al.,* 2014).

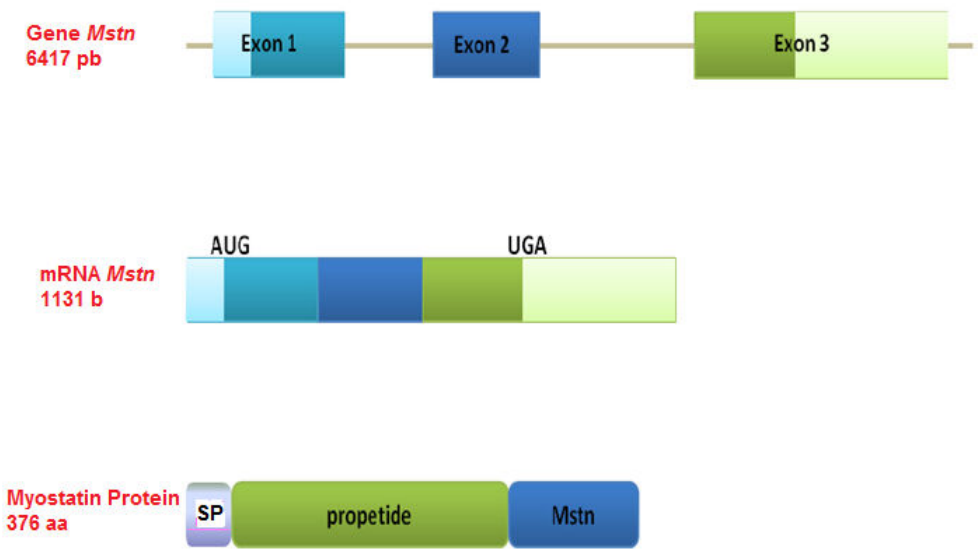


FIGURE 3: STRUCTURE OF GENE, TRANSCRIPT AND PROTEIN OF MYOSTATIN

The myostatin gene consists of three exons. The *Mstn* transcript encodes a protein of 376 amino acids in mice that is composed of a signal peptide, the propeptide and the mature form of myostatin. The light blue and green parts on the gene and the *Mstn* transcript correspond to the non-coding regions. Mstn, myostatin; SP, signal peptide (McPherron *et al.,* 1997).

As shown in Figure 4, myostatin is originally produced as a precursor protein and undergoes proteolytic processing to form the N-terminal prodomain and the biologically active C-terminal disulfide-linked dimer (Ohsawa *et al.,* 2008). The active form of myostatin is identical in human, mouse, rat, chicken, dog, pig and turkey which indicates its physiological importance (Lee, 2004). Unlike in serum where myostatin is mainly found as a latent complex with the propeptide, it is present as the pro-myostatin form bound to LTBP-3 in skeletal muscle (Anderson *et al.,* 2008).

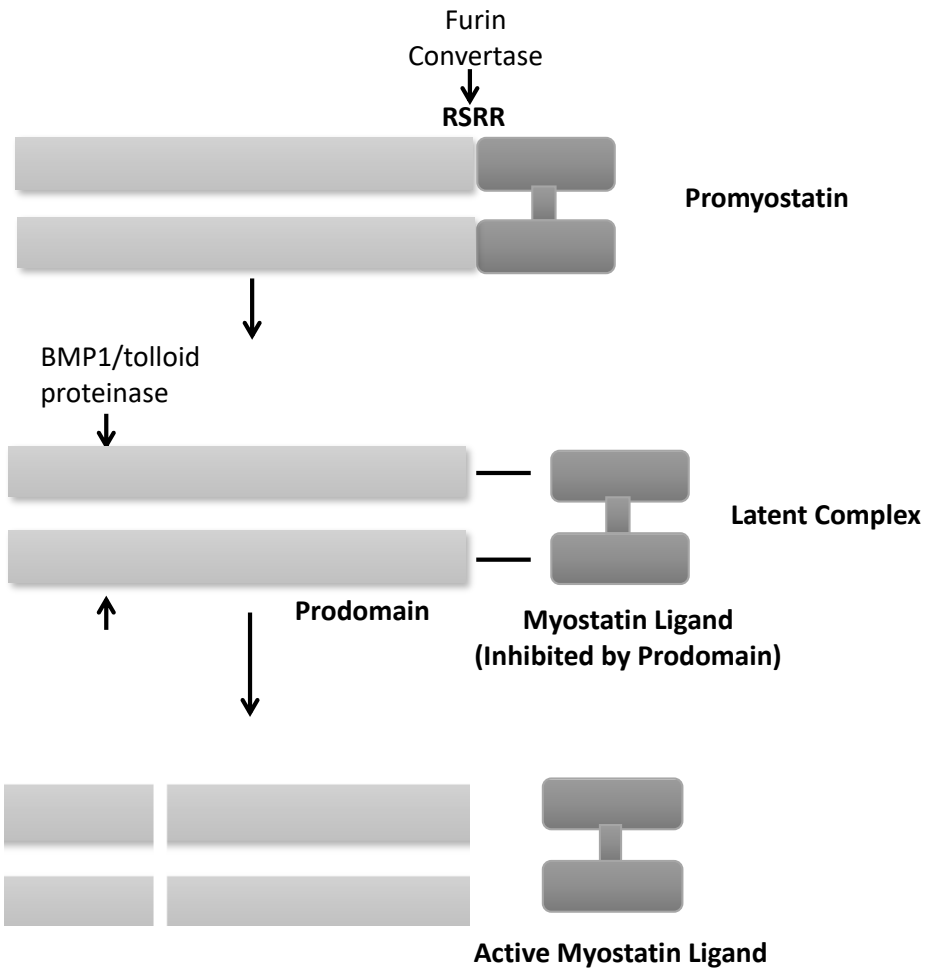


FIGURE 4: MYOSTATIN BIOSYNTHESIS AND MATURATION

Myostatin is synthesized as a precursor which requires two proteolytic cleavages to generate the active form; the first cleavage allows elimination of the secretion signal peptide. The second cleavage carried out by furin-like proteases, separates the propeptide of the active form of myostatin which possesses the binding activity to the receptor. The propeptide bounds to the non-covalently active dimer producing an inactive latent complex. Activation of myostatin occurs following a final proteolytic cleavage by members of the BMP-1/tolloid metalloprotease family, allowing dissociation of the latent complex (Breitbart *et al.,* 2011).

The mature form activates its canonical signaling SMAD pathway. It binds to ActRIIA and ActRIIB, with a better affinity for ActRIIB. These receptors are also targeted by other TGF-β such as GDF-11 (McPherron *et al.,* 1999; Thies *et al.,* 2001; Rebbapragada *et al.,* 2003). The type I receptors Alk4 and Alk5 are then recruited and lead to the activation of SMAD2/3 which once phosphorylated by the receptor will bind to SMAD4. These complexes will then be translocated into the nucleus where they can play their role as transcription factors (McPherron *et al*., 1999; Thies *et al.,* 2001; Langley *et al*., 2002; Rebbapragada *et al.,* 2003). For example, myostatin inhibits the

proliferation of myoblasts cells by regulating the expression of genes involved in the cell cycle (Thomas *et al.,* 2000). The importance of myostatin proteolytic processing is quite clear, as mutation of the myostatin RSRR processing site to the amino acids GLDG leads to prominent skeletal muscle hypertrophy in mice. The removal of the myostatin LAP region from the latent myostatin complex producing mature myostatin activation is one of the main functions of the bone morphogenetic protein-1/tolloid (BMP-1/TLD) family of metalloproteinases (Sharma *et al*., 2015).

I.2.2. Myostatin pathway and its regulations

The myostatin signaling pathway can be divided into a canonical SMAD 2/3 and a non-canonical MAPK pathways (Huang *et al.*, 2011).

I.2.2.1. The SMADs mediated canonical signaling pathway

The canonical myostatin signaling pathway, the SMAD pathway, constitutes a complex transcriptional pathway subject to numerous regulations. Myostatin binds to the receptors ActRIIA or ActRIIB, with a preferential binding to the ActRIIB receptor, leading to the recruitment of the corresponding type I receptor (Alk4/5, activin-like kinase 4/5), which results in the phosphorylation of SMAD2 and SMAD3 (Langley *et al.,* 2002; Zhu *et al.,* 2004). pSMAD2/3 then complex with SMAD4 (common-mediator-SMAD), to allow their translocation into the nucleus and control the expression of target genes (Zhu *et al.,* 2004; Sartori *et al.,* 2009). This intracellular signaling pathway can be inhibited by SMAD7 which prevents phosphorylation of R-SMADs by TGF-β receptors (Massague, 1998) **(Figure 5)**.

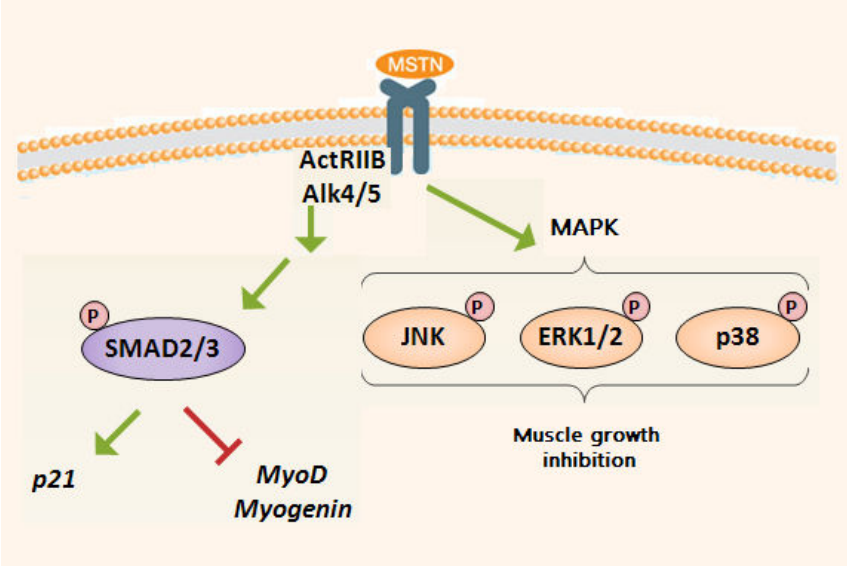


FIGURE 5 : CANONICAL AND NON CANONICAL PATHWAYS OF MYOSTATIN

After binding to its receptor, myostatin activates the canonical pathway of smads, inhibiting myogenic processes. It is also able to inhibit these myogenic processes via the activation of MAPKs (Brun C., PhD thesis 2013).

In addition to its canonical pathway, myostatin can act independently of SMADs via other signaling pathways such as MAPKs signaling pathways as well as the PI3K signaling pathway (Philip *et al.,* 2005; Yang *et al.,* 2006; Huang *et al.,* 2007; Trendelenburg *et al.,* 2009).

I.2.2.2. The non-canonical signaling pathway

Numerous *in vitro* studies have revealed the involvement of MAPK signaling pathways in the inhibition of muscle growth induced by myostatin **(Figure 5)**. Indeed, myostatin is able to activate the p38 MAPK protein by the TGF-β-activated kinase 1 - Mitogen-activated protein kinase kinase 6 (TAK1-MKK6) cascade which results in an inhibition of the proliferation of C2C12 myoblast cells and participates maintaining a basal p21 rate (Philip *et al.,* 2005). Myostatin is also able to activate the c-Jun N-terminal kinase (JNK) signaling pathway by the TAK1-MMK4 cascade in proliferating and differentiating C2C12 cells (Huang *et al.,* 2007). It has also been shown that the ActRIIB receptor is necessary for the activation of this pathway, the sub-expression of the receptor by small interfering RNA (siRNA) leading to a sharp decrease in JNK phosphorylation (Huang *et al.,* 2007). In addition, the use of a JNK-specific inhibitor decreases proliferative myostatin-induced p21 expression, and increases the expression of markers of differentiation (Huang *et al.,* 2007).

The extracellular signal-regulated kinase 1/2 (ERK1/2) pathway also appears to be involved in the myogenic processes (Yang *et al.,* 2006; McFarlane *et al.,* 2008). For example, its binding to the ActRIIB receptor activates the Ras / ERK1/2 signaling pathway, which leads to a decrease in the proliferation and fusion processes of C2C12 myoblasts (Yang *et al.,* 2006).

Myostatin is able to inhibit the IGF-1 activated Akt/mTOR signaling pathway, thereby enhancing its inhibitory capacity of myogenic processes. In addition, the inhibition of mTOR leads to an increase in SMAD2 phosphorylation establishing a true positive feedback loop (Glass, 2010). By inhibiting Akt phosphorylation, myostatin also induces the activation of the FoxO1 protein, which promotes the expression of the MuRF1 and MAFbx atrogens, independently of the NFκB pathway **(Figure 6)** (McFarlane *et al.,* 2006).

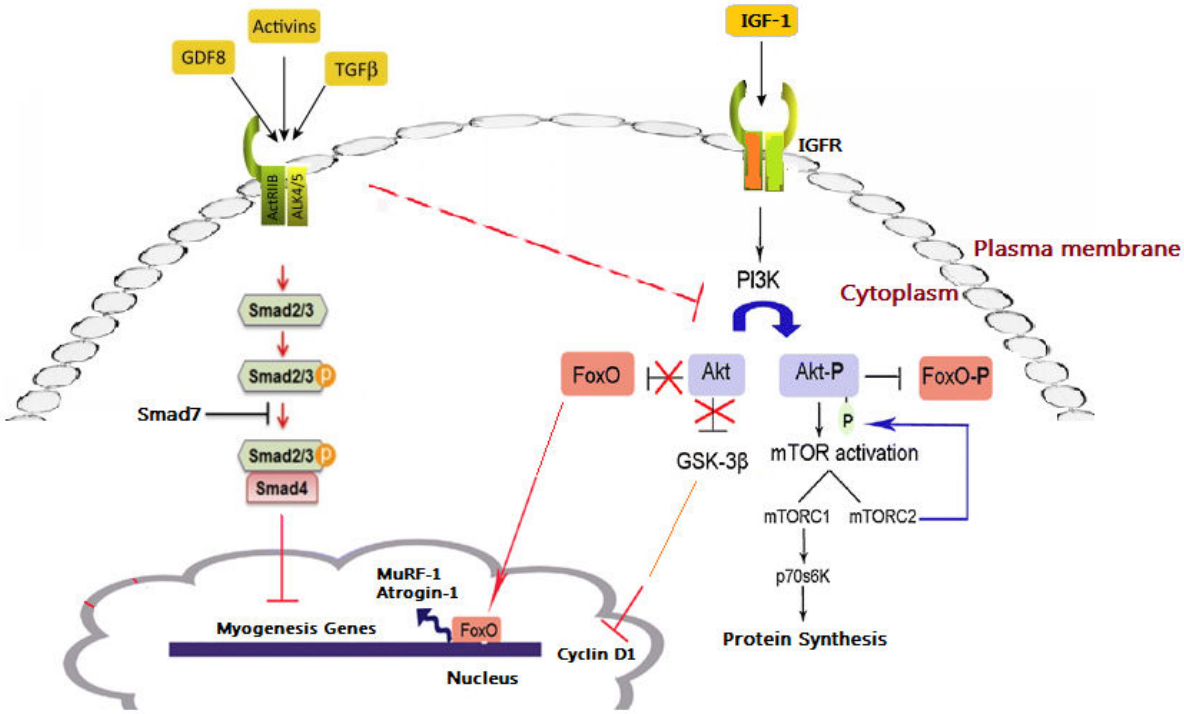


FIGURE 6: MYOSTATIN AND IGF-1 SIGNALING PATHWAYS INTERACTION

Myostatin signal pathway take place by a mechanism similar to that of TGF-β family members, their interaction with ActRII leads to phosphorylation of SMAD2 and SMAD3 receptors, which leads to the decrease of expression of myogenin. Also, Myostatin is able to inhibit the IGF-1 activated Akt/mTOR signaling pathway. By inhibiting Akt phosphorylation, myostatin also induces the activation of the FoxO1 protein, which promotes the expression of atrogenic MuRF1 involved in protein degradation. In addition, myostatin inhibits the PI3K pathway which promotes activation of the FoxO pathway (Rexford and Hyeong, 2015).

I.2.2.3. Regulation of myostatin expression by intracellular factors

The regulatory region of myostatin is conserved in mammals and has a large number of transcription factor binding motifs (Ma *et al.,* 2001). It includes SMAD binding sites demonstrating that myostatin is capable of self-regulation (Allen and Unterman, 2006). It has been shown that Smad3 deficient mice show an increase in myostatin expression suggesting a negative feedback of myostatin expression by this canonical pathway (Ge *et al.,* 2012). Indeed, SMAD7, once activated by SMAD2/3, will inhibit the activation of the myostatin promoter **(Figure 7)** (Allen and Unterman 2006).

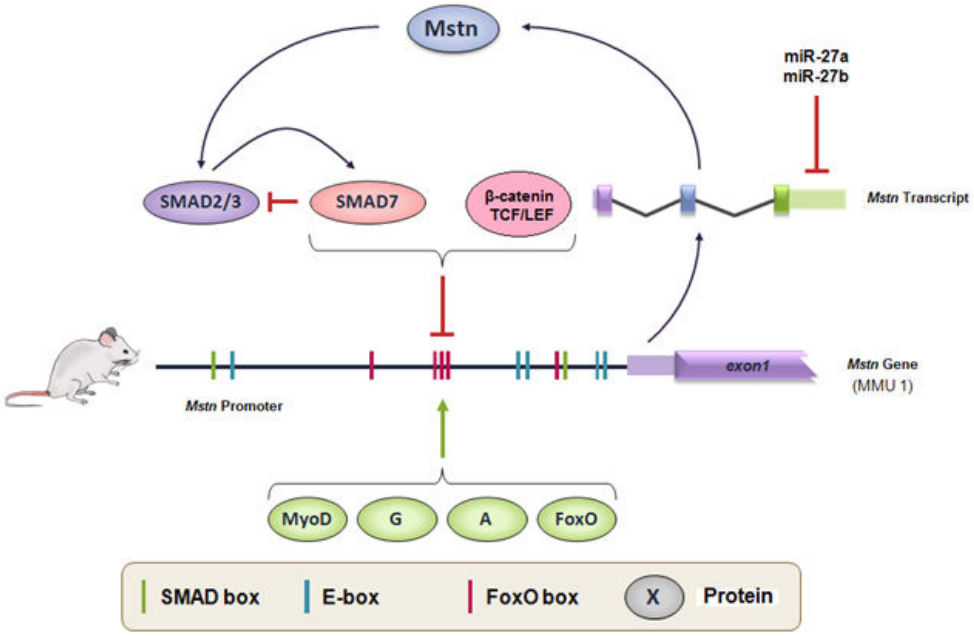


FIGURE 7: REGULATION OF MYOSTATIN EXPRESSION

Myostatin-activated SMAD2/3 can positively regulate the expression of SMAD7. SMAD7 in turn sets on the SMAD box of the myostatin promoter to inhibit its transcription. Myostatin promoter comprises different response elements allowing its activation, in particular "E-boxes" for MRFs, FoxO boxes or elements for glucocorticoid (G) or androgen (A) response. Finally, miR-27a and miR-27b microRNA bind to the 3'untranslated region of the myostatin transcription resulting in its degradation. A, androgenic; FoxO, forkhead box protein O; G, glucocorticoids; miR, microRNA; Mstn, myostatin; MyoD, myoblast determination 1; SMAD, sma mothers against decapentaplegic homolog; TCF / LEF, Tcell factor / Lymphoid enhancer factor (Brun C., PhD thesis 2013).

The regulatory region of myostatin also includes Ebox. Myoblast Determination 1 myogenic factor activates the transcription of myostatin at the level of Ebox 6 and

also at the level of Ebox 5 thus allowing specific activation of myostatin in muscle fibers (Spiller *et al.,* 2002; Salerno *et al.,* 2004). Sites of binding to Myocyte Enhancer Factor-2, Nuclear Factor kappa-light-chain-enhancer of activated B cells, Peroxisome Proliferator-Activated Receptor γ are also found on this regulatory region. Recently, it has been shown that the expression of myostatin is also regulated by miRNAs, such as miR-208a, miR-208b and miR-499 (Callis *et al.,* 2009, Bell *et al.,* 2010; Allen and Loh, 2011). In the same way, certain hormones will bind to the Androgen Response Element (ARE), the Glucocorticoid Response Element (GRE), and the Thyroid Response Element (TRE) to activate the myostatin promoter (Ma *et al.,* 2001; Carneiro *et al.,* 2008; Trendelenburg *et al.,* 2009).

I.2.2.4. Extracellular factors

Myostatin is a powerful negative regulator, subjected to intracellular complex regulation but also to a very fine extracellular regulation, mainly exercised by its inhibitors. Like the propeptide that is able to interact with the active form of myostatin (Thies *et al*., 2001) and thus inhibit its action, other proteins are able to bind and regulate myostatin **(Figure 8)**.

a) Follistatin (FST) was initially discovered for its ability to block the secretion of pituitary-stimulating hormone FSH by inhibiting activin (Ueno *et al.,* 1987; Nakamura *et al.,* 1990). Other studies have shown that it is also able to negatively regulate other members of the TGF-β superfamily, such as myostatin or GDF-11 (Lee and McPherron, 2001; Zimmers *et al*., 2002). FST inhibits myostatin binding to ActRIIB allowing for greater action of main Myogenic Regulatory Factor (MRF) like Pax3 and MyoD (Amthor *et al.,* 2004). Overexpression of FST using a skeletal muscle specific myosin light chain promoter in mice significantly increased muscle mass by 2 to 3-fold through hyperplasia and hypertrophy (Lee and McPherron, 2001). On the other hand, FST-knockout mice have decreased muscle size at birth and many other defects that cause death within a few hours; this indicates that FST controls the activity of numerous TGF-β members (Matzuk *et al.,* 1995). The muscle mass was quadrupled when the fst transgene was existing in MSTN-null mice; this indicated that FST inhibited other regulators of muscle growth that could be similar to MSTN.

A myostatin/follistatin interaction test demonstrates that two follistatin molecules surround the active dimer of myostatin and block all putative sites of interaction between TGF-β and its receptor (Cash *et al.,* 2009). This inhibitor is composed of 3 Follistatin domains (FS) which correspond to a cysteine repeat (Schneyer *et al.,*

1994; Sidis *et al.,* 2001). FS domain proteins are classified into two subfamilies according to their sequence similarity and their ability to bind activin (Schneyer *et al.,* 2001). The first subfamily contains the proteins capable of inhibiting activin, and the second proteins whose association with activin has not been demonstrated.

b) Like follistatin, Follistatin-Like 3 (FSTL3) also called Follistatin-Related Gene (FLRG) has been shown to bind and inhibit the activities of TGF-β family ligands including activin, BMP-2, BMP-6, BMP-7 and myostatin (Kunihiro *et al.,* 2009). FSTL3 was one of the first proteins found to be associated with myostatin in human and murine serum (Hill, 2002; Lee, 2007). The expression of FSTL3 is up regulated by TGF-β and activin signaling through SMAD proteins.

c) the GASP-1 and GASP-2 (GDF-Associated Serum Protein) proteins contain also a follistatin domain and are known to be antagonists of myostatin and GDF-11 but do not inhibit activin (Hill *et al.,* 2003; Kondas *et al.,* 2008; Szlama *et al.,* 2010).

d) extracellular matrix proteins, decorin and laminin, are able to inhibit the mature form of myostatin and its binding to its receptor (Miura *et al.,* 2006). Finally, at the cellular level, human Small Glutamine-rich Tetratricopeptide repeat-containing protein (hSGT) prevents the secretion and maturation of myostatin, and telethonin inhibits its secretion and latent complex formation (Nicholas *et al.,* 2002; Wang *et al.,* 2003).

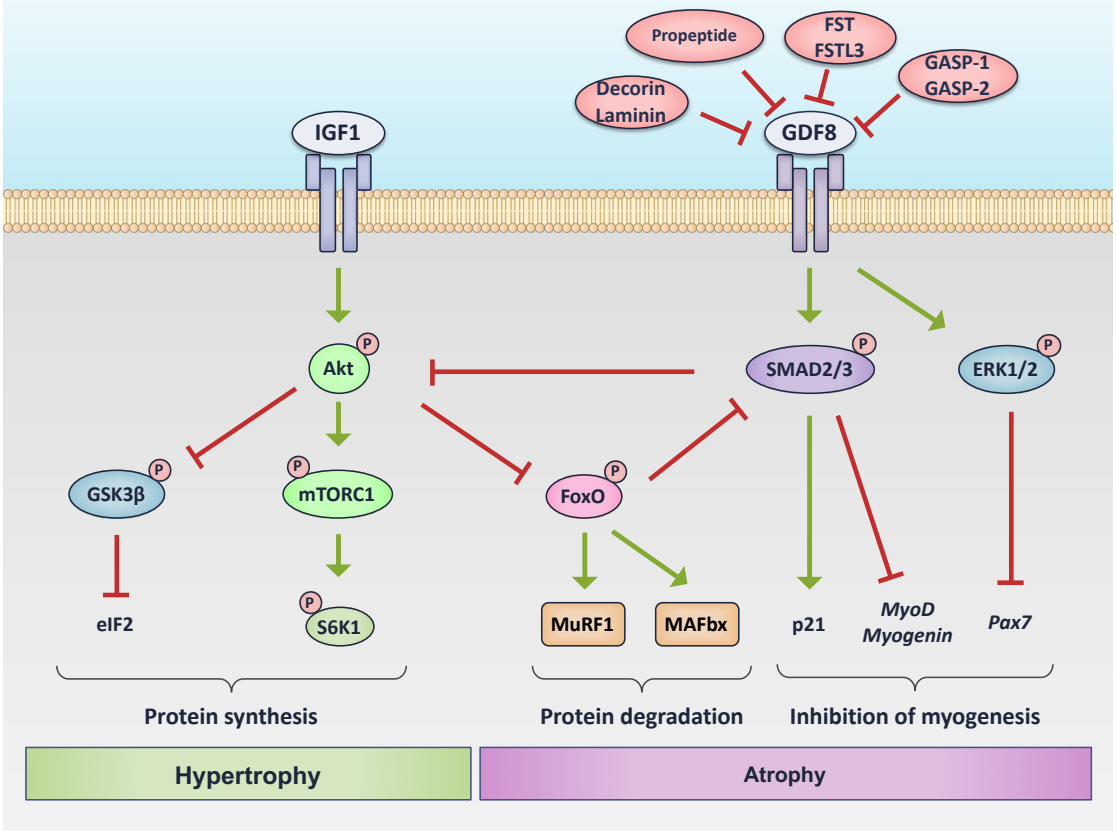


FIGURE 8: REGULATION OF MYOSTATIN ACTIVITY AND MUSCLE GROWTH.

Myostatin is negatively regulated by various naturally-occurring binding proteins. When not bound to these inhibitory proteins, myostatin signals by binding initially to the two activin type II receptors. After binding to its receptor, myostatin activates the canonical pathway of smads, inhibiting myogenic processes. It is also able to inhibit these myogenic processes via the activation of MAPKs (Brun C., PhD thesis 2013).

I.2.3. Roles of Myostatin

I.2.3.1. Myostatin and muscle development

McPherron *et al.* (1997) elucidated the inhibitory role of myostatin on muscle development through the knockout of this gene in mice. *Mstn* $^{-/-}$ mice show a dramatic increase in skeletal muscle mass 2 to 3 times greater than wild animals, resulting both from an increase in the number of fibers (hyperplasia) and their size (hypertrophy) (**Figures 9A and 9B**). This discovery immediately brought out the idea that the "hyper-muscularity" phenotypes observed in other species could be related to a defect in the expression of myostatin. Thus, subsequent studies have made it possible to specify the genotypes of cattle breeds selected for years for this trait (**Figure 9D**). Thereafter, "Loss of function" mutations in other species such as sheep, dogs or even the human,

confirm an extremely conserved role in the species studied (Lee and McPherron, 2001; Schuelke *et al.,* 2004; Shelton and Engvall, 2007). Many studies identified that the excessive muscle growth seen in Belgium Blue cattle was caused by natural mutation in the myostatin-coding gene (Kambadur *et al.,* 1997; McPherron and Lee, 1997). In a similar finding, whippet dogs (**Figures 9E and 9F**) with excessive muscle growth were found to have a heterozygous naturally occurring mutation (Mosher *et al.,* 2007). In addition, Myostatin mutations in Netherlands Texel sheep (**Figure 9C**) have a heavy muscle mass that produces a lean meat (Boman *et al.,* 2009).

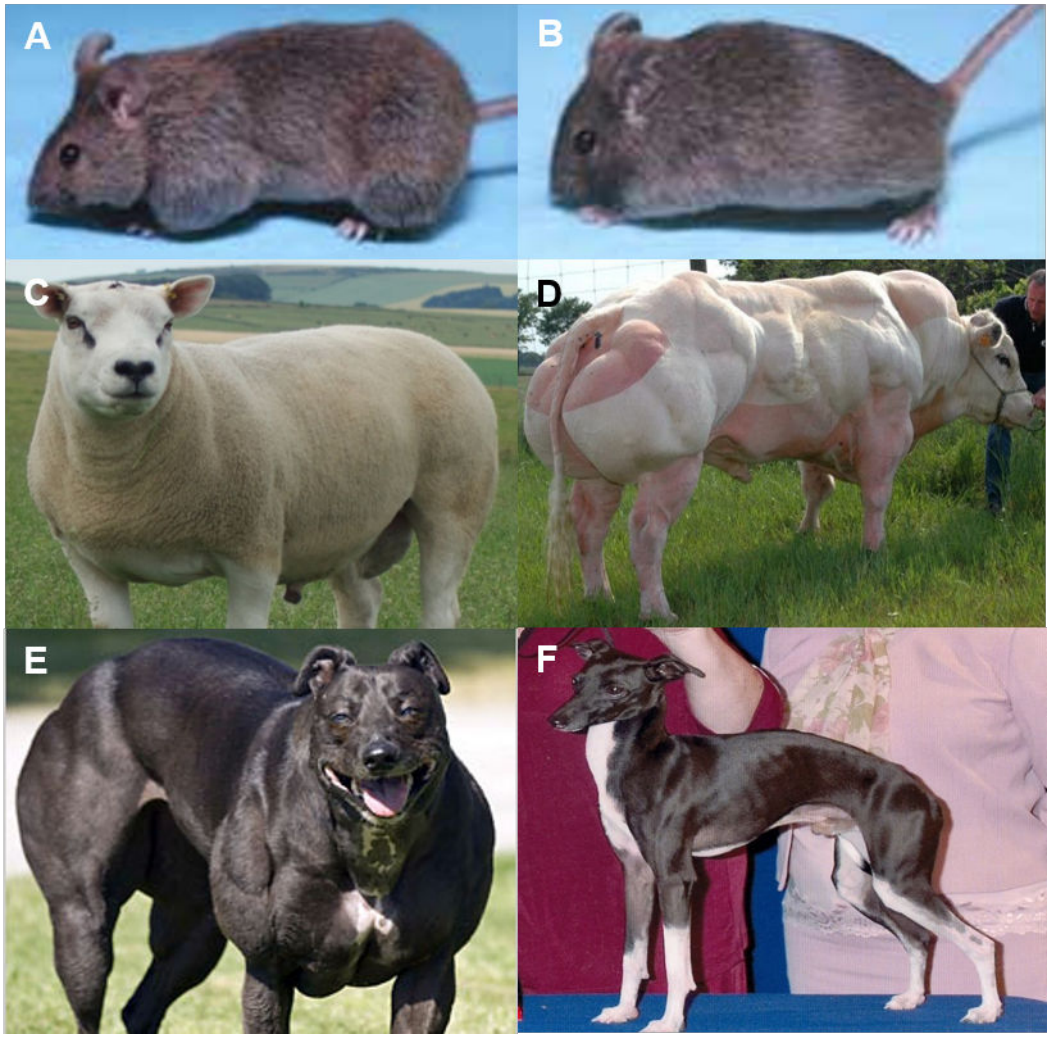


FIGURE 9: INACTIVATION OF MYOSTATIN GENE CAUSES A HYPER MUSCLE PHENOTYPE

(A) Myostatin-null mouse, (B) Wild mouse, (C) Texel sheep, (D) White Belgian Blue, (E) Greyhound; homozygous whippet for a mutation of myostatin, (F) heterozygous whippet for the same mutation (Lee, 2007).

Myostatin plays a role in pigs (Ji *et al.,* 1998), and in fish also (Weber *et al.,* 2005), but appears to demonstrate an evolutionary split in function between mammals and fish (Rodgers and Garikipati, 2008). Zhu *et al.* (2004) suggested that the expression of myostatin depends on the total ratio of R-SMADs (SMAD2 and SMAD3), co-SMADs (SMAD4) and I-SMADs (Inhibitory-SMADs, SMAD7) at a time given that this ratio can be modulated by both myostatin and other TGF-β. In addition, the mutation of cysteine to tyrosine resulting in the production of non-functional myostatin and the hyper-muscular phenotype, deregulation of its expression would have an additive effect on this phenotype (Miretti *et al.,* 2013).

Invalidation of myostatin in mice leads to significant overexpression of microRNAs which contributes to the massive increase in muscle mass by promoting proliferation and differentiation of myoblast. These data also suggest that myostatin can regulate microRNAs by its signaling pathway (Rachagani *et al.,* 2010).

Naturally mutated myostatin gene leads to a hyper muscular phenotype in mice, sheep, human, cattle, dogs, and some other breeds. Therefore, myostatin knockout or inhibition in postnatal life enhances muscle development and increases muscle mass and also play an essential role in regulating adult muscle growth (Camporez *et al.,* 2016).

I.2.3.2. Myostatin and metabolism

Not only does myostatin inhibit the suppression of muscle and skeletal growth, it also has a metabolism function because it improves insulin sensitivity *in vivo* (Chen *et al.,* 2010a). In addition to the positive metabolic effects of any rat with enlarged muscle mass, there is increasing evidence of the metabolic role of myostatin, separated from the role of muscle size regulation (McPherron, 2010). Myostatin is used as a muscle derivative (myokine) to maintain complete body balance (McPherron and Lee, 2002). Myokine as defined by Hjorth *et al.* (2016) is the endocrine hormone produced and released by muscle cells. Myokine regulates processes such as fat degradation, breakdown of glycogen, internal glucose production, and excretion of appetite or hormone in other tissues (Schnyder and Handschin, 2015).

Myostatin regulates glucose metabolism by promoting glucose uptake through the signaling pathway of adenosine monophosphate activated protein kinase (AMPK) in muscle cells (Huang *et al.,* 2011). The deletion of myostatin in the genetic models of obesity and diabetes has improved glucose metabolism, improved insulin sensitivity, and prevented obesity (Wilkes *et al.,* 2009). Bonala *et al.* (2014) showed that myostatin enhances the degradation of insulin receptor substrate -1. Gestational muscle expression of myostatin was almost positively associated with increased glucose use

in response to insulin in older persons (Ryan *et al.,* 2013). Hjorth *et al.* (2016) showed that myostatin increases glucose uptake by a separate signal pathway. In addition, myostatin increases the use of net energy in human myotubes by affecting oxidation and glucose uptake. They also showed that the expression of myostatin gene was associated with adverse effects on insulin sensitivity in fatty tissue and skeletal muscle in humans, but myostatin signals were unlikely to interact directly with the insulin signal pathway. However, this does not exclude the role of myostatin in meeting the energy requirements caused by deflation through other local mechanisms or endocrine systems. Mouisel *et al.* (2014) suggested that myostatin increases the oxidative metabolism of skeletal muscles by means of activated peroxisome proliferator-activated receptors (PPAR). The higher oxidative phosphorylation activity and lower respiratory exchange rate indicate increased burning up of fatty acids as a preferred fuel in the presence of myostatin and suggest higher energy efficiency compared to less effective glucose degradation in myostatin deficiency.

Matsakas (2014) showed that exercise training in myostatin-free mice has significantly reduced the phenotype of massive muscle fibres, improved muscle oxidation characteristics and metabolic gene characterization, increased capillary density and the restoration of power generation capacity deficit without changing the composition of fibre type.

Altogether, Myostatin is increasingly expressed in muscle atrophy and metabolic disorders such as diabetes and obesity, suggesting that changes in the internal expression of myostatin may provide therapeutic benefit to these disorders (Huang *et al.,* 2011; Hittel *et al.,* 2009). The blocking of the MSTN signal in the muscles appears to be a key factor in promoting improved metabolic patterns including normalization of high-fat feeding rather than any direct hypothetical effects of MSTN itself on the central nervous system (CNS) (McPherron *et al.,* 2013).

CHAPTER II : THE GASP PROTEINS

At the beginning of the 2000's, by bio-informatic analyses, two closely related multi-domain proteins were identified and named WFIKKN1 and WFIKKN2 in reference to their domains WAP (Whey Acidic Protein), Follistatin/Kazal, Immunoglobulin, Kunitz and Netrin-containing protein (Trexler *et al*., 2001; 2002). The proteins with a follistatin domain are classified in two subfamilies, ranging the GASPs protein to the subfamily whose members are able to bound the TGF-β (Schneyer *et al*., 2001). With the exception of the Immunoglobulin domain, all other domains are associated with protease inhibitors. This feature suggests that these proteins act as multivalent inhibitors of serine proteases and metalloproteases. In 2003, a major advance in the understanding of the function(s) of these proteins was made when WFIKKN2 was described as an associated protein found in serum that could specifically inhibit myostatin (Hill *et al.,* 2003). Because of this caracteristic (function), WFIKKN2 has been renamed GASP-1 for GDF-Associated Serum Protein-1 (GASP-1). Consecutively, WFIKKN1 was called GASP-2. Since 2003, many studies confirmed that the GASP proteins act as myostatin and GDF11 inhibitors (Kondas *et al.,* 2008; Szlàma *et al*., 2010), these proteins are implicated in muscle and skeletal tissues growth (Aoki *et al.,* 2009). However, only weak effects on muscle tissue were observed after overexpression or inactivation of *Gasp-1* and *Gasp-2* in mice (Lee and Lee*,* 2013). However, the role of this gene family in other tissues is not well understood despite potential action in the brain (Barua *et al.,* 2014), inner ear (Nishida *et al.,* 2004) and reproduction (Harris *et al.,* 2014). Both of the two proteins form asymmetric and symmetric complexes with myostatin, respectively (Walker *et al.,* 2015). Due to their predicted protein domain composition and their ubiquitous expression, these proteins are assumed to form a family of multivalent protease inhibitors (Monestier and Blanquet, 2016).

II.1. Evolution and structure of the *GASP* genes

In mammals, the *GASP* family is composed of two genes of 3-7 kbp structured in two exons and one intron. The first short exon (~ 570 bp) encodes the signal peptide for secretion and the WAP domain whereas the last long exon encodes the follistatin/Kazal domain, the IgC2 domain, the two Kunitz domains and the NTR domain (Monestier and Blanquet*,* 2016).

GASP-1 and *GASP-2* are paralogous genes as supported by paralogons and blocks of synteny (Monestier *et al.,* 2012a). These genes are present in chicken and mammals in single copy, while one copy of *GASP-2* and two copies of *GASP-1* are present in some fish species. A unique ancestral gene was found in invertebrate species. Based on phylogenetic data, Monestier *et al.* (2012a) previously hypothesized that the sequence of the GASP ancestral gene evolved from a sequence close to that found in the cnidarian and *Nematostella vectensis* included in the vertebrate GASP proteins. A comparison of the structural organization of the two genes in human and mouse is given in table 1.

TABLE 1: COMPARISON OF THE *GASP* GENES IN HUMAN AND MOUSE

	Chromosome	Gene size (kb)	Transcript size (b)	Exon number	Protein size (aa)
Human *GASP-1*	17	6.94	3588	2	576
Mouse *Gasp-1*	11	6.75	3516	2	571
Human *GASP-2*	16	3.19	2018	2	548
Mouse *Gasp-2*	17	3.23	2530	2	552

As seen in figure 10, the *Gasp-1* gene is located on the murine chromosome 11. It has two exons (353 bp and 3163 bp) separated by an intron of 3236 bp. The transcript of 3516 b encodes a protein of 571 amino acids (63.3 kDa). Its paralog, the *Gasp-2* gene, is found on the murine chromosome 17 and is composed of two exons of 442 bp and 1539 bp and one intron of 695 bp. The transcript of 2530 b encodes a protein of 552 amino acids (59.8 kDa) (Trexler *et al.,* 2001).

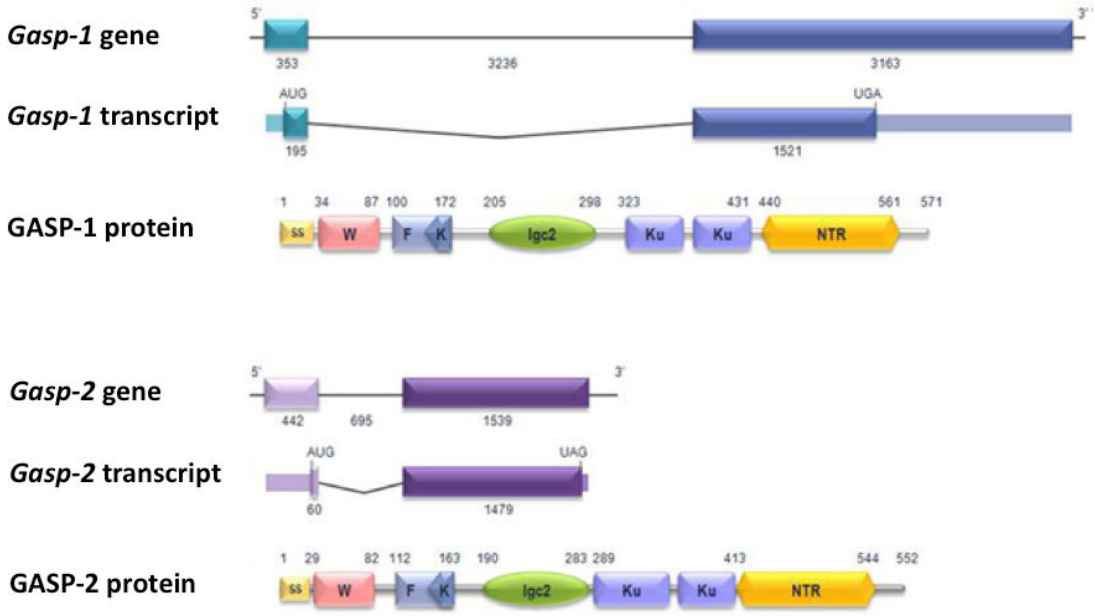


FIGURE 10: STRUCTURE OF GASP-1 AND GASP-2 GENES, TRANSCRIPTS AND PROTEINS IN MOUSE

Gasp-1 and *Gasp-2* genes are composed of two exons and one intron. The transcripts encode proteins of 576 aa for GASP-1 and 542 aa for GASP-2. Proteins are composed of a signal peptide (ss) and several domains: a WAP domain (W), a follistatin domain including a Kazal pattern (F/K), an immunoglobulin domain (IgC2), two Kunitz domains (Ku) and a Netrin domain (NTR) (Brun C., PhD thesis 2013).

II.2. Structural organization and evolution of the GASP proteins

II.2.1. Structural organization in domains of the GASP proteins

As shown in figure 10, GASP-1 and GASP-2 are two large extracellular multidomain proteins consisting of an N-terminal Whey Acidic Protein (WAP) domain, a Follistatin/Kazal (F/K) domain, an Immunoglobulin (IgC2) like domain, two tandem Kunitz (Ku) domains and a C-terminal Netrin-like (NTR) domain. GASP-1 and GASP-2 share only about 55% identity (**Table 2**), but the 6 modules including 38 cysteine residues are well conserved in all vertebrate GASP proteins (Monestier *et al.,* 2012a).

TABLE 2: PERCENT IDENTITY MATRIX FOR HUMAN (h) AND MOUSE(m) GASP PROTEINS (Monestier *et al.,* 2012a)

	hGASP-1	mGASP-1	hGASP-2	mGASP-2
hGASP-1	-	92.82%	55.70%	55.11%
mGASP-1	92.82%	-	55.47%	54.14%
hGASP-2	55.70%	55.47	-	87.77%
mGASP-2	55.11%	54.14%	87.77%	-

The GASP proteins are composed of six domains and five out of these domains exhibit homology-predicted antiprotease activity, so these proteins are classified as heterotypic compound inhibitors (I-90 family of peptidase inhibitors) along with only two other proteins: the Red Sea turtle chelonian and human Epididymal Protease Inhibitor (EPIN) (Rawlings *et al.,* 2004). From the N-terminal end, we distinguish:

- The WAP domain is composed of a characteristic sequence motif of eight cysteines found in a four disulfide core arrangement. The WAP domain is a functional motif found in many proteins that exhibit serine protease inhibition activity (Letunic *et al*., 2015).
- The Kazal-type serine protease inhibitor and the Follistatin-like domain are composed of ten cysteines. The follistatin domain is primarily responsible for the binding with mature growth factors (Kondas *et al*., 2008). In addition, the antagonist follistatin uses multiple domains and two FS molecules to encircle both myostatin and activin ligands (Walker *et al.*, 2015). Further, the role of the C-terminal region of the follistatin domain is similar to the Kazal domain (Hill *et al.,* 2002). The follistatin domain of GASP-1 is also responsible for mediating the interaction with myostatin (Hill *et al.,* 2003). One form of TGF-β family regulation is through inhibition by extracellular antagonists such as the follistatin type proteins (Cash *et al.,* 2012). Kondas *et al.* (2008) showed that myostatin binds to follistatin and the follistatin related protein with high affinity. Therefore, they assumed that the follistatin domains are responsible for these interactions. Moreover, GASP proteins bind mature growth factors firstly via interactions with the follistatin domain, also their NTR domains contribute to the interaction.
- The Immunoglobulin domain (IgC2) was found to contain a cysteine -rich region related to the follistatin/Kazal module family (Trexler *et al*., 2001).

- The two tandem Kunitz domains are usually present in proteins belonging to the MEROPS inhibitor family or the Kunitz/bovine pancreatic trypsin inhibitor family (Laskowski and Kato, 1980). Their size is about 50 residues with few secondary structures and their fold is constrained by three disulphide bonds (Letunic *et al*., 2015). The structure of the second Kunitz domain from human GASP-2 is very similar to that of the Bovine Pancreatic Trypsin Inhibitor (BPTI) (Liepinsh *et al.,* 2006). This domain blocks trypsin and a panel of trypsin-related serine proteases (Nagy *et al.,* 2003; Kondas *et al.,* 2008). Indeed, the second Kunitz domain of GASP-2 showed remarkable specificity for trypsin (Kondas *et al.,* 2011). Monestier *et al*. (2012a) showed the role of the second Kunitz domain in preventing myostatin proteolysis, acting as a functional antiprotease activity. Therefore, the trypsin inhibitory activity of GASP proteins was given to the second Kunitz domain. Nagy *et al.* (2003) reported that WAP and follistatin domains of GASP-1 didn't inhibit the activity of elastase, trypsin and chymotrypsin. They indicated that the second Kunitz-type domain of GASP-2 did not inhibit the activity of elastase, chymotrypsin and various proteases with trypsin. They added that only the second Kunitz domain of the GASP proteins exhibits a antiproteasic activity. Indeed, in GASP-2, this domain seems to have a typical Kunitz fold structure. They concluded that the second Kunitz domain can slightly inhibit the proteolytic action of trypsin (Nagy *et al.,* 2003). Further, Liepinsh *et al.* (2006) showed that this domain has some specificity in the protease- binding loop that is rarely present in other known Kunitz domains. Finally, Kondas *et al.* (2011) also tested the activity of the full-length GASP proteins. The authors indicate that GASP-1 and GASP-2 can inhibit the proteolytic activity of bovine trypsin and have no effect on other bovine proteins like furin or the metalloproteinase BMP1 (for Bone Morphogenetic Protein 1) but no experimental data have been reported.

- The NTR domain is an essential domain that includes six conserved cysteines; this domain may form internal disulphide bonds, and several conserved blocks of hydrophobic residues (Banyai and Patthy, 1999). Information on the role and the inhibitory activity of the NTR domains containing proteins is limited. In contrast, this domain is frequently involved in binding to heparin or related macromolecules (Bekhouche *et al.,* 2010). Kondas *et al.* (2008) revealed that the NTR domain contributes most significantly to the interaction with myostatin propeptide. Due to the supposed fixation function of NTR domain, myostatin inhibition is achieved by protein-protein interaction. Also the NTR domain is responsible for the inhibitory activity of tissue inhibitor of metalloproteinases (TIMPs) (Hill *et al*., 2003). Myostatin and GDF11 are produced from precursor proteins by proteolytic processing (**Figure 14**). Kondas *et al*. (2008) revealed that the propeptide region of myostatin binds to GASP-2

independently from mature myostatin. They revealed that the NTR domain of GASP-2 mediates the interaction with myostatin.

Finally, considering the post translational maturations, GASP-1 is a glycoprotein that contains two N-glycans; one located between the IgC2 and the first Kunitz domains and the second in the NTR domain. The loss of these two N-glycans leads to a decrease in protein secretion rate but does not appear to affect the functional activity of the protein in C2C12 cells (Brun *et al*., 2012). Furthermore, several mucin type *O*-glycans seems to be present between the Kazal and the IgC2 domains.

II.2.2. Evolution of the GASP proteins

Hill *et al.* (2003) reported that the common ancestor of GASP proteins was formed in the chordate lineage. The GASP protein of Ciona intestinal differs from vertebrate proteins because it doesn't have an immunoglobulin domain. Phylogenetic analyses of domains have revealed that the Ciona protein is basal to GASP-1 and GASP-2. After the duplication of the Kunitz domain, the protein is evolved by module/exon shuffling (**Figure 11**). Also the ancestral *GASP* gene probably replicated to *GASP-1* and *GASP-2* in the second whole genome duplication (WGD2) as groups of paralogues parallel to a unique chromosome or contiguousin Ciona or branchiostoma, and GASP-2 sequences only are found in sea lamprey (Monestier *et al*., 2012a). Finally, the structure was completed by addition of the NTR domain at the 3' end. It is known that the NTR domain has appeared several times in convergent evolution and that the NTR domain in GASP is relatively close to the NTR domain of a disintegrin-like and metalloprotease domain with thrombospondin type I motifs-like 5 (ADAMTSL5) and procollagen C-proteinase enhancer (PCoLCE) proteins (Leclere and Rentzsch, 2012). This exon shuffling evolution implies that each domain is encoded by a special exon as observed in the sea urchin/ acorn worm clade. Later in chordates, a progressive elimination of introns occurred as there is only one remaining in mouse and human sequences. Moreover, the presence of at least one GASP sequence in all deuterostomes indicates a major role of these proteins in this group (Monestier and Blanquet, 2016). In mammals, analysis of amino acid substitution rates revealed that the whole GASP-1 protein sequence is more conserved than the GASP-2 (Monestier *et al.,* 2012a).

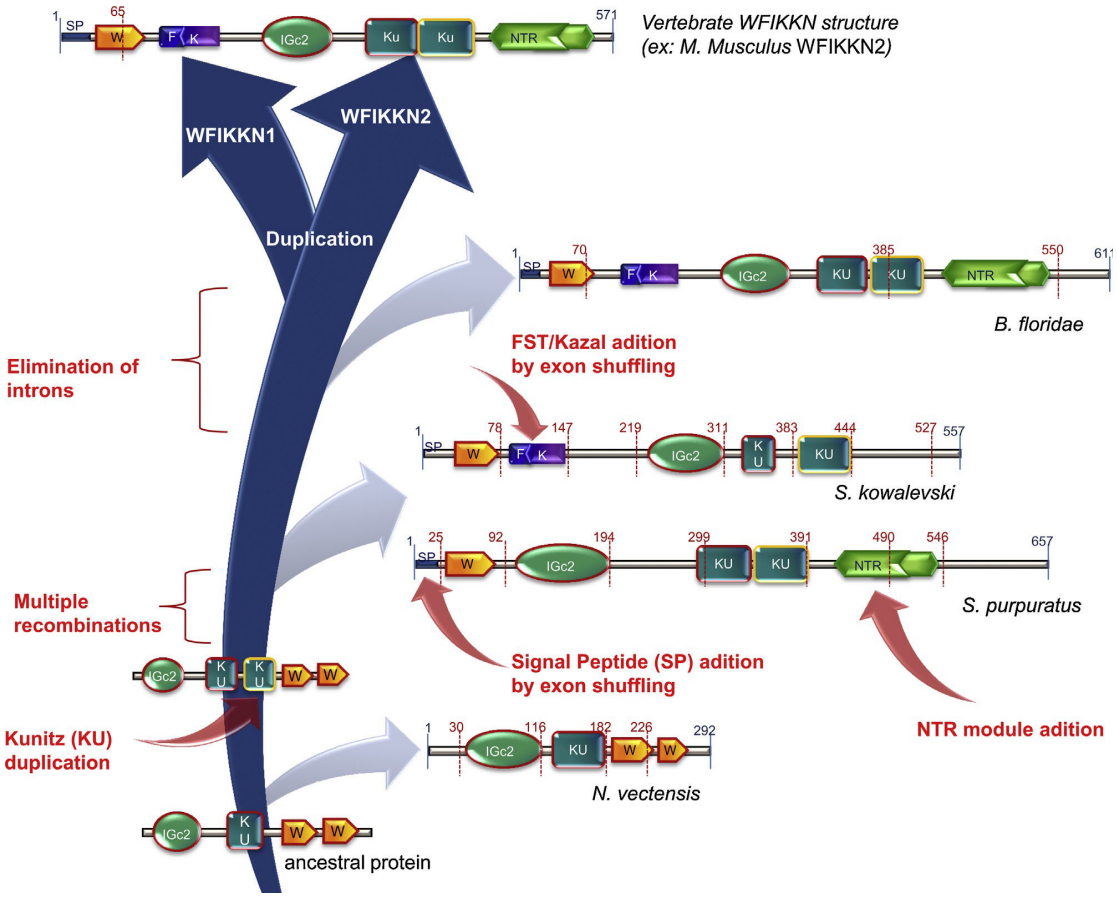


FIGURE 11: EVOLUTION OF THE GASP PROTEINS BY MODULE SHUFFLIN

Domains related to the starlet sea anemone protein shown in red, and the new duplicated Kunitz domain is shown in yellow. After the duplication of the Kunitz domain, the protein is evolved by module/exon shuffling, where the ancestral *GASP* gene probably replicated to *GASP-1* and *GASP-2* in the second whole genome duplication (WGD2) as groups of paralogues parallel to a unique chromosome or contiguous (Monestier *et al*., 2012a).

II.2.3. Tissular expression of the GASP proteins

In order to determine the physiological role(s) of the GASP proteins, the expression patterns of transcripts have been made at different stages of development in human fetal and adult tissues. Studies show that *GASP-1* and *GASP-2* have different profils. In fetal tissues, *GASP-1* is preferentially expressed in the brain, skeletal muscle, kidney and thymus. Its expression is weak in the lung and not detected in the liver. *GASP-2* is essentially expressed in the lung and more weakly in the liver, skeletal muscle and kidney (**Figure 12A**). In adult tissues, *GASP-1* is expressed in the ovary, testis, pancreas, brain and lung. *GASP-2* is expressed in the pancreas, thymus, liver, kidney, lung and testis (**Figure 12B**) (Trexler *et al*., 2001; 2002). In mouse, a similar profil is found for *Gasp-1* (Hill *et al*., 2003).

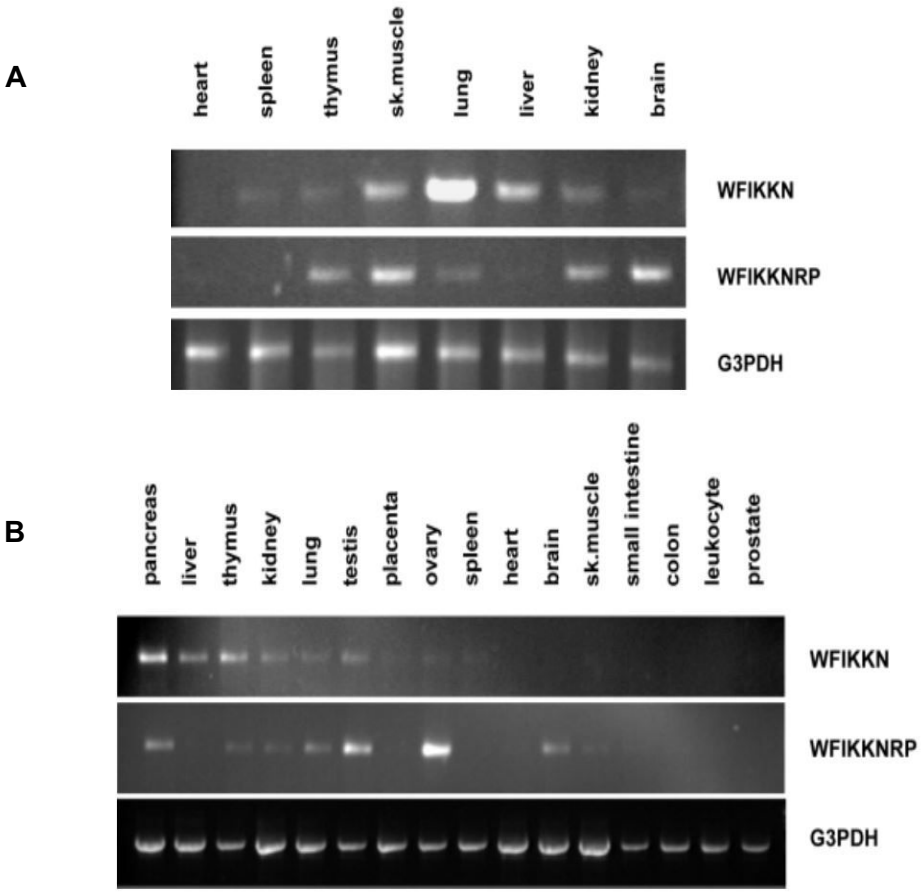


FIGURE 12: **EXPRESSION PATTERN OF THE HUMAN GENES *GASP-1* AND *GASP-2* IN FETAL (A) AND ADULT (B) TISSUES** (Trexler *et al*., 2002)
G3pdh, Glyceraldehyde-3-phosphate dehydrogenase

Since these first studies, the proteins GASP have been identified in large number of organs and tissues. GASP-1 was found in the pigmented neuro-epithelia ciliary body of the eye in humans (Janssen *et al.,* 2012). Regarding the regulation of the expression of the *Gasp* genes, a functional peroxisome proliferator activated receptor beta/gamma (PPARB/G) site in the promoter of murine *Gasp-1* was found to be involved in the regulation of the gene expression (Bonala *et al.,* 2012). The increased expression of Peroxisome proliferator activated receptor gamma (Pparg) in mice overexpressing *Gasp-1* provided some evidence of auto regulation of the gene expression by Pparg (Monestier *et al.,* 2012a). Estrogen could also has a role in *GASP-1* regulation as RNA sequencing revealed an up-regulation of the gene in osteoblasts in response to estrogen receptor alpha mediated estrogen stimulation (Roforth *et al.,* 2014).
Recently, Mille-Hamard *et al.* (2012) reported a decrease in the GASP-1 expression in soleus and tibialis anterior of erythropoietin (EPO) deficient mice. The expression of

GASP-1 was also up-regulated in hyperplasic growth zones of the late trout embryo myotome compared with adult myotomal muscle as shown by Rescan *et al*. (2013) using laser capture micro dissection and Agilent trout microarray. In addition, the concentration of GASP-2 in the human serum also decreased with age in a Duchenne muscular dystrophy patient while it increased with age in controls (Hathout *et al.,* 2015). GASP-1 is one of the 32 most highly enriched proteins in embryonic stem cells that are known to enhance the proliferation of mouse and human muscle progenitors (Yousef *et al.,* 2014). In the mouse brain, Liddelow *et al.* (2012) showed an up-regulation of *Gasp-1* in the embryonic versus adult lateral ventricular choroid plexus. Transcript was confirmed to be increased by 30–50 folds in the choroid plexus (Bowyer *et al.,* 2013); as well as by 15–30 folds in the meninges and associated vasculature (MAV) as compared to the striatum and the parietal cortex (Bowyer *et al.,* 2012).
Further, it was shown that *Gasp-2* sequences exhibit hypomethylated CpG sites in brain myeloma of progeny after high maternal folic acid diet in mice (Barua *et al.,* 2014). An up-regulation of *Gasp-2* was also observed in mouse neural crest cells specifically ablated for the integrin-linked kinase gene (Dai *et al.,* 2013).
In the liver, an up-regulation of *GASP-1* has been reported in response to aflatoxin B1 (Merrick *et al.,* 2013). Interestingly *Gasp-2* is down regulated in neonatal liver of mouse clones produced by somatic cell nuclear transfer or by intracytoplasmic sperm injection (Kohda *et al.,* 2012). In lymphocytes, Wirth *et al.* (2010) showed a down regulation of *GASP-1* in quaternary memory T cells as compared to primary memory T cells. Hughes *et al.* (2011) showed an up-regulation of *GASP-1* in the feather forming mesenchyme in adult chicken compared with that in embryo. Regarding *GASP-2*, it has been reported to be overexpressed in response to various compounds such as polycyclic aromatic hydrocarbons in zebrafish (Goodale *et al.,* 2013), retinoic acid in human embryonic stem cells (Cheong *et al.,* 2010), the caspase inhibitor in mice (Wu *et al.,* 2014) and the gastric inhibitory polypeptide in rats (Maino *et al.,* 2014).

II.3. Roles of the GASP proteins

II.3.1. Interactions with GDF and BMP family

Many studies have been reported that the proteins GASP are capable to bind several members of the TGF-β family. They are usually assigned to three main subfamilies: Activins, TGF-β and BMPs (Bone Morphogenic Proteins) (Walker *et al.*, 2015). In fact, *in vitro* interaction tests have shown that the proteins GASP are capable of interacting with the active dimers of TGF-β and GDF-11 by their follistatin domain,

and with their propeptide by their Netrin domain (Kondas *et al.,* 2008). As a result, the proteins GASP interact with the mature form of these TGF-βs as well as with the latent complex, where found GASP-2 having a better affinity for the latent complex than GASP-1 *in vitro* (Szlàma *et al.,* 2013). Indeed, binding of GASP proteins to the propeptide could alter the action of the members of the BMP-1/tolloid metalloproteinase family responsible for the proteolytic cleavage of the propeptide, and thus maintain myostatin in the form of a latent complex (Hill *et al.,* 2003). GASP-1 and GASP-2 could therefore play the role of TGF-β transporter in order to locate their actions. Moreover, the peculiarity of GASP-1 and GASP-2 compared to the follistatin is they do not bind activin (a protein of the TGF-β family involved in many physiological processes), making them more specific to myostatin (Szlàma *et al.,* 2010) (**Figure 13**).

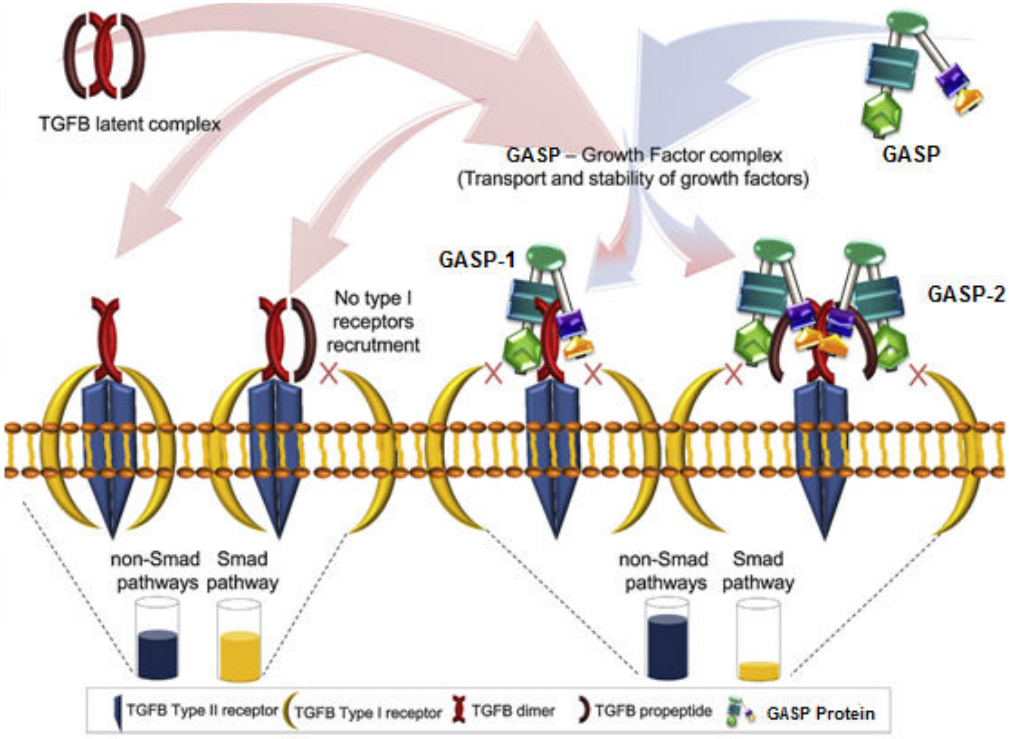


FIGURE 13 : THE INTERACTIONS OF GASP-1 AND GASP-2 WITH TGF-β

GASP proteins act as companion proteins for growth factors and do not allow the recruitment of type I receptors. GASP-1 interacts with mature dimer whereas GASP-2 mostly bindings inactive latent complexes. Therefore, the GASP proteins disrupt the canonical SMAD pathway and consequently favour the non-SMAD pathways. The interaction between TGF-β and GASP-2 might lead to the enrichment of semi-latent complexes (Monestier *et al.,* 2012a).

Szlàma *et al.,* (2010) demonstrated that the GASP proteins have a strong affinity for TGF-β, BMP-2 and BMP-4. However, these different interaction capacities with the TGF-β do not necessarily lead to the inhibition of their signaling pathway. Although these proteins interact with the TGF-β, only myostatin-mediated or GDF-11

mediated signal transduction is inhibited by this interaction (Kondas *et al.,* 2008). László Patthy's team suggested that the proteins GASP have an additional role as a "localizing" protein for TGF-β in serum and thus facilitate their specific actions. GASP-2 protein could modulate and localize the action of BMP-4 during the development of the inner ear, the two proteins being expressed in this organ (Gerlach *et al.,* 2000; Szlàma *et al.,* 2010).

II.3.2. GASPs: Myostatin antagonists

The first link between GASP proteins and myostatin was made in 2003. Hill *et al.* (2003) explained the close relationship between GASP-1 expression and myostatin with muscle development. Direct injections of GASP-1 using an adenovirus into the lower muscles of mice were performed. These studies show that overexpression of GASP-1 in these muscles increases muscle mass and strength. However, this increase is less important than that obtained with follistatin (Haidet *et al*., 2008; Rodino-Klapac *et al*., 2009). Recent studies showed that the GASP proteins could inhibit myostatin in its mature form and in its latent form also. It seemed that the inhibitory effect of GASP-1 to be specific for myostatin as it did not inhibit activin activity like other myostatin inhibitors (Hill *et al.,* 2003). Additionally, they showed an increase in muscle mass consistent with myostatin inhibition. Furthermore, the role of GASP-1 in the inhibition of myostatin *in-vivo* from the ubiquitous overexpression of GASP-1 in mice (SurGasp-1) showed an increase in muscle mass of about 35%; this was caused uniquely by hypertrophy of type I and type II fibers (Monestier *et al.,* 2012b). Overexpression of GASP-1 would promote proliferation and differentiation of myoblasts by inhibiting the myostatin pathway (Brun *et al*., 2014). Interestingly, these mice did not exhibit any modification of the fat mass ratio. This hyper muscular phenotype was only visible 21 days after birth, this seems to be different from the one observed in myostatin-null mice (Zhu *et al.,* 2000). More recently, it has been shown that *Gasp-2* overexpression or knockdown by sh-RNA increase C2C12 proliferation and differentiation of myoblasts by inhibiting the myostatin pathway (Périé *et al*., 2016). The majority expression of GASP-1 in fetal skeletal muscle is overlapping with the almost exclusive expression of Myostatin in this tissue, which supports the hypothesis proposed by Hill *et al.* (2003) of an interaction between the two proteins involved in the regulation of muscle mass. Thus, GASP-1 could not only inhibit the active form of myostatin, but also intervene in regulating the maturation of myostatin by its Netrin domain. Their Netrin domain also would allow the association with its propeptide which preventing the maturation of myostatin following the masking of the cleavage sites of

furin and BMP-1/tolloid enzymes. Their follistatin domain would interact with active myostatin and inhibiting its interaction with ActRIIB (**Figure 14**).

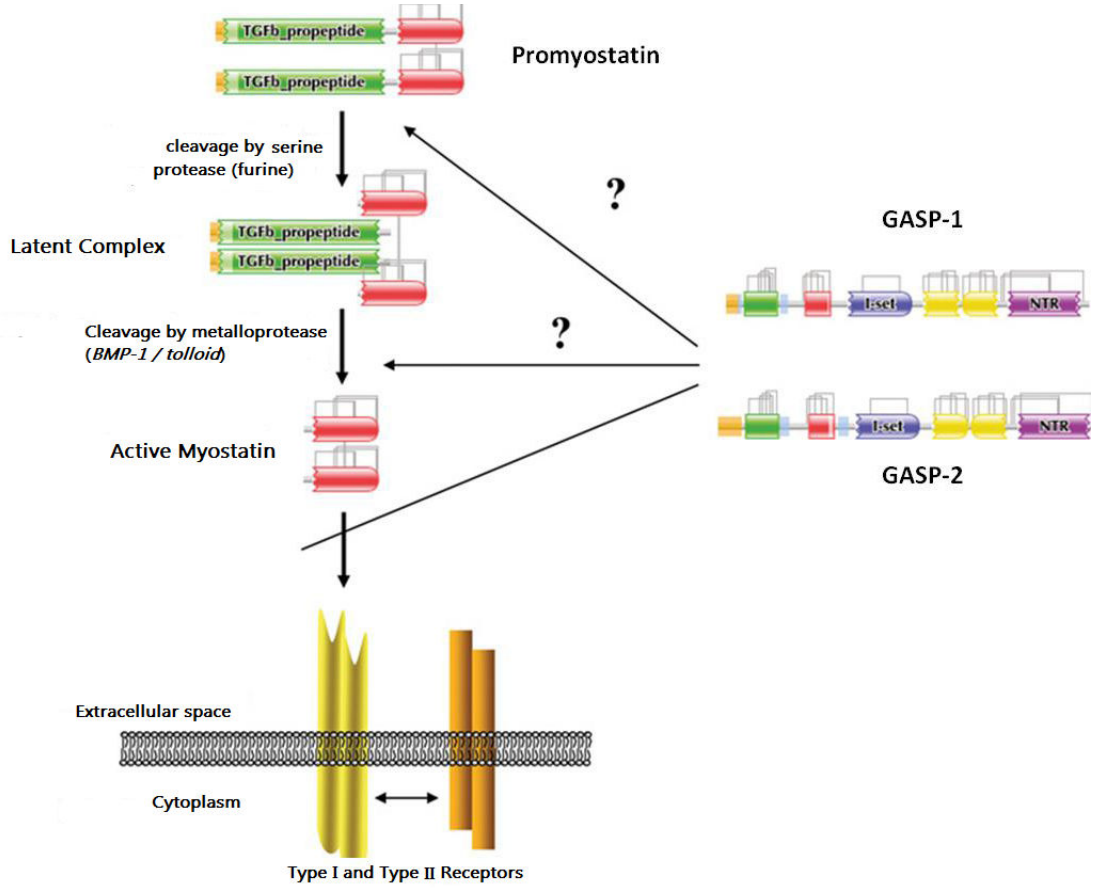


FIGURE 14 : BIOLOGICAL FUNCTIONS OF THE GASP PROTEINS AS ANTAGONISTS OF MYOSTATIN AND GDF-11

GASP-1 and GASP-2 stop the signaling activity of myostatin and GDF11 by sequestering these growth factors, thus preventing their binding to their receptors. The arrows with question marks show the possibility that GASP might interfere with the formation of mature active growth factors through binding the propeptide domains in the precursors or latent complexes, blocking the action of proteases involved in the release of mature growth factors (Kondas *et al.*, 2011).

The presence of the GASP proteins in skeletal muscle at the fetal stage and to a lesser extent at the adult stage would allow them to participate in the regulation of myostatin during the different phases of muscle development. Expression levels of myostatin inhibitors during longitudinal muscle growth indicate that GASP-1 may regulate myostatin during this growth (Aoki *et al.,* 2009). In addition, various studies indicated that GASP-1 is a more potent myostatin inhibitor than GASP-2 (Walker *et al.,* 2015). However, a recent work suggests that GASP-2 also has an effect on muscle growth and axial skeletal placement (Lee and Lee, 2013). The inactivation of GASP-1 and

GASP-2 leads to a modification of the axial skeleton pattern: a loss of the 13th rib in the majority of GASP-2 null mice.

II.3.3. Other roles of the GASP proteins

The GASP-1 and GASP-2 proteins show about 55% of identity and several conserved domains potentially involved in the inhibition of proteases correlating with an important and common expression in the pancreas as shown in figure 13. The different domains WAP, Kazal and Kunitz are regularly involved in the inhibition of serine proteases (Trexler *et al*., 2001). The WAP domain can also inhibit cysteine proteases. Finally, the Netrin domain which is found in TIMPs (Tissue Inhibitors of MetalloProteinases) is involved in the inhibition of the metalloproteinases (Banyai and Patthy, 1999). These domains may prevent the action of furine (a serine protease) and BMP-1/tolloid (metalloprotease) involved in the maturation of myostatin. More recent studies on whole GASP-1 and GASP-2 proteins would indicate their ability to inhibit trypsin (Kondás *et al.,* 2011). So far, only the *in vitro* measured kinetic parameters for the isolated second Kunitz domain of GASP-2 have been published (Nagy *et al.,* 2003; Liepinsh *et al.,* 2006). The authors indicate that an activity against trypsin is currently associated with this domain but no effect was detected on the peptidolytic activity of bovine elastase, chymotrypsin, tissue type plasminogen activator, urokinase type plasminogen activator, furin and BMP-1.

CHAPTER III : THE PROTEASE/ANTI-PROTEASE SYSTEMS

Most of the studies on the GASP proteins focus on their role(s) as inhibitors of myostatin. Although present in the MEROPS database, the GASP proteins are classified as putative protease inhibitors. As described in the chapter II, the GASP proteins are one of the rare heterotypic compound inhibitors. They are composed of several anti-proteasic domains that could be involved in the maturation of myostatin. The WAP, Kazal and the two Kunitz domains are serine protease inhibitor domains that are involved in the cleavage of myostatin, resulting in an inactive uncleaved myostatin. The “Netrin like” domain could directly inhibit the BMP-1/tolloid that are potentially involved in the maturation of the myostatin by cleavage of the proprotein as seen in figure 12.

Furthermore, if *Gdf-8* and *Gdf-11* are only expressed in muscle and axial skeleton respectively, *Gasp-1* and *Gasp-2* are expressed in many fetal and adult tissues. Thus, these proteins could participate in the protease/anti-protease equilibrium in many tissues and organs. To date, only the anti-trypsin activity of the second Kunitz domain of GASP-2 have been demonstrated (Nagy *et al*., 2003). No kinetic data have been published for the inhibitor activity of the other domains or the whole protein GASPs.

III.1. The different types of proteases

The proteases and their inhibitors form in the body an equilibrium that regulates many biological functions as digestion, reproduction, homeostasis inflammatory response or coagulation. Proteases are widespread enzymes in human, animals, plants and micro-organisms. Proteases regulate the fate, localization and activity of many substrates, create new bioactive molecules, contribute to process cellular information, transduce and amplify molecular signals and modulate protein-protein interaction (Carlos and Judith, 2008). The proteases are divided into exopeptidases, which cleave a carboxyl-terminal amino acid of a protein or peptide, and endopeptidases, which cleave peptide bonds of nonterminal amino acids, i.e. within the molecule. The exopeptidases can be further subdivided into the carboxypeptidases A and B and the amino-peptidases, comprising many different members with very similar specificities. Several isoforms of carboxypeptidases have been isolated.

According to “The International Union of Biochemistry”, proteases can be classified into four major families: cysteine proteases, aspartyl proteases, serine proteases and metalloproteases. This classification is based on the amino acid composition of their catalytic site (cysteine, aspartyl, serine) or the use of metal ions leading to different mechanisms of action (**Figure 15**). Furthermore, each family has its own substrates, active sites and cellular sources (Megan and Ilona, 2015).

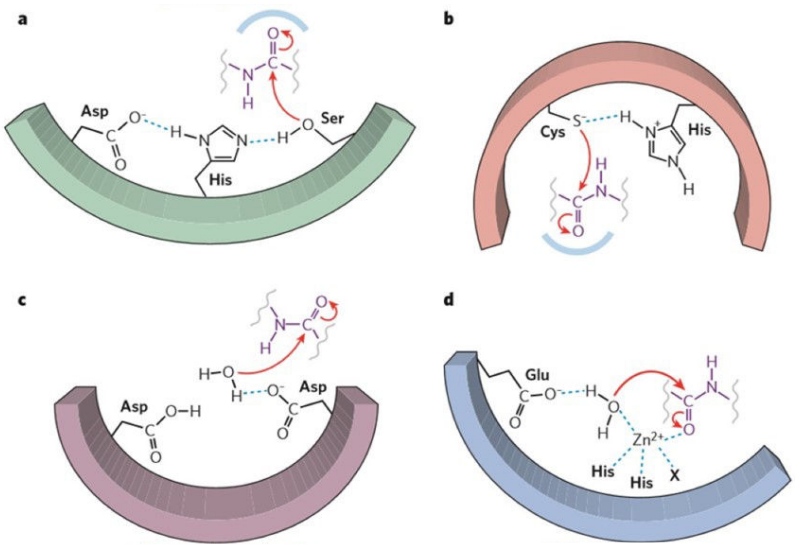


FIGURE 15: **MECHANISMS OF ACTION OF PROTEASES**

The different classes of proteases, serine proteases (**a**), cysteine proteases (**b**), aspartyl proteases (**c**) and metalloproteases (**d**) have a similar mechanism of action. Nevertheless, aspartyl proteases and metalloproteases use a water molecule for nucleophilic attack of the peptide bond (according to Erez *et al*., 2009).

For the different classes of proteases (cysteine, metallo and aspartyl), their inhibitors will be briefly presented. A special chapter will be dedicated to serine protease inhibitors, Kazal and Kunitz inhibitors.

III.1.1. Cysteine proteases

Cysteine proteases are one of the five major classes of proteolytic enzymes involved in a number of physiological and pathological processes and are widespread,

being found in viruses and almost every group of living organisms, including bacteria, fungi, protists, plants, invertebrates and vertebrates. When their synthesis, activity and localization in mammalian cells are altered, they may contribute to the development of many diseases, such as rheumatoid arthritis, osteoporosis or cancer (Kedzior *et al.,* 2016). Inhibitors of cysteine proteases are also produced by almost every group of living organisms and being responsible for the control of intracellular proteolytic activity. Along with aspartic, serine, threonine proteases and metalloproteinases, they constitute the biocatalysts which hydrolyze peptide bonds in various proteins (Kedzior *et al.,* 2016). Moreover, the hallmark of cysteine proteases is the presence of cysteine residue in the enzyme's active site. Papain-like cysteine proteases are mainly endopeptidases, yet some of them possess additional or exclusive exopeptidase activity (Chapman *et al.,* 1997). The activity of cysteine proteases is often controlled within a cell by their endogenous inhibitors in order to maintain physiological levels of proteolysis (Vergnolle, 2016). However, such a protective action is not the only attribute of the inhibitors, the differences in their structure, specificity, affinity and distribution indicate much more complex roles of these molecules, which have also been proven to interact with exogenous peptidases produced by other species.

There are 20 families of the proteinaceous inhibitors of cysteine proteases having representatives in viruses, microorganisms, plants and animals. These molecules are predominantly tight-binding and reversible inhibitors (Turk *et al.,* 2001; 2012).

III.1.2. Aspartyl proteases

Aspartyl proteases are eukaryotic protein enzymes which catalyze peptide substrates through aspartate residues. Aspartyl proteases include cathepsins, renin and pepsin (Rich, 1985). In their catalytic sites, aspartic proteinases possess two residues of aspartic acid, where the activated water molecule act as a nucleophile to attack the scissile peptide bond. In addition, within the proteinases belonging to this group, Cathepsin D is the most important aspartic proteinase implicated in Extracellular Matrix (ECM) degradation. Cathepsin D is likely responsible for intra-cellular degradation of phagocytosed ECM fragments that were degraded in the extra-cellular regions (Okada, 2017).

III.1.3. Metalloproteases

The metalloproteases also called Matrixines or MMPs constitute a group of proteins involved in the remodelling of the extracellular matrix and in the degradation of collagen. As a result, MMPs play an important role in tissue remodelling including tissue regeneration, wound healing, bone growth and regulation of organ development (Bourboulia and Stetler-Stevenson, 2010). These proteases are capable to regulate cellular functions, including survival, cell signalling and angiogenesis.

In human, 23 genes encoding metalloproteases have been identified. These proteases are endopeptidases with multiple Zn^{2+} binding domains required for the catalytic reaction. They also require other divalent ions such as Ca^{2+}. These proteases are synthetized with a 20 amino acids signal peptide and secreted as pro-proteinases which require activation by cleavage of the propeptide domain (Van Wart et Birkedal-Hansen, 1990; Romi *et al*., 2012). Structurally, they are characterized by a catalytic core called Zn^{2+} binding site of about 170 amino acids and consisting of a conserved pattern HEXXHXXGXXH. In most of these enzymes, there is a binding peptide of variable length, also called hinge region and a hemopexin domain of about 20 amino acids contributing to substrate and endogenous inhibitors specificities (Cauwe *et al*., 2007). The MMPs activity is very low in the normal steady-state-tissues and the majority of the MMPs are activated by other MMPs or serine proteases outside the cell. The active MMPs are tightly inhibited by its own endogenous inhibitors. Between all MMPs inhibitors, tissues inhibitors of metalloproteases are essential in the regulation of the proteolytic activity of MMPs (For review see Fassina *et al*., 2000).

Members of the MMPs family were subdivided into five subgroups: stromelysins, “membrane-type” MMPs, gelatinases, collagenases and matrilysins (Brew and Nagase, 2010). They are therefore key players in many pathophysiological processes such as morphogenesis, tissue repair and tumor progression.

Vertebrate bone morphogenetic protein 1 (BMP-1) and Drosophila Tolloid (TLD) are prototypes of a family of metalloproteases with important roles in various developmental events. BMP-1 is a zinc-dependent metallo-proteinase, originally identified in bone extracts, which is capable of inducing bone formation at ectopic sites and morphogenesis (Vadon-Le Goff *et al.,* 2015). BMP-1 molecule is composed of several structural motifs. These include a protease domain common to the astacin family, an EGF-like motif, several copies of a domain present in complement components C1r/C1s and thought to mediate protein-protein interactions. The latter domain has been referred to a CUB (Complement-Uegf-BMP-1) domain. This protein remains one of the major proteases involved in extracellular matrix assembly.

However, it has also established its roles in the activation of growth factors, including TGF-β1, BMP-2/4, IGFs and GDF-8/11. Active myostatin is released from latent myostatin by proteases of the BMP-1/tolloid family that cleave the prodomain of myostatin, thereby disrupting the noncovalent myostatin/prodomain complex (Szláma *et al*., 2016). The key importance of BMP-1-mediated cleavage of latent myostatin for the release of mature myostatin is supported by the observation that mice carrying BMP-1-resistant prodomain myostatin exhibit significant increase in muscle mass (Lee, 2008).

Tissue Inhibitors of Metalloproteinases (TIMPs) include four members; TIMP1, 2, 3, 4. TIMP3 is bound to ECM, whereas TIMP1, 2, and 4 are secreted in the soluble form (Bode *et al.,* 1999). Furthermore, the TIMPs molecule consists of 184 - 194 amino acids with two structural and functional domains: C-terminal and N-terminal. C-terminal region with pro-forms of MMP-2 and MMP-9 stabilize this complex, while N-terminal domain interacts with the enzyme catalytic domain. TIMPs inhibit the proteolytic activity of MMPs in a 1:1 molar stoichiometry (Lukaszewicz *et al.,* 2014). Therefore, TIMP-1 is the most effective inhibitor for MMP-1, MMP-3, MMP-7 and MMP-9; TIMP-2 is the most effective inhibitor for MMP-2; TIMP-3 is the most effective inhibitor for MMP-2 and MMP-9, whereas TIMP-4 reduces the activity of MMP-2 (Nagase *et al.,* 2006; Bourboulia and Stetler-Stevenson, 2010; Brew and Nagase, 2010). TIMPs have essentially an inhibitory function, while they can serve as growth-like factors in addition to anti-angiogenic agents. TIMPs are produced by alveolar macrophages and alveolar epithelial cells. TIMPs inhibit MMP activity by forming an N-terminal reactive ridge domain which inserts into the active site of the target MMP. Additionally, the TIMP family shares around 40% homology and has overlapping anti-MMP activity. TIMPs encourage cell growth and proliferation. TIMP1 and TIMP2 activate growth of erythroid cells, fibroblasts, and keratinocytes (Megan and Ilona, 2015).

III.1.4. Serine proteases

Among the enzymes with proteolytic activity, serine proteases represent more than 30% of the 26,000 referenced proteases (Barrett et Rawlings, 1995). Serine proteases are present in all living organisms including viruses and bacteria. Based on their three-dimensional structure, these proteins are divided into 13 clans and 40 families (**Table 3**). The PA clan, found exclusively in eukaryotes, is the most studied. It is the most important in number and includes trypsin, chymotrypsin or elastase.

TABLE 3: CLANS, FAMILIES AND CATALYTIC RESIDUES OF SERINE PROTEASES

Clan	Number of families	Example	Catalytic residues
PA	12	Trypsin	His, Asp, Ser
SB	2	Subtilisin	Asp, His, Ser
SC	6	Prolyl oligopeptidase	Ser, Asp, His
SE	3	D-A-D-A carboxypeptidase	Ser, Lys
SF	2	LexA repressor	Ser, Lys/His
SH	1	Cytomegalovirus Assemblin	His, Ser, His
SJ	3	Lon	Ser, Lys
SK	3	Clp	Ser, His, Asp
SP	1	Nucleoporin	His, Ser
SQ	1	Aminopeptidase DmpA	Ser
SR	1	Lactoferrin	Lys, Ser
SS	1	L,D-carboxypeptidase	Ser, Glu, His
ST	1	Rhomboid	His, Ser

Catalytic mechanism of serine proteases

The name of these proteases is due to the presence of a nucleophilic serine residue located in the active site of the enzyme and responsible for the catalytic activity. Other residues also occur in the active site, thus constituting the catalytic dyad or triad. For the PA clan, it is a triad formed of histidine, aspartic acid and serine residues. Most of the serine proteases of this clan are synthetized as an inactive precursor called zymogen. This precursor is then cleaved between residues 15 and 16. This cleavage causes a conformational change of the protein to form the triad. As example, for the chymotrypsin, the triad is formed between residues serine 195, histidine 57 and aspartic acid 102. This structuration also promotes the formation of a pocket that will stabilize the triad during the enzymatic reaction. During the enzymatic reaction, catalysis of the peptide bound requires the formation of two tetrahedral intermediates (**Figure 16**).

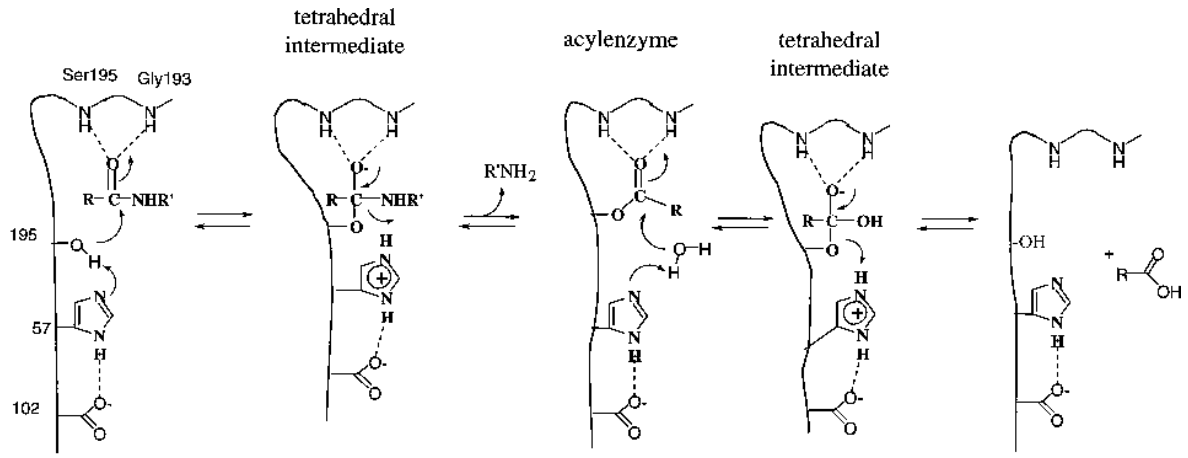


FIGURE 16: SCHEMATIC REPRESENTATION OF THE CATALYTIC MECHANISM OF SERINE PROTEASE OF THE PA CLAN

In this representation, only the side chains of the amino acids of the catalytic pocket are indicated. The proteolysis of the substrate RCONHR' requires the contribution of an oxygen atom of the serine. The cleavage of the substrate passes through the formation of two tetrahedral reaction intermediates. The proteolysis leads to the release of the two molecules R'NH2 and RCOOH (Adapted from Hedstrom, 2002).

In a first step, the oxygen atom of the hydroxyl function of serine attacks the substrate RCONHR' with the aid of histidine, forming a first tetrahedral intermediate which generates the acylenzyme and releases the first part of the substrate. In a second step, a molecule of H_2O displaces the free peptide fragment R'NH_2, attacks the acyl-enzyme and forms the second tetrahedral intermediate stabilized by the pocket. This results in the release of the second part of the substrate RCOOH and the regeneration of serine and histidine residues.

Normal regulation of serine proteases activity is critical for physiological activities of the cell and tissue. However, abnormal regulation of these proteases activity can cause pathological conditions such as cancer (DeClerck *et al.,* 2004).

III. 2. The serine proteases inhibitors

The proteases are subject to regulations that modulates their activities. In the body, proteases and their inhibitors form an equilibrium that regulates many physiological mechanisms such as homeostasis, inflammatory response, digestion, reproduction or even blood coagulation. The protease inhibitors constitute the third functional group of proteins in human plasma after albumin and immunoglobulins. This

class of proteins represents about 10% of the circulating proteins (Carrell and Travis, 1985). They are different types of inhibitors and most of them are specific to a class of proteases. However, some inhibitors can act on proteases of different classes. This phenomenon is described as “cross class inhibition”.

Three well-known mechanistic types of protease inhibition have been described: The first type, considered as the standard mechanism involves a broad group including many families (*e.g*., Kunitz and Kazal families), functions mostly as tight competitive binders to the active site (Laskowski *et al.,* 2000). These inhibitors usually have small rigid reactive-center loops (RCLs) affixed to the body by a disulfide link (Read *et al.,* 1986). They bind to the active site of the protease reversibly, forming a high affinity complex in which the reactive site bond is hydrolyzed slowly. Due to the smallness and rigidity of the RCL, cleavage of the scissile bond does not result in a significant displacement of the P1 and P’1 residues. As a result, the native and cleaved forms of the inhibitor are in equilibrium and are both functionally inactivated. These inhibitors reach a thermodynamic equilibrium (Schechter and Plotnick, 2004). The second mechanistic type is the macroglobulin (Travis and Salvesen, 1983). The primary example of this inhibitor type is the human alpha 2-macroglobulin. These inhibitors are large multi-subunit proteins, each subunit has a specialized segment of polypeptide chain termed the bait region, which is accessible to the protease. Cleavage of the bait section induces a conformational change that traps the protease in a cage-like structure. The trap is relatively non-specific and any protease cleaving the bait region may be trapped. Within the trap, the active site of the protease is free and capable of hydrolyzing small synthetic substrates. The trap stability is not permanent and could differ for different proteases. The third mechanistic type of inhibitor is represented by the Serpins. The recent crystal structure of the complex formed upon inhibition of trypsin by alpha 1-antitrypsin supports a mechanism involving the trapping of a catalytic intermediate formed during the course of hydrolysis of the reactive site. Unlike the protease alpha 2-macroglobulin complexes, protease Serpin complexes are inactive (Huntington *et al.,* 2000).

III.2.1. The serine proteases inhibitors

Serine protease inhibitors were first studied in human blood plasma (Travis et Salvesen, 1983) before being found in other tissues and organisms. The serine protease inhibitors are divided into several families (**Table 4**). This classification is

based on sequence homology and topology and takes into account their protease binding mechanisms.

TABLE 4: PARTIAL LIST OF SERINE PROTEASE INHIBITORS (adapted from Rawlings *et al.*, 2004).

Common name	**Family number in MEROPS database**
Kazal	I1
Kunitz (animal)	I2
Kunitz (vegetal)	I3
Serpins	I4
Cereals	I6
Squash	I7
Ascaris	I8
Protease B inhibitor	I9
Bowman-Birk	I12
Hirudin	I14
Substilisin inhibitor (SSI)	I16
Elafin	I17
Potato carboxypeptidase inhibitor	I37
Alpha2-macroglobulin	I39
Latexin	I47

III.2.1.1. Serpins

Physiological roles of serpins

Serpins, acronym for "**SER**ine **P**rotease **In**hibitors" are considered the major family of serine protease inhibitors. To date, more than 1,500 serpins have been found in animals, plants, bacteria and some viruses (Rawlings *et al.*, 2008; 2012). By phylogenetic analysis, Irving *et al.*, (2000) classified the serpins in 16 clades (A to P). In vertebrates, a phylogenetic analysis based on 110 sequences and based on protein sequences and structural gene organization confirmed the distribution of these proteins in 6 groups (Ragg *et al.*, 2001). In human, the two largest groups of serpins are the antitrypsin-like (Serpin-A) and the ovalbumin-like (Serpin-B). No evidence exists for a horizontal transfer of serpin genes. Serpins would be ancient proteins for which most procaryotes would have lost the use (Roberts *et al.*, 2004). They are involved in many highly regulated physiological processes such as blood coagulation, fibrinolysis, cell migration and inflammation (Potempa *et al.*, 1994).

Most of serpins inactivate serine protease targets (van Gent *et al*., 2003). Despite their initial name, the specificity of serpins is not limited to serine proteases. It has been reported that some serpins are able to inactivate cysteine proteases. This phenomenon called "cross-inhibition" was initially observed by Salvesen (1993) and now, many examples are published (Zhou *et al*., 1997; Schick *et al*., 1998; Blanchet *et al*., 2009). The mechanism of inhibition of cysteine proteases is not fully understood. However, it appears that several serpins inhibit these proteases by an irreversible mechanism similar to that described for serine proteases.

However, some serpins are non-inhibitory. Serpin A8 is a cortisol-binding globulin that mediates the transport of hormones. In plant, their exact roles are still uncertain. Pumpkin CmPS appears to regulate the immune response after parasite infection (Yoo *et al*., 2000). Finally, in procaryotes, target proteases and biological functions of these proteins remain to be discovered.

Structural characteristics of serpins

In mammals, genes encoding serpins are often clustered, presumably as a result of gene duplications. Most of the serpins are glycoproteins composed of oligosaccharide moieties and a single polypeptide chain of 350 to 500 amino acids (Nakashima *et al*., 2000). Despite the diversity of amino acid composition and functions, serpins share the same overall fold and demonstrate a highly conserved structure consisting of three β-sheets (A, B, C) and nine α-helices. The β-sheet A is the largest and plays an important role in conformational dynamics (Yamasaki *et al*., 2002).

At the C-terminal end Serpins have a mobile sequence of about 20 amino acids recognized by the target protease and called RCL for “Reactive Center Loop” or RSL for “Reactive Site Loop”. This native or stressed “S” conformation is very thermodynamically unstable. The RCL is a very mobile and flexible fragment. It interacts specifically with the catalytic site of the target protease acting as a "pseudo-substrate". This interaction most often results in the cleavage of the RCL and the insertion of its remaining part in the β-sheet A forming a new strand. The serpin adopts a relaxed “R” conformation more stable (Im *et al*., 2000) (**Figure 17**).

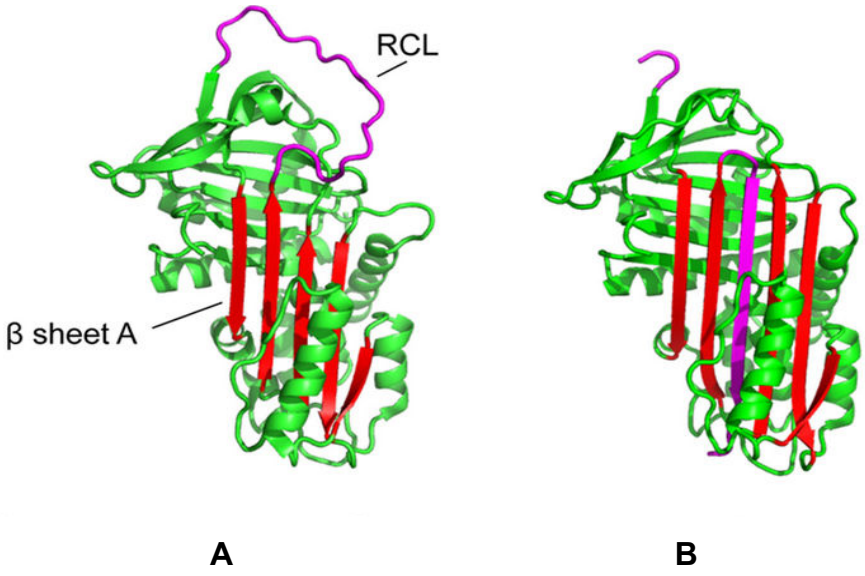


FIGURE 17: STRUCTURE OF NATIVE/STRESSED (A) AND CLEAVED/RELAXED (B) ANTI-THROMBIN
In the cleaved or covalently complexed form of the serpin, the RCL forms an extra strand in β-sheet A. β-Sheet A is in red and the RCL is in magenta (Adapted from Whisstock *et al*., 2010).

Crystallographic studies have shown that serpins are able to adopt other conformations: latent, δ and polymerized (Whisstock *et al*., 1998; Dafforn *et al*., 2004; Marszal and Shrake, 2006).

The flexibility and metastability of serpins makes them highly sensitive to aberrant conformational transitions that may result from point mutations. These transitions can disrupt their inhibitory function and lead to various pathologies such as cirrhosis, dementia or emphysema. All these pathologies have been grouped under the term of “serpinopathy”.

Inhibition mechanism of serpins

Except for few special cases, serpins inhibit their target protease by a "suicide-substrate" mechanism (Patston *et al*., 1991). This mechanism leads to a complete inactivation of the protease and the inhibitor and is shown in figure 18.

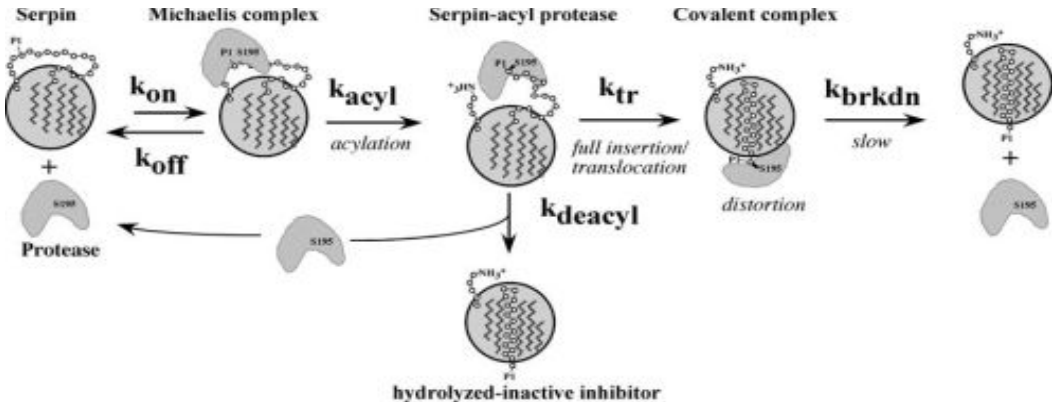


FIGURE 18: MECHANISM OF PROTEASE INHIBITION BY SERPINS

The inhibition pathway is demonstrated in the horizontal steps showing the slow breakdown of the covalent complex. k_{on} and k_{off} are the bimolecular and dissociation rate constants for the formation or dissociation of the Michaelis complex. k_{acyl} is the rate constant in the chemical step producing the Serpin–acyl-protease intermediate. k_{tr} is the rate constant for translocation of the protease to the distal pole. k_{brkdn} is the rate constant for slow complex breakdown. k_{deacyl} is the rate constant for substrate-like deacylation of the Serpin–acyl-protease intermediate. Competition between these k_{tr} and k_{deacyl} steps accounts for apparent stoichiometries of inhibition greater than 1 (Schechter and Plotnick, 2004).

At the beginning of the reaction, the protease forms with the inhibitor a non-covalent Michaelis-type complex by interactions with amino acids located on either side of the P1-P'1 cleavage site of the RCL (a). P1 represents the residue after which serine 195 of the active site of the protease cleaves the inhibitor and releases the C-terminal portion of the RCL. A covalent acyl-ester bond is then formed between the serine 195 of the protease and the carbonyl function of the residue P1 (b). The complex thus formed becomes thermodynamically stable. The serpin changes from conformation "S" to conformation "R". The RCL is inserted into the sheet A training the protease. The complete insertion of the RCL allows the protease to pass to the other end of the serpin (c). This switch causes a distortion of the active site of the protease, due to its compression. The catalytic pocket formed by the triad is destabilized preventing the end of catalysis and the release of the substrate (Huntington *et al.*, 2000; Gettins, 2002). Hydrolysis of the peptide bond may allow release of the inactive protease and the cleaved inhibitor (d). In some cases, proteolysis of serpin may occur before or during protease insertion and lead to release of intact protease and inactive

cleaved inhibitor (e). However, the protease has a weak functional activity and this metabolic pathway has not a biological significance.

III.2.1.2. Kazal-type inhibitors

The Kazal-type serine protease inhibitors belong to MEROPS inhibitor family I1 (Table 4). This family is named in reference on the work on the pancreatic secretory trypsin inhibitor, first isolated by Kazal *et al*. (1948).

Kazal domain is structurally highly conserved, similar to the conformation of small serine protease inhibitors. A typical or canonical Kazal domain is composed of 40-60 amino acids and its structure consists of a central α-helix which is inserted between two β-strands, a third that is toward the C-terminus and three loops A, B and C. Kazal inhibitors are characterized by the presence of six cysteine residues engaged in three disulfide bounds necessary to the inhibitory activity and arranged as shown in the following schematic representation (**Figure 19**).

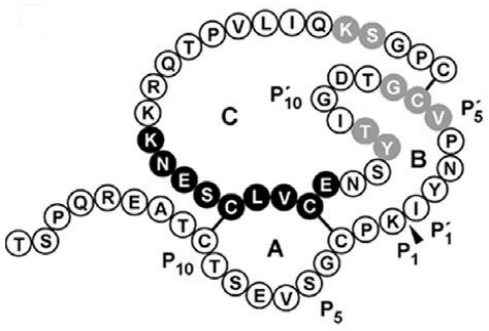


FIGURE 19: SCHEMATIC REPRESENTATION OF THE COVALENT PRIMARY STRUCTURE OF THE PORCINE PANCREATIC SECRETORY INHIBITOR

α-helix and β-sheets are shown in black and gray circles respectively. The scissile peptide bound is indicated by arrow. The structure can be viewed as consisting of three loops A, B and C. The B loop harbours the specificity determining P1 amino acid and the scissile peptide bond, and is, thus, called the reactive site loop (Rimphanitchayakit and Tassanakajon, 2010).

Loop B displays the P1 and P1′ residues which represent the protease inhibitor. There is a standard mechanism to inhibit the proteinase by the Kazal protease inhibitory domain (Magert *et al.,* 2002). Each Kazal domain acts as a substrate analogue that stoichiometrically binds competitively during its reactive site loop to the active site of cognate protease forming a relatively stable protease-protease inhibitor complex,

much more stable than the Michaelis enzyme–substrate complex (Li *et al.,* 2009). However, the binding is non-covalent, it is very tight as the association constant is extremely high (10^7–10^{13} M^{-1}) and therefore, the inhibition is so strong. In addition, despite that the inhibitory specificity is mainly characterized by the P1 amino acid residue, amino acid residues in other contact sites influence the potency of the binding as well as the specificity of a serine proteinase inhibitor to its cognate protease (Rimphanitchayakit and Tassanakajon, 2010).

Kazal-type protease inhibitors are distributed in a wide range of organisms from all kingdoms of life and play crucial role(s) in various physiological mechanisms. They inhibit a number of serine proteases such as trypsin or elastase and can contain several Kazal repeats. There is a large number of proteins with Kazal domain. Excepted GASP-1 and GASP-2, we can give two other examples:

-The membrane-anchored glycoprotein RECK is a cysteine-rich protein with Kazal motifs that inhibits expression and activity of some matrix metalloproteinases (MMP-2 and -9), thereby suppressing tumor cell metastasis (Takahashi *et al.*, 1998; Oh *et al.*, 2001).

- The HTRA1 protein is composed of four distinct domains ; an Insulin-like growth factor binding domain, a kazal domain, a trypsin-like peptidase domain and a PDZ domain. This protein is a secreted enzyme that is proposed to regulate the availability of insulin-like growth factors (IGFs) by cleaving IGF-binding proteins. It has also been suggested to be a regulator of cell growth (Zumbrunn and Trueb*,* 1997*)*.

However, kazal-like domains are also seen in the extracellular part of agrins, which are not known to be protease inhibitors.

III.2.1.3. Protease inhibitor Kunitz-type domain

Majority of the sequences having a Kunitz-type domain are classically classified as serine protease inhibitors and belong to the MEROPS database I2, clan IB (Table 4). Kunitz-type proteins are an important group of ubiquitous protease inhibitors found in viruses up to mammals (Bode and Huber, 2000). The prototype for this family is the basic trypsin inhibitor of bovine pancreas also known aprotinin (Kassell, 1970) but the family includes numerous other members such as the Alzheimer's amyloid precursor protein (APP) (Tanzi *et al*., 1988) or the tissue factor pathway inhibitor (TFPI) (Crawley and Lane, 2008). If Kunitz-type inhibitors are naturally associated with trypsin, other serine proteases are usually inhibited including the neutrophil elastase, the digestive chymotrypsin and several proteases which are involved in the blood coagulation cascade such as kallikrein, thrombin and many other tissue factors. These proteins

also play an important role in inflammatory and fibrinolysis processes (Wan *et al.*, 2013). In invertebrates, they are involved in a range of various functions.

These proteins can have single or multiple Kunitz inhibitory units linked together or associated with other domain types. In their monomeric form, Kunitz domains are low molecular mass (6~8 kDa) peptides of 50 to 70 amino acids adopting a conserved structural fold with two antiparallel β-sheets and one or two helical regions (**Figure 20**) (Ranasinghe and McManus, 2013; Smith *et al.*, 2016). This arrangement is stabilized by three conserved disulphide bounds that maintains structural integrity of the inhibitor and allows presentation of a protease-binding loop at the surface of the molecule.

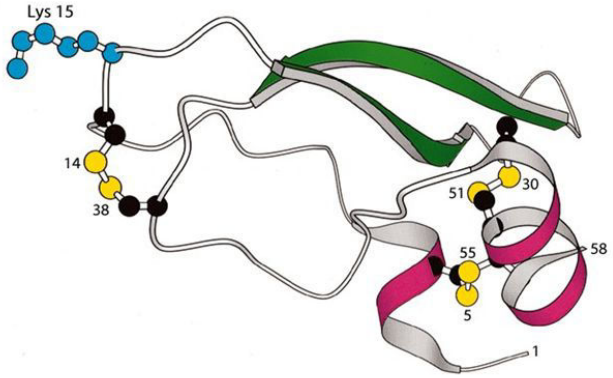


FIGURE 20: 3D STRUCTURE OF BOVINE PANCREATIC TRYPSIN INHIBITOR (BPTI)

The two antiparallel β-sheets are in green and the two α-helix are in magenta. The three disulphide bounds symbolized by yellow circles are also represented. BPTI inhibits trypsin by inserting Lys-15 into the specificity pocket of the enzyme.

A highly exposed P1 active site residue which inserts into the S1 site of the cognate protease is located at the peak of the binding loop and is of prime importance in determining the specificity of serine protease inhibition (Krowarsch *et al.*, 2003). Within this inhibitor family, arginine or lysine, two residues with a positively charge side chain, occupy the P1 site in the reactive site and are the preferential site of interaction for the trypsin. Thus, Kunitz-type protease inhibitors are classically associated with trypsin inhibition (Ascenzi *et al.*, 2003; Farady and Craik, 2010).

The mechanism of action of Serpins also called the "suicide substrate" mechanism is the process for which the serine or cysteine protease forms a covalent complex with the inhibitor at the level of the RCL. Yet, other research has showed examples of

Serpins acting in a different mechanism, a reversible mechanism and therefore as a non-covalent inhibition such as for Kunitz protease inhibitors (Gulley *et al.*, 2013). The typical mechanism of Kunitz inhibition (canonical inhibition) includes a tight, non-covalent interaction like the enzyme–substrate Michaelis complex. Kunitz inhibitors would directly block the active site of the serine protease without any structural changes and all together form an antiparallel β-sheet between enzyme and inhibitor. The protease-binding loop is the section responsible for protease inhibition (Krowarsch *et al*., 2003). Residues that precede or follow this section and residues from a sequentially remote segment can also contact the enzyme and influence the association energy (Ascenzi *et al*., 2003).

RESULTS

Studies have shown that the second Kunitz domain of human GASP-2 has an inhibitory activity against trypsin (Nagy *et al*., 2003; Liepinsh *et al*., 2006), however, no kinetic parameter has been published for the second Kunitz domain of GASP-1. In the same way, no inhibitory activity was published for the two whole proteins. During this work, it seemed interesting to determine for both proteins if, in their native conformation, the second Kunitz domain retains its inhibitory properties. We were able to show that the GASP-2, produced *in E. coli*, is capable of inhibiting trypsin according to a competitive mechanism. Like GASP-2, GASP-1 has also anti-trypsin properties. However, although these two proteins are closely structural related, there is a difference in specificity between them. In order to determine whether this difference is only due to the second Kunitz domain or to the molecular environment of each protein, we have carried out analyses on two chimeric proteins in which only the second Kunitz domains have been interchanged. Therefore, in addition to the results submitted in *PLOS Biology*, some supplementary experimental data are detailled below.

I. Construction and production of the different recombinant proteins GASP used in this study

I.1. Choice of the heterologous expression system

To produce a heterologous protein, there are different expression systems. Although proteins GASP have *N*- and *O*- glycosylations, previous studies in the laboratory have shown that these oligosaccharidic chains are not essential for the anti-myostatin activity of GASP-1 and GASP-2 (Brun *et al*., 2012; Périé *et al*., 2016). These results allowed us to choose a prokaryotic expression system. This system can provide a high protein yield necessary to perform the various subsequent anti-protease assays. Indeed, we need about 100 µg of purified recombinant protein to perform an enzymatic triplicate test.

Before my arrival in the laboratory, two constructs had been made to allow the production of the recombinant proteins GASP-1 and GASP-2 in *E. coli* BL21 strain (Parenté A., personal communication). Full-length mature cDNAs encoding GASP-1

and GASP-2, excluding the signal peptide, were cloned in pGEX-4T-1 vector between the *Bam*H1 and *Not*1 restriction sites of the polylinker (**Figure 21**).

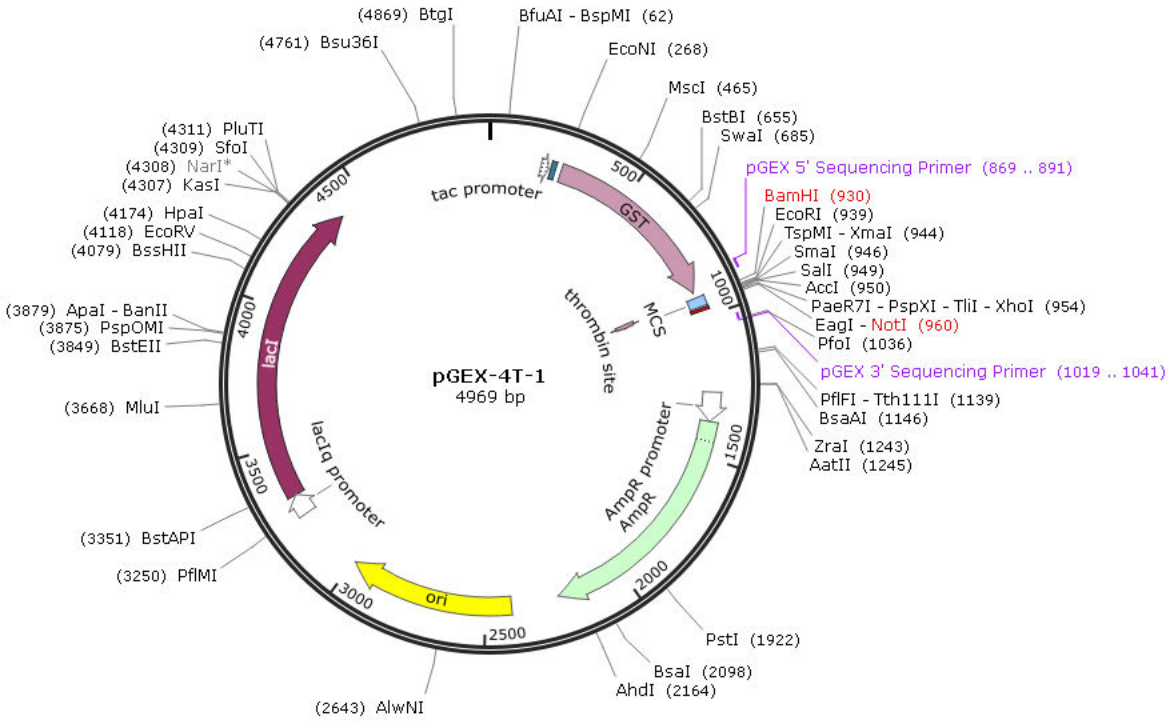


FIGURE 21: MAP OF THE GLUTHATIONE S-TRANFERASE FUSION pGEX-4T-1 VECTOR

PGEX-4T-1 is a double-stranded circular plasmid of 4.9 kbp in which the cDNA of interest is cloned under the control of the IPTG-inducible *tac* promoter. It allows the production of a fusion protein carrying a GST (Gluthatione S-transferase) tag at the N-terminus allowing its purification by affinity chromatography without denaturation. The removal of the GST tag is accomplished by using a site-specific thrombin recognition sequence located between the GST moiety and the target protein.

In this work, we also constructed two chimeric proteins corresponding to the proteins GASP-1 and GASP-2 in which the second Kunitz domains were interchanged. Although structurally close and possessing about 55% identity, proteins GASP show significant differences, especially in their second Kunitz domain. Of the 51 amino acids that make up this area, 20 differ. As a result, a strategy of modification by site-directed mutagenesis was not conceivable. We therefore opted for a method to replace the entire sequence encoding the second Kunitz domain in each of the constructs using

an assembly PCR strategy (NEBuilder HiFi DNA Assembly Master Mix, New England Biolabs). For example, the construction encoding the chimeric GASP-1/$K_{2\text{-}2}$ is schematized in figure 22. According to this strategy, only the second Kunitz domain is modified in both proteins; no other amino acid has been replaced.

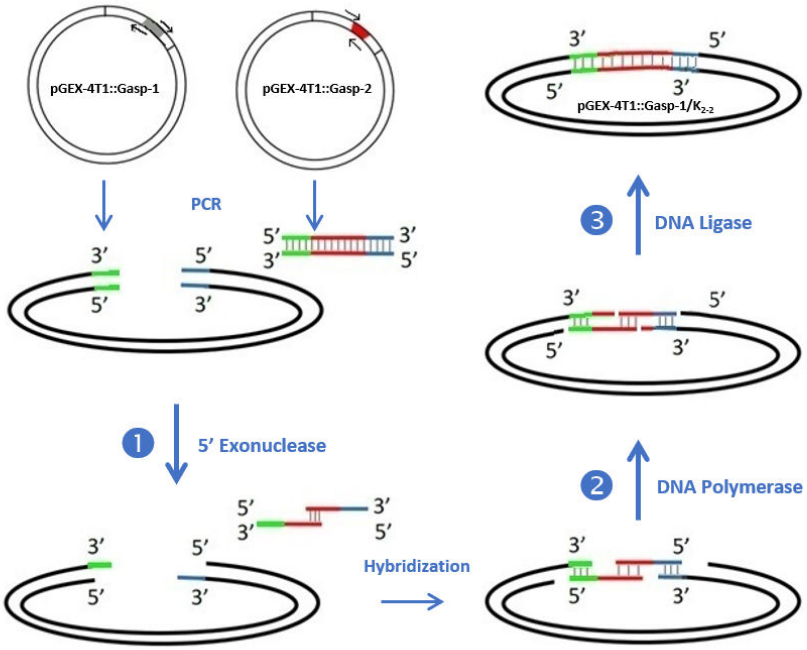


FIGURE 22: **OVERVIEW OF THE NEBuilder HiFi DNA ASSEMBLY METHOD ADAPTED TO THE CONSTRUCTION OF THE CHIMERIC PROTEIN GASP-1/$K_{2\text{-}2}$**

A large fragment (corresponding to GASP-1 without the second Kunitz domain) and a short fragment (encoding the second Kunitz domain of GASP-2) were amplified using sets of primers designed with an overlap region of about 20 bases. The assembly of the two PCR fragments was performed using a mix of three enzymes:

❶ 5' exonuclease creates single-stranded 3´ overhangs that facilitate the annealing of fragments that share complementarity at one end (the overlap region)

❷ DNA polymerase fills in gaps within each annealed fragment

❸ DNA ligase seals nicks in the assembled DNA

The end results is a double-stranded fully sealed DNA molecule that can serve as template for direct transformation of *E coli* BL21 strain.

The same strategy was used to construct the chimeric GASP-2/$K_{2\text{-}1}$.

I.2. Construction of the different recombinant proteins

As for GASP-1 and GASP-2, the two chimeric proteins were expressed into the plasmid pGEX-4T-1. These two chimeric constructs were named GASP-1/$K_{2\text{-}2}$ corresponding to the GASP-1 protein with the second Kunitz domain of GASP-2 and GASP-2/$K_{2\text{-}1}$ for the reciprocal protein corresponding to GASP-2 with the second Kunitz domain of GASP-1. The coding sequences of the two fusion proteins GST-GASP-1, GST-GASP-2 and the two chimeric proteins GST-GASP-1/$K_{2\text{-}2}$ and GST-GASP-2/$K_{2\text{-}1}$ are given in Appendix.

I.3. Optimization of production conditions of recombinant fusion proteins in soluble form

For each recombinant protein, the first production tests were performed under standard induction conditions; 3 h at 37°C, with stirring (250 rpm) and 1 mM IPTG. Under these conditions, the proteins are produced almost exclusively in the insoluble fraction in the form of inclusion bodies (**Figure 23A**). The different induction parameters (IPTG concentration, temperature and induction time) have been modified in order to optimize the amount of recombinant protein in the soluble fraction. The best yields were obtained after 15 h of culture at 20°C with 0.025 mM IPTG (**Figure 23B**). These culture conditions make it possible to produce the protein in a low quantity but mostly in soluble form (supernatant). The lowering of the temperature and concentration of IPTG induces a slowing of the bacterial expression system, which favours the production of more consistent and thus soluble proteins.

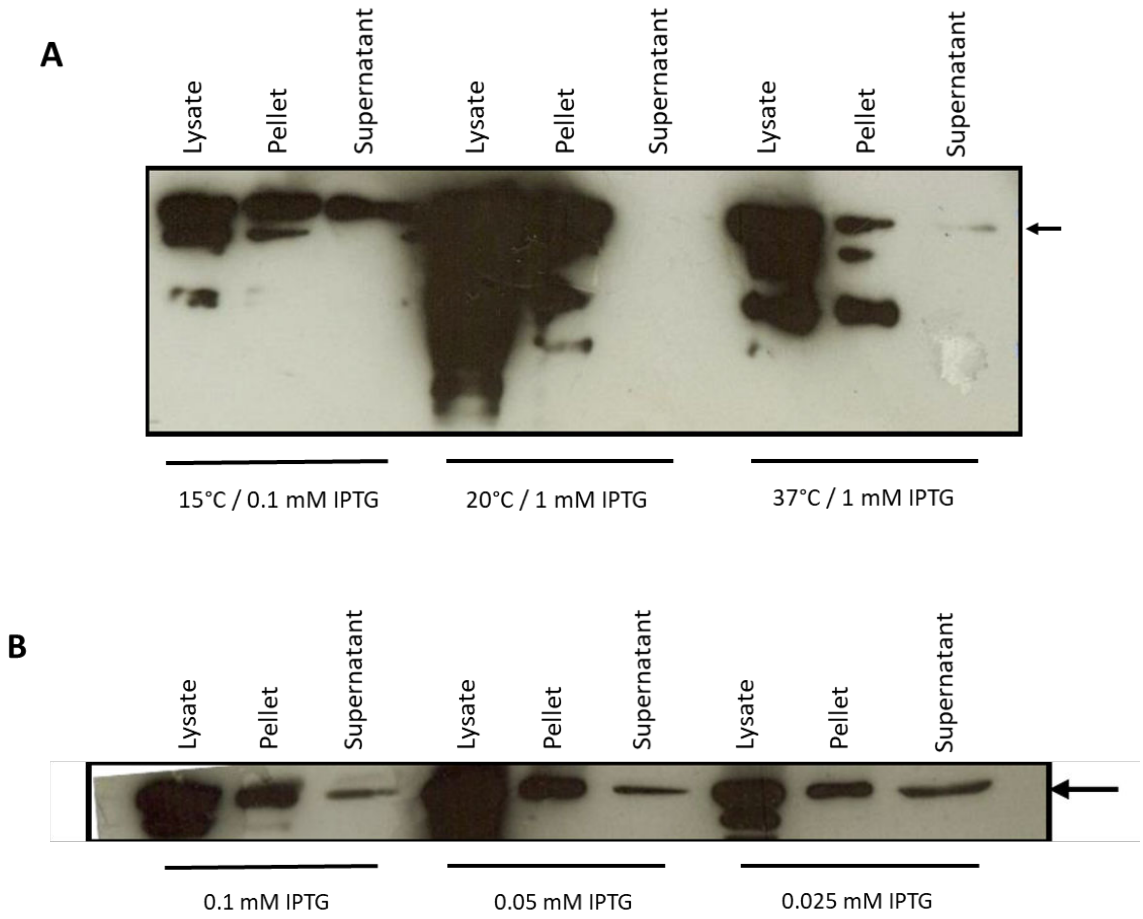


FIGURE 23: OPTIMIZATION OF THE PRODUCTION CONDITIONS OF THE RECOMBINANT FUSION PROTEIN GASP-1

The presence of the recombinant fusion protein GASP-1 in the bacterial lysate and in the insoluble (pellet) and soluble (supernatant) fractions was verified by Western blotting using a specific antibody directed against the protein.
The protein was produced during 15 h:
A – at different induction temperatures and different IPTG concentrations
B – at 15°C and different IPTG concentrations
the arrows indicate the recombinant fusion protein GASP-1 with an estimated size about 90 kDa.

All the recombinant proteins used in this study were produced under the same conditions: at 20°C for 15 h and with stirring (250 rpm) in the presence of IPTG (0.025 mM).

I.4. Purification of recombinant fusion proteins

Often, for their subsequent use in various biochemical studies, the recombinant fusion proteins produced by *E. coli* must be purified. The different fusion proteins GASP used

in this study were purified from the supernatant obtained after centrifugation of the bacterial lysate, by affinity chromatography on Glutathione Sepharose column using the ÄKTAprime plus system. As an example, the recombinant protein GASP-1/$_{K2-2}$ fixed on the column was eluted by competition using a reduced glutathione solution (**Figure 24**). Fractions 1, 2, 3 and 4 contain a majority band of about 90 kDa corresponding to the size of the fusion protein and confirmed by western blotting using a specific antibody directed against the GASP-1 protein (data not shown). For each recombinant protein produced and purified, the GST tag was then cleaved with thrombin following the instructions provided by the supplier. A protein assay makes it possible to estimate that one liter of bacterial culture produces approximately 300 to 500 µg for each partially purified recombinant protein.

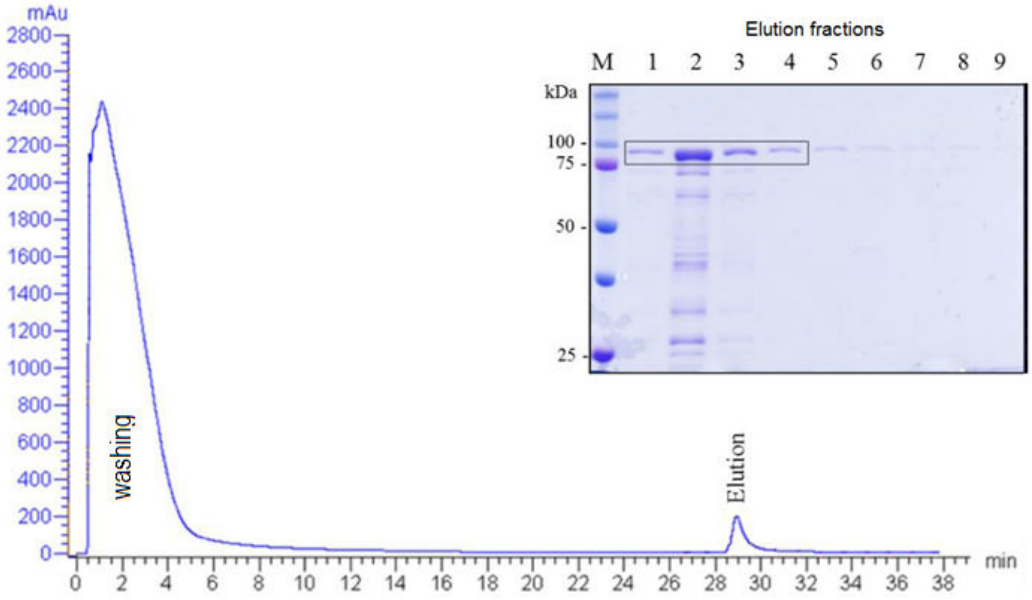


FIGURE 24: PURIFICATION OF THE RECOMBINANT FUSION PROTEIN GASP-1 / K_{2-2} BY AFFINITY CHROMATOGRAPHY ON THE ÄKTAprime-plus SYSTEM

After washing away the contaminating bacterial proteins, the proteins attached to the GST-Sepharose column were eluted in the presence of reduced glutathione. Nine elution fractions of 1 ml were collected (peak elution). For each fraction, an aliquot is analyzed by electrophoresis on 10% SDS-PAGE and stained with Coomassie blue. A protein of expected size (~ 90 kDa) is observed in several eluted fractions. **M**: protein marker (Precision Plus Protein ™ Dual Color Standards, Bio-Rad).

II. Functional activities of the different recombinant proteins GASP

II.1. Effect of the different recombinant proteins GASP on proliferation and differentiation of C2C12 myoblast cells

As previously described, GASP-1 and GASP-2 have been shown to bind to myostatin and are considered as inhibitors of TGFβ signaling pathway (Szlama *et al*., 2010). These two proteins might act as regulators of the myogenic proliferation and differentiation. In previous publications, we demonstrated that both the murine GASP-1 protein produced in an eukaryotic system and lacking of *N*- linked glycosylation (Brun *et al*., 2012) and the murine GASP-2 produced in a prokaryotic system (Périé *et al*., 2016) enhanced the C2C12 proliferation and differentiation by preventing myostatin signaling as reported by Bonala *et al*., 2012. Considering this function, we supposed that the chimeric GASP-1/$K_{2\text{-}2}$ and GASP-2/$K_{2\text{-}1}$ would conserve their ability to activate myogenesis. To test this hypothesis, C2C12 myoblast cells were treated with 1 $\mu g.ml^{-1}$ of GASP-1/$K_{2\text{-}2}$ or GASP-2/$K_{2\text{-}1}$ for 72 h as described in the “Materials and Methods” section of the publication. GASP-1/$K_{2\text{-}2}$ treatment improves significantly myoblast proliferation rate compared to control cells. Similar results were obtained with a GASP-2/$K_{2\text{-}1}$ treatment.

In a second step, we analyzed the effect of GASP-1/$K_{2\text{-}2}$ or GASP-2/$K_{2\text{-}1}$ during C2C12 myoblast differentiation. As previously, we treated C2C12 myoblast cells at 80% confluence with 1 $\mu g.ml^{-1}$ of recombinant protein and we followed the cell fusion for 96 hours We showed that the two chimeric proteins enhanced C2C12 differentiation by increasing the fusion index at 96 hours.

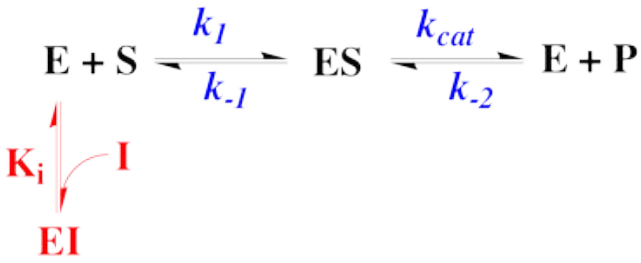
II.2. Determination of kinetic parameters of anti-trypsin activity

In order to check if the recombinant proteins GASP produced *in E. coli* retain the inhibitory properties, as already described for the second Kunitz domain of the human GASP-2, activity tests were carried out using a commercial trypsin. Trypsin is an endoprotease that hydrolyzes, in the C-terminal position, the peptide bonds in which a basic amino acid, lysine (K) or arginine (R) engages its acidic function.

II.2.1. Mechanism of a reversible competitive inhibition

According to the study published by Nagy *et al*. (2003), the second Kunitz domain of human GASP-2 inhibits trypsin by a reversible competitive-type mechanism.

The binding of the inhibitor (I) prevents the binding of the substrate (S) on the catalytic site of the enzyme (E) and *vice versa*. Competitive inhibition can be expressed according to the following reaction mechanism:



According to this reaction mechanism, binding of the substrate and the inhibitor to the enzyme are therefore mutually exclusive. There are several models to illustrate competitive inhibition as shown in figure 25.

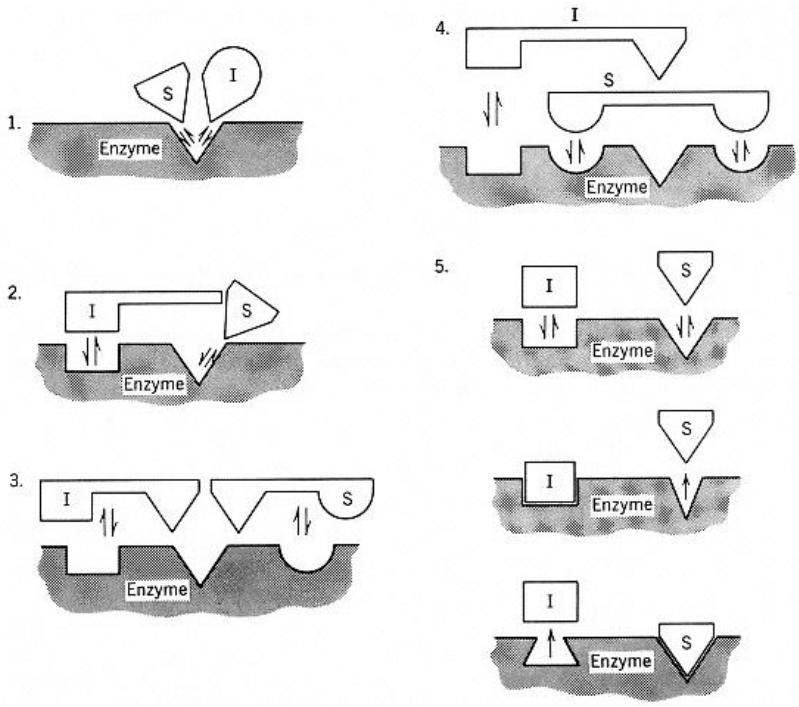


FIGURE 25: CLASSICAL AND ALTERNATIVE MODELS OF A REVERSIBLE COMPETITIVE INHIBITION (Segel I. 1975)

In majority of cases, (S) and (I) compete for binding to the active site of the enzyme **(1)**. This requires that (S) and (I) have a structural analogy. However, there are other mechanisms of competitive inhibition in which the binding sites of (S) and (I) are distinct:

- the binding of (I) on an effector site causes a steric hindrance preventing the binding of (S) on the catalytic site of the enzyme **(2)**.
- (S) and (I) have a group in common which bind to a third site of the enzyme **(3)**.
- the binding sites of (S) and (I) overlap each other **(4)**.
- the binding of (I) induces a conformational change in the enzyme that deforms or masks the catalytic site. (S) loses its affinity for this site and *vice versa* **(5)**. Many multimeric allosteric enzymes can be inhibited according to this later model.

In the presence of a competitive inhibitor, the apparent affinity of the enzyme decreases: *K*m is increased and Vmax is unchanged.

A reversible competitive inhibition can be represented by the Lineweaver-Burk double reciprocal plot (**Figure 26**).

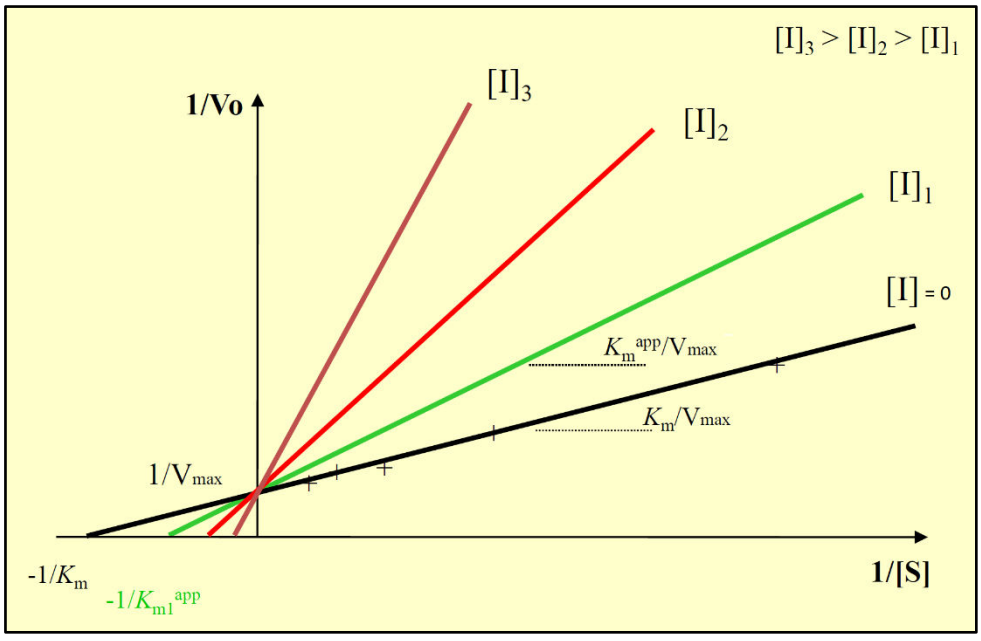


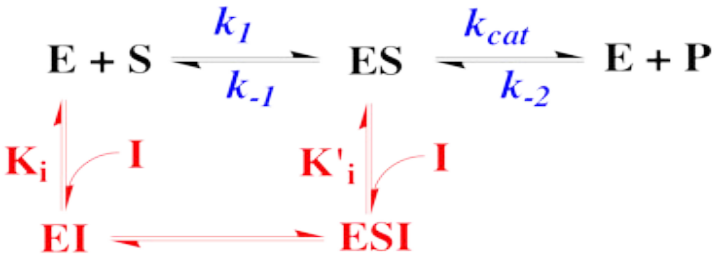
FIGURE 26: REPRESENTATION OF THE LINEWEAVER-BURK DOUBLE RECIPROCAL PLOT OF A REVERSIBLE COMPETITIVE-TYPE INHIBITION

The intercept of the lines on the y-axis corresponds to the value $1/V_{max}$. The maximum speed is not modified. For increasing concentrations of inhibitor, the intercept on the x-axis corresponds to the value of $-1/K_m^{app}$ which increases, so K_m^{app} increases. The representation K_m^{app} as a function of [I] makes it possible to determine the inhibition constant *K*i which corresponds to the absolute value of the point of intersection of the line with the x-axis.

Competitive-type inhibition may be “leavened” using a saturating concentration of substrate. Indeed, since the binding of the substrate and that of the inhibitor are mutually exclusive, the addition of a high concentration of substrate displaces the equilibrium E + S ⇆ ES in favour of ES; the equilibrium EI ⇆ E + I is displaced in favour of E.

II.2.2. Mechanism of a reversible non-competitive inhibition

Non-competitive inhibitors do not exhibit structural homology with the substrate. A classical non-competitive inhibitor has no influence on the binding of the substrate (and *vice versa*); the binding sites of the substrate and the inhibitor are distinct. As a result, the inhibitor can bind to the free enzyme (E) or the complex (ES). In the same way, the substrate can bind to the free enzyme (E) or the complex (EI).
The reaction mechanism of a non-competitive inhibition can be expressed according to the model:



Conventionally, during non-competitive inhibition, the binding of the inhibitor does not alter the way the substrate binds, but it prevents the conformational adjustments of the active site that must occur for catalysis. Thus, the ternary complex (ESI) is inactive.
A reversible non-competitive inhibition can be schematized as shown in figure 27.

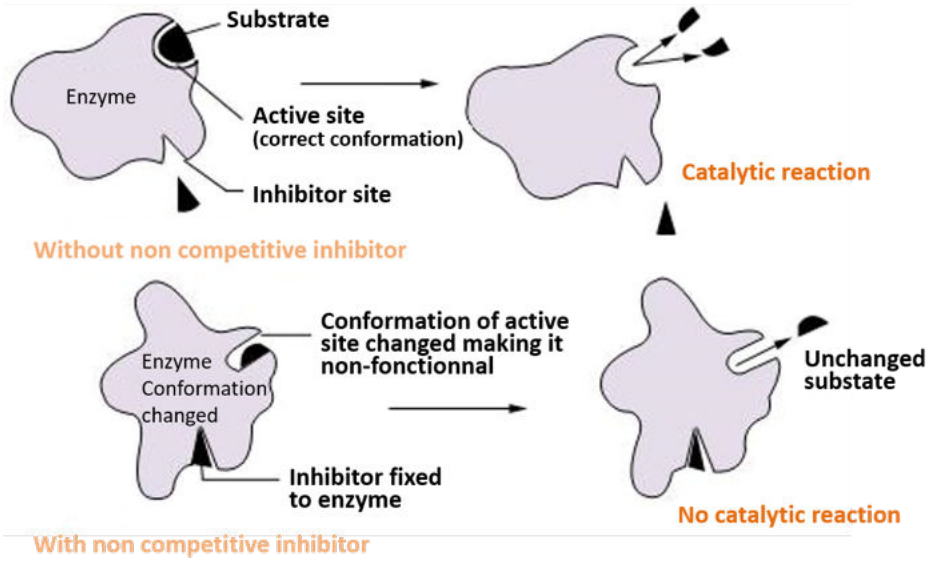


FIGURE 27: SCHEMATIC REPRESENTATION OF A REVERSIBLE NON-COMPETITIVE INHIBITION

In the presence of a non-competitive inhibitor, a portion of the enzyme molecules are in the form (EI) and (ESI), together with the (ES) complex; so there are fewer active enzyme molecules and Vmax decreases. On the other hand, since the binding sites of (I) and (S) are distinct, the saturation of the free enzyme molecules by the substrate is not modified: *K*m is not modified.

A reversible non-competitive inhibition can be represented by the Lineweaver-Burk double reciprocal plot (**Figure 28**).

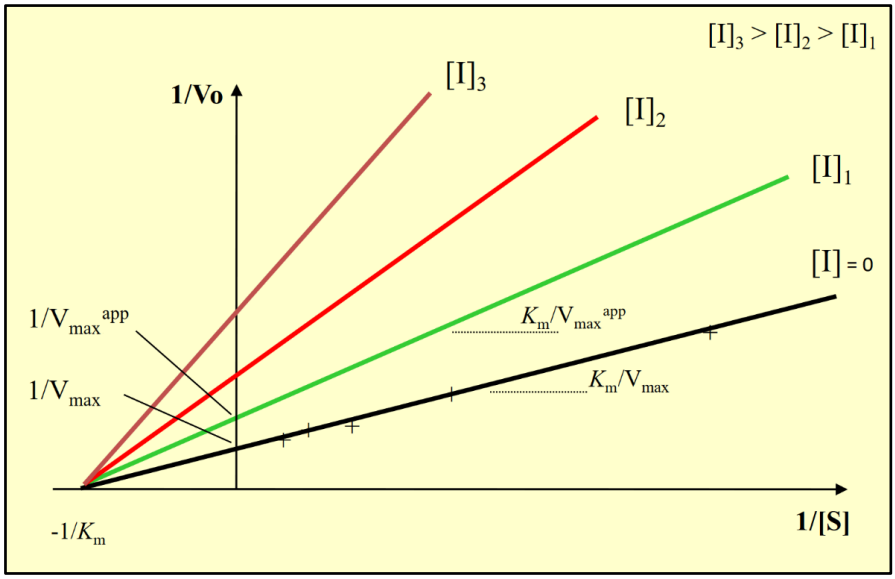


FIGURE 29: REPRESENTATION OF THE LINEWEAVER-BURK DOUBLE RECIPROCAL PLOT OF A REVERSIBLE NON COMPETITIVE-TYPE INHIBITION

For increasing inhibitor concentrations, the intercept of the straight lines on the y-axis corresponds to the value $1/V_{max}^{app}$ which increases; so V_{max}^{app} decreases. The intercept of the straight lines on the x-axis corresponds to $-1/K_M$; K_M is not changed. The representation $1/V_{max}^{app}$ as a function of [I] makes it possible to determine the inhibition constant *K*i which corresponds to the absolute value of the intercept of the line with the x-axis.
This inhibition can not be leavened in the presence of high substrate concentrations.

In practice, during this work and for each purified recombinant protein, the various inhibition tests were carried out according to the following protocol. Constant concentrations of Trypsin (5 nM) are pre-incubated with increasing concentrations of recombinant protein (0-9.4 µM) in buffer (100 mM Tris-HCl pH 8.3, 100 mM $CaCl_2$) for 15 minutes at 25°C. Different concentrations (12.5-200 µM) of fluorescent substrate (Z-Phe-Arg7-amido-4-methylcoumarin hydrochloride) are added. After excitation of the substrate ($\lambda_{excitation}$ = 355 nm), the fluorescence emitted by the fluorescent product ($\lambda_{emission}$ = 460 nm) and corresponding to the residual activity of the trypsin is measured by spectrofluorometry every 20 seconds for 12 minutes.

In conclusion, GASP-2 and GASP-1 have anti-trypsin properties. However, although these two proteins are closely structural related, there is a difference in specificity between them. Indeed, our work reveals that GASP-1 inhibits trypsin according to a non-competitive mechanism while GASP-2 according to a competitive mechanism. In order to determine whether the difference of specificity is only due to the second Kunitz domain or to the molecular environment in each of the two proteins, we have carried out the construction and the production of two chimeric proteins in which only these two domains have been interchanged. We identified by site-directed mutagenesis the amino acids responsible for this functional duality. **All the results obtained during this study allowed the writing of a publication which is submitted to *PLOS Biology*.**

GASP-1 and GASP-2, two closely structural related proteins with a functional duality in anti-trypsin inhibition specificity: a mechanistic point of view.

Parenté A. *et al.*

INTRODUCTION

The Growth and differentiation factor Associated Serum Proteins (GASP-1 and GASP-2) are two closely related multi-domain glyco-proteins, discovered in the early 2000s and initially named WFIKKN proteins with respect to their structural composition made of a Whey acidic protein domain (W), a Follistatin/Kazal (FK) domain, an Immunoglobulin-related (I) domain, two tandem Kunitz (K) modules and a Netrin (N) module (1,2). With the exception of the immunoglobulin and follistatin domains, all others are associated with protease inhibitors. This suggests that these proteins could act as multivalent inhibitors of proteases (1). At the fetal stage, GASP-1 is mainly expressed in the skeletal muscle, brain, kidney and thymus. GASP-2 is highly expressed in the lung and to a lower extent in the skeletal muscle and liver. At the adult stage, GASP-1 is mainly expressed in the ovaries, testes and pancreas whereas GASP-2 is mainly expressed in the pancreas, liver, thymus and lung (1–3). Despite globally distinct expression patterns, their common high level of expression in the pancreas, which secretes a variety of proteases, as trypsin, reinforces the putative antiprotease function of the GASP proteins.

Follistatin-type proteins as GASPs are factors that function by binding and/or inhibiting members of the TGF-β superfamily (Transforming Growth Factor β) including the activin, TGF-βs, BMPs (Bone Morphogenetic Proteins) and GDFs (Growth Differentiation Factor) subfamilies (4–6). If GASP proteins play a role of chaperone for some TGF-βs members, they have both chaperone and inhibitor functions for GDF-11 and GDF-8 (myostatin) (7–9). By co-immunoprecipitation from murine serum, a direct interaction between GASP-1 and myostatin has been shown, which results in inactivation of myostatin signaling pathway (3). GASP-2 was also shown to exhibit high affinity to myostatin, acting as an inhibitor *in vitro* (8). This myostatin inhibition has also been demonstrated *in vivo* using different transgenic models of invalidation or overexpression of GASPs. As myostatin is a powerful negative regulator of muscle growth, GASP-1 or GASP-2 murine knockout leads to muscle atrophy (10), while reversely, GASP-1 overexpression leads to an increase in muscle mass (11–13). Furthermore,

submitted to PLOS Biology

overexpression of GASP-1 or GASP-2 promotes proliferation and differentiation of myoblastic cells by inhibiting the myostatin pathway (14,15).

To date, most studies on these two GASP genes concern their function during myogenesis and not their potential role as protease inhibitor. Although present in the resource for peptidases and their inhibitors (MEROPS database), GASP proteins are classified as putative protease inhibitors (16). GASPs are among the rare heterotopic compound inhibitors together with the Eppin proteins (17). Under physiological conditions, a balance exists between proteases and their inhibitors regulating many mechanisms such as digestion, homeostasis inflammatory response, reproduction or coagulation (18). Proteases are divided into four major families: serine proteases, cysteine proteases, aspartyl proteases and metalloproteases. This classification considers the amino acid composition of their catalytic site (serine, aspartate or cysteine) or the use of metal ions which lead to different mechanisms of action (19). The protease inhibitors are grouped into several families according to their sequence homology with "inhibitory units" defined as an amino acid sequence containing a single reactive site carrying the inhibitory activity (20). Several domains present in GASP proteins are frequently involved in the inhibition of proteases. Many studies have shown that the WAP (Whey Acidic Protein), Kazal or Kunitz domains are involved in the inhibition of serine proteases. In addition, the WAP domain can act on cysteine proteases (21). The Netrin domain is associated with the inhibition of metalloproteases, as found in the TIMPs (Tissue Inhibitor of MetalloProteinases) inhibiting MMPs (Matrix MetalloProteinases) or sFRPs (secreted Frizzled-Related Proteins), involved in regulation of the Wnt pathway (Wingless/Integrated) (22–24).

This knowledge has suggested that GASP-1 and GASP-2 are multivalent protease inhibitors. To date, only the second Kunitz domain of human, GASP-2, has been described as capable of inhibiting trypsin, the most extensively studied serine protease. No effect on the proteasic activities of the elastase, chymotrypsin, tissue-type plasminogen activator, urikinase-type plasminogen activator, has been showed (25,26). No kinetics of enzyme inhibition has been reported neither for GASP-2 nor for GASP-1.

submitted to PLOS Biology

We investigated the anti-trypsin activity of the full-length GASP-1 and GASP-2 proteins and we determined the role of the second Kunitz (Ku_2) domain in this activity. Using a bacterial expression system, several native and chimeric murine GASPs (*m*GASP) proteins were produced to perform the *in vitro* enzymatic activities. We showed for the first time that the full-length *m*GASP-1 protein has a non-competitive anti-trypsin activity whereas the *m*GASP-2 protein presents a competitive anti-trypsin activity. The production of two chimeric proteins, *m*GASP-1/$Ku_{2\text{-}2}$ and *m*GASP-2/$Ku_{2\text{-}1}$, in which only the second Kunitz (Ku_2) domain has been interchanged, showed that the difference in inhibition specificity is only attributed to the second Kunitz domain. Molecular models and molecular dynamic simulations supported this biological data, rationalizing how the second Kunitz domains of GASP-1 and GASP-2 were implicated in the anti-trypsin inhibition specificity, based on the description of the highly dynamic modes of binding.

MATERIALS AND METHODS

Construction of expression vectors for different murine recombinant GASPs.

Wildtype recombinant mGASP-1 and mGASP-2: expression plasmids encoding full-length mature wild-type mGASP-1 and *m*GASP-2 excluding the signal peptide were constructed in pGEX-4T1 plasmid vector (Sigma-Aldrich, Saint Louis, MO) as previously described (15).

Chimeric mGASP-1/$Ku_{2\text{-}2}$ and GASP-2/$Ku_{2\text{-}1}$: two other recombinant vectors encoding two chimeric GASP proteins, *m*GASP-1/$Ku_{2\text{-}2}$ and *m*GASP-2/$Ku_{2\text{-}1}$, in which the second Kunitz (Ku_2) domain was interchanged. *m*GASP-1/$Ku_{2\text{-}2}$ corresponds to the *m*GASP-1 protein with the *m*GASP-2 second Kunitz domain and the reciprocal protein *m*GASP-2/$Ku_{2\text{-}1}$ includes the second Kunitz domain of *m*GASP-1. Briefly, for each construction, we amplified two independent PCR fragments using the Q5® High-Fidelity DNA polymerase (New England Biolabs): one corresponding to the plasmid vector pGEX-4T1 with the sequence encoding the GASP protein without its second Kunitz domain and the other corresponding to the second Kunitz domain of the reciprocal GASP protein. Primers and PCR conditions are given in **Table 1**. We designed each set of primers with an overlap region using the primer analysis

submitted to PLOS Biology

105 software (New England Biolabs). After purification using the QIAquick Gel Extraction kit (Qiagen,
106 Hilden, Germany), the amplified fragment concentration was measured with the NanoDrop® ND-1000
107 spectrophotometer (Thermo Scientific, Wilmington, DE). For each construct, successful assembly of
108 the two PCR fragments was performed using the NEBuilder HiFi DNA Assembly Master Mix (New
109 England Biolabs, Inc., Ipswich, MA) according to the manufacturer's instructions.

110 **Expression and purification of recombinant *m*GASP proteins.**

111 The expression vectors encoding the murine recombinant proteins, *m*GASP-1 and 2, *m*GASP-1/Ku_{2-2}
112 and GASP-2/Ku_{2-1} were transformed in the *E. coli* BL21 strain competent cells (New England Biolabs,
113 Beverly, MA) in LB medium supplemented with 100 µg.mL^{-1} ampicillin and 0.2% glucose. After
114 induction by adding 0.025 mM Isopropyl β-D-1-thiogalactopyranoside for 16 h at 20°C, bacterial
115 growth was stopped 15 min at 4°C before centrifugation at 6000xg during 15 min at 4°C. Pellet was
116 resuspended in lysis buffer (20 mM Tris-HCl, pH 7.5, 250 mM NaCl, 1 mM EDTA, 0.05% Triton X-100,
117 10 µg.mL^{-1} lysozyme). Recombinant proteins contained a GST tag at the N-terminus which allows
118 affinity purification with the ÄKTAprime system (GE Healthcare Bio-Science, Uppsala, Sweden). Briefly,
119 supernatant was loaded into a GST-Trap HP 1ml column (GE Healthcare) equilibrated in 20 mM Sodium
120 Phosphate pH 7.3, 0.15 M NaCl. Elution was performed at a 1 mL.min^{-1} rate with 50 mM Tris-HCl pH
121 8.0, 10 mM reduced glutathion. Cleavage of the GST tag was performed with 10 units.mg^{-1} of
122 recombinant thrombin (Sigma-Aldrich) by incubation for 16 h at 4°C. Then, samples were concentrated
123 by centrifugation at 4000xg, 4°C on an Amicon Ultra 30K filter (Millipore). The protein concentration
124 was determined by the Bradford method (27). Eluted fractions were analyzed by SDS/PAGE with
125 Coomassie blue staining and by Western blot.

126 **SDS/PAGE and Western blot analyses.**

127 SDS/PAGE was performed as described previously (28) under reducing conditions on 10% acrylamide
128 separating gels. Proteins were analyzed in reducing loading buffer (2% SDS, 20% glycerol, 100 mM Tris-
129 HCl pH 6.8, 0.1% bromophenol blue, 5% β-mercaptoethanol). Molecular masses were estimated using

submitted to PLOS Biology

the Precision Plus Protein Standards calibration kit (Biorad, Hercules, CA). Proteins were revealed with 0.25% Coomassie Brillant R-250 solution. For Western blot analyses, separated proteins were then transferred onto a PVDF Western blotting membrane (Roche Diagnostics, Mannheim, Germany) and electroblotted for 1h30 at 200 mA. After 1h of saturation in TBS (Tris-Buffered Saline) with 0.1% Tween 0.2% and 5% non-fat dry milk, the membrane was first incubated overnight at 4°C under agitation with primary antibody diluted in 2% non-fat dry milk. After 4 washes of 15 min each in TBS-Tween 0.1% (v/v), membrane was incubated for 1 h at room temperature with a second antibody conjugated to horseradish peroxidase. After 4 more washes in TBS-Tween 20% (v/v), the immunoblot was processed by chemiluminescence detection (Chemiluminescence Blotting Substrate (POD), Roche Molecular Biochemicals, Mannheim, Germany). The membranes were analyzed using ImageQuant TL software (GE Healthcare). Commercial antibodies used are listed in **Table 2**.

Anti-trypsin activity assays.

Titration of bovine pancreatic trypsin (Sigma Aldrich) was carried out using 4-nitrophenyl p-guanidinobenzoate (Sigma Aldrich) (29). The activity was determined by using the fluorescent substrate Z-Phe-Arg-NHMec (Sigma Aldrich). The kinetic parameters and the value of the equilibrium constant were determined by incubating trypsin (10 nM final concentration) and inhibitor (0 - 10 µM final concentration) in 100 mM Tris / HCl buffer, pH 8.3, containing 100 mM $CaCl_2$ for 15 min at 25 °C, after which substrate (6.25 - 200 µM final concentration) was added. The Michaelis-Menten constant (K_m) and the apparent K_m were determined by the Lineweaver-Burk method using the reciprocal values of substrate concentration (1/[S]) and initial velocity ($1/V_0$) to obtain straight-line graphs through linear regressions, based on the kinetic equation: $1/V_0 = 1/V_{max} + K_m/V_{max} \times 1/[S]$, where V_{max} is the maximum reaction velocity. For a competitive inhibitor, the dissociation constant (K_i) of the trypsin-inhibitor complex was determined from the replot of the the apparent K_m values *vs.* the inhibitor concentration [I], in agreement with the following linear relationship: $K_{m}{}_{app} = K_m + K_m/K_i \times [I]$. For a non-competitive inhibitor, the dissociation constant (K_i) of the trypsin-inhibitor complex was determined from the

submitted to PLOS Biology

replot of the inverse of the apparent V_{max} values *vs*. the inhibitor concentration [I], in agreement with the following linear relationship: $1/V_{max app} = 1/V_{max} + 1/(V_{max} \times K_i) \times [I]$.

Proliferation assays.

C2C12 mouse myoblast cells (30) were obtained from American Type Culture Collection (ATCC-CRL-1772). Myoblasts were maintained at 37°C in Growth Medium consisting in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen-Life Technologies, Saint Aubin, France) supplemented with 10% heat-inactivated Fetal Bovine Serum (FBS) (Invitrogen-Life Technologies), 100 units.mL^{-1} penicillin and 100 µg.mL^{-1} streptomycin (Invitrogen-Life Technologies). To induce C2C12 differentiation, cells at 70% confluence were shifted to DMEM supplemented with 2% horse serum (Invitrogen-Life Technologies). Test of proliferation was assessed as previously described (31). For proliferation assays, cells were grown with or without 1 µg.mL^{-1} of each recombinant GASP protein. Three independent experiments were carried out. Each experiment corresponds to the analysis of 96 wells with 2,000 cells per well at t = 0 h. The CellTiter 96® Aqueous One Solution Cell Proliferation Assay (Promega, Madison, WI) was used during a period of 72 h to determine cell proliferation. 20 µL of MTS solution were added to the cells during 3 h. Absorbance at 490 nm of the formazan product from bioreduced MTS was read using an ELISA plate reader.

C2C12 differentiation kinetic.

Myoblast cells were grown to 70% confluence and were differentiated into myotubes after shifting growth medium in differentiation medium supplemented or not by 1 µg.mL^{-1} of GASP-1/$K_{2\text{-}2}$ or GASP-2/$K_{2\text{-}1}$ for 96 h. For each kinetic point analyzed, cells were fixed during 15 min in 4% paraformaldehyde, washed three times in PBS 1X and permeabilized with HEPES/Triton X-100 buffer (20 mM HEPES, 300 mM sucrose, 50 mM $NaCl_2$, 3 mM $MgCl_2$ and 0.5% Triton X-100 pH7.4). Then, cells were blocked for 1 h at room temperature in blocking buffer (10% goat serum, 1% bovine serum albumin (BSA) and 0.1% Triton X-100 in PBS). After a quick wash in PBS 1X containing 0,2% BSA, cells were incubated overnight at 4°C with the primary antibody anti-MyHC in PBS 1X containing 1% BSA. After 2 washes in BSA 0.2%,

submitted to PLOS Biology

Tween-20 0.1%, PBS 1X, cells were incubated for 15 min at 37°C with the secondary antibody diluted in BSA 1%, PBS 1X. The staining was completed with 3 washes in BSA 0.2%, PBS 1X, Tween-20 0.1% following by an incubation for 5 min at room temperature in DAPI (1mg.mL^{-1} diluted in PBS 1X) and 3 washes in PBS 1X. Images were acquired and analyzed with the automated Leica DMI6000B inverted epifluorescence microscope using the MetaMorph software (Molecular Devices, Sunnyvale, USA). The fusion index of C2C12 cells was calculated as the ratio of the number of nuclei in myotube to the total number of nuclei.

Construction of the Kunitz domains.

The murine GASP1 and GASP2 Kunitz domains (*m*Ku$_2^{GASP1}$ and *m*Ku$_2^{GASP2}$, respectively) were obtained from the "Ensembl" database (ENSMUSG00000044177 and ENSMUSG00000071192). Both *m*Ku$_2^{GASP1}$ and *m*Ku$_2^{GASP2}$ domains exhibit high sequence identity with the GASP2 human Kunitz domain *h*Ku$_2$ (88.2 and 64.7% identity, respectively, see **S1 Table**), for which the 3D structure was resolved in solution by NMR (PDB ID: 2DDJ) (26). The two domains were constructed by protein threading (an advanced homology modeling technique) using the I-Tasser webserver in which 2DDJ was used as a template (32–34). The protonation states of ^{69}His and ^{36}His were set as a ε-protonated state according to their neighboring chemical environment, *i.e.*, maximizing hydrogen bonding. The protonation states of titrable aminoacids (*i.e.*, aspartates, glutamates, arginines, lysines and threonines) were defined by PROPKA v3.0 (35,36). Expected disulfide bonds were manually constructed between facing cysteines, namely ^{10}Cys-^{60}Cys, ^{19}Cys-^{43}Cys and ^{35}Cys-^{56}Cys in both *m*Ku$_2^{GASP1}$ and *m*Ku$_2^{GASP2}$.

A (*N,P,T*) 200 ns molecular dynamics (MD) simulation was carried out for both *m*Ku$_2^{GASP1}$ and *m*Ku$_2^{GASP2}$ protein domains, solvated in a box of explicit water molecules (see below for details). This MD simulation corrected the structural errors associated with the protein threading construction, *i.e.*, correction of the secondary and tertiary structures. To check if the relaxation was sufficient, root mean square deviation (RMSD) analysis was achieved to check convergence. Five structures were extracted

submitted to PLOS Biology

from the last 50 ns to assess the binding to bovine trypsin (*b*Trypsin), which have constituted a set of five replicas in this study.

Description of the mode of binding to trypsin.

The binding between the Kunitz domains and *b*Trypsin was assessed from existing 3D structures, namely PDB ID: 5MNF and PDB ID: 2RA3. The former is a bovine cationic *b*Trypsin structure, which was recently elucidated by X-ray with a 0.99 Å resolution (37). The latter is the structure of human cationic trypsin (*h*Trypsin) complexed with bovine pancreatic trypsin inhibitor (BPTI), which was elucidated by X-ray at a resolution of 1.46 Å (38); this complex has been referred to [*h*Trypsin-BPTI] throughout this text. Considering the high structural similarity between the Kunitz domains and BPTI (**S1a Fig**) as well as between *h*Trypsin and *b*Trypsin (74,11 % identity, see **S1b Fig**), the three [*b*Trypsin-BPTI], [*b*Trypsin-*m*Ku_2^{GASP1}] and [*b*Trypsin-*m*Ku_2^{GASP2}] complex 3D-models were constructed. [*b*Trypsin-BPTI] was obtained by combining *b*Trypsin and [*h*Trypsin-BPTI] as initial templates. The [*b*Trypsin-BPTI] model served to construct five replicas of [*b*Trypsin-*m*Ku_2^{GASP1}] and [*b*Trypsin-*m*Ku_2^{GASP2}] (**S2 Fig**). The four complexes were solvated in boxes of water molecules (see below for details). They were minimized and equilibrated prior to (*N,P,T*) 200 ns MD simulations. The [*b*Trypsin-*m*Ku_2^{GASP1}] and [*b*Trypsin-*m*Ku_2^{GASP2}] models were directly dedicated to this study, while [*h*Trypsin-BPTI] and [*b*Trypsin-BPTI] were used as references for the sake of comparison in the binding regions.

MD simulations methodology.

Prior to MD simulations, all systems were solvated in water and they were neutralized using NaCl ([NaCl] = 0.154 M). The TIP3P explicit water model (39) and the corresponding ion parameters were used (40,41). The Amber protein FF14SB forcefield was used to describe the proteins (42). Each system was relaxed by minimization first of the water and ion system, then of the modes involving heavy atoms only, and finally of the whole system. Thermalization from 0 to 320 K was then achieved during a 250 ps (*N,V,T*) MD simulation, using a 0.5 fs integration time step. Pressure and density were then equilibrated during a 1 ns (*N,P,T*) MD simulations at 298.15 K, with a 1 fs integration time step

submitted to PLOS Biology

maintaining pressure at 1 atm. Finally, a 300 ns (*N,P,T*) MD simulations (298.15 K, 1 atm) was carried out with a 2 fs integration time step. The temperature was maintained using Langevin Dynamics with a damping coefficient of 1.0 ps^{-1}. Constant pressure was maintained using the Berendsen barostat (43), in which pressure relaxation time was set to 1 ps. Non-covalent interactions were explicitly calculated using a cutoff distance of 10 Å for both short Coulomb and van der Waals interactions. Long-range interactions were treated using the particle-mesh Ewald method with a 1 Å grid size (44–47). Bonds involving H-atoms were constrained using the SHAKE algorithm. Box sizes and component information are reported in **S2 Table**.

Input files were prepared using the AmberTools17 package. Simulations were performed using the GPU-version of AMBER PMEMD code (Amber16) (48). Analyses were performed using the CPPTRAJ software (49). The VMD 1.9.3 software (50–53) was used to visualize trajectories and to render pictures.

RESULTS AND DISCUSSION

Previous data have shown that the second Kunitz-type protease inhibitor domain of the human protein GASP-2 (hKu_2^{GASP2}) was found to inhibit trypsin with a Ki value of 9.6 nM (25). However, no experimental data have been reported concerning the antiprotease activity of the full-length GASP-2 and/or GASP-1. Although GASP proteins have *N*- and *O*- glycosylations, previous studies have shown that these oligosaccharidic chains are not essential for the inhibition of the myostatin (14,15). Here, we have studied the inhibitory property of the two proteins expressed in *E. coli* as GST-fusion proteins. This bacterial expression system allows providing high protein yields, as required for anti-protease assays.

Production and purification of wild-type *m*GASP-1 and *m*GASP-2 proteins

Full-length mature *m*GASP-1 and *m*GASP-2, excluding the signal peptide were cloned in pGEX-4T-1 vector and expressed in *E. coli* BL21 as GST-tagged fusion proteins (**Fig 1A**). After production, the total bacterial proteins and the recombinant proteins purified on a GST-Trap HP column were

submitted to PLOS Biology

visualized on a reducing 10% SDS/PAGE. For the wild-type *m*GASP-1, a major band was detected with an apparent molecular weight of about 90 kDa, which is consistent with the expected molecular weight (86.3 kDa) of GST-GASP-1 fusion protein (**Fig 1B**, lane 2). Similar results were obtained for the wild-type *m*GASP-2 protein, with a band of approximately 90 kDa (expected molecular weight 86.5 kDa) (data not shown). To optimize subsequent activity tests on the recombinant GASP proteins, removal of the GST tag was necessary. After a thrombin digestion of 16h at 4°C, collected proteins were analyzed on reducing 10% SDS/PAGE. As seen in **Fig 1B**, lane 3, after the cleavage of the GST-GASP-1 fusion protein, a band of approximately 60 kDa (theoretical molecular mass of 60.63 kDa) was released. The 27 kDa protein was assigned to the GST. *m*GASP-1 identification was confirmed by western blot analysis (**Fig 1C**, lanes 1 and 2). For *m*GASP-2, similar results were obtained after thrombin treatment releasing a band of approximately 60 kDa corresponding to *m*GASP-2 and confirmed by Western blotting (**Fig 1B**, lane 4; **Fig 1C**, lanes 3 and 4).

Functional characterization of the wild-type mGASP-1 and mGASP-2 proteins

It has been shown that the second kunitz domain of the human GASP-2 is a competitive trypsin inhibitor (25). Therefore, we tested if the full-length *m*GASP-2 has the same properties. Kinetic parameters (K_m and V_{max}) were determined by fitting a linear function between the inverse of the reaction rate and the inverse of the substrate concentration, at different inhibitor concentrations. It appeared that V_{max} was constant and K_m increased with increasing amounts of *m*GASP-2, demonstrating that *m*GASP-2 is a competitive inhibitor of trypsin. Furthermore, the GASP-2/trypsin dissociation constant (K_i of 50 nM) indicated a strong affinity between GASP-2 and trypsin (**Fig 2B**), as reported for the isolated second domain Kunitz of the human GASP-2 (K_i of 9.6 nM). For *m*GASP-1, K_m was constant and V_{max} decreased with increasing amounts of inhibitor (**Fig 2A**), revealing that *m*GASP-1 is a non-competitive inhibitor of trypsin. This different inhibition mechanisms for both GASP-1 and GASP-2 were unexpected since the two proteins are structurally very close. To rationalize whether this difference can only be attributed to amino-acid composition, we produced two chimeric recombinant

submitted to PLOS Biology

proteins, *m*GASP-1/Ku_{2-2} and *m*GASP-2/Ku_{2-1}, in which only the second Kunitz domain was interchanged (**Fig 1A**).

Construction, expression and purification of the two recombinant chimeric proteins *m*GASP-1/Ku_{2-2} and *m*GASP-2/Ku_{2-1}

Full-length mature *m*GASP-1/Ku_{2-2} and *m*GASP-2/Ku_{2-1}, excluding the signal peptide, were constructed using the NEBuilder HiFi DNA Assembly Master Mix. For each chimeric GASP, this method allows the assembly of a large fragment corresponding to GASP without the second Kunitz domain and a little fragment encoding the second Kunitz domain of the other GASP. For the two chimeric proteins, only the second Kunitz domain was interchanged. No other amino acid of the native *m*GASP-1 and *m*GASP-2 was modified. The set of primers and PCR conditions are given in Table 1. *m*GASP-1/Ku_{2-2} and *m*GASP-2/Ku_{2-1} were produced in *E. coli*, purified on a GST-Trap HP column and cleaved by thrombin using the same experimental conditions as for *m*GASP-1 and *m*GASP-2. As seen in **Fig 1B**, thrombin cleavage released a band of approximately 60 kDa corresponding to both *m*GASP-1/Ku_{2-2} and *m*GASP-2/Ku_{2-1} and confirmed by western blotting (data not shown).

Effect of the two chimeric proteins *m*GASP-1/Ku_{2-2} and *m*GASP-2/Ku_{2-1} on proliferation and differentiation of C2C12 myoblast cells

As previously described, GASP-1 and GASP-2 have been shown to bind myostatin and are considered as inhibitors of TGFβ signaling pathway (7). These two proteins might act as regulators of myogenic proliferation and differentiation. It was demonstrated that *m*GASP-1 produced in eukaryotic system and devoid of *N*- linked glycosylation (14) or *m*GASP-2 produced in prokaryotic system (15), enhanced C2C12 proliferation and differentiation by preventing myostatin signaling. Considering this function, we supposed that chimeric *m*GASP-1/Ku_{2-2} and *m*GASP-2/Ku_{2-1} would conserve their function to activate myogenesis. To test this hypothesis, C2C12 myoblast cells were first treated with 1 $\mu g.mL^{-1}$ of *m*GASP-1/Ku_{2-2} or *m*GASP-2/Ku_{2-1} for 72 h. The treatment to *m*GASP-1/Ku_{2-2} or *m*GASP-2/Ku_{2-1} significantly improved the myoblast proliferation rate compared to control cells (**Fig 3A**). In a second

submitted to PLOS Biology

304 step, we analyzed the effect of *m*GASP-1/$Ku_{2\text{-}2}$ or *m*GASP-2/$Ku_{2\text{-}1}$ during C2C12 myoblast
305 differentiation. As for proliferation, we treated C2C12 myoblast cells at 70% confluence with 1 µg.mL^{-1}
306 of recombinant protein and we followed the cell fusion for 96 hours (**Fig 3B** and **3C**). We showed that
307 the two chimeric proteins enhanced C2C12 differentiation by increasing the fusion index at 96 hours.

308 **Implication of the second Kunitz domain in the differential anti-trypsin activity of GASP-1 and**
309 **GASP- 2**

310 We studied the effect(s) of interchanging the second Kunitz domain between the two proteins
311 on the inhibition of the proteolytic action of trypsin. For each purified chimeric protein, kinetic
312 parameters were determined using the same conditions as described for both recombinant GASP-1
313 and GASP-2. *m*GASP-1/$Ku_{2\text{-}2}$ was a competitive inhibitor of trypsin, exhibiting a constant V_{max} value and
314 K_m values decreasing *vs*. increasing inhibitor concentration (**Fig 4A**). Following the same approach, the
315 chimeric *m*GASP-2/$Ku_{2\text{-}1}$ behaved as a non-competitive inhibitor of trypsin, exhibiting a constant K_m
316 value and V_{max} values decreasing *vs*. increase of *m*GASP-2/$Ku_{2\text{-}1}$ concentration (**Fig 4B**). As for GASP-1
317 and GASP-2, we measured inhibitor/trypsin dissociation K_i values of 220 nM and 290 nM for mGASP-
318 1/$Ku_{2\text{-}2}$ and mGASP-2/$Ku_{2\text{-}1}$, respectively. K_i values are of the same order of magnitude as those
319 obtained for the two recombinant GASP-1 and GASP-2 proteins. Our data showed that the anti-trypsin
320 activity of GASP proteins is attributed to the second Kunitz domain. The amino acid composition and /
321 or conformation of this domain is likely responsible for the specificity of this inhibition.

322 **Structural features of the *m*Ku_2^{GASP1} and *m*Ku_2^{GASP2} domains**

323 As expected from the high sequence identity between *m*Ku_2^{GASP1}, *m*Ku_2^{GASP2}, BPTI and *h*Ku_2 (**S1**
324 **Fig** and **S1 Table**), *m*Ku_2^{GASP1} and *m*Ku_2^{GASP2} exhibited similar topology than BPTI and *h*Ku_2 all along the
325 MD simulations. The structures are made of two main domains, namely the scaffold domain and the
326 canonical loop (54–56). The former (scaffold) domain consists of two antiparallel β-sheets (β_1 and β_2)
327 and one α-helix (**Fig 5**), which are conserved among many known Kunitz domains (*e.g.*, in human
328 amyloid precursor protein inhibitor APPI (57,58) or in bikunin (59)). As in BPTI, the two β-sheets of

submitted to PLOS Biology

329 *m*Ku$_2^{GASP1}$ and *m*Ku$_2^{GASP2}$ were maintained close to each other by a strong H-bonding network made of
330 six and seven H-bonds between the amide moieties of the backbones, respectively (**S3 Table**).

331 The latter domain (canonical loop) is also known as the binding loop to trypsin. It is made of
332 residues that are defined as the $P_{5\text{-}1}$ and $P_{1'\text{-}3'}$ residues (**Fig 5**), in which trypsin activity (hydrolysis)
333 targets the P_1-$P_{1'}$ peptide bond (56). Despite high sequence identity, several differences exist between
334 *m*Ku$_2^{GASP1}$, *m*Ku$_2^{GASP2}$, *h*Ku$_2$ and BPTI, which are reported in **Fig 5 and S1 Table**. The major difference
335 concerns the P_1 residue which is a glutamine in *m*Ku$_2^{GASP2}$ whereas it is cationic in the other three
336 (either lysine in BPTI and *m*Ku$_2^{GASP1}$ or arginine in *h*Ku$_2$). In conventional trypsin inhibitors, the region
337 defined by these residues is called Lys/Arg-Xaa, since P_1 is often a lysine or an arginine residue while
338 $P_{1'}$ stands for either alanine or glycine residue. It is worth noting that the cationic P_1 was described as
339 a key residue in the endopeptidase activity of tryspin (56,58).

340 MD simulations allowed to investigate the structural flexibility of the canonical loops of
341 *m*Ku$_2^{GASP1}$ and *m*Ku$_2^{GASP2}$, which was analyzed by the atomic root-mean-square fluctuation (RMSF) of
342 backbone atoms (namely C_α, N, O and C, see **S3 Fig**) and by the Ramachandran (ϕ,ψ-angle) plots (**S4**
343 **Fig** and **S4 Table**). The $P_{5\text{-}2}$ residues in *m*Ku$_2^{GASP1}$ exhibit a greater flexibility than in *m*Ku$_2^{GASP2}$ (**S4 Fig**),
344 two distinct rotamers were seen for P_5 and P_4 in *m*Ku$_2^{GASP1}$. As expected, the ϕ,ψ-angles of P_1 in
345 *m*Ku$_2^{GASP2}$ differ from that of the other three trypsin inhibitors due to different residue types (*i.e.*,
346 glutamine with respect to lysine or arginine).

347 **Non-covalent interactions between BPTI and bovine/human trypsins**

348 Interactions between trypsin and their their inhibitors have been extensively investigated over
349 the past decades (55,56,58,60). *h*Trypsin and *b*Trypsin are arranged into two six-stranded β-barrels,
350 mainly constituted of the catalytic site and the substrate-recognition domains. The catalytic site is
351 constituted of the catalytic triad (made of ^{63}His, ^{200}Ser and ^{107}Asp in *b*Trypsin), and the oxyanion hole
352 (^{198}Gly and ^{195}Ser in *b*Trypsin), stabilizing tetrahedral intermediates (**Fig 6**) (56). The substrate-
353 recognition domain is defined by the $S_{n\text{-}1}$ and $S_{1'\text{-}m'}$ sub-domains, where specific non-covalent contacts

submitted to PLOS Biology

with the substrate are formed. S_1 is the most important site since it features the trypsin-substrate interactions. It is defined by three different zones, namely A, B and C for the ^{194}Asp-^{197}Gln, ^{215}Ser-^{217}Gly and ^{225}Lys-^{229}Tyr regions, respectively (see **Fig 6**). After MD relaxation, the interactions between BPTI and *b*Trypsin observed in the constructed [*b*Trypsin-BPTI] model were very similar than those between BPTI and *h*Trypsin, in [*h*Trypsin-BPTI].

In agreement with a previous study (58), our MD simulations underline the key role of P_1 and $P_{2'}$ in [*b*Trypsin-BPTI] complex stabilization (**Fig 7**), the two sub-domains being involved in an H-bond network (**S5 Table**). P_1 binds the three S_1 zones and the catalytic site (**Table 3** and **Fig 7**) but in a dynamic way, mainly driven by the H-bonds which are swapping from one to another residue over time (**see S1 Movie**). For instance, the H-bonding interactions between P_1 and three trypsin backbone residues (NH group of ^{198}Gly or ^{200}Ser and CO group of ^{195}Ser) spend 71%, 35% and 61% of time, respectively, as calculated along the simulation time. P_1 is also involved in solvent-mediated H-bond bridges with S_1, either ^{194}Asp or ^{228}Val residues (*ca*. 69 % and 36 % of time, respectively).

Not only S_1 but also the $S_{1'}$-$S_{3'}$ domain participates in binding, however in an inhibitor dependent way. In the case of BPTI, $P_{2'}$ is directly involved in a strong H-bonding with $S_{2'}$ defined by ^{46}His and ^{47}Phe, as also observed in previous studies (55,56,58). Precisely, these H-bonds are formed either between (i) the guanidinium moiety of $P_{2'}$ (arginine) and the backbone CO group of ^{46}His of $S_{2'}$, or (ii) the backbone NH group of $P_{2'}$ and the backbone CO group of ^{47}Phe of $S_{2'}$; these two H-bonds spend 39% and 65% of time, respectively (**Table 3**).

H-bonds involving P_1 and $P_{2'}$ exhibit similar strengths as estimated by interatomic distances, *e.g.*, H-bond distances are in the range between 2.80 and 2.91 Å. To a lesser extent, P_2 (cysteine) and P_3 (proline) of BPTI are involved in the binding to trypsin, namely with ^{197}Gln of zone A and ^{217}Gly of zone B of S_1, respectively (**Table 3**).

Binding of murine Kunitz secondary domains to bovine trypsin

submitted to PLOS Biology

The residues of P_1 and $P_{2'}$ differ in both *m*Ku^{GASP1} and *m*Ku^{GASP2} from BPTI, which is a Lys/Arg-Xaa inhibitor. This is expected to affect their mode of binding with respect to this type of inhibitors. P_1 is a glutamine in *m*Ku^{GASP2} and $P_{2'}$ is aromatic in both *m*Ku^{GASP1} and *m*Ku^{GASP2} (namely tyrosine and tryptophan, respectively), which are no longer cationic thus lowering the electrostatic character of H-bond with trypsin residues if any.

As a first difference in binding, the [*b*Trypsin-*m*Ku$_2^{GASP1}$] and [*b*Trypsin-*m*Ku$_2^{GASP2}$] complexes show significant conformational differences with respect to [*b*Trypsin-BPTI] (**Fig 8**). While in [*b*Trypsin-BPTI], BPTI is strongly anchored to trypsin along MD simulation time, preventing rocking motion, both *m*Ku$_2^{GASP1}$ and *m*Ku$_2^{GASP2}$ significantly fluctuate more in the complexes (**Fig 8A**). This greater fluctuation of the *m*Ku$_2^{GASP1}$ and *m*Ku$_2^{GASP2}$ domains is well depicted by the higher variation of RMSD along time than what is seen for BPTI (**Fig 8B**). The following sequence in terms of flexibility is: *m*Ku$_2^{GASP1}$ > *m*Ku$_2^{GASP2}$ > BPTI. Similar conclusion is drawn from root-mean square fluctuation (RMSF) of the *m*Ku$_2^{GASP1}$ backbone compare to that of both BPTI and *m*Ku$_2^{GASP2}$ (**Fig 8C**). Such greater fluctuation of *m*Ku$_2^{GASP1}$, and to a lesser extent of *m*Ku$_2^{GASP2}$, compare to BPTI is attributed to conformational fluctuation of the side chains, especially in the β-sheet and α-helix domains (**Fig 8A**). This is likely correlated to a weaker binding to trypsin. Such a result has already been reported in the [*b*Trypsin-APPI] complex, showing greater RMSD variation than in [*b*Trypsin-BPTI] (58).

This clearly suggests that non-covalent interactions between trypsin and *m*Ku$_2^{GASP1}$ are weaker than between trypsin and both *m*Ku$_2^{GASP2}$ and BPTI. To rationalize this weaker binding, focus should be given to H-bonding interactions. Indeed, less H-bonds are seen between trypsin and *m*Ku^{GASP1} than between trypsin and *m*Ku^{GASP2} or BPTI, *i.e.*, 1.9, 4.5, and 4.1 in average, respectively (**S5 Table**). In *m*Ku^{GASP1}, P_1 is still involved in the binding to trypsin (1.2 H-bond, see **S5 Table**). The NH_3^+ group of P_1 still plays a key role, forming H-bonds that are swapping from zone A to zone B of S_1 (45%, 33%, 11% with the backbone CO group of ^{195}Ser and ^{194}Asp and backbone NH group of ^{217}Gly, respectively) and to the catalytic triad domain (with the CO group of ^{200}Ser). In *m*Ku^{GASP1}, only P_1 can contribute to the binding to trypsin, $P_{2'}$

submitted to PLOS Biology

being only very weakly involved (0.1 H-bonds, see **S5 Table**). This poor organization for binding rationalizes the non-competitive inhibition of trypsin by *m*Ku^{GASP1} (**Table 3** and **Fig 9**).

In *m*Ku^{GASP2}, P_1 is also involved in the binding to trypsin (1.5 H-bond, see **S5 Table**). This binding was unexpected, as in this case, the usual cationic P_1 is substituted by an electronegative polar glutamine, suggesting a different mode of binding as compare to BPTI or APPI. Here, glutamine mimics the mode of binding of BPTI by strongly interacting with the same residues (namely ^{195}Ser, ^{215}Ser and ^{198}Gly) as well as ^{196}Cys (**Fig 9B** and **Table 3**) by its two amide groups (two H-atom donors and two H-atom acceptors)(56). Additionally, *m*Ku^{GASP2} creates a strong H-bond network between trypsin and $P_{1'}$ and, to a lesser extent, P_2, $P_{2'}$, $P_{3'}$ (**Fig 9B** and **Table 3**). $P_{1'}$ strongly binds to the catalytic ^{195}Ser (through its OH group, 90%), while $P_{2'}$ (tryptophan) and $P_{3'}$ (glutamate) are two possible sites for binding to ^{66}Lys of trypsin. In *m*Ku^{GASP2} the local conformational organization suggests a new mechanism of trypsin inhibition, in which the reactive OH group of ^{200}Ser (56) is scavenged by the H-bond with $P_{1'}$. The very specific interaction between $P_{1'}$ and ^{195}Ser is thus expected to be responsible for the inhibition of trypsin activity in this case. This organization puts this group at 3.60 ± 0.69 and 4.30 ± 0.46 Å from the C=O group of ^{200}Ser and glutamine, respectively. Furthermore, the high inter-atomic distance between the OH-group of ^{200}Ser and the N-atoms of ^{63}His may also lead to decrease the reactivity at the active site. These findings rationalize the experimentally observed competitive inhibition of trypsin by the *m*Ku^{GASP2} domain.

submitted to PLOS Biology

1. Trexler M, Bányai L, Patthy L. A human protein containing multiple types of protease-inhibitory modules. Proc Natl Acad Sci USA. 2001 Mar 27;98(7):3705–9.

2. Trexler M, Bányai L, Patthy L. Distinct expression pattern of two related human proteins containing multiple types of protease-inhibitory modules. Biol Chem. 2002 Jan;383(1):223–8.

3. Hill JJ, Qiu Y, Hewick RM, Wolfman NM. Regulation of myostatin in vivo by growth and differentiation factor-associated serum protein-1: a novel protein with protease inhibitor and follistatin domains. Mol Endocrinol. 2003 Jun;17(6):1144–54.

4. Nakamura T, Takio K, Eto Y, Shibai H, Titani K, Sugino H. Activin-binding protein from rat ovary is follistatin. Science. 1990 Feb 16;247(4944):836–8.

5. Keutmann HT, Schneyer AL, Sidis Y. The Role of Follistatin Domains in Follistatin Biological Action. Mol Endocrinol. 2004 Jan 1;18(1):228–40.

6. Thompson TB, Lerch TF, Cook RW, Woodruff TK, Jardetzky TS. The Structure of the Follistatin:Activin Complex Reveals Antagonism of Both Type I and Type II Receptor Binding. Developmental Cell. 2005 Oct 1;9(4):535–43.

7. Szláma G, Kondás K, Trexler M, Patthy L. WFIKKN1 and WFIKKN2 bind growth factors TGFβ1, BMP2 and BMP4 but do not inhibit their signalling activity. FEBS J. 2010 Dec;277(24):5040–50.

8. Kondás K, Szláma G, Trexler M, Patthy L. Both WFIKKN1 and WFIKKN2 have high affinity for growth and differentiation factors 8 and 11. J Biol Chem. 2008 Aug 29;283(35):23677–84.

9. Monestier O, Blanquet V. WFIKKN1 and WFIKKN2: "Companion" proteins regulating TGFB activity. Cytokine Growth Factor Rev. 2016;32:75–84.

10. Lee Y-S, Lee S-J. Regulation of GDF-11 and myostatin activity by GASP-1 and GASP-2. Proc Natl Acad Sci USA. 2013 Sep 24;110(39):E3713-3722.

11. Monestier O, Brun C, Heu K, Passet B, Malhouroux M, Magnol L, et al. Ubiquitous Gasp1 overexpression in mice leads mainly to a hypermuscular phenotype. BMC Genomics. 2012 Oct 10;13:541.

12. Brun C, Périé L, Baraige F, Vernus B, Bonnieu A, Blanquet V. Absence of hyperplasia in Gasp-1 overexpressing mice is dependent on myostatin up-regulation. Cell Physiol Biochem. 2014;34(4):1241–59.

13. Périè L, Parenté A, Baraige F, Magnol L, Blanquet V. Alterations in Adiposity and Glucose Homeostasis in Adult Gasp-1 Overexpressing Mice. CPB. 2017;44(5):1896–911.

14. Brun C, Monestier O, Legardinier S, Maftah A, Blanquet V. Murine GASP-1 N-glycosylation is not essential for its activity on C2C12 myogenic cells but alters its secretion. Cell Physiol Biochem. 2012;30(3):791–804.

15. Pèrié L, Parenté A, Brun C, Magnol L, Pélissier P, Blanquet V. Enhancement of C2C12 myoblast proliferation and differentiation by GASP-2, a myostatin inhibitor. Biochem Biophys Rep. 2016 Mar 3;6:39–46.

16. Rawlings ND, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 2012 Jan;40(Database issue):D343-350.

17. Richardson RT, Sivashanmugam P, Hall SH, Hamil KG, Moore PA, Ruben SM, et al. Cloning and sequencing of human Eppin: a novel family of protease inhibitors expressed in the epididymis and testis. Gene. 2001 May 30;270(1–2):93–102.

18. van Eijk M, van Noorden CJF, de Groot C. Proteinases and their inhibitors in the immune system. Int Rev Cytol. 2003;222:197–236.

19. Rawlings ND, Barrett AJ. Evolutionary families of peptidases. Biochem J. 1993 Feb 15;290(Pt 1):205–18.

20. Rawlings ND, Tolle DP, Barrett AJ. Evolutionary families of peptidase inhibitors. Biochem J. 2004 Mar 15;378(Pt 3):705–16.

21. Chhikara N, Saraswat M, Tomar AK, Dey S, Singh S, Yadav S. Human Epididymis Protein-4 (HE-4): A Novel Cross-Class Protease Inhibitor. PLOS ONE. 2012 Nov 5;7(11):e47672.

submitted to PLOS Biology

22. Bányai L, Patthy L. The NTR module: domains of netrins, secreted frizzled related proteins, and type I procollagen C-proteinase enhancer protein are homologous with tissue inhibitors of metalloproteases. Protein Sci. 1999 Aug;8(8):1636–42.

23. Iyer RP, Patterson NL, Fields GB, Lindsey ML. The history of matrix metalloproteinases: milestones, myths, and misperceptions. Am J Physiol Heart Circ Physiol. 2012 Oct 15;303(8):H919-930.

24. von Maltzahn J, Chang NC, Bentzinger CF, Rudnicki MA. Wnt signaling in myogenesis. Trends Cell Biol. 2012 Nov;22(11):602–9.

25. Nagy A, Trexler M, Patthy L. Expression, purification and characterization of the second Kunitz-type protease inhibitor domain of the human WFIKKN protein. Eur J Biochem. 2003 May;270(9):2101–7.

26. Liepinsh E, Nagy A, Trexler M, Patthy L, Otting G. Second Kunitz-type protease inhibitor domain of the human WFIKKN1 protein. J Biomol NMR. 2006 May;35(1):73–8.

27. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 1976 May 7;72:248–54.

28. Laemmli UK. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature. 1970 Aug;227(5259):680–5.

29. Chase T, Shaw E. p-Nitrophenyl-p'-guanidinobenzoate HCl: a new active site titrant for trypsin. Biochem Biophys Res Commun. 1967 Nov 30;29(4):508–14.

30. Yaffe D, Saxel O. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. Nature. 1977 Dec 22;270(5639):725–7.

31. Oliver MH, Harrison NK, Bishop JE, Cole PJ, Laurent GJ. A rapid and convenient assay for counting cells cultured in microwell plates: application for assessment of growth factors. J Cell Sci. 1989 Mar;92 (Pt 3):513–8.

32. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008 Jan 23;9(1):40.

33. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols. 2010 Apr;5(4):725–38.

34. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nat Methods. 2015 Jan;12(1):7–8.

35. Søndergaard CR, Olsson MHM, Rostkowski M, Jensen JH. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. J Chem Theory Comput. 2011 Jul 12;7(7):2284–95.

36. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. J Chem Theory Comput. 2011 Feb 8;7(2):525–37.

Submitted to PLOS Biology

37. Schiebel J, Gaspari R, Wulsdorf T, Ngo K, Sohn C, Schrader TE, et al. Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes. Nat Commun. 2018 Sep 3;9(1):3559.

38. Salameh MA, Soares AS, Hockla A, Radisky ES. Structural basis for accelerated cleavage of bovine pancreatic trypsin inhibitor (BPTI) by human mesotrypsin. J Biol Chem. 2008 Feb 15;283(7):4115–23.

39. Price DJ, Brooks CL. A modified TIP3P water potential for simulation with Ewald summation. J Chem Phys. 2004 Nov 22;121(20):10096–103.

40. Joung IS, Cheatham TE. Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters. J Phys Chem B. 2009 Oct 8;113(40):13279–90.

41. Joung IS, Cheatham TE. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. J Phys Chem B. 2008 Jul 31;112(30):9020–41.

42. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015 Aug 11;11(8):3696–713.

43. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics. 1984 Oct 15;81(8):3684–90.

44. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics. 1993 Jun 15;98(12):10089–92.

45. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. The Journal of Chemical Physics. 1995 Nov 15;103(19):8577–93.

46. Crowley M, Darden T, Cheatham T, Deerfield D. Adventures in Improving the Scaling and Accuracy of a Parallel Molecular Dynamics Program. The Journal of Supercomputing. 1997 Nov 1;11(3):255–78.

47. Pratt LR, 1950-, Hummer G. Simulation and theory of electrostatic interactions in solution. In AIP; 1999 [cited 2018 Oct 12]. Available from: http://agris.fao.org/agris-search/search.do?recordID=US201300040464

48. Case DA, Cerutti DS, Cheatham III TE, Darden TA, Duke RE, Giese TJ, et al. AMBER17. 2017. San Francisco: University of California. 2018;

49. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J Chem Theory Comput. 2013 Jul 9;9(7):3084–95.

50. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. Journal of Molecular Graphics. 1996 Feb 1;14(1):33–8.

51. Roberts E, Eargle J, Wright D, Luthey-Schulten Z. MultiSeq: unifying sequence and structure data for evolutionary analysis. BMC Bioinformatics. 2006 Aug 16;7(1):382.

52. Eargle J, Wright D, Luthey-Schulten Z. Multiple Alignment of protein structures and sequences for VMD. Bioinformatics. 2006 Feb 15;22(4):504–6.

submitted to PLOS Biology

544 53. Stone J. An efficient library for parallel ray tracing and animation. Masters Theses [Internet].
545 1998 Jan 1; Available from: http://scholarsmine.mst.edu/masters_theses/1747

546 54. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, et al. Atomic-level
547 characterization of the structural dynamics of proteins. Science. 2010 Oct 15;330(6002):341–6.

548 55. Pendlebury D, Wang R, Henin RD, Hockla A, Soares AS, Madden BJ, et al. Sequence and
549 conformational specificity in substrate recognition: several human Kunitz protease inhibitor
550 domains are specific substrates of mesotrypsin. J Biol Chem. 2014 Nov 21;289(47):32783–97.

551 56. Hedstrom L. Serine protease mechanism and specificity. Chem Rev. 2002 Dec;102(12):4501–24.

552 57. Naftaly S, Cohen I, Shahar A, Hockla A, Radisky ES, Papo N. Mapping protein selectivity
553 landscapes using multi-target selective screening and next-generation sequencing of
554 combinatorial libraries. Nature Communications. 2018 Sep 26;9(1):3935.

555 58. Kayode O, Wang R, Pendlebury DF, Cohen I, Henin RD, Hockla A, et al. An Acrobatic Substrate
556 Metamorphosis Reveals a Requirement for Substrate Conformational Dynamics in Trypsin
557 Proteolysis. J Biol Chem. 2016 Dec 16;291(51):26304–19.

558 59. Xu Y, Carr PD, Guss JM, Ollis DL. The crystal structure of bikunin from the inter-alpha-inhibitor
559 complex: a serine protease inhibitor with two Kunitz domains. J Mol Biol. 1998 Mar;276(5):955–
560 66.

561 60. Salameh MA, Soares AS, Navaneetham D, Sinha D, Walsh PN, Radisky ES. Determinants of
562 Affinity and Proteolytic Stability in Interactions of Kunitz Family Protease Inhibitors with
563 Mesotrypsin. J Biol Chem. 2010 Nov 19;285(47):36884–96.

564 **FUNDING DISCLOSURE**

565 This project was co-financed by the European Union, the Limousin Regional Council and the French
566 National Institute for Agricultural Research. A.P. were supported by a PhD fellowship from
567 INRA/Region Limousin and M.A.M. by the French Ministry of Foreign Affairs. F.D.M and P.T. thank CALI
568 (CAlcul en LImousin) for the theoretical study. P.T. thanks the Czech Science Foundation
569 (P208/12/G016) and the Ministry of Education, Youth and Sports of the Czech Republic (project
570 LO1305).

571 **LEGENDS**

572 **Fig 1. Production of recombinant wild-type and chimeric GASP proteins.** (**A**) Domain architecture of
573 GST-fusion GASP proteins. The different proteins are expressed with GST (black box) in N-terminal
574 position, a WAP domain (W), a follistatin/Kazal domain (F/K), an Igc2 domain, two BPTI/kunitz domains
575 (Ku_1 and Ku_2) and a netrin domain (NTR). The domains of GASP-1 and GASP-2 are represented by gray

submitted to PLOS Biology

or white boxes respectively. In the two chimeric GASP proteins (GASP-1/K_{2-2} and GASP-2/K_{2-1}), the second Kunitz domain has been interchanged. The boxes representing domains are drawn to scale. The numbers above the boxes correspond to the amino-acids. (**B**) Electrophoretic analyses of recombinant wild type and chimeric GASP proteins. For each construction, the bacterial cytosolic extract was purified on a GST-Trap HP column. Proteins were analyzed before and after thrombin cleavage on 10% SDS/PAGE and revealed by Coomassie blue staining. Lane 1: total protein extract (0.1 UDO) of *E. coli* BL21 expressing the GST-GASP-1 after IPTG induction. Lane 2: Purified fraction of GST-GASP-1 before thrombin cleavage. Lanes 3 to 6: Purified fractions after thrombin treatment corresponding to GASP-1, GASP-2, GASP-1/Ku_{2-2} and GASP-2/Ku_{2-1} respectively. Lane MW: Pre-stained protein marker. (**C**) – Immunodetection of mGASP proteins before and after thrombin cleavage. Lane 1: non cleaved fraction of GST-GASP-1. Lane 2: cleaved fraction of GST-GASP-1. Lane 3: non cleaved fraction of GST-GASP-2. Lane 4: cleaved fraction of GST-GASP-2

Fig 2. Lineweaver-Burk plots of the trypsin activity. Trypsin activity was recorded using 10 nM of trypsin and different concentrations of (**A**) mGASP-1 protease inhibitor (0 to 400 nM)**,** (**B**) mGASP-2 protease inhibitor (0 to 400 nM). Hydrolysis of fluorescent substrate (Z-Phe-Arg-NHMec) (6.25 to 200 µM) was monitored during 15 min at 25°C in 100 mM Tris-HCl, 100 mM $CaCl_2$, pH 8.3 buffer. The GASP/trypsin dissociation constant (K_i) was calculated by replotting the inverse of the apparent V_{max} values for mGASP-1 or the replot of the apparent K_m for mGASP-2 *vs*. the inhibitor concentrations [I] (inset).Each test was performed 3 times independently.

Fig 3. Effects of chimeric GASP-1/K_{2-2} and GASP-2/K_{2-1} on proliferation and differentiation of C2C12 myoblast cells. (**A**) Proliferation analysis of C2C12 cells cultured for 72 h in the absence (PBS) or presence of 1 µg.ml^{-1} GASP-1/Ku_{2-2} or 1 µg.ml^{-1} GASP-2/Ku_{2-1} proteins produced *in E. coli* as measured by formazan assay. Each point corresponds to the mean ± S.D. of three independent experiments. Statistical significance was determined using a *t*-test analysis. ***: $p<0.005$. (**B**) Quantification of fusion index C2C12 myotubes treated with PBS, 1 µg.ml^{-1} GASP-1/K_{2-2} or 1 µg.ml^{-1} GASP-2/K_{2-1} proteins and

submitted to PLOS Biology

cultured for 96 h of differentiation. Each point corresponds to the mean ± S.D. of three independent experiments. Statistical significance was determined using a *t*-test analysis. **: $p<0.01$; ***: $p<0.005$. (**C**) C2C12 myotubes treated by PBS (a), 1 µg.ml^{-1} GASP-1/$K_{2\text{-}2}$ (b) or 1 µg.ml^{-1} GASP-2/$K_{2\text{-}1}$ (**C**) proteins were immunostained for MyHC protein and DAPI at 72 h after induction of differentiation.

Fig 4. Lineweaver-Burk plots of the activity of trypsin (10 nM) recorded at different concentrations of: (**A**) The recombinant GASP-1/$K_{2\text{-}2}$ protease inhibitor (0 to 10 µM)**.** (**B**) The recombinant GASP-2/K2-1 protease inhibitor (0 to 3.2 µM). Hydrolysis of fluorescent substrate (Z-Phe-Arg-NHMec) (12 to 200 µM) was monitored during 15 min at 25°C in 100 mM Tris-HCl, 100 mM CaCl2, pH 8.3 buffer. The GASP/trypsin dissociation constant (Ki) was calculated by replotting the apparent Km values for the recombinant GASP-1/K2-2 and the inverse of the apparent Vmax values for the recombinant GASP-2/K2-1 vs. the inhibitor concentrations [I] (inset). Each test was performed 3 times independently.

Fig 5. Sequence and structural patterns of trypsin inhibitors: (**A**) Sequence alignment of BPTI, mKu_2^{GASP1}, mKu_2^{GASP2}, hKu_2 and (**B**) related 3D structures.

Fig 6. **Structural patterns of bovine trypsin:** (**A**) Sequence and (**B**) 3D structure of bovine trypsin highlighting sites of interest in bovine trypsin (catalytic triad and oxyanion hole being colored in red and blue, respectively; S1-sites being depicted in green, yellow and pink, respectively for defined S1 site A, B and C. The color code of these sites is kept along manuscript. (**C**) Globally admitted mechanism of serine protease hydrolysis (Adapted from (56)) and (**D**) proposed inhibition mechanism of mKu_2^{GASP2}.

Fig 7. Binding modes of BPTI-trypsin complexes: Key residues of BPTI binding with (**A**) *b*Trypsin and (**B**) *h*Trypsin. BPTI and trypsin residues are represented in balls-and-sticks and licorice, respectively.

Fig 8. Structural features of trypsin-Kunitz domain complexes: (**A**) Superimposition of 10 representative snaphots along MD simulations of [*b*Trypsin-BPTI] (upper left), of the 5 replicas of [*b*Trypsin-mKu_2^{GASP1}] (upper right) and of [*b*Trypsin-mKu_2^{GASP2}] (lower center). (**B**) Time-dependent backbone RMSD of mKu_2^{GASP1} (left) and mKu_2^{GASP2} (right) replicas, including those of [*b*Trypsin-BPTI] and [*h*Trypsin-BPTI] as reference. (**C**) Per-residue RMSF of mKu_2^{GASP1} (left) and mKu_2^{GASP2} (right) replicas

submitted to PLOS Biology

including those of [*b*Trypsin-BPTI] and [*h*Trypsin-BPTI] as reference. For the sake of readability, [*b*Trypsin-BPTI] and [*h*Trypsin-BPTI] are colored in black and grey, respectively, while replicas are depicted in red, blue, green, cyan and purple.

Fig 9. Binding modes of Kunitz domain-trypsin complexes: Key residues of (**A**) *m*Ku_2^{GASP1} and (**B**) *m*Ku_2^{GASP1} binding with *b*Trypsin. mKu_2^{GASP} and trypsin residues are represented in ball and sticks and licorice, respectively.

Table 1: Primer sequences and amplification conditions used for PCR assembly.
Overlapping nucleotides between the large amplified fragment and the short amplified fragment are indicated in lowercase characters.

Primers	PCR conditions	Fragment lenght
pGEX-4T1::*Gasp-1* (amplification of GASP-1 without $K_{2\text{-}1}$) 5'GGTGGCCAATGGCCCACT3' 5'CCCTTCCCGAGGGGCAAC3' **pGEX-4T1::*Gasp-2* (amplification of $K_{2\text{-}2}$)** 5'TGAGTGGGCCATTGGCCACCtgtgcactgcctgcagtt3' 5'TGGTTGCCCCTCGGGAAGGGgcaagcatcctcacaagctc3'	98°C, 30 s 98°C, 30 s / 72°C, 4 min (35 cycles) 72°C, 2 min 98°C, 30 s 98°C, 10 s / 67°C, 30 s / 72°C, 30 s (35 cycles) 72°C, 2 min	6422 bp 193 bp
pGEX-4T1::*Gasp-2* (amplification of GASP-2 without $K_{2\text{-}2}$) 5'GACATCCCCGGGGCCACG3' 5'CCTGTACCACGCACACCACC3' **pGEX-4T1::*Gasp-1* (amplification of $K_{2\text{-}1}$)** 5'TTCGTGGCCCCGGGGATGTCtgcagcctgcctgccctg3' 5'CAGGGTGGTGTGCGTGGTACAGGacacgactcctcacaagcctcac3'	98°C, 30 s 98°C, 30 s / 72°C, 4 min (35 cycles) 72°C, 2 min 98°C, 30 s 98°C, 10 s / 72°C, 1 min (35 cycles) 72°C, 2 min	6377 bp 196 bp

submitted to PLOS Biology

Table 2: Primary and secondary antibodies

	Origin	References	Dilution	Coupled to
Primary Antibodies				
Anti GASP-1	Polyclonal goat	AF2070, R&D Systems Inc.	1 :1000	
Anti GASP-2	Polyclonal goat	AF2136, R&D Systems Inc.	1 :1000	
Anti SMAD2/3	Polyclonal goat	AF3797, R&D Systems Inc.	1 :1200	
Anti phospho-SMAD3	Polyclonal rabbit	AB3226, R&D Systems Inc.	1 :1200	
Anti myosin skeletal	Monoclonal mouse	M4276, Sigma-Aldrich	1 :400	
Secondary Antibodies				
Anti goat IgG	Polyclonal swine	P0449, DAKO	1 :1000	Horseradish peroxydase
Anti rabbit IgG	Polyclonal swine	P0399, DAKO	1 :1000	Horseradish peroxydase
Anti mouse IgG (H+L)	Polyclonal goat	R37120, Invitrogen	1 :1000	Alexa Fluor® 488

submitted to PLOS Biology

650 **Table 3: Description of H-bond networks between inhibitors and *b*Trypsin**

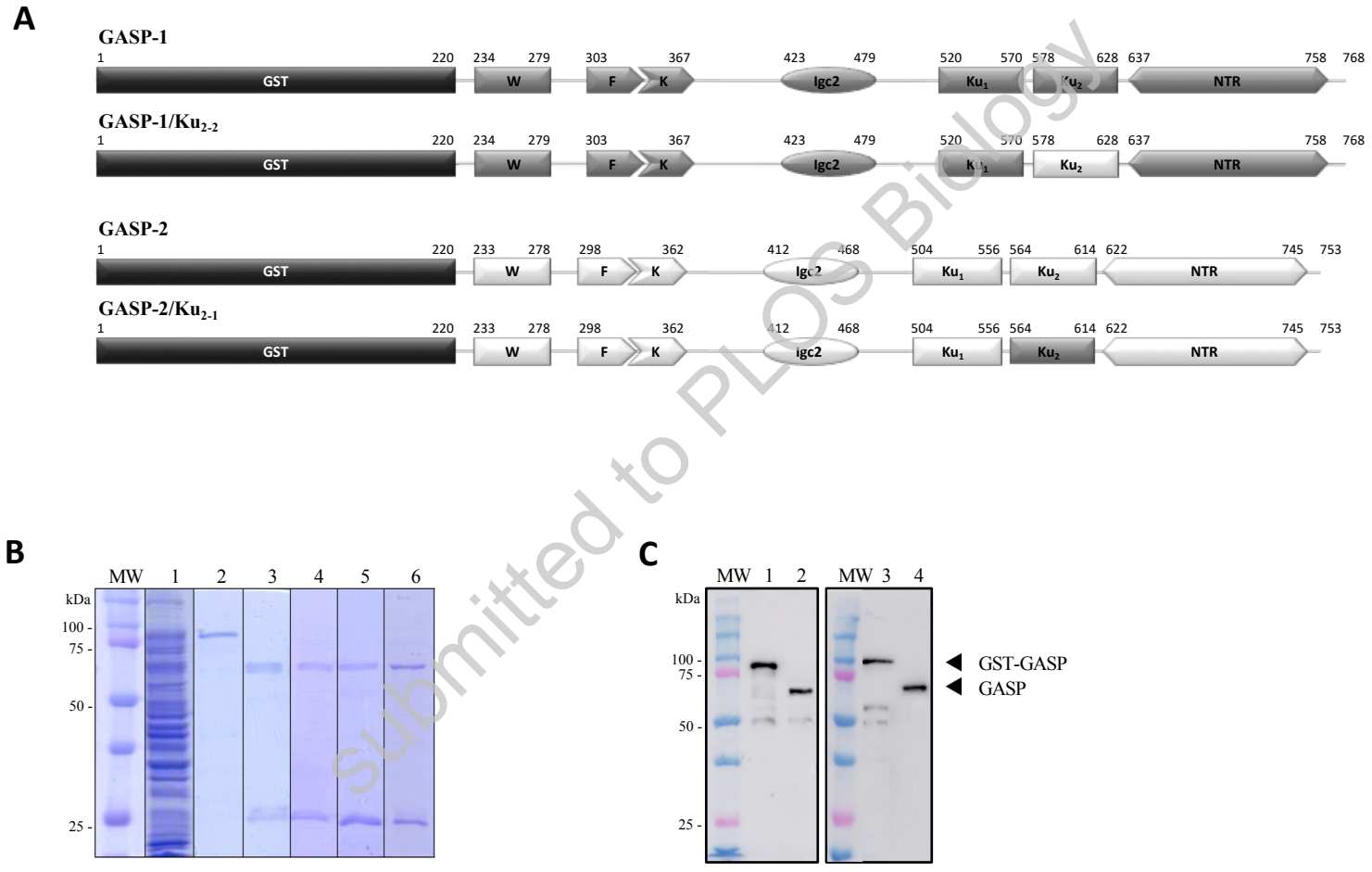
System	Moiety			Trypsin residue	Domain	d (Å)	θ (deg)	f
[*b*Trypsin-BPTI]	P_1	Lys	NH_3^+	^{195}Ser (C=O)	S_1-site A	2.82	150.2	0.61
			NH	^{215}Ser (C=O)	S_1-site B	2.91	146.3	0.15
			C=O	^{198}Gly (NH)	Oxyanion hole	2.81	152.0	0.71
				^{200}Ser (NH)	Catalytic triad	2.91	157.0	0.35
	$P_{2'}$	Arg	NH	^{47}Phe (C=O)	$S_{2'}$-site	2.87	156.6	0.65
			Guanidinium	^{41}His (C=O)	$S_{2'}$-site	2.86	155.2	0.39
	P_2	Cys	C=O	^{197}Gln ($CONH_2$)	S_1-site A	2.84	160.2	0.82
	P_3	Pro	C=O	^{217}Gly (NH)	S_1-site B	2.87	154.1	0.64
[*h*Trypsin-BPTI]	P_1	Lys	NH_3^+	^{190}Ser (C=O)	S_1-site A	2.81	157.8	0.72
			NH	^{214}Ser (C=O)	S_1-site B	2.92	155.6	0.13
			C=O	^{193}Gly (NH)	Oxyanion hole	2.79	149.2	0.60
				^{195}Ala[a] (NH)	Catalytic triad	2.91	156.7	0.44
	$P_{2'}$	Arg	Guanidinium	^{40}His (C=O)	$S_{2'}$-site	2.86	153.6	0.53
			NH	^{41}Phe (C=O)	$S_{2'}$-site	2.90	156.8	0.36
	P_2	Cys	C=O	^{192}Gln ($CONH_2$)	S_1-site A	2.84	159.1	0.81
	P_3	Pro	C=O	^{216}Gly (NH)	S_1-site B	2.87	151.9	0.55
[*b*Trypsin-*m*Ku_2^{GASP1}]	P_1	Lys	NH3+	^{194}Asp (COO^-)	S_1-site A	2.82	153.6	0.33
				^{217}Gly (NH)	S_1-site B	2.84	156.2	0.11
				^{195}Ser (C=O)	S_1-site A	2.84	153.5	0.45
			CO	^{200}Ser (OH)	Catalytic triad	2.77	158.7	0.42
[*b*Trypsin-*m*Ku_2^{GASP2}]	P_1	Gln	C=O/amide CO	^{198}Gly (NH)	Oxyanion hole	2.83	152.8	0.68
			$CONH_2$	^{196}Cys (C=O)	S_1-site A	2.86	162.4	0.38
			C=O/$CONH_2$	^{215}Ser (C=O)	S_1-site B	2.89	158.8	0.19
				^{195}Ser (OH)	S_1-site A	2.89	157.1	0.13
	$P_{1'}$	Gly	C=O	^{200}Ser (OH)	Catalytic triad	2.69	161.4	0.90
			C=O	^{197}Gln ($CONH_2$)	S_1-site A	2.83	151.3	0.48
	P_2	Cys	C=O	^{197}Gln ($CONH_2$)	S_1-site A	2.83	158.6	0.46
	$P_{2'}$	Trp	C=O	^{66}Lys (NH_3^+)		2.82	158.5	0.20
	$P_{3'}$	Glu	COO^-	^{66}Lys (NH_3^+)		2.78	155.7	0.33

[a] In [hTrypsin-BPTI] X-ray structure (PDB ID: 2RA3), catalytic ^{195}Ser were mutated into alanine. It is worth mentioning that binding mode is not expected to be significantly modified given the globally accepted catalytic mechanism.

651

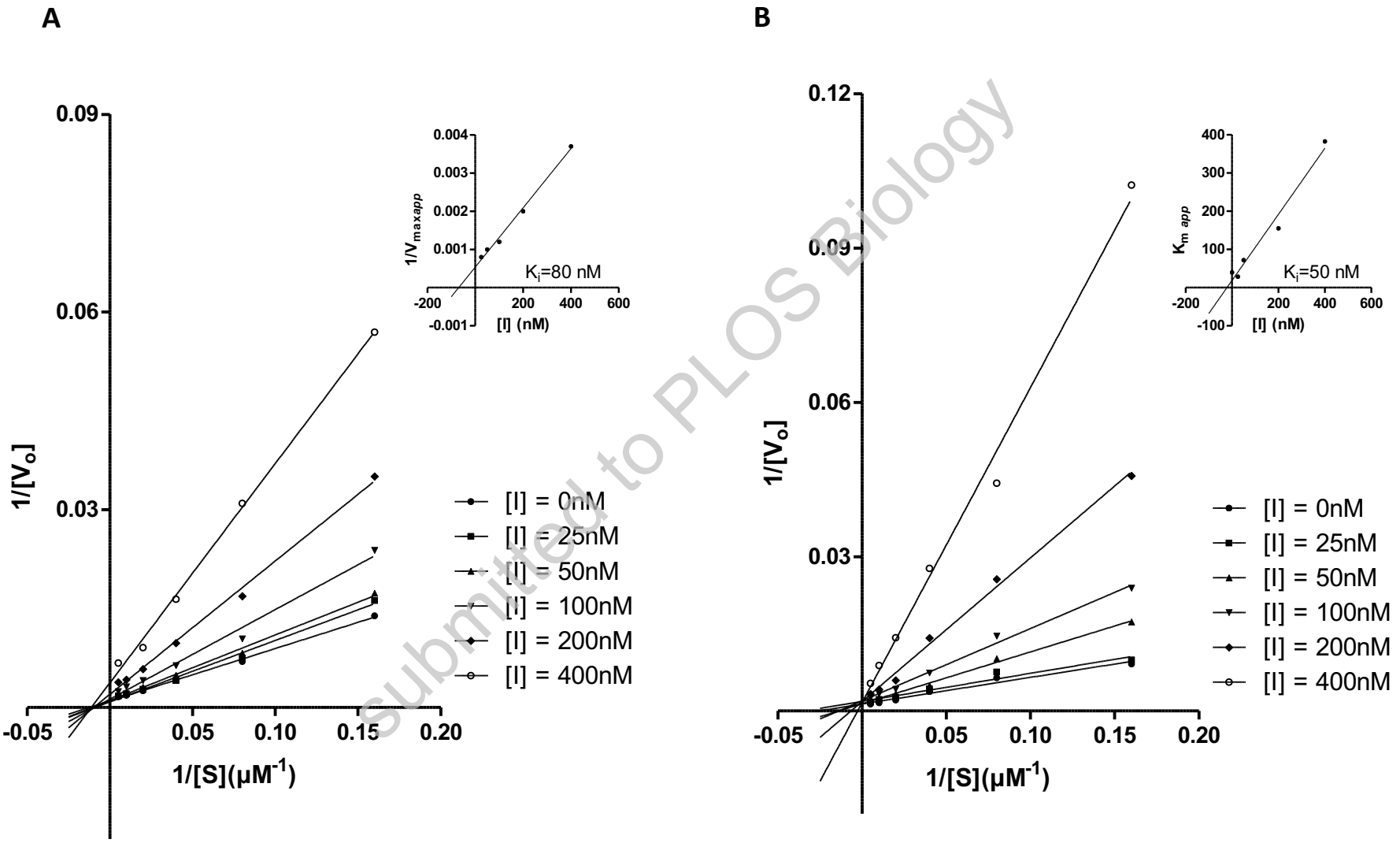
submitted to PLOS Biology

Fig 1



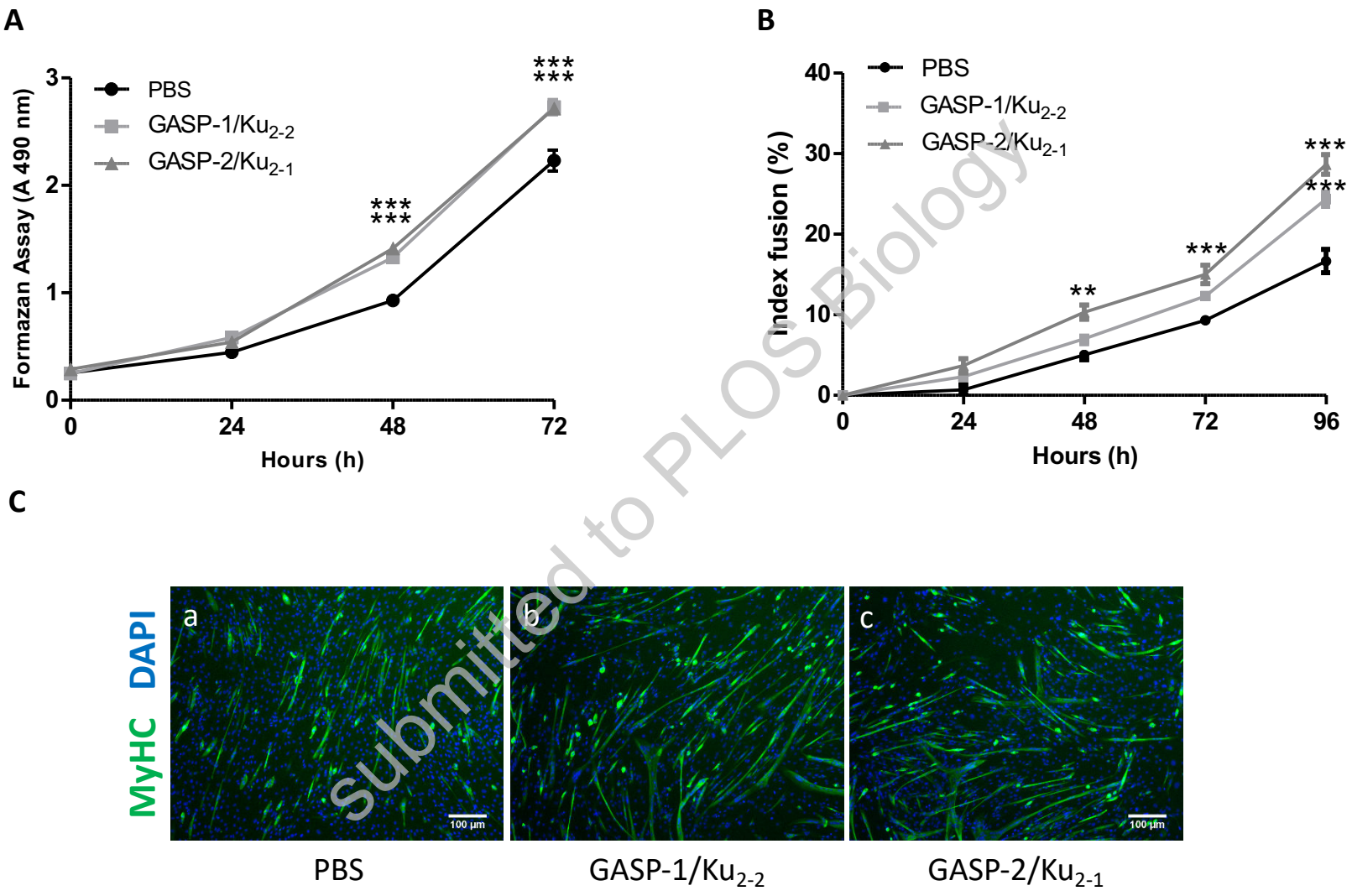
submitted to PLOS Biology

Fig 2



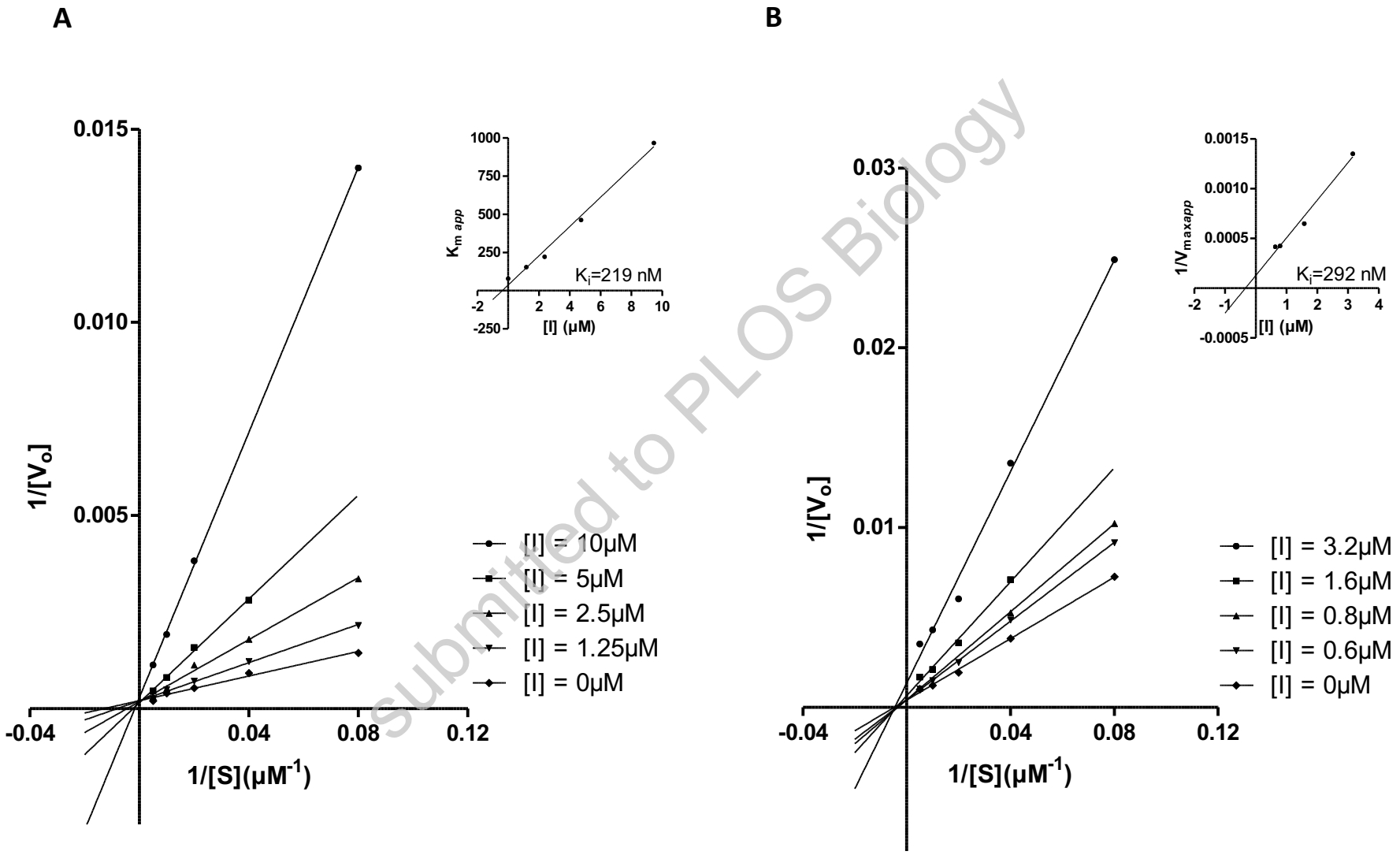
Submitted to PLOS Biology

Fig 3



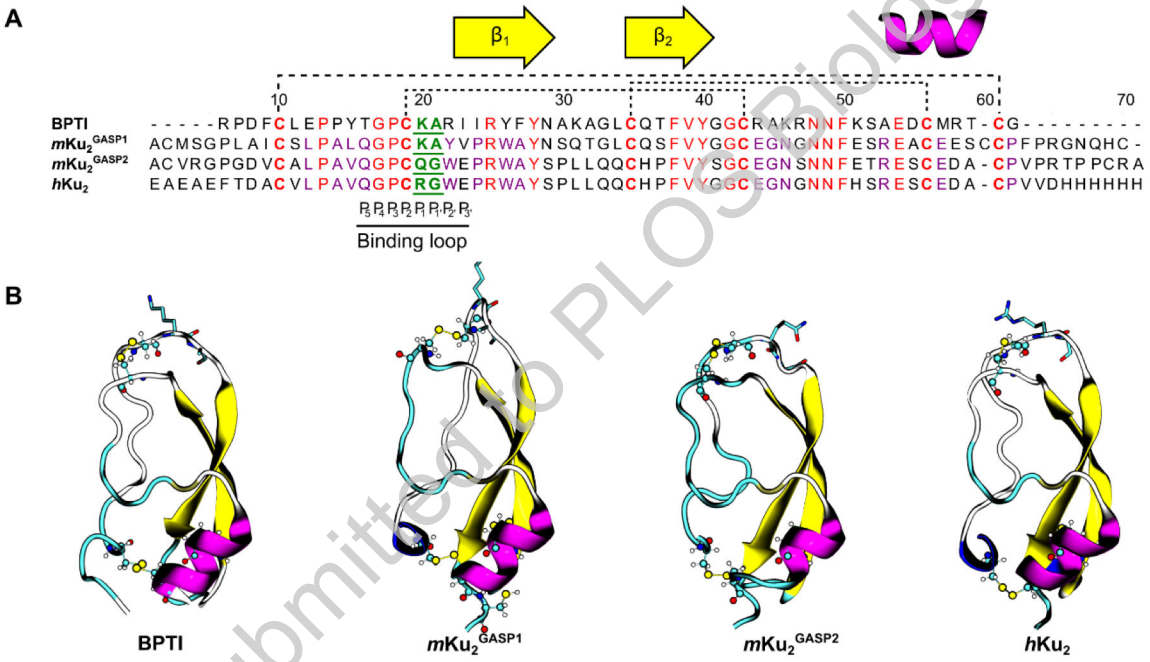
submitted to PLOS Biology

Fig 4



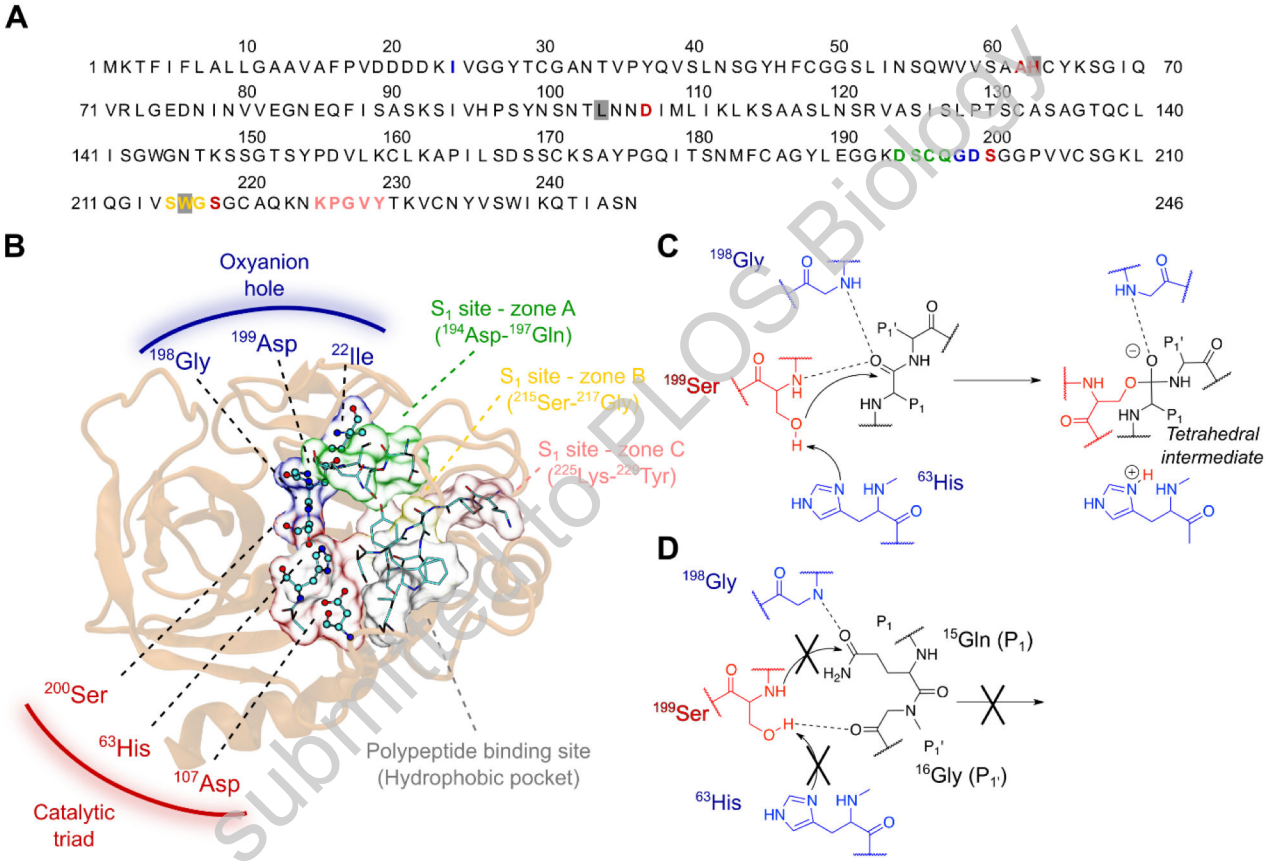
submitted to PLOS Biology

Fig 5



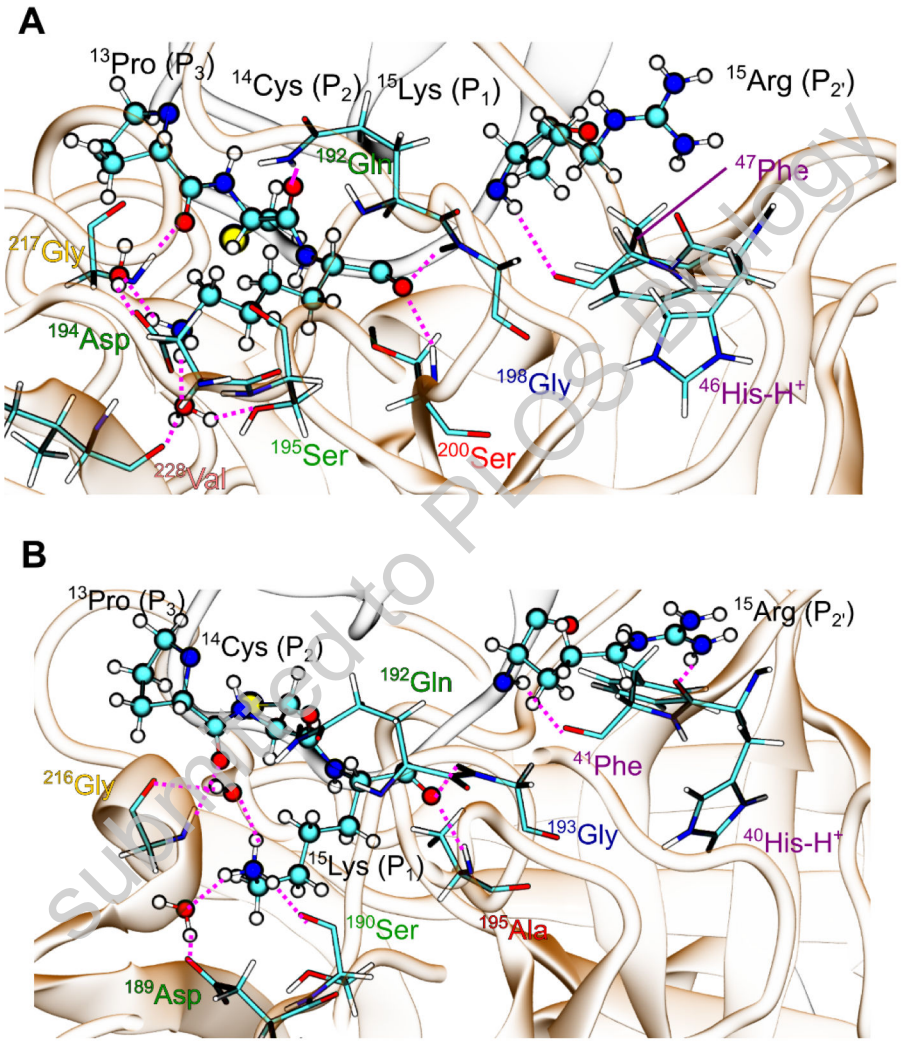
submitted to PLOS Biology

Fig 6



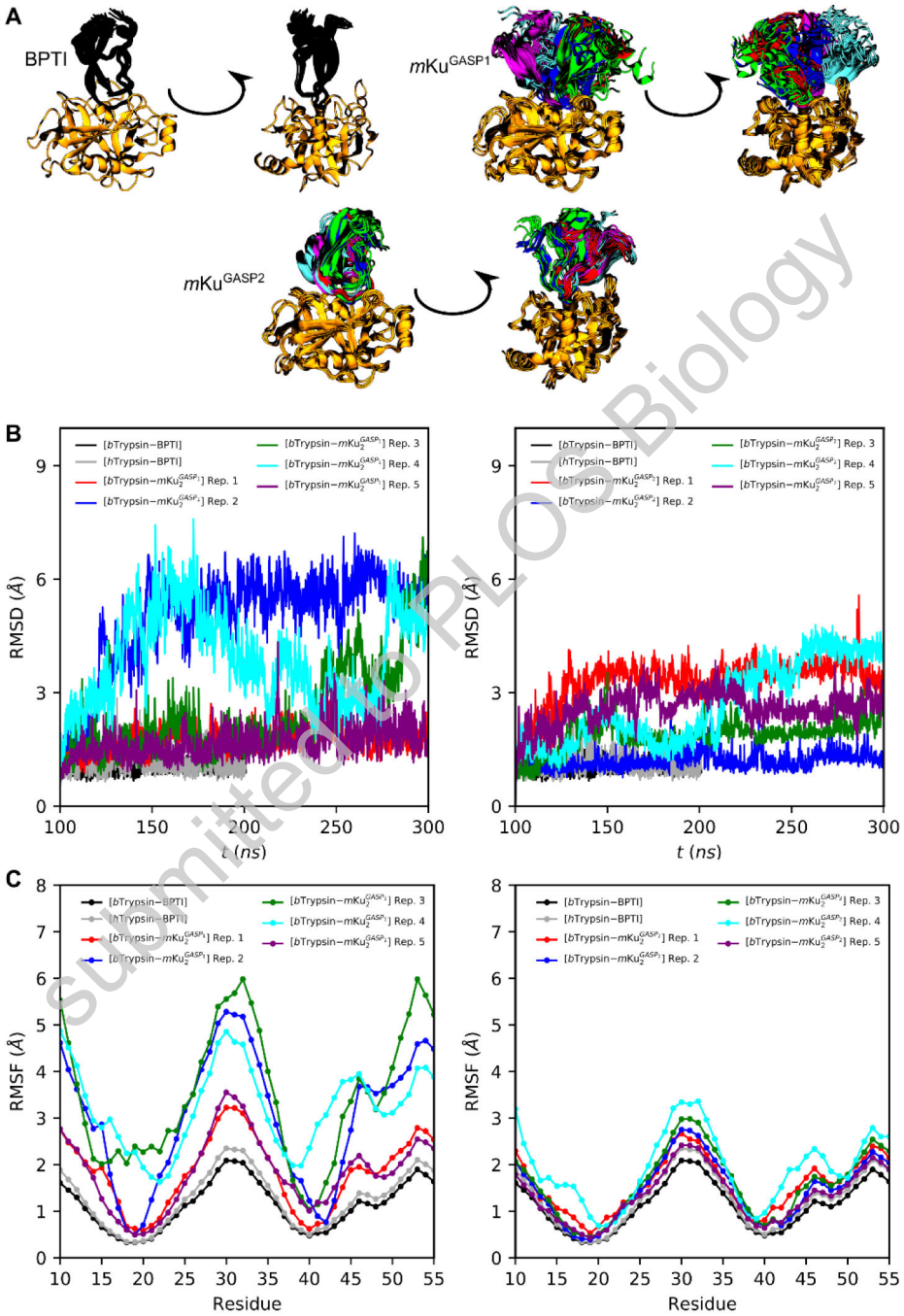
submitted to PLOS Biology

Fig 7



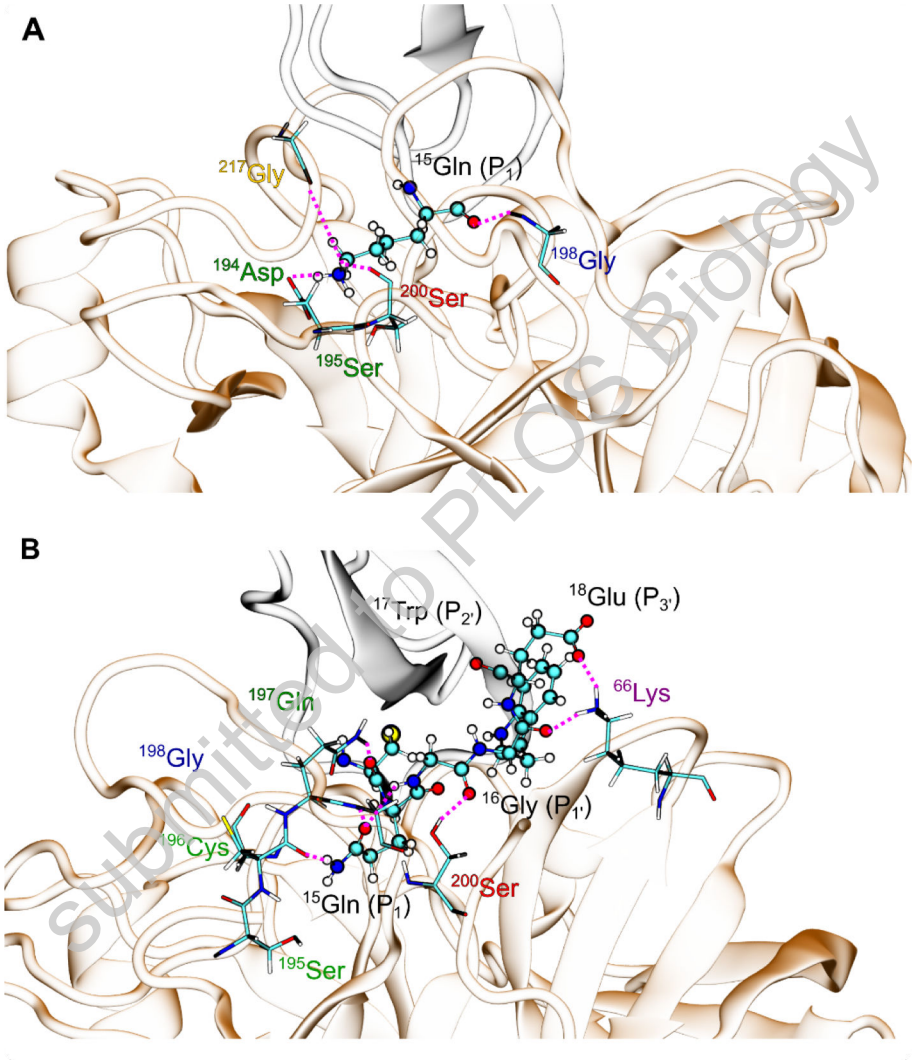
submitted to PLOS Biology

Fig 8



Submitted to PLOS Biology

Fig 9



submitted to PLOS Biology

Supplementary Tables

S1 Table. Sequence Identities between trypsin inhibitors.

	BPTI	*h*Ku$_2$	*m*Ku$_2^{GASP1}$	*m*Ku$_2^{GASP2}$
BPTI	-	41.18%	37.25%	52.94%
*h*Ku$_2^{GASP1}$	41.18%	-	88.24%	64.71%
*m*Ku$_2^{GASP1}$	37.25%	88.24%	-	60.78%
*m*Ku$_2^{GASP2}$	52.94%	64.71%	60.78%	-

S2 Table.

System		Box Size *Initial* x	y	z	*Equilibrated* x	y	z	N_{atom}
*m*Ku$_2^{GASP1}$		59.0	67.9	60.8	54.5	62.7	56.2	19017
*m*Ku$_2^{GASP2}$		64.7	62.5	61.6	59.6	57.5	56.7	19288
[*b*Trypsin-BPTI]		79.7	74.0	86.5	74.0	68.7	80.4	40878
[*h*Trypsin-BPTI]		82.1	75.3	87.1	76.3	70.0	80.9	43343
[*b*Trypsin-*m*Ku$_2^{GASP1}$]	Rep. 1	78.7	81.3	98.4	73.7	76.1	92.1	51593
	Rep. 2	78.7	77.3	95.9	73.5	72.3	89.6	47666
	Rep. 3	79.1	84.4	92.5	73.8	78.7	86.3	50171
	Rep. 4	79.1	81.0	99.9	74.1	75.9	93.6	52526
	Rep. 5	81.1	79.7	94.7	75.5	74.2	88.2	49430
[*b*Trypsin-*m*Ku$_2^{GASP2}$]	Rep. 1	77.7	86.7	87.7	72.9	81.3	82.2	48577
	Rep. 2	78.7	84.4	90.2	73.3	78.6	84.0	48403
	Rep. 3	77.4	79.2	95.9	72.3	74.0	89.6	47854
	Rep. 4	77.4	77.9	98.4	72.6	73.1	92.3	48997
	Rep. 5	78.7	78.6	95.4	73.4	73.4	89.1	48001

S3 Table. H-bond network between β_1- and β_2-sheets in *m*Ku$_2^{GASP1}$ and *m*Ku$_2^{GASP2}$.

System	H-Acceptor *Residue*	*Moiety*	H-Donor *Residue*	*Moiety*	Overall fraction	Distance *(Å)*	Angle *(deg)*
*m*Ku$_2^{GASP1}$	^{23}Val	C=O	^{40}Tyr	NH	0.81	2.8	160.4
	^{27}Ala	C=O	^{36}Gln	NH	0.81	2.8	161.6
	^{36}Gln	C=O	^{27}Ala	NH	0.73	2.9	160.8
	^{25}Arg	C=O	^{38}Phe	NH	0.67	2.9	159.7
	^{38}Phe	C=O	^{25}Arg	NH	0.57	2.9	159.7
	^{40}Tyr	C=O	^{23}Val	NH	0.30	2.9	154.3
*m*Ku$_2^{GASP2}$	^{23}Glu	C=O	^{40}Tyr	NH	0.90	2.8	160.8
	^{27}Ala	C=O	^{36}His	NH	0.78	2.9	162.2
	^{25}Arg	C=O	^{38}Phe	NH	0.76	2.9	160.4
	^{38}Phe	C=O	^{25}Arg	NH	0.72	2.9	160.9
	^{36}His	C=O	^{27}Ala	NH	0.51	2.9	161.5
	^{40}Tyr	C=O	^{22}Trp	NH	0.44	2.9	160.4
	^{40}Tyr	C=O	^{23}Glu	NH	0.34	2.9	159.8

submitted to PLOS Biology

S4 Table. Average backbone dihedral angles (deg) of trypsin inhibitors.

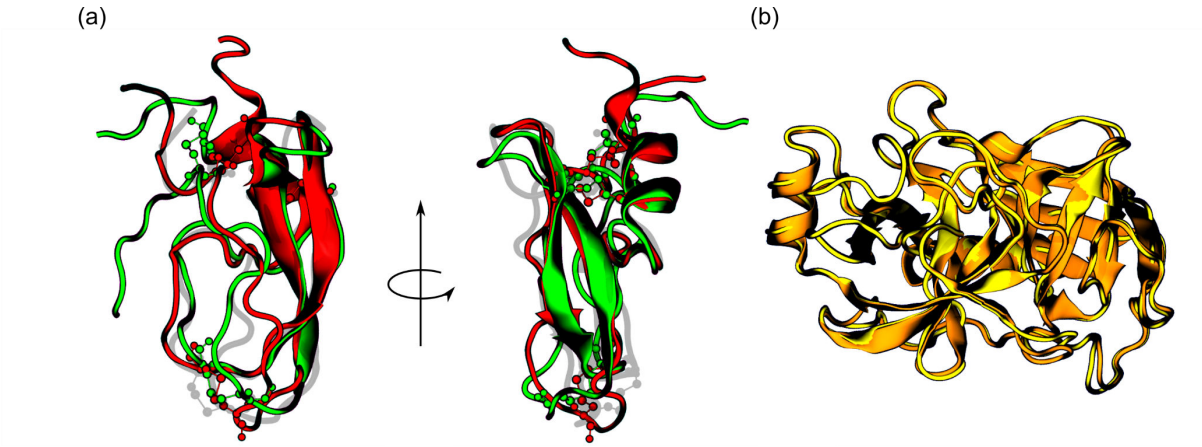
	*m*Ku2^{GASP1}		*m*Ku2^{GASP2}		*h*Ku2		BPTI	
	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
P_5	-119.1 ± 34.4	53.2 ± 86.4	-154.9 ± 21.0	166.3 ± 10.4	-83.1 ± 7.8	-35.7 ± 14.5	-68.3	-42.7
P_4	-166.2 ± 98.0	160.9 ± 23.7	-73.7 ± 12.7	164.5 ± 15.3	92.5 ± 12.8	172.3 ± 2.5	83.0	177.3
P_3	-72.1 ± 10.7	152.5 ± 32.1	-71.6 ± 9.3	34.4 ± 43.6	-75.9 ± 5.6	-14.6 ± 12.7	-84.5	-16.9
P_2	-133.6 ± 32.2	143.1 ± 11.4	-90.6 ± 40.0	155.0 ± 10.4	-72.1 ± 5.7	174.4 ± 10.8	-76.6	162.7
P_1	-61.2 ± 18.2	-17.1 ± 20.8	-71.7 ± 11.6	132.4 ± 23.3	-123.8 ± 10.5	4.2 ± 1.6	-112.8	22.2
$P_{1'}$	-83.0 ± 18.4	-1.2 ± 48.7	-114.9 ± 53.1	147.9 ± 13.8	-81.7 ± 8.7	156.3 ± 3.4	-71.9	166.6
$P_{2'}$	40.2 ± 54.7	37.4 ± 29.8	-142.8 ± 12.1	-6.5 ± 30.3	-128.8 ± 4.6	113.9 ± 9.4	-121.2	84.9
$P_{3'}$	-84.2 ± 22.7	122.5 ± 9.5	-99.4 ± 29.7	122.1 ± 8.7	-115.0 ± 14.1	117.2 ± 3.9	-105.9	123.0

S5 Table. Average intermolecular H-bond numbers involving trypsin inhibitor binding loop (i.e., from P_5 to $P_{3'}$) over MD simulations.

System		Binding Loop	Residues							
			P_5	P_4	P_3	P_2	P_1	$P_{1'}$	$P_{2'}$	$P_{3'}$
[*b*Trypsin-BPTI]		4.5 ± 1.3	0.0 ± 0.0	0.0 ± 0.0	0.6 ± 0.5	0.8 ± 0.4	1.9 ± 0.9	0.0 ± 0.1	1.1 ± 0.7	0.0 ± 0.0
[*h*Trypsin-BPTI]		4.5 ± 1.3	0.0 ± 0.0	0.0 ± 0.0	0.5 ± 0.5	0.8 ± 0.4	2.1 ± 0.9	0.0 ± 0.2	1.0 ± 0.7	0.0 ± 0.0
[*b*Trypsin-*m*Ku2^{GASP1}]	Overall	1.9 ± 1.2	0.3 ± 0.5	0.0 ± 0.1	0.0 ± 0.2	0.1 ± 0.3	1.2 ± 1.0	0.1 ± 0.4	0.1 ± 0.3	0.0 ± 0.0
	Rep. 1	1.2 ± 0.7	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.3	0.0 ± 0.1	1.1 ± 0.7	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	Rep. 2	2.6 ± 1.1	0.1 ± 0.4	0.0 ± 0.0	0.0 ± 0.1	0.3 ± 0.5	2.0 ± 0.9	0.1 ± 0.3	0.0 ± 0.2	0.0 ± 0.0
	Rep. 3	1.0 ± 0.8	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.2	0.0 ± 0.0	0.9 ± 0.7	0.0 ± 0.1	0.1 ± 0.3	0.0 ± 0.0
	Rep. 4	1.6 ± 1.0	0.5 ± 0.6	0.0 ± 0.2	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.3	0.6 ± 0.5	0.4 ± 0.5	0.0 ± 0.0
	Rep. 5	2.9 ± 1.2	0.7 ± 0.7	0.0 ± 0.2	0.0 ± 0.0	0.0 ± 0.0	2.0 ± 0.8	0.0 ± 0.0	0.1 ± 0.3	0.0 ± 0.0
[*b*Trypsin-*m*Ku2^{GASP2}]	Overall	4.1 ± 1.4	0.0 ± 0.2	0.0 ± 0.1	0.1 ± 0.3	0.6 ± 0.5	1.5 ± 0.8	1.0 ± 0.7	0.4 ± 0.5	0.4 ± 0.5
	Rep. 1	4.5 ± 1.3	0.0 ± 0.2	0.0 ± 0.1	0.1 ± 0.3	0.7 ± 0.5	1.4 ± 0.7	1.4 ± 0.6	0.4 ± 0.5	0.4 ± 0.6
	Rep. 2	3.5 ± 1.2	0.0 ± 0.0	0.0 ± 0.0	0.2 ± 0.4	0.9 ± 0.3	1.8 ± 0.8	0.4 ± 0.5	0.0 ± 0.1	0.3 ± 0.5
	Rep. 3	2.9 ± 1.2	0.1 ± 0.3	0.0 ± 0.0	0.1 ± 0.3	0.0 ± 0.1	1.5 ± 0.7	0.6 ± 0.5	0.1 ± 0.3	0.5 ± 0.5
	Rep. 4	4.7 ± 1.2	0.0 ± 0.1	0.1 ± 0.3	0.1 ± 0.3	0.8 ± 0.4	1.3 ± 0.8	1.3 ± 0.5	0.7 ± 0.5	0.5 ± 0.5
	Rep. 5	4.7 ± 1.2	0.0 ± 0.1	0.0 ± 0.0	0.1 ± 0.4	0.8 ± 0.4	1.4 ± 0.7	1.5 ± 0.6	0.8 ± 0.4	0.1 ± 0.3

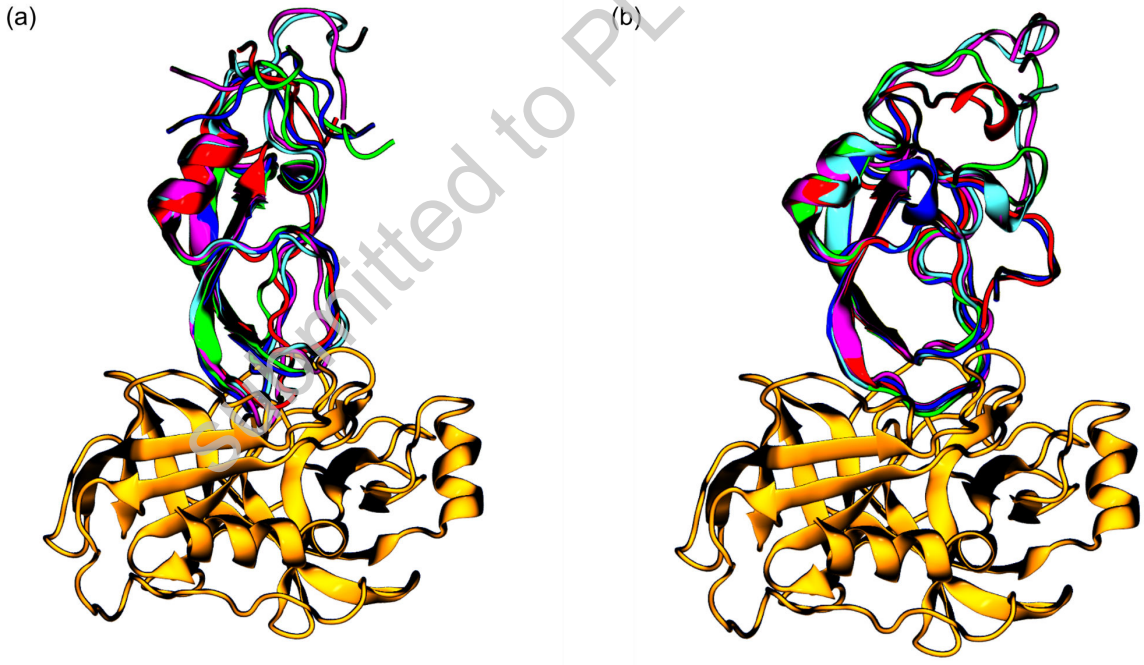
submitted to PLOS Biology

Supplementary Figures

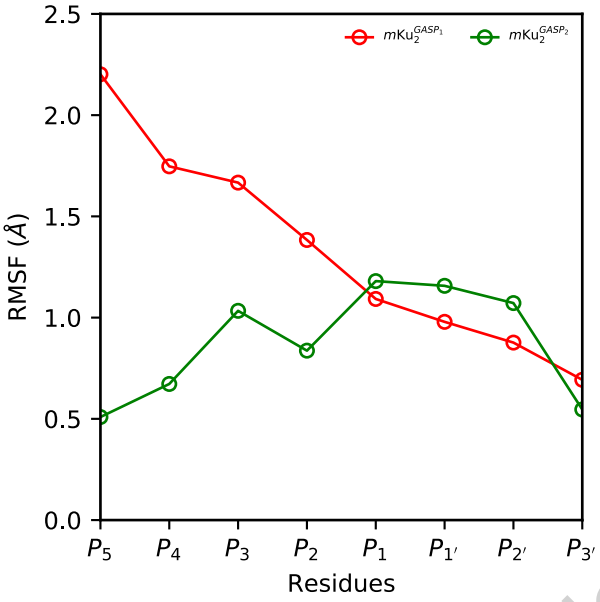


S1 Fig. This supplementary figure shows the superimposition of (a) *m*Ku$_2^{GASP1}$ (red, extracted snapshot from MD), *m*Ku$_2^{GASP2}$ (green, extracted snapshot from MD) and BPTI (shadow black, X-ray structure from PBD ID: 2RA3) and (b) *h*Trypsin (orange, X-ray structure from PBD ID: 2RA3) and *b*Trypsin (yellow, X-ray structure from PBD ID: 2RA3). Cystein bonds of *m*Ku$_2^{GASP1}$, *m*Ku$_2^{GASP2}$ and BPTI are depicted in CPK.

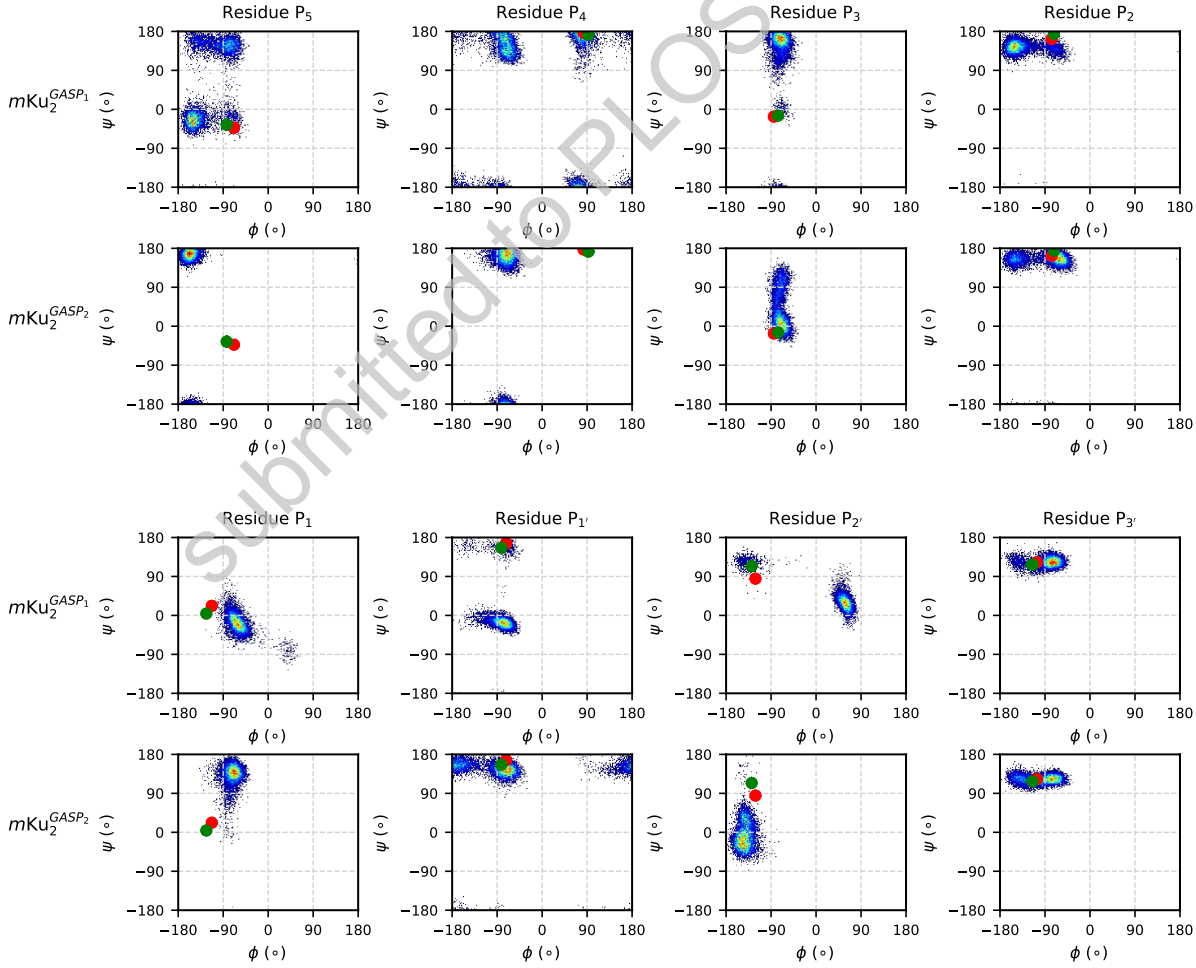
Submitted to PLOS Biology



S2 Fig. This supplementary figure shows the superimposition of five replica of *b*Trypsin complexed to (a) *m*Ku$_2^{GASP1}$ and (b) *m*Ku$_2^{GASP2}$. *b*Trypsin is depicted in orange, poses 1, 2, 3, 4 and 5 of Ku-domains are respectively colouls -red in red, blue and green, cyan and purple.



S3 Fig. This shows the atomic root-mean square fluctuations (RMSF, in Å) of mKu_2^{GASP1} and mKu_2^{GASP2} (in red and green, respectively).



S4 Fig. This shows the Ramachandran distribution of mKu_2^{GASP1} and mKu_2^{GASP2}. Reference value from X-ray and NMR structures of respectively BPTI (PDB ID: 2RA3) and hKu2 (average value from PDB ID: 2DDJ) are depicted in red and green, respectively.

Submitted to PLOS Biology

DISCUSSION

For the study of the structure / function relationships of the GASP-1 and GASP-2 proteins, we sought to highlight their anti-protease properties. This study required their prior production as recombinant proteins. While bacterial systems are still the low cost and easiest means of production, we have now many host cell / vector expression systems with the growth of genetic engineering and the development of large-scale cell culture technologies. The choice of a system depends on different criteria such as:

i)- the post-translational modifications that the protein must undergo to acquire its biological functionality. Although in recent years genetically modified bacteria offer the possibility of making glycosylation (Zhang *et al*., 2004), glycosylated recombinant proteins are often produced in the extra-cytoplasmic compartment of eukaryotic expression systems such as yeast (Gerngross, 2004), animal cells (Wurm, 2004) or plant cells (Hellwig *et al*., 2004).

ii)- the cell compartment (s) in which the protein of interest is produced or exerts its biological activity. Indeed, the location of the protein will guide the choice of the vector which must have the specific sequences for the correct addressing of the recombinant protein.

Finally, let us mention the development, in recent years, of acellular systems based on the *in vitro* coupling of transcription and translation reactions that make possible to overcome the various constraints due to the physiology of the cell. Although still very expensive, these systems often offer high production yields and allow the expression of toxic proteins.

The GASP proteins are secreted proteins carrying N- and O-glycosylated chains. At first, an eukaryotic expression system could have been chosen for the study. However, it has been shown in the laboratory that the absence of these post-translational modifications does not affect the anti-myostatin activity of GASP-1 (Brun *et al*., 2012) and GASP-2 (Périé *et al*., 2016). As a result, we opted for a prokaryotic expression system. In addition to its ease and low cost of use, this system often has the advantage of allowing the production of large quantities of the protein of interest. This last criterion has proved to be essential for our choice. Indeed, our study, based on tests of anti-trypsin activity and the determination of different kinetic parameters, requires on average 100 µg of partially purified protein, and this for each measurement done in triplicate. This quantity is not often obtained in animal cells.

The prokaryotic BL21 / pGEX4T-1 system made possible to produce each recombinant protein as a fusion protein with a N-terminal GST tag allowing a partial but sufficient purification by affinity chromatography. As shown in our results, during the first production tests carried out under standard conditions of growth temperature and IPTG concentration, the GASP proteins were found mainly in the insoluble fraction, indicating that they are probably not produced in their native state. They aggregate in inclusion bodies, suggesting their biological inactivity. The inclusion bodies correspond to the aggregation of proteins having adopted partially folded or unstable conformations. Their formation is partly related to the level of expression of the heterologous gene in the host cell. With a high level of expression, the hydrophobic intermolecular interactions are favored and the protein is mostly present as inclusion bodies (Georgiou and Valax, 1999). Under physiological conditions, the correct folding of the protein can be facilitated by several factors (Fischer *et al*., 1998):

i)- chaperone proteins which, in collaboration with cellular proteases, control this folding process. The incorrectly folded proteins are in fact cellular substrates of degradation systems which thus ensure their rapid elimination (Wickner *et al*., 1999).

ii)- oxidoreductases that catalyze the formation of di-sulfide bridges.

iii)- peptidyl-prolyl isomerases that catalyze *cis-trans* isomerization of peptide bonds in the polypeptide chain.

Thus, more the synthesis of a protein is increased, more these different factors are found in limiting quantities. Several methods are widely documented and can help to promote correct folding of a recombinant protein. Their success rate is however very variable from one protein to another. It is, for example, possible to renaturate *in vitro* the protein by acting directly on the resolubilization of easily purifiable inclusion bodies, by the addition of powerful denaturing agents such as 8M urea or 6M guanidine. This method, which requires an additional step of renaturation of the protein by dialysis (to remove the denaturant) may lead to an weak renaturation rate and a precipitation of the renatured protein.

Several other strategies may also influence the folding of the recombinant protein. In the literature, numerous experiments show that the construction of the protein in the form of a fusion protein with a tag in N-terminal position, added initially to facilitate its purification, also promotes its correct folding or increases its intracellular stability. *In E. coli*, the most widely used fusion protein systems are maltose affinity protein (MBP), glutathione S-transferase (GST) or NusA protein. In our study, the presence of the GST, in the N-terminal position in the different constructions, did not make possible to obtain a sufficient level of the recombinant proteins in the soluble fraction.

Finally, different culture parameters influence the expression and the folding of recombinant proteins. Among them, the temperature variation of the growth or the inducer concentration are recommended in the literature (Harper and Speicher, 2011) and are almost systematically tested. Generally, lowering the temperature decreases the rate of aggregation and promotes correct folding. In the same way, the decrease of the inducer concentration decreases the transcriptional level of the cDNA of interest. Although the recombinant protein is produced in smaller amounts, the proportion present in the soluble fraction is larger. For our study, the optimization of these different parameters (lowering the temperature from 37 to 15°C and decreasing the IPTG concentration from 1 to 0.025 mM) enabled us to obtain, from 1 liter of culture and after partial purification, between 300 to 500 µg depending on the protein produced, compatible quantities to pursue the study.

Since the goal of the protein production in a heterologous expression system is often to study its functionality, it is necessary to check after its purification, if the protein conformation is close to its native state. There are many techniques such as mass spectrometry or circular dichroism that can be used. However, the biological activity tests, when they exist, represent the best validation. Thus, for our study, we used the GASP proteins produced *in E. coli* to treat C2C12 myoblastic cells in culture. We have shown that each recombinant protein is able to activate myogenic proliferation and differentiation by inhibiting the myostatin signaling pathway. These results attest the biological functionality of our GASP proteins, and once again confirm that the presence of N- and O-glycosidic chains is not essential for their activity. In the future, to overcome the immunogenicity due to glycosylation, we could imagine to use these functional recombinant proteins as potential therapeutic tools for the study of certain associated pathologies including to muscle atrophy.

As seen in the chapter II, the GASP proteins present a certain structural complexity. In addition to their role as inhibitors of the myostatin signaling pathway and their ability to interact with several other members of the TGF-β family, the presence of several distinct structural domains suggests that they could have various other biological functions yet unknown. With the development and improvement of methods such as double-hybrid technique, co-immunoprecipitation or histidine pull-down assay, novel investigations may be considered in order to search potential new partners of GASP proteins.

In this study, we chose to explore the antiproteasic characteristics of GASP proteins. Their structural organization reveals the presence of several domains WAP, Kazal and Kunitz. While there are many studies in the literature that relate the involvement of these different domains in the inhibition of proteases, to date, very little is known of the

anti-protease properties of the GASP proteins. Only the second Kunitz domain isolated from the human GASP-2 protein and produced in *E. coli* has been shown to be capable to inhibiting trypsin *in vitro* (Nagy *et al*., 2003; Liepinsh *et al*., 2006). However, no kinetic parameters have been published for whole proteins or other domains. Our results showed that the recombinant GASP-1 and GASP-2 proteins produced and purified according to the methods described in the "Results" section are capable to inhibit trypsin. Trypsin is a serine endoprotease which hydrolyzes, on the C-terminal side, the peptide bonds in which a basic amino acid lysine or arginine engages its acidic function. In the body, trypsin is synthesized in the pancreas in the form of trypsinogen (inactive propeptide), then stored in vesicles of acinar cells from which it is excreted during the digestion. Activation of trypsinogen to trypsin is then the result of propeptide hydrolysis by enterokinase or by auto-activation. Although strongly expressed in the pancreas (Trexler *et al*., 2001; 2002), the role of GASP-1 and GASP-2 proteins in this organ is still unknown. It would be interesting to determine more precisely whether they are expressed in acinous cells or islet cells of Langerhans. However, it seems unlikely that the GASP proteins play a trypsin inhibitory role. Indeed, the mode of production and activation of this protease does not seem compatible with any intracellular regulation of its activity by anti-proteases in the pancreas. Although the anti-trypsin activity of the GASP proteins does not have any real biological reality in skeletal muscle, however, this enzyme represents, with chymotrypsin, the archetype of the most frequently used serine proteases in the different approaches of *in vitro* experimental research to study the activity of serine protease inhibitors. Several commercial fluorescent substrates are indeed available. In the same way, the protocols for titration and use of the enzyme are established. They allowed us to determine the stoichiometry as well as the mechanism of action of the inhibitor / enzyme complexes in order to calculate the different kinetic parameters K_m and V_{max}.

Like the second Kunitz domain of the human GASP-2 protein, the entire GASP-2 protein inhibits trypsin following a competitive-type mechanism. In this case, the GASP-2 protein could:

i)- compete with the substrate for binding to the catalytic site of the enzyme. This would imply a certain structural analogy of the GASP-2 protein with the substrate.

ii)- more probably, the binding of GASP-2 on a site distinct from the enzyme would cause a steric hindrance preventing the fixation of the substrate on the catalytic site of the trypsin.

We have shown that in the presence of increasing concentrations of GASP-2, K_m^{app} increases, thus indicating that the apparent affinity of the enzyme for the substrate decreases and V_{max} remains unchanged.

The GASP-1 protein is also capable to inhibit trypsin. Surprisingly, our results, indicate a different inhibition mechanism. GASP-1 is a non-competitive inhibitor of trypsin. Its fixation has no influence on the binding of the substrate. It only prevents the enzymatic reaction. We were able to observe that in the presence of increasing concentrations of GASP-1, trypsin is less active (V_{max} decreases) but the affinity of trypsin for the substrate remains the same (K_m is not modified).

These results seem to be surprising, considering both the similar multi-domain structural organization and the significant degree of identity (about 55% identical amino acids) between GASP-1 and GASP-2. However, there are several other examples in the literature that describe similar observations. Studies in plants have shown that some proteins that possess a Kunitz domain can inhibit trypsin by different mechanisms of action. Thus, the PdKI2 (Pithecellobium dumosum Kunitz Inhibitor 2) and CaTI (Cassia absus linn Trypsin Inhibitor) proteins inhibit trypsin using a competitive mechanism (Patel *et al*., 2015). In contrast, EvTI (Erythrina velutina Trypsin Inhibitor) and TTI (Tamarindus Trypsin Inhibitor) proteins are non-competitive inhibitors of protease (Machado *et al*., 2013; Araujo *et al*., 2005). This functional duality could reflect a cellular mechanism to regulate the activity of certain serine proteases and would depend of the concentration of the substrate. Thus, in the case of a competitive inhibitor, the inhibition can be removed when the substrate concentration increases. Conversely, this increase has no effect on a non-competitive inhibitor. In order to determine if this functional duality is due to the structural organization (amino acid composition, three-dimensional structure) of the second Kunitz domain, we realized the construction, production and purification of two other recombinant proteins in which only this domain has been interchanged. As already mentioned in the "Results" chapter, the construction of these two chimeric proteins was carried out following an assembly strategy based on the sequential use of three DNA modification enzymes. Both chimeric proteins were produced and purified according to the same experimental conditions described for the recombinant proteins GASP-1 and GASP-2. In a first step, the functional activity of these two recombinant proteins has been verified. As we could suppose, the exchange of the second Kunitz domain does not alter their antagonistic role of myostatin. In a second step, the anti-protease properties of GASP-1/$K_{2\text{-}2}$ and GASP-2/$K_{2\text{-}1}$ were determined with trypsin. Kinetic analyzes showed that the GASP-1/$K_{2\text{-}2}$ chimeric protein exhibits a competitive-type inhibition mechanism similar to that described for GASP-2. Conversely, GASP-2/$K_{2\text{-}1}$ is a noncompetitive inhibitor of trypsin comparable to GASP-1 protein. These results indicate that the second Kunitz domain present in the GASP proteins is responsible for the specificity of inhibition observed *in vitro* by the measurement of different kinetic

parameters. By modeling, we propose a structural model of the second Kunitz domain of GASP-1 and GASP-2 implicated in the anti-trypsin inhibition specificity. As expected from the high sequence identity between mKu2GASP1, mKu2GASP2, BPTI and hKu2, mKu2GASP1 and mKu2GASP2 exhibited similar topology than BPTI and hKu2. The residues of P1 and P2' differ in both mKuGASP1 and mKuGASP2 from BPTI, a Lys/Arg-Xaa inhibitor, which is expected to affect their mode of binding with respect to this type of inhibitors. As a difference in binding, the [bTrypsin-mKu2GASP1] and [bTrypsin-mKu2GASP2] complexes show significant conformational differences with respect to [bTrypsin-BPTI]. While in [bTrypsin-BPTI], BPTI is strongly anchored to trypsin, preventing rocking motion, mKu2GASP1 and mKu2GASP2 significantly fluctuate more in the complexes.

In the near future, this modeling could be complemented by a study of the structure of the Kunitz second domain or whole proteins by high resolution crystallographic analysis that would provide structural information at the atomic scale on the region(s) interacting with the protease. This study, however, would require improving the degree of purity of the recombinant produced proteins. It could be envisaged, for example, to couple the purification by affinity chromatography with ion exchange chromatography. Currently, low resolution modeling has already been obtained for GASP-1 and GASP-2 proteins complexed with myostatin (Walker *et al*., 2015).

One of the main research axes of the laboratory is the study of the role (s) of GASP proteins in a myogenic context. These results constitute a solid experimental basis for considering the identification of their target protease(s). As we have already stated, the maturation of TGF-β and more specifically myostatin involve different proteases. By following a similar experimental approach, we plan to test the inhibitory activity of GASP proteins against another serine protease, the furin, which is directly involved in the maturation of myostatin. In the same way, it would be interesting to determine whether GASP-1 and GASP-2 have anti-protease properties against certain metalloproteases such as the BMP-1 / tolloid protein which is involved in the cleavage of the latent complex of the myostatin.

More generally, the GASP proteins are defined as heterotypic multi-domain inhibitors. To date, no study has been published yet concerning the potential anti-protease properties of other WAP, Kazal and the first Kunitz domains. As done by Patthy's team for the second Kunitz domain, the expression of each isolated domain could be considered (Nagy *et al*., 2003).

Proteases are involved in a large number of physiological functions or pathological disorders. It seems unavoidable to explore all the mechanisms that regulate the activity of these enzymes. Among all these mechanisms, proteins that have anti-protease

properties present obviously a major role. Understanding the mode of action of these inhibitors and identifying their cellular target(s) constitute today an exciting challenge in the field of basic or applied research.

BIBLIOGRAPHIC REFERENCES

Allen D., Unterman T. (2006). Regulation of myostatin expression and myoblast differentiation by FoxO and SMAD transcription factors. *Am. J. Physiol. Cell. Physiol*. **292**: 188-199.

Allen, D.L., Loh, A.S. (2011). Posttranscriptional mechanisms involving microRNA-27a and b contribute to fast-specific and glucocorticoid-mediated myostatin expression in skeletal muscle. *Am. J. Physiol. Cell Physiol.* 300: C124-137.

Amirouche A., Durieux A.C., Banzet S., Koulmann N., Bonnefoy, R., Mouret C., Bigart X., Peinnequin A., Freyssenet D. (2009). Down-regulation of Akt/mammalian target of rapamycin signaling pathway in response to myostatin overexpression in skeletal muscle. *Endocrinol.* **150**(1): 286-294.

Amthor H., Nicholas G., McKinnell I., Kemp CF., Sharma M., Kambadur R., Patel K. (2004). Follistatin complexes myostatin and antagonises myostatin-mediated inhibition of myogenesis. *Dev. Biol*. **270**: 19-30.

Anderson S.B., Goldberg A.L., Whitman M. (2008). Identification of a novel pool of extracellular pro-myostatin in skeletal muscle. *J. Biol. Chem.* **283**(11): 7027-7035.

Aoki M.S., Soares A.G., Miyabara E.H., Baptista I.L., Moriscot A.S. (2009). Expression of genes related to myostatin signaling during rat skeletal muscle longitudinal growth. *Muscle Nerve.* **40**: 992-999.

Araùjo C.L., Bezerra I.W., Oliveira A.S., Moura F.T., Macedo L.L., Gomes C.E., Barbosa A.E., Macedo F.P., Souza T.M., Franco O.L. (2005). *In vivo* bioinsecticidal activity toward Ceratis capitate (fruit fly) and Callosobruchus maculatus (cowpea weevil) and *in vitro* bioinsecticidal activity toward different orders of insect pests of a trypsin inhibitor purified from tamarind tree (Tamarindus indica) seeds. *J. Agric. Food Chem*. **53**: 4381-4387.

Ascenzi P., Bocedi A., Bolognesi M., Spallarossa A., Coletta M., De Cristofaro R., Menegatti E. (2003). The bovine basic pancreatic trypsin inhibitor (Kunitz inhibitor): a milestone protein. *Curr. Prot. Pept. Sci*. **4**: 231-251.

Bailey A.J., Enser M.B., Dransfield E., Restall D.J., Avery N.C. (1982). Muscle and adipose tissue from normal and double muscled cattle: Collagen types, muscle fiber diameter, fat cell size and fatty acid composition and organoleptic properties. *Curr. Top. Vet. Med. Anim. Sci.* **16**: 178-202.

Banyai L., Patthy L. (1999). The NTR module: domains of netrins, secreted frizzled related proteins, and type I procollagen C-proteinase enhancer proteins are homologous with tissue inhibitors of metalloproteases. *Protein Sci.* **8**(8): 1636-1642.

Barrett A.J., Rawlings N.D. (1995) Families and clans of serine peptidases. *Arch. Biochem. Biophys*. **318**(2): 247-250.

Barua S., Kuizon S., Chadman K.K, Flory M.J., Brown W.T., Junaid M.A. (2014). Single-base resolution of mouse offspring brain methylome reveals epigenome modifications caused by gestational folic acid. *Epigenetics Chromatin* **7**: 3.

Bekhouche M., Kronenberg D., Vadon-Le Goff S., Bijakowski C., Lim N.H., Font B., Kessler E., Colige A., Nagase H., Murphy G., Hulmes D.J., Moali C. (2010). Role of the netrin-like domain of procollagen C-proteinase enhancer-1 in the control of metalloproteinase activity. *J. Biol. Chem.* **285**: 15950-15959.

Bell M.L., Buvoli M., Leinwand L.A. (2010). Uncoupling of Expression of an Intronic MicroRNA and Its Myosin Host Gene by Exon Skipping. *Mol. Cell Biol*. **30**: 1937-1945.

Bentzinger C.F., Wang Y.X., Rudnicki M.A. (2012). Building muscle: molecular regulation of myogenesis. *Cold Spring Harb. Perspect Biol*. **4**(2).

Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., Dinola A., Haak J.R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Physics*. **81**(8): 3684-3690.

Bernardi H., Gay S., Fedon Y., Vernus B., Bonnieu A., Bacou F. (2011). Wnt4 activates the canonical beta-catenin pathway and regulates negatively myostatin: functional implication in myogenesis. *Am. J. Physiol. Cell Physiol.* **300**(5): C1122-1138.

Berry C., Thomas M., Langley B., Sharma M., Kambadur R. (2002). Single cysteine to tyrosine transition inactivates the growth inhibitory function of Piedmontese myostatin. *Am. J. Physiol. Cell Physiol.* **283**: C135-C141.

Biga P.R., Schelling G.T., Hardy R.W., Cain K.D., Overturf K., Ott T.L. (2004). The effects of recombinant bovine somatotropin (rbST) on tissue IGF-1, IGF-1 receptor, and GH mRNA levels in rainbow trout, Oncorhynchus mykiss. *Gen. Comp. Endocrinol*. **135**: 324-333.

Blanchet X., Herrera-Mendez C.H., Becila S., Pélissier P., Delourme D., Coulis G., Sentandreu M.A., Boudjellal A., Brémaud L., Ouali A. (2009) Inhibition of human initiator caspase 8 and effector caspase 3 by cross-class inhibitory bovSERPINA3-1 and A3-3. *FEBS Lett*., **583**(17): 2743-2748.

Bode W., Huber R. (2000). Structural basis of the endoproteinase-protein inhibitor interaction. *Biochim. Biophys. Acta* **1477**(1-2): 241-252.

Boman I.A., Klemetsdal G., Blichfeldt T., Nafstad O., Våge D.I. (2009). A frameshift mutation in the coding region of the myostatin gene (MSTN) affects carcass conformation and fatness in Norwegian White Sheep. *Anim. Genet.* **40**(4): 418-422.

Bonala S., Lokireddy S., Arigela H., Teng S., Wahli W., Sharma M. (2012). Peroxisome proliferator-activated receptor b/d induces myogenesis by modulating myostatin activity. *J. Biol. Chem*. **287**: 12935-12951.

Bonala S., Lokireddy S., McFarlane C., Patnam S., Sharma M., Kambadur R. (2014). Myostatin induces insulin resistance via Casitas B-lineage lymphoma b (Cblb)-mediated degradation of insulin receptor substrate 1 (IRS1) protein in response to high calorie diet intake. *J. Biol. Chem.* **289**(11): 7654-7670.

Bongiorni S., Valentini A., Chillemi G. (2016). Structural and Dynamic Characterization of the C313Y Mutation in Myostatin Dimeric Protein, Responsible for the “Double Muscle” Phenotype in Piedmontese Cattle. *Front. Genet*. **7**: 14.

Bourboulia D., Stetler-Stevenson W.G. (2010). Matrix metalloproteinases (MMPs) and tissue inhibitors of metalloproteinases (TIMPs): positive and negative regulators in tumor cell adhesion. *Semin. Cancer Biol.* **20**: 161-168.

Bowyer J.F., Thomas M., Patterson T.A., George N.I., Runnells J.A., Levi M.S. (2012). A visual description of the dissection of the cerebral surface vasculature and associated meninges and the choroid plexus from rat brain. *J. Vis. Exp.* **69**: e4285.

Bowyer J.F., Patterson T.A., Saini U.T., Hanig J.P., Thomas M., Camacho L. (2013). Comparison of the global gene expression of choroid plexus and meninges and associated vasculature under control conditions and after pronounced hyperthermia or amphetamine toxicity. *BMC Genomics* **14**: 147.

Bradford M.M. (1976). A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**: 248-254.

Breitbart A., Auger-Messier M., Molkentin J.D., Heineke J. (2011). Myostatin from the heart: local and systemic actions in cardiac failure and muscle wasting. *Am. J. Physiol. Heart Circ. Physiol.* **300**(6): H1973-1982.

Brew K., Nagase H. (2010). The tissue inhibitors of metalloproteinases (TIMPs): an ancient family with structural and functional diversity. *Biochim. Biophys. Acta* **1803**: 555-571.

Brun C., Monestier O., Legardinier S., Maftah A., Blanquet V. (2012). Murine GASP-1 N-Glycosylation is not essential for its Activity on C2C12 Myogenic Cells but Alters its Secretion. *Cell Physiol. Biochem*. **30**(3): 791-804.

Brun C., Périé L., Baraige F., Vernus B., Bonnieu A., Blanquet V. (2014). Absence of hyperplasia in Gasp-1 overexpressing mice is dependent on myostatin up-regulation. *Cell Physiol. Biochem*. **34**(4): 1241-1259.

Brun C. (2013). Etude des conséquences fonctionnelles de l'inhibition de la myostatine par GASP-1 au cours du développement musculaire. Thèse N° 45-2013.

Buckingham M. (2006). Myogenic progenitor cells and skeletal myogenesis in vertebrates. *Curr. Opin. Genet. Dev*. **16**(5): 525-532.

Callis T.E., Pandya K., Seok H.Y., Tang R.H., Tatsuguchi M., Huang Z.P., Chen J.F., Deng Z., Gunn B., Shumate J., Willis M.S., Selzman C.H., Wang D.Z. (2009). MicroRNA-208a is a regulator of cardiac hypertrophy and conduction in mice. *J. Clin. Invest.* **119**(9): 2772-2786.

Camporez J., Max P., Abulizi A., Gabriela M., Michael J., Glenn F., Christopher H., Kitt P., Gerald S. (2016). Anti-myostatin antibody increases muscle mass and strength and improves insulin sensitivity in old mice. *Proc. Natl. Acad. Sci. USA* **113**(8): 2212-2217.

Carneiro I., Castro-Piedras I., Muñoz A., Labandeira-García J.L., Devesa J., Arce V.M. (2008). Hypothyroidism is associated with increased myostatin expression in rats. *J. Endocrinol. Invest.* **31**: 773-778.

Carrell R., Travis J. (1985). α1-Antitrypsin and the serpins: variation and countervariation. *Trends Biochem. Sci*. **10**: 20-24.

Carlos L., Judith S.B. (2008). Proteases: Multifunctional Enzymes in Life and Disease. *J. Biol. Chem.* **283**: 30433-30437.

Case D.A., Ben-Shalom I.Y., Brozell S.R., Cerutti D.S., Cheatham T.E., Cruzeiro V.W.D., Darden T.A., Duke R.E., Ghoreishi D., Gilson M.K., Gohlke H., Goetz A.W., Greene D., Harris R., Homeyer N., Izadi S., Kovalenko A., Kurtzman T., Lee T.S., LeGrand S., Li P., Lin C., Liu J., Luchko T., Luo R., Mermelstein D.J., Merz K.M., Miao Y., Monard G., Nguyen C., Nguyen H., Omelyan I., Onufriev A., Pan F., Qi R., Roe D.R., Roitberg A., Sagui C., Schott-Verdugo S., Shen J., Simmerling C.L., Smith J., Salomon-Ferrer R., Swails J., Walker R.C., Wang J., Wei H., Wolf R.M., Wu X., Xiao L., York D.M. and Kollman P.A. (2018). AMBER 2018, University of California, San Francisco.

Cash J.N., Rejon C.A., McPherron A.C., Bernard D.J., Thompson T.B. (2009). The structure of myostatin: follistatin 288: insights into receptor utilization and heparin binding. *EMBO J.* **28**(17): 2662-2676.

Cash J.N., Angerman E.B., Kattamuri C., Nolan K., Zhao H., Sidis Y., Keutmann H.T., Thompson T.B. (2012). Structure of Myostatin Follistatin-Like 3 N-Terminal Domains of Follistatin-Type Molecules Exhibit Alternate Modes of Bindings*. J. Biol. Chem.* **287**: 1043-1053.

Cauwe B., Van den Teen P.E., Opdenakker G. (2007). The biochemical, biological and pathological kaleidoscope of cell surface substrates processed by matrix metalloproteinases. *Crit. Rev. Biochem. Mol. Biol.* **42**(3): 113-185.

Chapman H.A., Riese R.J., Shi G.P. (1997). Emerging roles for cysteine proteases in human biology. *Annu. Rev. Physiol.* **59**: 63-88.

Chase T., Shaw E. (1967). P-nitrophenyl-p'-guanidinobenzoate HCl: a new active site titrant for trypsin. *Biochem. Biophys. Res. Comm*. **29**(4): 508-514.

Chen D., Zhao M., Mundy G.R. (2004). Bone morphogenetic proteins. *Growth Factors* **22**(4): 233-41.

Chen Y., Ye J., Cao L., Zhang Y., Xia W., Zhu D. (2010a). Myostatin regulates glucose metabolism via the AMP-activated protein kinase pathway in skeletal muscle cells. *Int. J. Biochem. Cell Biol.* **42**: 2072-2081.

Chen Y.W., Lee M.S., Lucht A., Chou F.P., Huang W., Havighurst T.C., Kim K., Wang J.K., Antalis T.M., Johnson M.D., Lin C.Y. (2010b). A serine protease expressed in the prostate on the apical surface of luminal epithelial cells and released into semen in prostasomes, is misregulated in prostate cancer cells. *Am. J. Pathol.* **176**: 2986-2996.

Chen X., Huang Z., Chen D., Yang T., Liu G. (2013). MicroRNA-27a is induced by leucine and contributes to leucine-induced proliferation promotion in C2C12 cells. *Int. J. Mol. Sci*. 14(7): 14076-14084.

Cheong H.S., Lee H.C., Park B.L., Kim H., Jang M.J., Han Y.M. (2010). Epigenetic modification of retinoic acid-treated human embryonic stem cells. *BMB Rep*. **43**: 830-835.

Chhikara N., Saraswat M., Tomar A.K., Dey S., Singh S., Yadav S. (2012). Human epididymis protein-4 (HE-4): a novel cross-class protease inhibitor. *PLoS ONE* **7**: e47672.

Crawley J.T.B., Lane D.A. (2008). The haemostatic role of Tissue Factor Pathway Inhibitor. *Arterioscler. Thromb. Vasc. Biol*. **28**: 233-242.

Crowley M., Darden T., Cheatham T., Deerfield D. (1997). Adventures in Improving the Scaling and Accuracy of a Parallel Molecular Dynamics Program. *J. Supercomputing*. **11**(3): 255-278.

Dafforn T.R., Pike R.N., Bottomley S.P. (2004). Physical characterization of serpin conformations. *Methods* **32**(2): 150-158.

Dai X., Jiang W., Zhang Q., Xu L., Geng P., Zhuang S. Petrich B.G., Jiang C., Peng L., Bhattacharya S., Evans S.M., Sun Y., Chen J., Liang X. (2013). Requirement for integrin-linked kinase in neural crest migration and differentiation and outflow tract morphogenesis. *BMC Biol.* **11**: 107.

Darden T., York D., Pedersen L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Physics* **98**(12): 10089-10092.

DeClerck Y.A, Mercurio A.M., Stack M.S., Chapman H.A., Zutter M.M., Muschel R.J., Raz A., Matrisian L.M., Sloane B.F., Noel A. (2004). Proteases, extracellular matrix, and cancer: a workshop of the path B study section. *Am. J. Pathol.* **164**: 1131-1139.

Derynck R., Zhang Y.E. (2003). Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature* **425**: 577-584.

Derynck R., Akhurst R.J. (2007). Differentiation plasticity regulated by TGF-beta family proteins in development and disease. *Nat. Cell Biol*. **9**(9): 1000-1004

Derynck R., Miyazono K. (2008). The TGF-β Family. *Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press;* pp. 29-43

Domingo G., Iglesias A., Monserrat L., Sanchez L., Cantalapiedra J., Lorenzo J.M. (2015). Effect of crossbreeding with Limousine, Rubia Gallega and Belgium Blue on meat quality and fatty acid profile of Holstein calves. *Anim. Sci. J.* **86**: 913-921.

Drummond M.J., Glynn E.L., Fry C.S., Dhanani S., Volpi E., Rasmussen B.B. (2009). Essential amino acids increase microRNA-499, -208b, and -23a and downregulate myostatin and myocyte enhancer factor 2C mRNA expression in human skeletal muscle. *J. Nutr.* **139**(12): 2279-2284.

Eargle J., Wright D., Luthey-Schulten Z. (2006). Multiple Alignment of protein structures and sequences for VMD. *Bioinformatics* **22**(4): 504-506.

Erez E., Fass D., Bibi E. (2009). How intramembrane proteases bury hydrolytic reactions in the membrane. *Nature* **459**: 371-378.

Essmann U., Perera L., Berkowitz M.L., Darden T., Lee H., Pedersen L.G. (1995). A smooth particle mesh Ewald method. *J. Chem. Physics*. **103**(19): 8577-8593.

Farady C.J., Craik C.S. (2010). Mechanisms of macromolecular protease inhibitors. *Chem. Biochem*. **11:** 2341-2346.

Fassina G., Ferrari N., Brigati C., Benelli R., Santi L., Noonan D.M., Albini A. (2000). Tissue inhibitors of metalloproteases: Regulation and biological activities. *Clin. Exp. Metastasis* **18**: 111-120.

Fiems L.O. (2012). Double muscling in cattle: Genes, husbandry, carcasses and meat. *Animals (Basel)* **2**: 472-506.

Fischer G., Tradler T., Zarnt T. (1998). The mode of action of peptidyl prolyl cis/trans isomerases *in vivo*: binding versus catalysis. *FEBS Lett*. **426:** 17-20.

Forbes D., Jackman M., Bishop A., Thomas M., Kambadur R., Sharma M. (2006). Myostatin auto-regulates its expression by feedback loop through Smad7 dependent mechanism. *J. Cell Physiol.* **206**(1): 264-272.

Fujita S., Dreyer H.C., Drummond M.J., Glynn E.L., Cadenas J.G., Yoshizawa F., Volpi E., Rasmussen B.B. (2007). Nutrient signalling in the regulation of human muscle protein synthesis. *J. Physiol.* **582**(Pt 2): 813-823.

Ge X., Vajjala A., McFarlane C., Wahli W., Sharma M., Kambadur R. (2012). Lack of Smad3 signaling leads to impaired skeletal muscle regeneration. *Am. J. Physiol. Endocrinol. Metab.* **303**(1): 90-102.

Georgiou G., Valax P. (1999). Isolating inclusion bodies from bacteria. *Meth. Enzymol*. **309**: 48-58.

Gerlach L.M., Hutson M.R., Germiller J.A., Nguyen-Luu D., Victor J.C., Barald K.F. (2000). Addition of the BMP4 antagonist, noggin, disrupts avian inner ear development. *Development* **127**: 45-54.

Gerngross T.U. (2004). Advances in the production of human therapeutic proteins in yeast and filamentous fungi. *Nat. Biotechnol*. **22**: 1409-1414.

Gettins P.G.W. (2002). Serpin structure, mechanism, and function. *Chem. Rev.* **102**: 4751-4803.

Glass D.J. (2010). PI3 kinase regulation of skeletal muscle hypertrophy and atrophy. *Curr. Top. Microbiol. Immunol.* **346**: 267-278.

Goodale B.C., Tilton S.C., Corvi M.M., Wilson G.R., Janszen D.B., Anderson K.A. (2013). Structurally distinct polycyclic aromatic hydrocarbons induce differential transcriptional responses in developing zebrafish, *Toxicol. Appl. Pharmacol.* **272**: 656-670.

Grobet L., Martin L.J., Poncelet D., Pirottin D., Brouwers B., Riquet J., Schoeberlein A., Dunner S., Ménissier F., Massabanda J., Fries R., Hanset R., Georges M. (1997). A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. *Nat. Genet.* **17**(1): 71-74.

Gulley M.M., Zhang X., Michel K. (2013). The roles of serpins in mosquito immunology and physiology. *J. Insect Physiol*. **59**: 138-147.

Haidet A.M., Rizo L., Handy C., Umapathi P., Eagle A., Shilling C. (2008). Long-term enhancement of skeletal muscle mass and strength by single gene administration of myostatin inhibitors. *Proc. Natl. Acad. Sci. USA.* **105**: 4318-4322.

Harper S., Speicher D.W. (2011). Purification of proteins fused to Glutathione S-transferase. *Methods Mol. Biol.* **681**: 259-280.

Harris R.A., Tardif S.D., Vinar T., Wildman D.E., Rutherford J.N., Rogers J. (2014). Evolutionary genetics and implications of small size and twinning in callitrichine primates. *Proc. Natl. Acad. Sci. USA*. **111**: 1467-1472.

Hathout Y., Brody E., Clemens. P.R., Cripe L., DeLisle R.K., Furlong P. (2015). Large-scale serum protein biomarker discovery in Duchenne muscular dystrophy. *Proc. Natl. Acad. Sci. USA*. **112**: 7153-7158.

Hedstrom L. (2002) Serine protease mechanism and specificity. *Chem. Rev*. **102**(12): 4501-4524.

Heldin C.H., Miyazono K., Ten Dijke P. (1997). TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature* **390**: 465-471.

Hellwig S., Drossard J., Twyman R.M., Fischer R. (2004). Plant cell cultures for the production of recombinant proteins. *Nat. Biotechnol*. **22**: 1415-1422.

Hill J.J., Davies M.V., Pearson A.A., Wang J.H., Hewick R.M., Wolfman N.M., Qiu Y. (2002). The myostatin propeptide and the follistatin-related gene are inhibitory binding proteins of myostatin in normal serum. *J. Biol. Chem.* **277**(43): 40735-41.

Hill J.J., Qiu Y., Hewick R.M. Wolfman N.M. (2003). Regulation of myostatin *in vivo* by growth and differentiation factor-associated serum protein-1: a novel protein with protease inhibitor and follistatin domains. *Mol. Endocrinol*. **17**(6): 1144-1154.

Hinck A.P. (2012). Structural studies of the TGF-βs and their receptors - insights into evolution of the TGF-β superfamily. *FEBS Lett.* **586**(14): 1860-1870.

Hittel D.S., Berggren J.R., Shearer J., Boyle K., Houmard J.A. (2009). Increased secretion and expression of myostatin in skeletal muscle from extremely obese women. *Diabetes* **58**(1): 30-38.

Hjorth M., Pourteymour S., Gorgens S.W., Langleite T.M., Lee S., Holen T., Gulseth H.L., Birkeland K.I., Jensen J., Drevon C.A., Norheim, F. (2016). Myostatin in relation to physical activity and dysglycaemia and its effect on energy metabolism in human skeletal muscle cells. *Acta Physiol*. **217:** 45-60.

Huang Z., Chen D., Zhang K., Yu B., Chen X., Meng J. (2007). Regulation of myostatin signaling by c-Jun N-terminal kinase in C2C12 cells. *Cell Signal.* **19**(11): 2286-2295.

Huang Z., Chen X., Chen D. (2011). Myostatin: A novel insight into its role in metabolism, signal pathways, and expression regulationI. *Cell Signal.* **23**(9): 1441-1446.

Huang Z., Chen X., Yu B., He J., Chen D. (2012). MicroRNA-27a promotes myoblast proliferation by targeting myostatin. *Biochem. Biophys. Res. Commun*. **423**(2): 265-269.

Hughes M.W., Wu P., Jiang T.X., Lin S.J., Dong C.Y., Li A. (2011). In search of the Golden Fleece: unraveling principles of morphogenesis by studying the integrative biology of skin appendages. *Integr. Biol. (Camb).* **3**: 388-407.

Humphrey W., Dalke A., Schulten K. (1996). VMD: Visual molecular dynamics. *J. Mol. Graphics* **14**(1): 33-38.

Huntington J.A., Read R.J., Carrell R.W. (2000). Structure of a Serpin-protease complex shows inhibition by deformation. *Nature* **407**: 923-926.

Im H., Ahn H.Y., Yu M.H. (2000). Bypassing the kinetic trap of serpin protein folding by loop extension. *Protein Sci.* **9**(8): 1497-1502.

Inman G.J., Nicolas F.J., Callahan J.F., Harling J.D., Gaster L.M., Reith A.D. (2002). SB-431542 is a potent and specific inhibitor of transforming growth factor- beta superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7. *Mol. Pharmacol.* **62**: 65-74.

Irving J.A., Pike R.N., Lesk A.M., Whisstock J.C. (2000). Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function. *Genome Res*. **10**(12): 1845-1864.

Iwasaki S., Miyake M., Watanabe H., Kitagawa E., Watanabe K., Ohwada S., Kitazawa H., Rose M.T., Aso H. (2013). Expression of myostatin in neural cells of the olfactory system. *Mol. Neurobiol*. **47**(1): 1-8.

Iyer R.P., Patterson N.L., Fields G.B., Lindsey M.L. (2012). The history of matrix metalloproteinases: milestones, myths and misperceptions. *Am. J. Physiol. Heart Circ. Physiol*. **303**(8): H919-H930.

Janssen S.F., Gorgels T.G., Bossers K., Ten Brink J.B., Essing A.H., Nagtegaal M. (2012). Gene expression and functional annotation of the human ciliary body epithelia. *PLoS ONE* **7**: e44973.

Jeanplong F., Bass J.J., Smith H.K. (2003). Prolonged underfeeding of sheep increases myostatin and myogenic regulatory factor Myf-5 in skeletal muscle while IGF-1 and myogenin are repressed. *J. Endocrinol.* **176**(3): 425-437.

Ji W.R, Castellino F.J., Chang Y., Deford M.E., Gray H., Villarreal X., Kondri M.E., Marti D.N., Llinás M., Schaller J., Kramer R.A., Trail P.A. (1998). Characterization of kringle domains of angiostatin as antagonists of endothelial cell migration, an important process in angiogenesis. *FASEB J.* **12**(15): 1731-1738.

Joulia D., Bernardi H., Garandel V., Rabenoelina F., Vernus B., Cabelloa G. (2003). Mechanisms involved in the inhibition of myoblast proliferation and differentiation by myostatin. *Exp. Cell Res.* **286**: 263-275.

Joung I.S., Cheatham T.E. (2008). Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*. **112**(30): 9020-9041.

Joung I.S., Cheatham T.E. (2009). Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters. *J. Phys. Chem. B*. 113(40): 13279-13290.

Kambadur R., Sharma M., Smith T.L., Bass J.J. (1997). Mutations in myostatin (GDF8) in double-muscled Belgian Blue and Piedmontese cattle. *Genome Res.* **7**: 910-916.

Kanzaki S., Ariyoshi W., Takahashi T., Okinaga T., Kaneuji T., Mitsugi S. (2011). Dual effects of heparin on BMP-2-induced osteogenic activity in MC3T3-E1 cells. *Pharmacol. Rep.* **63**: 1222-1230.

Kassell B. (1970). Bovine trypsin-kallikrein inhibitor (Kunitz inhibitor, basic pancreatic trypsin inhibitor, polyvalent inhibitor from bovine organs). *Methods Enzymol*. **19**: 844-852.

Kayode O., Wang R., Pendlebury D.F., Cohen I., Henin R.D., Hockla A., *et al*. (2016). An Acrobatic Substrate Metamorphosis Reveals a Requirement for Substrate Conformational Dynamics in Trypsin Proteolysis. *J. Biol. Chem*. **291**(51): 26304-26319.

Kazal L.A., Spicer D.S., Brahinsky R.A. (1948). Isolation of a Crystalline Trypsin Inhibitor-Anticoagulant Protein from Pancreas. *J. Am. Chem. Soc*. **70**: 3034-3040.

Kędzior M., Seredyński R., Gutowicz J. (2016). Microbial inhibitors of cysteine proteases. *Med. Microbiol. Immunol.* **205**(4): 275-296.

Keutmann H.T., Schneyer A.L., Sidis Y. (2004). The Role of Follistatin Domains in Follistatin Biological Action. *Mol. Endocrinol*. **18**(1): 228-240.

Kim Y.S., Bobbili N.K., Paek K.S., Jin H.J. (2006). Production of a monoclonal anti-myostatin antibody and the effects of in ovo administration of the antibody on posthatch broiler growth and muscle mass. *Poult. Sci.* **85**: 1062-1071.

Kingsley D.M. (1994). The TGF-beta superfamily: new members, new receptors, and new genetic tests of function in different organisms. *Genes Dev.* **8**(2): 133-146.

Kocamis H., Killefer J. (2002) Myostatin expression and possible functions in animal muscle growth. *Domest. Anim. Endocrinol.* **23**(4): 447-454.

Kohda T., Kishigami S., Kaneko-Ishino T., Wakayama T., Ishino F. (2012). Gene expression profile normalization in cloned mice by trichostatin A treatment, Cell Reprogram. *Cell Reprogram.* **14**: 45-55.

Kondas K., Szlama G., Trexler M., Patthy L. (2008). Both WFIKKN1 and WFIKKN2 have high affinity for growth and differentiation factors 8 and 11. *J. Biol. Chem.* **283**(35): 23677-23684.

Kondas K., Szlama G., Nagy A., Trexler M., Patthy L. (2011). Biological functions of the WAP domain-containing multidomain proteins WFIKKN1 and WFIKKN2. *Biochem. Soc. Trans.* **39**: 1416-1420.

Krowarsch D., Cierpicki T., Jelen F., Otlewski J. (2003). Canonical protein inhibitors of serine proteases. *Cell. Mol. Life Sci*. **60**: 2427-2444.

Kunihiro T., Nakatani M., Hitachi K., Uezumi A., Sunada Y., Ageta H., Inokuchi K. (2009). Activin signaling as an emerging target for therapeutic interventions. *Cell Commun. Signal.* **7**: 15.

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **277**: 680-685.

Langley B., Thomas M., Bishop A., Sharma M., Gilmour S., Kambadur R. (2002). Myostatin inhibits myoblast differentiation by down-regulating MyoD expression. *J. Biol. Chem.* **277**: 49831-49840.

Laskowski M., Kato I. (1980). Protein inhibitors of proteinases. *Annu. Rev. Biochem.* **49**: 593-626.

Laskowski M.J., Qasim M.A., Lu S.M., Kleanthous C. (2000). Protein-Protein Recognition. *FEBS Lett.* **31**: 228-270.

Leclere L., Rentzsch F. (2012). Repeated evolution of identical domain architecture in metazoan netrin domain-containing proteins. *Genome Biol. Evol.* **4**: 883-899.

Lee S.J. (2004). Regulation of muscle mass by myostatin. *Annu. Rev. Cell Dev. Biol*. **20**: 61-86.

Lee S.J. (2007). Quadrupling muscle mass in mice by targeting TGF-beta signaling pathway. *PLoS One* **2**: e789.

Lee S.J. (2008). Genetic Analysis of the Role of Proteolysis in the Activation of Latent Myostatin. *PLoS One* **3**: e1628.

Lee S.J., McPherron A.C. (2001). Regulation of myostatin activity and muscle growth. *Proc. Natl. Acad. Sci. USA* **98**(16): 9306-9311.

Lee Y.S., Lee S.J. (2013). Regulation of GDF-11 and myostatin activity by GASP-1 and GASP-2. *Proc. Natl. Acad. Sci. USA.* **110**(39): 3713-3722.

Letunic I., Doerks T., Bork P. (2015). SMART: recent updates, new developments and status in 2015, *Nucl. Acids Res.* **43**: 257-260.

Li Y., Qian Y.Q. Ma, W.M., Yang W.J. (2009). Inhibition Mechanism and the Effects of Structure on Activity of Male Reproduction-Related Peptidase Inhibitor Kazal-Type (MRPINK) of Macrobrachium rosenbergii. *Mar. Biotechnol*. ***11:*** 252-259.

Liddelow S.A., Temple S., Møllgård K., Gehwolf R., Wagner A., Bauer H. (2012). Molecular characterisation of transport mechanisms at the developing mouse blood-CSF interface: a transcriptome approach. *PLoS One* **7**: e33554.

Liepinsh E., Nagy A., Trexler M., Patthy L., Otting G. (2006). Second Kunitz-type protease inhibitor domain of the human WFIKKN1 protein. *J. Biomol. NMR* **35**(1): 73-78.

Ma K., Mallidis C., Artaza J., Taylor W., Gonzalez-Cadavid N., Bhasin S. (2001). Characterization of 5′-regulatory region of human myostatin gene: regulation by dexamethasone *in vitro. Am. J. Physiol. Endocrinol. Metab.* **281**: 1128-1136.

Ma K., Mallidis C., Bhasin S., Mahabadi V., Artaza J., Gonzales- Cadavid N., Arias J., Salehian B. (2003). Glucocorticoid-induced skeletal muscle atrphy is associated with upregulation of myostatin gene expression. *Am. J. Physiol. Endocrinol. Metab*. **285**(2): E363-371.

Machado R.J., Monteiro N.K., Migliolo L., Silva O.N., Pinto M.F., Oliveira A.S., Franco O.L., Kiyota S., Bemquerer M.P., Uchoa A.F. (2013). Characterization and pharmacological properties of a novel multifunctional Kunitz inhibitor from Erythrina velutina seeds. *PLoS ONE* **8**: e63571.

Mägert H.J., Kreutzmann P., Ständker L., Walden M., Drögemüller K., Forssmann W.G. (2002). LEKTI: a multidomain serine proteinase inhibitor with pathophysiological relevance. *Int. J. Biochem. Cell Biol.* **34**: 573-576.

Maier J.A., Martinez C., Kasavajhala K., Wickstrom L., Hauser K.E., Simmerling C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**(8): 3696-3713.

Maino B., Ciotti M.T., Calissano P., Cavallaro S. (2014). Transcriptional analysis of apoptotic cerebellar granule neurons following rescue by gastric inhibitory polypeptide. *Int. J. Mol. Sci.* **15**: 5596-5622.

Marszal E., Shrake A. (2006). Serpin crystal structure and serpin polymer structure. *Arch. Biochem. Biophys*. **453**(1): 123-129.

Matsakas A. (2014). Myostatin tilts the balance between skeletal muscle size, function and metabolism. *Exp. Physiol.* **99**(3): 469-470.

Massague J. (1998). TGF-beta signal transduction. *Annu. Rev. Biochem.* **67**: 753-791.

Matzuk M.M., Lu N., Vogel H., Sellheyer K., Roop D.R., Bradley A. (1995). Multiple defects and perinatal death in mice deficient in follistatin. *Nature* **374**(6520): 360-363.

McFarlane C., Plummer E., Thomas M., Hennebry A., Ashby M., Ling N., Smith H., Sharma M., Kambadur R. (2006). Myostatin induces cachexia bu activating the ubiquitin proteolytic system through an NF-kappaB-independent, FoxO1-dependent mechanism. *J. Cell Physiol.* **209**(1): 501-514.

McFarlane C., Hennebry A., Thomas M., Plummer E., Ling N., Sharma M., Kambadur R. (2008). Myostatin signals through Pax7 to regulate satellite cell self-renewal. *Exp. Cell. Res.* **314**(2): 317-329.

McPherron A.C. (2010). Metabolic functions of myostatin and GDF11. *Immunol. Endocr. Metab. Agents Med. Chem.* **10**(4): 217-231.

McPherron A.C., Lee S.J. (1997). Double muscling in cattle due to mutations in the myostatin gene. *Proc. Natl. Acad. Sci. USA* **94**(23): 12457-12461.

McPherron A.C., Lee S.J. (2002). Suppression of body fat accumulation in myostatin-deficient mice. *J. Clin. Invest.* **109**: 595-601.

McPherron A.C., Lawler A.M., Lee. S.J. (1997). Regulation of skeletal muscle mass in mice by a new TGF-β superfamily member. *Nature* **387**: 83-90.

McPherron A.C., Guo T., Bond N., Gavrilova O. (2013). Increasing muscle mass to improve metabolism. *Landes Bioscience, Adipocyte* **2**: 92-98.

Megan M., Ilona J. (2015). Respiratory protease/antiprotease balance determines susceptibility to viral infection and can be modified by nutritional antioxidants. *Am. J. Physiol. Lung Cell Mol. Physiol.* **308**: 1189-1201.

Ménissier F. (1982). General survey of the effect of double muscling on cattle performance. In: Muscle hypertrophy of genetic origin and its use to improve beef production. King J.W.B., Ménissier F. (Eds). *Curr. Top. Vet. Med. Anim. Sci., Martinus Nijhoff, The Netherlands* **16**: 21-53.

Merrick B.A., Phadke D.P., Auerbach S.S., Mav D., Stiegelmeyer S.M., Shah R.R. (2013). RNA-Seq profiling reveals novel hepatic gene expression pattern in aflatoxin B1 treated rats. *PLoS One* **8**: e61768.

Miar Y., Salehi A., Kolbehdar D., Aleyasin A. (2014). Application of myostatin in sheep breeding programs. *Mol. Biol. Res. Commun.* **3**(1): 33-43.

Mille-Hamard L., Billat V.L., Henry E., Bonnamy B., Joly F., Benech P. (2012). Skeletal muscle alterations and exercise performance decrease in erythropoietin-deficient mice: a comparative study. *BMC Med. Genomics.* **29**: 5-29.

Miretti S., Martignani E., Accornero P., Baratta M. (2013). Functional effect of mir-27b on myostatin expression: a relationship in piedmontese cattle with double-muscled phenotype. *BMC Genomics.* **14**: 194.

Miura T., Kishioka Y., Wakamatsu J., Hattori A., Hennebry A., Berry C.J., Sharma M., Kambadur R., Nishimura T. (2006). Decorin binds myostatin and modulates its activity to muscle cells. *Biochem. Biophys. Res. Commun.* **340**: 675-680.

Monestier O., Brun C., Heu K., Passet B., Malhouroux M., Magnol L., Vilotte J.L., Blanquet V. (2012a). Ubiquitous *Gasp1* overexpression in mice leads mainly to a hypermuscular phenotype. *BMC Genomics.***13**: 541.

Monestier O., Brun C., Cocquempot O., Petit D., Blanquet V. (2012b). GASP/WFIKKN Proteins: Evolutionary Aspects of Their Functions. *PLoS One* **7**: e43710.

Monestier O. and Blanquet V. (2016). WFIKKN1 and WFIKKN2: "Companion" proteins regulating TGFβ activity. C*ytokine Growth Factor Rev.* **32**: 75-84.

Mori Y., Ishida W., Bhattacharyya S., Li Y., Platanias L.C., Varga J. (2004). Selective inhibition of activin receptor-like kinase 5 signaling blocks profibrotic transforming growth factor beta responses in *skin fibroblasts, Arthritis Rheum*. **50**: 4008-4021.

Mosher D.S., Quignon P., Bustamante C.D., Sutter N.B., Mellersh C.S., Parker H.G., Ostrander E.A. (2007). A mutation in the myostatin gene increases muscle mass and enhance racing performance in heterozygote dogs. *PLoS Genet.* **3**: e79.

Mouisel E., Relizani K., Mille-Hamard L., Denis R., Hourdé C., Agbulut O., Patel K., Arandel L., Morales-Gonzalez S., Vignaud A., Garcia L., Ferry A., Luquet S., Billat V., Ventura-Clapier R., Schuelke M., Amthor H. (2014). Myostatin is a key mediator between energy metabolism and endurance capacity of skeletal muscle. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **307**: R444-R454.

Naftaly S., Cohen I., Shahar A., Hockla A., Radisky E.S., Papo N. (2018). Mapping protein selectivity landscapes using multi-target selective screening and next-generation sequencing of combinatorial libraries. *Nature Commun.* **9**(1): 3935.

Nagaraj N.S., Datta P.K. (2010). Targeting the transforming growth factor-beta signaling pathway in human cancer. *Expert Opin. Investig. Drugs* **19**(1): 77-91.

Nagy A., Trexler M., Patthy L. (2003). Expression, purification and characterization of the second Kunitz-type protease inhibitor domain of the human WFIKKN protein. *J. Biochem*. **270**(9): 2101-2107.

Nakamura T., Takio K., Eto Y., Shibai H., Titani K., Sugino H. (1990). Activin-binding protein from rat ovary is follistatin. *Science* **247**(4944): 836-838.

Nakashima T., Pak S.C., Silverman G.A., Spring P.M., Frederick M.J., Clayman G.L. (2000). Genomic cloning, mapping, structure and promoter analysis of HEADPIN, a Serpin which is down-regulated in head and neck cancer cells. *Biochim. Biophys. Acta* **1492**: 441-446.

Nicholas G., Thomas M., Langley B., Somers W., Patel K., Kemp C.F., Sharma M. Kambadur R. (2002). Titincap associates with and regulates secretion of Myostatin. *J. Cell. Physiol*. **193**: 120-131.

Nishida A.T., Kobuke K., Kojima K., Ito J., Honjo T., Tashiro K. (2004). OC29 is preferentially expressed in the presumptive sensory organ region of the otocyst. *Dev. Dyn.* **231**: 766-774.

Oh J, Takahashi R, Kondo S, Mizoguchi A, Adachi E, Sasahara RM, Nishimura S, Imamura Y, Kitayama H, Alexander DB, Ide C, Horan TP, Arakawa T, Yoshida H, Nishikawa S, Itoh Y, Seiki M, Itohara S., Takahashi C, Noda M. (2001). The membrane-anchored MMP inhibitor RECK is a key regulator of extracellular matrix integrity and angiogenesis. *Cell* **107**: 789-800.

Ohsawa Y., Okada T., Kuga A., Hayashi S., Murakami T., Tsuchida K., Noji S., Sunada Y. (2008) Caveolin-3 regulates myostatin signaling. *Acta Myol.* **27**(1): 19-24.

Okada Y. (2017). Proteinases and Matrix Degradation. In: *Kelley and Firestein's Textbook of Rheumatology* (Elsevier) pp. 106-125.

Oliver M.H., Harrison N.K., Bishop J.E., Cole P.J., Laurent G.J. (1989). A rapid and convenient assay for counting cells cultured in microwell plates: application for assessment of growth factors. *J. Cell Sci*. **92**(Pt 3): 513-518.

Olsson M.H.M., Søndergaard C.R., Rostkowski M., Jensen J.H. (2011). PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput*. **7**(2): 525-537.

Patston P.A. Gettins P. Beechem J. and Schapira M. (1991) Mechanism of serpin action: evidence that C1 inhibitor functions as a suicide substrate. *Biochemistry* **30**: 8876-8882.

Patel G.K., Gupta A.K., Gupta A., Mishra M., Singh P.K., Saxena A.K., Sharma A.K. (2014). Purification and physiocochemical characterization of a trypsin inhibitor from Cassia absus Linn. *Protein Pept. Lett*. **21**: 108-114.

Pendlebury D., Wang R., Henin R.D., Hockla A., Soares A.S., Madden B.J., *et al*. (2014). Sequence and conformational specificity in substrate recognition: several human Kunitz protease inhibitor domains are specific substrates of mesotrypsin. *J. Biol. Chem*. 289(47): 32783-32797.

Périé L. (Thèse **2015).** La myostatine et ses partenaires GASP-1 et GASP-2: Implications dans le développement musculaire et le métabolisme du glucose.

Périé L., Parenté A., Brun C., Magnol L., Pélissier P., Blanquet V. (2016). Enhancement of C2C12 myoblast proliferation and differentiation by GASP-2, a myostatin inhibitor. *Biochem. Biophys. Rep*. **6**: 39-46.

Périé L., Parenté A., Baraige F., Magnol L., Blanquet V. (2017). Alterations in adiposity and glucose homeostasis in adult Gasp-1 overexpressing mice. *Cell Physiol. Biochem*. **44**(5): 1896-1911.

Philip B., Lu Z., Gao Y. (2005). Regulation of GDF-8 signaling by the p38 MAPK. *Cell. Signal.* **17**: 365-375.

Piek E., Moustakas A., Kurisaki A., Heldin C.H., Dijke P. (1999). TGF-β type I receptor/ALK-5 and Smad proteins mediate epithelial to mesenchymal transdifferentiation in NMuMG breast epithelial cells. *J. Cell Sci.* **112**: 4557-4568.

Potempa J., Korzus E, Travis J. (1994). The serpin superfamily of proteinase inhibitors: structure, function, and regulation. *J. Biol. Chem*. **269**(23): 15957-15960.

Pratt L.R. 1950-, Hummer G. Simulation and theory of electrostatic interactions in solution. In AIP; 1999 [cited 2018 Oct 12]. Available from: http://agris.fao.org/agris-search/search.do?recordID=US201300040464

Price D.J., Brooks C.L. (2004). A modified TIP3P water potential for simulation with Ewald summation. *J. Chem. Phys*. **121**(20): 10096-10103.

Rachagani S., Cheng Y., Reecy J.M. (2010). Myostatin genotype regulates muscle-specific miRNA expression in mouse pectoralis muscle. *BMC Res. Notes.* **11**(3): 297.

Ragg H., Lokot T., Kamp P.B., Atchley W.R., Dress A. (2001). Vertebrate serpins: construction of a conflict free phylogeny by combining exon-intron and diagnostic site analyses. *Mol. Biol. Evol*. **18**(4): 577-584.

Ranasinghe S., McManus D.P. (2013). Structure and function of invertebrate Kunitz serine protease inhibitors. *Dev. Comp. Immunol*. **39**(3): 219-227.

Rawlings N.D., Barrett A.J. (1993). Evolutionary families of peptidases. *Biochem. J*. **290**(Pt 1): 205-218.

Rawlings N.D., Tolle D.P., Barrett A.J. (2004). Evolutionary families of peptidase inhibitors. *Biochem. J*. **378**(3): 705-716.

Rawlings N.D., Morton. F.R., Kok C.Y., Kong J., Barrett A.J. (2008). MEROPS: the peptidase database. *Nucl. Acids Res.* **36**: 320-325.

Rawlings N.D., Barrett A.J., Bateman A. (2012). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucl. Acids Res.* **40**: 343-350.

Read R.J., James M.N.G., Salvesen G.S., Barrett A.J. (1986). Proteinase Inhibitors. *Elsevier, New York*. **12**: 301-359.

Reardon K.A., Davis J., Kapsa R.M., Choong P., Byrne E. (2001). Myostatin, insulin-like growth factor-1, and leukemia inhibitory factor mRNAs are upregulated in chronic human disuse muscle atrophy. *Muscle Nerve* **24**(7): 893-899.

Rebbapragada A., Benchabane H., Wrana J.L., Celeste A. J., Attisano L. (2003). Myostatin signals through a transforming growth factor beta-like signaling pathway to block adipogenesis. *Mol. Cell. Biol.* **23**: 7230-7242.

Rescan P.Y., Montfort J., Fautrel A., Ralliere C., Lebret V. (2013). Gene expression profiling of the hyperplastic growth zones of the late trout embryo myotome using laser capture microdissection and microarray analysis. *BMC Genomics* **14**: 173.

Rexford S.A., Hyeong K.P. (2015). Connecting Myokines and Metabolism. *Endocrinol. Metab. (Seoul).* **30**(3): 235-245.

Rich D. (1985). Pepstatin-derived inhibitors of aspartic proteinases. A close look at an apparent transition-state analog inhibitor. *J. Med. Chem.* **28**: 263-273.

Richardson R.T., Sivashanmugam P., Hall S.H., Hamil K.G., Moore P.A., Ruben S.M., *et al*. (2001). Cloning and sequencing of human Eppin: a novel family of protease inhibitors expressed in the epididymis and testis. *Gene* **270**(1-2): 93-102.

Rimphanitchayakit V., Tassanakajon A. (2010). Structure and function of invertebrate Kazal-type serine proteinase inhibitors. *Dev. Comp. Immunol*. **34**(4): 377-386.

Rios R., Carneiro I., Arce V.M., Devesa J. (2001). Myostatin regulates cell survival during C2C12 myogenesis. *Biochem*. *Biophys. Res. Commun.* **280**: 561-566.

Rios R., Carneiro I., Arce V.M., Devesa J. (2002). Myostatin is an inhibitor of myogenic differentiation. *Am. J. Physiol. Cell Physiol.* **282**: C993-C999.

Roberts T.H., Hejgaard J., Saunders N.F., Cavicchioli R., Curmi P.M. (2004). Serpins in unicellular Eucarya, Archaea, and Bacteria: sequence analysis and evolution. *J. Mol. Evol.* **59**(4): 437-447.

Roberts E., Eargle J., Wright D., Luthey-Schulten Z. (2006). MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinforma*tics **7**(1): 382.

Rodgers B.D., Garikipati D.K. (2008). Clinical, agricultural, and evolutionary biology of myostatin: a comparative review. *Endocr. Rev*. **29**: 513-534.

Rodino-Klapac L.R., Haidet A. M., Kota J., Handy C., Kaspar B.K., Mendell R. (2009). Inhibition of myostatin with emphasis on follistatin as a therapy for muscle disease. *Muscle Nerve* **39**(3): 283-296.

Roe D.R., Cheatham T.E. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput*. **9**(7): 3084-3095.

Roforth M.M., Atkinson E.J., Levin E.R., Khosla S., Monroe D.G. (2014). Dissection of estrogen receptor alpha signaling pathways in osteoblasts using RNA- sequencing. *PLoS One* **9**: e95987.

Romi F., Helgeland G., Gilhus N.E. (2012). Serum levels of matrix metalloproteinases: implications in clinical neurology. *Eur. Neurol*. **67**: 121-128.

Roy A., Kucukural A., Zhang Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**(4): 725-738.

Ryan M.C., Itsiopoulos C., Thodis T., Ward G., Trost N., Hofferberth S., O'Dea K., Desmond P.V., Johnson N.A., Wilson A.M. (2013). The Mediterranean diet improves hepatic steatosis and insulin sensitivity in individuals with non-alcoholic fatty liver disease*. J. Hepatol*. **59**(1): 138-143.

Salameh M.A., Soares A.S., Hockla A., Radisky E.S. (2008). Structural basis for accelerated cleavage of bovine pancreatic trypsin inhibitor (BPTI) by human mesotrypsin. *J. Biol. Chem*. **283**(7): 4115-4123.

Salameh M.A., Soares A.S., Navaneetham D., Sinha D., Walsh P.N., Radisky E.S. (2010). Determinants of Affinity and Proteolytic Stability in Interactions of Kunitz Family Protease Inhibitors with Mesotrypsin. *J. Biol. Chem*. 285(47): 36884-36896.

Salehian B., Mahabadi V., Bilas J., Taylor W.E., Ma K. (2006). The effect of glutamine on prevention of glucocorticoid-induced skeletal muscle atrophy is associated with myostatin suppression. *Metabolism* **55**(9): 1239-1247.

Salerno M.S., Thomas M., Forbes D., Watson T., Kambadur R., Sharma M. (2004). Molecular analysis of fiber type-specific expression of murine myostatin promoter. *Am. J. Physiol. Cell Physiol.* **287**(4): C1031-1040.

Salvesen G.S. (1993) Can serine proteinase inhibitors inhibit cystein proteinases? Proteolysis and protein turnover. Edited by Bond J.S. & Barret A.J. London : *Portland Press*. pp. 57-64.

Sarbassov D.D., Ali S.M., Sabatini D.M. (2005). Growing roles for the mTOR pathway. *Curr. Opin. Cell Biol*. **17**(6): 596-603.

Sartori R., Milan G., Patron M., Mammucari C., Blaauw B., Abraham R., Sandri M. (2009). Smad2 and 3 transcription factors control muscle mass in adulthood. *Am. J. Physiol. Cell Physiol.* **296**(6): C1248-C1257.

Sartori R., Schirwis E., Blaauw B., Bortolanza S., Zhao J., Enzo E., Stantzou A., Mouisel E., Toniolo L., Ferry A. (2013). BMP signaling controls muscle mass. *Nat. Genet.* **45**:1309-1318.

Schechter N.M., Plotnick M.I. (2004). Measurement of the kinetic parameters mediating protease-Serpin inhibition. *Methods*. **32**: 159-168.

Schick C., Pemberton P.A., Shi G.P., Kamachi Y., Cataltepe S., Bartuski A.J., Gornstein E.R., Bromme D., Chapman H.A., Silverman G.A. (1998). Cross-class inhibition of the cysteine proteinases cathepsins K, L, and S by the Serpin squamous cell carcinoma antigen 1: a kinetic analysis. *Biochemistry* **37**: 5258-5266.

Schiebel J., Gaspari R., Wulsdorf T., Ngo K., Sohn C., Schrader T.E., *et al.* (2018). Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes. *Nat. Commun*. **9**(1): 3559.

Schneyer, A.L., Rzucidlo, D.A., Sluss, P.M., Crowley, W.F. (1994). Characterization of unique binding kinetics of follistatin and activin or inhibin in serum. *Endocrinol.* **135**: 667-674.

Schneyer A., Tortoriello D., Sidis Y., Keutmann H., Matsuzaki T., Holmes W. (2001). Follistatin-related protein (FSRP): a new member of the follistatin gene family. *Mol. Cell Endocrinol*. **180**: 33-38.

Schnyder S., Handschin, C. (2015). Skeletal muscle as an endocrine organ: PGC-1α, myokines and exercise. *Bone* **80**: 115–125.

Schuelke M., Wagner K.R., Stolz L.E., Hubner C;, Riebel T., Komen W., Braun T; Tobin J.F., Lee S.J. (2004). Myostatin mutation associated with gross muscle hypertrophy in a child. *N. Engl. J. Med*. **350**(26): 2682-2688.

Segel I. (1975). "Enzyme Kinetics" *Ed. J. Wiley & Sons, Inc*.

Sharma M., Kambadur R., Matthews K.G., Somers W.G., Devlin G.P., Conaglen J.V., Fowke P.J., Bass J. (1999). Myostatin, a transforming growth factor-beta superfamily member, is expressed in heart muscle and is upregulated in cardiomyocytes after infarct. *J. Cell. Physiol*. **180**: 1-9.

Sharma M., McFarlane C., Kambadur R., Kukreti H., Bonala S., Srinivasan S. (2015). Myostatin: Expanding horizons. *IUBMB Life.* **67**(8): 589-600.

Shaw D.E., Maragakis P., Lindorff-Larsen K., Piana S., Dror R.O., Eastwood M.P., *et al*. (2010). Atomic-level characterization of the structural dynamics of proteins. *Science* **330**(6002): 341-346.

Shelton G.D., Engvall E. (2007). Gross muscle hypertrophy in whippet dogs is caused bt a mutation in the myostatin gene. *Neuromuscul. Disord*. **17**(9-10): 721-722.

Shi M., Zhu J., Wang R., Chen X., Mi L., Walz T. (2011). Latent TGF-β structure and activation. *Nature* **474**: 343-349.

Short R.E., MacNeil M.D., Grosz M.D., Gerrard D.E., Grings E.E. (2002). Pleiotropic effects in Hereford, Limousin, and Piedmontese F2 crossbred calves of genes controlling muscularity including the Piedmontese myostatin allele. *J. Anim. Sci.* **80**: 1-11.

Sidis Y., Schneyer A.L., Sluss P.M., Johnson L.N., Keutmann H.T. (2001). Follistatin: essential role for the N-terminal domain in activin binding and neutralization. *J. Biol. Chem.* **276**(21): 17718-17726.

Smith D., Tikhonova I., Jewhurst H., Drysdale O., Dvorak J., Robinson M., Cwiklinski K., Dalton J. (2016). Unexpected Activity of a Novel Kunitz-type Inhibitor: Inhibition of Cysteine Proteases but not Serine Proteases. *J. Biol. Chem.* **291**: 19220-19234.

Søndergaard C.R., Olsson M.H.M., Rostkowski M., Jensen J.H. (2011). Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **7**(7): 2284-2295.

Sorrentino A., Liu CG, Addario A., Peschle C., Scambia G., Ferlini C. (2008). Role of microRNAs in drug-resistant ovarian cancer cells. *Gynecol. Oncol.* **111**: 478-486.

Souza T.A., Chen X., Guo Y., Sava P., Zhang J., Hill J.J., Yaworsky P.J., Qiu Y. (2008) Proteomic identification and functional validation of activins and bone morphogenetic protein 11 as candidate novel muscle mass regulators. *Mol. Endocrinol*. **22**(12): 2689-2702.

Spiller M., Kambadur R., Jeanplong F., Thomas M., Martyn J., Bass J., Sharma M. (2002). The Myostatin Gene Is a Downstream Target Gene of Basic Helix-Loop-Helix Transcription Factor MyoD. *Mol. Cell. Biol.* **22**(20): 7066-7082.

Stone J. (1998). An efficient library for parallel ray tracing and animation. Masters Theses [Internet]. Available from: http://scholarsmine.mst.edu/masters_theses/1747

Strom S.P., Stone J.L., Ten Bosch J.R., Merriman B., Cantor R.M., Geschwind D.H. (2010). High-density SNP association study of the 17q21 chromosomal region linked to autism identifies CACNA1 G as a novel candidate gene. *Mol. Psychiatry* **15**: 996-1005.

Suren P., Roth C., Bresnahan M., Haugen M., Hornig M., Hirtz D. (2013). Association between maternal use of folic acid supplements and risk of autism spectrum disorders in children. *JAMA* **309**: 570-577.

Suryawan A., Escobar J., Frank J.W., Nguyen H.V., Davis T.A. (2006). Developmental regulation of the activation of signaling components leading to translation initiation in skeletal muscle of neonatal pigs. *Am. J. Physiol. Endocrinol. Metab*. **291**(4): E849-859.

Szlàma G., Kondas K., Trexler M., Patthy L. (2010). WFIKKN1 and WFIKKN2 bind growth factors TGFβ1, BMP2 and BMP4 but do not inhibit their signalling activity. *FEBS J.* **277**(24): 5040-5050.

Szlàma G., Trexler M., Patthy L. (2013). Latent myostatin has significant activity and this activity is controlled more efficiently by WFIKKN1 than by WFIKKN2. *FEBS J.* **280**: 3822-3839.

Szlàma G., Vasarhelyi V., Trexler M. Patthy L. (2016). Influence of WFIKKN1 on BMP1-mediated activation of latent myostatin. *FEBS J.* **283**: 4515-4527.

Takahashi C., Sheng Z., Horan T.P., Kitayama H., Maki M., Hitomi K., Kitaura Y., Takai S., Sasahara R.M., Horimoto A., Ikawa Y., Ratzkin B.J., Arakawa T., Noda M. (1998). Regulation of matrix metalloproteinase-9 and inhibition of tumor invasion by the membrane-anchored glycoprotein RECK. *Proc. Natl. Acad. Sci. USA*. **95**: 13221-13226.

Tanzi R.E., McClatchey A.I., Lamperti E.D., Villa-Komaroff L., Gusella J.F., Neve R.L. (1988). Protease inhibitor domain encoded by an amyloid protein precursor mRNA associated with Alzheimer's disease. *Nature* **331**: 528-530.

Taylor W.E., Bhasin S., Artaza J., Byhower F., Azam M., Willard D.H., Kull F.C., Gonzalez-Cadavid N. (2001). Myostatin inhibits cell proliferation and protein synthesis in C2C12 muscle cells. *Am. J. Physiol. Endocrinol. Metab.* **280**(2): E221-228.

Thies R.S., Chen T., Davies M.V., Tomkinson K.N., Pearson A.A., Shakey Q.A., Wolfman N.M. (2001). Gdf-8 propeptide binds to Gdf-8 and antagonizes biological activity by inhibiting Gdf-8 receptor binding. *Growth Factors.* **18**: 251-259.

Thomas M., Langley B., Berry C., Sharma M., Kirk S., Bass J., Kambadur R. (2000). Myostatin, a negative regulator of muscle growth, functions by inhibiting myoblast proliferation. *J. Biol. Chem.* **275**(51): 40235-40243.

Thompson T.B., Lerch T.F., Cook R.W., Woodruff T.K., Jardetzky T.S. (2005) The structure of the follistatin: activing complex reveals antagonism of both type I and type II receptor binding. *Dev. Cell* **9**(4): 535-543.

Tintignac L.A., Leibovitch M.P., Kitzmann M., Fernandez A., Ducommun B., Meijer L., Leibovitch S.A. (2000). Cyclin E-cdk2 phoshorylation promotes late G1-phase degradation of MyoD in muscle cells. *Exp. Cell Res*. **259**(1): 300-307.

Travis J., Salvesen G.S. (1983). Human plasma proteinase inhibitors. *Annu. Rev. Biochem.* **52**: 655-709.

Trendelenburg A.U., Meyer A., Rohner D., Boyle J., Hatakeyama S., Glass D.J. (2009). Myostatin reduces Akt/TORC1/p70S6K signaling, inhibiting myoblast differentiation and myotube size. *Am. J. Physiol. Cell Physiol.* **296**(6): 1258-1270.

Trexler M., Banyai L., Patthy L. (2001). A human protein containing multiple types of protease-inhibitory modules. *Proc. Natl. Acad. Sci. USA.* **98**(7): 3705-3709.

Trexler M., Banyai L., Patthy L. (2002). Distinct expression pattern of two related human proteins containing multiple types of protease-inhibitory modules. *Biol. Chem.* **383**(1): 223-228.

Turk V., Turk B., Turk D. (2001). Lysosomal cysteine proteases: facts and opportunities. *EMBO J.* **20**: 4629-4633.

Turk V., Stoka V., Vasiljeva O., Renko M., Sun T., Turk B., Turk D. (2012). Cysteine Cathepsin: From structure, function and regulation to new frontiers. *Biochim. Biophys. Acta* **1824**: 68-88.

Ueno N., Ling N., Ying S.Y., Esch F., Shimasaki S., Guillemin R. (1987). Isolation and partial characterization of follistatin: a single-chain Mr 35,000 monomeric protein that inhibits the release of follicle-stimulating hormone. *Proc. Natl. Acad. Sci. USA.* **84**: 8282-8286.

Vadon-Le Goff S., Hulmes D., Moali C. (2015). BMP-1/tolloid-like proteinases synchronize matrix assembly with growth factor activation to promote morphogenesis and tissue remodeling. *Matrix Biol.* **44**: 14-23.

Van Eijk M., van Noorden C.J.F., de Groot C. (2003). Proteinases and their inhibitors in the immune system. *Int. Rev. Cytol*. **222**: 197-236.

Van Gent D., Sharp P., Morgan K., Kalsheker N. (2003). Serpins : structure, function and molecular evolution. *Int. J. Biochem. Cell Biol*. **35**(11): 1536-1547.

Van Wart H.E., Birkedal-Hansen H. (1990). The cysteine switch: a principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family. *PNAS* **87**(14): 5578-5582.

Vergnolle N. (2016). Protease inhibition as new therapeutic strategy for GI diseases. *Vergnolle N. Gut*. **0**: 1-10.

Vinet A., Drouilhet L., Bodin L., Mulsant P., Fabre S., Phocas F. (2012). Genetic control of multiple births in low ovulating mammalian species. *Mamm*. *Genome* **23**: 727-740.

Von Maltzahn J., Chang N.C., Bentzinger C.F., Rudnicki M.A. (2012). Wnt signaling in myogenesis. *Trends Cell Biol*. **22**(11): 602-609.

Walker R., Angerman E.B., Kattamuri C., Lee Y.S., Lee S.J., Thompson T.B. (2015). Alternative Binding Modes Identified for Growth and Differentiation Factor-associated Serum Protein (GASP) Family Antagonism of myostatin. *J. Biol. Chem.* **290**: 7506-7516.

Wan H., Lee K.S., Kim B.Y., Zou F.M., Yoon H.J., Je Y.H., Ll J., Jin B.R. (2013). A spider-derived Kunitz-type serine protease inhibitor that acts as a plasmin inhibitor and an elastase inhibitor. *PLoS One* **8**(1): e53343.

Wang H., Zhang Q., Zhu D. (2003). hSGT interacts with the N-terminal region of myostatin. *Biochem. Biophys. Res. Commun*. **311**: 877-883.

Weber K., Johnson N., Champlin D., Patty A. (2005). Many P-element insertions affect wing shape in Drosophila melanogaster. *Genetics* **169**(3): 1461-1475.

Wegner J., Albrecht E., Fiedler I., Teuscher F., Papstein H.J., Ender K. (2000). Growth- and breed-related changes of muscle fiber characteristics in cattle. *J. Anim. Sci.* **78**: 1485-1496.

Welle S., Bhatt K., Pinkert C.A., Tawil R., Thornton C.A. (2007). Muscle growth after postdevelopmental myostatin gene knockout. *Am. J. Physiol. Endocrinol. Metab*. **292**(4): E985-991.

Wheeler T.L., Shackelford S.D., Casas E., Cundiff L.V., Koohmaraie M. (2001). The effects of Piedmontese inheritance and myostatin genotype on the palatability of longissimus thoracis, gluteus medius, semimembranosus, and biceps femoris*. J. Anim. Sci.* **79**: 3069-3074.

Whisstock J.C., Skinner R., Lesk A.M. (1998). An atlas of serpin conformations. *Trends Biochem. Sci*. **23**(2): 63-67.

Whisstock J.C., Silverman G.A., Bird P.I., Bottomley S.P., Kaiserman D., Luke C.J., Pak S.C., Reichhart J.M., Huntington J.A. (2010) Serpins flex their muscle: II. Structural insights into target peptidase recognition, polymerization, and transport functions. *J. Biol. Chem.* **285**: 24307-24312.

Wickner S., Maurizi M.R., Gottesman S. (1999). Posttranslational quality control: folding, refolding, and degrading proteins. *Science* **286**: 1888-1893.

Wilkes J.J., Lloyd D.J., Gekakis N. (2009). Loss-of-function mutation in myostatin reduces tumor necrosis factor alpha production and protects liver against obesity-induced insulin resistance. *Diabetes* **58**(5): 1133-1143.

Wirth T.C., Xue H.H., Rai D., Sabel J.T., Bair T., Harty J.T. (2010). Repetitive antigen stimulation induces stepwise transcriptome diversification but preserves a core signature of memory CD8 (+) T cell differentiation. *Immunity* **33**: 128-140.

Wolfman N.M., McPherron A.C., Pappano W.N., Davies M.V., Song K., Tomkinson K.N., Wright J.F., Zhao L., Sebald S.M., Greenspan D.S., Lee S.J. (2003). Activation of latent myostatin by the BMP-1/tolloid family of metalloproteinases. *Proc. Natl. Acad. Sci. USA* **100**(26): 15842-15846.

Wu M.Y., Hill C.S. (2009). TGF-beta superfamily signaling in embryonic development and homeostasis. *Dev. Cell Mar.* **16**(3): 329-343.

Wu J., Wang Y., Liang S., Ma H. (2014). Cytoprotective effect of selective small-molecule caspase inhibitors against staurosporine-induced apoptosis. *Drug Des. Dev. Ther*. **8**: 583-600.

Wurm F.M. (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.* **22**: 1393-1398.

Xu Y., Carr P.D., Guss J.M., Ollis D.L. (1998). The crystal structure of bikunin from the inter-alpha-inhibitor complex: a serine protease inhibitor with two Kunitz domains. *J. Mol. Biol*. **276**(5): 955-966.

Yaffe D. Saxel O. (1977). A myogenic cell line with altered serum requirements for differentiation. *Differentiation* **7**: 159-166.

Yaffe D., Saxel O. (1977). Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature* **270**(5639): 725-727.

Yamasaki M. Arii Y. Mikami B. and Hirose M. (2002). Loop-inserted and thermostabilized structure of P1-P1' cleaved ovalbumin mutant R339T. *J. Mol. Biol.* **315**: 113-120.

Yamashita M., Mizusawa N., Hojo M., Yabu T. (2008). Extensive apoptosis and abnormal morphogenesis in pro-caspase-3 transgenic zebrafish during development. *J. Exp. Biol.* **211**: 1874-1881.

Yang W., Chen Y., Zhang Y., Wang X., Yang N., Zhu D. (2006). Extracellular signal-regulated kinase 1/2 mitogen-activated protein kinase pathway is involved in myostatinregulated differentiation repression. *Cancer Res*. **66**: 1320-1326.

Yang W., Zhang Y., Li Y., Wu Z., Zhu D. (2007). Myostatin induces cyclin D1 degradation to cause cell cycle arrest through a phosphatidylinositol 3-kinase/AKT/GSK-3 beta pathway and is antagonized by insulin-like growth factor 1. *J. Biol. Chem.* **282**: 3799-3808.

Yang J., Yan R., Roy A., Xu D., Poisson J., Zhang Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nat. Methods*. **12**(1): 7-8.

Yoo B.C., AoKI K., Xiang Y., Campbell L.R., Hull R.J., Xoconostle-Cazares B;, Monzer J., Lee J.Y., Ullman D.E., Lucas W.J. (2000). Characterization of *cucurbita maxima phloem* serpin-1 (CmPS-1). A developmentally regulated elastase inhibitor. *J. Biol. Chem*. **275**(45): 35122-35128.

Yousef H., Conboy M.J., Mamiya H., Zeiderman M., Schlesinger C., Schaffer D.V. (2014). Mechanisms of action of hESC-secreted proteins that enhance human and mouse myogenesis. *Aging (Albany NY)* **6:** 602-620.

Zhang Z., Gildersleeve J., Yang Y.Y., *et al*. (2004)**.** A new strategy for the synthesis of glycoproteins. *Science* **303**: 371-373.

Zhang J., Liu M., Su Y., Du J., Zhu A.J. (2012). A Targeted *In Vivo* RNAi Screen Reveals Deubiquitinases as New Regulators of Notch Signaling. *G3 (Bethesda)* **2**(12): 1563-1575.

Zhang Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**(1): 40.

Zhou Q., Snipas S., Orth K., Muzio M., Dixit V.M., Salvesen G.S. (1997). Target protease specificity of the viral serpin CrmA. Analysis of five caspases. *J. Biol. Chem*. **272**(12): 7797-7800.

Zhu X., Hadhazy M., Wehling M., Tidball J.G., McNally E.M. (2000). Dominant negative myostatin produces hypertrophy without hyperplasia in muscle. *FEBS Lett.* **474**: 71-75.

Zhu X., Topouzis S., Liang L.F., Stotish R.L. (2004). Myostatin signaling through Smad2, Smad3 and Smad4 is regulated by the inhibitory Smad7 by a negative feedback mechanism. **26**(6): 262-272.

Zimmers T.A., Davies M.V., Koniaris L.G., Haynes P., Esquela A.F., Tomkinson K.N., McPherron A.C., Wolfman N.M., Lee S.J. (2002). Induction of cachexia in mice by systemically administered myostatin. *Science* **296**: 1486-1488.

Zumbrunn J., Trueb B. (1997). Primary structure of a putative serine protease specific for IGF-binding proteins. *FEBS Lett*. **398**(2-3): 187-192.

APPENDIX

Nucleic and Protein Sequences derived from the 4 GASP proteins

The GST tag at the N-terminus of the fusion protein is indicated in red, the thrombin cleavage site in green, and the GASP protein (without the peptide signal) in blue and black. The black residuals correspond to the different conserved domains: W, WAP domain; F / K, Follistatine / Kazal domain; Ig, type C2 immunoglobulin domain; Ku1, first Kunitz estate; Ku2, second Kunitz domain of GASP-1; NTR, Netrin domain. The numbers on the left refer to the +1 position of the vector.

(A) GASP-1
(B) GASP-1 / $K_{2\text{-}2}$
(C) GASP-2
(D) GASP-2 / $K_{2\text{-}1}$

(A)

```
   258 atgtcccctatactaggttattggaaaattaagggccttgtgcaacccactcgacttcttttggaatatcttgaa
        M  S  P  I  L  G  Y  W  K  I  K  G  L  V  Q  P  T  R  L  L  L  E  Y  L  E
   333 gaaaaatatgaagagcatttgtatgagcgcgatgaaggtgataaatggcgaaacaaaaagtttgaattgggtttg
        E  K  Y  E  E  H  L  Y  E  R  D  E  G  D  K  W  R  N  K  K  F  E  L  G  L
   408 gagtttcccaatcttccttattatattgatggtgatgttaaattaacacagtctatggccatcatacgttatata
        E  F  P  N  L  P  Y  Y  I  D  G  D  V  K  L  T  Q  S  M  A  I  I  R  Y  I
   483 gctgacaagcacaacatgttgggtggttgtccaaaagagcgtgcagagatttcaatgcttgaaggagcggttttg
        A  D  K  H  N  M  L  G  G  C  P  K  E  R  A  E  I  S  M  L  E  G  A  V  L
   558 gatattagatacggtgtttcgagaattgcatatagtaaagactttgaaactctcaaagttgattttcttagcaag
        D  I  R  Y  G  V  S  R  I  A  Y  S  K  D  F  E  T  L  K  V  D  F  L  S  K
   633 ctacctgaaatgctgaaaatgttcgaagatcgtttatgtcataaaacatatttaaatggtgatcatgtaacccat
        L  P  E  M  L  K  M  F  E  D  R  L  C  H  K  T  Y  L  N  G  D  H  V  T  H
   708 cctgacttcatgttgtatgacgctcttgatgttgttttatacatggacccaatgtgcctggatgcgttcccaaaa
        P  D  F  M  L  Y  D  A  L  D  V  V  L  Y  M  D  P  M  C  L  D  A  F  P  K
   783 ttagtttgttttaaaaaacgtattgaagctatcccacaaattgataagtacttgaaatccagcaagtatatagca
        L  V  C  F  K  K  R  I  E  A  I  P  Q  I  D  K  Y  L  K  S  S  K  Y  I  A
   858 tggcctttgcagggctggcaagccacgtttggtggtggcgaccatcctccaaaatcggatctggttccgcgtgga BamHI
        W  P  L  Q  G  W  Q  A  T  F  G  G  G  D  H  P  P  K  S  D  L  V  P  R//G
   933 tccctgccacccatccgatactcccatgcgggcatctgccccaacgacatgaaccccaacctctgggtggatgcc
        S  L30 P  P  I  R  Y  S  H  A  G  I  C  P  N  D  M  N  P  N  L  W  V  D  A
  1008 cagagcacctgcaagcgggagtgtgaaacagaccaggaatgtgagacctatgagaaatgctgccccaatgtgtgt
W(37-82)  Q  S  T  C  K  R  E  C  E  T  D  Q  E  C  E  T  Y  E  K  C  C  P  N  V  C
  1083 gggaccaagagctgtgtggcagcccgctacatggatgtgaaagggaagaaggggcctgtaggcatgcccaaggag
        G  T  K  S  C  V  A  A  R  Y  M  D  V  K  G  K  K  G  P  V  G  M  P  K  E
  1158 gccacatgtgaccatttcatgtgcctgcagcagggctctgagtgtgacatctgggacggccagcccgtgtgtaag
        A  T  C  D  H  F  M  C  L  Q  Q  G  S  E  C  D  I  W  D  G  Q  P  V  C  K
  1233 tgcaaagatcgctgtgagaaggagcccagcttcacctgggcctctgatggccttacctactacaaccgttgcttc
F/K(106-170) C  K  D  R  C  E  K  E  P  S  F  T  W  A  S  D  G  L  T  Y  Y  N  R  C  F
  1308 atggacgccgaagcctgctccaagggcatcacactgtctgtggtcacctgtcgttatcacttcacctggcctaac
        M  D  A  E  A  C  S  K  G  I  T  L  S  V  V  T  C  R  Y  H  F  T  W  P  N
  1383 accagccctccaccgcctgagaccacggtgcatcccaccaccgcctctccggagactctcgggctggacatggca
        T  S  P  P  P  P  E  T  T  V  H  P  T  T  A  S  P  E  T  L  G  L  D  M  A
  1458 gccccggccctgctcaaccaccctgtccatcagtcagtcaccgtgggtgagactgtgagtttcctctgcgacgtg
        A  P  A  L  L  N  H  P  V  H  Q  S  V  T  V  G  E  T  V  S  F  L  C  D  V
  1533 gtaggccggcctcggccagagctcacttgggagaaacagctggaggaccgagagaatgttgtcatgaggcccaac
Ig(226-282)  V  G  R  P  R  P  E  L  T  W  E  K  Q  L  E  D  R  E  N  V  V  M  R  P  N
  1608 cacgtgcgtggtaatgtggtggtcactaacattgcccagctggtcatctacaacgtccagccccaggatgctggc
        H  V  R  G  N  V  V  V  T  N  I  A  Q  L  V  I  Y  N  V  Q  P  Q  D  A  G
  1683 atatacacctgtacagctcgaaatgtcgctggtgtcctgagggctgacttcccgttgtcggtggtcaggggtggt
        I  Y  T  C  T  A  R  N  V  A  G  V  L  R  A  D  F  P  L  S  V  V  R  G  G
  1758 caggccagggccacttcagagagcagtctcaatggcacagcttttccagcaacagagtgcctgaagcccccagac
        Q  A  R  A  T  S  E  S  S  L  N  G  T  A  F  P  A  T  E  C  L  K  P  P  D
  1833 agtgaggactgtggagaggagcagacacgctggcacttcgacgcccaggctaacaactgcctcactttcaccttt
Ku1(323-373) S  E  D  C  G  E  E  Q  T  R  W  H  F  D  A  Q  A  N  N  C  L  T  F  T  F
  1908 ggccactgccaccacaatctcaaccactttgagacctacgaggcctgtatgctggcttgtatgagtgggccattg
        G  H  C  H  H  N  L  N  H  F  E  T  Y  E  A  C  M  L  A  C  M  S  G  P  L
  1983 gccacctgcagcctgcctgccctgcaagggccttgcaaagcttatgtcccacgctgggcctacaacagccagaca
        A  T  C  S  L  P  A  L  Q  G  P  C  K  A  Y  V  P  R  W  A  Y  N  S  Q  T
  2058 ggcctatgccagtccttcgtctatggcggctgtgagggcaacggtaacaactttgaaagccgtgaggcttgtgag
Ku2(381-431) G  L  C  Q  S  F  V  Y  G  G  C  E  G  N  G  N  N  F  E  S  R  E  A  C  E
  2133 gagtcgtgtcccttcccgaggggcaaccagcactgccgggcctgcaagccccggcaaaaacttgttaccagcttc
```

```
      E  S  C  P  F  P  R  G  N  Q  H  C  R  A  C  K  P  R  Q  K  L  V  T  S  F
2208  tgtcggagtgactttgtcatcctgggcagggtctctgagctgaccgaggagcaagactcgggccgtgccctggtg
      C  R  S  D  F  V  I  L  G  R  V  S  E  L  T  E  E  Q  D  S  G  R  A  L  V
2283  accgtggatgaggtcttaaaagatgagaagatgggcctcaagtttctgggccgggagcctctggaagtcaccctg
NTR(440-561) T  V  D  E  V  L  K  D  E  K  M  G  L  K  F  L  G  R  E  P  L  E  V  T  L
2358  cttcatgtagactggacctgtccttgccccaacgtgacagtgggtgagacaccactcatcatcatgggggaggtg
      L  H  V  D  W  T  C  P  C  P  N  V  T  V  G  E  T  P  L  I  I  M  G  E  V
2433  gacggcggcatggccatgctgagacccgatagctttgtgggggcatcgagcacacggcgggtcaggaagctccgt
      D  G  G  M  A  M  L  R  P  D  S  F  V  G  A  S  S  T  R  R  V  R  K  L  R
2508  gaggtcatgtacaagaaaacctgtgacgtcctcaaggacttcctgggcttgcaatga 2564
      E  V  M  Y  K  K  T  C  D  V  L  K  D  F  L  G  L  Q571*
```

(B)

```
 258  atgtcccctatactaggttattggaaaattaagggccttgtgcaacccactcgacttcttttggaatatcttgaa
      M  S  P  I  L  G  Y  W  K  I  K  G  L  V  Q  P  T  R  L  L  L  E  Y  L  E
 333  gaaaaatatgaagagcatttgtatgagcgcgatgaaggtgataaatggcgaaacaaaaagtttgaattgggtttg
      E  K  Y  E  E  H  L  Y  E  R  D  E  G  D  K  W  R  N  K  K  F  E  L  G  L
 408  gagtttcccaatcttccttattatattgatggtgatgttaaattaacacagtctatggccatcatacgttatata
      E  F  P  N  L  P  Y  Y  I  D  G  D  V  K  L  T  Q  S  M  A  I  I  R  Y  I
 483  gctgacaagcacaacatgttgggtggttgtccaaaagagcgtgcagagatttcaatgcttgaaggagcggttttg
      A  D  K  H  N  M  L  G  G  C  P  K  E  R  A  E  I  S  M  L  E  G  A  V  L
 558  gatattagatacggtgtttcgagaattgcatatagtaaagactttgaaactctcaaagttgattttcttagcaag
      D  I  R  Y  G  V  S  R  I  A  Y  S  K  D  F  E  T  L  K  V  D  F  L  S  K
 633  ctacctgaaatgctgaaaatgttcgaagatcgtttatgtcataaaacatatttaaatggtgatcatgtaacccat
      L  P  E  M  L  K  M  F  E  D  R  L  C  H  K  T  Y  L  N  G  D  H  V  T  H
 708  cctgacttcatgttgtatgacgctcttgatgttgttttatacatggacccaatgtgcctggatgcgttcccaaaa
      P  D  F  M  L  Y  D  A  L  D  V  V  L  Y  M  D  P  M  C  L  D  A  F  P  K
 783  ttagtttgttttaaaaaacgtattgaagctatcccacaaattgataagtacttgaaatccagcaagtatatagca
      L  V  C  F  K  K  R  I  E  A  I  P  Q  I  D  K  Y  L  K  S  S  K  Y  I  A
 858  tggcctttgcagggctggcaagccacgtttggtggtggcgaccatcctccaaaatcggatctggttccgcgtgga BamHI
      W  P  L  Q  G  W  Q  A  T  F  G  G  G  D  H  P  P  K  S  D  L  V  P  R//G
 933  tccctgccacccatccgatactcccatgcgggcatctgccccaacgacatgaaccccaacctctgggtggatgcc
      S  L30 P  P  I  R  Y  S  H  A  G  I  C  P  N  D  M  N  P  N  L  W  V  D  A
1008  cagagcacctgcaagcgggagtgtgaaacagaccaggaatgtgagacctatgagaaatgctgccccaatgtgtgt
W(37-82)  Q  S  T  C  K  R  E  C  E  T  D  Q  E  C  E  T  Y  E  K  C  C  P  N  V  C
1083  gggaccaagagctgtgtggcagcccgctacatggatgtgaaagggaagaaggggcctgtaggcatgcccaaggag
      G  T  K  S  C  V  A  A  R  Y  M  D  V  K  G  K  K  G  P  V  G  M  P  K  E
1158  gccacatgtgaccatttcatgtgcctgcagcagggctctgagtgtgacatctgggacggccagcccgtgtgtaag
      A  T  C  D  H  F  M  C  L  Q  Q  G  S  E  C  D  I  W  D  G  Q  P  V  C  K
1233  tgcaaagatcgctgtgagaaggagcccagcttcacctgggcctctgatggccttacctactacaaccgttgcttc
F/K(106-170) C  K  D  R  C  E  K  E  P  S  F  T  W  A  S  D  G  L  T  Y  Y  N  R  C  F
1308  atggacgccgaagcctgctccaagggcatcacactgtctgtggtcacctgtcgttatcacttcacctggcctaac
      M  D  A  E  A  C  S  K  G  I  T  L  S  V  V  T  C  R  Y  H  F  T  W  P  N
1383  accagccctccaccgcctgagaccacggtgcatcccaccaccgcctctccggagactctcgggctggacatggca
      T  S  P  P  P  P  E  T  T  V  H  P  T  T  A  S  P  E  T  L  G  L  D  M  A
1458  gccccggccctgctcaaccaccctgtccatcagtcagtcaccgtgggtgagactgtgagtttcctctgcgacgtg
      A  P  A  L  L  N  H  P  V  H  Q  S  V  T  V  G  E  T  V  S  F  L  C  D  V
1533  gtaggccggcctcggccagagctcacttgggagaaacagctggaggaccgagagaatgttgtcatgaggcccaac
Ig(226-282)  V  G  R  P  R  P  E  L  T  W  E  K  Q  L  E  D  R  E  N  V  V  M  R  P  N
1608  cacgtgcgtggtaatgtggtggtcactaacattgcccagctggtcatctacaacgtccagccccaggatgctggc
      H  V  R  G  N  V  V  V  T  N  I  A  Q  L  V  I  Y  N  V  Q  P  Q  D  A  G
1683  atatacacctgtacagctcgaaatgtcgctggtgtcctgagggctgacttcccgttgtcggtggtcaggggtggt
      I  Y  T  C  T  A  R  N  V  A  G  V  L  R  A  D  F  P  L  S  V  V  R  G  G
1758  caggccagggccacttcagagagcagtctcaatggcacagcttttccagcaacagagtgcctgaagcccccagac
      Q  A  R  A  T  S  E  S  S  L  N  G  T  A  F  P  A  T  E  C  L  K  P  P  D
1833  agtgaggactgtggagaggagcagacacgctggcacttcgacgcccaggctaacaactgcctcactttcaccttt
Ku1(323-373)  S  E  D  C  G  E  E  Q  T  R  W  H  F  D  A  Q  A  N  N  C  L  T  F  T  F
1908  ggccactgccaccacaatctcaaccactttgagacctacgaggcctgtatgctggcttgtatgagtgggccattg
      G  H  C  H  H  N  L  N  H  F  E  T  Y  E  A  C  M  L  A  C  M  S  G  P  L
1983  gccacctgtgcactgcctgcagttcaggggccctgccagggctgggagccacgctgggcctacagcccactgcta
      A  T  C  A  L  P  A  V  Q  G  P  C  Q  G  W  E  P  R  W  A  Y  S  P  L  L
2058  cagcagtgccacccctttgtatacagtggctgtgaaggaaacagcaataactttgagacccgggagagctgtgag
Ku2-2(381-431) Q  Q  C  H  P  F  V  Y  S  G  C  E  G  N  S  N  N  F  E  T  R  E  S  C  E
2133  gatgcttgcccttcccgaggggcaaccagcactgccgggcctgcaagccccggcaaaaacttgttaccagcttc
      D  A  C  P  F  P  R  G  N  Q  H  C  R  A  C  K  P  R  Q  K  L  V  T  S  F
```

```
2208 tgtcggagtgactttgtcatcctgggcagggtctctgagctgaccgaggagcaagactcgggccgtgccctggtg
      C  R  S  D  F  V  I  L  G  R  V  S  E  L  T  E  E  Q  D  S  G  R  A  L  V
2283 accgtggatgaggtcttaaaagatgagaagatgggcctcaagtttctgggccgggagcctctggaagtcaccctg
NTR(440-561) T  V  D  E  V  L  K  D  E  K  M  G  L  K  F  L  G  R  E  P  L  E  V  T  L
2358 cttcatgtagactggacctgtccttgccccaacgtgacagtgggtgagacaccactcatcatcatgggggaggtg
      L  H  V  D  W  T  C  P  C  P  N  V  T  V  G  E  T  P  L  I  I  M  G  E  V
2433 gacggcggcatggccatgctgagacccgatagctttgtgggggcatcgagcacacggcgggtcaggaagctccgt
      D  G  G  M  A  M  L  R  P  D  S  F  V  G  A  S  S  T  R  R  V  R  K  L  R
2508 gaggtcatgtacaagaaaacctgtgacgtcctcaaggacttcctgggcttgcaatga 2564
      E  V  M  Y  K  K  T  C  D  V  L  K  D  F  L  G  L  Q571*
```

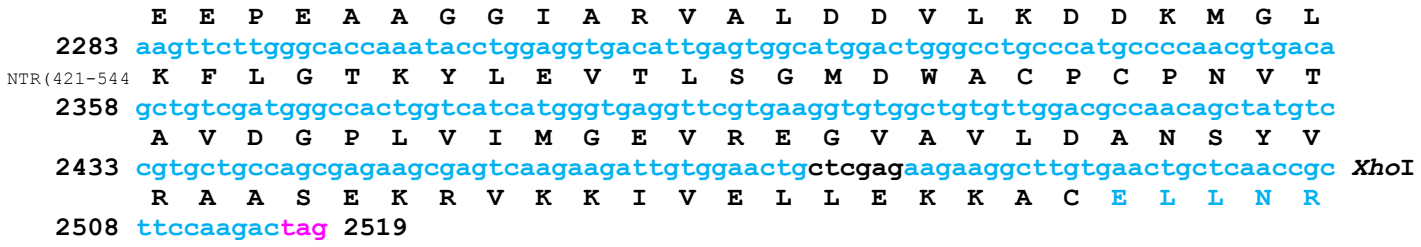
(C)

```
 258 atgtcccctatactaggttattggaaaattaagggccttgtgcaacccactcgacttcttttggaatatcttgaa
      M  S  P  I  L  G  Y  W  K  I  K  G  L  V  Q  P  T  R  L  L  L  E  Y  L  E
 333 gaaaaatatgaagagcatttgtatgagcgcgatgaaggtgataaatggcgaaacaaaaagtttgaattgggtttg
      E  K  Y  E  E  H  L  Y  E  R  D  E  G  D  K  W  R  N  K  K  F  E  L  G  L
 408 gagtttcccaatcttccttattatattgatggtgatgttaaattaacacagtctatggccatcatacgttatata
      E  F  P  N  L  P  Y  Y  I  D  G  D  V  K  L  T  Q  S  M  A  I  I  R  Y  I
 483 gctgacaagcacaacatgttgggtggttgtccaaaagagcgtgcagagatttcaatgcttgaaggagcggttttg
      A  D  K  H  N  M  L  G  G  C  P  K  E  R  A  E  I  S  M  L  E  G  A  V  L
 558 gatattagatacggtgtttcgagaattgcatatagtaaagactttgaaactctcaaagttgattttcttagcaag
      D  I  R  Y  G  V  S  R  I  A  Y  S  K  D  F  E  T  L  K  V  D  F  L  S  K
 633 ctacctgaaatgctgaaaatgttcgaagatcgtttatgtcataaaacatatttaaatggtgatcatgtaacccat
      L  P  E  M  L  K  M  F  E  D  R  L  C  H  K  T  Y  L  N  G  D  H  V  T  H
 708 cctgacttcatgttgtatgacgctcttgatgttgttttatacatggacccaatgtgcctggatgcgttcccaaaa
      P  D  F  M  L  Y  D  A  L  D  V  V  L  Y  M  D  P  M  C  L  D  A  F  P  K
 783 ttagtttgttttaaaaaacgtattgaagctatcccacaaattgataagtacttgaaatccagcaagtatatagca
      L  V  C  F  K  K  R  I  E  A  I  P  Q  I  D  K  Y  L  K  S  S  K  Y  I  A
 858 tggcctttgcagggctggcaagccacgtttggtggtggcgaccatcctccaaaatcggatctggttccgcgtgga Bam HI
      W  P  L  Q  G  W  Q  A  T  F  G  G  G  D  H  P  P  K  S  D  L  V  P  R//G
 933 tccctgccagatcccggaagccatcctggcatgtgccccaacgagctcagcccccacctgtgggtcgacgcccag
      S  L26 P  D  P  G  S  H  P  G  M  C  P  N  E  L  S  P  H  L  W  V  D  A  Q
1008 agcacctgtgagcgtgagtgtaccggggaccaggactgtgcggcatccgagaagtgctgcaccaatgtgtgtggg
W(32-77)  S  T  C  E  R  E  C  T  G  D  Q  D  C  A  A  S  E  K  C  C  T  N  V  C  G
1083 ctgcagagctgcgtggctgcccgctttcccagtggtggcccagctgtacctgagacagcagcctcctgtgaaggc
      L  Q  S  C  V  A  A  R  F  P  S  G  G  P  A  V  P  E  T  A  A  S  C  E  G
1158 ttccaatgcccacaacagggttctgactgtgacatctgggatgggcagccagtttgtcgctgccgtgaccgctgt
      F  Q  C  P  Q  Q  G  S  D  C  D  I  W  D  G  Q  P  V  C  R  C  R  D  R  C
1233 gaaaaagaacccagcttcacatgtgcttctgatggccttacctattacaaccgctgctacatggacgcagaagcc
F/K(97-161) E  K  E  P  S  F  T  C  A  S  D  G  L  T  Y  Y  N  R  C  Y  M  D  A  E  A
1308 tgcctgcggggtctccacctgcacgttgtaccctgtaagcacattctcagttggccgcccagcagcccgggacca
      C  L  R  G  L  H  L  H  V  V  P  C  K  H  I  L  S  W  P  P  S  S  P  G  P
1383 cccgagaccactgctcgcccaacccctggggctgctcccatgccacctgccctgtacaacagcccctcaccacag
      P  E  T  T  A  R  P  T  P  G  A  A  P  M  P  P  A  L  Y  N  S  P  S  P  Q
1458 gcagtgcatgttgggggacagccagcctccactgtgatgttagtggccgtccaccacctgctgtgacctgggag
      A  V  H  V  G  G  T  A  S  L  H  C  D  V  S  G  R  P  P  P  A  V  T  W  E
1533 aagcagagccatcagcgggagaacctgatcatgcgccctgaccaaatgtatggcaacgtggttgtcaccagtatc
      K  Q  S  H  Q  R  E  N  L  I  M  R  P  D  Q  M  Y  G  N  V  V  V  T  S  I
1608 ggacagctagtcctctacaatgctcagttggaggatgcgggcctgtatacctgcactgcacgaaacgctgccggc
Ig(211-267) G  Q  L  V  L  Y  N  A  Q  L  E  D  A  G  L  Y  T  C  T  A  R  N  A  A  G
1683 ctgctgcgggccgactttcccctttccgttttacagcgggcaactactcaggacagggacccaggtatcccagcc
      L  L  R  A  D  F  P  L  S  V  L  Q  R  A  T  T  Q  D  R  D  P  G  I  P  A
1758 ttggctgagtgccaggccgacacacaagcctgtgttgggccacctactccccatcatgtcctttggcgctttgac
      L  A  E  C  Q  A  D  T  Q  A  C  V  G  P  P  T  P  H  H  V  L  W  R  F  D
1833 ccacagagaggcagctgcatgacattcccagccctcagatgtgatggggctgcccggggctttgagacctatgag
Ku1(303-355) P  Q  R  G  S  C  M  T  F  P  A  L  R  C  D  G  A  A  R  G  F  E  T  Y  E
1908 gcatgccagcaggcctgtgttcgtggccccggggatgtctgtgcactgcctgcagttcaggggccctgccagggc
      A  C  Q  Q  A  C  V  R  G  P  G  D  V  C  A  L  P  A  V  Q  G  P  C  Q  G
1983 tgggagccacgctgggcctacagcccactgctacagcagtgccacccctttgtatacagtggctgtgaaggaaac
Ku2(363-413) W  E  P  R  W  A  Y  S  P  L  L  Q  Q  C  H  P  F  V  Y  S  G  C  E  G  N
2058 agcaataactttgagacccgggagagctgtgaggatgcttgccctgtaccacgcacaccaccctgtcgtgcctgc
      S  N  N  F  E  T  R  E  S  C  E  D  A  C  P  V  P  R  T  P  P  C  R  A  C
2133 cgcctcaagagcaagctggctctgagcttgtgccgcagtgactttgccatcgtggggagactcacagaggtcctg
      R  L  K  S  K  L  A  L  S  L  C  R  S  D  F  A  I  V  G  R  L  T  E  V  L
```

```
2208 gaggagcccgaggctgcaggcggcatagctcgtgtggccttggatgatgtgctaaaggacgacaagatgggcctc
      E  E  P  E  A  A  G  G  I  A  R  V  A  L  D  D  V  L  K  D  D  K  M  G  L
2283 aagttcttgggcaccaaatacctggaggtgacattgagtggcatggactgggcctgcccatgccccaacgtgaca
NTR(421-544  K  F  L  G  T  K  Y  L  E  V  T  L  S  G  M  D  W  A  C  P  C  P  N  V  T
2358 gctgtcgatgggccactggtcatcatgggtgaggttcgtgaaggtgtggctgtgttggacgccaacagctatgtc
      A  V  D  G  P  L  V  I  M  G  E  V  R  E  G  V  A  V  L  D  A  N  S  Y  V
2433 cgtgctgccagcgagaagcgagtcaagaagattgtggaactgctcgagaagaaggcttgtgaactgctcaaccgc  Xho I
      R  A  A  S  E  K  R  V  K  K  I  V  E  L  L  E  K  K  A  C  E  L  L  N  R
2508 ttccaagactag 2519
      F  Q  D552*
```

(D)

```
 258 atgtcccctatactaggttattggaaaattaagggccttgtgcaacccactcgacttcttttggaatatcttgaa
      M  S  P  I  L  G  Y  W  K  I  K  G  L  V  Q  P  T  R  L  L  L  E  Y  L  E
 333 gaaaaatatgaagagcatttgtatgagcgcgatgaaggtgataaatggcgaaacaaaaagtttgaattgggtttg
      E  K  Y  E  E  H  L  Y  E  R  D  E  G  D  K  W  R  N  K  K  F  E  L  G  L
 408 gagtttcccaatcttccttattatattgatggtgatgttaaattaacacagtctatggccatcatacgttatata
      E  F  P  N  L  P  Y  Y  I  D  G  D  V  K  L  T  Q  S  M  A  I  I  R  Y  I
 483 gctgacaagcacaacatgttgggtggttgtccaaaagagcgtgcagagatttcaatgcttgaaggagcggttttg
      A  D  K  H  N  M  L  G  G  C  P  K  E  R  A  E  I  S  M  L  E  G  A  V  L
 558 gatattagatacggtgtttcgagaattgcatatagtaaagactttgaaactctcaaagttgattttcttagcaag
      D  I  R  Y  G  V  S  R  I  A  Y  S  K  D  F  E  T  L  K  V  D  F  L  S  K
 633 ctacctgaaatgctgaaaatgttcgaagatcgtttatgtcataaaacatatttaaatggtgatcatgtaacccat
      L  P  E  M  L  K  M  F  E  D  R  L  C  H  K  T  Y  L  N  G  D  H  V  T  H
 708 cctgacttcatgttgtatgacgctcttgatgttgttttatacatggacccaatgtgcctggatgcgttcccaaaa
      P  D  F  M  L  Y  D  A  L  D  V  V  L  Y  M  D  P  M  C  L  D  A  F  P  K
 783 ttagtttgttttaaaaaacgtattgaagctatcccacaaattgataagtacttgaaatccagcaagtatatagca
      L  V  C  F  K  K  R  I  E  A  I  P  Q  I  D  K  Y  L  K  S  S  K  Y  I  A
 858 tggcctttgcagggctggcaagccacgtttggtggtggcgaccatcctccaaaatcggatctggttccgcgtgga  BamHI
      W  P  L  Q  G  W  Q  A  T  F  G  G  G  D  H  P  P  K  S  D  L  V  P  R//G
 933 tccctgccagatcccggaagccatcctggcatgtgccccaacgagctcagccccacctgtgggtcgacgcccag
      S  L26 P  D  P  G  S  H  P  G  M  C  P  N  E  L  S  P  H  L  W  V  D  A  Q
1008 agcacctgtgagcgtgagtgtaccggggaccaggactgtgcggcatccgagaagtgctgcaccaatgtgtgtggg
W(32-77)   S  T  C  E  R  E  C  T  G  D  Q  D  C  A  A  S  E  K  C  C  T  N  V  C  G
1083 ctgcagagctgcgtggctgcccgctttcccagtggtggcccagctgtacctgagacagcagcctcctgtgaaggc
      L  Q  S  C  V  A  A  R  F  P  S  G  G  P  A  V  P  E  T  A  A  S  C  E  G
1158 ttccaatgcccacaacagggttctgactgtgacatctgggatgggcagccagtttgtcgctgccgtgaccgctgt
      F  Q  C  P  Q  Q  G  S  D  C  D  I  W  D  G  Q  P  V  C  R  C  R  D  R  C
1233 gaaaaagaacccagcttcacatgtgcttctgatggccttacctattacaaccgctgctacatggacgcagaagcc
F/K(97-161)  E  K  E  P  S  F  T  C  A  S  D  G  L  T  Y  Y  N  R  C  Y  M  D  A  E  A
1308 tgcctgcggggtctccacctgcacgttgtaccctgtaagcacattctcagttggccgcccagcagcccgggacca
      C  L  R  G  L  H  L  H  V  V  P  C  K  H  I  L  S  W  P  P  S  S  P  G  P
1383 cccgagaccactgctcgcccaacccctggggctgctcccatgccacctgccctgtacaacagcccctcaccacag
      P  E  T  T  A  R  P  T  P  G  A  A  P  M  P  P  A  L  Y  N  S  P  S  P  Q
1458 gcagtgcatgttgggggacagccagcctccactgtgatgttagtggccgtccaccacctgctgtgacctgggag
      A  V  H  V  G  G  T  A  S  L  H  C  D  V  S  G  R  P  P  P  A  V  T  W  E
1533 aagcagagccatcagcgggagaacctgatcatgcgccctgaccaaatgtatggcaacgtggttgtcaccagtatc
      K  Q  S  H  Q  R  E  N  L  I  M  R  P  D  Q  M  Y  G  N  V  V  V  T  S  I
1608 ggacagctagtcctctacaatgctcagttggaggatgcgggcctgtatacctgcactgcacgaaacgctgccggc
Ig(211-267)  G  Q  L  V  L  Y  N  A  Q  L  E  D  A  G  L  Y  T  C  T  A  R  N  A  A  G
1683 ctgctgcgggccgactttccccttccgttttacagcgggcaactactcaggacagggacccaggtatcccagcc
      L  L  R  A  D  F  P  L  S  V  L  Q  R  A  T  T  Q  D  R  D  P  G  I  P  A
1758 ttggctgagtgccaggccgacacacaagcctgtgttgggccacctactccccatcatgtcctttggcgctttgac
      L  A  E  C  Q  A  D  T  Q  A  C  V  G  P  P  T  P  H  H  V  L  W  R  F  D
1833 ccacagagaggcagctgcatgacattcccagccctcagatgtgatggggctgcccggggctttgagacctatgag
Ku1(303-355)  P  Q  R  G  S  C  M  T  F  P  A  L  R  C  D  G  A  A  R  G  F  E  T  Y  E
1908 gcatgccagcaggcctgtgttcgtggccccggggatgtctgcagcctgcctgccctgcaagggccttgcaaagct
      A  C  Q  Q  A  C  V  R  G  P  G  D  V  C  S  L  P  A  L  Q  G  P  C  K  A
1983 tatgtcccacgctgggcctacaacagccagacaggcctatgccagtccttcgtctatggcggctgtgagggcaac
Ku2-1(363-413)  Y  V  P  R  W  A  Y  N  S  Q  T  G  L  C  Q  S  F  V  Y  G  G  C  E  G  N
2058 ggtaacaactttgaaagccgtgaggcttgtgaggagtcgtgcctgtaccacgcacaccaccctgtcgtgcctgc
      G  N  N  F  E  S  R  E  A  C  E  E  S  C  P  V  P  R  T  P  P  C  R  A  C
2133 cgcctcaagagcaagctggctctgagcttgtgccgcagtgactttgccatcgtggggagactcacagaggtcctg
      R  L  K  S  K  L  A  L  S  L  C  R  S  D  F  A  I  V  G  R  L  T  E  V  L
2208 gaggagcccgaggctgcaggcggcatagctcgtgtggccttggatgatgtgctaaaggacgacaagatgggcctc
```


E E P E A A G G I A R V A L D D V L K D D K M G L
2283 aagttcttgggcaccaaatacctggaggtgacattgagtggcatggactgggcctgcccatgccccaacgtgaca
NTR(421-544 K F L G T K Y L E V T L S G M D W A C P C P N V T
2358 gctgtcgatgggccactggtcatcatgggtgaggttcgtgaaggtgtggctgtgttggacgccaacagctatgtc
A V D G P L V I M G E V R E G V A V L D A N S Y V
2433 cgtgctgccagcgagaagcgagtcaagaagattgtggaactgctcgagaagaaggcttgtgaactgctcaaccgc XhoI
R A A S E K R V K K I V E L L E K K A C E L L N R
2508 ttccaagactag 2519
F Q D552*

ABSTRACT

Skeletal muscles, responsible for voluntary movements such as locomotion or posture maintenance, represent about 40% of body mass. This muscle mass is maintened by several signaling pathways that regulate , among other things, the balance between synthesis and degradation of myofibrillar proteins. By targeting the Akt/mTOR pathway, myostatin is a negative regulator of myogenesis. It inhibits myogenic differentiation and cell turnover. Among the various endogenous molecular factors that regulate myostatin, proteins GASP (Growth and differentiation factor Associated Serum Protein) have been described as antagonists of its activity. The Animal Genetics Unit has developed several strategies to understand the molecular mechanisms that govern the role (s) of GASP proteins during muscle development. Thus, the creation of the transgenic mouse line named surGasp-1-20 has shown that overexpression of *Gasp*-1 results in a hypermuscular phenotype associated with myofibril hypertrophy. An analysis of gene expression in myoblasts derived from satellite cells showed overexpression of myostatin correlating with an absence of hyperplasia in Gasp-1-20 mice. Similar studies currently underway for the protein GASP-2 should clarify its role in the muscular context. Proteins GASP are also defined as compound heterotypic inhibitors characterized by several inhibitory domains that can modulate the activity of different proteases. Among these different modules, the second Kunitz domain of GASP-2 was previously been described as able to inhibit trypsin. In this work, we have shown that the two whole proteins conserve this capacity of inhibition. However, our results indicate that GASP-1 and GASP-2 exhibit a difference in specificity due to the composition of the second Kunitz domain and not to the molecular environment present in each of the proteins. Finally, by modeling, we propose a structural model of the second Kunitz domain of GASP-1 and GASP-2 implicated in the anti-trypsin inhibition specificity.

Keywords : myogenesis, myostatin, GASP proteins, Kunitz domain, anti-protease activity