Aristides Lopes de Andrade Mendes

Graduated in Cellular and Molecular Biology

Characterization of MsmX: A multitask ATPase from *Bacillus subtilis*

Dissertation presented to obtain the
Master degree in Biotechnology

Supervisor: Isabel M. G. de Sá-Nogueira, Associate Professor w. Habilitation (Agregação), FCT/UNL

Jury Members:

President: Prof. Doutora Susana Filipe Barreiros
Main Examiner: Prof. Doutor Adriano O. Henriques
Supervisor: Prof. Doutora Isabel M. G. de Sá-Nogueira

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

September 2014

Characterization of MsmX: A multitask ATPase

Acknowledgments

"No matter what people tell you,
Words and ideas can change the world"
Robin Williams

First of all I want to thank my Professor and Supervisor, Isabel de Sá-Nogueira, for all the support and guidance throughout this year. Thank you for all the effort in the moments I needed help and for pushing me forward in the pursuit of a scientific career.

To Centro de Recursos Microbiológicos (CREM) da Faculdade de Ciências e Tecnologias da Universidade Nova de Lisboa for providing good work conditions and a great environment to develop science.

To Fundação para a Ciência e Tecnologia for partially supporting this work through grant PEst-OE/BIA/UI0457/2013 to CREM.

To my lab mates Isabel, Mário, Lia and Viviana for the advices and teachings that helped me become a better scientist. Thank you not only for all your patience to my constant questions and doubts that interrupted your work and thoughts, but mainly for the support and encouragement in the bad moments throughout the year. I could not ask for better colleagues!

My thanks to our neighbours from lab 333, especially to Raquel and Bárbara, for the good environment and jokes at work and lunch time.

A especial thank to my biology teacher Graça Silva that was always able to motivate me and had a determinant role in my future.

Thanks Michael for the company and friendship in this adventure of five years we have pursued since we entered FCT; Alicia for your constant good mood and contagious laugh; Luis and Daniel for all the support; Diana for the special person you are; Tânia for all the times you listened to me; Rita for the good you do to me; Pedreira, Syd, Nobre, Pedro, Rute, Francisca and many more that are always in my thoughts! To all my amazing friends that have always supported me throughout my life, in the good and the bad times, and that make it all worth. Thanks for all the moments that you made me laugh and smile!

Most important, I wish to thank my family! Not only because without them none of this would be possible, but mostly because the person I am today, all my principles and beliefs… my understanding of life… comes from them. A especial thank to my Dad for all he had to pass through to support me and my Grandfather that always believed I would be capable of achieving great things. Today you would be very proud of me, I will never forget you!

Abstract

Transport across biological membranes is fundamental for cell survival and is mainly accomplished by specialized membrane proteins known as transporters. The ABC-type transporters constitute one of the largest and most diverse transporter superfamilies and are found in all three domains of life (Archaea, Bacteria and Eukarya). These permeases utilize the binding and hydrolysis of ATP to power the directional transport of a wide variety of substrates across membranes, ranging from ions to macromolecules. The ABC transporters have a modular architecture comprising two transmembrane domains (TMDs) that form the translocation pore and two nucleotide-binding domains (NBDs) that hydrolyse the necessary ATP for the translocation process.

Until the last few years, NBDs were thought to be exclusive of each transporter complex, however it was discovered that several bacterial species ABC transporters share the same energy-generating component. In the Gram-positive model organism *Bacillus subtilis* ten ABC transporters are predicted to be involved in sugar uptake and eight of these systems do not display a gene encoding a putative NBD protein in their vicinity. Recent studies, demonstrated that MsmX interacts with several of these eight distinct ABC sugar importers. Thus, unlike other NBDs, MsmX was shown to be multitask serving as energy-generating component to several sugar importers. Other examples of multitask ATPases are also found in *Streptococcus* and *Streptomyces* species and are involved in the uptake of carbohydrates. So a better understanding of these multitask ATPases may have an important impact in therapy development for pathogenic bacteria since carbohydrate utilization is essential for microorganisms survival.

In this work we demonstrate that *B. subtilis* can be used as a model to study multitask ATPases from other Gram-positive pathogenic species. A genetic system was developed to test *in vivo* the functionality of MsmX homologs from pathogenic species: *Bacillus thuringiensis*, *Staphylococcus aureus* and *Streptococcus pneumoniae*. The results show that the proteins are capable to play the role of MsmX in the cell, although with a different degree of efficiency. Functional analysis studies of MsmX indicate that modifications in the N- and C-terminal amino acid sequence do not affect its function. Moreover, by mutagenesis we show both *in vitro* an *in vivo* that the Lys 43 of MsmX, a conserved amino acid of NBDs, is essential for ATP hydrolysis. Additionally, the MsmX production and purification protocol was improved for further crystallography studies and biochemical characterization.

Keywords:

Bacillus subtilis, ABC Transporters, Multitask ATPases, MsmX, Sugars, Pathogenic bacteria.

Resumo

O transporte através de membranas biológicas é fundamental para a sobrevivência das células e é realizado, principalmente, por proteínas especializadas da membrana designadas como transportadores. Os transportadores do tipo ABC constituem uma das maiores e mais diversificadas superfamílias de transportadores existindo em todos os três domínios dos seres vivos (Archaea, Bacteria e Eukarya). Estas permeases utilizam a ligação e hidrólise do ATP como força motriz para o transporte de uma ampla variedade de substratos através da membrana, desde iões a macromoléculas. Os transportadores ABC têm uma arquitetura modular que compreende dois domínios transmembranares (TMDs), que formam o poro de translocação, e dois domínios de ligação a nucleótidos (NBDs), que hidrolisam o ATP necessário para o processo de translocação.

Até recentemente, pensava-se que os NBDs eram exclusivos de cada complexo transportador, no entanto descobriu-se que vários transportadores ABC de procariontes partilham o mesmo componente de produção de energia. Na espécie *Bacillus subtilis*, organismo modelo de bactérias Gram-positivas, existem dez transportadores ABC que se prevêem estar envolvidos no transporte de açúcares, sendo que oito desses sistemas não exibem um gene codificante para uma NBD putativa na sua vizinhança. Estudos recentes demonstraram que o MsmX interage com vários desses oito importadores. Assim, ao contrário de outros NBDs, o MsmX mostrou ser multitarefa servindo como "motor" para vários importadores de açúcares. Outros exemplos de ATPases multitarefa são também encontrados em espécies de *Streptococcus* e *Streptomyces*, estando envolvidos no transporte de açúcares. Assim sendo, uma melhor compreensão destas ATPases multitarefa pode ter um impacto importante no desenvolvimento de tratamentos para infecções bacterianas, dado que a utilização de hidratos de carbono é essencial para a sobrevivência dos microrganismos.

Neste trabalho demonstramos que *B. subtilis* pode ser usado como modelo para o estudo de ATPases multitarefa de outras espécies Gram-positivas patogénicas. De forma a testar *in vivo* a funcionalidade dos homólogos do MsmX provenientes de espécies patogénicas (*Bacillus thuringiensis*, *Staphylococcus aureus* e *Streptococcus pneumoniae*) foi desenvolvido um sistema genético. Os resultados mostram que as proteínas são capazes de desempenhar o papel do MsmX na célula, embora com diferentes graus de eficácia. Estudos funcionais do MsmX indicam que modificações na sequência de aminoácidos do N- e C-terminal não afectam a sua função. Além disso, mostramos através de mutagénese, tanto *in vitro* e *in vivo*, que a Lys 43 do MsmX, um aminoácido conservado de NBDs, é essencial para a hidrólise de ATP. Além disso, o protocolo de produção e purificação do MsmX foi melhorado de modo a realizar estudos cristalográficos e caracterizar bioquimicamente a proteína.

Palavras-chave:

Bacillus subtilis, Transportadores ABC, ATPases Multitarefa, MsmX, Açúcares, Bactérias Patogénicas.

Contents

Figures Index

Tables Index

Abbreviations, Symbols and Notations

aa – Amino acid

ABC – ATP-binding cassette

ADP – Adenosine diphosphate

Amp – Ampicillin

ATP – Adenosine triphosphate

A3 – 1,5-α-L-Arabinotriose

BLAST – Basic Local Alignment Search Tool

bp – Base pairs

cat – chloramphenicol acetyltransferase gene

Cm – Chloramphenicol

CRD – C-terminal regulatory domain

CSK – C medium supplemented with potassium succinate

CUT1 – Carbohydrate Uptake Transporter-1

DMSO – Dimethyl Sulfoxide

dNTP - Deoxynucleoside triphosphate

DNA – Deoxyribonucleic acid

ECF – Energy coupling factor

EDTA - Ethylenediaminetetraacetic acid

IPTG – Isopropyl-β-D-galactopyranoside

kDa - KiloDalton

Km - Kanamycin

LB – Luria-Bertani medium

mRNA – messenger ribonucleic acid

NBD – Nucleotide-binding domain

OD – Optical Density

ORF – Open Reading Frame

PBS – Phosphate buffer saline

PCR – Polymerase Chain Reaction

PMSF - phenylmethylsulphonyl fluoride

PTS – Phosphotransferase system

RBS – ribosome binding site

RNA - Ribonucleic acid

rpm – Revolutions per minute

SBP – Solute-binding domain

SDS-PAGE – Sodium dodecyl sulphate polyacrilamide gel electrophoresis

Spc - Spectinomycin

Tc – tetracycline

TE – Tris-EDTA

TMD – Transmembrane domain

Tris - Tris(hydroxymethyl)aminomethane

UV – Ultraviolet light

Amino acids – three and one letter code

Amino acid	Three letter code	One letter code
alanine	ala	A
arginine	arg	R
asparagine	asn	N
aspartic acid	asp	D
cysteine	cys	C
glutamic acid	glu	E
glutamine	gln	Q
glycine	gly	G
histidine	his	H
isoleucine	ile	I
leucine	leu	L
lysine	lys	K
methionine	met	M
phenylalanine	phe	F
proline	pro	P
serine	ser	S
threonine	thr	T
tryptophan	try	W
tyrosine	tyr	Y
valine	val	V

Bases - one letter code

Base	Letter
Adenine	A
Citosyne	C
Guanine	G
Thymine	T

Chapter 1

General Introduction

1. General Introduction

1.1. Membrane Transport Systems

Living cells are separated from the environment by lipid membranes, whereas other subcellular membranes play essential roles in the architecture of eukaryotic cells (Seyffer and Tampé, 2014). Transport across biological membranes is fundamental for cell survival, which is accomplished, in the majority of the cases, by specialized membrane proteins that are known as transporters. These proteins are responsible for the regulated and selective passage of small molecules, ions and even some macromolecules into the cell or in the opposite direction (Higgins, 1992; Rees *et al.*, 2009; Eitinger *et al.*, 2010).

The importance of membrane transport to the cell is exemplified by the fact that ~10% of the *Escherichia coli* genome has been classified as participating in transport processes and, overall, more than 550 different types of transporters have been identified (Rees *et al.*, 2009). *Bacillus subtilis* possesses a similar percentage of transporters in the genome with minor differences. While *E. coli* has more sugar transporters, *B. subtilis* encodes more drug/hydrophobic exporters (Saier *et al.*, 2002).

Accordingly to the way solute transport is energized, membrane transporters are classified in three major groups: primary active transporters, secondary transporters and group translocators. Primary active transporters comprise very diverse protein families that use chemical, electrical or solar energy sources to transport substrates across the membrane against a concentration gradient. On the other hand secondary transporters often accumulate substrates in – or deplete them from – cells by using the energy stored in ion gradients, frequently Na^+ or H^+, to drive transport. This includes uniporters, antiporters, and symporters. Group translocators chemically modify the substrate during the transport reaction, as for example the phosphotransferase system (PTS) which transfers a phosphate group to the substrate once it enters the cell (Davidson *et al.*, 2008; Jaehme and Slotboom, 2014).

This dissertation will focus in one of the primary active transporters families: the ATP-binding cassette (ABC) transporters.

1.2. ABC-type Transport Systems

The ABC transporters constitute one of the largest and most diverse transporter superfamilies (seven distinct subfamilies) and are found in all three domains of life (Archaea,

Bacteria and Eukarya) (Yoshida *et al.*, 1996; Rees *et al.*, 2009; Seyffer and Tampé, 2014). These transporters use the binding and hydrolysis of ATP (phosphate bond between the γ- and the β-phosphate) to power the directional transport of a wide variety of substrates across membranes, ranging from ions to macromolecules (Rees *et al.*, 2009; Seyffer and Tampé, 2014; ter Beek *et al.*, 2014).

The designation ABC transporters acknowledges a highly conserved ATP-binding cassette, which is the most characteristic feature of this superfamily (Higgins, 1992). These systems are widespread among living organisms and have been found in all species from the microbe to man, with a high conservation of the primary sequence of the ATP-hydrolyzing domain and in the modular architecture (Higgins, 2001; Davidson *et al.*, 2008).

The first systems discovered and characterized in detail were the high-affinity histidine (HisJQMP) and maltose (MalEFGK) uptake systems of *Salmonella enterica serovar typhimurium* and *E. coli* in the 1970s (Davidson *et al.*, 2008). Since then, a variety of systems have been identified both in prokaryotes and eukaryotes. In *E. coli* and *B. subtilis* ~80 and 59 distinct systems were reported, respectively, while ~50 systems (five subfamilies) exist in humans. Plants genome, like *Arabidopsis thaliana*, has 4120 ABC systems. However, in relation to genome size, the highest number of ABC systems exists in bacteria (Rees *et al.*, 2009; Eitinger *et al.*, 2010; Seyffer and Tampé, 2014).

In prokaryotes, these transporters are present in the plasma membrane with the ATP being hydrolyzed in the cytoplasmic side and transport a wide variety of substrates: sugars, amino acids, opines, peptides, phosphates, sulphates, vitamins, metallic cations, molybdenum and organo-iron complexes. Many of these transporters are central to antibiotic and antifungal resistance. In eukaryotes, the ABC transporters are found in organellar membranes with the hydrolysis taking place in the cytosolic side. The transporters from mitochondria and chloroplasts are an exception, because the domains responsible for the ATP hydrolysis are present in the matrix or stroma side. The ABC transporters found in humans handle antibiotics, toxins, vitamins, drugs, metals, peptides, ions, lipids, bile acids, polycarbonates and sterols. Dysfunctions in these systems are associated with genetic diseases such as cystic fibrosis, Tangier disease, obstetric cholestasis, and drug resistance of cancers. Here we will designate *cis*-side as the side where the ATP is bound and hydrolysed and the opposite side termed as *trans*-side (Quentin *et al.*, 1999; Higgins, 2001; Seyffer and Tampé, 2014; ter Beek *et al.*, 2014).

The ABC-type transporters are importers, which move substrates from the *trans*-side to the *cis*-side, or exporters that transport the molecules from the *cis*-side to the *trans*-side. (ter Beek *et al.*, 2014)**.** Despite this distinction between importers and exporters, they all share a common structural organization. They have a modular architecture comprising two transmembrane

domains (TMDs) that form the translocation pore and two nucleotide-binding domains (NBDs) that hydrolyze ATP (Eitinger *et al.*, 2010).

The TMDs are highly hydrophobic and each one, normally, contains six membrane-spanning segments. These domains create the pathway for substrate transport and determine the substrate specificity of the transporter through substrate-binding sites (Higgins, 1992, 2001; Quentin *et al.*, 1999). The NBDs are peripherally located at the *cis*-side of the membrane, where they associate with the TMDs, and are highly conserved in structure and sequence. There are some NBDs that do not interact with TMDs, being involved in other functions such as mRNA translation and DNA repair. However, for these exceptions the term “ABC transporter” is not applied and thus will not be discussed here (Higgins, 1992; Davidson *et al.*, 2008; ter Beek *et al.*, 2014).

The individual domains of an ABC transporter are frequently expressed as individual polypeptides, often found in prokaryotic species, or are fused into larger, multifunctional polypeptides. When one of the four domains is absent, one of the remaining domains functions as a homodimer (Higgins, 1992, 2001; Seyffer and Tampé, 2014).

In many ABC transporters auxiliary domains have been recruited for specific functions, such as the substrate-binding protein (SBP) a domain that binds to the substrate in the *trans*-side of the membrane and delivers it to the membrane-associated transport complex. The SBPs are soluble and periplasmic in Gram-negative bacteria, while in Gram-positive bacteria they are anchored to the membrane via an N-terminal hydrophobic lipid extension due to inexistence of an outer membrane. Their presence gives specificity and a high degree of affinity for substrates to the transport systems (Quentin *et al.*, 1999; Higgins, 2001).

ABC Transporters Classes

In contrast to the conserved NBDs, several unrelated folds of the TMDs have been found that do not share significant sequence similarity. Based in these different folds we can classify the transporters in four different classes: Importers Type I and II, Exporters and Energy Coupling Factor (ECF) transporters (Type III importers) (Figure 1.1; ter Beek *et al.*, 2014).

All three types of ABC importers are found only in prokaryotes, for the exception of ECF transporters which are also present in some plant organelles. Normally Type I importers are dedicated to the transport of compounds required in massive amounts (such as sugars and amino acids), whereas Type II importers and ECF transporters are responsible for the transport of compounds needed in small quantities (metal chelates, vitamins). Importers generally are selective for a single or a few related water-soluble substrates (Seyffer and Tampé, 2014; ter Beek *et al.*, 2014).

ABC transporters that play exporter functions can be found both in prokaryotes and eukaryotes, where they are involved in the transport of hydrophobic compounds such as lipids, fatty acids, cholesterol, and drugs. Some are called multidrug-resistant transporters because they can secret a large variety of drugs out of the cell. Also, they translocate larger molecules such as proteins (toxins, hydrolytic enzymes, S-layer proteins, lantibiotics, bacteriocins, and competence factors) (ter Beek *et al.*, 2014). Note that it is not excluded that the identical fold exerts import and export functions in homologous proteins (Seyffer and Tampé, 2014).

From now on, only to the ABC Importers Type I will be refered.

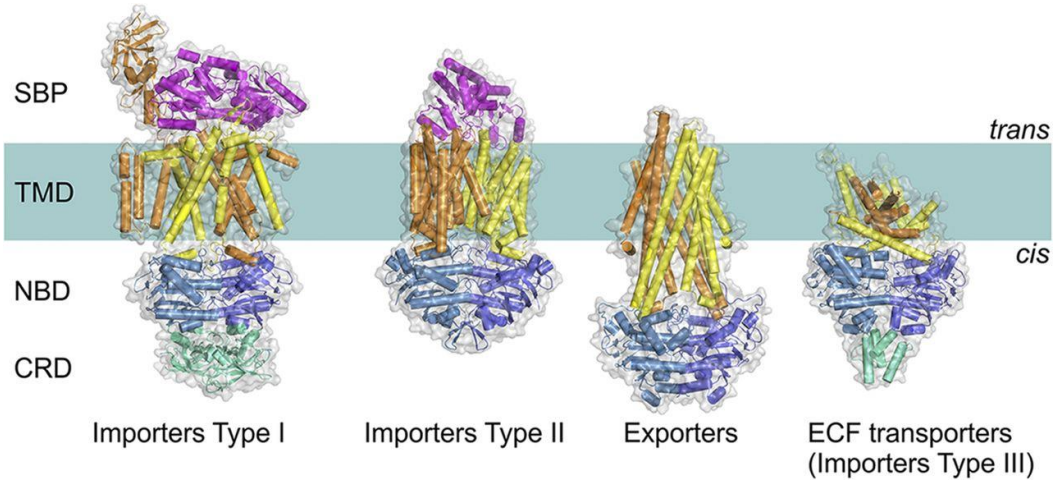


Figure 1.1 - Classes of ABC transporters. The four classes share a common structural organization: two NBDs (blue and sky blue) are attached to two TMDs (orange and yellow). Type I and II importers can have additional domains (green), which often have a regulatory function (C-terminal regulatory domain [CRD] fused to NBD). Type I and II importers have SBPs (magenta) located in the periplasm (Gram-negative bacteria) or external space (Gram-positive bacteria and Archaea) which deliver the substrates to the TMDs. ECF - energy coupling factor (ter Beek *et al.*, 2014).

Transport Mechanism

When the transporter is not translocating any substrate, it is in a resting-state conformation (Inward Open). In this conformation, the NBD dimer interface is open and the two TMDs line a cavity that is accessible from the cytoplasmic side of the membrane and is sealed by a hydrophobic gate on the outside. When substrate is present it binds to the SBP, which creates conformational changes in the TMDs and the NBDs. The NBDs come into closer proximity (the Pre-Translocation state) allowing ATP binding (Oldham *et al.*, 2007; ter Beek *et al.*, 2014).

When ATP binds, the NBD dimer closes, with the two ATP molecules located at the interface, and the cavity between the two TMDs is open toward the outside (Outward Open). Then the SBP, which stabilizes in a conformation with decreased affinity to the substrate, releases the substrate into the cavity between the TMDs where it binds to a specific site. ATP hydrolysis does not appear to affect this state, but subsequent release of Pi and ADP is expected to allow the TMDs to return to a conformation with the cavity exposed to the cytoplasm (Post-hydrolysis open), which then results in the release of the substrate and completion of the transport cycle (Oldham *et al.*, 2007; ter Beek *et al.*, 2014).

A key concept of this model is that the transition from the inward open to the outward open conformation is driven by substrate binding to the SBP instead of the binding of ATP to the NBDs (Oldham *et al.*, 2007; ter Beek *et al.*, 2014). The transport mechanism is summarized in the Figure 1.2.

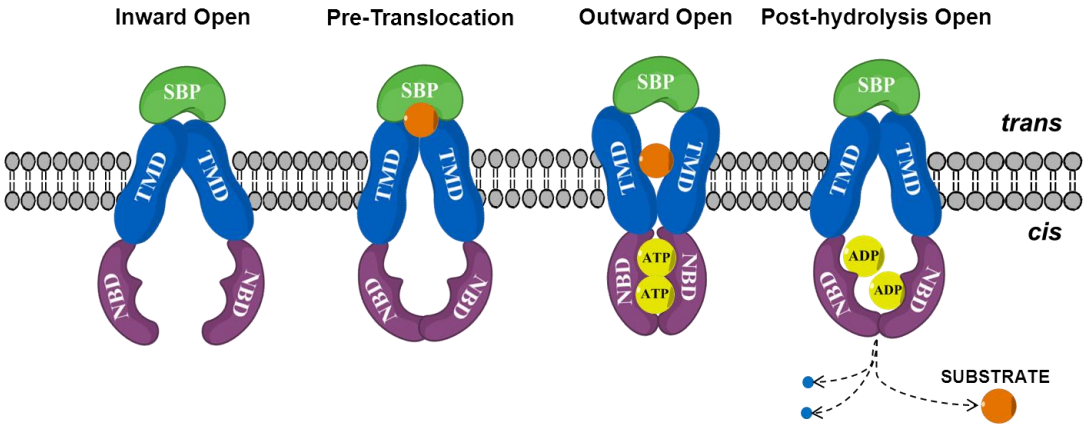


Figure 1.2 - The transport mechanism of Type I importers. First the substrate (orange) binds to the SBP (green), which creates conformational changes in the TMDs (blue) and NBDs (purple), allowing binding of ATP (yellow). When ATP binds, the NBD dimer closes and the cavity between the two TMD is open toward the outside. Subsequent ATP hydrolysis and Pi release (blue circles) return the TMDs to the initial position, releasing the substrate to the cytoplasm.

The Solute Binding Protein (SBP)

The SBP is a soluble constituent of ABC importers that is located on the *trans*-side of the membrane. They vary in size from 25 kDa to 59 kDa and there is little sequence conservation between binding proteins for different substrates (Higgins, 1992; ter Beek *et al.*, 2014). Depending on the type of cell wall the bacteria possess, the binding proteins can diffuse freely in the periplasm between the inner and the outer membrane (Gram-negative bacteria), or be

anchored to the outer surface of the cell membrane via an N-terminal lipid moiety (Gram-positive bacteria and archaea) or a Transmembrane helix (archaea). In some cases, SBPs are linked in a single polypeptide with the TMD of the transporter (Eitinger *et al.*, 2010; ter Beek *et al.*, 2014).

Despite their little sequence conservation, SBPs display a similar folding pattern composed of two symmetrical lobes or domains (termed N- and C-lobe according to the protein's termini). Each lobe is composed of an alpha-beta fold consisting of pleated β-sheets surrounded by α-helices and connected by loops, but the number and order of β-strands vary from lobe to lobe. They are connected by a hinge region and the substrate binding site is located in a cleft between the two lobes (Davidson *et al.*, 2008; Eitinger *et al.*, 2010).

In the absence of a ligand, the two lobes exist predominantly in an open conformation, but upon substrate binding, the molecule occupies the cleft and induces the lobes to rotate towards each other and close (the "Venus fly trap" model). In this new conformation, the SBP is able to interact with the appropriate transporter, with each lobe interacting with one of the TMDs (Higgins, 1992; Davidson *et al.*, 2008; ter Beek *et al.*, 2014).

These domains bind their substrates with high affinities, in the range of 0.01 to 1μM. This high-affinity binding is responsible for the capture and accumulation of substrate in the vicinity of the transporter at submicromolar concentrations (cells can concentrate nutrients up to 10^6-fold). However, SBPs are still essential for transport even at high substrate concentrations, as demonstrated by deletion of the gene encoding the maltose-SBP in *E. coli*. So they probably also play an important functional role in the catalytic cycle of the transporter. Substrate specificity is achieved mostly by differential H-bonding (Davidson *et al.*, 2008; Eitinger *et al.*, 2010).

ABC importers are specific for a single substrate or for a family of structurally related substrates, such as maltose and maltodextrins. However, some are more versatile, handling structurally unrelated substrates. This is possible either because a single SBP can recognize various substrates, as illustrated by the MalEFG1 transport system of *Thermus thermophiles* HB27 (which recognizes maltose, trehalose, sucrose and palatinose), or the importer is able to connect to multiple SBPs with different binding specificities, as illustrated by the His/Lys/Arg transport system in *Enterobacteriaceae* (Higgins and Ames, 1981; Silva *et al.*, 2005; Davidson *et al.*, 2008; ter Beek *et al.*, 2014).

The Transmembrane Domains (TMD)

The TMDs are integrated in the membrane forming a channel, located at the interface between the two domains, which is alternately accessible from the *cis*-side and *trans*-side of the membrane. The two domains of one transport system can be either homodimers or heterodimers (*e.g.*, MalF and MalG of the maltose transporter MalEFGK2 share only 13% sequence identity but are structurally related; ter Beek *et al.*, 2014).

The TMDs are highly hydrophobic and each one usually consists of six α-helices that span the membrane with the N- and C-termini on the *cis*-side, however this number varies from transporter to transporter ranging from five to eight. Probably all the transporters maintain a core number of α-helices essential for functionality and the remaining exist simply to facilitate correct folding, packing, and orientation within the membrane. The transmembrane domains contacts the other domains (SBPs and NBDs) through intra- and extracellular protrusions that connect the transmembrane segments (Higgins, 1992; Eitinger *et al.*, 2010).

Indeed, in one of those intracellular protrusions there is an α-helix that plays an essential role in the interaction between the TMDs and the NBDs. This helix is called the coupling helix, existing in each TMD, and is a conserved element located between the transmembrane helices 3 and 4 that fits into a groove of a NBD monomer (Figure 1.3A; Higgins, 2001; Eitinger *et al.*, 2010; ter Beek *et al.*, 2014).

The coupling helices are characterized by a conserved sequence of amino acids: EAA-X_3-G-X_9-I-X-LP (where X is any amino acid). In the NBD, the region that interacts with the coupling helices is the Q-loop and the groove in which the helix fits is located at the interface between the helical subdomain and the RecA-like subdomain. However, other interactions are being discovered. In the $MalFGK_2$ maltose transporter of *E. coli* it was found that the C-terminal segment of MalG (one of the TMDs) is partially inserted between the two MalK (NBDs), as illustrated in Figure 1.3B (Higgins, 2001; Eitinger *et al.*, 2010; ter Beek *et al.*, 2014).

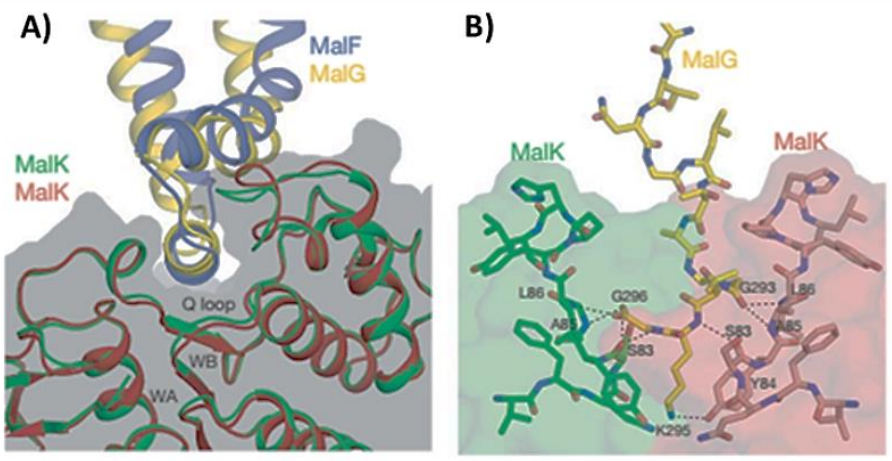


Figure 1.3 – TMDs and MalK interactions. A) Docking of the coupling helix into a surface cleft of MalK. The EAA loops of MalF and MalG are compared by superposition of the two MalK subunits. The MalK dimer is also shown as a transparent surface model. WA - Walker A motif; WB - Walker B motif. **B)** Insertion of the MalG C-terminal segment into the MalK dimer interface. The two MalK subunits are represented as a transparent surface model except for the interacting Q-loops, which are shown in stick model. Hydrogen bonds and salt bridges are indicated by black dashed lines (adapted from Oldham *et al.*, 2007).

The Nucleotide Binding Domains (NBD)

The NBDs can be considered as the 'motor domains' of ABC transporters. These domains are a subgroup of the diverse superfamily of P-loop NTPases and depend on magnesium ions for catalysis. Each NBD has a core of ~200 amino acids with a highly conserved architecture and sequence identity, varying between 30 and 50% identity depending on the transporters being compared. These proteins consist of two subdomains: the larger RecA-like domain, which is also found in other P-loop ATPases, and a structurally more diverse α-helical domain, which is unique to ABC transporters. The two subdomains are interconnected by two flexible loop regions (Higgins, 1992; Eitinger *et al.*, 2010; ter Beek *et al.*, 2014).

NBDs have a specific set of six highly conserved motifs:

1) The **P-loop** or **Walker A** motif (GXXGXGK(S/T)) forms a loop structure, containing a highly conserved lysine residue that binds to the β- and γ-phosphate of ATP.

2) The **Walker B** motif (φφφφDE, where φ is a hydrophobic amino acid) is involved in the coordination of Mg^{2+} via the conserved aspartate residue and polarizes the attacking water molecule via the glutamic acid.

3) The **D-loop** helps to form the ATP hydrolysis site.

4) The **H-loop** (name that derives from the highly conserved histidine residue present in this motif) assists with the positioning of the attacking water and Mg^{2+}.

5) The **Q-loop** has approximately eight residues with a conserved glutamine residue that binds to Mg^{2+}. Its involved in the interaction between the TMDs and NBDs.

6) The **LSGGQ signature** motif is located in the α-helical subdomain and is exclusive to the ATPase from the ABC superfamily. (Figure 1.4; Oldham *et al.*, 2007; Eitinger *et al.*, 2010; ter Beek *et al.*, 2014).

In the assembled transporter, NBDs are present as a dimer but can adopt different conformations: packed tightly against each other (closed conformation) or partially dissociated (open conformation). ATP binding promotes the closure in a “tweezer-like” fashion with the ATP bound between the Walker A residues of one monomer and the Signature motif of the other monomer. In this state ATP hydrolysis occurs and the consequent release of Pi destabilizes the dimer which reopens. When the NBD dimer is bound to ADP it resides in a semi-open state which presumably returns to the open apo-state after the dissociation of ADP. These conformational changes promotes the alternating access of the translocation pathway to the two sides of the membrane, through the interactions of the Q-loops of the NBDs and the coupling helixes of the TMDs (Rees *et al.*, 2009; Eitinger *et al.*, 2010; ter Beek *et al.*, 2014). These interactions must be specific as the ATP-binding domain from one transporter cannot normally replace that of another (Higgins, 1992).

Additionally, the NBDs activity may be regulated through protein interactions with the C-Terminal domain (C-Terminal Regulatory Domain) which has a common tertiary fold. The maltose/maltodextrins transport system can be regulated at the NBD (MalK) activity level. The C-terminal domain of MalK can interact with $EIIA^{glc}$ (involved in the carbon catabolite repression in *E. coli*) and MalT (*mal* regulon activator) (Biemans-Oldehinkel *et al.*, 2006; Cui and Davidson, 2011).

When glucose is present in the cell, MalK is inactived by EIIAglc (unphosphorylated form) binding to the C-Terminal. This regulatory process determines the hierarchy of sugar utilization and is known as inducer exclusion mechanism. On the other hand, MalK can inactivate *mal* genes expression through binding to the maltotriose binding site of MalT. Binding of maltotriose to MalT is essential for oligomerization and consequent DNA binding. This MalT inactivation only occurs when there is no transport activity. Finally, the C-Terminal domain may also be involved in a process called *trans*-inhibition. The accumulation of substrate in the cytoplasm can decrease transport activity and this is called *trans*-inhibition. In some ATPases, the regulatory domain forms substrate-binding pockets where the substrate can bind to the dimer locking the

transporter in the inward open conformation (Biemans-Oldehinkel *et al.*, 2006; Cui and Davidson, 2011).

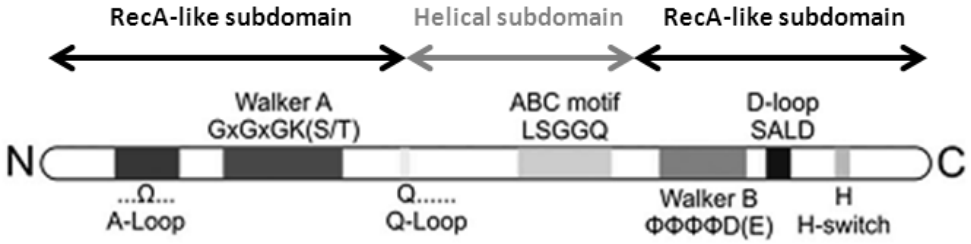


Figure 1.4 – Conserved motifs in the NBDs domain and subdomains. The RecA-like domain comprises the Walker A and B motifs responsible for binding ATP, the H-loop which has a highly conserved residue that allows contact with γ-phosphate of ATP and the D loop that contacts the Walker A of the other monomer. In the Helical subdomain is localized the signature motif, which is unique to ABC ATPases and also contacts ATP. The Q-loop, like the H-loop, has a highly conserved residue and is responsible for the formation and disruption of the catalytic site. This motif is present in one of the connective loops. The C-Terminal regulatory domain is not shown (adapted from ter Beek *et al.*, 2014).

Multitask ATPases

Initially it was thought that each transporter complex had an exclusive ATPase protein, with each TMDs couple having a unique interaction site only recognised by one ATPase in the cell (Schneider and Hunke, 1998). However with the increase of knowledge about these transport systems new observations lay some doubts on this idea.

In *S. mutans*, the MsmEFGK system is a multiple sugar transporter responsible for the transport of raffinose, melibiose, stachyose, isomaltose, and isomaltotriose (referred in the SBP subsection). There is also another ABC transporter named MalXFGK which is responsible for the uptake of maltodextrins. Both transport systems have their own ATPases (MsmK and MalK) to energize the translocation process, which are encoded in the same operon as the other protein domains of the respective transport system. Surprisingly it was shown that both ATPases can interact with either their own or the alternative transporter complex (Russell *et al.*, 1992; Webb *et al.*, 2008).

Streptomyces species have an ATPase named MsiK, which is involved in oligosaccharide uptake systems. MsiK was shown to energize the transport systems responsible for the uptake of cellobiose, xylobiose and maltose in *Streptomyces lividans* (Hurtubise *et al.*, 1995; Kampers *et al.*, 1997), while in *Streptomyces reticuli*, MsiK energizes the trehalose uptake system

(Schlösser, 2000). In *Streptomyces coelicolor* beside being required for the transport of N,N'-diacetylchitobiose, some of the data suggest that MsiK plays a role in the utilization of the disaccharides maltose and cellobiose (Saito *et al.*, 2008).

In the Gram-positive *B. subtilis* an *in silico* analysis identified at least 78 ABC transporters based on the identification of 86 NBDs in 78 proteins, 103 TMDs proteins, and 37 SBP proteins. From these 78 ABC systems at least 10 are predicted to be involved in sugar uptake, however 8 of these systems do not have a NBD protein being coded in the same operon. The *B. subtilis* genome encodes two potential ATPases capable of energizing the uptake of sugars: *msmX* and *yurJ* (Quentin *et al.*, 1999; Ferreira and de Sá-Nogueira, 2010). Recent work, demonstrated that MsmX interacts with several of these 8 distinct ABC sugar importers orphans of an NBD, thus energizing the transport of several sugar oligomers: maltodextrins (Schönert *et al.*, 2006), arabinooligosaccharides (Ferreira and de Sá-Nogueira, 2010), galactooligosaccharides, galaturonic acid oligomers and rhamnogalacturonic acid (Ferreira and de Sá-Nogueira, unpublished data). Thus, unlike other NBDs, MsmX was shown to be multitask serving as energy-generating component to several sugar importers (Figure 1.5).

This type of multitask ATPases was also observed in the pathogenic species *Streptococcus pneumoniae*. It was demonstrated that the MsmK ATPase, from *S. pneumoniae* TIGR4, is the NBD protein partner for the raffinose, maltotetraose, fructooligosaccharides and sialic acid ABC importer systems. Thus, MsmK has been show to energize four of the six Carbohydrate Uptake Transporter-1 (CUT1) family transporters found in this species (Marion *et al.*, 2011; Linke *et al.*, 2013).

These evidences demonstrate that bacteria evolved towards the sharing of energy-generating components between ABC importers, from the same organism, involved in carbohydrate utilization. Whether this evolved for genome minimization or transport regulation remains unknown (Buckwalter and King, 2012). Therefore a better knowledge about these multitask ATPases may have an important impact in therapy development for pathogenic bacteria.

First of all, carbohydrate utilization is crucial for the survival of many microorganisms. If the utilization of these substrates is impaired most probably it will be possible to prevent the proliferation of pathogenic organisms in certain specific conditions (Buckwalter and King, 2012). One example is MsmK which has been linked to airway colonization maintenance (Marion *et al.*, 2011). Secondly, biofilm formation, which confers resistance to antibiotics, has been shown to be linked to carbohydrate metabolism. Recently it was demonstrated that *B. subtilis* biofilm formation was induced by polysaccharides like pectin, arabinogalactan and xylan (Beauregard *et al.*, 2013). Also galactan consumption and galactose metabolism play an important role in biofilm production (Chai *et al.*, 2012). Furthermore, the ABC importers are a mechanism of transport exclusive to prokaryotes, thus these ATPases may represent potential targets for therapy since they do not exist in human cells.

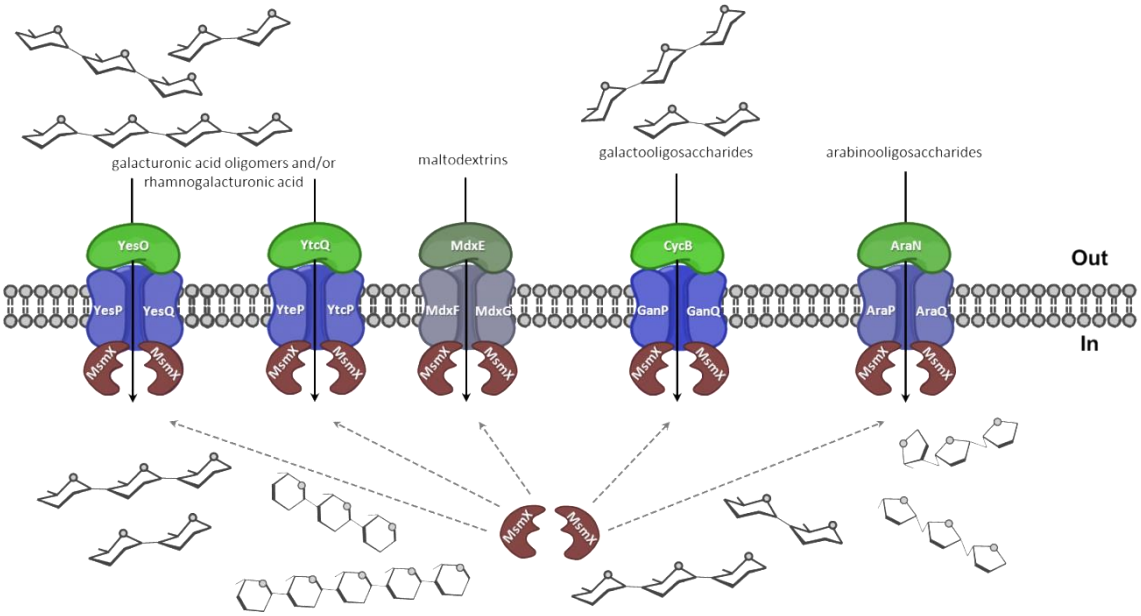


Figure 1.5 - MsmX-dependent ABC importers in *B. subtilis*. The ABC-type importer AraNPQ is involved in the uptake of α-1,5-arabinooligosaccharides (arabinotriose, arabinotetraose and some arabinobiose). The ABC-type importer MdxEFG is involved in the uptake of maltodextrins. The CycB/GanPQ ABC-type importer is responsible for the uptake of galactooligosaccharides. The YesOPQ and YtcQ/YesPQ transport systems are involved in the uptake of galacturonic acid oligomers and/or rhamnogalaturonic acid (adapted from Schönert *et al.*, 2006 and Ferreira and Sá-Nogueira, 2010 and unpublished data).

1.3. Scope of the Thesis

In this work we will show that *B. subtilis* can be used as a model organism to study these multitask ATPases from other Gram-positive pathogenic species, facilitating the study of these proteins since *B. subtilis* is a well-known organism and easily manipulated genetically. For this purpose, we demonstrate that these multitask ATPases from other species are able to replace the *B. subtilis* MsmX function in the cell. Furthermore, we show that MsmX is in fact a protein capable to hydrolize ATP and that amino acid K43 is essential for MsmX functionality both *in vivo* and *in vitro*.

Chapter 2

Development of an in vivo system to test the functionality of msmX Alleles and MsmX homologs

2. Development of an *in vivo* system to test the functionality of *msmX* Alleles and MsmX homologs

2.1 Introduction

In the Gram-positive model organism *Bacillus subtilis* the MsmX ATPase was shown to energize multiple ABC carbohydrate importers. One of those importers is the AraNPQ transport system, encoded by the *ara* operon, which is responsible for the uptake of arabino-oligosaccharides and that requires the presence of MsmX for energizing the system (Ferreira and de Sá-Nogueira, 2010).

The gene encoding MsmX seems to belong to an operon located at 340ºC where it is co-transcribed with *yxkF*, a putative PucR regulatory protein (http://www.ncbi.nlm.nih.gov; Kunst *et al.*, 1997). However monocistronic messages were detected in both rich and minimal medium (Yoshida *et al.*, 2000). An *in silico* survey identified MsmX homologs in several pathogenic bacteria.

The aim of the work presented in this chapter is to develop an *in vivo* system in order to test if those homologs are capable of playing the role of MsmX in the cell, hence demonstrating that the *B. subtilis* multitask ATPase can be used as a model to study those proteins. Genes from *Bacillus thuringiensis*, *Staphylococcus aureus* and *Streptococcus pneumoniae* were introduced in a *B. subtilis msmX*-null mutant genome under the control of an inducible promoter. The capacity of each gene to complement MsmX deficiency was evaluated through their ability to energize the AraNPQ system in *B. subtilis*. We show that these proteins from different pathogenic bacteria are able to substitute MsmX and energize the *B. subtilis* transporter. Moreover, we showed that modifications in the N- and C-terminal amino acids of MsmX do not affect its functionality or its ability to interact with the AraNPQ transporter system.

2.2 Materials and Methods

Substrates

1,5-α-L-Arabinotriose (sugar beet, purity 95%) was purchased from Megazyme International Ireland Ltd., arabinose and glucose from Sigma-Aldrich Co.

Bioinformatic Analysis

For the identification of MsmX homologs in other Gram-positive bacteria bioinformatics tools were used. BLASTp algorithm was used to compare MsmX amino acid sequence with the sequence database from the National Center for Biotechnology Information at the National Institutes of Health, Bethesda, Maryland (http://www.ncbi.nlm.nih.gov). Proteins with an identity superior to 60% were considered as targets. An amino acid sequences alignment was made using the multiple sequence alignment algorithm ClustalW2 (EMBL-EBI). Translation rates were calculated using RBS Calculator$_{v2.0}$ created by Salis Lab, Pennsylvania State University (Salis *et al.*, 2009; Salis, 2011).

DNA manipulation and sequencing

Routine DNA manipulations were performed as described by Sambrook *et al*. (1989). All restriction enzymes were purchased from Thermo Fisher Scientific Inc. and used according to the manufacturers' recommendations. PCR amplifications were carried out using Phusion® High-Fidelity DNA Polymerase (Thermo Fisher Scientific Inc.). DNA from agarose gels and PCR products were purified with the illustra™ GFX™ PCR DNA and Gel Band Purification kit (GE Healthcare). All DNA ligations were performed using T4 DNA Ligase (Thermo Fisher Scientific Inc.). Plasmids were purified using the QIAGEN® Plasmid Midi kit (Qiagen) or NZYMiniprep kit (NZYTech, Lda.). DNA sequencing was performed with the ABI PRISM BigDye® Terminator Cycle Sequencing Kit (Applied Biosystems). The sequencing reaction was purified by gel filtration and resolved in an ABI 3730XL sequencer.

Construction of plasmids and strains

Plasmid pAM4 was obtained by amplification of the *msmX* gene from chromosomal DNA of the wild-type strain *B. subtilis* 168T$^+$, with the oligonucleotides ARA741 and ARA742, which

contain unique restriction sites *Sal*I and *Sph*I, and the resulting fragment (1224bp) inserted between the *Sal*I and *Sph*I sites of pDR111 (gift from David Rudner, Harvard University). pAM5 was obtained by amplification of the *HD73_4301* gene, from chromosomal DNA of strain *B. thuringiensis* serovar kurstaki str. HD73 (Bacillus Genetic Stock Center, BGSC, Ohio State University), with the oligonucleotides ARA744 and ARA745, bearing unique restriction sites *Sal*I and *Sph*I, and subsequent cloning of this fragment (1272bp) into pDR111 digested with *Sal*I and *Sph*I. The amplification of the *ugpC* gene, from chromosomal DNA of the pathogenic strain *Staphylococcus aureus subsp. aureus* ST398 (a gift from Hermínia de Lencastre, ITQB, Universidade Nova de Lisboa), using oligonucleotides ARA746 and ARA747, which contain unique restriction sites *Sal*I and *Sph*I, and cloning this fragment (1215bp) into pDR111 *Sal*I-*Sph*I, yielded plasmid pAM6. The amplification of the *msmK* gene, from chromosomal DNA of strain *Streptococcus pneumoniae* TIGR4 (a gift from Hermínia de Lencastre, ITQB, Universidade Nova de Lisboa), with the oligonucleotides ARA748 and ARA749, harbouring unique restriction sites *Sal*I and *Sph*I, and subsequent cloning of this fragment (1295bp) between the *Sal*I and *Sph*I sites of pDR111, yielded plasmid pAM7. pAM9 and pAM10 were obtained by site-directed mutagenesis and their construction is described below. Plasmid pAM12 was obtained by amplification of the *msmX* sequence using pAM9 as template with the oligonucleotides ARA741 and ARA742, which contain unique restriction sites *Sal*I and *Sph*I, and then the resulting fragment (1224bp) was inserted between the *Sal*I and *Sph*I sites of pDR111. Plasmids and oligonucleotides used in this work are listed in Table 2.1 and 2.2, respectively.

Table 2.1 - List of plasmids used or constructed during the course of this work.

Plasmids	Relevant construction	Source or Reference
pDR111	Derivative of the Pspac(hy) plasmid pJQ43; contains an additional *lacO* binding site	David Rudner
pAM4	PDR111 derivate, with *msmX* under the control of Pspank(hy)	This work*
pAM5	PDR111 derivate, with *HD73_4301* under the control of Pspank(hy)	This work*
pAM6	PDR111 derivate, with *ugpC* under the control of Pspank(hy)	This work*
pAM7	PDR111 derivate, with *msmK* under the control of Pspank(hy)	This work*
pMJ1	pBluescript II KS(+)-based plasmid harboring the *msmX* coding region	Ferreira and de Sá-Nogueira, 2010
pAM9	pMJ1 derivate, with a new *Bgl*II restriction site in *msmX* coding region	This work*
pAM10	pAM10 derivate, with a new *Nhe*I restriction site in *msmX* coding region	This work*
pAM12	PDR111 derivate, with *msmX* from pAM10 under the control of Pspank(hy)	This work*

* See Appendices from to 6.2 to 6.7 and 6.9.

Table 2.2 - List of oligonucleotides used in this work.

Oligonucleotides	Sequence[a]
ARA741	CTTGTGTCGACAGGGAATTGCTG
ARA742	CCCTTGCATGCGGTTTGATTCTGAG
ARA744	CGGCGTCGACTGAATCTATTCG
ARA745	TCATGGCATGCAGTAGAAGCCC
ARA746	CGGGTCGACACGAAGTGTATTGC
ARA747	ATACGCATGCCACGGCTAACGTG
ARA748	GTATCGTCGACTGGTCATCTTGCC
ARA749	GCACGCATGCACTGATATCTCTCC
ARA769	GACAGAAGTGAGATCTCGATAAGATC
ARA770	CTTATCGAGATCTCACTTCTGTCTC
ARA771	TTAGATGGCTAGCTTGCGGATGG
ARA772	CCGCAAGCTAGCCATCTAACATC

[a] Sequence orientation is 5'→3'. Restriction sites are underlined.

Plasmids pDR111, pAM4, pAM5, pAM6, and pAM7, were used to transform the IQB495 strain, according to the method described by Anagnostopoulos and Spizizen, 1961, resulting in strains IQB672 (pDR111), IQB673 (pAM4), IQB674 (pAM7), IQB677 (pAM5) and IQB678 (pAM6) strains (Table 2.3). The transformants were screened for *amyE*$^-$ phenotype. The *amyE* phenotype was tested on plates of solid LB medium containing 1% (w/v) potato starch; after overnight incubation, plates were flooded with a solution of 0.5% (w/v) I_2–5.0% (w/v) KI for detection of starch hydrolysis.

Table 2.3 - List of *B. subtilis* strains used or constructed during the course of this work.

Strain	Relevant genotype	Sources or Reference[a]
168T$^+$	Prototroph	F. E. Young
IQB495	Δ*msmX*::*cat*	Ferreira and de Sá-Nogueira, 2010
IQB672	Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)	pDR111 → IQB495
IQB673	Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmX*	pAM4 → IQB495
IQB674	Δ*msmX*::*cat* ΔamyE::Pspank(hy)-*msmK*	pAM7 → IQB495
IQB676	Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmX*(Glu3Ser, Ile364Ser)	pAM12 → IQB495
IQB677	Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*HD73_4301*	pAM5 → IQB495
IQB678	Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*ugpC*	pAM6 → IQB495

[a] The arrows indicate transformation and point from donor DNA to recipient strain

Site-Directed Mutagenesis

Vector pMJ1 was used as template for site-directed mutagenesis experiments using the mutagenic oligonucleotides ARA769 and ARA770. This pair of primers allowed the generation of an unique *Bgl*II restriction site in the 4th nucleotide, downstream from the start codon This originated a mutation in the residue at position 3 (Glu to Ser) in the resulting plasmid pAM9. Afterwards, pAM9 was used as a template with mutagenic oligonucleotides ARA771 and ARA772, which created an unique *Nhe*I restriction site in the 11th nucleotide, upstream from the stop codon. The resulting plasmid pAM10 contains a mutation in the residue at position 364 (Ile to Ser). For both site-directed mutagenesis experiments, a polymerase chain reaction was carried on using 1x Phusion® GC Buffer (Thermo Fisher Scientific Inc.), 0.2 µM primers, 200 µM dNTPs, 3% DMSO, 0.8 ng/µL of Template DNA (first pMJ1 and in the second pAM9) and 0.02 U/µL of Phusion® High-Fidelity DNA Polymerase in a total volume of 50 µL. The PCR product was digested with 10 U of *Dpn*I, at 37 ºC, overnight. Both mutations were confirmed by DNA sequencing.

Growth conditions

E. coli XL1Blue (Strategene) was used as host for the construction of all pDR111 derives and *E. coli DH5α* (Gibco-BRL) used for pMJ1 derives. All *E. coli* strains were grown in liquid Luria-Bertani (LB) medium (Miller, 1972) and on LB solidified with 1.6% (w/v) agar, ampicillin (100 µg/mL) and/or Tetracycline (12 µg/mL) were added as required. *B. subtilis* was grown in liquid LB medium, LB medium solidified with 1.6% (w/v) agar or SP medium (Martin *et al.*, 1987) with chloramphenicol (5 µg/mL) and spectinomycin (50 µg/mL) being added as required. Growth kinetics parameters of the wild-type and mutant *B. subtilis* strains were determined in liquid minimal medium. *B. subtilis* strains 168T^{+}, IQB672, IQB673, IQB674, IQB676, IQB677 and IQB678 were grown overnight (37 ºC, 150 rpm), from freshly streaked colonies, in C minimal medium (Pascal *et al.*, 1971) supplemented with L-tryptophan (100 µg/mL), potassium glutamate (8 µg/mL) and potassium succinate (6 µg/mL) (CSK medium; Debarbouille *et. al*, 1990). The cell cultures were washed and resuspended, to an initial OD_{600nm} of 0.05, in 1.5 mL of CSK medium without potassium succinate and supplemented with different carbon and energy sources (glucose, arabinose and arabinotriose) at a final concentration of 0.1% (w/v) and IPTG (expression inducer) at 1 mM was added when appropriated. The cultures were grown in sterile 50 mL Falcon tubes (Sarstedt), incubated at 37 ºC and 150 rpm in an OLS200 orbital/linear shaking bath (Grant Instruments) and the OD_{600nm} periodically read in an Ultrospec 2100 pro UV/Visible Spectrophotometer (GE Healthcare Life Sciences).

2.3 Results and Discussion

In silico Survey

An *in silico* survey was performed for the identification of MsmX homologs, in other Gram-positive bacteria, using BLASTp tool. From this analysis *Bacillus thuringiensis*, *Bacillus cereus*, *Staphylococcus aureus* and *Streptococcus pneumoniae* proteins were selected (Table 2.4). The proteins from *B. thuringiensis* and *B. cereus* have the higher identity (74%) while UgpC from *S. aureus* exhibits a homology of 66%. MsmK from *S. pneumoniae* possess the lower homology (64%). Although we identified proteins from *B. thuringiensis* and *B. cereus* in these studies we only used the first because they both have identical amino acid sequence. A multiple alignment of the proteins primary sequence was performed using ClustalW2 and is shown in the Appendix 6.11. The genomic context of each gene selected for study is displayed in Appendices 6.36, 6.37 and 6.38.

Table 2.4 – MsmX homologs selected. List of MsmX homologs selected for this work, with respective NCBI Database Reference and identity with MsmX.

Species	Gene name	NCBI Reference	Protein Length	Identities
***Bacillus thuringiensis* serovar *kurstaki* str. HD73**	*HD73_4301*	YP_007423400.1	366	271 (74%)
***Bacillus cereus* ATCC 14579**	*BC4016*	NP_833734.1	366	271 (74%)
***Staphylococcus aureus* subsp. *aureus* ST398**	*ugpC (or SAPIGO223)*	YP_005733033.1	365	240 (66%)
***Streptococcus pneumoniae* TIGR4**	*msmK (or SP_1580)*	NP_346026.1	375	241 (64%)

Development of a Genetic System for Complementation Analysis

In order to test if other multitask ATPases (MsmX homologs), found in pathogenic species, are able to play the roles of MsmX in the cell, we developed an *in vivo* system for their expression in a *B. subtilis msmX*-null mutant (IQB495). For this, we placed the genes under the control of an inducible promoter and used an integrative *B. subtilis* vector.

The incorporation of the genes in the chromosome guarantees no variations in the number of copies present in the cell during the experiments and for comparison analysis. The use of an inducible promoter allows a comparable and controlled expression of the different genes. The

capability of the other ATPases to fully (or partially) complement the *msmX* absence was evaluated using the AraNPQ transport system, which cannot transport arabinotriose in the absence of MsmX. Determination of the growth kinetics parameters of the distinct strains, bearing the genes encoding each one of the ATPases selected, using arabinotriose as the sole carbon and energy source was the physiological parameter used in order to evaluate the functionality of each protein in the cell.

The integrative vector chosen to construct this system was pDR111 (Appendix 6.1) that enables gene expression from a modified version of the Pspac promoter, the Phyper-spank promoter (Pspank(hy)), which is stronger and bears an additional lacO binding site to achieve a better repression in the absence of inducer (Quisel *et al.*, 2001; Britton *et al.*, 2002). Gene expression can be induced with IPTG, a compound that mimics allolactose but is not hydrolysable, therefore maintaining sustainable levels of expression along time. The plasmid possesses two *amyE* gene fragments for integration at the *amyE* locus of the *B. subtilis* chromosome. However this plasmid does not have any ribosomal binding site (RBS) thus the genes were cloned into this vector with their own RBS and for practical reasons also their terminator.

A schematic illustration of the *in vivo* system developed and the strains constructed for this work are represented in Figure 2.1 and 2.2, respectively.

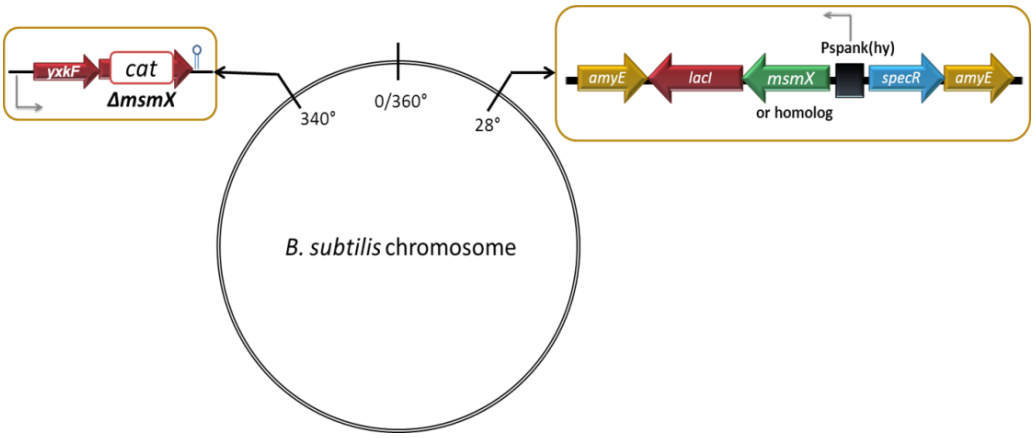


Figure 2.1 – Schematic illustration of the *in vivo* system developed. The *amyE* locus at 28º is represented (top left) with the genes introduced by a double-recombination event with pDR111 derivatives: a *lacI* copy (red arrow), a spectinomycin resistance cassette (blue arrow), both from pDR111, a copy of the *msmX* gene, or homolog (green arrow) and the *amyE* gene fragments (yellow arrows). At 340º, is shown the *yxkF-msmX* operon with the *msmX* inactivated by an insertion-deletion mutation with a chloramphenicol resistance cassette (*cat*).

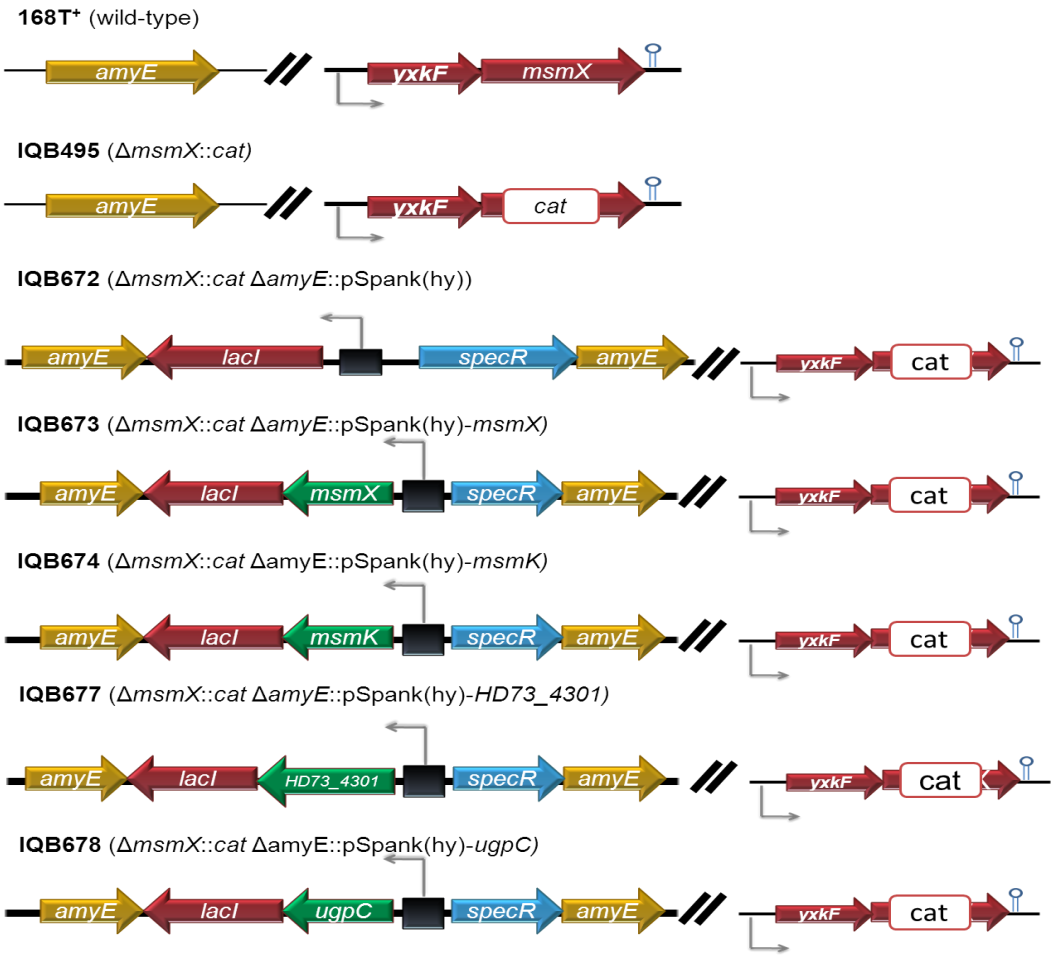


Figure 2.2 – Schematic representation of the *amyE* locus and the *yxkf-msmX* operon in the chromosome of several *B. subtilis* strains used in this work (not drawn to scale).

MsmX Functional Studies

First the functionality of the system was tested using the *msmX* under the control of Pspank(hy) in a *msmX*-null mutant genetic background (strain IQB673; Figure 2.2). A negative control (strain IQB492; Figure 2.2) was constructed by transformation of strain IQB495, *msmX*-null mutant genetic background, with plasmid pDR111. The wild-type strain 168T⁺ was used as positive control. Growth kinetics parameters of the three strains in minimal medium using arabinotriose (MsmX dependent uptake), glucose and arabinose (both MsmX independent uptake) as the sole carbon and energy source, were determined and the results are summarized in Table 2.5.

The doubling time of strains IQB672 and IQB673 in a medium supplemented with glucose and arabinose are similar to those of the wild-type strain 168T⁺ and the presence of IPTG did

not cause any variation in these values. Variations were not expected since both saccharides are transported by systems that are not dependent of MsmX for functionality. Moreover, the addition of IPTG does not have any toxic effect on the cells, and the presence of a copy of the *lacI* gene and the spectinomycin resistance gene do not affect growth in these conditions.

Table 2.5 – Growth of different strains 168T⁺, IQB672 and IQB673 in the presence of several saccharides as sole carbon/energy.

Carbon Source	Doubling Time (minutes)*		
	168T⁺ (wild-type)	**IQB672** (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy))	**IQB673** (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmX*)
Glucose 0.1%	60.65±0.58	58.58±1.45	58.38±2.24
Glucose 0.1% + IPTG 1mM	-	59.81±0.68	59.07±1.97
Arabinose 0.1%	78.75±3.09	79.76±1.64	78.47±3.47
Arabinose 0.1% + IPTG 1mM	-	76.61±2.94	77.53±2.95
Arabinotriose 0.1%	114.25±3.70	No Growth	No Growth
Arabinotriose 0.1% + IPTG 1mM	-	No Growth	107.60±3.85

* Cells were grown in CSK minimal medium (see Materials and Methods). Growth kinetics parameters were determined and the results represent the average of three independent experiments, except the glucose and arabinose growths for strains IQB672 and IQB673 without IPTG, which were performed only twice (see Annexes for the complete data used in the determination of doubling time).

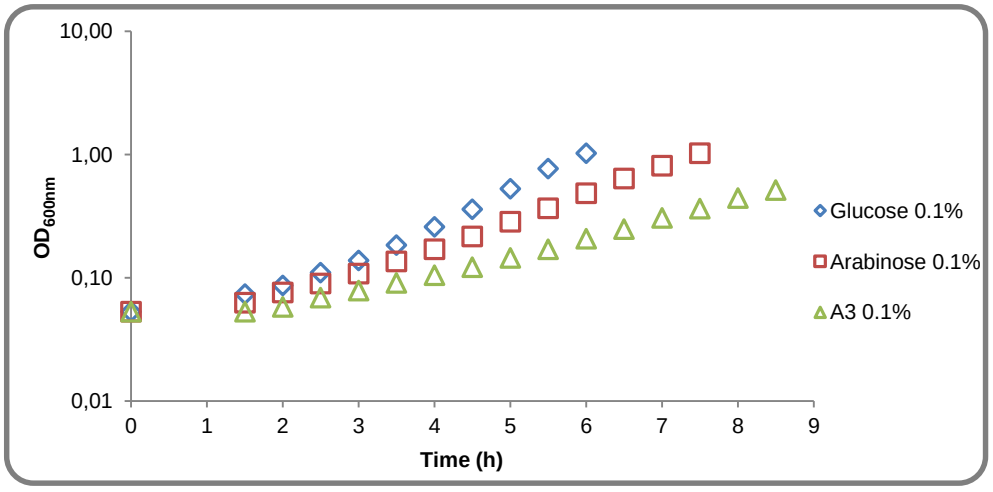


Figure 2.3 - Growth of *B. subtilis* 168T⁺ (wild-type) in CSK medium using glucose, arabinose and arabinotriose as the sole carbon and energy source.

The results obtained using only arabinotriose as a carbon/energy source display different behaviours. As expected, the negative control, strain IQB672, was not able to use the saccharide to support growth, even in the presence of IPTG (Table 2.5 and Figure 2.4). In this strain, there is no MsmX present in the cells to energize the AraNPQ transport system. Strain IQB673 was also not able to grow in the presence of arabinotriose, however when IPTG was added to the medium, the wild-type phenotype was restored since the doubling time is similar to that displayed by strain 168T⁺ (Table 2.5 and Figure 2.5).

These results show that IPTG induces expression of the *msmX* gene present at the *amyE* locus, which allows the cell to transport arabinotriose from the medium to the cytoplasm through the AraNPQ transporter.

Furthermore, when we compare the doubling time of strains 168T⁺ and IQB673 with arabinotriose the results are very similar (considering the deviations; Table 2.5). Nonetheless, a very slight decrease in IQB673 doubling time is observed and this variation could be due to a small increase in the amount of MsmX present in the cell, due to the fact that the copy of *msmX* present in *trans* is being transcribed from a different promoter. This observation however, is not relevant for the purpose of this genetic system, which is to test the functionality of MsmX homologs, or different alleles of *msmX* bearing mutations, because kinetics of growth will be compared to strain IQB673 and not to the 168T⁺ strain.

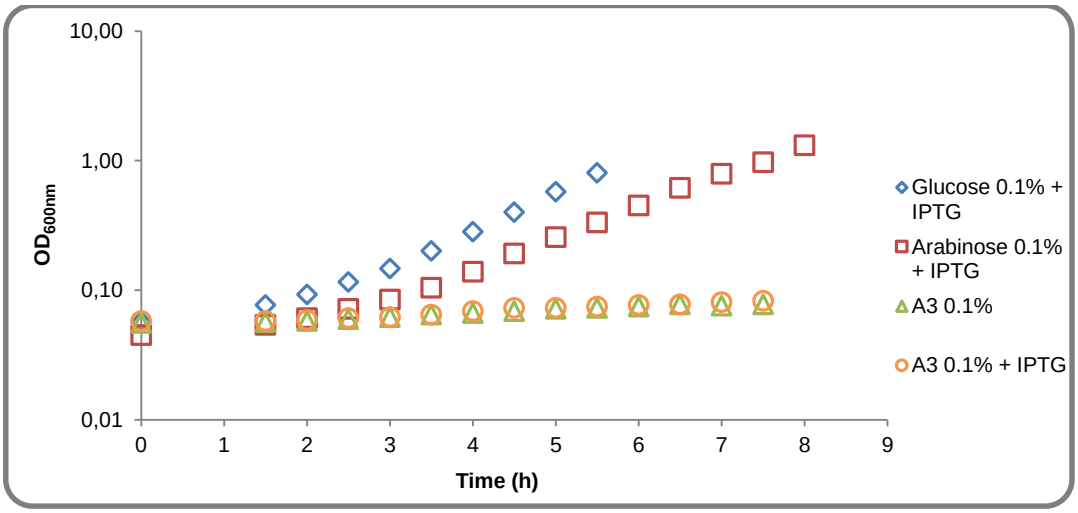


Figure 2.4 - Growth of IQB672 (Δ*msmX*::*cat* Δ*amyE*::pSpank(hy)) in CSK medium using glucose, arabinose and arabinotriose as the sole carbon and energy source.

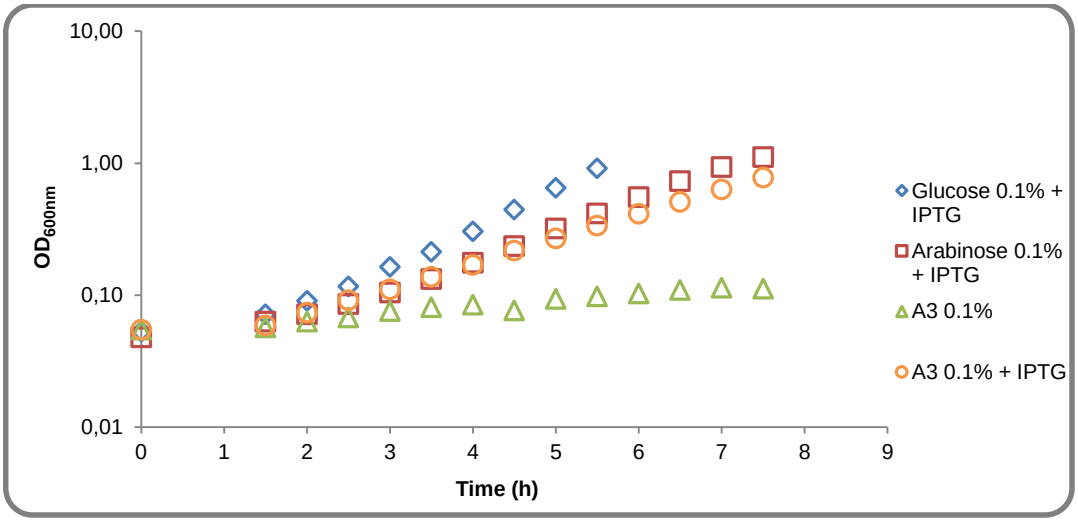


Figure 2.5 - Growth of IQB673 (Δ*msmX*::*cat* Δ*amyE*::pSpank(hy)-*msmX*) in CSK medium using glucose, arabinose and arabinotriose as the sole carbon and energy source.

In sum, the results obtained indicate that the expression of *msmX* in *trans* using the system developed does not affect the bacteria phenotype in the presence of arabinotriose *i.e*. we may conclude that using this system the intracellular concentration of MsmX is sufficient to energize the AraNPQ transporter as observed in the wild-type strain. Cells expressing *msmX* in *trans*, under the control of the inducible Pspank(hy) promoter, and the wild-type cells expressing the *msmX* in the original locus, under the control of its own promoter, display a similar phenotype under the tested conditions. Thus, the genetic system constructed is suitable for studying the effect of mutations in the *msmX* allele and to test the functionality of MsmX homologs in *B. subtilis*.

Functional Analysis of MsmX Homologs

As described above, we identified MsmX homologs in several pathogenic species and select proteins from *B. thuringiensis*, *S. aureus* and *S. pneumoniae*. Using the system developed, we tested if the homologs chosen were able to complement MsmX absence in *B. subtilis*.

The different genes were placed under the control of the inducible Pspank(hy) promoter at the *amyE* locus and resulted in the construction of strains IQB674 (*msmK* gene from *S. pneumoniae*), IQB677 (*HD73_4301* from *B. thuringiensis*) and IQB678 (*ugpC* from *S. aureus*) is depicted in Figure 2.2. Growth kinetics parameters of the three strains in minimal medium using arabinotriose (MsmX dependent uptake), glucose and arabinose (both MsmX independent

uptake) as the sole carbon and energy source, were determined and the results are summarized in Table 2.6.

Table 2.6 - Growth of different *B. subtilis* strains in the presence of several saccharides as sole carbon/energy.

Carbon Source	Doubling Time (minutes)*			
	IQB673 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmX*)	**IQB674** (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmK* [*S. pneumonia*])	**IQB677** (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*HD73_4301* [*B. thuringiensis*])	**IQB678** (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*ugpC* [*S. aureus*])
Glucose 0.1%	58.38±2.24	59.17±1.81	59.98±1.18	59.38±0.50
Glucose 0.1% + IPTG 1mM	59.07±1.97	59.88±2.52	60.62±0.71	61.49±2.13
Arabinose 0.1%	78.47±3.47	80.97±3.81	79.01±0.59	79.98±1.65
Arabinose 0.1% + IPTG 1mM	77.53±2.95	81.40±2.20	77.22±1.98	81.40±2.20
Arabinotriose 0.1%	No Growth	260.87±11.18	No Growth	No Growth
Arabinotriose 0.1% + IPTG 1mM	107.60±3.85	226.05±9.55	141.18±3.76	205.47±8.28

* Cells were grown in CSK minimal medium (see Materials and Methods). Growth kinetics parameters were determined and the results represent the average of three independent experiments, except the glucose and arabinose growths without IPTG, which were performed only twice (see Annexes for the complete data used in the determination of doubling time). IQB673 results are the same presented in table 5 and are shown here to facilitate comparison.

The doubling time of strains IQB674, IQB677, and IQB678, in a medium supplemented with glucose and arabinose are similar to those of the wild-type strain 168T$^+$ and the presence of IPTG did not cause any variation in these values, as previously observed with strain IQB673 (Table 2.5 and Table 2.6).

Regarding the growth with arabinotriose in inducible conditions, plus IPTG, all the three strains were able to growth in the presence of arabinotriose as the sole carbon and energy source. This phenotype is not observed in the absence of IPTG in strains IQB677, and IQB678 (Table 2.6 and Figures 2.7 and 2.8), unexpectedly strain IQB674 was able to growth in the presence of arabinotriose even in the absence of IPTG (Table 2.6 and Figure 2.6). A closer look to the sequence of *msmK* amplified from *S. pneumonia* and cloned in pDR111 revealed that in addition to the RBS we also cloned a putative *msmK* promoter (PmsmK). Thus, transcription of *msmK* is initiated at this promoter and is not subjected to regulation by LacI. Interestingly, the growth of strain IQB674 in arabinotriose in the absence of IPTG is slower than that observed in the presence of IPTG (a difference of about 34 min), suggesting that in inducing conditions transcription of *msmK* is also driven by the Pspank(hy) promoter.

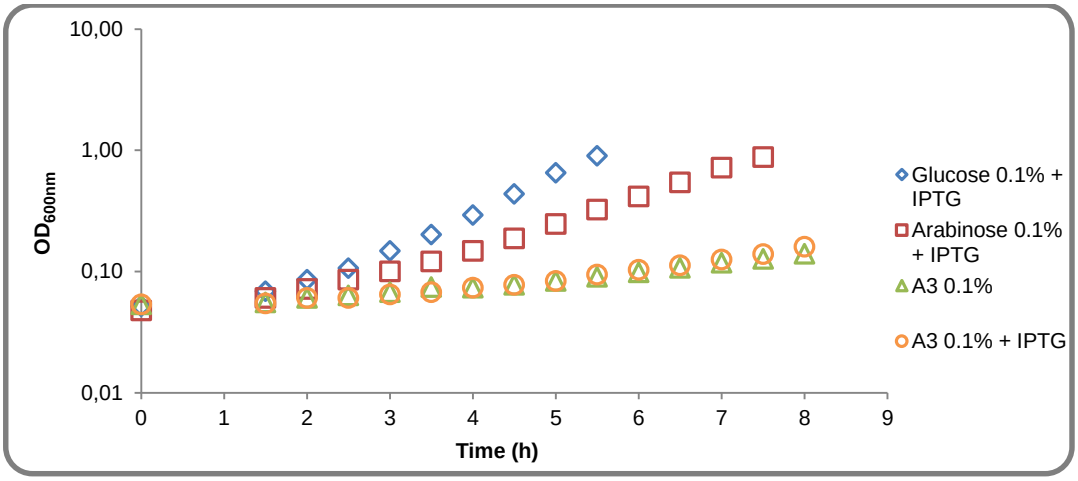


Figure 2.6 - Growth of IQB674 (Δ*msmX*::*cat* Δ*amyE*::pSpank(hy)-*msmK*) in CSK medium using glucose, arabinose and arabinotriose as the sole carbon and energy source.

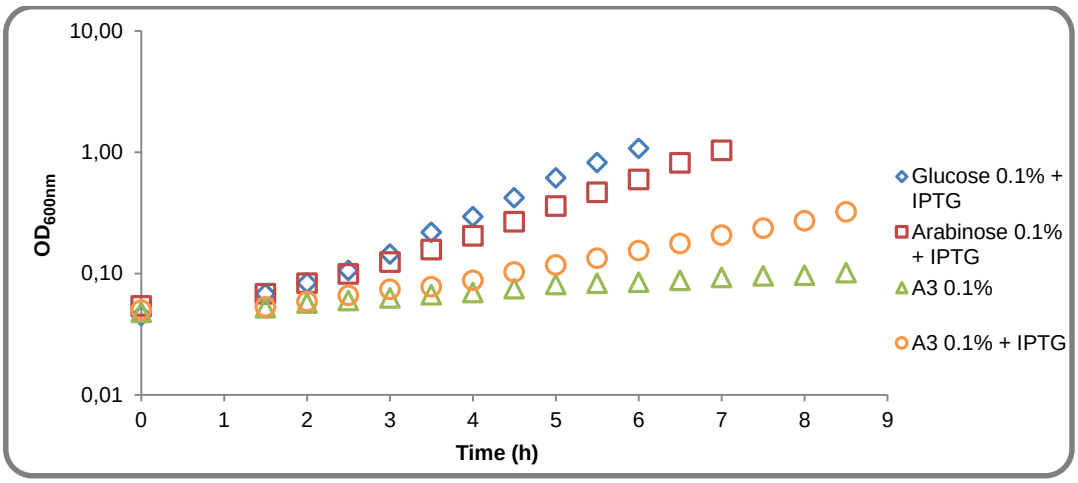


Figure 2.7 - Growth of IQB677 (Δ*msmX*::*cat* Δ*amyE*::pSpank(hy)-*HD73_4301*) in CSK medium using glucose, arabinose and arabinotriose as the sole carbon and energy source.

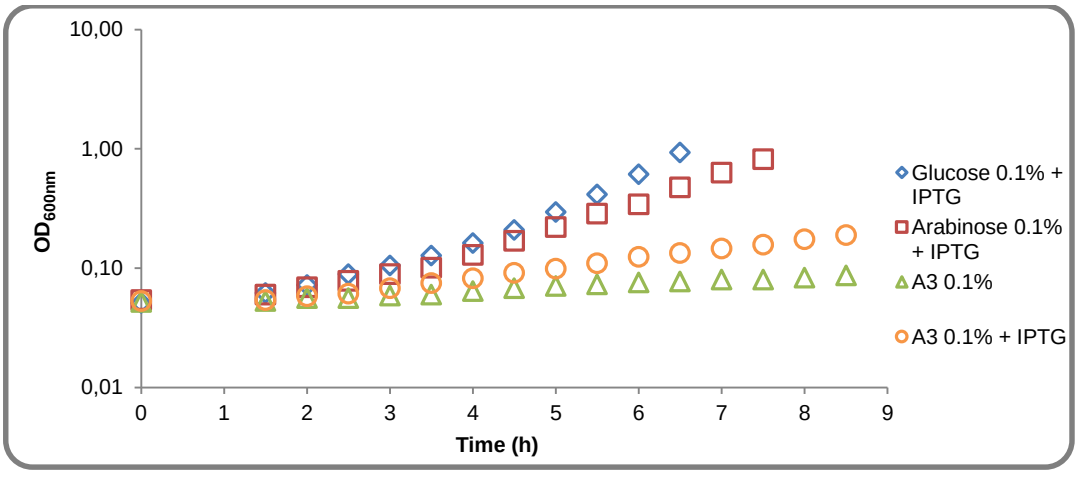


Figure 2.8 - Growth of IQB678 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*ugpC*) in CSK medium using glucose, arabinose and arabinotriose as the sole carbon and energy source.

In conclusion these results show that *msmK* from *S. pneumoniae*, *HD73_4301* from *B. thuringiensis* and *ugpC* from *S. aureus* are able to play the role of *msmX* in energizing the *B. subtilis* AraNPQ transporter, and suggest that MsmK, HD73_4301, and UgpC are able to interact with the TMDs from the AraNPQ transporter in the absence of MsmX.

Nevertheless, the complementation of MsmX deficiency observed is only partial, the doubling time, determined in inducing conditions in the presence of arabinotriose in strains IQB674 (*msmK* from *S. pneumoniae*), IQB677 (*HD73_4301* from *B. thuringiensis*) and IQB678 (*ugpC* from *S. aureus*) increased 2.1-fold, 1.3 fold, and 1.9-fold, respectively, when compared with strain IQB673 (*msmX* from *B. subtilis*). Comparing these results it seems that the protein HD73_4301 from *B. thuringiensis* is the one who reaches a value closer to full complementation of MsmX, while between the other two proteins, UgpC from *S. aureus* achieves a higher complementation degree than MsmK from *S. pneumoniae*. If we compare the complementation degree with protein identity (or homology), or phylogenetic relationship between the species used in this work, we can observe some correlation between them. The protein displaying higher identity to MsmX is HD73_4301 (74%), which also came from the species phylogenetically more close to *B. subtilis* (same genus), achieved the best complementation degree, while the protein with less identity (64%) and also from the further species, in phylogenetic terms, attained the worst complementation degree. So it is tempting to speculate that a higher protein identity corresponds to a greater ability to play the MsmX role.

Interestingly, the region of the proteins predicted to interact with the TMDs of AraPQ is highly conserved (see Appendix 6.11). However, differences in one or two amino acids near the Q-loop, may be sufficient to alter the interaction between the ATPase and one of the TMDs. Changes in the electrochemical properties in those positions could affect the interactions protein-protein directly or indirectly through interaction with amino acids important for the connection with the TMD. These changes could not prevent connection between the two proteins, but be enough to slow down the connection and interfering with the uptake rate of arabinotriose and, consequently, increasing the doubling time of the strains when using this saccharide as carbon/energy source.

Another aspect that could contribute for these results is possible differences in post-transcriptional regulation that result in variations in the total NBD protein amount produced in each strain. Since each gene was cloned together with its own RBS, and belong to different species, we calculated the translation initiation rates for the mRNAs transcribed from each gene to estimate the impact of the different RBS in translation initiation. The results obtained are summarized in Table 2.7.

Table 2.7 – RBS Calculator algorithm results for each RBS. The translation initiation rate is a relative value and is represented in arbitrary units (au). ΔG_{total} – Total energy necessary for translation initiation.

Gene	Translation Initiation Rate (au)	ΔG_{total} (kcal/mol)
msmX	42339.64	-6.29
HD73_4301	7519.89	-2.45
ugpC	229874.35	-10.05
msmK	18759.16	-4.48

These results clearly show us that exists differences in the post-transcriptional regulation processes, but they do not correlate with the complementation results. In addition to differences in the RBS strength, there are other processes that can affect post-transcriptional regulation, such as the mRNA stability.

Since this system was created not only to study these proteins, but also other possible multitask ATPases from other bacteria or even from *B. subtilis* we decided to improve our design in order to eliminate those variables, which could influence the results obtained in the future experiments.

Fine-tuning the Genetic System for Functional Analysis

In order to eliminate transcriptional and the post-transcriptional regulation variables, we improved the system's design. For this purpose we modified the N-terminal and C-terminal of MsmX in order to allow the sub-cloning of only the Open Reading Frame (ORF) of the gene to be tested.

The integrative vector was engineered to harbour the RBS and the terminator of *B. subtilis msmX*. The *msmX* coding sequence was mutated in order to introduce two new unique restriction sites at the 5'-end and 3'-end of the gene, one right after the RBS (*Nhe*I) and the other before the stop codon (*Bgl*II). The first one was introduced between the second (Ala) and third amino acid (Glu), and led to the transition of the Glu residue to a Ser. The second restriction site was introduced between amino acids 363 (Arg) and 364 (Ile), substituting Ile by a Ser residue (Figure 2.9). The resulting plasmid pAM12, a pDR111 derivative, containing the modified version of *msmX*, in future studies will be digested with *Nhe*I and *Bgl*II, which enable the removal of the *msmX* coding region, leaving behind the RBS and terminator, then the ORF of the gene object of study will be cloned between those restriction sites.

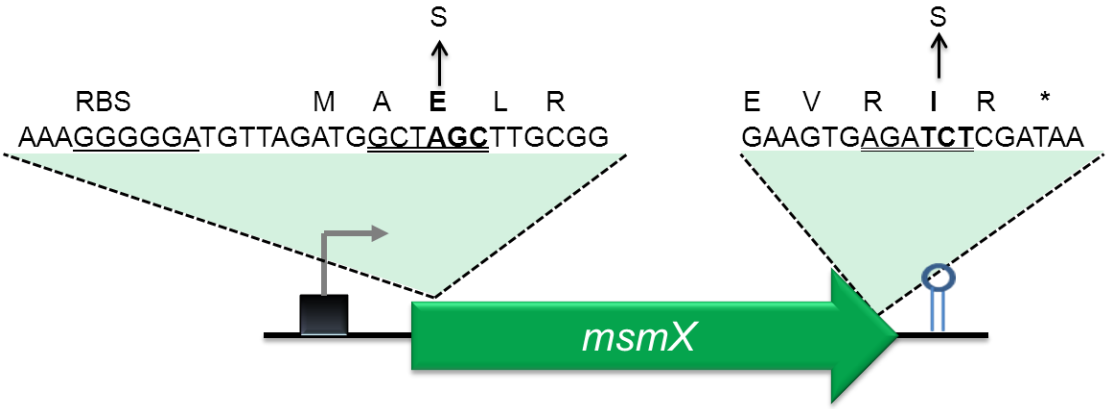


Figure 2.9 - Schematic representation of the modifications made in the genetic system. The *msmX* gene is represented by a green arrow and the promoter of the transcriptional unit is depicted by a grey arrow and a black box. Above is displayed the sequence of the 5' and 3'-end of the *msmX* gene. The ribosome-binding site, RBS, is underlined. The nucleotides that were mutated are represented in bold and the corresponding modification in the primary sequence of MsmX is depicted above. The new restriction sites, *Nhe*I (5'-end) and *Bgl*II (3'-end), are double underlined.

Although the substitutions were made in the N-terminal and C-terminal MsmX regions, which are known not to be important for the function of the NBDs, we tested if the functionality and/or capability to interact with the AraNPQ transporter components where not affected. For this purpose, we constructed strain IQB676, which harbours the modified version of *msmX* under the control of the inducible Pspank(hy) promoter at the *amyE* locus, and determined the growth kinetic parameters using arabinotriose, glucose and arabinose as the sole carbon and energy source. The results are summarized in Table 2.8.

Table 2.8 - Growth of IQB673 and IQB676 strains in the presence of several saccharides as sole carbon/energy.

Carbon Source	Doubling Time (minutes)*	
	IQB673 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmX*)	**IQB676** (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmX*(Glu3Ser, Ile364Ser))
Glucose 0.1%	58.38±2.24	60.21±0.34
Glucose 0.1% + IPTG 1mM	59.07±1.97	58.68±1.73
Arabinose 0.1%	78.47±3.47	76.99±0.28
Arabinose 0.1% + IPTG 1mM	77.53±2.95	76.80±0.78
Arabinotriose 0.1%	No Growth	No Growth
Arabinotriose 0.1% + IPTG 1mM	107.60±3.85	110.27±4.78

* Cells were grown in CSK minimal medium (see Materials and Methods). Growth Kinetics parameters were determined and the results represent the average of three independent experiments, except the glucose and arabinose growths without IPTG, which were performed only twice (see Annexes for the complete data used in the determination of doubling time). IQB673 results are the same presented in table 5 and are shown here to facilitate comparison.

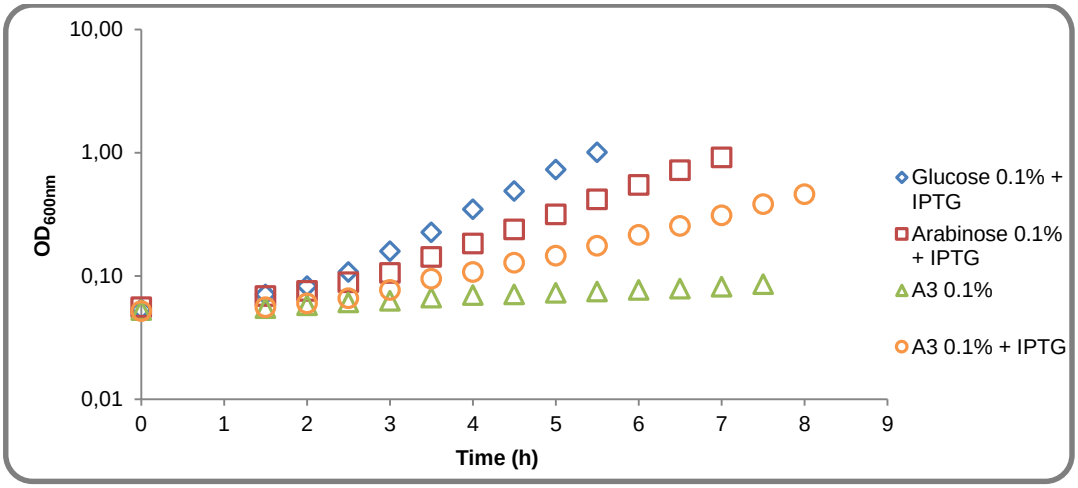


Figure 2.10 - Growth of IQB676 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmX*(Glu3Ser, Ile364Ser)) in CSK medium using glucose, arabinose and arabinotriose as the sole carbon and energy source.

The doubling time obtained with this strain when in a medium supplemented with glucose and arabinose is similar to IQB673 and the presence of IPTG also did not cause any variation in these values. In the assays with arabinotriose, but without IPTG added to the medium, there was no growth. When IPTG was added the cells were able to grow, with a doubling time similar to IQB673. These results indicated that the amino acids substitutions introduced in MsmX do not affect neither its functionality nor the interaction with the TMDs, and this improved system can be used in future complementation tests.

Chapter 3

Biochemical Characterization of MsmX

3. Biochemical Characterization of MsmX

3.1 Introduction

The first ATPase with a resolved crystal structure was HisP, from *Salmonella typhimurium*, belonging to the well-characterized histidine ABC transporter $HisQMP_2$ (Hung *et al.*, 1998). Since then the crystal structure of several ATPases was solved, such as the MJ0796 and MJ1267 proteins from *Methanocaldococcus jannaschii* (Karpowich *et al.*, 2001; Smith *et al.*, 2002). Other protein structures were determined with the NBD attached to the TMDs from the correspondent transporter like MalK from *E. coli*, ModC from *Archaeoglobus fulgidus* and *Methanosarcina acetivorans* (Chen *et al.*, 2003; Hollenstein *et al.*, 2007; Gerber *et al.*, 2008). More recently, the structure from ATPases MetNI from *E. coli*, involved in methionine uptake, and LIC12079 from *Leptospira spp.*, involved in manganese utilization, were also determined (Rees *et al.*, 2009; Benaroudj *et al.*, 2013).

Beside crystallography studies, dynamic monomer/dimer equilibrium and ATPase activity studies were performed to unravel other important structural information (Horn *et al.*, 2003; Verdon *et al.*, 2003; Zoghbi *et al.*, 2012; Benaroudj *et al.*, 2013). Some of these studies revealed residues important for ATP hydrolysis like the Glu residue from Walker B motif, the second Gly residue from the Signature motif and the conserved Lys from the Walker A motif, while the Ser residue from the Signature motif was not essential for the enzymatic activity (Verdon *et al.*, 2003; Benaroudj *et al.*, 2013).

In spite of all these studies, the only well-characterized ATPase of ABC transporters involved in the uptake of carbohydrates is MalK from *E. coli*. To better comprehend the energy-generating component sharing phenomenon among distinct sugar ABC importers, structural and biochemical characterization of multitask ATPases like MsmX must be performed. MsmX is highly similar (>40% identity) to MalK from *E. coli* and MsmK from *S. mutans*, ATPases of transport systems for maltose and multiple sugars, respectively (Ferreira and de Sá-Nogueira, 2010).

The aim of the work presented in this chapter was to optimize the production of MsmX in *E. coli* and further purification in order to obtain the necessary amount of protein for functional and structure characterization. Additionally functional analysis was performed by site-directed mutagenesis. The Lys43 residue of MsmX, predicted to be essential for ATP hydrolysis, was substituted by an Ala and the effects were observed both *in vivo* and *in vitro*.

3.2 Materials and Methods

DNA manipulation and sequencing

Routine DNA manipulations were performed as described by Sambrook *et al*. (1989). All restriction enzymes were purchased from Thermo Fisher Scientific Inc. and used according to the manufacturers recommendations. PCR amplifications were carried out using Phusion® High-Fidelity DNA Polymerase (Thermo Fisher Scientific Inc.). DNA from agarose gels and PCR products were purified with the illustra™ GFX™ PCR DNA and Gel Band Purification kit (GE Healthcare). All DNA ligations were performed using T4 DNA Ligase (Thermo Fisher Scientific Inc.). Plasmids were purified using the QIAGEN® Plasmid Midi kit (Qiagen) or NZYMiniprep kit (NZYTech, Lda.). DNA sequencing was performed with the ABI PRISM BigDye® Terminator Cycle Sequencing Kit (Applied Biosystems). The sequencing reaction was purified by gel filtration and resolved in an ABI 3730XL sequencer.

Site-Directed Mutagenesis

Vectors pMJ22 and pAM4 were used in independent experiments as template for site-directed mutagenesis using the mutagenic oligonucleotides ARA757 and ARA758 (Table 3.1). This pair of primers introduced the mutation of codon AAA (Lys at position 43) to GCA (Ala) in the resulting plasmids pAM11 and pAM13, respectively. A polymerase chain reaction was carried on using 1x Phusion® GC Buffer (Thermo Fisher Scientific Inc.), 0.2 µM primers, 200 µM dNTPs, 3% DMSO, 0.6 ng/µL of pMJ22 DNA (or 0.5 ng/µL of pAM4 DNA) and 0.02 U/µL of Phusion® High-Fidelity DNA Polymerase in a total volume of 50 µL. The PCR product was digested with 10 U of *Dpn*I, at 37 ºC, overnight. The mutation was confirmed by DNA sequencing.

Table 3.1 - List of oligonucleotides used in this work.

Oligonucleotides	Sequence[a]
ARA757	TGCGGGGCATCAACGACGCTGCGAATGG
ARA758	CATTCGCAGCGTCGTTGATGCCCCGCAG

[a] Sequence orientation is 5'→3'. Mutated nucleotides are underlined

Construction of strains

pAM13 was used to transform IQB495 strain (Table 3.2), according to the method described by Anagnostopoulos and Spizizen, 1961, resulting in the IQB675 strain (Table 3.3). The transformants were screened for *amyE*$^-$ phenotype. The *amyE* phenotype was tested in plates of solid LB medium containing 1% (w/v) potato starch; after overnight incubation, plates were flooded with a solution of 0.5% (w/v) I_2–5.0% (w/v) KI for detection of starch hydrolysis.

Table 3.2 - List of plasmids used or constructed during the course of this work.

Plasmids	Relevant construction	Source or Reference
pT-GroE	Vector for expression of GroESL chaperones under the control of T7 promoter	(Yasukawa *et al.*, 1995)
pMJ22	pET30(a)-based vector for the expression of *msmX* in *E. coli*	Ferreira and de Sá-Nogueira, unpublished data
pAM4	PDR111 derivate, with *msmX* under the control of Pspank(hy)	This work*
pAM11	pMJ22 derivate with msmX sequence with mutated AAA (Lys at position 43) to GCA (Ala)	This work*
pAM13	pAM4 derivate with msmX sequence with mutated AAA (Lys at position 43) to GCA (Ala)	This work*

* See Appendices 6.2, 6.8 and 6.10.

Table 3.3 - List of *B. subtilis* strains used or constructed during the course of this work.

Strain	Relevant genotype	Sources or Reference[a]
IQB675	Δ*msmX*::*cat* Δ*amyE*::pSpank(hy)-*msmX*(Lys43Ala)	pAM13 → IQB495

[a] The arrows indicate transformation and point from donor DNA to recipient strain

Growth conditions

E. coli DH5α (Gibco-BRL) was used as host for routine DNA manipulations for plasmids with pET-30(a) backbone and *E. coli* BL21 (DE3) pLysS (Studier *et al*., 1990) and *E. coli* BL21 (DE3) (Novagen) were used as hosts for the overproduction of recombinant MsmX and MsmX-Lys43Ala. *E. coli* XL1Blue (Strategene) was used for the construction of pDR111 derivatives. All

E. coli strains were grown in liquid Luria-Bertani (LB) medium (Miller, 1972) and on LB solidified with 1.6% (w/v) agar, and ampicillin (100 µg/mL), Tetracycline (12 µg/mL), kanamycin (30 µg/mL) and/or chloramphenicol (25 µg/mL) were added as required.

B. subtilis was grown in liquid LB medium, LB medium solidified with 1.6% (w/v) agar or SP medium (Martin *et al.*, 1987) and chloramphenicol (5 µg/mL) and spectinomycin (50 µg/mL) were added as required.

Growth kinetics parameters were determined in liquid minimal medium. *B. subtilis* strain IQB675 was grown overnight (37 ºC, 150 rpm), from freshly streaked colonies, in C minimal medium (Pascal *et al*., 1971) supplemented with L-tryptophan (100 µg/mL), potassium glutamate (8 µg/mL) and potassium succinate (6 µg/mL) (CSK medium; Debarbouille *et al.*, 1990). The cell cultures were washed and resuspended, to an initial OD_{600nm} of 0.05, in 1.5 mL of CSK medium without potassium succinate and supplemented with different carbon and energy sources (glucose, arabinose and arabinotriose) at a final concentration of 0.1% (w/v) and with IPTG (expression inducer) at 1mM was added when appropriated. The cultures were grown in sterile 50 mL Falcon tubes (Sarstedt), incubated at 37 ºC and 150 rpm in an OLS200 orbital/linear shaking bath (Grant Instruments) and the OD_{600nm} periodically measured in an Ultrospec 2100 pro UV/Visible Spectrophotometer (GE Healthcare Life Sciences)..

Production and purification of recombinant proteins

E. coli BL21 (DE3) cells harbouring both pMJ22 and pT-GroE were grown at 37 ºC and 150 rpm in 10 mL of LB medium with antibiotics (small-scale production). When the OD_{600nm} of the cultures reached 0.6, 5 mL of each culture were induced by the addition of IPTG to a final concentration of 1 mM. Both the induced and non-induced cultures were grown for a further 18 hours at 25 ºC, and then the cells were harvested by centrifugation at 16000 *g* for 5 minutes and resuspended in 100 µL of Lysis Buffer (20 mM sodium phosphate buffer, pH 7.4, 62.5 mM NaCl, 100 mM imidazole, glycerol 10% (v/v)). To disrupt the cells, incubation with lysozyme (1 mg/mL) at 37 ºC for 10 minutes and three cycles of freezing in liquid nitrogen and thawing for 5 minutes at 37 ºC were performed, followed by incubation with 0.02 U/mL of benzonase (Sigma) and PMSF (10 mM) at 37 ºC for 10 minutes. After centrifugation at 9000 *g* and 4 ºC for 15 minutes, the soluble fraction was recovered and the insoluble fraction was resuspended in 100 µL of Lysis Buffer.

E. coli BL21 (DE3) cells harbouring pMJ22 were grown at 37 ºC and 150 rpm in 10 mL of LB medium with antibiotics (small-scale production). When the OD_{600nm} of the cultures reached 0.6, 5 mL of each culture was induced by the addition of IPTG to a final concentration of 0.1 mM. Both the induced and non-induced cultures were grown for a further 16 hours at 16 ºC, and then cells were harvested and lysed as described above.

To purify MsmX, a large-scale induction was performed. *E. coli* BL21 (DE3) harbouring pMJ22 was incubated at 37 ºC and 150 rpm in 1 L of LB medium, using antibiotics. When the OD_{600nm} of the cultures reached 0.6, the culture was induced by the addition of IPTG to a final concentration of 0.1 mM and grown for a further 16 hours at 16 ºC. The cells were harvested by centrifugation at 12000 *g*. The pellets were resuspended in 10mL of Start Buffer (20 mM Tris-HCl, pH 7.4, 50 mM NaCl and 10 mM imidazole) and 5 µl of benzonase at 2.5 U/mL were added. The cells were lysed by sonication, and then centrifuged at 23500 *g* (30 minutes and 4 ºC). The supernatant was filtered using a 0.45 µm filter (Sarstedt) and loaded into a 1mL Hitrap Chelating HP column (GE Healthcare). The bound proteins were eluted with a discontinuous gradient of imidazole (40 mM, 100 mM, 200 mM, 300 mM and 500 mM).

To purify MsmX-Lys43Ala, a large-scale induction was performed with *E. coli* BL21 (DE3) harbouring pAM11, using the same conditions as described to MsmX. However the purification protocol was slightly changed instead of using the 20 mM Tris-HCl buffers, a 10 mM sodium phosphate and 500 mM NaCl, were used.

Protein analysis

The analysis of the production, purification and molecular mass of the enzymes (monomers and/or oligomers) was performed with SDS-PAGE (stained with Coomassie Blue), using Low Molecular Weight Protein Marker (NZYTech). The protein quantification was determined using Bio-Rad Protein Assay (Bio-Rad Laboratories, Inc).

ATPase activity assay

Purified MsmX and MsmX-Lys43Ala were assayed for ATP hydrolysis by monitoring the production of inorganic phosphate at 37 °C in 20 mM Tris-HCl, pH 7.4, 50 mM NaCl and 10 mM $MgCl_2$. MsmX-Lys43Ala was stored in sodium phosphate buffer which was changed to the assays buffer using PD-10 Desalting Columns (GE Healthcare Life Sciences). Activity assays were performed with 5 µg of total protein and 0.1 mM ATP.

The assay mixture was incubated at 37ºC for 3h, with inorganic phosphate being measured at 0 h, 1 h, 2 h and 3 h. The quantification of inorganic phosphate was determined using the Malachite Green Phosphate Detection Kit (R&D Systems, Minneapolis, MN, USA) in accordance with the manufacturer's instructions. Background phosphate levels were monitored in parallel using a control reaction without protein. OD_{620nm} was measured in an Ultrospec 2100 pro UV/Visible Spectrophotometer (GE Healthcare Life Sciences).

3.3. Results and Discussion

Sequence Analysis

MsmX, which can be transcribed monocistronicaly, encodes a 365 aa with a molecular weight of ~41 kDa and an isoelectric point of 7.385 (http://subtiwiki.uni-goettingen.de; Michna *et al.*, 2014). MsmX is highly similar (>40% identity) to MalK from *E. coli* and a protein sequence alignment of MsmX and MalK highlights the most conserved residues (Figure 2.1).

The characteristic motifs of ATP-binding proteins family are indicated in the alignment: Walker A and B, Q-loop, Signature, D-loop and H-loop. Although these motifs are highly conserved in both proteins, as observed in this type of ATPases, there are some minor differences. In the Q-loop there is a semi-conservative substitution on position 84 (Asn to Ser), while in Walker B occurs a conservative substitution in position 158 (Met to Leu), maintaining the hydrophobic characteristic necessary for that position. Three amino acids in the H-loop were substituted by amino acids with distinct characteristics but maintaining the conserved histidine.

The online tool PSIpred, which predicts secondary structures, was used to compare MsmX and MalK (Jones, 1999; McGuffin *et al.*, 2000; Bryson *et al.*, 2005) and the results are shown in Figure 2.2. The structures predicted with the program are highly similar from the N-terminus to the H-loop motif. From this motif towards the C-terminal some differences are noticed in the structure, with MsmX displaying three additional α-helices. The C-terminal domain of MalK is known to be involved in regulatory processes through interactions with regulatory proteins (Biemans-Oldehinkel *et al.*, 2006).

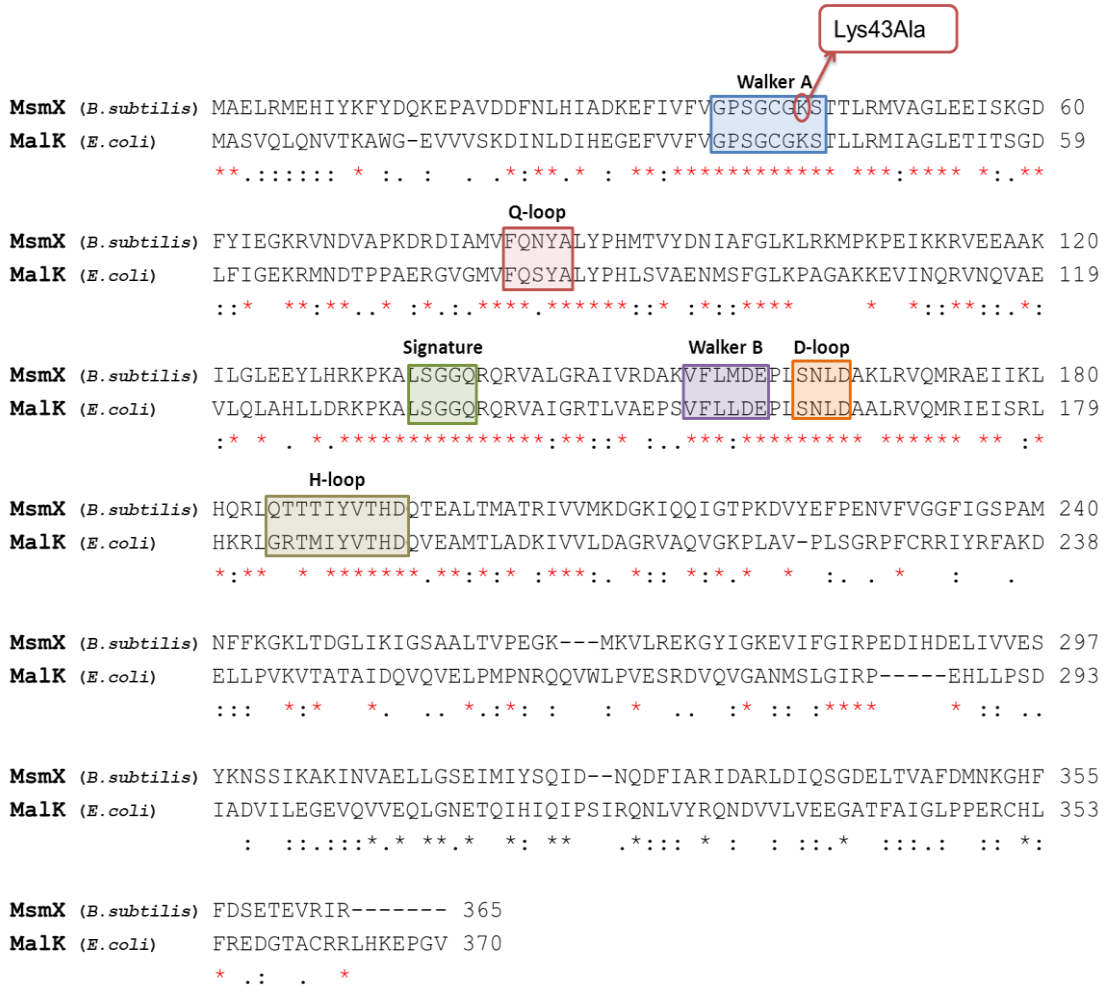


Figure 3.1 - MsmX and MalK sequence alignment. The amino acid sequences of MsmX and MalK from *E. coli* were aligned using ClustalW2 (Thompson *et al.*, 1994). Identical amino acids are marked with a red asterisk, while substitutions with residues displaying the same properties are marked with ":" and semi-conservative substitutions are marked with a ".". Gaps in the amino acid sequence are represented by a "–". The conserved motifs are labeled in colors. The red arrow indicates the mutation performed in MsmX (Lys43A) and described latter in this chapter.

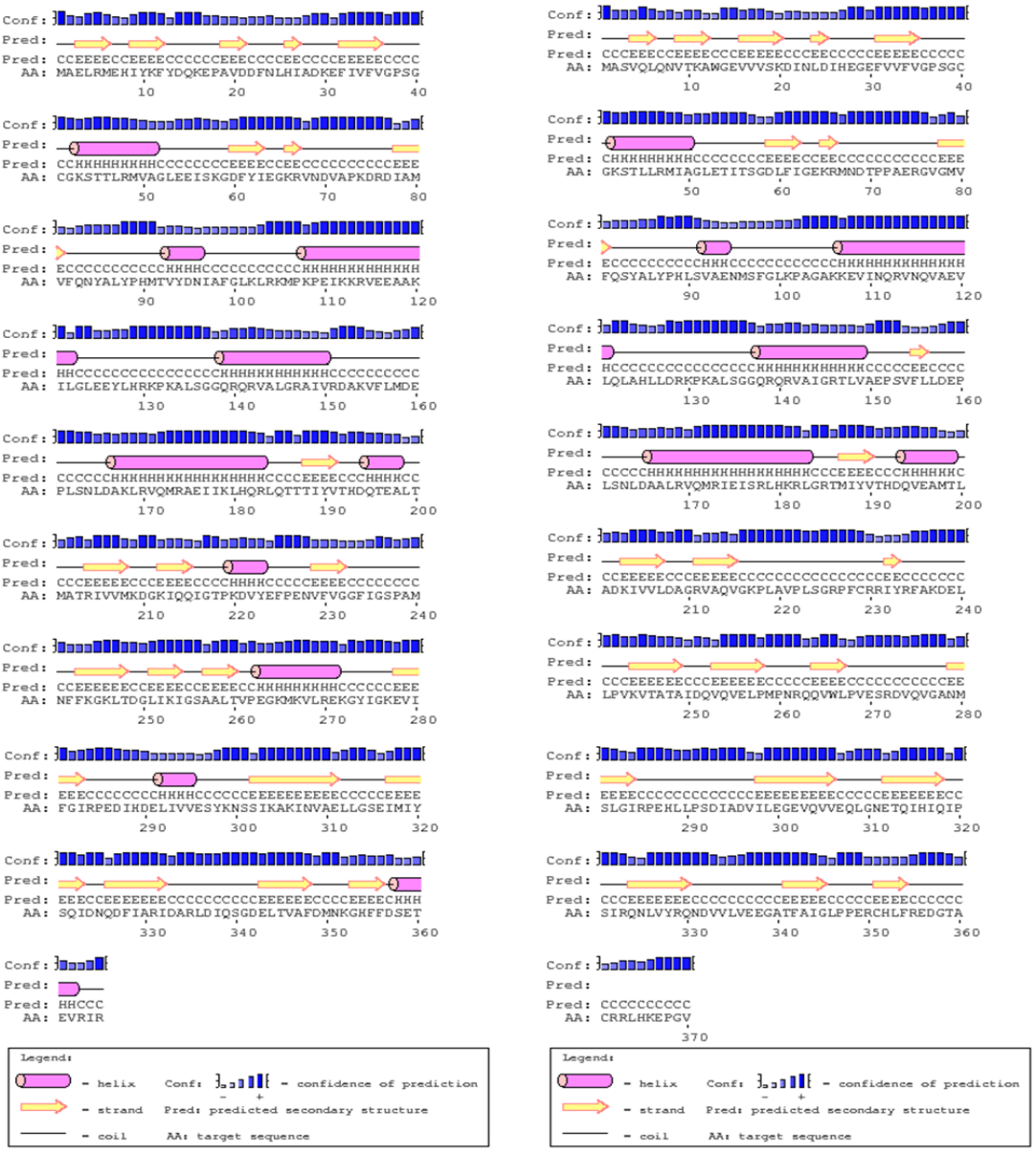


Figure 3.2 – MsmX secondary structure prevision by PSIpred. On the left is the predicted secondary structure for MsmX and on the right for MalK.

Expression of *msmX* in *E. coli*

Previously, a recombinant version of MsmX (rMsmX) with a H_6-tag at the C-terminal was constructed and expressed in *E. coli* using the pET30 vector. The preliminary data showed that rMsmX was functional in *B. subtilis*, however the soluble amount of protein obtained when expressed *in E. coli* was low (Ferreira and Sá-Nogueira, unpublished results). When producing this protein at 37 ºC with a 4 hour induction a considerable amount of protein was produced, but almost all the protein was in the insoluble fraction.

The low levels of soluble protein could be due to: i) the protein is not being able to achieve its natural folding which is making the protein to be targeted to inclusion bodies, ii) although this protein is cytoplasmic it can interact with membrane proteins from *E. coli* or iii) because this type of proteins usually have a residual ATPase activity, the production of high amounts of protein could be disrupting the energetic balances in the cell. In order to improve solubility we tested different expression conditions. First, we co-expressed GroESL chaperones, encoded by the pT-GroE plasmid, together with MsmX, and secondly we reduced the protein synthesis rate, by lowering the growth temperature, and the IPTG (inducer) concentration (Figure 3.3).

The results indicate that both strategies improve solubility, however the reduction of protein synthesis seems to have a higher effect on protein solubility. Although the total amount of protein is lower in these conditions, the proportion of soluble protein is higher and should result in greater protein concentration after purification.

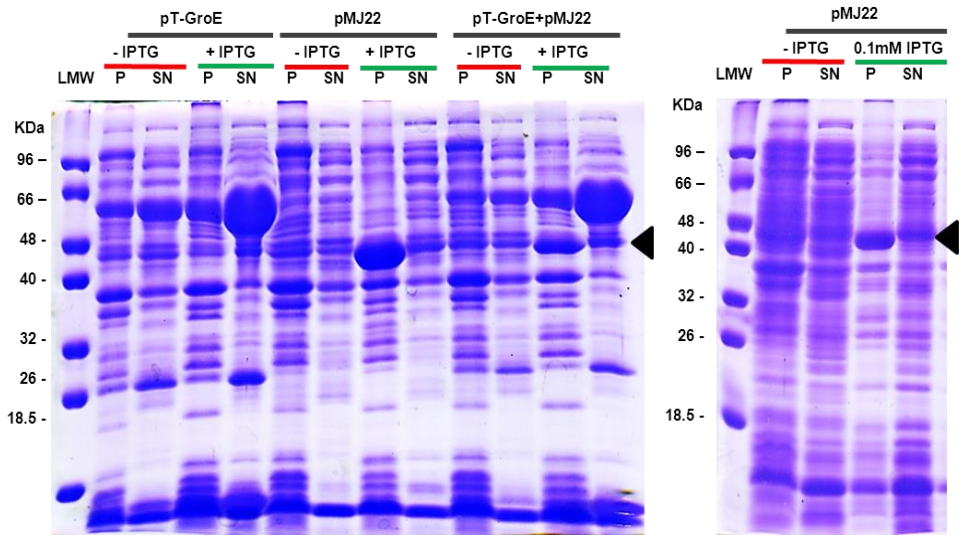


Figure 3.3 – Overproduction of rMsmX in *E. coli* BL21 (DE3). On the left are shown the cell extracts from co-expression of rMsmX and GroESL chaperones at 30 ºC for 4 h with 1 mM IPTG. On the right are displayed the cell extracts from expression of rMsmX at 16 ºC for 16 h with 0.1 mM IPTG. P – Insoluble Fraction; SN – Soluble Fraction; - IPTG – IPTG not added; +IPTG – IPTG added; LMW - Low Molecular Weight Protein Marker (NZYTech). Extracts run in a 12.5% SDS-PAGE, stained with Coomassie Blue. Soluble rMsmX is indicated with black triangle.

These results suggest that the protein insolubility may be derived from a defensive response of the cell against the energetic balances disruption induced by over-expression of a heterologous ATPase. Decreasing the rate of protein synthesis lead to smaller amounts of ATPase and, consequently, decreased the toxic effect in the cell. However, we cannot discard

possible interactions with membrane proteins because in the new conditions we still have more insoluble protein than soluble, which could be explained by these interactions. In conclusion, large-scale overproduction of rMsmX, for protein purification, was conducted by induction with 0.1 mM IPTG during 16 h at 16 ºC.

Purification of MsmX in *E. coli*

Usually, we perform protein purifications in a sodium phosphate buffer (PBS), however with this protein we decided to use Tris-HCl buffer because inorganic phosphate will interfere with the assay for ATPase activity determination. In the purification it was used a Ni-NTA column, as described in the Materials and Methods section, which exclusively binds to the H_6-Tag from rMsmX.

As shown in Figure 3.4, it was possible to obtain rMsmX with a high degree of purity, although some contaminant proteins are present. These persistent contaminants could be due to interactions of those proteins with exposed hydrophobic regions of rMsmX. Elution of rMsmX initiated at 200 mM of imidazole and persisted until 500 mM of imidazole. In the flow-through and the wash fraction we observe the rMsmX band, which represents loss of protein, in addition it is possible that some protein remained in the column since in the last elution fraction there was still protein. These observations indicate that it is necessary to optimize the purification protocol. Possible changes are the increase of Tris-HCl concentration and/or alter the number of fractions eluted with each imidazole concentration.

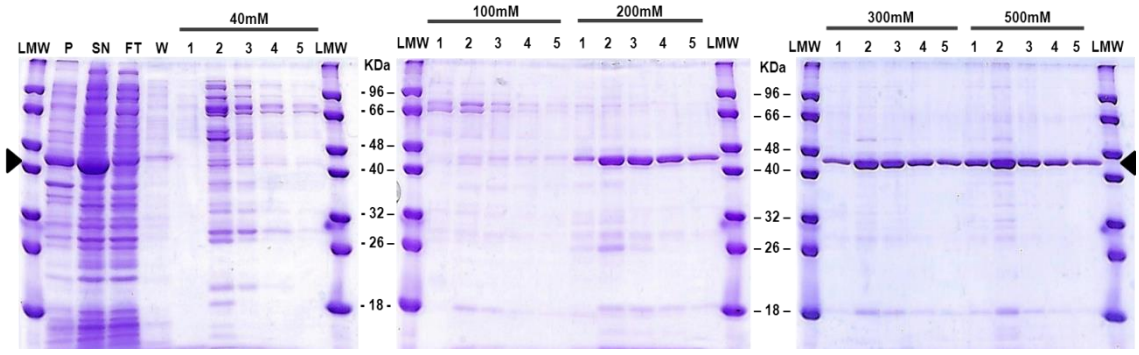


Figure 3.4 – Purification of rMsmX. rMsmX was overexpressed, by largescale induction, in *E. coli* BL21 (DE3) harbouring pMJ22 and resulting protein extracts were purified in Ni-NTA column. P and SN represent the non-purified insoluble and soluble fractions, respectively; FT represents unbound proteins from the soluble fraction (flow-through); W represents washed and weakly bound proteins; 40 mM (1 to 5), 100 mM (1 to 5), 200 mM (1 to 5), 300 mM (1 to 5) and 500 mM (1 to 5) represent fractions eluted with, respectively, 40 mM, 100 mM, 200 mM, 300 mM and 500 mM imidazole. LMW represents the Low Molecular Weight Protein Marker (NZYTech). Fractions run in a 12.5% SDS-PAGE, stained with Coomassie Blue. rMsmX is indicated with black triangle.

Analysis of the MsmX-Lys43Ala variant *in vivo* and *in vitro*

In the Walker A motif there is a highly conserved Lys at position 43 thought to be responsible for ATP stabilization in the active site (see Chapter I). To evaluate its impact in the ATPase activity of MsmX, we exchanged this residue to an Ala. The *msmX* allele present in both pAM4 and pMJ22 was mutated to evaluate the effects of such mutation *in vivo* and *in vitro*.

To study the effects *in vivo* we applied the same methodology used to test the MsmX homologs (Chapter II), and a strain with the *msmX(*Lys43Ala*)* mutant allele under the control of the pSpank(hy), inserted in *amyE* locus, was constructed (IQB675). The kinetic parameters of growth were calculated and are shown in Table 3.4.

The doubling time of IQB675 in the presence of glucose and arabinose are similar to the strain IQB673, which contains the wild-type *msmX* in the *amyE* locus. However, when IPTG was added to the medium in the presence of arabinotriose, while IQB673 is able to grow in these conditions, no growth was observed in strain IQB675 (Figure 3.5).

Table 3.4 – Growth of IQB673 and IQB675 in the presence of several saccharides as sole carbon/energy.

Carbon Source	Doubling Time (minutes)*	
	IQB673 (Δ*msmX*::*cat* Δ*amyE*::pSpank(hy)-*msmX*)	**IQB675** (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmX*(Lys43Ala))
Glucose 0.1%	58.38±2.24	59.55±0.58
Glucose 0.1% + IPTG 1mM	59.07±1.97	59.81±0.68
Arabinose 0.1%	78.47±3.47	77.21±1.68
Arabinose 0.1% + IPTG 1mM	77.53±2.95	76.80±1.10
Arabinotriose 0.1%	No Growth	No Growth
Arabinotriose 0.1% + IPTG 1mM	107.60±3.85	No Growth

* Cells were grown in CSK minimal medium (see Materials and Methods). Growth Kinetics parameters were determined and the results represent the average of three independent experiments, except the glucose and arabinose growths without IPTG, which were performed only twice (see Annexes for the complete data used in the determination of doubling time). IQB673 results are the same presented in table 5 and are shown here to facilitate comparison.

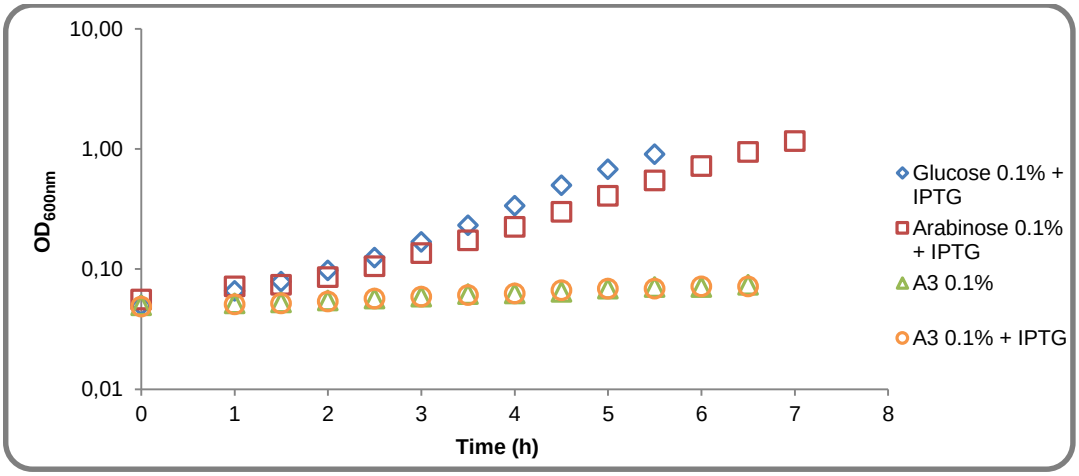


Figure 3.5 – Growth of IQB676 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmX*(Lys43Ala)) in CSK medium using glucose, arabinose and arabinotriose as the sole carbon and energy source.

This result confirms the importance of Lys43 in the functionality of MsmX. Most probably the ATPase is being produced by the cell and could interact with the transport system AraNPQ in the membrane. However, due to the mutation in Lys43, the ATP hydrolysis necessary for the transportation of arabinotriose is impaired. The MsmX Lys43 residue is the conserved Lys present in the Walker A motif and corresponds to the amino acid Lys42 of MalK. This amino acid is responsible for direct binding with ATP phosphate groups between which the hydrolysis occurs (β- and γ-phosphate). Without this amino acid the phosphodiester bond may not be placed in the right position relative to the His in the H-loop, making impossible to break it. This way the transporter remains in the outward open position. Benaroudj *et al.* (2013) mutated the same Lys of the ATPase LIC12079 from *Leptospira interrogans* into an Ala (Lys45Ala), which resulted in the loss of ATPase activity.

For the characterization *in vitro* of the rMsmX(Lys43Ala) variant, large-scale overexpression of rMsmX(Lys43Ala) was conducted in the same conditions used for the production of rMsmX. Cell extracts analysis by SDS-PAGE showed that the mutation resulted in an increase of protein solubility (Figure 3.6). This observation supports the possible existence of energetic balances disruption caused by the overexpression of the rMsmX form discussed above.

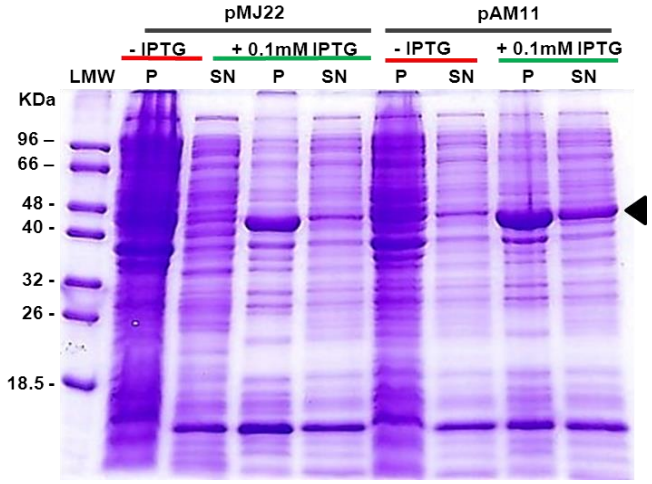


Figure 3.6 - Overproduction of rMsmX(Lys43Ala) in *E. coli* BL21 (DE3). Cell extracts from expression of rMsmX(LYS43ALA) at 16 ºC for 16 h with 0.1 mM IPTG. P – Insoluble Fraction; SN – Soluble Fraction; - IPTG – IPTG not added; +IPTG – IPTG added; LMW - Low Molecular Weight Protein Marker (NZYTech). Extracts run in a 12.5% SDS-PAGE, stained with Coomassie Blue. Soluble rMsmX(Lys43Ala) is indicated with black triangle.

The methodology used to purify this rMsmX variant was the same applied in the purification of the rMsmX, however we experience some complications in the process. The protein concentrations were too high for the buffer used in the purification, resulting in total precipitation inside the column or immediately after recovery from the column. Due to these results we changed the elution buffer from Tris-HCl to PBS. When using the PBS buffer protein precipitation did not occur and the results are shown in Figure 3.7. Additionally it was added glycerol to the buffer to improve protein stability.

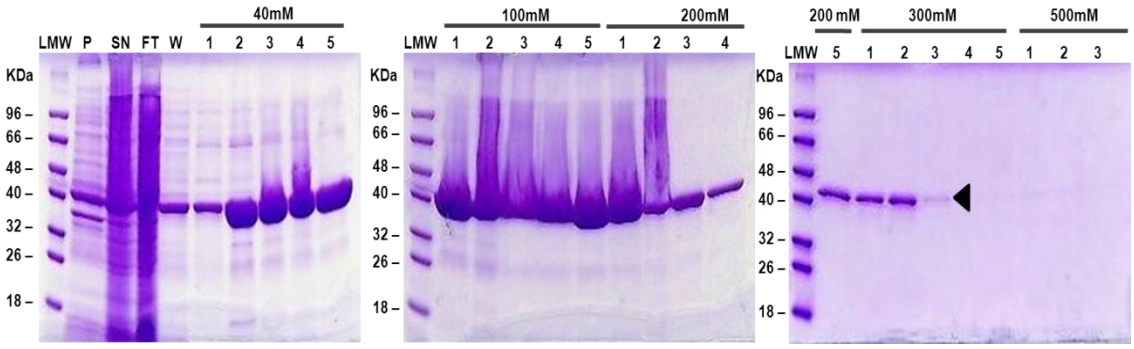


Figure 3.7 - Purification of rMsmX(Lys43Ala). rMsmX(Lys43Ala) was overexpressed, by largescale induction, in *E. coli* BL21 (DE3) harbouring pMJ22 and resulting protein extracts were purified in Ni-NTA column. P and SN represent the non-purified insoluble and soluble fractions, respectively; FT represents unboundproteins from the soluble fraction (flow-through); W represents washed and weakly bound proteins; 40 mM (1 to 5), 100 mM (1 to 5), 200 mM (1 to 5), 300 mM (1 to 5) and 500 mM (1 to 5) represent fractions eluted with, respectively, 40 mM, 100 mM, 200 mM, 300 mM and 500 mM imidazole. LMW represents the Low Molecular Weight Protein Marker (NZYTech). Fractions run in a 12.5% SDS-PAGE, stained with Coomassie Blue. rMsmX(Lys43Ala) indicated with black triangle.

Preliminary assays were performed to determine the ATPase activity of both rMsmX and rMsmX(Lys43Ala). The assays were made with 5µg of protein in TrisHCl buffer, supplemented with 10 mM $MgCl_2$ and 0.1 mM ATP. rMsmX displayed ATPase capability (Figure 3.8), however the ATP rate of ATP hydrolysis is 10 times lower than the activity calculated for other proteins of the same family (Benaroudj *et al.*, 2013) and optimization of assays conditions is required to obtain maximum activity values. The results obtained with the rMsmX(Lys43Ala) showed a drastic decrease in ATP hydrolysis, which confirms that the results obtained *in vivo* with this protein resulted from ATPase activity decrease. Nevertheless, more independently assays in these conditions must be conducted to confirm these results.

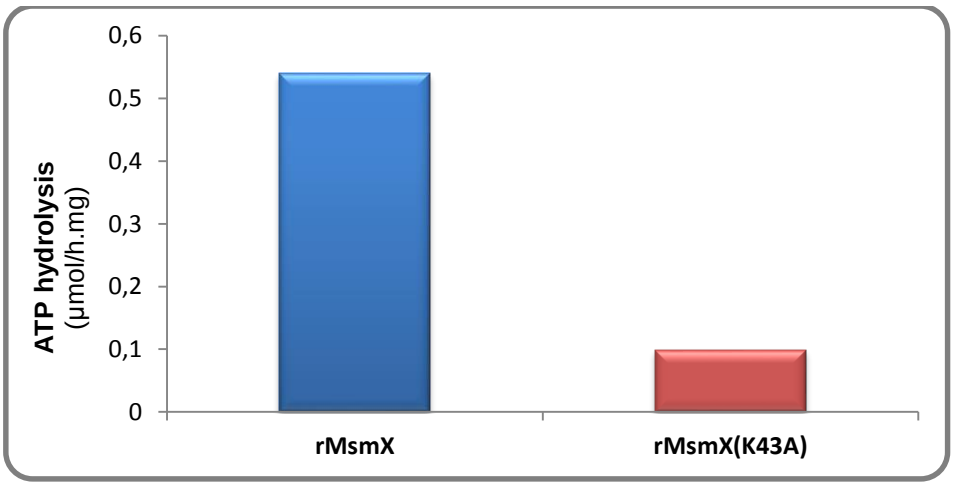


Figure 3.8 - ATP hydrolysis by rMsmX and rMsmX(Lys43Ala). ATP hydrolysis was determined by measuring the production of inorganic phosphate, as described in Materials and Methods. 5 µg of protein were incubated with 0.1 mM ATP and 10 mM $MgCl_2$. Only one assay was performed for each protein.

Chapter 4

Concluding Remarks and Future Perspectives

4. Concluding Remarks and Future Perspectives

The introduction of antibiotics in the treatment of infectious disease had a great impact in medicine saving countless lives. However, the use of antibiotics creates a phenomenon called antibiotic pressure which is the emergence of resistant pathogens. Although this side-effect was predicted by Fleming, over the years there has been a negligent and abusive use of antibiotics (French, 2010). This led to development of multidrug-resistance pathogens like *Clostridium difficile*, *Pseudomonas aeruginosas*, *S. aureus* or *S. pneumoniae*. Also, in some cases, these multidrug-resistance pathogens have developed more virulent traits and acquired enhanced transmissibility (Davies and Davies, 2010).

We have reached a point where we must direct our attentions to the discovery of new antimicrobials and therapeutic targets. The carbohydrate metabolism of bacteria plays an important role in their survival and capacity for colonization and pathogenesis (Buckwalter and King, 2012). One potentially attractive target to act against these pathogens is the multitask ATPases recently discovered whose inactivation leads to inability to utilize multiple carbon sources. This could have a great impact in the pathogen adaptability to the host niche or even kill the bacterial by starvation. So a better understanding of these proteins functionality, structure and regulation may have a huge impact in combat against multidrug-resistance strains.

In the work we presented, it was shown that MsmX homologs in other pathogenic species are able to substitute MsmX in the cell by restoring arabinooligosaccharides uptake in a *B. subtilis msmX*-null mutant strain. The complementation degree varied depending on the protein introduced in the strain genome. The *B. thuringiensis* protein showed to be more capable of fully complement MsmX absence while *S. aureus* and *S. pneumoniae* proteins were only able to partially complement it. Whether this resulted from differences in amino acid sequence that affected the interaction with AraNPQ complex or rather from differences in the amount of protein produced by the cell still remains unknown. Thus, here we demonstrated that *B. subtilis* and MsmX could be used as a model to better comprehend this sharing of energy-generating components between distinct ABC sugar importers.

Additionally, we proved that MsmX has ATPase activity and that the conserved Lys 43 present in the Walker A motif is essential for this activity both *in vivo* and *in vitro*. Moreover, modifications in the N- and C-terminal primary sequence do not affect MsmX functionality or its ability to interact with the AraNPQ transporter. In this work we also achieved important results towards the improvement of MsmX production and purification processes, which is crucial for crystallography studies and biochemical characterization of MsmX that will provide important clues or even answers for the main questions arising from the discovery of multitask ATPases.

The work presented here is just the tip of the iceberg regarding to the knowledge about these multitask ATPases and the main forces that drove bacteria to develop these type of sugar uptake mechanisms/regulation. First, we need to test the proteins used in this work with the new system developed because it will allow us to conclude if the differences in the complementation results observed are due to dissimilarities in the proteins or in the amount produced in the cell. In addition, we must test the YurJ ATPase from *B. subtilis,* which displays 58% identity with MsmX and is suggested to be multitask, as well as other *B. subtilis* ATPases thought to be exclusive to one transporter. The conjunction of all these results may indicate which MsmX residues are important for this multi-transporters interaction capability. Simultaneously, the initiation of crystallography studies is of major importance to provide insights into this mechanism. Finally, it would be interesting to ascertain if MsmX suffers any post-translational regulation like MalK does (Biemans-Oldehinkel *et al.*, 2006). The discovery of possible regulatory mechanisms that control MsmX activity and their understanding may also reveal MsmX regions for possible targeting.

Chapter 5

References

5. References

Anagnostopoulos, C., Spizizen, J. (1961). Requirements for Transformation in *Bacillus subtilis. J. Bacteriol.* **81**(5): 741–746.

Beauregard, P.B., Chai, Y., Vlamakis, H., Losick, R., Kolter, R. (2013). *Bacillus subtilis* biofilm induction by plant polysaccharides. *Proc. Natl. Acad. Sci. U. S. A.* **110**(17): E1621–30.

Benaroudj, N., Saul, F., Bellalou, J., Miras, I., Weber, P., Bondet, V., Murray, G.L., Adler, B., Ristow, P., Louvel, H., Haouz, A., Picardeau, M. (2013). Structural and Functional Characterization of an Orphan ATP-Binding Cassette ATPase Involved in Manganese Utilization and Tolerance in *Leptospira* spp. *J. Bacteriol.* **195**(24): 5583–91.

Biemans-Oldehinkel, E., Doeven, M.K., Poolman, B. (2006). ABC transporter architecture and regulatory roles of accessory domains. *FEBS Lett.* **580**(4): 1023–35.

Britton, R.A., Eichenberger, P., Gonzalez-Pastor, J.E., Fawcett, P., Monson, R., Losick, R., Grossman, A.D. (2002). Genome-Wide Analysis of the Stationary-Phase Sigma Factor (Sigma-H) Regulon of *Bacillus subtilis. J. Bacteriol.* **184**(17): 4881–4890.

Bryson, K., McGuffin, L.J., Marsden, R.L., Ward, J.J., Sodhi, J.S., Jones, D.T. (2005). Protein structure prediction servers at University College London.. *Nucleic Acids Res.* . **33** (suppl 2): W36–W38.

Buckwalter, C.M., King, S.J. (2012). Pneumococcal carbohydrate transport: food for thought. *Trends Microbiol.* **20**(11): 517–22.

Chai, Y., Beauregard, P.B., Vlamakis, H., Losick, R., Kolter, R. (2012). Galactose Metabolism Plays a Crucial Role in Biofilm Formation by *Bacillus subtilis. MBio*. **3**(4): 1–10.

Chen, J., Lu, G., Lin, J., Davidson, A.L., Quiocho, F. a. (2003). A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. *Mol. Cell*. **12**(3): 651–61.

Cui, J., Davidson, A.L. (2011). ABC solute importers in bacteria. *Biochem. Soc. Essays Biochem*. **50**: 85–99.

Davidson, A.L., Dassa, E., Orelle, C., Chen, J. (2008). Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol. Mol. Biol. Rev.* **72**(2): 317–64.

Davies, J., Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* **74**(3): 417–33.

Debarbouille, M., Arnaud, M., Fouet, A., Klier, A., Rapoport, G. (1990). The *sacT* gene regulating the *sacPA* operon in *Bacillus subtilis* shares strong homology with transcriptional antiterminators. *J. Bacteriol.* **172**(7): 3966–73.

Eitinger, T., Rodionov, D. a, Grote, M., Schneider, E. (2010). Canonical and ECF-type ATP-binding cassette importers in prokaryotes: diversity in modular organization and cellular functions. *FEMS Microbiol. Rev.* **35**(1): 3–67.

Ferreira, M.J., de Sá-Nogueira, I. (2010). A multitask ATPase serving different ABC-type sugar importers in *Bacillus subtilis* . *J. Bacteriol.* **192**(20): 5312–8.

French, G.L. (2010). The continuing crisis in antibiotic resistance. *Int. J. Antimicrob. Agents*. **36 Suppl 3**: S3–7.

Gerber, S., Comellas-Bigler, M., Goetz, B. a, Locher, K.P. (2008). Structural basis of trans-inhibition in a molybdate/tungstate ABC transporter. *Science*. **321**(5886): 246–50.

Higgins, C.F. (1992). ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol.* **8**: 67–113.

Higgins, C.F. (2001). ABC transporters: physiology, structure and mechanism - an overview. *Res. Microbiol.* **152**(3-4): 205–10.

Higgins, C.F., Ames, G.F. (1981). Two periplasmic transport proteins which interact with a common membrane receptor show extensive homology: complete nucleotide sequences. *Proc. Natl. Acad. Sci. U. S. A.* **78**(10): 6038–42.

Hollenstein, K., Frei, D.C., Locher, K.P. (2007). Structure of an ABC transporter in complex with its binding protein. *Nature*. **446**(7132): 213–6.

Horn, C., Bremer, E., Schmitt, L. (2003). Nucleotide Dependent Monomer/Dimer Equilibrium of OpuAA, the Nucleotide-binding Protein of the Osmotically Regulated ABC Transporter OpuA from *Bacillus subtilis. J. Mol. Biol.* **334**(3): 403–419.

Hung, L.W., Wang, I.X., Nikaido, K., Liu, P.Q., Ames, G.F., Kim, S.H. (1998). Crystal structure of the ATP-binding subunit of an ABC transporter. *Nature*. **396**(6712): 703–7.

Hurtubise, Y., Shareck, F., Kluepfel, D., Morosoli, R. (1995). A cellulase/xylanase-negative mutant of *Streptomyces lividans* 1326 defective in cellobiose and xylobiose uptake is mutated in a gene encoding a protein homologous to ATP-binding proteins. *Mol. Microbiol.* **17**(2): 367–377.

Jaehme, M., Slotboom, D.J. (2014). Diversity of membrane transport proteins for vitamins in bacteria and archaea. *Biochim. Biophys. Acta*.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**(2): 195–202.

Kampers, T., Schlösser, A., Schrempf, H. (1997). The *Streptomyces* ATP-binding component MsiK assists in cellobiose and maltose transport. *J. Bacteriol.* **179**(6): 2092–2095.

Karpowich, N., Martsinkevich, O., Millen, L., Yuan, Y.R., Dai, P.L., MacVey, K., Thomas, P.J., Hunt, J.F. (2001). Crystal structures of the MJ1267 ATP binding cassette reveal an induced-fit effect at the ATPase active site of an ABC transporter. *Structure*. **9**(7): 571–86.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, a M., Alloni, G., Azevedo, V., Bertero, M.G., Bessières, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruschi, C. V, Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Danchin, A. (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis* . *Nature*. **390**(6657): 249–56.

Linke, C.M., Woodiga, S. a, Meyers, D.J., Buckwalter, C.M., Salhi, H.E., King, S.J. (2013). The ABC transporter encoded at the pneumococcal fructooligosaccharide utilization locus determines the ability to utilize long- and short-chain fructooligosaccharides. *J. Bacteriol.* **195**(5): 1031–41.

Marion, C., Aten, A.E., Woodiga, S. a, King, S.J. (2011). Identification of an ATPase, MsmK, which energizes multiple carbohydrate ABC transporters in *Streptococcus pneumoniae* . *Infect. Immun.* **79**(10): 4193–200.

Martin, I., Débarbouillé, M., Ferrari, E., Klier, A., Rapoport, G. (1987). Characterization of the levanase gene of *Bacillus subtilis* which shows homology to yeast invertase*. Mol. Gen. Genet. MGG*. **208**: 177–184.

McGuffin, L.J., Bryson, K., Jones, D.T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*. **16**(4): 404–5.

Michna, R.H., Commichau, F.M., Tödter, D., Zschiedrich, C.P., Stülke, J. (2014). SubtiWiki-a database for the model organism Bacillus subtilis that links pathway, interaction and expression information. *Nucleic Acids Res.* **42**(Database issue): D692–8.

Miller, J.H. (1972). Experiments in Molecular Genetics.. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Oldham, M.L., Khare, D., Quiocho, F. a, Davidson, A.L., Chen, J. (2007). Crystal structure of a catalytic intermediate of the maltose transporter. *Nature*. **450**(7169): 515–21.

Pascal, M., Kunst, F., Lepesant, J. A. and Dedonder, R. (1971). Characterization of two sucrase activities in *Bacillus subtilis* Marburg. *Biochimie*. **53**: 1059–1066.

Quentin, Y., Fichant, G., Denizot, F. (1999). Inventory, assembly and analysis of *Bacillus subtilis* ABC transport systems. *J. Mol. Biol.* **287**(3): 467–84.

Quisel, J.D., Burkholder, W.F., Grossman, A.D. (2001). *In Vivo* Effects of Sporulation Kinases on Mutant Spo0A Proteins in *Bacillus subtilis* . *J. Bacteriol.* **183**(22): 6573–6578.

Rees, D.C., Johnson, E., Lewinson, O. (2009). ABC transporters: the power to change. *Nat. Rev. Mol. Cell Biol.* **10**(3): 218–27.

Russell, R.R., Aduse-Opoku, J., Sutcliffe, I.C., Tao, L., Ferretti, J.J. (1992). A binding protein-dependent transport system in *Streptococcus mutans* responsible for multiple sugar metabolism. *J. Biol. Chem.* **267**(7): 4631–7.

Saier, M.H., Goldman, S.R., Maile, R.R., Moreno, M.S., Weyler, W., Yang, N., Paulsen, I.T. (2002). Transport capabilities encoded within the *Bacillus subtilis* genome. *J. Mol. Microbiol. Biotechnol.* **4**(1): 37–67.

Saito, A., Fujii, T., Shinya, T., Shibuya, N., Ando, A., Miyashita, K. (2008). The *msiK* gene, encoding the ATP-hydrolysing component of N,N'-diacetylchitobiose ABC transporters, is essential for induction of chitinase production in *Streptomyces coelicolor* A3(2). *Microbiology*. **154**(Pt 11): 3358–65.

Salis, H.M. (2011). The ribosome binding site calculator. *Methods Enzymol.* **498**: 19–42.

Salis, H.M., Mirsky, E. a, Voigt, C. a. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**(10): 946–50.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular cloning: A laboratory manual., 2nd ed.*, Cell*. Cold Spring Harbor: Cold Spring Harbor Laboratory, NY.

Schlösser, a. (2000). MsiK-dependent trehalose uptake in *Streptomyces reticuli* . *FEMS Microbiol. Lett.* **184**(2): 187–92.

Schneider, E., Hunke, S. (1998). ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolyzing subunits/domains. *FEMS Microbiol. Rev.* **22**(1): 1–20.

Schönert, S., Seitz, S., Krafft, H., Feuerbaum, E.-A., Andernach, I., Witz, G., Dahl, M.K. (2006). Maltose and maltodextrin utilization by *Bacillus subtilis* . *J. Bacteriol.* **188**(11): 3911–22.

Seyffer, F., Tampé, R. (2014). ABC transporters in adaptive immunity. *Biochim. Biophys. Acta*.

Silva, Z., Sampaio, M.-M., Henne, A., Böhm, A., Gutzat, R., Boos, W., da Costa, M.S., Santos, H. (2005). The high-affinity maltose/trehalose ABC transporter in the extremely thermophilic bacterium *Thermus thermophilus* HB27 also recognizes sucrose and palatinose. *J. Bacteriol.* **187**(4): 1210–8.

Smith, P.C., Karpowich, N., Millen, L., Moody, J.E., Rosen, J., Thomas, P.J., Hunt, J.F. (2002). ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol. Cell*. **10**(1): 139–49.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. and Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes*. Meth. Enzym.* **185**: 60–89.

Ter Beek, J., Guskov, A., Slotboom, D.J. (2014). Structural diversity of ABC transporters. *J. Gen. Physiol.* **143**(4): 419–35.

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

Verdon, G., Albers, S.-V., van Oosterwijk, N., Dijkstra, B.W., Driessen, A.J.M., Thunnissen, A.-M.W.H. (2003). Formation of the Productive ATP-Mg^{2+}-bound Dimer of GlcV, an ABC-ATPase from *Sulfolobus solfataricus. J. Mol. Biol.* **334**(2): 255–267.

Webb, A.J., Homer, K. a, Hosie, A.H.F. (2008). Two closely related ABC transporters in *Streptococcus mutans* are involved in disaccharide and/or oligosaccharide uptake. *J. Bacteriol.* **190**(1): 168–78.

Yasukawa, T., Kanei-Ishii, C., Maekawa, T., Fujimoto, J., Yamamoto, T., Ishii, S. (1995). Increase of Solubility of Foreign Proteins in *Escherichia coli* by Coproduction of the Bacterial Thioredoxin*.. J. Biol. Chem.* **270**(43): 25328–25331.

Yoshida, K., Ishio, I., Nagakawa, E., Yamamoto, Y., Yamamoto, M., Fujita, Y. (2000). Systematic study of gene expression and transcription organization in the *gntZ-ywaA* region of the *Bacillus subtilis* genome. *Microbiology*. **146 (Pt 3**: 573–9.

Yoshida, K., Shindo, K., Sano, H., Seki, S., Fujimura, M., Yanai, N., Miwa, Y., Fujita, Y. (1996). Sequencing of a 65 kb region of the *Bacillus subtilis* genome containing the lic and cel loci, and creation of a 177 kb contig covering the *gnt-sacXY* region*.. Microbiology*. **142**(11): 3113–3123.

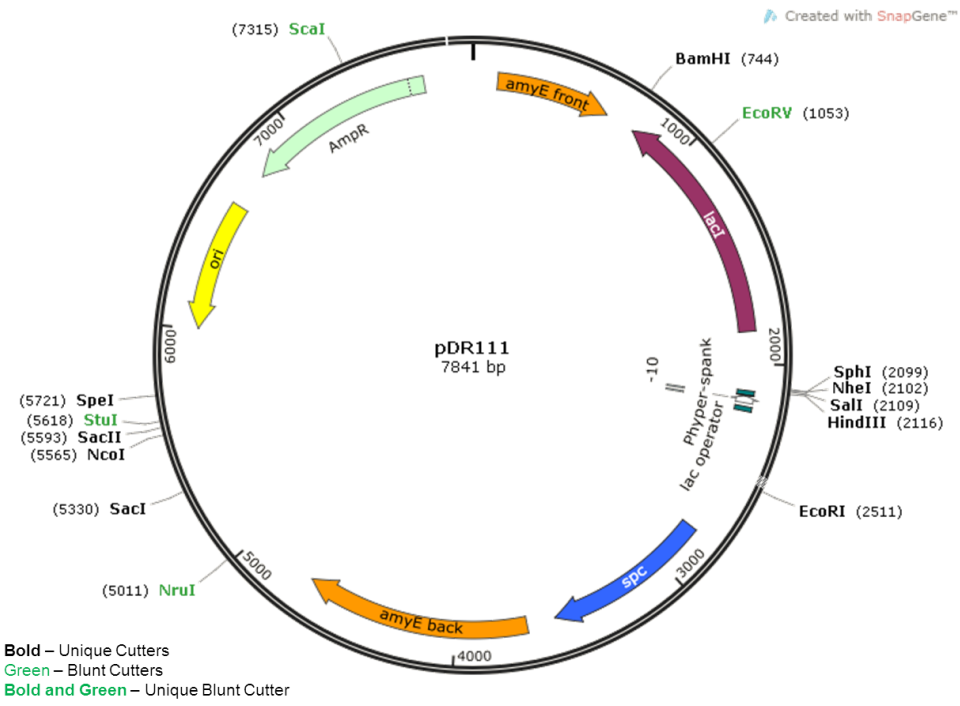
Zoghbi, M.E., Fuson, K.L., Sutton, R.B., Altenberg, G. a. (2012). Kinetics of the association/dissociation cycle of an ATP-binding cassette nucleotide-binding domain. *J. Biol. Chem.* **287**(6): 4157–64.

Chapter 6

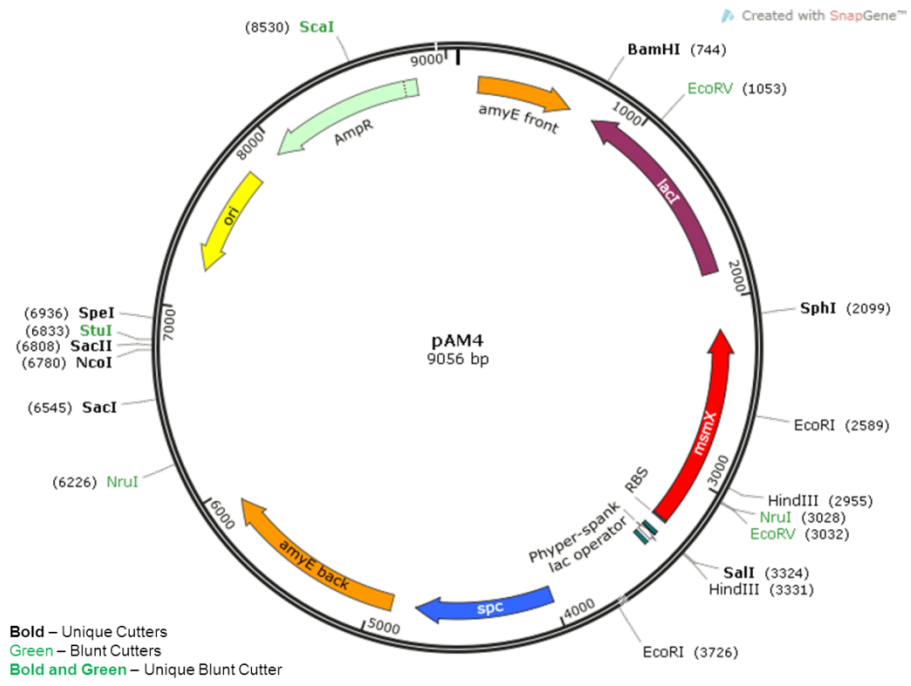
Appendices

6. Appendices

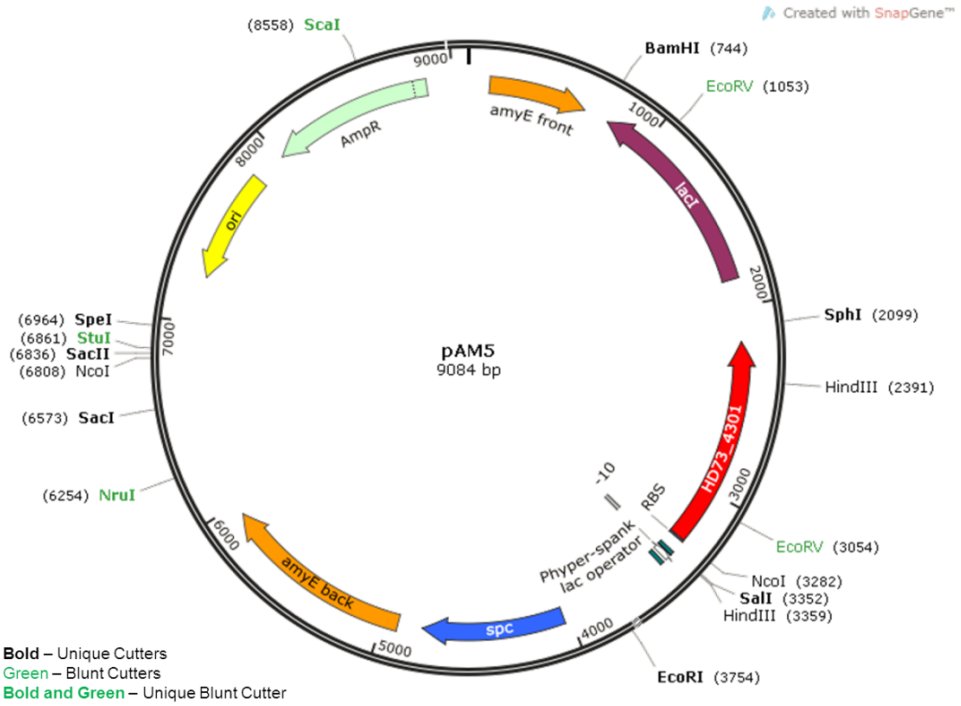
Appendix 6.1 - Map of pDR111 (gift from David Rudner, Harvard University**.**



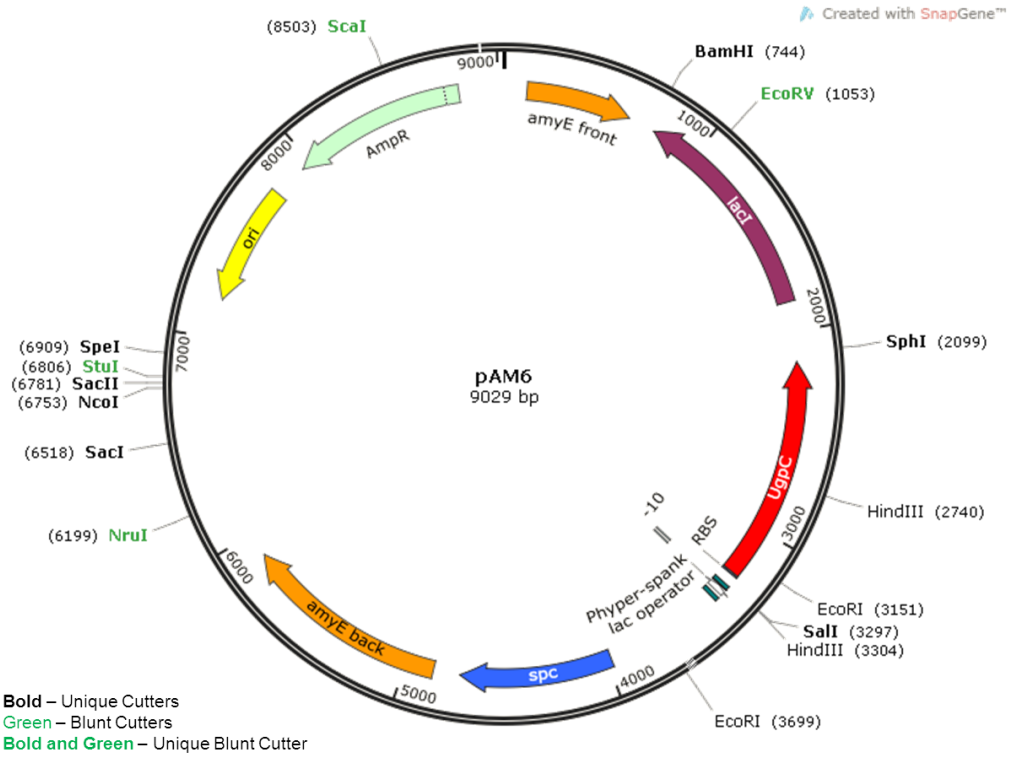
Appendix 6.2 - Map of pAM4**.** Insertion of a 1224bp fragment containing *msmX* gene (amplified with primers ARA741 and ARA742 and digested with *Sal*I and *Sph*I) between the *Sal*I and *Sph*I sites of pDR111.



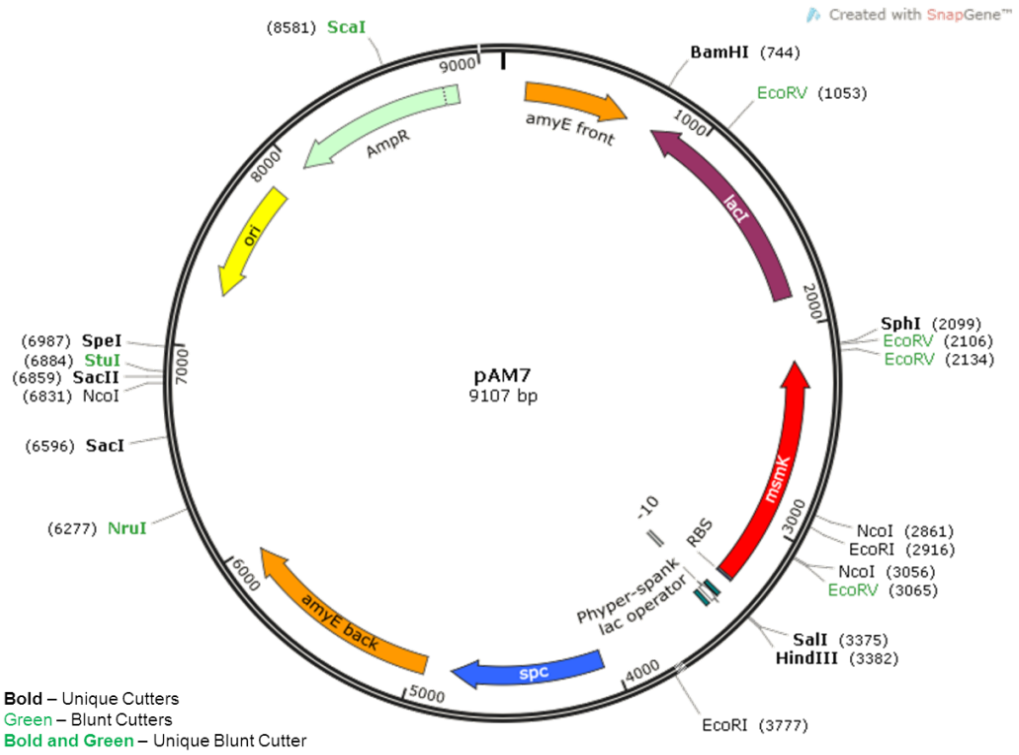
Appendix 6.3 - Map of pAM5. Insertion of a 1272bp fragment containing *HD73_4301* gene, from *B. thuringiensis* serovar kurstaki str. HD73 (amplified with the primers ARA744 and ARA745 and digested with *Sal*I and *Sph*I) between the *Sal*I and *Sph*I sites of pDR111.



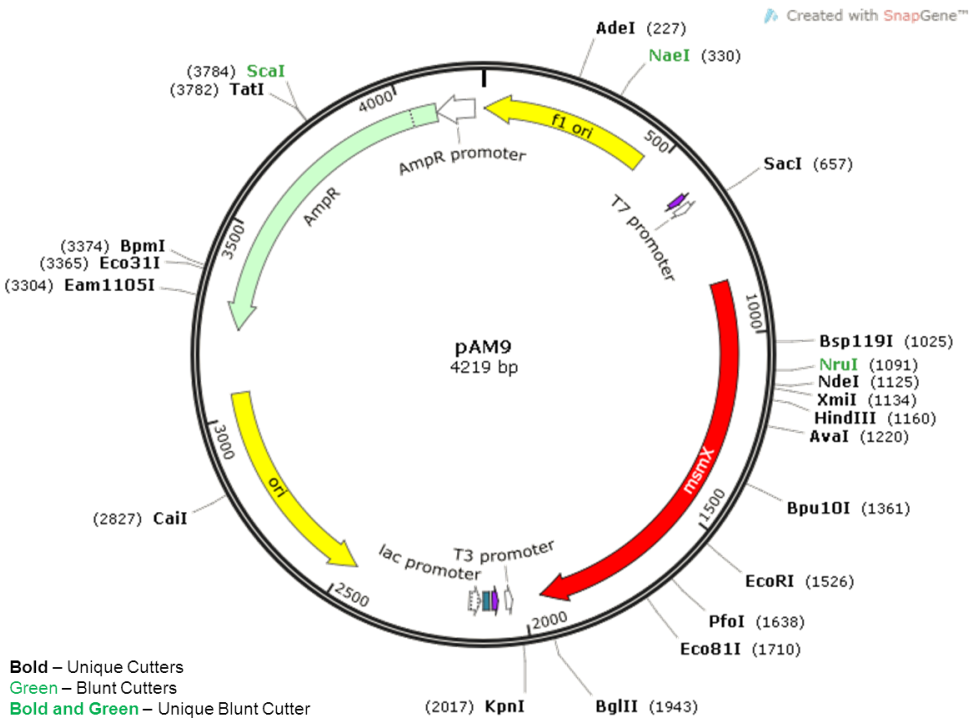
Appendix 6.4 - Map of pAM6. Insertion of a 1215bp fragment containing *ugpC*, from *S. aureus subsp. aureus* ST398, (amplified with the primers ARA746 and ARA747 and digested with *Sal*I and *Sph*I) between the *Sal*I and *Sph*I sites of pDR111.



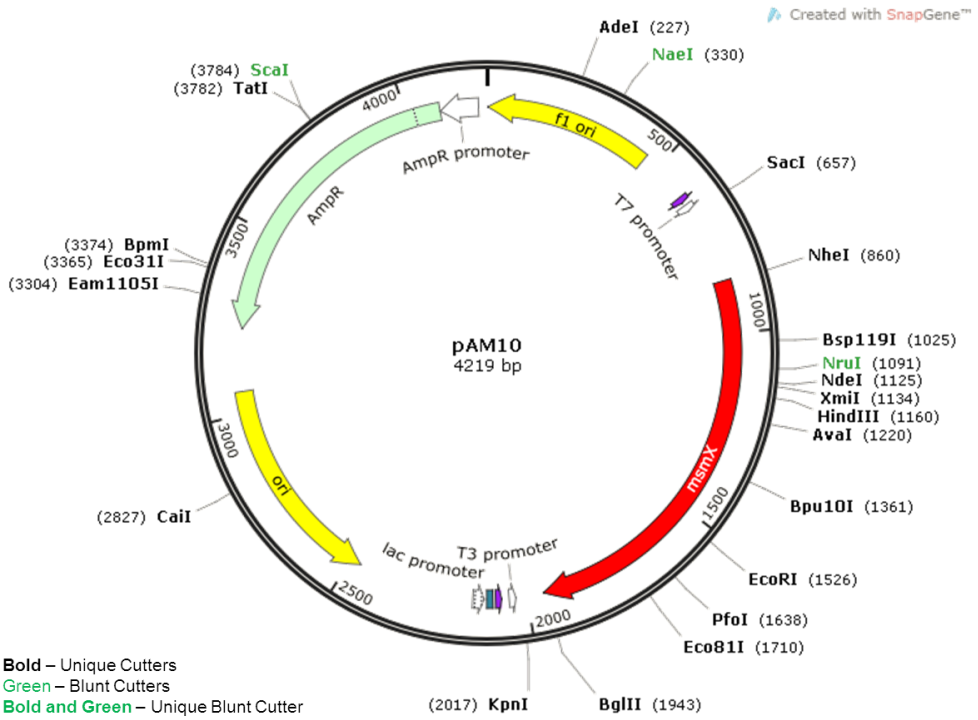
Appendix 6.5 - Map of pAM7. Insertion of a 1295bp fragment containing *msmK,* from *S. pneumoniae* TIGR4, (amplified with the primers ARA748 and ARA749 and digested with *Sal*I and *Sph*I) between the *Sal*I and *Sph*I sites of pDR111.



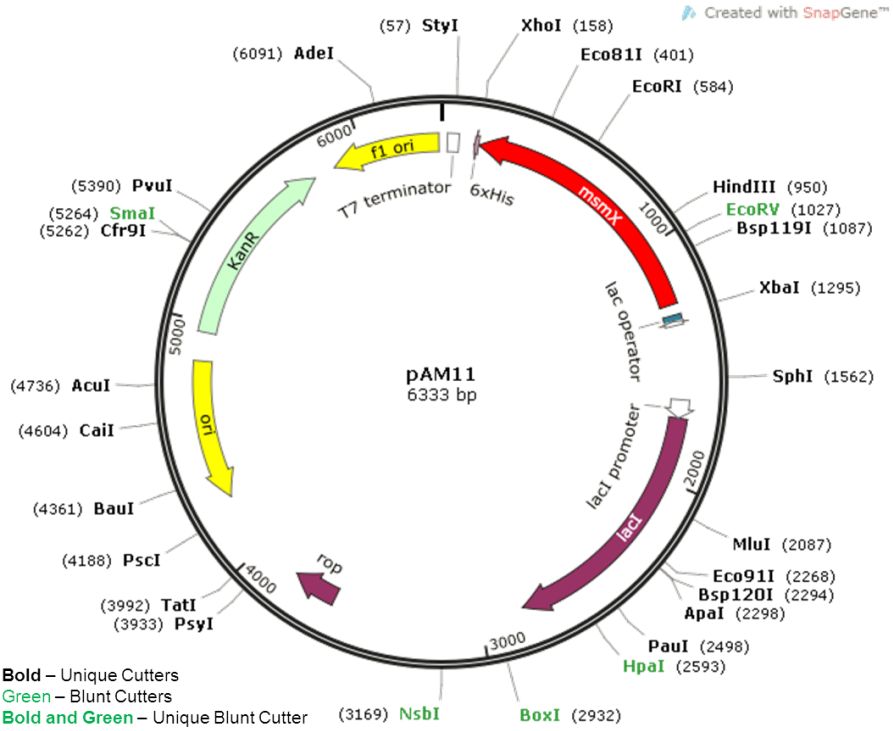
Appendix 6.6 - Map of pAM9. Site-directed mutagenesis of pMJ1 using the mutagenic primers set ARA769 and ARA770. Mutagenesis generated a unique BglII restriction site in the 4th nucleotide, downstream from the start codon.



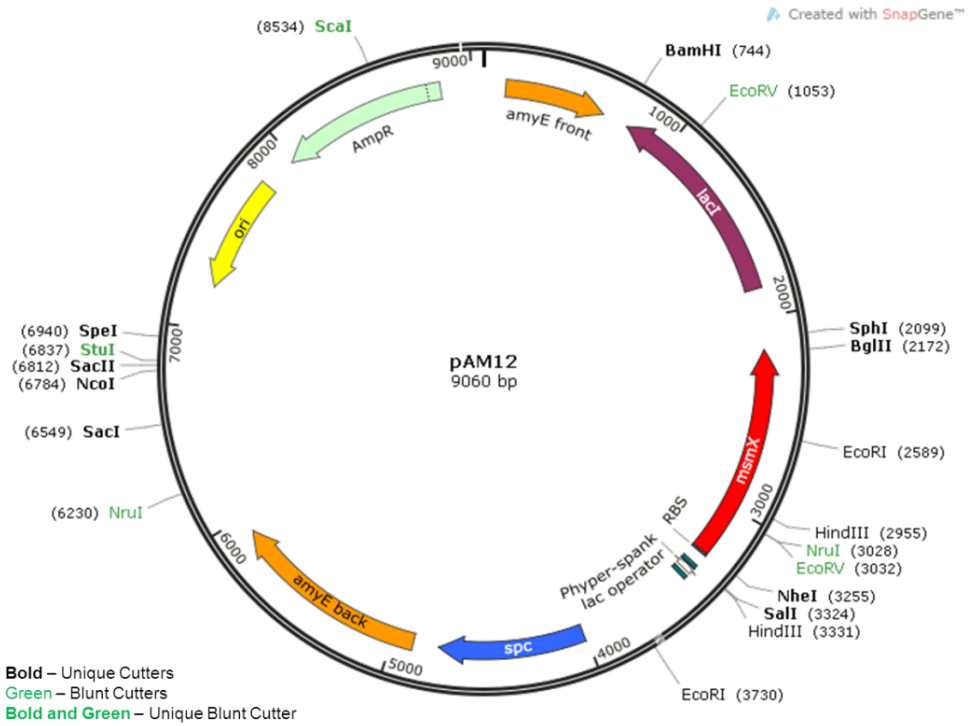
Appendix 6.7 - Map of pAM10. Site-directed mutagenesis of pAM9 using the mutagenic primers set ARA771 and ARA772. Mutagenesis generated a unique *Nhe*I restriction site in the 11th nucleotide, upstream from the stop codon.



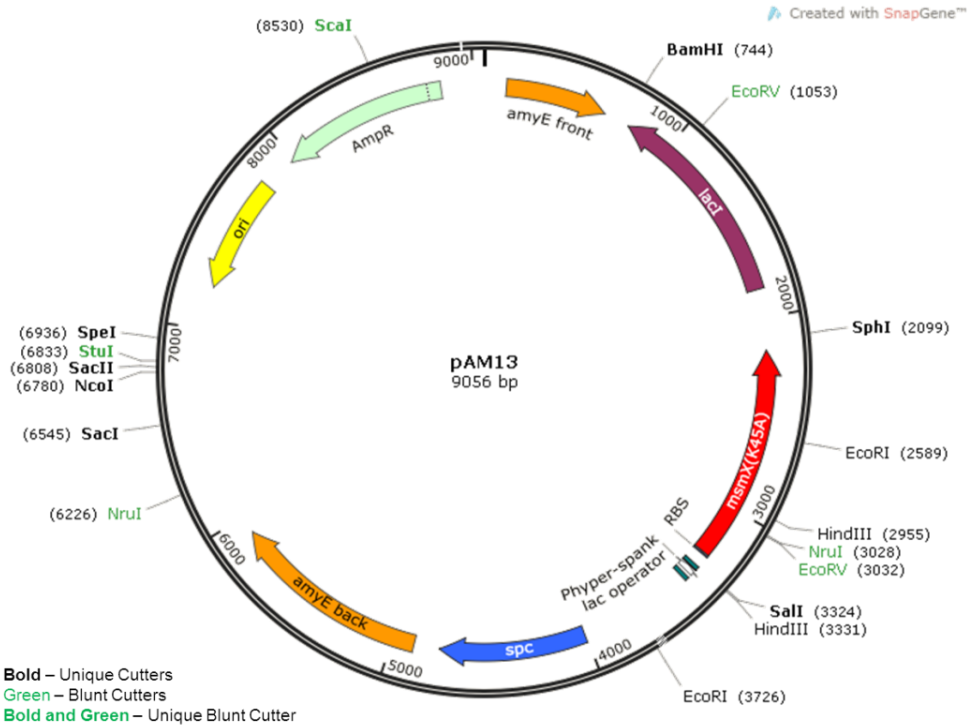
Appendix 6.8 - Map of pAM11. Site-directed mutagenesis of pMJ22 using the mutagenic primers set ARA757 and ARA758. This pair of primers allowed the mutation of codon AAA (lysine at position 45) to GCA (alanine).



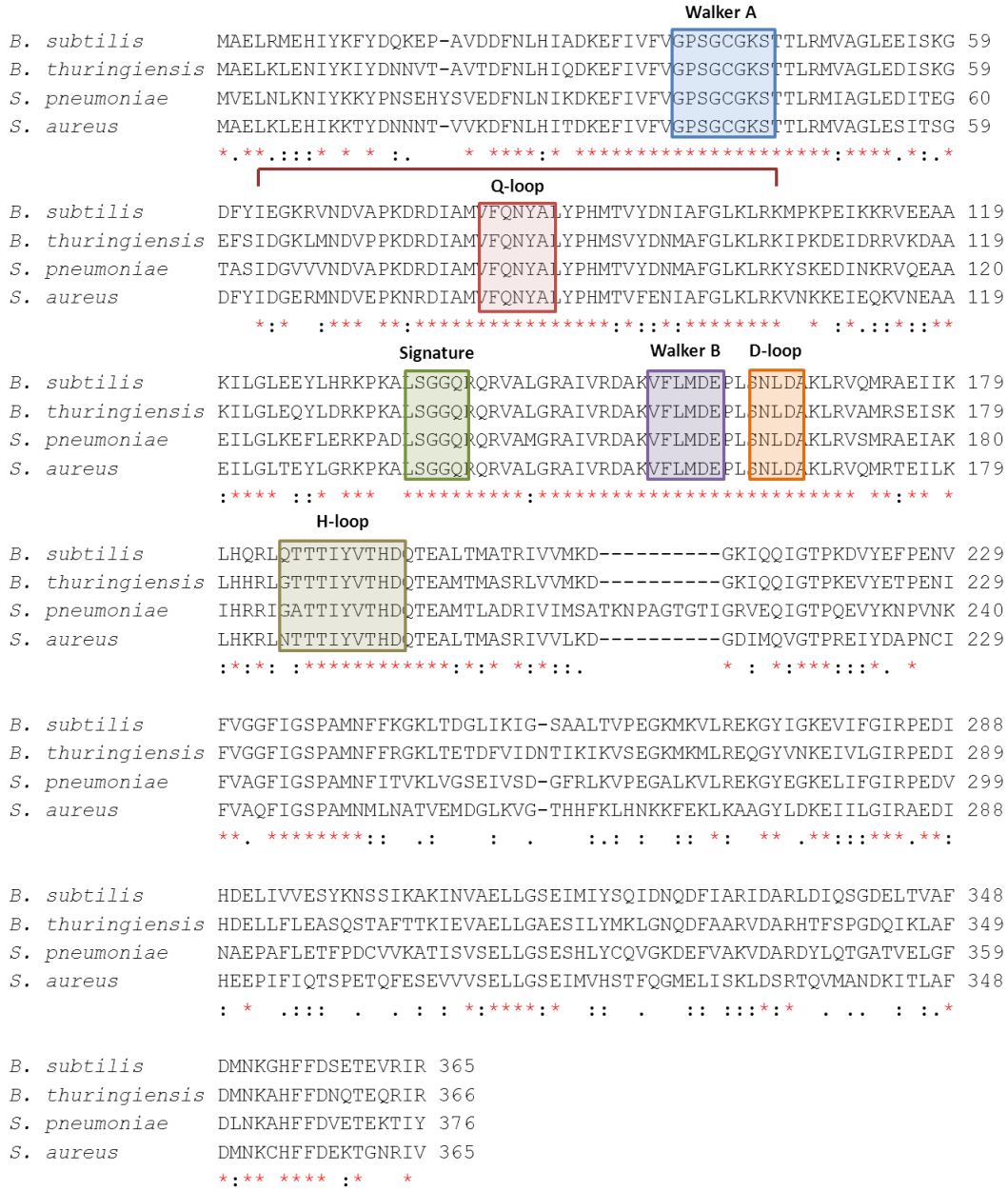
Appendix 6.9 - Map of pAM12. Insertion of a 1224bp fragment containing *msmX* gene from pAM10 (amplified with primers ARA741 and ARA742 and digested with *Sal*I and *Sph*I) between the *Sal*I and *Sph*I sites of pDR111.



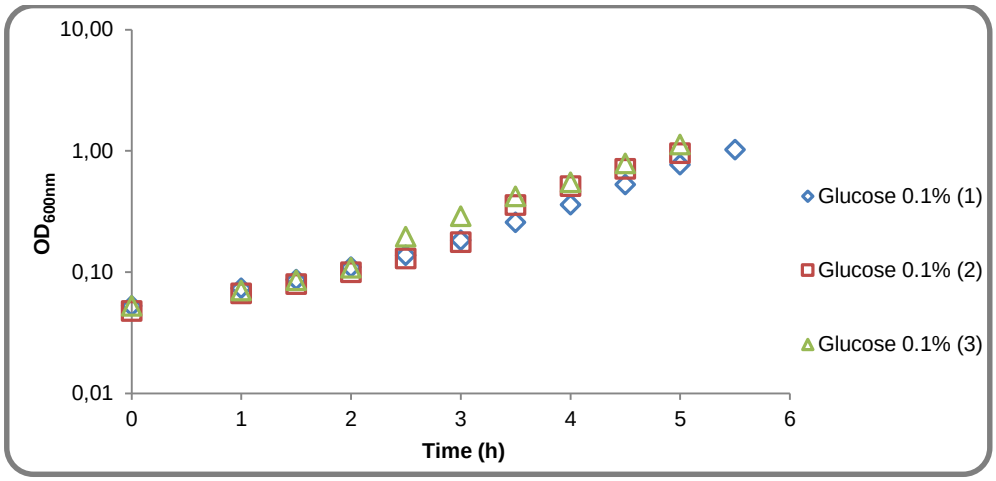
Appendix 6.10 - Map of pAM13. Site-directed mutagenesis of pAM4 using the mutagenic primers set ARA757 and ARA758. This pair of primers allowed the mutation of codon AAA (lysine at position 43) to GCA (alanine).



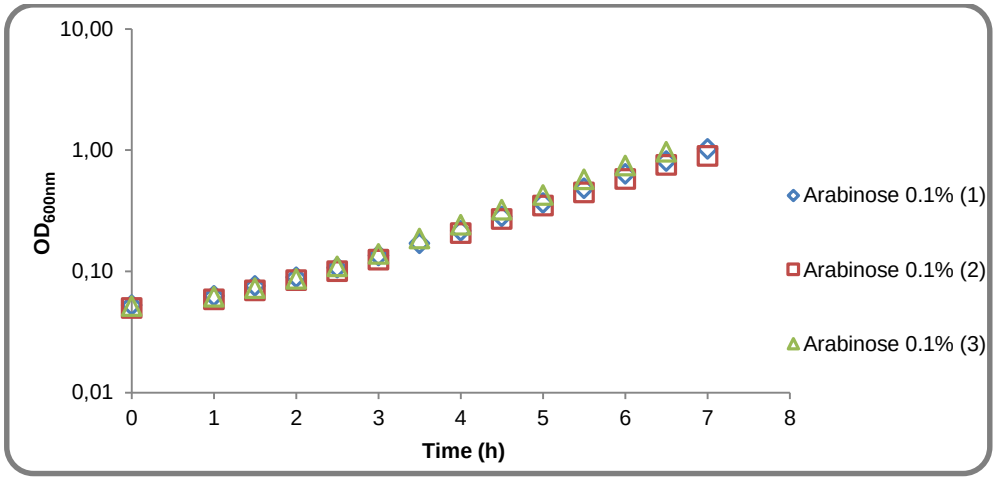
Appendix 6.11 - MsmX and MsmX Homologs sequence alignment. The amino acid sequences of MsmX and MalK from *E. coli* were aligned using ClustalW2 (Thompson *et al.*, 1994). Identical amino acids are marked with a red asterisk, while substitutions with residues displaying the same properties are marked with ":" and semi-conservative substitutions are marked with a ".". Gaps in the amino acid sequence are represented by a "–". The conserved motifs are labeled in colors. The red bar indicates the region of the proteins predicted to interact with the TMDs of AraPQ.



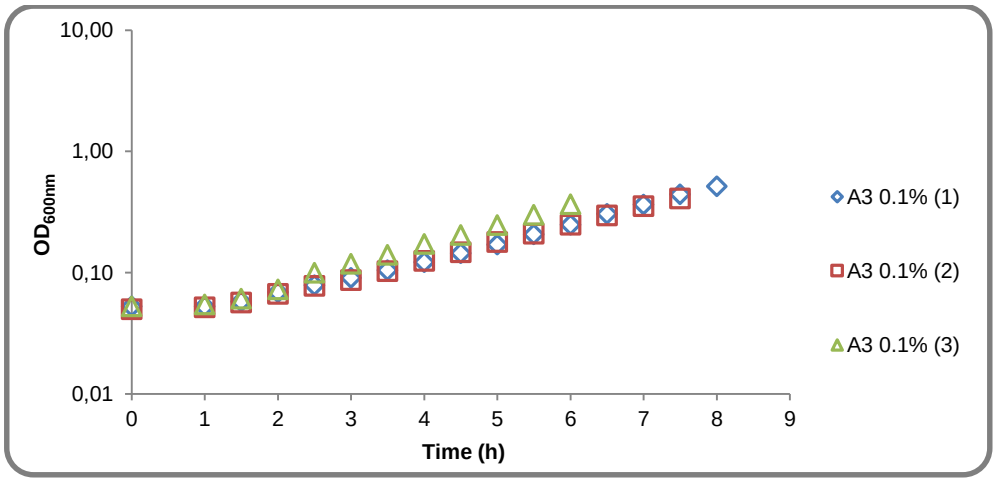
Appendix 6.12 - Growth assays of *B. subtilis* 168T⁺ (wild-type) using glucose 0.1% (w/v). OD_{600nm} plotted *vs*. time (h).



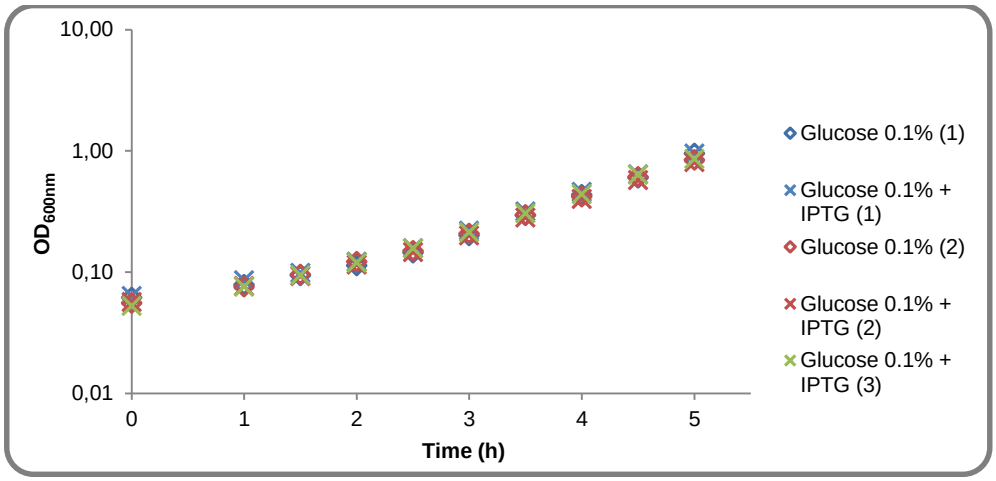
Appendix 6.13 - Growth assays of *B. subtilis* 168T⁺ (wild-type) using arabinose 0.1% (w/v). OD_{600nm} plotted *vs*. time (h).



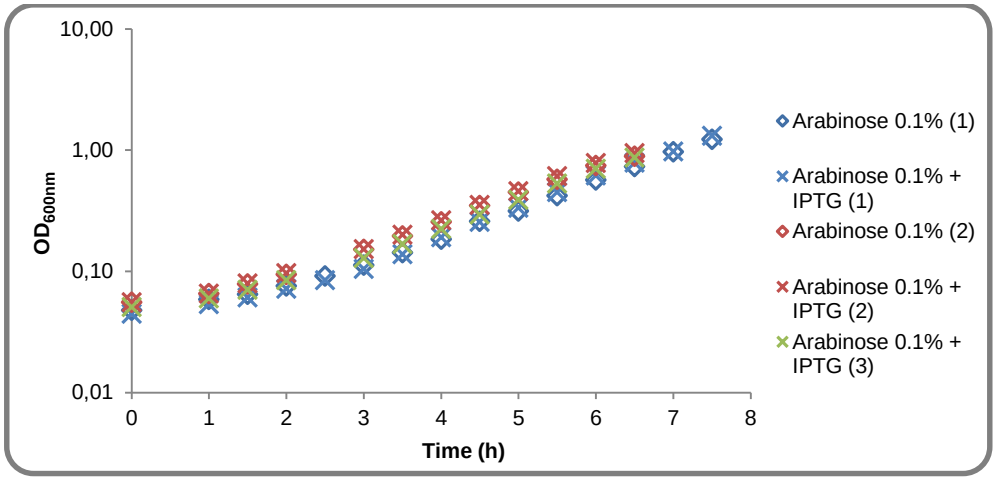
Appendix 6.14 - Growth assays of *B. subtilis* 168T⁺ (wild-type) using arabinotriose (A3) 0.1% (w/v). OD_{600nm} plotted *vs*. time (h).



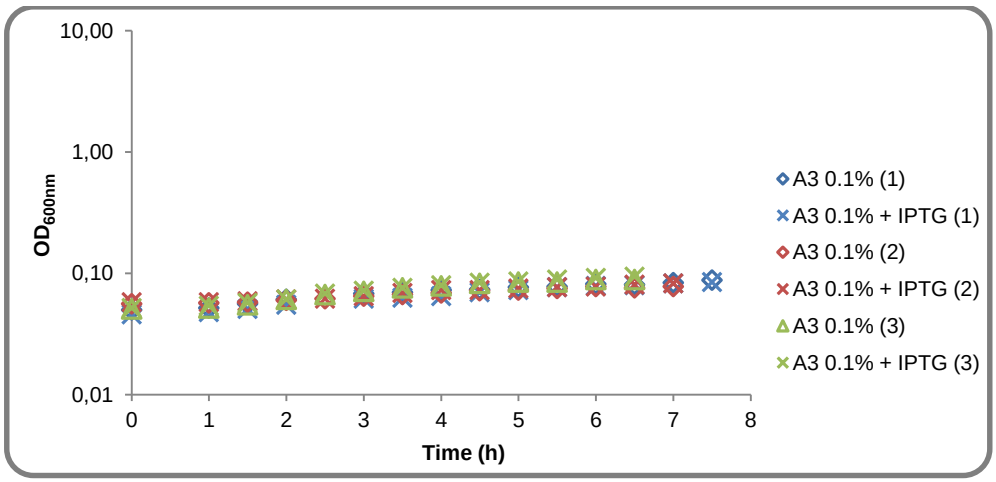
Appendix 6.15 - Growth assays of *B. subtilis* IQB672 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)) using glucose 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



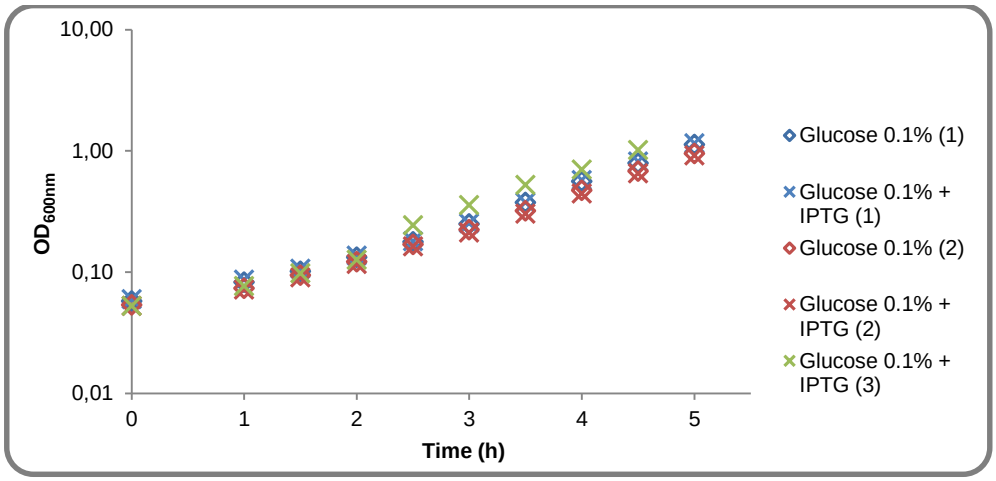
Appendix 6.16 - Growth assays of *B. subtilis* IQB672 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)) using arabinose 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



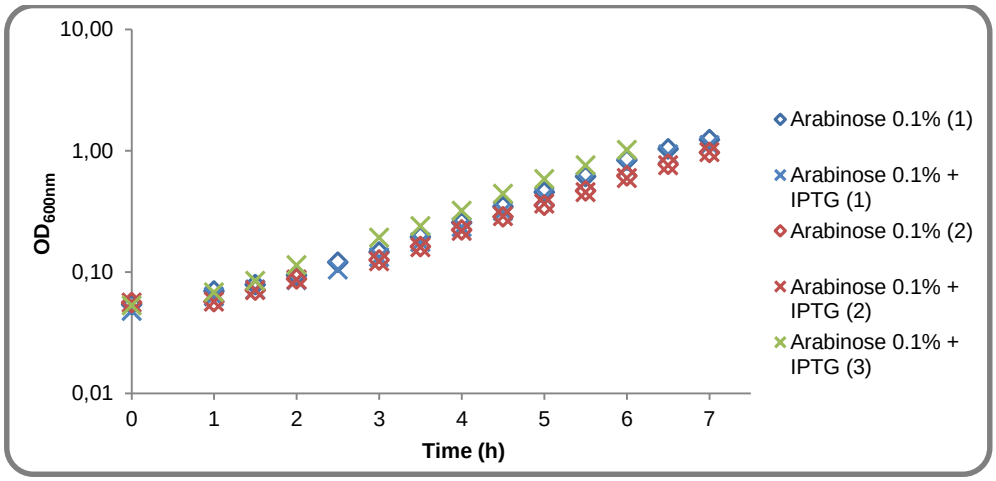
Appendix 6.17 - Growth assays of *B. subtilis* IQB672 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)) using arabinotriose (A3) 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



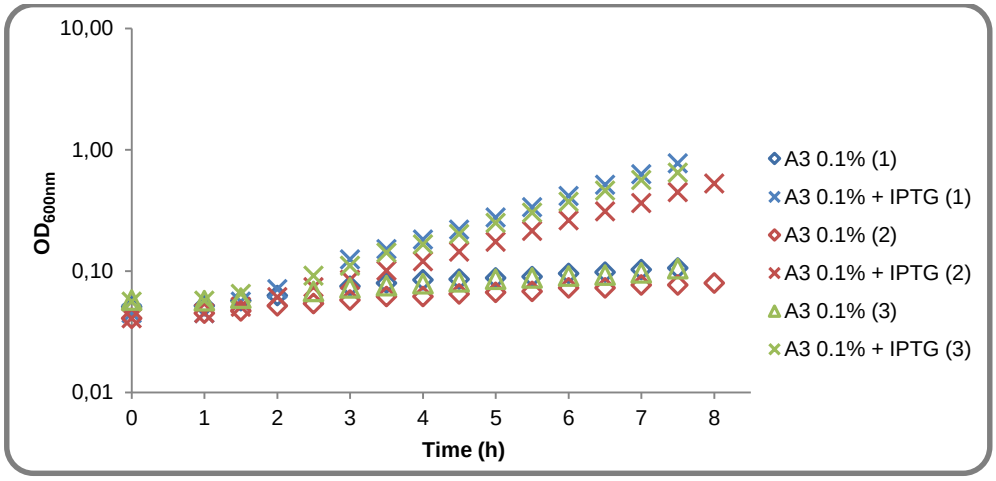
Appendix 6.18 - Growth assays of *B. subtilis* IQB673 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmX*) using glucose 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



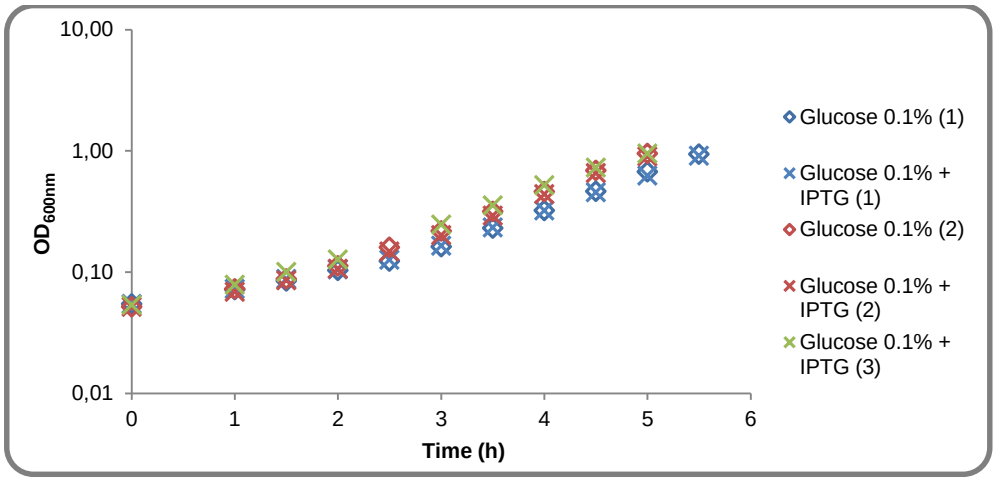
Appendix 6.19 - Growth assays of *B. subtilis* IQB673 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmX*) using arabinose 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



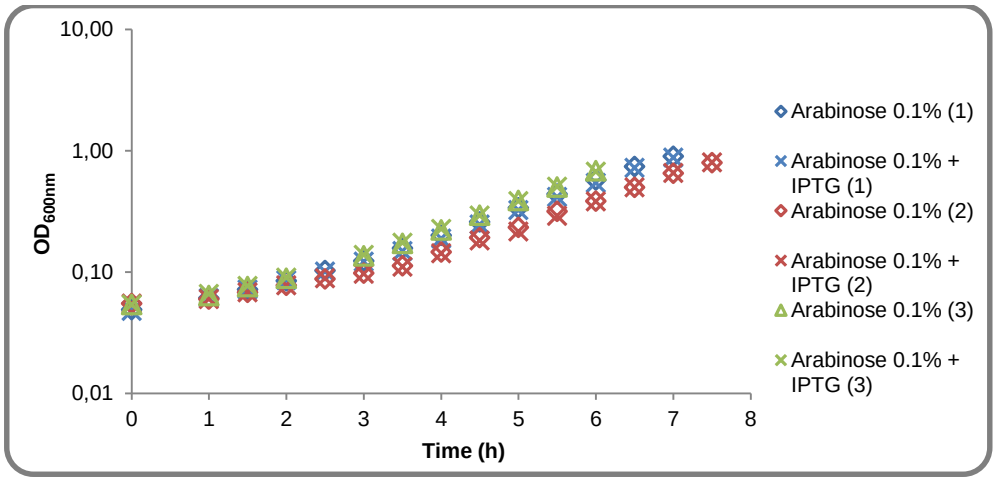
Appendix 6.20 - Growth assays of *B. subtilis* IQB673 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmX*) using arabinotriose (A3) 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



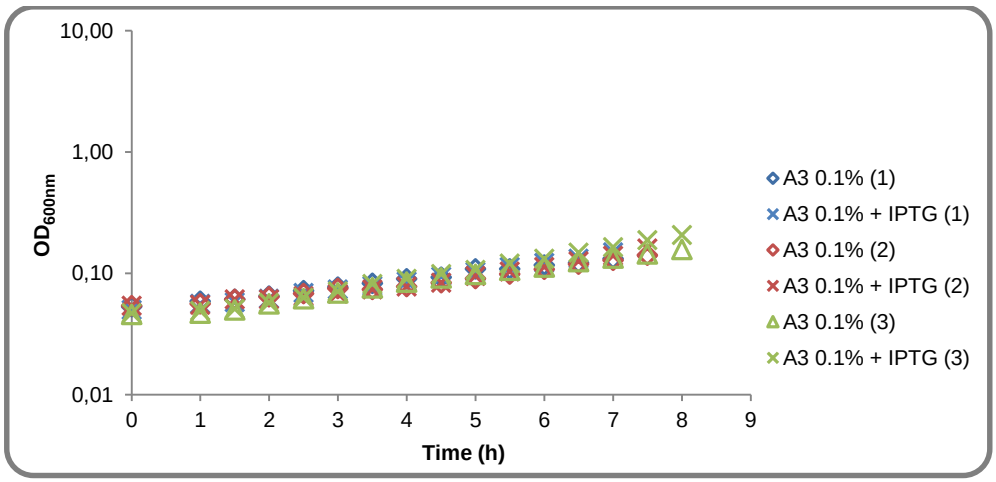
Appendix 6.21 - Growth assays of *B. subtilis* IQB674 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmK*) using glucose 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



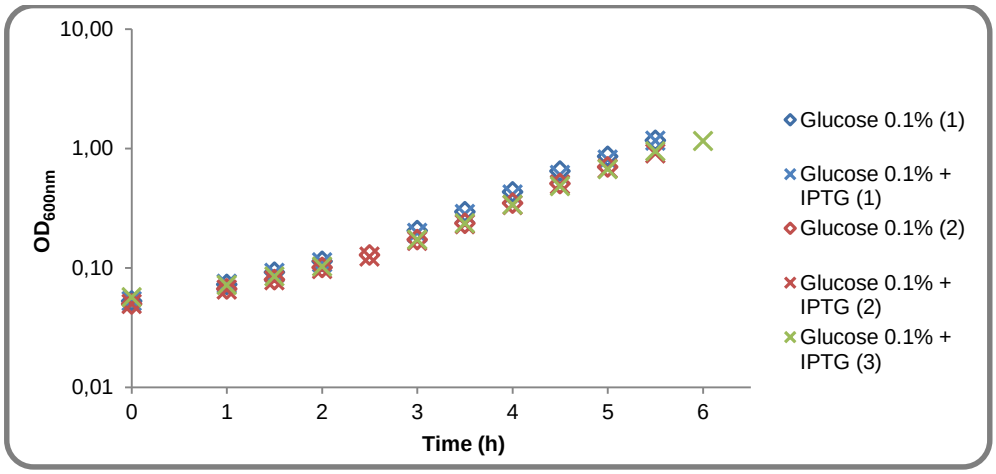
Appendix 6.22 - Growth assays of *B. subtilis* IQB674 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmK*) using arabinose 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



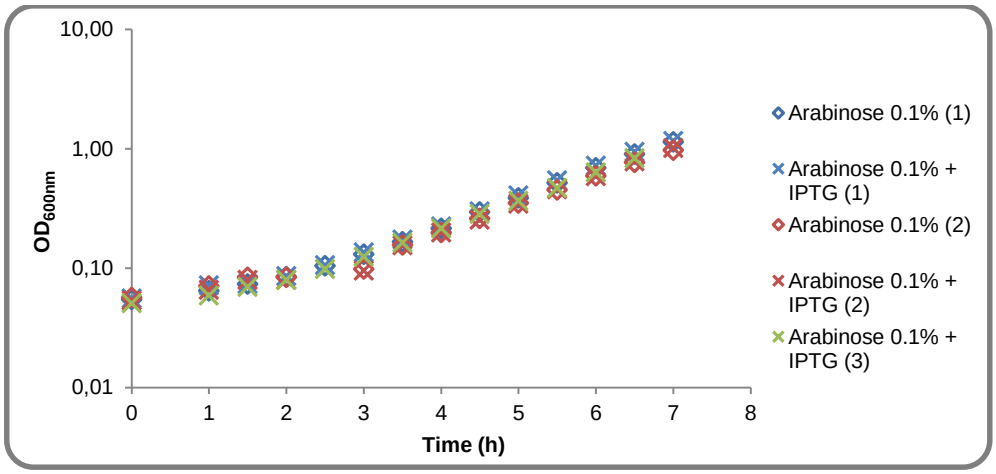
Appendix 6.23 - Growth assays of *B. subtilis* IQB674 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmK*) using arabinotriose (A3) 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



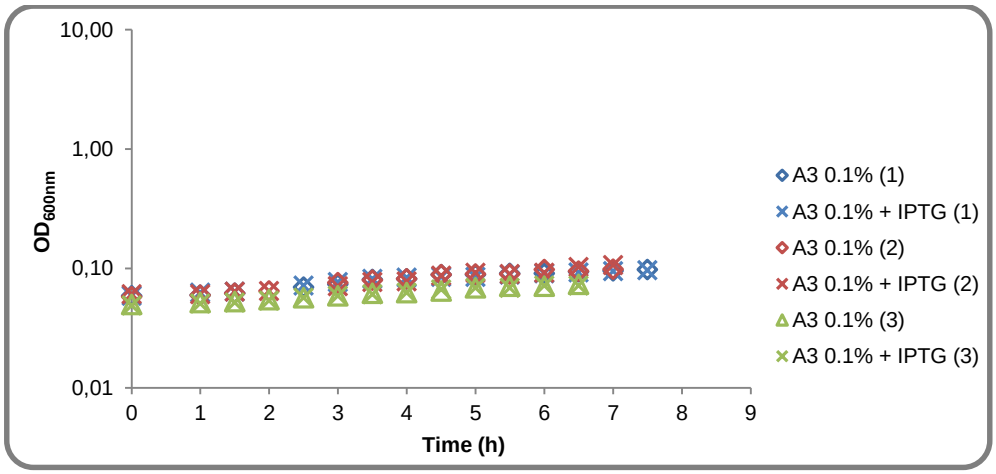
Appendix 6.24 - Growth assays of *B. subtilis* IQB675 (Δ*msmX*::*cat* Δ*amyE*::pSpank(hy)-*msmX*(Lys45Ala)) using glucose 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



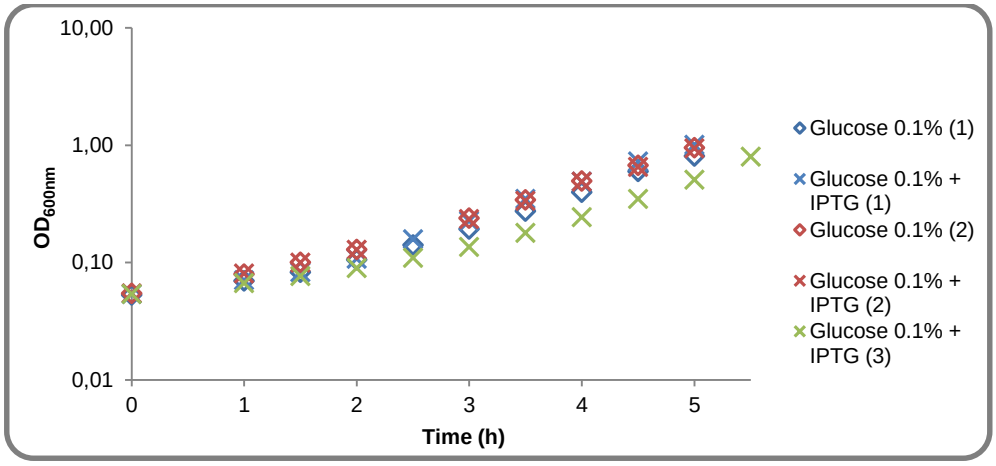
Appendix 6.25 - Growth assays of *B. subtilis* IQB675 (Δ*msmX*::*cat* Δ*amyE*::pSpank(hy)-*msmX*(Lys45Ala)) using arabinose 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



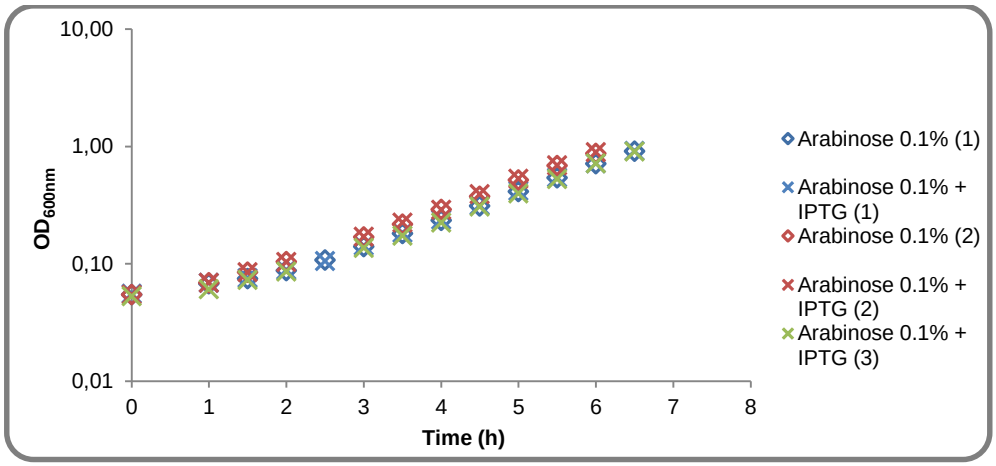
Appendix 6.26 - Growth assays of *B. subtilis* IQB675 (Δ*msmX*::*cat* Δ*amyE*::pSpank(hy)-*msmX*(Lys45Ala)) using arabinotriose (A3) 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



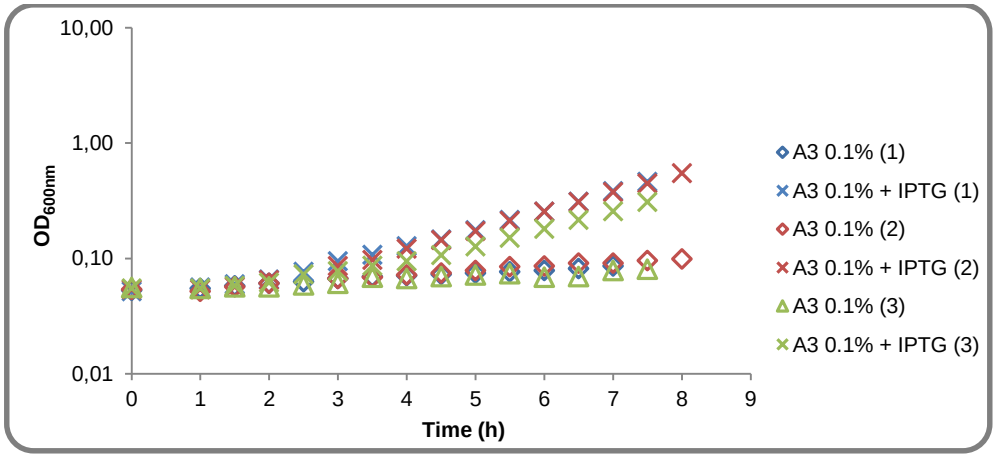
Appendix 6.27 - Growth assays of *B. subtilis* IQB676 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmX*(Glu3Ser, Ile364Ser)) using glucose 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



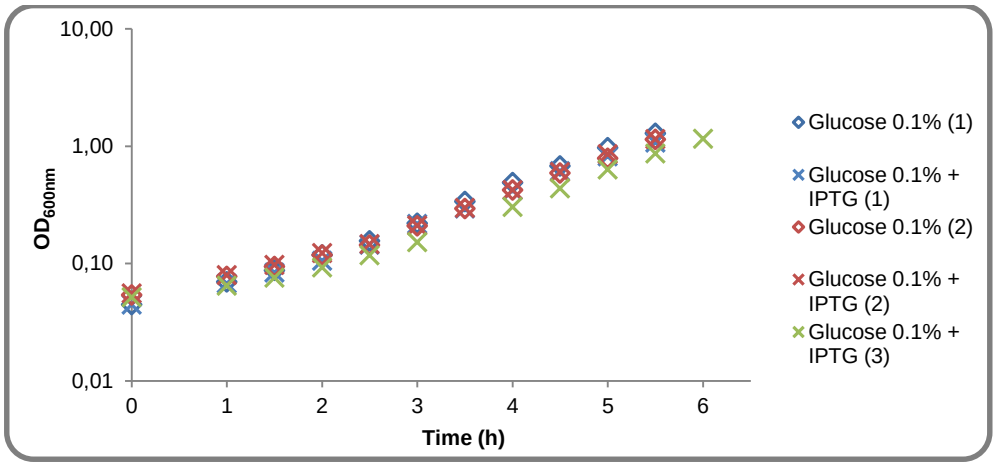
Appendix 6.28 - Growth assays of *B. subtilis* IQB676 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmX*(Glu3Ser, Ile364Ser)) using arabinose 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



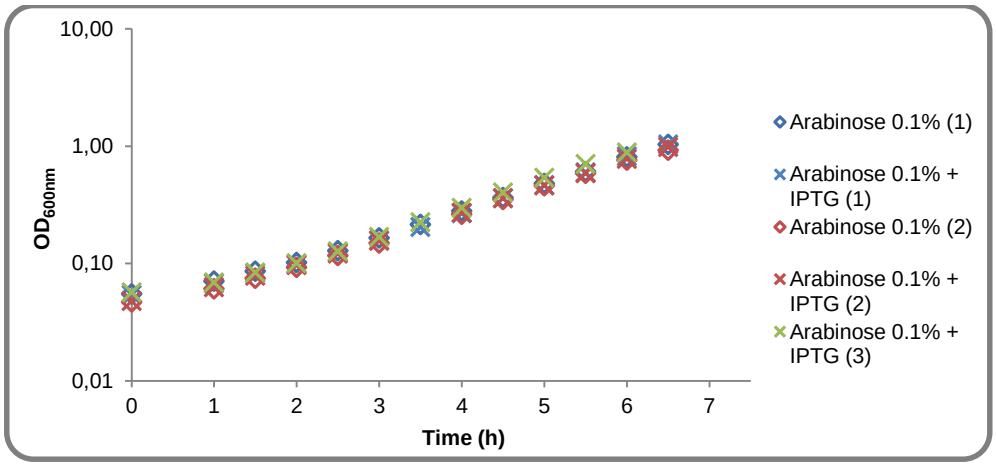
Appendix 6.29 - Growth assays of *B. subtilis* IQB676 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*msmX*(Glu3Ser, Ile364Ser)) using arabinotriose (A3) 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



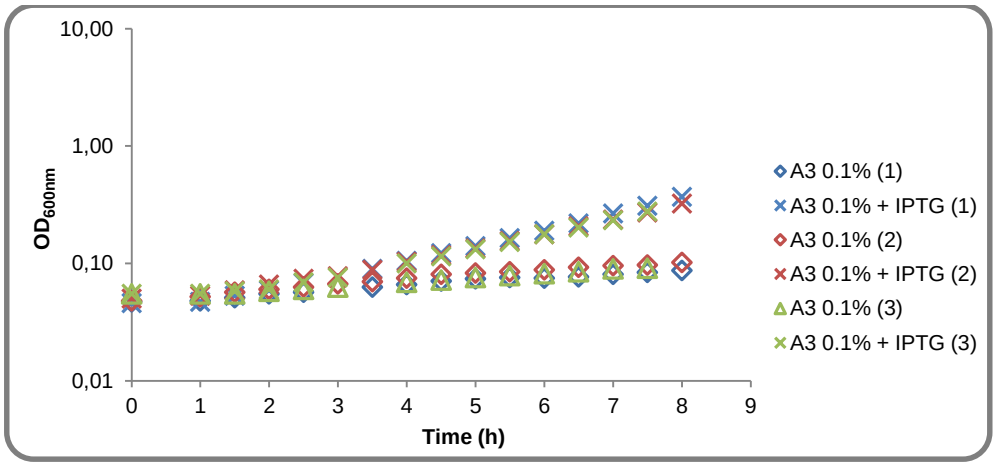
Appendix 6.30 - Growth assays of *B. subtilis* IQB677 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*HD73_4301*) using glucose 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



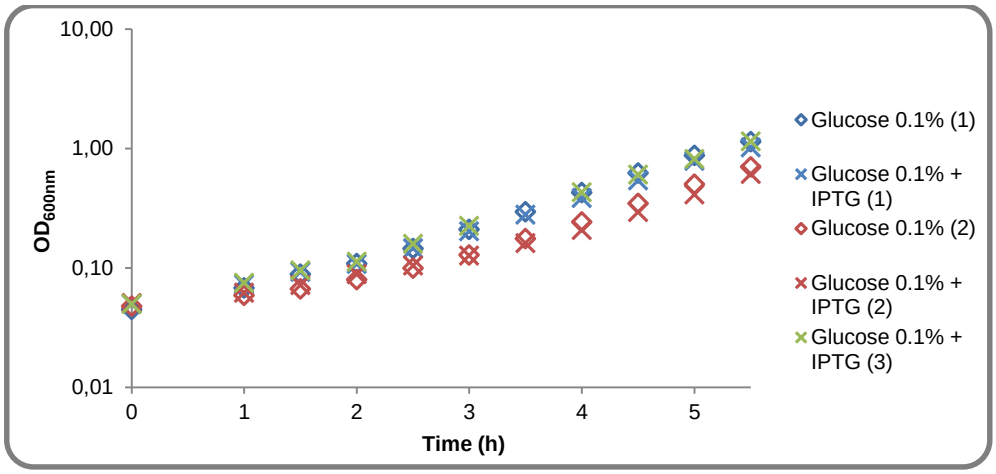
Appendix 6.31 - Growth assays of *B. subtilis* IQB677 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*HD73_4301*) using arabinose 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



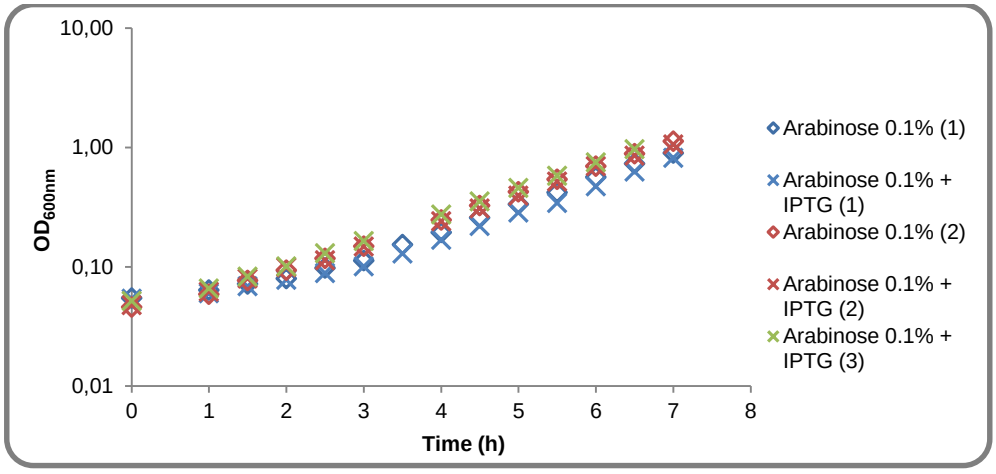
Appendix 6.32 - Growth assays of *B. subtilis* IQB677 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*HD73_4301*) using arabinotriose (A3) 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



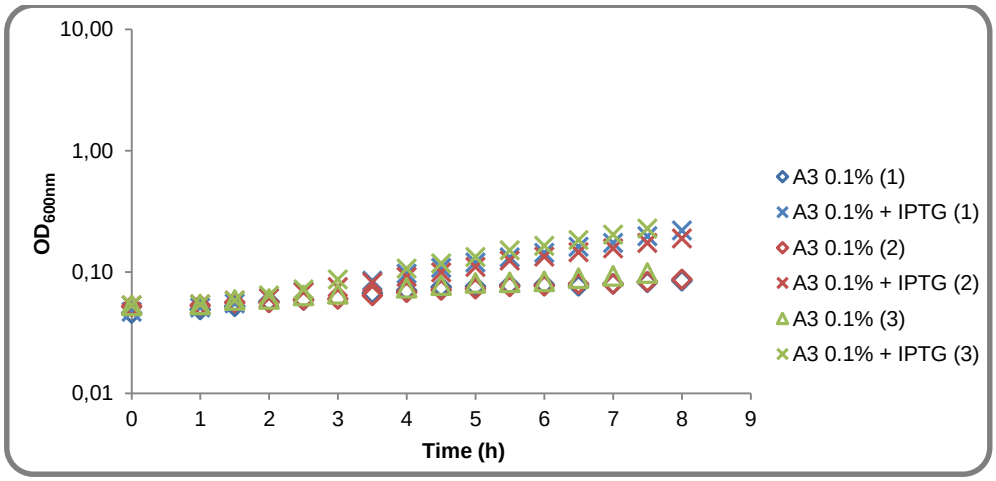
Appendix 6.33 - Growth assays of *B. subtilis* IQB677 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*ugpC*) using glucose 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



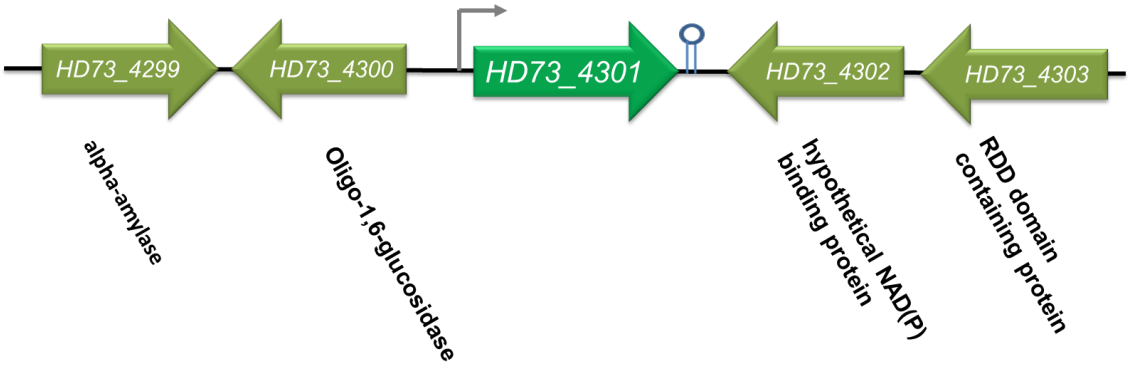
Appendix 6.34 - Growth assays of *B. subtilis* IQB677 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*ugpC*) using arabinose 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



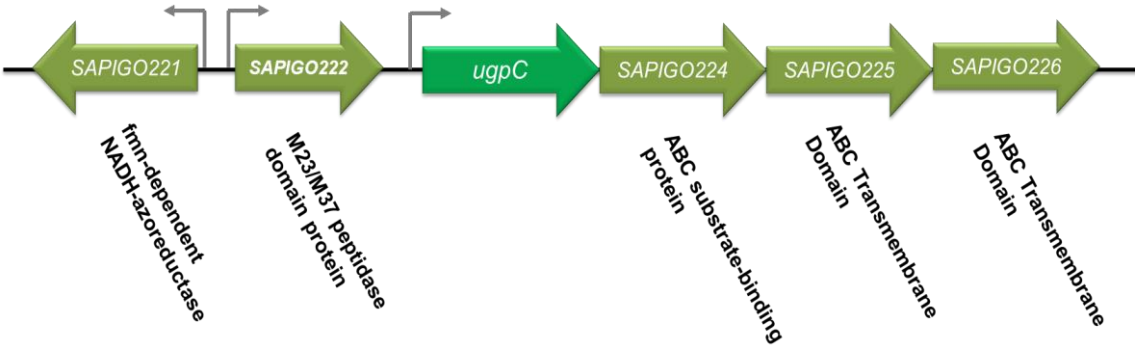
Appendix 6.35 - Growth assays of *B. subtilis* IQB677 (Δ*msmX*::*cat* Δ*amyE*::Pspank(hy)-*ugpC*) using arabinotriose (A3) 0.1% (w/v) and IPTG 1 mM. OD_{600nm} plotted *vs*. time (h).



Appendix 6.36 - Genomic context of *HD73_4301* from *Bacillus thuringiensis*. The arrows denote the genes present in the *msmK* gene region. Below each arrow is indicated the putative function of encoded protein based on BLAST results.



Appendix 6.37 - Genomic context of *ugpC* from *Staphylococcus aureus*. The arrows represent the genes present in the *ugpC* gene region. Below each arrow is indicated the putative function of encoded protein based on BLAST results.



Appendix 6.38 - Genomic context of *msmK* from *Streptococcus pneumoniae*. The arrows indicate the genes present in the *HD73_4301* gene region. Below each arrow is indicated the putative function of encoded protein based on BLAST results.