



Sara Raquel Nascimento Santos Pereira da Silva

Graduated in Biochemistry

**MsmX as model for functional studies of
Multitask ATPases from pathogenic bacteria**

Dissertation presented to obtain the
Master degree in Biochemistry

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

Júri:

Presidente: Prof. Doutor Pedro António de Brito Tavares
Arguente: Prof. Doutor Luís Jaime Gomes Ferreira da Silva Mota
Vogal: Prof. Doutora Isabel Maria Godinho de Sá Nogueira

MsmX as model for functional studies of Multitask ATPases from pathogenic bacteria

Copyright © reserved to Sara Pereira da Silva, FCT/UNL, UNL

The Faculty of Science and Technology and the New University of Lisbon have the perpetual right, and without geographical limits, to archive and publish this dissertation through press copies in paper or digital form, or by other known form or any other that will be invented, and to divulgate it through scientific repositories and to admit its copy and distribution with educational or research objectives, non-commercial, as long as it is given credit to the author and editor.

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Acknowledgments

First of all I would like to profoundly acknowledge my supervisor, Isabel de Sá Nogueira, for all the support, advice and knowledge provided throughout this master thesis, which allowed me to grow not only as a young scientist but also as a person. Having arrived in a time where my colleagues were physically absent from the laboratory, the support and dedication of my supervisor made all the difference during this astonishing scientific journey.

To Research Unit on Applied Molecular Biosciences (UCIBIO), which is financed by national funds from FCT/MEC (UID/Multi/04378/2013) and co-financed by the ERDF under the PT2020 Partnership Agreement (POCI-01-0145-FEDER-007728), for partially supporting this work.

I thank my lab colleagues for the support they were able to give me in spite of being absent from the lab. Thank you Lia for making me comfortable and part of the group since the beginning, as well as for your useful tips. Thanks Mário for your help in this project and for being available to answer my questions through the telephone. I specially thank you Viviana for your constant support since my graduation project, it has been a pleasure.

I would also like to thank to Barbara, Cinthia and Raquel from lab 333, for the good mood and companionship, and in particular to Raquel for clarify my questions. Thank you Cláudia for the good times that we had during lunch hour and the breaks from work, I feel grateful to have meet you. Thanks Nicole for making me laugh and for maintaining the good mood in the department. Finally, I thank to all the people from the DCV department for making me feel welcome.

To my friends, who are a life's treasure, I thank you all for being present in my life. Thanks to Filipa Inácio, Filipa Trovão, Joana Marques, Susana Filipa ("Marias" group) for all the support and that special moments filled with joy, with a special thanks to Susana Filipa who in this group heard me the most (and vice-versa). Thanks Gonçalo for the solid friendship we have had through the last 6 years, I feel thankful for being your friend. Thank you Nádia for being my best friend in the last 13 years, for all the unconditional support you give me, for listening me every day and saying "You will make it", I feel truly blessed for having you in my life.

To my family, thank you for all the good moments through my life, in particularly to my parents for making this dream came true. To my Dad for his academic advices and relaxing posture, making me believe I could do everything through dedication. To my Mom, my biggest supporter in life, I thank everything and especially all the concern and love that you have for me.

And finally I would like to thank you João for all the support you were able to give me in spite of being far, and for all the love that we have shared in the past 6 years.

Abstract

The ABC-type transporters constitute one of the largest and most diverse transporter superfamilies characterized by a highly conserved ATP-binding cassette, and are widespread among all domains of life. Recent studies, performed in our laboratory demonstrated that MsmX ATPase from *Bacillus subtilis* interacts with several distinct ABC sugar importers thus, unlike other NBDs MsmX was shown to be multitask serving as energy-generating component to several sugar importers. Sharing of an ATPase among carbohydrate ABC transporters in both Gram-positive and Gram-negative bacteria seems to be a common strategy for adaption and survival and may represent novel therapeutic approaches for targeting since ABC importers are exclusive to prokaryotes.

To characterize multipurpose ATPases and to assess their intra- and interspecies interchangeability, we fine-tuned a genetic system in *B. subtilis* for controlled ectopic gene expression. The functionality of distinct multitask ATPases alleles was determined by their ability to complement the role of MsmX in a *B. subtilis* *msmX*-null mutant. Moreover, this genetic system allowed the determination of intracellular accumulation of the tested ATPases by Western-Blot analysis.

The results show that an ATPase from *B. thuringiensis* was able to fulfill the role of MsmX in its absence, while another ATPase from *B. subtilis* YurJ was only able to partially play MsmX role. In addition to intra- and interspecies interchangeability of *Bacillus* ATPases, we found that ATPases from *Streptococcus pneumoniae* and *Staphylococcus aureus* were also able to complement to a certain degree the *B. subtilis* MsmX function in vivo. In contrast ATPases from the Gram-negative bacterium *Escherichia coli* were not functional in *B. subtilis*. Furthermore, all the tested ATPases accumulate in the cells.

Our study shows that *B. subtilis* can be use as model for the study of bacterial multitask ATPases. Furthermore, it provides a genetic tool for the characterization of this phenomenon in bacterial carbohydrate transport and particularly in bacterial pathogens.

Keywords: *Bacillus subtilis*, ABC sugar importers, Multitask ATPases, MsmX, Interchangeability, Bacterial pathogens.

Resumo

Os transportadores do tipo ABC constituem uma das maiores e mais diversificadas superfamílias de transportadores que são caracterizadas por possuírem uma cassete de ligação ao ATP com elevado grau de conservação, encontrando-se amplamente distribuídos em todos os domínios da vida. Estudos recentes, desempenhados no nosso laboratório, demonstraram que a ATPase MsmX proveniente da espécie *Bacillus subtilis* interage com diversos importadores ABC de açúcares distintos, sendo que, ao contrário de outros NBDs, MsmX revelou ser multitarefa servindo como componente gerador de energia de vários importadores de açúcares. A partilha de uma ATPase entre transportadores ABC de carboidratos em bactérias Gram-positivas e Gram-negativas aparenta ser uma estratégia comum de adaptação e sobrevivência, e pode representar um alvo para novas abordagens terapêuticas dado que os importadores ABC são exclusivos dos procariontes.

De modo a caracterizar as ATPases multipropósito e no sentido de determinar a sua permutabilidade intra- e inter-espécies, aperfeiçoou-se um sistema genético em *B. subtilis* para obter uma expressão ectópica controlada do gene. A funcionalidade de alelos distintos das ATPases multitarefa foi determinada através da sua capacidade de complementar o papel de MsmX num mutante de *B. subtilis* sem o gene *msmX*. Além disso, este sistema genético permitiu a determinação da acumulação intracelular das ATPases testadas através de uma análise do tipo Western-Blot.

Os resultados demonstraram que uma ATPase proveniente de *B. thuringiensis* é capaz de realizar na totalidade o papel de MsmX na célula durante a sua ausência, enquanto outra ATPase de *B. subtilis* (YurJ) foi apenas capaz de desempenhar parcialmente o papel de MsmX. Para além da permutabilidade intra e inter-espécies de ATPases em *Bacillus*, descobriu-se que, ATPases provenientes de *Streptococcus pneumoniae* e *Staphylococcus aureus* também são capazes de complementar em determinado grau a função *in vivo* de MsmX de *B. subtilis*. Em contraste, ATPases provenientes da bactéria Gram-negativa *Escherichia coli* não demonstraram esta funcionalidade em *B. subtilis*. Adicionalmente, verificou-se que todas as ATPases testadas são produzidas e acumulam no interior das células do hospedeiro.

O nosso trabalho de investigação demonstra que *B. subtilis* pode ser utilizado como um modelo para o estudo de ATPases multitarefa bacterianas. Além disso, o estudo fornece uma ferramenta genética para a caracterização deste fenómeno no transporte bacteriano de carboidratos, e em particular em agentes patogénicos bacterianos.

Palavras-chaves: *Bacillus subtilis*, Importadores ABC de açúcares, ATPases multitarefa, MsmX, Permutabilidade, Agentes patogénicos bacterianos.

Contents

Acknowledgments.....	V
Abstract.....	VII
Resumo.....	IX
Contents.....	XI
Figures Index.....	XIII
Tables Index.....	XVII
Abbreviations, Symbols and Notations.....	XIX
1. General Introduction.....	3
1.1. <i>Bacillus subtilis</i> - An overview.....	3
1.1.1. AraNPQ ABC importer and the Multitask MsmX ATPase.....	4
1.2. Membrane Transports.....	5
1.3. ABC Transporters.....	6
1.3.1. An overview.....	6
1.3.2. Types of ABC Transporters.....	7
1.3.3. Transport Mechanism.....	9
1.3.4. The solute binding protein (SBP).....	11
1.3.5. The transmembrane domains (TMDs).....	12
1.3.6. The nucleotide binding domains (NBDs).....	14
1.3.7. Multitask ATPases.....	16
1.3.7.1. MsmK, a multitask ATPase in <i>Streptococcus</i> species.....	17
1.3.7.2. Multitask ATPases from <i>Streptomyces</i> species and <i>Thermus thermophilus</i>	19
1.4. Scope of the Thesis.....	20
2. Materials and Methods.....	23
2.1. Substrates.....	23

2.2. Bioinformatic Analysis.....	23
2.3. Isolation of chromosomal DNA.....	23
2.4. DNA manipulation and sequencing.....	24
2.5. Site-directed mutagenesis by Overlapping PCR.....	24
2.6. Construction of plasmids and strains.....	25
2.7. Growth conditions.....	29
2.8. Protein extracts of <i>B. subtilis</i>	30
2.9. Western-Blot assay.....	31
2.10. Protein analysis.....	31
3. Results and Discussion.....	35
3.1. Complementation analysis of MsmX homologs in <i>msmX</i> -null mutant <i>B. subtilis</i> strains.....	35
3.1.1. In silico Analysis.....	35
3.1.2. Fine-tuning of a Genetic System for Complementation Analysis.....	36
3.1.3. Functional studies of different ATPases in <i>B. subtilis</i>	40
3.1.3.1. Wild-type MsmX and MsmX Homologs.....	40
3.1.3.2. Recombinant MsmX and MsmX Homologs with a C-terminal His-tag.....	46
3.1.4. Analysis of the predicted proteins structures.....	52
3.1.5. Redesigning the genetic system for functional analysis.....	55
3.1.5.1. Functional studies.....	57
3.1.5.2. Analysis of the 3-D structure models.....	62
3.2. Detection of the in vivo accumulation of the recombinant ATPases in the cells.....	64
4. Concluding Remarks and Future Perspectives.....	71
5. References.....	75
6. Appendices.....	87

Figures Index

Figure 1.1 - MsmX-dependent ABC importers in <i>B. subtilis</i>	5
Figure 1.2 - Four distinct folds of ABC transporters.....	7
Figure 1.3 - Conformations of the MalEFGK ₂ transporter (class I importer).....	9
Figure 1.4 - The transport mechanism of Type I importers (exemplified by MalEFGK ₂).....	10
Figure 1.5 - Rearrangements in SBP MalE upon the substrate binding.....	11
Figure 1.6 - Transfer of the maltose from MBP to the TM binding site.....	13
Figure 1.7 - The TMD–MalK interface.....	13
Figure 1.8 - The structure of the NBDs, as exemplified by the MalK dimer of the maltose transporter MalEFGK ₂	15
Figure 1.9 - Schematic presentation of the <i>msm</i> and <i>mal</i> loci encoding relevant carbohydrate ABC transporters in selected <i>streptococci</i>	19
Figure 1.10 - Schematic summary of carbohydrates utilization by ABC transporters in <i>S. mutans</i> (A), <i>S. pneumoniae</i> (B) and <i>S. suis</i> (C).....	19
Figure 3.1 - Schematic illustration of the in vivo system fine-tuned.....	37
Figure 3.2 - Schematic representation of the <i>amyE</i> locus and the <i>yxf-msmX</i> operon in the chromosome of several <i>B. subtilis</i> strains used in the work.....	38
Figure 3.3 - Growth of <i>B. subtilis</i> ISN1 ($\Delta amyE::P_{spank(hy)}-spec$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	42
Figure 3.4 - Growth of <i>B. subtilis</i> IQB672 ($\Delta msmX::cat \Delta amyE::P_{spank(hy)}$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	42
Figure 3.5 - Growth of <i>B. subtilis</i> ISN673 ($\Delta msmX::cat \Delta amyE::P_{spank(hy)}-msmX$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	42
Figure 3.6 - Growth of <i>B. subtilis</i> ISN8 ($\Delta msmX::cat \Delta amyE::P_{spank(hy)}-msmK$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	44
Figure 3.7 - Growth of <i>B. subtilis</i> ISN9 ($\Delta msmX::cat \Delta amyE::P_{spank(hy)}-malk$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	45
Figure 3.8 - Growth of <i>B. subtilis</i> ISN16 ($\Delta msmX::cat \Delta amyE::P_{spank(hy)}-ycjV$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	46
Figure 3.9 - Growth of <i>B. subtilis</i> IQB676 ($\Delta msmX::cat \Delta amyE::P_{spank(hy)}-msmX(Glu3Ser, Ile364Ser)$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	48
Figure 3.10 - Growth of <i>B. subtilis</i> ISN2 ($\Delta msmX::cat \Delta amyE::P_{spank(hy)}-msmX-His_6$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	48

Figure 3.11 - Growth of <i>B. subtilis</i> ISN3 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmK-His_6$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	50
Figure 3.12 - Growth of <i>B. subtilis</i> ISN4 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-HD73_RS21400-His_6$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	50
Figure 3.13 - Growth of <i>B. subtilis</i> ISN5 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ugpC-His_6$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	50
Figure 3.14 - Growth of <i>B. subtilis</i> ISN6 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-malk-His_6$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	51
Figure 3.15 - Growth of <i>B. subtilis</i> ISN7 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-yurJ-His_6$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	51
Figure 3.16 - Representation of MsmX-His ₆ and MsmX-LEHis ₆ 3-D structure model.....	53
Figure 3.17 - Ribbon representation of the Malk dimer.....	53
Figure 3.18 - Representation of UgpC-His ₆ and YurJ-His ₆ 3-D structure model.....	54
Figure 3.19 - Schematic representation of the modifications made in the genetic system.....	55
Figure 3.20 - Schematic representation of the <i>amyE</i> locus and the <i>yxkf-msmX</i> operon in the chromosome of several <i>B. subtilis</i> strains, constructed with the redesigned genetic system, used in the work.....	56
Figure 3.21 - Growth of <i>B. subtilis</i> ISN10 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmX-LEHis_6$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	58
Figure 3.22 - Growth of <i>B. subtilis</i> ISN13 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ugpC-LEHis_6$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	58
Figure 3.23 - Growth of <i>B. subtilis</i> ISN15 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-yurJ-LEHis_6$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	59
Figure 3.24 - Growth of <i>B. subtilis</i> ISN11 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-HD73_RS21400-LEHis_6$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	60
Figure 3.25 - Growth of <i>B. subtilis</i> ISN12 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmK-LEHis_6$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	61
Figure 3.26 - Growth of <i>B. subtilis</i> ISN14 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-malk-LEHis_6$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	61
Figure 3.27 - Growth of <i>B. subtilis</i> ISN17 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ycjV-LEHis_6$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.....	61
Figure 3.28 - Representation of YurJ-H ₆ and YurJ-LEH ₆ 3-D structure model.....	62
Figure 3.29 - Representation of UgpC-H ₆ and UgpC-LEH ₆ 3-D structure model.....	63
Figure 3.30 - Western-Blot analysis of <i>B. subtilis</i> cell extracts, obtained from CSK medium with arabinose 0.1% in inducing conditions.....	64

Figure 3.31 - Western-Blot analysis of *B. subtilis* cell extracts, obtained from CSK medium with arabinose 0.1% in non-inducing conditions.....65

Figure 3.32 - Western-Blot analysis of *B. subtilis* cell extracts, obtained from CSK medium with arabinotriose 0.1% in inducing conditions.....66

Figure 3.33 - Western-Blot analysis of *B. subtilis* cell extracts, obtained from CSK medium with arabinotriose 0.1% in non-inducing conditions.....67

Tables Index

Table 2.1 - List of plasmids used in this work.....	27
Table 2.2 - List of oligonucleotides used in this work.....	28
Table 2.3 - List of <i>B. subtilis</i> strains used or constructed during this work.....	29
Table 3.1 - List of MsmX homologs selected for this work.....	35
Table 3.2 - Growth of different <i>B. subtilis</i> strains ISN1, IQB672 and IQB673 in the presence of distinct saccharides as sole carbon and energy source.....	41
Table 3.3 - Growth of different <i>B. subtilis</i> strains IQB642, IQB677 and IQB678 in the presence of distinct saccharides as sole carbon and energy source.....	43
Table 3.4 - Growth of different <i>B. subtilis</i> strains ISN8, ISN9 and ISN16 in the presence of distinct saccharides as sole carbon and energy source.....	44
Table 3.5 - Growth of different <i>B. subtilis</i> strains IQB673, IQB676 and ISN2 in the presence of distinct saccharides as sole carbon and energy source.....	47
Table 3.6 - Growth of different <i>B. subtilis</i> strains ISN3, ISN4, ISN5, ISN6 and ISN7 in the presence of distinct saccharides as sole carbon and energy source.....	49
Table 3.7 - Growth of different <i>B. subtilis</i> strains IQB676, ISN10, ISN13 and ISN15 in the presence of distinct saccharides as sole carbon and energy source.....	57
Table 3.8 - Growth of different <i>B. subtilis</i> strains ISN11, ISN12, ISN14 and ISN17 in the presence of distinct saccharides as sole carbon and energy source.....	59

Abbreviations, Symbols and Notations

a. a. – Amino acid	LMW - Low Molecular Weight
ABC - ATP-binding cassette	mRNA – Messenger ribonucleic acid
ADP – Adenosine diphosphate	MW – Molecular weight
ATP – Adenosine triphosphate	NBD – Nucleotide-binding domain
A3 – 1,5- α -L-Arabinotriose	OD – Optical Density
bla - Beta-lactamase gene	ORF – Open Reading Frame
BLAST – Basic Local Alignment Search Tool	PBS – Phosphate buffer saline
bp – Base pairs	PCR – Polymerase Chain Reaction
cat – chloramphenicol acetyltransferase gene	Pi - Inorganic phosphate
CRD – C-terminal regulatory domain	PMSF – Phenylmethylsulfonyl fluoride
CSK – C medium supplemented with potassium succinate	PTS – Phosphotransferase system
CUT1 – Carbohydrate Uptake Transporter-1	RBS – Ribosome binding site
DTT - Dithiothreitol	RNA – Ribonucleic acid
dNTP – Deoxyribonucleic acid	rpm – Revolutions per minute
ECF – Energy coupling factor	SBP – Solute-binding domain
EDTA - Ethylenediaminetetraacetic acid	SDS-PAGE - Sodium dodecyl sulphate polyacrylamide gel electrophoresis
HRP - Horseradish peroxidase	<i>SpecR</i> – Spectinomycin resistance gene
IPTG – Isopropyl- β -D-galactopyranoside	TIR - Translation initiation rate
I-TASSER - Iterative Threading ASSEmbly Refinement	TAE - Tris-acetate-EDTA
kDa - KiloDalton	TMD – Transmembrane domain
LB – Luria-Bertani medium	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	trp	W
tyrosine	tyr	Y
valine	val	V

Bases – one letter code

Base	Letter
Adenine	A
Cytosine	C
Guanine	G
Thymine	T

Chapter 1

General Introduction

1. General Introduction

1.1. *Bacillus subtilis* – An overview

Bacillus subtilis is a Gram-positive and rod-shaped bacteria that like other members of the genus *Bacillus*, is able to form an endospore in order to survive in extreme environmental conditions. This bacterium is found in the soil, water sources or in association with plants and animal gastrointestinal tract (Priest, 1993; Casula and Cutting, 2002). In these environments the major source of carbohydrates for microorganisms is plant biomass, which is constituted by cellulose and hemicellulose polymers. Therefore *B. subtilis*, and the other microorganisms, from these habitats possess a wide variety of extracellular polysaccharide degrading enzymes, cellulases and hemicellulases. The resulting degradation products (mono-, di- and oligosaccharides) are imported to the cells through different transport systems, including specific ABC transporters.

Kunst *et al* in 1997 sequenced the entire genome of *B. subtilis*, opening the doors for a full use of this bacterium as a model of Gram-positive bacteria in fundamental and applied research. For example in industrial microbiology, specifically in “white biotechnology”, which relies on microorganisms and enzymes to synthesize products that are easily degradable, and during their production require less energy and create less waste (Chauhan *et al*, 2012). Carbohydrate-, lipid- and protein-degrading enzymes, antibiotics, fine biochemicals (vitamins) and insecticides (Harwood, 1992) are examples of the use of *Bacillus* spp in the “white biotechnology” associated with the vast diversity of the metabolism of this organism.

Hemicellulases and cellulases are the major industrially important enzymes right after the proteases (Polizeli *et al*, 2005; Dhawan and Kaur, 2007), due to the wide abundance of hemicellulose and cellulose in nature. Microbial mannanases, which degrade hemicellulose, are mainly produced by Gram-positive *Bacillus* species (Mabrouk and Ahwany, 2008; Meenakshi *et al*, 2010) in extracellular medium. These enzymes can act in wide range of pH and temperature, being used in multiple applications in pulp and paper, pharmaceutical, food, feed, oil and textile industries (Chauhan *et al*, 2012).

1.1.1. AraNPQ ABC importer and the Multitask MsmX ATPase

Quentin *et al*, 1999 performed in silico an inventory and assembly of the ABC transporter systems in the complete genome of *Bacillus subtilis* with an in silico analysis, estimating the existence of at least 78 ABC transporters based on the identification of 86 NBDs in 78 proteins, 103 MSD proteins and 37 BPD proteins representing 5% of the protein-coding genes of this organism. The uptake of sugars is undertaken by at least 10 ABC systems, being one of this importers the AraNPQ. The AraNPQ system is encoded by the *araABDLMNPQ-abfA* operon (Sá Nogueira *et al*, 1997), where AraN is the high-affinity substrate-binding protein (BPD) and AraP and AraQ are the two transmembrane domains (TMDs). In this operon are present genes encoding for enzymes that play a role in arabinose catabolism and degradation of arabinooligosaccharides (Inácio *et al*, 2008). This system is regulated at the transcriptional level by induction in the presence of arabinose and repression by glucose (Sá-Nogueira *et al*, 1997; Sá-Nogueira and Mota, 1997). AraNPQ is the sole transporter for α -1,5-arabinotriose, and α -1,5-arabinotetraose is only partially accountable for the uptake of α -1,5-arabinobiose (Ferreira and Sá-Nogueira, 2010)

The AraNPQ transporter in the context of the arabinose operon lacks a gene encoding for the nucleotide-binding domains (NBDs) or ATPase, which is necessary to provide energy to the system. Ferreira and Sá-Nogueira (2010) identified the ATPase MsmX, encoded in a monocistronic gene in another locus of the chromosome, as the NBD partner of the AraNPQ transporter. Later, the MsmX was also identified as the NBD partner of the CycB-GanPQ transporter responsible for the uptake of galactooligosaccharides, and YesOPQ and YtcQP-YteQ transporters that are involved in the transport of galacturonic acid oligomers and/or rhamnose-galacturonic acid disaccharides (Ferreira and Sá-Nogueira, unpublished data). This ATPase is also related with another importer in *B. subtilis*, the maltose and maltodextrins transporter MdxEFG, in a previous study conducted by Schönert *et al* (2006). Clearly the data show that MsmX is a multitask ATPase shared by multiple sugar ABC transporters (Ferreira and Sá-Nogueira, 2010 and unpublished data; Schönert *et al*, 2006) (Figure 1.1).

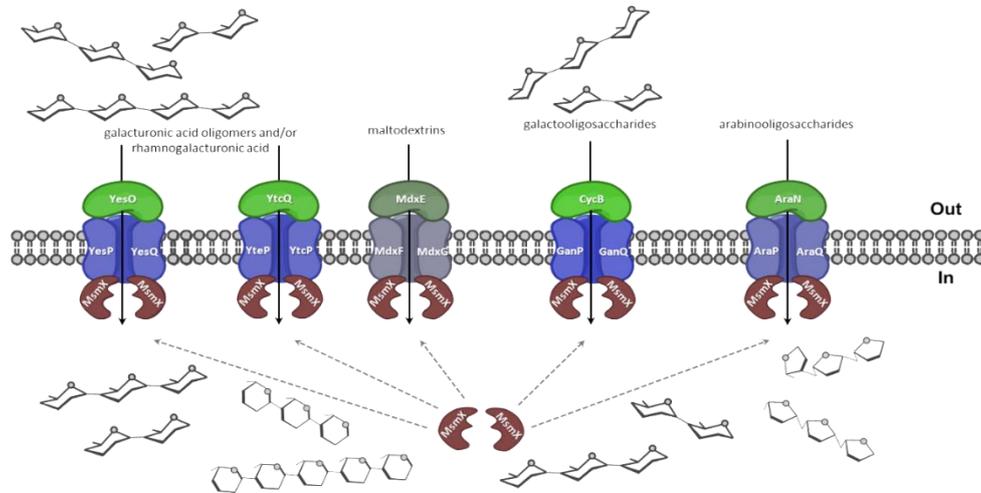


Figure 1.1 – MsmX-dependent ABC importers in *B. subtilis*. The ABC-type importer AraNPQ is involved in the uptake of α -1,5-arabinooligosaccharides (arabinotriose, arabinotetraose and some arabinobiose). The ABC-type importer MdxEFG is involved in the uptake of maltodextrins. The CycB-GanPQ ABC-type importer is responsible for the uptake of galactooligosaccharides. The YesOPQ and YtcQP-YtcQ transport systems are involved in the uptake of galacturonic acid oligomers and/or rhamnogalacturonic acid (adapted from Ferreira and Sá-Nogueira, unpublished data).

1.2. Membrane Transports

Selective permeability to nutrients and metabolites is an essential feature for cell survival, requiring transporters in the membrane with different characteristics in the function and structure, each one adapted to the type of solute translocated. Membrane transporters are distinguished based on the energy source used and therefore are classified in four major groups: protein channels, primary active transporters, secondary transporters and group translocators.

Channels allow the passage of solutes through facilitated diffusion in a process that is energy-independent. In Gram-negative bacteria, porin proteins form a TM-spanning aqueous pore constituting a channel in the outer-membrane that allow diffusion of several substrates. Another example are the cytoplasmic membrane channels which are gated and controlled by voltage (Ren *et al*, 2001) or membrane tensions, for instance the MscS channel in *E. coli* that function as a protector of the cells from hypo-osmotic shock (Levina *et al*, 1999; Bass *et al*, 2002; Davidson *et al*, 2008).

Primary active transporters constitute a large and diverse protein family that transport substrates across the membrane against a concentration gradient, which depends upon the energy withdrawn from chemical, electrical or solar energy sources. The largest and most widespread family of this class is the ATP-binding cassette (ABC) transporters that obtain energy through the hydrolysis of ATP molecules for the translocation of the substrate.

Secondary active transporters use energy provided by ion gradients to drive transport, being examples uniporters, antiporters and symporters (Davidson *et al*, 2008). Group translocation causes chemical changes in the substrates, being one example of this family the phosphotransferase system (PTS), which is exclusive to prokaryotes and plays an important role in the uptake of sugars. In this system the transported carbohydrate is phosphorylated once it reaches the cytoplasmic side of the membrane by phosphoenolpyruvate (PEP) that act as both the phosphate donor for sugar phosphorylation and the energy source for sugar accumulation (Postma *et al*, 1993; Robillard and Broos, 1999; Tchieu *et al*, 2001; Saier *et al*, 2002; Jaehme *et al*, 2015).

1.3. ABC Transporters

1.3.1. An overview

ABC-type transporters constitute a large and diverse superfamily of ATP-dependent protein complexes, which play an important role in organisms from the three domains of life Bacteria, Archaea and Eukarya (Eitinger *et al*, 2011). They are characterised by a highly conserved ATP-binding cassette that hydrolyses ATP molecules and provide free energy that is converted into trans-bilayer movement of substrates, by the transmembrane domains, as import to the cytoplasm or export from the cytoplasm (Locher, 2009). The type of substrates transported by this systems varies in a wide range from small inorganic and organic molecules, such as amino acids, sugars, nucleosides, vitamins and metal clusters to larger organic compounds, as peptides, lipid molecules, oligonucleotides and polysaccharides (Wilkins, 2015). The importance of the ABC transporters goes beyond the uptake of nutrients or export of toxic waste, for example ABC importers have important roles in the maintenance of cell integrity, responses to environmental stresses, cell-to-cell communication and cell differentiation, and in pathogenicity (Eitinger *et al*, 2011) and ABC exporters are involved in the drug resistance of bacteria and cancer cells (ter Beek *et al*, 2014).

All ABC transporters have a core with the same modular architecture: two transmembrane domains (TMDs) or subunits and two nucleotide-binding domains (NBDs) or subunits. So far four types of ABC transporters have been identified based on the TMDs folds as determined by the crystal structures: importers type I and type II, ECF transporters (importer type III) and exporters. The available structures of ABC transporters provided fundamental knowledge about the transport mechanism and the great structural diversity of this systems (ter Beek *et al*, 2014).

1.3.2. Types of ABC Transporters

Canonical (protein-dependent) ABC transporters can be distinguish in two major groups: importers and exporters. Within the importers group there are two classes (I and II) and the energy coupling factor (ECF) transporters considered the third class of importers, although it is structurally and functionally more distinct (Eitinger *et al*, 2011 and Erkens *et al*, 2011). Until a few years ago it was believed that ABC importers only existed in bacteria and archaea, however recent studies showed evidence of the presence of class I and II Importers in plants, namely *Arabidopsis thaliana* and *Oryza sativa* (Eitinger *et al*, 2011). ABC exporters exists in bacteria, archaea and eukaryotes, being the only type of ABC transporters present in higher eukaryotes, and are involved in the transport of hydrophobic compounds such as lipids, fatty acids, cholesterol, drugs and large molecules as proteins (ter Beek *et al*, 2014). This type of transporters plays an important role in mammals since its defects are associated with several diseases such as, immune deficiency and cancer, cystic fibrosis, genetic conditions including Tangier and Stargardt disease (Wilkins, 2015). The following image (Figure 1.2) illustrates the four types of ABC transporters.

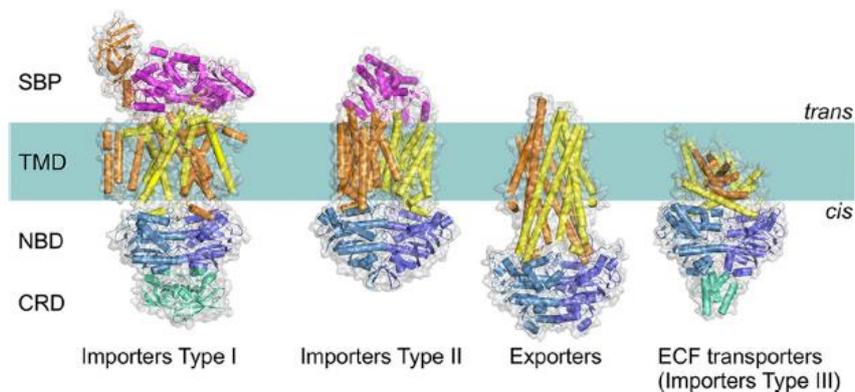


Figure 1.2 - Four distinct folds of ABC transporters. The components of the general architecture are the two NBDs (blue and sky blue) that are attached to two TMDs (orange and yellow). Additional domains (green) that often have a regulatory function (C-terminal regulatory domain [CRD]) are present in some transporters. In Type I and II importers, the SBPs (or SBDs; magenta). ECF, energy coupling factor (Adapted from ter Beek *et al*, 2014).

ABC Importers were originally divided in two categories, type I and type II, due to differences in the size and overall architecture of the core of the transporters (Locher, 2009; Oldham *et al*, 2008), which is associated with the type of substrates transported. In general, importers type I contain less transmembrane helices than the type II, are smaller and use a single type of mechanism of transport the “alternating access model” (Jardetzky, 1966; Rice *et al*, 2014). This is associated with the transport of substances in bulk (although not exclusively) as amino acids and sugars (ter Beek *et al*, 2014), being examples the methionine transporter MetNI, the maltose transporter MalFGK₂ and molybdate transporters ModBC (from *Archaeoglobus fulgidus* and *Methanosarcina acetivorans*) (Rice *et al*, 2014).

Importers of type II are more specialized in the transport of compounds in small quantities such as, metal chelates and vitamins (Davidson *et al*, 2008; Eitinger *et al*, 2011). Complexes that fall in this category are the vitamin B12 transporter BtuCD, the heme transporter HmuUV and the molybdate transporter MoIBC (HI1470/1) from *Haemophilus influenzae* (Rice *et al*, 2014), which possess differences in the transport mechanism that are also associated with the substrate size ranging from vitamin B12 to molybdate (Rice *et al*, 2013 and 2014).

Energy coupling factor (ECF) transporters were recently considered as a third class of ABC importers (Rodionov *et al*, 2009), due to the energy withdrawn from ATP hydrolyse, although displaying significant difference with the other classes for instance in lacking a substrate-binding protein (SBP) presenting instead a EcfS or S component (ter Beek *et al*, 2014). These complexes have a critical role in micronutrient uptake in bacteria and archaea, being examples the folate and the hydroxymethyl pyrimidine transporters from *Lactobacillus brevis* (Wang *et al*, 2013; Xu *et al*, 2013; Rice *et al*, 2014).

Briefly there are ABC transporters that evolved to perform different functions than membrane transport, the so-called non-canonical ABC transporters, being examples the chloride channel CFTR, and the sulfonyleurea receptor SUR (Wilkins, 2015).

The AraNPQ transport system, the system evaluated in this study, is an ABC importer type I so our focus is in this class of transporters.

1.3.3. Transport Mechanism

Type I importers mechanism of transport was first proposed by Jardetzky in 1966 (and further elaborated by Tanford, 1982) as the “alternating access” model and ever since other mechanisms with a few shared features were proposed, namely “switch” model (Higgins *et al*, 2004), and “constant contact” model (Sauna *et al*, 2007; Siarheyeva *et al*, 2010).

The maltose importer (MalEFGK₂) from *E. coli* is one of the best-characterized Type I systems, and may be used to explain the transport mechanism employed by this type of transporters. Chen (2013) proposed a mechanism for the maltose transporter based on the “alternating access” model, deduced from the interpretation of X-Ray crystallography experiments that captured the complex in several conformations: inward-facing, pre-translocation, and outward-facing conformations (Figure 1.3). Biochemical experiments and structural studies using spectroscopic techniques such as, electron paramagnetic resonance, (Davidson *et al*, 1992; Chen *et al*, 2001; Lu *et al*, 2005; Grote *et al*, 2008, 2009; Orelle *et al*, 2008, 2010; Bordignon *et al*, 2010; Jacso *et al*, 2012; Böhm *et al*, 2013; Chen, 2013), support this proposed mechanism (ter Beek *et al*, 2014).

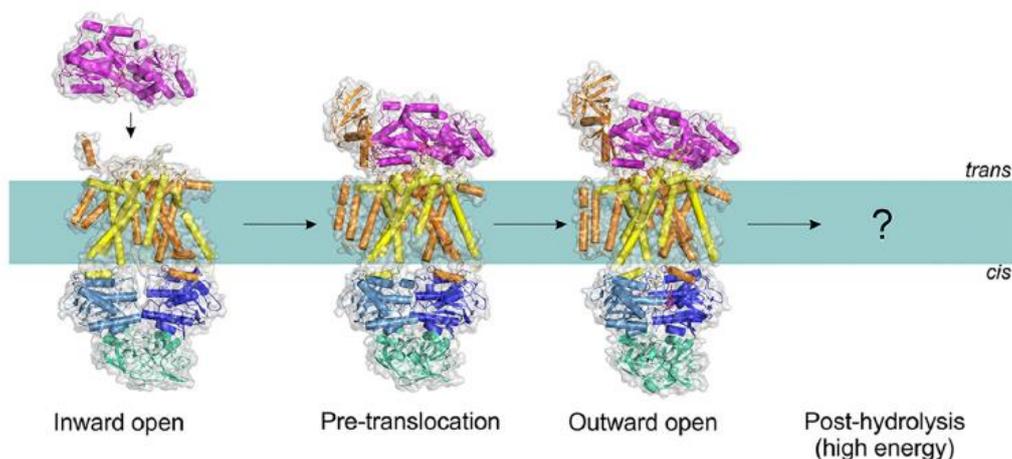


Figure 1.3 – Conformations of the MalEFGK₂ transporter (class I importer). Structures have been determined for the inward-facing, pre-translocation, and outward-facing conformations (Protein Data Bank accession n°: 4JBW, 4KHZ, and 4KI0). Adapted from ter Beek, *et al* (2014).

The proposed mechanism for type I importers, exemplified by the MalEFGK₂ transporter, is the following (Figure 1.4): the binding of a substrate-loaded MBP (MalE) to the TMDs causes a conformational change in these domains that is propagated to the NBDs, bringing the two monomers (NBDs) into closer proximity (the pre-translocation state). This proximity allows the ATP to bind to the NBDs causing the dimer to close with two ATP molecules located at the interface, which in turn causes conformational changes in the TMDs allowing the cavity to open toward the outside (outward open/facing) and the substrate to bind to a specific site on the TM (MalF). The NBDs promote the hydrolysis of the ATP molecules and the consequent release of its products (Pi and ADP) triggers the dimer to move apart, that in turn propagate the conformational change to the TMDs allowing them to expose their cavity to the cytoplasm with subsequent release of the substrate (ter Beek, *et al* 2014).

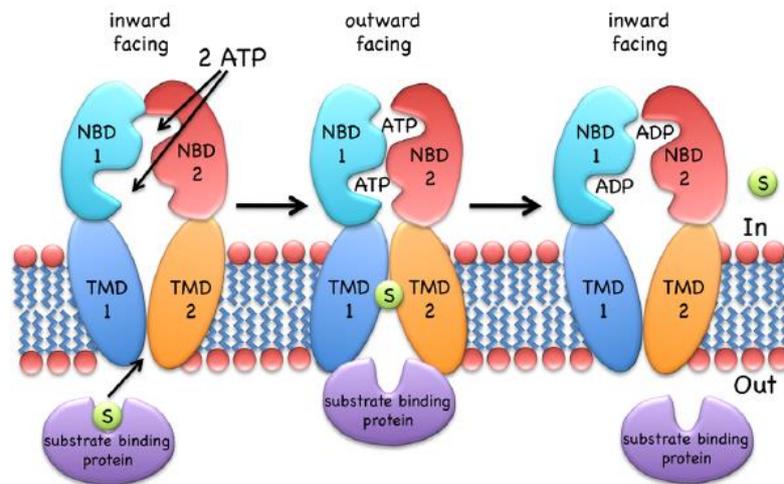


Figure 1.4 – The transport mechanism of Type I importers (exemplified by MalEFGK₂). The inward-facing type I transporter (e.g., MalFGK₂) binds to the substrate through the SPB. NBDs dimerize and result in the outward-facing conformation, allowing the substrate to contact with the TMDs. ATP is hydrolyzed and product release, together with NBD dissociation, resets the transporter to the inward-facing conformation. Adapted from Wilkens (2015).

1.3.4. The solute binding protein (SBP)

In Type I transporters (and also in Type II) the substrate is delivered to the transmembrane domains through a soluble substrate-binding protein (SBP) located in the trans-side of the membrane (Quioco and Ledvina, 1996; Berntsson *et al*, 2010). The SBPs present in Gram-negative bacteria constitute soluble proteins with 30-40 kDa that are freely diffused through the periplasm, while in Gram-positive bacteria and archaea they are anchored to the membrane through a lipid or a separate TM helix (Sutcliffe and Russell, 1995; Biemans-Oldehinkel *et al*, 2006; Eitinger *et al*, 2011; Wilkens, 2015).

SBPs are able to bind to a wide variety of substrates therefore displaying different affinities that range from the nanomolar to the micromolar. The sequence and size of the SBPs are also determined by the type of binding substrate, however the general architecture is highly conserved with two symmetrical domains or lobes that are connected via a hinge region (Quioco and Ledvina, 1996; Davidson *et al*, 2008; ter Beek *et al*, 2014; Eitinger *et al*, 2011).

The maltose-binding protein (MalE), a constituent of the maltose importer (MalEFGK₂) from *E. coli*, share the same mechanism of substrate binding than the others SBPs: the Venus fly trap model (Figure 1.5; Quioco and Ledvina, 1996; ter Beek *et al*, 2014). Basically in the absence of a ligand, the two lobes adopt predominantly an open conformation that changes to a close conformation when the substrate binds to the SBPs becoming trapped inside. The substrate is released into the TMDs through the interaction of each lobe of the SBPs with the respective two domains of the TMDs.

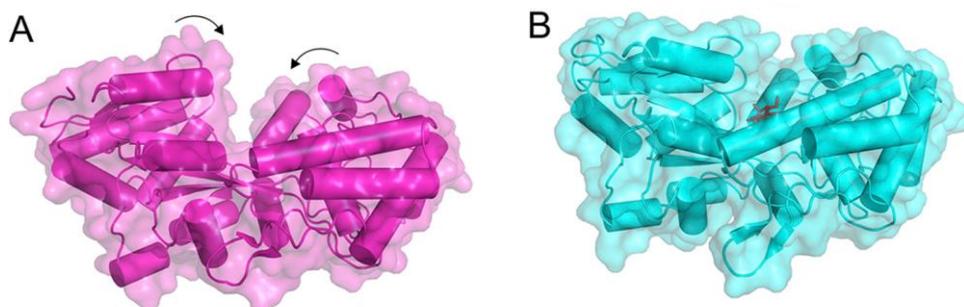


Figure 1.5 - Rearrangements in SBP MalE upon the substrate binding. (A) In the substrate-free form (Protein Data Bank accession no. 1ANF), the cavity between two protein lobes connected by the hinge is accessible. (B) Upon the binding of substrate maltose (dark sticks; Protein Data Bank accession no. 1EZ9), the cavity becomes occluded. Adapted from ter Beek *et al*, 2014.

1.3.5. The transmembrane domains (TMDs)

The two transmembrane domains form a translocation pore in the membrane that allows the passage of the substrate from the cis-side to the trans-side. These domains are mainly constituted by 4 to 10 membrane-spanning α -helices (Eitinger *et al*, 2011) and in type I importers the two domains can be either identical (homodimers) or structurally similar (heterodimers). In the maltose transporter MalFGK₂ the TMDs, MalF and MalG, displaying 8 and 6 helices, respectively, form a heterodimer, which are structurally related but not sequence related since they only share 13% of amino acid sequence identity (ter Beek *et al*, 2014; Rice *et al*, 2014). The TMDs primary sequence is less conserved than the NBDs but they share a similar topology characteristic from each transporter class.

The transmembrane domain MalF plays an important role in the maltose transporter, besides the formation of the pore, due to two unique features: an additional loop (P2) that interacts with the SBP MalE and a substrate-binding site for the maltose (Oldham *et al*, 2007). The P2 loop acts like a receptor that recognizes the SBP inducing an activated conformational change of the MalE, while maintaining the MalE and MalF in close contact throughout the catalytic cycle (Daus *et al*, 2009; Jacso *et al*, 2009 and 2012; Rice *et al*, 2014).

Upon the release of the substrate into the transmembrane domains, one maltose molecule binds to a unique site on the MalF domain composed by 10 residues that interact with the molecule through H-bonds, van der Waals interactions and aromatic ring stacking (Oldham *et al*, 2007; Eitinger *et al*, 2011). These residues were identified from crystallographic studies and mutagenesis experiments (Chen, 2013; Oldham and Chen, 2011).

MalG, the other transmembrane domain, also displays two important functions alongside with the formation of the translocation pathway. MalG P3 loop, the “scoop”, is inserted in the SPB binding site and promotes the displacement of the sugar, facilitating an efficient transfer into the membrane pore (Figure 1.6; Oldham *et al*, 2007). Another important interaction of the MalG is the insertion of its C-terminal tail into the MalK dimer interface (the ATPase), more specifically through the interaction with the Q-loop of each monomer, which may represent an important factor for the formation of the catalytic intermediate conformation of the entire transporter (Oldham *et al*, 2007).

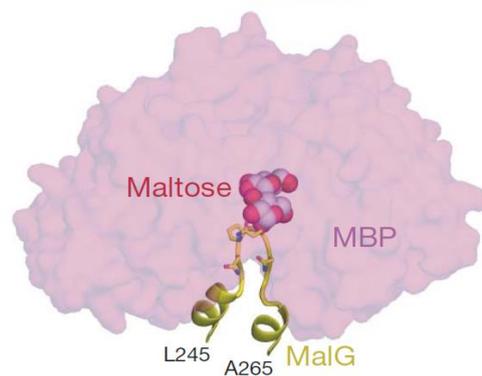


Figure 1.6 - Transfer of the maltose from MBP to the TM binding site. Insertion of the MalG scoop loop into the substrate-binding site of MBP. A maltose molecule is modelled into the binding site on the basis of the crystal structure of open maltose-bound MBP (PDB accession number 1JW5). Adapted from Oldham *et al*, 2007.

An essential feature of the two TMDs MalF and MalG is the so called “coupling helices” that consists in two short helices per domain with a characteristic “EAA” motif. This feature is an architecturally conserved element that forms the NBD–TMD interface, where it contacts the Q-loop in the NBDs grooves (Figure 1.7; Locher, 2009; Oldham *et al*, 2007).

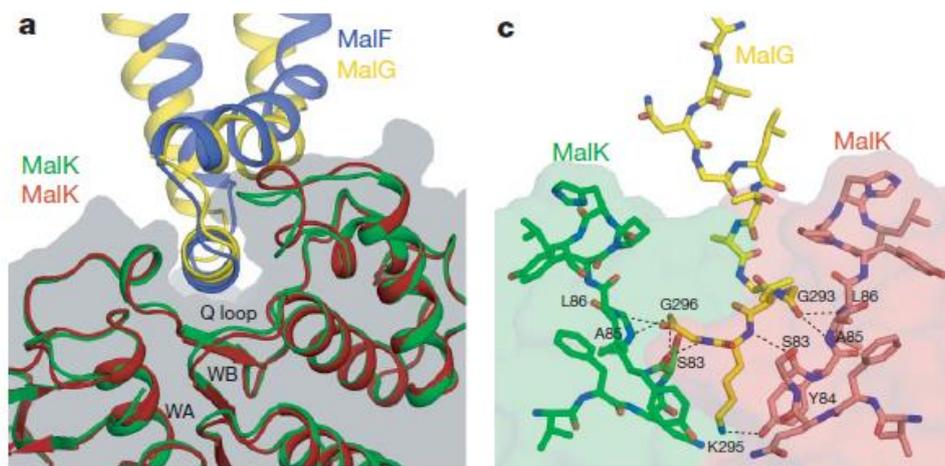


Figure 1.7 - The TMD–MalK interface. a) Docking of the EAA loops into a surface cleft of MalK. The EAA loops of MalF and MalG are compared by superposition of the two MalK subunits. The MalK dimer is also shown as a transparent surface model. WA, Walker A motif; WB, Walker B motif. b) Insertion of the MalG C-terminal tail 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.

1.3.6. The nucleotide binding domains (NBDs)

The nucleotide binding domains (NBD), also called ATPases, are considered the “motor domains” of the ABC-type transporters, since they provide the necessary energy to induce the conformational changes in the TMDs. The NBDs exhibit a highly conserved structure and sequence with several conserved motifs among all ABC transporters, being considered the hallmark of the family. They function in the dimeric form and depend on magnesium ions for catalysis, each monomer consists of two subdomains: a RecA-like and an α -helical subdomain that are interconnected by two flexible loop regions. The RecA-like subdomain is found in other P-loop ATPases while the α -helical subdomain is unique to the ABC transporters and presents a more structural diversity (Davidson *et al*, 2008). In the maltose transporter MalFGK₂, the NBDs are the two MalK units, whose genes are encoded in the transporter operon. MalK ATPase is a model for the NBDs structure in type I importers, as the whole transporter is a model for this type of importers.

NBDs can be identified at the sequence level by a specific set of seven highly conserved motifs (Figure 1.8; ter Beek *et al*, 2014):

(1) The A-loop helps to position the ATP molecule through stacking with the adenine ring of the conserved aromatic residue (usually a tyrosine);

(2) The P-loop or Walker A motif (GXXGXGK(S/T)) is a phosphate-binding loop that contains the highly conserved lysine residue, which form a network of interactions with two phosphates of the ATP molecule;

(3) The Walker B motif ($\phi\phi\phi\phi$ DE, where ϕ is a hydrophobic amino acid) seems to perform two functions: helps to coordinate the magnesium ion via the conserved aspartate residue and polarizes the attacking water through the glutamate residue that functions as a general base (ex: maltose transport; Oldham and Chen, 2011);

(4) The D-loop (motif: SALD) directly follows the Walker B motif, and affect the geometry of the catalytic site helping to form the ATP hydrolysis site through conformational changes;

(5) The H-loop (or switch region) contains a highly conserved histidine residue that forms a hinge between a beta strand and an alfa helix near the C terminus of the NBD, holding together, through interaction the γ -phosphate, the attacking water and the catalytic glutamate for catalysis (Oldham and Chen, 2011);

(6) The Q-loop possess a conserved glutamine residue that binds to the Mg^{2+} cofactor in the active site, therefore its mobility is essential to complete the catalytic cycle of the transporter (Daus *et al*, 2007). It is located at the interface between the RecA-like subdomain and the alpha-helical subdomain, as well as at the interface to the TMDs constituting a major site of interaction with the “coupling helix” of the TMDs (Dawson *et al*, 2007; Eitinger, *et al* 2011; ter Beek *et al*, 2014).

(7) The ‘LSGGQ’ motif (C-loop) is the ABC signature motif, representing a characteristic feature of the ABC superfamily (Schneider and Hunke, 1998). This motif is located in the α -helical subdomain, in the N-terminal end of a long helix that directs the positive charge of the helical dipole toward the γ -phosphate of ATP.

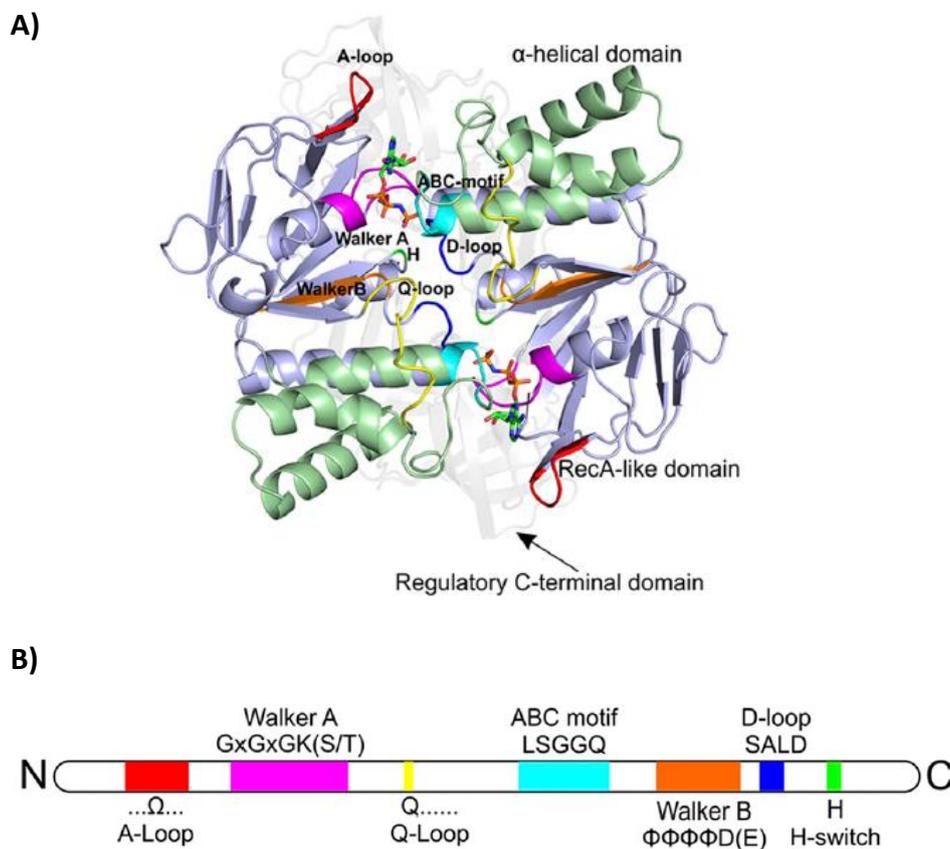


Figure 1.8 - The structure of the NBDs, as exemplified by the MalK dimer of the maltose transporter MalEFGK₂ (Protein Data Bank accession no. 3RLF). **(A)** View along an axis perpendicular to the membrane plane from the trans-side onto the NBDs (The TMDs and SBP have been removed for clarity). Domains and highly conserved sequence motifs are color-coded: green, α -helical domain; light blue, RecA-like domain; faded gray, regulatory C-terminal domain; red, A-loop; magenta, Walker A; orange, Walker B; blue, D-loop; green, H-loop; cyan, ABC motif; yellow, Q-loop. The ATP analogue AMP-PNP is shown in sticks. **(B)** The relative positions of sequence motifs in NBDs. Adapted from ter Beek *et al*, 2014.

In the outward-facing conformation of the maltose transporter the NBDs form a “sandwich dimer” with two molecules of ATP, where one molecule is bound to NBD1 coordinated by P-loop residues from NBD1 and from residues of the signature sequence of NBD2 and vice versa for the second ATP. The number of ATP molecules that need to be hydrolysed to accomplish a complete transport cycle is not universal for all ATP transporters (Davidson and Sharma, 1997).

The interconnection between the NBDs with the TMDs depend on two key elements: the Q-loop from the NBDs and the “coupling helix” from the TMDs (Figure 1.7; Wilkens, 2015). This two elements allow the transference of motion from the NBDs to the TMDs, that come from the rotational movement of the RecA-like domain with respect to the NBD helical domain during the catalysis, which was shown through crystallographic and EPR spectroscopy experiments with the maltose ATP-binding cassette transporter (Khare *et al*, 2009; Orelle *et al* 2010; Wilkens, 2015).

Maltose transporter is regulated by IIA^{glc}, an enzyme from PEP-dependent sugar PTS system, that inhibits transport activity or through the binding of MalT, a transcription factor (activator) that belongs to the maltose transporter (Boos and Shuman, 1998). The inhibition of the transport is accomplish by the binding of two IIA^{glc} with the MalK dimer (Chen *et al*, 2013), which prevents the closure of the MalK dimer and maintain MalFGK₂ in the inward-facing resting state (Rice *et al*, 2014).

1.3.7. Multitask ATPases

ATPases or the NBDs are essential components to the ABC-type transporters, however a closer look to the gene clusters of some species, that encode sugar ABC transporters, demonstrated that the nucleotide-binding domains sequences are occasionally absent from the cluster. Early evidence that an ATPase was capable of energizing more than one ABC transporter system was observed with the MsiK ATPase, whose gene was not encoded in an ABC transporter operon. This ATPase is shared by the cellobiose and the maltose ABC transport systems in *Streptomyces reticuli* and *S. lividans* (Schlösser *et al*, 1997).

In *B. subtilis*, the MsmX ATPase is the multitask ATPase that provides energy to multiple ABC transporters (Ferreira and Sá-Nogueira, 2010 and unpublished data), while the *msmX* gene

is not encoded in the transporters operon but in a monocistronic message (see above, section 1.1.1).

Several reports have demonstrated that some sugar ABC transporters systems share an ATPase with greater incidence in Gram-positive bacteria (Schlösser *et al*, 1997; Ferreira and Sá-Nogueira, 2010; Marion *et al*, 2011b; Tyx *et al*, 2011; Tan *et al*, 2015;) however, the existence of this phenomenon in a Gram-negative bacteria is also observed (Silva *et al*, 2005; Chevance *et al*, 2006). So far the only examples of a multitask ATPase in a pathogenic bacteria is the MsmK ATPase in different species of the *genus Streptococcus* namely, *S. pneumoniae* (Marion *et al*, 2011b; Tyx *et al*, 2011), *S. suis* (Tan *et al*, 2015) and *S. mutans* (Webb *et al*, 2008).

1.3.7.1. MsmK, a multitask ATPase in *Streptococcus* species

S. pneumoniae or pneumococcus (a Gram-positive bacteria), is an opportunistic respiratory human pathogen that can cause diseases as otitis media, meningitis and pneumonia, being the last two diseases a major cause of death. This pathogen possesses a vast ability in the utilization of carbohydrates which may provide a competitive advantage in the bacterial population of the nasopharynx. Since this organism uses as a carbon source only carbohydrates their import is ensured by 30% of the transport mechanisms of the organism encoded by the genome, being ABC transporters particularly important (Tyx *et al*, 2011, Buckwalter *et al*, 2012). There are six or seven (depending on the strain) predicted carbohydrate uptake transporter family 1 (CUT1) ABC importers within the pneumococcal genome, which lack in each locus a gene coding for an ATPase required to energize the transporter (Buckwalter *et al*, 2012). In *S. pneumoniae* TIGR4 there are three ABC importers from CUT1 family, the RafEFG, SatABC and MalXCD transporters, for which MsmK ATPase is the component that provides energy to the systems (Marion *et al*, 2011b). In *S. pneumoniae*, RafEFG is responsible for the uptake of raffinose (Rosenow *et al*, 1999), while SatABC transports sialic acid (Marion *et al*, 2011) and MalXCD provides the uptake of maltooligosaccharides (Puyet *et al*, 1993; Abbott *et al*, 2010). The MsmK ATPase also contributes to pneumococcal colonization of the *S. pneumoniae*, which suggests that transport of at least one of the carbohydrate substrates is important during colonization (Marion *et al*, 2011b). An example is the transport of sialic acid that was associated with cell signalling during the chain of events of the biofilm formation, colonization and host invasion of the organism (Trappetti *et al*, 2009; Marion *et al*, 2011).

In *S. suis*, an emerging important pathogen that is causing deadly infections in pigs and in humans, MsmK was also identified as a multitask ATPase that provides energy to MsmEFG and MalXCD transporters. Likewise, the ATPase gene is not encoded in the operons of these ABC transporters, and is responsible for the uptake of raffinose and melibiose through MsmEFG transporter and maltotriose, maltotetraose and maltodextrins (product of glycogen degradation) by MalXCD transporter, thereby contributing to the colonization and the *in vivo* survival of *S. suis* (Tan *et al*, 2015).

In *Streptococcus* species, namely *S. pneumoniae*, *S. pyrogens* and *S. suis*, the CUT1 family of ABC importers are usually encoded in operons lacking the gene that codes for an ATPase. However, in *S. mutans* the MsmEFGK and MalXFGK carbohydrates transporters of the CUT1 family both possess a gene encoding for the respective ATPase (MsmK and MalK) together with the genes that encode for the two transmembrane proteins and the solute binding domain in the same operon (Webb *et al*, 2008). The multiple sugar metabolism system (MsmEFGK) is responsible for the uptake of melibiose, raffinose, isomaltotriose, stachyose and isomaltose (Russel *et al*, 1992; Tao *et al*, 1993), and the MalXFGK is involved in the uptake of maltotriose, maltotetraose and other maltodextrins which are zymolytic products of pullulan and glycogen (Webb *et al*, 2008). The special feature about these transporters is that in spite of having their own ATPase, when suppression of one of these proteins occurs the other protein is capable of providing energy to the alternative system (Webb *et al*, 2008). A schematic presentation of the *msm* and *mal* loci encoding ABC transporters and a schematic of the carbohydrates utilized in *S. mutans*, *S. pneumoniae* and *S. suis*, are represented in Figure 1.9 and 1.10, respectively.

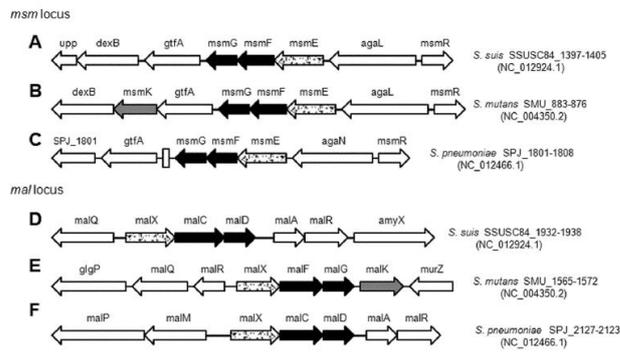


Figure 1.9 - Schematic presentation of the *msm* and *mal* loci encoding relevant carbohydrate ABC transporters in selected *streptococci*. The genes within the loci encoding components of the *S. suis* MsmEFG (A), *S. mutans* MsmEFGK (B), *S. pneumoniae* MsmEFG (C), *S. suis* MalXCD (D), *S. mutans* MalXFGK (E) and *S. pneumoniae* MalXCD (F) are represented. Arrows indicate the direction of transcription. Gray arrows, genes encoding the ATPase of ABC transporters; spotted arrows, genes encoding solute binding proteins; black arrows, genes encoding permeases of ABC transporters; blank arrows, other genes adjacent. Adapted from Tan *et al.* 2015

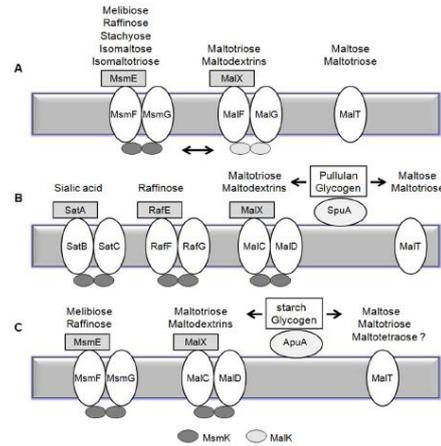


Figure 1.10 - Schematic summary of carbohydrates utilization by ABC transporters in *S. mutans* (A), *S. pneumoniae* (B) and *S. suis* (C). Above each ABC complexes is a list of known or putative carbohydrates transported by each ABC transporter. Bidirectional arrow means the MsmK and MalK ATPases can energize permeases interactively. Unidirectional arrows mean the degradation of pullulan and glycogen by pullulanases SpuA or ApuA. Adapted from Tan *et al.*, 2015

1.3.7.2. Multitask ATPases from *Streptomyces* species and *Thermus thermophilus*.

Streptomyces species hydrolyse chitin to oligosaccharides being (GlcNAc)₂ and chitosan the main products. In *S. coelicolor* A3(2) the (GlcNAc)₂ uptake occurs via the ABC transporter DasABC (Saito *et al.*, 2007) and the chitosan-derived oligosaccharides uptake is accomplished by the ABC transporter CsnEFG (Viens *et al.*, 2015). The respective operons of both transporters lack a gene encoding an ATPase, however it was demonstrated that the MsiK protein is the ATPase responsible for providing energy to the system (Saito *et al.*, 2008; Viens *et al.*, 2015). In *S. lividans* and *S. reticuli* the MsiK ATPase is shared by two distinct ABC transporters responsible for the uptake of the disaccharides cellobiose and maltose (Schlösser *et al.*, 1997). Additionally in *S. reticuli* the MsiK ATPase was also associated with the ABC transporter responsible for the uptake of trehalose (Schlösser *et al.*, 2000).

To date the phenomenon of a multitask ATPase is mainly associated with Gram-positive bacteria however, in spite of being less frequent, it also occurs in Gram-negative bacteria, as reported in *Thermus thermophilus*. The MalK1 ATPase, which is homologous to MalK from *E. coli*, and is encoded in a monocistronic gene is responsible for energizing two distinct ABC importers: the trehalose/maltose/sucrose/palatinose (TMSP) ABC transporter (Silva *et al*, 2005) and the glucose/mannose ABC transport system (Chevance *et al*, 2006).

1.4. Scope of the Thesis

In this study we will characterize multipurpose ATPases from both Gram-positive and Gram-negative bacteria and assess their intra- and interspecies interchangeability in the host *B. subtilis*. A genetic system is fine-tuned to test the ability of ATPases from other species to complement MsmX function and establish *B. subtilis* as model for the study of bacterial multitask ATPases.

Chapter 2

Materials and Methods

2. Materials and Methods

2.1. Substrates

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

2.2. Bioinformatic Analysis

For the identification of MsmX homologs in other Gram-positive and Gram-negative bacteria, bioinformatics tools were used. BLASTp algorithm was the tool chosen 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 40% were considered as targets. An amino acid sequences alignment was made using the multiple sequence alignment program Clustal Omega (EMBL-EBI). Protein 3-D structures were predicted using the online program I-TASSER created by Zang Lab, University of Michigan (Zhang, 2008; Roy *et al*, 2010; Yang *et al*, 2015).

2.3. Isolation of chromosomal DNA

Chromosomal DNA was extracted from *Escherichia coli* K-12 strain based on the method described by Ferrari *et al* (1982). The strain was grown overnight (37 °C, 180 rpm), in liquid Luria-Bertani (LB) medium (Miller, 1972). All subsequent centrifugations were performed at 16060 *g*. The cells were harvested in two tubes with 2 mL of culture each by centrifugation for 2 minutes, washed once with 50 mM Tris and 5 mM EDTA, and resuspended in 175 μ L of 50 mM Tris, 5 mM EDTA, lysozyme 1 mg/mL and RNase 20 μ g/mL with incubation at 37 °C for 30 minutes. After, the solutions were vigorously agitated for 5 minutes and further incubated at the same temperature for at least 15 minutes, followed by a centrifugation step of 10 minutes. 100 μ L of phenol (saturated with Tris-HCl, pH=8) was added to each tube, followed by centrifugation for 5 minutes. Each aqueous phase was recovered and mixed with 100 μ L of chloroform:isoamyl alcohol (24:1), followed by centrifugation for 3 minutes. The upper phase from the two tubes, was mixed and two volumes of absolute ethanol were added. The precipitated DNA was

collected by centrifugation for 20 minutes at 4 °C, dried and further resuspended in 1X TAE buffer.

2.4. 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 manufacturer's recommendations. PCR amplifications were carried out using Phusion® high-fidelity DNA polymerase (Thermo Fisher Scientific Inc.). Oligonucleotides designed in this work or in previous experiments performed in our laboratory, were purchased from Metabion International AG or StabVida, Lda (Table 2.2). 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 NZYMiniprep kit (NZYTech, Lda). DNA was sequenced using the method of Sanger performed at StabVida, Lda.

2.5. Site-directed mutagenesis by Overlapping PCR

Chromosomal DNA of strain *Escherichia coli* K-12 was used as template for site-directed mutagenesis by primer extension, using mutagenic oligonucleotides ARA862 and ARA863 and flanking oligonucleotides. This pair of mutagenic primers allowed the insertion of a nucleotide C in codon 320, thereby restoring the wild-type phenotype of the protein YcjV with a total length of 360 amino acids. Primers ARA860 and ARA863 created fragment AB and primers ARA862 and ARA861 generated fragment CD, which through overlapping PCR resulted in a fragment of *ycjV* gene with *HindIII* and *SphI* restriction sites, provided by the flanking primers ARA860 and ARA861. Another fragment containing *ycjV* gene and *NheI* and *BglII* restriction sites, harbored in the flanking primers ARA845 and ARA846, was generated through the junction of fragment AB, created by primers ARA845 and ARA863, and fragment CD created by primers ARA862 and ARA846. For both experiments, two polymerase chain reactions were carried on using 1x Phusion®HF Buffer (Thermo Fisher Scientific Inc.), 0.5 µM primers, 200 µM dNTPs, 1.2 ng/µl of template genomic DNA (in the first reaction) and 0.2 ng/µl of Template DNA (in the second reaction: fragments AB, CD) and 0.05 U/µl of Phusion®High-Fidelity DNA Polymerase (Thermo

Fisher Scientific Inc.) in a total volume of 50 µl. The insertion of the nucleotide was confirmed by DNA sequencing.

2.6. Construction of plasmids and strains

Plasmid pPS1 was obtained by amplification of the *HD73_RS21400* gene from chromosomal DNA of strain *Bacillus thuringiensis* serovar *kurstaki* str. HD73 (Bacillus Genetic Stock Center, BGSC, Ohio State University), with the oligonucleotides ARA851 and ARA852, bearing unique restriction sites *NheI* and *BglII*, and subsequent cloning of this fragment (1082 bp) into pSN74 digested with *NheI* and *BglII*. pPS2 was obtained by 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 ARA843 and ARA844, which harbor unique restriction sites *NheI* and *BglII*, and then the resulting fragment (1079 bp) was inserted between the *NheI* and *BglII* sites of pSN74. The amplification of the *malk* gene, from chromosomal DNA of strain *Escherichia coli* K-12, with oligonucleotides ARA847 and ARA848, which contain unique restriction sites *NheI* and *BglII*, and subsequent cloning of this fragment (1097 bp) into pSN74 *NheI*-*BglII*, yielded plasmid pPS3. Another amplification of the *malk* gene was made using oligonucleotides ARA858 and ARA859, which contain unique restriction sites *HindIII* and *SphI*, and cloning this fragment (1220 bp) into pDR111 (a gift from David Rudner, Harvard University) digested with *HindIII* and *SphI*, yielded plasmid pPS6. Plasmid pPS4 was obtained by amplification of the *yurJ* gene, from chromosomal DNA of the wild-type strain *Bacillus subtilis* 168T⁺, with oligonucleotides ARA837 and ARA838, bearing unique restriction sites *NheI* and *BglII*, and subsequent cloning of this fragment (1085 bp) into pSN74 digested with *NheI* and *BglII*. Plasmid pPS5 was obtained by amplification of *msmK* gene using pAM7 as template with oligonucleotides ARA855 and ARA749, which contain unique restriction sites *Sall* and *SphI*, and subsequent cloning of this fragment (1241 bp) into pDR111 digested with *Sall* and *SphI*.

Plasmid pPS7 was obtained by amplification of a DNA fragment from pSN74 containing the *lacI* gene and the terminal end of the *msmX* gene with oligonucleotides ARA854 (mutagenic primer contain a unique restriction site *BglII* and two novel codons) and ARA632. This procedure introduced two novel amino acids (Leu, Glu) into the *msmX* coding region preceding the C-terminal His-tag were introduced into the *msmX* coding region of plasmid pSN74. The resulting amplification product was digested with *BglII* and *BamHI* and the product (1452 bp) sub cloned

into pSN74 *Bgl*II-*Bam*HI, yielding pPS7. *msmK* gene was amplified from pAM7 with the oligonucleotides ARA773 and ARA774 (Sá-Nogueira, unpublished), which harbour unique restriction sites *Nhe*I and *Bgl*II, respectively, and digested with *Nhe*I and *Bgl*II resulting in a fragment with 1112bp. All genes of *msmX* homologs amplified as described above and digested with *Nhe*I and *Bgl*II were further cloned into pPS7 *Nhe*I-*Bgl*II, yielding the following plasmids: pPS8 (*HD73_RS21400*), pPS9 (*msmK*), pPS10 (*ugpC*), pPS11 (*malk*), and, pPS12 (*yurJ*). The amplification of *ycjV* gene from chromosomal DNA of strain *E. coli* K-12, obtained by site-directed mutagenesis (described in the previous topic), yielded a fragment with unique restriction sites *Hind*III and *Sph*I and another fragment with unique restriction sites *Nhe*I and *Bgl*II. Plasmid pPS13 was obtained by cloning a fragment (1191bp) containing *ycjV* gene, digested with *Hind*III and *Sph*I, into pDR111 *Hind*III-*Sph*I. *ycjV* gene digested with *Nhe*I and *Bgl*II resulted in a fragment with 1064bp, being subsequently cloned into pPS7 *Nhe*I-*Bgl*II which yielded plasmid pPS14.

Plasmids and oligonucleotides used in this work are listed in Table 2.1 and 2.2, respectively. Oligonucleotides ARA411, ARA430, ARA442, ARA662, ARA741 and ARA841 were used for DNA sequencing (see Appendices 6.2 to 6.17)

Table 2.1 – List of plasmids used in this work.

Plasmids	Relevant construction	Source or Reference
pDR111	Derivative of the Pspac(hy) plasmid pJQ43; contains an additional <i>lacO</i> binding site	David Rudner*
pGS1	pMS38 derivate used for the fusion of a C-terminal His-tag to <i>msmX</i> at his own locus, <i>bla</i> , <i>cat</i>	Ferreira and Sá-Nogueira, unpublished
pMJ2	pMJ1 derivate used for the creation of <i>msmX</i> -null mutations, <i>bla</i> , <i>cat</i>	Ferreira and Sá-Nogueira, 2010
pAM7	pDR111 derivate, with <i>msmK</i> under the control of Pspank(hy)	Mendes and Sá-Nogueira, unpublished
pAM12	pDR111 derivate, with <i>msmX</i> under the control of Pspank(hy), and with <i>Bgl</i> III and <i>Nhe</i> I restriction sites in the <i>msmX</i> coding region	Mendes and Sá-Nogueira, unpublished
pSN74	pDR111 derivate, with an additional C-terminal His-tag in the coding region of <i>msmX</i> from pAM10 under the control of Pspank(hy)	Sá-Nogueira, unpublished*
pSN75	pSN74 derivate, with <i>msmK</i> under the control of Pspank(hy) and with a C-Terminal His-tag	Sá-Nogueira, unpublished*
pPS1	pSN74 derivate, with <i>HD73_RS21400</i> under the control of Pspank(hy) and with a C-Terminal His-tag	This work*
pPS2	pSN74 derivate, with <i>ugpC</i> under the control of Pspank(hy) and with a C-Terminal His-tag	This work*
pPS3	pSN74 derivate, with <i>malk</i> under the control of Pspank(hy) and with a C-Terminal His-tag	This work*
pPS4	pSN74 derivate, with <i>yurJ</i> under the control of Pspank(hy) and with a C-Terminal His-tag	This work*
pPS5	pDR111 derivate, with <i>msmK</i> under the control of Pspank(hy)	This work*
pPS6	pDR111 derivate, with <i>malk</i> under the control of Pspank(hy)	This work*
pPS7	pSN74 derivate, with the insertion of two amino acids (Leu, Glu) in <i>msmX</i> sequence, followed by the C-Terminal His-tag	This work*
pPS8	pPS7 derivate, with <i>HD73_RS21400</i> under the control of Pspank(hy), and with a C-Terminal LEHis-tag	This work*
pPS9	pPS7 derivate, with <i>msmK</i> under the control of Pspank(hy), and with a C-Terminal LEHis-tag	This work*
pPS10	pPS7 derivate, with <i>ugpC</i> under the control of Pspank(hy), and with a C-Terminal LEHis-tag	This work*
pPS11	pPS7 derivate, with <i>malk</i> under the control of Pspank(hy), and with a C-Terminal LEHis-tag	This work*
pPS12	pPS7 derivate, with <i>yurJ</i> under the control of Pspank(hy), and with a C-Terminal LEHis-tag	This work*
pPS13	pDR111 derivate, with <i>ycjV</i> under the control of Pspank(hy)	This work*
pPS14	pPS7 derivate, with <i>ycjV</i> under the control of Pspank(hy), and with a C-Terminal LEHis-tag	This work*

*See Appendices 6.1 to 6.17

Table 2.2 – List of oligonucleotides used in this work.

Oligonucleotides	Sequence ^a
ARA411	CTCTTGCCAGTCACGTTACG
ARA430	GCGACATCGTATAACGTTACTGG
ARA442	GAGCTGCCTGCCGCGTTTCGGTG
ARA632	AGAGCTGTACGCAGCCGCTG
ARA662	ACATCGGCAAAGAGGTCATC
ARA741	CTTG <u>TGTCGAC</u> AGGGAATTGCTG
ARA749	GCACGCATGC <u>ACTGATATCTCTCC</u>
ARA773	CACCTCATGGCTAGCTTGAATCTTA
ARA774	GCCGCCAGATCTTTTTTTCAGTTTCTAC
ARA837	CTGATGGCTAGCTTAACATTTGAACACG
ARA838	TAAGATCTCTCTCCGTTCCGCATCG
ARA841	GTACATAATGGATTCCTTACG
ARA843	GCCAACATGGCTAGCTTAAAGTTAG
ARA844	GCCGCCAGATCTATTTCTGTTTTTTC
ARA845	ATCAACATGGCTAGCCTTTCGTTAC
ARA846	TTGGCGAGATCTTATTTCCGTTTCTGC
ARA847	AAGTTTATGGCTAGCGTACAGCTGC
ARA848	TTAGCCAGATCTCTCCTTATGCAGTCG
ARA851	GGTACCATGGCTAGCCTTAAATTAG
ARA852	TTAGCGAGATCTTTGTTTCAGTTTGG
ARA854	GAAGATCTCGACTCGAGCACCACCATCACCACCACTAAGATC
ARA855	CGCGCCGTCGACTATATAATATAATTATC
ARA858	CGGAAGCTTGCTGTCGATGACAGG
ARA859	CAGTGCATGCCTCCTGAGTCATTGC
ARA860	TGGCCAAGCTTATCGGCCTTCTG
ARA861	CACATCAGGCATGCGGTACAGGG
ARA862	TTGGGGGCACGAGTTAGTGG
ARA863	ACTAACTCGTGCCCCCAACC

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

Plasmid pDR111 was used to transform the *B. subtilis* wild-type strain 168T⁺, according to the method described by Anagnostopoulos and Spizizen (1961), yielding strain ISN1. Plasmids pSN74, pSN75 and pPS1 to pPS14 were used to transform the *B. subtilis* strain IQB495, resulting in strains ISN2 to ISN17 (Table 2.3), respectively. The *amyE*⁻ phenotype of the transformants was confirmed 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) I₂-5.0% (w/v) KI for detection of starch hydrolysis.

Table 2.3 – List of *B. subtilis* strains used or constructed during this work.

Strain	Relevant genotype	Source or Reference ^a
168T ⁺	Prototroph	F. E. Young
IQB495	$\Delta msmX::cat$	pMJ2 ^b → 168T ⁺ (Ferreira and Sá-Nogueira, 2010)
IQB622	<i>msmX</i> -LEHis ₆ <i>cat</i>	pGS1 → 168T ⁺ (Ferreira and Sá-Nogueira, unpublished)
IQB672	$\Delta msmX::cat \Delta amyE::Pspank(hy)$	pDR111 ^b → IQB495 (Mendes and Sá-Nogueira, unpublished)
IQB676	$\Delta msmX::cat \Delta amyE::Pspank(hy)$ - <i>msmX</i> (Glu3Ser, Ile364Ser)	pAM12 ^b → IQB495 (Mendes and Sá-Nogueira, unpublished)
ISN1	$\Delta amyE::Pspank(hy)$ - <i>spec</i>	pDR111 ^b → 168T ⁺
ISN2	$\Delta msmX::cat \Delta amyE::Pspank(hy)$ - <i>msmX</i> -His ₆	pSN74 ^b → IQB495
ISN3	$\Delta msmX::cat \Delta amyE::Pspank(hy)$ - <i>msmK</i> -His ₆	pSN75 ^b → IQB495
ISN4	$\Delta msmX::cat \Delta amyE::Pspank(hy)$ - HD73_RS21400-His ₆	pPS1 ^b → IQB495
ISN5	$\Delta msmX::cat \Delta amyE::Pspank(hy)$ - <i>ugpC</i> -His ₆	pPS2 ^b → IQB495
ISN6	$\Delta msmX::cat \Delta amyE::Pspank(hy)$ - <i>malk</i> -His ₆	pPS3 ^b → IQB495
ISN7	$\Delta msmX::cat \Delta amyE::Pspank(hy)$ - <i>yurJ</i> -His ₆	pPS4 ^b → IQB495
ISN8	$\Delta msmX::cat \Delta amyE::Pspank(hy)$ - <i>msmK</i>	pPS5 ^b → IQB495
ISN9	$\Delta msmX::cat \Delta amyE::Pspank(hy)$ - <i>malk</i>	pPS6 ^b → IQB495
ISN10	$\Delta msmX::cat \Delta amyE::Pspank(hy)$ - <i>msmX</i> -LEHis ₆	pPS7 ^b → IQB495
ISN11	$\Delta msmX::cat \Delta amyE::Pspank(hy)$ - HD73_RS21400-LEHis ₆	pPS8 ^b → IQB495
ISN12	$\Delta msmX::cat \Delta amyE::Pspank(hy)$ - <i>msmK</i> -LEHis ₆	pPS9 ^b → IQB495
ISN13	$\Delta msmX::cat \Delta amyE::Pspank(hy)$ - <i>ugpC</i> -LEHis ₆	pPS10 ^b → IQB495
ISN14	$\Delta msmX::cat \Delta amyE::Pspank(hy)$ - <i>malk</i> -LEHis ₆	pPS11 ^b → IQB495
ISN15	$\Delta msmX::cat \Delta amyE::Pspank(hy)$ - <i>yurJ</i> -LEHis ₆	pPS12 ^b → IQB495
ISN16	$\Delta msmX::cat \Delta amyE::Pspank(hy)$ - <i>ycjV</i>	pPS13 ^b → IQB495
ISN17	$\Delta msmX::cat \Delta amyE::Pspank(hy)$ - <i>ycjV</i> -LEHis ₆	pPS14 ^b → IQB495

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

^b The DNA was linearized with *ScaI*

2.7. Growth conditions

E. coli DH5 α (Gibco-BRL) was used as host for the construction of all plasmids. The strain was grown in liquid Luria-Bertani (LB) medium and on LB solidified with 1.5% (w/v) agar and ampicillin (100 μ g/mL). *B. subtilis* strains were grown in liquid LB medium, LB medium solidified with 1.5% (w/v) agar or SP medium (Martin *et al*, 1987) with chloramphenicol (5 μ g/mL) and/or

spectinomycin (60 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 IQB672, IQB676, IQB622, and ISN1 to ISN17, were grown overnight (37 °C, 180 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 approximately 0.05, in 1.5 mL of CSK medium without potassium succinate and supplemented with different carbon and energy sources (arabinose and arabinotriose) to a final concentration of 0.1% (w/v) and 1 mM of IPTG added when appropriated. The cultures were grown in sterile 50 mL Falcon tubes (Sarstedt), incubated at 37 °C and 180 rpm in an Aquatron® Waterbath Rotary Shaker and the OD_{600nm} periodically measured in an Ultrospec™ 2100 pro UV/Visible Spectrophotometer (GE Healthcare Life Sciences).

2.8. Protein extracts of *B. subtilis*

B. subtilis strains ISN10 to 15, ISN17 and IQB622 were grown in CSK minimal medium supplemented with 0.1% (w/v) of arabinose or arabinotriose, in the presence or absent of the inducer IPTG, as previously described for growth kinetic parameters (see previous topic). Cells from the strains ISN12-15 and ISN17 grown in CSK medium with arabinotriose 0.1%, in the presence or absent of the inducer IPTG, and cells from the strains ISN10 and ISN11 grown with arabinotriose 0.1% without IPTG, were collected after 8 h of experiment with O.D_{600nm} <0.4. Cells of the remaining growing cultures were collected when O.D_{600nm} reached 0.65-0.8. A total of 1.1 to 1.4 mL of each culture was centrifuged at 6000 *g* and 4° C for 5 minutes, and the sediment resuspended in 50 to 100 µL (depending on the O.D_{600nm} of the culture) of Lysis Buffer (500 mM KCl, 20 mM Hepes K⁺ pH 7.6, 10 mM EDTA, 1 mM DTT, 10% glycerol). Lysozyme (1 mg/mL) was added to the mixture and incubated for 10 minutes at 37 °C, followed by three cycles of freezing in liquid nitrogen and thawing for 5 minutes at 37 °C. 10 mM PMSF and 7.5 U of Benzonase® Nuclease (Sigma-Aldrich Co.) were added followed by an incubation for 15 minutes at 37 °C. The total protein content for each extract was determined using Bio-Rad Protein Assay (Bio-Rad Laboratories, Inc).

2.9. Western-Blot assay

Immunoblot assay was performed using the cell free extracts, to detect the presence of the recombinant ATPase in the cells. In this procedure, first it was made an electrophoretic separation of all proteins contained in each extract, from strains grown in arabinose or arabinotriose 0.1% (w/v), obtain in inducible or non-inducible conditions. 10 µg of total protein from each extract [20 µg of total protein from the extract of IQB622 grown in arabinose 0.1% (w/v)], and 0.6 µg of purified MsmX-LEHis₆ were load in 12.5% SDS-PAGE and run at a constant voltage (150 V) for approximately 1 h. The fractionated proteins were then electrotransferred into a nitrocellulose membrane (Bio-Rad) in a semi-dry Bio-Rad Trans-Blot® Turbo™ Transfer System, for 30 minutes at constant voltage (25 V) and constant amperage (1.0 A). Afterwards the membrane was stained with Ponceau S to verify the transference efficient. All subsequent incubation and wash steps were performed at room temperature with mild shaking. The membrane was incubated in blocking solution [powdered non-fat milk solution in 1X PBS-Tween20 0.1% (4% w/v)] for 1 h, followed by incubation with 10 mL of the primary antibody (mouse monoclonal Anti-6X His-tag® antibody [HIS.H8; Abcam], diluted 1:1000 in blocking solution) for 1 h. The membrane was then washed three times with PBS 1X-Tween20 0.1% (with two incubation periods of 10 minutes), and then incubated in 10 mL of the secondary antibody (HRP-conjugated donkey anti-mouse IgG antibody [Jackson ImmunoResearch Europe Ltd.], diluted 1:10000 in blocking solution) for 1 h. Finally, the membrane was washed three times with PBS 1X-Tween20 0.1% (with two incubation periods of 10 minutes). All subsequent steps were performed in a dark room. Blots were developed using 1 mL of SuperSignal™ West Pico Chemiluminescent Substrate (Thermo Scientific). Amersham Hyperfilm plates (GE Healthcare) were exposed to luminescence for 30 s or 40 s, 1 minute, 3 minutes and/or 4 minutes inside a Hypercassette Autoradiography Cassette (GE Healthcare).

2.10. Protein analysis

The analysis of the production, purification and molecular mass of the proteins was performed with SDS-PAGE (stained with Coomassie Blue), using Low Molecular Weight Protein Marker (NZYTech) as standard.

Chapter 3

Results and Discussion

3. Results and Discussion

3.1. Complementation analysis of MsmX homologs in *msmX*-null mutant *B. subtilis* strains

3.1.1 In silico Analysis

An in silico survey was conducted for the identification of MsmX homologs, in Gram-positive and Gram-negative bacteria, using BLASTp tool. From this analysis *Bacillus subtilis*, *Bacillus thuringiensis*, *Staphylococcus aureus*, *Streptococcus pneumoniae* and *Escherichia coli* proteins were selected (Table 3.1). The protein from *Bacillus thuringiensis* displays the higher identity (74%), while UgpC from *Staphylococcus aureus* exhibits an identity of 66%. MsmK from *Streptococcus pneumoniae* (Gram-positive bacteria) and YcjV from *Escherichia coli* (Gram-negative bacteria) display the same identity (64%) to MsmX from *B. subtilis*. On the other hand, the protein YurJ also from *B. subtilis*, has a lower identity (58%). Since Malk from *Escherichia coli* is a model ATPase it was included in this study, however this protein possesses the lower identity (45%). A multiple alignment of the proteins primary sequence was performed using ClustalOmega and is shown in the Appendix 6.18 (wild-type form) and 6.19 (recombinant form). The genomic context of each gene selected for study is illustrated in Appendices 6.20 to 6.25.

Table 3.1 – List of MsmX homologs selected for this work.

Species	Gene name	NCBI Reference	Protein Length	Identity ^a
<i>Bacillus thuringiensis</i> serovar <i>kurstaki</i> str. HD73	HD73_RS21400	WP_000818931.1	366	271 (74%)
<i>Staphylococcus aureus</i> subsp. <i>aureus</i> ST398	<i>ugpC</i> (or <i>SAPIG0223</i>)	CAQ48661.1 (GeneBank)	365	240 (66%)
<i>Streptococcus pneumoniae</i> TIGR4	<i>msmK</i> (or <i>SP_1580</i>)	AAK75666.1 (GeneBank)	376	241 (64%)
<i>Escherichia coli</i> str. K-12	<i>ycjV</i> (or <i>b4524</i>)	P77481.2 (UniProtKB)	360	233 (64%)
<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168	<i>yurJ</i> (or <i>BSU32550</i>)	NP_391135.1	367	212 (58%)
<i>Escherichia coli</i> str. K-12	<i>malk</i> (or <i>b4035</i>)	NP_418459.1	371	162 (45%)

^a Amino acid identity to *B. subtilis* MsmX is indicated

3.1.2 Fine-tuning of a Genetic System for Complementation Analysis

An *in vivo* system for the expression of MsmX and MsmX homologs was constructed in a *B. subtilis* *msmX*-null mutant (IQB495) by Mendes and Sá-Nogueira (unpublished work), in order to test if other ATPases found in pathogenic species from Gram-positive and Gram-negative bacteria, are able to play the role of MsmX in the cell. The genetic system consists of an integrative *B. subtilis* vector with the target gene under the control of the inducible Phyper-spink promotor. The integrative vector allows the incorporation of the genes in the chromosome and guarantees no variations in the number of copies present in the cell during the experiments and comparison analysis. A comparable and controlled expression of the different genes is achieved through the use of an inducible promotor.

The genetic system was improved through the creation of new restriction sites in the *msmX* gene, which allowed the cloning of the *msmX* homologs coding sequences in the plasmid flanked by the sequences of the ribosome binding site (RBS) and the terminator of the *msmX* gene. These modifications were introduced by Mendes and Sá-Nogueira (unpublished results) in order to eliminate transcriptional and post-transcriptional regulation variables. A third generation of improvement was constructed by Sá-Nogueira (unpublished results) and consists in the incorporation of a C-terminal His-tag, which allows the detection of the *in vivo* intracellular accumulation of the target ATPase.

All the vectors derive from the integrative plasmid pDR111 (Appendix 6.1). This plasmid possesses a modified and stronger version of the Pspac promoter, the Phyper-spink promoter (Pspink(hy)), that enables a controlled gene expression and additionally bears a lacO binding site that allows 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. The plasmid used in this study contains the ribosomal binding site (RBS) and the terminator of the *msmX* gene, in order to eliminate transcriptional and post-transcriptional regulation variables. This vector allows the cloning of only the open reading frame (ORF) of the ATPase gene to be tested, due to modifications made in the N-terminal and C-terminal of *msmX* gene that created two unique restriction sites *NheI* and *BglII*, respectively (pAM12; Mendes and Sá-Nogueira, unpublished work). Furthermore, it was also introduced a C-terminal His-tag that

allows the detection of the intracellular amount of the proteins in the cell with Western-Blot assays (pSN74; Sá-Nogueira, unpublished results).

A schematic illustration of the fine-tuned *in vivo* system is shown in Figure 3.1.

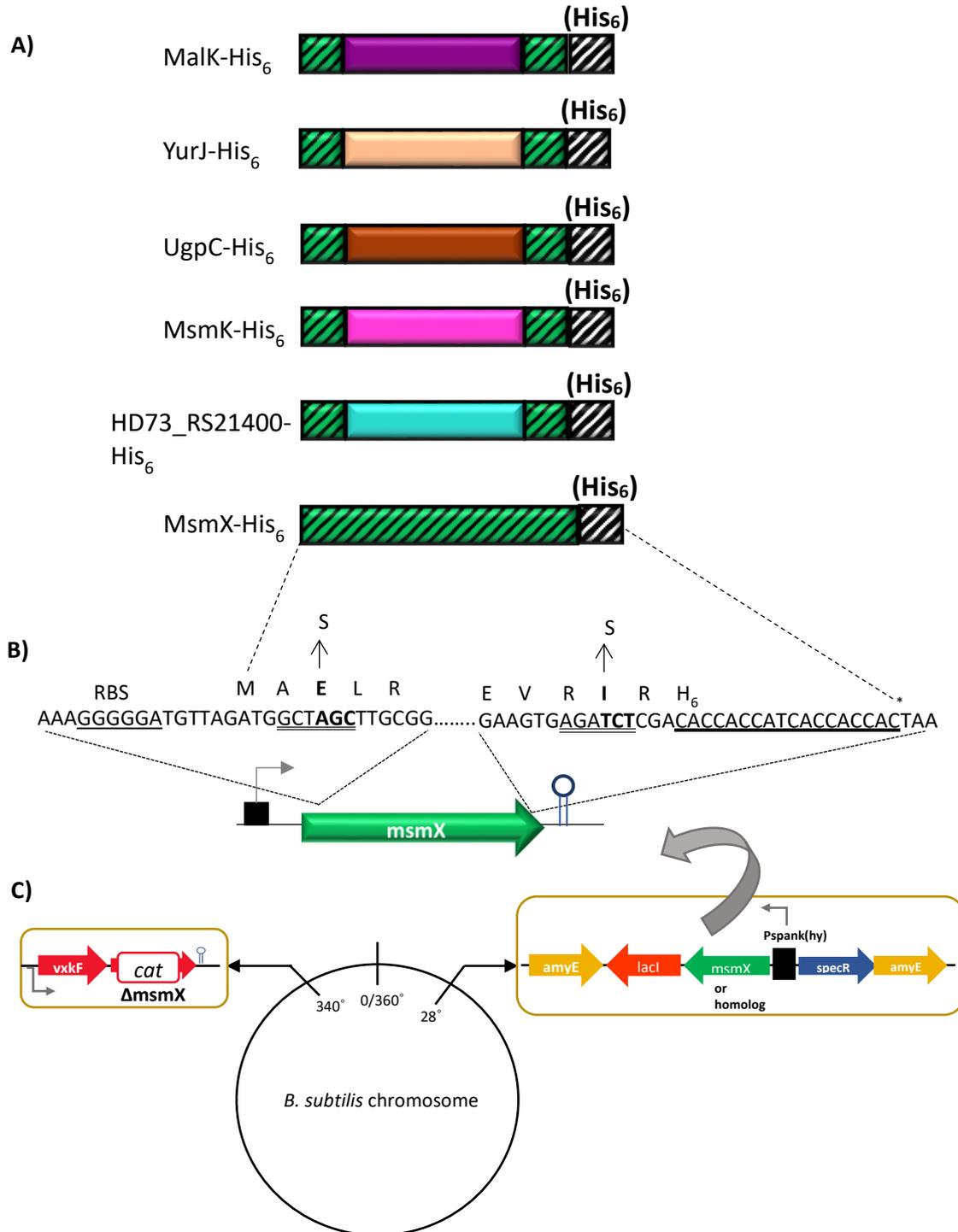
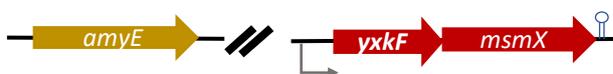


Figure 3.1 - Schematic illustration of the *in vivo* system fine-tuned. A) Representation of the translated protein MsmX (by a green rectangle with stripes) and the translated MsmX homologs (by rectangles with different colours for each homolog). MsmX protein's C- and N-terminal modifications are represented in the homologs by small green squares with stripes; the C-terminal His₆-tag is represented

in all proteins by a small white square with stripes. **B)** Representation of the modifications made in the *msmX* gene of plasmid pSN74. 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 *msmX* gene was modified through the mutation of residues (represented in bold), in order to introduced new restriction sites, *NheI* (5'-end) and *BglIII* (3'-end), both double underlined (pAM12; Mendes and Sá-Nogueira, unpublished work). The system was fine-tuned through the addition of a His-tag (3'-end), which is underlined in bold (pSN74; Sá-Nogueira, unpublished work). **C)** In *B. subtilis* chromosome, the *amyE* locus at 28° is represented (top left) with the genes introduced by double-recombination event with pDR111 derivatives: a *lacI* copy (red arrow), a spectinomycin resistance cassette (blue arrow), both from pDR111, a copy of *msmX* gene, or homolog (green arrow) and the *amyE* fragments (yellow arrows). At 340°, is shown the *yxkF-msmX* operon with the *msmX* inactivated by an insertion-deletion mutation with a chloramphenicol resistance cassette (*cat*).

The strains constructed or used in this work, resulting from the partial integration of pDR111 (ISN1 and IQB672) and its derivatives (ISN8, ISN9 and ISN16), pAM12 (IQB676), or pSN74 and its derivatives (ISN2 and ISN3 to ISN7), are represented in a schematic illustration shown in Figure 3.2.

168T⁺ (wild-type)



ISN1 ($\Delta amyE::Pspank(hy)-spec$)



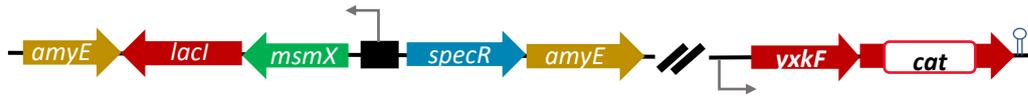
IQB495 ($\Delta msmX::cat$)



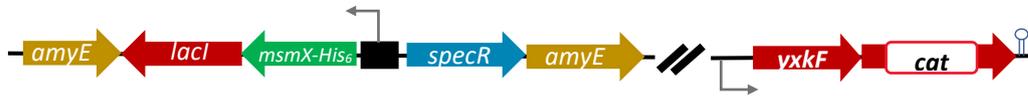
IQB672 ($\Delta msmX::cat \Delta amyE::Pspank(hy)$)



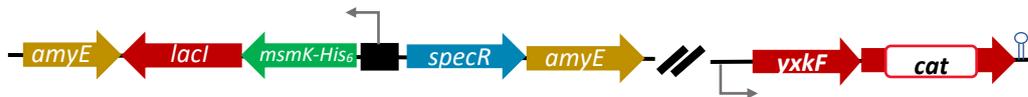
IQB676 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmX(Glu3Ser, Ile364Ser)$)



ISN2 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmX-His_6$)



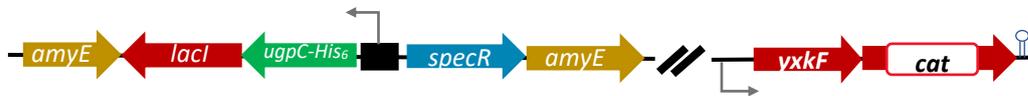
ISN3 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmK-His_6$)



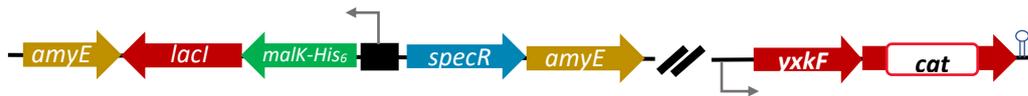
ISN4 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-HD73_RS21400-His_6$)



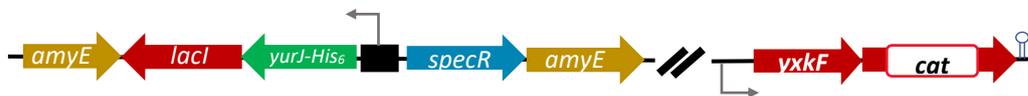
ISN5 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ugpC-His_6$)



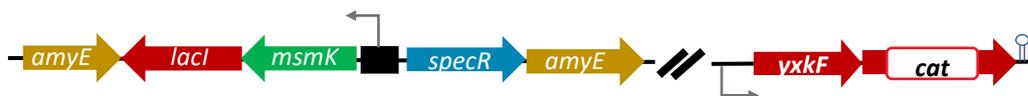
ISN6 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-malk-His_6$)



ISN7 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-yurJ-His_6$)



ISN8 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmK$)



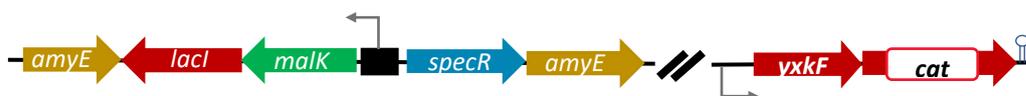
ISN9 ($\Delta msmX::cat \Delta amyE::P_{spank(hy)}-malk$)ISN16 ($\Delta msmX::cat \Delta amyE::P_{spank(hy)}-ycjV$)

Figure 3.2 – Schematic representation of the *amyE* locus and the *yxkF-msmX* operon in the chromosome of several *B. subtilis* strains used in the work (not drawn to scale).

3.1.3 Functional studies of different ATPases in *B. subtilis*

The AraNPQ uptake system cannot transport arabinotriose in the absence of MsmX (Ferreira and Sá-Nogueira, 2010), thus it was chosen to evaluate the capability of the other ATPases to fully (or partially) complement MsmX absence. In order to assess the functionality of the ATPases growth kinetic assays were performed in minimal medium supplemented with arabinose or arabinotriose as the sole carbon and energy source, and in the presence or absence of the inducer IPTG.

3.1.3.1. Wild-type MsmX and MsmX Homologs

B. subtilis strains constructed with the genetic system developed by Mendes and Sá-Nogueira (unpublished work; see above 3.1.2 subsection) were evaluated to test the functionality of the different wild-type ATPases. Growth kinetic parameters of several *B. subtilis* strains in minimal medium using arabinose (MsmX independent uptake) and arabinotriose (MsmX dependent uptake) as the sole carbon and energy source, were determined and the results are summarized in Tables 3.2, 3.3 and 3.4. The *B. subtilis* IQB672 strain (Mendes and Sá-Nogueira, unpublished work; Figure 3.2) possesses a *msmX*-null genetic background, and therefore was used as negative control. *B. subtilis* ISN1 strain (Figure 3.2) was constructed by transformation of the wild-type 168T⁺ strain with plasmid pDR111, and was used as a positive control.

Table 3.2 – Growth of different *B. subtilis* strains ISN1, IQB672 and IQB673 in the presence of distinct saccharides as sole carbon and energy source.

Carbon Source	Doubling Time (minutes)*		
	ISN1 ($\Delta amyE::Pspank(hy)$ - <i>spec</i>)	IQB672 ($\Delta msmX::cat$ $\Delta amyE::Pspank(hy)$)	IQB673 ($\Delta msmX::cat$ $\Delta amyE::Pspank(hy)$ - <i>msmX</i>)
Arabinose 0.1%	62.25±0.98	70.66±5.83	78.47±3.47
Arabinotriose 0.1%	86.84±0.76	No Growth	No Growth
Arabinotriose 0.1% + IPTG 1mM	-	No Growth	107.60±3.85

*Cells were grown in CSK medium (see Materials and Methods). Growth kinetics parameters were determined and the results represent the average of three independent experiments (see Annexes for the complete data used to determine the doubling time). The results of strain IQB673 displayed were obtained from Mendes and Sá-Nogueira (unpublished work), and the arabinose growth for this strain was performed only twice by the authors.

B. subtilis IQB673 strain, which contains *msmX* gene placed at the *amyE* locus, under the control of an inducible promoter is unable to grow in minimal medium with arabinotriose as the sole carbon and energy source in the absence of the inducer IPTG displays like the negative control IQB672 strain (Table 3.2, Figure 3.4 and Figure 3.5). However, when IPTG was added to the medium, the ability to utilize arabinotriose as carbon and energy source was restored and the doubling time is similar to the wild-type ISN1 strain (considering the deviations; Table 3.2, Figure 3.3). These results show the efficacy of the regulation of expression in the used genetic system by the effector molecule IPTG, and confirm that the AraNPQ transport system depends on the presence of the ATPase MsmX (Ferreira and Sá-Nogueira, 2010). The uptake of arabinose is not dependent on this transport system, or the multitask ATPase MsmX (Ferreira and Sá-Nogueira, 2010), which is confirmed by the results obtained in the presence of arabinose and in the absence of the inducer: the doubling time of the strain IQB673 is similar to that of the wild-type ISN1 strain (considering the deviations; Table 3.2).

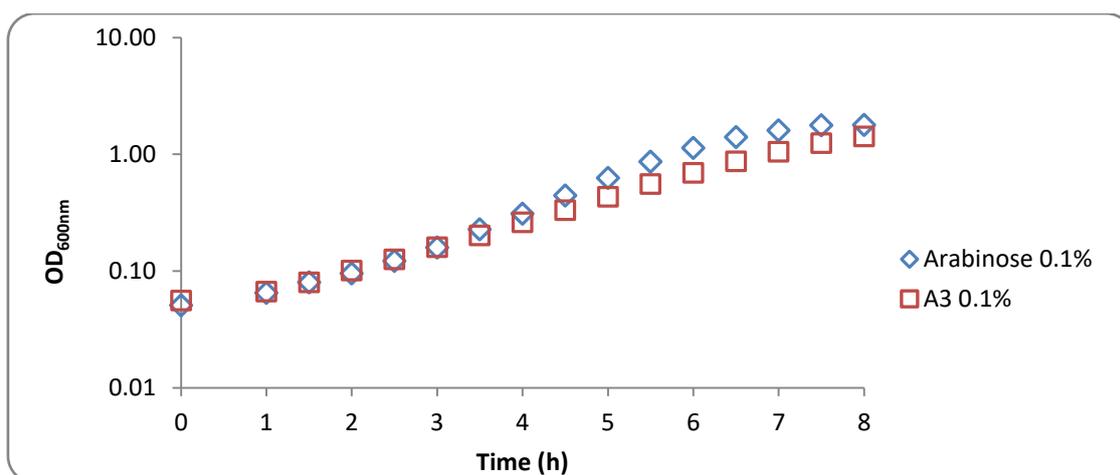


Figure 3.3 – Growth of *B. subtilis* ISN1 ($\Delta amyE::P_{spank}(hy)-spec$) in CSK medium using arabinose and arabinotriose as the sole carbon and energy source.

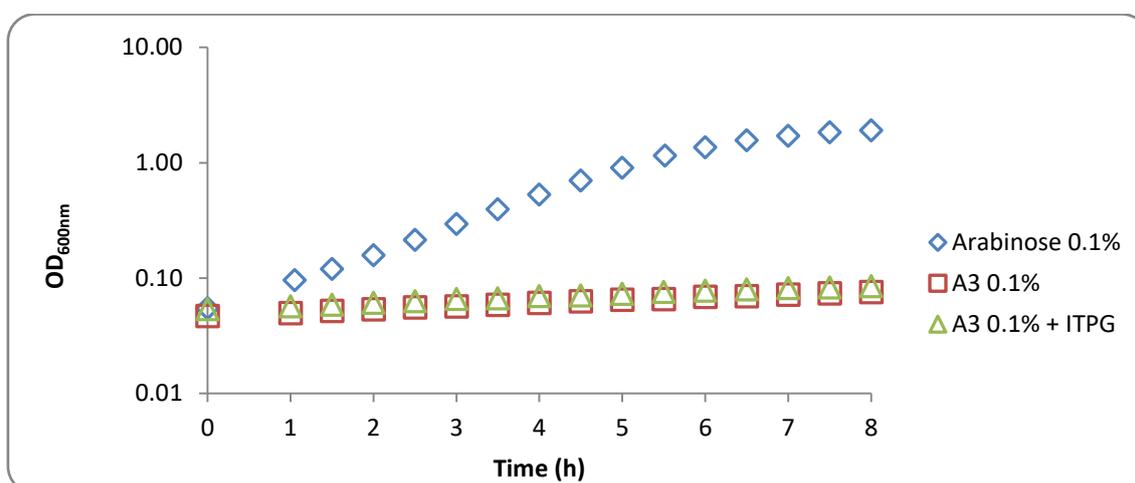


Figure 3.4 – Growth of *B. subtilis* IQB672 ($\Delta msmX::cat \Delta amyE::P_{spank}(hy)$) in CSK medium using arabinose and arabinotriose as the sole carbon and energy source.

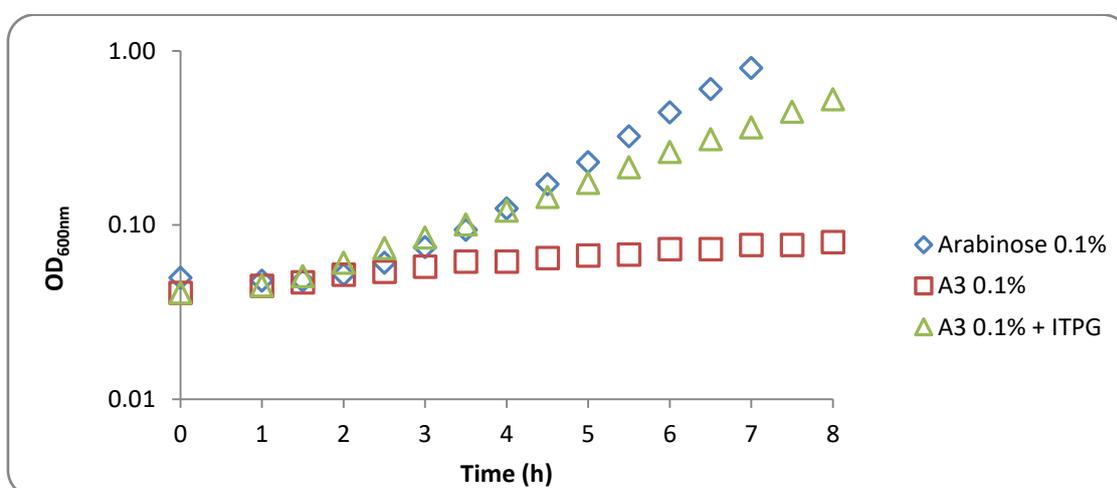


Figure 3.5 – Growth of *B. subtilis* ISN673 ($\Delta msmX::cat \Delta amyE::P_{spank}(hy)-msmX$) in CSK medium using arabinose and arabinotriose as the sole carbon and energy source.

The growth rate of the *B. subtilis* strains bearing the different MsmX homologs grown in the presence of arabinose is similar to the wild-type ISN1 strain (Table 3.3 and Table 3.4). However, in the presence of arabinotriose and IPTG are quite distinct.

Table 3.3 – Growth of different *B. subtilis* strains IQB642, IQB677 and IQB678 in the presence of distinct saccharides as sole carbon and energy source.

Carbon Source	Doubling Time (minutes)*		
	IQB642 ($\Delta msmX::cat$ <i>amyE::</i> [Phyper- spank- <i>yurJ spec</i>] [<i>B. subtilis</i>])	IQB677 ($\Delta msmX::cat$ $\Delta amyE::Pspank(hy)$ - <i>HD73_RS21400</i> [<i>B. thuringiensis</i>])	IQB678 ($\Delta msmX::cat$ $\Delta amyE::Pspank(hy)$ - <i>ugpC</i> [<i>S. aureus</i>])
Arabinose 0.1%	-	79.01±0.59	79.98±1.65
Arabinotriose 0.1%	No Growth	No Growth	No Growth
Arabinotriose 0.1% + IPTG 1mM	94.60±0.50	141.18±3.76	205.47±8.28

* Results adapted from Ferreira and Sá-Nogueira (strain IQB642, unpublished data) and Mendes and Sá-Nogueira (strains IQB677 and IQB678, unpublished data). Cells were grown in CSK medium (see Materials and Methods). Growth kinetics parameters were determined and the results represent the average of three independent experiments, except the arabinose growth for IQB677 and IQB678 without IPTG, which were performed only twice by the authors.

In previous studies, here presented for clarity, Ferreira and Sá-Nogueira (unpublished data) cloned *yurJ* gene from *B. subtilis* in pDR111 and transformed the plasmid in *B. subtilis* IQB495, yielding strain IQB642. This strain displays a doubling time very similar to that of *B. subtilis* IQB673, which possesses *msmX* gene, suggesting that YurJ is able to fulfil the role of MsmX in the AraNPQ transport system (Table 3.2 and 3.3). In previous experiments performed by Mendes and Sá-Nogueira (unpublished data) *HD73_RS21400* gene from *B. thuringiensis* was cloned in pDR111, and posteriorly transformed in *B. subtilis* IQB495 resulting in strain IQB677. The growth rate of *B. subtilis* IQB677 is 1.4-fold higher than *B. subtilis* IQB673 (Table 3.2 and 3.3), indicating that the ATPase from *B. thuringiensis* (HD73_RS21400) is able to partially complement the MsmX deficiency. Although the ATPase HD73_RS21400 shares 74% of amino acid (a. a.) identity to MsmX, it shows a lower degree of complementation than that of YurJ which only shares 58% of a. a. identity to MsmX.

Similarly, the *ugpC* gene from *S. aureus* was cloned in pDR111 and transformed in *B. subtilis* IQB495 resulting in strain IQB678, constructed by Mendes and Sá-Nogueira (unpublished

data), is only able to partially complement the *msmX*-null mutation, since the strain IQB678 grows with a doubling time approximately 2-fold higher than *B. subtilis* IQB673 (Table 3.2 and 3.3).

Table 3.4 – Growth of different *B. subtilis* strains ISN8, ISN9 and ISN16 in the presence of distinct saccharides as sole carbon and energy source.

Carbon Source	Doubling Time (minutes)*		
	ISN8 ($\Delta ms m X::cat$ $\Delta amy E::Pspank(hy)-$ $ms m K$ [<i>S.</i> <i>pneumoniae</i>])	ISN9 ($\Delta ms m X::cat$ $\Delta amy E::Pspank(hy)-$ $mal K$ [<i>E. coli</i>])	ISN16 ($\Delta ms m X::cat$ $\Delta amy E::Pspank(hy)-$ $yc j V$ [<i>E. coli</i>])
Arabinose 0.1%	78.60±4.15	72.10±1.79	69.48±4.19
Arabinotriose 0.1%	No Growth	No Growth	No Growth
Arabinotriose 0.1% + IPTG 1mM	168.46±8.50	No Growth	No Growth

*Cells were grown in CSK medium (see Materials and Methods). Growth kinetics parameters were determined and the results represent the average of three independent experiments (see Annexes for the complete data used to determine the doubling time).

Likewise, MsmK from *S. pneumoniae* also exhibits a partial degree of complementation, as observed in the *B. subtilis* ISN8 strain (Figure 3.2) displaying a doubling time approximately 1.7-fold higher than the IQB673 strain (Table 3.4, Figure 3.6). UgpC and MsmK share a similar percentage of amino acids identity to MsmX, 66% and 64%, respectively, which is higher than that observed for YurJ, however in this genetic system complementation is less efficient.

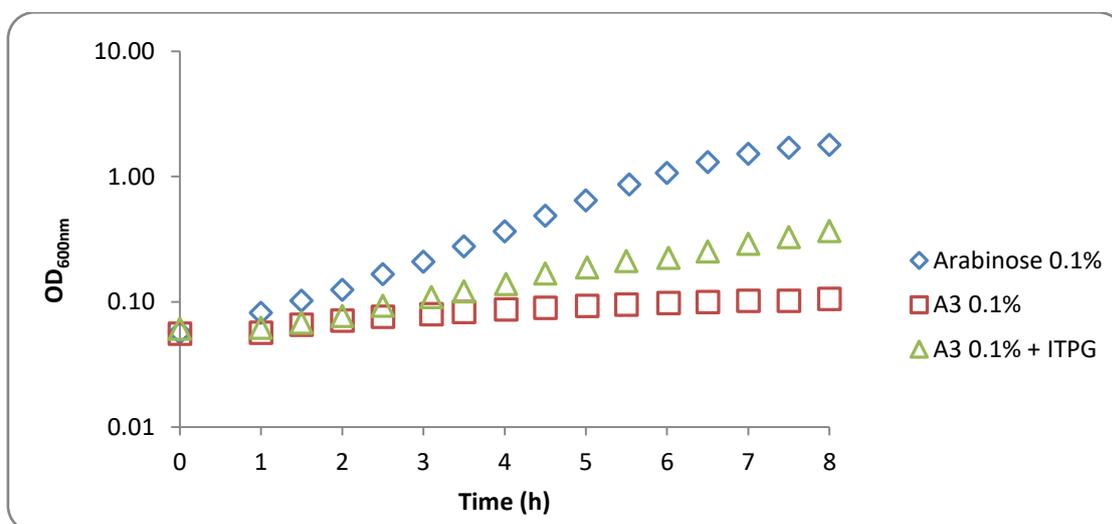


Figure 3.6 - Growth of *B. subtilis* ISN8 ($\Delta ms m X::cat$ $\Delta amy E::Pspank(hy)-ms m K$) in CSK medium using arabinose and arabinotriose as the sole carbon and energy source.

The different complementation degrees observed suggest that there is not a direct correlation between the percentage of amino acids identity that the homologs share with MsmX and the efficiency of functional complementation. However differences in one or two amino acids near the conserved region of the Q-loop (see Appendix 6.18) may affect the interaction with the TMDs through the “coupling helix” (ter Beek *et al*, 2014) and thus slowing down the uptake of arabinotriose and consequently the ability to growth. Another possible explanation for the observed results may be the differences in post-transcriptional regulation such as mRNA stability and/or RBS strength since the genes were cloned with their own RBS. Mendes and Sá-Nogueira (unpublished data) verified differences in the RBS strength of the genes through the calculation of the translation initiation rates (RBS Calculator_{v2.0} tool; Salis *et al*, 2009; Salis, 2011), however no correlation was found with the complementation results.

B. subtilis strains ISN9 and ISN16 (Figure 3.2) possess the ATPases MalK and YcjV respectively, two proteins from the Gram-negative bacteria *E. coli*, and are unable to utilize arabinotriose as the sole carbon and energy source (Table 3.4, Figures 3.7 and 3.8). MalK displays the lowest a. a. identity to MsmX however, ATPase YcjV that does not substitute MsmX function in the transport system shares 64% of amino acid identity with MsmX, which is the same percentage shared between MsmK and MsmX. A closer look into the primary sequences alignment (see Appendix 6.18) highlights the absence of a small region of five amino acids in these two ATPases from *E. coli*, which is present in MsmX and all the other tested ATPases from Gram-positive bacteria.

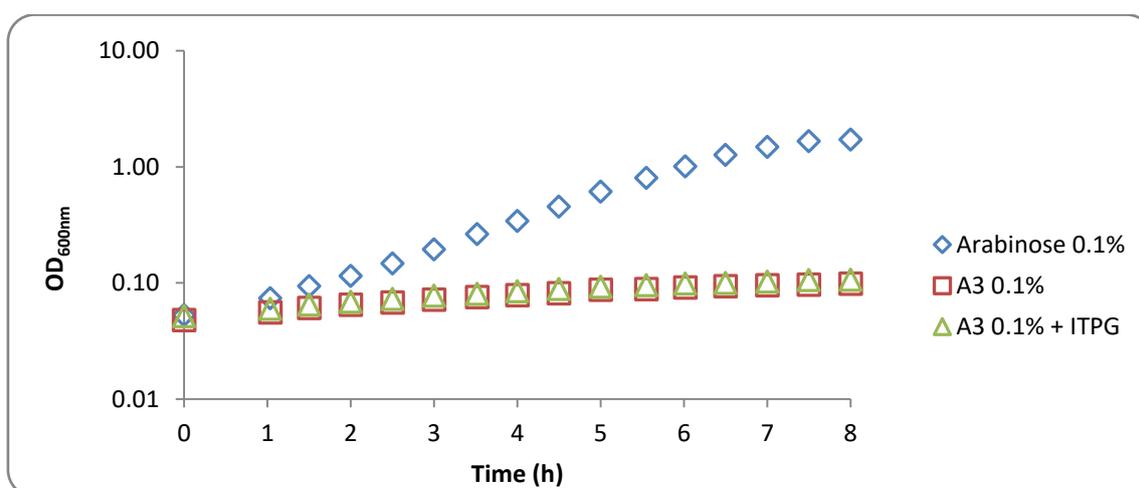


Figure 3.7 - Growth of *B. subtilis* ISN9 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-malk$) in CSK medium using arabinose and arabinotriose as the sole carbon and energy source.

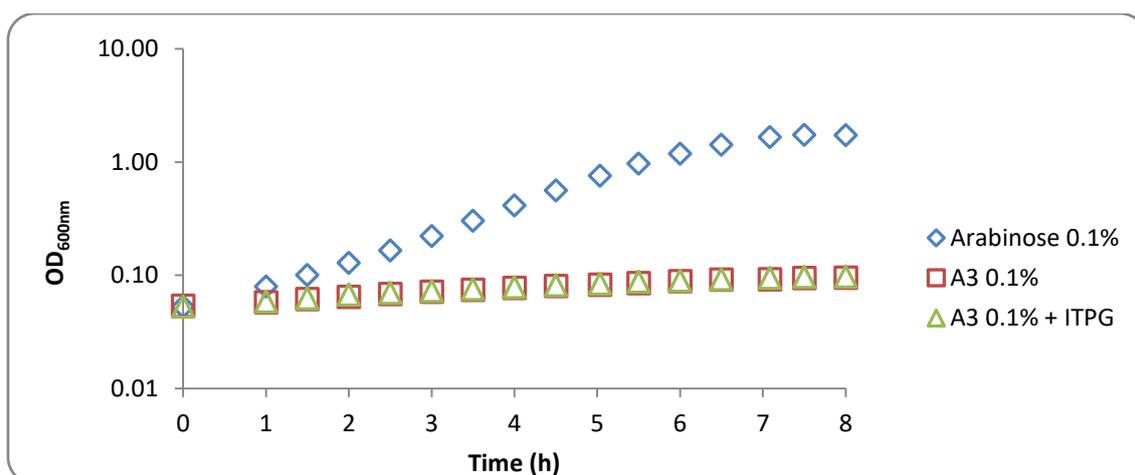


Figure 3.8 - Growth of *B. subtilis* ISN16 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ycjV$) in CSK medium using arabinose and arabinotriose as the sole carbon and energy source

3.1.3.2. Recombinant MsmX and MsmX Homologs with a C-terminal His-tag

To assess the intracellular level of all the ATPases expressed in *B. subtilis* modified versions of the proteins were constructed, strains ISN2-ISN7, and their functionality was examined by determination of the growth kinetic parameters of the different strains as described above. The eight strains were grown in minimal medium using arabinose (MsmX independent uptake) and arabinotriose (MsmX dependent uptake) as the sole carbon and energy source, and the results are summarized in Tables 3.5 and 3.6. In the presence of arabinose the *B. subtilis* strains ISN2-ISN7 present the same growth rate as the strains with the wild-type proteins. However, in the minimal medium with arabinotriose and the inducer IPTG, strains ISN2-5 and ISN7 with the recombinant proteins show different kinetic parameters relatively to the wild-type proteins strains IQB673, ISN8, IQB677, IQB678 and IQB642, respectively (Tables 3.2, 3.3 and 3.4).

Table 3.5 – Growth of different *B. subtilis* strains IQB673, IQB676 and ISN2 in the presence of distinct saccharides as sole carbon and energy source.

Carbon Source	Doubling Time (minutes)*		
	IQB673 ($\Delta msmX::cat$ $\Delta amyE::Pspank(hy)-$ $msmX$)	IQB676 ($\Delta msmX::cat$ $\Delta amyE::Pspank(hy)-$ $msmX(Glu3Ser,$ $Ile364Ser)$)	ISN2 ($\Delta msmX::cat$ $\Delta amyE::Pspank(hy)-$ $msmX-His_6$)
Arabinose 0.1%	78.47±3.47	70.48±5.70	69.73±7.68
Arabinotriose 0.1%	No Growth	No Growth	No Growth
Arabinotriose 0.1% + IPTG 1mM	107.60±3.85	110.57±2.26	204.95±7.10

*Cells were grown in CSK medium (see Materials and Methods). Growth kinetics parameters were determined and the results represent the average of three independent experiments, except the arabinose growth for strain IQB676, which was performed only twice (see Annexes for the complete data used to determine the doubling time). The results of the strain IQB673 (Mendes and Sá-Nogueira, unpublished data) are the same presented in table 3.2 and are shown here to facilitate comparison.

During the process of fine-tuning the genetic system, Mendes and Sá-Nogueira (unpublished results) constructed *B. subtilis* IQB676 strain (Figure 3.2) which possesses the *msmX* gene with modified C- and N-terminal sequences to generate the restriction sites of *NheI* and *BglII* to facilitate sub-cloning of the coding region of other ATPases (see 3.1.2 subsection). In the presence of arabinotriose and the inducer IPTG this strain displays a doubling time similar to the *B. subtilis* IQB673 strain (Table 3.5; Figure 3.9), demonstrating that the modifications introduced did not affect the function of the protein MsmX. However, in *B. subtilis* ISN2 strain (Figure 3.2), which possesses, in addition to the previous modifications, a C-terminal H₆-tag, in the presence of arabinotriose with the inducer IPTG displays a doubling time 2-fold higher than that of *B. subtilis* IQB676 strain (Table 3.5, Figure 3.10). This observation suggests that the addition of the His-tail is affecting the ATPase function.

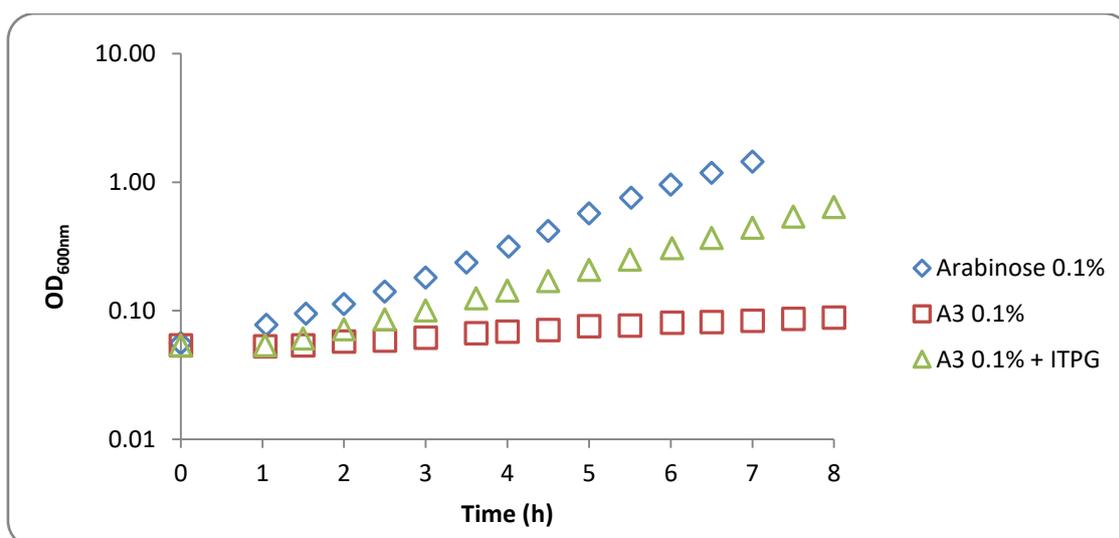


Figure 3.9 - Growth of *B. subtilis* IQB676 ($\Delta msmX::cat \Delta amyE::P_{spank(hy)}-msmX(Glu3Ser, Ile364Ser)$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.

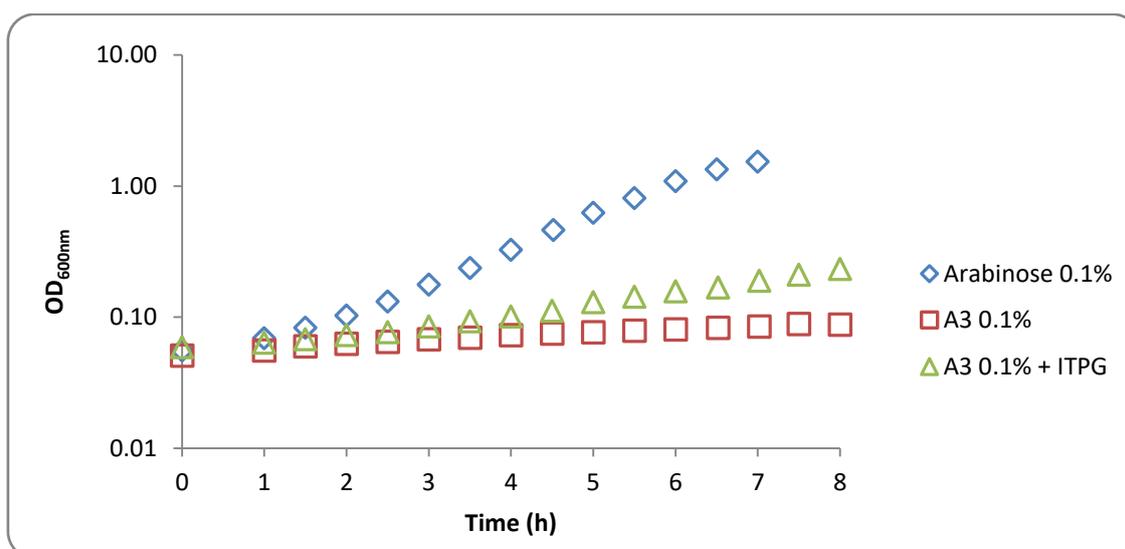


Figure 3.10 - Growth of *B. subtilis* ISN2 ($\Delta msmX::cat \Delta amyE::P_{spank(hy)}-msmX-His_6$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.

Growth kinetic parameters of the five strains bearing the other ATPases under study (ISN3 to ISN7) in minimal medium using arabinose (MsmX independent uptake) and arabionotriose (MsmX dependent uptake) as the sole carbon and energy source, were also determined and the results are summarized in Table 3.6.

Table 3.6 – Growth of different *B. subtilis* strains ISN3, ISN4, ISN5, ISN6 and ISN7 in the presence of distinct saccharides as sole carbon and energy source.

Carbon Source	Doubling Time (minutes)*				
	ISN3 (<i>ΔmsmX::cat</i> <i>ΔamyE::Pspank(hy)-msmK-His₆</i> [<i>S. pneumoniae</i>])	ISN4 (<i>ΔmsmX::cat</i> <i>ΔamyE::Pspank(hy)-HD73_RS21400-His₆</i> [<i>B. thuringiensis</i>])	ISN5 (<i>ΔmsmX::cat</i> <i>ΔamyE::Pspank(hy)-ugpC-His₆</i> [<i>S. aureus</i>])	ISN6 (<i>ΔmsmX::cat</i> <i>ΔamyE::Pspank(hy)-malK-His₆</i> [<i>E. coli</i>])	ISN7 (<i>ΔmsmX::cat</i> <i>ΔamyE::Pspank(hy)-yurJ-His₆</i> [<i>B. subtilis</i>])
Arabinose 0.1%	68.68±2.31	69.98±1.79	74.19±3.53	71.24±3.79	69.79±2.42
Arabinotriose 0.1%	No Growth	No Growth	No Growth	No Growth	No Growth
Arabinotriose 0.1% + IPTG 1mM	150.26±7.33	111.95±2.71	No Growth	No Growth	No Growth

*Cells were grown in CSK medium (see Materials and Methods). Growth kinetics parameters were determined and the results represent the average of three independent experiments, except the arabinotriose growth for strains ISN5 and ISN6 without IPTG, which were performed only twice (see Annexes for the complete data used to determine the doubling time).

The recombinant ATPase HD73_RS21400 (from *B. thuringiensis*) cloned in *B. subtilis* ISN4 (Figure 3.2) fully complements MsmX function in the uptake of arabinotriose, since the doubling time of ISN4 (Table 3.6, Figure 3.12) decreased approximately 1.25-fold relatively to value displayed by the strain IQB677 with the wild-type protein (Table 3.3), being very similar to the value determined from the IQB673 strain growth (Table 3.5). It was also observed an improvement in the growth of *B. subtilis* ISN3 strain (Figure 3.2), with the recombinant MsmK from *S. pneumoniae* (Table 3.6, Figure 3.11), when compared to strain ISN8 with the wild-type protein (Table 3.4). However, in the strains ISN5 and ISN7 (Figure 3.2) with the ATPases UgpC-His₆ and YurJ-His₆, respectively, it was observed a significantly increase in the doubling time values (Table 3.6; Figure 3.13 (strain ISN5) and Figure 3.15 (strain ISN7)). The strain ISN6 (Figure 3.2) with the ATPase MalK-His₆ maintained the same growth rate (Table 3.6, Figure 3.14) compared to that of strain ISN9 with the wild-type protein (Table 3.4).

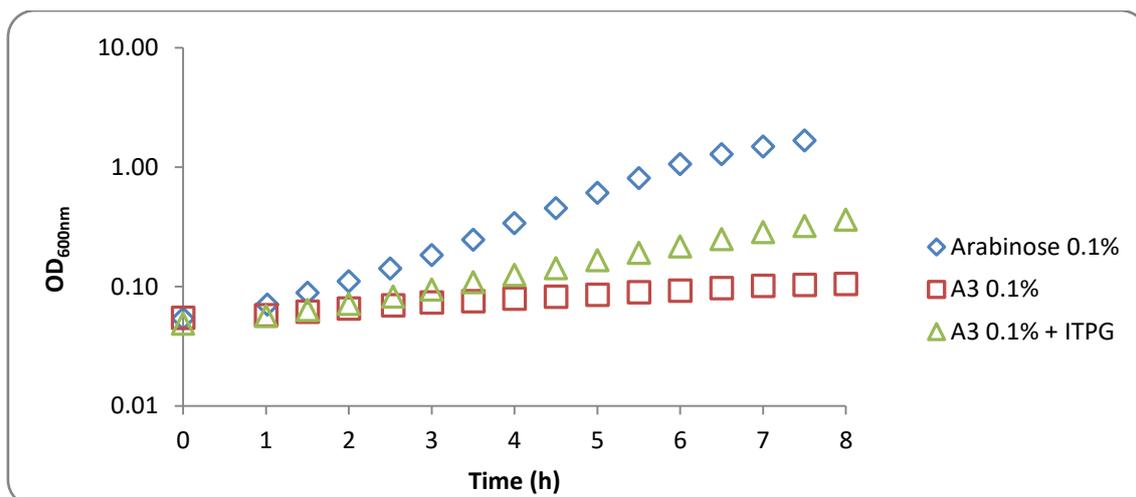


Figure 3.11 - Growth of *B. subtilis* ISN3 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmK-His_6$) in CSK medium using arabinose and arabinotriose as the sole carbon and energy source.

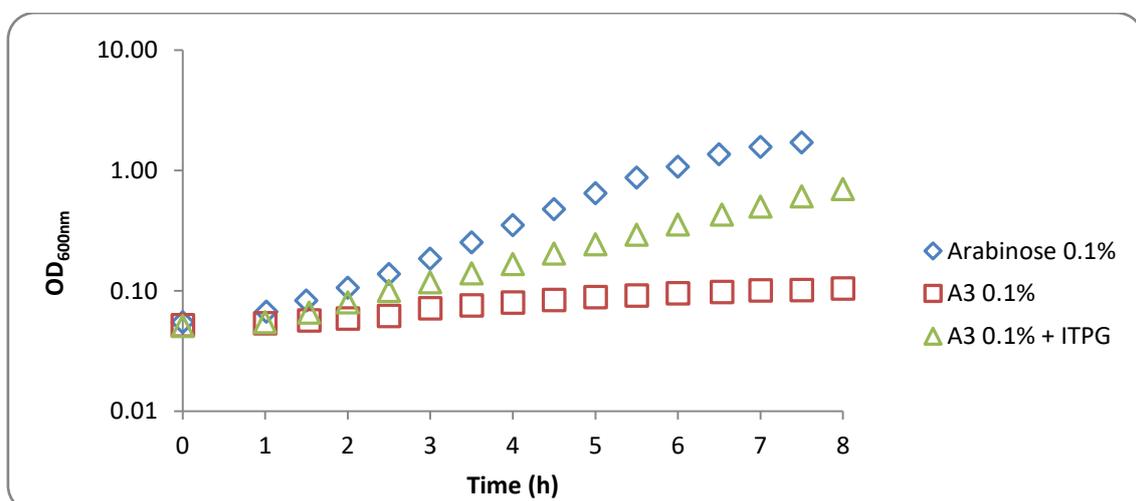


Figure 3.12 - Growth of *B. subtilis* ISN4 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-HD73_RS21400-His_6$) in CSK medium using arabinose and arabinotriose as the sole carbon and energy source.

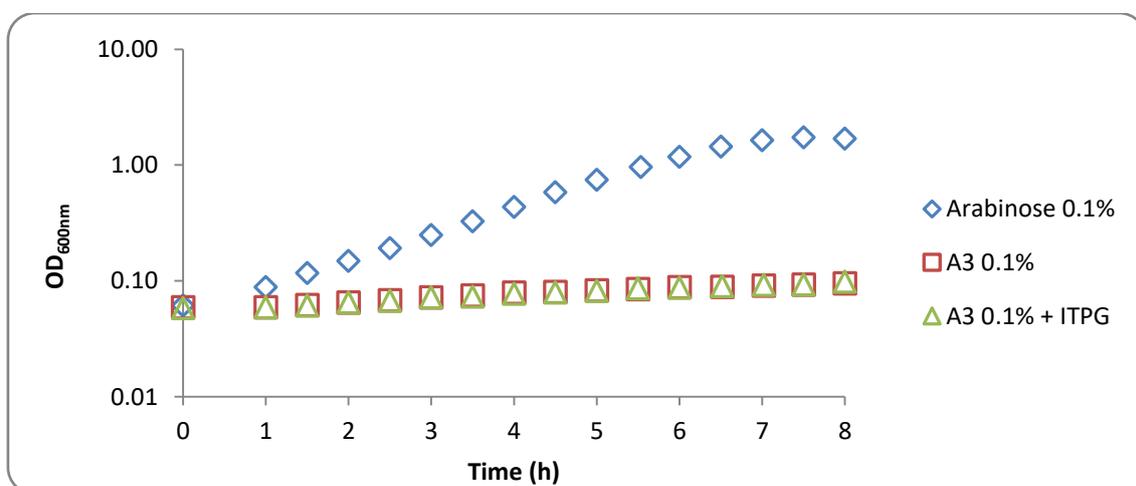


Figure 3.13 - Growth of *B. subtilis* ISN5 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ugpC-His_6$) in CSK medium using arabinose and arabinotriose as the sole carbon and energy source.

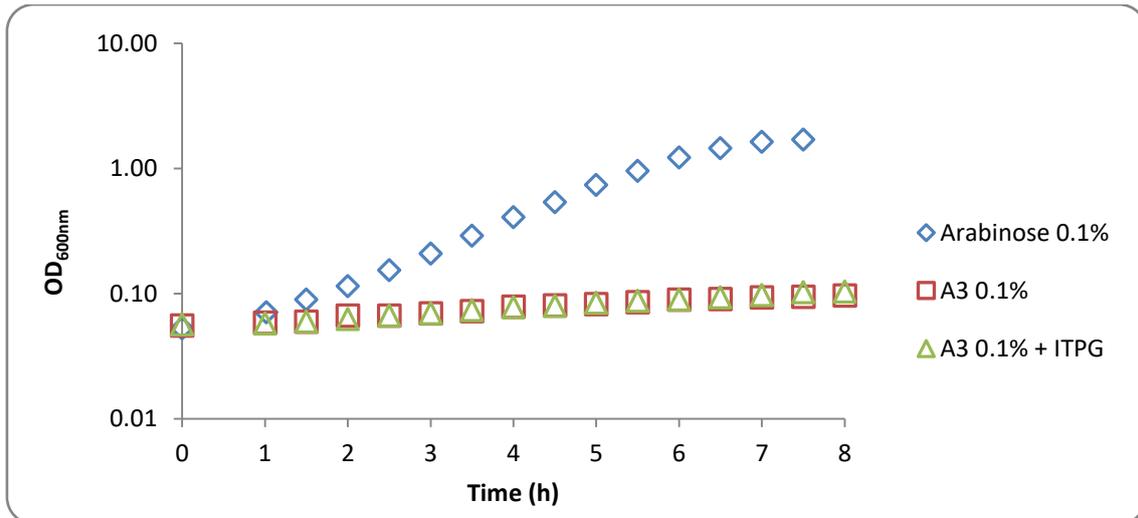


Figure 3.14 - Growth of *B. subtilis* ISN6 ($\Delta msmX::cat \Delta amyE::P_{spank(hy)}-malk-His_6$) in CSK medium using arabinose and arabinotriose as the sole carbon and energy source.

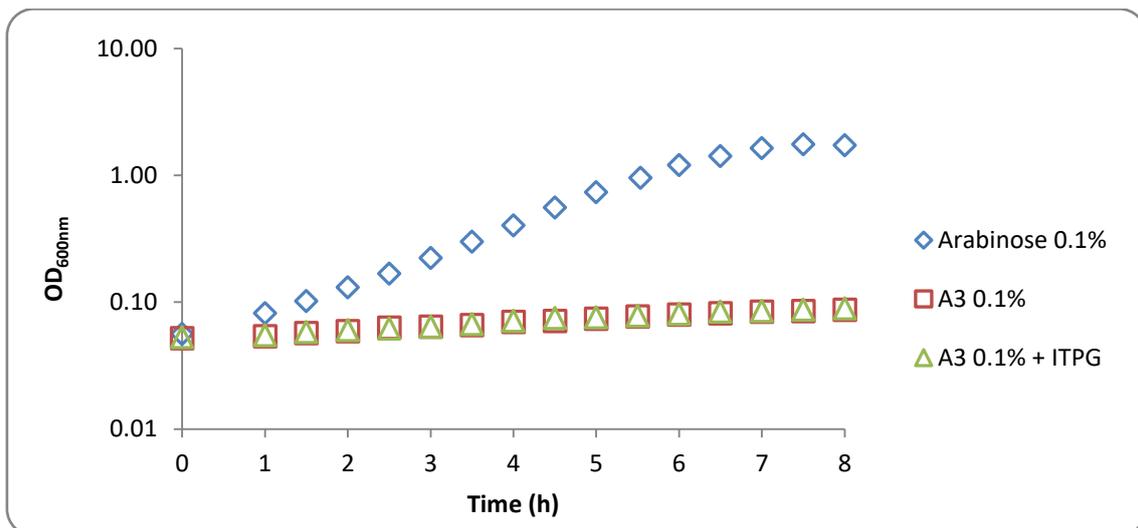


Figure 3.15 - Growth of *B. subtilis* ISN7 ($\Delta msmX::cat \Delta amyE::P_{spank(hy)}-yurJ-His_6$) in CSK medium using arabinose and arabinotriose as the sole carbon and energy source.

The different complementation degrees observed in the recombinant proteins versus the wild-type proteins may be explained by a distinct translation initiation rate and/or conformational alterations due to the presence of the His-tag. The translation initiation rate (TIR) changed in the fine-tuned genetic system since the ribosome binding site (RBS) sequence is from *msmX* gene instead of the homologs genes as was the case in the first generation genetic system. The *B. subtilis* ISN4 (ATPase from *B. thuringiensis*) enhanced growth rate could be due to an improvement of the translation initiation rate with the RBS of the *msmX* gene, by increasing the production of the protein resulting in a higher level of copies of the ATPase

present in the cells. A similar improvement of MsmK-His₆ protein production due to the TIR of the *msmX* gene could also explain the decrease of the doubling time of the *B. subtilis* ISN3 strain. *B. subtilis* ISN5 represents a different situation due to a significantly poor growth relatively to that of strain IQB678 (UgpC, wild-type protein), which could mean that the RBS of the *ugpC* gene was stronger than the RBS of the *msmX* gene, representing a slower TIR and consequently a lower level of ATPase. However, in this case, it is also possible that the C-terminal His-tag may be interfering with UgpC function in the transport system. Less likely but also possible is the inference of the a. a. modified in the C- and N-terminal, although they are not involved in the main domains of the ATPase. The same hypotheses maybe be applied to *B. subtilis* ISN7 strain (with the ATPase YurJ-His₆) that also exhibits non-growth phenotype in arabinotriose.

3.1.4. Analysis of the predicted proteins structures

To understand the differences between the wild-type and the recombinant ATPases we performed a protein structure prevision. The online program I-TASSER was used to obtain automated three-dimensional protein structure from the input of the amino acid sequence, through multiple threading alignments and iterative structural assembly simulations.

B. subtilis MsmX, like other ABC ATPases most likely will hydrolyse ATP as a homodimer, and analysing the prevision of the 3-D structure of the recombinant MsmX-His₆ (Figure 3.16, B) it is possible to infer that the conformation of the 6xHis residues added to the C-terminal may be affecting the interface of the two monomers through steric effects, preventing the dimer to hydrolyse the ATP molecules and consequently not transmitting the conformational changes to the transmembrane domains. This conclusion is supported by the comparison with the available structures of the MalK dimer captured in different conformations (Figure 3.17). This analysis is further supported by comparison with the prevision of the 3-D structure of another recombinant ATPase MsmX (Figure 3.16, A) constructed by Ferreira and Sá-Nogueira (unpublished work), from the *B. subtilis* IQB622 strain, where two additional amino acids, Leucine and a Glutamic acid (LE) were inserted upstream of the 6xHis tail. In this structure it is visible a different conformation of the LE6xHis terminal end, with a torsion due to the addition of the two amino acids and possibly by the polar nature of the glutamic acid. In this situation the ATPase works normally because the *B. subtilis* IQB622 strain is able to grow in arabinotriose as the sole carbon and energy source as the wild-type ISN1.

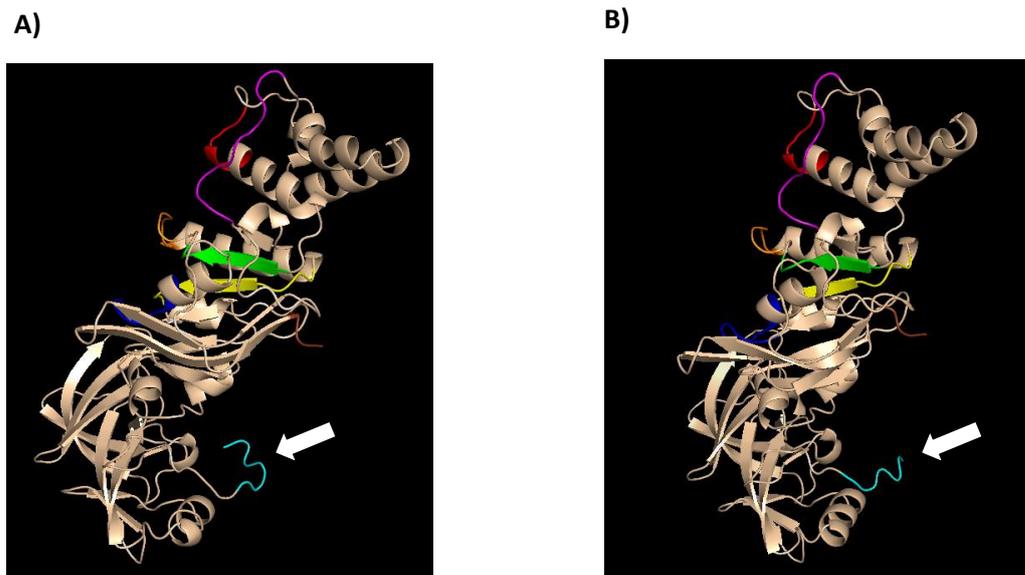


Figure 3.16 – Representation of MsmX-His₆ and MsmX-LEHis₆ 3-D structure model. The structures were drawn in Pymol with the data obtained from the online program I-TASSER. **A)** Model of MsmX-LEHis₆ protein from the strain IQB622 (Ferreira and Sá-Nogueira, unpublished data); **B)** Model of MsmX-His₆ protein from the strain ISN2; The N-terminal, the seven conserved domains and the C-terminal tag are highlighted with the following colours scheme: N-terminal (Brown), Walker A (blue), Q-loop (magenta), Signature (red), Walker B (green), D-loop (orange), H-loop (yellow) and the C-terminal tag (cyan), which is highlighted by a white arrow.

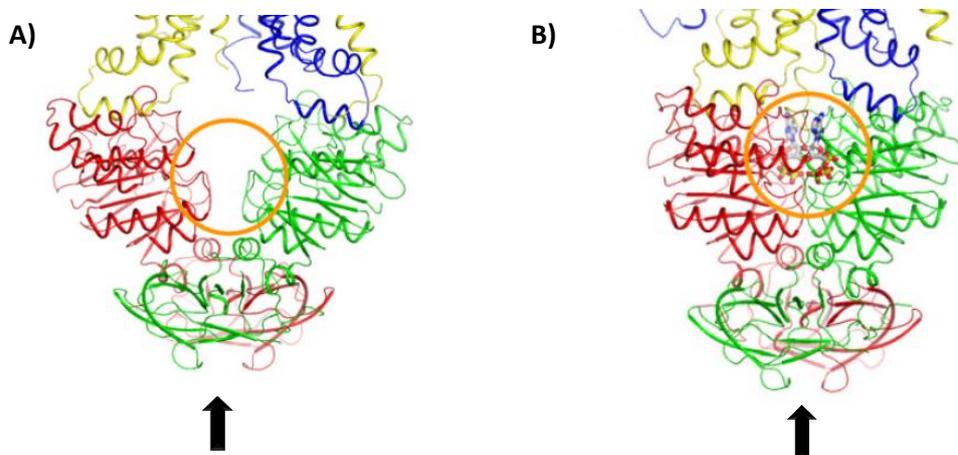


Figure 3.17 - Ribbon representation of the MalK dimer. **A)** inward-facing conformation in the absence of ATP. **B)** outward-facing conformation with ATP in the active site of the NBDs. The NBD dimer interface is indicated by an orange circle and the NBDs C-terminal region is indicated by a black arrow, which highlights the increased dimer proximity between conformations. NBDs MalK are coloured in green and red; the TMDs MalG and MalF are coloured in yellow and blue, respectively (only displayed a small region, including the coupling helix). The nucleotides are displayed in ball-and-stick models. Adapted from Chen, 2013.

B. subtilis ISN5 (UgpC-His₆) and ISN7 (YurJ-His₆) strains exhibited an extreme change in the kinetic growth, relatively to the behaviour of the strains IQB678 and IQB642 respectively, carrying the wild-type proteins. Analysing the predicted structure of the two ATPases UgpC-His₆ and YurJ-His₆ (Figure 3.18, A) and B) respectively) in similar manner to that observed with MsmX-His₆ (ISN2 strain; Figure 3.16, B), and we infer that the tail conformation may be causing steric effects in the junction of the homodimers, thus completely abolishing the role of the two ATPases in providing energy to the transport system. With the ATPases MsmK-His₆ and HD73_RS21400-His₆ we did not observe significant changes in their function upon the His-tag addition, since the kinetic growth of the respective strains ISN3 and ISN4, respectively, is very similar to the strains ISN8 and IQB677 with the wild-type proteins, respectively, which suggests that in these cases the conformation of the tail does not affect the dimerization process. The ATPase MalK remained non-functional upon the His-tag addition, since ISN6 strain with the recombinant protein exhibit non-growth phenotype like ISN9 strain with the wild-type protein. Therefore, MsmK-His₆, HD73_RS21400-His₆ and MalK-His₆ 3-D structure models are not relevant here for the discussion and are shown in the Annexes 6.64, 6.65, and 6.66.

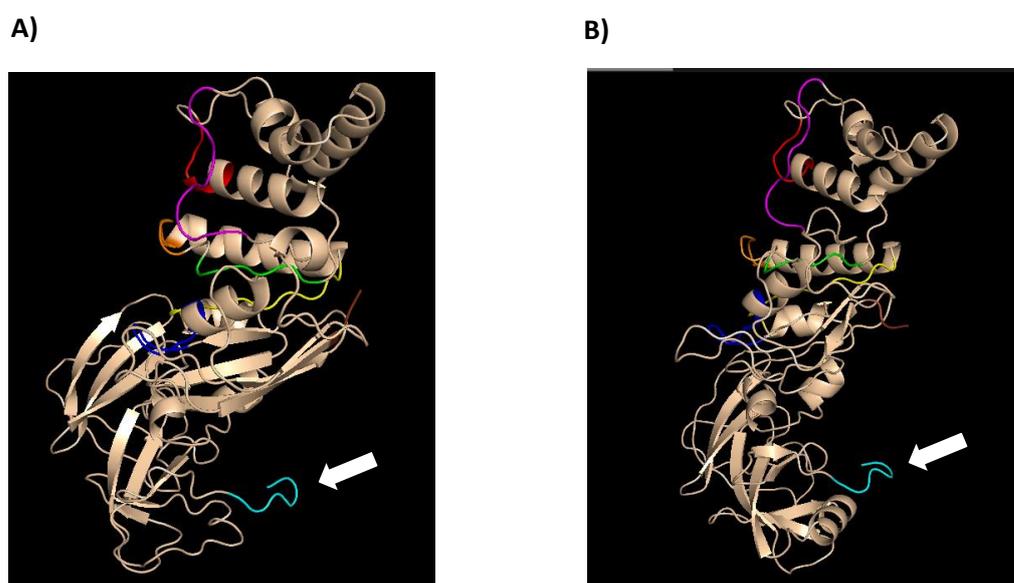


Figure 3.18 – Representation of UgpC-His₆ and YurJ-His₆ 3-D structure model. The structures were drawn in Pymol with the data obtained from the online program I-TASSER. **A)** Model of UgpC-His₆ protein from the strain ISN5; **B)** Model of YurJ-His₆ protein from the strain ISN7.; The N-terminal, the seven conserved domains and the C-terminal tag were highlighted with the following colours scheme: N-terminal (brown), Walker A (blue), Q-loop (magenta), Signature (red), Walker B (green), D-loop (orange), H-loop (yellow) and the C-terminal tag (cyan), which is highlighted by a white arrow.

3.1.5. Redesigning the genetic system for functional analysis

The predicted models of the modified ATPases (His-tag) together with the observation made by Ferreira and Sá-Nogueira (*B. subtilis* IQB622) that the addition of two extra amino acids upstream of the 6xHis does not affect the function of recombinant MsmX, we decided to redesign the genetic system by adding the same amino acids, Leucine and Glutamic acid, before the C-terminal 6xHis-tag. A schematic illustration of the modifications made in the genetic system and the new strains constructed are represented in Figures 3.19 and 3.20, respectively.

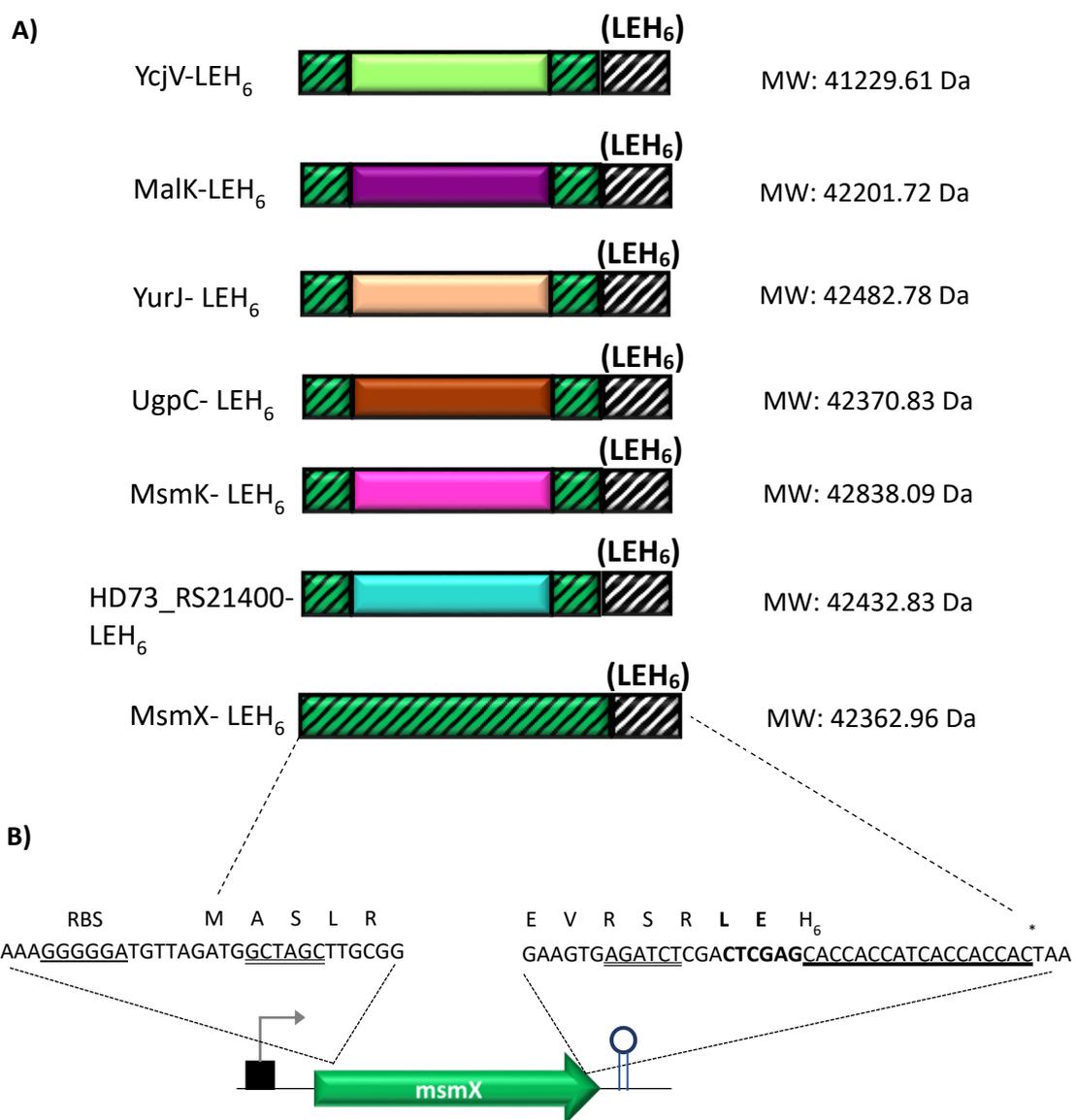
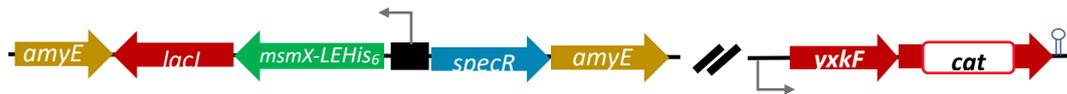


Figure 3.19 – Schematic representation of the modifications made in the genetic system. A) Representation of the translated protein MsmX (green rectangle with stripes) and the translated MsmX homologs (rectangles with different colours for each homolog). MsmX protein's C- and N-terminal are represented in the homologs by small green squares with stripes; the C-terminal LEH₆-tag is represented

in all proteins with a small white rectangle with stripes. Protein sequence and the respective DNA sequence of all ATPases here studied are shown in Appendices 6.68 to 6.74. **B)** 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 and the restriction sites, *NheI* (5'-end) and *BglIII* (3'-end) are both double underlined. The system was redesigned through the insertion of the two amino acids Leu (L) and Glu (E), which are represented in bold, before the C-terminal H₆-tag (3'-end) that is underlined in bold.

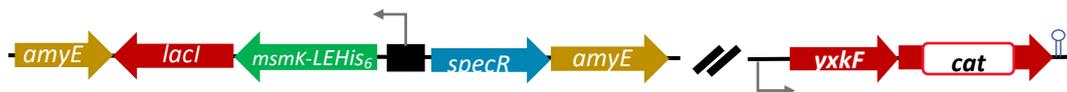
ISN10 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmX-LEHis_6$)



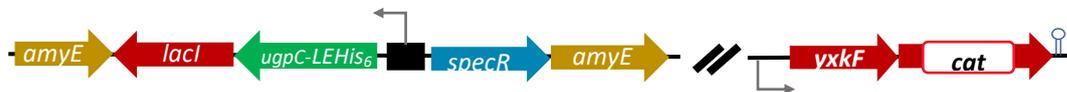
ISN11 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-HD73_RS21400-LEHis_6$)



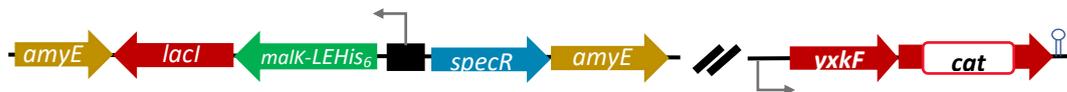
ISN12 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmK-LEHis_6$)



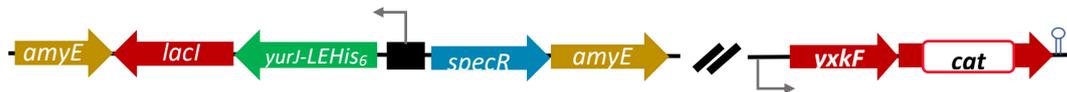
ISN13 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ugpC-LEHis_6$)



ISN14 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-malK-LEHis_6$)



ISN15 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-yurJ-LEHis_6$)



ISN17 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ycjV-LEHis_6$)



Figure 3.20 – Schematic representation of the *amyE* locus and the *yxkF-msmX* operon in the chromosome of several *B. subtilis* strains, constructed with the redesigned genetic system, used in the work (not drawn to scale).

3.1.5.1 Functional studies

Growth kinetic parameters of the seven new strains in minimal medium using arabinose (MsmX independent uptake) and arabionotriose (MsmX dependent uptake) as the sole carbon and energy source, were determined and the results are summarized in Tables 3.7 and 3.8. The growth kinetics of the strains in arabinose remained unchanged in the three genetic systems used because the transport of this sugar is independent of MsmX and MsmX homologs. Thus, the analysis will focus on the growth kinetics of the strains in the presence of arabionotriose, which uptake is MsmX-dependent.

Table 3.7 – Growth of different *B. subtilis* strains IQB676, ISN10, ISN13 and ISN15 in the presence of distinct saccharides as sole carbon and energy source.

Carbon Source	Doubling Time (minutes)*			
	IQB676 ($\Delta msmX::cat$ $\Delta amyE::Pspank(hy)-$ $msmX(Glu3Ser,$ $Ile364Ser)$)	ISN10 ($\Delta msmX::cat$ $\Delta amyE::Pspank(hy)-$ $msmX-LEHis_6$)	ISN13 ($\Delta msmX::cat$ $\Delta amyE::Pspank(hy)-$ $ugpC-LEHis_6$ [<i>S. aureus</i>])	ISN15 ($\Delta msmX::cat$ $\Delta amyE::Pspank(hy)-$ $yurJ-LEHis_6$ [<i>B. subtilis</i>])
Arabinose 0.1%	70.48±5.70	69.02±2.96	69.10±2.31	71.05±2.43
Arabinose 0.1% + IPTG 1mM	74.15	69.85±3.42	69.51±2.17	71.39±0.43
Arabinotriose 0.1%	No Growth	No Growth	No Growth	No Growth
Arabinotriose 0.1% + IPTG 1mM	110.57±2.26	115.98±9.05	373.72±45.76	211.54±12.83

*Cells were grown in CSK medium (see Materials and Methods). Growth kinetics parameters were determined and the results represent the average of three independent experiments, except the growth in the presence of arabinose strain IQB676 with IPTG, which was performed once (see Annexes for the complete data used to determine the doubling time). The results of strain IQB676 are the same presented in table 6 (except the arabinose growth with IPTG) and are shown here to facilitate comparison.

B. subtilis ISN10 (Figure 3.20), which harbours the *msmX* gene with the C-terminal LEH₆-tag, grown in arabionotriose and with the inducer IPTG displays a doubling time similar to that of strain IQB676 (*msmX* gene) (Table 3.7; Figure 3.21). The recombinant ATPase YurJ-LEH₆ was able to partially complement MsmX function in AraNPQ transport system, as observed by the significant value of the growth rate of the strain ISN15 (Figure 3.20, Figure 3.23; Table 3.7) when compared to the inability to growth in the presence of arabionotriose displayed by strain ISN7 (YurJ-His₆) (Table 3.6). *B. subtilis* ISN13 (Figure 3.20), with the ATPase UgpC-LEH₆, exhibit growth

although the doubling time is elevated (Table 3.7; Figure 3.22). Nevertheless, is an improvement when compared to strain ISN5 (*UgpC-His₆*, Table 3.6).

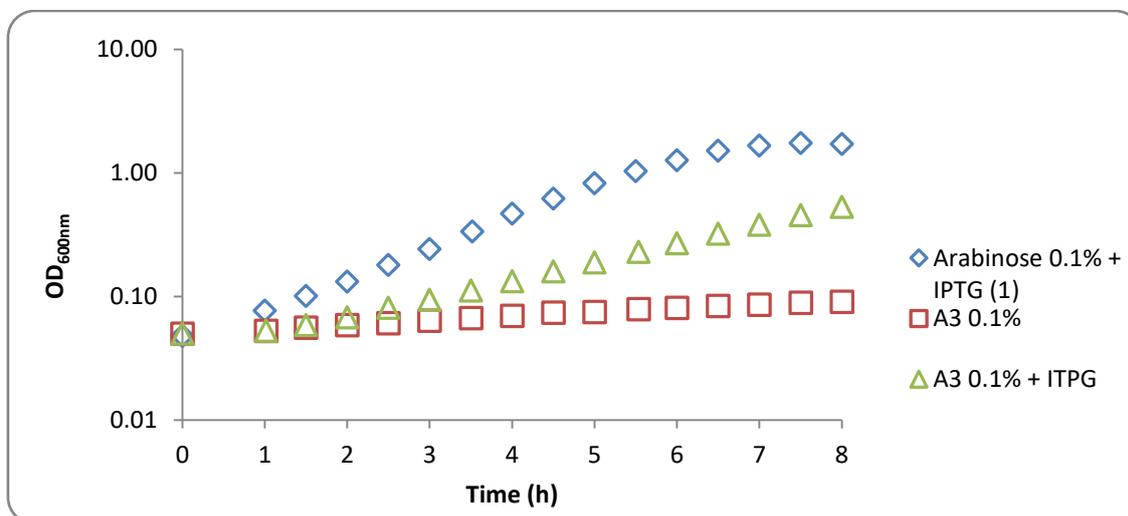


Figure 3.21 - Growth of *B. subtilis* ISN10 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmX-LEHis_6$) in CSK medium using arabinose and arabinotriose as the sole carbon and energy source.

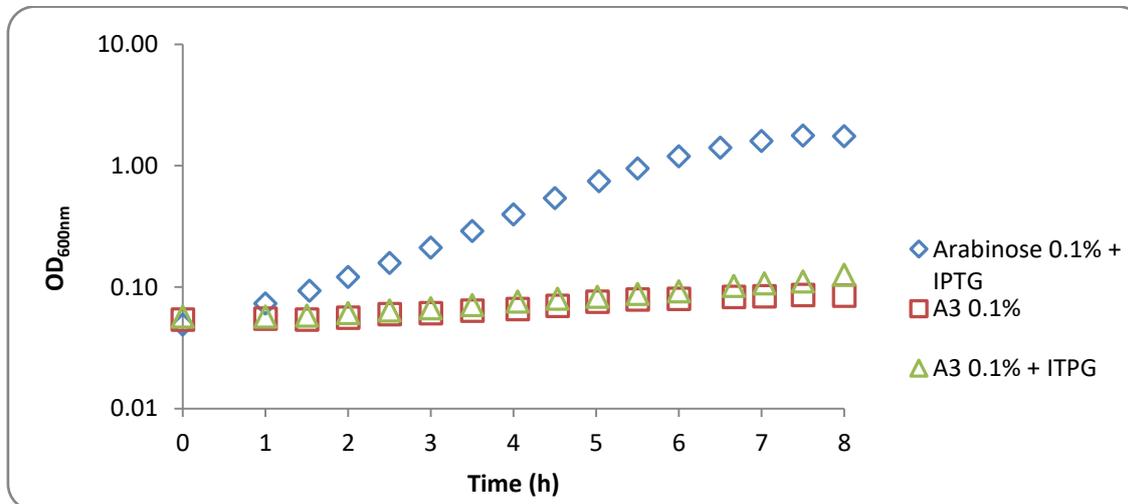


Figure 3.22 - Growth of *B. subtilis* ISN13 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ugpC-LEHis_6$) in CSK medium using arabinose and arabinotriose as the sole carbon and energy source.

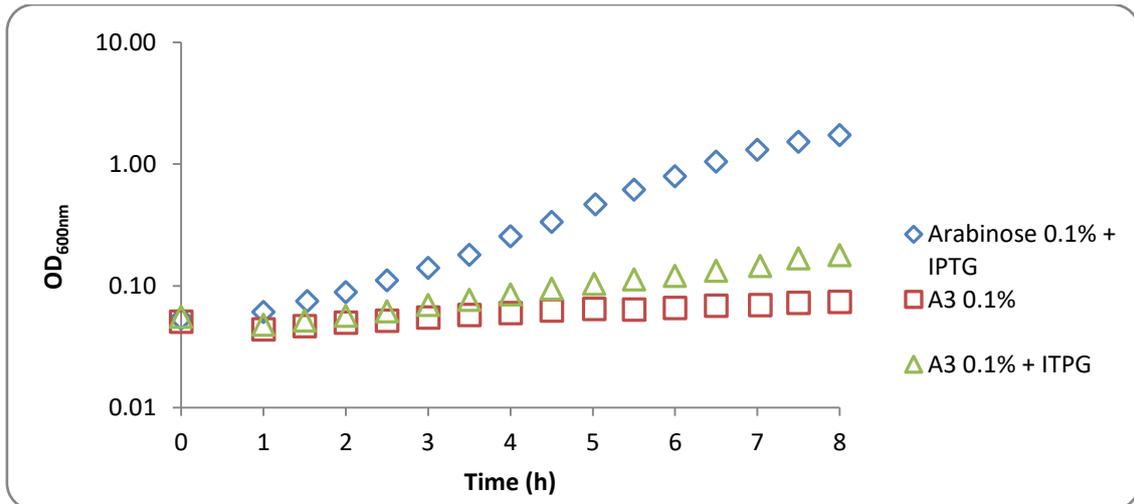


Figure 3.23 - Growth of *B. subtilis* ISN15 ($\Delta msx::cat \Delta amyE::Pspank(hy)-yurJ-LEHis_6$) in CSK medium using arabinose and arabinotriose as the sole carbon and energy source.

Growth kinetic parameters of the strains ISN11, ISN12, ISN14 and ISN17 in minimal medium using arabinose and arabinotriose as the sole carbon and energy source are summarized in the Table 3.8.

Table 3.8 – Growth of different *B. subtilis* strains ISN11, ISN12, ISN14 and ISN17 in the presence of distinct saccharides as sole carbon and energy source.

Carbon Source	Doubling Time (minutes)*			
	ISN11 ($\Delta msx::cat \Delta amyE::Pspank(hy)-HD73_RS21400-LEHis_6$ [<i>B. thuringiensis</i>])	ISN12 ($\Delta msx::cat \Delta amyE::Pspank(hy)-msmK-LEHis_6$ [<i>S. pneumoniae</i>])	ISN14 ($\Delta msx::cat \Delta amyE::Pspank(hy)-malK-LEHis_6$ [<i>E. coli</i>])	ISN17 ($\Delta msx::cat \Delta amyE::Pspank(hy)-ycjV-LEHis_6$ [<i>E. coli</i>])
Arabinose 0.1%	72.80±5.33	71.01±8.99	71.40±5.43	68.51±5.69
Arabinose 0.1% + IPTG 1mM	72.88±2.39	70.57±8.92	72.48±5.99	68.35±2.99
Arabinotriose 0.1%	No Growth	No Growth	No Growth	No Growth
Arabinotriose 0.1% + IPTG 1mM	113.81±6.19	148.81±2.58	No Growth	No Growth

*Cells were grown in CSK medium (see Materials and Methods). Growth kinetics parameters were determined and the results represent the average of three independent experiments (see Annexes for the complete data used to determine the doubling time).

The ATPases HD73_RS21400-LEH₆ and MsmK-LEH₆ were able to complement MsmX function with the same degree as the previous proteins construction, since the doubling time values of the strains ISN11 (Figure 3.24) and ISN12 (Figure 3.25), respectively, presented in Table 3.8 (construction of the strains are depicted in Figure 3.20), are very similar to the values of the strains ISN4 and ISN3, respectively (Table 3.6). The MalK-LEH₆, strain ISN14 (Figure 3.20), was not able to complement the MsmX deficiency (Table 3.8; Figure 3.26), as previously observed with MalK-His₆ (strain ISN6, Table 3.6). The ATPase YcjV was only designed with LEH₆-tag and is produced in the *B. subtilis* ISN17 strain (Figure 3.20), which displays inability to growth in the medium with arabionotriose and ITPG (Table 3.8; Figure 3.27), the same result was observed with the *B. subtilis* ISN16 strain (wild-type protein; Table 3.4), indicating that the ATPase remains non-functional.

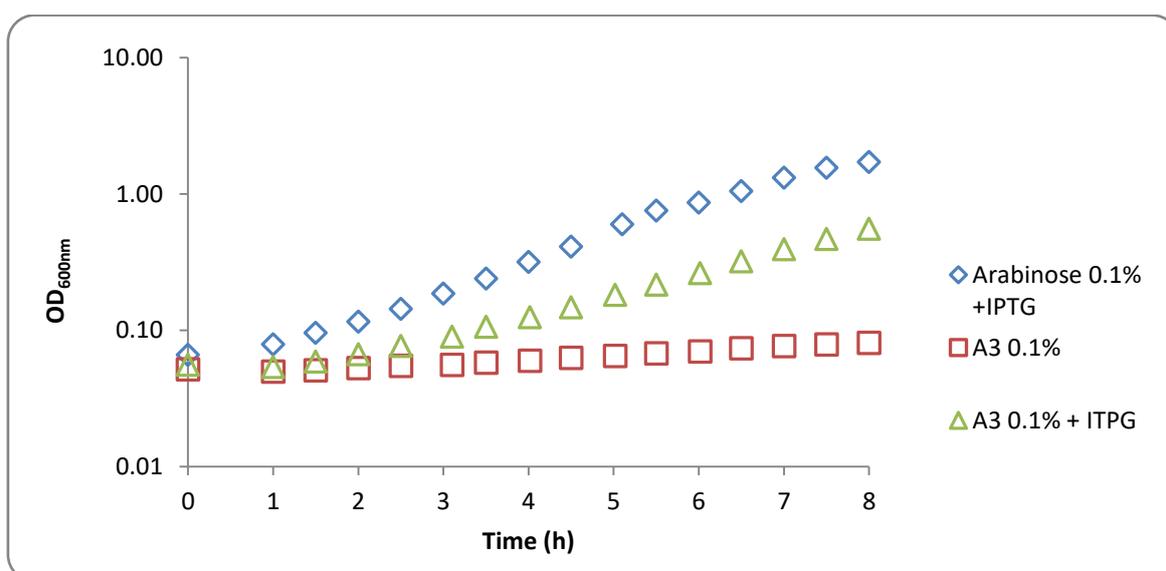


Figure 3.24 - Growth of *B. subtilis* ISN11 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-HD73_RS21400-LEHis_6$) in CSK medium using arabinose and arabionotriose as the sole carbon and energy source.

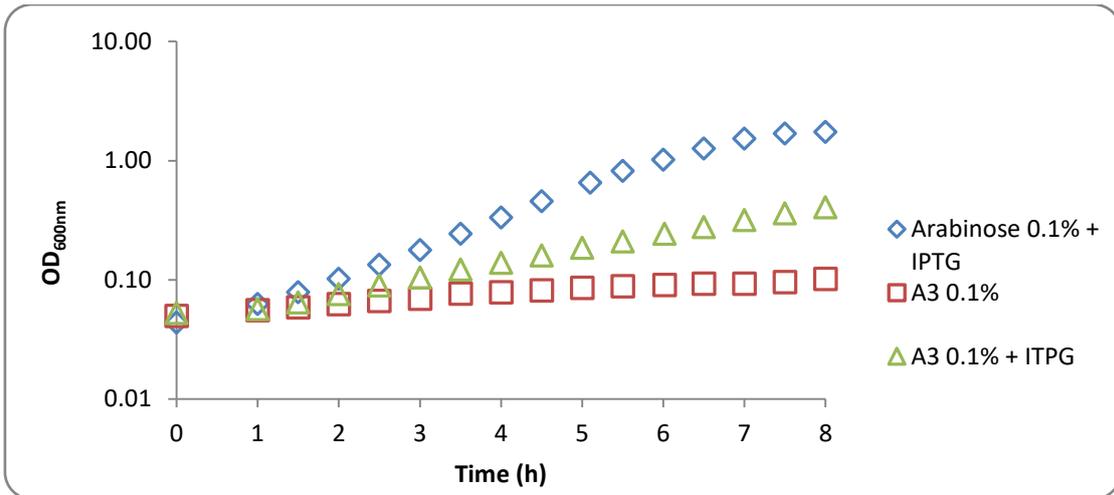


Figure 3.25 - Growth of *B. subtilis* ISN12 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmK-LEHis_6$) in CSK medium using arabinose and arabinotriose as the sole carbon and energy source.

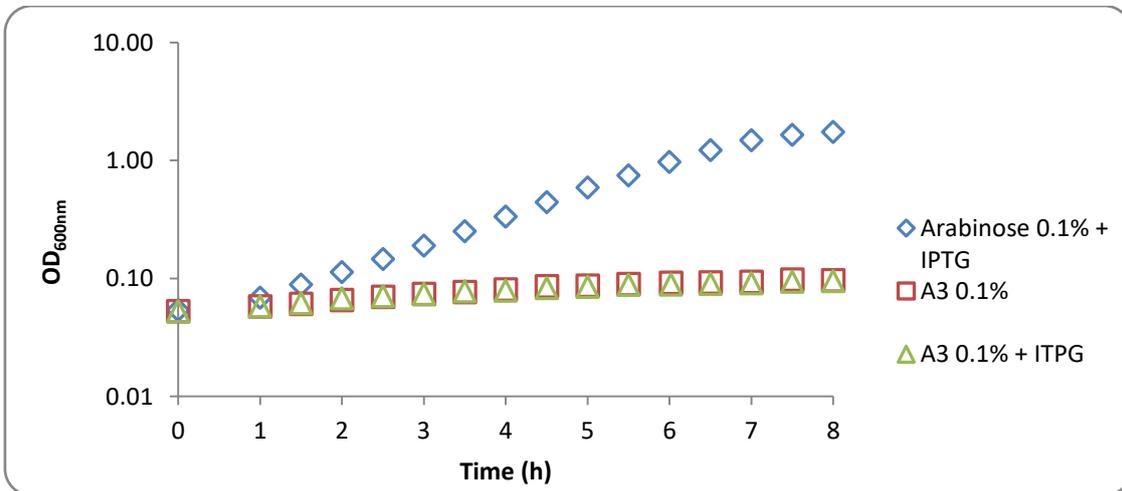


Figure 3.26 - Growth of *B. subtilis* ISN14 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-malk-LEHis_6$) in CSK medium using arabinose and arabinotriose as the sole carbon and energy source.

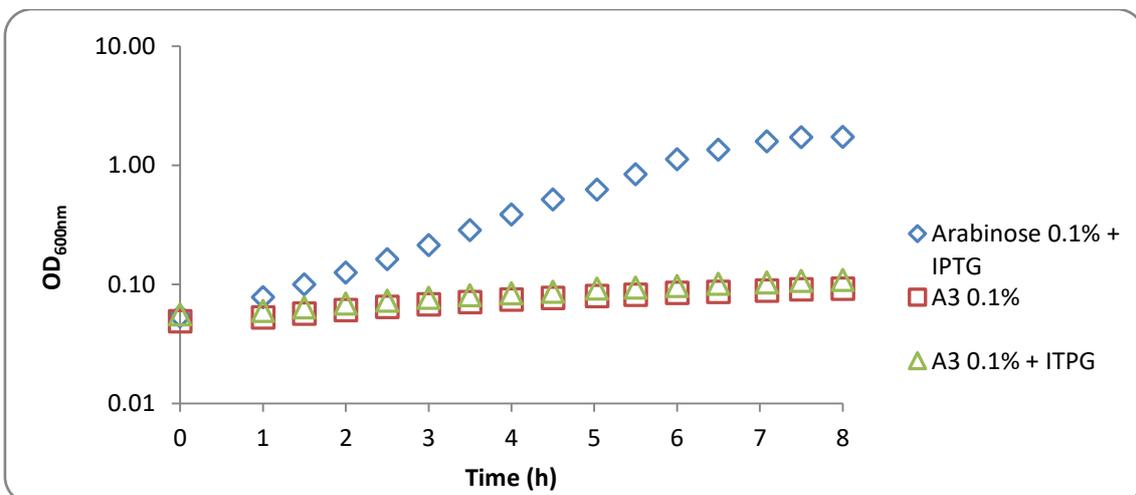


Figure 3.27 - Growth of *B. subtilis* ISN17 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ycjV-LEHis_6$) in CSK medium using arabinose and arabinotriose as the sole carbon and energy source.

3.1.5.2 Analysis of the 3-D structure models

The results of the growth kinetic assays support the analysis of the predicted 3-D protein structures and the comparison with the protein structures from the previous construction. MsmX ATPase structure differences were already analysed above (3.1.4 subsection) (Figure 3.16), and corroborate the normalization observed in the doubling time of the strain ISN10 with the redesigned genetic system. In the ATPase YurJ-LEH₆ model (from strain ISN15; Figure 3.28, B) the structure of the improved tail is spatially more close to the protein in comparison with the structure of the tail (H₆) in the ATPase YurJ-H₆ model (strain ISN7; Figure 3.28, A), and therefore may allow the formation of the homodimer, improving the hydrolyse of the ATP molecules and consequently the transmission of the conformational changes to the transmembrane domains that permit the entrance of arabinotriose into the cell (ter Beek *et al*, 2014).

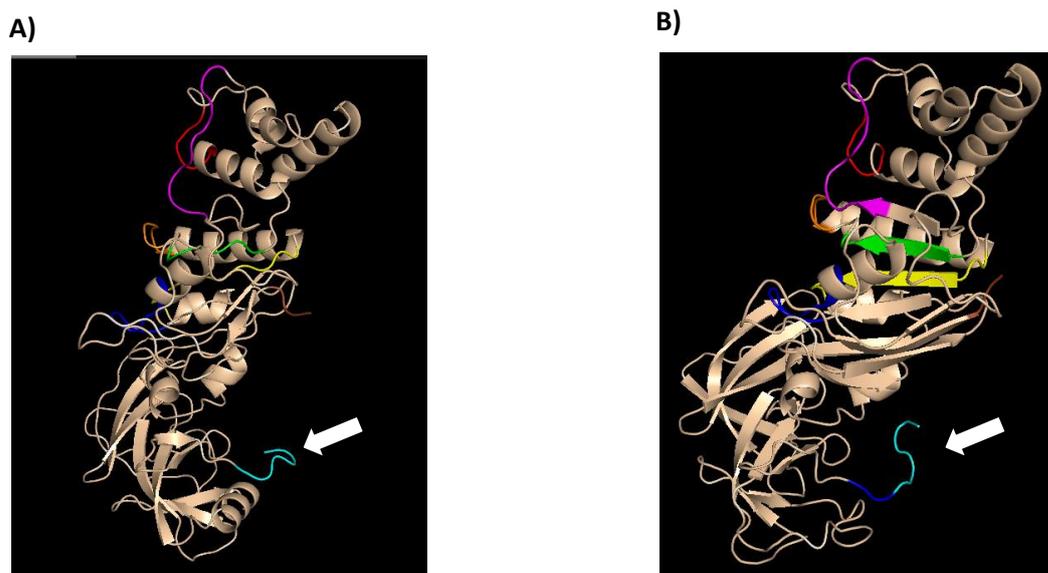


Figure 3.28 – Representation of YurJ-H₆ and YurJ-LEH₆ 3-D structure model. The structures were drawn in Pymol with the data obtained from the online program I-TASSER. **A)** Model of YurJ-H₆ protein from the strain ISN7; **(B)** Model of YurJ-LEH₆ protein from the strain ISN15; The N-terminal, the seven conserved domains and the C-terminal tag were highlighted with the following colours scheme: N-terminal (brown), Walker A (blue), Q-loop (magenta), Signature (red), Walker B (green), D-loop (orange), H-loop (yellow) and the C-terminal tag (cyan), which is highlighted by a white arrow.

The *B. subtilis* ISN13 strain (UgpC-LEH₆) shown the ability to grown, with arabionotriose in the medium and the inducer ITPG, although displaying a very high value of doubling time. This result is supported with the analysis of the 3-D structure of the ATPase with the improved tail (UgpC-LEH₆; Figure 3.29, B) versus the first recombinant ATPase (UgpC-H₆ from strain ISN5; Figure 3.29, A), where it is visible a change in the position of the tail, relatively to the protein. However, the improvement of the ATPase function was not that significant as it was observed for the ATPase YurJ, and even in this case the improvement is not ideal (wild-type protein). These observations may indicate that the tail construction needs further modifications or that the translation initiation rate is in fact interfering with the number of protein copies in the cell (discussed in 3.1.3.2 subsection) or that the mRNA structure may be less stable. The recombinant ATPase YcjV-LEH₆ is non-functional like the wild-type ATPase YcjV since the respective strains ISN17 and ISN16 display a non-growth phenotype, therefore YcjV-LEH₆ 3-D structure model is not relevant here for the discussion and is shown in the Appendix 6.67.

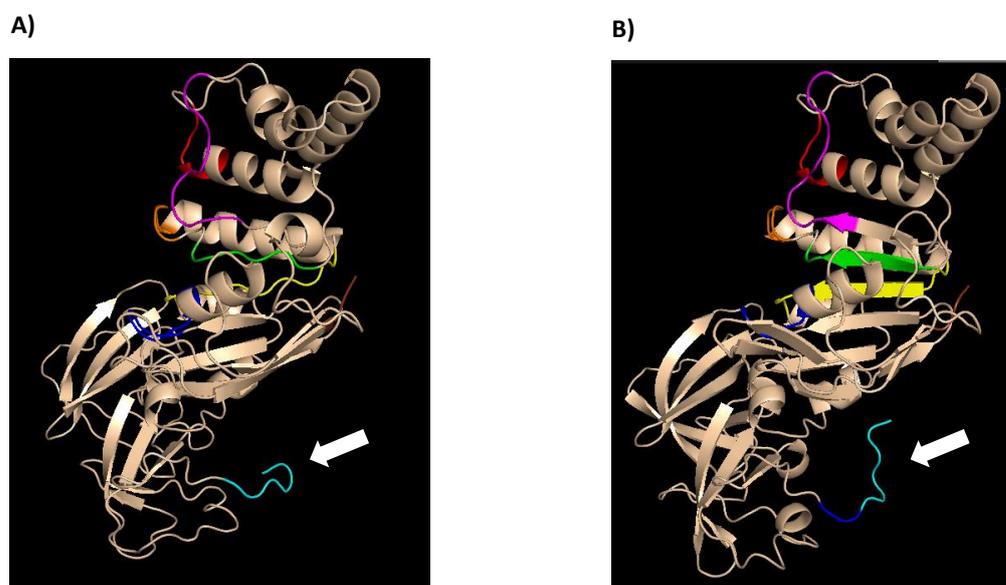


Figure 3.29 – Representation of UgpC-H₆ and UgpC-LEH₆ 3-D structure model. The structures were drawn in Pymol with the data obtained from the online program I-TASSER. **A)** Model of UgpC-H₆ protein from the strain ISN5; **(B)** Model of UgpC-LEH₆ protein from the strain ISN13; The N-terminal, the seven conserved domains and the C-terminal tag were highlighted with the following colours scheme: N-terminal (brown), Walker A (blue), Q-loop (magenta), Signature (red), Walker B (green), D-loop (orange), H-loop (yellow) and the C-terminal tag (cyan), which is highlighted by a white arrow.

3.2. Detection of the *in vivo* accumulation of the recombinant ATPases in the cells.

To confirm the production of the different recombinant ATPases in the *B. subtilis* strains ISN10-15 and ISN17, constructed with the redesigned system (LEH₆), cell extracts were prepared from these strains grown in minimal medium supplemented with arabinose or arabinotriose and in the presence or absence of the inducer IPTG, and analysed by Western-blot using an anti-His₆-specific antibody. Purified MsmX-LEHis₆ was used as a positive control. The results of western-blot analysis (films and the respective transference membrane) are shown in Figures 3.30-3.33.

In the experiment to detect the accumulation of ATPase in cells grown in the presence of arabinose and the inducer IPTG it is visible the presence of a single band in wells 2 to 8, with a position that corresponds to molecular weight between 40 kDa and 48 kDa.

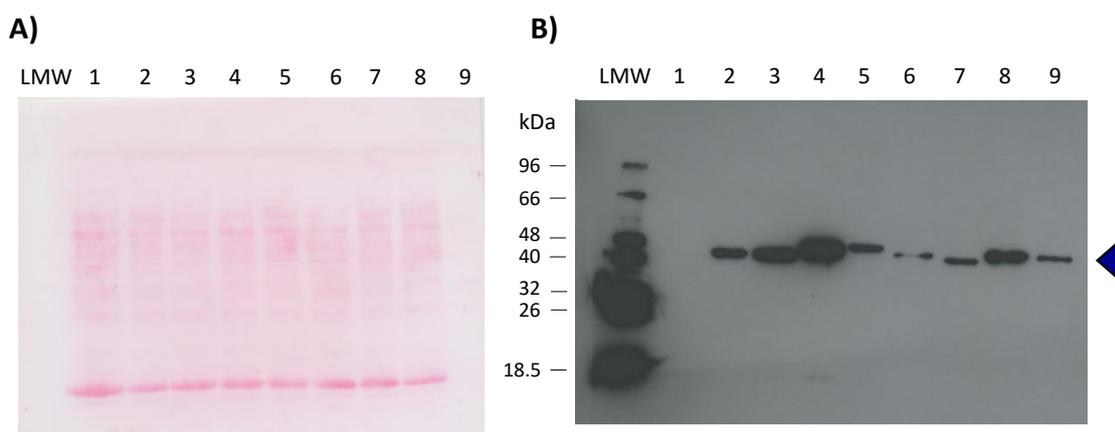


Figure 3.30–Western-Blot analysis of *B. subtilis* cell extracts, obtained from CSK medium with arabinose 0.1% in inducing conditions. A) Nitrocellulose membrane with the fractionated proteins transferred from the gel of SDS-PAGE (12.5%) (10 μ g of total extract, except 20 μ g of total extract of IQB622, and 0.6 μ g of purified MsmX-LEHis₆ were loaded); **B)** Autoradiography plate after His-tag detection using an anti-His₆-specific antibody (3 min exposure). LMW represents the Low Molecular Weight Protein Marker (NZYTech). **1)** total extract of IQB622 (*msmX*-His₆ *cat*); **2)** total extract of ISN10 (Δ *msmX*::*cat* Δ *amyE*::Pspank(hy)-*msmX*-LEHis₆); **3)** total extract of ISN11 (Δ *msmX*::*cat* Δ *amyE*::Pspank(hy)-*HD73_RS21400*-LEHis₆); **4)** total extract of ISN12 (Δ *msmX*::*cat* Δ *amyE*::Pspank(hy)-*msmK*-LEHis₆); **5)** total extract of ISN13 (Δ *msmX*::*cat* Δ *amyE*::Pspank(hy)-*ugpC*-LEHis₆); **6)** total extract of ISN15 (Δ *msmX*::*cat* Δ *amyE*::Pspank(hy)-*yurJ*-LEHis₆); **7)** total extract of ISN17 (Δ *msmX*::*cat* Δ *amyE*::Pspank(hy)-*ycjV*-LEHis₆); **8)** total extract of ISN14 (Δ *msmX*::*cat* Δ *amyE*::Pspank(hy)-*malk*-LEHis₆); **9)** purified MsmX-LEHis₆ (MW= 42.431 kDa) is indicated with a blue triangle.

Those bands correspond to the different ATPases in study, which predicted molecular weight is displayed in Appendix 6.75, when compared to the 42.431 kDa band of the purified MsmX-LEHis₆ protein detected in well 9. The MsmX-LEHis₆ from *B. subtilis* IQB622 strain was detected in previous experiments in the same minimal medium with arabinose (without IPTG) by Ferreira and Sá-Nogueira (unpublished work), however in this experiment we failed to detect the band corresponding to the protein (well 1). The transference process of the proteins from the gel to the membrane was successful, as is indicated by the membrane revealed with Ponceau (Figure 3.30, A). The different intensity of the signal may correspond to distinct concentration of ATPases in the cell and/or better exposition to the His-tag to the antibody.

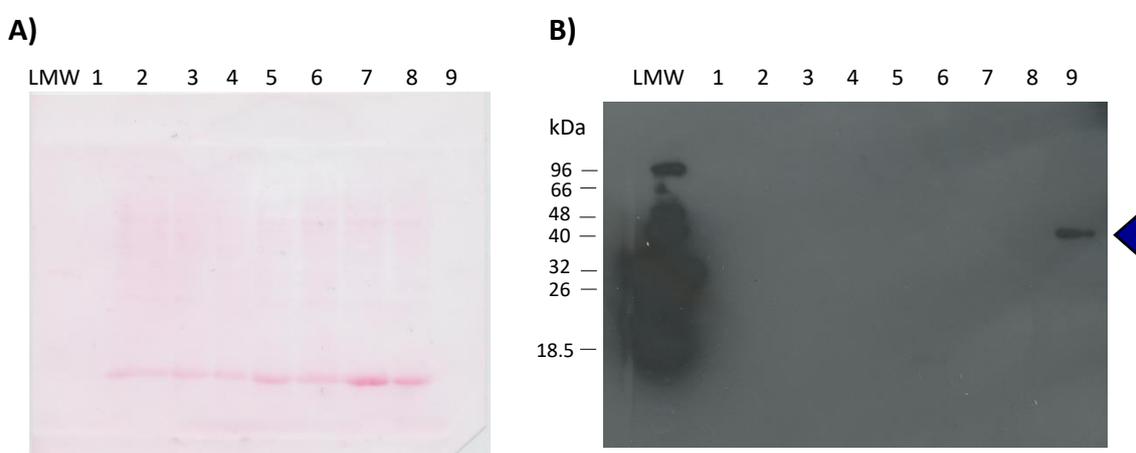


Figure 3.31—Western-Blot analysis of *B. subtilis* cell extracts, obtained from CSK medium with arabinose 0.1% in non-inducing conditions. **A)** Nitrocellulose membrane with the fractionated proteins transferred from the gel of SDS-PAGE (12.5%) (10 µg of total extract and 0.6 µg of purified MsmX-LEHis₆ were loaded); **B)** Autoradiography plate after His-tag detection using an anti-His₆-specific antibody (4 min exposure). LMW represents the Low Molecular Weight Protein Marker (NZYTech). **1)** Empty; **2)** total extract of ISN10 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmX-LEHis_6$); **3)** total extract of ISN11 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-HD73_RS21400-LEHis_6$); **4)** total extract of ISN12 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmK-LEHis_6$); **5)** total extract of ISN13 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ugpC-LEHis_6$); **6)** total extract of ISN15 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-yurJ-LEHis_6$); **7)** total extract of ISN17 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ycjV-LEHis_6$); **8)** total extract of ISN14 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-malk-LEHis_6$); **9)** purified MsmX-LEHis₆. The position of the purified MsmX-LEHis₆ (MW= 42.431 kDa) is indicated with a blue triangle.

In non-inducing conditions (absence of IPTG) the ATPases are not produced (Figure 3.31, B), since the only band observed corresponds to the purified MsmX-LEHis₆ protein (42.431 kDa). Again, the transference process was successful as observed by the coloured bands in the membrane after staining with Ponceau (Figure 3.31, A). These results demonstrate the effectiveness of the control by the inducer over the expression and production of the ATPases in this genetic system.

The uptake of arabinotriose through the AraNPQ transport is dependent of the multitask MsmX (Ferreira and Sá-Nogueira, 2010), and the same is valid to MsmX homologs HD73_RS21400, MsmK, UgpC and YurJ, which are able to complement in vivo the MsmX function. However, the homologs Malk and YcjV are not able to complement MsmX function, neither one of the recombinant forms of YurJ and UgpC (YurJ-His₆ and UgpC-His₆, respectively; see previous sections). Therefore, is essential to verify the present of the ATPases in the cells in order to clarify this aspect and for this purpose western-blot analysis was also performed with extracts of cells grown in the presence of arabinotriose with (Figure 3.32) or without IPTG (Figure 3.33).

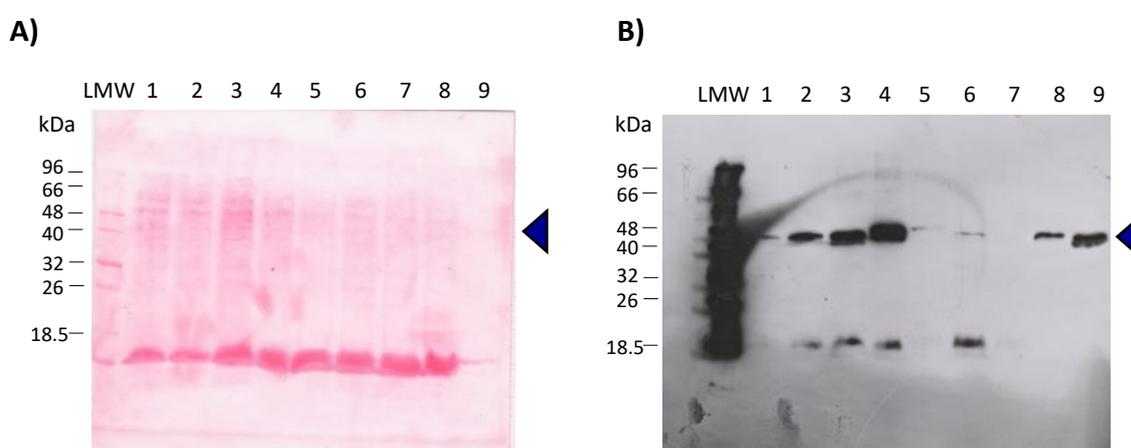


Figure 3.32–Western-Blot analysis of *B. subtilis* cell extracts, obtained from CSK medium with arabinotriose 0.1% in inducing conditions. A) Nitrocellulose membrane with the fractionated proteins transferred from the gel of SDS-PAGE (12.5%) (10 µg of total extract and 0.6 µg of purified MsmX-LEHis₆ were loaded); **B)** Autoradiography plate after His-tag detection using an anti-His₆-specific antibody (30 sec exposure). LMW represents the Low Molecular Weight Protein Marker (NZYTech). **1)** total extract of IQB622 (*msmX*-His₆ *cat*); **2)** total extract of ISN10 (Δ *msmX*::*cat* Δ *amyE*::Pspank(hy)-*msmX*-LEHis₆); **3)** total extract of ISN11 (Δ *msmX*::*cat* Δ *amyE*::Pspank(hy)-*HD73_RS21400*-LEHis₆); **4)** total extract of ISN12 (Δ *msmX*::*cat* Δ *amyE*::Pspank(hy)-*msmK*-LEHis₆); **5)** total extract of ISN13 (Δ *msmX*::*cat* Δ *amyE*::Pspank(hy)-*ugpC*-LEHis₆); **6)** total extract of ISN15 (Δ *msmX*::*cat* Δ *amyE*::Pspank(hy)-*yurJ*-LEHis₆); **7)** total extract of ISN17 (Δ *msmX*::*cat* Δ *amyE*::Pspank(hy)-*ycjV*-LEHis₆); **8)** total extract of ISN14 (Δ *msmX*::*cat* Δ *amyE*::Pspank(hy)-*malk*-LEHis₆); **9)** purified MsmX-LEHis₆. The position of the purified MsmX-LEHis₆ (MW= 42.431 kDa) is indicated with a blue triangle.

In the presence of arabinotriose and IPTG the western-blot film (figure 3.32, B)) presents a single band that corresponds to a molecular mass between 40 kDa and 48 kDa in well 2 to 6 and well 8, which parallel to all the different ATPases studied in this work thereby demonstrating the presence of the proteins in the respective strains. The band corresponding to YcjV-LEHis₆ is not visible in this film (well 7), however after increasing the exposure to 3 min it is possible to visualize a discreet signal (data not shown). The results are supported by the positive control

(purified MsmX-LEHis₆) detected in well 9 with a corresponding molecular weight of 42.431 kDa. The MsmX-LEH₆ from IQB622 (grown in non-induction conditions) also displays a band in well 1, although with low intensity. The transference process of the proteins from the gel to the membrane was successful, as is indicated by staining with Ponceau (Figure 3.32, A)). As discussed above the differences in the intensity of the signal may correspond to distinct concentration of ATPases in the cell and/or better exposition to the His-tag to the antibody.

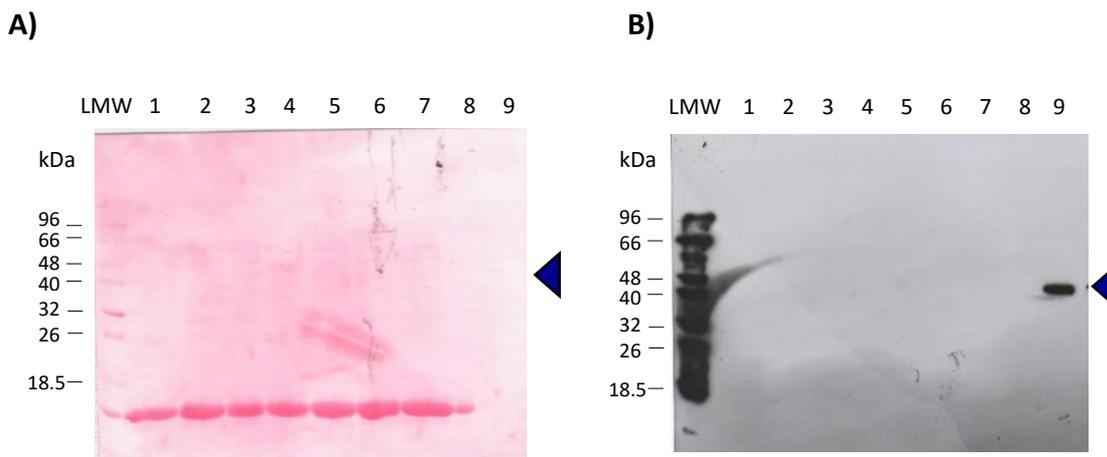


Figure 3.33—Western-Blot analysis of *B. subtilis* cell extracts, obtained from CSK medium with arabinotriose 0.1% in non-inducing conditions. A) Nitrocellulose membrane with the fractionated proteins transferred from the gel of SDS-PAGE (12.5%) (10 µg of total extract and 0.6 µg of purified MsmX-His were loaded); **B)** Autoradiography plate after His-tag detection using an anti-His₆-specific antibody (30 sec exposure). LMW represents the Low Molecular Weight Protein Marker (NZYTech). **1)** total extract of ISN10 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmX-LEHis_6$); **2)** total extract of ISN11 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-HD73_RS21400-LEHis_6$); **3)** total extract of ISN12 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmK-LEHis_6$); **4)** total extract of ISN13 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ugpC-LEHis_6$); **5)** total extract of ISN15 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-yurJ-LEHis_6$); **6)** total extract of ISN17 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ycjV-LEHis_6$); **7)** total extract of ISN14 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-malk-LEHis_6$); **8)** Empty; **9)** purified MsmX-LEHis₆. The position of the purified MsmX-LEHis₆ (MW= 42.431 kDa) is indicated with a blue triangle.

In non-inducing conditions the *B. subtilis* strains ISN10-15 and ISN17 do not accumulate the respective recombinant ATPases (Figure 3.33, B)) as observed by the absence of bands corresponding to the different ATPases. The positive control (MsmX-LEHis₆ purified) is visible in well 9 with a MW = 42.431 kDa, demonstrating that the technique was successful. The transference process of the proteins from the gel to the membrane was also successful, as seen by the bands revealed with Ponceau (Figure 3.33, A)).

These results obtained in this section need to be validate by repeating the experiments.

Chapter 4

Concluding Remarks and Future Perspectives

4. Concluding Remarks and Future Perspectives

The discovery of antibiotics and its fabrication constitute a great accomplishment in the field of medicine, by providing the necessary tools to patients that fight infections caused by pathogenic microbes. However, pathogenic bacteria remains a constant concern to the population due to the antibiotic resistance of this organisms through constant mutations in their genome, leading to a pressing need in the research for new anti-infective strategies.

Bacterial metabolism has been closely related with virulence of pathogenic strains, namely in *Salmonella Typhimurium*, *Vibrio cholera*, *Helicobacter pylori*, etc (Rohmer *et al*, 2011). Within the metabolic components, ABC importers constitute an attractive target for vaccine or therapy strategies since they are exclusive to prokaryotes (few exceptions in plants). A therapeutic approach could be the use of ABC transporters as a novel drug delivery specific system through the recognition and bound of the transporters to engineered antimicrobials or, another approach, could be the direct disruption of the transport system. A protein-based vaccine with transporter components is a strong candidate in providing a selective pressure in the medium to eliminate strains with the target transporter (Buckwalter *et al*, 2015).

In this work we have studied the functionality of distinct ATPases from pathogenic species through their ability to complement MsmX ATPase function in the AraNPQ ABC-type importer, in a *B. subtilis* *msmX*-null mutant. The results indicated that an ATPase from *B. thuringiensis* was able to fulfill the role of MsmX in its absence, while the YurJ ATPase from *B. subtilis* displayed only a partial degree of complementation. In addition to intra- and interspecies interchangeability of *Bacillus* ATPases, we show the existence of different genera interchangeability through the ability of the ATPases from *Streptococcus pneumoniae* and *Staphylococcus aureus* to complement to a certain degree the *B. subtilis* MsmX function in vivo. This phenomenon was observed only with ATPases from Gram-positive bacteria since the two ATPases from the Gram-negative bacterium *Escherichia coli* were not functional in *B. subtilis*. Western-blot experiments showed that all the different ATPases were present in the samples of the respective *B. subtilis* strains, which indicates that the proteins are being translated in the cells. In summary, our study shows that *B. subtilis* can be use as model for the study of bacterial multitask ATPases.

Here, we have redesigned a genetic system that provides a tool for the characterization of interspecies interchangeability in bacterial carbohydrate transport and particularly in bacterial pathogens. However, it is still possible to further improve the genetic system and hence the results. Two ATPases, YurJ from *B. subtilis* and UgpC from *S. aureus*, exhibited significantly different results when comparing the wild-type form to the recombinant form. In order to clarify this aspect, different genetic systems could be constructed to isolate and study each of the variables that were introduced in the actual genetic system: C- and N-terminal modifications, RBS and terminator of the *msmX* gene, and the His-tag construction.

Regarding the non-functional ATPases from *E. coli*, we have noticed in the multiple sequence alignments an absence of five amino acids in a specific region of only these two proteins. In order to determine the importance of this region to the function of the protein function, site-directed mutagenesis experiments may be conducted in order to add the nucleotides of the *msmX* gene that correspond to that five amino acids, and see if the protein became functional in *B. subtilis*, in particular YcjV, which displays high amino acid identity to MsmX. We may also evaluate the complementation degree of these two ATPases from the Gram-negative bacteria *E. coli*, in another sugar ABC transporter energized by MsmX ATPase in *B. subtilis*. Additionally, we may also test ATPases from other Gram-negative bacteria rather than *E. coli*. Furthermore, crystallographic studies of the MsmX currently in progress will allow new insight into the molecular mechanism of this type of transport and versatile ATPases, together with future mutagenic and protein-protein interaction studies.

Chapter 5

References

5. References

- Abbott, D.W., Higgins, M.A., Hyrnuik, S., Pluvinage, B., Lammerts Van Bueren, A., Boraston, A.B.** (2010). The molecular basis of glycogen breakdown and transport in *Streptococcus pneumoniae*. *Mol. Microbiol.* **77**(1): 183–99.
- Bass, R.B., Strop, P., Barclay, M., Rees, D.C.** (2002). Crystal structure of *Escherichia coli* MscS, a voltage-modulated and mechanosensitive channel. *Science.* **298**(5598): 1582–7.
- Berntsson, R.P.A., Smits, S.H.J., Schmitt, L., Slotboom, D.J., Poolman, B.** (2010). A structural classification of substrate-binding proteins. *FEBS Lett.* **584**(12): 2606–17.
- Biemans-Oldehinkel, E., Doeven, M.K., Poolman, B.** (2006). ABC transporter architecture and regulatory roles of accessory domains. *FEBS Lett.* **580**(4): 1023–35.
- Böhm, S., Licht, A., Wuttge, S., Schneider, E., Bordignon, E.** (2013). Conformational plasticity of the type I maltose ABC importer. *Proc. Natl. Acad. Sci. USA.* **110**(14): 5492–7.
- Boos, W., Shuman, H.** (1998). Maltose/Maltodextrin System of *Escherichia coli*: Transport, Metabolism, and Regulation. *Microbiol. Mol. Biol. Rev.* **62**(1): 204–29.
- Bordignon, E., Grote, M., Schneider, E.** (2010). The maltose ATP-binding cassette transporter in the 21st century - towards a structural dynamic perspective on its mode of action. *Mol. Microbiol.* **77**(6): 1354–66.
- 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): 4881–90.
- Buckwalter, C.M., King, S.J.** (2012). Pneumococcal carbohydrate transport: Food for thought. *Trends Microbiol.* **20**(11): 517–22.
- Casula, G., Cutting, S.M.** (2002). *Bacillus* Probiotics: Spore Germination in the Gastrointestinal Tract. *Appl. Environ. Microbiol.* **68**(5): 2344–52.
- Chauhan, P.S., Puri, N., Sharma, P., Gupta, N.** (2012). Mannanases: Microbial sources, production, properties and potential biotechnological applications. *Appl. Microbiol. Biotechnol.* **93**(5): 1817–30.

Chen, J., Sharma, S., Quioco, F.A., Davidson, A.L. (2001). Trapping the transition state of an ATP-binding cassette transporter: evidence for a concerted mechanism of maltose transport. *Proc. Natl. Acad. Sci. USA.* **98**(4): 1525–30.

Chen, J. (2013). Molecular mechanism of the *Escherichia coli* maltose transporter. *Curr. Opin. Struct. Biol.* **23**(4): 492–8.

Chevance, F.F.V., Erhardt, M., Lengsfeld, C., Lee, S.J., Boos, W. (2006). Mlc of *Thermus thermophilus*: A glucose-specific regulator for a glucose/mannose ABC transporter in the absence of the phosphotransferase system. *J. Bacteriol.* **188**(18): 6561–71.

Davidson, A.L., Shuman, H.A., Nikaido, H. (1992). Mechanism of maltose transport in *Escherichia coli*: transmembrane signaling by periplasmic binding proteins. *Proc. Natl. Acad. Sci. USA.* **89**(6): 2360–4.

Davidson, A.L., Sharma, S. (1997). Mutation of a single MalK subunit severely impairs maltose transport activity in *Escherichia coli*. *J. Bacteriol.* **179**(17): 5458–64.

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.

Daus, M.L., Grote, M., Müller, P., Doebber, M., Herrmann, A., Steinhoff, H-J., et al. (2007). ATP-driven MalK dimer closure and reopening and conformational changes of the “EAA” motifs are crucial for function of the maltose ATP-binding cassette transporter (MalFGK2). *J. Biol. Chem.* **282**(31): 22387–96.

Daus, M.L., Grote, M., Schneider, E. (2009). The MalF P2 loop of the ATP-binding cassette transporter MalFGK2 from *Escherichia coli* and *Salmonella enterica* serovar typhimurium interacts with maltose binding protein (MalE) throughout the catalytic cycle. *J. Bacteriol.* **191**(3): 754–61.

Dawson, R.J.P., Hollenstein, K., Locher, K.P. (2007). Uptake or extrusion: Crystal structures of full ABC transporters suggest a common mechanism. *Mol. Microbiol.* **65**(2): 250–7.

Dhawan, S., Kaur, J. (2007) Microbial mannanases: an overview of production and applications. *Crit. Rev. Biotechnol.* **27**(4): 197–216.

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. (2011). Canonical and ECF-type ATP-binding cassette importers in prokaryotes: Diversity in modular organization and cellular functions. *FEMS Microbiol. Rev.* **35**(1): 3–67.

Erkens, G.B., Berntsson, R.P.-A., Fulyani, F., Majsnerowska, M., Vujičić-Žagar, A., ter Beek, J., et al. (2011). The structural basis of modularity in ECF-type ABC transporters. *Nat. Struct. Mol. Biol.* **18**(7): 755–60.

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.

Grote, M., Bordignon, E., Polyhach, Y., Jeschke, G. (2008) A Comparative Electron Paramagnetic Resonance Study of the Nucleotide-Binding. *Biophys. J.* **95**(6): 2924–38.

Grote, M., Polyhach, Y., Jeschke, G., Steinhoff, H.J., Schneider, E., Bordignon, E. (2009). Transmembrane signaling in the maltose ABC transporter MalFGK2-E: periplasmic MalF-P2 loop communicates substrate availability to the ATP-bound MalK dimer. *J. Biol. Chem.* **284**(26): 17521–6.

Harwood, C.R. (1992). *Bacillus subtilis* and its relatives: molecular biological and industrial workhorses. *Trends Biotechnol.* **10**(7): 247-56.

Higgins, C.F., Linton, K.J. (2004). The ATP switch model for ABC transporters. *Nat. Struct. Mol. Biol.* **11**(10): 918–26.

Inácio, J.M., Correia, I.L., de Sá-Nogueira, I. (2008). Two distinct arabinofuranosidases contribute to arabino-oligosaccharide degradation in *Bacillus subtilis*. *Microbiology.* **154**: 2719–29.

Jacso, T., Grote, M., Daus, M.L., Schmieder, P., Keller, S., Schneider, E., et al. (2009). Periplasmic Loop P2 of the MalF Subunit of the Maltose ATP Binding Cassette Transporter Is Sufficient To Bind the Maltose Binding Protein MalE. *Biochemistry.* **48**(10):2216–25.

Jacso, T., Schneider, E., Rupp, B., Reif, B. (2012). Substrate transport activation is mediated through second periplasmic loop of transmembrane protein Malf in maltose transport complex of *Escherichia coli*. *J. Biol. Chem.* **287**(21): 17040–9.

Jaehme, M., Slotboom, D.J. (2015). Diversity of membrane transport proteins for vitamins in bacteria and archaea. *Biochim. Biophys. Acta.* **1850**(3): 565–76.

Jardetzky, O. (1966). Simple allosteric model for membrane pumps. *Nature.* **211**(5052): 969–70.

Khare, D., Oldham, M.L., Orelle, C., Davidson, A.L., Chen, J. (2009). Alternating Access in Maltose Transporter Mediated by Rigid-Body Rotations. *Mol. Cell.* **33**(4): 528–36.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., et al. (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature.* **390**(6657): 249–56.

Levina, N., Töttemeyer, S., Stokes, N.R., Louis, P., Jones, M.A., Booth, I.R. (1999). Protection of *Escherichia coli* cells against extreme turgor by activation of MscS and MscL mechanosensitive channels: identification of genes required for MscS activity. *EMBO J.* **18**(7): 1730–7.

Locher, K.P. (2009). Review. Structure and mechanism of ATP-binding cassette transporters. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **364**(1514): 239–45.

Lu, G., Westbrook, J.M., Davidson, A.L., Chen, J. (2005). ATP hydrolysis is required to reset the ATP-binding cassette dimer into the resting-state conformation. *Proc. Natl. Acad. Sci. USA.* **102**(50): 17969–74.

Mabrouk, M.E.M., El-Ahwany, A.M.D. (2008). Production of β -mannanase by *Bacillus amylolequifaciens* 10A1 cultured on potato peels. *African J. Biotechnol.* **7**(8): 1123–8.

Marion, C., Burnaugh, A.M., Woodiga, S.A., King, S.J. (2011) Sialic acid transport contributes to pneumococcal colonization. *Infect. Immun.* **79**(3): 1262–9.

Marion, C., Aten, A.E., Woodiga, S.A., King, S.J. (2011b). 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.* **208**(1–2): 177–84.

Meenakshi, Singh, G., Bhalla, A., Hoondal, G.S. (2010). Solid state fermentation and characterization of partially purified thermostable mannanase from *Bacillus* sp. MG-33. *BioResources.* **5**(3): 1689–701.

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

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

Oldham, M.L., Davidson, A.L., Chen, J. (2008). Structural insights into ABC transporter mechanism. *Curr. Opin. Struct. Biol.* **18**(6): 726–33.

Oldham, M.L., Chen J. (2011). Snapshots of the maltose transporter during ATP hydrolysis. *Proc. Natl. Acad. Sci. USA.* **108**(37): 15152–6.

Orelle, C., Ayvaz, T., Everly, R.M., Klug, C.S., Davidson, A.L. (2008). Both maltose-binding protein and ATP are required for nucleotide-binding domain closure in the intact maltose ABC transporter. *Proc. Natl. Acad. Sci. USA.* **105**(35): 12837–42.

Orelle, C., Alvarez, F.J., Oldham, M.L., Orelle, A., Wiley, T.E., Chen, J., et al. (2010). Dynamics of alpha-helical subdomain rotation in the intact maltose ATP-binding cassette transporter. *Proc. Natl. Acad. Sci. USA.* **107**(47): 20293–8.

Pascal, M., Kunst, F., Lepesant, J.A., Dedonder, R. (1971). Characterization of two sucrase activities in *Bacillus subtilis* marburg. *Biochimie.* **53**(10): 1059–66.

Polizeli, M.L.T.M., Rizzatti, A.C.S., Monti, R., Terenzi, H.F., Jorge, J.A., Amorim, D.S. (2005). Xylanases from fungi: Properties and industrial applications. *Appl. Microbiol. Biotechnol.* **67**(5): 577–91.

Postma, P.W., Lengeler, J.W., Jacobson, G.R. (1993). Phosphoenolpyruvate: Carbohydrate phosphotransferase systems of bacteria. *Microbiol. Rev.* **57**(3): 543–94.

- Priest, F.** (1993). Systematics and Ecology of *Bacillus*. In Sonenshein, A. L., Hoch, J. A. and Losick, R. (ed.), *Bacillus subtilis* and other Gram-positive Bacteria. ASM Press, Washington, DC.
- Puyet, A., Espinosa, M.** (1993). Structure of the maltodextrin-uptake locus of *Streptococcus pneumoniae*. Correlation to the *Escherichia coli* maltose regulon. *J. Mol. Biol.* **230**(3): 800–11.
- Quentin, Y., Fichant, G., Denizot, F.** (1999). Inventory, assembly and analysis of *Bacillus subtilis* ABC transport systems. *J. Mol. Biol.* **287**(3): 467–84.
- Quiocho, F.A., Ledvina, P.S.** (1996). Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol. Microbiol.* **20**(1): 17-25.
- Quisel, J.D., Burkholder, W.F., Alan, D., Grossman, A.D.** (2001). In Vivo Effects of Sporulation Kinases on Mutant Spo0A Proteins in *Bacillus subtilis*. *J. Bacteriol.* **183**(22): 6573–8.
- Ren, D., Navarro, B., Xu, H., Yue, L., Shi, Q., Clapham, D.E.** (2001). A prokaryotic voltage-gated sodium channel. *Science.* **294**(5550): 2372-5.
- Rice, A.J., Alvarez, F.J.D., Schultz, K.M., Klug, C.S., Davidson, A.L., Pinkett, H.W.** (2013). EPR spectroscopy of MolB₂C₂-A reveals mechanism of transport for a bacterial type II molybdate importer. *J. Biol. Chem.* **288**(29): 21228–35.
- Rice, A.J., Park, A., Pinkett, H.W.** (2014). Diversity in ABC transporters: Type I, II and III importers. *Crit. Rev. Biochem. Mol. Biol.* **49**(5): 426–37.
- Robillard, G.T., Broos, J.** (1999). Structure/function studies on the bacterial carbohydrate transporters, enzymes II, of the phosphoenolpyruvate-dependent phosphotransferase system. *Biochim. Biophys. Acta.* **1422**(2): 73–104.
- Rodionov, D.A., Hebbeln, P., Eudes, A., ter Beek, J., Rodionova, I.A., Erkens, G.B., et al.** (2009). A novel class of modular transporters for vitamins in prokaryotes. *J. Bacteriol.* **91**(1): 42–51.
- Rohmer, L., Hocquet, D., Miller, S.I.** (2011). Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. *Trends Microbiol.* **19**(7): 341-8

- Rosenow, C., Maniar, M., Trias, J. (1999). Regulation of the alpha-galactosidase activity in *Streptococcus pneumoniae*: Characterization of the raffinose utilization system. *Genome Res.* **9**(12): 1189–97.
- Roy, A., Kucukural, A., Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**(4): 725–38.
- Russell, R.R.B., 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.
- Sá-Nogueira, I., Nogueira, T.V., Soares, S., de Lencastre, H. (1997). The *Bacillus subtilis* L-arabinose (*ara*) operon: nucleotide sequence, genetic organization and expression. *Microbiology.* **143**: 957–69.
- Sá-Nogueira, I., Mota, L.J. (1997). Negative regulation of L-arabinose metabolism in *Bacillus subtilis*: Characterization of the *araR* (*araC*) gene. *J. Bacteriol.* **179**(5): 1598–608.
- Saier, M.H., Goldman, S.R., Maile, R.R., Moreno, M.S., Weyler, W., Yang, N., *et al.* (2002). Transport capabilities encoded within the *Bacillus subtilis* genome. *J. Mol. Microbiol. Biotechnol.* **4**(1): 37–67.
- Saito, A., Shinya, T., Miyamoto, K., Yokoyama, T., Kaku, H., Minami, E., *et al.* (2007). The *dasABC* gene cluster, adjacent to *dasR*, encodes a novel ABC transporter for the uptake of N,N'-diacetylchitobiose in *Streptomyces coelicolor* A3(2). *Appl Environ Microbiol.* **73**(9): 3000–8.
- 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**(11): 3358–65.
- Salis, H.M., Mirsky, E. A., Voigt, C. A. (2009). Automated Design of Synthetic Ribosome Binding Sites to Precisely Control Protein Expression. *Nat. Biotechnol.* **27**(10): 946–50.
- Salis, H.M. (2011). The ribosome binding site calculator. *Methods Enzymol.* **498**: 19–42.
- Sambrook, J., Frisch, E.F. and Maniatis, T. (1989). *Molecular Cloning: A laboratory manual.*, 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

- Sauna, Z.E., Kim, I.W., Nandigama, K., Kopp, S., Chiba, P., Ambudkar, S.V.** (2007). Catalytic cycle of ATP hydrolysis by P-glycoprotein: evidence for formation of the E.S reaction intermediate with ATP-gamma-S, a nonhydrolyzable analogue of ATP. *Biochemistry*. **46**(48): 13787–99.
- Schlösser, A.** (2000). MsiK-dependent trehalose uptake in *Streptomyces reticuli*. *FEMS Microbiol. Lett.* **184**(2): 187–92.
- Schlösser, A., Kampers, T., Schrempf, H.** (1997). The *Streptomyces* ATP-binding component MsiK assists in cellobiose and maltose transport. *J. Bacteriol.* **179**(6): 2092–5.
- Schneider, E., Hunke, S.** (1998). ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolysing subunits/domains. *FEMS Microbiol. Rev.* **22**(1): 1-20.
- Schönert, S., Seitz, S., Krafft, H., Feuerbaum, E.A., Andernach, I., Witz, G., et al.** (2006). Maltose and maltodextrin utilization by *Bacillus subtilis*. *J. Bacteriol.* **188**(11): 3911–22.
- Siarheyeva, A., Liu, R., Sharom, F.J.** (2010). Characterization of an asymmetric occluded state of P-glycoprotein with two bound nucleotides: Implications for catalysis. *J. Biol. Chem.* **285**(10): 7575–86.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., et al.** (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**(539).
- Silva, Z., Sampaio, M., Henne, A., Gutzat, R., Boos, W., Costa, M.S., et al.** (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.
- Sutcliffe, I.C., Russell, R.R.** (1995). Lipoproteins of gram-positive bacteria. *J. Bacteriol.* **177**(5): 1123–8.
- Tan, M.F., Gao, T., Liu, W.Q., Zhang, C.Y., Yang, X., Zhu, J.W., et al.** (2015). MsmK, an ATPase, contributes to utilization of multiple carbohydrates and host colonization of *Streptococcus suis*. *PLoS One.* **10**(7): 1–19.
- Tanford, C.** (1982). Simple model for the chemical potential change of a transported ion in active transport. *Proc. Natl. Acad. Sci. USA.* **79**(9): 2882–4.

Tao, L., Sutcliffe, I.C., Russell, R.R., Ferretti, J.J. (1993). Transport of sugars, including sucrose, by the msm transport system of *Streptococcus mutans*. *J. Dent. Res.* **72**(10): 1386–1390.

Tchieu, J.H., Norris, V., Edwards, J.S., Saier, M.H. (2001). The complete phosphotransferase system in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.* **3**(3): 329–46.

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

Trappetti, C., Kadioglu, A., Carter, M., Hayre, J., Iannelli, F., Pozzi, G., et al. (2009). Sialic acid: a preventable signal for pneumococcal biofilm formation, colonization, and invasion of the host. *J. Infect. Dis.* **199**(10): 1497–505.

Tyx, R.E., Roche-Hakansson, H., Hakansson, A.P. (2011). Role of dihydrolipoamide dehydrogenase in regulation of raffinose transport in *Streptococcus pneumoniae*. *J. Bacteriol.* **193**(14): 3512–24.

Viens, P., Dubeau, M.P., Kimura, A., Desaki, Y., Shinya, T., Shibuya, N., et al. (2015). Uptake of chitosan-derived D-glucosamine oligosaccharides in *Streptomyces coelicolor* A3(2). *FEMS Microbiol. Lett.* **362**(9): 1–9.

Wang, T., Fu, G., Pan, X., Wu, J., Gong, X., Wang, J., et al. (2013). Structure of a bacterial energy-coupling factor transporter. *Nature.* **497**(7448): 272–6.

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.

Wilkens, S. (2015). Structure and mechanism of ABC transporters. *F1000Prime Rep.* **7**(14).

Xu, K., Zhang, M., Zhao, Q., Yu, F., Guo, H., Wang, C., et al. (2013). Crystal structure of a folate energy-coupling factor transporter from *Lactobacillus brevis*. *Nature.* **497**(7448): 268–71.

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

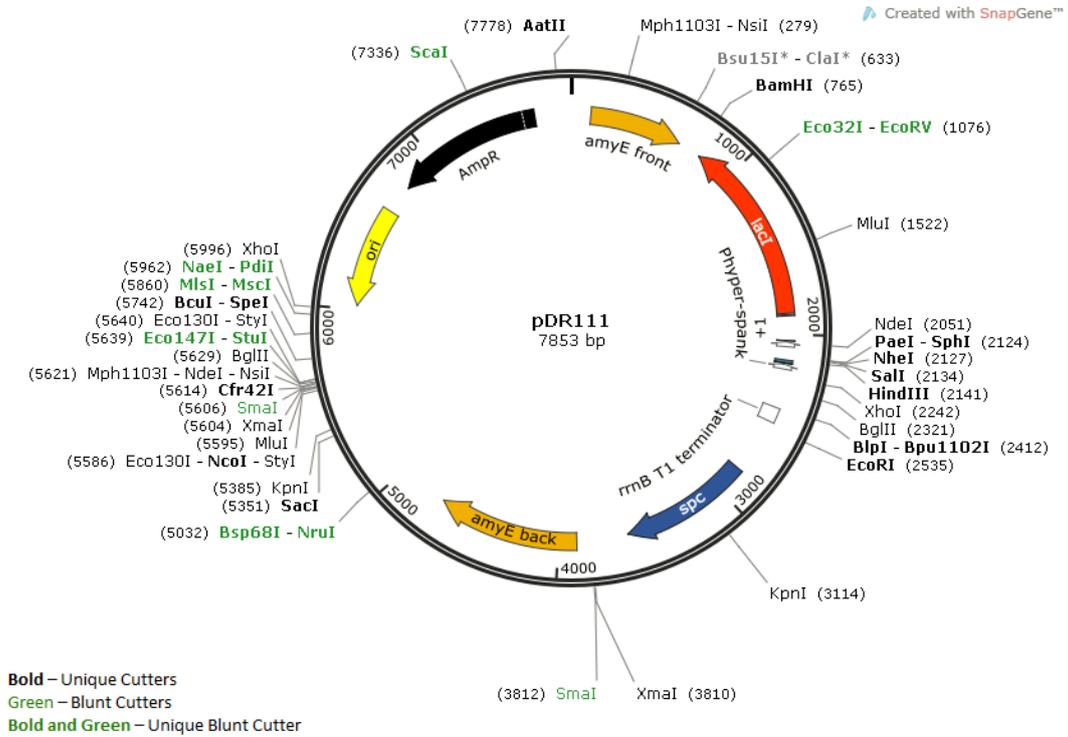
Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics.* **9**(40).

Chapter 6

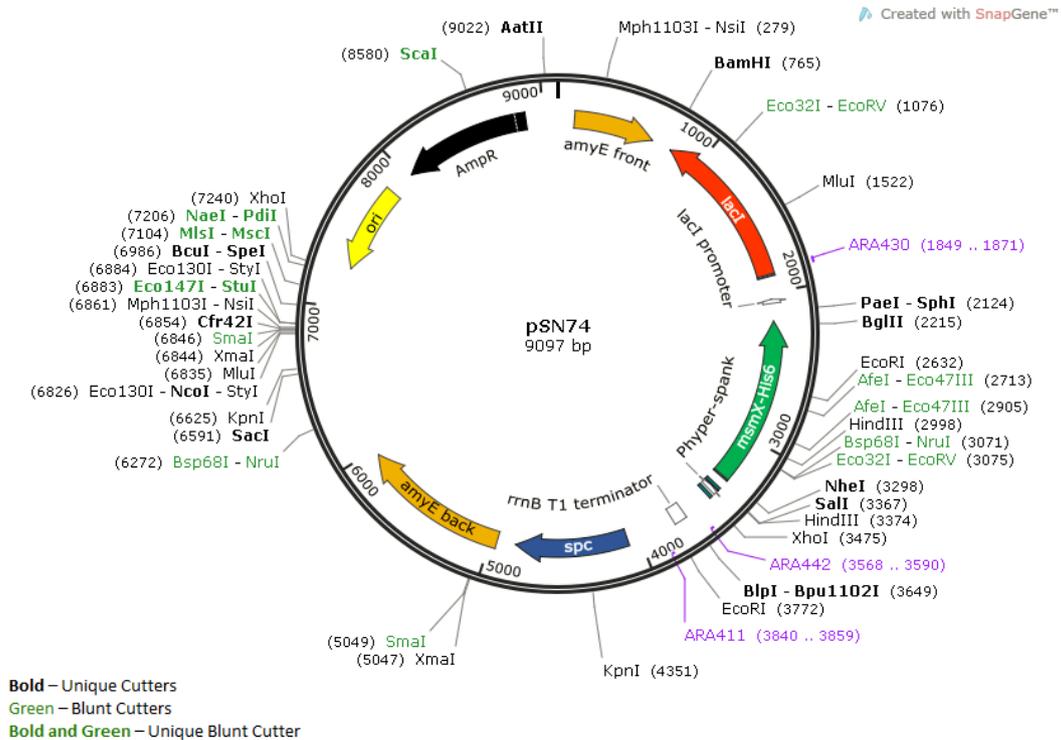
Appendices

6. Appendices

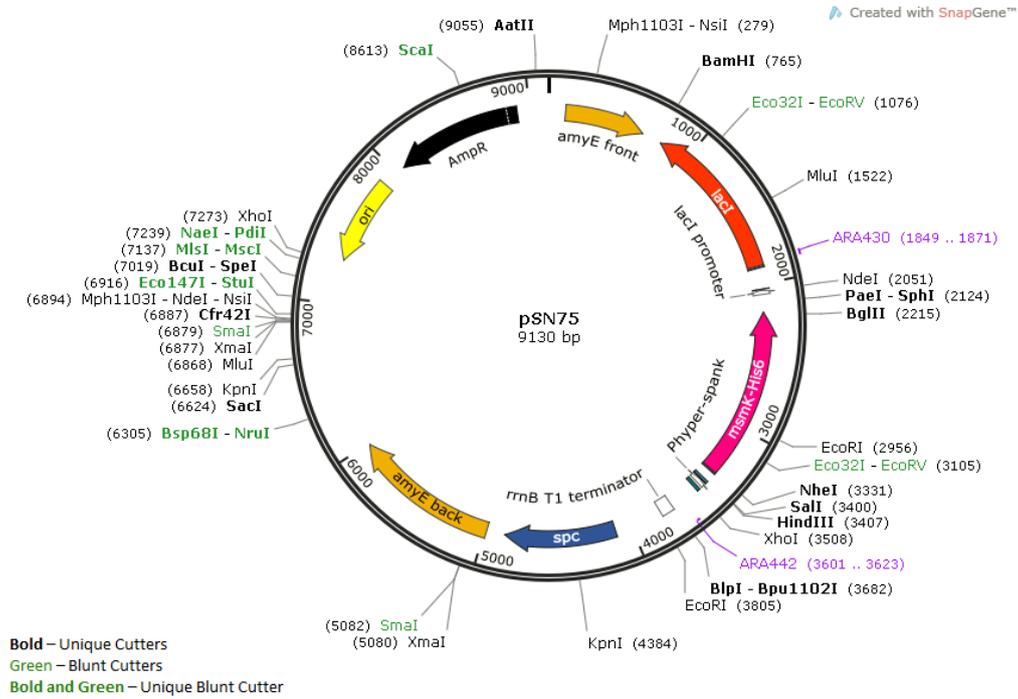
Appendix 6.1 – Map of pDR111 (gift from David Rudner, Harvard University).



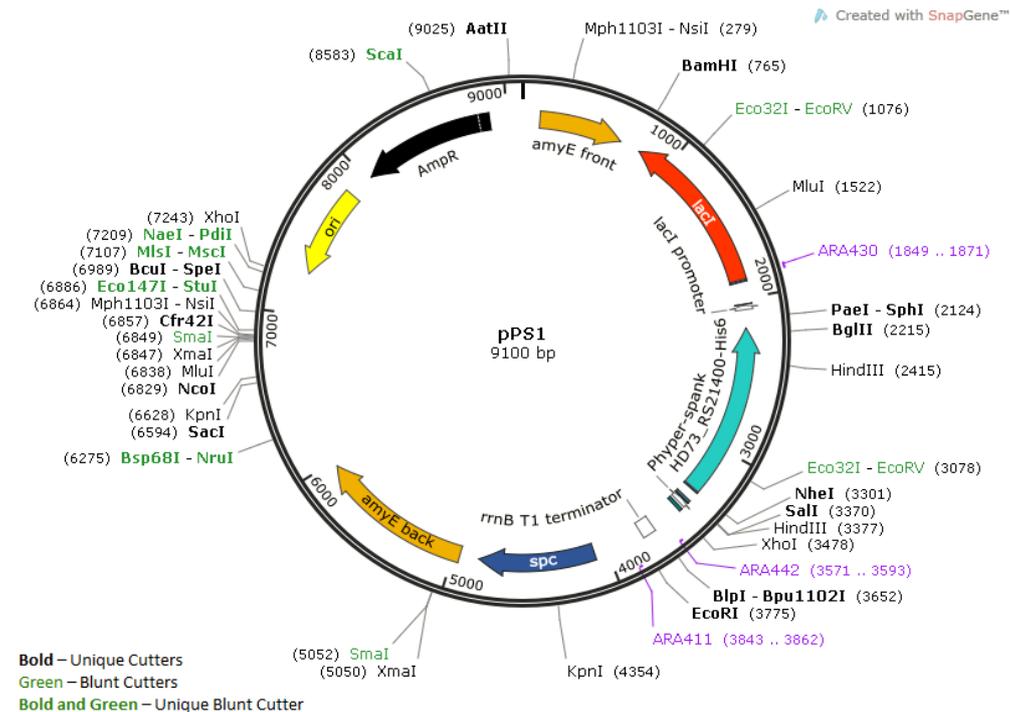
Appendix 6.2 – Map of pSN74 (Sá-Nogueira, unpublished). Primers ARA411, ARA442 and ARA430 were used for DNA sequencing.



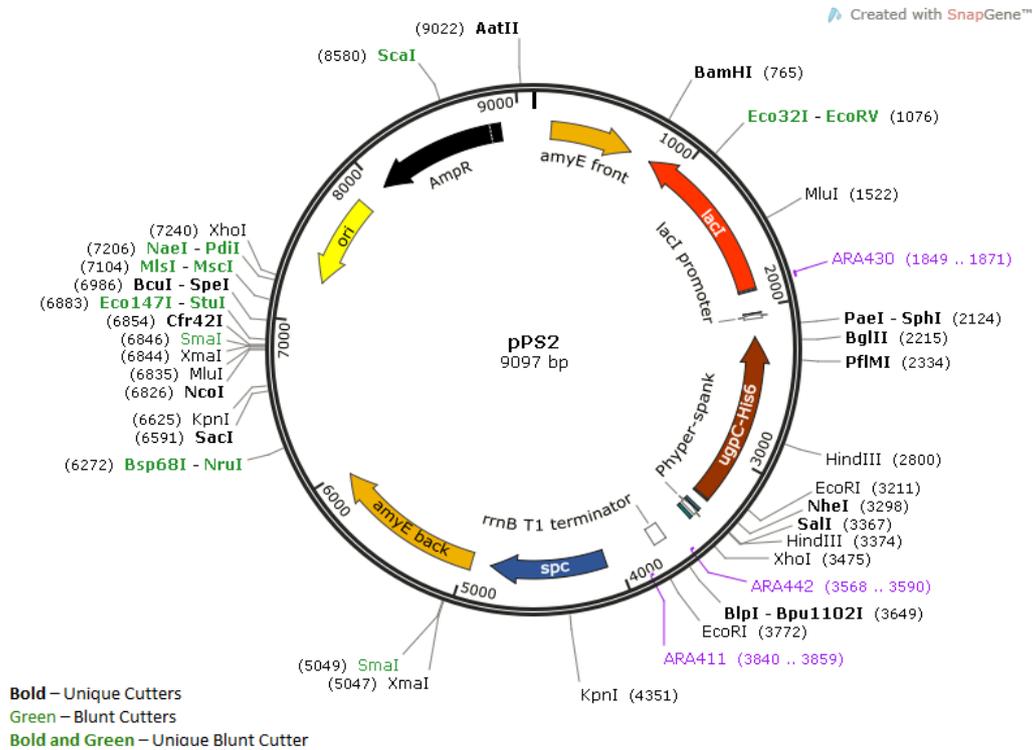
Appendix 6.3 – Map of pSN75. Insertion of a 1112bp fragment containing *msmK* gene from pAM7 (amplified with primers ARA773 and ARA774, and digested with *NheI* and *BglII*) between the *NheI* and *BglII* sites of pSN74 (Sá-Nogueira, unpublished). Primers ARA442 and ARA430 were used for DNA sequencing.



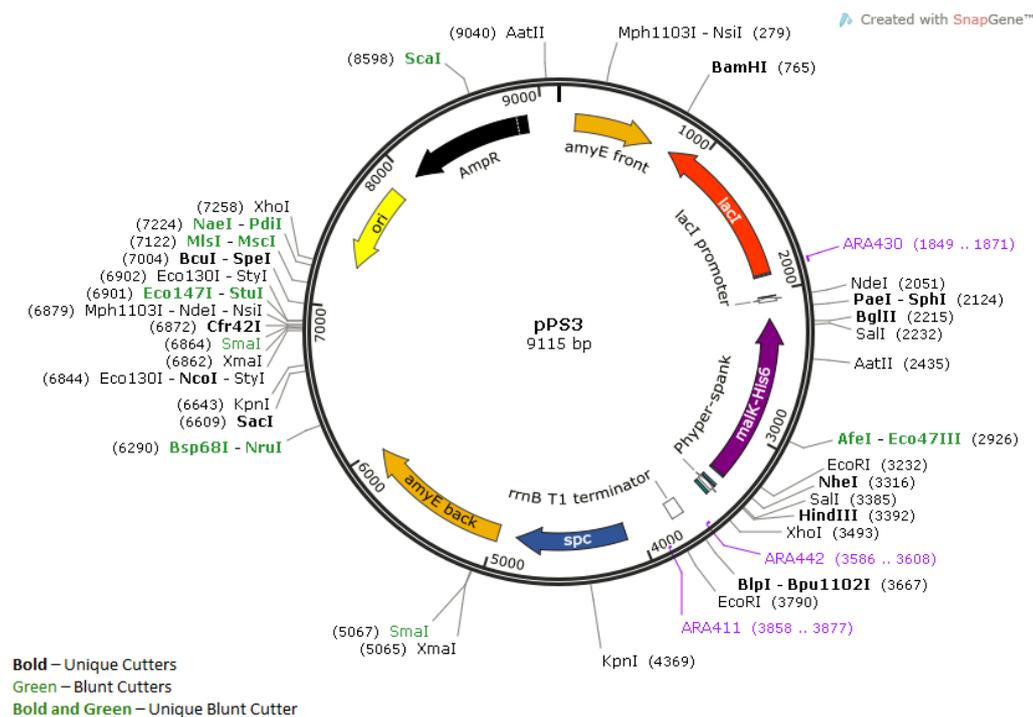
Appendix 6.4 – Map of pPS1. Insertion of a 1082bp fragment containing *HD73_RS21400* gene, from *B. thuringiensis* serovar *kurstaki* str. HD73 (amplified with primers ARA851 and ARA852, and digested with *NheI* and *BglII*) between the *NheI* and *BglII* sites of pSN74. Primers ARA411, ARA442 and ARA430 were used for DNA sequencing.



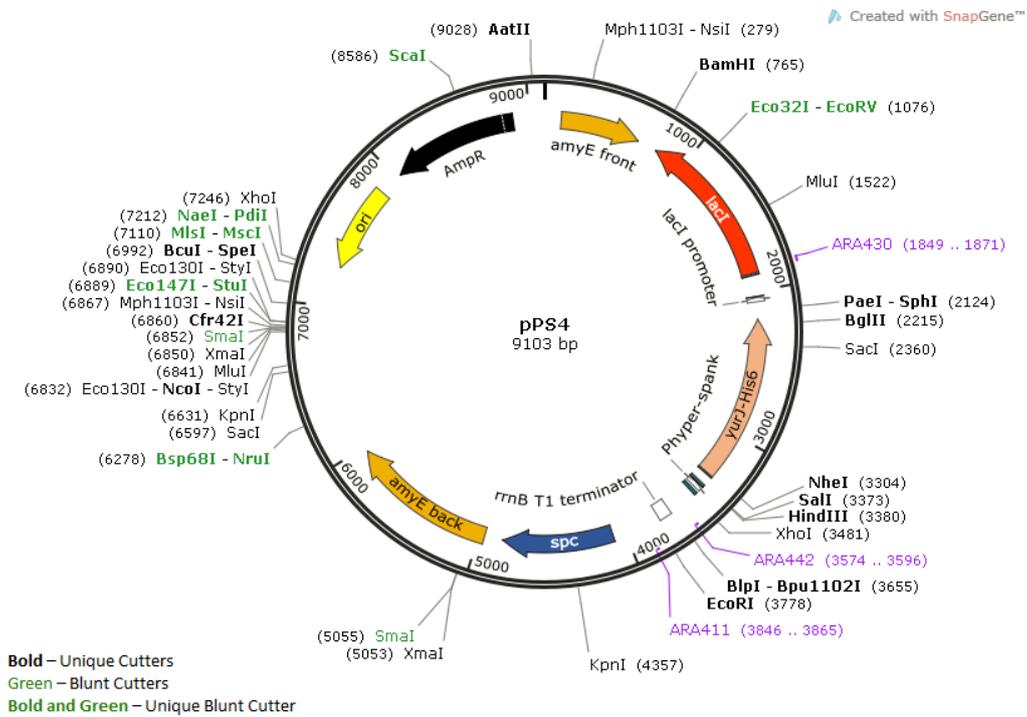
Appendix 6.5 – Map of pPS2. Insertion of a 1079bp fragment containing *ugpC* gene from *S. aureus* subsp. *aureus* ST398 (amplified with primers ARA843 and ARA844, and digested with *NheI* and *BglII*) between the *NheI* and *BglII* sites of pSN74. Primers ARA411, ARA442 and ARA430 were used for DNA sequencing.



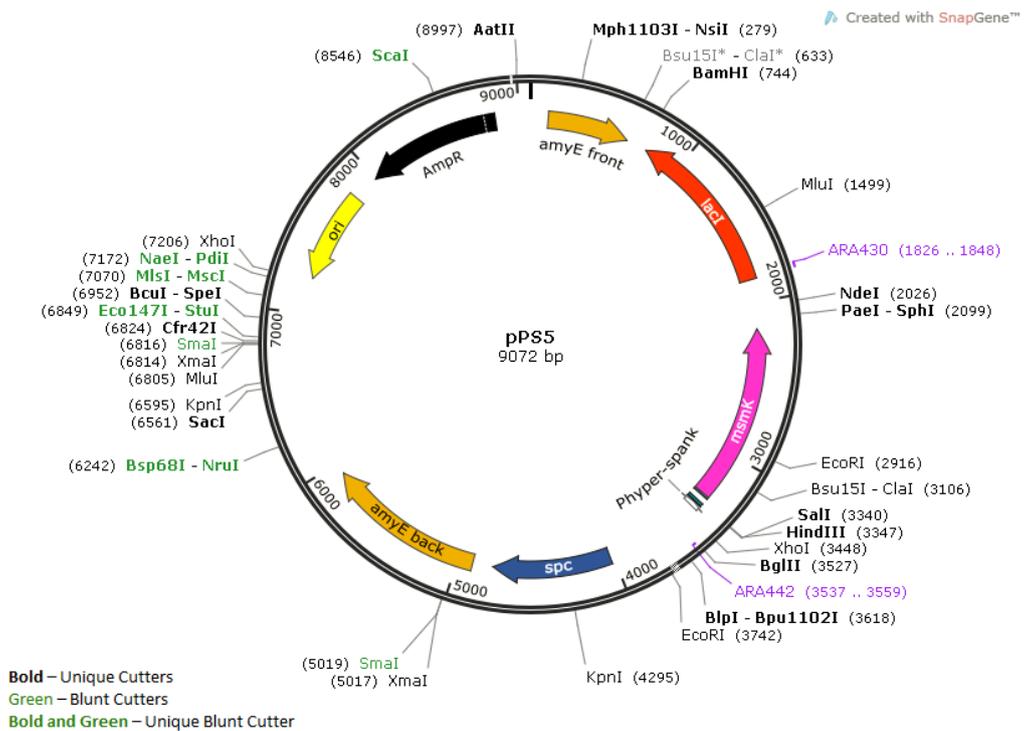
Appendix 6.6 – Map of pPS3. Insertion of a 1097bp fragment containing *malk* gene from *E. coli* K-12 (amplified with primers ARA847 and ARA848, and digested with *NheI* and *BglII*) between the *NheI* and *BglII* sites of pSN74. Primers ARA411, ARA442 and ARA430 were used for DNA sequencing.



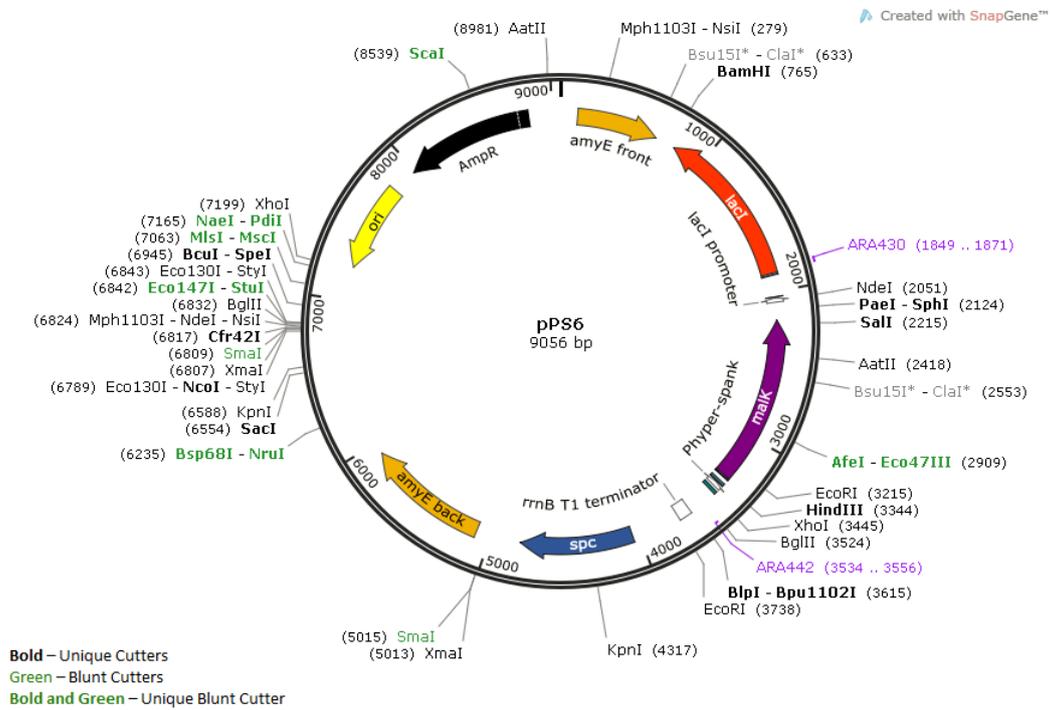
Appendix 6.7 – Map of pPS4. Insertion of a 1085bp fragment containing *yrjJ* gene from *B. subtilis* 168T⁺, (amplified with primers ARA837 and ARA838, and digested with *NheI* and *BglII*) between the *NheI* and *BglII* sites of pSN74. Primers ARA411, ARA442 and ARA430 were used for DNA sequencing.



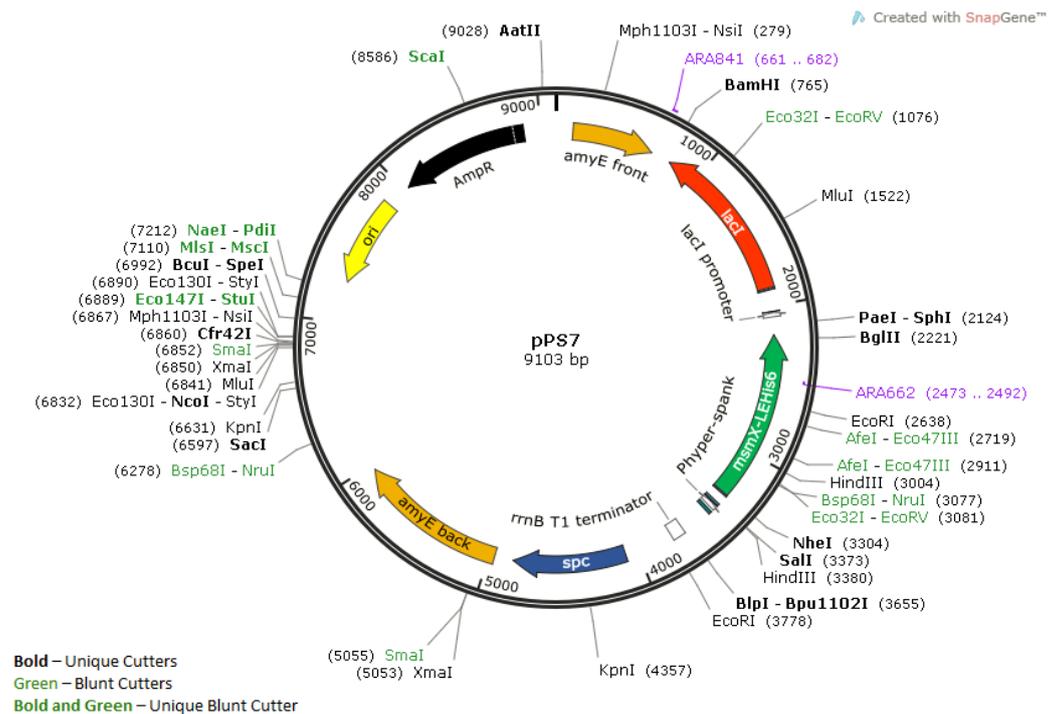
Appendix 6.8 – Map of pPS5. Insertion of a 1241bp fragment containing *msmK* gene (amplified from pAM7 with primers ARA855 and ARA749, and digested with *SalI* and *SphI*) between the *SalI* and *SphI* sites of pDR111. Primers ARA442 and ARA430 were used for DNA sequencing.



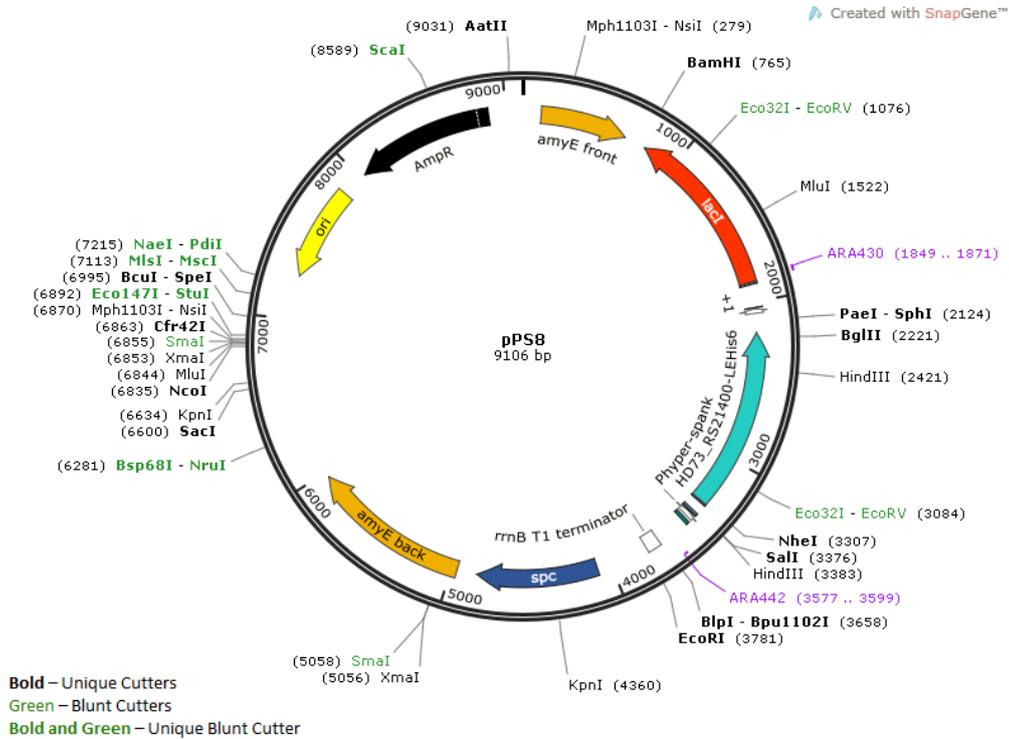
Appendix 6.9 – Map of pPS6. Insertion of a 1220bp fragment containing *malk* gene from *E. coli* K-12 (amplified with primers ARA858 and ARA859, and digested with *Hind*III and *Sph*I) between the *Hind*III and *Sph*I sites of pDR111. Primers ARA442 and ARA430 were used for DNA sequencing.



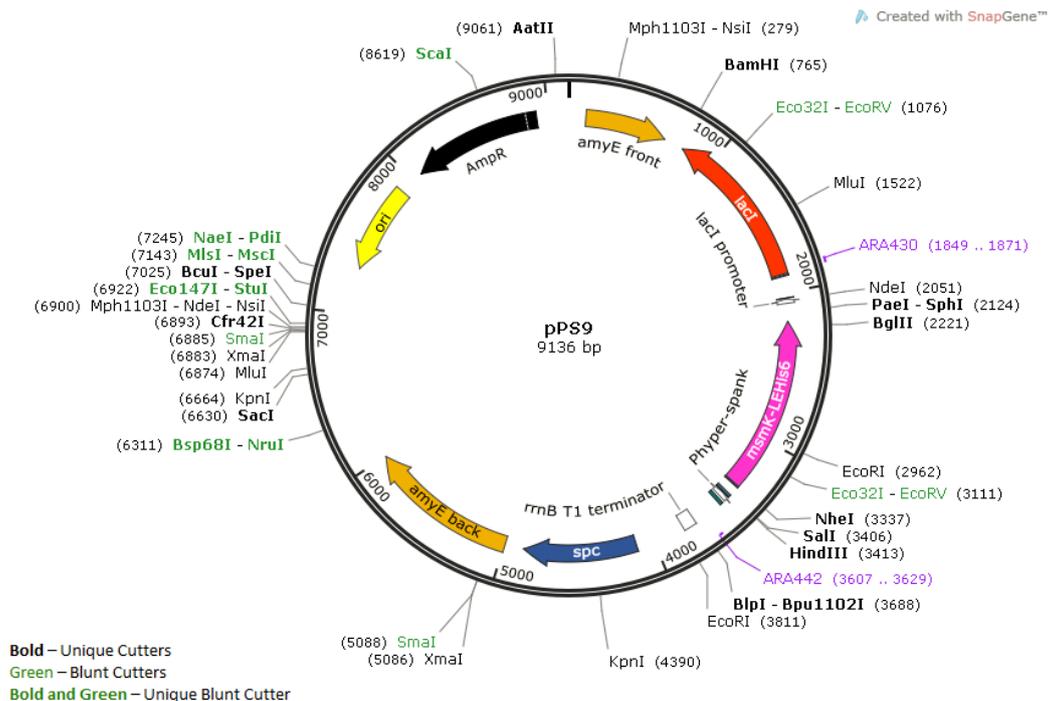
Appendix 6.10 – Map of pPS7. Insertion of a 1452bp fragment, containing *LacI* gene and the terminal part of *msmX* gene with two novel amino acids Leu and Glu followed by the C-terminal His-tag (amplified from pPS74 with primers ARA54 and ARA632, and digested with *Bgl*II and *Bam*HI), between the *Bgl*II and *Bam*HI sites of pSN74. Primers ARA662 and ARA841 were used for DNA sequencing.



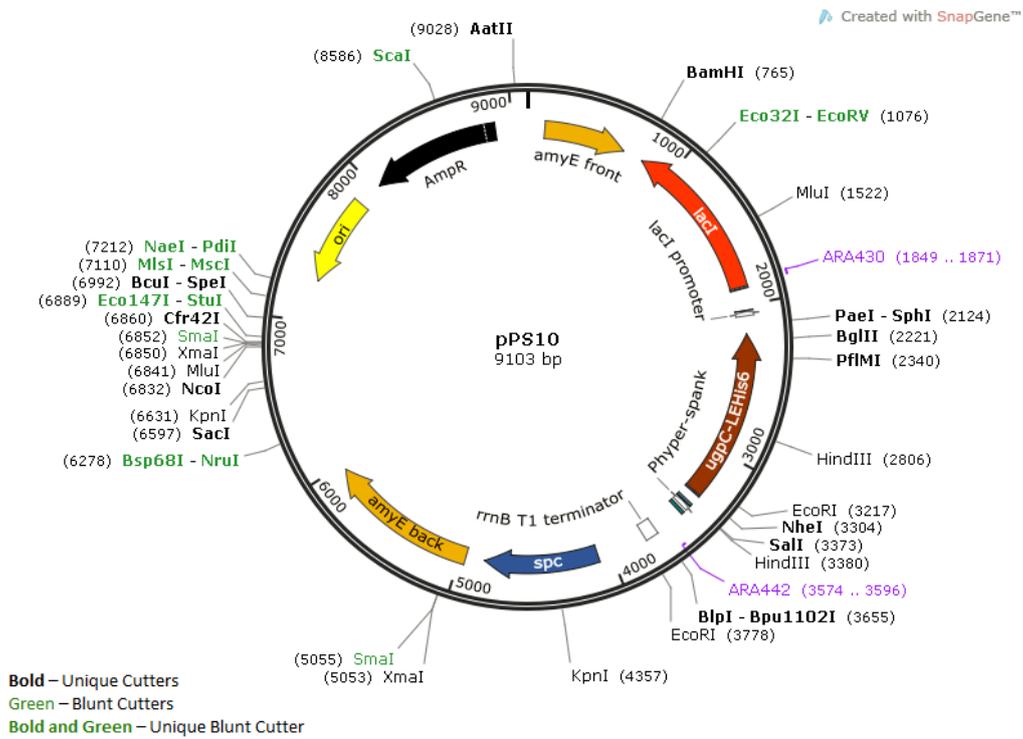
Appendix 6.11 – Map of pPS8. Insertion of a 1082bp fragment containing *HD73_RS21400* gene, from *B. thuringiensis* serovar *kurstaki* str. HD7 (amplified with primers ARA851 and ARA852, and digested with *NheI* and *BglII*) between the *NheI* and *BglII* sites of pPS7. Primers ARA442 and ARA430 were used for DNA sequencing.



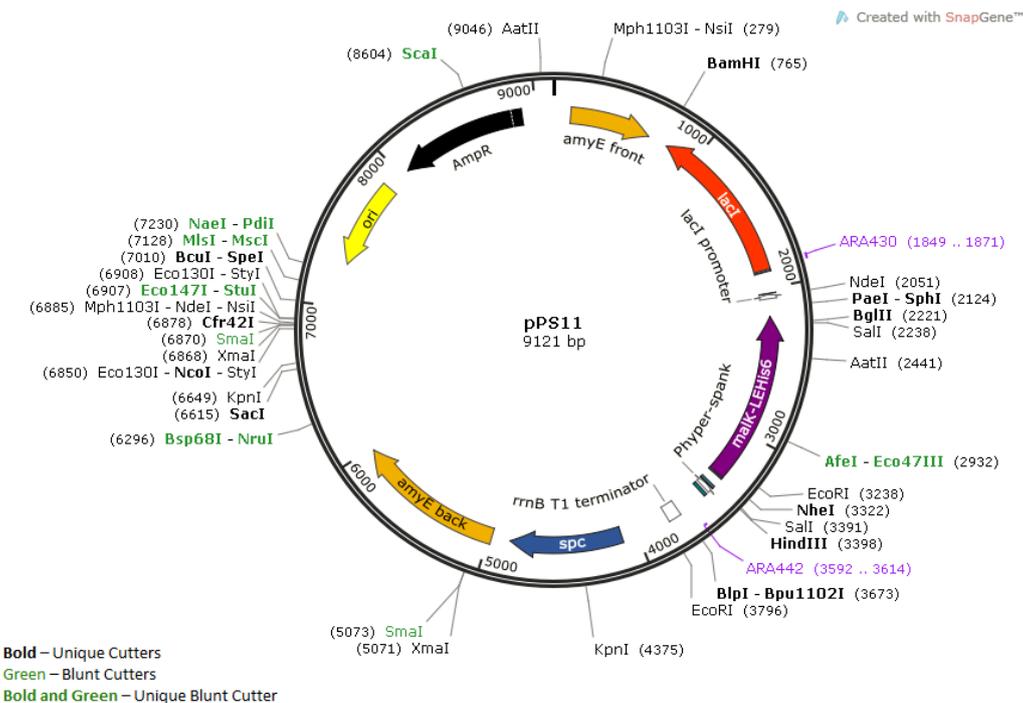
Appendix 6.12 – Map of pPS9. Insertion of a 1112bp fragment containing *msmK* gene from pAM7 (amplified with primers ARA773 and ARA774, and digested with *NheI* and *BglII*) between the *NheI* and *BglII* sites of pPS7. Primers ARA442 and ARA430 were used for DNA sequencing.



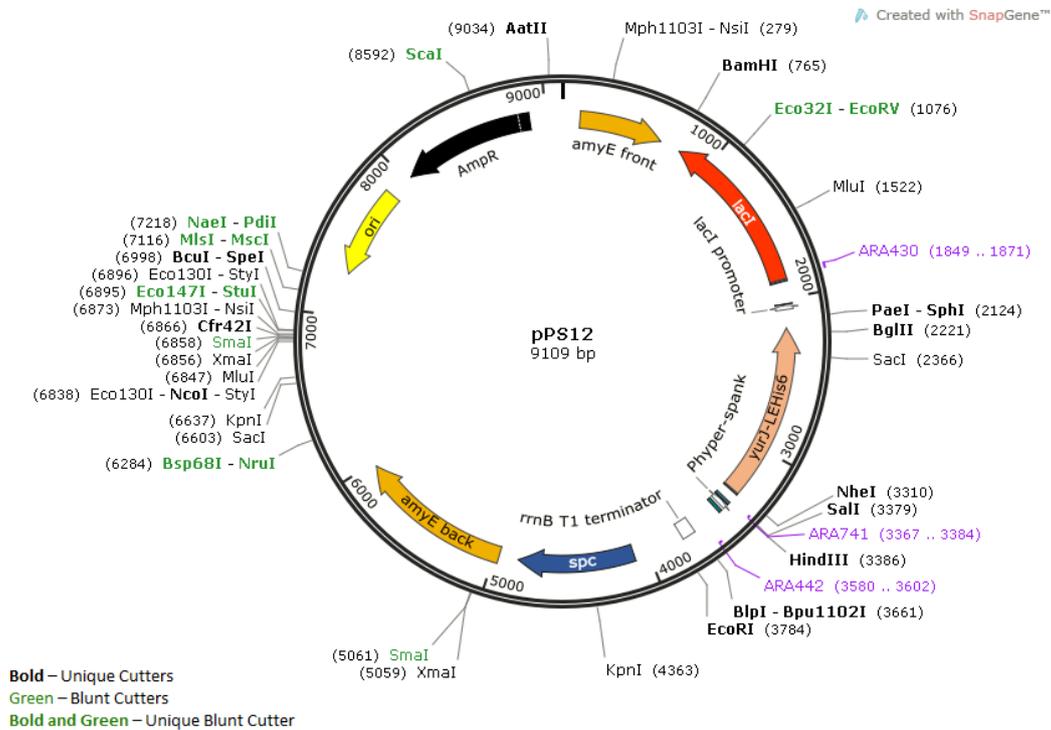
Appendix 6.13 – Map of pPS10. Insertion of a 1079bp fragment containing *ugpC* gene from *S. aureus* subsp. *aureus* ST398 (amplified with primers ARA843 and ARA844, and digested with *NheI* and *BglII*) between the *NheI* and *BglII* sites of pPS7. Primers ARA442 and ARA430 were used for DNA sequencing.



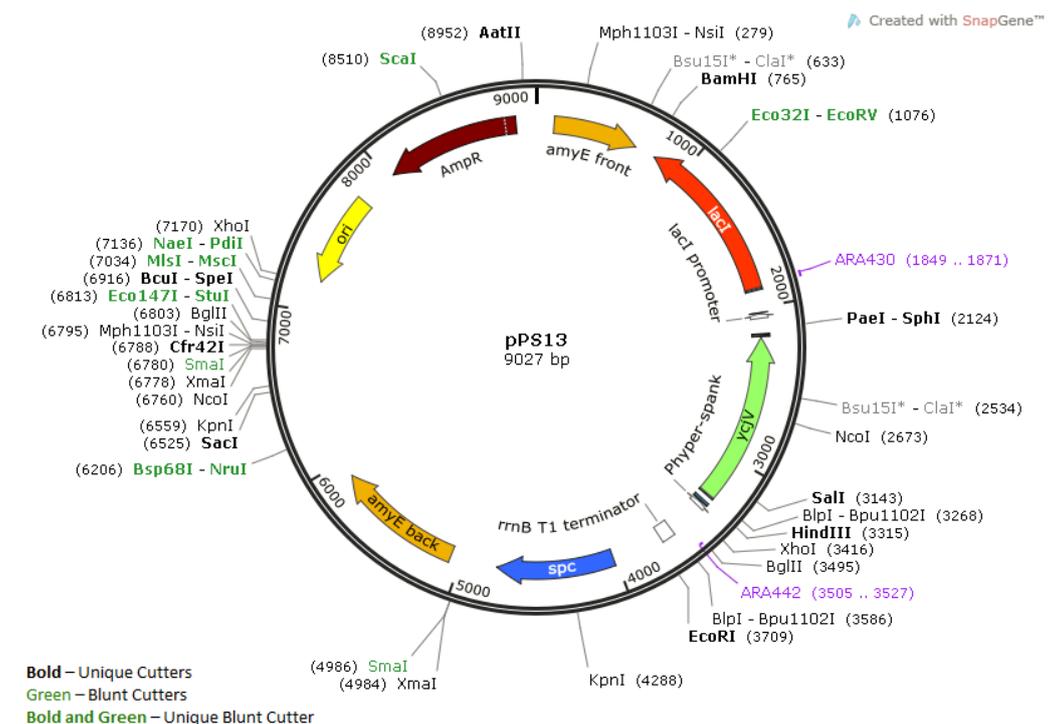
Appendix 6.14 – Map of pPS11. Insertion of a 1097bp fragment containing *malk* gene from *E. coli* K-12 (amplified with primers ARA847 and ARA848, and digested with *NheI* and *BglII*) between the *NheI* and *BglII* sites of pPS7. Primers ARA442 and ARA430 were used for DNA sequencing.



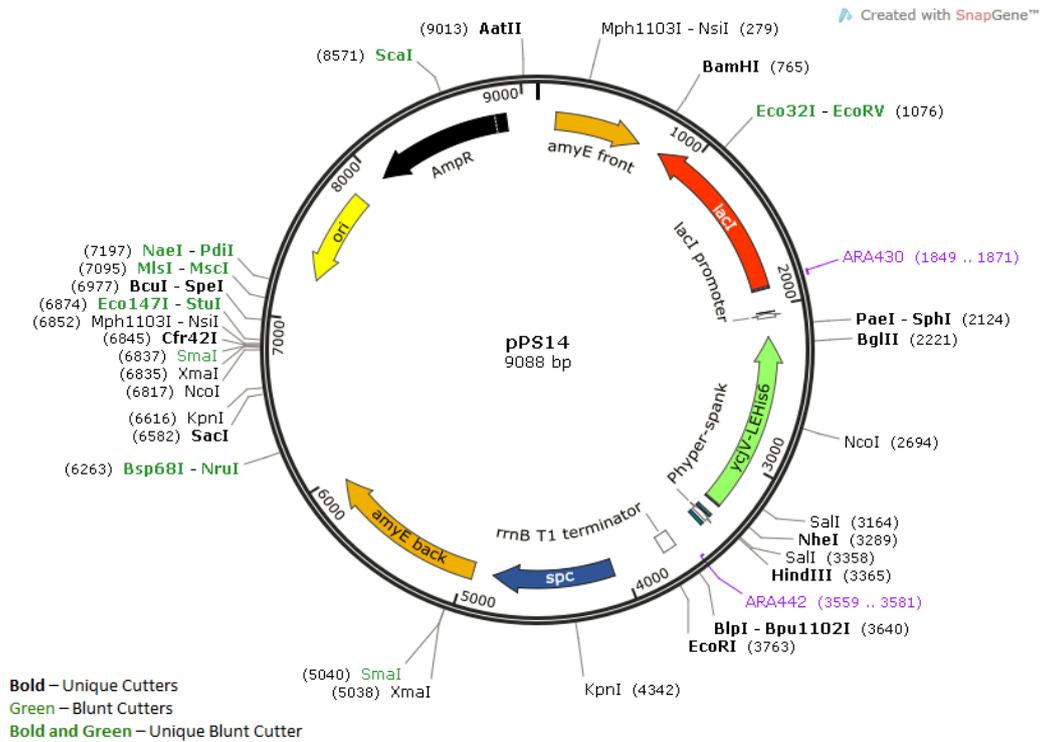
Appendix 6.15 – Map of pPS12. Insertion of a 1085bp fragment containing *yurJ* gene from *B. subtilis* 168T* (amplified with primers ARA837 and ARA838, and digested with *NheI* and *BglII*) between the *NheI* and *BglII* sites of pPS7. Primers ARA442, ARA741 and ARA430 were used for DNA sequencing.



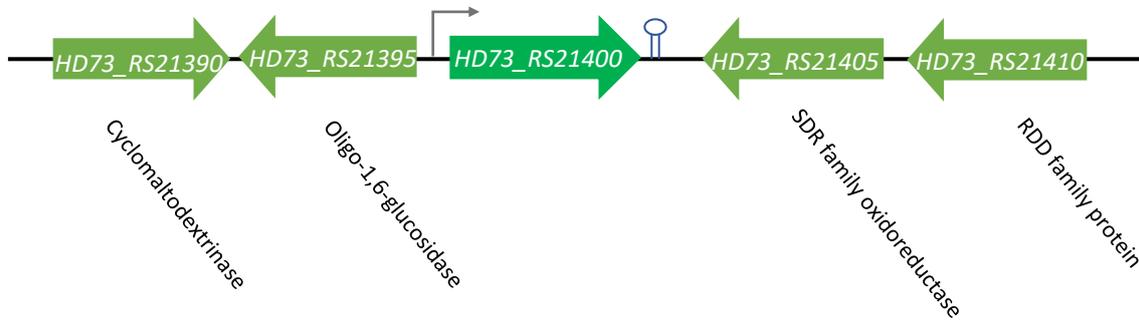
Appendix 6.16 – Map of pPS13. The fragment (1191bp) containing *ycjV* gene, obtained from *E. coli* K-12 through site-directed mutagenesis by overlapping PCR (amplified with the internal mutagenic primers ARA862 and ARA863, and the flanking primers ARA860 and ARA861), was digested with *HindIII* and *SphI* and subsequently cloned in pDR111 *HindIII-SphI*. Primers ARA442 and ARA430 were used for DNA sequencing.



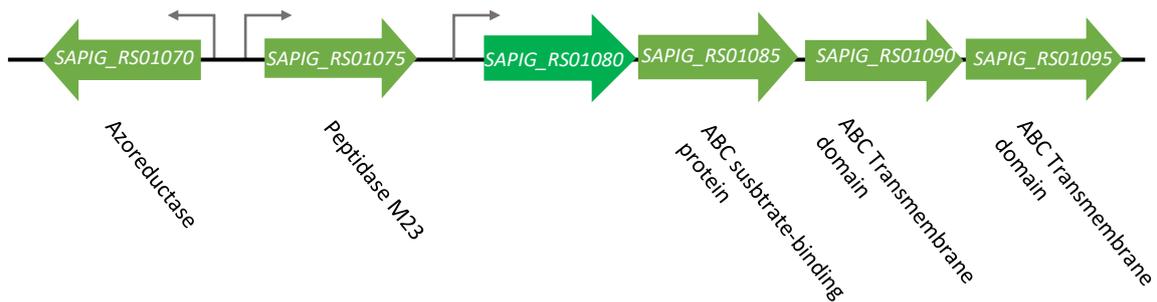
Appendix 6.17 – Map of pPS14. The fragment (1064bp) containing *ycjV* gene, obtained from *E. coli* K-12 through site-directed mutagenesis by overlapping PCR (amplified with the internal mutagenic primers ARA862 and ARA863, and the flanking primers ARA845 and ARA846), was digested with *NheI* and *BglIII* and subsequently cloned in pPS7 *NheI*-*BglIII*. Primers ARA442 and ARA430 were used for DNA sequencing.



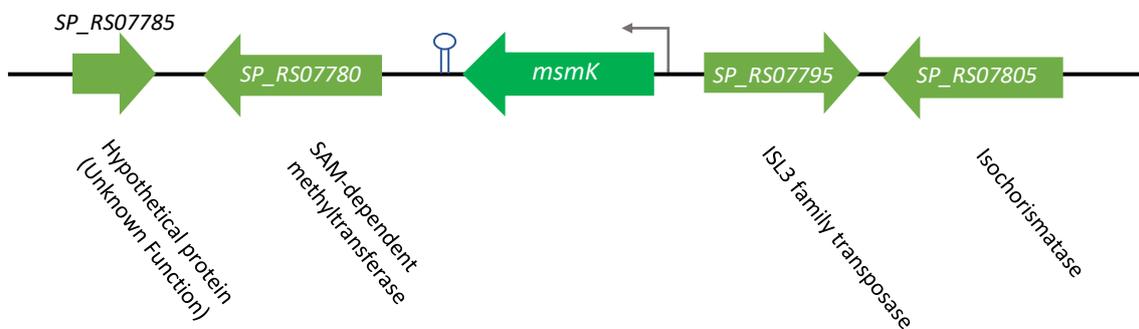
Appendix 6.20 – Genomic context of *HD73_RS21400* from *Bacillus thuringiensis*. The arrows denote the genes present in the *HD73_RS21400* gene region. Below each arrow is indicated the putative function of encoded protein based on BLAST results.



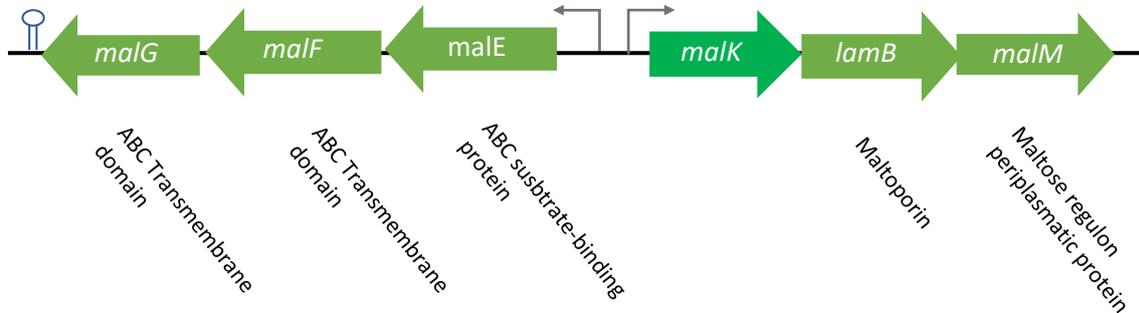
Appendix 6.21 – Genomic context of *SAPIG_RS01080* (or *ugpC*) from *Staphylococcus aureus*. The arrows denote the genes present in the *SAPIG_RS01080* gene region. Below each arrow is indicated the putative function of encoded protein based on BLAST results.



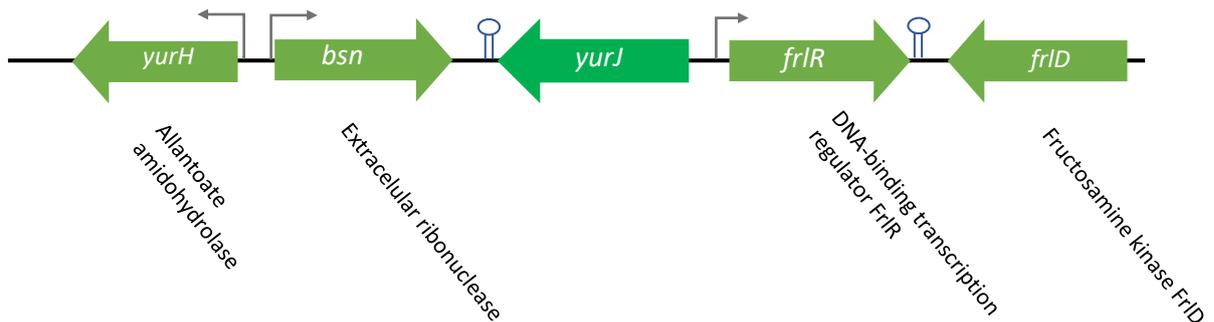
Appendix 6.22 – Genomic context of *msmK* from *Streptococcus pneumoniae*. The arrows denote the genes present in the *msmK* gene region. Below each arrow is indicated the putative function of encoded protein based on BLAST results.



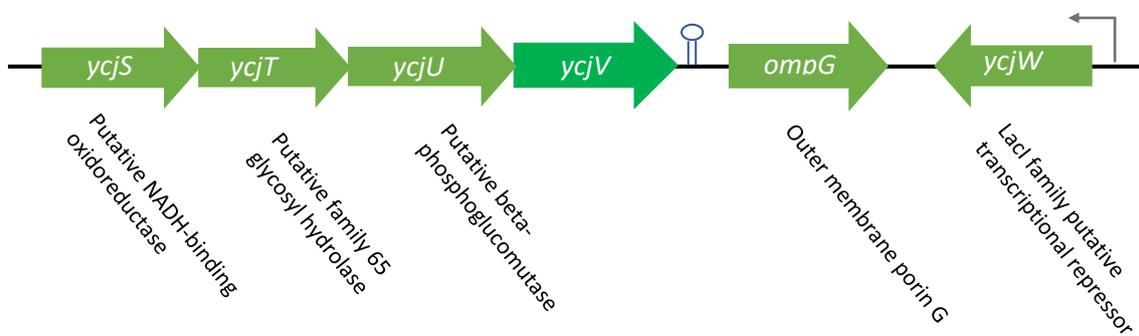
Appendix 6.23 – Genomic context of *malK* from *Escherichia coli*. The arrows denote the genes present in the *malK* gene region. Below each arrow is indicated the putative function of encoded protein based on BLAST results.



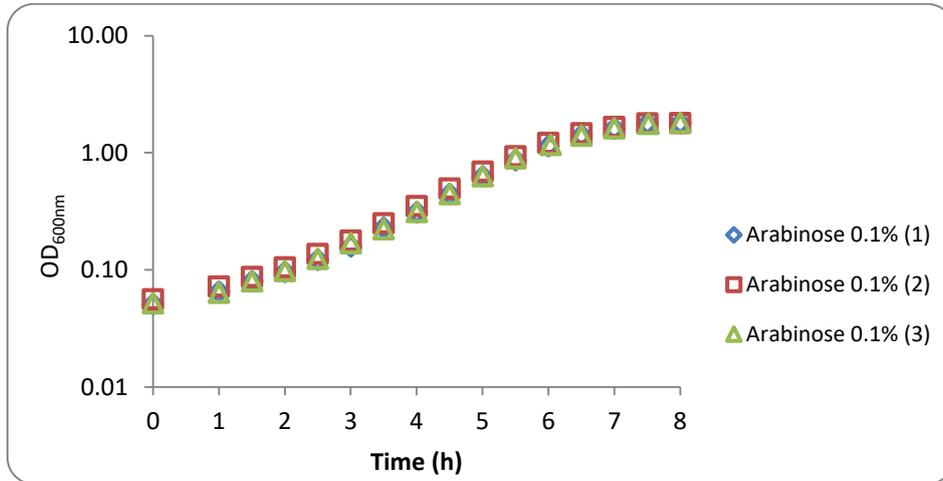
Appendix 6.24 – Genomic context of *yurJ* from *Bacillus subtilis*. The arrows denote the genes present in the *yurJ* gene region. Below each arrow is indicated the putative function of encoded protein based on BLAST results.



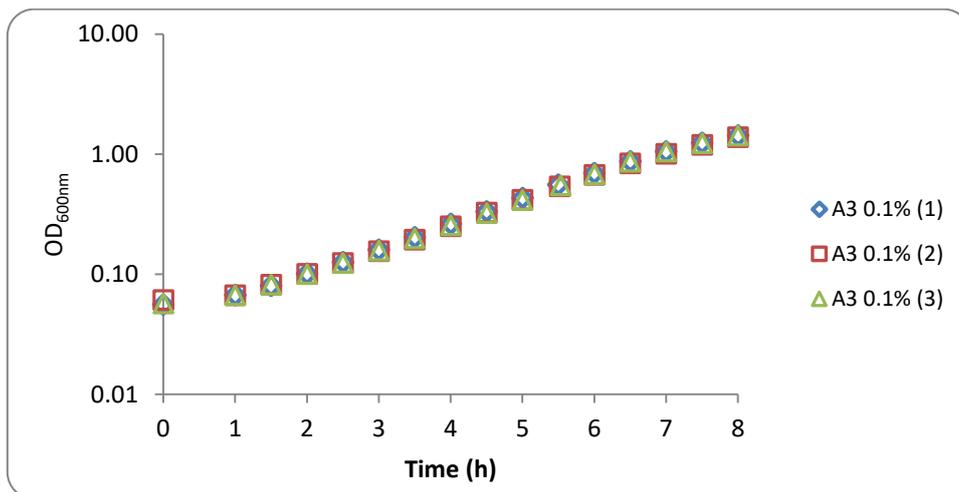
Appendix 6.25 – Genomic context of *ycjV* from *Escherichia coli*. The arrows denote the genes present in the *ycjV* gene region, which belongs in a large operon that is not displayed in full. Below each arrow is indicated the putative function of encoded protein based on BLAST results.



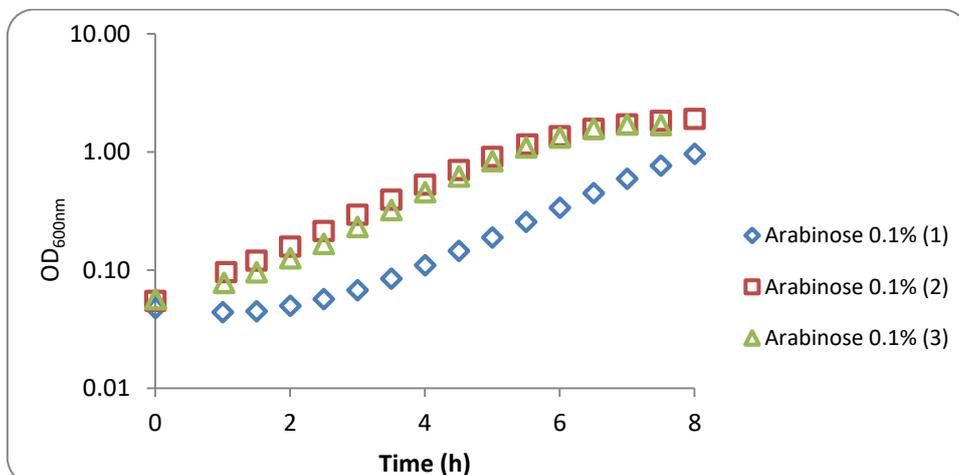
Appendix 6.26 - Growth assays of *B.subtilis* ISN1 ($\Delta amyE::Pspank(hy)$ -spec) using arabinose 0.1% (w/v). OD_{600nm} plotted vs. time(h).



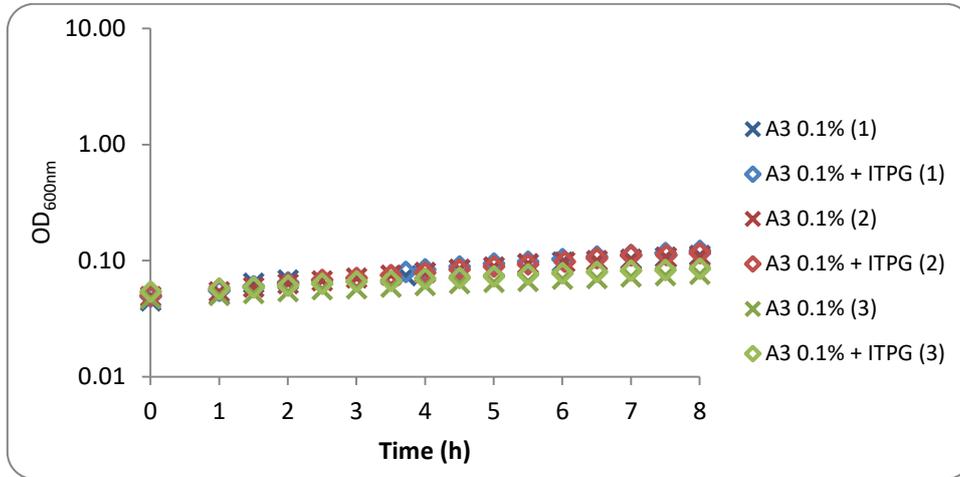
Appendix 6.27 - Growth assays of *B.subtilis* ISN1 ($\Delta amyE::Pspank(hy)$ -spec) using arabinotriose (A3) 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



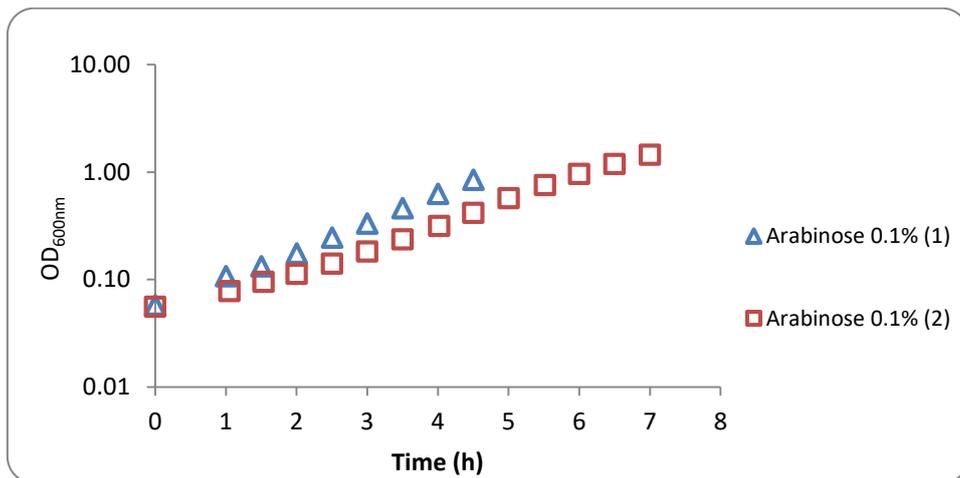
Appendix 6.28 - Growth assays of *B.subtilis* IQB672 ($\Delta msmX::cat$ $\Delta amyE::Pspank(hy)$) using arabinose 0.1% (w/v). OD_{600nm} plotted vs. time(h).



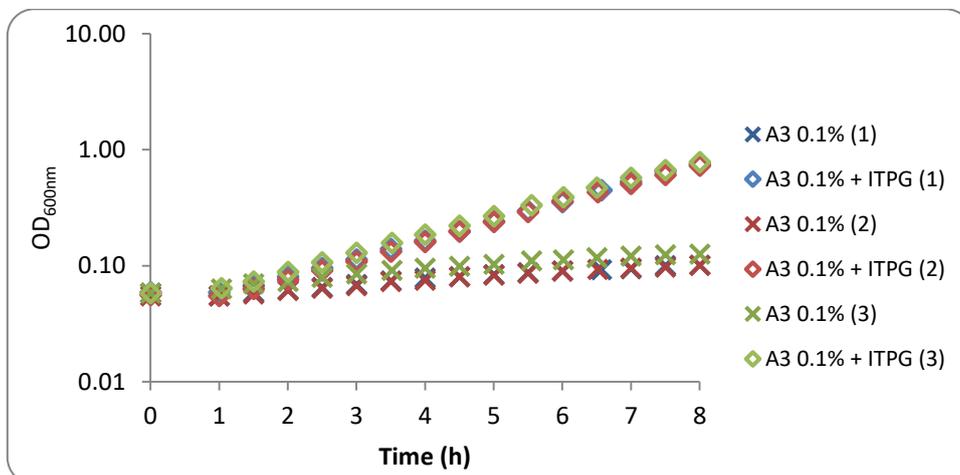
Appendix 6.29 - Growth assays of *B. subtilis* IQB672 (Δ msmX::cat Δ AmyE::Pspank(hy)) using arabinose 0.1% (w/v). OD_{600nm} plotted vs. time(h).



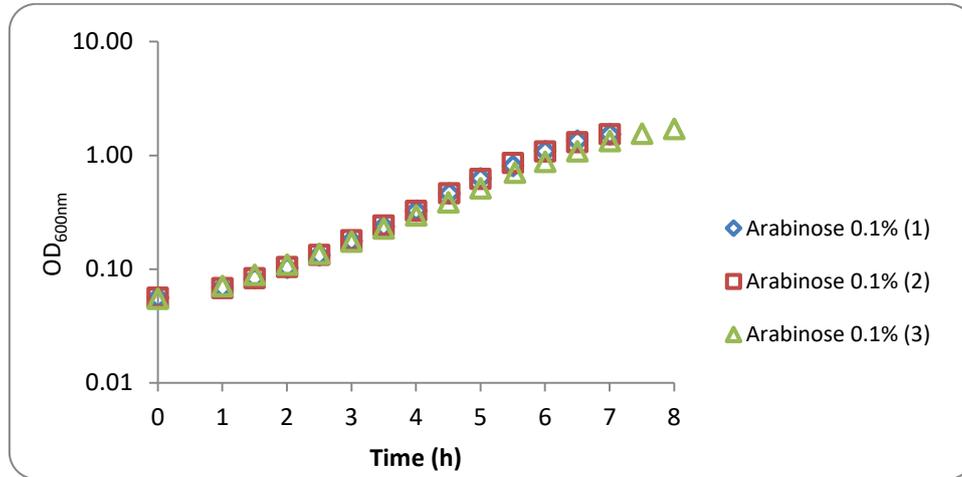
Appendix 6.30 - Growth assays of *B. subtilis* IQB676 (Δ msmX::cat Δ AmyE::Pspank(hy)-msmX(Glu3Ser, Ile364Ser)) using arabinose 0.1% (w/v). OD_{600nm} plotted vs. time(h).



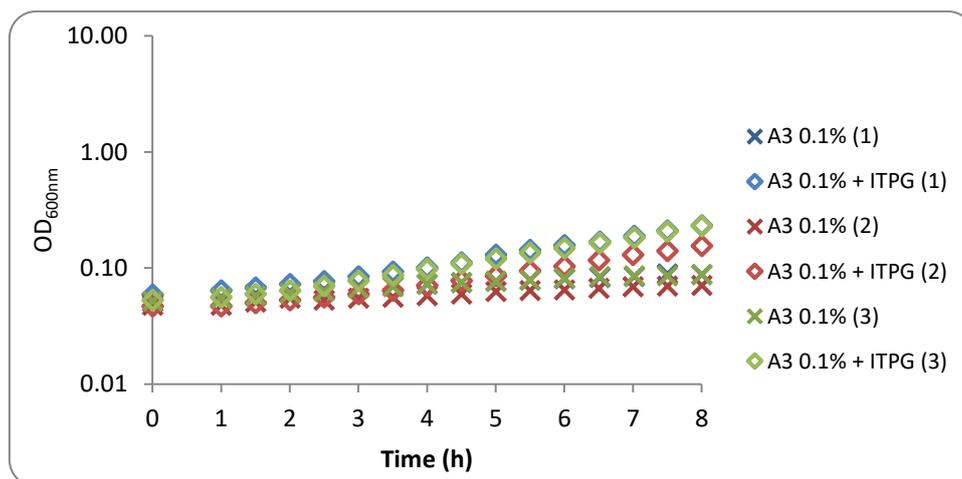
Appendix 6.31 - Growth assays of *B. subtilis* IQB676 (Δ msmX::cat Δ AmyE::Pspank(hy)-msmX(Glu3Ser, Ile364Ser)) using arabinotriose (A3) 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



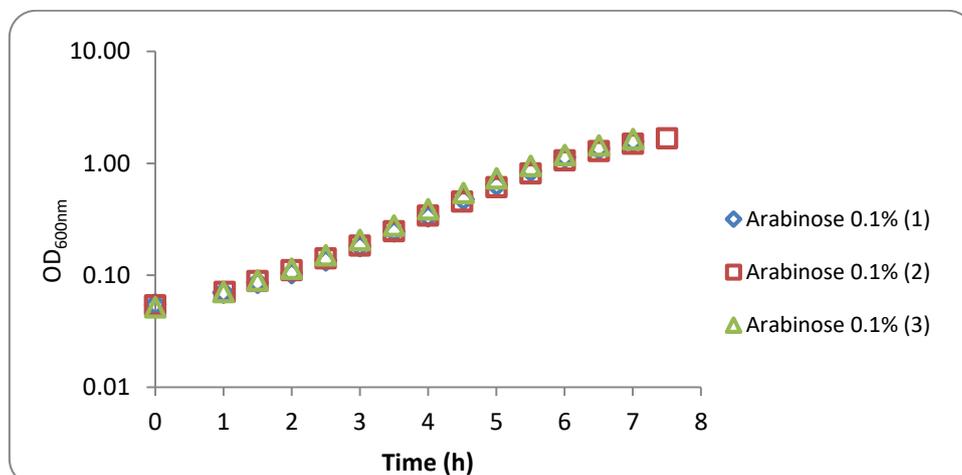
Appendix 6.32 - Growth assays of *B. subtilis* ISN2 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmX-His_6$) using arabinose 0.1% (w/v). OD_{600nm} plotted vs. time(h).



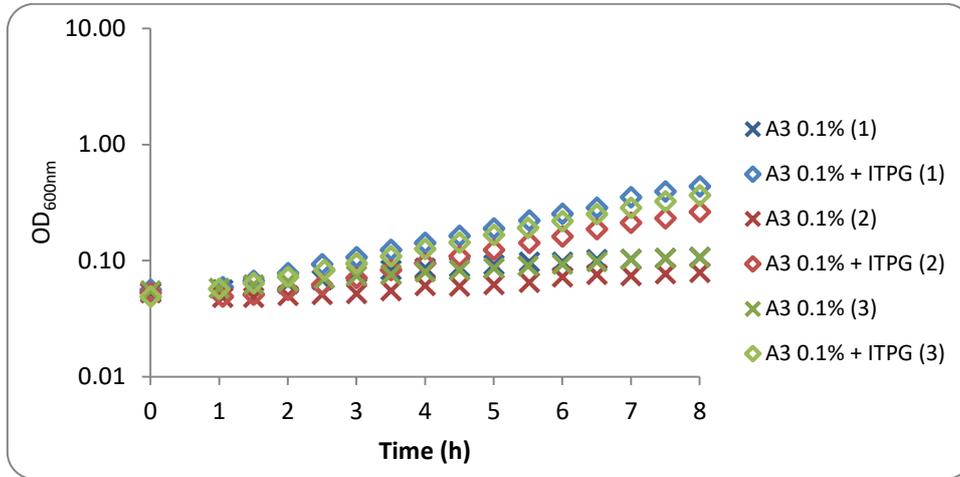
Appendix 6.33 - Growth assays of *B. subtilis* ISN2 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmX-His_6$) using arabinotriose (A3) 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



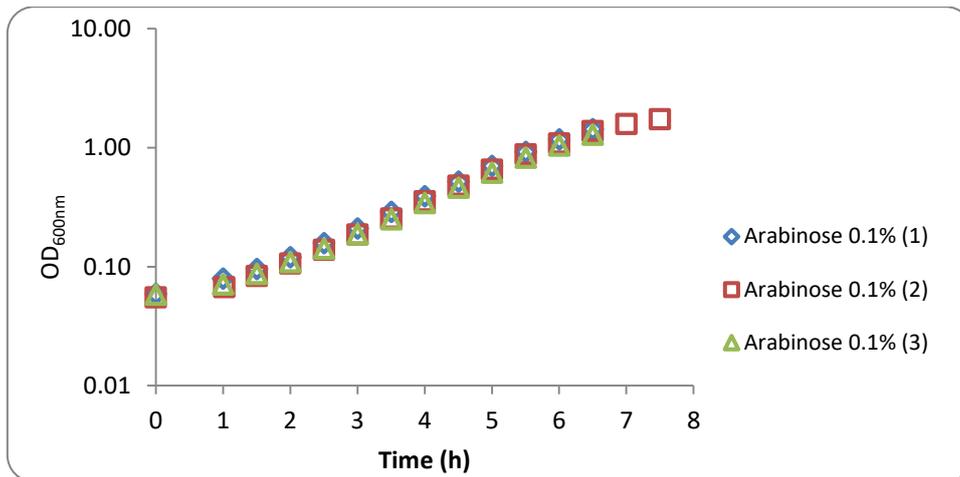
Appendix 6.34 - Growth assays of *B. subtilis* ISN3 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmK-His_6$) using arabinose 0.1% (w/v). OD_{600nm} plotted vs. time(h).



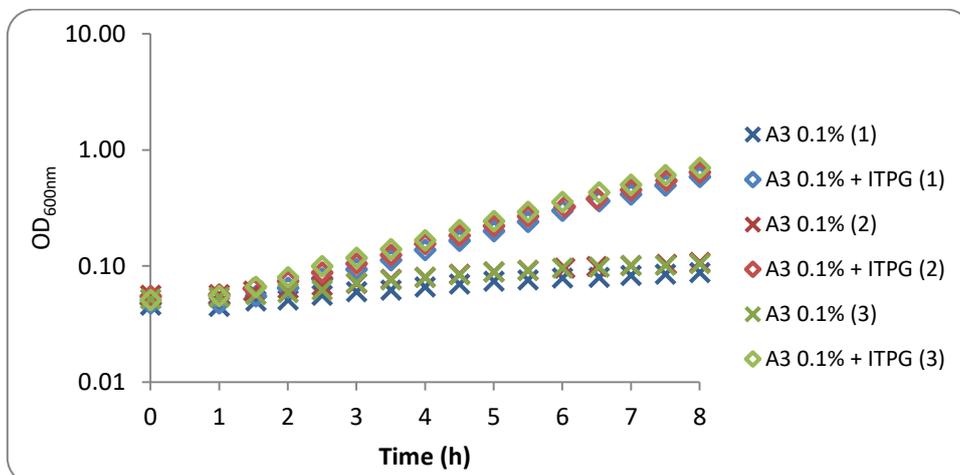
Appendix 6.35 - Growth assays of *B. subtilis* ISN3 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmK-His_6$) using arabinotriose (A3) 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



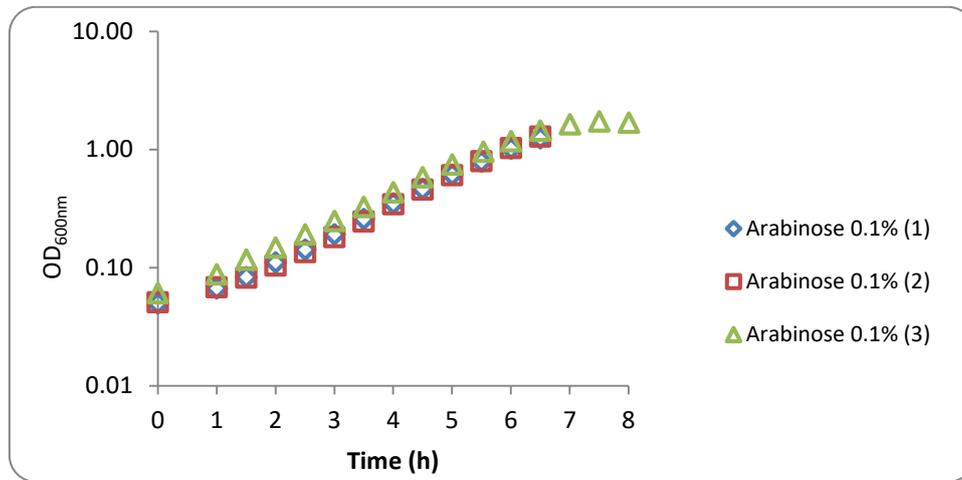
Appendix 6.36 - Growth assays of *B. subtilis* ISN4 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-HD73_RS21400-His_6$) using arabinose 0.1% (w/v). OD_{600nm} plotted vs. time(h).



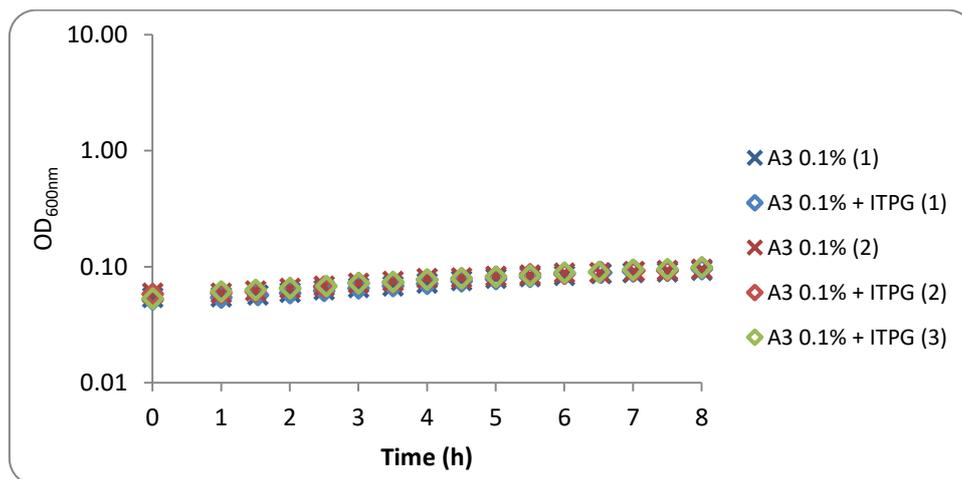
Appendix 6.37 - Growth assays of *B. subtilis* ISN4 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-HD73_RS21400-His_6$) using arabinotriose (A3) 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



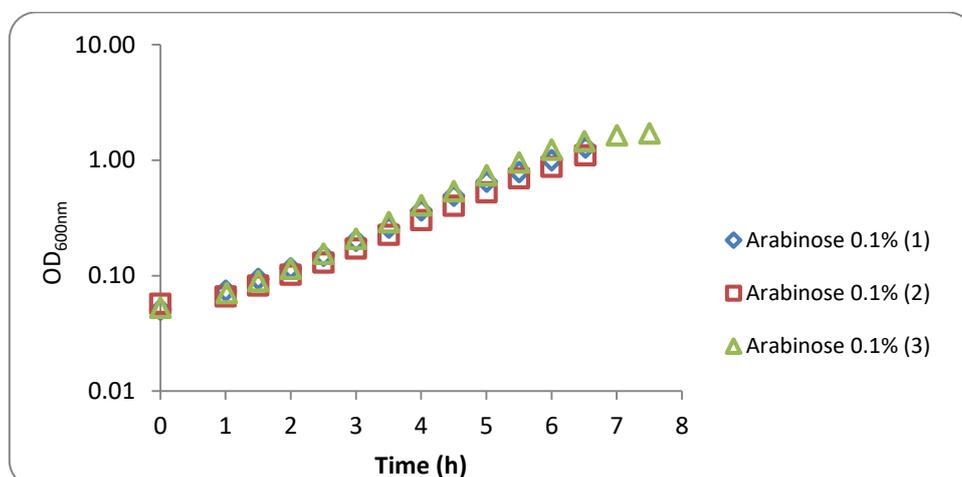
Appendix 6.38 - Growth assays of *B. subtilis* ISN5 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ugpC-His_6$) using arabinose 0.1% (w/v). OD_{600nm} plotted vs. time(h).



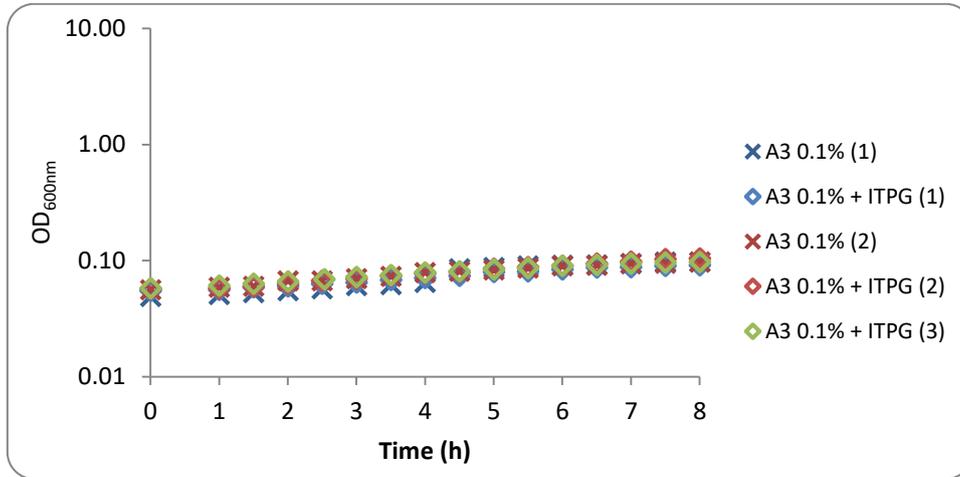
Appendix 6.39 - Growth assays of *B. subtilis* ISN5 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ugpC-His_6$) using arabinotriose (A3) 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



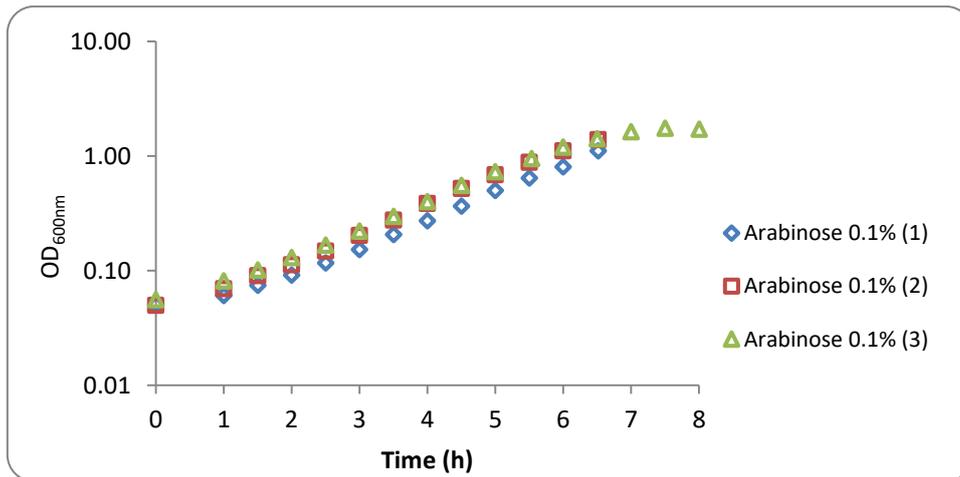
Appendix 6.40 - Growth assays of *B. subtilis* ISN6 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-malk-His_6$) using arabinose 0.1% (w/v). OD_{600nm} plotted vs. time(h).



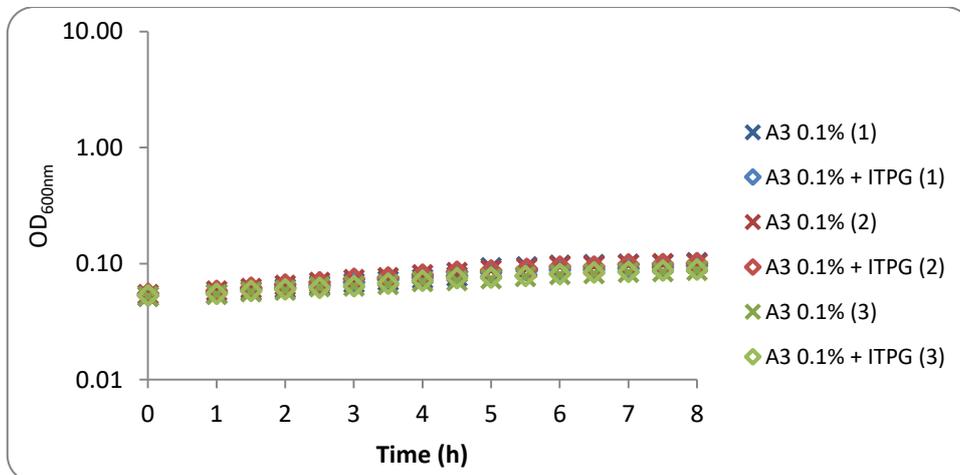
Appendix 6.41 - Growth assays of *B. subtilis* ISN6 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-malk-His_6$) using arabinotriose (A3) 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



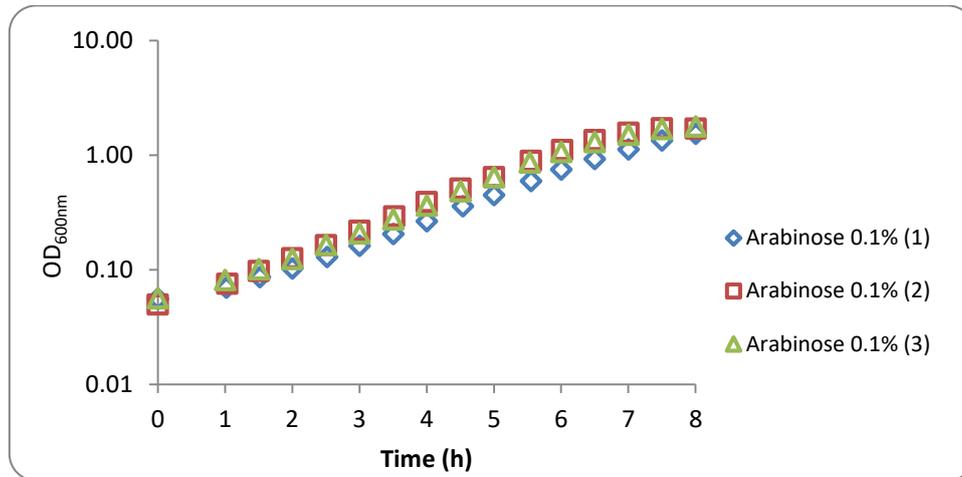
Appendix 6.42 - Growth assays of *B. subtilis* ISN7 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-yurJ-His_6$) using arabinose 0.1% (w/v). OD_{600nm} plotted vs. time(h).



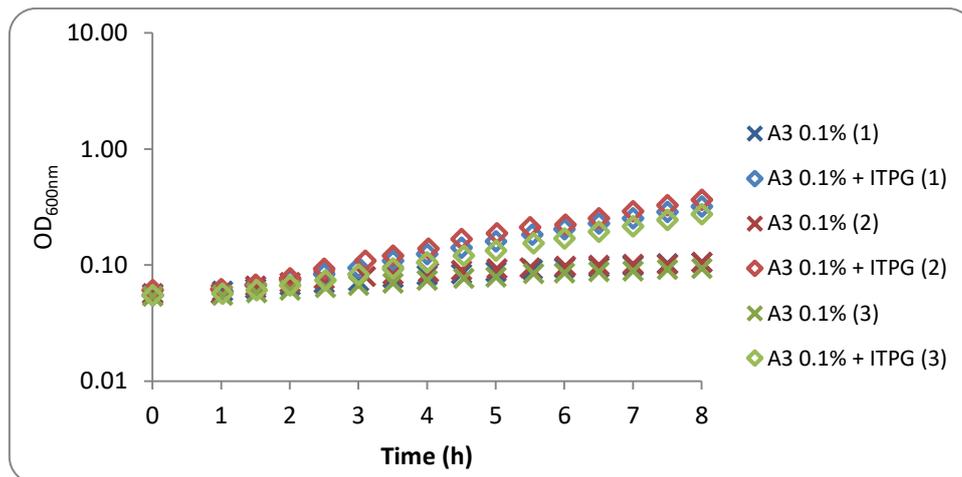
Appendix 6.43 - Growth assays of *B. subtilis* ISN7 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-yurJ-His_6$) using arabinotriose (A3) 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



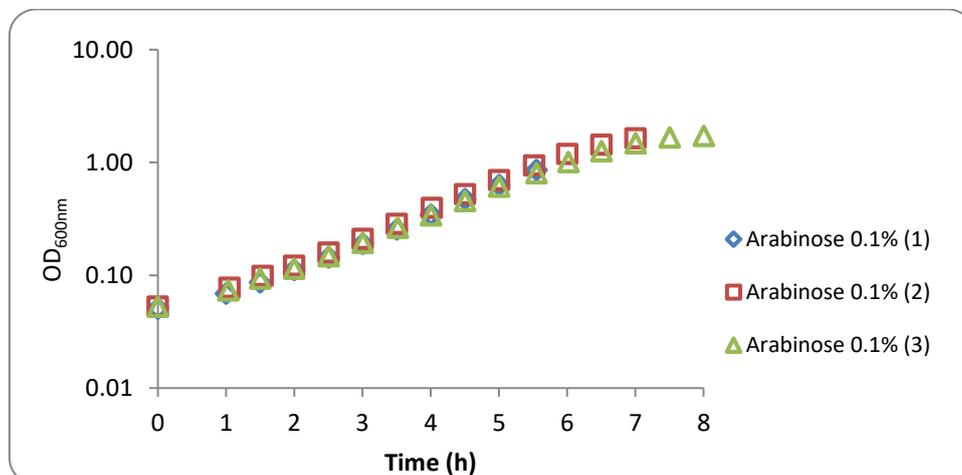
Appendix 6.44 - Growth assays of *B. subtilis* ISN8 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmK$) using arabinose 0.1% (w/v). OD_{600nm} plotted vs. time(h).



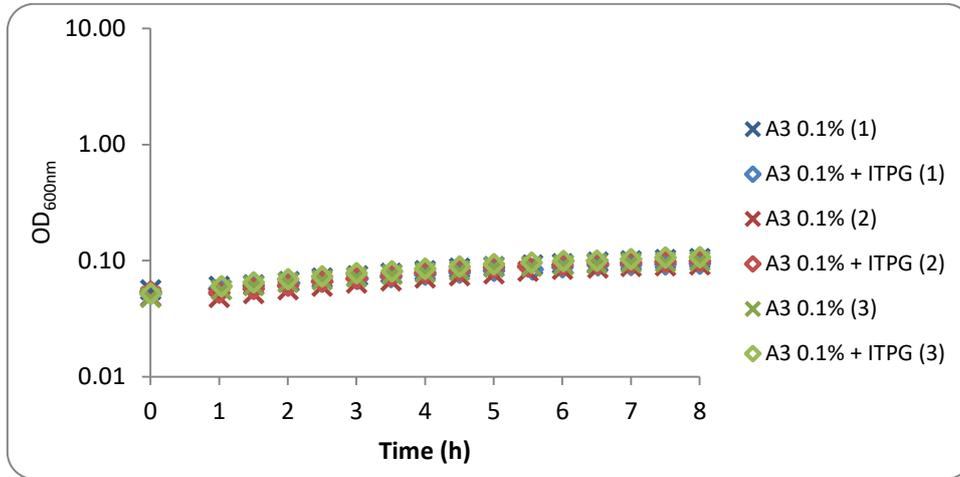
Appendix 6.45 - Growth assays of *B. subtilis* ISN8 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmK$) using arabinotriose (A3) 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



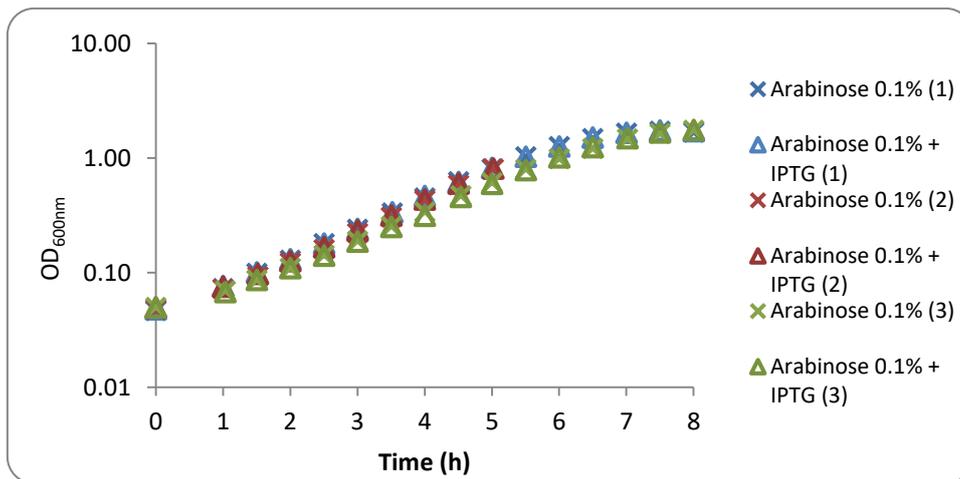
Appendix 6.46 - Growth assays of *B. subtilis* ISN9 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-malk$) using arabinose 0.1% (w/v). OD_{600nm} plotted vs. time(h).



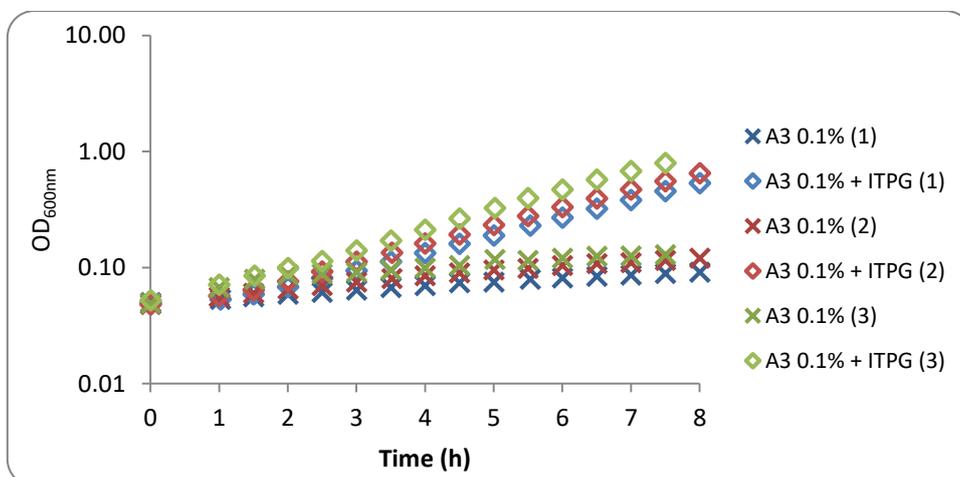
Appendix 6.47 - Growth assays of *B. subtilis* ISN9 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-malk$) using arabinotriose (A3) 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



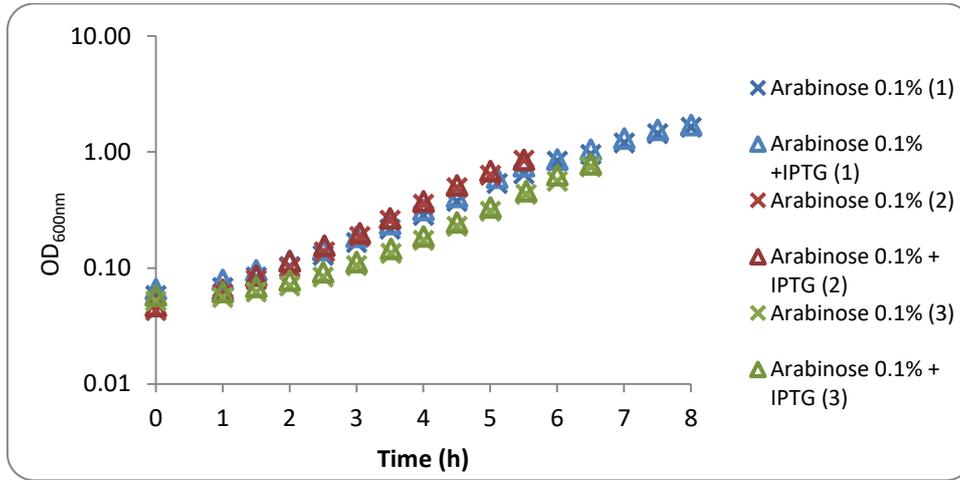
Appendix 6.48 - Growth assays of *B. subtilis* ISN10 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmX-LEHis_6$) using arabinose 0.1% (w/v). OD_{600nm} plotted vs. time(h).



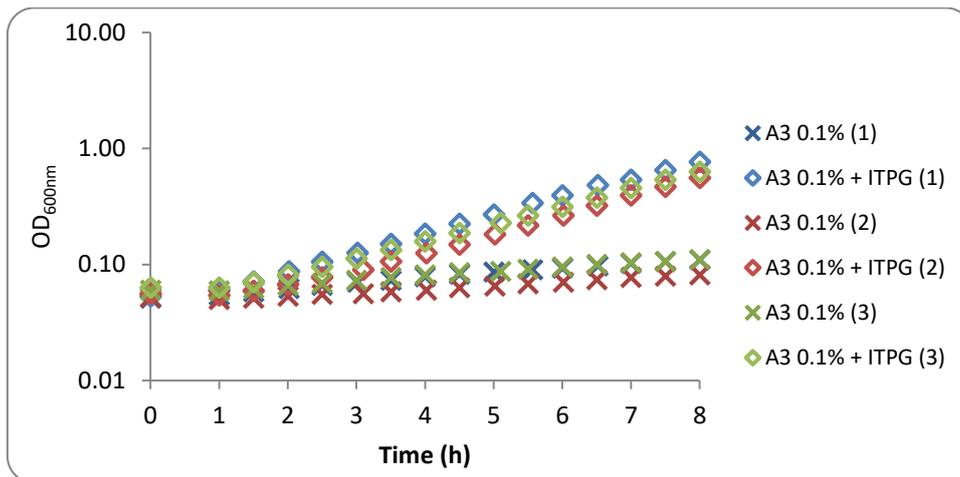
Appendix 6.49 - Growth assays of *B. subtilis* ISN10 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmX-LEHis_6$) using arabinotriose (A3) 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



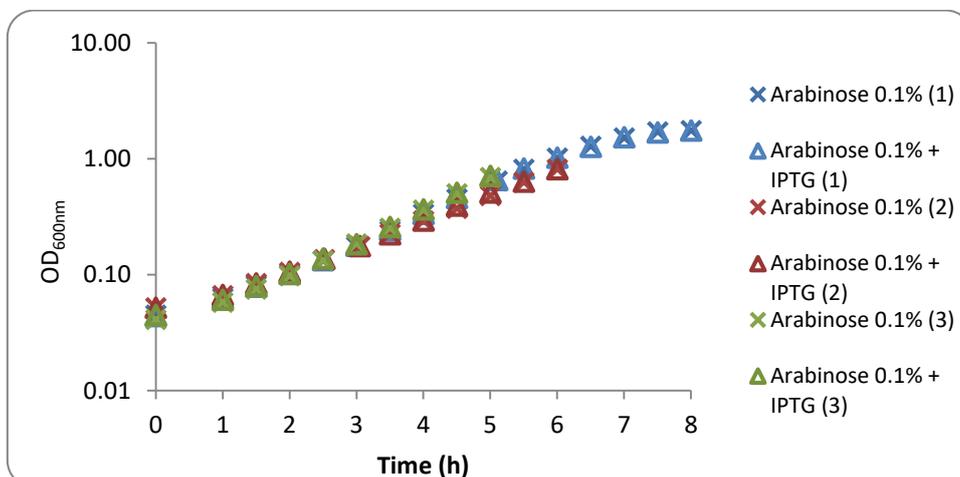
Appendix 6.50 - Growth assays of *B. subtilis* ISN11 ($\Delta ms m X::cat \Delta amy E::Pspank(hy)-HD73_RS21400-LEH_{i6}$) using arabinose 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



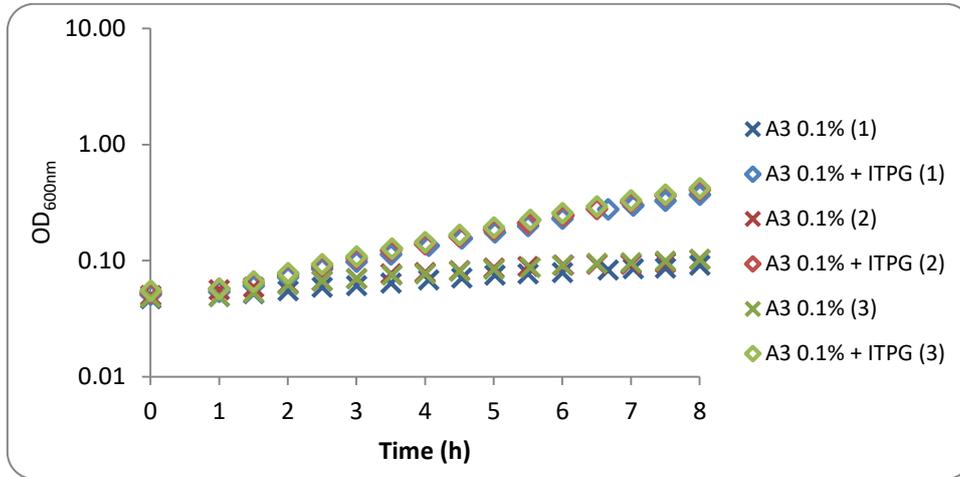
Appendix 6.51 - Growth assays of *B. subtilis* ISN11 ($\Delta ms m X::cat \Delta amy E::Pspank(hy)-HD73_RS21400-LEH_{i6}$) using arabinotriose (A3) 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



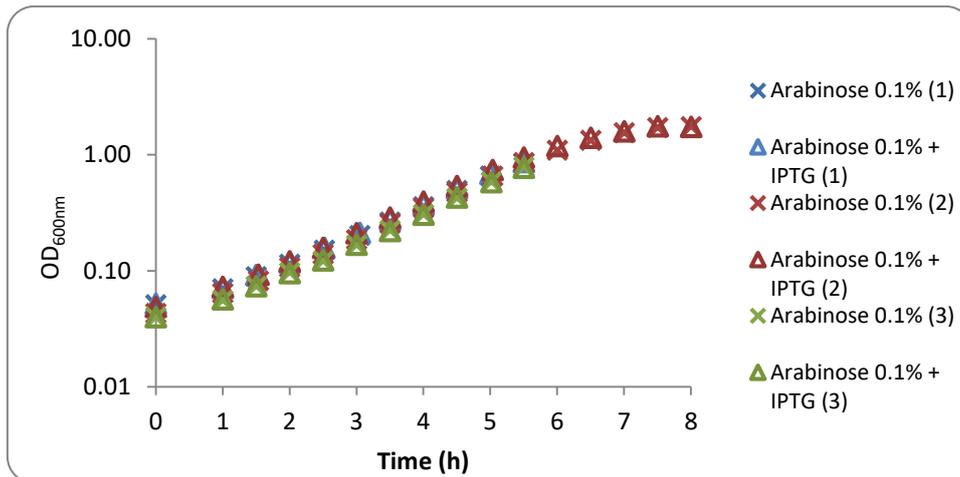
Appendix 6.52 - Growth assays of *B. subtilis* ISN12 ($\Delta ms m X::cat \Delta amy E::Pspank(hy)-msmK-LEH_{i6}$) using arabinose 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



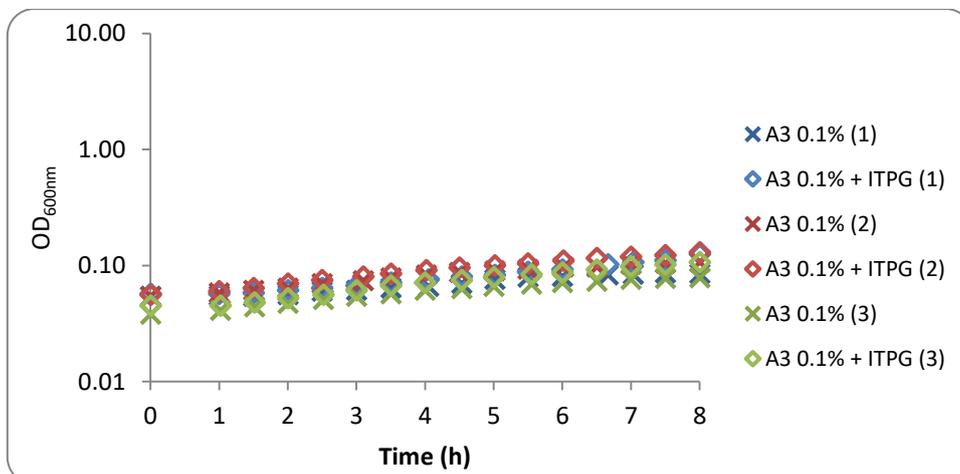
Appendix 6.53 - Growth assays of *B. subtilis* ISN12 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-msmK-LEHis_6$) using arabinotriose (A3) 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



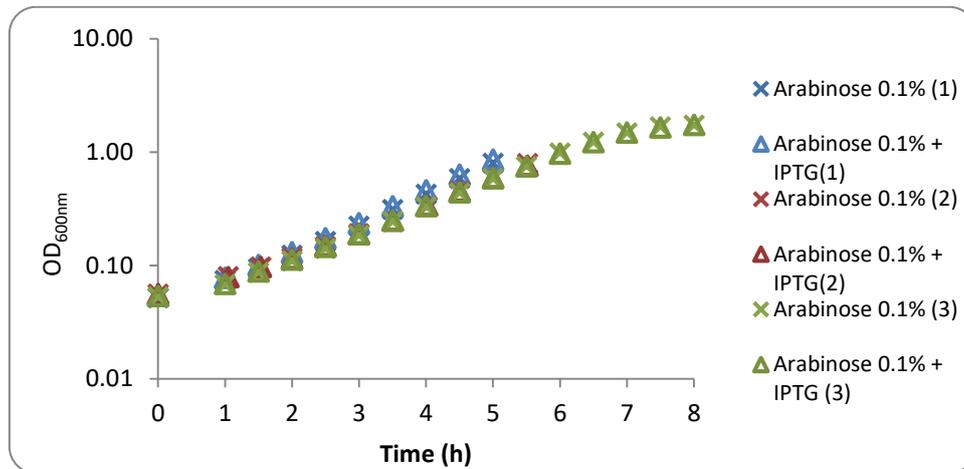
Appendix 6.54 - Growth assays of *B. subtilis* ISN13 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ugpC-LEHis_6$) using arabinose 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



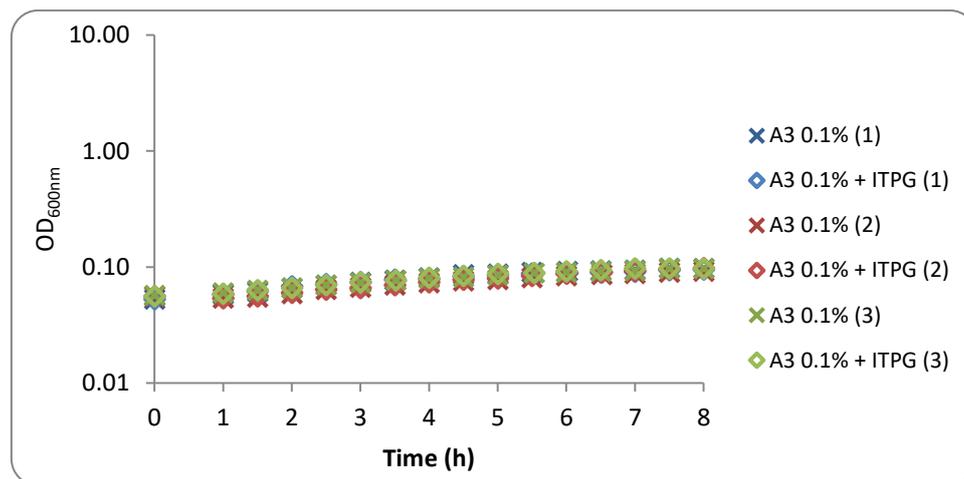
Appendix 6.55 - Growth assays of *B. subtilis* ISN13 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ugpC-LEHis_6$) using arabinotriose (A3) 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



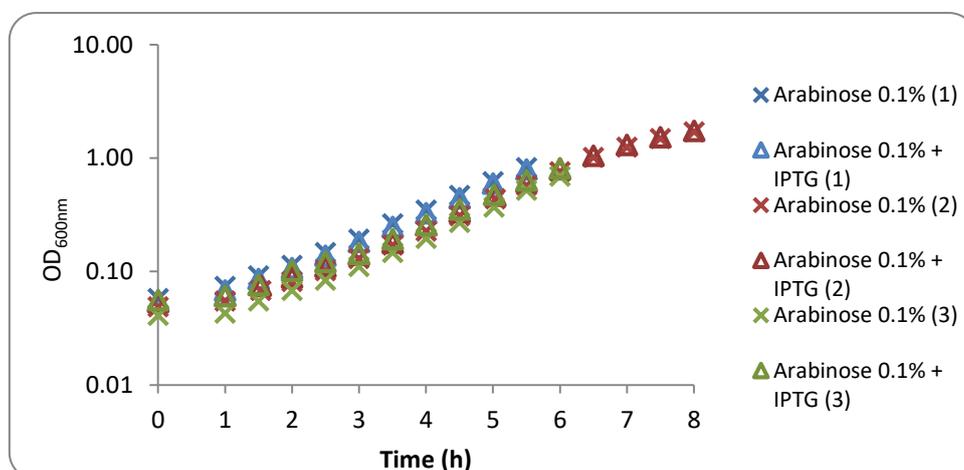
Appendix 6.56 - Growth assays of *B. subtilis* ISN14 ($\Delta ms m X::cat \Delta amy E::Psp ank(hy)-malk-LEHis_6$) using arabinose 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



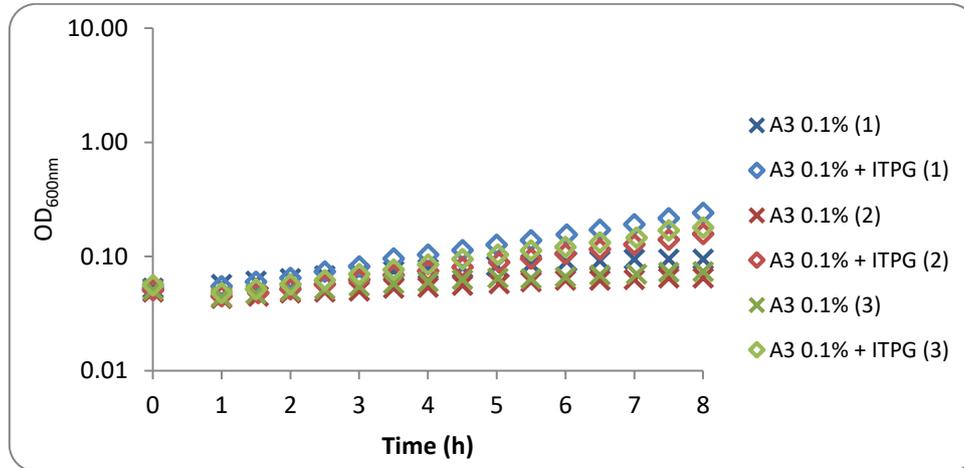
Appendix 6.57 - Growth assays of *B. subtilis* ISN14 ($\Delta ms m X::cat \Delta amy E::Psp ank(hy)-malk-LEHis_6$) using arabinotriose (A3) 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



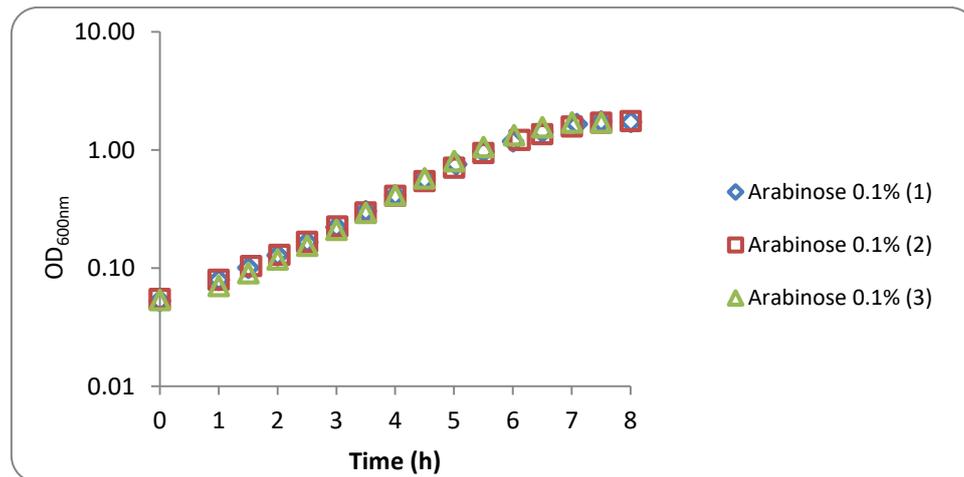
Appendix 6.58 - Growth assays of *B. subtilis* ISN15 ($\Delta ms m X::cat \Delta amy E::Psp ank(hy)-yurJ-LEHis_6$) using arabinose 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



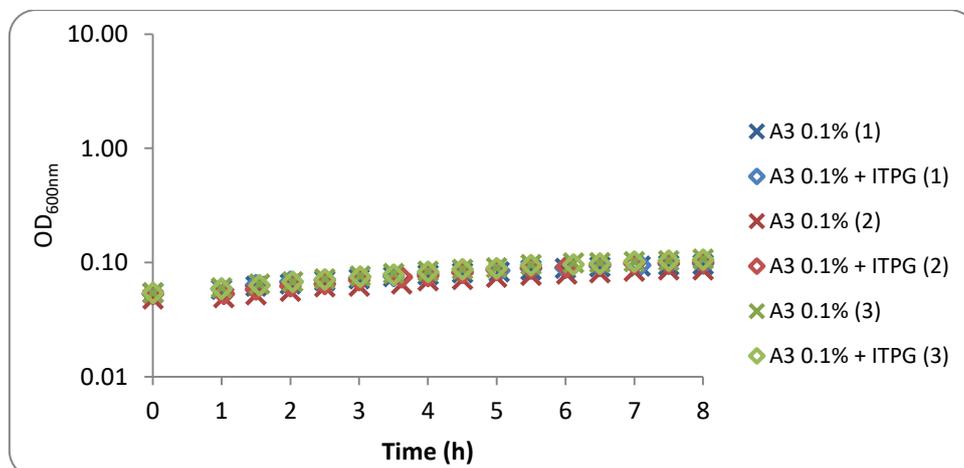
Appendix 6.59 - Growth assays of *B. subtilis* ISN15 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-yurJ-LEHis_6$) using arabinotriose (A3) 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



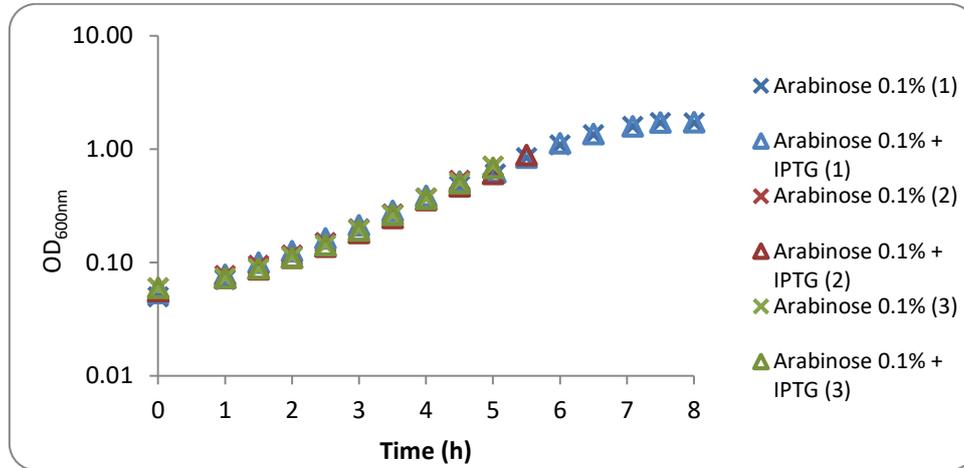
Appendix 6.60 - Growth assays of *B. subtilis* ISN16 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ycjV$) using arabinose 0.1% (w/v). OD_{600nm} plotted vs. time(h).



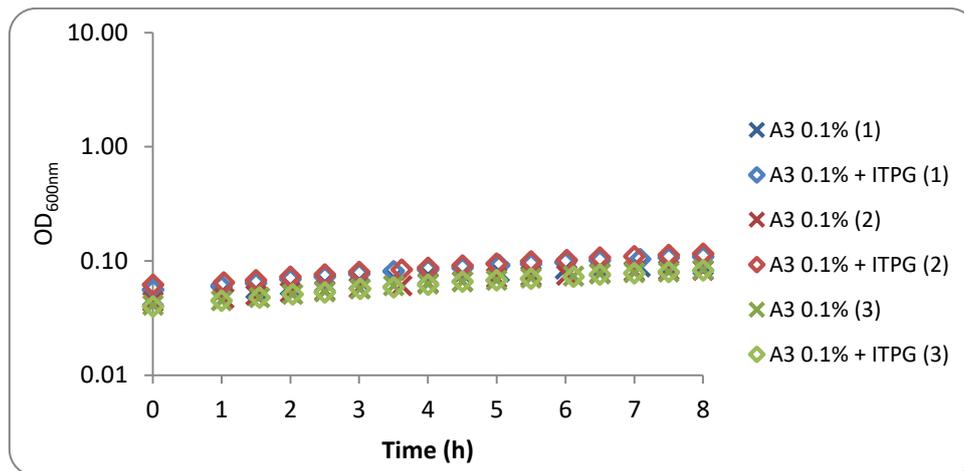
Appendix 6.61 - Growth assays of *B. subtilis* ISN16 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ycjV$) using arabinotriose (A3) 0.1% (w/v). OD_{600nm} plotted vs. time(h).



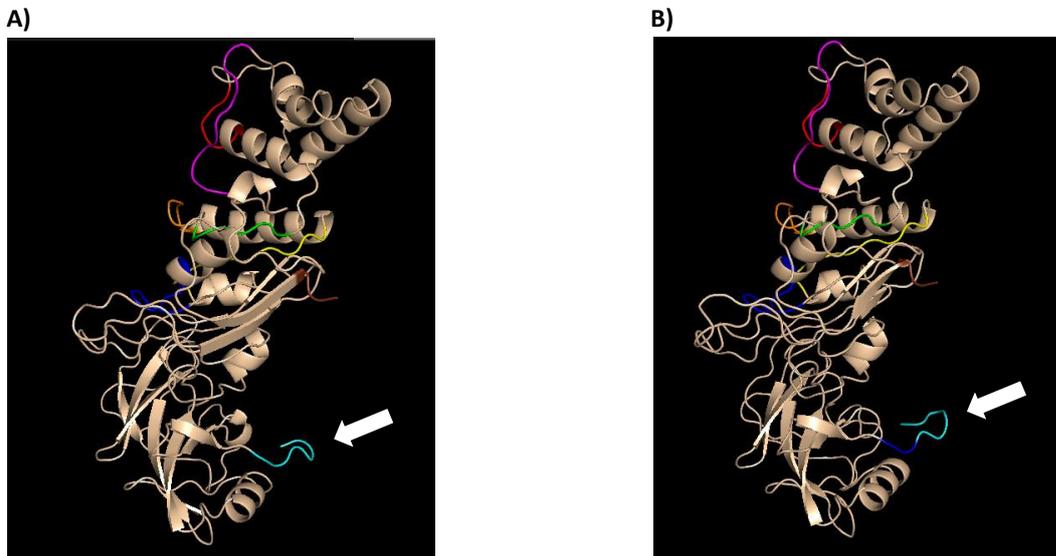
Appendix 6.62 - Growth assays of *B. subtilis* ISN17 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ycjV-LEHis_6$) using arabinose 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



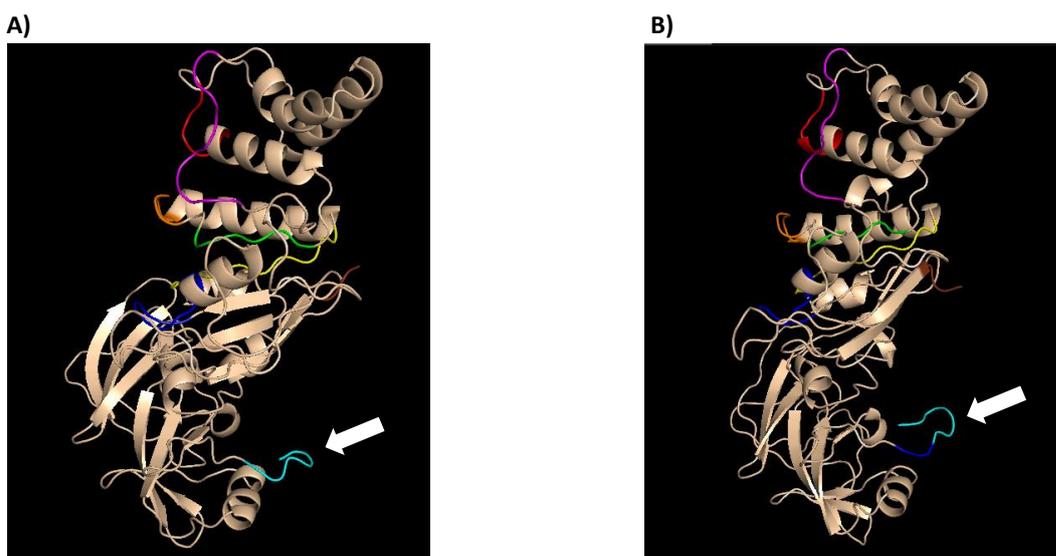
Appendix 6.63 - Growth assays of *B. subtilis* ISN17 ($\Delta msmX::cat \Delta amyE::Pspank(hy)-ycjV-LEHis_6$) using arabinotriose (A3) 0.1% (w/v) and IPTG 1mM. OD_{600nm} plotted vs. time(h).



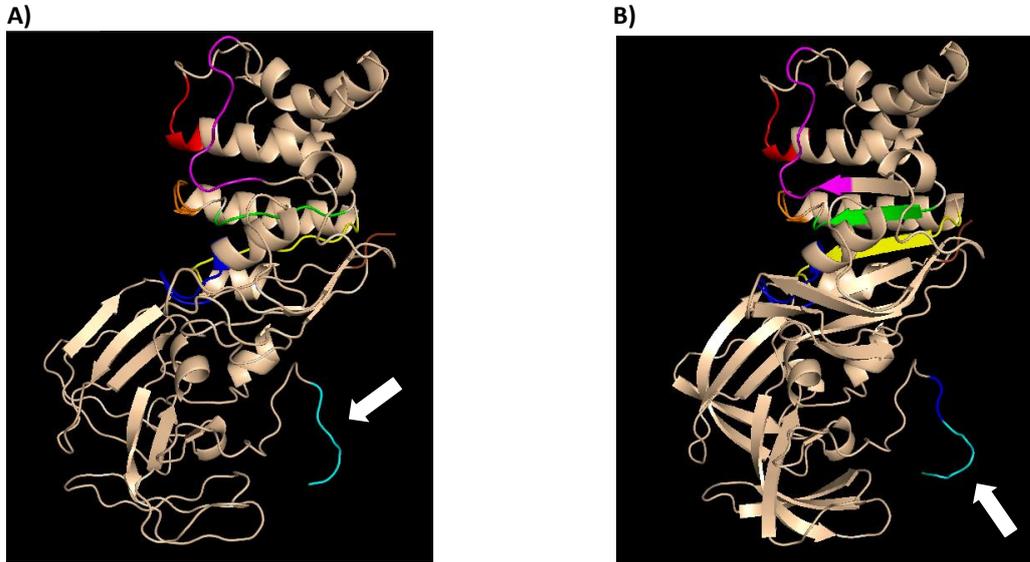
Appendix 6.64 – Representation of MsmK-His₆ and MsmK-LEHis₆ 3-D structure model. The structures were drawn in Pymol with the data obtained from the online program I-TASSER. **(A)** Model of MsmK-His₆ protein from the strain ISN3; **(B)** Model of MsmK-LEHis₆ protein from the strain ISN12.; The N-terminal, the seven conserved domains and the C-terminal tag were highlighted with the following colours scheme: N-terminal (brown), Walker A (blue), Q-loop (magenta), Signature (red), Walker B (green), D-loop (orange), H-loop (yellow) and the C-terminal tag (cyan), which is highlighted by a white arrow.



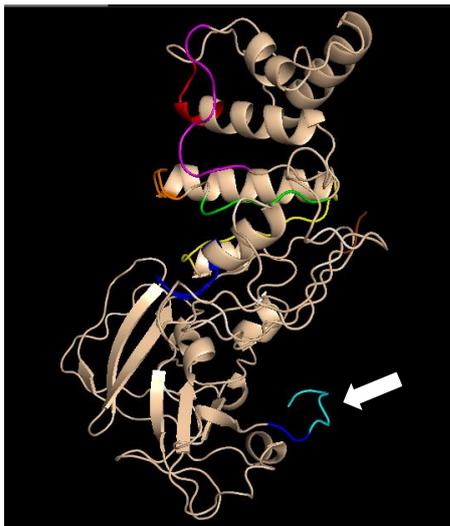
Appendix 6.65 – Representation of HD73_RS2140-His₆ and HD73_RS2140-LEHis₆ 3-D structure model. The structures were drawn in Pymol with the data obtained from the online program I-TASSER. **(A)** Model of HD73_RS2140-His₆ protein from the strain ISN4; **(B)** Model of HD73_RS2140-LEHis₆ protein from the strain ISN11; The N-terminal, the seven conserved domains and the C-terminal tag were highlighted with the following colours scheme: N-terminal (brown), Walker A (blue), Q-loop (magenta), Signature (red), Walker B (green), D-loop (orange), H-loop (yellow) and the C-terminal tag (cyan), which is highlighted by a white arrow.



Appendix 6.66 – Representation of Malk-His₆ and Malk-LEHis₆ 3-D structure model. The structures were drawn in Pymol with the data obtained from the online program I-TASSER. **A)** Model of Malk-His₆ protein from the strain ISN6; **(B)** Model of Malk-LEHis₆ protein from the strain ISN14; The N-terminal, the seven conserved domains and the C-terminal tag were highlighted with the following colours scheme: N-terminal (brown), Walker A (blue), Q-loop (magenta), Signature (red), Walker B (green), D-loop (orange), H-loop (yellow) and the C-terminal tag (cyan), which is highlighted by a white arrow.



Appendix 6.67 – Representation of YcjV-LEHis₆ 3-D structure model. The structure was drawn in Pymol with the data obtained from the online program I-TASSER. Model of YcjV-LEHis₆ protein from the strain ISN17. The N-terminal, the seven conserved domains and the C-terminal tag were highlighted with the following colours scheme: N-terminal (brown), Walker A (blue), Q-loop (magenta), Signature (red), Walker B (green), D-loop (orange), H-loop (yellow) and the C-terminal tag (cyan), which is highlighted by a white arrow.



Appendix 6.68 –MsmX-LEHis₆ protein sequence and the respective DNA sequence, from *B. subtilis* (Bioinformatic tool “Show Translation”). The protein sequence is shown in blue and the DNA sequence is shown in black; the modified amino acids (and respective codons) are highlighted with a green arrow and box.



```

1 M A S L R M E H I Y K F Y D Q K E P A V
1 ATGGCTAGCTTGC GGATGGAGCACATTTATAAATTTTATGATCAGAAGGAACCGGCGTT
21 D D F N L H I A D K E F I V F V G P S G
61 GATGACTTTAACCTTCATATTGCCGATAAGGAATTTATCGTATTCGTCGGCCCGTCCGGC
41 C G K S T T L R M V A G L E E I S K G D
121 TGC GGAAATCAACGACGCTGCGAATGGTCG CAGGACTTGAAGAAATTTCGAAAGGTGAT
61 F Y I E G K R V N D V A P K D R D I A M
181 TTTTATATTGAAGGAAAACGGGTCAATGATGTAGCGCAAAGGACAGGGATATCGCGATG
81 V F Q N Y A L Y P H M T V Y D N I A F G
241 GTATTT CAGAACTACGCGCTTTATCCGCATATGACGGTCTACGATAATATCGCGTTCGGG
101 L K L R K M P K P E I K K R V E E A A K
301 CTCAAGCTTCGGAAAATGCCGAAGCCTGAAATCAAAAAAGAGTCGAAGAAGCCGCTAAA
121 I L G L E E Y L H R K P K A L S G G Q R
361 ATTCTCGGGCTTGAGGAATATTTGCACCGTAAACCGAAAAGCGCTGTCAGGCGGACAGAGA
141 Q R V A L G R A I V R D A K V F L M D E
421 CAGCGGGTTGCGCTGGGCCGGCAATCGTGCGGGATGCAAAGGTGTTCTGATGGATGAG
161 P L S N L D A K L R V Q M R A E I I K L
481 CCTTTGTCAAACCTGGACGCGAAGCTGAGGGTGCAAATGCGGGCGGAAATCATTAAGCTC
181 H Q R L Q T T T I Y V T H D Q T E A L T
541 CACCAGAGATTGCAGACTACAACGATTTATGTGACGCATGACCAGACAGAAGCGCTGACA
201 M A T R I V V M K D G K I Q Q I G T P K
601 ATGGCGACACGGATTGTAGTCATGAAAGATGGGAAAATTCAGCAGATCGGGACGCCGAAG
221 D V Y E F P E N V F V G G F I G S P A M
661 GATGTATATGAATTCCTGAAAACGTCTTTGTCGGCGGGTTTATCGGATCACCGGCGATG
241 N F F K G K L T D G L I K I G S A A L T
721 AATTTTTTCAAAGGAAAGCTCACGGATGGCTTAATCAAAATCGGTTCTGCGGCATTAACC
261 V P E G K M K V L R E K G Y I G K E V I
781 GTCCCCGGAAGGAAAAATGAAAGTGCTGCGTGAAAAAGGCTACATCGGCAAAGAGGTCATC
281 F G I R P E D I H D E L I V V E S Y K N
841 TTCGGCATCCGTCCTGAGGATATACATGATGAATTGATTGTCGTGGAATCATATAAGAAC
301 S S I K A K I N V A E L L G S E I M I Y
901 TCTTCGATTAAGGCGAAAATTAATGTTGCAGAGCTGCTCGGTTCTGAAATTATGATTTAT
321 S Q I D N Q D F I A R I D A R L D I Q S
961 TCGCAGATAGACAACCAAGACTTTATTGCGCGGATCGACGCCCGCCTCGATATTC AATCG
341 G D E L T V A F D M N K G H F F D S E T
1021 GGCGATGAGCTGACGGTTGCATTTGATATGAATAAAGGCCATTTCTTTGACAGTGAGACA
361 E V R S R L E H H H H H H *
1081 GAAGTGAGATCTCGA CTCGAGCACCACCATCACCACCACTAA

```



Appendix 6.69 –HD73_RS21400-LEHis₆ protein sequence with the respective DNA sequence, *B. thuringiensis* (Bioinformatic tool “Show Translation”). The protein sequence is shown in blue and the DNA sequence is shown in black; the modified amino acids (and respective codons) are highlighted with a green arrow and box.



```

1 M A S L K L É N I Y K I Y D̄ N N V T A V
1 ATGGCTAGCCTTAAATTAGAAAACATTTATAAAATATATGATAATAACGTAACAGCTGTA
21 T D F N L H I Q D K E F I V F V G P S G
61 ACTGATTTTAAATTTACACATTCAAGATAAAGAATTTATTGTATTTGTCGGTCCTTCTGGA
41 C G K S T T L R M V A G L E D I S K G E
121 TGCGGAAAACTCAACATTACGAATGGTAGCTGGACTTGAAGATATTTCAAAAGGCGAG
61 F S I D G K L M N D V P P K D R D I A M
181 TTTTCAATTGATGGTAAACTAATGAATGATGTTCCCTCCAAAAGATCGAGATATCGCTATG
81 V F Q N Y A L Y P H M S V Y D N M A F G
241 GTATTCAAAACACTACGCTCTTTATCCACATATGAGTGTATATGATAATATGGCATTGGGA
101 L K L R K I P K D E I D R R V K D A A K
301 TTA AA ACTTCGAAAAATACAAAAGATGAAATCGATCGTCGTGTGAAAGATGCAGCAAAA
121 I L G L E Q Y L D R K P K A L S G G Q R
361 ATTTTAGGACTTGAACAATATTTAGATAGAAAACCGAAAGCATTATCAGGTGGACAGCGC
141 Q R V A L G R A I V R D A K V F L M D E
421 CAACGTGTTGCGTTAGGTAGAGCAATCGTTCGAGATGCAAAAAGTTTTCTTAATGGATGAA
161 P L S N L D A K L R V A M R S E I S K L
481 CCATTATCAAACCTTAGATGCTAAACTCCGTGTTGCAATGCGCTCAGAAATTTCTAAGCTA
181 H H R L G T T T I Y V T H D Q T E A M T
541 CATCATCGCCTTGAACAACAACAATTTATGTAACACATGACCAAAACAGAAGCAATGACT
201 M A S R L V V M K D G K I Q Q I G T P K
601 ATGGCTTCGCGTCTTGTGCGTTATGAAGGATGGAAAAATCCAACAAATTGGAACCTCAAAA
221 E V Y E T P E N I F V G G F I G S P A M
661 GAAGTATACGAAACGCCAGAAAACATTTTCGTGGGTGGATTTATCGGCTCACCAGCTATG
241 N F F R G K L T E T D F V I D N T I K I
721 AATTTCTCCGTGGTAAATTGACTGAAACCGATTTTCGTTATAGACAATACTATAAAAATT
261 K V S E G K M K M L R E Q G Y V N K E I
781 AAAGTATCTGAAGGAAAAATGAAGATGCTACGTGAACAAGGCTATGTAAATAAGGAAATT
281 V L G I R P E D I H D E L L F L E A S Q
841 GTTTTAGGTATTCGCCCAGAGGACATTCATGATGAACTACTATTTTTAGAAGCTTCGCAA
301 S T A F T T K I E V A E L L G A E S I L
901 TCTACTGCCTTCACAACAAAATTGAAGTCGCTGAATTATTAGGTGCTGAATCCATTTTA
321 Y M K L G N Q D F A A R V D A R H T F S
961 TATATGAAACTTGGAAATCAAGATTTTGCAGCACGTGTAGATGCAAGACATACATTTTCA
341 P G D Q I K L A F D M N K A H F F D N Q
1021 CCTGGTGATCAAATTAATTTAGCATTGATATGAATAAAGCACATTTCTTTGATAACCAA
361 T E Q R S R L E H H H H H H *
1081 ACTGAACAAAGATCTCGA CTCTGAGCACCACCATCACCACCACTAA

```



Appendix 6.70 –MsmK-LEHis₆ protein sequence with the respective DNA sequence, *S. pneumoniae* (Bioinformatic tool “Show Translation”). The protein sequence is shown in blue and the DNA sequence is shown in black; the modified amino acids (and respective codons) are highlighted with green boxes.

```

1 M A S L N L K N I Y K K Y P N S E H Y S
1 ATGGCTAGCTTGAATCTTAAAAATATTTACAAAAATATCCAAACAGCGAACACTATTCA
21 V E D F N L N I K D K E F I V F V G P S
61 GTTGAAGATTTCAACTTGAACATCAAAGATAAAGAATTTATCGTTTTCGTAGGACCTTCA
41 G C G K S T T L R M I A G L E D I T E G
121 GGATGTGGTAAATCAACTACACTCCGTATGATTGCTGGTCTTGAAGACATTACAGAAGGT
61 T A S I D G V V V N D V A P K D R D I A
181 ACTGCATCTATCGATGGCGTAGTTGTCAACGACGTAGCTCCAAAAGACCGTGATATCGCC
81 M V F Q N Y A L Y P H M T V Y D N M A F
241 ATGGTATTCCAAAACACTACGCTCTTACCCACACATGACTGTTTATGACAACATGGCTTTC
101 G L K L R K Y S K E D I N K R V Q E A A
301 GGTGGTAAATTGCGTAAATACAGCAAAGAAGACATTAACAAACGTGTCAAGAAGCAGCT
121 E I L G L K E F L E R K P A D L S G G Q
361 GAAATACTTGGATTGAAAGAATTCTTGAACGTAAACCAGCTGACCTTTCAGGTGGTCAA
141 R Q R V A M G R A I V R D A K V F L M D
421 CGTCAACGTGTTGCCATGGGGCGTGCGATTGTCCGTGATGCGAAAGTATTCTTGATGGAC
161 E P L S N L D A K L R V S M R A E I A K
481 GAACCTTTGTCAAACCTGGATGCCAAAACCTCGTGATCAATGCGTGCTGAAATCGCTAAA
181 I H R R I G A T T I Y V T H D Q T E A M
541 ATTCACCGTCGTATCGGAGCTACAACATCTATGTAACACAGACCAAACAGAAGCGATG
201 T L A D R I V I M S A T K N P A G T G T
601 ACACTTGACAGACCGTATCGTTATTATGTCAGCTACTAAGAACCCTGCTGGTACAGGTA
221 I G R V E Q I G T P Q E V Y K N P V N K
661 ATCGGACGTGTAGAACAAATCGGTACTCCTCAAGAAGTTTACAAAAATCCAGTTAAACAAA
241 F V A G F I G S P A M N F I T V K L V G
721 TTCGTTGACAGGATTCATCGGAAGCCCAGCTATGAACTTCATCACCGTGAAATTGGTTGGT
261 S E I V S D G F R L K V P E G A L K V L
781 AGCGAAATTGTTTCTGACGGTTTCCGTTTGAAGTGCCAGAAGGAGCATTGAAAGTTCTT
281 R E K G Y E G K E L I F G I R P E D V N
841 CGTGAAAAAGGCTACGAAGGAAAAGAATTGATCTTTGGTATCCGTCCAGAAGACGTGAAT
301 A E P A F L E T F P D C V V K A T I S V
901 GCAGAACCTGCTTTCCTTGAACATTCCCAGACTGTGTTGTAAGGACGACTATCTCTGTA
321 S E L L G S E S H L Y C Q V G K D E F V
961 TCAGAACTGCTTGGTTGAGAATCTCACCTTTACTGTCAAGTTGGTAAAGACGAGTTTGT
341 A K V D A R D Y L Q T G A T V E L G F D
1021 GCAAAAAGTTGATGCTCGTACTACTTGCAAAACAGGTGCAACAGTTGAGCTTGGATTTGAC
361 L N K A H F F D V E T E K R S R L E H H
1081 TTGAACAAAGCACACTTCTTCGATGTAGAAACTGAAAAAGATCTCGACTCGAGCACCAC
381 H H H H *
1141 CATCACCACCACTAA

```

Appendix 6.71 –UgpC-LEHis₆ protein sequence with the respective DNA sequence, *S. aureus* (Bioinformatic tool “Show Translation”). The protein sequence is shown in blue and the DNA sequence is shown in black; the modified amino acids (and respective codons) are highlighted with a green arrow and box.



```

1 M A S L K L E H I K K T Y D N N N T V V
1 ATGGCTAGCTTAAAGTTAGAGCATATTA AAAAGACGTATGATAACAACAATACTGTAGTG
21 K D F N L H I T D K E F I V F V G P S G
61 AAAGATTTTAACTACATATTACTGACAAAGAATTCATTGTATTTGTTGGACCATCGGGA
41 C G K S T T L R M V A G L E S I T S G D
121 TGTGGTAAATCAACAAC TTTACGAATGGTTGCTGGACTAGAGTCTATCACATCTGGAGAT
61 F Y I D G E R M N D V E P K N R D I A M
181 TTTTATATTGATGGGGAACGCATGAACGATGTTGAACCAAAGAATAGAGATATTGCGATG
81 V F Q N Y A L Y P H M T V F E N I A F G
241 GTATTTCAAAC TATGCATTATATCCACATATGACTGTTTTTAAAAATATAGCATTTGGG
101 L K L R K V N K K E I E Q K V N E A A E
301 CTAAAGCTACGTA AAGTGAATAAAAAAGAGATTGAACAAAAAGTTAATGAAGCAGCAGAA
121 I L G L T E Y L G R K P K A L S G G Q R
361 ATATTAGGATTA AACTGAGTATCTTGGTCGTAACCAAAGCGTTATCTGGTGGACAGCGT
141 Q R V A L G R A I V R D A K V F L M D E
421 CAACGTGTTGCT TTTGGGCAGAGCTATTGTTAGGGATGCGAAAGTCTTTTAAATGGATGAA
161 P L S N L D A K L R V Q M R T E I L K L
481 CCATTATCGAAT CTTGATGCGAAGCTTCGAGTACAAATGCGTACAGAAATATTGAAATTA
181 H K R L N T T T I Y V T H D Q T E A L T
541 CATAAGCGACTT AACTACGACAATTTATGTTACACATGATCAAAC TGAAGCATTGACG
201 M A S R I V V L K D G D I M Q V G T P R
601 ATGGCTAGTCGA ATTTGTTGTTTTGAAAGATGGCGACATTATGCAAGTCGGCACACCTAGA
221 E I Y D A P N C I F V A Q F I G S P A M
661 GAAATATATGAT GCCCTAATTGCATATTTGTGGCGCAATTTATCGGCTCACCAGCAATG
241 N M L N A T V E M D G L K V G T H H F K
721 AATATGTTGAAT GCTACAGTTGAAATGGACGGATTGAAGGTAGGAACACACCATTTTAAA
261 L H N K K F E K L K A A G Y L D K E I I
781 TTACATAATAAAA AATTTGAAAAGTTAAAAGCTGCTGGCTACTTAGACAAGGAAATTATT
281 L G I R A E D I H E E P I F I Q T S P E
841 TTAGGTATTCGAG CTGAAGACATTCATGAAGAACCAATATTTATTCAAACCTCTCCAGAG
301 T Q F E S E V V V S E L L G S E I M V H
901 ACACAATTTGAAT CTGAAGTAGTTGTATCCGAAGTGTAGGTT CAGAAATTATGGTACAT
321 S T F Q G M E L I S K L D S R T Q V M A
961 AGCACATTCGAAG GAATGGAATTTGATTTCTAAATTAGATTCAAGAACACAAGTGATGGCG
341 N D K I T L A F D M N K C H F F D E K T
1021 AATGACAAGATT AACTAGCATTGATATGAATAAGTGTCAC TTTTTTGATGAAAAACA
361 G N R S R L E H H H H H H *
1081 GGAAATAGATCTCGACTCGAGCACCACCATCACCACCACTAA

```

Appendix 6.72 –YurJ-LEHis₆ protein sequence with the respective DNA sequence, *B. subtilis* (Bioinformatic tool “Show Translation”). The protein sequence is shown in blue and the DNA sequence is shown in black; the modified amino acids (and respective codons) are highlighted with a green box.

```

1 M A S L T F E H V K K S Y H S Q L T V K
1 ATGGCTAGCTTAACATTTGAACACGTCAAAAAATCATATCACTCACAATTAACCGTGAAG
21 D F D L D V K D K E L L V L V G P S G C
61 GACTTTGATTTAGATGTAAAGGATAAGGAGCTTTTGGTGCTGGTGGGGCCGTCAGGATGC
41 G K S T T L R M V A G L E S I S E G N L
121 GGAAAATCAACAACGCTGCGGATGGTGGCTGGATTAGAAAGCATTCTGAAGGAAATCTT
61 L I D G E R V N D L P P K E R D I A M V
181 CTTATTGATGGGGAGCGGGTCAATGATCTTCCGCCGAAAGAACGTGATATTGCGATGGTA
81 F Q N Y A L Y P H M T V F D N M A F G L
241 TTCCAGAACTATGCGCTTTATCCGCATATGACGGTTTTTGATAATATGGCGTTTGGATTA
101 K L R K M A K Q E I A E R V H A A A R I
301 AAGCTCAGAAAGATGGCGAAACAGGAAATTGCCGAGCGGGTTCATGCGGCAGCCAGAATT
121 L E I E H L L K R K P K A L S G G Q R Q
361 TTAGAAAATTGAACATTTGCTGAAGAGAAAACCGAAGGCGCTCTCCGGCGGGCAGCGGCAA
141 R V A L G R S I V R E P K V F L M D E P
421 AGGGTGGCGCTTGGCAGGTCGATTGTGAGGGAGCCGAAGGTCTTTTAAATGGACGAGCCG
161 L S N L D A K L R V T M R T E I S K L H
481 CTTTCTAATCTGGATGCAAACTAAGAGTGACGATGCGGACAGAAATCAGCAAGCTGCAC
181 Q R L E A T I I Y V T H D Q T E A M T M
541 CAGCGATTAGAGGCAACGATCATTATGTACGCATGACCAGACCGAAGCAATGACAATG
201 G D R I V V M N E G E I Q Q V A K P H D
601 GGAGATCGGATTGTCGTCATGAATGAAGGGGAAATTCAGCAAGTGGCGAAACCTCATGAC
221 I Y H Y P A N L F V A G F I G S P G M N
661 ATTTATCATTACCCGGCCAATTTGTTTGTGCGAGGCTTCATTGGCTCTCCCGGAATGAAC
241 F L K G I I E Q Q H G E L F F T N S S I
721 TTTTGAAGGCATCATAGAACAGCAGCATGGCGAGCTGTTTTTACAAAATCTTCGATC
261 R L H I P E E K A K R L K E K G Y A G E
781 CGGCTTCATATTCCTGAAGAAAAAGCAAAGCGGCTGAAGGAAAAGGGATATGCCGGAGAA
281 Q M I A G V R P E H I T Q M T G N D Q L
841 CAGATGATTGCCGGTGTACGGCCGGAGCACATCACACAAATGACAGGAAATGACCAGTTG
301 F D S V F Q A N V E V N E N L G S E L I
901 TTTGATTCAGTGTTC AAGCGAACGTGGAAGTGAATGAAAATCTGGGCTCGGAGCTCATA
321 V H V M A G D E R L K V R L D G N T R I
961 GTGCATGTCATGGCTGGGGACGAGCGGCTCAAGGTCCGCCTGGACGGAAACACGCGAATC
341 D A G D S I Q L S V K M D H V V F F D A
1021 GATGCAGGCGATTCAATCCAGCTCTCAGTGAAAATGGACCACGTCGTCTTTTTCGATGCG
361 E T E E R S R L E H H H H H H *
1081 GAAACGGAAGAGAGATCTCGACTCGAGCACCACCATCACCACCACTAA

```

Appendix 6.73 –MalK-LEHis₆ protein sequence with the respective DNA sequence, *E. coli* (Bioinformatic tool “Show Translation”). The protein sequence is shown in blue and the DNA sequence is shown in black; the modified amino acids (and respective codons) are highlighted with a green box.

```

1 M A S V Q L Q̇ N V T K A W Ġ E V V V S K
1 ATGGCTAGCGTACAGCTGCAAAATGTAACGAAAGCCTGGGGCGAGGTCGTGGTATCGAAA
21 D I N L D I H E G E F V V F V G P S G C
61 GATATCAATCTCGATATCCATGAAGGTGAATTCGTGGTGTGTTGTCGGACCGTCTGGCTGC
41 G K S T L L R M I A G L E T I T S G D L
121 GGTAAATCGACTTTACTGCGCATGATTGCCGGGCTTGAGACGATCACCAGCGGGCGACCTG
61 F I G E K R M N D T P P A E R G V G M V
181 TTCATCGGTGAGAAACGGATGAATGACACTCCGCCAGCAGAACGCGGGCTTGGTATGGTG
81 F Q S Y A L Y P H L S V A E N M S F G L
241 TTTAGTCTTACGCGCTCTATCCCCACCTGTCAGTAGCAGAAAACATGTCATTTGGCCTG
101 K L A G A K K E V I N Q R V N Q V A E V
301 AAAGTGGCTGGCGCAAAAAAGAGGTGATTAACCAACGCGTTAACCAGGTGGCGGAAGTG
121 L Q L A H L L D R K P K A L S G G Q R Q
361 CTACAACGGCGCATTGCTGGATCGAAACCGAAAGCGCTCTCCGGTGGTCAGCGTCAG
141 R V A I G R T L V A E P S V F L L D E P
421 CGTGTGGCGATTGGCCGTACGCTGGTGGCCGAGCCAAGCGTATTTTGTCTGATGAACCG
161 L S N L D A A L R V Q M R I E I S R L H
481 CTCTCAACCTCGATGCTGCACTGCGTGTGCAAATGCGTATCGAAATCTCCCGTCTGCAT
181 K R L G R T M I Y V T H D Q V E A M T L
541 AAACGCTGGGCCGACAATGATTTACGTCACCCACGATCAGGTGCAAGCGATGACGCTG
201 A D K I V V L D A G R V A Q V G K P L E
601 GCCGACAAAATCGTGGTGTGGACGCCGGTTCGCGTGGCGCAGGTTGGGAAACCGCTGGAG
221 L Y H Y P A D R F V A G F I G S P K M N
661 CTGTACCACTATCCGGCAGACCGTTTTGTCGCCGGATTTATCGGTTCCGCAAAGATGAAC
241 F L P V K V T A T A I D Q V Q V E L P M
721 TTCCTGCCGGTAAAAGTGACCGCCACCGCAATCGATCAAGTGCAGGTGGAGCTGCCGATG
261 P N R Q Q V W L P V E S R D V Q V G A N
781 CCAAATCGTCAGCAAGTCTGGCTGCCAGTTGAAAGCCGTGATGTCCAGGTTGGAGCCAAT
281 M S L G I R P E H L L P S D I A D V I L
841 ATGTCGCTGGGTATTCGCCCGAACATCTACTGCCGAGTGATATCGCTGACGTATCCTT
301 E G E V Q V V E Q L G N E T Q I H I Q I
901 GAGGGTGAAGTTCAGGTCGTCGAGCAACTCGGCAACGAAACTCAAATCCATATCCAGATC
321 P S I R Q N L V Y R Q N D V V L V E E G
961 CCTCCATTGTCAAAACCTGGTGTACCGCCAGAACGACGTGGTGTGGTAGAAGAAGGT
341 A T F A I G L P P E R C H L F R E D G T
1021 GCCACATTGCTATCGGCCTGCCCGCAGAGCGTTGCCATCTGTTCCGTGAGGATGGCACT
361 A C R R L H K E R S R L E H H H H H H *
1081 GCATGTCGTCGACTGCATAAGGAGAGATCTCGACTCGAGCACCACCATCACCACCACTAA

```

Appendix 6.74 –YcjV-LEHis₆ protein sequence with the respective DNA sequence, *E. coli* (Bioinformatic tool “Show Translation”). The protein sequence is shown in blue and the DNA sequence is shown in black; the modified amino acids (and respective codons) are highlighted with a green box and arrow.



```

1 M A S L S L Q H I Q K I Y D N Q V H V V
1 ATGGCTAGCCTTTCGTTACAACATATTCAAAAAATCTACGATAACCCAGGTGCATGTGGTG
21 K D F N L E I A D K E F I V F V G P S G
61 AAGGACTTCAACCTGGAAATTGCCGATAAAGAGTTCATCGTGTTCGCGCCCGTCGGGC
41 C G K S T T L R M I A G L E E I S G G D
121 TGCGGTAAGTCGACCACCTGCGCATGATTGCCGGGCTTGAGGAGATCAGCGGCGGCGAT
61 L L I D G K R M N D V P A K A R N I A M
181 CTGTTGATCGACGGCAAACGAATGAATGACGTTCCAGCCAAAGCACGCAATATAGCGATG
81 V F Q N Y A L Y P H M T V Y D N M A F G
241 GTGTTCCAGAACTACGCGTTGTATCCGCATATGACGGTTTACGACAACATGGCGTTTGGT
101 L K M Q K I A K E V I D E R V N W A A Q
301 CTGAAGATGCAAAAAATCGCCAAAGAGGTGATTGATGAGCGGGTGAACTGGGCGGGCGCAA
121 I L G L R E Y L K R K P G A L S G G Q R
361 ATTCTCGGCCTGCGTGAGTACCTGAAACGTAAGCCGGGGGCGCTTTCGCGGGGCAACGT
141 Q R V A L G R A I V R E A G V F L M D E
421 CAGCGAGTGGCGCTTGGGCGGGCGATTGTACGCGAAGCGGGCGTGTTTTAAATGGATGAA
161 P L S N L D A K L R V Q M R A E I S K L
481 CCGCTCTAACCCTTGATGCCAAGCTGCGCGTGCAAATGCGCGCAGAGATCAGCAAGCTG
181 H Q K L N T T M I Y V T H D Q T E A M T
541 CATCAGAAACTGAACACCACCATGATCTACGTGACCCACGATCAGACCGAAGCGATGACC
201 M A T R I V I M K D G I V Q Q V G A P K
601 ATGGCGACGCGGATTGTGATTATGAAAGACGGGATTGTTCAAGCAAGTAGGTGCGCCGAAA
221 T V Y N Q P A N M F V S G F I G S P A M
661 ACCGTTTATAACCAACCCGCGAATATGTTTGTTCGGATTTATTGGATCACCAGCGATG
241 N F I R G T I D G D K F V T E T L K L T
721 AATTTTATTCGCGGCACGATCGATGGCGATAAATTCGTTACGGAAACGCTTAAATTAACC
261 I P E E K L A V L K T Q E S L H K P I V
781 ATTCCCGAAGAGAAAATTAGCGTTCTGAAAACACAGGAAAGTTTGCATAAGCCCATCGTG
281 M G I R P E D I H P D A Q E E N N I S A
841 ATGGGAATACGACCAGGAAAGATATTCATCCGGACGCGCAAGAGGAAAATAACATTTCCGCC
301 K I S V A E L T G A E F M L Y T T V G G
901 AAAATTAGCGTGGCAGAATTAACCGGTGCGGAATTTATGCTCTACACCACGGTTGGGGGG
321 H E L V V R A G A L N D Y H A G E N I T
961 CACGAGTTAGTGGTCCGTGCTGGTGCGTAAATGATTATCATGCAGGAGAAAATATCACT
341 I H F D M T K C H F F D A E T E I R S R
1021 ATTCATTTTGATATGACGAAATGTCATTTCTTTGATGCAGAAACGGAAATAAGATCTCGA
361 L E H H H H H H *
1081 CTCGAGCACCACCATCACCACCACTAA

```

Appendix 6.75 –Molecular weight of the recombinant proteins calculated with the bioinformatic tool ProtParam from ExPASy.

Proteins	Number of amino acids	Molecular weight (Da)
MsmX-LEHis ₆	373	42362.96
HD73_RS21400-LEHis ₆	374	42432.83
MsmK-LEHis ₆	384	42838.09
UgpC-LEHis ₆	373	42370.83
YurJ-LEHis ₆	375	42482.78
MalK-LEHis ₆	379	42201.72
YcjV-LEHis ₆	368	41229.61