UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA VEGETAL

Functional analysis of RTX proteins of *Sinorhizobium meliloti* and role in biological nitrogen fixation symbiosis

Sílvia Filipa Lopes de Matos

MESTRADO EM MICROBIOLOGIA APLICADA

2013

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA VEGETAL

Functional analysis of RTX proteins of *Sinorhizobium meliloti* and role in biological nitrogen fixation symbiosis

Dissertação orientada por Prof. Doutora Leonilde M. Moreira (IST) e Prof. Doutor Mário Santos (FCUL)

Sílvia Filipa Lopes de Matos

MESTRADO EM MICROBIOLOGIA APLICADA

2013

Functional analysis of RTX proteins of *Sinorhizobium meliloti* and role in biological nitrogen fixation symbiosis

Sílvia Filipa Lopes de Matos

MASTER THESIS

2013

This thesis was fully performed at the Biological Sciences Research Group of Instituto Superior Técnico of the University of Lisbon under the direct supervision of Prof. Dra. Leonilde M. Moreira.

Prof. Dr. Mário Santos was the internal designated supervisor in the scope of the *Master in Applied Microbiology* of the Faculty of Sciences of the University of Lisbon.

Acknowledgements

Firstly I want to thank Prof. Dra. Leonilde Moreira for her patience and for all the support given during the laboratory work and during the writing of the thesis. I am grateful for her careful guidance, key to my scientific path.

I would like to acknowledge the financial support of Fundação para a Ciência e a Tecnologia through the research project PTDC/BIA-MIC/113733/2009.

To Mário Santos, Ph.D. student and my lab supervisor, thank you is not enough. For all he taught me all the tips, suggestions and especially for all the support either on the bench or on the desk. I will always remember with care not only the labor but also the hang out moments. Also to Inês Silva and Sofia Ferreira, thanks for your availability to help me, to integrate me in the work team and to the other members of BSRG for making me feel welcomed.

My thanks to Prof. Dr. Mário Santos, designated as internal supervisor, for the willingness.

To my great friends, from life and work, thank you for sticking by my side at this stage, even if with that have been living with my sorrows and anxieties. To "muita lady", for the friendship, the camaraderie, the laughs that make my belly hurt!

The last but not the least, the family! Huge thanks for always being by my side and especially during these three years. Because not always seemed easy but you have made it achievable. Because only with your support, your words or just your presence, it was possible to get here. To Diogo one additional thank for the sleepless nights during this final stretch! To my mother, father and sister and to my little star, a heartfelt THANK YOU!

Resumo

Sinorhizobium meliloti é uma bactéria fixadora de azoto que estabelece simbiose com plantas leguminosas do género *Medicago*. No decorrer da simbiose, ocorre uma troca de sinais entre a planta e a bactéria, levando à formação de nódulos nas raízes. Este processo tem como objetivo a fixação biológica de azoto surgindo, do ponto de vista agrícola, como uma alternativa ao uso de fertilizantes industriais. O genoma de *S. meliloti* codifica para inúmeras proteínas RTX, caracterizadas pela presença de sequências nonapeptídicas repetitivas ricas em resíduos de glicina e aspartato, que irão favorecer a ligação ao ião cálcio. Algumas destas proteínas estão implicadas no processo simbiótico mediando a biossíntese e/ou modificação de exopolissacáridos. Muitas ainda estão por caracterizar, apesar de algumas delas apresentarem um fenótipo simbiótico, como é o caso de SMc04171, SMb20079 e SMb20838. Uma mutação no gene *SMc04171* origina um menor número de nódulos, enquanto mutações em *SMb20079* e *SMb20838* parecem afectar a colonização dos mesmos. Tendo estas informações em mente, o objetivo deste trabalho foi caracterizar estas três proteínas.

Delineou-se inicialmente uma estratégia para construção de mutantes de eliminação em cada um dos genes, o que constituiria uma vantagem face aos de inserção por serem mais estáveis. Não obstante uma série de passos de clonagem bem sucedidos, nenhum dos mutantes de eliminação chegou a ser obtido.

Usaram-se então mutantes de inserção disponíveis para os ensaios conjuntos da estirpe selvagem e mutante de modo a avaliar se uma possível redução de competitividade poderia ser a responsável pelo fenótipo simbiótico anteriormente identificado. Para isso, inocularam-se as raízes de *Medicago* sativa com os diferentes mutantes e estirpe selvagem nas proporções 10:90, 50:50 e 90:10. Após contagem das unidades formadoras de colónias presentes nos nódulos com 5 semanas de desenvolvimento, os níveis de competição foram semelhantes entre mutantes, independentemente da proporção testada. No caso do mutante *SMc04171*, a média do número de nódulos foi maior em condições de dominância da estirpe selvagem, resultado que está de acordo

com os obtidos previamente. Em estudos anteriores, os nódulos de plantas inoculadas com os mutantes para *SMb20079* e *SMb20838* não apresentavam células viáveis. Contudo, neste trabalho verificaram-se proporções idênticas de células da estirpe selvagem e mutantes no inóculo inicial e colonização final, o que pode ser explicado pelo fornecimento da proteína RTX extracelular por parte da estirpe selvagem, compensado assim a deficiência imposta pelos mutantes.

Para caracterizar a atividade bioquímica das proteínas em estudo, os genes *SMb20079* e *SMb20838* foram clonados no vetor pWH844. Esta clonagem originou uma cauda His-tag na região N-terminal. Verificou-se que todas as proteínas recombinantes são sobre-expressas em *Escherichia coli*, ainda que só His-SMb20079 tenha sido purificada até à homogeneidade por cromatografia de afinidade. A migração em SDS-PAGE destas proteínas aconteceu para valores cerca de duas vezes superiores ao seu peso molecular, comportamento característico de algumas proteínas acídicas.

Na tentativa de identificar uma possível função destas proteínas, foram pesquisadas as homologias com outras proteínas RTX e foram construídos os modelos das estruturas tridimensionais de cada proteína. Apesar das ferramentas bioinformáticas apontarem para um possível papel na proteólise, na modificação de hidratos de carbono e até mesmo na adesão celular, a função destas proteínas ainda tem de ser determinada.

Palavras chave: *Sinorhizobium meliloti, Medicago sativa*, simbiose, fixação de azoto, proteínas de ligação a cálcio, RTX.

Abstract

Sinorhizobium meliloti is a nitrogen fixing bacterium establishing symbiosis with leguminous plants of the genus *Medicago*. During symbiosis, an intense signal exchange between bacteria and the infected plant occurs, leading to the formation of root nodules. Biological nitrogen fixation takes place in these new organs, emerging as an alternative to the use of industrial fertilizers. *S. meliloti* has one of the highest numbers of repeats in toxin (RTX) proteins, characterized by having aspartate and glycine-rich nonapeptide repeats involved in Ca^{2+}-binding. Some of these RTX proteins have been implicated in symbiosis by mediating exopolysaccharide biosynthesis and/or modification. Others, despite having a symbiotic phenotype, remain uncharacterized. Such is the case of SMc04171 whose mutation gives rise to fewer nodules, and of SMb20079 and SMb20838 which seem to have a role during nodule colonization. The aim of this work was the functional characterization of these three proteins. The first attempt was the construction of deletion mutants by gene replacement with an interposon cassette. Despite many successful cloning steps, no deletion mutant in any gene was achieved. Since there were insertion mutants available, their co-inoculation with the wild-type strain in ratios 10:90, 50:50 and 90:10 was performed in *Medicago sativa* roots. Data showed similar competitiveness for each mutant when compared with the wild-type. This can be explained by the fact that these RTX proteins are extracellular and therefore its secretion by the wild-type strain may complement mutant's deficiency in co-inoculation experiments. The three proteins under study were expressed in *E. coli* as fusions to a his-tag and protein His-SMb20079 was successfully purified to homogeneity. Despite bioinformatics tools predict that SMb20079, SMb200838 and SMc04171 proteins may have a role in proteolysis, carbohydrate modification or even cell adhesion, the function of these proteins still needs to be determined.

Keywords: *Sinorhizobium meliloti, Medicago sativa*, symbiosis, nitrogen-fixation, calcium-binding proteins, RTX.

Table of Contents

Abbreviations

(v/v)	Volume per Volume
(w/v)	Weight per Volume
µg	10^{-6} g
µm	10^{-6} m
ABC	ATP-Binding Cassette
AC	Adenylate Cyclase
Amp^r	resistance to ampicillin
BNM	Buffered Nod Medium
bp	base pair
cAMP	cyclic Adenosine MonoPhosphate
CCRH	Colonized Curly Root Hair
CFU	Colony Forming Unit
DNA	DeoxyriboNucleic Acid
EDTA	EthyleneDiamineTetracetic Acid
EPS	ExoPolySaccharide
Fig.	Figure
Gm^r	resistance to gentamicin
IPTG	IsoPropyl-β-D-ThioGalactopyranoside
Kb	thousand of base pairs
kDa	thousands of daltons
Km^r	resistance to kanamycin
LB	Luria Bertani
LPS	LipoPolySaccharide
MARTX	Multifunctional Autoprocessing RTX
MFP	Membrane Fusion Protein
mg	10^{-3} g
mM	10^{-3} moles per litre
Nm^r	resistance to neomycin
NTA	NitriloTriacetic Acid
^oC	Degrees Celsius
OMP	Outer Membrane Protein
PCR	Polymerase Chain Reaction
rpm	rotations per minute
RTX	Repeats in ToXin
SB	Super Broth
SD	Standard Deviation
SDS-PAGE	Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis
Sm^r	resistance to streptomycin
Tc^r	resistance to tetracicline
TCA	TriCarboxylic Acid
TISS	Type I Secretion Systems
TY	Tryptone-Yeast extract
WT	Wild-Type

1. Introduction

1.1. The nitrogen cycle and importance of biological nitrogen fixation

Despite of the large amount of free dinitrogen in the atmosphere, this cannot be used directly by living organisms. Therefore, decomposers play a key role in converting organic nitrogen products into ammonia and finally into very soluble nitrates, which will be absorbed by plants for the formation of amino acids, nucleic acids, among others. Still, nitrogen availability is quite limited causing a problem in agriculture. To overcome this obstacle, nitrogen enriched fertilizers of industrial origin that increase productivity are used. However, this overuse has serious consequences not only in the consumption of fossil fuels, but also in the increased pollution of ground water and the greenhouse effect (Zahran, 1999). It is therefore extremely important to reduce this use through alternative methods, such as biological nitrogen fixation. In this process, N_2 is reduced to NH_3, a reaction catalyzed by a nitrogenase enzyme of bacterial origin. *Rhizobium* is one of the genus that in symbiosis with leguminous plants plays an important role in nitrogen fixation. When it happens, the symbiosis leads to the formation of structures called nodules in the roots of the host plant where nitrogen fixation takes place. *Sinorhizobium meliloti* (also known by *Ensifer meliloti*) is the symbiont of plants from the *Medicago* genus which present indeterminate nodules, a classification based on the pattern of meristem growth that, by being persistent and continuous, allows individualization of the various developmental stages. Therefore the model system for the biological nitrogen fixation symbiosis is *Medicago truncatula* and *Sinorhizobium meliloti* (Gibson et al., 2006). Further details on the interaction steps will be given bellow.

1.2. The symbiosis between rhizobia and leguminous plants

1.2.1 The symbionts *Sinorizhobium meliloti* and *Medicago truncatula*

Medicago truncatula is a small Mediterranean plant with a small genome. Apart from these, characteristics such as self-fertilization, rapid generation time, prolific seed production, and amenable to genetic transformation, make it an

excellent model organism within the group of leguminous plants for the study of nitrogen fixation processes. *Sinorizhobium meliloti* is a alpha-proteobacterium capable of fixing atmospheric nitrogen when establishing symbiosis with leguminous plants of the *Medicago*, *Melilotus* and *Trigonella* genus (Galibert et al., 2001).

The symbiotic system has evolved in a way that when the plant is infected, it starts an exchange of signals that lead to the formation of a root nodules, which will be responsible for nitrogen fixation. Once the bacterium is endocytosed by nodule plant cells, it differentiates and becomes able of converting atmospheric nitrogen into ammonia. These processes are dependent on a variety of environmental conditions, including soil pH, temperature, salinity and oxidative stress. In turn, the plant can grow in the absence of external nitrogen, as the fixation made by the bacterium ensures the existence of this compound. This mechanism is very important in an ecological and agricultural point of view because it allows the plant to grow without the use of nitrogen-based fertilizers (Becker et al., 2009).

The *S. meliloti* genome contains a circular chromosome (3.65 Mb) and two megaplasmids pSymA (1.35 Mb) and pSymB (1.68 Mb) (Galibert et al., 2001) with a 6204 protein-encoding genes: 3341 on the chromosome (Capela et al., 2001), 1293 on pSymA (Barnett et al., 2001) and 1570 on pSymB (Finan et al., 2001). More recently, 86 new putative genes were identified, removed 66 previously predicted orphan genes and adjusted the start positions of 360 coding regions (Becker et al., 2009). In this last work, putative functions were assigned to 313 proteins formerly classified as hypothetical or conserved hypothetical. As a result, more than 71% of genes have now a predicted function (Becker et al., 2009).

1.2.2 Steps of symbiotic development

Although the plant cell wall is an important barrier against most bacterial species, the intense signal exchange between *S. meliloti* and *M. sativa* allow bacterial entrance through the plant root hairs. Flavonoids are the first compounds secreted by plants to the rhizosphere to signal rhizobial bacteria so that they produce Nod factors. For example, *M. sativa*-derived flavonoid luteolin

stimulates binding of an active form of the transcriptional regulator NodD1 to *S. meliloti* “nod-box” promoters, which activates transcription of the downstream *nod* genes (Jones et al., 2007). The expression of the *nod* genes results in the biosynthesis of the signaling molecules called Nod factors (Fig. 1a, b). This signaling molecules are perceived by root epidermal cells and a genetic program leading to nodulation is initiated. Raising intracellular calcium levels is the plant’s first response to the production of Nod factors. This is followed by severe oscillations in its concentration as well as changes in the root hair structure. Simultaneously, mitosis restart in root cells induced by Nod factors giving rise to the nodule primordium (Downie, 2010). At the same time, at the tip of the root hair, a new membrane is formed which causes a reversal in the growth and subsequent reorganization of polarity. Bacteria are then internalized on the root, through the production of infection threads (Fig. 1c). Numerous mutations, such as the inability to produce cyclic β-glucans or deficient production of low molecular weight exopolysaccharides, can prevent bacterium entry through CCRH (colonized curly root hair). *S. meliloti* produces the exopolysaccharides succinoglycan (also known as exopolysaccharide I, EPSI) and galactoglucan (EPSII), which facilitate infection thread formation. During infection thread progression, new cycles of infection are stimulated in successive layers of the root cells. The plant hormone cytokinin and the Nod-factor dependent reinitiation of the cell cycle are involved in directing infection threads to the plant cortex (Jones et al., 2007). The nodule meristem is formed and it is responsible for the continuous cellular division and consequently for nodule growth. These meristematic layers become polyploid, through cycles of genomic endoreduplication without cytokinesis, indispensable for the nodules to become functional. This mechanism allows transcription and metabolic processes at higher rates. Once bacteria reach the target nodule tissue, they are internalized by a layer of cortical cells through endocytosis that puts them in unique compartments, the symbiosomes (Fig. 1d, e). The lipopolysaccharide (LPS), a component of the outer membrane, is a major cellular defense mechanisms against external agents. It consists in a lipid A membrane anchor attached to a polysaccharide core that in turn is attached to an O-antigen repeating unit. Some *S. meliloti* genes are necessary to enable the production of LPS components essential for survival within the plant host cell. BacA gene

product for example, is required for lipid A production so that bacteria survive within the host. A mutation in this gene modifies the morphology of a characteristic bacteroid and induces its lysis as soon as bacteria are internalized and differentiate. A mutation in the *lpsB* gene changes the sugar composition of the LPS such that the bacteroids do not complete the differentiation process and thus no nitrogen fixation occurs (Jones et al., 2007). Mutations in non-LPS factors may also form deficient bacteroids, which do not fix nitrogen. In indetermine nodules, the internalized bacteria and the symbiosome membrane divides simultaneously, before bacteroid differentiation. In cases like this, there is an endoreduplication program on the invading bacterial cells during which bacteroids are able to increase their size and DNA content, increasing their metabolic rate which translates into an advantage for nitrogen fixation (Prell et al., 2010).

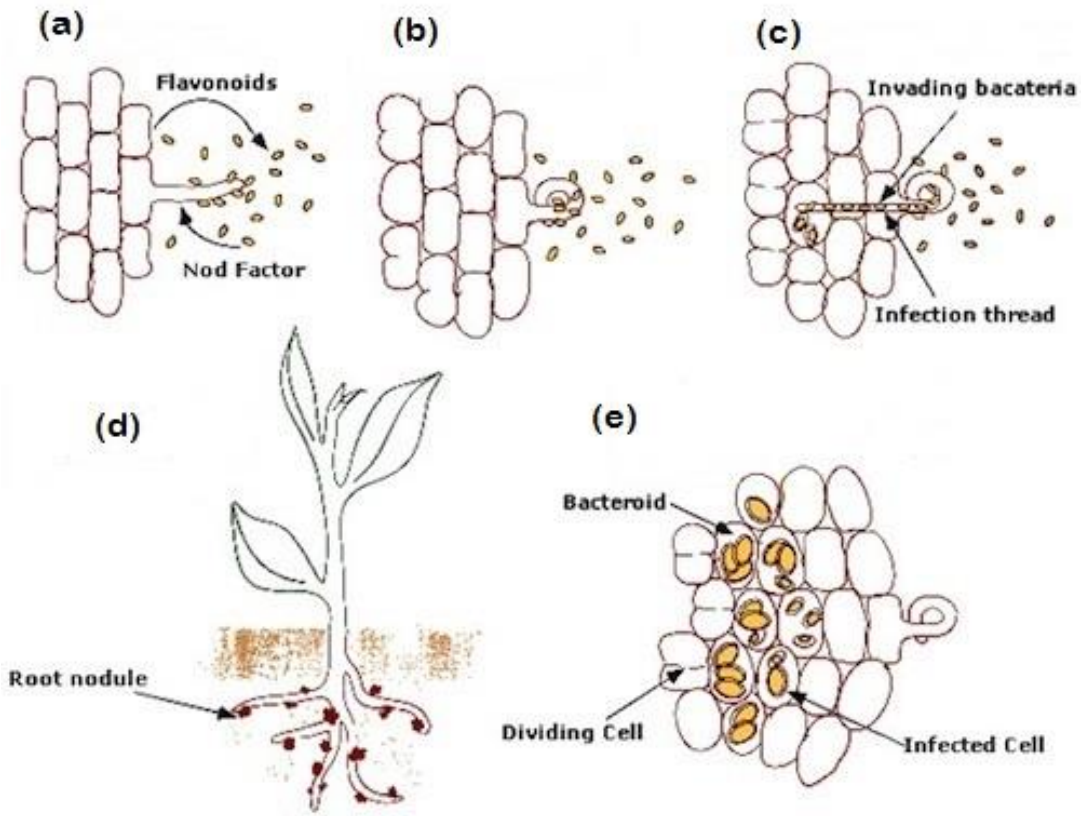


Figure 1 – Stages in the biology of the nodulation process. (a) Chemical recognition between the symbionts with flavonoids produced by the plant and the Nod factors by the bacteria. (b) Deformation of root hair and root cortex cell division initiation. (c) Formation of the infection thread containing bacteria. (d) Legume provides Rhizobia with carbon sources and Rhizobia provides the legume with NH_4^+. (e) Nodule tissue formation and bacteroid differentiation leading to nitrogenase and leghemoglobin synthesis (adapted from Spaink et al., 1998).

1.2.3. Nitrogen fixation within bacteroids and assimilation by the plant

After bacterial internalization, in an environment with low oxygen content, enzymes of the nitrogenase family are expressed and the chain reactions that lead to nitrogen fixation starts. The biosynthesis of nitrogenase complex is controlled by an oxygen sensor, which also controls the microaerobic respiratory enzymes that are required to provide energy and reductant to nitrogenase. This sensor contains the oxygen-sensing two-component regulatory system FixL and FixJ, NifA, σ54 and FixK (Bauer et al., 1998). Most of the changes that occur during bacteroid differentiation are regulated by this sensor, which is sensitive to low concentrations of oxygen. The respiratory activity provides the necessary energy for nitrogenase to reduce N_2 to NH_4^+, which is then secreted by bacteroids to the plant cells cytoplasm. When this happens, it is conducted through the peribacteroid membrane NH_4^+ channels and then assimilated. There is a need of a carbon source that provides metabolites and energy for nitrogen fixation. The fixed carbon during plant photosynthesis is the source available to the bacteria in the form of dicarboxylic acids (malate, e.g) (Maunoury et al., 2010). The NAD^+-malic enzyme responsible for the direct production of pyruvate is needed for nitrogen fixation by *S. meliloti*. This suggests that the production of acetyl-CoA from malate using malic enzyme and pyruvate dehydrogenase is important for the funneling of carbon into the TCA cycle in bacteroids (Jones et al., 2007).

Nitrogen assimilation into amino acids is carried out by the plant glutamate synthase, but it is also recognized the role of Leghaemoglobins, oxygen-binding proteins produced by plants and responsible for the reddish hue present in the functional nodules. It is thought that they are responsible for monitoring the microaerobical environment within the nodules. It is also necessary a sucrose synthase that catabolizes sucrose to glucose and fructose which will be metabolized to malate and transported to bacteroids (Jones et al., 2007).

1.3. The highly diverse family of secreted proteins containing RTX domains

RTX proteins are produced by numerous Gram-negative bacteria and are characterised by having nonapeptide repeats of glycine and aspartate-rich sequences, located at the C-terminal of the protein. These repeats form numerous sites for the binding of Ca^{2+} ions and are at the origin of the protein family name, where RTX stands for Repeats in ToXin (Welch, 2001). The number of RTX repeats vary among RTX proteins from <10 to >40 (with a consensus sequence X-(L/I/F)-X-G-G-X-G-(N/D)-D, where X means any residue). The RTX proteins contain an ~60-residue-long C-terminal secretion signal that is not processed during secretion by type I secretion systems (TISS), promoting direct translocation from the cytoplasm to the extracellular space (Linhartová et al., 2010). Type I secretion is dependent on three proteins: a polytopic inner membrane protein with a cytoplasmatic ATPase domain operating as an ABC exporter, a membrane fusion protein (MFP) and an outer membrane protein (OMP) (Fig. 2). The MFP spans out from the inner membrane into the periplasm and contacts both the inner membrane ABC exporter and the OMP. The OMP trimeric export channel in the outer membrane is TolC protein (Koronakis et al., 2000). Binding of calcium ions to the repeats of RTX proteins occurs only upon secretion, as the intracellular cytoplasmic Ca^{2+} concentration in bacteria is quite low (Gangola & Rosen, 1987). To be secreted, the RTX proteins need to remain unfolded before translocation out of the cell by TISS. Once outside, calcium binding to the nonapeptide repeats promotes the correct functional conformation (reviewed in Linhartová et al., 2010).

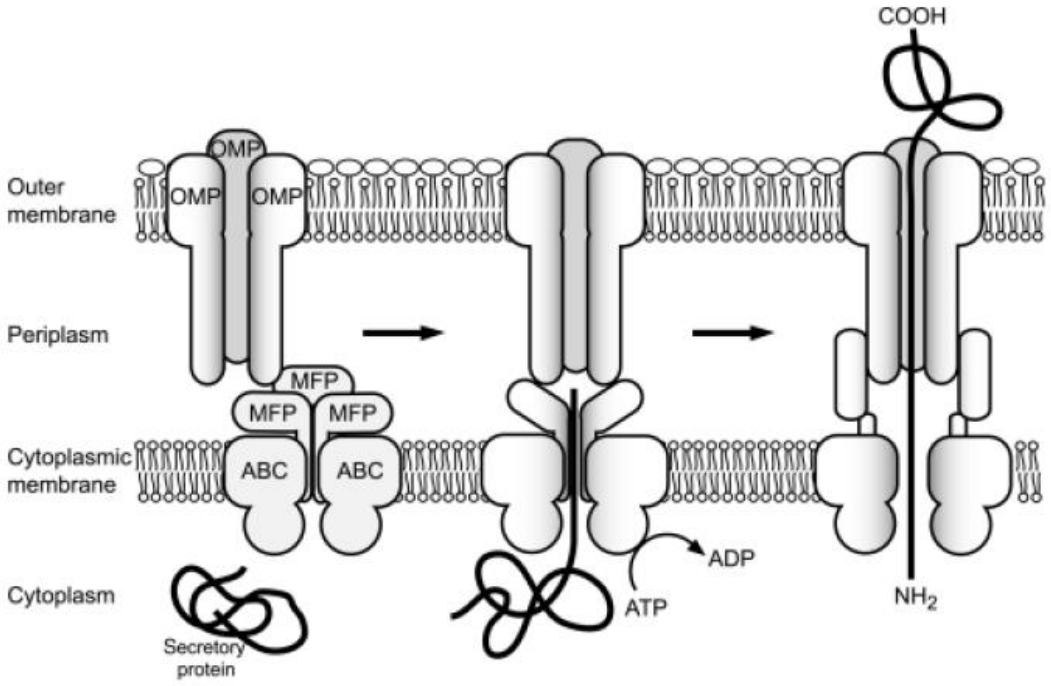


Figure 2 – TISS assembly operation. The C-terminal signal recognition of a RTX-protein allows the formation of a complex between an ABC and MFP transporter that connects with the trimeric OMP. A channel is formed through which the protein is exported. The export occurs in a single step, with a direct passage from the bacterial cytoplasm to the outer surface. After secretion, Ca^{2+} ions responsible for the loading of RTX repeats induce the folding and acquisition of biological activity (adapted from Linhartová et al., 2010).

One of the most well characterized RTX leukotoxins is HlyA of *E. coli*, characterized by exhibiting a cytotoxic pore-forming activity, detected as a hemolytic halo surrounding bacterial colonies grown on blood agar plates (Welch, 1991). The C-terminal signal of HlyA is recognized by a complex formed between the MFP HlyD and the ABC transporter HlyB. Upon binding of HlyA, the HlyD protein interacts with the OMP TolC, causing a conformational change and exporting HlyA (Andersen et al., 2001). This protein complex is transient and reverts to a resting state once the RTX protein is secreted. The C-terminal calcium-binding site of HlyA contains more than a dozen of glycine and aspartate rich nonapeptide β-strand repeats, whereas the N-terminal has nine amphipathic α-helices. Although initially assumed the N-terminal region as the main responsible for the interaction with the eukaryotic cell membrane, it is now known that this interaction occurs at both ends, with the calcium binding domain being the responsible for the initial steps of docking.

1.3.1 Types and biological functions of RTX proteins

The most studied class of RTX proteins are cytotoxins produced by a broad range of Gram-negative pathogens from the genera *Escherichia*, *Bordetella*, *Proteus*, *Vibrio*, among others (reviewed in Linhartová et al., 2010). RTX cytotoxins include the broadly studied pore-forming leukotoxins (e.g. HlyA from *E. coli* and CyaA from *Bordetella pertussis*) and the more recently discovered MARTX (multifunctional autoprocessing RTX toxin) with the prototype being $MARTX_{VC}$ from *Vibrio cholera*. HlyA is a polypeptide of 107 kDa secreted by uropathogenic, as well as many commensal fecal isolates of *E. coli*. This cytolytic (pore forming) RTX leukotoxin is synthesized as inactive protoxin that undergoes post-translational activation before export. This consists in post-translational modification of ε-amino group from internal lysine residues by covalent attachment of amide-linked fatty acyl residues. HlyA exhibits cytotoxic activity on a wide spectrum of cells from various species, including erythrocytes, granulocytes, monocytes, endothelial cells or renal epithelial cells from mice, ruminants and primates. CyaA from *Bordetella pertussis* is a 1706-residue-long polypeptide being a bifunctional toxin in which a cell-invasive adenylate cyclase (AC) domain has been fused to a pore-forming RTX moiety. CyaA has post-translational palmitoylation of lysine residues and ~40 calcium binding sites (Rose et al., 1995). CyaA primary targets are the leukocytes expressing the $\alpha_M\beta_2$ integrin (CD11b/CD18). The AC domain of CyaA is delivered directly across the cytoplasmic membrane and binds to calmodulin, which then catalyzes uncontrolled conversion of cellular ATP to cAMP, subverting several signaling cascades essential in the phagocytic cells (Vojtova et al., 2006).

The other group of RTX-cytotoxins are the very large proteins of the MARTX. They differ from all other RTX by the size (3000-5000 amino acids), molecular structure and RTX gene cluster organization. These MARTX have been identified in species of *Vibrio*, *Yersinia*, *Proteus*, and a few other (Satchell, 2007). In contrast to other toxins, the C-terminal repeats exhibited an 18-residue-long consensus sequence motif. Little is known about the post-translational modifications, but recent studies seem to show that the acylation may not be essential for all the toxins of this group. The prototype *Vibrio cholera*

$MARTX_{VC}$ was shown to cause rounding of epithelial cells by catalyzing covalent cross-linking of cellular actin (Cordero et al., 2006).

Besides RTX-cytotoxins, there is also a wide range of other RTX proteins including (1) a particular class of RTX proteins containing more than one cadherin domain; (2) RTX proteases with a zinc metallopeptidase domain secreted by various pathogens; (3) RTX bacterial lipases which present numerous advantages in the food, biotechnology and biomedical industry; (4) RTX bacteriocins which are able to inhibit the growth of other strains; (5) S-layer RTX proteins capable of establish one of the simplest systems of self-assembly; (6) and RTX proteins involved in the motility of *Cyanobacteria* (Linhartová et al., 2010).

RTX proteases are part of a subgroup of metalloproteases, the metalloendopeptidases, and contain histidine residues involved in binding zinc ions. They are described in six genera: *Serratia*, *Erwinia*, *Pseudomonas*, *Proteus*, *Caulobacter* and *Photorhabdus*. The genetic organization of RTX proteases loci varies and is dependent on the number of secreted proteases. There are examples of species where one protease is secreted (*Erwinia amylovora*), two (*P. mirabilis*) and up to four (*E. chrysanthemi*). There are still species in which the operon can be combined with genes for RTX lipases, such as *P. fluorescens*. In vitro, the activity of these proteases is inhibited either by the action of EDTA, or by *o*-phenanthroline, a specific inhibitor (Linhartova et al., 2010). Metalloproteases from species like *P. aeruginosa* and *S. marcescens* have a structure in which the proteolytic N-terminal has a folding topology similar to astacin and the C-terminus contains a parallel β roll, with successive β bands connected to Ca^{2+} ions by the repetitive sequence GGXGXD (Bode et al., 1996). Cell own protection against the activity of these metalloproteases is achieved by synthesizing a RTX protease inhibitor. For example, *Erwinia chrysanthemi* synthesizes the protein Inh, exported into the periplasm by sec-dependent secretion (Letoffé et al., 1989). This inhibitor protein is presumed to protect periplasmic proteins against proteases that might potentially leak out from the TISS channel.

One of the best characterized metalloproteases is the AprA protease of *P. aeruginosa* which has hydrolytic and anticoagulant activity in human plasma, contributing to its most important action during septicemia that is hemorrhagic lesions lacking thrombi (Fetzer et al., 1967). Purified *P. aeruginosa* AprA also plays a direct role in the invasion and subsequent necrosis of tissues, as well as degrading the human γ-interferon and inhibiting its biological activity (Horvat & Parmely, 1988). AprA is also capable of degrading some opsonins, facilitating the survival and proliferation of *Pseudomonas* in plasma.

Another example of RTX proteins are the RTX lipases. These enzymes have been extensively characterized and show several applications for biomedical, biotechnology and food industry. *Pseudomonas* and *Serratia* are responsible for the production RTX lipases of I.3 subfamily which do not have cysteine residues, do not require additional products for their activity, and are secreted by the TISS system (Jones et al., 2007). For example, in *P. fluorescens* B52, lipase production is repressed by iron and regulated by temperature, with production occurring far below the optimum growth temperature (Woods et al., 2001). NaCl also intervenes in the process, reducing lipase secretion (McCarthy et al., 2004).

1.3.2. RTX proteins of rhizobia

A survey of several sequenced genomes from rhizobia species (Linhartová et al., 2010) indicates that the number of genes encoding putative RTX proteins varies between 3 in *Mesorhizobium loti* and 17 in *S. meliloti*. Symbiotically important species such as *Bradyrhizobium japonicum* and *Rhizobium leguminosarium* have 8 and 10 RTX protein enconding genes, respectively. Despite their identification in the genomes, only a few have been experimentally characterized. The first protein to be characterized was the 30 kDa RTX protein, NodO, from *Rhizobium leguminosarium* bv. *viciae* and was shown to be important in pea and vetch nodulation (Economou et al., 1990). NodO forms cation-selective channels and there are two hypotheses for how it may increase nodulation. The first suggests that the passage of nodulation factors may be facilitated by pore formation in root cell membrane. The other possibility could be an amplification of the signaling resulting from calcium fluxes through NodO

channels across the plasma membrane, thus increasing plant response to Nod factors (Sutton et al., 1994). In *S. meliloti* three RTX proteins have been characterized. These are the endoglycanases ExsH and EglC acting on depolymerization of succinoglycan and WgeA (formerly ExpE1) required for galactoglucan biosynthesis. The two endoglycanases are 99% identical at the amino acid level and contain a single nonapeptide Ca^{2+}-binding repeat. Both proteins influence symbiotic efficiency as they contribute to the production of the symbiotically active low molecular weight succinoglycan (Sharypova et al., 1999; York & Walker, 1998). WgeA has been shown to be secreted by TISS composed of WgdAB (formerly ExpD1, ExpD2) and to bind Ca^{2+} ions (Moreira et al., 2000). The deletion of *wgeA* gene has no influence in symbiosis. WgeA is required for galactoglucan biosynthesis although the exact mechanism is not determined.

2. Objectives

Although most of the genes that are specific and crucial for establishment of a successful symbiosis between *Sinorhizobium meliloti* and *Medicago* plants have been characterized, other genes influencing symbiosis in more subtle ways still need to be identified. With that in mind, Santos (2008) prepared insertion mutants in each of the putative RTX proteins from *S. meliloti* 1021 and studied their effect on several nodulation properties. The number of nodules elicited by each mutant was 20 to 40% lower in the cases of SMc04171, SMc00286 and SMa0034 disruption, and it was 20-30% higher for SMa2111 and SMb21543 gene disruption. Furthermore, colony forming units (CFU) count inside mature pink nodules showed a residual amount of viable non-differiated bacteria when colonization was done with mutants of SMa0034, SMb20079, SMb20838, SMb21543 and SMb21397 (Santos, 2008). These results, together with expression data showing increased expression of SMc00286, SMc04171, SMa2111 and SMa0034 (Ampe et al., 2003) during early nodulation, suggested that some RTX encoding genes may have a role in symbiosis. For this reason we have chosen SMc04171, SMb20079 and SMb20838 to further characterize their involvement in *S. meliloti* symbiosis. To do that it was planned to prepare deletion mutants in each of the genes under study to further evaluate nodule number and nodule occupancy with or without competition with the wild-type strain. In addition, it was also envisaged to purify each of the recombinant proteins (fused to a His tag) overexpressed in *E. coli* to perform biochemical tests and elucidate the biological function of each of them.

3. Materials and Methods

3.1. Bacterial strains and plasmids

Bacterial strains and plasmids used are listed in table I.

Table I – Strains and plasmids used in this study.

Strain or plasmid	Relevant characteristics	References
S. meliloti		
Sm1021	Wild-type, SU47, Sm^r	Meade et al., 1982
Sm2011	Wild-type, Nx^r, Sm^r	Casse et al., 1979
SmAC78-2	Sm2011, SMc04171::pAC78-2	Santos, 2008
SmMAS710-2	Sm2011, SMb20079::pMAS710-2	Santos, 2008
SmAC78-4	Sm2011,SMb20838::pAC78-4	Santos, 2008
E. coli		
DH5α	*recA1* Δ*lac*U169,φ80 *lac*ZΔM15	Hanahan, 1983
S-17	*E. coli* 294, *thi* RP4-2-Tc::Mu-Km::Tn7 chromosomally integrated	Simon et al., 1983
XL1-Blue	*recA1 lac* [F' *proAB* $lacI^q$ *ZαM15* Tn*10* (Tc^r)]	Bullock et al., 1987
Sure	*thi* $e14^-$ (*mcrA*) Δ(*mcrCB-hsdSMR-mrr)171 endA1 supE44 thi-1 gyrA96 relA1 lac recB recJ sbsC umuC:*Tn5 (Km^r) *uvrC* [F' *proAB* $lacI^q$ *ZαM15* Tn*10* (Tc^r)]	Stratagene
Plasmids		
pMS252	pSVB30 derivative carrying the *aacC1* gene of Tn1996, Amp^r, Gm^r.	Becker et al., 1995
pAS137-1	pK19mob sac containing the flanking regions of *SMb20079* gene.	This work
pAS137-2	pAS137-1 derivative containing the *aacC1* cassette between the *SMb20079* flanking regions	This work
pWH844	pQE9 derivative, $lacI^q$, Amp^r	Schirmer et al., 1997
pSM129	pWH844 containing the 1083 bp fragments of *SMb20838* gene, Amp^r	This work
pSM1210	pWH844 containing the 3240 bp fragments of *SMb20079* gene, Amp^r	This work
pAC712	pwH844 containing the 1676 bp fragments of *SMc04171* gene, Amp^r	
pK19mobG	oriColE1 mob^+, $lacZ^+$, gusA, Km^r	Schäfer et al., 1994
pK19mobsac	oriColE1 mob^+, $lacZ^+$, sac^+, Km^r	Schäfer et al., 1994

Sm^r, streptomycin resistance; Tc^r, tetracycline resistance; Amp^r, ampicillin resistance; Km^r, kanamycin resistance; Gm^r, gentamicin resistance.

Fresh cultures were obtained by transferring a cellular frost portion at -80ºC to a plate with selective medium, incubated at indicated temperature and then kept at 4ºC.

3.2. Plant growth conditions

Medicago sativa genotype Europa seeds were surface sterilized with concentrated sulphuric acid for 10 minutes and then washed several times with sterile water for 5 minutes each and plated on agar 1.5%. The plates with the seeds were kept at 4ºC for 2 days, then allowed to germinate for 24 hours in the dark at room temperature.

For the co-inoculation assay, three plantlets were placed on each plate containing BNM medium (Rolfe et al., 1980) and inoculated with 100 µl of *S. meliloti* suspension (10^8 – 10^9 cells). Plants were grown for 4 weeks under 16h-8h cycles (light-dark). During this period, white (immature) and pink (mature) nodules of each plant were counted. After 4 weeks mature nodules were removed, sterilized with ethanol 70% for 1 minute followed by wash with sterile water and crushed in TY medium. The resulting suspensions were plated on TY and on TY supplemented with neomycin.

3.3. Bacteria growth conditions

E. coli strains were grown in LB medium (Sambrook et al., 1989) at 37ºC with orbital agitation of 250 rpm. *S. meliloti* strains were grown in TY medium (Beringer, 1974), at 30ºC with orbital agitation of 180 rpm. For solid media 20 g/L of agar were added. In overexpression assays, *E. coli* Sure harbouring the different plasmids was grown in SB medium (Botstein et al., 1975), at 37ºC.

Media were supplemented with antibiotics at the following concentrations: for *S. meliloti*, streptomycin 600 µg/ml, neomycin 120 µg/ml and gentamicin 40 µg/ml; *for E. coli*, kanamycin 50 µg/ml, ampicillin 100 µg/ml and gentamicin 10 µg/ml. The antibiotics were dissolved on water and sterilized by filtration (filter with a 0.2 µm pore).

3.4. Extraction of genomic and plasmid DNA

Genomic DNA of *S. meliloti* was extracted by cell lysis with glass beads, purification with phenol and chloroform and precipitation with isopropanol according to Sambrook & Russell (2001). Plasmid DNA was extracted using the minipreps method (Sambrook & Russell, 2001). For analyzes in which greater purity was required, the QIAprepR Spin Miniprep Kit (Qiagen) was used, following the procedures recommended by the manufacturer.

3.5. Construction of *S. meliloti* deletion mutants

In the preparation of *S. meliloti* deletion mutants for *SMb20079*, *SMb20838* and *SMc04171* genes, the flanking regions of each gene were cloned in pK19mobG or pK19mobsac. Table II shows the primers used in the amplification of each flanking region.

Table II – Primers used to amplify the flanking regions of the *SMb20079*, *SMb20839* and *SMc04171* genes.

ORF	Fragments	Forward Primer*	Reverse Primer*	Annealing Temperature (ºC)	Expected product Size (bp)
SMb20079	Left	GAG AAGCTT AAT CTT GCC AGG	GAT GGATCC CTC GTC GTG TGA	64.2	1895
	Right	GAT GGATCC CTC GTC GTG TGA	CGG GAATTC GAG CAA GTT GAA	67.3	2042
SMb20838	Left	CGG GGTACC TGC CCT TCA ATG	CGC TCTAGA ATT GCC CGA CCT	65.5	2098
	Right	CTG TCTAGA GAC GAT TTC TTC	CTA AAGCTT TCG TCG GTT ATC	62.1	2047
SMc04171	Left	GCG GGTACC CAT ATT TGG TGA	ATG TCTAGA ATC CCC TGA GTT	64.2	1994
	Right	GTC TCTAGA GTT TGA GGG ATG	TCG AAGCTT ACG AAG CCT ACA	64.2	2026

* Restriction sites are underlined.

The amplification reactions started with an initial step of denaturation at 96ºC for 30 seconds, followed by 30 cycles of: 10 seconds at 96ºC (denaturation), 30 seconds at an optimized annealing temperature for primer annealing (Table II), and 2 minutes at 72ºC (polymerization). Finally, an elongation step at 72ºC for 7 minutes was included. PCR products were then loaded into a 0.8% agarose gel that after completion of electrophoresis was stained with GelRed (Biotium). Each PCR product was digested with restriction enzymes as shown in Table II and cloned into appropriate vectors. The plasmid

for *SMc20079* deletion (pAS137-1) was further digested with BamHI to insert the *aacC1* gentamicin resistance cassette between the two flanking regions. The plasmid obtained (pAS137-2) (Table I) was mobilized from strain *E. coli* S17-1 to *S. meliloti* 1021 by conjugation and colonies selected in the presence of gentamicin. To distinguish colonies where single crossover or double crossover had occurred, each colony was grown in medium containing neomycin. The ones unable to grow in neomycin were potential deletion mutants with the *aacC1* cassette replacing the gene of interest. The candidate deletion mutants were further confirmed by PCR amplification.

3.6. Construction of plasmids expressing recombinant proteins

To obtain the recombinant proteins of SMb20079, SMb20838 and SMc04171, each gene coding region was cloned into pWH844 expression vector. On table III are represented the primers used in the amplification.

Table III – Primers used to amplify *SMb20079*, *SMb20839* and *SMc04171* genes.

ORF	Forward Primer*	Reverse Primer*	Annealing Temperature (ºC)	Product Size (bp)
SMb20079His	ATT CTGCAG CAG GTC ATC AAC	CCT AAGCTT GGG CGA TCA CAC	67.8	3240
SMb20838His	ATT GGATCC GTT TTG CCG AAA	CCA AAGCTT CTG CGA TCA AAG	65.5	1083
SMc04171His	ATA CTGCAG GCC TTC TTC GCA	CCA AAGCTT CGC GAA GGC ATC	66.5	1676

* Restriction sites are underlined.

The amplification reactions started with an initial step of denaturation at 94ºC for 3 minutes, followed by 34 cycles of 45 seconds at 94ºC (denaturation), 30 seconds at an optimized annealing temperature for primer annealing (Table III) and 3 minutes at 72ºC (polymerization). Finally, an elongation step at 72ºC for 6 minutes was performed. PCR products were digested with appropriate restriction endonucleases and cloned into pWH844. Plasmids obtained were named pSM129, pSM1210 and pAC712 (Table I) and were further confirmed by DNA sequence determination.

3.7. Protein overexpression and purification

E. coli Sure cells containing pSM129, pSM1210 and pAC712 were grown in superbroth medium containing 100 mg/L of ampicillin, at 37ºC, to an optical density at 640 nm of 0.5. Induction of gene expression was started by adding 1mM IPTG, followed by incubation for several hours. To search for changes in protein expression patterns, an appropriate culture volume was harvested at different times and cell pellets were solubilized in SDS sample buffer (100 mM Tris-HCl pH 6.8; 4% (w/v) SDS; 20% (v/v) glycerol; 0.2 M β-mercaptoethanol; 0.2% (w/v) bromophenol blue) to yield a preparation of total cellular proteins. Proteins were then visualized by vertical SDS-PAGE in 12.5% acrylamide (Sambrook & Russell, 2001) and stained with Coomassie Brilliant Blue R-250.

The His-tagged proteins were purified from cell-free lysates of *E. coli* obtained after sonication in START buffer containing 10 mM imidazole. After removal of cell debris by centrifugation at 17600 *xg* for 60 minutes at 4ºC, His-tagged proteins were purified by Ni^{2+}-affinity chromatography. After loading column with the protein extract, the Ni^{2+}-nitrilotri-acetic acid matrix was washed with 10 volumes of START buffer containing 20 mM imidazole. Bound proteins were eluted with START buffer containing increasing concentrations of imidazole up to 250 mM. Protein concentration of the purified His-tagged protein solutions were determined by the method of Bradford (Bradford, 1976), using bovine serum albumin as the standard.

3.8. Western Immunoblot analysis

Protein sample were separated on 15% SDS-PAGE gels prior to being transferred onto nitrocellulose membranes according to the procedure of Towbin et al (1979). The membranes were incubated with a monoclonal antibody against the His-tag (GE Healthcare Life Sciences) at a dilution of 1:2000 for 2 hours. The membranes were then incubated another 2 hours with alkaline phosphatase-conjugated rabbit anti-mouse (Santa Cruz Biotechnology) and reacted with ECL reagent (GE Healthcare Life Sciences) for signal detection. Broad-range SDS-PAGE standard from Bio-Rad was used as a molecular weight marker (annex I).

4. Results

4.1. *In silico* analysis of the RTX proteins SMc04171, SMb20079 and SMb20838

Gene *SMc04171* is located in the chromosome of *S. meliloti* 1021 and encodes a protein of approximately 55 kDa. Upstream, and in the opposite direction of transcription, is located a gene encoding putative adenylate/guanylate cyclase (*SMc04172*). The downstream region, in the same transcription orientation, encodes a putative two-component receiver domain (*SMc04170*). The analysis of this flanking region in other genome sequenced strains of *S. meliloti* (now *Ensifer meliloti*) shows conservation in both upper and lower regions of *SMc04171* gene homologues. When this analysis was extended to other species of *Ensifer arboris* and *Ensifer medicae* the result varies between full conservation and variability. This result suggests that *SMc04171* gene product function may be unrelated to *SMc04170* and *SMc04172* gene products function. The search for identities at the amino acid level of SMc04171 showed that besides other *Ensifer meliloti* strains, the best identities were obtained for proteins of *Ensifer arboris*, *Ensifer medicae* and *Ochrobactrum anthropi* (Table IV). The search for conserved motifs using Pfam database (http://pfam.sanger.ac.uk/) identified 7 putative hemolysin-type calcium-binding repeats (2 copies) corresponding to a total of 14 nonapeptide repeats (Fig. 3). Using the I-TASSER server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) for 3D structure and function prediction it was built the model shown in Fig. 4A with a TM-score of 0.68±0.12 (c-score is typically in the range of [-5,2], where a c-score of higher value signifies a model with a high confidence). The templates used by I-tasser were lipases from *Pseudomonas sp.* and *Serratia mascescens*, polysaccharide lyases from *Streptomyces coelicolor*, *Proteus vulgaris* and *Bacillus sp.* and a mannuronan C-5 epimerse from *Azotobacter vinelandii*. Based on that, the biological processes with higher gene ontology scores are lipid metabolic processes with 0.44 and carbohydrate metabolic processes with 0.32 (values range in between [0-1], where a higher value indicates a better confidence in

predicting the function using the template). Despite being 28% identical to *Pseudomonas sp*. MIS 38 lipase, SMc04171 does not have the conserved [LIV] – {KG} – [LIVFY] – G – [HYWV] – S – {YAG} – G – [GSTAC] motif characteristic of serine active site lipases. Gene expression data shows that the transcript of *SMc04171* is increased during infection of young nodules of *Medicago* plants (Ampe et al., 2003), but decreased under heat shock (Barrett et al., 2012). In addition, *SMc04171* was detected in the supernatant of *S. meliloti* grown in MOPS-buffered medium (Cosme et al., 2008). These observations suggest a role of protein *SMc04171* both in symbiosis and free-living, but the function of this enzyme remains to be determined.

Table IV – RTX proteins similarity to amino acid sequences in databases.

Protein	Organism	% Identity (Similarity)	Accession n.º
SMc04171	*Ensifer meliloti* (16 strains)	98-100 (99-100)	WP_ 003530081.1
	Ensifer arboris	93 (96)	2512533422
	Ensifer medicae	83 (91)	WP_018012380.1
	Ochrobactrum anthropi	38 (50)	WP_010660505.1
SMb20079	*Ensifer meliloti*	87-100 (91-100)	650898204
	Ensifer arboris	84 (91)	2512599720
	Ensifer medicae	86 (91)	2513984635
	Ensifer terangae	79 (87)	2509112877
	Geminococcus roseus	38 (50)	2517135553
SMb20838	*Ensifer meliloti*	99 (99)	15965468
	Sinorhizobium fredii	90 (96)	YP_006396004.1
	Ensifer terengae	89 (95)	2509107333
	Rhizobium giardini	68 (81)	WP_018329408.1
	Paracoccus aminophilus	38 (50)	YP_008406331.1

SMb20079 gene is located in megaplasmid pSymB and encodes a protein of 1072 amino acids. This gene is the first of a putative operon including gene *SMb20080* encoding a formaldehyde dehydrogenase. *SMb20079* flanking regions are conserved in other sequenced stains of *Sinorhizobium meliloti*, *E.*

arboris and *E. medicae*. The search for identities at the amino acid level showed that the best homologues were within the *Ensifer*/*Sinorhizobium* genus, followed by a protein of *Geminococcus roseus* (Table IV). The search for conserved motifs identified 12 putative hemolysin-type calcium-binding repeats corresponding to a total of 24 RTX-toxin nonapeptide repeats (Fig. 3). I-tasser server built the 3D model for *SMb20079* show in Fig. 4B based on *Pseudomonas sp.* MIS 38 lipase, *Haemophilus influenza* Hap adhesin, an *E. coli* autotransporter hemoglobin protease and the extracellular lipase LipA from *Serratia marcescens*. According to gene ontology predictions, the molecular function with higher score (0.63 [0.1]) is serine-type endopeptidase and the biological process is proteolysis (0.63) and biological adhesion (0.48). The expression of *SMb20079* gene was decreased in the *tolC* mutant grown in GMS medium (Santos et al., 2010), being the only microarrays data set with reference to this gene.

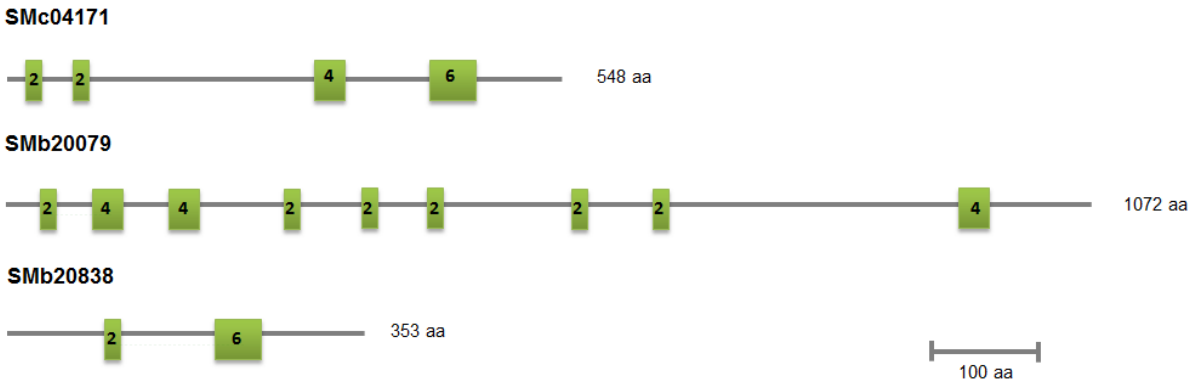


Figure 3 – Domains found in RTX proteins predicted to be secreted via TISS. Colored boxes represent clusters of repeated RTX nonapeptides with the number of repeats indicated. Bar represents 100 amino acid residues.

SMb20838 gene encodes a protein of 353 amino acids and is the first gene of a putative operon that includes *SMb20837* encoding an hypothetical protein. *SMb20838* is duplicated in *S. meliloti* 1021 genome with the second copy being located in the chromosome (*SMc00286*). The identity between the two proteins is 99%. Table IV shows the best homologues of SMb20838. As the two other proteins under study, SMc20838 has 4 putative hemolysin-type calcium-binding repeats corresponding to a total of 8 RTX-toxin nonapeptide repeats (Fig. 3). The 3D model structure for this protein is depicted in Fig. 4C

and it was based on the structures of LipA lipase from *S. marcescens*, a Ca^{2+}-binding dependent antifreeze protein from an Antarctic bacterium and *Pseudomonas sp*. MIS 38 lipase. The predicted molecular function is serine-type endopeptidase (0.42) and cytoskeletal protein binding (0.33). The predicted biological process is cell adhesion (0.42) and proteolysis (0.36). The differential expression of *SMb20838* has been detected in several microarrays data sets. For example, it was up-regulated in early nodules of *M. truncatula* (Ampe et al., 2003); when *nodD3* gene was overexpressed, in bacteroids (Barnett et al., 2004) and in the Δ*podJ1* and Δ*tolC* mutants (Santos et al., 2010; Fields et al., 2012). This data suggests a role of *SMb20838* both in symbiosis and free-living.

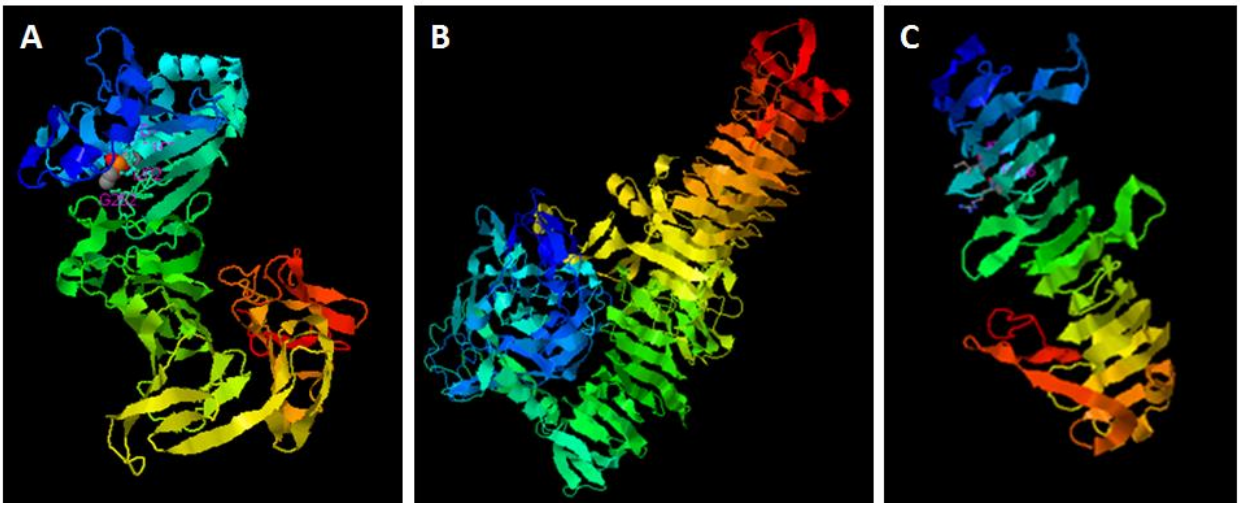


Figure 4 – 3D model prediction of SMc04171 (A), SMb20079 (B) and SMb20838 (C) proteins obtained at server I-Tasser.

4.2. Construction of *S. meliloti* deletion mutants

The general strategy for deletion mutant construction was to amplify approximately 2 Kb of each gene flanking region and clone them into a suicide vector for gene replacement. To facilitate mutant selection, a gentamicin resistance interposon cassette was inserted between the two flanking regions. Figure 5 shows the recombinant plasmid needed to delete *SMb20079* gene.

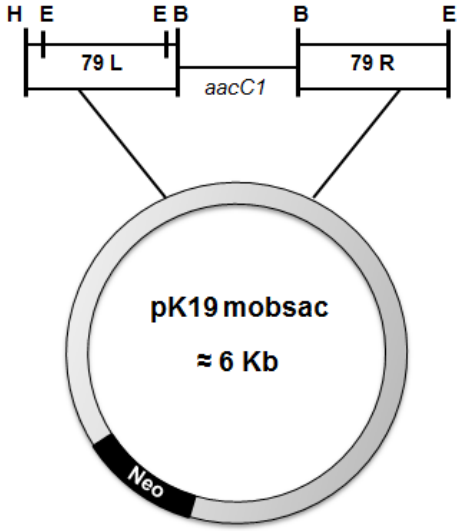


Figure 5 – Schematic representation of the plasmid harboring the *aacC1* cassette replacing *SMb20079* gene and the flanking regions needed for recombination and named pAS137-2. E, EcoRI; B, BamHI; H, HindIII. Vector and fragments are not at scale.

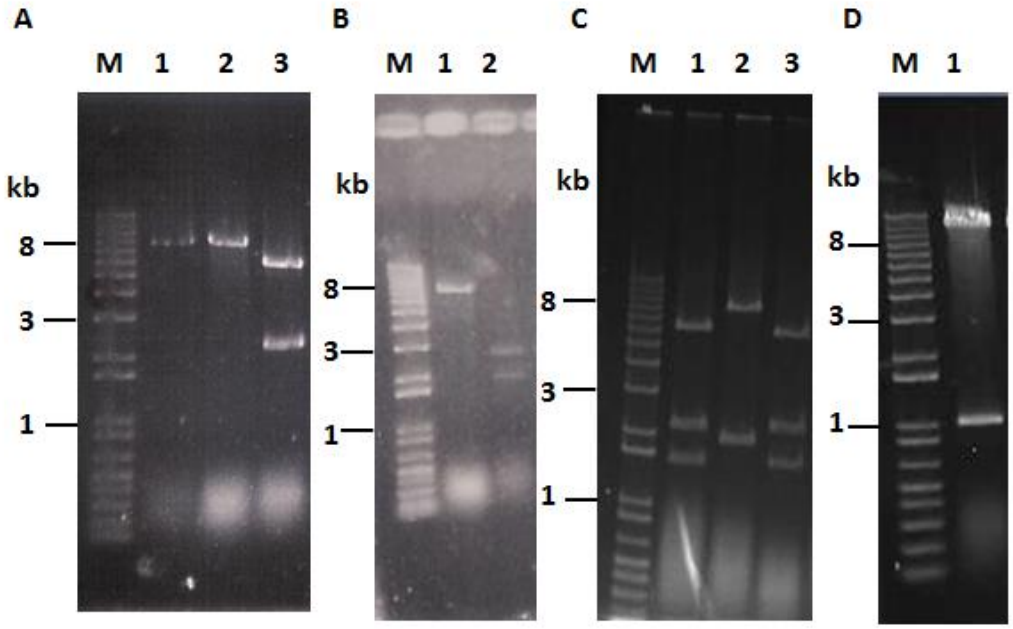


Figure 6 – Electrophoretic separation in 0.8% agarose gels of: (A) pK19mobsac + 79R digested with *EcoRI* (1), *BamHI* (2), and *EcoRI* + *BamHI* (3); (B) pK19mobsac + 79R digested with *BamHI* + *HindIII* (1) and pUC18 + 79L digested with *BamHI* + *HindIII* (2); (C) pAS137-1 digested with *EcoRI* + *BamHI* (1), *BamHI* + *HindIII* (2) and *EcoRI* + *HindIII* (3); (D) pAS137-2 digested with *BamHI*.

Following that strategy, the approximately 2 Kb right flanking region (79R) was cloned directly into pK19mobsac and the left flanking region (79L) into pUC18 vector (Fig. 6A, B). The 79L fragment was then transferred to pk19mobsac + 79R giving rise to pAS137-1 (Fig. 6C). Figure 6D shows the

insertion of the 1 Kb *aacC1* cassette into pAS137-1 and this final plasmid was named pAS137-2 (Fig. 5). This plasmid was introduced into *E. coli* S-17 and then mobilized to *S. meliloti* 1021. Selection was made in TY medium supplemented with streptomycin and gentamicin. Colonies were then tested for neomycin sensitivity, an indication of vector loss due to double cross over. Several candidate colonies were identified, but the confirmation of gene *SMb20079* loss by PCR amplification showed the presence of the gene in all of them. Most likely these colonies are single recombination mutants that lost resistance to neomycin or are wild-type cells with spontaneous resistance to gentamicin. As an alternative approach, selection can be done in medium containing gentamicin and saccharose since this vector has a gene that when expresses is lethal to cells. These are ongoing experiments. Following the same strategy, the flanking regions of genes *SMb20838* and *SMc04171* were cloned in pK18mobG or intermediate plasmids, but the final construct, similar to the one presented in Figure 5, was not obtained. Therefore, this data will not be presented here. As it was not possible to obtain the deletion mutants for the genes under study, the following symbiotic tests were done using the insertion mutants described by Santos (2008).

4.3. Competition between wild-type and mutant strains for nodule occupancy

The symbiotic properties of *S. meliloti* strains used in this study were analyzed in plant assays with aseptically grown *Medicago sativa* plants. In each assay, 5 replica plates were used for each strain tested, in a total of 15 plants. The nodulation was followed for 4 weeks and nodule number, as well as nodule occupancy determined. As previous data had shown a reduction in the nodule number elicited by the *SMc04171* insertion mutant (Fig. 7A), we investigated whether this mutant strain was less competitive than the wild-type strain in co-inoculation experiments. For that, three different ratios of wild-type vs. mutant were tested: 10:90, 50:50 and 90:10. As shown in Fig. 7B the average number of pink and total nodules per plant is higher in conditions where the wild-type strain is dominant. This result is in line with the previous symbiotic phenotype observed for *SMc04171* insertion mutant.

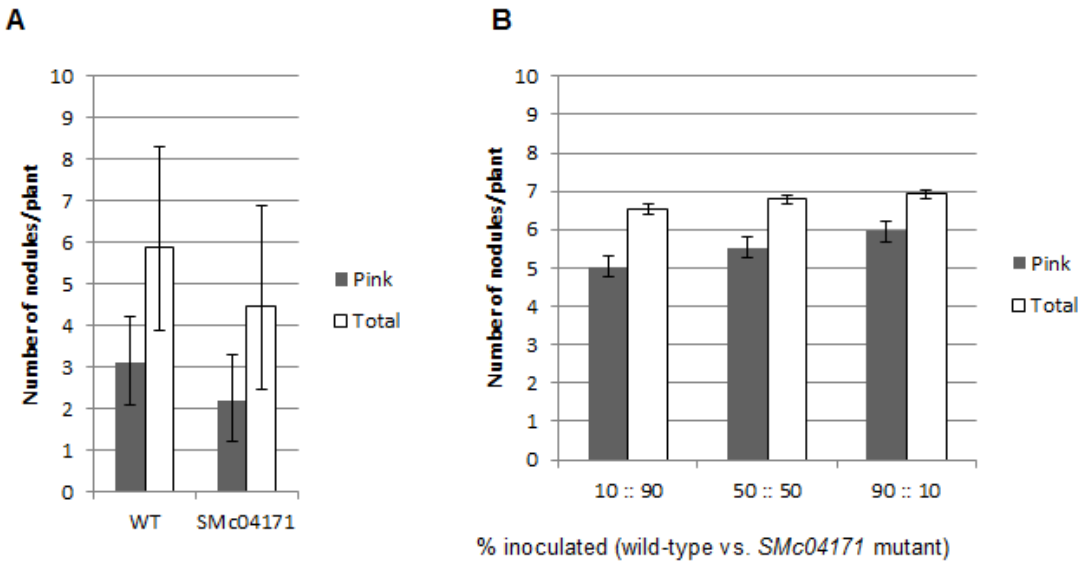


Figure 7 – Average number of pink (■) and total (□) nodules after 4 weeks post-inoculation of (A) *S. meliloti* 2011 and *SMc04171* insertion mutant (data obtained from Santos, 2008); (B) different ratios of co-inoculated bacteria. This data was collected from 15 plants in each experiment and the values are the mean ± SD.

To assess competition between the wild-type and mutant strains for nodule colonization, four weeks old nodules were crushed and CFU determined in TY medium supplemented with streptomycin and neomycin (for mutant evaluation) or only streptomycin (for total CFU evaluation). The proportion of bacteria obtained in the colonized nodule was then compared to the proportion of the initial inoculum. Results obtained for the co-inoculation of the wild-type strain and the *SMc04171* mutant showed that both strains are similarly competitive in nodule invasion (Fig. 8).

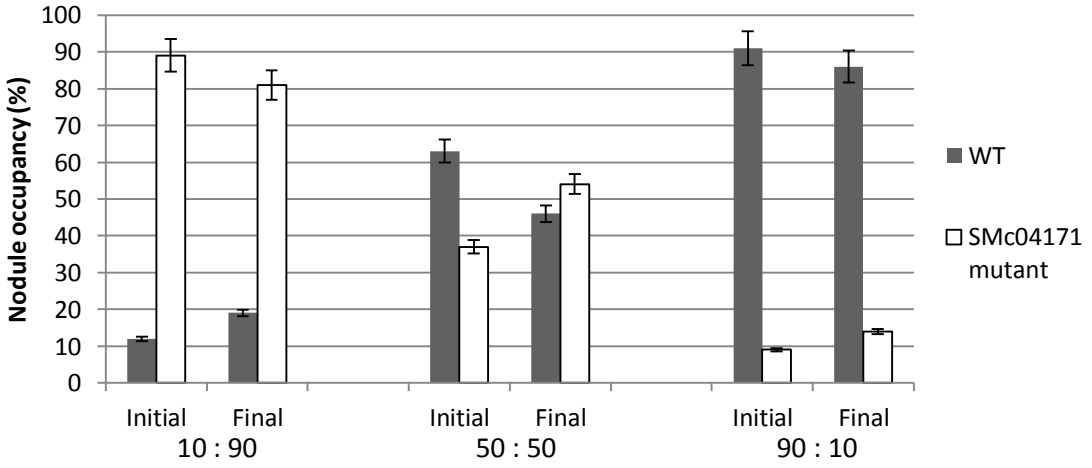


Figure 8 – Competition for 4 weeks old nodule occupancy between the wild-type Sm2011 strain and the *SMc04171* mutant. Data shown includes the initial proportion between the strains at the inoculum and the final proportion within the 4-weeks nodule. Values are the mean ± SD of at least three independent experiments.

The *SMb20079* insertion mutant did not induce a significantly different number of nodules than the wild-type strain (Fig. 9A), but no viable bacteria was obtained from crushed nodules (Santos, 2009). To assess whether this could be explained by reduced competitiveness, co-inoculation experiments were performed and nodules counted. The results shown in Fig. 9B are inconclusive since the amount of nodules obtained for the ratio 50:50 and 90:10 are to low.

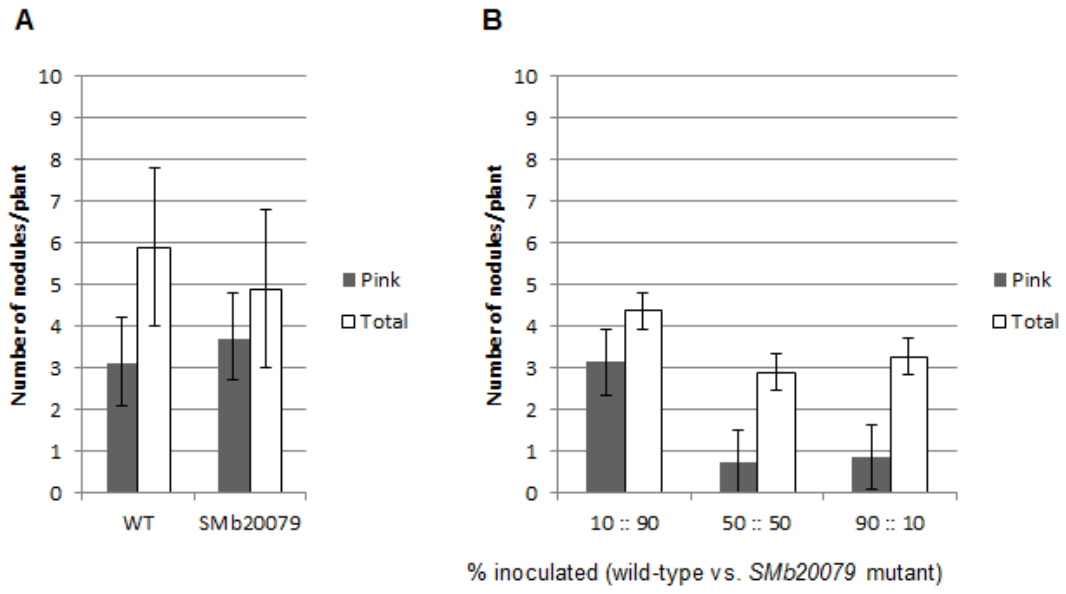


Figure 9 – Average number of pink (■) and total (□) nodules after 4 weeks post-inoculation of (A) *S. meliloti* 2011 and *SMb20079* insertion mutant (data obtained from Santos, 2008); (B) different ratios of co-inoculated bacteria. This data was collected from 15 plants in each experiment and the values are the mean ± SD.

Nodule occupancy by the wild-type and mutant bacteria showed the same proportions as in the initial inoculum for each of the tested ratios (Fig. 10). This result suggests the same competitiveness of the *SMb20079* mutant when compared to the wild-type strain.

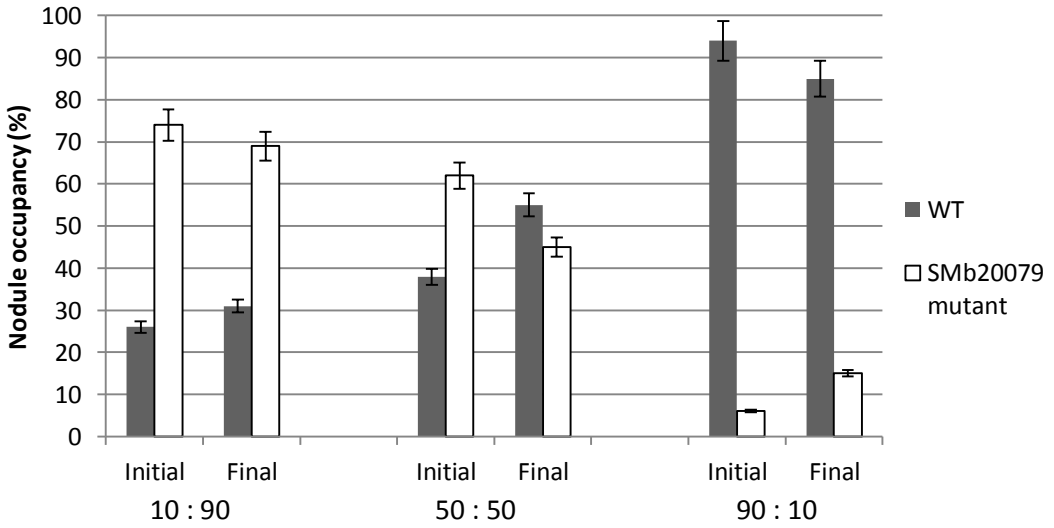


Figure 10 – Competition for 4 weeks old nodule occupancy between the wild-type Sm2011 strain and the *SMb20079* mutant. Data shown includes the initial proportion between the strains at the inoculum and the final proportion within the 4-weeks nodule. Values are the mean ± SD of at least three independent experiments.

The last tested insertion mutant for gene *SMb20838* induces a similar number of nodules as the wild-type strain, but no viable bacteria were detected inside nodules (Santos, 2008). As it can be seen in Fig. 11, the number of nodules induced by the three strains ratio was identical.

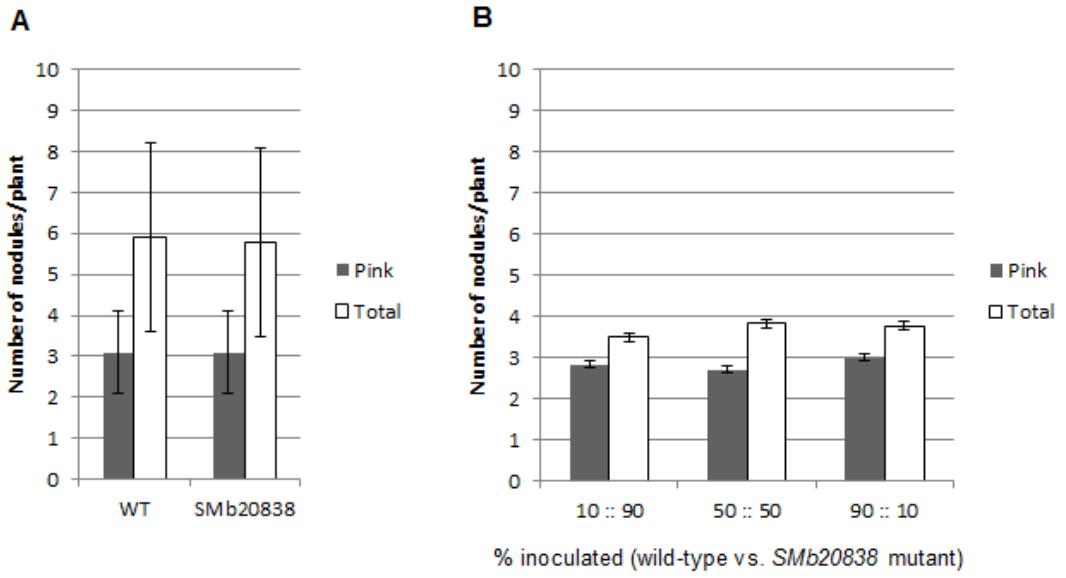


Figure 11 – Average number of pink (■) and total (□) nodules after 4 weeks post-inoculation of (A) *S. meliloti* 2011 and *SMb20838* insertion mutant (data obtained from Santos, 2008); (B) different ratios of co-inoculated bacteria. This data was collected from 15 plants in each experiment and the values are the mean ± SD.

As regards nodule colonization, in the proportion of wild-type vs. mutant (10:90) the *SMb20838* mutant seems to be less competitive than the wild-type strain, but at 50:50 ratio no differences were observed (Fig. 12).

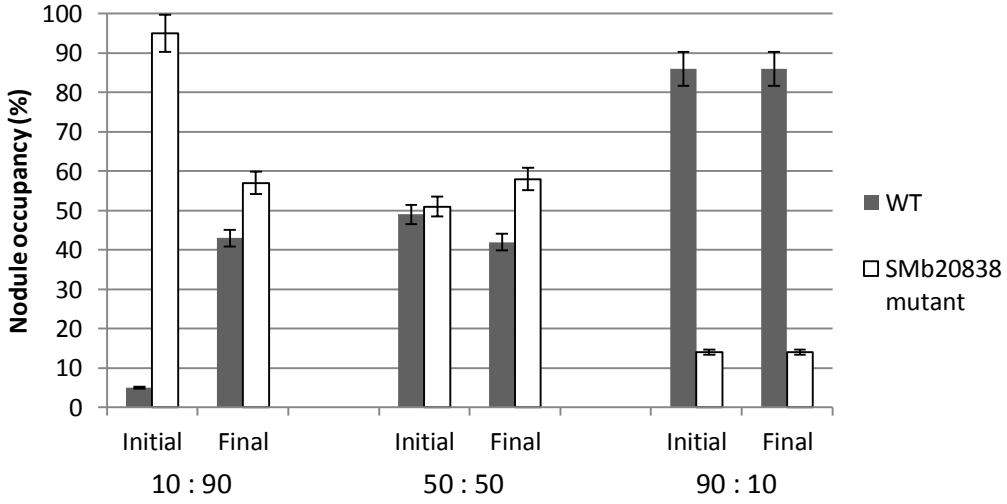


Figure 12 – Competition for 4 weeks old nodule occupancy between the wild-type Sm2011 strain and the *SMb20838* mutant. Data shown includes the initial proportion between the strains at the inoculum and the final proportion within the 4-weeks nodule. Values are the mean ± SD of at least three independent experiments.

4.4. Recombinant protein overexpression and purification

With the aim of characterizing the biochemical activity of the proteins codified by *SMc04171*, *SMb20079* and *SMb20838*, each of these genes was cloned in frame with the his-tag coding region present in pWH844 vector giving rise to a N-terminal his-tag fusion. The construct for expression of *SMc04171* gene from pWH844 was already available (Cosme, 2010). For the two missing genes, the strategy involved PCR amplification, digestion with appropriate enzymes and ligation to pWH844 vector. Fig. 13 shows a schematic representation of the obtained recombinant plasmids.

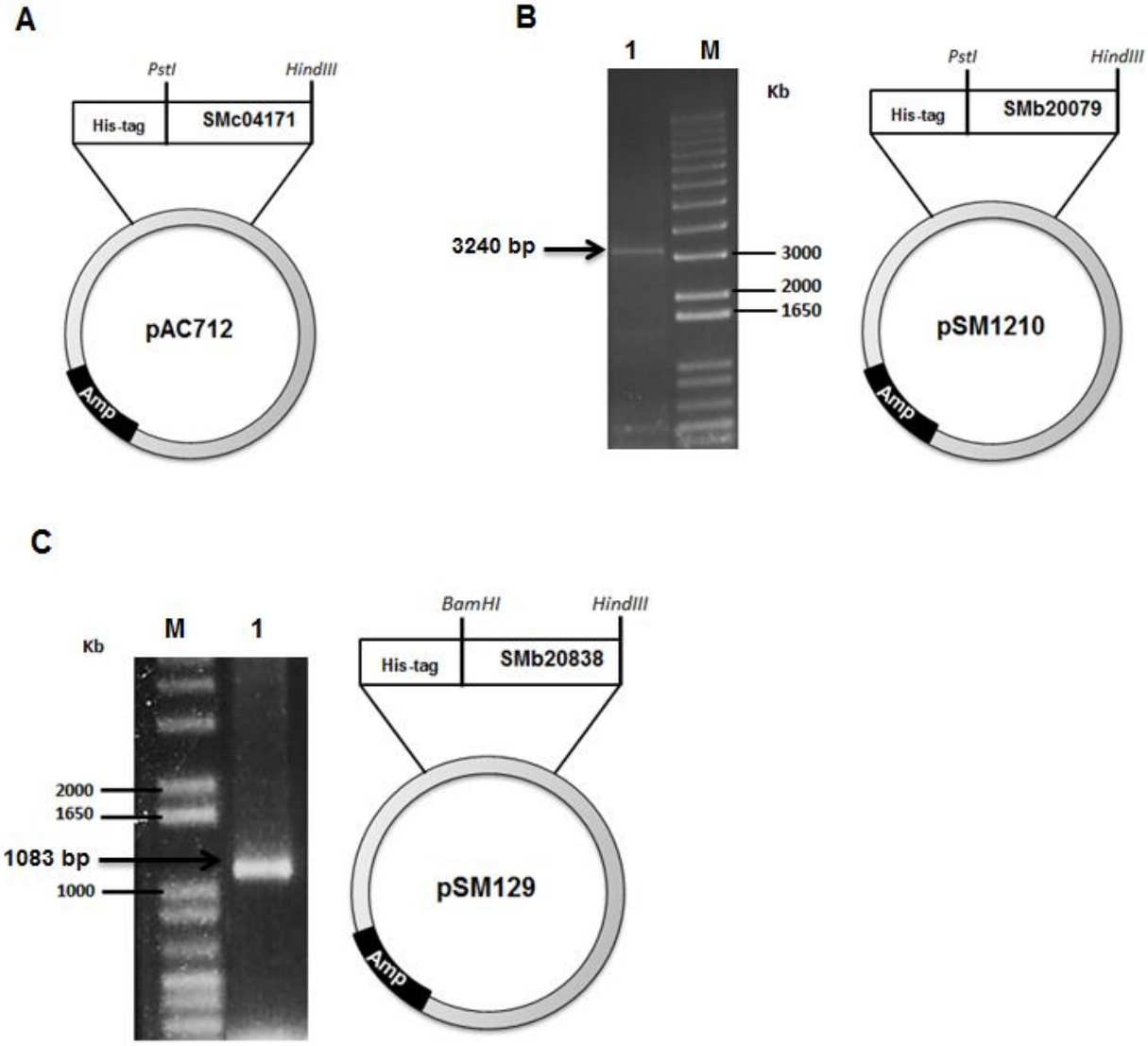


Figure 13 – Schematic representation of the recombinant plasmids generated by cloning the *SMc04171* gene (A), the *SMb20079* gene (B) and the *SMb20838* gene (C) into vector pWH844. The PCR amplification of *SMb20079* (B) and *SMb20838* (C) are also shown. Vector and fragments are not at scale.

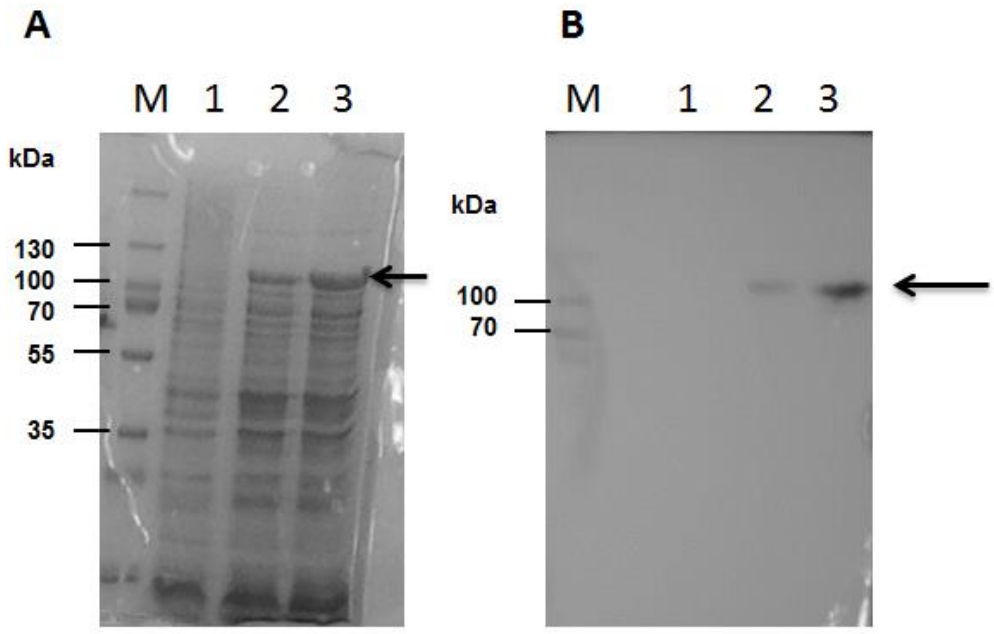


Figure 14 – Expression of *SMc04171* gene from vector pWH844. (A) SDS-PAGE of total protein extracts from whole cells before induction with IPTG (lane 1), one hour after IPTG induction (lane 2) or two hours after induction (lane 3). (B) Immunodetection of recombinant His-SMc04171 using a monoclonal antibody against the his-tag. Protein extracts analyzed were the same as in A. Lanes M represent the protein marker. Arrows represent the recombinant protein.

The next step in this work was to overexpress His-SMc04171 in *E. coli* and confirm its presence in crude extracts of whole cells by immunodetection of the His-tag by a commercial antibody. Using standard growth conditions for *E. coli* harbouring pAC712, namely, 37ºC at 250 rpm and induction with 1 mM IPTG, it was observed that cells are producing the recombinant protein (Fig. 14A) (indicated by an arrow). Although the expected size of His-SMc04171 is ≈ 56 kDa, the overexpressed protein band migrates slightly above 100 kDa. This type of behaviour is not unusual for this type of acidic proteins (SMc04171 pI: 3.68) as reported by others (Moreira et al., 2000; Economou et al., 1990). Data from the immunodetection analysis confirmed the nature of this band as corresponding to the recombinant His-SMc04171 (Fig. 14B). Purification of His-SMc04171 by affinity chromatography was carried out as shown in Fig. 15, but no fraction containing the recombinant protein was obtained.

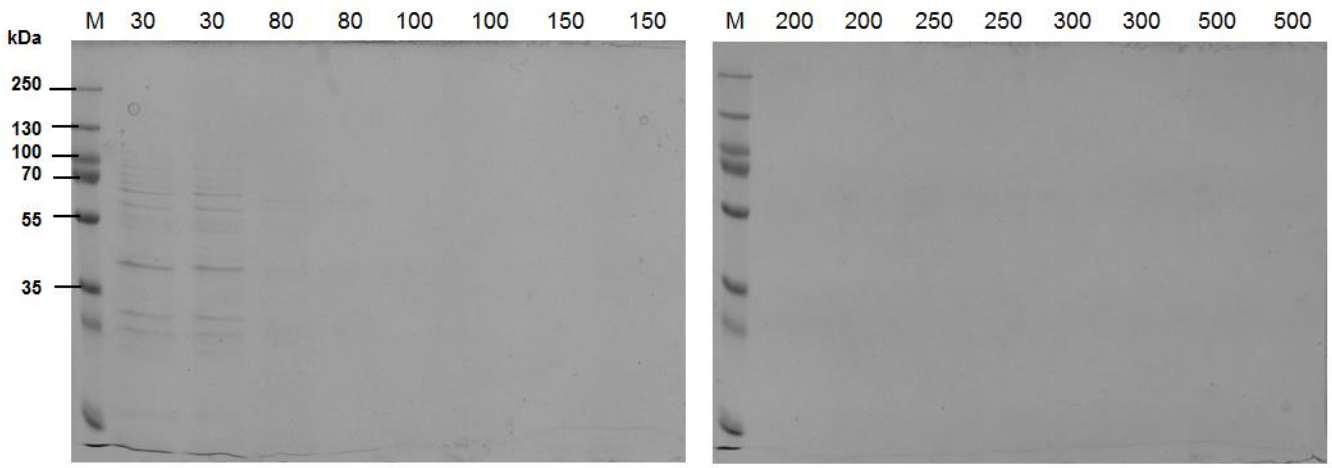


Figure 15 – Electrophoretic separation by SDS-PAGE followed by coomassie blue staining of protein samples eluted from the Ni^{2+}-NTA resin in the presence of imidazole. The numbers present on the top of each lane indicate the concentration of imidazole (mM) used to elute the proteins bound to the chromatographic resin. M indicates the protein marker.

Overexpression in *E. coli* of His-SMb20079, a protein with a predicted molecular mass of 106 kDa (pI: 3.78), was successful as shown in Fig. 16A. Once again the band migrates much higher than the expected. Immunodetection of the his-tag confirmed that this band corresponds to the recombinant protein (Fig. 16B).

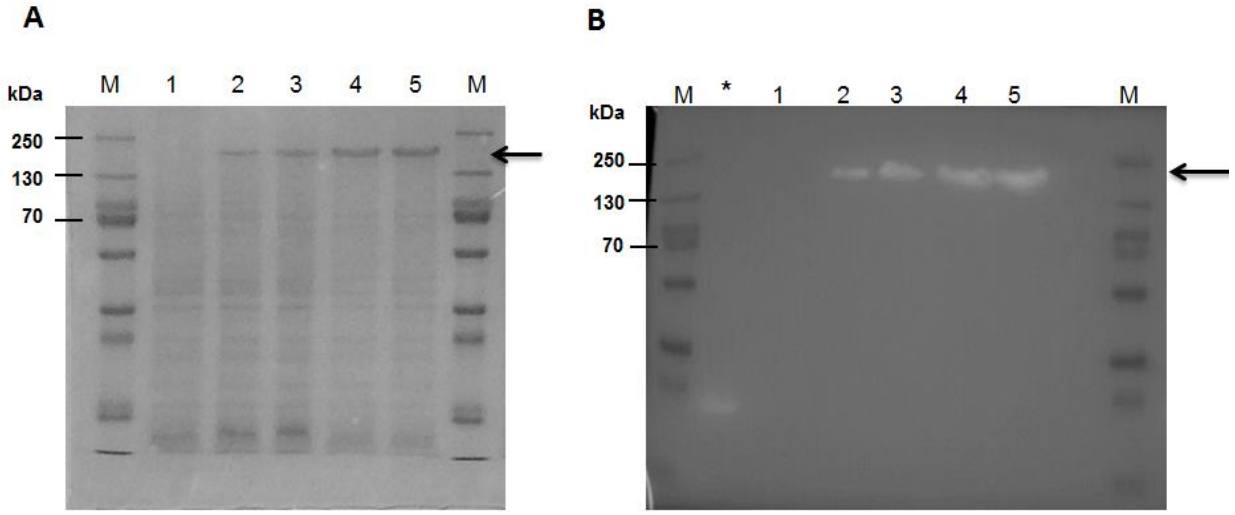


Figure 16 – Expression of *SMb20079* gene from vector pWH844. (A) SDS-PAGE of whole cell protein extracts before (lane 1) or after IPTG induction for 0.5 hours (lane 2), 1 hour (lane 3), 2 hours (lane 4) and 3 hours (lane 5). (B) Immunodetection of recombinant His-SMb20079 using a monoclonal antibody against the his-tag. Protein extracts analyzed were the same as in A, but additionally another protein known to have the his-tag was included as positive control (lane *). Lanes M represent the protein marker. Arrows indicate the recombinant protein.

Purification of His-SMb20079 by affinity chromatography showed that protein eluted in all concentrations of imidazole between 80 and 200 mM as shown in Fig. 17.

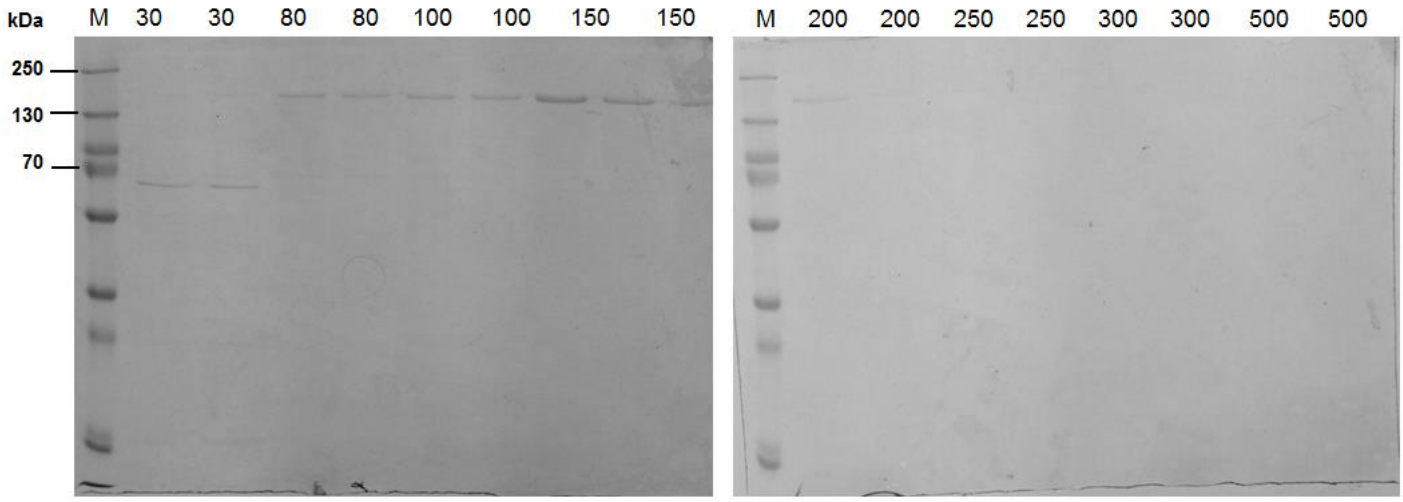


Figure 17 – Electrophoretic separation by SDS-PAGE followed by coomassie blue staining of protein samples eluted from the Ni^{2+}-NTA resin in the presence of imidazole. The numbers above each lane indicate the concentration of imidazole (mM) used to elute the proteins bound to the chromatographic resin. M indicates the protein marker.

The last protein to be tested for overexpression and purification was His-SMb20838 with an expected size of approximately 38 kDa (pI:3.97). Whole cell protein extracts separated by SDS-PAGE allow the visualization of a band running between 55 and 70 kDa, showing once again the abnormal migration of these acidic proteins (Fig. 18A). This protein band was confirmed to react with the anti-his-tag antibody (Fig. 18B). Regarding the purification of recombinant His-SMb20838 by affinity chromatography, it was not possible to obtain fractions with high degree of purification. As it is shown in Fig. 19, the His-SMb20838 eluted with low concentration of imidazole, suggesting that it binds weakly to the Ni^{2+}-NTA resin.

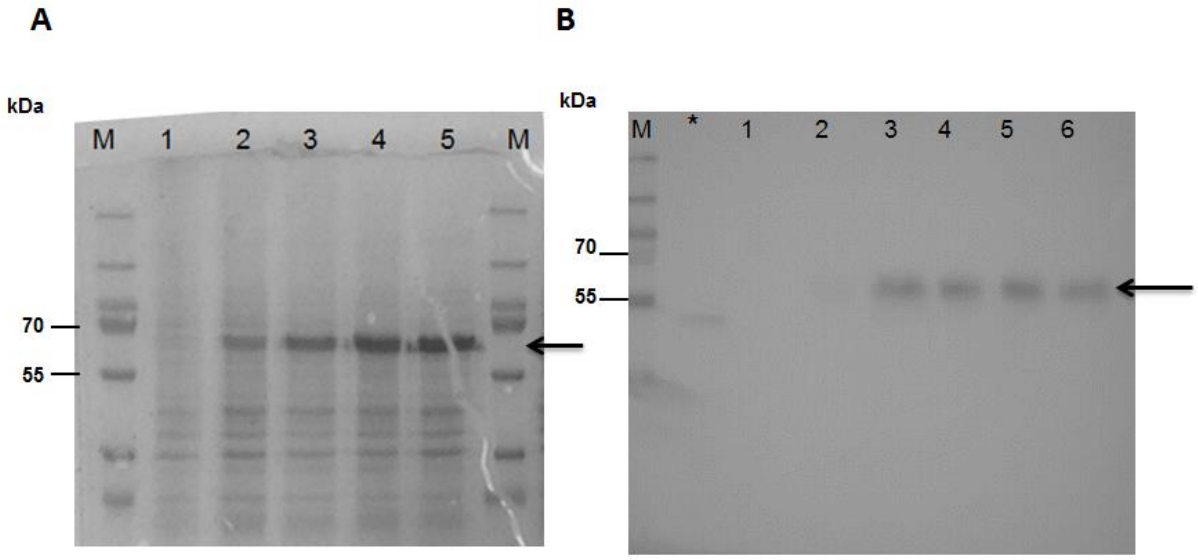


Figure 18 – Expression of *SMb20838* gene from vector pWH844. (A) SDS-PAGE of whole cell protein extracts before (lane 1) or after IPTG induction for 0.5 hours (lane 2), 1 hour (lane 3), 2 hours (lane 4) and 3 hours (lane 5). (B) Immunodetection of recombinant His-SMb20838 using a monoclonal antibody against the his-tag. Protein extracts analyzed were the same as in A, but additionally lane 6 has the 4 hours post-induction extract and lane * has another his-tag protein included as positive control. Lanes M represent the protein marker. Arrows indicate the recombinant protein.

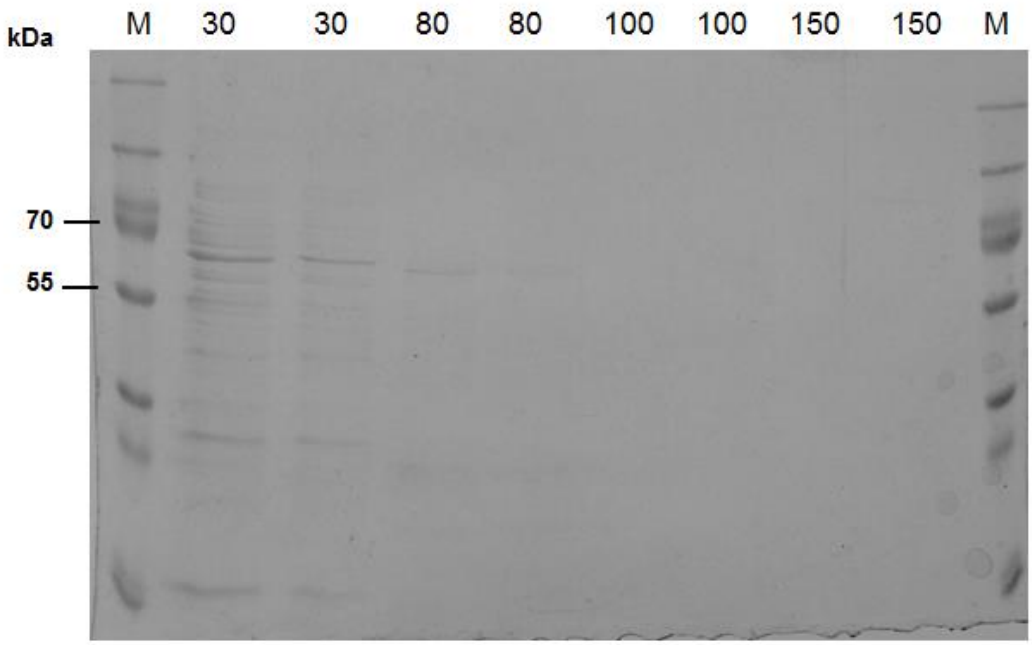


Figure 19 – Electrophoretic separation by SDS-PAGE followed by coomassie blue staining of protein samples eluted from the Ni^{2+}-NTA resin in the presence of imidazole. The numbers above each lane indicate the concentration of imidazole (mM) used to elute the proteins bound to the chromatographic resin. M indicates the protein marker.

5. Discussion

In this study we reported the characterization of genes *SMc04171*, *SMb20079* and *SMb20838* from *Sinorhizobium meliloti* which encode putative secreted calcium-binding proteins. These proteins belong to the RTX family which includes members from diverse bacterial genus (Linhartová et al., 2010). Although the function of RTX proteins can be as diverse as lipases, proteases, leukotoxins, and glycanases, they share structural features such as the amphipathic α-helix close to the C-terminus that is thought to be involved in secretion (Duong et al., 1992; Ghigo and Wandersman, 1992) and the Ca^{2+}-binding domains. These Ca^{2+}-binding regions may have a role in folding of the molecule after transmembrane translocation. In the absence of Ca^{2+} this structure is unstable and could facilitate membrane translocation of the polypeptide in an unfolded form. The presence of Ca^{2+} in the extracellular medium could induce the polypeptide to fold in the right tertiary structure (Baumann et al., 1993).

S. meliloti 1021 genome encodes one of the highest numbers of RTX proteins with a total of 17. Three of them were implicated in exopolysaccharide biosynthesis/modification and include ExsH and EglC endoglycanases acting on succinoglycan (Sharypova et al., 1999; York & Walker, 1998) and ExpE1 involved in EPSII biosynthesis (Moreira et al., 2000). The remaining RTX proteins have unknown function, although a few of them seem to have a role in symbiosis with leguminous plants (Santos, 2008). That is the case of *SMc04171* whose absence induces a lower number of nodules in *Medicago sativa*, and *SMb20079* and *SMb20838* mutants that induce nodules apparently with no viable bacteria (Santos, 2008).

To further understand the role of *SMc04171*, *SMb20079* and *SMb20838* gene products in symbiosis it would be helpful to have deletion mutants. Contrastingly to insertion mutants which may be unstable during a symbiotic experiment resulting in mutation loss, deletion mutants are stable (Cosme, 2009). Flanking regions of the genes under study were successfully obtained by PCR amplification and cloned in the final or intermediary vectors, but only from *SMb20079* gene deletion strategy was obtained the final construct. Using this

construct it were obtained several candidate deletion mutants, but genotype analysis indicated that *SMb20079* gene was still present. Further experiments have to be done, possibly increasing antibiotic concentration in the selection steps.

In another experiment, the available insertion mutants were used in co-inoculation with the wild-type strain to determinate if the symbiotic phenotype previously identified could be explained by reduced competitiveness. Overall, our experiments showed similar competition levels between mutants and wild-type, independent of the wild-type:mutant ratio tested. Previous results with *SMb20079* and *SMb20838* mutants indicated that nodules colonized by these mutants had no viable cells (Santos, 2008). In this work we had shown isolation of these mutant strains from co-inoculation plant nodules, in proportions similar to the initial inoculum. This result can be explained by the fact that the missing protein can be provided by the wild-type strain (as they are likely extracellular) and an eventual deficiency of the mutants during symbiosis is then complemented in these co-inoculation experiments. Specially from *SMb20079* and *SMb20838* mutants will be interesting to determine if the colonized nodules are fix^+ and whether these mutants are more sensitive under stress conditions such as presence of antimicrobial peptides, oxidative stress and others.

With the goal of investigating a possible in vitro catalytic activity for the proteins under study, a strategy to obtain recombinant his-tagged proteins was followed. *E. coli* was able to overexpress each of the three proteins under study, but only SMb20079 was successfully purified to homogeneity. For the other two proteins, further experiments have to be made and different purification methods applied. That can include different buffers or the addition of Ca^{2+} ions.

One feature common to the three recombinant proteins was their abnormal migration in SDS-PAGE when compared to the estimated molecular weight. In all cases, this value was almost double than the predicted one. This discrepancy in the predicted and observed molecular weight in SDS-PAGE was also reported for ExpE1 (Moreira et al., 1990), and NodO from *R. leguminosarum* (Economou et al., 1990). Size fractionation of the His-ExpE1

and NodO by gel filtration indicated that these proteins purify as dimers (Moreira et al., 2000; Economou et al., 1990).

Despite all the bioinformatics tools available that, based on homology to other characterized RTX proteins, helped us to determine a 3D-model for the proteins under study, the possible role of each protein remains unknown. Some hypothesis to follow in the future will include determining the activities as peptidases/proteases, polysaccharide modifying enzymes, or even involvement in cell adhesion to surfaces including host-cells interaction.

6. References

Ampe, F., Kiss, E., Sabourdy, F. & Batut, J. (2003). Transcriptome analysis of *Sinorhizobium meliloti* during symbiosis. *Genome Biology,* **4**, R15.

Andersen C., Hughes C. & Koronakis V. (2001) Protein export and drug efflux through bacterial channel–tunnels. *Current Opinion in Cell Biology,* **13**:412–416.

Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P., Barloy-Hubler F., Bowser L., Capela D.,Galibert F., Gouzy J. *et al.* (2001). Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. *Proceeding of the National Academy of Sciences USA*, **98(17)**:9883-9888.

Bauer, E., Kaspar, T., Fischer, H.-M. & Hennecke, H. (1998). Expression of the *fixR-nifA* operon in *Bradyrhizobium japonicum* depends on a new response regulator, RegR. *Journal of Bacteriology,* **180**, 3853-3863.

Becker A., Schmidt M., Jäger W., Pühler A. (1995). New gentamicin resistance and *lacZ* promoter probe cassettes suitable for insertion mutagenesis and generation of transcriptional fusions. *Gene,* **162**: 37-39.

Becker A., Barnett M.J., Capela D., Dondrup M., Kamp P.B., Krol E., Linke B., Ruberg S., Runte K., Schroeder B.K., *et al.* (2009) A portal for rhizobial genomes: RhizoGATE integrates a *S. meliloti* genome annotation update with postgenome data. *Journal of Biotechnology,* **140(1-2)**:45-50.

Beringer J.E. (1974). R factor transfer in *Rhizobium leguminosarum*. *Journal of General Microbiology*, **84(1)**:188-198.

Bode W., Grams F., Reinemer P., Gomis-Ruth F. X., Baumann U., McKay D.B. & Stocker W. (1996) The metzincin-superfamily of zinc-peptidases. *Advances in Experimental Medicine and Biology,* **389**:1–11.

Botstein, D., Lew, K.K., Jarvik, V.M. and Swanson, C.A., Jr. (1975) Role of antirepressor in the bipartite control of repression and immunity by bacteriophage P22. *Journal of Molecular Biology,* **91**: 439-462.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry,* **72**, 248-254.

Bullock W.C., Fernandez J.M., Short J.M. (1987). XL1-Blue: a high efficient plasmid transforming *reca Escherichia coli* strain with beta-galactosidase selection. *Biotechniques*, **5(4)**:376-379.

Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J., Boistard P., Becker A., Boutry M., Cadieu E. *et al.* (2001). Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti*. *Proceeding of the National Academy of Science USA*, **98(17)**:9877-9882.

Cordero C. L., Sozhamannan S., Satchell K. J. (2007) RTX toxin actin cross-linking activity in clinical and environmental isolates of *Vibrio cholerae*. *Journal of Clinical Microbiology,* **45**:2289–92.

Cosme A., (2009). "Functional analysis of the outer membrane protein TolC from *Sinorhizobium meliloti*: role in protein secretion,polysaccharide biosynthesis and symbiosis", PhD Thesis, IST, UTL, Lisbon.

Downie J. A. (2010). The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. *FEMS Microbiology Reviews*, **34(2)**:150-170.

Duong, F., Lazdunski, A., Cami, B. & Murgier, M. (1992). Sequence of a cluster of genes controlling synthesis and secretion of alkaline protease in *Pseudomonas aeruginosa*: relationship to other secretory pathways. *Gene,* **121**, 47-54.

Economou, A., Hamilton, W. D. O., Johnston, A. W. B. & Downie, J. A. (1990). The *Rhizobium* nodulation gene *nodO* encodes a Ca^{2+}-binding protein that is exported without N-terminal cleavage and is homologous to haemolysin and related proteins. *EMBO Journal,* **9**, 349-354.

Fetzer A. E., Werner A. S. & Hagstrom J. W. (1967) Pathologic features of pseudomonal pneumonia. *American Review Respiratory Diseases,* **96**:1121–1130.

Finan T.M., Weidner S., Chain P., Becker A., Wong K., Cowie A., Buhrmester J., Vorholter F.J., Golding B., Puhler A. (2001). The complete sequence of the *Sinorhizobium meliloti* pSymB megaplasmid. *Proceeding of the National Academy of Science USA*, **98(17)**:9889-9894.

Galibert F., Finan T.M., Long S.R., Puhler A., Abola P., Ampe F., Barloy-Hubler F., Barnett M.J., Becker A., Boistard P. *et al.* (2001). The composite genome of the legume symbiont *Sinorhizobium meliloti. Science*, **293(5530)**:668-672.

Gangola, P. & Rosen, B. P. (1987). Maintenance of intracellular calcium in *Escherichia coli*. *Jornal of Biological Chemistry* **262**, 12570-12574.

Ghigo, J. M. and Wandersman (1992). Cloning nucleotide sequence and characterization of the gene encoding the *Erwinia chrysanthemi* B374 metalloprotease secreted via a C-terminal secretion signal. *Molecular and General Genetics,* **236**, 135-144.

Gibson, K. E., Campbell, G. R., Lloret, J., and Walker, G. C. (2006). CbrA is a stationary-phase regulator of cell surfasse physiology and legume symbiosis in *Sinorhizobium meliloti*. *Journal of Bacteriology,* **188**:4508-4521.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology,* **166:** 557-580.

Horvat RT & Parmely MJ (1988) *Pseudomonas aeruginosa* alkaline protease degrades human gamma interferon and inhibits its bioactivity. *Infection and Immunity,* **56**: 2925–2932.

Jones, M. K., Kobayashi, H., Davies, B. W., Taga, E. & Walker G. C. (2007). How rhizobial symbionts invade plants: the *Sinorhizobium-Medicago* model. *Microbiology,* **1705 (Vol 5)**, 619-633.

Koronakis V., Sharff A., Koronakis E., Luisi B. & Hughes C. (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature,* **405**:914–919.

Letoffe, S., Delepelaire, P. & Wandersman, C. (1989). Characterization of a protein inhibitor of extracellular proteases produced by*Erwinia chrysanthemi*. *Molecular Microbiology,* **3**, 79-86.

Linhartová I., Bumba L., Mašín J., Basler M., Osička R., Kamanová J., Procházková K., Adkins I., Hejnová-Holubová J., Sadílková L., Morová J., Sebo P. (2010). RTX proteins: a highly diverse family secreted by a commom mechanism. *FEMS Microbiology Reviews,* **34**, 1076-1112.

Maunoury N., Redondo-Nieto M., Bourcy M., Van de Velde W., Alunni B., et al. (2010) Differentiation of symbiotic cells and endosymbionts in *Medicago truncatula* nodulation are coupled to two Transcriptome-switches. *PLoS ONE,* **5(3)**: e9519.

McCarthy C. N., Woods R. G. & Beacham I. R. (2004) Regulation of the *aprX-lipA* operon of *Pseudomonas fluorescens* B52:differential regulation of the proximal and distal genes, encoding protease and lipase, by ompR-envZ. *FEMS Microbiology Letters,* **241**: 243–248.

Meade, H. R., Long, S. R., Ruvkun, G. B., Brown, S. E., Ausubel, F. M. (1972). Physical and Genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn5 mutagenesis. *Journal of bacteriology,* **149**: 114-122.

Moreira, L. M., Becker, J. D., Puhler, A. & Becker, A. (2000). The Sinorhizobium meliloti ExpE1 protein secreted by a type I secretion system involving ExpD1 and ExpD2 is required for biosynthesis or secretion of the exopolysaccharide galactoglucan. *Microbiology,* **146 (Pt 9)**, 2237-2248.

Prell J., Bourdès A., Kumar S., Lodwig E., Hosie A., et al. (2010) Role of symbiotic auxotrophy in the *Rhizobium*-Legume symbioses. *PLoS ONE,* **5(11)**: e13933.

Rolfe B.G., Gresshoff P.M., Shine J., Vincent J.M. (1980). Interaction between a non-nodulating and an ineffective mutant of *Rhizobium trifolli* resulting in effective (Nitrogen-Fixing) nodulation. *Applied and Environmental Microbiology*, **39(2)**:449-452.

Rose, T., Sebo, P., Bellalou, J. & Ladant, D. (1995). Interaction of calcium with *Bordetella pertussis* adenylate cyclase toxin. Characterization of multiple calcium-binding sites and calcium-induced conformational changes. *Journal of Biological Chemistry,* **270**, 26370-26376.

Sambrook & Russell. (2001). Molecular Cloning: A Laboratory Manual. *Cold Spring Harbor Lab. Press* 3rd ed., Plainview NY.

Santos M.R. (2008). "Assessment of the role of proteins with Ca^{2+}- binding repeats in the symbiosis between *Sinorhizobium meliloti* and *Medicago sativa*", MSc Thesis, IST, UTL, Lisbon.

Santos M.R., Cosme A.M., Becker J.D., Medeiros J.M., Mata M.F., Moreira L.M. (2010). Absence of functional TolC protein causes increased stress response gene expression in *Sinorhizobium meliloti*. *BMCMicrobiology*, **10**:180.

Satchell K. J. (2007) MARTX, multifunctional autoprocessing repeats-in-toxin toxins. *Infection and Immunity,* **75**: 5079–5084.

Schäfer, A., Tauch, A., Jäger, W., Kalinowski, j., Thierbach, G. & Pühler, A. (1994). Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene,* **145**, 69-73.

Schirmer, K., Chan, A.G.J., Greenberg, B.M., Dixon, D.G., Bols, N.C., (1997). Methodology for demonstrating and measuring the photocytotoxicity of fluoranthene to fish cells in culture. *Toxicology in* Vitro, **11,** 107–119.

Sharypova, L. A., Yurgel, S. N., Keller, M., Simarov, B. V., Pühler, A. & Becker, A. (1999). The *eff*-482 locus of *Sinorhizobium meliloti* CXM1-105 that influences symbiotic effectiveness consists of three genes encoding an endoglycanase, a transcriptional regulator and an adenylate cyclase. *Molecular and General Genetics,* **261**, 1032-1044.

Simon R., Priefer U., Puhler, A. (1983). A broad-host-range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram-negative bacteria. *Biotechnology*, **1(9):**784-791.

Spaink, H. P. (2000). Root nodulation and infection factors produced by rhizobial bacteria. *Annual Review of Microbiology,* **54**, 257-288.

Spaink, H. P., Kondorosi, A., Hooykaa, P. J. J. (1998) The Rhizobiaceae: Molecular Biology of Model Plant-associated Bacteria., Kluwer, Dordrecht, The Netherlands

Sutton, J. M., Lea, E. J. A. & Downie, J. A. (1994). The nodulation signalling protein NodO from *Rhizobium leguminosarum* biovar *viciae* forms ion channels in membranes. *Proceeding of the National Academy of Sciences USA,* **91**, 9990-9994.

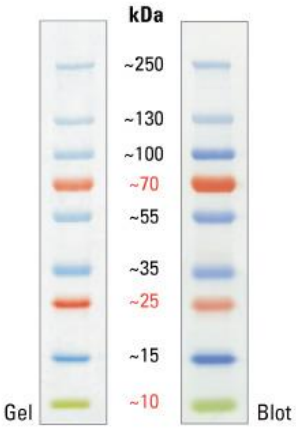
Vojtova J., Kamanova J. & Sebo P. (2006) *Bordetella* adenylate cyclase toxin: a swift saboteur of host defense. *Current Opinion in Microbiology,* **9**: 69–75.

Welch R. A. (2001) RTX toxin structure and function: a story of numerous anomalies and few analogies in toxin biology. *Current Topics in Microbiology and Immunology,* **257**: 85–111.

York, G. M. & Walker, G. C. (1998). The Rhizobium meliloti ExoK and ExsH glycanases specifically depolymerize nascent succinoglycan chains. *Proceeding of the National Academy of Sciences USA,* **95**, 4912-4917.

Zahran H.H. (1999). *Rhizobium*-Legume symbiosis and nitrogen fixation under severe conditions and in an arid climate. *Microbiology and Molecular Biology Reviews*, **63(4):**968-989.

7. Annex I



SDS-Page profile of the Thermo Scientific PageRuler Plus Prestained Protein Ladder

SMc04171

Compute pI/Mw

Theoretical pI/Mw (average) for the user-entered sequence:

```
        10         20         30         40         50         60
MAFFATQGMS DMTTVVGTAG NDTLSGTEEK DRIWGLAGAD EIDAGGGDDL VDGGAGNDRL

        70         80         90        100        110        120
TSRSGYDRLD GGEGDDLISL IGTGGAVTGG AGVDTLVVDL SNVSSDVRFS GEHGHGIIGY

       130        140        150        160        170        180
NTANWEHIFF NRIERLVLTT GSGNDRIFGA ATDDIISTGA GNDVVGPYGV DGDDGTSMLG

       190        200        210        220        230        240
DDAIDTGSGG DIINDTVGAN RIFAGDHNDI IITTLSSAVI DGGNGWDKLS IFDEGRTDGV

       250        260        270        280        290        300
TIDFGQGFAS TGTLISGIEV ASVDLGGGSD TLIAGNLSSL TAHMGDGDNY VAGGSGRDYV

       310        320        330        340        350        360
ASGSGDDALY GGAGDDILIS AGGNDVLAGG DGNDEIYDGG TSFGDGETYI DGGAGDDLIQ

       370        380        390        400        410        420
VVAPSGFIDG GGGSDTLRVV DPLPGTNFDA STGTLGTTLI FTNIERFELS GGSGNDSIRT

       430        440        450        460        470        480
LAGDDQLAGN DGNDRLDGGA GNDVLWGGAG DDVMTGGAGA DTFLWSSDTF SFAGVDRITD

       490        500        510        520        530        540
FDADGGDLLR FTGYAPDTTR IDSFADLVAA ATETEDGLYI AFNGSDTFGL LLDKVALADL


SADDVVFV
```

Theoretical pI/Mw: 3.68 / 55196.14

SMb20079

Compute pI/Mw

Theoretical pI/Mw (average) for the user-entered sequence:

```
        10         20         30         40         50         60
MAVINGTAGN NVLIGTDLDD VISGFGGDDF IQGLGGADVI NGGAGVDTVD YSEKTTSVVV

        70         80         90        100        110        120
TLTGANAATV FVNGVADDTL SNVENVYGGS GNDTITGDAQ NNLFRGGGGN DVLDGGAGND

       130        140        150        160        170        180
TADYADKTTS VVVTLMGATP ATVFVNGIAE DTISNFENVY GGSGNDILTG DDRSNILRGE

       190        200        210        220        230        240
AGNDILNGGA DDDLLSGGAG NDTADGGTGI DTFDLREKTS SVVVQLSGAN AATVFVGGVA

       250        260        270        280        290        300
EDTIRNVENI VGGTADDTLS GDAAANKLSG ARGNDWLKGG GGADTLDGGE DSDTADYSDK

       310        320        330        340        350        360
AAAIAVALNG GNPVTVTVGG IAEDLIAKIE NIVGGSGGDT IIGDAAANAF RGGLGADVLD

       370        380        390        400        410        420
GGGGSDTADF SDKVQSVVLA LNGAVDAIAA VGGTAEDTVR NIENITGGSG NDQFTGDAAA

       430        440        450        460        470        480
NTFRGGLGAD VLDGGDGSDT ADYGDKTVSV VVTLAGANPT TAFVDGMAED SLRNIENIIG

       490        500        510        520        530        540
GSGNDVLAGD GFQNVLDGGA GTDMADYSAS AKGIAVMLNG ANDAKVIVGS AAEDTLRNIE

       550        560        570        580        590        600
NVTGSAFADV ITGDAQANIL LGGSGSDILK GDGGQDVIDG GAGTDTADFS EKTAAVVLAL

       610        620        630        640        650        660
AGAANAIATV GGLAEDTVRN IESIFGGIGA DVLTGDGNSN TIRGGAGADS LDGGAGVDTV

       670        680        690        700        710        720
DYRDKTKSVA VTLDGATPVT VKVGGVIEDT IRNFENISGG SAGDMLTGDG LANVLVGNDG

       730        740        750        760        770        780
ADTLRGGLGK DVLDGGNGVD TADYLEKTDA ISVTLNGTAS AAVLVGGTAE DTIRGVENIL

       790        800        810        820        830        840
SGSGADTLVG DTASNMFRGA LGADFIDGGA GVDTADYREK TGSVEVALSG ASDSFVFVGG

       850        860        870        880        890        900
VVEDTIRNIE NVFGGKGNDT LTGDGLANTL NGNDGKDLLT GGGGADILDG GAASDTASYR

       910        920        930        940        950        960
DKTASVSVTL DGATYTTVTV GGAAEDTIRN VENIWGGTGN DSLSGDANAN LLSGGGGSDI

       970        980        990       1000       1010       1020
LFGGAGADIF QFDFALGSTN VDTVLDFTAG DRLFLSKSIF TSLSGGTLAA TQFYAAADAT

      1030       1040       1050       1060       1070
AAQNVNQKII YDTTSGALYY DADGSLSGHT AVQFAVFSTH PGLTAGDFVL VV
```

Theoretical pI/Mw: 3.78 / 105264.67

SMb20838

Compute pI/Mw

Theoretical pI/Mw (average) for the user-entered sequence:

```
        10         20         30         40         50         60
MVLPKGRAMA TRIYGTNYAD VIKQNGYIAV EIYAYDGDDD IYLNRTDSYG GYNFVDAGYG

        70         80         90        100        110        120
NDLVVNSYEG GNDIYLGGGN DIYVADIRAR DASSYDIVYG GSGNDRFEIE GYASDYYGES

       130        140        150        160        170        180
GNDTFFSVGF NNYFNGGTGT DTISYQFQDD WSAERGKGVT IDLGYNYATT GSGRREDLIS

       190        200        210        220        230        240
IENATGTNYG HDDITGSAVA NTLRGLGGHD IIEGLGGNDV LDGGSGDDDL YGGSGADILR

       250        260        270        280        290        300
GGTGFDYLVG GTGTDSFDFN SISESAVGSR RDVITDFHRS EFDVIDLSTI DADTTWSGNQ

       310        320        330        340        350
SFTYIGGNAF SGEAGELNFR SGIISGDVNG DGYADFQVRV NGVTSLRVDD FFL
```

Theoretical pI/Mw: 3.97 / 37728.23