Mário José Rodrigues Ferreira
MSc. in Biotechnology

Genetic and biochemical characterization of hemicellulose polysaccharides utilization in *Bacillus subtilis*

Dissertation to obtain a PhD degree in Biology

Supervisor: Isabel Maria Godinho de Sá Nogueira, Associate Professor with Habilitation (Agregação), FCT-UNL.

Jury:

President: Maria João Lobo de Reis Madeira Crispim Romão, Full Professor, FCT-UNL.

Examiners: Carlos Mendes Godinho de Andrade Fontes, Associate Professor with Habilitation (Agregação), FMV-UL.

Adriano José Alves de Oliveira Henriques, Associate Professor, ITQB-UNL.

Members: Carlos Jorge Sousa de São José, Assistant Professor, FF-UL.

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

June 2016

Mário José Rodrigues Ferreira
MSc. in Biotechnology

Genetic and biochemical characterization of hemicellulose polysaccharides utilization in *Bacillus subtilis*

Dissertation to obtain a PhD degree in Biology

Supervisor: Isabel Maria Godinho de Sá Nogueira, Associate Professor with Habilitation (Agregação), FCT-UNL.

Jury:

President: Maria João Lobo de Reis Madeira Crispim Romão, Full Professor, FCT-UNL.

Examiners: Carlos Mendes Godinho de Andrade Fontes, Associate Professor with Habilitation (Agregação), FMV-UL.

Adriano José Alves de Oliveira Henriques, Associate Professor, ITQB-UNL.

Members: Carlos Jorge Sousa de São José, Assistant Professor, FF-UL.

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

June 2016

Genetic and biochemical characterization of hemicellulosic polysaccharides utilization in *Bacillus subtilis*

Acknowledgments

To Prof. Isabel de Sá-Nogueira, for accepting me as a PhD student and allowing me to learn from her. A lot of her love for *Bacillus subtilis* and gene regulation has most definitely passed on to me.

To Centro de Recursos Microbiológicos (CREM) and to Unidade de Ciências Biomoleculares Aplicadas (UCIBIO) from Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa for providing the necessary work conditions for the completion of this work.

To Fundação para a Ciência e Tecnologia for financial support through grants number PTDC/AGR-AAM/102345/2008 to I.S.N., PEst-OE/BIA/UI0457/2011 to Centro de Recursos Microbiológicos, UID/Multi/04378/2013 to Unidade de Ciências Biomoleculares Aplicadas, and fellowship SFRH/BD/73039/2010 to M.J.F.

To all my lab mates who simply could not have been better. Lia, we have been in this together since 2008. Thanks for all the support through these years, for all those times I conveniently used your many recently prepared stock solutions and buffers, for encouraging me to start running, for all the talk about “Lost” and “Game of Thrones” and “Dexter”… Oh, wait, not for “Dexter”. I will never forget those words: “A mulher do Dexter morre”.

Maria Isabel, even though we could not reach an agreement in about 90% of the times we were discussing something, you taught me a lot from the very beginning. So many tips and tricks and the greatest knowledge of all: when working in the lab there is no such thing as a minute waiting.

Viviana, your sarcasm is absolutely priceless. So often you could win the day with one of your timely sarcastic comments. Thank you for all those moments and also for showing me some things that will surely never be erased from my mind.

Aristides (not de Sousa) Mendes, a.k.a. Tides, after so many years begging to have someone in the lab with whom I could speak about football I got you… Can’t complain. You on the other hand were always complaining. Oh I’m tired, uh I’m hungry, I’m getting old, my back hurts. Man up! ☺

To Raquel and Damien for the foundation of 330. That legacy will always live on.

To everyone else who spent some time with me at the Microbial Genetics Lab, Joana Pedro, Joana Lima, Inês, Diana, Renato, Graça, Liliana, Maria João, Michael, in many different ways all of you are a little bit part of this Thesis.

To Rita and Márcia from the X-tal lab for all your help during these years and for sharing the so many AraN-induced headaches. Also to Teresa Santos Silva and Prof. Maria João Romão for embarking in this challenging AraN journey.

To Pedro and Prof. Eurico Cabrita for helping us to better understand the “moody” AraN.

To Prof. Carlos Fontes for allowing the use of the ITC machine at his lab.

To Larguinho, Rosa, Gonçalo, Márcia, Marco, Carla, Lia D., Bárbara, Cynthia and everyone else at DCV for your companionship throughout these years.

To all the guys from Thursday’s football matches for all the fun and relaxing moments and for calling me “Messi da Caparica” when I got a nice pass after missing nine.

To all my friends and especially to my family. Mom and dad, in so many different ways you helped shaping me into what I am today. I am so grateful for you. Hélder, like good brothers we’ve had our fights, but we were always able to leave them behind us ever since we were kids (I miss the time when you allowed me to punch you 10 times whenever you hurt me in our wrestling matches).

Finally, Ana for being loving, supporting and demanding, and for putting up with me for the past 12 years. I wouldn’t have been here without you. And Madalena, for being there to hug me after a tough day, for making all problems disappear with a simple smile, and even for those moments when I had to write just one more sentence and you would put on your sweet look and insistently say “Papá, anda brincar comigo”. How could I say no to that?

Abstract

Plant biomass is the major source of available carbohydrates in Nature. The walls of plant cells are mainly composed by homo- and heteropolysaccharides, usually found intertwined with each other and making a highly complex and heterogeneous structure, which displays a strong resistance to biological degradation. Still, many microorganisms are able to breakdown this complex structure and use the available sugars as carbon and energy source. Several saprophytic bacteria, including *Bacillus subtilis*, release a high number of saccharolytic enzymes capable of cleaving most polysaccharides in plant cell walls, and are then able to transport and metabolize the resulting mono- and oligosaccharides. Thus, transport is a key step in sugar utilization. Here, we characterize four ABC-type transporters involved in the uptake of arabino- and galactooligosaccharides, and galacturonic acid oligomers and/or rhamnose-galacturonic acid disaccharides. We show that MsmX is the sole ATPase responsible for energizing these four ABC importers, thus playing a key role in pectin mobilization by *B. subtilis*. AraN, the solute-binding protein (SBP) from the ABC importer, AraNPQ, is shown by *in vivo* and *in vitro* analyses to bind arabinooligosaccharides. However, AraN displays a significantly lower affinity than that previously reported for other SBPs. In addition, AraN also binds cello-, xylo- and maltooligosaccharides with similar affinity but is shown to be non-essential for growth in the presence of these oligosaccharides. We speculate that binding of these sugars may prevent arabinooligosaccharides uptake when more favorable carbon sources, such as glucose-rich cello- and maltodextrins, are available. The mechanisms of expression of *abn2*, a gene encoding an endo-arabinanase, were investigated and we provide evidence that this AraR-independent gene is temporally regulated. The transition state regulator AbrB and the master sporulation regulator Spo0A are implicated in the regulation of *abn2*. Finally, we demonstrate that a metallic cluster close to the catalytic site of Abn2 is essential for its arabinanase activity.

Keywords: *Bacillus subtilis*; plant cell wall sugar mobilization; ABC importer; multitask ATPase; sugar-protein interactions.

Sumário

A biomassa vegetal é a maior fonte de carbo-hidratos na Natureza. As paredes de células vegetais são compostas principalmente por homo- e heteropolissacáridos, que se encontram habitualmente interligados entre si constituindo uma estrutura altamente complexa e heterogénea, que apresenta uma grande resistência à degradação biológica. Ainda assim, muitos microorganismos conseguem quebrar esta estrutura complexa e utilizar os açúcares disponíveis como fonte de carbono e energia. Várias bactérias saprófitas, incluindo *Bacillus subtilis*, libertam um elevado número de enzimas sacarolíticas capazes de clivar a maioria dos polissacáridos das paredes de células vegetais, conseguindo depois transportar e metabolizar os mono- e oligossacáridos resultantes. Deste modo, o transporte é um passo chave na utilização de açúcares. Aqui, caracterizamos quatro transportadores do tipo ABC envolvidos no transporte de oligossacáridos de arabinose, galactose e ácido galacturónico e/ou dissacáridos de ramnose-ácido galacturónico. Demonstramos que MsmX é a única ATPase responsável pelo fornecimento de energia a estes quatro transportadores ABC, desempenhando assim um papel chave na mobilização de pectina por *B. subtilis*. É demonstrado por análises *in vivo* e *in vitro* que AraN, a proteína de ligação ao substrato (SBP) do importador ABC AraNPQ, liga arabinooligossacáridos. No entanto, AraN apresenta uma afinidade significativamente inferior em relação ao anteriormente reportado para outras SBPs. Para além disso, AraN também liga celo-, xilo- e maltooligossacáridos com afinidades semelhantes, mas não é essencial para o crescimento na presença destes oligossacáridos. Especulamos que a ligação destes açúcares pode prevenir o transporte de oligossacáridos de arabinose quando fontes de carbono mais favoráveis, como celo- e maltodextrinas, se encontram disponíveis. Os mecanismos de expressão de *abn2*, um gene que codifica uma endo-arabinanase, foram investigados e apresentamos evidências de que este gene não dependente de AraR é regulado temporalmente. O regulador do estado de transição AbrB e o regulador principal da esporulação Spo0A são implicados na regulação de *abn2*. Finalmente, demonstramos que um centro metálico perto do local catalítico de Abn2 é essencial para a sua actividade de arabinanase.

Termos chave: *Bacillus subtilis*; mobilização de açúcares da parede de células vegetais; importadores ABC; ATPase multifunções; interacções açúcar-proteína.

Contents

Figures Index

Tables Index

Abbreviations and Symbols

ABC – ATP-binding cassette

ABN – α-1,5-arabinanase

AF – α-L-arabinofuranosidase

ATP – Adenosine triphosphate

bp – Base pairs

BSA – Bovine Serum Albumin

CCR – Carbon catabolite repression

CDR – C-terminal regulatory domain

DNA – Deoxyribonucleic acid

EDTA – ethylenediaminetetraacetic acid

EGTA – ethylene glycol-bis(2-aminoethylether)tetraacetic acid

G1PDH – glycerol-1-phosphate dehydrogenase

GH – Glycosyl Hydrolase

HEPES – 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid

IPTG – Isopropyl-β-D-galactopyranoside

LB – Luria-Bertani

MF – Major facilitator

MWCO – Molecular weight cut-off

MBP – Maltose binding protein

NBD – Nucleotide-binding domain

OD – Optical density

ORF – Open reading frame

PCR – Polymerase chain reaction

PEP – Phosphoenolpyruvate

PPP – Pentose phosphate pathway

PTS – Phosphotransferase system

RBS – Ribosome binding site

RG-I – Rhamnogalacturonan type I

RG-II – Rhamnogalacturonan type II

RNA – Ribonucleic acid

RT-qPCR – Quantitative reverse transcription polymerase chain reaction

SBP – Solute-binding protein

SDR – Structurally diverse region

SDS-PAGE – Sodium dodecyl sulfate polyacrylamide gel electrophoresis

TMD – Transmembrane domain

X-Gal – 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Thesis Outline

The Thesis is divided in seven Chapters.

In Chapter I, “General Introduction”, a brief overview on the organism studied in this thesis, *Bacillus subtilis*, is given. Several concepts on the utilization of various carbohydrates by Bacilli, including sugar hydrolysis and uptake, are introduced with an emphasis on *B. subtilis* and arabinose utilization.

In Chapter II, “A multitask ATPase serving different ABC-type sugar importers in *Bacillus subtilis*”, the AraNPQ system is shown to be responsible for arabinooligosaccharides uptake. Additionally, the results presented here prove that MsmX is the ATPase responsible for energizing the AraNPQ transporter and at least two more ABC-type sugar importers.

In Chapter III, “Uptake of pectin by MsmX-dependent ABC-type importers in *Bacillus subtilis*”, the MsmX ATPase is implicated in the uptake of several different carbohydrates. New ABC-type sugar importers are functionally characterized and MsmX is established as a key player in pectin mobilization. Additionally, it is also shown that another *B. subtilis* ATPase is able to functionally replace MsmX when placed in *trans* but not when expressed from its own locus in the chromosome.

In Chapter IV, “Characterization of the sugar-binding protein AraN of *Bacillus subtilis*”, the solute-binding protein from the AraNPQ-MsmX arabinooligosaccharides ABC-type importer is studied. Sugar-protein interactions determined using STD-NMR and site-directed mutagenesis analyses are presented.

In Chapter V, “Mechanisms controlling the expression of the *abn2* gene of *Bacillus subtilis*”, the AraR-independent regulation of the gene *abn2*, encoding the endo-arabinanase Abn2, is assessed. A model for the regulation of *abn2* at the transcription level is proposed.

In Chapter VI, “The importance of the Abn2 calcium cluster in the endo-1,5-arabinanase activity from *Bacillus subtilis*”, a divalent metal cluster is shown to be critical for protonation of catalytic residues and arabinanase activity.

Finally Chapter VII, “Concluding Remarks”, contains a brief general discussion about the relevance of the results obtained.

Chapter I

General Introduction

General Introduction

An overview on *Bacillus subtilis*

Bacillus subtilis is a saprophytic, rod-shaped, endospore-forming, Gram-positive bacterium, usually found in the soil, for long considered to be its natural habitat and reservoir, from which it is often transferred to various environments such as plants and plant materials, foods, animals, and marine or freshwater habitats (Priest, 1993; Hong *et al.*, 2009). The ability of *B. subtilis* to form a highly differentiated structure, the endospore, which is able to indefinitely resist starvation, dissection and radiation, is the ultimate survival strategy and also facilitates the geographical dispersion of this microorganism in dust and air currents (Priest, 1993; Earl *et al.*, 2008). Spores of *B. subtilis* are normally found in the same ecological niches as vegetative cells, and have recently been shown to be able to germinate and grow in the gastrointestinal tract of animals, including humans, which led to the proposal of a bimodal life cycle of growth and sporulation in both the environment and gastrointestinal tracts (Tam *et al.*, 2006; Serra *et al.*, 2014). This ability has been commercially exploited and *B. subtilis* spores are currently being sold for use as probiotics, for both human consumption and as animal food supplements, with beneficial effects on several gastrointestinal disorders and on the prevention of colonization of the gastrointestinal tract by pathogenic bacteria (Casula and Cutting, 2002; Hong *et al.*, 2005).

B. subtilis plays an important role in the global carbon cycle through its ability to degrade plant cell wall polysaccharides by the concerted action of several sugar hydrolases and sugar-dedicated transporters. Of particular interest are enzymes involved in the degradation of the backbone of polysaccharides from the hemicellulosic and pectic fractions of plant biomass: arabinans (Leal and Sá-Nogueira, 2004; Inácio *et al.*, 2008; Inácio and Sá-Nogueira, 2008), xylans (Roncero, 1983; Rhee *et al.*, 2013), galactans (Shipkowski and Brenchley, 2006; Chai *et al.*, 2012), mannans (Sadaie *et al.*, 2008; Sun and Altenbuchner, 2010) and rhamnogalacturonan (Ochiai *et al.*, 2007). These enzymes may have a significant positive impact for the bioethanol industry in the saccharification process of hemicellulosic biomass. The industrial process in which wild-type *Saccharomyces cerevisiae* is used for the production of ethanol does not take full advantage of the available plant biomass. Instead, only hexoses like glucose and mannose are fermented and the remaining non-cellulosic components (mainly pentoses such as xylose and arabinose) are accountable for approximately 70% (w/w) of the total residues from bioethanol production. Although pentose-fermenting *S. cerevisiae* strains have been engineered, and are industrially employed, these strains are unable to breakdown large polysaccharides and can simply ferment monosaccharides such as L-arabinose or D-xylose (reviewed in Hahn-Hägerdal *et al.*, 2007). Moreover, the need to take advantage of lignocellulosic biomass has been increased by the ethical issues regarding the use of substrates like corn for the biofuel industry. This is likely

to further strengthen the utility of *B. subtilis* and/or its enzymes for bioethanol production using substrates which are not competing with food production (Farrel *et al.*, 2006; Numan and Bhosle, 2006; van Maris *et al.*, 2006; Nevoigt, 2008; Pauly and Keegstra, 2010).

Bacillus species, including *B. subtilis*, have a high impact in industry, not only due to their capability for naturally releasing large amounts of several commercially interesting enzymes – namely cellulases, hemicellulases, proteases, lipases or amylases – and other increased value biomolecules, such as antibiotics, but also by their use for the production of several recombinant proteins, in which human proteins are included (Arbige *et al.*, 1993; Priest, 1993). Although the recent financial and economic crisis diverted attentions from the need to address the problems caused by the still increasing global carbon emissions (Le Quéré *et al.*, 2015), a renewed demand for a more sustainable fuel source will eventually arise due to the growing environmental awareness by the general public. Therefore, the capability of *B. subtilis* to hydrolyze most of the polysaccharides in plant biomass, particularly that from agricultural and forest industry waste, is likely to turn this bacterium into an essential part of the biofuels industry and of global carbon economy. Finally, the fact that the complete genome of *B. subtilis* has been sequenced (Kunst *et al*., 1997) – and made available online (http://genolist.pasteur.fr/SubtiList/) – with essential genes having been identified (Kobayashi *et al.*, 2003), and that this bacterium is easily manipulated in the laboratory, have made *B. subtilis* a highly attractive microorganism for both fundamental and applied research, with the potential for providing a significant and positive economical impact if used as a tool for several biotechnological industrial processes.

Utilization of carbohydrates by Bacilli

Biological diversity in our planet is astonishing, with many different organisms living on it. Nevertheless, even when so many differences can be found, some features and needs are common to all life. Be it a unicellular microorganism or a multicellular higher animal, whether it can be classified as Bacteria, Archaea or Eukarya, all life on Earth depends on energy and carbon to be able to grow and reproduce. Plants, algae and photoautotrophic bacteria are able to use light as energy source and to store this energy in the form of carbohydrates through photosynthesis. In turn, these carbohydrates are used not only by the organisms which synthesize them, but also, directly or indirectly, by most of the known heterotrophic life forms. In plants, these carbohydrates can be stored and later used for various metabolic processes and also as the primary building blocks for cell wall synthesis as a plant grows. When plant cell tissues eventually die, saprophytic organisms (including some bacteria) are able to degrade these carbohydrates, by releasing a large number of saccharolytic enzymes, and use them as their carbon and energy sources (McNeil *et al.*, 1984; Thomson, 1993).

Carbohydrates in plant cell walls

All plant cells are encased by a structurally complex wall which confers rigidity to cells and some level of protection against environmental stress, such as infection by pathogens or extreme weather conditions. Additionally, walls play an important role in plant cell adhesion and cell-cell communication. The structural organization of the plant cell wall is exemplified in Figure 1.1. During cell growth and elongation, primary walls surround the cells offering structural support, then, when growth stops, cells from various plant tissues develop a thicker secondary wall which provides further strength. Primary and secondary walls are considerably different, but, in both, polysaccharides are responsible for the larger part of their dry weight. In primary walls, water content may reach up to 60% (w/v) of the total mass, but polysaccharides may be accountable for 90% of the total dry weight. On the other hand, water is virtually absent from secondary walls mostly due to the presence of the hydrophobic polyphenol lignin, and polysaccharides represent around 60% of the dry weight (McNeil *et al.*, 1984; Thomson, 1993; Caffall and Mohnen, 2009; Scheller and Ulvskov, 2010; Pettolino *et al.*, 2012; Pauly *et al.*, 2013; Loqué *et al.*, 2015).

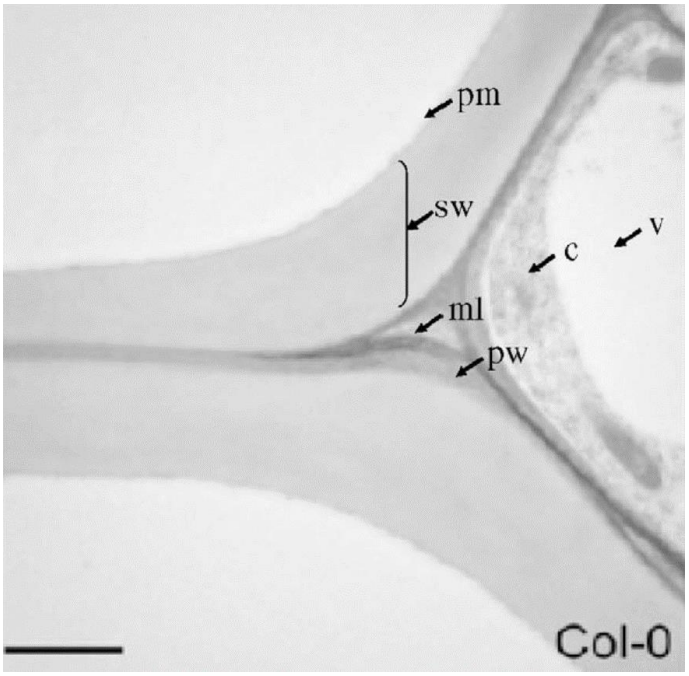


Figure 1.1 – Structural organization of the *Arabidopsis thaliana* cell wall. Transmission electron micrograph of wild-type *Arabidopsis thaliana* Columbia-O (Col-O) transverse root section showing the middle lamella (*ml*), primary wall (*pw*), and secondary wall (*sw*) of the metaxylem. The plasma membrane (*pm*), cytoplasm (c) and vacuole (v) are also indicated. Bar = 2 µm (Caffall and Mohnen, 2009).

The plant cell wall as a whole is the major reservoir of carbohydrates to be found on our planet. Its composition is highly variable, not only between species but also between different tissues of the same plant (Table 1.1). Generally, the two most abundant carbohydrates in plant biomass are glucose and xylose, but arabinose, mannose and galacturonic acid may also have a significant contribution. Even though some glycoproteins can be found as an integrant part of the

cell wall of plants, its four main components are cellulose, hemicellulose, pectin and lignin, which are intertwined with each other forming a complex mesh (Figure 1.2). Cellulose is a linear polymer made of variable molecular weight chains of β-1,4-linked D-glucosyl units which displays low solubility in water. Up to approximately 36 of these glucan chains can associate with each other, by hydrogen bonds and van der Waals interactions, to form partially crystallized microfibrils which are aligned parallel to each other and are highly resistant to hydrolysis (Figure 1.2). Cellulose is generally the most common polymer in plant cell walls (up to about 45% of the total dry weight in tobacco callus cell walls) and its relative abundance is higher in secondary walls than in primary walls. Additionally, the average molecular weight of cellulose chains and the thickness of cellulose microfibrils are much higher in secondary walls when compared to primary walls (McNeil *et al.*, 1984; van Maris *et al.*, 2006; Caffall and Mohnen, 2009; Pauly and Keegstra, 2010; Edwards and Doran-Peterson, 2012; Pauly *et al.*, 2013; Loqué *et al.*, 2015).

Table 1.1 – Composition of common agricultural lignocellulosic byproducts/leftovers. Dry weight percentage of the most important carbohydrates and lignin in some agricultural biomass (adapted from van Maris *et al.*, 2006).

	Corn stover	Wheat straw	Bagasse	Sugar beet pulp	Switch grass
D-Glucose	34.6	32.6	39.0	24.1	31.0
D-Mannose	0.4	0.3	0.4	4.6	0.2
D-Galactose	1.0	0.8	0.5	0.9	0.9
D-Xylose	19.3	19.2	22.1	18.2	0.4
L-Arabinose	2.5	2.4	2.1	1.5	2.8
Uronic Acids	3.2	2.2	2.2	20.7	1.2
Lignin	17.7	16.9	23.1	1.5	17.6

Unlike cellulose, which is exclusively composed by linear chains of β-1,4-linked glucose units, non-cellulosic polysaccharides, such as hemicelluloses and pectins, are highly heterogeneous in composition and structure. Together, both types of polysaccharides contain a high number of different sugars including hexoses (D-glucose, D-galactose, D-mannose, L-fucose, D-glucuronic acid, D-galacturonic acid, L-rhamnose) and pentoses (D-xylose, L-arabinose, D-apiose). Hemicelluloses and pectins can also be distinguished from cellulose by their higher water-solubility (Ochiai *et al.*, 2007; Caffall and Mohnen, 2009; Scheller and Ulvskov, 2010).

Hemicellulose polysaccharides can be linear or branched (with a high degree of substitution) and their backbones – which, like cellulose, typically display an equatorial configuration – are mainly homopolymeric xylans and mannans. Hemicelluloses are often interwoven with cellulose microfibrils, especially in primary walls, generally conferring increased strength to the cellulose network. Xylans are found in the walls of many plant species – especially monocots, and predominantly as part of secondary walls – and are often substituted by glucuronosyl (glucuronoxylan) and arabinosyl (arabinoxylan) units (Figure 1.2). Hemicellulosic polysaccharides with a glucose backbone highly substituted with xylosyl residues (xyloglucan) have also been reported in most studied species (Charophytes being the exception). Depending

on the tissue and plant, these xyloglucans may have additional branching (arabinose or fucose) and are predominantly composed by either a repeating oligosaccharide unit of three xylose-substituted glucose units followed by one unsubstituted glucose or by a repeating unit of two substituted units followed by two with no substitution. Xyloglucans are more abundant in the primary walls of dicots and conifers, while arabinoxylans are the major hemicellulose in monocot walls. Xyloglucan-cellulose interactions in the primary wall are thought to be very important for growing cells (Figure 1.2).

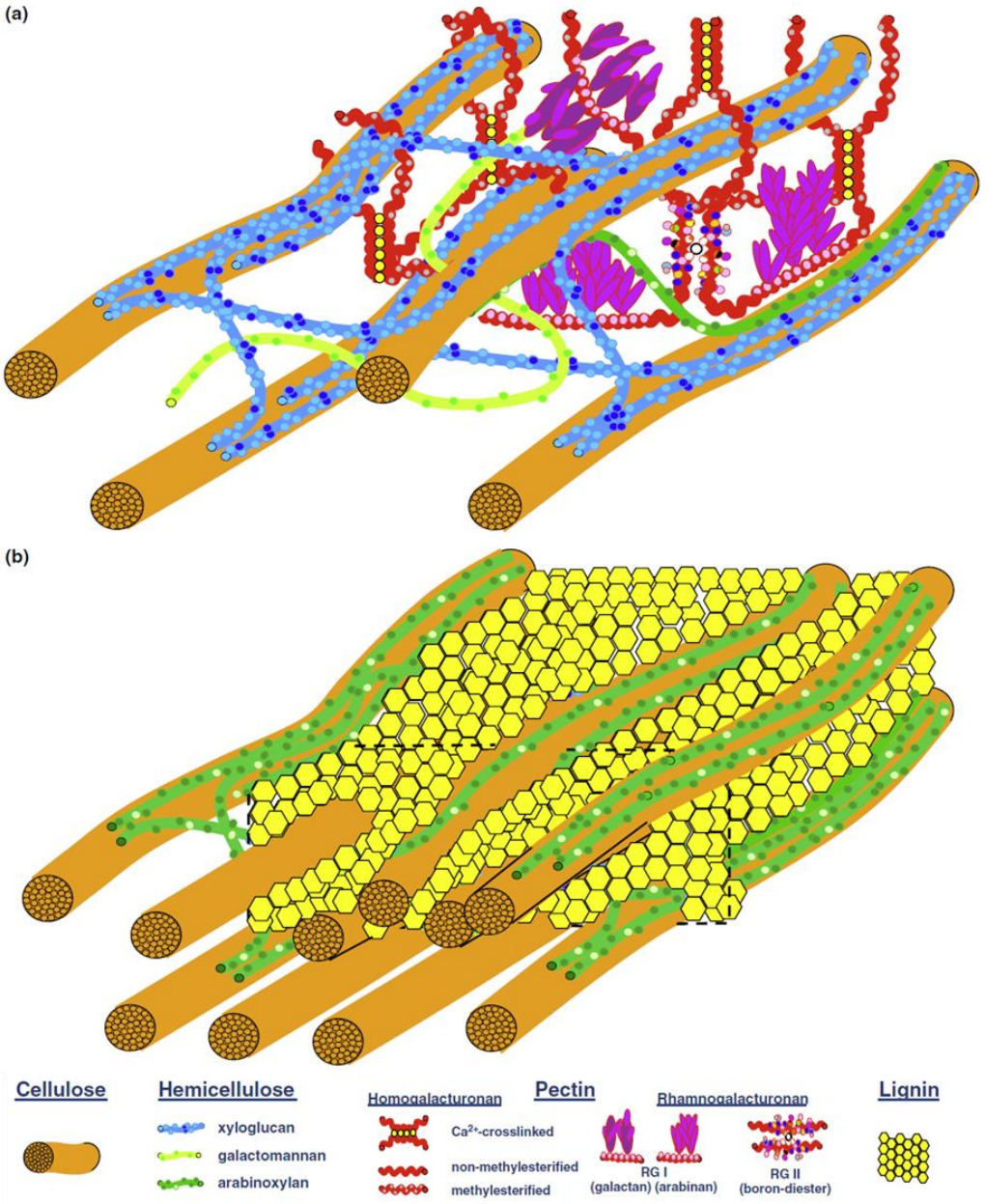


Figure 1.2 – Schematic representation of the plant cell wall composition and organization. A) The primary wall is a more heterogeneous and flexible structure, more suited to sustain the stresses of expanding cells. The cellulose network is highly tethered by mainly arabinoxylan and pectins **B)** The secondary wall, which develops in some cells after they stop growing, is more rigid due to the close association between the more abundant cellulose microfibrils and the hydrophobic lignin polymer (adapted from Loqué *et al.*, 2015).

Although these interactions between xyloglucan chains and cellulose microfibrils weaken cellulose networks in the primary wall when compared to the secondary wall, they confer a higher flexibility to expanding cell walls (Caffall and Mohnen, 2009; Scheller and Ulvskov, 2010; Pauly *et al.*, 2013; Rhee *et al.*, 2013).

Mannans are β-1,4-linked polysaccharides with a mannose backbone, often substituted with galactose (galactomannans), and constitute the main hemicellulose in Charophytes. Due to their abundance in algae, mannans are considered to have been the major hemicellulosic polysaccharide in earlier plants, having since been replaced by other hemicelluloses in modern plants. Mannans can have a mixed backbone of a non-repeating pattern of mannose and glucose (glucomannans). Although not as abundant as other hemicelluloses, (gluco)mannans have been reported to play an important role in the embryogenesis process of *Arabidopsis thaliana* (Goubet *et al.*, 2003; Scheller and Ulvskov, 2010; Pauly *et al.*, 2013).

Mixed linkage glucans are also classified as hemicellulases and can be found in the primary walls of some plant species, particularly in the Poaceae family – which includes cereal crop plants such as rice, wheat and corn – where they play an important role in wall elongation during cell growth. These glucans are generally constituted by small cellotriosyl and cellotetrasyl units linked between them by β-1,3 linkages, thus displaying a non-linear configuration in contrast with the linear configuration of β-1,4-linked cellulose chains (Scheller and Ulvskov, 2010; Pauly *et al.*, 2013).

Pectins are a class of polysaccharides, sometimes grouped together with hemicelluloses, which are characterized by its high galacturonic acid content (approximately 70% of pectin), and are often more abundant in low-lignin biomass like citrus waste and sugar beet pulp (Figure 1.3). There are three main types of pectins: homo- or polygalacturonan, rhamnogalacturonan type I (RG-I), and rhamnogalacturonan type II (RG-II), which may be covalently bound to each other in a single polysaccharide chain (McNeil *et al.*, 1984; Mohnen, 2008; Caffall and Mohnen, 2009; Edwards and Doran-Peterson, 2012).

Polygalacturonan has a linear backbone of α-1,4-linked galacturonic acid units and comprises about 65% of all pectin. Residues of this pectic polysaccharide are partially methylesterified which has implications in the crosslinking between chains of polygalacturonan. Unmethylated units are negatively charged and can interact with Ca^{2+} ions to form a stable gel-like structure with unmethylated regions of other polygalacturonan chains (Figure 1.2). Polygalacturonan backbones can also be substituted by mono- or disaccharides of D-apiose (apiogalacturonan) or D-xylose (xylogalacturonan). RG-I is the second most abundant pectic polysaccharide and is characterized by a backbone of alternating L-rhamnose and D-galacturonic acid residues. Depending on the plant or tissue RG-I may be scarcely branched or about 50% of the rhamnose units may be substituted. The branches attached to rhamnose residues are mainly pure arabinans, galactans and arabinogalactans, which can be further branched themselves (Figure 1.2). Finally, the least common and most structurally complex pectic polysaccharide is

RG-II. The backbone of RG-II is similar to polygalacturonan, i.e., is composed by α-1,4-linked galacturonic acid units. The chains of this polysaccharide usually self-associate through boron diester bonds to form RG-II dimers (Figure 1.2). The complexity of RG-II is due to the high heterogenicity of its branching side chains. The heterogeneous side chains are composed by 12 different sugars, including arabinose, galactose, rhamnose, and other rare species such as apiose, methylated xylose or fucose, and aceric acid (McNeil *et al.*, 1984; Ochiai *et al.*, 2007; Mohnen, 2008; Caffall and Mohnen, 2009; Loqué *et al.*, 2015).

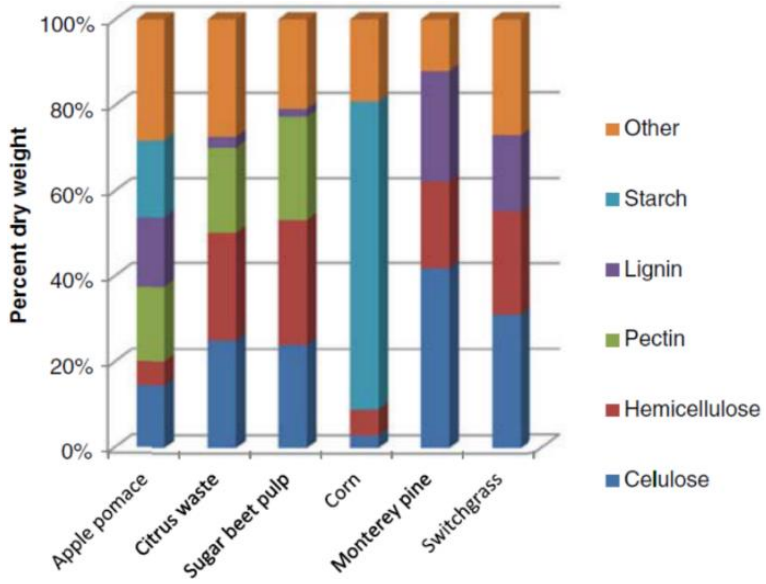


Figure 1.3 – Comparison of the dry weight composition of pectin-rich substrates to starch-rich and lignocellulosic biomasses. Pectin-rich biomass includes apple pomace, citrus waste and sugar beet pulp, which have a considerably lower lignin content, making these substrates less recalcitrant and more attractive for bioconversion to ethanol (Edwards and Doran-Peterson, 2012).

Lignin is an aromatic polymer consisting of various different phenolic units, and is typically accountable for approximately 10-30% of total biomass dry weight – although this value drops to <2% in pectin-rich biomass (Table 1.1 and Figure 1.3). This polymer is more commonly found in secondary walls where it is tightly associated with cellulose microfibrils (Figure 1.2), but can also be linked to hemicellulosic polysaccharides. Although lignin cannot be fermentable, much research on this polymer has been conducted due to its impact in the recalcitrance of plant biomass. The association between lignin and other plant polysaccharides increases the enzymatic load needed for the release of fermentable sugars for the bioethanol industry, which consequently increases the production costs. Therefore, low-lignin substrates such as the pectin-rich citrus waste or sugar beet pulp are an attractive alternative to lignocellulosic biomass for industrial processes (van Maris *et al.*, 2006; Pauly and Keegstra, 2010; Edwards and Doran-Peterson, 2012).

Carbohydrate metabolism

Being a highly complex structure composed mainly, but not exclusively, by homo- and heteropolysaccharides of many different sugars, and which are often linked to each other by covalent, hydrogen or diester bonds, the plant cell wall displays a high recalcitrance. Microorganisms, particularly saprophytes, that are able to degrade plant cell wall polysaccharides, and use the resulting sugars as their carbon and energy sources, release a wide array of enzymes (glycosyl hydrolases, GHs) which act synergistically for the complete breakdown of plant cell wall carbohydrates (Beldman *et al.*, 1997; Shallom and Shoham, 2003). These enzymes allow the utilization of most non-cellulosic sugars. Examples of the metabolic pathways/networks for the most relevant sugars present in hemicelluloses and pectins (arabinose, xylose, galactose and mannose) will be given, and include enzymes for polysaccharide breakdown, sugar uptake systems and regulation mechanisms, with focus on the known pathways in *B. subtilis*.

Arabinose/arabinan metabolism

In order to metabolize L-arabinose, which is not readily available from arabinans, arabinogalactans and arabinoxylans, *B. subtilis* depends on the concerted action of several enzymes. In this enzymatic consortium there are two α-1,5-arabinanases (ABNs, EC 3.2.1.99), that act in an endo-fashion and cleave the backbone of arabinose-containing polysaccharides, yielding arabinose and arabinooligosaccharides. In addition, two α-L-arabinofuranosidases (AFs, EC 3.2.1.55), which act in an exo-fashion on the oligosaccharides, release arabinose residues that can easily be further metabolized by *B. subtilis* (Raposo *et al*., 2004; Leal and Sá-Nogueira, 2004; Inácio and Sá-Nogueira, 2008; Inácio *et al*., 2008). An extracellular arabinoxylan arabinofuranohydrolase, XynD, which specifically releases arabinosyl units from arabinoxylan as also been characterized (Bourgois *et al*., 2007). The resulting L-arabinose is further catabolized according to the metabolic pathway identified by Lepesant and Dedonder (1967). L-arabinose isomerase converts L-arabinose to L-ribulose, which is then phosphorylated by L-ribulokinase, with the resulting L-ribulose-5-phosphate being finally converted by L-ribulose-5-phosphate 4-epimerase into D-xylulose-5-phosphate (Figure 1.4).

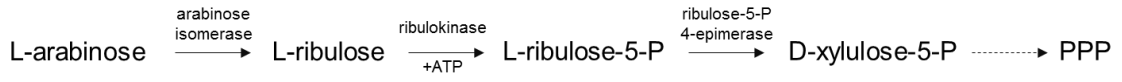


Figure 1.4 – The metabolic pathway of L-arabinose in *B. subtilis*. L-arabinose is converted to D-xylulose-5-phosphate by a sequence of reactions catalyzed by arabinose isomerase, ribulokinase and riblulose-5-phosphate 4-epimerase. The final product of this pathway enters the pentose phosphate pathway (PPP).

The latter product, D-xylulose-5-phosphate, is then incorporated and further catabolized via the Pentose Phosphate Pathway (PPP). Several other microorganisms are known to have the same metabolic pathway, namely *Escherichia coli* (Englesberg *et al*., 1962), *Lactobacillus plantarum* (Heath *et al*., 1958), *Aerobacter aerogenes* (Simpson *et al*., 1958), *Salmonella typhimurium* (Pueyo and Lopez-Barea, 1979), and *Geobacillus stearothermophilus* (Shulami *et al.*, 2011).

In *B. subtilis*, the three enzymes responsible for the sequential conversion of L-arabinose to D-xylulose-5-phosphate are encoded by the genes *araA* (L-arabinose isomerase), *araB* (L-ribulokinase) and *araD* (L-ribulose-5-phosphate 4-epimerase), which are located at about 256º in the genetic map of *B. subtilis*. The products of these genes present high similarity to the ones found in *E. coli* (52.9%, 25.7% and 57.1%, respectively; Sá-Nogueira *et al*. 1997). Additionally, *araA*, *araB* and *araD* from *B. subtilis* were shown to be functional in *E. coli* (Sá-Nogueira and Lencastre, 1989). However, two significant differences are noted in the organization of these genes in both organisms. In *E. coli* they are ordered *araBAD* and are the only genes of this arabinose operon (Englesberg *et al*., 1969), whereas in *B. subtilis* the genes are arranged as *araABD* and are the first of a nine-cistron operon, which also includes *araLMNPQ-abfA* (Sá-Nogueira and Lencastre, 1989; Sá-Nogueira *et al*., 1997). In *G. stearothermophilus* the homologous genes for the conversion of L-arabinose to D-xylulose-5-phosphate are ordered *araDBA* and are part of an operon with two additional ORFs (Shulami *et al.*, 2011).

The remaining products of the *araABDLMNPQ-abfA* (or simply *ara*) operon have been functionally characterized. AraL from *B. subtilis* was shown to have phosphatase activity towards several sugar phosphate substrates (Godinho and Sá-Nogueira, 2011) and AraM was demonstrated to function as a glycerol-1-phosphate dehydrogenase (G1PDH), which makes it the first ever reported G1PDH in Bacteria, although it's biological role is yet to be determined (Guldan *et al*., 2008). AraN has been previously proposed to be the extracellular solute-binding protein (SBP) of a putative ABC-type transport system responsible for the uptake of arabinooligosaccharides and also involved in the transport of arabinose, with AraP and AraQ being the transmembrane domains (TMDs) that make the channel through which the substrates recognized by AraN are transported (Sá-Nogueira *et al*. 1997; Inácio *et al*., 2008). The functional characterization of this transport system will be presented in Chapter II and Chapter III. The last gene of the *ara* operon encodes an intracellular α-1,5-L-arabinofuranosidase, AbfA, which is able to release L-arabinose residues from arabinooligosaccharides (Inácio *et al*., 2008). Genes from the *ara* operon which are located downstream of *araD* were shown to be non-essential for the growth of *B. subtilis* on arabinose as the sole carbon and energy source (Sá-Nogueira *et al*. 1997).

In the genome of *B. subtilis* there are at least four encoded enzymes implicated in the degradation of arabinose-containing polysaccharides. AbnA and Abn2 are two extracellular endo-α-1,5-arabinanases, belonging to the Glycosyl Hydrolase (GH) 43 family, which are able to hydrolyze the α-1,5 bonds from the backbone of arabinans. Both enzymes are encoded by

monocistronic genes, *abnA*, which is located immediately upstream of the *ara* operon, and *abn2* (formerly *yxiA*), which is found considerably far from the operon at about 340º on the genetic map of *B. subtilis* (Raposo *et al*., 2004; Leal and Sá-Nogueira, 2004; Inácio and Sá-Nogueira, 2008). AbfA (encoded by *abfA* from the *ara* operon) and Abf2 (encoded by the monocistronic *abf2*, formerly named *xsa*, located about 23kb downstream of the operon) are two intracellular α-L-arabinofuranosidases from the GH 51 family with exo-activity towards small oligosaccharides and also the side chains of arabinose-containing polysaccharides. Contrary to that observed with the ABNs, these two AFs hydrolyze different bonds. While AbfA is predominantly more active towards α-1,5 bonds of arabinan and arabinooligosaccharides, Abf2 acts preferentially on α-1,2 and α-1,3 bonds of branched arabinan and arabinoxylan, which suggests the concerted action of these enzymes as being essential for the complete degradation of arabinose-containing polysaccharides (Inácio *et al*., 2008).

Located at about 298º on the genetic map of *B. subtilis*, the *araR* locus has two monocistronic and divergent transcription units (*araR* and *araE*). The *araR* gene encodes a transcription factor, AraR, which is responsible for the negative regulation of five transcription units involved in the utilization of arabinose and arabinose-containing polysaccharides (Figure 1.5). The N-terminal region of this protein comprises a helix-turn-helix consensus signature sequence of the GntR family of bacterial regulators (Haydon and Guest, 1991), while the larger C-terminal region presents a high similarity to the bacterial regulators of the LacI/GalR family (Weickert and Adhya, 1992). The *araE* gene encodes an integral membrane permease belonging to the Major Facilitator (MF) superfamily, AraE, which is involved in the H^+ symport-coupled uptake of L-arabinose (Sá-Nogueira and Ramos, 1997), D-xylose and D-galactose by *B. subtilis* (Krispin and Allmansberger, 1998).

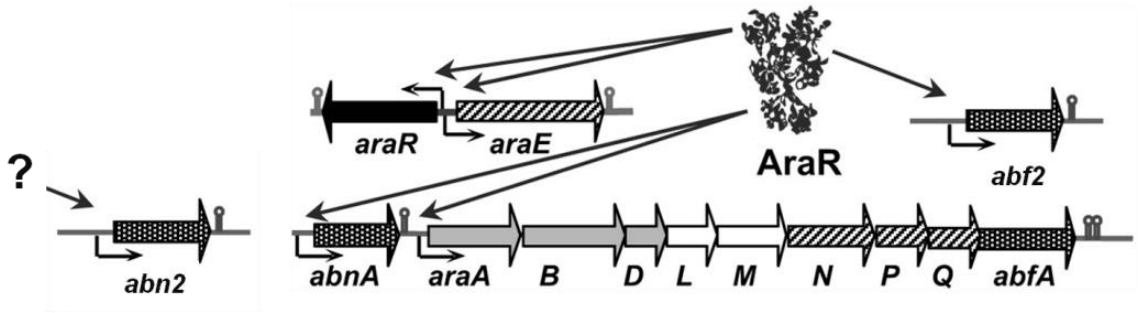


Figure 1.5 – Representation of the arabinose (*ara*) regulon of *B. subtilis*. The AraR repressor is responsible for the negative regulation of five transcription units related to the utilization of arabinose by *B. subtilis*. The expression of *abn2* is not under the control of AraR and mechanisms responsible for mediating its transcription are not known (adapted from Franco *et al*., 2006).

The regulation of genes involved in arabinose utilization at the transcription level is mainly achieved through the AraR transcription factor. Among all six transcription units known to be related with arabinose utilization, *abn2* is the only not under the control of AraR (Figure 1.5) and

its regulatory mechanism is still unknown (Inácio and Sá-Nogueira, 2008). The repression of *araABDLMNPQ-abfA*, *araE*, *araR*, *abnA* and *abf2* expression is achieved by specific binding of AraR homodimers to palindromic operator sequences (Figure 1.6) located in the promoter regions of these five units. The binding of AraR to DNA is accomplished by two distinct mechanisms. In the promoter region of the *ara* operon, and the *araE* and *abf2* genes there are two in-phase operators that allow the formation of a DNA loop upon cooperative-binding of two AraR homodimers, leading to a highly repressed transcription. A looser control is observed for *abnA* and *araR* since both genes display a single operator sequence in their promoter regions, leading to an increased basal level of expression (Mota *et al*., 1999; Mota *et al*., 2001; Raposo *et al*., 2004). The expression of the *ara* regulon is achieved in the presence of arabinose. When L-arabinose binds to AraR the protein changes its conformation, thus becoming unable to bind DNA (Mota *et al*., 1999; Franco *et al*., 2006; Franco *et al*., 2007).

Figure 1.6 – Position weight matrix of the *B. subtilis* AraR-binding site. The consensus palindromic sequence of the AraR-binding site is At(T/A)tGTaCGTAcaa(A/T)T (adapted from Correia *et al.*, 2014).

In addition to the regulation mediated by AraR, the arabinose system of *B. subtilis* is also controlled by a Carbon Catabolite Repression (CCR) mechanism. Several *cre* sites, associated to *araE*, *abnA*, *abn2*, *abf2* and the *ara* operon, allow the binding of the *trans*-acting repressor protein CcpA (Henkin *et al*., 1991) in the presence of glucose. Consequently the expression of the *ara* genes is repressed even when arabinose is available. This regulation mechanism allows the cells to preferentially utilize the more favorable glucose as its carbon and energy source (Inácio *et al*., 2003; Inácio and Sá-Nogueira, 2007; Inácio and Sá-Nogueira, 2008).

The *G. stearothermophilus* arabinan utilization system, although functionally similar to that of *B. subtilis*, is of increased complexity and has a clearly distinct gene organization. At least 23 genes of *G. stearothermophilus*, organized in five transcription units, were proposed to be involved in the utilization of arabinose and arabinose-containing polysaccharides: *araPST*, *araEGH*, *araR*, *araDBA-abp-abnB*, and *abnEFJA-abfBA-araJKLMN*. In contrast to that observed in *B. subtilis*, these transcription units are clustered together in a relatively small 38 Kb segment of its chromosome (Shulami *et al.*, 2011). The genes encoding the three enzymes responsible for the conversion of L-arabinose to D-xylulose-5-phosphate are ordered *araDBA*. The breakdown of arabinan or arabinose-containing polysaccharides is achieved by the concerted action of five enzymes. One extracellular endo-arabinanase (AbnA) and an intracellular endo-arabinanase (AbnB), both belonging to the GH43 family, cleave the backbone of arabinan or large arabinooligosaccharides. AbfA and AbfB, two intracellular GH51 α-L-arabinofuranosidases, and

Abp, a GH27 β-L-arabinopyranosidase, complete the breakdown of arabinose oligomers (Shulami *et al.*, 2011). Sugar uptake is achieved through two distinct ABC-type transporters. AraEGH and AbnEFJ are responsible for the uptake of arabinose and arabinooligosaccharides, respectively. In the AraEGH system, AraE is the SBP which recognizes arabinose, AraH is the single protein responsible for making the transporter pore, and AraG serves as the ATPase which energizes this system. In the AbnEFJ transporter, the recognition of arabinose oligomers is accomplished by the SBP AbnE which delivers its substrates to the channel formed by AbnF and AbnJ, the two transmembrane proteins. AraG was also proposed to be able to energize the AbnEFJ system (Shulami *et al.*, 2011). Additionally, a NAD(P) sugar dehydrogenase (AraJ), an aldose-1-epimerase (AraK), a sugar phosphatase (AraL), a glycerol-1-phosphate dehydrogenase (AraM) and a hypothetical protein (AraN) were proposed to be involved in an alternative pathway for arabinose utilization, previously identified in *Azospirillum brasiliense* (Watanabe *et al.*, 2006; Shulami *et al.*, 2011).

The *ara* genes from *G. stearothermophilus* are regulated by two different mechanisms. A unique three component regulatory system is encoded in the *araPST* operon, which is constitutively expressed at low levels. AraP is an extracellular lipoprotein responsible for recognizing arabinose which, when bound to this sugar, activates the histidine sensor kinase AraS. In turn, AraS phosphorylates the response regulator AraT which then activates the transcription of the adjacent arabinose ABC transporter genes, *araEGH* (Shulami *et al.*, 2011). In addition, a self-regulated negative regulator, AraR, is functionally similar to its *B. subtilis* counterpart and is responsible for the repression of the remaining *ara* genes in the absence of arabinose. Finally, a putative binding site for the catabolite repressor CcpA was identified only in the promoter region of the putative *abnEFJA-abfBA-araJKLMN* operon (Shulami *et al.*, 2011).

Xylose/xylan metabolism

In several *Bacillus* spp., including *B. subtilis*, the metabolic pathway for xylose begins with the conversion of D-xylose to D-xylulose by action of a xylose isomerase. A xylulokinase then phosphorylates D-xylulose and the resulting D-xylulose-5-phosphate is further metabolized by entering the PPP (Figure 1.7). The enzymes responsible for the sequential conversion of D-xylose to D-xylulose-5-phosphate are encoded by *xylA* (xylose isomerase) and *xylB* (xylulokinase), which are organized as an operon located at about 162º in the chromosome of *B. subtilis* (Gärtner *et al.*, 1988; Scheler *et al.*, 1991; Rygus *et al.*, 1991). Alternative pathways for xylose metabolism have also been described (Weimberg, 1961; Dahms, 1974; Stephens *et al.*, 2007). Several strains of *B. subtilis* are unable to grow in the presence of xylose as the sole carbon and energy source due to the lack of a dedicated xylose transporter (Lindner *et al.*, 1994). However, xylose uptake can be carried out by the arabinose H^+-symporter AraE (Sá-Nogueira and Ramos, 1997; Krispin and Allmansberger, 1998). Since both sugars are commonly found together, xylose uptake in the natural habitats of *B. subtilis* is unlikely to be a hindering factor in its ability to metabolize xylose.

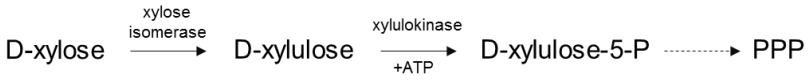


Figure 1.7 – The metabolic pathway of D-xylose in *B. subtilis*. D-xylose is converted to D-xylulose-5-phosphate by a sequence of reactions catalyzed by xylose isomerase and xylulokinase. The final product of this pathway enters the pentose phosphate pathway (PPP).

Encoded in the genome of *B. subtilis* are several enzymes capable of completely hydrolyzing the backbone of homoxylans. Two extracellular endo-xylanases, belonging to GH11 (XynA, EC 3.2.1.8) and GH30 (XynC, EC 3.2.1.136) families, cleave the β-1,4-linkages between the xylose units of the xylan backbone, thus releasing smaller oligosaccharides (Rhee *et al.*, 2013). An intracellular β-xylosidase (EC 3.2.1.99), XynB from GH43 family, displays exo-activity and is able to release xylosyl residues from the non-reducing end of xylans and xylooligosaccharides (Roncero, 1983). The presence of two extracellular endo-xylanases and an intracellular xylosidase suggests that an unknown transporter is involved in the uptake of xylooligosacharides, possibly one of eight ABC-type importers proposed to be involved in carbohydrate uptake (Quentin *et al.*, 1999).

Expression of the *xylAB* operon and *xynB* is controlled at the transcription level by the XylR repressor protein and mediated by the presence of D-xylose. XylR recognizes and binds to palindromic operator sequences in the promoter regions of the *xyl* operon, *xynB* and its own gene, *xylR*, thus blocking their expression. When xylose becomes available, the sugar interacts with the repressor and XylR releases the DNA which, in turn, leads to transcription from the units belonging to the *xyl* regulon (Hastrup, 1988; Gärtner *et al.*, 1988; Kreuzer *et al.*, 1989). Additionally, these genes are regulated by the global carbon catabolite repressor CcpA and have a decreased expression in the presence of glucose (Kraus *et al.*, 1994; Galinier *et al.*, 1999). XylR- and CcpA-binding sequences were not identified in the promoter regions of the genes encoding both endo-xylanases, *xynA* and *xynC*, and their expression is not induced by the presence of xylose nor repressed when glucose is available. Instead, both genes are apparently constitutively expressed at low levels (St. John *et al.*, 2006).

The *G. stearothermophilus* xylan utilization system resembles that of *B. subtilis* but some differences are worth noting. Similarly to its arabinan utilization system, the *G. stearothermophilus* xylan utilization genes are all clustered together in a relatively small region of the chromosome with approximately 40 Kb (Shulami *et al.*, 2007; Shulami *et al.*, 2014). In this bacterium, the xylose isomerase and xylulokinase responsible for the conversion of D-xylose to D-xylulose-5-phosphate are encoded by the *xylAB* genes, which are organized as an operon (Shulami *et al.*, 2014). Like *B. subtilis*, *G. stearothermophilus* possesses two endo-xylanases which act on the backbone of xylans and release xylooligosaccharides. However, only one of these xylanases, XynA, is extracellular, while the other, XynA2, is intracellular (Teplitsky *et al.*, 2000; Shulami *et al.*, 2007). Additionally, *G. stearothermophilus* has three intracellular β-xylosidases – XynB1, XynB2 and

XynB3 belonging to GH families 39, 52 and 43, respectively – responsible for cleaving xylooligosaccharides and releasing xylose monomers (Bravman *et al.*, 2001; Bravman *et al.*, 2003; Czjzek *et al.*, 2005; Brüx *et al.*, 2006). Interestingly, a *Paenibacillus* sp. strain was shown to possess two endoxylanases similar to those from *G. stearothermophilus* (extra- and intracellular), and only a single β-xylosidase like *B. subtilis* (Sawhney *et al.*, 2015).

An ABC-type importer is responsible for the uptake of xylooligosaccharides in *G. stearothermophilus* and its protein components are encoded by the *xynEFG* genes. XynE is the SBP that recognizes xylooligosaccharides and XynFG are the TMDs making the channel through which the substrates enter the cell (Shulami *et al.*, 2014). An ABC system dedicated to the uptake of xylose oligomers was also found in *Paenibacillus* sp. (Sawhney *et al.*, 2015). The genes for the *G. stearothermophilus* ABC transporter are cotranscribed with two other and are organized as the *xynDCEFG* operon, although a putative promoter between *xynC* and *xynE* has also been identified, and transcription from this promoter may occur (Shulami *et al.*, 2007; Shulami *et al.*, 2014).

The *xynDC* genes encode a two-component regulatory system in which XynD is the sensor kinase and XynC serves as the response regulator. The XynDC system is responsible for the activation of transcription from the *xynE* promoter when xylose becomes available. XylR is the major xylose repressor that, in the absence of this sugar, represses transcription from five cistrons involved in xylose/xylan utilization, including the *xynDCEFG* and *xylAB* operons and the monocistronic *xynA*. CcpA was also shown to repress the extracellular endo-xylanase gene *xynA* in the presence of glucose (Shulami *et al.*, 2007; Shulami *et al.*, 2014).

Galactose/galactan metabolism

The Leloir pathway for galactose catabolism is highly conserved in all organisms, including humans. In this pathway four enzymatic reactions are needed for the conversion of β-D-galactose to glucose-1-phosphate (Holden *et al.*, 2003). In the first step of the Leloir pathway (Figure 1.8) a galactose mutarotase catalyzes the conversion of β-D-galactose into α-D-galactose, which is then phosphorylated by a galactokinase, resulting in galactose-1-phosphate. The third enzyme of this pathway, galactose-1-phosphate uridylyltransferase, transfers a UMP group from UDP-glucose to galactose-1-phosphate, which simultaneously generates glucose-1-phosphate and UDP-galactose. Finally, UDP-glucose consumed in the previous step is regenerated from UDP-galactose by action of a UDP-galactose-4-epimerase. Glucose-1-phosphate generated in the third step can be further metabolized by entering glycolysis (Holden *et al.*, 2003). Via an alternative pathway controlled by several repressors and anti-repressors, including the master sporulation regulator Spo0A, UDP-galactose is shunted towards the synthesis of exopolysaccharides, thus playing a critical role in biofilm formation by *B. subtilis* (Hamon and Lazazzera, 2001; Chai *et al.*, 2012).

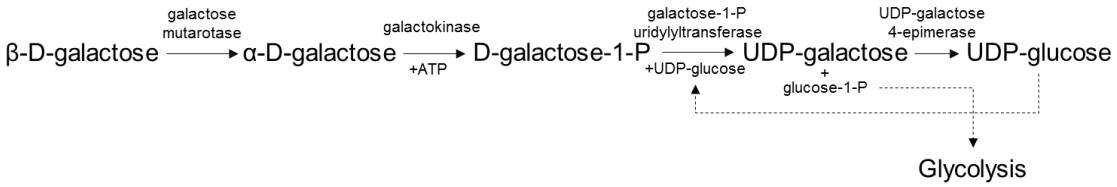


Figure 1.8 – The metabolic pathway of D-galactose in *B. subtilis*. D-galactose is converted to glucose-1-phosphate by a sequence of reactions catalyzed by galactose mutarotase, galactokinase, galactose-1-phosphate uridylyltransferase and UDP-galactose-4-epimerase. The final product of this pathway enters the glycolytic pathway.

In *B. subtilis*, the genes encoding the enzymes of the Leloir pathway are *galM* (galactose mutarotase or aldose-1-epimerase), *galK* (galactokinase), *galT* (galactose-1-phosphate uridylyltransferase), and *galE* (UDP-galactose-4-epimerase) (Duez *et al.*, 2009; Chai *et al.*, 2012). *galM* is found at about 171º in the chromosome of *B. subtilis* and is organized as the *galM-dacC* operon. The second gene of this operon, *dacC*, encodes a penicillin-binding protein with DD-endo-peptidase activity towards peptidoglycan cross-bridges (Duez *et al.*, 2009). The galactokinase and galactose-1-phosphate uridylyltransferase are encoded in the last two genes of the *ywcC-gtcA-galKT* operon, which is located at about 335º in the chromosome. The other genes of this operon encode a TetR-type repressor (*ywcC*) and a putative glycosyltransferase (*gtcA*) (Chai *et al.*, 2012). Finally, *galE* is a monocistronic gene found at about 341º in the chromosome.

A putative galactan utilization operon, *cycB-ganPQAB*, is located at about 300º in the chromosome. *cycB-ganPQ* encode the components of an ABC-type importer. CycB is the SBP – shown to interact with cyclodextrins (Kamionka and Dahl, 2001) – and GanPQ are the TMDs in this system. Although CycB interacts with cyclodextrins, this transporter is likely to be involved in the uptake of galactooligosaccharides (see Chapter III). *ganA* and *ganB* encode an intracellular β-galactosidase (EC 3.2.1.23, GH42) and an extracellular endo-β-1,4-galactanase (EC 3.2.1.89, GH53), respectively (Daniel *et al.*, 1997; Shipkowsky and Brenchley, 2006). GanB probably cleaves the backbone of galactans and arabinogalactans releasing smaller oligosaccharides, which are then transported inside the cell through CycB-GanPQ. Once inside the cell, GanA completes the breakdown by releasing galactosyl residues from galactooligosaccharides (Shipkowsky and Brenchley, 2006; Chai *et al.*, 2012). The *cycB-ganPQAB* operon is negatively regulated by the LacI-like repressor protein LacR (Daniel *et al.*, 1997). The gene encoding this negative regulator, *lacR*, lies just upstream of the operon. It has been suggested that galactooligosaccharides and not galactose may act as the inducer of this operon by binding to LacR (Shipkowsky and Brenchley, 2006; Chai *et al.*, 2012). In fact, a *G. stearothermophilus* galactan utilization operon was shown to be induced in the presence of galactooligosaccharides but not galactose (Tabachnikov and Shoham, 2013).

In other polysaccharide utilization systems, like those previously described for arabinan and xylan, *G. stearothermophilus* possesses a single mega-cluster containing all genes for each system. In contrast, the *G. stearothermophilus* galactan utilization system is more closely related to that of *B. subtilis* and the genes for galactan breakdown and galactose metabolism are encoded by two unrelated and distant operons (Tabachnikov and Shoham, 2013). Interestingly, several other Bacilli have most of the genes involved in galactose/galactan utilization clustered together. In *B. licheniformis*, *B. amyloliquefaciens* and *L. acidophilus* the enzymes for the Leloir pathway are encoded by genes clustered together with those for galactan utilization or even as part of the same operon. Also, in *B. amyloliquefaciens* and *L. acidophilus* the sugar transporters are from different classes, a phosphotransferase system (PTS) and a permease, respectively (Chai *et al.*, 2012).

Mannose/mannan metabolism

Although mannose is not as abundant as other non-cellulosic carbohydrates, *B. subtilis* is able to grow using this hexose as the sole carbon and energy source (Sun and Altenbuchner, 2010). To achieve this, *B. subtilis* possesses two gene clusters involved in the utilization of mannose (Sun and Altenbuchner, 2010) and (gluco)mannans (Sadaie *et al.*, 2008). The mannose (*man*) operon *manPA-yjdF*, located at about 109º in the *B. subtilis* chromosome, encodes a PTS mannose-specific EIIBCA component (ManP), a mannose-6-phosphate isomerase (ManA), and the uncharacterized protein YjdF, which was shown to be non-essential for mannose utilization (Sun and Altenbuchner, 2010). ManP is responsible for the uptake of D-mannose concomitantly with its phosphorylation. ManA then catalyzes the conversion of the resulting D-mannose-6-phosphate to D-fructose-6-phosphate which can be further metabolized by entering the glycolytic pathway (Figure 1.9).

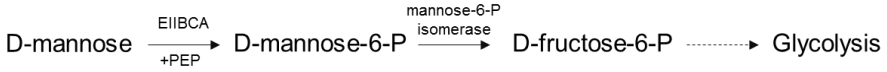


Figure 1.9 – The metabolic pathway of D-mannose in *B. subtilis*. D-mannose is converted to D-fructose-6-phosphate by a sequence of reactions catalyzed by the EIIBCA component of a PTS and mannose-6-phosphate isomerase. The final product of this pathway enters the glycolytic pathway.

Upstream and adjacent to the *man* operon is the *manR* gene which encodes a transcriptional regulator. This regulator, ManR, functions as a transcriptional activator to the *man* operon and *manR*. In addition to an N-terminal DNA-binding domain, ManR possesses two PTS regulation domains (PRDs), which become phosphorylated in turn, depending on the presence or absence of mannose, consequently leading to ManR activation or inactivation, respectively (Joyet *et al.*, 2013; Wenzel and Altenbuchner, 2013). The *man* operon is also under indirect glucose-mediated repression. A CcpA-binding site in the promoter region of the *manR* gene leads

to repression of the transcriptional activator in the presence of glucose and, consequently, to a reduced expression of the *man* operon genes (Sun and Altenbuchner, 2010). Located at about 54º in the *B. subtilis* chromosome lies the *gmuBACDREFG* (or simply *gmu*) operon involved in glucomannan utilization (Sadaie *et al.*, 2008). The first three ORFs encode enzymes IIB (GmuB), IIA (GmuA), and IIC (GmuC) of a mannooligosaccharide-specific PTS transporter. The next gene of the *gmu* operon encodes the 6-phospho-β-glucosidase GmuD (EC 3.2.1.86, GH1). *gmuR* codes for a negative regulator, responsible for the repression of the *gmu* operon in the absence of mannooligosaccharides. *gmuE* and *gmuF* encode a fructokinase and a mannose-6-phosphate isomerase, respectively. Finally, the last gene of this operon codes for the extracellular endo-1,4-β-mannanase GmuG (EC 3.2.1.78, GH26), which acts on the backbone of (gluco)mannans releasing smaller oligosaccharides (Sadaie *et al.*, 2008). These oligosaccharides are transported and phosphorylated by GmuBAC and further cleaved by GmuD. In addition to negative regulation by GmuR, expression of the *gmu* operon is also repressed in the presence of glucose by binding of CcpA to two *cre* sites. The first of these sites is located in the operon promoter region and the second is located within the operon itself in the *gmuC* gene (Sadaie *et al.*, 2008).

Bacillus subtilis transport systems

All cells and subcellular compartments are isolated from the external medium by lipidic membranes. Thus membrane transporters are essential in cell physiology. Transporters enable the uptake of nutrients, ions and metabolites, allow the expulsion of toxic compounds, the release of extracellular proteins and end products of metabolism, participate in energy generation and interconversion and also facilitate cell-cell and cell-environment communication (Higgins, 1992; Saier *et al*., 2002a).

The utilization of carbohydrates and carbohydrate-containing polysaccharides by *B. subtilis* relies on the existence of several systems that allow the uptake of these substrates to the cytoplasm, making them available for metabolism, providing energy or being used in the synthesis of biological relevant molecules, such as amino acids or antibiotics (Saier *et al*., 2002a). Since *B. subtilis* possesses a large number of extracellular endo-carbohydrases, which release short sugar oligosaccharides from longer polysaccharides, having multiple different transport systems capabile of transporting mono- and oligosaccharides provides a significant competitive advantage in comparison to other microorganisms that can only uptake monosaccharides.

The Different Classes of Transport Systems

Transport systems of living organisms can be classified as belonging to one of four different major types of transporters (Figure 1.10). Channel proteins (Figure 1.10a) allow the transport of

solutes through an aqueous pore or channel by an energy-independent process without evidence for a carrier-mediated mechanism. Membrane channels are gated, opening or closing in response to voltage (Ren *et al.*, 2001; Davidson *et al.*, 2008). Although these systems are typically energy-independent, a few examples of energized channel-mediated transport are known, in which solute movement is against a large concentration gradient (reviewed in Saier *et al.*, 2002b). In prokaryotes, the number of this type of transport systems is relatively low when compared to what is observed in eukaryotes where they are found in large numbers. The *B. subtilis* genome codes for as few as six channel proteins which are involved in potassium ion translocation, glycerol uptake, and osmoregulation (Saier *et al.*, 2002b).

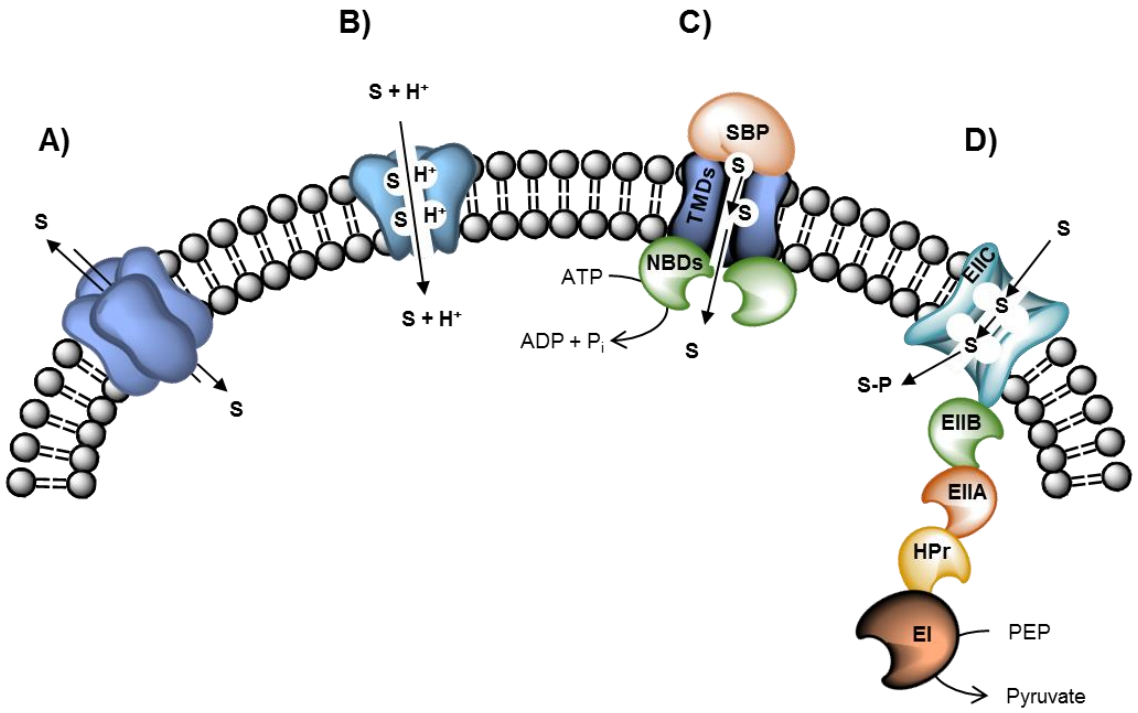


Figure 1.10 – Schematic representation of the four major types of transport systems in living organisms. A) Cannels (usually oligomeric). A voltage-gated ion channel (VIC) family member is represented. The tetrameric channel allows solute (S) to flow freely across the membrane with no energy required. **B)** Secondary active transporters (usually monomeric or dimeric). A major facilitator (MF) superfamily porter is depicted. A functional hetero- or homodimeric carrier makes use of both domains for solute and cation recognition. The majority of secondary carriers has both domains fused in a single polypeptide chain. A solute-proton symport mechanism is portrayed, but solute uniport, solute-proton antiport and solute-solute antiport are also observed in secondary transport systems. **C)** Primary active transporters (usually multidomain and multicomponent). An ATP-binding cassette (ABC) superfamily uptake permease is illustrated. The functionally hetero- or homodimeric pump allows active transport of solute into or out of the cell against a concentration gradient. A single extracellular receptor (the solute-binding protein, SBP) recognizes a specific solute and feeds it into the dimeric membrane channel (the transmembrane domains, TMDs) and the transport into the cell is energized by ATP hydrolysis, catalyzed by the dimeric ATPase (the nucleotide-binding domains, NBDs). **D)** Group translocators (always multidomain). A phosphoenolpyruvate (PEP) dependent phosphotransferase system (PTS) superfamily group translocator is shown. The functional homodimeric membrane transporter (EIIC) is energized by phosphoryl transfer reactions sequentially catalyzed by EI, HPr, EIIA and EIIB. The sugar substrate is phosphorylated during its transport (adapted from Saier *et al.*, 2002a).

Secondary active transporters, a group that includes the *B. subtilis* arabinose H^+-symporter AraE, use chemiosmotic energy in the form of transmembrane ion or solute electrochemical gradients to drive transport across the cytoplasmic membrane. *B. subtilis* possesses close to 200 recognized secondary active transporters belonging to 42 different families. The largest of this families is the Major Facilitator (MF) superfamily (Figure 1.10b) – includes roughly one third of all secondary active transporters – of which AraE is a part of (Pao *et al.*, 1998; Saier *et al.*, 2002b). Characterized MFs are involved in sugar uptake, drug efflux, glycerol-3-phosphate uptake, and nitrite and nitrate uptake or efflux. Solute movement of these transporters is energy coupled and typically associated with H^+-symport (uptake) or H^+-antiport (efflux). As a whole, secondary active transporters are involved in the transport of many different solutes, such as sugars, amino acids, organic and inorganic ions, nucleotides and nucleosides, and drugs (Saier *et al.*, 2002b).

Primary active transporters, which include ABC-type transporters (Figure 1.10c), namely AraNPQ, use chemical, electrical or solar energy to drive transport against a concentration gradient. The majority of chemically driven active transporters is energized through ATP hydrolysis, but decarboxylation or methyltransfer – in some bacteria and archaea, respectively – can drive uptake or extrusion of solutes via other systems. *B. subtilis* has 78 predicted ABC transporters (38 importers and 40 exporters) which are mainly involved in the uptake of carbohydrates, oligopeptides, and several micronutrients, drug efflux, and osmoregulation (Quentin *et al.*, 1999). Oxidoreduction and light absorption are specifically used for ion pumping mediated by primary active transporters (Saier *et al.*, 2002b). A more in-depth structural and functional description of ABC sugar importers is given below.

Finally, group translocating systems (Figure 1.10d), such as the bacterial phosphotransferase systems (PTS), transport and concomitantly phosphorylate their sugar substrates using phosphoenolpyruvate (PEP) as both the phosphate donor for sugar phosphorylation and the energy source for sugar accumulation. These systems, which couple transport and metabolic functions, were only recently described in Archaea and are yet to be identified in eukaryotes (Pickl *et al.*, 2012; Cai *et al.*, 2014; Saier, 2015). PTS transporters belong to four different superfamilies and, although they do not show high sequence conservention among them, all display structural homology (McCoy *et al.*, 2015). Bacterial PTS systems are involved in the uptake of several metabolically relevant sugars like glucose, fructose, lactose and mannose. Phosphorylation associated to sugar uptake is indicative of intracellular carbohydrate levels and has also been proposed to prevent the efflux of the sugar back through the permease (McCoy *et al.*, 2015). Although their main function is the transport and phosphorylation of sugar molecules, many cellular functions, such as nitrogen and phosphorous utilization, biofilm formation, and virulence, are directly or indirectly regulated by PTS systems, specifically via its EIIA component. Some transcription factors, carbohydrate catabolic enzymes and non-PTS permeases may also be functionally controlled by these systems (Lengeler and Jahreis, 2009; Saier, 2015).

ABC-type transporters

The ATP-binding cassette (ABC) systems are best known for their ability to couple the hydrolysis of ATP to the uptake or expulsion of several different solutes, which led to their initial nomenclature of traffic ATPases. All ABC systems dedicated to solute translocation share a similar general architecture: two transmembrane domains (TMDs) that allow the import or export of solutes, and two nucleotide-binding domains (NBDs) that couple the hydrolysis of ATP to transport (Figure 1.11).

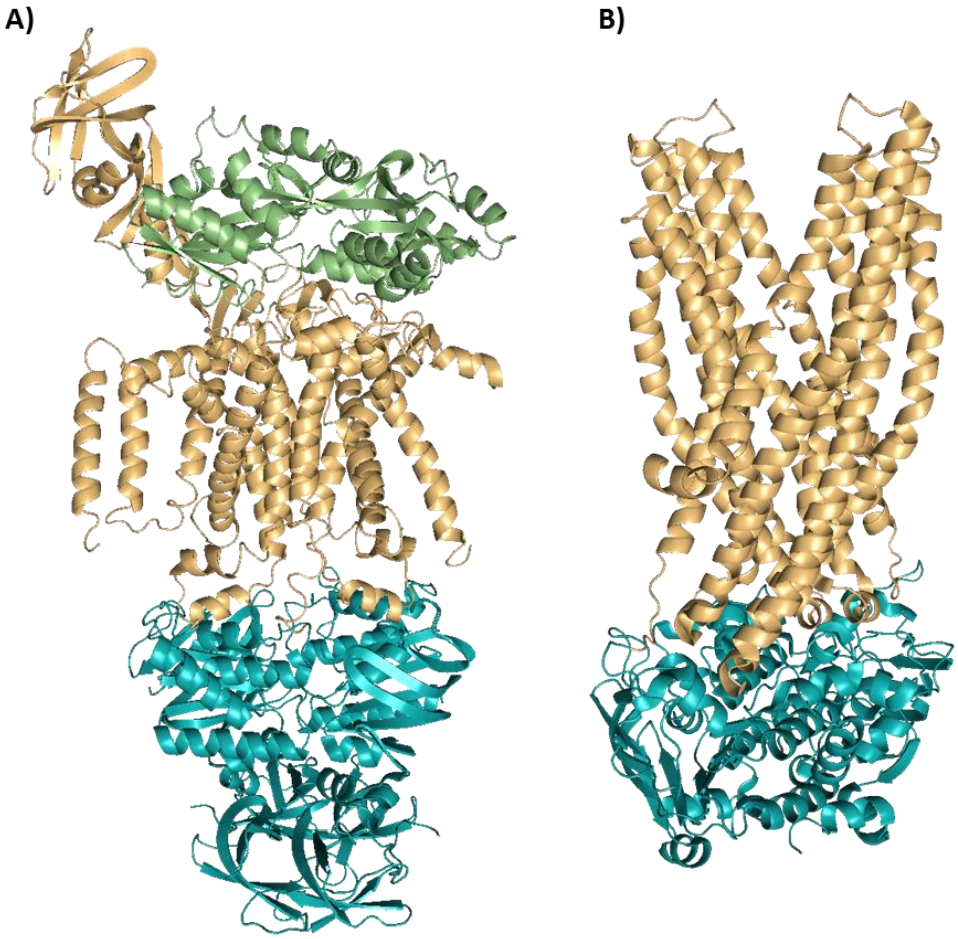


Figure 1.11 – The architecture of ABC-type transporters. Importers and exporters are similarly characterized by two NBDs (*blue*) and two TMDs (*orange*). **A)** The *E. coli* maltose transporter (Oldham *et al.*, 2007; PDB accession number: 2R6G) has heterodimeric TMDs and homodimeric NBDs and an additional solute-binding protein for substrate recognition (*green*). **B)** The *Staphylococcus aureus* multidrug exporter Sav1866 is a homodimeric transporter in which the NBDs are fused to the TMDs (Dawson and Locher, 2006; PDB accession number: 2HYD).

However, ABC systems are not exclusively involved in transport and are associated to a substantial variety of cellular processes like translation regulation or DNA repair (Davidson *et al.*, 2008). ABC-type systems are essential for antibiotic and antifungal resistance in microorganisms and, in humans, they are known to be associated to several genetic diseases, including cystic

fibrosis or Tangier disease, and also to drug resistance evidenced by cancer cells (Higgins, 2001). The nucleotide-binding domain (NBD), also known as the ATP-binding domain, is responsible for the hydrolysis of ATP and characterizes these systems. ABC systems are found in all species of all three domains of life (Archaea, Bacteria and Eukarya) and show an impressive conservation in their NBDs and in the organization of constitutive domains or subunits (Higgins, 1992; Davidson *et al*., 2008). The number of single ABC systems is well correlated with a given species genome size. However, while the ABC-count per genome-size ratio is similar between bacteria and archaea it is considerably lower in eukaryotes. Among bacteria, intracellular parasites like *Mycobacterium tuberculosis* have a relatively low number of ABC transporters, which can be related to its intracellular lifestyle, where available nutrients are less complex. In contrast, bacteria such as *Agrobacterium tumefaciens* and *Mesorhizobium loti* have a very high number of ABC systems, which is probably related to the highly competitive soil and plant nodule environments where these species are found (Davidson *et al*., 2008).

ABC systems can be divided in three classes, which are thought to have already been present in the last universal common ancestor (LUCA) of the three domains of life (Davidson *et al.*, 2008). Class 1 ABC systems are mostly exporters with fused NBDs and transmembrane domains (TMDs), which are rare in Bacteria and virtually absent in Archaea, but represent the majority of ABC systems in eukaryotes (Davidson *et al*., 2008). The systems from Class 2 are present in all domains of life and are involved in various processes such as translation elongation, DNA repair or resistance to antibiotics. ABC systems belonging to this class are not associated to the plasma membrane and are therefore found in the cytoplasm. They are characterized by two tandemly-repeated NBDs and by the absence of TMDs (Davidson *et al*., 2008). Class 3 ABC systems are almost exclusive of prokaryotes and are mostly importers that rely on a solute-binding protein (SBP), which typically displays high-affinity and high-specificity to their substrates. The NBDs and TMDs of these systems are carried by independent polypeptide chains (Davidson *et al*., 2008). These binding-protein dependent systems were for a long time considered to be exclusive to Gram-negative bacteria, however they have since been shown to be widely present in Gram-positive bacteria as well (Eitinger *et al.*, 2011). Some Class 3 importers have an additional C-terminal regulatory domain (CRD), which is associated to the NBDs. CRDs have a substrate binding site which becomes occupied when the internal concentration of the substrate reaches an upper threshold. When this occurs, the NBDs are kept separated, inhibiting ATP hydrolysis and substrate transport itself (ter Beek *et al.*, 2014). Other Class 3 systems with no associated SBP have been described and are thought to be related to drug and antibiotic resistance or lipoprotein trafficking mechanisms (Davidson *et al*., 2008).

The *E. coli* maltose transporter $MalEFGK_2$ is, by far, the most studied ABC-type sugar importer (Eitinger *et al.*, 2011). The available structural, biochemical and biophysical data on this transporter led to the proposal of a generally accepted model for oligosaccharide uptake by ABC importers (Oldham *et al.*, 2007; Davidson *et al.*, 2008; Eitinger *et al.*, 2011; Oldham and Chen 2011). According to this model, known as the alternating access model (Figure 1.12), when

maltose is not present $MalFGK_2$ is in a resting state in which the TMDs (MalF and MalG) are in an inward-facing conformation, i.e., open to the cytoplasm and closed to the periplasm. In this resting state, ATP can be bound to the NBDs (MalK dimer) but is not hydrolyzed due to the distance between the catalytic sites of both ATPases. When maltose becomes available and binds to MalE (the maltose-binding protein) closure of the NBD dimer is promoted by the interaction of the closed maltose-SBP complex with the TMDs (pre-translocation). Then ATP binds to the NBDs, which simultaneously promotes the positioning of ATP to the catalytic site for hydrolysis, reorients the TMDs to an outward-facing conformation and opens the maltose-binding protein. At the same time, maltose is delivered to the binding pocket in one of the TMDs (MalF). Finally, hydrolysis of both ATP molecules brings the transporter back to an inward-facing resting state and maltose is released to the cytoplasm (Oldham and Chen 2011). This model is based on the assumption that MalE only interacts with the transporter when bound to maltose or maltodextrins but it has been suggested that the maltose-free SBP can be permanently associated to the transport complex (Eitinger *et al.*, 2011).

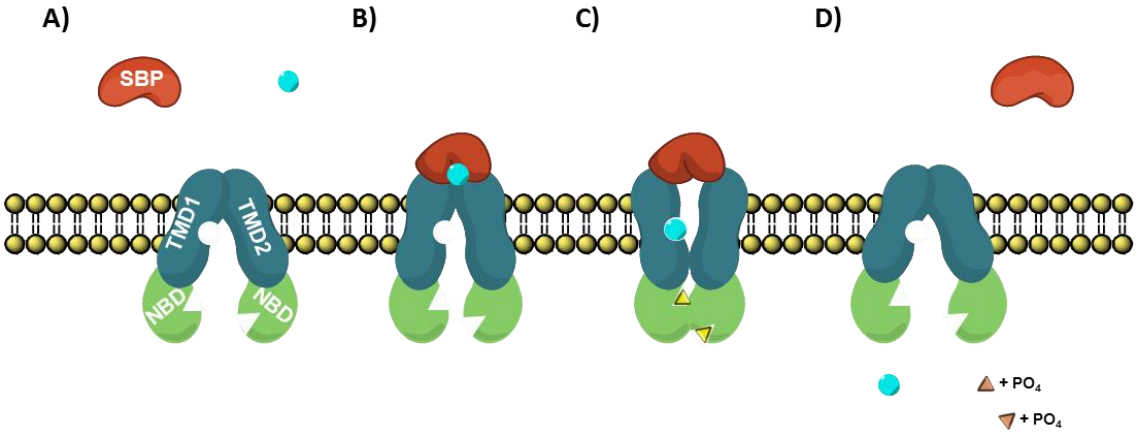


Figure 1.12 – Alternating access model for substrate uptake by ABC-type importers. A) Inward-facing resting state. When substrate (*cyan circle*) is not bound to the SBP, this protein remains in an open conformation (and free in the periplasm of Gram-negative bacteria). The transporter complex displays an inward-facing conformation and the NBSs are separated from each other (ATP molecules may bind but hydrolysis does not occur). **B) Pre-translocation.** The substrate-bound SBP changes to a closed conformation (and binds to the transporter complex in Gram-negative bacteria). This also promotes a conformational change in the transporter complex which brings the NBDs closer together. **C) Outward-facing.** When ATP (*yellow triangles*) binds to the NBDs, an additional conformation change switches the orientation of the TMDs from inward-facing to outward-facing, and the transfer of the substrate from the SBP to the recognition site of the TMDs. This conformation change also allows the formation of the NBD dimer which promotes ATP hydrolysis. **D) Post-translocation.** ATP hydrolysis returns the transporter complex to its initial inward-facing resting state, since ADP (*orange triangles*) is unable to maintain the outward-facing conformation. The substrate is then allowed to enter the cell and the SBP returns to an open conformation (and to the periplasm in Gram-negative bacteria).

Being strongly related to the utilization of multiple carbohydrates by many microorganisms, including pathogenic bacteria such as *Streptococcus mutans* (Russel *et al*., 1992), understanding ABC transport mechanisms may be of extreme relevance for therapeutic purposes. Also, the ability to transport most of the sugars available from plant biomass may be exploited for the

creation of more efficient ethanologenic species that can take full advantage of the available substrates.

The Solute-Binding Protein (SBP)

In Gram-positive bacteria, the SBP of ABC-type importers is often a lipoprotein bound to the external side of the cytoplasmic membrane by N-terminal acyl-glyceryl cysteines. Sometimes the SBP is found fused to the membrane transporter, contrary to what is observed in Gram-negative bacteria where the SBP is a soluble protein located in the periplasmic space between the inner and outer membranes. In some archaea the SBP can be attached to the surface of the membrane via an N-terminal hydrophobic helix (Eitinger *et al.*, 2011). All SBPs which have been structurally characterized are arranged as two globular lobes (named N- and C-lobes since they contain the N- and C-terminal amino acids), connected by a flexible hinge of one or more polypeptide chains, with the substrate binding pocket between these lobes (Davidson *et al*., 2008).

When SBPs are not bound to their substrates, the protein is in an open configuration and the N- and C-lobes are separated, leaving the binding pocket exposed. Solute binding occurs exclusively to one of the lobes – the side of the binding pocket with exposed side chains of aromatic amino acids – and promotes closure of the SBP (Sharff *et al.*, 1992). Closure of the binding protein is essential for solute uptake (Davidson *et al*., 2008). Binding proteins of ABC sugar importers typically display high-affinities to their substrates, with K_d values in the range of 0.1-10 µM (Nataf *et al.*, 2009; Shulami *et al.*, 2007 and 2011; Berntsson *et al.*, 2010; Tabachnikov and Shoham, 2013). However, K_d values 100-fold higher have been reported for solute-SBP interactions in *B. subtilis*. The CycB binding protein was shown to bind cyclodextrins with associated K_d values ranging from 100-500 µM but it is not clear that, in this case, the interacting substrates are effectively transported via the ABC system to which CycB belongs (Kamionka and Dahl, 2001).

Most of the SBP-dependent ABC importers are specific for a single substrate or a family of closely related substrates, such as in the transport system for maltose and maltodextrins in *E. coli*. Nonetheless, some of these importers are more versatile and the SBP displays a broad specificity to several different substrates, as is observed in the Msm transport system of *S. mutans*, which is involved in the uptake of melibiose, sucrose, raffinose and isomaltotriose (Russel *et al*., 1992; Tao *et al*., 1993). In addition, in other systems such as the histidine, lysine and arginine transport system of *S. typhimurium* (Higgins and Ames, 1981), different SBPs, with specificity to different substrates, were shown to be able to associate to the same transporter complex.

The Transmembrane Domains (TMDs)

All ABC-type transporters have two hydrophobic TMDs, each of them comprising several α-helices that span the membrane. In *B. subtilis* each TMD ranges from four to 12 of these membrane-spanning helices, and those from systems involved in sugar uptake typically display six (Quentin *et al.*, 1999). In most cases, the arrangement of TMDs creates an aqueous pore, through which highly hydrophilic substrates, such as sugars or inorganic ions, are transported. Some other transporters have been reported as being able to transport hydrophobic substrates by a lipid-filled pathway (Davidson *et al.*, 2008; Poelarends and Konings, 2002). In ABC sugar importers, the two TMDs can either be encoded by the same gene, as illustrated by the MglAC galactose transporter of *E. coli*, or by two distinct genes, homologous but non-identical, as exemplified by the maltodextrins transporter of *E. coli* (Nikaido and Hall, 1998; ter Beek *et al.*, 2014) or the arabinooligosaccharide transporter of *B. subtilis* (Sá-Nogueira *et al*., 1997; Ferreira and Sá-Nogueira, 2010). Although sequence conservation between TMDs is considerably low, most ABC importers, including those dedicated to oligosaccharide uptake, share a conserved motif known as the EAA loop or coupling helix (Quentin *et al.*, 1999; Eitinger *et al.*, 2011). The EAA loop is characterized by two short α-helices – positioned in the last cytoplasmic-side loop between the membrane-spanning helices (Oldham *et al.*, 2007) – with the conserved consensus sequence EAAx_3Gx_9IxLP (Saurin *et al.*, 1994; Mourez *et al.*, 1997; Davidson *et al.*, 2008). These conserved motifs were shown to be inserted in a surface cleft of the NBD (Figure 1.13) where they establish multiple interactions with amino acids from the conserved Q loop (Daus *et al.*, 2007; Oldham *et al.*, 2007).

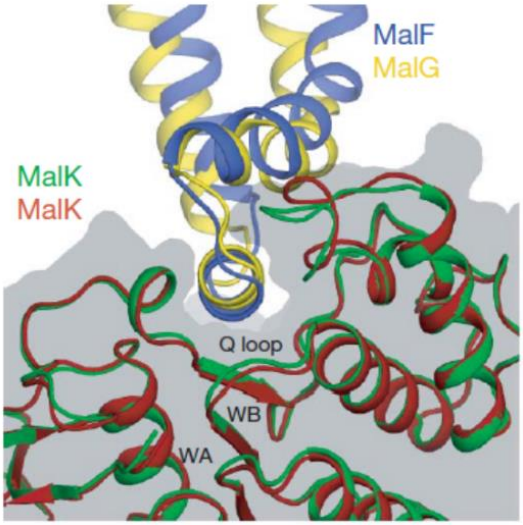


Figure 1.13 – TMD-NBD interactions in the *E. coli* maltose importer. The EAA loop from each transmembrane domain is inserted in a surface cleft of one of the NBDs where it establishes contacts with several amino acids from the Q loop (adapted from Oldham *et al.*, 2007).

The interactions between the EAA loops from both TMDs and the Q loops from the NBD dimer were proposed to be responsible for communicating ATP coupling and hydrolysis to the other components of the transporter complex (Locher *et al.*, 2002; Daus *et al.*, 2007). In the *E. coli* maltose transporter the EAA motifs are not the only sites able to contact the NBDs, since a C-terminal tail from one of the TMDs, MalG, was shown to be inserted along the interface between the MalK dimer, where it establishes several interactions with amino acids from both NBD proteins (Oldham *et al.*, 2007).

In addition to allowing the transport of substrates across the cell membrane, TMDs from ABC importers also play a role in substrate specificity. The *E. coli* maltose importer can only transport maltodextrins with up to seven glucosyl units (maltoheptaose), even though its SBP (MalE) can recognize and bind longer oligosaccharides (Davidson *et al.*, 2008; Eitinger *et al.*, 2011). A binding cleft in one of the membrane domains, MalF, was shown to be able to interact with three glucosyl residues from the non-reducing end of maltoheptaose when the transporter is in an outward-facing conformation (Oldham *et al.*, 2013). However, in the absence of the SBP the transporter complex is unable to undertake maltose uptake. This is consistent with the conformation changes promoted in the complex by the sugar bound SBP, which are mandatory for ATP hydrolysis and substrate translocation (Davidson *et al.*, 2008).

SBP-independent mutants of the *E. coli* maltose transporter and the *S. typhimurium* histidine transporter, which were shown to be able to stimulate ATP hydrolysis and transport their substrates, are thought to be able to undertake conformational changes without needing the energy from the coupling of the binding protein to the transporter complex (Petronilli and Ames, 1991; Davidson *et al.*, 1992; Davidson *et al.*, 2008). Although the ATP hydrolysis rate is unaffected in these mutants, substrate specificity was reduced by 1000-fold (Panagiotidis and Shuman, 1998).

The Nucleotide-Binding Domains (NBDs)

The ATPases are the main common components of all ABC systems and, like the TMDs, are always present as dimers that can be either a homo- or heterodimer, such as the ATPases of the $MalEFGK_2$ maltose transporter of *E. coli* or the OppDFBC oligopeptide transporter of *S. typhimurium*, respectively (Nikaido and Hall, 1998). These domains are the energizing components of ABC systems and, in transporters, are responsible for catalyzing ATP hydrolysis and transferring that energy to the other components of the system (ter Beek *et al.*, 2014). For some time it was considered that each NBD would correspond to a single ABC system. However, it has since been proposed and shown that a single ATPase is responsible for interacting with and energizing multiple ABC transporters, exclusively those involved in solute uptake and particularly sugar importers (Quentin *et al.*, 1999; Ferreira and Sá-Nogueira, 2010; Marion *et al.*, 2011; Tan *et al.*, 2015).

All ABC ATPases contain two subdomains connected by two flexible loop regions: a larger RecA-like motor domain, also found in other P-loop ATPases, and a smaller α-helical domain characteristic of this family of proteins (Davidson *et al.*, 2008; Eitinger *et al.*, 2011; ter Beek *et al.*, 2014). Their sequence is highly conserved in all classes of ABC systems and is characterized by several short sequence motifs: the Walker A and B motifs, the LSGGQ signature (also known as the C loop), the A loop, the H loop (also called the switch region), the D loop, and the Q loop (Davidson *et al.*, 2008; Figure 1.14).

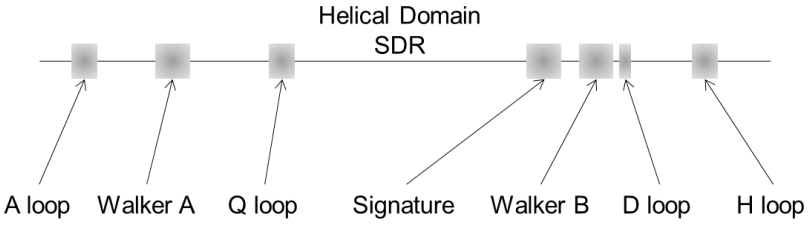


Figure 1.14 – Conserved motifs in the NBDs. The A loop has a conserved aromatic residue which helps in the positioning of ATP. The Walker A and B motifs are responsible for the formation of the ATP binding pocket. The signature motif, which is unique to ABC ATPases, also contacts ATP. The Q loop is involved in the NBD-TMD interactions and has a highly conserved glutamine which is important in the formation of the active site. The H loop has a highly conserved histidine residue that allows contact with the γ-phosphate of ATP, the Walker B motif and the D loop. The D loop contacts with Walker A of the other NBD monomer and contributes to the formation of the active site. The structural diverse region (SDR) is responsible for the interactions between the NBDs and the TMDs (adapted from Davidson *et al.*, 2008, and ter Beek *et al.*, 2014).

Walker A is characterized by the conserved GxxGxGK(S/T) sequence motif in which the lysine residue is involved in interactions with β- and γ-phosphate of ATP. Walker B has the ϕϕϕϕDE (where ϕ is a hydrophobic amino acid) motif in which the glutamate residue polarizes the attacking water molecule. The signature motif, with the highly conserved LSGGQ sequence, is found in the α-helical domain and is exclusive to the ABC ATPases superfamily. The side chain of the serine and the backbone amide groups of the glycine residues in the LSGGQ motif coordinate the γ-phosphate. The A loop contains a conserved aromatic residue (typically a tyrosine) to help in the positioning of ATP. The D loops of both NBD monomers are found alongside each other in the dimer structure and participate in the formation of the catalytic site. The H loop is characterized by a highly conserved histidine residue closer to the C-terminal region of the NBDs. This histidine residue interacts with the aspartate from the D-loop, the glutamate residue of the Walker B motif, and the γ-phosphate of ATP. Finally, the Q loop has a conserved glutamine residue that is important for the formation of the catalytic site and also participates in the binding between NBDs and TMDs (Davidson *et al.*, 2008; ter Beek *et al.*, 2014). The NBD dimer is structurally arranged in a way that makes ATP hydrolysis occur in a cooperative manner. ATP is trapped between the Walker A motif from one monomer and the LSGGQ signature sequence from the other (Oldham *et al.*, 2007; Eitinger *et al.*, 2011). This feature only allows ATP hydrolysis when the NBD dimer is formed, thus preventing hydrolysis by isolated monomers. The hydrolysis of ATP changes the conformation of the NBDs, which consequentially changes the

conformation of the TMDs, thus allowing transport to occur. The SBP also has an enhancing effect on ATP hydrolysis in ABC importers, contributing to an effective mechanism that couples transport and ATP hydrolysis (Davidson *et al*., 1992; Oldham *et al.*, 2011).

Chapter II

A multitask ATPase serving different ABC-type sugar importers in *Bacillus subtilis*

Most of the work presented in this chapter was published in “Ferreira M. J. and Sá-Nogueira I. (2010). A Multitask ATPase Serving Different ABC-Type Sugar Importers in *Bacillus subtilis*. *J. Bacteriol.* 192: 5312–5318”. All the experimental work presented in this chapter was performed by the author of this Thesis, except the construction of some plasmids.

A multitask ATPase serving different ABC-type sugar importers in *Bacillus subtilis*

Abstract

Bacillus subtilis is able to utilize arabinopolysaccharides derived from plant biomass. Here, by combining genetic and physiological analyses we characterize the AraNPQ importer and identify primary and secondary transporters of *B. subtilis* involved in the uptake of arabinosaccharides. We show that the ABC-type importer AraNPQ is involved in the uptake of α-1,5-arabinooligosaccharides, at least up to four L-arabinosyl units. Although this system is the key transporter for α-1,5-arabinotriose and α-1,5-arabinotetraose, the results indicate that α-1,5-arabinobiose is also translocated by the secondary transporter AraE. This broad-specificity proton symporter is the major transporter for arabinose and also is accountable for the uptake of xylose and galactose. In addition, MsmX is shown to be the ATPase that energizes the incomplete AraNPQ importer. This study assigns MsmX as a multipurpose *B. subtilis* ATPase required to energize different saccharide transporters, the arabinooligosaccharide-specific AraNPQ-MsmX system and the previously characterized maltodextrin-specific MdxEFG-MsmX system.

Introduction

Transport across biological membranes is a fundamental process for life and is accomplished by channels, primary and secondary transporters, and group translocators (Saier, 2000). ATP-binding cassette (ABC) transporters constitute one of the largest families of translocation facilitators and are distributed across all domains of life. Although they are involved in a variety of distinct processes, such as nutrient uptake, resistance to antibiotics and other drugs, lipid trafficking, cell division, sporulation, immune response, and pathogenesis, all ABC transporters, regardless of the polarity of transport (exporters and importers), share a structure and a general mechanism (reviewed in Davidson *et al.*, 2008, and Higgins, 1992). Their organization comprises two transmembrane domains (TMDs) coupled to two cytosolic nucleotide-binding domains (NBDs), or ATP-binding cassettes, responsible for ATP binding and hydrolysis-driven conformational changes.

The majority of the eukaryotic ABC transporters are exporters facilitating translocation from the cytoplasm. In contrast, prokaryotic ABC permeases are involved mainly in the import of nutrients, vitamins, and trace elements (Davidson *et al.*, 2008; Eitinger *et al.*, 2011; Higgins, 1992). Canonical bacterial ABC importers are dependent predominantly on high-affinity substrate-

binding proteins (BPDs) that capture the substrate and deliver it to the transporter. The canonical maltose/maltodextrin importer $MalEFGK_2$ of *Escherichia coli*/*Salmonella enterica* serovar Typhimurium (Ehrmann *et al.*, 1998) is one of the most-characterized members of this superfamily of translocation facilitators, serving as model for ABC importers in general (Oldham *et al.*, 2007; Oldham *et al.*, 2008). In Gram-negative bacteria, BPDs are proteins located in the periplasmic space between the inner and outer membranes. In Gram-positive organisms, which are devoid of this type of cellular compartment, BPDs often are lipoproteins that are anchored to the extracellular side of the cytoplasmic membrane via its N-terminal domain (Davidson *et al.*, 2008; Higgins, 1992).

In nature, the major source of carbohydrates for microorganisms to utilize is plant biomass. Thus, in their natural habitat, such as soil, aquatic environments, or animal digestive tracts, bacteria secrete a vast number of polysaccharolytic enzymes for the degradation of plant-derived polysaccharides. The resulting mono-, di-, and oligosaccharides enter the cell mainly through specific ABC transporters. The number of well-characterized ABC transporters devoted to the uptake of products resulting from the degradation of hemicellulose is very scarce (Eitinger *et al.*, 2011), and in Gram-positive organisms only two systems, BxlEFG of *Streptomyces thermoviolaceus* (Tsujibo *et al.*, 2004) and XynEFG (Shulami *et al.*, 2007) of *Geobacillus staerothermophilus*, both dedicated to the transport of xylodextrins, have been characterized in detail.

An *in silico* analysis of the *Bacillus subtilis* genome estimated the existence of at least 78 ABC transporters based on the identification of 86 NBDs in 78 proteins, 103 TMD proteins, and 37 BPD proteins, which account for about 5% of the protein-coding genes of this model organism (Quentin *et al.*, 1999). At least 10 ABC systems are predicted to be involved in the uptake of sugars (Saier *et al.*, 2002a). One of these ABC importers, AraNPQ, is clustered together with genes encoding enzymes involved in arabinose catabolism and the degradation of arabinooligosaccharides in a large operon, *araABDLMNPQ-abfA* (Sá-Nogueira *et al.*, 1997). AraN is the BPD, and AraP and AraQ are the TMDs. This transporter, which lacks the NBD protein partner, was proposed to be involved in the uptake of arabinose oligomers mainly by genomic context and *in silico* analysis (Inácio *et al.*, 2008; Inácio and Sá-Nogueira, 2008; Sá-Nogueira *et al.*, 1997).

Here, by combining genetic and physiological analyses, we characterize the AraNPQ importer and identify primary and secondary transporters of *B. subtilis* involved in the uptake of arabinooligosaccharides. Furthermore, this study assigns the role of MsmX as a multipurpose *B. subtilis* ATPase required to energize different saccharide transporters.

Materials and Methods

Substrates

α-1,5-arabinooligosaccharides (arabinobiose, arabinotriose and arabinotetraose), arabinan and debranched arabinan (sugar beet) were purchased from Megazyme International Ireland. L-arabinose, maltotriose, maltotetraose and maltopentaose were purchased from Sigma-Aldrich Co. D-(+)-glucose was purchased from BDH.

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 or New England Biolabs and used according to the manufacturers' recommendations. PCR amplifications were carried out using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) or NZYDNAChange DNA Polymerase (NZYTech). DNA from agarose gels and PCR products were purified with the illustra™ GFX™ PCR DNA and Gel Band Purification kit (GE Healthcare Life Sciences). All DNA ligations were performed using T4 DNA Ligase (Thermo Fisher Scientific). DNA phosphorylation was performed using T4 Polynucleotide Kinase (Thermo Fisher Scientific) and DNA dephosphorylation with FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific). Plasmids were purified using the QIAGEN Plasmid Midi Kit (QIAGEN), QIAprep Spin Miniprep Kit (QIAGEN) or NZYMiniprep kit (NZYTech). DNA sequencing was performed with the ABI PRISM BigDye® Terminator Cycle Sequencing Kit (Applied Biosystems). The sequencing reaction was purified by gel filtration and resolved in an ABI 3730XL sequencer.

Construction of plasmids and strains

To construct a *B. subtilis msmX* insertion-deletion mutation by the insertion of a chloramphenicol resistance (Cm^R) cassette, *msmX* was amplified by PCR from chromosomal DNA of *B. subtilis* $168T^+$ using primers ARA422 and ARA423 (Table 2.1), and the amplification product was digested with SacI and KpnI. The resulting 1360 bp DNA fragment was cloned in the phagemid pBluescript II KS(+) (Table 2.2), also digested with SacI and KpnI, yielding pMJ1. A Cm^R cassette was obtained by the digestion of pMS38 (Table 2.2) with HindIII and EcoRI and was subcloned in pMJ1 digested with the same restriction enzymes, resulting in pMJ2. This plasmid was linearized with ScaI and used to transform *B. subtilis* $168T^+$, yielding strain IQB495 (Table 2.3). To create *B. subtilis* mutant strains with in-frame deletions of *araN* and *araNPQ*, plasmids pMJ6 and pMJ11 were constructed. These plasmids were generated using the pMAD plasmid (Arnaud *et al.*, 2004; Table 2.2). Regions immediately upstream and downstream of *araN* were amplified by two independent PCR experiments from chromosomal DNA of *B. subtilis* $168T^+$

using primers ARA426 and ARA427 (PCR1) and ARA428 and ARA429 (PCR2). The products from PCR1 and PCR2 were joined by overlapping PCR with primers ARA426 and ARA429 (Table 2.1), and the resulting fragment was digested with ScaI and BamHI and cloned into pMAD, digested with SmaI and BamHI, yielding pMJ6.

Table 2.1 – List of all oligonucleotides used during the course of this work.

Oligonucleotide	Sequence (5'→3')	Complementary sequence
ARA422	TATTGAGCTCTCAGGGATAGATATCAAATCG	*msmX*
ARA423	CATCGGTACCATTCTGAGATTTTCAATAGC	*msmX*
ARA426	GCGGATCCCTTTGGTGACATGCTCGG	*araN*
ARA427	CGCTTTCACTTTTTGAATGGGCTGCGTTACCCCTTCCATTCGAGGTGCGGG	*araN*
ARA428	CCCGCACCTCGAATGGAAGGGGTAACGCAGCCCATTCAAAAAGTGAAAGCG	*araN*
ARA429	GCTTAGTACTGCTCTTTCGGCACATTTTGC	*araN*
ARA448	GATAAAGTACTTTTCGAAAAAAGTCATTTTTTTCATCTGCGTTACCCCTTC	*araN*
ARA449	GAAGGGGTAACGCAGATGAAAAAAATGACTTTTTTCCAAAAGTACTTTATC	*araQ*
ARA450	GGCATGCGCATGTTTGAGCTGCCGCAGGC	*araQ*
ARA467	CCAGCAATATTTTATCC	*araE*
ARA488	CAAATTTCGGTACCTCACAGCG	*araE*
ARA714	GTGTTCATCGATTCAACTCC	*cycB*
ARA779	GCTTTCTCTTTATTTTCCGCTGTGTGCCATTTTCATTGATTGTCACCC	*cycB*
ARA780	GGGTGACAATCAATGAAAATGGCACACAGCGGAAAATAAAGAGAAAGC	*cycB*
ARA781	TCTAACGCGTAGAAAGCCAAG	*cycB*
ARA718	ATGAGAGCTCTGCAGCATATGATCTGC	*galK*
ARA719	TTTGGTACCAGGCCATACTCAAGCAGC	*galK*

Restriction sites in the primer sequences are underlined.

Using an identical technique, regions immediately upstream of *araN* and downstream of *araQ* were amplified with primers ARA426 and ARA448 (PCR3) and ARA449 and ARA450 (PCR4) and were joined by overlapping PCR with primers ARA426 and ARA450 (Table 2.1). The resulting product was phosphorylated and then cloned into pMAD digested with SmaI and dephosphorylated, yielding pMJ11. These two plasmids harboring in-frame deletions of *araN* and *araNPQ*, respectively, were used in separate experiments for the integration and generation of clean deletions in the *B. subtilis* chromosome following the published procedure described in Arnaud *et al*., 2004. Both in-frame deletions were confirmed by DNA sequencing, and the resulting strains were named IQB496 and IQB611, respectively (Table 2.3). To construct an *araE*-null mutant by disruption with the insertion of a kanamycin resistance (Km^R) cassette, the *araE* gene was amplified by PCR from chromosomal DNA of *B. subtilis* 168T^+ using primers ARA467 and ARA488 (Table 2.1). The amplification product was digested with SspI and XmnI, and the resulting 1445 bp DNA fragment was cloned into pBluescript II KS(+) digested with SmaI, yielding pMJ7. A Km^R cassette was obtained by the digestion of pJL3 (Table 2.2) with XbaI and EcoRI. Plasmid pJL3 is a pLitmus29 derivative carrying a Km^R cassette from pAH248 (A. O. Henriques, unpublished data) cloned between the BamHI and PstI sites. The cassette was filled-in using Klenow Fragment (Thermo Fisher Scientific), and the resulting 1504 bp DNA fragment was subcloned into pMJ7 digested with Eco47III and dephosphorylated, creating pMJ10. This plasmid was linearized with ScaI and used to transform *B. subtilis* strains 168T^+, IQB495, and IQB496, yielding strains IQB608 (*araE*::*kan*), IQB609 (Δ*msmX*::*cat araE*::*kan*), and IQB610 (Δ*araN araE*::*kan*), respectively (Table 2.3).

Table 2.2 – List of plasmids used or constructed during the course of this work.

Plasmid	Relevant features	Source or Reference
pBluescript II KS(+)	Integrative phagemid with a *lacZ* gene for blue/white selection of clones, *bla*	Stratagene
pLitmus29	Integrative plasmid with a *lacZ* gene for blue/white selection of clones, *bla*	New England Biolabs
pDG1731	Plasmid used as a source of erythromycin resistance cassette, *bla*, *erm*, *spec*	Guérout-Fleury *et al*., 1996
pMAD	Plasmid used for allelic replacement in Gram-positive bacteria, *bla*, *erm*	Arnaud *et al*., 2004
pMJ1	pBluescript II KS(+)-derivative harboring the *msmX* coding region, *bla*	This work
pMJ2	Integrative plasmid used for the construction of *msmX* null-mutations, *bla*, *cat*	This work
pMJ6	Integrative plasmid used for the construction of *araN* in-frame deletion mutants, *bla*, *erm*	This work
pMJ7	pBluescript II KS(+)-derivative harboring the promoterless *araE* coding region, *bla*	This work
pMJ10	Integrative plasmid used for the construction of *araE* null-mutations, *bla*, *kan*	This work
pMJ11	Integrative plasmid used for the construction of *araNPQ* in-frame deletion mutants, *bla*, *erm*	This work
pMJ23	pBluescript II KS(+)-based plasmid used as a source of erythromycin resistance cassette, *bla*, *erm*	This work
pMJ31	pBluescript II KS(+)-based plasmid harboring the coding region of *galK*, bla	This work
pMJ32	Integrative plasmid used for the construction of *galK* null-mutations, *bla*, *erm*	This work
pMJ38	Integrative plasmid used for the construction of *cycB* in-frame deletion mutants, *bla*, *erm*	This work
pAH248	pGEM-7Zf(+)-based plasmid used as a source of kanamycin resistance cassette, *bla*, *kan*	A. O. Henriques, unpublished results
pJL1	pLitmus29-based plasmid used as a source of erythromycin resistance cassette, *bla*, *erm*	This work
pJL3	pLitmus29-based plasmid used as a source of kanamycin resistance cassette, *bla*, *kan*	This work
pMS38	pLitmus29-based plasmid used as a source of chloramphenicol resistance cassette, *bla*, *cat*	Zilhão *et al*., 2004

For the generation of *B. subtilis* mutant strains with an in-frame deletion of *cycB*, plasmid pMJ38 was constructed. The regions upstream and downstream of the *cycB* gene were independently amplified by PCR, from chromosomal DNA of *B. subtilis* 168T$^+$, with primers ARA714 and ARA779 (PCR5) and ARA780 and ARA781 (PCR6), respectively. The products from PCR5 and PCR6 were joined by overlapping PCR using primers ARA714 and ARA781 (Table 2.1). The final product was digested with MluI and the resulting 1579 bp DNA fragment was cloned into pMAD digested with SmaI and MluI. The obtained plasmid, pMJ38, was used for the generation of clean deletions of *cycB* in the chromosome of *B. subtilis* strains 168T$^+$ and IQB611, following the published procedure described by Arnaud *et al*., 2004. These deletions were confirmed by sequencing and the new strains were named IQB638 and IQB639, respectively (Table 2.3). To construct a *B. subtilis galK* insertion-deletion mutant, *galK* was amplified by PCR from chromosomal DNA of *B. subtilis* 168T$^+$ using primers ARA718 and ARA719, and the obtained product was digested with SacI and KpnI. The resulting 1300 bp DNA fragment was cloned into the phagemid pBluescript II KS(+) (Table 2.2), also digested with SacI and KpnI, yielding pMJ31. An erythromycin resistance (Erm^R) cassette was obtained by the digestion of pMJ23 with NaeI and EcoRV and subcloned between the StuI sites of pMJ31, resulting in pMJ32. pMJ23 is a pBluescript II KS(+) derivative carrying an Erm^R cassette obtained from the digestion

of pJL1 with BamHI and PstI. pJL1 is a pLitmus29 derivative carrying an Erm^R cassette, obtained from the digestion of pDG1731 with SmaI and NaeI, cloned in its EcoRV site. pMJ32 was linearized with ScaI and used to transform *B. subtilis* $168T^+$ and IQB611 yielding strains IQB630 and IQB631, respectively (Table 2.3). The transformations of *B. subtilis* strains were performed according to the method described by Anagnostopoulos and Spizizen, 1961.

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

Strain	Genotype	Source or Reference
$168T^+$	Prototroph	F. E. Young
IQB495	Δ*msmX*::*cat*	pMJ2→$168T^+$ #
IQB496	Δ*araN*	pMJ6→$168T^+$
IQB608	*araE*::*kan*	pMJ10→$168T^+$ #
IQB609	*araE*::*kan* Δ*msmX*::*cat*	pMJ10→IQB495 #
IQB610	*araE*::*kan* Δ*araN*	pMJ10→IQB496 #
IQB611	Δ*araNPQ*	pMJ11→$168T^+$
IQB630	Δ*galK*::*erm*	pMJ32→$168T^+$ #
IQB631	Δ*galK*::*erm* Δ*araNPQ*	pMJ32→IQB611 #
IQB638	Δ*cycB*	pMJ38→$168T^+$
IQB639	Δ*cycB* Δ*araNPQ*	pMJ38→IQB611

Arrows indicate transformation and point from donor DNA to the recipient strain.
Transformation was carried out with linearized plasmid.

Growth conditions

E. coli DH5α (Gibco-BRL) was used for the construction of all plasmids. All *E. coli* strains were grown in liquid Luria-Bertani (LB) medium (Miller, 1972) and on LB solidified with 1.6% (w/v) agar, where ampicillin (100 $\mu g.mL^{-1}$) was added as appropriate. *B. subtilis* strains were grown in liquid LB medium, LB medium solidified with 1.6% (w/v) agar or liquid SP medium (Martin *et al.*, 1987), with the addition of chloramphenicol (5 $\mu g.mL^{-1}$), kanamycin (10 $\mu g.mL^{-1}$), erythromycin (1 $\mu g.mL^{-1}$) and X-Gal (80 $\mu g.mL^{-1}$) as appropriate. Growth kinetics parameters of *B. subtilis* strains $168T^+$, IQB495, IQB496, IQB608, IQB609, IQB610, IQB611, IQB630, IQB631, IQB638 and IQB639 (Table 2.3) were determined in liquid minimal medium. Cells from freshly streaked *B. subtilis* strains were grown overnight (37°C, 150 rpm) in C minimal medium (Pascal *et al.*, 1971) supplemented with L-tryptophan (100 $\mu g.mL^{-1}$; Sigma Aldrich Co.), potassium glutamate (8 $\mu g.mL^{-1}$; Sigma Aldrich Co.), and potassium succinate (6 $\mu g.mL^{-1}$; Sigma Aldrich Co.) (CSK medium; Débarbouillé *et al.*, 1990). The cell cultures were washed and resuspended to an initial OD_{600nm} of 0.05 in 1.5 mL of CSK medium without potassium succinate and supplemented with different carbon and energy sources (glucose, arabinose, arabinobiose, arabinotriose, arabinotetraose, arabinan, debranched arabinan, maltotriose, maltotetraose, and maltopentaose) at a final concentration of 0.1% (w/v). The cultures were grown at 37°C and 180 rpm in an Aquatron® water bath rotary shaker, and the OD_{600nm} was read periodically with an Ultrospec 2100 *pro* UV/visible spectrophotometer.

Results

Uptake of arabinose oligomers by AraNPQ

The AraNPQ proteins are the BPD and TMD components of an ABC-type importer (Sá-Nogueira *et al.*, 1997). The three proteins, which are encoded by the arabinose metabolic operon *araABDLMNPQ-abfA*, were previously shown to be dispensable for arabinose utilization in a strain bearing a large deletion comprising all genes downstream of *araD* (Sá-Nogueira *et al.*, 1997). However, growth experiments indicated a possible involvement in the utilization of arabinose oligomers (Inácio *et al.*, 2008). To confirm this hypothesis, an in-frame deletion mutation in the *araN* gene was generated by allelic replacement to minimize the polar effect on the genes of the *araABDLMNPQ-abfA* operon which are located downstream of *araN* (Figure 2.1).

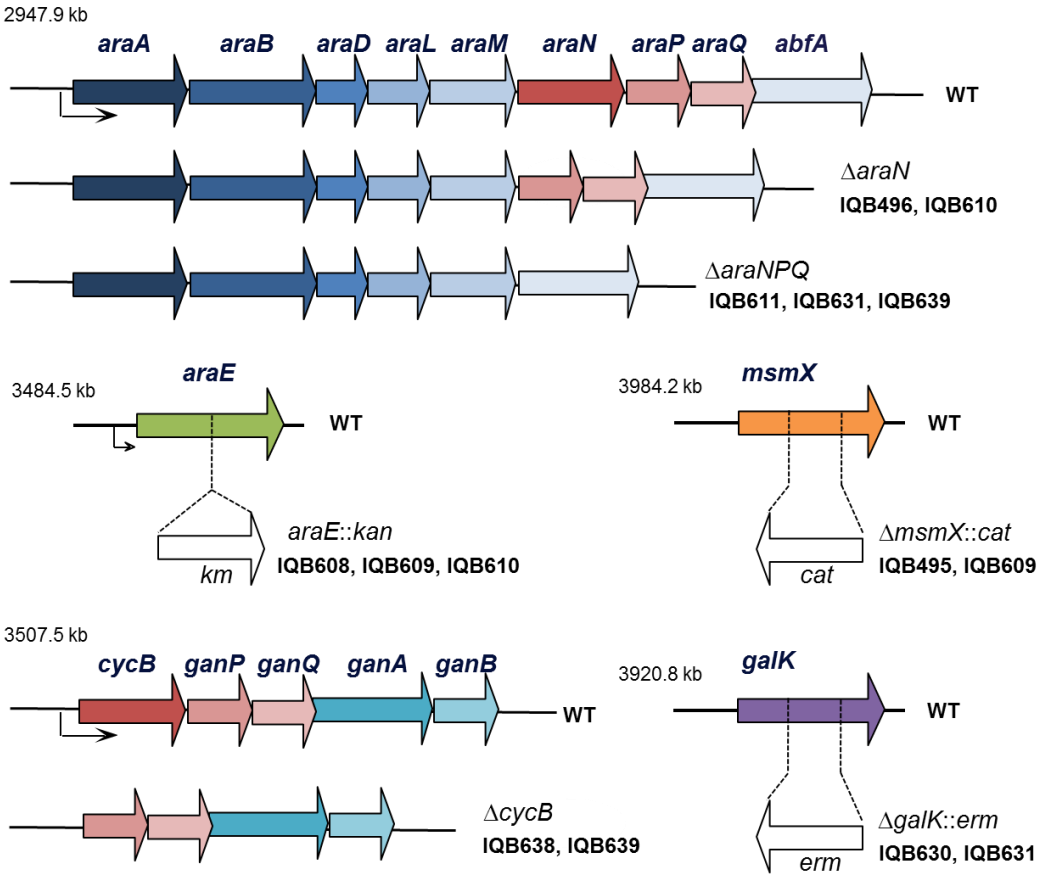


Figure 2.1 – Organization of the *araABDLMNPQ-abfA*, *araE*, *msmX*, *cycB-ganPQAB* loci of the wild-type *B. subtilis* chromosome (WT). The location of the five regions is indicated in kilobase (kb) pairs. The genes are represented by arrows pointing in the direction of transcription. The constructs bearing the mutations used in this work are displayed below each region and the strains harboring each mutation are indicated in front of the construct. The three in-frame deletions generated by allelic replacement, Δ*araN*, Δ*araNPQ* and Δ*cycB*, are depicted below the *araABDLMNPQ-abfA* and *cycB-ganPQAB* loci. The mutation generated by the insertion of a kanamycin resistance cassette (*kan*) is represented in the *araE* locus. The insertion-deletion mutation created by a deletion in *msmX* followed by the insertion of a chloramphenicol resistance cassette (*cat*) is presented below the *msmX* locus. The insertion-deletion mutation created by a deletion in *galK* followed by the insertion of an erythromycin resistance cassette (*erm*) is presented below the *galK* locus.

The physiological effect of this knockout mutation in *B. subtilis* (strain IQB496; Table 2.3) was assessed by the determination of the growth kinetic parameters using various saccharides as the sole carbon and energy source (Table 2.4). In the presence of glucose and arabinose, the doubling time of the mutant is comparable to that of the wild-type strain, indicating both the stability of the strain bearing the in-frame deletion and that the AraNPQ system is not responsible for the uptake of arabinose. In contrast, the absence of AraN has a negative effect on the ability of strain IQB496 (Δ*araN*) to grow on the α-1,5-arabinose oligomers (α-1,5-arabinobiose, α-1,5-arabinotriose, and α-1,5-arabinotetraose; Table 2.4). Similar results were obtained in the growth experiments (Table 2.4) with a strain harboring an in-frame deletion of the three genes *araNPQ* (IQB611 Δ*araNPQ*; Table 2.3), which were constructed by allelic replacement as describe below for the in-frame deletion of *araN* (Figure 2.1). Taken together, the data implicate the AraNPQ system in the uptake of arabinose oligomers.

Table 2.4 – Effect of different mutations in the uptake of α-1,5-arabinooligosaccharides by *B. subtilis*. Doubling times (minutes) for different strains in liquid minimal medium (CSK) using glucose, arabinose, arabinobiose, arabinotriose or arabinotetraose as sole carbon and energy source. Results are the averages of three independent assays and their respective standard deviations. NG, no growth; -, not determined.

	168T⁺ (wild-type)	**IQB496 (Δ*araN*)**	**IQB611 (Δ*araNPQ*)**	**IQB608 (*araE*::*kan*)**	**IQB610 (*araE*::*kan* Δ*araN*)**
Glucose 0.1%	55.5±1.3	53.4±1.0	52.5±0.8	57.3±0.3	54.1±1.1
Arabinose 0.1%	67.0±0.2	66.6±0.9	63.7±2.9	NG	NG
Arabinobiose 0.1%	112.4±10.8	263.8±37.2	258.2±46.1	178.6±19.9	NG
Arabinotriose 0.1%	98.2±10.0	NG	NG	-	-
Arabinotetraose 0.1%	92.4±4.5	NG	NG	-	-

Interestingly, the absence of a functional AraNPQ system (strains IQB496 and IQB611) impaired growth in the presence of α-1,5-arabinotriose and α-1,5-arabinotetraose; however, in the presence of α-1,5-arabinobiose, the growth rate decreased only 2.3-fold when compared to that of the wild-type strain (Table 2.4). This observation suggested the existence of an additional transporter for this substrate. In *B. subtilis*, arabinose enters the cell mainly through the AraE transporter, a proton symport-type permease (Sá-Nogueira and Ramos, 1997) with broad substrate specificity that also is responsible for the uptake of xylose and galactose (Krispin and Allmansberger, 1998). Thus, AraE, due to its capacity to transport different substrates, was considered a potential candidate for the transport of α-1,5-arabinobiose. To test this hypothesis, the *araE* gene was disrupted by an insertion mutation (Figure 2.1) as described in Materials and Methods. The resulting strain IQB608 (*araE*::*kan*) was unable to grow in the presence of arabinose (Table 2.4), although it displayed normal growth in the presence of glucose. On the other hand, in the presence of α-1,5-arabinobiose it showed a 1.6-fold decrease of its growth rate when compared to that of the wild-type strain (Table 2.4). This effect is similar to that observed with strain IQB496 (Δ*araN*), suggesting the involvement of AraE in the uptake of arabinobiose. To further support this evidence, a double mutant, Δ*araN araE*::*kan*, was constructed (strain IQB610) and shown to be unable to grow in the presence of α-1,5-arabinobiose (Table 2.4). These

data revealed that both AraE and AraNPQ are able to transport the disaccharide arabinobiose and represent the only transport systems accountable for its uptake in *B. subtilis*.

Transport via the AraNPQ system is energized by MsmX

The nucleotide-binding domain (NBD) protein partner of the AraNPQ ABC importer is missing in the *araABDLMNPQabfA* transcriptional unit (Sá-Nogueira *et al.*, 1997). Quentin *et al.* (1999) carried out, *in silico*, the inventory and assembly of the ATP binding cassette (ABC) transporter systems in the complete genome of *B. subtilis*. The authors identified five genes encoding ATPases without any gene coding for TMD protein in their neighborhood. On the basis of similarities, three ATP-binding proteins were proposed to energize 10 incomplete systems. Among these ATPases, MsmX is highly similar (>40% identity) to MalK of *E. coli* (Gilson *et al.*, 1982) and MsmK of *Streptococcus mutans* (Russell *et al.*, 1992), ATPases of transport systems for maltose and multiple sugars, respectively. Thus, the authors proposed that MsmX would be an ATPase involved in the transport of carbohydrates and would be one candidate to energize several incomplete systems in operons that encode only the membrane and solute-binding proteins (Quentin *et al.*, 1999).

To test the possible involvement of MsmX in the transport of AraNPQ substrates, an insertion-deletion mutation was generated in the *msmX* gene (Figure 2.1). The resulting strain, IQB495 (Δ*msmX*::*cat*), was grown as described above, and the physiological response to the presence of different saccharides was determined. The *msmX*-null mutation does not affect the kinetic growth parameters in the presence of glucose and arabinose. The negative impact of this mutation in the doubling time of strain IQB495 (Δ*msmX*::*cat*) grown in the presence of α-1,5-arabinobiose, α-1,5-arabinotriose, and α-1,5-arabinotetraose (Table 2.5) was comparable to that observed for the Δ*araN* mutant strain (IQB496; Table 2.4).

Table 2.5 – Effect of a *msmX*-null mutation in the uptake of α-1,5-arabinooligosaccharides by *B. subtilis*. Doubling times (minutes) for different strains in liquid minimal medium (CSK) using glucose, arabinose, arabinobiose, arabinotriose or arabinotetraose as sole carbon and energy source. Results are the averages of three independent assays and their respective standard deviations. NG, no growth; -, not determined.

	IQB495 (Δ*msmX*::*cat*)	**IQB609 (Δ*msmX*::*cat araE*::*kan*)**
Glucose 0.1%	57.7±2.7	54.1±1.1
Arabinose 0.1%	68.2±2.3	NG
Arabinobiose 0.1%	286.7±33.1	NG
Arabinotriose 0.1%	NG	-
Arabinotetraose 0.1%	NG	-

This observation clearly indicates that MsmX, like AraN, is essential for the functionality of AraNPQ, playing the role of the ATPase that energizes the unidirectional substrate transport via

AraNPQ. Furthermore, since both IQB495 and IQB496 exhibit a parallel growth rate in the presence of α-1,5-arabinobiose, the additional transporter responsible for the uptake of this substrate must be MsmX independent. Moreover, the phenotype of the double mutant IQB609 (Δ*msmX*::*cat araE*::*kan*; Table 2.5) was very similar to that observed with strain IQB610 (Δ*araN araE*::*kan*; Table 2.4). Taken together, these findings corroborate the evidence that AraE is involved in the uptake of α-1,5-arabinobiose.

MsmX contributes to the transport of multiple saccharides

In addition to its role as an essential partner in the AraNPQ-MsmX system, MsmX previously was shown to be the NBD protein partner of the MdxEFG ABC importer, where MdxE is a maltodextrin-binding protein with high affinities for maltodextrins and a low affinity for maltose, and MdxF and MdxG are the membrane components (Schönert *et al.*, 2006). In this study, we assayed the physiological impact of the mutation Δ*msmX*::*cat* (strain IQB495) on the growth on maltotriose, maltotetraose, and maltopentaose. Although a negative effect in the doubling time was observed, the *msmX*-null mutation does not completely impair growth in the presence of maltooligosaccharides (data not shown). Studies of transport performed by Schönert et al. (2006) using radiolabeled maltotriose showed that the uptake was not prevented in the *msmX*-null mutation, although in the *malP msmX* double mutant this transport was drastically reduced (Schönert *et al.*, 2006). Thus, our results correlate with these observations, and the growth observed in the presence of maltooligosaccharides is most probably due to its extracellular degradation and uptake of resulting products, such as maltose and glucose, by other transport systems (Schönert *et al.*, 2006).

Mutant and wild-type strains were also grown in the presence of two different types of the homopolysaccharide arabinan: branched arabinan, a molecule comprising a backbone of α-1,5-linked L-arabinofuranosyl residues decorated with α-1,2-, and α-1,3-linked L-arabinofuranosyl units, and debranched arabinan, the linear homopolysaccharide α-1,5-L-arabinan. In both mutant strains IQB495 (Δ*msmX*::*cat*) and IQB611 (Δ*araNPQ*), a slow but steady growth rate on debranched arabinan was observed compared to that of the wild type (Table 2.6). Since we showed that a functional AraNPQ-MsmX system is required for the transport of α-1,5-arabinotriose and α-1,5-arabinotetraose and is partially responsible for the uptake of α-1,5-arabinobiose, the slow growth observed most probably is due to the metabolism of some arabinobiose and arabinose residues resulting from the extracellular degradation of linear α-1,5-L-arabinan, which enters the cell through the AraE permease. Interestingly, in the presence of branched arabinan, strain IQB611 (Δ*araNPQ*) displays a marginal increase in the doubling time compared to that of the wild-type strain (Table 2.6); in contrast, the *msmX*-null mutation (strain IQB495) has a severe impact on growth (Table 2.6). This fact suggested that most of the arabinooligosaccharides resulting from the extracellular degradation of branched arabinan were nonlinear, i.e., arabinooligosaccharides with α-1,2- and/or α-1,3-linked arabinosyl residues, and

that these products were not transported by the AraNPQ system. The slight decrease of the growth rate observed in strains lacking a functional AraNPQ compared to that of the wild type could represent the reduced amount of linear arabinooligosaccharides resulting from the extracellular degradation of arabinan that, in the wild-type strain, would be transported through AraNPQ and metabolized.

Table 2.6 – Uptake of arabinan by the AraNPQ-MsmX system. Doubling times (minutes) for different strains in liquid minimal medium (CSK) using glucose, arabinan or debranched arabinan as sole carbon and energy source. Results are the averages of three independent assays and their respective standard deviations. NG, no growth.

	168T⁺ (wild-type)	**IQB495 (Δ*msmX*::*cat*)**	**IQB611 (Δ*araNPQ*)**
Glucose 0.1%	55.5±1.3	57.7±2.7	52.5±0.8
Arabinan (branched) 0.1%	149.0±6.8	NG	188.8±15.3
Debranched Arabinan 0.1%	84.0±4.6	332.5±6.8	330.6±30.8

Taken together, the growth kinetic parameters of the different strains grown in the presence of two different types of arabinan molecules suggested the existence of an additional transporter for nonlinear arabinooligosaccharides. Furthermore, this putative importer would be MsmX dependent due to the lack of growth of the *msmX*-null mutant in the presence of branched arabinan (sugar beet).

To identify a transporter responsible for the uptake of these nonlinear arabinooligosaccharides, we targeted the CycB-GanPQ system (previously YvfKLM), which is part of the group of ABC transporters proposed to be energized by MsmX (Quentin *et al.*, 1999) and, displays higher identity to AraNPQ. An in-frame deletion mutation in the *cycB* gene was generated by allelic replacement, as shown in Figure 2.1. The physiological effect of this mutation in the utilization of arabinan by *B. subtilis* strains IQB638 (Δ*cycB*) and IQB639 (Δ*cycB* Δ*araNPQ*) was assessed by the determination of the growth kinetic (Table 2.4). Similarly to that observed for the Δ*araNPQ* mutant, only a marginal decrease in the growth rate of the Δ*cycB* mutant was detected (Table 2.7) when compared do the wild-type strain grown on arabinan. Nonetheless, the ability for the double Δ*cycB* Δ*araNPQ* mutant (IQB639) to grow on the same substrate was completely abolished (Table 2.7), suggesting that the putative MsmX-dependent transporter for nonlinear arabinooligosaccharides was the CycB-GanPQ system.

Table 2.7 – Effect of mutations in galactose utilization genes in the uptake of arabinan. Doubling times (minutes) for different strains in liquid minimal medium (CSK) using glucose or arabinan as sole carbon and energy source. Results are the averages of three independent assays and their respective standard deviations. NG, no growth.

	IQB638 (Δ*cycB*)	**IQB639 (Δ*cycB* Δ*araNPQ*)**	**IQB630 (Δ*galK*::*erm*)**	**IQB631 (Δ*araNPQ* Δ*galK*::*erm*)**
Glucose 0.1%	52.6±0.1	52.1±1.2	45.8±1.2	53.9±1.7
Arabinan (branched) 0.1%	168.7±39.4	NG	300.4±28.1	NG

However, we later became aware that both the branched and debranched sugar beet arabinans purchased from Megazyme International Ireland, which were supposed to contain about 90% (w/w) of arabinose and less than 4% (w/w) of galactose, had in fact about 25% (w/w) galactose. Thus, to test if the growth observed in the Δ*araNPQ* mutant was due to the presence of galactose (or most likely galactooligosaccharides) and not due to nonlinear arabinooligosaccharides, an insertion-deletion mutation of the galactose kinase gene (*galK*) was generated (Figure 2.1). The resulting Gal⁻ strains, IQB630 (Δ*galK*::*erm*) and IQB631 (Δ*galK*::*erm* Δ*araNPQ*), were grown as previously described in the presence of arabinan. As expected, there was a decrease in the growth rate of the Δ*galK*::*erm* mutant (Table 2.7) when compared to the wild-type strain. Moreover, the double Δ*galK*::*erm* Δ*araNPQ* mutant completely lost its ability to grow in the presence of arabinan as the sole carbon and energy source (Table 2.7). These results confirm that the used arabinan is highly contaminated with galactose which is, consequently, responsible for maintaining growth of *B. subtilis* when a functional AraNPQ system is unavailable. Additionally, these results implicate that the CycB-GanPQ system plays a role in the uptake of galactooligosaccharides as previously suggested (Shipkowsky and Brenchley, 2006; Chai *et al.*, 2012).

Discussion

Bacillus subtilis participates in the enzymatic dissolution of plant biomass. Thus, it is able to synthesize a vast variety of glycoside hydrolases capable of the depolymerization of plant cell wall polysaccharides, such as cellulose, hemicellulose, or pectin (Shallom and Shoham, 2003; Stülke and Hillen, 2000). L-Arabinose, the second most-abundant pentose in plant biomass, next to D-xylose, is found in homopolysaccharides (branched and debranched arabinans) and heteropolysaccharides (arabinoxylans, arabinogalactans, etc.). *B. subtilis* produces exo- and endo-acting arabinases capable of releasing arabinosyl oligomers and L-arabinose from plant cell walls (Inácio *et al.*, 2008; Inácio and Sá-Nogueira, 2008; Weinstein and Albersheim, 1979). Although many polysaccharolytic glycoside hydrolases have been purified from bacteria, including several *Bacillus* spp., information on ABC transporters devoted to the uptake of products resulting from the degradation of hemicellulose is limited (Eitinger *et al.*, 2011).

The results presented here lead to the characterization of an ABC primary transporter of *B. subtilis* involved in the uptake of products of the degradation of arabinans. The AraNPQ system is encoded by the *araABDLMNPQ-abfA* operon, which is regulated at the transcriptional level by induction in the presence of arabinose and repression by glucose (Sá-Nogueira *et al.*, 1997). AraN is the high-affinity substrate-binding protein (BPD) that captures the substrate and delivers it to the two transmembrane components (TMDs), AraP and AraQ. The genetic and physiological impact of in-frame deletions on both the *araN* gene and the entire transporter *araNPQ* lead us to the conclusion that AraNPQ is involved in the uptake of α-1,5-arabinose oligomers (up to four

units) but not arabinose. AraNPQ is the sole transporter for α-1,5-arabinotriose and α-1,5-arabinotetraose, but is only partially accountable for the uptake of α-1,5-arabinobiose. The latter is also transported by the AraE permease, the main permease involved in the uptake of arabinose (Sá-Nogueira and Ramos, 1997), and also is responsible for the transport of xylose and galactose (Krispin and Allmansberger, 1998). This evidence is supported by the observation that both single mutations in the *araN* (or *araNPQ*) and *araE* genes result in slower growth on α-1,5-arabinobiose; on the other hand, growth on this substrate is impaired by the double mutation *araN araE*.

In Gram-positive bacteria gene clusters that encode sugar ABC transporters, the BPDs and TMDs frequently lack the nucleotide-binding domains (NBDs). Thus, the NBDs responsible for providing energy for the corresponding transporters are not linked. The *B. subtilis* genome encodes three ATP-binding proteins without any gene coding for TMD protein in their neighborhood (Quentin *et al.*, 1999). Among these ATPases, MsmX is a strong candidate for supplying energy to the incomplete AraNPQ transporter (Quentin *et al.*, 1999); moreover, the *msmX* gene is constitutively expressed in both complex and minimal medium (Yoshida *et al.*, 2000). MsmX is highly similar to MsmK from *Streptococcus mutans* (61% identity; Russell *et al.*, 1992) and MsiK from *Streptomyces lividans* (55% identity; Hurtubise *et al.*, 1995). In *Streptomyces* species the MsiK protein is an ATPase involved in several oligosaccharide uptake systems. MsiK energizes the cellobiose, xylobiose, and maltose uptake systems in *Streptomyces lividans* (Hurtubise *et al.*, 1995; Schlösser *et al.*, 1997), the trehalose uptake system in *Streptomyces reticuli* (Schlösser, 2000), and the *N*,*N*'-diacetylchitobiose uptake in *Streptomyces coelicolor* (Saito *et al.*, 2008). In *Streptococcus mutans*, the MsmEFGK system transports raffinose, melibiose, and stachyose, and the MalXFGK transporter is specific for maltodextrins (Webb *et al.*, 2008). The two ATPase domains, MsmK and MalX, were shown to interact with either transporter system to energize uptake (Webb *et al.*, 2008). In *B. subtilis* it was previously shown that MsmX energizes MdxEFG, a transporter specific for maltodextrins, as the cognate ABC domain (Schönert *et al.*, 2006). Here, we show that MsmX is required for the uptake of α-1,5-arabinose oligomers (up to four units) but not arabinose, thus being the ATPase that supplies energy to AraNPQ. Contrarily to our first hypothesis, an additional MsmX-dependent transporter for nonlinear arabinooligosaccharides is unlikely to exist and these sugars are imported by the AraNPQ system, which was shown to be responsible for the uptake of α-1,5-arabinooligosaccharides. Furthermore, our findings lead us to infer that the CycB-GanPQ (previously YvfKLM) ABC-type system is likely to be involved in the uptake of galactooligosaccharides as suggested in previous work by other groups (Shipkowsky and Brenchley, 2006; Chai *et al.*, 2012). As predicted by Quentin *et al.* (1999), our results also point to MsmX being the ATPase responsible for energizing this system. These findings concerning the functional characterization of CycB-GanPQ will be further discussed in Chapter III. Taken together, these data led to our current model of the transport and utilization of arabinooligosaccharides by *B. subtilis* (Figure 2.2).

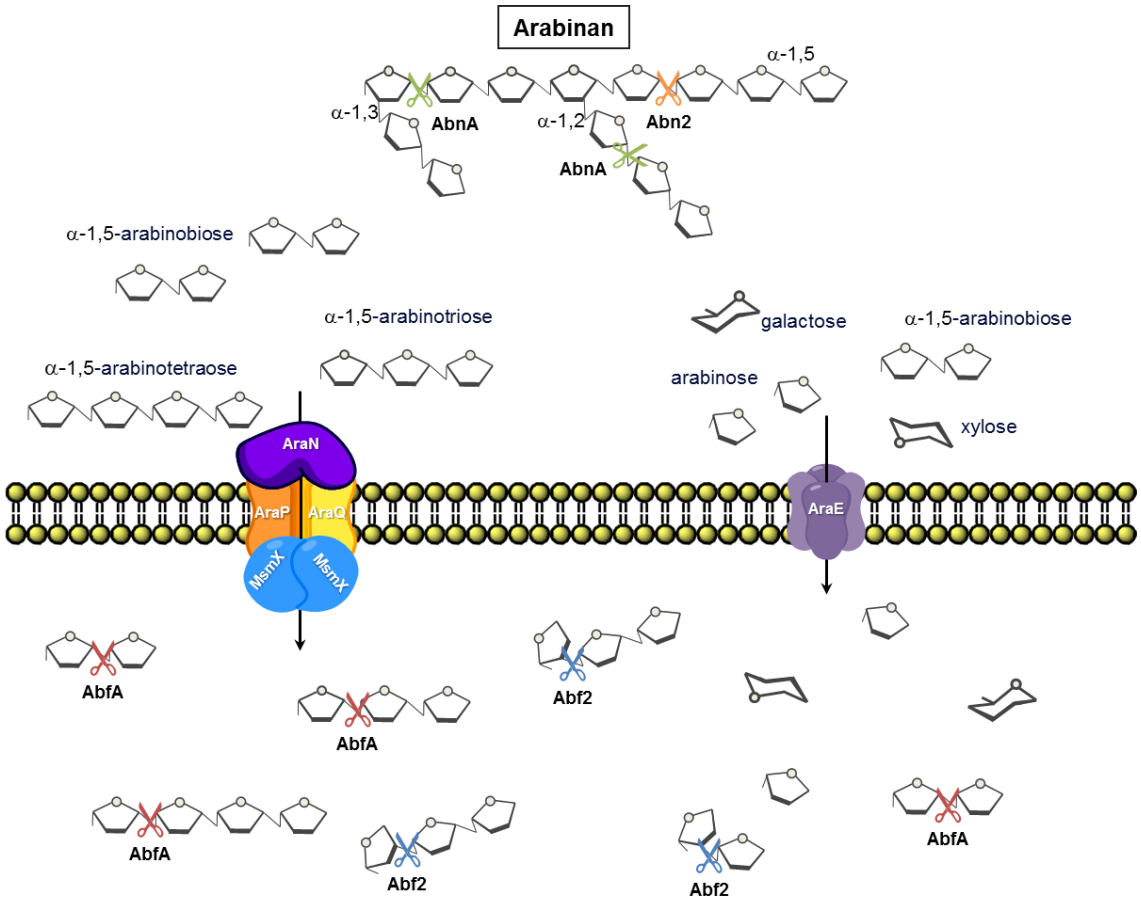


Figure 2.2 – Primary and secondary transporters involved in the uptake of arabinooligosaccharides in *B. subtilis* and the role of MsmX in their transport. Arabinan, a homopolymer of arabinose, is extracellularly degraded by two endo-α-1,5-arabinanases, AbnA and Abn2. The resulting products, mainly arabinooligosaccharides, enter the cell through different transport systems. The ABC-type importer AraNPQ is involved in the uptake of α-1,5-arabinooligosaccharides and is energized by the ATPase MsmX. The AraNPQ-MsmX system is the key transporter for α-1,5-arabinotriose and α-1,5-arabinotetraose. The AraE permease, a proton symporter, is the major transporter for arabinose, and is additionally responsible for the uptake of xylose and galactose. This broad-specificity transporter also displays capacity for the uptake of α-1,5-arabinobiose. Once inside the cell, arabinooligosaccharides are further degraded to L-arabinose by the action of two α-L-arabinofuranosidases, AbfA and Abf2 (see the text for a more detailed explanation).

Arabinan, a homopolymer of arabinose, is extracellularly degraded by two GH43 endo-α-1,5-arabinanases, AbnA and Abn2 (Inácio and Sá-Nogueira, 2008; Leal and Sá-Nogueira, 2004). The resulting products, mainly arabinooligosaccharides, enter the cell through different transport systems. The ABC-type importer AraNPQ is involved in the uptake of α-1,5-arabinooligosaccharides, at least up to four L-arabinosyl units, and is energized by the ATPase MsmX. The primary AraNPQ-MsmX system is the key transporter for α-1,5-arabinotriose and α-1,5-arabinotetraose. The AraE permease, a proton symporter, the major transporter for arabinose (Sá-Nogueira and Ramos, 1997), is additionally responsible for the uptake of xylose and galactose (Krispin and Allmansberger, 1998) from other hemicellulosic heteropolysaccharides like arabinoxylan, arabinogalactan, or pectin. This broad-specificity secondary transporter also displays capacity for the uptake of α-1,5-arabinobiose, although this sugar enters the cell

preferentially through the primary AraNPQ-MsmX importer. Once inside the cell, α-1,5-, α-1,2- and α-1,3-arabinooligosaccharides are further degraded by the action of two GH51 family α-L-arabinofuranosidases, AbfA and Abf2 (Inácio *et al.*, 2008). The resulting L-arabinose is then metabolized by AraA, AraB, and AraD, which sequentially convert L-arabinose to D-xylulose 5-phosphate, which is further catabolized through the pentose phosphate pathway.

In this work we provide evidence that MsmX, either as a homo- or heterodimer, is responsible for providing energy for at least two, and possibly three, ABC oligosaccharide transport systems. To our knowledge, the capacity of an ATPase to associate and energize distinct transport systems with different specificities in *B. subtilis* is described here for the first time. These findings support the establishment of this peculiar broad capacity of protein-protein interaction in the general bacterial mechanism of ABC-type sugar import.

Acknowledgments

Michel Débarbouillé provided us plasmid pMAD as a gift and gave us some advice for the construction of the in-frame deletion mutations. Adriano O. Henriques and Mónica Serrano provided plasmids pAH248 and pMS38, respectively, as gifts. Joana Lima constructed plasmids pJL1 and pJL3. This work was supported by the Fundação para a Ciência e a Tecnologia [grant numbers PTDC/AGR-AAM/102345/2008 to I.S.N.; PEst-OE/BIA/UI0457/2011 to Centro de Recursos Microbiológicos; fellowship SFRH/BD/73039/2010 to M.J.F.].

Chapter III

Uptake of pectin by MsmX-dependent ABC-type importers in *Bacillus subtilis*

This chapter contains data to be published in “The MsmX ATPase plays a crucial role in pectin mobilization by *Bacillus subtilis*” (manuscript in preparation). All the experimental work was performed by the author of this Thesis, except the construction of some plasmids and strains.

Uptake of pectin by MsmX-dependent ABC-type importers in *Bacillus subtilis*

Abstract

Carbohydrates from plant cell walls are often found as heteropolysaccharides intertwined with each other. To have a competitive advantage against other microorganisms, and be able to fully exploit the available carbon and energy sources, *Bacillus subtilis* possesses a high number of proteins dedicated to the uptake of mono- and oligosaccharides. Here, we identify new transporter complexes, belonging to the ABC superfamily, involved in the uptake of oligosaccharides commonly found in pectins. Additionally, these newly identified carbohydrate importers are also shown to be MsmX-dependent, Thus MsmX a multipurpose ATPase serving several distinct ABC-type sugar importers plays a crucial role in pectin mobilization. MsmX is shown to be essential for energizing various ABC transporters however, a second *B. subtilis* ATPase, YurJ, is capable of complementing its function when placed in *trans* at a different locus of the chromosome.

Introduction

The physiological need for energy is common to all known living organisms and thus, the continued search for suitable energy sources demands constant adaptation and efficiency improvement in the utilization of the available resources. Although many organisms, especially Bacteria and Archaea, use a wide array of energy sources such as light or inorganic compounds (like molecular hydrogen, ammonia, nitrogen dioxide or hydrogen sulfide) and depend on carbon dioxide as a carbon source, most Bacteria rely on a variety of organic compounds, from simple methane molecules to more complex sugars, as both energy and carbon sources.

The plant cell wall is a highly complex structure, with a variable species- and tissue-dependent composition, and a major reservoir of carbohydrates in the form of cellulose, hemicellulose, lignin and pectin (van Maris *et al.*, 2006; Pauly and Keegstra, 2010; Kalluri *et al.*, 2014; Loqué *et al.*, 2015). Cellulose is exclusively composed of D-glucose units, linked by β-1,4-glycosidic bonds and organized as linear parallel polymers connected via hydrogen bonds (van Maris *et al.*, 2006). Hemicellulose and pectin are complex mixtures of branched polysaccharides composed of many different sugar monomers such as glucose, galactose, xylose, arabinose, mannose, rhamnose and galacturonic acid (van Maris *et al.*, 2006; Ochiai *et al.*, 2007; Edwards and Doran-Peterson, 2012).

To be able to thrive and proliferate, and quickly adapt to changes in the environment and nutrient availability, microorganisms have successfully developed mechanisms that enable the use of most of these complex polysaccharides as carbon and energy sources, providing a competitive advantage when nutrient resources are scarce. To achieve this, many species developed concerted mechanisms for the degradation of these composite polysaccharides and the subsequent uptake of smaller sugar monomers and oligomers, which can be easily metabolized for energy or for the biosynthesis of other molecules. One such concerted mechanism is the *Bacillus subtilis* transport and utilization system for arabinan (Inácio and Sá-Nogueira, 2008), a hemicellulose polysaccharide commonly found in low-lignin and pectin-rich substrates such as sugar beet pulp and citrus waste (Edwards and Doran-Peterson, 2012). This bacterium in its natural environment – the soil or the gastrointestinal tract of animals (Tam *et al.*, 2006; Serra *et al.*, 2014) – possesses two endo-α-1,5-arabinanases, AbnA and Abn2, responsible for breaking down the backbone of the homopolysaccharide arabinan (Inácio and Sá-Nogueira, 2008; Leal and Sá-Nogueira, 2004), releasing arabinose monomers and oligomers. Two intracellular α-L-arabinofuranosidases, AbfA and Abf2 (Inácio *et al.*, 2008), are accountable for further degrading the arabinooligosaccharides, which are transported inside the cell through two different types of transport systems. The AraE proton symporter is responsible for the uptake of arabinose (Sá-Nogueira and Ramos, 1997) and also plays a role in the transport of α-1,5-arabinobiose (Chapter II); the AraNPQ ABC-type system is responsible for the uptake of arabinooligosaccharides up to at least 4 arabinosyl units (Chapter II). This system, along with the *B. subtilis* ABC-type maltodextrins importer MdxEFG, is energized by the multipurpose ATPase MsmX (Chapter II; Schönert *et al.*, 2006). MsmX together with another putative orphan ATPase are predicted to be the nucleotide-binding domain (NBD) that energizes five additional putative ABC-type importers (Quentin *et al.*, 1999). Although multipurpose ABC ATPases have been described in other organisms like *Streptomyces lividans* (Hurtubise *et al.*, 1995), *Streptococcus pneumoniae* (Marion *et al.*, 2011) and *Streptococcus suis* (Tan *et al.*, 2015), the capability to interact with and energize multiple transporters is rather uncommon among this type of ATPases. In fact, to date the ability of a single ATPase to energize several ABC-systems appears to be exclusive of sugar importers.

Here, we investigate the role of MsmX in the utilization of plant cell wall polysaccharides by *B. subtilis* and the biological significance of sharing a single ATPase among multiple ABC transport systems. The expression of both ATPases MsmX and YurJ was analized. New ABC-type sugar importers responsible for the uptake of pectin-based polysaccharides have been functionally characterized and established MsmX as a key player in plant cell wall polysaccharides mobilization. Furthermore, important amino acid residues for the contacts between MsmX and the transmembrane domains (TMDs) of the AraNPQ importer were also identified.

Materials and Methods

Substrates

α-1,5-arabinotriose, galactan (ex. Lupin), polygalacturonan and rhamnogalacturonan I were purchased from Megazyme International Ireland. D-(+)-glucose, pectin (from apple) and D-(+)-galacturonic acid were purchased from Sigma-Aldrich Co.

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 and used according to the manufacturers' recommendations. PCR amplifications were carried out using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific). DNA from agarose gels and PCR products were purified with the illustra™ GFX™ PCR DNA and Gel Band Purification kit (GE Healthcare Life Sciences). All DNA ligations were performed using T4 DNA Ligase (Thermo Fisher Scientific). DNA phosphorylation was performed using T4 Polynucleotide Kinase (Thermo Fisher Scientific) and DNA dephosphorylation with FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific). Plasmids were purified using the NZYMiniprep kit (NZYTech). DNA sequencing was performed with the ABI PRISM BigDye® Terminator Cycle Sequencing Kit (Applied Biosystems). The sequencing reaction was purified by gel filtration and resolved in an ABI 3730XL sequencer.

Construction of plasmids and strains

For the construction of a *B. subtilis yurJ* deletion mutant, the regions upstream and downstream of the *yurJ* gene were independently amplified by PCR, from chromosomal DNA of *B. subtilis* 168T$^+$, with primers ARA517 and ARA518 (PCR1) and ARA519 and ARA520 (PCR2), respectively. The products from PCR1 and PCR2 were joined by overlapping PCR using primers ARA517 and ARA520 (Table 3.1). The final product was digested with EcoRV and the resulting 1620 bp fragment cloned into pMAD digested with SmaI and subsequently dephosphorylated. The obtained plasmid, pMJ14, was used for the generation of clean deletions of *yurJ* in the chromosome of *B. subtilis* strains 168T$^+$ and IQB613, following the procedure described by Arnaud *et al*., 2004. The deletions were confirmed by sequencing and the new strains named IQB618 and IQB635, respectively. To obtain a Δ*yurJ* Δ*msmX*::*cat* mutant, pMJ2 (Table 3.2) was linearized with ScaI and used to transform *B. subtilis* IQB618, yielding strain IQB634. To construct a *B. subtilis* strain with a copy of *yurJ* under the control of an IPTG-inducible promoter, in the *amyE* locus of the *B. subtilis* chromosome, this gene was amplified by PCR from *B. subtilis* 168T$^+$ chromosomal DNA using primers ARA820 and ARA821 (Table 3.1). The amplification product was digested with SalI and SphI and the resulting 1222 bp DNA fragment was cloned into pDR111

(Table 3.2) digested with the same restriction enzymes. The obtained plasmid, pMJ40, was linearized with ScaI and used for the transformation of *B. subtilis* IQB495, yielding strain IQB642. For the construction of a *B. subtilis* strain with a C-terminal His_6-tag fused to YurJ, the 3' end of *yurJ* was amplified by PCR, from chromosomal DNA of *B. subtilis* $168T^+$, using primers ARA831 and ARA832 (Table 3.1). The obtained product of 819 bp was phosphorylated and subcloned into pMS38 (Table 3.2) digested with EcoRV and subsequently dephosphorylated, resulting in pMJ43. This plasmid was used for the transformation of *B. subtilis* $168T^+$. The fusion of the His_6-tag to *msmX* was confirmed by sequencing and the new strain was named IQB644.

Table 3.1 – List of all oligonucleotides used during the course of this work.

Oligonucleotide	Sequence (5'→3')	Complementary sequence
ARA55	CTATCATTAGTACGTATCTTTTGTATT	*araA*
ARA422	TATTGAGCTCTCAGGGATAGATATCAAATCG	*msmX*
ARA423	CATCGGTACCATTCTGAGATTTTCAATAGC	*msmX*
ARA517	TCCCCGATATCACAATCCAGG	*yurJ*
ARA518	AGCCTCTTCCGTTTCCGCATCGAATGTTAATGAAGCCATCAGAAAAGTCCC	*yurJ*
ARA519	GGGACTTTTCTGATGGCTTCATTAACATTCGATGCGGAAACGGAAGAGGCT	*yurJ*
ARA520	ATGACCGATATCCGTCTTCCC	*yurJ*
ARA544	GGATGTCATATGGCTGAATTGCGG	*msmX*
ARA545	CCGGTTTTTTTGCTCGAGTCGGATTCTCAC	*msmX*
ARA566	GGGTTTGCCGGTGCAGGAGATTGGAAGAC	*araA*
ARA567	GTCTTCCAATCTCCTGCACCGGCAAACCC	*araA*
ARA573	CATACGCATAAATCTGCATG	*araA*
ARA583	TCGCGGTTTCGCTGCCCTTT	*16S*
ARA584	AAGTCCCGCAACGCGCGCAA	*16S*
ARA630	TACATGAAGGATCCTGCCAAGGC	*araP*
ARA631	CAGCGGCTGCGTACAGCTCT	*araP*
ARA632	AGAGCTGTACGCAGCCGCTG	*araP*
ARA633	TTGGAAGGCCATGGGCATATTGC	*araP*
ARA634	ATGTGCTGACCATGGGCATCATCG	*araQ*
ARA637	CGGGATCCATGCCTCTTGTG	*araQ*
ARA638	TTCGATCCGGCTTCATATTC	*yurJ*
ARA639	CGTTCGCTTGAAACACTGAA	*yurJ*
ARA640	CCGAAGCCTGAAATCAAAAA	*msmX*
ARA641	ACAAAGGCTCATCCATCAGG	*msmX*
ARA642	AAACACGCGTTGGACCATATCC	*yesO*
ARA645	CCGGATCCTCCAAAATTGTATGCC	*yesQ*
ARA652	CTGTGCCATCTGCAACACTCCTTTTTAGTAACAAATCTTCTTCAAACCAATCCC	*yesO*
ARA653	GGGATTGGTTTGAAGAAGATTTGTTACTAAAAAGGAGTGTTGCAGATGGCACAG	*yesQ*
ARA667	GCTTCGTACAGCGCTTTCGGC	*araP*
ARA668	GCCGAAAGCGCTGTACGAAGC	*araP*
ARA669	GTATTCGCGCCGGCTATATCAGC	*araP*
ARA670	GCTGATATAGCCGGCGCGAATAC	*araP*
ARA671	CGTACAGCCGGCCATCCTTGC	*araQ*
ARA672	GCAAGGATGGCCGGCTGTACG	*araQ*
ARA673	CGTTTATAGGATCCCGCTGACC	*araQ*
ARA674	CATCACACGTTTCCTCCTTCATTACCCTACCGTCAGGCCGGAGATAAA	*araQ*
ARA675	TTTATCTCCGGCCTGACGGTAGGGTAATGAAGGAGGAAACGTGTGATG	*araQ*
ARA676	GCCAACGCGTTTGGCAGATTGT	*araQ*
ARA733	TTCGCTAGATGCTTATCGC	*ytcQ*
ARA734	TGGGTTTTTCCAATGTTATTTTGCTTGCACCCTCCATTTGTTTCCCATCGTACC	*ytcQ*
ARA735	GGTACGATGGGAAACAAATGGAGGGTGCAAGCAAAATAACATTGGAAAAACCCA	*ytcQ*
ARA736	TCAGACGCGTTGAAATATCG	*ytcQ*
ARA820	ACCAAGTCGACTCCAGTGACG	*yurJ*
ARA821	GCAATGCATGCATCGACCACC	*yurJ*
ARA831	CTCAGAAAGATGGCGAAACAGG	*yurJ*
ARA832	CTAGTGGTGGTGGTGGTGGTGATACACAGCCTCTTCCGTTTCCG	*yurJ*
ARA833	GTGCGTCGACACAACTTTGTATTGCTG	*yurJ*
ARA834	GGAAACAACTTGTTCTTTCTGTTA	*yurJ*
ARA835	TAACAGAAAGAACAAGTTGTTTCC	*yurJ*
ARA836	AAAGATCTATGTCATGAGGTTTCGC	*yurJ*

Restriction sites in the primer sequences are underlined.

For the construction of a *B. subtilis* strain with a single C→A nucleotide substitution at position -54 from the coding sequence of *yurJ*, pMJ42 was generated. Two partially overlapping regions upstream and downstream of the desired nucleotide were independently amplified by PCR, from *B. subtilis* 168T$^+$ chromosomal DNA, using primers ARA833 and ARA834 (PCR3) and primers ARA835 and ARA836 (PCR4), respectively. The products from PCR3 and PCR4 were joined by overlapping PCR using primers ARA833 and ARA836 (Table 3.1). The 1391 bp product was phosphorylated and cloned into pMAD digested with SmaI and subsequently dephosphorylated, yielding pMJ42. The obtained plasmid was used for the generation of a C→A nucleotide substitution at position -54 from the coding sequence of *yurJ* in the chromosome of *B. subtilis* strains 168T$^+$ and IQB495 following the published procedure described by Arnaud *et al*., 2004. This substitution was confirmed by sequencing and the new strains were named IQB647 and IQB648, respectively.

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

Plasmid	Relevant features	Source or Reference
pBluescript II KS(+)	Integrative phagemid with a *lacZ* gene for blue/white selection of clones, *bla*	Stratagene
pET30a(+)	Expression vector with an inducible T7 promoter for the expression of proteins in *E. coli*, *kan*	Novagen
pDR111	Integrative plasmid which allows the insertion of genes under the control of Phyper-spank in the *amyE* locus, *bla*, *spec*	David Rudner
pMAD	Plasmid used for allelic replacement in Gram-positive bacteria, *bla*, *erm*	Arnaud *et al*., 2004
pUC18	Integrative plasmid with a *lacZ* gene for blue/white selection of clones, *bla*	Yanisch-Perron *et al*., 1985
pMJ2	Integrative plasmid used for the construction of *msmX* null-mutations, *bla*, *cat*	Chapter II
pMJ14	Integrative plasmid used for the construction of *yurJ* in-frame deletion mutants, *bla*, *erm*	This work
pMJ18	Integrative plasmid used for the construction of *yesOPQ* in-frame deletion mutants, *bla*, *erm*	This work
pMJ21	pET30a(+)-based vector with a truncated version of *msmX*-His_6, *kan*	This work
pMJ22	pET30a(+)-based vector for the expression of *msmX*-His_6, *kan*	This work
pMJ23	pBluescript II KS(+)-based plasmid used as a source of erythromycin resistance cassette, *bla*, *erm*	This work
pMJ25	Integrative plasmid used for the construction of *araP* E205A mutants, *bla*, *erm*	This work
pMJ26	Integrative plasmid used for the construction of *araP* D213A mutants, *bla*, *erm*	This work
pMJ27	Integrative plasmid used for the construction of *araQ* D180A mutants, *bla*, *erm*	This work
pMJ28	Integrative plasmid used for the construction of *araQ* mutants with an in-frame deletion of 4 C-terminal codons, *bla*, *erm*	This work
pMJ33	Integrative plasmid used for the construction of *ytcQ* in-frame deletion mutants, *bla*, *erm*	This work
pMJ38	Integrative plasmid used for the construction of *cycB* in-frame deletion mutants, *bla*, *erm*	This work
pMJ40	Integrative plasmid used for the insertion of *yurJ* under the control of Phyper-spank in the *amyE* locus, *bla*, *spec*	This work
pMJ42	Integrative plasmid used for the construction of C→A mutants at position -54 from *yurJ*, *bla*, *erm*	This work
pMJ43	Integrative plasmid used for the fusion of a C-terminal Histidine tag to *yurJ*, *bla*, *cat*	This work
pGS1	Integrative plasmid used for the fusion of a C-terminal Histidine tag to *msmX*, *bla*, *cat*	This work
pLB1	Integrative plasmid used for the construction of *araP* E208A mutants, *bla*, *erm*	This work
pLG31	Integrative plasmid used for the construction of *araA* E305A mutants, *bla*, *erm*	Godinho, unpublished results
pMS38	pLitmus29-based plasmid used as a source of chloramphenicol resistance cassette, *bla*, *cat*	Zilhão *et al*., 2004

For the generation of *B. subtilis* mutant strains with in-frame deletions of *yesOPQ* and *ytcQ*, plasmids pMJ18 and pMJ33 were constructed. The regions upstream of the *yesO* gene and downstream of the *yesQ* gene were independently amplified by PCR, from chromosomal DNA of *B. subtilis* 168T$^+$, with primers ARA642 and ARA652 (PCR5) and ARA653 and ARA645 (PCR6), respectively. The products from PCR5 and PCR6 were joined by overlapping PCR using primers ARA642 and ARA645 (Table 3.1). The final product was digested with MluI and the resulting 1586 bp DNA fragment was cloned into pMAD digested with SmaI and MluI. The obtained plasmid, pMJ18, was used for the generation of clean deletions of *yesOPQ* in the chromosome of *B. subtilis* strains 168T$^+$ and IQB611 (Table 3.2), following the published procedure described by Arnaud *et al*., 2004. These deletions were confirmed by sequencing and the new strains were named IQB628 and IQB629, respectively. The regions upstream and downstream of the *ytcQ* gene were independently amplified by PCR, from chromosomal DNA of *B. subtilis* 168T$^+$, with primers ARA733 and ARA734 (PCR7) and ARA735 and ARA736 (PCR8), respectively. The products from PCR7 and PCR8 were joined by overlapping PCR using primers ARA733 and ARA736 (Table 3.1). The final product was digested with MluI and the resulting 1541 bp DNA fragment was cloned into pMAD digested with SmaI and MluI. The obtained plasmid, pMJ33, was used for the generation of clean deletions of *ytcQ* in the chromosome of *B. subtilis* strains 168T$^+$ and IQB628, following the published procedure described by Arnaud *et al*., 2004. These deletions were confirmed by sequencing and the new strains were named IQB632 and IQB633, respectively. For the construction of a *B. subtilis* strain with a single mutation leading to a E305A amino acid substitution in the AraA protein, *araA* was amplified by PCR, from *B. subtilis* 168T$^+$ chromosomal DNA, with primers ARA55 and ARA573 (Table 3.1). The obtained product was digested with EcoRI and KpnI and the resulting 2313 bp fragment cloned into pUC18 digested with the same restriction enzymes. The obtained pUC18::*araA* was used as template for a PCR with primers ARA566 and ARA567 (Table 3.1). The resulting pUC18::*araA*(E305A) was digested with SalI and EcoRI and the 2328 bp fragment with the mutagenized *araA* gene was subcloned into pMAD digested with the same restriction enzymes, yielding pLG31. This new plasmid was used for the generation of a mutation leading to an E305A amino acid substitution in AraA. *B. subtilis* strain IQB638 was transformed following the published procedure described by Arnaud *et al*., 2004. The mutation was confirmed by sequencing and the new strain was named IQB641. For the generation of single E208A, E205A and D213A substitutions in AraP of *B. subtilis*, pLB1, pMJ25 and pMJ26 were constructed. For the E208A substitution two partially overlapping regions upstream and downstream of the desired codon were independently amplified by PCR, from *B. subtilis* 168T$^+$ chromosomal DNA, using primers ARA630 and ARA631 (PCR9) and primers ARA632 and ARA633 (PCR10), respectively. The products from PCR9 and PCR10 were joined by overlapping PCR using primers ARA630 and ARA633 (Table 3.1). The obtained product was digested with BamHI and NcoI and the resulting 1595 bp DNA fragment was cloned into pMAD digested with the same restriction enzymes, yielding pLB1. A similar approach was used for the E205A and D213A substitutions. For the E205A substitution the initial upstream and downstream amplifications were performed with primers ARA630 and ARA667 (PCR11) and primers ARA668

and ARA633 (PCR12), respectively. The products from PCR11 and PCR12 were joined by overlapping PCR using primers ARA630 and ARA633 (Table 3.1). The obtained product was digested with BamHI and NcoI and the resulting 1595 bp DNA fragment was cloned into pMAD digested with the same restriction enzymes, yielding pMJ25.

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

Strain	Genotype	Source or Reference
168T$^+$	Prototroph	F. E. Young
IQB495	Δ*msmX*::*cat*	Chapter II
IQB611	Δ*araNPQ*	Chapter II
IQB618	Δ*yurJ*	pMJ14→168T$^+$
IQB622	*msmX*-His$_6$ *cat*	pGS1→168T$^+$
IQB623	*araP**E208A	pLB1→168T$^+$
IQB624	*araP**E205A	pMJ25→168T$^+$
IQB625	*araP**D213A	pMJ26→168T$^+$
IQB626	*araQ**D180A	pMJ27→168T$^+$
IQB627	Δ*araQ* (4 codon 3' deletion)	pMJ28→168T$^+$
IQB628	Δ*yesOPQ*	pMJ18→168T$^+$
IQB629	Δ*yesOPQ* Δ*araNPQ*	pMJ18→IQB611
IQB632	Δ*ytcQ*	pMJ33→168T$^+$
IQB633	Δ*ytcQ* Δ*yesOPQ*	pMJ33→IQB628
IQB638	Δ*cycB*	pMJ38→168T$^+$
IQB639	Δ*cycB* Δ*araNPQ*	pMJ38→IQB611
IQB641	Δ*cycB araA**E305A	pLG31→IQB638
IQB642	*amyE*::[Phyper-spank-*yurJ spec*] Δ*msmX*::*cat*	pMJ40→IQB495 #
IQB644	*yurJ*-His$_6$ *cat*	pMJ43→168T$^+$
IQB647	*yurJ**C→A (position-54)	pMJ42→168T$^+$
IQB648	*yurJ**C→A (position-54) Δ*msmX*::*cat*	pMJ42→IQB495

Arrows indicate transformation and point from donor DNA to the recipient strain.

Transformation was carried out with linearized plasmid.

For the D213A substitution the initial upstream and downstream amplifications were performed with primers ARA630 and ARA669 (PCR13) and primers ARA670 and ARA633 (PCR14), respectively. The products from PCR13 and PCR14 were joined by overlapping PCR using primers ARA630 and ARA633 (Table 3.1). The obtained product was digested with BamHI and NcoI and the resulting 1595 bp DNA fragment was cloned into pMAD digested with the same restriction enzymes, yielding pMJ26. The obtained plasmids, pLB1, pMJ25 and pMJ26, were used for the generation of mutations in *araP* in the chromosome of *B. subtilis* 168T$^+$, leading to single amino acid substitutions in the AraP protein, following the published procedure described by Arnaud *et al*., 2004. The E208A, E205A and D213A substitutions were confirmed by sequencing and the new strains were named IQB623, IQB624 and IQB625, respectively. For the generation of a single D180A substitution in AraQ of *B. subtilis*, pMJ27 was constructed. Two independent and partially overlapping regions upstream and downstream of the desired codon were amplified by PCR, from *B. subtilis* 168T$^+$ chromosomal DNA, using primers ARA634 and ARA671 (PCR15) and primers ARA672 and ARA637 (PCR16), respectively. The products from PCR15 and PCR16 were joined by overlapping PCR using primers ARA634 and ARA637 (Table 3.1). The obtained

product was digested with BamHI and NcoI and the resulting 1489 bp DNA fragment was cloned into pMAD digested with the same restriction enzymes, resulting in pMJ27. The obtained plasmid was used for the generation of a mutation in *araQ* in the chromosome of *B. subtilis* 168T$^+$, leading to a single D180A amino acid substitution in the AraQ protein, following the published procedure described by Arnaud *et al*., 2004. The amino acid substitution was confirmed by sequencing and the new strain was named IQB626. For the construction of a *B. subtilis* strain with an in-frame deletion of the 4 codons at the 3' end of *araQ*, the regions upstream and downstream of the target codons were independently amplified by PCR, from *B. subtilis* 168T$^+$ chromosomal DNA, using primers ARA673 and ARA674 (PCR17) and primers ARA675 and ARA676 (PCR18), respectively. The products from PCR17 and PCR18 were joined by overlapping PCR using primers ARA673 and ARA676 (Table 3.1). The obtained product was digested with MluI and the resulting 1605 bp DNA fragment was cloned into pMAD digested with SmaI and MluI, yielding pMJ28. The obtained plasmid was used for the generation of a clean deletion of 4 codons of *araQ* in the chromosome of *B. subtilis* strains 168T$^+$, following the published procedure described by Arnaud *et al*., 2004. This deletion was confirmed by sequencing and the new strain was named IQB627. For the fusion of a C-terminal Histidine-tag to MsmX, a DNA fragment containing part of the *msmX* gene was obtained, from chromosomal DNA of *B. subtilis* 168T$^+$, by PCR with primers ARA544 and ARA545 (Table 3.1) and digestion with NdeI and XhoI. The resulting 829 bp was cloned into pET30a(+) (Table 3.2) digested with the same enzymes, yielding pMJ21. A 270 bp DNA fragment containing the 5' end of the *msmX* gene (amplified with primers ARA544 and ARA545 and digested with NdeI) was subcloned into NdeI-digested pMJ21, resulting in pMJ22. To construct a *B. subtilis* strain with a C-terminal His_6-tag fused to MsmX, pMJ22 was digested with SspI and the resulting 1329 bp DNA fragment, containing the 3' end of *msmX* fused to the histidine tag, was subcloned into pMS38 (Table 3.2), resulting in pGS1. This plasmid was used for the transformation of *B. subtilis* 168T$^+$. The fusion of the His_6-tag to *msmX* was confirmed by sequencing and the new strain was named IQB622. The transformations of *B. subtilis* strains were performed according to the method described by Anagnostopoulos and Spizizen, 1961.

Growth conditions

E. coli DH5α (Gibco-BRL) or *E. coli* XL1-Blue (Stratagene) were used for the construction of all plasmids. All *E. coli* strains were grown in liquid Luria-Bertani (LB) medium (Miller, 1972) and on LB solidified with 1.6% (w/v) agar, where ampicillin (100 µg.mL^{-1}), chloramphenicol (25 µg.mL^{-1}), kanamycin (30 µg.mL^{-1}) or tetracycline (12 µg.mL^{-1}) were added as appropriate. *B. subtilis* strains were grown in liquid LB medium, LB solidified with 1.6% (w/v) agar or liquid SP medium (Martin *et al*., 1987), and chloramphenicol (5 µg.mL^{-1}), kanamycin (10 µg.mL^{-1}), erythromycin (1 µg.mL^{-1}), spectinomycin (60 µg.mL^{-1}), X-Gal (80 µg.mL^{-1}) or IPTG (1 mM) were added as appropriate. Growth kinetics parameters of *B. subtilis* strains 168T$^+$, IQB495, IQB611, IQB618, IQB622, IQB623, IQB624, IQB625, IQB626, IQB627, IQB628, IQB629, IQB632, IQB633,

IQB638, IQB639, IQB641, IQB642 and IQB648 (Table 3.3) were determined in CSK minimal medium in the presence of 0.1% (w/v) of several carbon sources (glucose, arabinose, arabinotriose, arabinan, galactan, pectin, galacturonic acid, polygalacturonan or rhamnogalacturonan I), as described in Chapter II. The optical density (OD_{600nm}) of growing cultures was used for the determination of the doubling time for each strain in the presence of the tested sugars as sole carbon and energy sources.

RNA extraction

B. subtilis 168T$^+$ was grown in CSK minimal medium supplemented with 0.1% (w/v) of glucose, arabinose or arabinotriose, as previously described for growth kinetics parameters determination (see above). When the OD_{600nm} of growing cultures reached 0.7-0.8, cells were collected by centrifuging 1.4 mL of each culture at 6000*g* and 4ºC for 5 minutes. Total RNA extraction was performed using the Absolutely Total RNA Miniprep kit (Agilent Technologies) according to the manufacturer's instructions and using RNase-free technique. The integrity of RNA extracts was analyzed in a 1% agarose gel (ran in and prepared with TBE 1x). DNA contamination of RNA samples was assessed by PCR using primers ARA422 and ARA423. Total RNA was quantified using a Nanodrop™ 1000 Spectrophotometer (Thermo Fisher Scientific) and stored in 10 µL aliquots at -80ºC.

RT-qPCR experiments and data analyses

The primers used for RT-qPCR experiments – ARA583 and ARA584 (16S gene), ARA638 and ARA639 (*yurJ*), and ARA640 and ARA641 (*msmX*) – were designed with the help of the online tool Primer3Plus (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi). Primer efficiency was assessed using the Rotor-Gene SYBR® Green PCR Kit (QIAGEN) and RT-qPCR experiments were performed using the SensiFAST™ SYBR No-ROX One-Step Kit (Bioline), both in a Rotor-Gene 6000 (Corbett) real-time cycler. RT-qPCR experiments were performed according to the kit manufacturer's instructions, using 40 ng of total RNA and 0.1 µM of each primer, in a volume of 12.5 µL. Statistical analyses were performed with GraphPad Prism (version 5.00) using C_t values obtained from three independent assays. *p* values were determined using an unpaired *t* test.

Protein extracts of *B. subtilis* and Western Blot analysis

B. subtilis strains 168T$^+$, IQB622 and IQB644 were grown in CSK minimal medium supplemented with 0.1% (w/v) of glucose, arabinose or arabinotriose, as previously described for

growth kinetics parameters determination (see above). Cells of growing cultures were collected as described for RNA extraction. The collected cells were resuspended in 100 µL of Lysis Buffer (500 mM KCl, 20 mM HEPES K^+, pH 7.6, 10 mM EDTA, 1 mM DTT, 10% glycerol). Lysozyme (1 $mg.mL^{-1}$) was added and the mixture was incubated for 10 minutes at 37ºC, followed by three cycles of freezing in liquid nitrogen and thawing for 5 minutes at 37ºC. 10 mM PMSF and 7.5 U of Benzonase® Nuclease (Sigma-Aldrich Co.) were added followed by an incubation for 15 minutes at 37ºC. Total protein content for each extract was determined using Bio-Rad Protein Assay (Bio-Rad Laboratories, Inc.). 25 µg of total protein from each extract and 2 µg of purified MsmX-His_6 were loaded in a 12.5% SDS-PAGE and run at constant electrical current (30 mA) for 50 minutes. The fractionated proteins were then electrotransfered into a nitrocellulose membrane (Bio-Rad), for 1h30 at constant voltage (90 V) and 4ºC. Protein-blocking to the membrane was then performed using 10 mL of a powdered milk solution in TBS-Tween (4% w/v) for 30 minutes at room temperature with mild shaking. The membrane was then washed with TBS-Tween solution for 15 minutes at room temperature with mild shaking, followed by overnight incubation, at 4ºC, with 10 mL of primary antibody solution (mouse monoclonal Anti-6X His-tag® antibody [HIS.H8; Abcam], diluted 1:1000 in PBS-Tween with 4% w/v powdered milk). All subsequent incubation and wash steps were performed at room temperature with mild shaking. The membrane was washed twice with 10 mL of TBS-Tween solution, for 10 minutes, followed by incubation with 10 mL of secondary antibody solution (HRP-conjugated goat anti-mouse IgG antibody [Jackson ImmunoResearch Europe Ltd.], diluted 1:10000 in TBS-Tween with 4% w/v powdered milk) for 30 minutes. Finally, the membrane was washed three times with 10 mL of TBS-Tween solution. All subsequent steps were performed in a dark room. Blots were developed using 1 mL Western Lightning™ ECL Pro kit (PerkinElmer). Amersham Hyperfilm plates (GE Healthcare) were exposed to luminescence for 10 seconds inside a closed Amersham Hypercassette Autoradiography Cassette (GE Healthcare).

Protein analysis

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

Results

Functional analysis of *yurJ*

The *B. subtilis* chromosome encodes two highly similar (>55%) ATPases, MsmX and YurJ (>55% amino acid identity), belonging to sub-family 5a of ABC-type importers associated to

carbohydrate uptake (Quentin *et al.*, 1999). MsmX also displays a high identity (>40%) to the MalK ATPase (Gilson *et al.*, 1982), which is responsible for energizing the *E. coli* maltose/maltodextrins ABC-type importer. The MalEFGK$_2$ transporter is the most studied member of this class of transporters and the determination of its complete structure (Oldham *et al.*, 2007) allowed the identification of amino acids accountable for contacts between the TMDs and the NBDs (highlighted in Figure 3.1).

An alignment of the primary structure of the three proteins, MsmX, YurJ and MalK, revealed that most of the MalK amino acids contacting the TMDs are conserved in both MsmX and YurJ suggesting that the same residues might be responsible for similar contacts in the two *B. subtilis* ATPases (Figure 3.1). Moreover, the 21 amino acid residues putatively responsible for MsmX-TMD contacts are conserved in YurJ, which implies the possibility of YurJ being able to associate to MsmX-energized ABC transporters.

Figure 3.1 – MsmX, YurJ and MalK protein sequence alignment. The sequence alignment was obtained using the online tool ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Conserved ABC ATPase motifs are depicted by colored boxes (Walker A, *blue*; Q-loop, *red*; Signature motif, *green*; Walker B, *purple*; D-loop, *orange*; H-loop, *yellow*). Residues known to be involved in the contacts between MalK and the TMDs of the E. coli maltose/maltodextrins transporter are underlined in *green*.

However, the results obtained using a *B. subtilis* strain without a functional MsmX indirectly show that YurJ is unable to functionally complement an *msmX*-null mutation in the conditions tested (Chapter II). Nevertheless, we constructed a deletion of *yurJ* from the *B. subtilis* chromosome (Δ*yurJ* strain IQB618) and the new strain was tested. The results shown in Table 3.4 indicate no negative impact on the ability to grow on arabinotriose as the sole carbon and energy source.

Table 3.4 – Effect of *msmX* and *yurJ* deletions in the uptake of arabinotriose by *B. subtilis*. Doubling times (minutes) for different strains in liquid minimal medium (CSK) using glucose or arabinotriose as sole carbon and energy source. Results are the averages of three independent assays and their respective standard deviations. # IPTG was added to a final concentration of 1 mM. NG, no growth.

	168T⁺ (wild-type)	**IQB495 (Δ*msmX*::*cat*)**	**IQB618 (Δ*yurJ*)**	**IQB642 (Δ*msmX*::*cat* *amyE*::[Phyper-spank-*yurJ spec*])**
Glucose 0.1%	55.5±1.3	57.7±2.7	50.4±5.5	52.9±3.2 #
Arabinotriose 0.1%	98.2±10.0	NG	92.8±10.9	94.6±0.5 #

These observations led to two likely hypotheses: i) YurJ is present in the cell but it is not functional and thus unable to energize transport; ii) in the conditions tested YurJ is not present in the cell. To test these hypotheses, a new *msmX*-null mutant strain harboring a copy of *yurJ* under the control of an IPTG-inducible promoter was constructed (IQB642). This strain was grown in minimal medium with arabinotriose as the sole carbon and energy source and, in the presence of IPTG, displayed a similar growth rate to that of the wild-type strain (Table 3.4). This result indicates that YurJ is in fact capable to functionally complement MsmX and energize the uptake of arabinooligosaccharides by AraNPQ. Additionally, this strongly suggests that in the conditions tested YurJ is not present in the cell when expressed from its own locus.

Transcriptional analysis of *msmX* and *yurJ* expression

Since YurJ is able to energize the importer AraNPQ in an *msmX*-null mutant strain when expressed in *trans* under the control of an IPTG-inducible promoter, but not from its own locus, we analyzed its expression at the transcriptional level in the wild-type strain. The expression levels of both *yurJ* and *msmX* was determined in the conditions previously used for growth assays (CSK minimal medium) by measuring the mRNA levels of both gene transcripts using quantitative reverse transcription polymerase chain reaction (RT-qPCR). The obtained data were used to calculate gene expression fold-change and the results are presented in Table 3.5 and Figure 3.2. For each gene the mRNA levels are similar in the presence of both arabinose and arabinotriose. However, in the presence of glucose an approximately 75-fold repression is observed for *msmX* but not for *yurJ*. This observation was expected since a catabolite-responsive element (*cre*) is present in *msmX* (Miwa *et al.*, 2000) but none have been predicted for *yurJ*. Furthermore, when directly comparing the transcript levels from both genes it is clear that there is no significant

difference in their expression in the presence of arabinose or arabinotriose (Figure 3.2). In fact, when arabinotriose is used as the sole carbon and energy source in growth assays a slightly higher expression level (about 3-fold) is detected for *yurJ* when compared to *msmX* (Figure 3.2).

Table 3.5 – Quantification of *yurJ* and *msmX* mRNA level in the B. subtilis wild-type strain. The results represent the fold-change of the expression in the target conditions versus the control conditions. (glu), (ara) and (A3) represent the presence of 0.1% of glucose, arabinose or arabinotriose, respectively. Primers ARA638 and ARA639 were used for *yurJ* and primers ARA640 and ARA641 for *msmX*. Fold-change was normalized using primers ARA583 and ARA584 for the 16S gene. Statistical analyses were performed with GraphPad Prism (version 5.00) using C_t values obtained from three independent assays. *p* values were determined using an unpaired *t* test (^{ns}, non-significant difference; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

	yurJ		***msmX***	
Target / Control	**(glu)**	**(A3)**	**(glu)**	**(A3)**
(ara)	0.993±0.054^{ns}	1.782±0.323*	0.013±0.001***	0.899±0.051^{ns}
(glu)	-	1.794±0.311**	-	68.120±3.683***

These results clearly show that in these conditions *yurJ* is expressed at the transcriptional from its locus in the *B. subtilis* chromosome suggesting the existence of a post-transcriptional regulatory mechanism responsible for the inability of YurJ to complement an *msmX*-null mutant.

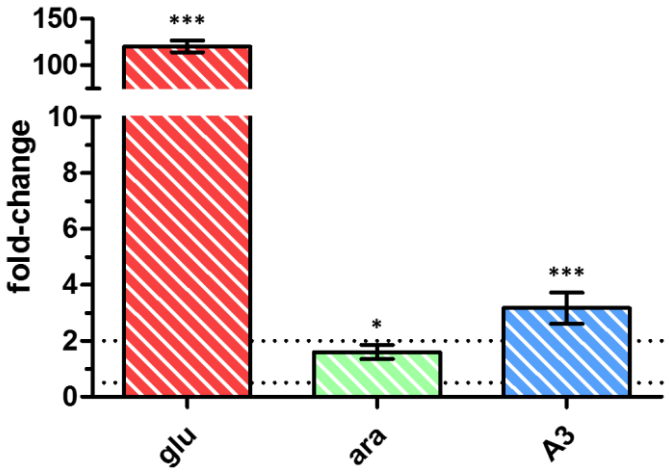


Figure 3.2 – mRNA level of *yurJ* versus *msmX* in the B. sutilis wild-type strain. The results represent the relative expression of *yurJ* versus *msmX*. Upper and lower relative expression threshold of 2.0 and 0.5, respectively, are represented by dotted lines. (glu, *red*-striped bar), (ara, *green*-striped bar) and (A3, *blue*-striped bar) represent the presence of 0.1% of glucose, arabinose or arabinotriose, respectively. Primers ARA638/ARA639 and ARA640/ARA641 were used for *yurJ* and *msmX*, respectively. Fold-change was normalized using primers ARA583 and ARA584 for the 16S gene. Statistical analyses were performed with GraphPad Prism (version 5.00) using C_t values obtained from three independent assays. *p* values were determined using an unpaired *t* test (^{ns}, non-significant difference; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

A post-transcriptional regulatory mechanism controls *yurJ* expression

Since the mRNA levels of *yurJ* and *msmX* in the cell are similar when *B. subtilis* is grown in minimal medium supplemented with arabinotriose, *yurJ* is being transcribed but a functional protein product is not present. Thus, in order to determine if YurJ is translated two new mutant strains in which C-terminal His_6-tags were fused to *msmX* (IQB622) and *yurJ* (IQB644) in their own loci were constructed. We then performed a Western Blot analysis to detect the accumulation of MsmX-His_6 and YurJ-His_6 in cells grown in the same conditions previously used in RT-qPCR experiments. Similar amounts of total protein extracts from strains IQB622 and IQB644, grown in CSK medium supplemented with arabinose, glucose or arabinotriose, were used for the detection of MsmX-His_6 and YurJ-His_6, respectively. Additionally, a protein extract from the wild-type strain ($168T^+$) grown in CSK with arabinose was used as negative control and purified recombinant MsmX-His_6 was used as positive control for His-tag detection (Figure 3.3).

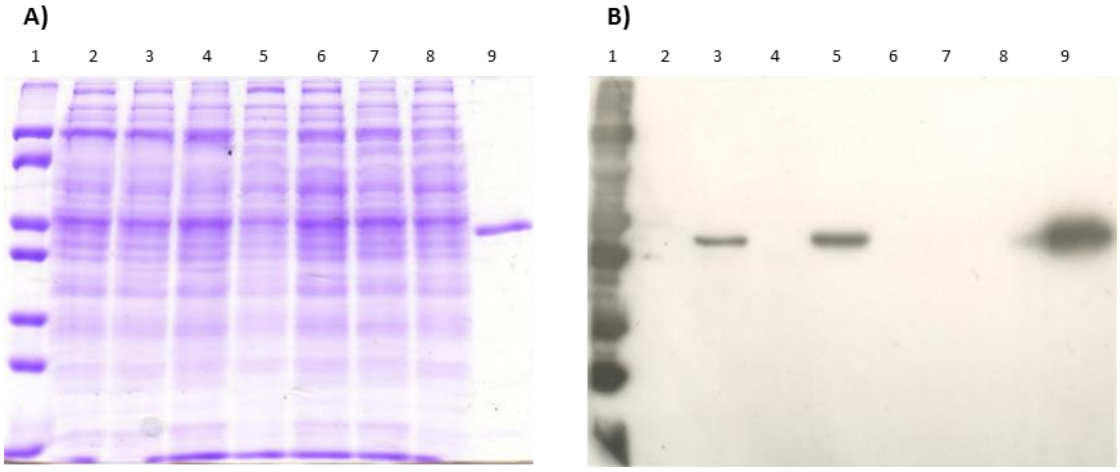


Figure 3.3 – Identification of MsmX and YurJ by Western Blot. A) SDS-PAGE (12.5%) with samples used for Western Blot (25 µg of total extract and 2 µg of purified MsmX-His_6 were loaded); **B)** Autoradiography plate after His-tag detection using an anti-His_6-specific antibody. ***1*)** LMW Protein Marker; ***2*)** wild-type strain + arabinose 0.1%; ***3*)** IQB622 (*msmX*-His_6) + arabinose 0.1%; ***4*)** IQB622 (*msmX*-His_6) + glucose 0.1%; ***5*)** IQB622 (*msmX*-His_6) + arabinotriose 0.1%; ***6*)** IQB644 (*yurJ*-His_6) + arabinose 0.1%; ***7*)** IQB644 (*yurJ*-His_6 *cat*) + glucose 0.1%; ***8*)** IQB644 (*yurJ*-His_6 *cat*) + arabinotriose 0.1%; ***9*)** purified MsmX-His_6.

The obtained results confirm the presence of MsmX-His_6 in the cell when grown in minimal medium with either arabinose or arabinotriose, however in these conditions YurJ-His_6 is not detected (Figure 3.3). In accordance with the results obtained from the RT-qPCR experiments where a strong glucose-mediated repression of *msmX* is observed, no MsmX-His_6 is detected when glucose is present in the medium (Figure 3.3). Thus, by combining the RT-qPCR results and Western Blot analysis, we show that although *msmX* and *yurJ* are being similarly expressed in the presence of arabinotriose, the latter translation product is not detected. This observation explains why there is no functional complementation by YurJ of an *msmX*-null mutation when the gene is expressed from its own locus in the *B. subtilis* chromosome. Together the results indicate

that in minimal medium with either glucose, arabinose or arabinotriose as sole carbon and energy source, *yurJ* is transcribed but not translated. To investigate this phenomenon the secondary structure of the 5' region in the *yurJ* mRNA was thoroughly analyzed for the presence of putative hairpins that could be hindering translation initiation. The analysis revealed a putative hairpin structure located eight bp upstream from the RBS, and predicted to have a free energy change (Δ*G*) of -13.3 Kcal.mol^{-1} (Figure 3.4).

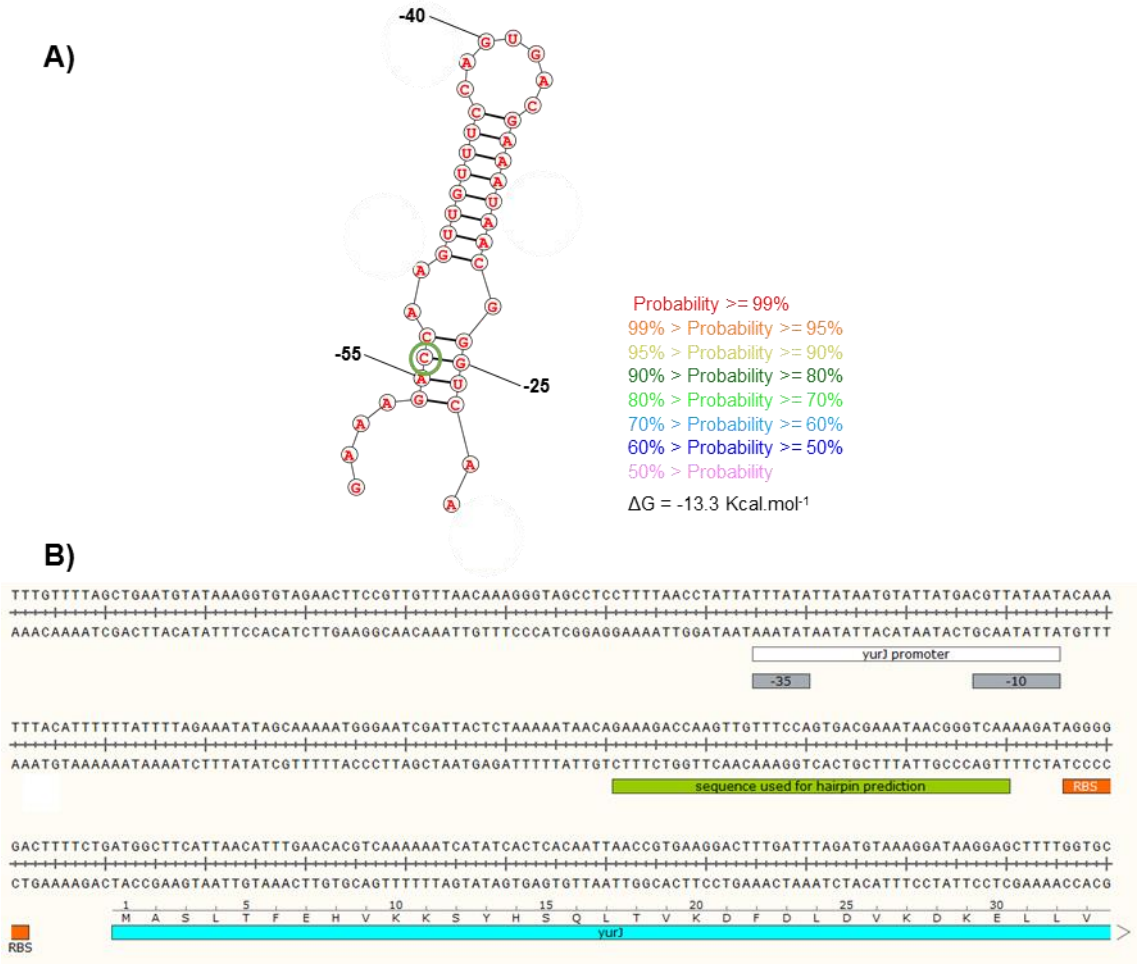


Figure 3.4 – Secondary structure prediction of the *yurJ* transcript in the wild-type strain. A) A putative hairpin in the mRNA from *yurJ* is predicted with a free energy change of -13.3 Kcal.mol^{-1}. This structure is located eight bp upstream from the RBS. The secondary structure prediction was obtained using the online tool RNAstructure and base-pair positions indicated are relative to translation initiation (http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html; Mathews, 2006). The probability for base-pairing or unpairing nucleotides is indicated. The substituted nucleotide (**C→A)** is highlighted with a *green* circle. **B)** Promoter region of the *yurJ* gene in the wild-type strain (168T^{+}). The *yurJ* ORF (*cyan*), the putative *yurJ* promoter (*white*, with the -10 and -35 boxes signaled in *gray*), predicted using the online tool Virtual Footprint (http://www.prodoric.de/vfp/vfp_promoter.php; version 3.0), the putative ribosome-binding site (RBS, *orange*) and the sequence used for hairpin prediction (*green*) are highlighted.

Due to its considerably low Gibbs free energy Δ*G* (-13.3 Kcal.mol^{-1}) and position close to the RBS, we hypothesized that this structure could interfere with ribosome-binding to the mRNA and, consequently, blocking translation. Furthermore, a similar secondary structure analysis was

performed for strain IQB642, in which functional complementation of MsmX by YurJ had been observed, and failed to detect a hairpin structure with such low free energy (see Supplementary Figure 1 in Annexes). To investigate the potential role of the secondary structure located at the 5' end of the *yurJ* transcript in a mechanism of post-translational regulation of this gene we disrupted the putative hairpin in the wild-type strain by substituting a single cytosine by an adenine (Figure 3.4). This mutation increases the free energy of the structure by about 4.9 Kcal.mol^{-1} when compared to the wild-type transcript. The single point change was introduced by allelic replacement in the chromosome of the *B. subtilis* Δ*msmX*::*cat* background, and the new double mutant strain (IQB648) was grown in minimal medium in the presence of arabinotriose as sole carbon and energy source. Although the putative hairpin containing the C→A substitution displays a higher calculated free energy (about -8.4 Kcal.mol^{-1}), strain IQB648 was unable to grow, presenting a similar phenotype to that observed for the single Δ*msmX*::*cat* mutant (data not shown). This observation suggests that this structure is not involved in the mechanism of post-transcriptional regulation of *yurJ* or, that the single mutation may have had a negligible impact in the overall stability of the putative hairpin, which may still be accountable for blocking translation in the conditions tested.

Identification of new carbohydrates imported via MsmX-dependent systems

We previously established MsmX as a multipurpose ATPase, which is solely responsible for energizing distinct ABC-type systems involved in the uptake of oligosaccharides (Chapter II). Since MsmX has been proposed to associate with 7 distinct *B. subtilis* ABC-type importers (Quentin *et al.*, 1999), we used an *msmX*-null mutant strain to determine the impact this mutation on the ability of the cell to grow on different sugars as sole carbon and energy source. Both the wild-type strain (168T^{+}) and the Δ*msmX*::*cat* strain (IQB495) were grown in the presence of sugars commonly found in the hemicellulosic fraction of plant cell wall, namely pectin, galactan, rhamnogalacturonan type I, polygalacturonan and galacturonic acid, and the results are summarized in Table 3.6. The ability to utilize pectin was completely abolished in the *msmX*-null mutant when compared to the wild-type. Since this polysaccharide is highly complex, branched, and composed by various different sugars, mainly arabinose, galactose, rhamnose and galacturonic acid, this result strongly suggests that MsmX is involved in the uptake of most of these sugars found in pectin. Arabinose uptake as arabinosyl oligomers was previously shown to be MsmX-dependent (Chapter II) and the uptake of galactose oligosaccharides was inferred from results presented and discussed in Chapter II. Here, we show that growth in the presence of galactan is abolished in the absence of a functional MsmX (Table 3.6), assigning this ATPase as a critical component for the utilization of galactan by *B. subtilis*. MsmX is likely to be responsible for energizing an ABC-type importer dedicated to the uptake of galactooligosaccharides resulting from the extracellular degradation of galactan.

Table 3.6 – Growth of a *B. subtilis msmX*-null mutant in the presence of different mono- and polysaccharides. Doubling time (minutes) for different strains in liquid minimal medium (CSK) using pectin, galactan, rhamnogalacturonan I, polygalacturonan or galacturonic acid as sole carbon and energy source. Results are the averages of three independent assays and their respective standard deviations. NG, no growth.

	168T⁺ (wild-type)	**IQB495 (Δ*msmX*::*cat*)**
Pectin 0.1%	153.1±12.4	NG
Galactan 0.1%	130.7±6.5	NG
Rhamnogalacturonan I 0.1%	189.3±15.3	603.9±32.6
Polygalacturonan 0.1%	123.5±4.4	365.6±44.7
Galacturonic Acid 0.1%	77.3±0.6	79.5±1.1

The absence of a functional MsmX also has a significant negative impact on the growth of *B. subtilis* in minimal medium supplemented with rhamnogalacturonan typeI or polygalacturonan (Table 3.6). Unlike the complete loss of growth observed when arabinan or galactan are used as sole carbon and energy source, a slower but steady growth rate is detected for the *msmX*-null strain (IQB495) in the presence of either rhamnogalacturonan type I or polygalacturonan. Interestingly, the growth rate decrease in the mutant strain (IQB495) when compared to the wild-type is very similar for both polysaccharides, about 3-fold (Table 3.6). Since galacturonic acid is the major component of both polysaccharides, the effect of the Δ*msmX*::*cat* mutation in the uptake of this monosaccharide was also assessed, and the results presented in Table 3.6 show that the utilization of galacturonic acid is MsmX-independent. Thus, the negative impact observed when rhamnogalacturonan type I or polygalacturonan were used is likely due to at least one MsmX-dependent ABC-type importer responsible for the uptake of galacturonic acid oligomers and/or rhamnose-galacturonic acid disaccharides but not galacturonic acid.

Functional characterization of new ABC-type sugar importers

The results presented so far establish MsmX as a key element for the transport of multiple sugars, mainly oligosaccharides, found in pectin. The utilization of arabinan, galactan, rhamnogalacturonan type I and polygalacturonan by *B. subtilis* is completely or partly MsmX-dependent. Although only the specific transport system for arabinan, more specifically for arabinosyl oligomers, has been identified as the ABC-type importer AraNPQ (Chapter II), our results imply that additional transporters from this class are involved the uptake of the other polysaccharides. The CycB (formerly YvfK) solute-binding protein (SBP) was previously described as a cyclodextrin-binding protein (Kamionka and Dahl, 2001). It was reported that this *B. subtilis* SBP displayed affinity for all three naturally occurring cyclodextrins (α, β and γ) and, contrarily to what had been shown for the cyclodextrin-binding protein CymE from *Klebsiella oxytoca* (Pajatsch *et al.*, 1998), its affinity increases with the increasing size of the cyclodextrin (Kamionka and Dahl, 2001). CycB is encoded by the *cycB-ganPQAB* operon, which is thought to

be involved in the utilization of galactan, also encoding a putative β-galactosidase (GanA) and a putative arabinogalactan endo-β-1,4-galactanase (GanB), as well as two ABC-type TMD proteins (GanP and GanQ), which are likely to be part of the same system as CycB (Shipkowsky and Brenchley, 2006; Chai *et al.*, 2012). Thus, we tested the impact of deleting the *cycB* gene in the ability to utilize galactan. The obtained results show that the Δ*cycB* mutant (IQB638) displays a 2-fold increase of its doubling time in minimal medium supplemented with galactan as the sole carbon and energy source, when compared to the wild-type strain (Table 3.7). Therefore, CycB is involved in the utilization of galactan by *B. subtilis*, and is likely to be part of a putative CycB-GanPQ ABC-type importer responsible for the uptake of galactooligosaccharides released by the enzymatic breakdown of galactan by the extracellular β-galactosidase GanB.

Table 3.7 – Uptake of galactan in *B. subtilis*. Doubling times (minutes) for different strains in liquid minimal medium (CSK) using glucose or galactan as sole carbon and energy source. Results are the averages of three independent assays and their respective standard deviations. NG, no growth.

	168T⁺ (wild-type)	**IQB638 (Δ*cycB*)**	**IQB639 (Δ*cycB* Δ*araNPQ*)**	**IQB641 (Δ*cycB araA**E305A)**
Glucose 0.1%	55.5±1.3	52.6±0.1	52.1±1.2	54.2±2.1
Galactan 0.1%	130.7±6.5	258.3±29.6	NG	532.8±2.2

We also tested a double Δ*cycB* Δ*araNPQ* mutant strain (IQB639) in the presence of galactan and, surprisingly, its ability to grow was completely lost (Table 3.7). Although unexpected, these results are similar to what had been previously observed for the Δ*araNPQ* and Δ*cycB* Δ*araNPQ* mutants grown in the presence of branched arabinan (discussed in Chapter II), and again like arabinan was shown to have a large fraction of galactose in its composition, probably the used galactan has a higher arabinose content. To confirm that the observed growth of the Δ*cycB* mutant in the presence of galactan was due to the presence of arabinose, we constructed an Ara⁻ strain by substituting the key catalytic residue E305 of L-arabinose isomerase (AraA) by an alanine. This glutamate residue from the active site is putatively involved in proton transfer, and its substitution by an alanine was previously shown to completely block the isomerization of L-arabinose to L-ribulose, thus preventing its metabolism (Kim *et al.*, 2010). When grown in minimal medium supplemented with galactan, the Δ*cycB araA**E305A strain (IQB641) presents a very slow growth rate (Table 3.7), which confirms that the growth observed for the Δ*cycB* mutant strain (IQB638) was most probably due to the presence of arabinose.

Polygalacturonan and rhamnogalacturonan type I are the two major components of pectin. The former is a linear chain of D-galacturonic acid units, while the latter has a backbone composed of alternating L-rhamnose and D-galacturonic acid units highly decorated with mainly arabinans and galactans as side chains (Ochiai *et al.*, 2007; Edwards and Doran-Peterson, 2012). Ochiai *et al.* (2007) previously showed that the genes encoding the putative YesOPQ and YtcQP-YteQ ABC-type importers were induced in the presence of rhamnogalacturonan type I. Since these transporters were proposed to be energized by MsmX (Quentin *et al.*, 1999), and this ATPase

was shown to be essential for the utilization of these polysaccharides by *B. subtilis* (Table 3.6), we constructed in-frame Δ*yesOPQ* and Δ*ytcQ* mutations and assessed their effect in the uptake of polygalacturonan and rhamnogalacturonan type I. The results presented in Table 3.8 show that both mutations had a negative impact in the utilization of both polysaccharides, and that this effect was larger in the Δ*ytcQ* mutant (IQB632) than in the Δ*yesOPQ* strain (IQB628). This indicates that both YesOPQ and YtcQP-YteQ are likely to be involved in the uptake of polygalacturonan and rhamnogalacturonan. However, the double mutant strain (Δ*ytcQ* Δ*yesOPQ,* IQB633) displayed no increase in the doubling time in the presence of either substrate when compared to the single Δ*ytcQ* mutant (Table 3.8). This observation may be due to Δ*yesOPQ* having only a marginal effect in the uptake of these substrates, which is likely to be negligible in the double Δ*ytcQ* Δ*yesOPQ* mutant since the Δ*ytcQ* mutation reveals a significantly higher impact in the utilization of both polysaccharides.

Table 3.8 – Uptake of pectin-based polysaccharides in *B. subtilis*. Doubling time (minutes) for different strains in liquid minimal medium (CSK) using glucose, pectin, polygalacturonan or rhamnogalacturonan I as sole carbon and energy source. Results are the averages of three independent assays and their respective standard deviations.

	168T⁺ (wild-type)	**IQB628 (Δ*yesOPQ*)**	**IQB632 (Δ*ytcQ*)**	**IQB633 (Δ*ytcQ* Δ*yesOPQ*)**
Glucose 0.1%	55.5±1.3	54.5±0.4	53.3±1.8	54.7±3.4
Pectin 0.1%	153.1±12.4	147.7±14.7	139.0±19.9	154.7±27.8
Polygalacturonan 0.1%	123.5±4.4	153.0±19.0	186.8±16.8	180.6±17.8
Rhamnogalacturonan I 0.1%	189.3±15.3	225.5±23.6	495.9±54.3	491.9±61.8

Additionally, for both polysaccharides, the growth rate of the double Δ*ytcQ* Δ*yesOPQ* mutant (Table 3.8) is higher than what had been observed for the Δ*msmX*::*cat* mutant (Table 3.6). This was expected for rhamnogalacturonan I, since the uptake of the oligosaccharides released from the breakdown of its arabinan and galactan side chains is MsmX-dependent. However, the same was not expected for polygalacturonan because is an unbranched homopolysaccharide composed of galacturonic acid units. It is thus probably that this substrate might have been contaminated during its purification process with a mixture of other sugars commonly found in pectin-rich substrates. Nevertheless, it is clear that the multitask ATPase MsmX is also able to energize the YesOPQ and YtcQP-YteQ systems (Table 3.6 and Table 3.8).

Identification of key residues for the interaction between MsmX and the TMDs of AraNPQ

Most ABC-type systems of *B. subtilis* possess its own dedicated NBD, but a few, particularly those involved in sugar uptake, share a single ATPase among them (Quentin *et al.*, 1999; Schönert *et al.*, 2006; Chapter II). This characteristic is not exclusive of *B. subtilis*, having been also described for other ABC sugar importers in *Streptomyces lividans* (Hurtubise *et al.*,

1995), *Streptococcus pneumoniae* (Marion *et al.*, 2011) and *Streptococcus suis* (Tan *et al.*, 2015). Systems relying on the same ATPase must have similar amino acid (and/or structural) determinants that mediate NBD-TMD interactions. In ABC importers, these interactions occur mainly, but not exclusively, between specific conserved motifs: the "Q loop" of the NBD, which is characterized by a highly conserved glutamine residue; and the "EAA" sequence motif (consensus $EAAX_3GX_9IXLP$) in the last cytoplasmic loop of the TMD (Daus *et al.*, 2007; Davidson *et al.*, 2008; Bordignon *et al.*, 2010). In the *E. coli* ABC-type maltose importer ($MalEFGK_2$) the determination of its complete structure allowed the identification of the amino acids directly responsible for the NBD-TMD contacts (Oldham *et al.*, 2007). Complete crystal structures for *B. subtilis* MsmX-dependent systems are yet unavailable, but it is known that members belonging to the ABC family share a similar structural arrangement. Thus, the general sequence homology between $MalEFGK_2$ and *B. subtilis* systems energized by MsmX (AraNPQ, MdxEFG and CycB-GanPQ), combined with the analyses of sequence alignments between the TMDs (Figure 3.5) and NBDs (see Supplementary Figure 2 in Annexes) of these systems and the glycine betaine importer OpuA of *B. subtilis* (Kempf and Bremer, 1995), allowed the identification of residues in AraP and AraQ which are likely to be involved in the contacts with MsmX.

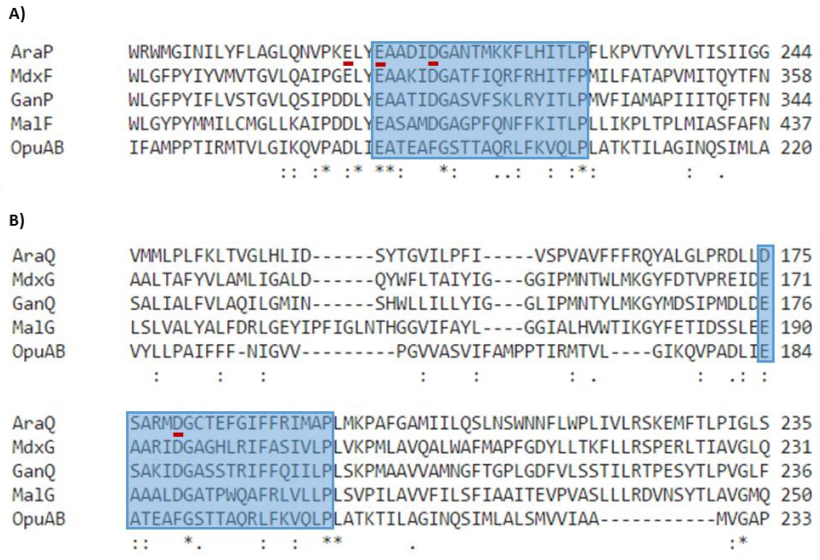


Figure 3.5 – Partial sequence alignment of the TMDs from the Ara, Mdx, Gan, Mal and OpuA ABC transporters. The sequence alignments were obtained with the online tool Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). The "EAA" sequence motifs are highlighted by *blue* boxes and residues selected for mutagenesis are underlined in *red*. **A)** TMD1 and **B)** TMD2 from the Ara, Mdx, Gan, Mal and OpuA.

Similarities between the MsmX-dependent systems and MalK from *E. coli* and differences between these and the OpuA system possessing its own NBD were key for determining which amino acids should be targeted. Single alanine substitutions of residues E205, E208 and D213 in AraP and D180 in AraQ were generated and their effect in binding between MsmX and the TMDs was inferred by the impact in arabinotriose uptake (Table 3.9).

Table 3.9 – Effect of mutations in AraP and AraQ in the uptake of α-1,5-arabinotriose. Doubling times (minutes) for different strains in liquid minimal medium (CSK) using glucose or arabinotriose as sole carbon and energy source. Results are the averages of three independent assays and their respective standard deviations.

	168T⁺ (wild-type)	**IQB623 (*araP**E208A)**	**IQB624 (*araP**E205A)**	**IQB625 (*araP**D213A)**	**IQB626 (*araQ**D180A)**	**IQB627 (Δ*araQ*)**
Glucose 0.1%	55.5±1.3	56.5±1.2	53.1±0.1	53.2±1.0	54.9±1.7	56.1±2.3
Arabinotriose 0.1%	98.2±10.0	114.0±4.8	94.1±5.0	357.0±25.3	149.3±12.8	472.8±22.6

These results show that the single E205A mutation in AraP had no effect on the growth rate of *B. subtilis*, while the E208A substitution resulted in a small negative impact in the uptake of arabinotriose, has shown by the slight increase of about 10% in the doubling time of strain IQB623 (*araP**E208A) when compared to the wild-type (Table 3.9). This is in accordance with results published by Mourez *et al.* (1997) for the *E. coli* maltose transporter, where, for each of the TMDs, a single alanine substitution of the residue homologous to E208 from AraP (E401 in MalF and E190 in MalG) had little to no impact in the uptake of maltose. Although E205 and E208 from AraP are still likely to establish contacts with MsmX, they are shown not to be essential for TMD-NBD interaction. Additionally, the same residues are also conserved in the MsmX-independent OpuAB, which further evidences that these amino acids are unlikely to be responsible for the specific binding of the ATPase to the transporter. Contrarily to these, the D213A mutation in AraP severely impacted the uptake of arabinotriose and, consequently, the growth rate of *B. subtilis* decreased approximately 3.6-fold. Likewise, although not as adversely, the homologous D180A substitution in AraQ also negatively impacted the growth rate of *B. subtilis* in the presence of arabinotriose by about 1.5-fold (Table 3.9). In the maltose transporter these aspartate residues are conserved in MalF (D406) and MalG (D195) and both contact four amino acids of MalK: Y87, F98, G99 and R146 (Oldham *et al.*, 2007). In comparison, E401 in MalF interacts with three amino acids from MalK (R47, L52 and F81) and E190 in MalG only contacts two (R47 and L52). The reduced number of putative contacts established between the TMDs and MsmX via E208 in AraP, when compared to those expected to be established through D213 in AraP, might explain why the E208A mutation had a negligible impact in the ability to grow in the presence of arabinotriose. According to these results, the loss of E208 in AraP is not critical to keep the MsmX dimer bound to the TMDs, but when the aspartate in position 213 is lost the remaining contacts are probably not able to keep the NBD dimer in its correct position, thus hindering sugar uptake.

In the *E. coli* maltose transporter the Q loop of MalK establishes additional contacts with the C-terminal end of MalG. This tail, which is not found in MalF, inserts itself along the MalK dimer interface and its four terminal amino acids (-GVKG) interact with residues from the Q loop of both NBDs (Oldham *et al.*, 2007). Interestingly, the primary sequence of this C-terminal tail from MalG is highly conserved in AraQ, MdxG and GanQ but not in the MsmX-independent OpuAB (Figure 3.6).

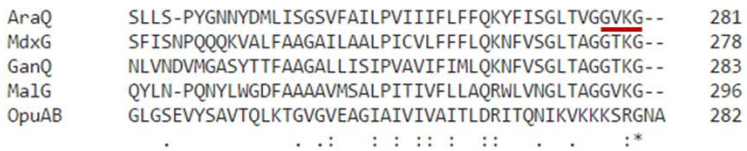


Figure 3.6 – C-terminal end of the sequence alignment of TMD2 from the Ara, Mdx, Gan, Mal and OpuA ABC transporters. The sequence alignments were obtained with the online tool Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Residues selected for deletion are underlined in *red*.

Since this feature was only found in systems that are associated with MsmX we deleted the four terminal amino acids of AraQ and assessed the impact of their absence in the uptake of arabinotriose. This deletion resulted in a severe increase of almost 5-fold in the doubling time of *B. subtilis* strain IQB627(Δ*araQ*) when compared to that of the wild-type (Table 3.9). It is possible that like in the maltose system there is an insertion of the C-terminal tail of AraQ along the MsmX dimer interface playing an important role in the overall organization and stabilization of the entire AraNPQ-MsmX complex. Furthermore, in *B. subtilis* this feature is not found in other sub-families of ABC-type transporters (with the exception of sub-family 5a) and might be a requirement for the interaction between the ATPase MsmX and the TMDs from systems it energizes. However, we cannot exclude the possibility that the small deletion generated in AraQ causes a destabilization of the protein leading to degradation.

Discussion

The search for carbon is a constant in the life cycle of every known species. Heterotrophic microorganisms as a whole are capable of using the vast majority of carbohydrates available in Nature and, as such, play a key role in the global carbon cycle. Saprotrophic bacteria like *B. subtilis*, commonly found on the soil, participate in the decomposition of plant biomass and are able to take advantage from the large variety of available sugars released by this decomposition process. In the soil, a highly competitive habitat, microorganisms must be able to efficiently uptake the available carbohydrates in order to proliferate – they either thrive or perish. The available sugar sources are usually highly complex polysaccharides with extremely high molecular mass, and its immediate uptake is virtually unattainable. To be able to use these carbon sources, organisms like *B. subtilis* must first be able to release smaller sugar monomers and/or oligomers that can then be transported. And at this crucial stage, after enzymatically breaking down the large polysaccharides, they must be highly efficient in the uptake of those sugars, or risk losing them for another opportunistic microorganism. As such, it is of major importance that, as soon as the less complex sugars become available, they are rapidly transported inside the cell. For that purpose, *B. subtilis* (Chapter II) and other bacterial species (Hurtubise *et al.*, 1995; Marion *et al.*, 2011; Tan *et al.*, 2015) developed carbohydrate transport networks which are responsible for the

uptake of various different sugars using similar systems – energized by a common component – differing only in substrate specificity.

In *B. subtilis* there are three identified ABC-type sugar importers that are energized by the same ATPase, MsmX: AraNPQ, responsible for the uptake of arabinooligosaccharides (Chapter II); MdxEFG, involved in the transport of maltodextrins (Schönert *et al.*, 2006); and CycB-GanPQ, which is the major transporter of galactooligosaccharides. Additionally, the results presented here also strongly suggest that the previously uncharacterized YtcQP-YteQ and YesOPQ systems are also energized by MsmX, as proposed by the work of Quentin *et al.* (1999). Furthermore, we confirmed that these systems play a role in the transport of polygalacturonan (or galacturonic acid oligomers) and rhamnogalacturonan type I (or rhamnose-galacturonic acid disaccharides), as previously hypothesized by Ochiai *et al.* (2007). According to a microarray analysis on the utilization of carbohydrates by an alkaliphilic *Bacillus* sp., genes with high identity to *yesOPQ* and *ytcQP-yteQ* were shown to have increased expression levels in the presence of pectin (Song *et al.*, 2013), further supporting the role of these transporters in the uptake of polygalacturonan and rhamnogalacturonan. The work presented here assigns MsmX as a key player in the uptake of pectin-rich substrates by *B. subtilis*. MsmX is now associated to five different ABC-type sugar importers (Figure 3.7).

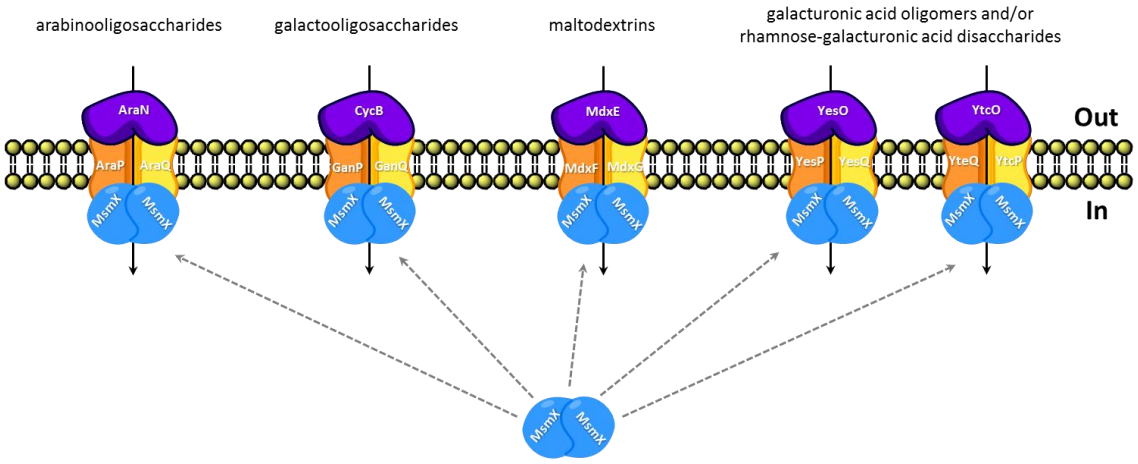


Figure 3.7 – *B. subtilis* ABC-type sugar importers energized by MsmX. With the exception of the MdxEFG system, which is involved in the uptake of maltodextrins, the remaining four transporters known to be energized by MsmX play a role in the uptake of carbohydrates resulting from the breakdown of pectin. AraNPQ and CycB-GanPQ are respectively responsible for the uptake of arabino- and galactooligosaccharides from the side chains of pectin; YesOPQ and YtcQP-YteQ are involved in the uptake of galacturonic acid oligomers and/or rhamnose-galacturonic acid disaccharides.

To understand why carbohydrate ABC-type importers share a common ATPase, in contrast with most other importers and all exporters encoded in the *B. subtilis* chromosome, we may speculate it represents an advantage in the constant cellular need of carbon and to the fact that in plant biomass sugars are often found as heteropolysaccharides, i.e., sugars like arabinose, galactose or xylose are usually present in the same substrates. Unlike glucose, which is the

premier source of energy for most heterotrophic organisms and effectively downregulates the expression of most genes associated with the utilization of other sugars, a clear hierarchy among those is not as well defined. Although a preference of one carbohydrate over another has been reported for both *B. subtilis* (Singh *et al.*, 2008) and *E. coli* (Desai and Rao, 2010) that might not be the case for sugars transported via the MsmX-dependent ABC importers. Being often simultaneously available, oligomers of arabinose, galactose and galacturonic acid, and rhamnose-galacturonic acid disaccharides are likely to be transported at the same time. Thus, a single ATPase capable of serving the various carbohydrate-specific transport systems, like MsmX, always present in the cell would be more effective because when a certain carbohydrate becomes unavailable, the already existing constitutively expressed ATPase would be rapidly mobilized to a different transporter.

Even though MsmX appears to be the only ATPase responsible for energizing the different carbohydrate ABC transporters in *B. subtilis*, there is an additional ATPase (YurJ) encoded in its genome with the capacity to functionally replace MsmX when placed in *trans* under the control of an inducible promoter (Table 3.4). Additionally, the results presented here show that YurJ is not present in the cell in the conditions tested (Figure 3.3), despite the presence of similar levels of *yurJ* and *msmX* mRNA (Figure 3.2). Thus, a post-transcriptional regulatory mechanism controls the production of YurJ. We found a putative hairpin in the secondary structure of the *yurJ* mRNA, located just upstream of the RBS (Figure 3.4), that could be responsible for blocking the translation of YurJ. However, our attempt to disrupt this structure, through a single nucleotide substitution in this region (Figure 3.4), proved ineffective as this mutation failed to suppress the growth defect in a Δ*msmX*::*cat* mutant when grown in the presence of arabinotriose as the sole carbon and energy source. The occurrence of a post-transcriptional mechanism involving secondary structure of the mRNA molecule requires further characterization in future work. In addition, other post-transcriptional mechanisms such as anti-sense RNA should be considered.

In *Streptococcus mutans*, two ABC ATPases, MsmK and MalK, are capable of substituting each other and energize both transporters for raffinose/stachyose and for maltodextrins, even though they are encoded by genes located in the vicinity of the remaining components of their partner ABC transporter (Webb *et al.*, 2008). In *B. subtilis*, no genes for SBPs or TMDs of ABC-type importers are found close to the gene encoding MsmX, but *yurJ* is located in the close proximity of the *yurONM* genes, which encode an uncharacterized putative ABC-type importer. Although no association between YurJ and YurONM has been established, due to its genomic context this ATPase is likely to energize its neighboring system. Nevertheless, interchangeability between MsmX and YurJ similar to that observed with MsmK and MalK in *Strep. mutans* is not discarded and may occur in specific conditions which allow translation of YurJ.

YurJ was shown to be able to interact and energize the AraNPQ transporter in the absence of MsmX. Moreover, when we analyze the primary sequence of both ATPases the 21 amino acids likely to establish contacts with the TMDs of the transporters are conserved (Figure 3.1).

Therefore, it is expected that YurJ can also functionally replace MsmX in the remaining ABC sugar importers associated to it. In order to understand why some ATPases are able to bind to various ABC-type systems when most are restricted to interactions with a single transporter, several of the residues expected to mediate the TMD-NBD contacts were mutated and its effect on sugar uptake assessed. An E208A substitution in AraP displayed a small negative impact in the uptake of arabinotriose, in accordance with previously published results (Mourez *et al.* 1997) that report a similar substitution in MalF of the *E. coli* maltose transporter also showing a very limited impact. Likewise, the same mutation in MalG had identical effect, and only when both glutamate residues in MalF and MalG were substituted a significant decrease in the uptake of maltose was observed. Accordingly, a *B. subtilis* strain harboring an E208A and a D175A substitution in AraP and AraQ, respectively, is expected to be severely impaired in the transport of arabinooligosaccharides. In contrast to the E208A mutation in AraP, single D213A and D180A substitutions in AraP and AraQ, respectively, had a significant negative impact in the uptake of arabinotriose. Based on the crystal structure obtained for the complete maltose transporter of *E. coli* (Oldham *et al.*, 2007) and the sequence homology with the AraNPQ system of *B. subtilis*, we hypothesize that both D213 in AraP and D180 in AraQ are likely to establish more contacts with MsmX than those expected for E208 of AraP, and consequently more important for maintaining the correct assembly of all components in the AraNPQ-MsmX system. Another key structural feature in the ABC systems energized by MsmX is a highly conserved C-terminal tail in the second TMD of each system. This feature is also found in MalG of the *E. coli* maltose system in which the tail is inserted along the ATPase dimer interface, possibly contributing to keep the dimer in a correct position (Oldham *et al.*, 2007). In fact, upon deletion of this C-terminal tail in AraQ, a severe decrease in the growth rate of *B. subtilis* in the presence of arabinotriose was observed (Table 3.9). This tail is not found in ABC systems belonging to other sub-families, which further emphasizes its putative significance in the specific binding between MsmX and the ABC sugar importers of *B. subtilis*.

Acknowledgments

Lia Godinho constructed pLG31 and gave me a hand with Western Blotting, Graça Susete constructed pGS1 and strain IQB622, Liliana Brito constructed pLB1 and strain IQB623 and Joana Lima constructed pJL1. This work was supported by the Fundação para a Ciência e a Tecnologia [grant numbers PTDC/AGR-AAM/102345/2008 to I.S.N.; PEst-OE/BIA/UI0457/2011 to Centro de Recursos Microbiológicos; UID/Multi/04378/2013 to Unidade de Ciências Biomoleculares Aplicadas; fellowship SFRH/BD/73039/2010 to M.J.F.].

Chapter IV

Characterization of the sugar-binding protein AraN of *Bacillus subtilis*

This chapter contains data to be published in “Functional analysis of the solute binding protein AraN from *Bacillus subtilis*” (manuscript in preparation). The author of this Thesis participated in all experiments with the exception of STD-NMR (Pedro Henriques and Ângelo Figueiredo), and Molecular docking (Pedro Henriques).

Characterization of the sugar-binding protein AraN of *Bacillus subtilis*

Abstract

ABC-type importers rely on high affinity solute-binding proteins (SBP) to sense and capture substrates from the external medium, even if these are available at very low concentrations. In this work, we used STD-NMR to characterize the binding interactions between AraN, the SBP of the AraNPQ-MsmX system from *Bacillus subtilis*, and arabinooligosaccharides. We show that AraN is capable of binding arabinose oligomers with up to at least five arabinosyl units but not arabinose. The K_d values determined for these interactions are about 100-fold higher than those observed with other SBP from ABC-type sugar importers. AraN is also shown to interact with cello-, xylo- and maltooligosaccharides, although in a different site than that for the recognition of arabinose oligomers. Furthermore, by site-directed mutagenesis and *in vivo* and *in vitro* analyses a tryptophan residue (W273) in AraN was found to be critical for binding of arabinotriose.

Introduction

Bacillus subtilis is naturally found in the soil and in the digestive tract of animals where it participates in the decomposition of the carbohydrate-rich plant biomass. Sugar uptake by *B. subtilis* is accomplished through different types of transporters including ABC-type importers. Although ABC transporters are found in all three domains of life, importers mediating the uptake of nutrients are almost exclusive to prokaryotes – importers have recently been identified in plants being mainly responsible for the transport of biologically relevant molecules between the different plant tissues (Shitan *et al.*, 2003; Terasaka *et al.*, 2005; Eitinger *et al.*, 2011). Bacterial ABC sugar importers rely on high-affinity solute-binding proteins (SBP), which are capable of sensing and binding their substrates at sub-micromolar concentrations (Dippel and Boos, 2005; Nataf *et al.*, 2009; Shulami *et al.*, 2007 and 2011; Tabachnikov and Shoham, 2013). However, even at higher sugar concentrations SBPs are essential for transport to take place, since their deletion was shown to effectively prevent sugar uptake (Shuman, 1982; Chapter II).

In Gram-negative bacteria, SBPs are found roaming the periplasmic space between the inner and outer membranes. Lacking a periplasm, Gram-positive bacteria keep SBPs in close proximity to the transporters through a lipidic anchor, which binds the N-terminal cysteine of the binding protein to the external side of the membrane (Davidson *et al.*, 2008; Eitinger *et al.*, 2011). Although most SBPs exclusively bind a single substrate or a family of closely related substrates,

some have been described to display broad substrate specificity (Davidson *et al.*, 2008). The Msm system of *Streptococcus mutans* is able to recognize structurally different sugars such as melibiose, sucrose, raffinose, isomaltotriose, and isomaltotetraose (Russell *et al.*, 1992; Tao *et al.*, 1993).

In a previous study (Chapter II) we showed that the ABC-type importer AraNPQ-MsmX from *B. subtilis* is responsible for the uptake of of α-1,5-arabinooligosaccharides, at least up to four L-arabinosyl units. In this work, we demonstrate that the SBP AraN is able to specifically bind arabinooligosaccharides with up to five arabinosyl units, albeit the displayed affinity being roughly 100-fold lower than that observed for other SBPs from ABC sugar importers. Cello-, xylo- and maltooligosaccharides are also recognized by AraN, and cellotriose and xylotriose were shown to be able to interact with the binding protein simultaneously with arabinotriose, indicating a different binding site. By combined *in vitro* and *in vivo* analyses a tryptophan residue in AraN was found to be essential for binding of arabinotriose.

Materials and Methods

Substrates

α-1,5-arabinooligosaccharides (arabinobiose, arabinotriose, arabinotetraose and arabinopentaose), cellotriose, 1,4-β-D-xylotriose, maltohexaose, 1,4-β-D-mannotriose and arabinan (sugar beet) were purchased from Megazyme International Ireland, and D-(+)-glucose and L-arabinose were purchased from Sigma-Aldrich Co.

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 and used according to the manufacturers' recommendations. PCR amplifications were carried out using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific). DNA from agarose gels and PCR products were purified with the illustra™ GFX™ PCR DNA and Gel Band Purification kit (GE Healthcare Life Sciences). All DNA ligations were performed using T4 DNA Ligase (Thermo Fisher Scientific). DNA phosphorylation was performed using T4 Polynucleotide Kinase (Thermo Fisher Scientific). Plasmids were purified using the NZYMiniprep kit (NZYTech). DNA sequencing was performed with the ABI PRISM BigDye® Terminator Cycle Sequencing Kit (Applied Biosystems). The sequencing reaction was purified by gel filtration and resolved in an ABI 3730XL sequencer.

Construction of plasmids and strains

For expression and purification of AraN (without the 21 amino acids predicted as the signal peptide) fused to a C-terminal Histidine tag pZI37 was constructed. *araN* was amplified by PCR, with primers ARA145 and ARA146 (Table 4.1), from *B. subtilis* 168T$^+$ chromosomal DNA. The amplification product was then digestion with NdeI, and the two resulting products were ligated as a single 1254 bp DNA fragment. This fragment was cloned into pET30a(+) (Table 4.2) digested with XhoI, filled-in with Klenow Fragment, and then digested with NdeI (Thermo Fisher Scientific).

Table 4.1 – List of oligonucleotides used during the course of this work.

Oligonucleotide	Sequence (5'→3')	Complementary sequence
ARA145	AGTCATTCATATGTGTTCAGCAGAAAAATCATCCG	*araN*
ARA146	GGTGGTGGTGGTGGTGGTGTTTCTGTTTCAGTTCACCTGCG	*araN*
ARA775	GTTTGAAAAAGCGGCCAATTC	*araN*
ARA776	GAATTGGCCGCTTTTTCAAAC	*araN*
ARA777	TCCCATATACGCAATCGGCAT	*araN*
ARA778	ATGCCGATTGCGTATATGGGA	*araN*
ARA822	GAAGATCTCTGACATACCGCTTGCC	*araN*
ARA823	TTCAGTCGACCTGCCGCTCTGTCC	*araN*

Restriction sites in the primer sequences are underlined.

pMJ36 and pMJ37 were constructed for expression and purification of AraN-His$_6$ with single R101A and W273A amino acid substitutions, respectively. For both constructions a site-directed mutagenesis approach was used. pZI37 was used as template for two independent amplifications with primers ARA775 and ARA776 (Table 4.1; for R101A), and ARA777 and ARA778 (Table 4.1; for W273A), yielding pMJ36 and pMJ37, respectively.

Table 4.2 – List of plasmids used or constructed during the course of this work.

Plasmid	Relevant features	Source or Reference
pET30a(+)	Expression vector with an inducible T7 promoter for the expression of proteins in *E. coli*, *kan*	Novagen
pMAD	Plasmid used for allelic replacement in Gram-positive bacteria, *bla*, *erm*	Arnaud *et al*., 2004
pZI37	pET30a(+)-based vector for the expression of *araN*-His$_6$, *kan*	Franco, 2007
pMJ36	pET30a(+)-based vector for the expression of *araN**R101A-His$_6$, *kan*	This work
pMJ37	pET30a(+)-based vector for the expression of *araN**W273A-His$_6$, *kan*	This work
pMJ41	Integrative plasmid used for the construction of *araN**W273A mutants, *bla*, *erm*	This work

For the generation of a single W273A substitution in AraN of *B. subtilis*, pMJ41 was constructed. The coding sequence of *araN* with a mutation leading to a W273A substitution was partially amplified by PCR, from pMJ37, using primers ARA822 and ARA823 (Table 4.1). The obtained product was digested with SalI and BglII and the resulting 963 bp DNA fragment was cloned into pMAD digested with the same restriction enzymes. The obtained plasmid, pMJ41, was used for the generation of a mutation in *araN* in the chromosome of *B. subtilis* 168T$^+$, leading

to a single W273A amino acid substitution in the AraN protein, following the published procedure described by Arnaud *et al*., 2004. The amino acid substitution was confirmed by sequencing and the new strain was named IQB643. The transformation of *B. subtilis* was performed according to the method described by Anagnostopoulos and Spizizen, 1961.

Table 4.3 – List of *B. subtilis* strains used or constructed during the course of this work.

Strain	Genotype	Source or Reference
168T$^+$	Prototroph	F. E. Young
IQB495	Δ*msmX*::*cat*	Chapter II
IQB496	Δ*araN*	Chapter II
IQB643	*araN**W273A	pMJ41→168T$^+$

Arrows indicate transformation and point from donor DNA to the recipient strain.

Growth conditions

E. coli DH5α (Gibco-BRL) was used as host for routine DNA manipulations and *E. coli* BL21 (DE3) pLysS (Studier *et al.*, 1990) was used as host for the overproduction of recombinant AraN. All *E. coli* strains were grown in liquid Luria-Bertani (LB) medium (Miller, 1972) and on LB solidified with 1.6% (w/v) agar. Kanamycin (30 μg.mL^{-1}) and ampicillin (100 μg.mL^{-1}) were added as appropriate. *B. subtilis* strains were grown in liquid LB medium, LB medium solidified with 1.6% (w/v) agar or liquid SP medium (Martin *et al*., 1987). Erythromycin (1 μg.mL^{-1}) and X-Gal (80 μg.mL^{-1}) were added as appropriate. Growth kinetics parameters of *B. subtilis* strains 168T$^+$, IQB495, IQB496 and IQB643 (Table 4.3) were determined in CSK minimal medium in the presence of 0.1% (w/v) of glucose, arabinotriose, cellotriose or xylotriose, as described in Chapter II. The optical density (OD_{600nm}) of growing cultures was used for the determination of the doubling time for each strain in the presence of the tested sugars as sole carbon and energy sources.

Production and purification of recombinant proteins

E. coli BL21 (DE3) pLysS harboring pZI37, pMJ36 and pMJ37 were used for the purification of AraN-His_6, AraN*(R101A)-His_6 and AraN*(W273A)-His_6, respectively. Overexpression from *E. coli* BL21 (DE3) pLysS harboring pZI37 was performed by auto-induction based on the method described by Studier (Studier, 2005). Cells were grown in LB medium supplemented with 1 mM $MgSO_4$, 1x NPS and 1x 5052 (Studier, 2005), for 16 hours at 37ºC and 150 rpm, using the appropriate antibiotics. After induction, growth was stopped by chilling on ice for 15 minutes and cells were harvested by centrifugation at 7000*g* (5 minutes and 4ºC). The obtained pellets were resuspended in Start Buffer (20 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.5) and 25 U of Benzonase® Nuclease (Sigma-Aldrich Co.) were added to degrade nucleic acids and to reduce cell suspension viscosity. Cell lysis was performed by sonication with a UP100H

(Hielscher) ultrasonic processor with a coupled MS7 (Hielscher) sonotrode. Cell suspensions were placed in a plastic container surrounded by an ice bath and subjected to 15 cycles of sonication (1 minute at 80% power and 0.5 seconds pulses) followed by 1 minute rests. After sonication the resulting lysate was centrifuged at 23500*g* (45 minutes and 4ºC) and the obtained supernatant was filtered using a 0.45 µm-pore diameter filter (Sarstedt). Purification was performed by loading samples onto a 1 mL Hitrap™ Chelating HP column (GE Healthcare Life Sciences). Non-specifically bound proteins were removed with Start Buffer and elution of bound His_6-tagged proteins was performed with a discontinuous gradient of imidazole (40 mM, 100 mM, 200 mM and 500 mM). The buffer of fractions in which AraN was more than 95% pure was changed to Storage Buffer (50 mM sodium phosphate, pH 7.5) using PD-10 Desalting Columns (GE Healthcare Life Sciences). Purified proteins were lyophilized and stored dry at -20ºC.

Protein analysis

The analysis of the production, purification and molecular mass of the proteins was performed with SDS-PAGE (stained with Coomassie Blue) and using the Low Molecular Weight Protein Marker (NZYTech) as standard. Protein quantification was performed using Bio-Rad Protein Assay (Bio-Rad Laboratories, Inc). Native-PAGE was performed for analysis of the mobility of AraN in different conditions. 10 µg, 25 µg and 40 µg of AraN-His_6 and Bovine Serum Albumin (BSA; Sigma-Aldrich Co.) were run in non-denaturing 6.25% Polyacrylamide gels, with and without 0.5% sugar beet Arabinan (Megazyme International Ireland).

Isothermal Titration Calorimetry

Isothermal titration calorimetry experiments were performed using a MicroCal™ VP-ITC System calorimeter (GE Healthcare Life Sciences). Prior to experiments, protein was dialyzed extensively against the reaction buffer (20 mM Sodium phosphate, pH 7.5, 50 mM NaCL). Binding protein (25 µM) was equilibrated in reaction buffer at 30°C in the cell of the calorimeter, and subsequently, 20 or 23 injections of 10 µL of 300 µM arabinotriose, 300 µM arabinotetraose or 0.5% arabinan were performed and the heat response recorded. After subtraction of the baseline, the integrated heat responses were fitted to the single binding site model using the ORIGIN software package (Northampton, MA, USA) supplied with the calorimeter.

NMR spectra acquisition and data processing

All spectra for characterization of carbohydrates and protein-sugar interactions by STD-NMR were acquired with a 600 MHz Bruker Avance III spectrometer, equipped with a cryogenic

probe (CP-TCI), operating at 600.13 MHz (1H), and with 400 MHz Bruker Avance III spectrometer, equipped with a TXI probe, operating at 400.15 MHz (1H). Pulse sequences found in the library of the Bruker TopSpin 3.1 software were used. Spectra acquisition temperature was 25ºC, except when stated otherwise. All data was processed using the program Bruker TopSpin 3.2.

Saturation-Transfer Difference (STD)-NMR

For acquisition of STD-NMR spectra, pulse programs from the Bruker library, with or without water suppression, were used: i) *stddiff* at 25ºC for optimization of saturation time (t_{sat}), and for titulation of AraN with arabinotriose for K_d determination; ii) *stdiffesgp.3* for binding epitopes mapping. t_{sat} optimization (without suppression of residual signal from water) were performed with the 600 MHz spectrometer with 16 scans (each scan repeated 32 times), with 14336 data points having been acquired in a spectral window of 9615.4 Hz (16.0221 ppm) centered at 2823.65 Hz (4.705 ppm). Various spectra with different saturation times were acquired: 0.1, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 2.5, 3.0 and 3.5 seconds. For the on-resonance experiment, proteins were irradiated at 150 Hz, using an Eburp2.1000 selective pulse (90º pulse for 0.05 seconds). The off-resonance experiment was performed with a 20000 Hz irradiation. A mixture of 8.9 mM arabinotriose and 58 µM AraN (154:1 ligand excess), dissolved in 92.6% D_2O and 7.4% DMSO, was analyzed. 1D STD-NMR spectra were obtained by subtracting the on-resonance spectrum from the corresponding off-resonance spectrum. During protein titulation with ligand for K_d determination, protein concentration was kept constant at 25 µM, being titulated with ligand up to a 500:1 ligand excess. Spectra acquisition conditions for STD-NMR experiments were the same as described for saturation time optimization using a constant saturation time of 1.5 seconds. For binding epitopes mapping, the solution was prepared using a protein:ligand ratio of 1:100 (25 µM AraN and 2.5 mM carbohydrate), in 93.7% D_2O and 6.3% DMSO. All experiments were performed at 10ºC, except when stated otherwise. STD-NMR experiments performed using the pulse program *stddiffesg* for arabinose, arabinobiose, arabinotriose (at 10ºC, 25ºC and 37ºC), arabinotetraose, arabinopentaose, cellotriose, maltohexaose, manotriose and xylotriose. Competition experiments were performed between i) arabinotriose and maltohexaose, ii) arabinotriose and cellotriose, iii) arabinotriose and xylotriose, and iv) cellotriose and xylotriose, and keeping protein concentration constant at 28 µM.

***In silico* AraN structure prediction (comparative modeling)**

A model with a prediction of the tertiary structure of AraN was obtained using I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/, version 3.0; Zhang, 2007), by inputting the amino acid sequence of AraN-His_6 encoded in pZI37. The highest scoring of all predicted models

was used for structural homology analysis and molecular docking simulations with arabinose oligomers.

Molecular docking

Molecular docking simulations were performed using Autodock 4.2 (Morris *et al.*, 2009). The structure prediction for AraN obtained with I-TASSER (see above) was used as receptor. Arabinotriose (PDB: 2W5O; de Sanctis *et al.*, 2010) was used as ligand in *α* conformation.

Results

Determining AraN-sugar interactions specificity

Isothermal Titration Calorimetry (ITC) is the premier and most commonly used technique for the determination of the binding constants of protein-solute interactions as it allows the determination of dissociation constants (K_d) for high-affinity ligands, typically in the sub-micromolar (nano- or even picomolar) range (Velazquez-Campoy and Freire, 2006). By directly measuring the enthalpy (heat; ΔH_a) associated to the interactions between a protein and a solute, this technique enables the determination of the binding parameters, such as the binding constant (K_a), entropy (ΔS_a), free energy (ΔG_a) and binding stoichiometry (n). Thus, ITC has been widely utilized to study the thermodynamics of protein-solute interactions in several SBPs from different organisms, such as *Pyrococcus furiosus* (Evdokimov *et al.*, 2001), *Clostridium thermocellum* (Nataf *et al.*, 2009) or *Geobacillus stearothermophilus* (Shulami *et al.*, 2007 and 2011; Tabachnikov and Shoham, 2013) belonging to ABC systems dedicated to the uptake of carbohydrates.

We used ITC to determine the association constant between AraN and its known substrates, arabinooligosaccharides. In a classical ITC experiment in which a protein is titrated with a high-affinity ligand, intense heat signals are measured for each ligand injection. These signals, whether associated to the absorption or release of heat, gradually become less intense as the protein in solution becomes saturated by its ligand. Eventually, when all protein is saturated, only small heat signals, intrinsically related to injections and with constant intensity (heat of dilution), are observed (Velazquez-Campoy and Freire, 2006). Although we previously demonstrated that AraN was essential for the uptake of arabinose oligomers with up to at least four arabinosyl units (Chapter II) the results obtained from ITC experiments (data not shown) did not confirm our previous work. The small and constant heat signals detected were likely due to a “heat of dilution” effect, and as a result no heat signal associated to ligand binding was measured when titrating AraN with either arabinotriose, arabinotetraose or arabinan, indicating an apparent absence of protein-sugar interaction.

In order to confirm that the AraN-His_6 recombinant protein was able to bind arabinose oligomers, we analyzed the mobility of this protein by Native-PAGE in the presence and absence of arabinan in the gel mesh. The obtained results indicate a decrease in the mobility of AraN-His_6 when 0.5% arabinan was added to the gel (Figure 4.1).

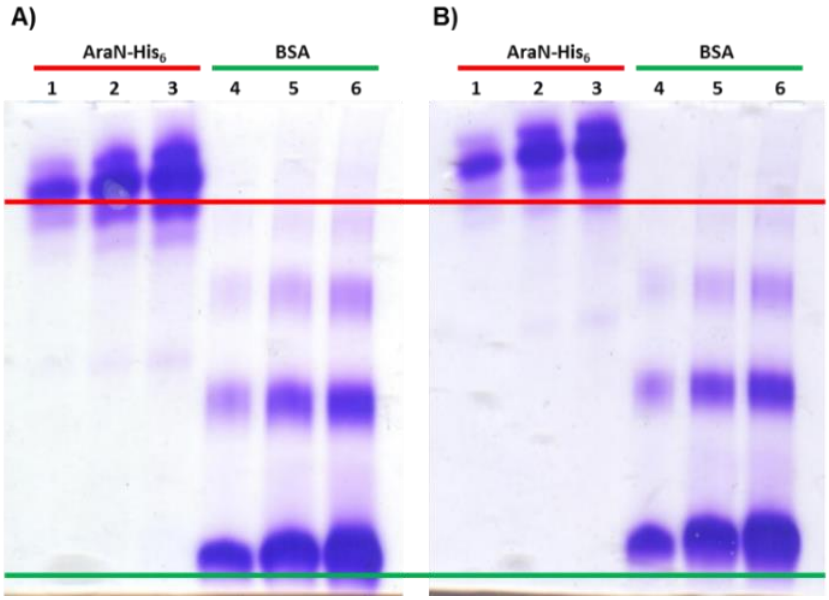


Figure 4.1 – AraN mobility assays by Native-PAGE. A) Native-PAGE (6.25%); **B)** Native-PAGE (6.25%) with 0.5% arabinan. ***1)*** AraN-His_6 (10 µg); ***2)*** AraN-His_6 (25 µg); ***3)*** AraN-His_6 (40 µg); ***4)*** BSA (10 µg); ***5)*** BSA (25 µg); ***6)*** BSA (40 µg). The *red* and *green* lines show the distance traveled by AraN-His_6 and BSA, respectively, in the Native-PAGE without sugar.

Although BSA (used as negative control) also had a slight decrease of its mobility in the presence of arabinan, that effect was significantly higher for AraN. Measuring the distance traveled by each protein in both gels we determined that without arabinan the relative mobility of AraN was 20% that of BSA. Comparatively, when arabinan was present in the gel mesh the relative mobility of AraN was 14.5% that of BSA, which strongly suggests that the presence of the polysaccharide led to a decrease of approximately 25% in the distance traveled by AraN when compared to that observed in a sugar-free Native-PAGE.

This result points to interactions between the recombinant version of AraN and arabinan. Since AraN-His_6 is able to interact with arabinan the absence of detection of interactions by ITC could be explained by the presence of glycerol in the binding pocket of AraN. During protein purification and concentration we found that the addition of glycerol or arabinotriose was essential for stabilization of AraN. In fact, when concentrating AraN in the absence of either molecule, it would rapidly reach a concentration threshold and start to precipitate. This observation led us to hypothesize that AraN could be able to non-specifically bind glycerol, affecting the utilization of ITC to determine the thermodynamics of AraN-sugar binding. AraN is likely to have higher affinity towards arabinooligosaccharides than towards glycerol, but if the binding enthalpies associated

to both ligands are very similar the thermal effect from binding of arabinose oligomers would be compensated by the thermal effect from the dissociation of glycerol, leading to a very small heat signal measurement (Velazquez-Campoy and Freire, 2006). A second hypothesis is that the binding constant between AraN and the substrates tested is too low to be determined with ITC. For a direct measurement of the interactions between a protein and a ligand using ITC, binding constants must be between 10^4 and 10^8 M^{-1} (Velázquez-Campoy *et al.*, 2004). Affinity constants (K_a) of SBPs from ABC-type importers are typically within this range, varying from 10^5 to 10^7 M^{-1}, and are therefore susceptible to measurement by ITC. However, previous reports indicated that a *B. subtilis* ABC-type SBP (CycB) is able to bind cyclodextrins, with K_d values of 100–500 µM (K_a = 0.2–1x10^4 M^{-1}) depending on the size of the cyclodextrin (Kamionka and Dahl, 2001). Although there is no indication that CycB is actually implicated in the uptake of cyclodextrins, and we previously demonstrated that this protein is part of a system involved in the utilization of galactan (Chapter III), it is possible that the affinity between AraN and arabinose oligomers is significantly lower than expected.

AraN binds arabinooligosaccharides with low affinity

In order to determine the interactions between AraN and the saccharides known to be transported by the AraNPQ-MsmX system we resorted to a different technique, Saturation-Transfer Difference (STD)-NMR (Mayer and Meyer, 1999). This technique relies on identifying hydrogen nuclei from solute molecules that become saturated, i.e., nuclei likely to be involved in protein-solute interactions. Using STD-NMR we can identify ligand and/or protein epitopes involved in binding, thus contributing to understand how arabinose oligomers interact with AraN, and, in addition, estimate the binding constants associated to these interactions – it provides a range of possible K_d values (in the same order of magnitude) in which the real dissociation constant of the sugar-AraN interaction must lie.

STD-NMR experiments were conducted using arabinose and arabinooligosaccharides with up to five arabinosyl units. The obtained results confirmed that AraN binds arabinose oligomers (with up to at least five arabinosyl units) but not arabinose (Figure 4.2). In STD-NMR spectra the absence of any signal (“flat-line” spectra) indicates that no contact between the protein and the substrate was established. According to the results displayed in Figure 4.2, AraN is unable to bind arabinose, but clearly interacts with all tested oligosaccharides (arabinobiose, -triose, -tetraose and -pentaose). Additionally, the analysis of the saturation levels for each of the tested carbohydrates and the identification of substrate epitopes involved in the binding process (data not shown) demonstrate that interactions are more prominent with arabinosyl units from the non-reducing end of arabinooligosaccharides. This evidence suggests that AraN preferably establishes contacts with the non-reducing end of arabinooligosaccharides, which is in contrast with the *E. coli* maltose transporter where the SBP (MalE) contacts the glucosyl residues from the reducing end of maltodextrins (Oldham *et al.*, 2013). In fact, the non-reducing end of maltodextrins

with five or more glucosyl units is kept outside of the binding pocket although they are still important for substrate recognition by one of the TMDs, MalF (Oldham *et al.*, 2013). The results obtained from STD-NMR experiments also indicate that arabinotriose is the strongest among all tested ligands and the one which is better stabilized when in contact with AraN. This is probably due to two main aspects, first arabinotriose establishes more contacts than arabinobiose, and secondly arabinooligosaccharides longer than three units do not establish additional interactions with AraN.

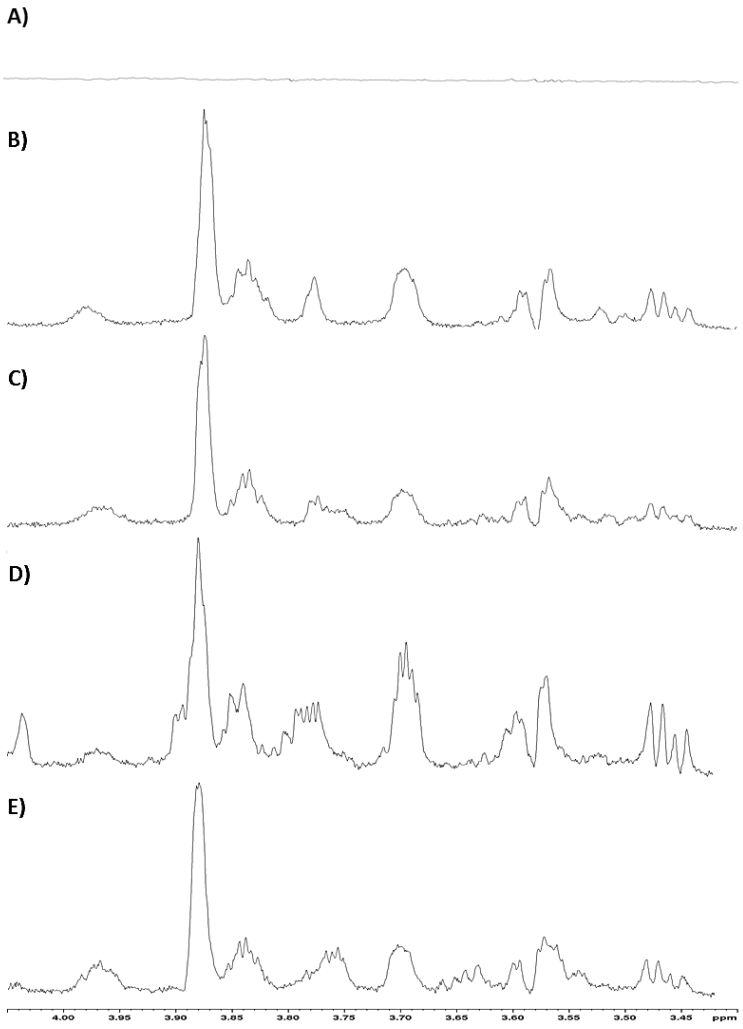


Figure 4.2 – ^{1}H-STD-NMR spectra for AraN with arabinose and arabinooligosaccharides. All spectra obtained using 2.5 mM of sugar and 25 µM of AraN. **A)** arabinose; **B)** arabinobiose; **C)** arabinotriose; **D)** arabinotetraose; **E)** arabinopentaose.

To estimate the dissociation constant (K_d) between arabinotriose and AraN, the binding protein was titulated with the sugar and saturation was determined as the ligand concentration increased. The saturation for each sugar epitope was plotted versus ligand concentration and an average K_d value of 200–500 µM (K_a = 0.2–0.5x10^4 M^{-1}) was determined from the multiple curves obtained using a non-linear fit for each epitope. The results from four of the epitopes are shown

in Figure 4.3 and confirm an association constant lower than the 10^4 M^{-1} minimum threshold for ITC measurements. This preliminary K_d estimation indicates that AraN displays a much lower affinity towards arabinotriose than expected, since other proteins from the same family are known to strongly bind their ligands.

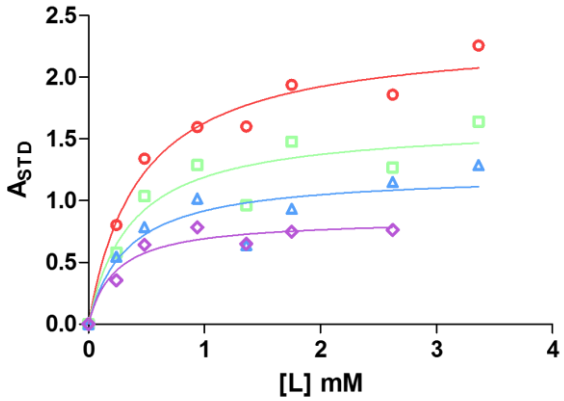


Figure 4.3 – Saturation (A_{STD}) versus ligand concentration ([L]). Saturation values from four epitopes of arabinotriose were plotted versus sugar concentration. Curves used for K_d determination for each epitope were obtained with a non-linear regression and the least squares fitting method using GraphPad Prism (version 5.00).

The role of key AraN amino acid residues for sugar binding

Although the primary structure among SBPs from ABC-type importers is not highly conserved, crystallographic data on these proteins has shown that the secondary and tertiary structures are generally similar. ABC solute-binding proteins are usually characterized by two globular lobes linked by flexible hinge regions, which create a pocket between them accessible for substrate binding (Davidson *et al.*, 2008). Being typically highly specific for a given substrate or group of substrates, solute binding by SBPs is determined by the presence of several polar and charged amino acids facing the inside of the binding pocket. In the *E. coli* maltose transporter, the amino acid residues determining substrate specificity of MalE towards maltose and maltodextrins were recently identified by Oldham *et al.* (2013). Through a number of polar residues, MalE is able to establish 14 direct hydrogen bonds and 118 van der Waals interactions with four glucosyl units from the reducing end of maltotetraose but makes no additional contacts with other glucosyl units of longer maltooligosaccharides. Considering the known contacts established by MalE and the overall structure homology found in ABC-type SBPs, in order to determine amino acids from AraN implicated in binding arabinooligosaccharides, we attempt to identify residues that participate in binding of MalE to maltodextrins and are conserved in AraN by performing an alignment between the primary sequences of both proteins (Figure 4.4). From the 10 amino acids of MalE known to contact maltooligosaccharides, only two are conserved in AraN, R101 and W273 (Figure 4.4). Additionally, three other amino acids involved in these contacts in MalE are substituted in AraN by residues with similar properties: two glutamate

residues of MalE are replaced by aspartates D79 and D198 in AraN, and one tryptophan of MalE is replaced by a tyrosine in position 391 of AraN (Figure 4.4). Thus, we hypothesized D79, R101, D198, W273 and Y391 as being putative binding sites in AraN responsible for recognizing arabinooligosaccharides and delivering them to AraPQ for transport across the membrane.

Figure 4.4 – Sequence alignment between MalE of *E. coli* and AraN of *B. subtilis*. MalE residues known to contact the first (*blue*), second (*green*), third (*red*) and fourth (*purple*) units of maltooligosaccharides (according to Oldham *et al.*, 2013) are highlighted. The sequence alignment was obtained with the online tool Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/).

To further support the hypothesis a structure prediction for AraN-His_6 was performed using the online tool I-TASSER and the most probable of the predicted structures obtained is shown in Figure 4.5. The overall structural arrangement in this prediction for AraN (Figure 4.5a, Figure 4.5c) is quite similar to the structure of MalE (Figure 4.5b, Figure 4.5d) and several conserved helices are easily detected along the structure. The two lobe arrangement and an internal cleft between them is also observed in this prediction. Additionally, residues R101 and W273 are found around what is likely to be the binding site of AraN for arabinooligosaccharides, and are located in opposite sides of this pocket, with remarkable similarity to the relative positions of the homologous residues in MalE. Although, contrarily to the arginine and tryptophan in MalE, the side chains of R101 and W273 are not pointing towards the center of the binding pocket in this prediction for AraN. However, this might be due to differences between the substrates recognized by each protein and their spatial arrangement inside the binding pocket, which would necessarily require a different positioning of the amino acids contacting the sugar, or simply to small errors resulting from the prediction itself.

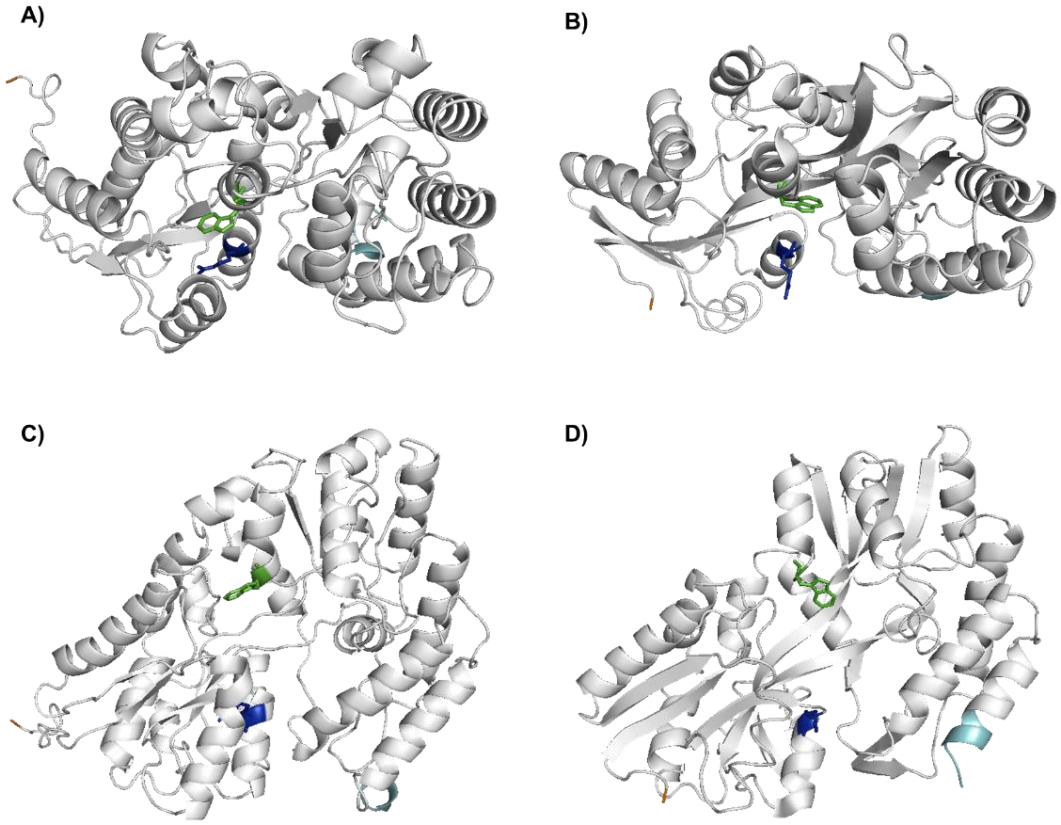


Figure 4.5 – Structure prediction for AraN-His$_6$. A) Forward-facing view of the AraN structure prediction; **B)** Forward-facing view of the MalE structure; **C)** Bottom-facing view of the AraN structure prediction; **D)** Bottom-facing view of the MalE structure. The conserved amino acids R101 (*blue*) and W273 (*green*) of AraN – as well as the homologous residues of MalE – are highlighted and their respective side chains shown. The N-terminal (*orange*) and C-terminal (*cyan*) ends are also highlighted. The structure prediction for AraN was obtained using the online tool I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/; version 3.0), from the sequence of recombinant AraN-His$_6$ used for STD-NMR experiments. The structure for MalE was obtained from PDB (accession number: 2R6G).

Molecular docking simulations between AraN (using the same structure prediction for AraN-His$_6$) and arabinotriose were also performed to determine if the sugar could be accommodated by the putative binding pocket, as well as to identify amino acids potentially involved in protein-sugar binding. According to this *in silico* analysis (data not shown), arabinotriose appears to be accommodated along the C-terminal lobe side of the binding pocket, and very close to one of the residues previously hypothesized as being involved in the interactions between AraN and arabinose oligomers, W273.

Thus, considering the multiple bioinformatics analyses performed, we initially targeted the AraN residues R101 and W273 for mutagenesis. Using a site-directed mutagenesis approach, we independently substituted both residues by alanines and purified the two AraN variants carrying the R101A or the W273A mutations. STD-NMR experiments were performed with both mutated versions of AraN, and the obtained results show that the tryptophan in position 273 is essential for the binding of arabinotriose, since no binding was observed for the W273A mutant

(Figure 4.6). On the other hand, the R101A mutation did not lead to any significant difference in the STD-NMR spectrum when compared to that of the wild-type protein (Figure 4.6), suggesting that this arginine is not an essential residue for binding of arabinotriose by AraN.

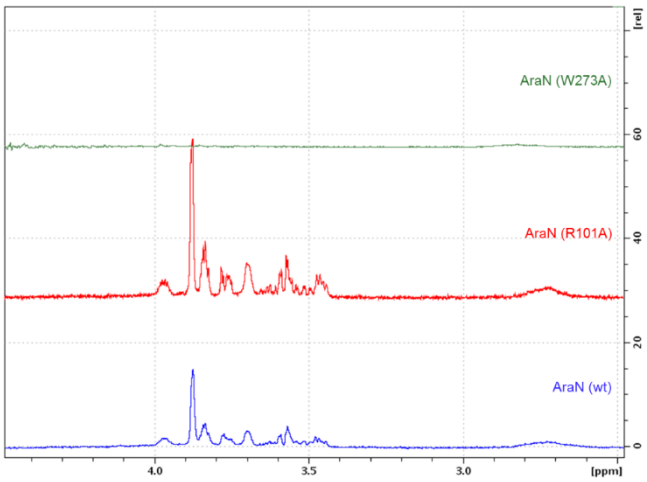


Figure 4.6 – ^{1}H-STD-NMR spectra for wild-type and mutant AraN in the presence of arabinotriose. All spectra obtained using 2.5 mM of sugar and 25 µM of AraN.

Since the W273A mutation in AraN led to a complete loss of measurable interactions between this protein and arabinotriose in the conditions tested – AraN in solution and free from the remaining proteins of the AraNPQ-MsmX complex – we decided to test the impact of this mutation *in vivo*. A new *B. subtilis* strain carrying the W273A substitution (IQB643) was constructed by allelic replacement and grown in CSK minimal medium supplemented with arabinotriose as the sole carbon and energy source. The results presented in Table 4.4 show that the W273A mutation completely hinders the transport of arabinotriose, thus supporting the findings from STD-NMR experiments and confirming this tryptophan residue as a critical element for the recognition and binding of arabinooligosaccharides by AraN.

Table 4.4 – Effect of a single W273A amino acid substitution in AraN in the uptake of arabinotriose. Doubling times (minutes) for different strains in liquid minimal medium (CSK) using glucose or arabinotriose as sole carbon and energy source. Results are the averages of three independent assays and their respective standard deviations. NG, no growth.

	168T^{+} (wild-type)	**IQB495 (Δ*msmX*::*cat*)**	**IQB643 (*araN**W273A)**
Glucose 0.1%	55.5±1.3	57.7±2.7	56.7±1.9
Arabinotriose 0.1%	98.2±10.0	NG	NG

AraN also interacts with cello- and xylooligosaccharides

STD-NMR was also used to test additional saccharides, namely cellotriose, xylotriose, maltohexaose and mannotriose, as ligands for AraN. According to the results obtained from these

experiments, AraN is also able to bind cellotriose, xylotriose and maltohexaose, but not mannotriose (data not shown). Moreover, when compared to arabinotriose, similar levels of saturation were obtained for cellotriose and xylotriose, which indicates that the K_d for these interactions is possibly in the same range of the value estimated for the AraN-arabinotriose interaction. However, AraN apparently displays lower affinity towards maltohexaose, which can be caused by the nature of the carbohydrate or simply due to its larger size (six units versus three), which may lead to a decreased binding stability.

To better assess the nature of these interactions (specific or non-specific) between AraN and other carbohydrates, STD-NMR competition experiments between these sugars were performed. These experiments consisted in the addition of an excess of "sugar 2" to a mixture already containing AraN and "sugar 1", and then checking for variations in the NMR spectra peaks which are characteristic of "sugar 1". If the addition of "sugar 2" leads to a decrease in the intensity of the peaks for "sugar 1" then both ligands are competing for the same binding site. On the other hand, if no change is detected in the peaks for "sugar 1" and, at the same time, peaks for "sugar 2" become observable, then the sugars are not competing with each other and probably bind to different sites of the protein. Thus, three different STD-NMR competition assays between arabinotriose, cellotriose and xylotriose were performed. The results for the arabinotriose/cellotriose competition assays are shown in Figure 4.7. According to these results, AraN is able to simultaneously bind arabinotriose and cellotriose, since a 10-fold excess of the latter had no significant impact in the STD-NMR peaks that are characteristic of arabinotriose, and at the same time saturation peaks from cellotriose become observable. This evidence strongly suggests that arabinotriose and cellotriose do not compete for the same binding site in AraN.

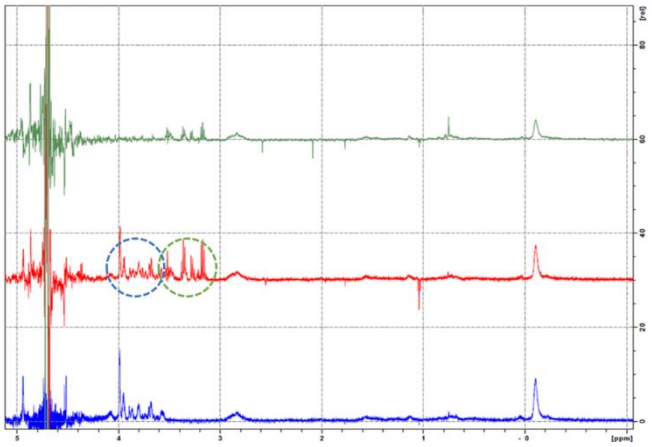


Figure 4.7 – ^{1}H-STD-NMR spectra for AraN in the presence of arabinotriose and/or cellotriose. All spectra obtained using 25 µM of AraN. *blue*, 500 µM of arabinotriose; *green*, 500 µM cellotriose; *red*, 500 µM of arabinotriose, followed by the addition of 5 mM of cellotriose. *Blue* and *green* dashed-circles highlight the characteristic peaks for arabinotriose and cellotriose, respectively.

Likewise, in arabinotriose/xylotriose competition assays the obtained results show that the addition of a 10-fold excess of xylotriose does not lead to a decrease in the STD-NMR peaks from AraN-arabinotriose interactions (Figure 4.8). Similarly to the arabinotriose/cellotriose competition assays, molecular displacement of arabinotriose by xylotriose is not observed and AraN is likely to simultaneously interact with both sugars, which are not competing for the same binding site.

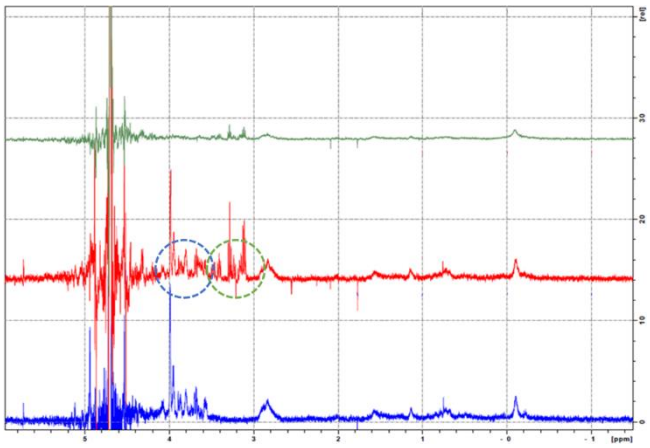


Figure 4.8 – ^{1}H-STD-NMR spectra for AraN in the presence of arabinotriose and/or xylotriose. All spectra obtained using 25 µM of AraN. *blue*, 500 µM of arabinotriose; *green*, 500 µM xylotriose; *red*, 500 µM of arabinotriose, followed by the addition of 5 mM of xylotriose. *Blue* and *green* dashed-circles highlight the characteristic peaks for arabinotriose and xylotriose, respectively.

Furthermore, to demonstrate that cellotriose and xylotriose bind to a different site than that for arabinotriose, binding of both saccharides by the AraN mutant with a W273A substitution was analyzed using STD-NMR. The obtained results show that, for both sugars, the interactions with AraN are unaffected by this mutation (Figure 4.9). Thus, unlike what had been observed with arabinotriose for the W273A AraN, this tryptophan residue is not essential for binding of cellotriose and xylotriose by AraN.

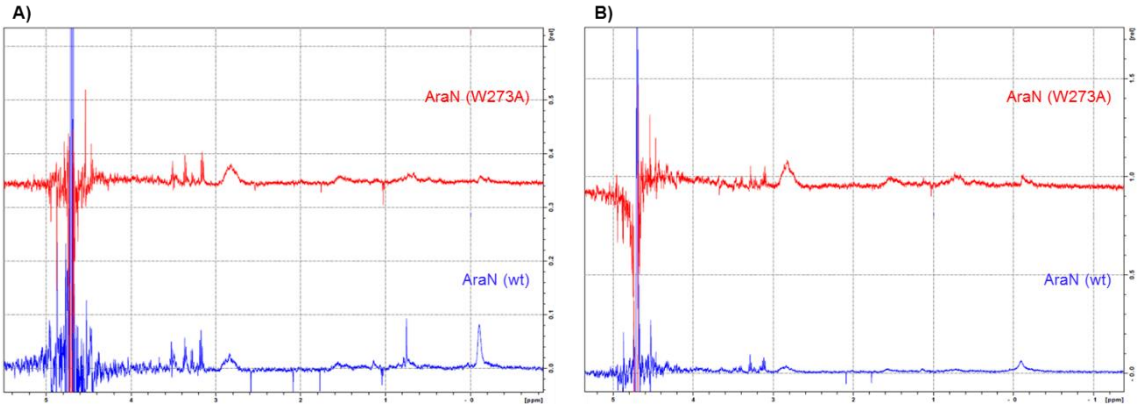


Figure 4.9 – ^{1}H-STD-NMR spectra for AraN (wild-type and W273A mutant) in the presence of cellotriose or xylotriose. All spectra obtained using 2.5 mM of sugar and 25 µM of AraN. **A)** cellotriose; **B)** xylotriose.

Since both cellotriose and xylotriose apparently bind to a different site of AraN than arabinotriose, we tried to determine if both would share a common binding site. Thus, similar competition experiments were performed, but this time AraN was initially incubated with xylotriose and then progressively an excess of cellotriose was added.

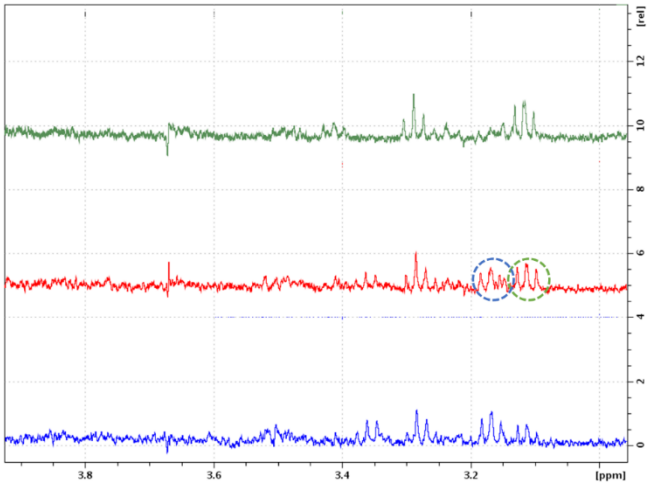


Figure 4.10 – ^{1}H-STD-NMR spectra for AraN in the presence of xylotriose and/or cellotriose. All spectra obtained using 25 µM of AraN. *green*, 500 µM of xylotriose; *red*, 500 µM xylotriose, followed by the addition of 5 mM of cellotriose; *blue*, 500 µM of xylotriose + 5 mM cellotriose, followed by the addition of 5 mM of cellotriose. *Blue* and *green* dashed-circles highlight the characteristic peaks for cellotriose and xylotriose, respectively.

According to the results shown in Figure 4.10, when a cellotriose:xylotriose ratio of 10:1 is used (*red* spectrum), apparently both sugars are simultaneously bound to AraN, and consequently, not sharing the same binding site. However, when the excess of cellotriose was increased to 20-fold (*blue* spectrum), a clear decrease in the peaks for xylotriose is observed. This suggests that both sugars are actually binding to the same site of AraN, even though it would be expected that, this being true, a 20-fold excess of cellotriose could virtually eliminate all AraN-xylotriose interactions. Moreover, AraN could have higher affinity towards xylotriose than towards cellotriose, which would explain why a large excess of cellotriose was unable to completely remove its competing sugar from their shared binding site.

The biological significance of AraN binding cello- and xylooligosaccharides

The results obtained from STD-NMR experiments indicate that AraN is able to bind cellotriose and xylotriose, albeit through a different mechanism than that for arabinotriose, i.e., these sugars do not interact with the AraN binding site for arabinooligosaccharides. In fact, our evidences point towards this SBP being capable of interacting with cellotriose or xylotriose while bound to arabinotriose. Thus, we attempt to address the biological significance of AraN binding cello- or xylotriose simultaneously with arabinotriose by performing *in vivo* analyses of the effect

of single Δ*araN* or Δ*msmX*::*cat* mutations in the uptake of cellotriose and xylotriose. *B. subtilis* strains harboring these mutations were grown in minimal medium in the presence of each of the oligosaccharides, and the obtained results are presented in Table 4.5. According to these results, neither of the tested mutations had a negative impact in the uptake of cellotriose, which means that the AraNPQ-MsmX system is not implicated in the transport of this sugar, and also that the uptake of cellotriose by *B. subtilis* is accomplished via an MsmX-independent mechanism. Nevertheless, the growth rate of the Δ*araN* mutant (IQB496) is slightly higher (roughly 10%) than that observed for both the wild-type strain and for the Δ*msmX*::*cat* mutant, and this small increase may in fact be attributed to the absence of AraN in strain IQB496.

Table 4.5 – AraN and MsmX role in the uptake of cellotriose and xylotriose. Doubling times (minutes) for different strains in liquid minimal medium (CSK) using glucose, cellotriose or xylotriose as sole carbon and energy source. Results are the averages of three independent assays and their respective standard deviations.

	168T⁺ (wild-type)	IQB495 (Δ*msmX*::*cat*)	IQB496 (Δ*araN*)
Glucose 0.1%	55.5±1.3	57.7±2.7	53.4±1.0
Cellotriose 0.1%	70.2±1.3	69.5±1.9	64.7±2.2
Xylotriose 0.1%	66.0±0.2	78.7±0.8	75.1±2.1

For xylotriose however, there was a slight decrease of 15–20% in the growth rate of the Δ*msmX*::*cat* and Δ*araN* mutant strains when compared to the wild-type (Table 4.5). These results suggest that, although an additional importer for xylotriose is likely to exist in *B. subtilis*, the AraNPQ-MsmX system may also be capable of transporting this sugar.

Discussion

B. subtilis possesses a significant number of transporters, which are exclusively dedicated to carbohydrate uptake, mostly primary and secondary active transporters and phosphotransferase systems (PTS). ABC-type sugar importers are an important class of transporters that greatly contribute for the ability of *B. subtilis* to utilize the majority of the available sugars found in hemicellulosic biomass. This class of importers is characterized by an extracellular solute-binding protein (SBP), which is attached to the outside of the cellular membrane of Gram-positive Bacteria by a lipid anchor and has generally been accepted to have very high affinity, in the sub-micromolar range (Davidson *et al.*, 2008; Berntsson *et al.*, 2010, ter Beek *et al.*, 2014), to a single substrate or family of substrates (e.g., sugar homooligomers)(Dippel and Boos, 2005; Nataf *et al.*, 2009; Shulami *et al.*, 2011). However, the results presented here strongly challenge this widely accepted concept of high affinity for SBPs from ABC-type sugar importers.

AraN, the SBP of the AraNPQ-MsmX system, was previously shown to be indispensable for the uptake of arabinooligosaccharides by *B. subtilis* (Chapter II). In this chapter, using STD-NMR, we show that AraN is in fact able to interact with arabinose oligomers with up to at least five arabinosyl residues. An attempt to determine the affinity constants for the interactions between AraN and arabinooligosaccharides by Isothermal titration calorimetry (ITC) was initially was performed without success. Although an exact K_d value was not obtained, STD-NMR experiments allowed the estimation of the dissociation constant within a range of 200–500 µM. The estimated K_d for the AraN-arabinooligosaccharides interactions is significantly higher (one to three orders of magnitude) than those from previously studied sugar-binding proteins belonging to this family. For example, high-affinity binding proteins which are involved in the recognition and uptake of various oligosaccharides by other Gram-positive Bacteria, such as *Clostridium thermocellum* (Nataf *et al.*, 2009) or *Geobacillus stearothermophilus* (Shulami *et al.*, 2007 and 2011; Tabachnikov and Shoham, 2013), typically display K_d values in the range of 0.1–10 µM. However, such a low affinity as the one observed for AraN was previously reported for *B. subtilis* binding protein CycB (Kamionka and Dahl, 2001). CycB is able to interact with α-, β-, and γ-cyclodextrins with K_d values of 100–500 µM, even though no evidence for the role of this binding protein in the uptake of cyclodextrins was provided (Kamionka and Dahl, 2001). Furthermore, we have previously shown that CycB is involved in the uptake of galactan (Chapter III), most likely through the recognition of galactooligosaccharides enzymatically released from the polysaccharide by action of the putative extracellular endo-β-1,4-galactanase GanB, which is, like CycB, encoded in the *cycB-ganPQAB* operon (Chai *et al.*, 2012). Whether CycB is involved in the uptake of cyclodextrins or can simply bind these substrates without delivering them to the cognate TMDs remains to be understood. Similarly, AraN is known to transport arabinose oligomers and was also shown to bind cellooligosaccharides without playing a role in their uptake. In *B. subtilis* cellobiose is most likely translocated via two PTS transporters LicBCA and GmuBAC (Tobisch *et al.*, 1997; Sadaie *et al.*, 2008). In addition, the uptake of cellotriose appears to have been slightly enhanced in a Δ*araN* mutant strain (Table 4.5), further evidencing binding without AraN-mediated transport. Moreover, arabino- and cellooligosaccharides are able to bind simultaneously to AraN suggesting a mechanism of inhibition in which binding of cellooligosaccharides prevents the uptake of arabinose oligomers. This is also valid for maltodextrins extending this type of inhibition to the presence of glucose-based oligosaccharides.

In the *E. coli* maltose system, the binding protein is found in an open configuration when ligand-free. Upon substrate binding, exclusively to the protein lobe that establishes contacts with the sugar through aromatic residues, a conformational change takes place leading to the closure of the binding pocket. Consequently, ATP hydrolysis is stimulated which, in turn, leads to the opening of the TMDs and allows substrate uptake (Davidson *et al.*, 2008, ter Beek 2014). In the case of AraN, we may speculate that the presence of more favorable carbon sources like cellooligosaccharides and/or maltodextrins would effectively block the uptake of arabinose oligomers by interacting with a different site inside the binding pocket and, consequently, preventing its closure and substrate delivery to the TMDs (Figure 4.11). It is possible that a similar

mechanism for regulating the uptake of galactose oligomers by CycB-GanPQ exists, in which the binding of cyclodextrins would prevent galactooligosaccharides delivery to the TMDs.

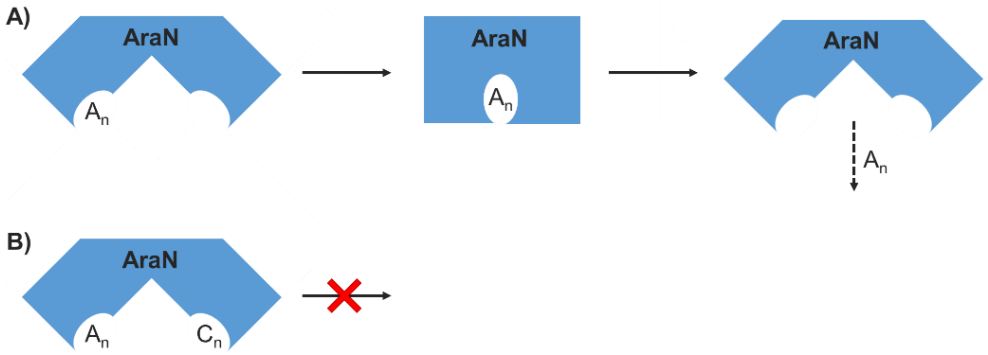


Figure 4.11 – Substrate binding and delivery by AraN. A) Ligand-free AraN is predominantly in an open conformation and is able to bind arabinooligosaccharides (A_n) through contacts established by aromatic amino acids found exclusively in one of the protein lobes, i.e., one side of the binding pocket. Upon substrate binding, the protein switches to its closed conformation, which stimulates ATP hydrolysis by MsmX and substrate uptake. **B)** When cellooligosaccharides (C_n) are available, they bind to AraN in a different site of the binding pocket, thus allowing simultaneous binding of A_n. When this happens, however, closure of the AraN binding pocket is not possible, due to spatial constraints from the presence of both sugars. As a result, ATP hydrolysis is not triggered and substrate delivery to the TMDs does not take place.

For the AraNPQ system the ability to recognize and interact with distinct sugars is different from the reports for multiple sugar ABC-type transporters from other organisms. In *Streptococcus mutans* the Msm system is involved in the uptake of structurally different substrates, like melibiose, sucrose, raffinose and isomaltotriose (Russel *et al*., 1992; Tao *et al*., 1993), but AraN might be able to bind other sugars as a regulatory mechanism for maximizing the utilization of more favorable carbon sources by *B. subtilis*.

STD-NMR experiments show that AraN can bind arabinooligosaccharides from two to five arabinosyl units, but the results obtained further indicate that arabinotriose is the favored substrate for AraN, and that the protein only establishes contacts with up to three arabinosyl units. In addition, the mode of recognition and binding by AraN is apparently opposite to that observed for the *E. coli* maltose binding protein, MalE. In the maltose system maltodextrin binding is achieved through contacts with up to four glucosyl units from the reducing end of the saccharide (Oldham *et al*, 2013), while in the AraN data indicate that substrate recognition is achieved through the non-reducing end of arabinooligosaccharides. This mode of binding has been previously observed in the cellodextrin transporter from the extremophile *Thermotoga maritima*, which is able to recognize and specifically bind the two glucosyl units from the non-reducing end of cellodextrins (Cuneo *et al.*, 2009).

Although our results suggest the preferred ligand for AraN is arabinotriose, the same may not be true for the entire AraNPQ system. In the *E. coli* maltose transporter specificity is not defined exclusively by the maltose binding protein. One of the TMDs from the maltose system,

MalF, has a cavity known to be involved in substrate recognition, and, more importantly, it is this binding site which establishes a threshold for the length of transportable maltodextrins (Oldham *et al.*, 2013). While the binding protein is capable of binding longer maltodextrins, transport is limited to molecules with no more than seven glucosyl units, due to constraints imposed by the TMDs recognition site (Davidson *et al.*, 2008). It is thus possible that a similar cavity exists in AraP (the homologous domain of MalF) and that better accommodates arabinose oligomers with a different number of arabinosyl units than arabinotriose. It is also expected that this putative cavity in AraP interacts with the reducing end of arabinooligosaccharides. Again, in the maltose transporter, interactions between MalF and maltoheptaose were shown to be achieved through contacts with the three glucosyl units from the non-reducing end of the sugar (Oldham *et al.*, 2013). If the same mechanism is responsible for substrate delivery to AraPQ, it is expected that the recognition site in the membrane domains interacts with the reducing end of arabinose oligomers, since the opposite end is involved in recognition by AraN.

Results obtained from STD-NMR experiments and *in vivo* assays showed that the tryptophan residue in position 273 of AraN is critical for binding and uptake of arabinose oligomers. According to the structure prediction for AraN (Figure 4.5), this tryptophan is found in the C-terminal lobe. Also, according to the same prediction, the other tested residue, arginine in position 101, is found in the N-terminal lobe of AraN and was shown not to be essential for arabinotriose binding. These results are in agreement with the model for the *E. coli* maltose binding protein in which sugars bind exclusively to one of the lobes of an open-form MBP, the lobe with aromatic residues (Davidson *et al.*, 2008). Nevertheless, R101 may still play a role in the binding of arabinooligosaccharides upon closure of the binding pocket and/or in the delivery of substrates to the TMDs.

In this chapter we present several contributions to the characterization and functional analysis of the AraNPQ-MsmX system from *B. subtilis*, including substrate specificity and affinity, putative transport regulation, and identification of critical residues for binding of arabinooligosaccharides. However, in order to fully understand the uptake of arabinose oligomers by AraNPQ attempts to determine the crystal structure of this transporter are currently under way in collaboration with the XTAL group at UCIBIO/REQUIMTE.

Acknowledgments

José Inácio constructed pZI37, Pedro Henriques and Ângelo Figueiredo performed the STD-NMR experiments, and Márcia Correia, Rita Otrelo, Teresa Santos Silva, Eurico Cabrita and Maria João Romão provided helpful input from our various scientific discussions about AraN. Carlos Fontes kindly allowed the use of ITC at the Animal Nutrition and Biotechnology lab. This work was supported by the Fundação para a Ciência e a Tecnologia [grant numbers PTDC/AGR-AAM/102345/2008 to I.S.N.; PEst-OE/BIA/UI0457/2011 to Centro de Recursos Microbiológicos;

UID/Multi/04378/2013 to Unidade de Ciências Biomoleculares Aplicadas; fellowship SFRH/BD/73039/2010 to M.J.F.].

Chapter V

Mechanisms controlling the expression of the *abn2* gene of *Bacillus subtilis*

All the experimental work presented in this chapter was performed by the author of this Thesis with the exception of the β-galactosidase assays.

Mechanisms controlling the expression of the *abn2* gene of *Bacillus subtilis*

Abstract

Among all genes known to be involved in the utilization of arabinose and arabinose-containing polysaccharides by *Bacillus subtilis*, *abn2* is the only not under the control of the negative regulator AraR. The results presented in this chapter show that *abn2* is constitutively expressed, albeit at a low level. Furthermore, its expression is temporally regulated, gradually increasing at the end of the exponential growth stage and during the early stages of the stationary phase. Our results also indicate that AbrB, a transition state regulator responsible for repressing many functions not needed during vegetative growth, and Spo0A, the key regulatory protein for triggering sporulation, are involved in the regulation of *abn2*. Finally, although the protein encoded by *abn2* is an arabinan-degrading enzyme, the presence of arabinose appears to repress its expression.

Introduction

Arabinose is one of the most abundant pentose in Nature and can be found as arabinans, arabinoxylans and arabinogalactans in the hemicellulosic fraction of plant cell walls and in the side chains of branched pectin polysaccharides (Caffall and Mohnen, 2009; Scheller and Ulvskov, 2010; Pauly *et al.*, 2013). Arabinose is not commonly found readily available as a monomer and to be used as a carbon and energy source by microorganisms must first be released from the more structurally complex arabinose-containnig polysaccharides.

Arabinan is a homopolysaccharide with a linear backbone of α-1,5 linked L-arabinosyl residues, branched by α-1,2 and/or α-1,3 linked L-arabinosyl chains, which can be broken-down by the concerted action of multiple enzymes (Beldman *et al.*, 1997). *Bacillus subtilis* possesses an enzymatic consortium of at least two extracellular endo-α-1,5-arabinanases, AbnA and Abn2, which cleave the backbone of arabinan and other arabinose-containing polysaccharides, releasing arabinose and arabinooligosaccharides (Raposo *et al*., 2004; Leal and Sá-Nogueira, 2004; Inácio and Sá-Nogueira, 2008). Additionally, two intracellular α-L-arabinofuranosidases, AbfA and Abf2, are able to complete the breakdown of arabinooligosaccharides to L-arabinose which can then be further metabolized (Inácio *et al*., 2008). These enzymes are encoded in the *araABDLMNPQ-abfA*, *abf2*, *abnA* and *abn2* transcriptional units of the *B. subtilis* chromosome which, with the exception of *abn2*, are negatively controlled by the arabinose metabolism regulator

AraR (Sá-Nogueira *et al.*, 1997; Raposo *et al*., 2004; Inácio and Sá-Nogueira, 2008, Inácio *et al.*, 2008). Although AbnA and Abn2 display comparable activity towards the same substrates, the mechanisms controlling their gene expression are not the same. Both *abnA* and *abn2* are controlled by the global regulator involved in carbon catabolite repression (CCR), CcpA, but in contrast to *abnA*, *abn2* is not under the negative control of AraR. However, previous work by our group suggested that pectin and arabinan were able to induce the expression of *abn2* (Inácio and Sá-Nogueira, 2008).

In this chapter we investigated the mechanisms responsible for regulating the expression of *abn2* at the transcription level. Our results indicate that the expression of *abn2* is growth phase dependent, increasing at the onset of the stationary growth stage. Additionally, the global transition state regulator AbrB (Strauch, 1993), which negatively regulates genes involved in functions not needed during vegetative growth, and the master regulator Spo0A (Strauch *et al.*, 1990; Hoch, 1993), the key protein for regulating entry in sporulation, are also implicated in the mechanisms that modulate the expression of *abn2*. Furthermore, we show that *abn2* expression is not triggered by the presence of pectin and arabinan. Finally, our results suggest that the presence of arabinose is responsible for a decrease in *abn2* expression.

Materials and Methods

Substrates

Arabinan (sugar beet) and polygalacturonan were purchased from Megazyme International Ireland. L-arabinose, D-(+)-glucose and pectin (from apple) and were purchased from Sigma-Aldrich Co.

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 and used according to the manufacturers' recommendations. PCR amplifications were carried out using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific). DNA from agarose gels and PCR products were purified with the illustra™ GFX™ PCR DNA and Gel Band Purification kit (GE Healthcare Life Sciences). All DNA ligations were performed using T4 DNA Ligase (Thermo Fisher Scientific). DNA dephosphorylation was performed with FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific). Plasmids were purified using the NZYMiniprep kit (NZYTech). DNA sequencing was performed with the ABI PRISM BigDye® Terminator Cycle Sequencing Kit (Applied Biosystems). The sequencing reaction was purified by gel filtration and resolved in an ABI 3730XL sequencer.

Construction of plasmids and strains

For the construction of a *B. subtilis* strain harboring an *abn2'-lacZ* fusion and a *lacR* mutantion, *lacR* was amplified by PCR from chromosomal DNA of *B. subtilis* 168T$^+$ using primers ARA479 and ARA480, and the obtained product was digested with SmaI and EcoRI. The resulting 1254 bp DNA fragment was cloned into pLitmus29 (Table 5.2), digested with EcoRV and EcoRI, yielding pJL4. An erythromycin resistance (Erm^R) cassette was obtained from the digestion of pDG1731 with SmaI and NaeI and subcloned into pJL4 digested with XmnI and NruI and dephosphorylated, resulting in pJL5, which was linearized with ScaI and used to transform *B. subtilis* IQB483, yielding strain IQB497 (Table 5.3).

Table 5.1 – List of oligonucleotides used during the course of this work.

Oligonucleotide	Sequence (5'→3')	Complementary sequence
ARA479	TGGAAGATGTGCTTTCG	*lacR*
ARA480	AGAATTCATAGCATGCCTGAG	*lacR*
ARA583	TCGCGGTTTCGCTGCCCTTT	*16S*
ARA584	AAGTCCCGCAACGCGCGCAA	*16S*
ARA731	GCGGACGGCAAATACTACAT	*abn2*
ARA732	TGGTTGCATCATACGGTGTT	*abn2*

Restriction sites in the primer sequences are underlined.

A *B. subtilis* strain harboring an *abn2'-lacZ* fusion and a *lacA* mutation was constructed by transformation of strain IQB497 with chromosomal DNA of a strain carrying Δ*lacA*::*spec*, IHA01 (Table 5.3), yielding strain IQB499.

Table 5.2 – List of plasmids used or constructed during the course of this work.

Plasmid	Relevant features	Source or Reference
pLitmus29	Integrative plasmid with a *lacZ* gene for blue/white selection of clones, *bla*	New England Biolabs
pDG1731	Plasmid used as a source of erythromycin resistance cassette, *bla*, *erm*, *spec*	Guérout-Fleury *et al.*, 1996
pJL4	pLitmus29-derivative harboring the *lacR* coding region, *bla*	This work
pJL5	Integrative plasmid used for the construction of *lacR* null-mutations, *bla*, *erm*	This work

Strains IQB636 and IQB637 (Table 5.3), respectively carrying *abrB* and *spo0A* null-mutations, were obtained by transformation of strain 168T$^+$ with chromosomal DNA from strains IQB425 and IQB434. In turn, IQB425 and IQB434 were obtained by transformation of *B. subtilis* MB24 with chromosomal DNAs from other *B. subtilis* strains carrying Δ*abrB*::*neo* and Δ*spo0A*::*erm* mutations, respectively. The transformations of *B. subtilis* strains were performed according to the method described by Anagnostopoulos and Spizizen, 1961.

Table 5.3 – List of *B. subtilis* strains used or constructed during the course of this work.

Strain	Genotype	Source or Reference
168T⁺	Prototroph	F. E. Young
IHA01	Δ*lacA*::*spec leuB8 metA5*	Härtl *et al.*, 2001
MB24	*trpC2 metC3*	P. J. Piggot
IQB425	Δ*abrB*::*neo trpC2 metC3*	Raposo, unpublished results
IQB434	Δ*spo0A*::*erm trpC2 metC3*	Raposo, unpublished results
IQB483	*amyE*::[*abn2-lacZ cat*]	Inácio and Sá-Nogueira, 2008
IQB497	*amyE*::[*abn2-lacZ cat*] Δ*lacR*::*erm*	pJL5→IQB483 #
IQB499	*amyE*::[*abn2-lacZ cat*] Δ*lacR*::*erm* Δ*lacA*::*spec*	DNAcr IHA01→IQB497
IQB636	Δ*abrB*::*neo*	DNAcr IQB425→168T⁺
IQB637	Δ*spo0A*::*erm*	DNAcr IQB434→168T⁺

Arrows indicate transformation and point from donor DNA to the recipient strain.
Transformation was carried out with linearized plasmid.

Growth conditions

E. coli DH5α (Gibco-BRL) or *E. coli* XL1-Blue (Stratagene) were used for the construction of all plasmids. All *E. coli* strains were grown in liquid Luria-Bertani (LB) medium (Miller, 1972) and on LB solidified with 1.6% (w/v) agar, where ampicillin (100 μg.mL^{-1}) and tetracycline (12 μg.mL^{-1}) were added as appropriate. *B. subtilis* strains were grown in liquid LB medium, LB medium solidified with 1.6% (w/v) agar, liquid SP medium (Martin *et al*., 1987), or liquid C minimal medium (Pascal *et al*., 1971) supplemented with L-tryptophan (Sigma-Aldrich Co.; 100 μg.mL^{-1}) and Casein Hydrolysate (Sigma-Aldrich Co.; 1% w/v). Chloramphenicol (5 μg.mL^{-1}), erythromycin (1 μg.mL^{-1}), spectinomycin (60 μg.mL^{-1}) and neomycin (3 μg.mL^{-1}) were added as appropriate.

B-galactosidase assays

Strains of *B. subtilis* strains harboring the transcriptional *lacZ* fusions (IQB483, IQB497 and IQB499) and the wild-type strain (168T⁺) were grown in liquid C minimal medium, supplemented with L-tryptophan and casein hydrolysate, with an initial OD_{600nm} of 0.050. When OD_{600nm} reached 0.120, arabinan or pectin were added to a final concentration of 0.4% (w/v) to induce growing cultures. Samples of cell cultures were collected 2 hours after induction (exponential growth phase, T_2) and 4 hours after induction (late exponential growth phase, T_4), and the level of β-galactosidase activity was determined as described previously (Raposo *et al.*, 2004).

RNA extraction

B. subtilis strains 168T⁺, IQB636 and IQB637 (Table 5.3) were grown in C minimal medium, at 37ºC and 180 rpm, with an initial OD_{600nm} of 0.050, and 0.4% (w/v) of arabinose, glucose,

arabinan, pectin or polygalacturonan were added when the OD_{600nm} reached 0.120. Cells were collected by centrifuging 2 mL from each of the sugar-induced and non-induced cultures (6000*g* and 4ºC for 5 minutes) at three distinct growth stages: exponential growth (1 hour before the end of exponential growth, T_{-1}), early stationary stage (1 hour after the end of exponential growth, T_1) and stationary stage (3 hours after the end of exponential growth, T_3). Total RNA extraction was performed using the Absolutely Total RNA Miniprep kit (Agilent Technologies) according to the manufacturer's instructions and using RNase-free technique. The integrity of RNA extracts was analyzed in a 1% agarose TBE 1x gel. DNA contamination of RNA samples was assessed by PCR using primers ARA422 and ARA423. Total RNA was quantified using a Nanodrop™ 1000 Spectrophotometer (Thermo Fisher Scientific) and 10 µL aliquots were stored at -80ºC.

RT-qPCR experiments and data analyses

The primers used for RT-qPCR experiments – ARA583 and ARA584 (16S gene), and ARA731 and ARA732 (*abn2*) – were designed with the help of the online tool Primer3Plus (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi). Primer efficiency was assessed using the Rotor-Gene SYBR® Green PCR Kit (QIAGEN) and RT-qPCR experiments were performed using the SensiFAST™ SYBR No-ROX One-Step Kit (Bioline), both in a Rotor-Gene 6000 (Corbett) real-time cycler. RT-qPCR experiments were performed according to the kit manufacturer's instructions, using 40 ng of total RNA and 0.1 µM of each primer, in a final volume of 12.5 µL. Statistical analyses were performed with GraphPad Prism (version 5.00) using C_t values obtained from three independent assays. *p* values were determined using an unpaired *t* test.

Results

LacR is not involved in the negative regulation of *abn2*

LacA (also known as GanA) and GanB are a β-galactosidase (GH family 42) and an endo-β-1,4-galactanase (GH family 53), respectively, proposed to be involved in the hydrolysis of galactan and arabinogalactan by *B. subtilis*. Both proteins are encoded in the *cycB-ganPQAB* operon, which is negatively regulated by LacR (Daniel *et al.*, 1997; Shipkowsky and Brenchley, 2006; Chai *et al.*, 2012). Since the *cycB-ganPQAB* operon was associated to the utilization of arabinogalactan, an arabinose-containing polysaccharide, we investigated if LacR was could also involved in the regulation of *abn2*. For this purpose, a *lacR* null-mutation was constructed in a *B. subtilis* strain carrying an *abn2'-lacZ* transcriptional fusion (IQB497; Table 5.3). , Since this mutation would lead to the constitutive expression of the *cycB-ganPQAB* operon and, consequently, to an increased level of β-galactosidase activity, an additional mutation to disrupt

the *lacA* gene was constructed (IQB499; Table 5.3) so that the effect of the Δ*lacR*::*erm* mutation on the β-galactosidase activity from the *abn2'-lacZ* fusion could be correctly assessed.

The β-galactosidase activity levels measured during exponential growth (T_2) and early stationary phase (T_4) show that LacR is not responsible for repressing transcription from the *abn2'-lacZ* fusion (Table 5.4). Even though there is an apparent derepression of the *abn2'-lacZ* fusion in strain IQB497 (Δ*lacR*::*erm*) when compared to strain IQB483, the results obtained from the double Δ*lacR*::*erm* Δ*lacA*::*spec* mutant (IQB499) show that the increased β-galactosidase activity in strain IQB497 is most probably due to the derepression of the *cycB-ganPQAB* operon, which is no longer under negative regulation by LacR and, consequently, to a higher level of ONPG-hydrolyzing LacA.

Table 5.4 – Expression from *abn2'-lacZ* transcriptional fusions on different *B. subtilis* strains. These strains were grown on C minimal medium supplemented with casein hydrolysate with or without the addition of the different sugars. Samples were collected and analyzed 2 hours and 4 hours after the addition of the inducer sugars. Levels of accumulated β-galactosidase activity (Miller units) are the averages from three independent experiments, each performed with triplicate measurements, and their respective standard deviations.

Strain	Inducer	β-galactosidase activity (Miller units)	
		T_2	T_4
IQB483 (*abn2'-lacZ*)	No sugar	1.25±0.14	1.73±0.30
	Arabinan	13.17±0.86	8.47±1.37
	Pectin	17.49±3.89	11.50±0.15
IQB497 (*abn2'-lacZ* Δ*lacR*::*erm*)	No sugar	30.02±1.89	33.94±7.15
	Arabinan	19.24±1.14	26.14±3.88
	Pectin	23.80±2.97	32.87±6.87
IQB499 (*abn2'-lacZ* Δ*lacR*::*erm* Δ*lacA*::*spec*)	No sugar	1.04±0.17	0.68±0.01
	Arabinan	0.92±0.09	0.86±0.06
	Pectin	1.41±0.02	1.03±0.08
168T$^+$ (wild-type)	No sugar	1.00±0.01	0.91±0.32
	Arabinan	11.72±0.71	14.48±6.95
	Pectin	13.11±1.52	12.60±0.19

In previous work by our group, results obtained from β-galactosidase activity assays using an *abn2'-lacZ* transcriptional fusion (strain IQB483) and total RNA Northern blot analyses suggested that the presence of either pectin or arabinan could lead to an increased gene expression (Inácio and Sá-Nogueira, 2008). However, β-galactosidase activity levels determined for the wild-type strain (168T$^+$) are very similar to those for IQB483, regardless of the presence or absence of an inducing carbohydrate (Table 5.4). This confirms that both arabinan and pectin are unable to induce expression from the *abn2'-lacZ* fusion, but are both likely to be able to derepress the *cycB-ganPQAB* operon.

The transcription level of *abn2* increases after exponential growth stage

To understand the mechanisms of growth-phase dependency of *abn2* transcription, RT-qPCR was used to measure mRNA levels at three different growth stages: exponential growth, early stationary phase, and stationary phase. The measured *abn2* transcript levels for each of the conditions tested were compared against each other and the determined expression fold-changes are presented in Table 5.5. According to these results, one hour after the end of exponential growth (T_1) the expression of *abn2* had increased by approximately 3-fold. Two hours later, when the culture was already in stationary phase (T_3), *abn2* transcript levels further increased and were over 4-fold higher when compared to the levels measured during exponential growth (T_{-1}). These results confirm temporal regulation of *abn2* transcription.

Table 5.5 – Quantification of *abn2* transcription product in strain 168T⁺. These results represent the fold-change of the expression in the target conditions versus the control conditions. (T_{-1}), (T_1) and (T_3) represent the number of hours before or after the end of exponential growth. (+ara) represents the addition of 0.4% arabinose. Primers ARA731 and ARA732 were used for *abn2*. Fold-change was normalized using primers ARA583 and ARA584 for the 16S gene. Statistical analyses were performed with GraphPad Prism (version 5.00) using C_t values obtained from three independent assays. *p* values were determined using an unpaired *t* test (^{ns}, non-significant difference; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

Control \ Target	**(T_1)**	**(T_3)**	**(T_{-1}+ara)**	**(T_1+ara)**	**(T_3+ara)**
(T_{-1})	2.856±0.249***	4.242±0.987***	0.426±0.089**	0.807±0.144^{ns}	0.145±0.045***
(T_1)	-	1.485±0.356*	-	0.282±0.052***	-
(T_3)	-	-	-	-	0.034±0.011***
(T_{-1}+ara)	-	-	-	1.895±0.335**	0.340±0.106**
(T_1+ara)	-	-	-	-	0.179±0.056***

Expression of *abn2* in the presence of different carbohydrates

Differences in the expression of *abn2* in the presence of several carbohydrates were also analyzed by RT-qPCR. In accordance with the results obtained from β-galactosidase assays, no significant variation of *abn2* mRNA levels during exponential growth was observed in the presence of arabinan or pectin (data not shown). Likewise, polygalacturonan was unable to lead to any change in the expression of *abn2*. Arabinan was, however, responsible for decreasing the expression of *abn2* by about 3-fold after the end of exponential growth (data not shown). Our results from RT-qPCR experiments using RNA extracted from *B. subtilis* cells grown in the presence of glucose also show a 5-fold decrease of *abn2* expression in the presence of this sugar (data not shown). These results are in accordance with previously published results, reporting that the expression of *abn2* (*yxiA*) is under the control of a carbon catabolite repression (CCR) mechanism (Blencke *et al.*, 2003).

The effect of arabinose on the expression of *abn2* was also analyzed by RT-qPCR (Table 5.5). The presence of arabinose results in a 2-fold decrease in *abn2* expression during exponential growth (T_{-1}+ara vs T_{-1}). A more significant negative impact was observed one hour after the end of exponential growth (T_1+ara), in which a 3.5-fold repression was observed when compared to the same growth stage in the absence of arabinose (T_1). Moreover, when the culture reached stationary phase the arabinose-mediated repression of *abn2* expression led to a 30-fold decrease in measured transcript levels (T_3+ara vs T_3). These results indicate that, contrarily to the other endo-1,5-arabinanase encoding gene in the *B. subtilis* genome, *abnA*, which is induced by the presence of arabinose, the expression of *abn2* is repressed when arabinose is available. Furthermore, arabinose-mediated repression of *abn2* significantly increases when cultures reach stationary phase, similarly to what had been observed in the presence of arabinan.

The major transition state regulator AbrB represses *abn2* during exponential growth

Gene expression from the *abn2* locus of *B. subtilis* was previously shown to be driven from two different promoters, each dependent on a different sigma factor, σ^A and σ^H (Inácio and Sá-Nogueira, 2008). A catabolite responsive element (*cre*) site was also identified and associated to CcpA binding and glucose-mediated repression. Here, using the bioinformatics tool Virtual Footprint (Münch *et al.*, 2005), a putative binding-site for the transition state regulator AbrB positioned between the σ^A-dependent transcription start site and the *abn2* start codon was identified (Figure 5.1).

Figure 5.1 – Promoter region of the *abn2* gene. The *abn2* ORF (*blue*), the *cre* site (*orange*), the final eight bases of the putative AbrB-binding site (*yellow*), the putative transcription start sites (+1, *red*), and the σ^A- and σ^H-dependent promoters (*white*, with the -10 and -35 boxes signaled in *gray*) are highlighted. AbrB-binding site prediction performed using Virtual Footprint (http://www.prodoric.de/vfp/vfp_promoter.php; version 3.0).

AbrB is a negative regulator that binds DNA, recognizing the proposed consensus sequence WAWWTTTWCAAAAAAW (Strauch, 1995) and blocking transcription of genes not

needed before the transition between vegetative growth and stationary phase in *B. subtilis* (Strauch, 1993). To determine if AbrB was responsible for repressing *abn2* expression during exponential growth, a Δ*abrB*::*neo* mutant strain was constructed (IQB636; Table 5.3). *abn2* mRNA levels in strain IQB636 were determined using RT-qPCR, at T_{-1}, T_1 and T_3, and compared to those measured for the wild-type strain ($168T^+$). The obtained results show a 2.5-fold increase of *abn2* expression in the Δ*abrB*::*neo* mutant during exponential growth (T_{-1}) when compared to the wild-type strain (Figure 5.2). This effect, however, is no longer observed when the culture reaches stationary phase (T_3) and expression of *abn2* is maximal for both the wild-type and mutant strains. These results, together with the identification of a putative AbrB-binding site, implicate AbrB in the negative regulation of *abn2* during vegetative growth.

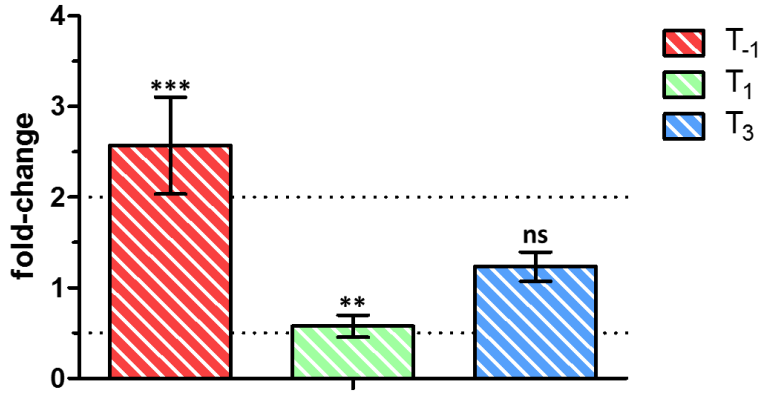


Figure 5.2 – *abn2* transcription product in strain IQB636(Δ*abrB*::*neo*) versus $168T^+$(wild-type). These results represent the fold-change of the expression in strain IQB636 versus strain $168T^+$. Upper and lower fold-change thresholds of 2.0 and 0.5, respectively, are represented by dotted lines. (T_{-1}), (T_1) and (T_3) represent the number of hours before or after the end of exponential growth. Primers ARA731 and ARA732 were used for *abn2*. Fold-change was normalized using primers ARA583 and ARA584 for the 16S gene. Statistical analyses were performed with GraphPad Prism (version 5.00) using C_t values obtained from three independent assays. *p* values were determined using an unpaired *t* test (^{ns}, non-significant difference; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

Since *abrB* is negatively regulated by the master sporulation regulator Spo0A (Strauch *et al.*, 1990) when the levels of its phosphorylated form (Spo0A~P) increase after the end of exponential growth, a new Δ*spo0A*::*erm* mutant strain (IQB637; Table 5.3) was constructed to determine the role of Spo0A in *abn2* regulation. *abn2* expression in strain IQB637 was analyzed by RT-qPCR, at T_{-1}, T_1 and T_3, and compared to the values obtained for the wild-type strain. The results indicate that the absence of Spo0A – and consequently its active phosphorylated form – leads to a 2-fold stronger AbrB-mediated repression of *abn2* during exponential growth (Figure 5.3). During early stationary growth (T_1), *abn2* repression in the Δ*spo0A*::*erm* mutant is 5-fold stronger than in the wild-type strain, but when the culture reached stationary phase (T_3) an expression decrease in the mutant strain is no longer significant (Figure 5.3). These results further support the role of AbrB in the modulation of *abn2* expression, and simultaneously show that Spo0A is needed for controlling the level of AbrB-mediated repression.

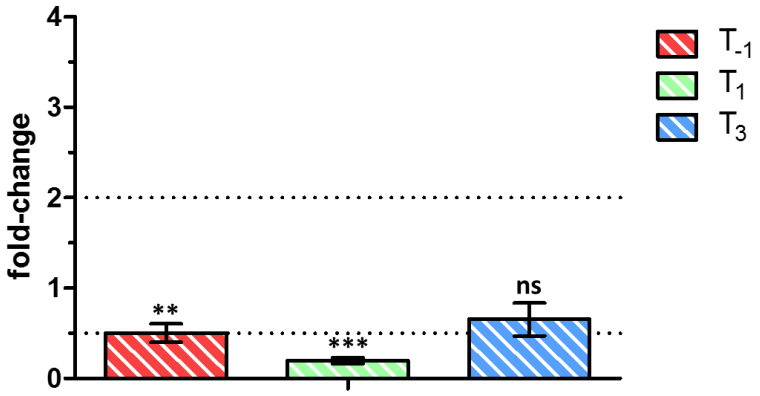


Figure 5.3 – *abn2* transcription product in strain IQB637(Δ*spo0A*::*erm*) versus 168T⁺(wild-type). These results represent the fold-change of the expression in strain IQB637 versus strain 168T⁺. Upper and lower fold-change thresholds of 2.0 and 0.5, respectively, are represented by dotted lines. (T_{-1}), (T_1) and (T_3) represent the number of hours before or after the end of exponential growth. Primers ARA731 and ARA732 were used for *abn2*. Fold-change was normalized using primers ARA583 and ARA584 for the 16S gene. Statistical analyses were performed with GraphPad Prism (version 5.00) using C_t values obtained from three independent assays. *p* values were determined using an unpaired *t* test (^{ns}, non-significant difference; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

Discussion

The *B. subtilis* chromosome encodes multiple proteins dedicated to the breakdown and uptake of arabinan and arabinooligosaccharides, and to the subsequent metabolism of arabinose. Proteins known to be involved in arabinan utilization are encoded by several genes organized as monocistronic transcription units or as an operon: *araR*, *araE*, *araABDLMNPQ-abfA*, *abf2*, *abnA* and *abn2*. Of these six transcription units, *abn2* is the only which is not under the control of AraR and not responding to the presence of arabinose (Franco *et al*., 2006; Inácio and Sá-Nogueira, 2008). The last ORF of the LacR-regulated *cycB-ganPQAB* operon, GanB, was proposed to be involved in the utilization of arabinogalactan (Daniel *et al.*, 1997; Shipkowsky and Brenchley, 2006). As such, we hypothesized that LacR could also be responsible for the regulation of *abn2*. However, our results not only discard a role of LacR in the regulation of *abn2*, but also show that the previously reported arabinan- and pectin-mediated induction of this gene does not occur. In a strain carrying an *abn2'-lacZ* fusion, an increased β-galactosidase activity had been observed upon addition of either arabinan or pectin (Inácio and Sá-Nogueira, 2008), however we show here that this increase was not due to the derepression of *abn2*. Instead, this increase of β-galactosidase activity was exclusively due to a higher level of GanA (LacA) in the presence of arabinan or pectin. LacR-mediated repression of the *cycB-ganPQAB* operon has been proposed to be modulated by the presence of galactooligosaccharides (Shipkowsky and Brenchley, 2006). Since pectin is known to be decorated by galactan side chains (Ochiai *et al.*, 2007), and a highly galactan-contaminated arabinan was used (see Chapter II), galactooligosaccharides are either already available when the polysaccharides are added to growing cultures, or become available from the breakdown of galactan by GanB. This relieves repression of the *cycB-ganPQAB* operon

and, consequently, leads to the increase of β-galactosidase activity from LacA, thus masking the level of LacZ and the determination of *abn2* expression variations.

The results presented here show that *abn2* expression is temporally controlled, progressively increasing during the transition from exponential growth to stationary phase. Two distinct promoters (σ^A- and σ^H-dependent) were previously identified upstream of the *abn2* coding region (Inácio and Sá-Nogueira, 2008). The *B. subtilis* σ^H transcription factor is responsible for the transcription of many genes needed during the transition from vegetative growth to stationary phase, and the amount of this alternative transcription factor increases after the end of exponential growth (Grossman, 1995; Britton *et al.*, 2002). The σ^H transcription factor is encoded by the *spo0H* gene and is negatively regulated by the transition state regulator AbrB. The amount of AbrB in *B. subtilis* cells decreases at the end of vegetative growth mainly due to Spo0A-mediated repression of its gene (Strauch *et al.*, 1990), thus allowing a higher *abn2* transcription during this stage. However, we propose that the AbrB-dependent regulation of *abn2* is not simply due to its role in the repression of *spo0H*. Although an AbrB-binding site is not well established, and AbrB was also proposed to recognize specific three-dimensional structures in the DNA (Xu and Strauch, 1996), the A/T-rich consensus sequence WAWWTTTWCAAAAAAW (W = A or T) has been suggested (Strauch, 1995). A representation of the position weight matrix for the final eight bases of the consensus sequence is shown in Figure 5.4, and is highly identical to the putative AbrB-binding site identified upstream of the *abn2* coding region (Figure 5.1).

Figure 5.4 – Position weight matrix of the *B. subtilis* AbrB-binding site. Weight matrix obtained from Virtual Footprint (http://www.prodoric.de/vfp/vfp_promoter.php; version 3.0). The putative AbrB-binding site for *abn2* is AAATTTAATAAAAAAG, in which the underlined bases correspond to the positions shown in the figure.

Although *abn2* (*yxiA*) has not been referenced yet as a gene belonging to the AbrB regulon (Banse *et al.*, 2008; Chumsakul *et al.*, 2011; Chumsakul *et al.*, 2013), our results suggest that AbrB is able to bind to this putative binding site thus repressing *abn2* during vegetative growth of *B. subtilis* cells. These observations lead to the proposal of a model for the regulation of *abn2* (Figure 5.5). When nutrients are abundantly available, Spo0A~P levels in the cell are low and AbrB is able to repress genes not needed during this phase, including *abn2* and *spo0H* (Figure 5.5a). When nutrients capable of maintaining an exponentially growing culture become scarce, several signals and cellular processes activate a phosphorelay cascade, which ultimately leads to the activation of Spo0A (Hoch, 1993). The increasing level of active Spo0A~P activates many genes, mostly involved in sporulation but not exclusively, and, simultaneously, represses *abrB*

expression (Strauch *et al.*, 1990; Molle *et al.*, 2003; Fujita *et al.*, 2005). Additionally, the *abbA* gene, which encodes an AbrB antirepressor, is not expressed during vegetative growth and is directly activated by Spo0A~P in the early stages of the transition to stationary phase (Fujita *et al.*, 2005; Banse *et al.*, 2008). The AbbA antirepressor binds to AbrB and prevents its binding to DNA. As a result, transcription of both *abn2* and *spo0H* is no longer under the negative control of AbrB (Figure 5.5b). Furthermore, during this transition from vegetative growth to stationary phase, AbrB levels slowly decrease due to protein degradation (Banse *et al.*, 2008). The combined effect of *abrB* repression by Spo0A~P, AbrB DNA-binding blocking by AbbA, and AbrB degradation results in a quick increase in *abn2* expression after the end of exponential growth.

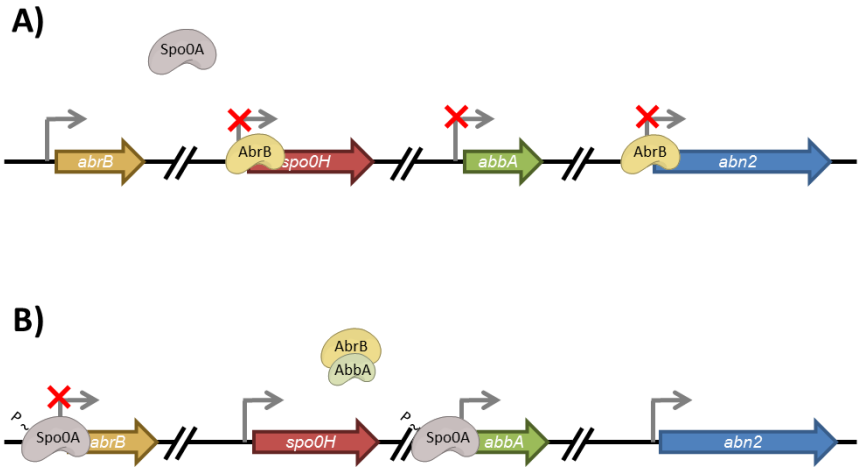


Figure 5.5 – Proposed model for the regulation of *abn2*. A) During vegetative growth, when rapidly metabolizable nutrients are available and Spo0A~P levels are low, AbrB binds to DNA and represses *abn2* expression. **B)** At the onset of stationary phase, in response to nutrient depletion, Spo0A is phosphorylated by a multicomponent phosphorelay system. The active Spo0A~P then binds to DNA and blocks *abrB* transcription. Simultaneously, Spo0A~P activates transcription of the *abbA* gene encoding an AbrB antirepressor. AbbA forms a complex with AbrB blocking its binding to DNA, which, together with degradation of AbrB, leads to a rapid derepression of *spo0H* and *abn2*.

Here, we also report that in the presence of arabinose or arabinan there is a decrease in *abn2* expression. *B. subtilis* encodes two extracellular endo-α-1,5-arabinanases, AbnA and Abn2, which display similar activities to the same substrates but are regulated by different mechanisms (Raposo *et al*., 2004; Leal and Sá-Nogueira, 2004; Inácio and Sá-Nogueira, 2008). *abnA* is part of the AraR regulon and, as such, is expressed when arabinose is present and binds to AraR. Since *abn2* appears to be repressed in conditions more favorable for *abnA*, and its transcription is enhanced during a period when growth conditions are not ideal and *B. subtilis* is triggering several mechanisms which will allow survival (sporulation, competence development, biofilm formation, cannibalism), we may speculate that Abn2 represents a probing enzyme responsible for searching new carbon sources before cells irreversibly initiate the sporulation process. A scavenging role for other *B. subtilis* extracellular enzymes has been previously proposed, such as amylase, also synthesized at the end of vegetative growth and under the control of a temporal regulation mechanism (Ferrari *et al.*, 1993). Further supporting the exploratory role of Abn2, the

expression of *abnA* during exponential growth and in the presence of arabinose, measured by RT-qPCR, is over 100-fold higher than that for *abn2* in the same conditions (data not shown). Also, during stationary phase and in the absence of arabinose the maximum expression of *abn2* is still over 10-fold lower than that of *abnA* during exponential growth on arabinose.

Combining all the available data, we hypothesize that in the absence of a suitable carbon source a low amount of Abn2 is released to the medium by *B. subtilis* cells, so that when arabinans become available their breakdown is rapidly initiated by Abn2. Then, when arabinose enters the cell it binds to AraR and triggers the other genes needed for arabinose utilization, including AbnA. At the same time, arabinose availability leads to a decrease of *abn2* expression, since *B. subtilis* cells are already committed to the production of another extracellular endo-α-1,5-arabinanase. In conclusion, AbnA and Abn2, although enzymatically similar, are not functionally redundant and their differentiated regulation mechanisms provide an ingenious strategy that allows *B. subtilis* to respond efficiently to arabinose availability changes.

Acknowledgments

Joana Lima and Inês Martins constructed some of the plasmids and strains used for this work and performed β-galactosidase assays. This work was supported by the Fundação para a Ciência e a Tecnologia [grant numbers PTDC/AGR-AAM/102345/2008 to I.S.N.; PEst-OE/BIA/UI0457/2011 to Centro de Recursos Microbiológicos; UID/Multi/04378/2013 to Unidade de Ciências Biomoleculares Aplicadas; fellowship SFRH/BD/73039/2010 to M.J.F.].

Chapter VI

The importance of the Abn2 calcium cluster in the endo-1,5-arabinanase activity from *Bacillus subtilis*

The work presented in this chapter was published in “McVey C. E.*, Ferreira M. J.*, Correia B., Lahiri S., de Sanctis D., Carrondo M. A., Lindley P. F., Sá Nogueira I., Soares C. M. and Bento I. (2014). The importance of the Abn2 calcium cluster in the endo-1,5-arabinanase activity from *Bacillus subtilis*. *J. Biol. Inorg. Chem.* 19: 505–513”. The author of this Thesis performed the preparation and purification of recombinant wild-type Abn2 and all the arabinanase activity assays.

* Equal contribution to the work.

The importance of the Abn2 calcium cluster in the endo-1,5-arabinanase activity from *Bacillus subtilis*

Abstract

Arabinanase is a glycosyl hydrolase that is able to cleave the glycosidic bonds of α-1,5-L-arabinan, releasing arabinooligosaccharides and L-arabinose. The enzyme has two domains, an N-terminal catalytic domain with a characteristic β-propeller fold and a C-terminal domain whose function is unknown. A calcium ion, located near the catalytic site, serves to stabilize the N-terminal domain, but it has also been proposed to play a key role in the enzyme mechanism. The present work describes the structure of an inactive mutant of the wild-type enzyme (H318Q) and in which the calcium ion has been adventitiously replaced by nickel. These structural studies, together with functional and modelling studies, clearly support the role of the calcium ion in the overall reaction mechanism.

Introduction

The plant cell wall is a complex composite of structural polysaccharides that represents the most abundant source of organic molecules in the biosphere. Most of the biomass synthesized by photosynthetic CO_2 fixation is stored in the plant cell wall as polymeric carbohydrates, mainly in the form of cellulose and hemicellulose. Degradation of these plant polysaccharides is caused by the synergistic action of several glycosyl hydrolases (GHs) produced by various microorganisms, and is one of the key steps of the carbon cycle (Beldman *et al.*, 1997; Shallom and Shoham, 2003). Owing to their increased application in several biotechnological processes, namely, enzymatic saccharification of (hemi)cellulosic biomass to fermentable sugars and concomitant production of economically competitive biofuel (bioethanol), obtaining detailed structural and functional knowledge of (hemi)cellulolytic enzymes together with their substrates is an important goal for the successful understanding of the breakdown of the structural components of plants (Saha, 2003).

Hemicellulose is a complex heterogeneous polysaccharide of pentoses (xylose and arabinose), hexoses (mannose, glucose and galactose) and sugar acid. The xylan backbone contains D-xylose as its basic unit, but its side groups can be replaced with hexose and sugar acids (Bastawde, 1992). The heterogeneous nature of xylan means that it requires a variety of

enzymes for degradation, including endoxylanases, which cleave the β-1,4 glycosidic linkage in the xylan backbone (GH families 10 and 11), arabinofuranosidases, which remove arabinose side chains (GH families 43, 51 and 53), β-xylosidases (GH families 3 and 43), which release xylose from xylooligosaccharides, acetyl xylan esterases, which remove acetyl groups from the xylan backbone, and feruloyl and ferulic acid esterases, which remove ferulic acid from xylan side chains (Dodd and Cann, 2009).

L-arabinose, the second most abundant pentose in nature, is found in significant amounts in branched and debranched arabinans, as well as in arabinoxylans and arabinogalactans. Arabinoxylan has a β-1,4-linked xylopyranose backbone and generally contains heterogeneous substituents such as L-arabinose, *O*-acetyl ferulic acid, *p*-coumaric acid and 4-*O*-methylglucuronic acid (Puls and Schuseil, 1993). Arabinan is an L-arabinose homoglycan present in plant tissues and is generally associated with pectins. This polysaccharide is composed of α-1,5-linked L-arabinofuranosyl units, some of which are substituted with α-1,3- and α-1,2-linked chains of L-arabinofuranosyl residues (Beldman *et al.*, 1997). The two major enzyme families that hydrolyze arabinan are α-L-arabinofuranosidases (AFs, EC 3.2.1.55) and endo-1,5-α-L-arabinanases (ABNs, EC 3.2.1.99). α-L-arabinofuranosidases remove arabinose side chains, allowing ABNs to attack the glycosidic bonds of the arabinan backbone, releasing a mixture of arabinooligosaccharides and L-arabinose as the products of the reaction. Arabinanases have several applications, in particular in the food industry and nutritional medical research (Beldman *et al.*, 1997).

The Gram-positive bacterium *Bacillus subtilis* participates in the enzymatic dissolution of plant biomass in the soil. Thus, it is able to synthesize a large variety of GHs capable of depolymerizing plant cell wall polysaccharides, including two extracellular ABNs belonging to GH family 43 (Coutinho and Henrissat, 1999), AbnA and Abn2, that hydrolyze sugar beet arabinan, linear α-1,5-L-arabinan and pectin (Inácio and Sá-Nogueira, 2008). Previous studies describing the detailed structural and functional characterization of Abn2 showed that this enzyme comprises 443 amino acids organized into two different domains, an N-terminal catalytic domain with the common β-propeller fold observed in the other ABNs and a C-terminal domain with a β-barrel configuration (de Sanctis *et al*., 2010). The active site includes three essential carboxylate residues. Two of these, D38 and E224, are directly involved in the enzymatic mechanism, and a third, D171, has been proposed to influence the protonation state of the other two (Pons *et al.*, 2004). Located approximately 6.5 Å from the catalytic site, a calcium cluster has also been identified, and it may have a functional role (de Sanctis *et al*., 2010). In this article, aiming to further clarify the role of the calcium metal center, the enzymatic and structural characterization of an Abn2 inactive mutant (H318Q) is reported and this structure together with simulations of equilibrium protonation unveils the role and importance of this metal center for the enzymatic mechanism of these types of enzymes.

Materials and Methods

Preparation of recombinant arabinanase

E. coli BL21 (DE3) pLysS (Studier *et al.*, 1990) was used as the host for recombinant protein expression. *E. coli* strains harboring the desired plasmids for the overproduction of the recombinant wild-type Abn2 (Inácio and Sá-Nogueira, 2008) or the H318Q mutant (de Sanctis *et al*., 2010) were grown at 37ºC and 150 rpm, for 18 hours, in liquid Luria-Bertani (LB) medium (Miller, 1972) supplemented with solutions for auto-induction (Studier, 2005), with the addition of chloramphenicol (25 $\mu g.mL^{-1}$) and kanamycin (20 $\mu g.mL^{-1}$). Cells were harvested by centrifugation for 10 minutes at 8000*g* and 4ºC. All subsequent steps were performed at 4ºC. The periplasmic protein fraction (PPF) was prepared by osmotic shock as described previously (Leal and Sá-Nogueira, 2004). The PPF was loaded onto a 1 mL Hitrap™ Chelating HP column (GE Healthcare Life Sciences) and the bound proteins were eluted with a discontinuous gradient of imidazole. Fractions in which recombinant Abn2 was more than 95% pure were dialyzed overnight against Storage buffer (20 mM sodium phosphate, 62.5 mM NaCl and 10% glycerol, pH 7.4) and then frozen in liquid nitrogen and kept at -80ºC until further use.

Enzyme activity assays

The source of the enzyme was purified arabinanases (Abn2-His_6 or Abn2*H318Q-His_6). Enzyme activity was assayed in a mixture containing 0.5% (w/v) sugar beet Arabinan (Megazyme International Ireland) in 25 mM bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane (Bis-Tris) buffer, pH 6.6, with incubation for 30 minutes at 37ºC. The reducing sugar content after hydrolysis was determined according to the Nelson-Somogyi method (Somogyi, 1952), with L-arabinose as standard. One unit of activity was defined as the amount of enzyme that produces 1 µmol of arabinose equivalents per minute. The enzyme activity towards sugar beet Arabinan was determined by measuring the initial rates obtained from the linear portion of the progress curve and was expressed in units per milligram of protein. Enzymatic activity was also determined after incubation, for 30 minutes at 4ºC, with 10 mM EDTA and/or EGTA and 20 mM calcium chloride, nickel chloride, manganese chloride or magnesium chloride, using the same conditions.

Crystallization and data collection

Crystals of the H318Q protein were obtained with a protein concentration of 14 $mg.mL^{-1}$, using the vapor diffusion method and as crystallization conditions 10 mM $NiCl_2$, 100 mM tris(hydroxymethyl)aminomethane–HCl, pH 8.5, and 20% PEG 2000 monomethyl ether. Data collection was performed at the European Synchrotron Radiation Facility (Grenoble, France), and the data were integrated using iMOSFLM (Battye *et al.*, 2011) and scaled using SCALA (Evans,

2006). H318Q crystals belong to the monoclinic space group $P2_1$, with one molecule per asymmetric unit. Data collection statistics are listed in Table 6.1.

Table 6.1 – Data collection statistics for Abn2 H318Q.

Beam line at the European Synchrotron Radiation Facility	ID14-EH4
Wavelength (Å)	0.95
Detector	ADSC Q315r
Distance between detector and crystal (mm)	299
Resolution (Å)	1.9
Space group	$P2_1$
Cell parameters – a, b, c (Å)	51.2, 76.5, 58.1
Cell parameters – α, β, γ (º)	90, 112.6, 90
Number of unique *hkl*	31,885 (4,339)
Completeness (%)	97.7 (91.6)
Mean *I*, σ(*I*)	12.6 (3.7)
R_{merge}	0.051 (0.287)
Multiplicity	3.1 (2.6)

Values in *parentheses* refer to the highest-resolution shell (1.90-2.00 Å).

Structure determination and refinement

The crystal structure of the H318Q mutant protein was solved by the molecular replacement method, using the program PHASER (McCoy *et al*., 2007) and the coordinates of the native protein (Protein Data Bank ID 2X8F; de Sanctis *et al*., 2010) without the solvent molecules, as the search model. Iterative model building and refinement were performed using the programs COOT (Emsley and Cowtan, 2004) and PHENIX (Adams *et al.*, 2010), respectively.

Table 6.2 – Refinement statistics for Abn2 H318Q.

Atoms	3507
Ligand atoms	9
Water molecules	306
Protein residues	443
Final *R* factor (%)	17.10
Final free *R* factor (%)[a]	21.40
Mean *B* values (Å²) – Wilson *B*	26.13
Mean *B* values (Å²) – Protein	33.10
Mean *B* values (Å²) – Solvent	36.30
Mean *B* values (Å²) – Ligands	27.80
Distance and angle deviations – Bond distances (Å)	0.005
Distance and angle deviations – Bond angles (Å)	0.93
Distance and angle deviations – Ramachandran favoured (%)	97.3
Distance and angle deviations – Ramachandran outliers (%)	0.2

[a] Calculated with 5% of reflections excluded from refinement; Coordinates were deposited in the Protein Data Bank under the ID code 4COT.

The program PHENIX was used to refine atomic coordinates together with individual isotropic atomic displacement parameters. TLS thermal anisotropic parameterization was also included in the final stages of refinement, with each molecule divided into six TLS groups, as suggested by TLS Motion Determination (Painter and Merritt, 2005; Winn *et al.*, 2001). Structure refinement statistics and details regarding the quality of the final structural model are listed in Table 6.2. Figures were drawn with PyMOL (DeLano, 2002).

Theoretical calculations

The equilibrium protonation of ionizable groups in Abn2 and mutants (the X-ray structures were used for the wild type, H318A and H318Q mutants, whereas the mutants D38A, D171A, and E224A were done *in silico*) was simulated using a combination of continuum electrostatic and Monte Carlo methods, as described elsewhere (Baptista and Soares, 2001; Teixeira *et al.*, 2002; Teixeira *et al.*, 2005). The continuum electrostatic calculations were done with the package MEAD, version 2.2.0 (Bashford and Gerwert, 1992), and the Monte Carlo simulations were done with PETIT, version 1.3 (Baptista and Soares, 2001; Teixeira *et al.*, 2002). Details of these calculations can be found in the electronic supplementary material. Briefly, these calculations take into account the whole structure of the protein to generate, with atomic detail, a low dielectric region corresponding to the protein, which is then immersed in a high dielectric region mimicking water. Partial charges are positioned at the atomic positions reflecting the charge distribution of the protein molecule and the charge distribution of the included ions. For ionizable groups of the protein, these charges reflect either the protonated or the deprotonated states (including tautomeric arrangements of protons), and the electrostatic interactions of these states are precalculated by solving the linear Poisson-Boltzmann equation of the system using a finite-difference method. The energies obtained are then used to sample ionizable states using a Monte Carlo procedure that accounts for the joint equilibrium protonation of the different ionizable groups at a given pH value. These calculations reflect, in a physically realistic way, the interactions between charges in the protein (including ions) and the solvation effect generated by the protein and the surrounding water.

Results

The crystal structure of the H318Q mutant

The three-dimensional structure of the H318Q mutant shows an overall fold very similar to that of the native Abn2, divided in two domains, an N-terminal catalytic domain with the characteristic β-propeller fold and a C-terminal domain with a distorted β-barrel configuration (de Sanctis *et al*., 2010) (Figure 6.1a). The major difference observed is close to the active site. In

the native Abn2, a calcium cluster (Figure 6.1b) is observed approximately 6.5 Å from the catalytic residues (de Sanctis *et al*., 2010), whereas in the H318Q mutant a nickel cluster is observed instead (Figure 6.1c). In this case, the nickel is coordinated by two protein residues, Q318 and H173. To coordinate the nickel ion, H173 adopts a different rotamer pointing towards the metal cluster, whereas in the wild-type enzyme this histidine residue is facing a sodium cluster that is located further down in the solvent channel.

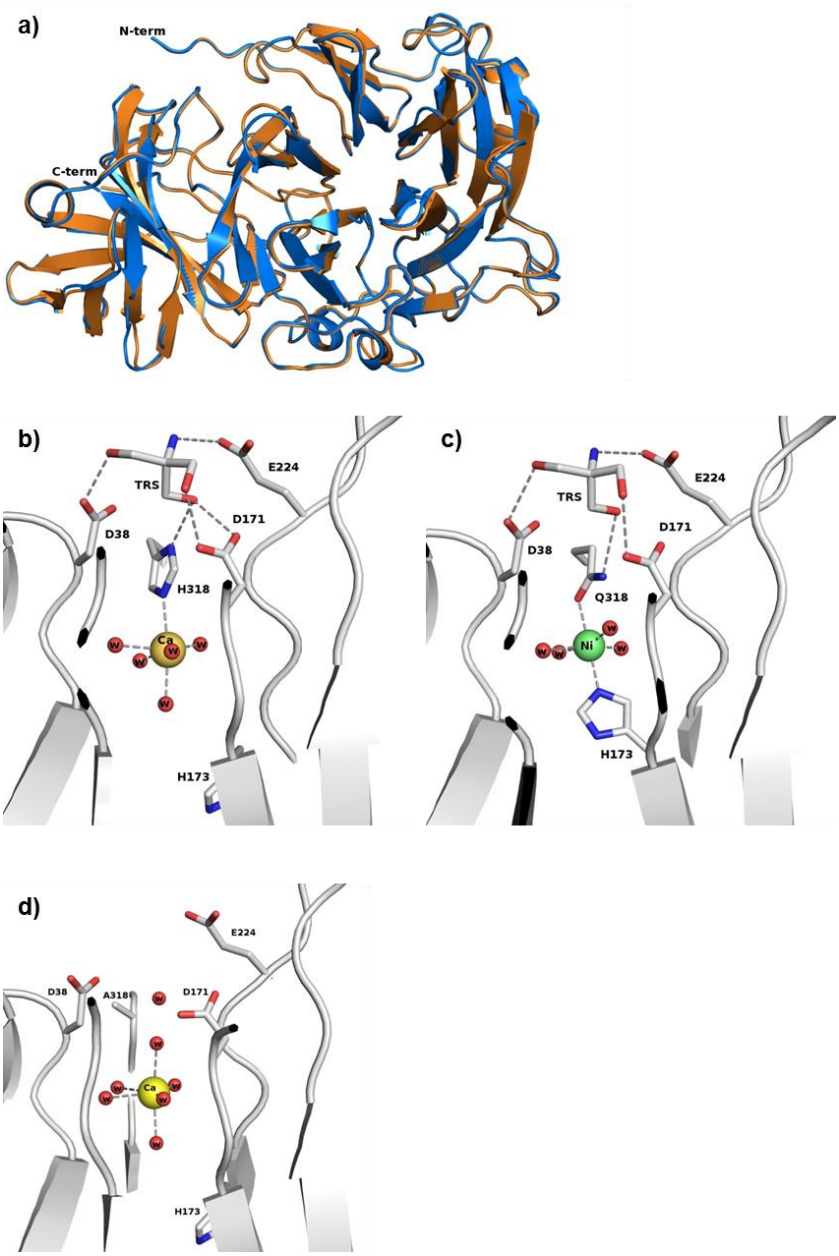


Figure 6.1 – Close-up view of the metal cluster in wild-type and mutant Abn2. a) Structural superposition of the H318Q mutant (*gray*) with wild-type Abn2 (Protein Data Bank ID 2X8F) (*blue*). Superposition of the Cα trace of Abn2 H318Q (Protein Data Bank ID 4COT) with the Cα trace of wild-type Abn2 (Protein Data Bank ID 2X8F) – root mean square deviation value of 0.37 Å for 442 Cα pairs. **b)** Close-up view of the active site in native Abn2. **c)** Close-up view of the metal cluster site in the H318Q mutant enzyme. **d)** Close-up view of the metal cluster site in the H318A mutant enzyme. *TRS* tris(hydroxymethyl)aminomethane.

In the mutant enzyme this sodium cluster, which seems to be an artefact of the crystallization conditions, is no longer observed and H173 turns to the other side, coordinating the nickel ion. The rest of the octahedral coordination sphere is made by the water molecules, similarly to what is observed for the calcium cluster in the native enzyme (Figure 6.1b, c). In the same way, a tris(hydroxymethyl)aminomethane molecule was modelled in the -1 subsite, within the hydrogen-bond distance of the catalytic carboxylates. To confirm the presence of the nickel in the crystal structure, X-ray emission spectra of the H318Q crystal and the mutant protein solution were obtained prior to data collection (see Supplementary Figure 3 in Annexes). The spectra clearly show the presence of nickel in the H318Q mutant crystal and its absence from the protein solution. The spectra also show that calcium is absent in both situations, whereas it was present in the spectrum of native Abn2 (de Sanctis *et al*., 2010).

Enzymatic assays using different metal ions

Enzymatic assays were performed with samples of the native Abn2 and the H318Q mutant. Native Abn2 was incubated with EGTA and EDTA in order to completely remove the calcium content and was then incubated for 30 minutes with several ions prior to the assay being performed (Table 6.3). In addition, this sample was incubated with the same metals, without previous treatment with any chelating agent. The H318Q mutant protein was incubated with calcium and nickel and tested for its enzymatic activity.

Table 6.3 – Role of metal ion binding in the catalytic activity of wild-type Abn2 and H318Q mutant of Abn2.

	Activity [c] ($U.mg^{-1}$)
Abn2 [a]	19.47±2.08
Abn2 + 10 mM EGTA [a]	2.65±0.41
Abn2 + 10 mM EDTA [a]	0.20±0.01
Abn2 + 20 mM $CaCl_2$ [b]	20.57±3.21
Abn2 + 20 mM $NiCl_2$ [b]	31.28±0.96
Abn2 + 20 mM $MnCl_2$ [b]	32.76±1.11
Abn2 + 20 mM $MgCl_2$ [b]	19.63±4.19
Abn2 + 10 mM EGTA + 20 mM $CaCl_2$ [b]	21.96±0.82
Abn2 + 10 mM EGTA + 20 mM $NiCl_2$ [b]	25.73±0.62
Abn2 + 10 mM EGTA + 20 mM $MnCl_2$ [b]	35.49±0.66
Abn2 + 10 mM EGTA + 20 mM $MgCl_2$ [b]	3.23±0.64
Abn2 (H318Q) [a]	No activity
Abn2 (H318Q) + 20 mM $CaCl_2$ [a]	No activity
Abn2 (H318Q) + 20 mM $NiCl_2$ [a]	No activity

[a] Abn2 was incubated for 30 minutes on ice prior to the assay.

[b] Abn2 was incubated for an additional 30 minutes on addition of $CaCl_2$, $NiCl_2$, $MnCl_2$ or $MgCl_2$.

[c] Activity is the average of three independent assays with the respective standard deviation.

Table 6.3 illustrates that either EGTA or EDTA is able to remove the calcium from both protein samples. Furthermore, it is also clear from these assays that magnesium is not able to replace calcium efficiently, having no effect on the activity of the enzyme when added to it. However, the nickel ion allows the enzyme to fully recover its efficiency and even become slightly more active when it is incubated with nickel (Table 6.3). Furthermore, the addition of manganese chloride, with or without a previous incubation with EGTA for calcium removal, has a similar effect in the recovery of the enzymatic activity to that observed when nickel is added. On the other hand, the H318Q mutant enzyme is not able to recover its activity when it is incubated with any of the metal ions chosen (Table 6.3).

Theoretical calculations

Simulations of the protonation equilibria of all titratable groups in the protein were undertaken in the native enzyme, with the aim of analyzing the actual protonation characteristics of the essential acidic groups of the active site. The titration behavior of the three essential acids can be found in Figure 6.2.

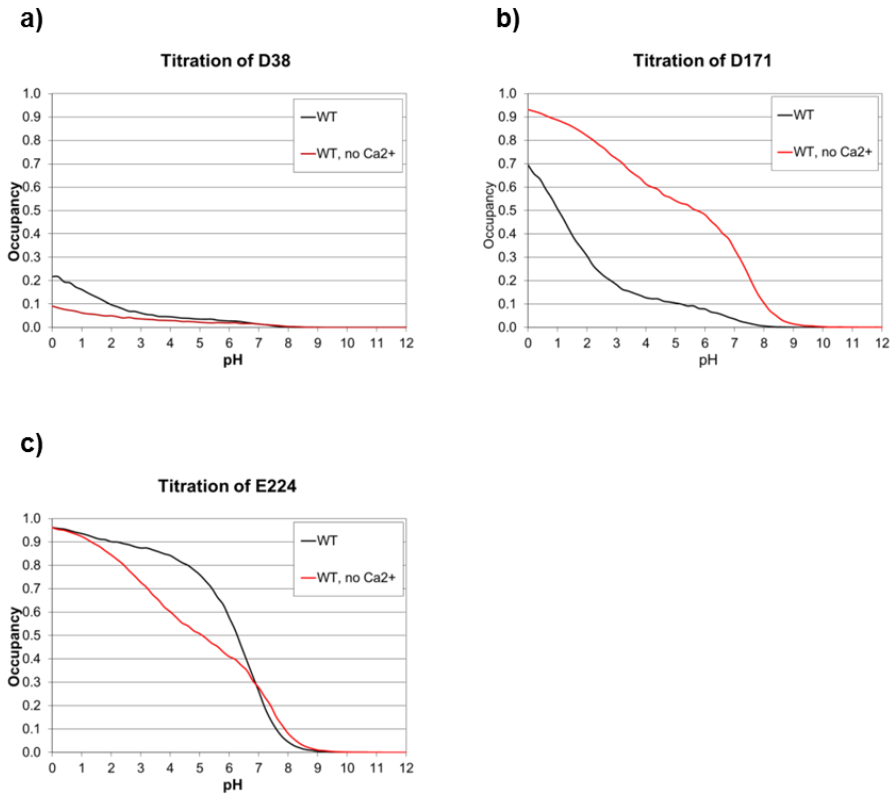


Figure 6.2 – Simulation of the pH titration of D38 (a), D171 (b) and E224 (c). *Black lines* correspond to the simulation of the wild-type protein and *red lines* correspond to the simulation of the wild-type protein but without the calcium ion (while keeping its water ligands). *WT* wild-type.

The simulation results with the wild-type structure show that D38 is mostly deprotonated over the pH range analyzed (and specifically at pH 6.5, where the enzyme displays the maximum activity), therefore being available to function as the general base. In contrast, the protonation behavior of E224 shows that it can function as a general acid, since it has considerable protonation at pH 6.5 (approximately 0.42), being able to provide protons to the substrate. Finally, the titration behavior of D171 is quite abnormal, with a long and nonstandard titration curve, evidence that it is substantially coupled with the other acidic groups. In fact, all these groups are substantially coupled, with the most important result being the extended and increased protonation of E224. *In silico* mutants of these three residues, one at a time, show that the D171A mutation yields an almost normal titration curve for E224 (data not shown), presumably reducing the catalytic activity, as seen experimentally for this mutant (de Sanctis *et al*., 2010).

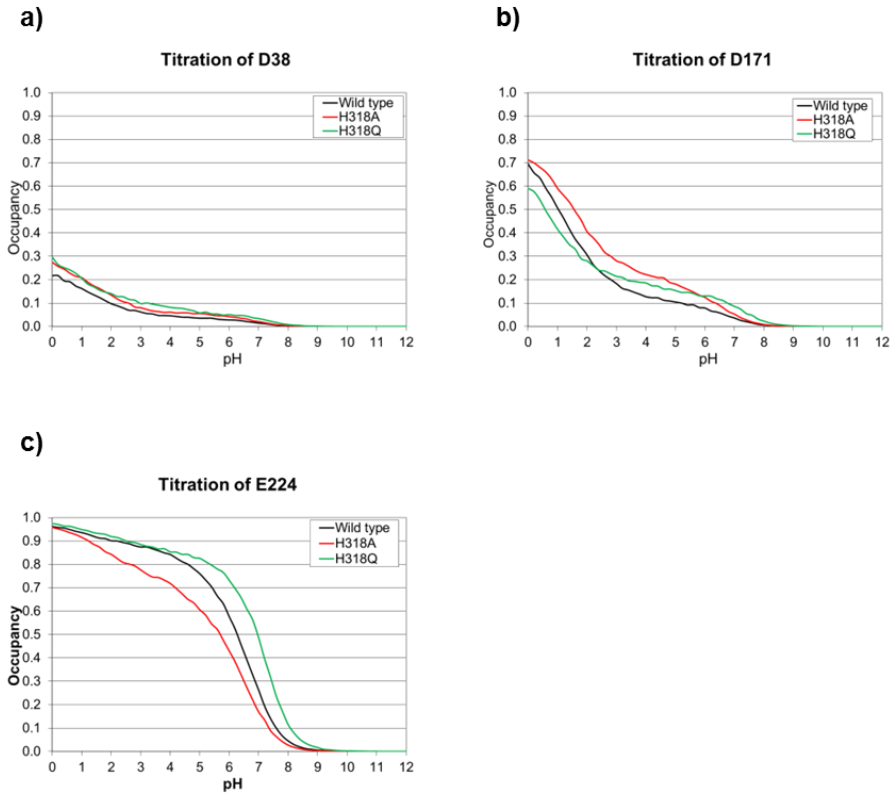


Figure 6.3 – Simulation of the pH titration of D38 (a), D171 (b) and E224 (c) in the wild-type Abn2 (*black*) and in the H318A (*red*) and H318Q (*green*) mutants.

Similar protonation equilibrium simulations were then performed in the absence of the calcium ion. The results are also shown in Figure 6.2. The effect on D38 is not very noticeable, but the effect on D171 and E224 is quite significant. D171 becomes substantially protonated at pH 6.5 (0.42 versus 0.06 in the wild type), whereas the equilibrium protonation of E224 is reduced (0.36 versus 0.42 in the wild-type). To evaluate the effect induced by mutations performed in H318, simulations of the equilibrium protonation were also done on the mutant enzymes (H318A and H318Q). The results are shown in Figure 6.3, and are compared with the results for the wild-

type. As can be seen, the major effect of the mutations (and associated conformational changes) at pH 6.5 is on the titration of E224. The mutants have opposite effects: H318A generates lower protonation of E224 (0.30 versus 0.42 in the wild type), whereas H318Q generates greater protonation (0.63).

Discussion

The biochemical and structural characterization of the native ABN (Abn2) from *B. subtilis* and of two mutant enzymes, H318A and D171A, has been previously undertaken in our laboratory. de Sanctis *et al*. (2010) have shown that Abn2 has a three-dimensional fold somewhat different from that in other arabinanases in GH family 43. In addition to the common catalytic domain comprising the characteristic β-propeller fold, there is an extra C-terminal domain whose precise function is yet to be determined. Furthermore, these studies showed the presence of a calcium cluster (rather than a chlorine ion as originally proposed by Nurizzo *et al.*, 2002) close to the catalytic site, which seems to be important not only for the overall stability of the enzyme, but also for the enzyme activity. With the specific objective of clarifying the role of this calcium cluster, the biochemical and structural characterization of a second mutant of H318 (H318Q) was performed. When the histidine residue that coordinates the calcium ion is mutated into an alanine, the calcium cluster close to the catalytic site is still observed and the catalytic efficiency of the mutant enzyme is significantly reduced (de Sanctis *et al*., 2010), but not entirely lost. The H318A mutant shows essentially the same structure as that of the native enzyme, and the mutation does not unduly disrupt the cluster; the calcium ion coordination is maintained by an extra water molecule that is positioned where NE2 of the histidine imidazole ring is found in the native enzyme (de Sanctis *et al*., 2010) (Figure 6.1d). On the other hand, when the histidine residue is mutated into a glutamine (H318Q), the mutant enzyme does not show any measurable activity, as reported previously (de Sanctis *et al*., 2010) (Table 6.3). The structure showed that although the three-dimensional fold of the mutant enzyme is maintained similar to that in the native Abn2, the calcium cluster is no longer observed near the active site (Figure 6.1b); instead, it appears that the calcium cluster is replaced by one containing nickel (Figure 6.1c). In fact, whereas the H318A and native Abn2 crystals appear from the same crystallization conditions, good diffracting crystals of the H318Q mutant only appear in the presence of 0.01 M $NiCl_2$. X-ray emission spectra clearly show that only the H318Q crystals have the metal present (see Supplementary Figure 3 in Annexes). These results not only corroborate the fact that in solution the mutant enzyme does not have the calcium cluster, or any other metal cluster, but also indicate that the lack of enzymatic activity observed for the H318Q mutant enzyme is associated with the absence of this metal cluster. Indeed, a few cases in GH family 43 where the presence of a divalent metal center has stimulated an increase in the enzyme efficiency have already been reported (Jordan *et al.*, 2013; Lee *et al.*, 2013; Moraïs *et al.*, 2012). To understand if any divalent cation could have a similar effect on the enzyme activity, enzymatic assays were performed with the native and H318Q mutant enzymes,

using other divalent cations as described above (Table 6.3). The results clearly show that the native enzyme is able to regain its full activity when incubated with calcium, nickel or manganese (Table 6.3), whereas in the presence of magnesium such recovery is not observed. In addition, it is also shown that the H318Q mutant, which in solution does not show any metal cluster (see Supplementary Figure 3 in Annexes), is not able to recover its activity when soaked with either calcium or nickel (Table 6.3). These results are in agreement with the dramatic decrease in the enzyme activity while maintaining the calcium cluster that was observed previously for the H318A mutant (de Sanctis *et al*., 2010) (Figure 6.1d), suggesting that in the mutant structures besides the metal cluster, other functional key determinants were affected. Furthermore, the results obtained when the enzyme was incubated with Mg^{2+} were also unexpected, as the enzyme shows similar activity when incubated with other divalent metal ions (Table 6.3). Moreover, previous studies performed by Lee *et al.* (2013) with a β-xylosidase from the same family (GH family 43) of enzymes revealed a similar enzymatic activity when the β-xylosidase was incubated with either calcium or magnesium. Therefore, it was expected that a fully active enzyme would be produced when Abn2 was incubated with magnesium. However, this was not observed (Table 6.3). These findings seem to suggest that with Abn2 there are other key determinants influencing the enzyme activity. Most likely, in the case of magnesium, the coordination sphere adopted by the calcium cluster in Abn2 is not efficient for binding and stabilizing the magnesium.

The actual protonation characteristics of catalytic residues in the native enzyme and mutant enzymes were also calculated by simulating the protonation equilibria of all titratable groups in the presence and absence of the metal cluster. Altogether, these simulation results support the role of D171 in the modulation of the titration behavior of the catalytic site, especially of E224, by keeping it substantially protonated and allowing it to work as a general acid in the mechanism. On the other hand, when the calcium was removed from the calculations, a substantial change in the intrinsic characteristics of D171 and E224 was observed (Figure 6.2). D171 becomes substantially protonated at pH 6.5 and the equilibrium protonation of E224 is reduced and consequently its capacity to give a proton to the leaving aglycon is reduced. This may explain the inactivation of the enzyme upon calcium removal with EGTA and EDTA. Additionally, the effect of calcium removal simulated here is mostly electrostatic owing to the intrinsic nature of the continuum electrostatic/Monte Carlo calculations; neighboring groups feel the absence of this strong electrostatic interaction with a +2 charge. Similar results would be obtained if the ion were Ni^{2+}, given that it has the same charge. This may explain the catalytic activity evidenced with the Ni^{2+}-soaked enzyme (Table 6.3). Surprisingly, the H318A and H318Q mutant enzymes show a much reduced (de Sanctis *et al*., 2010) or no activity (Table 6.3), respectively; this is observed in the presence of the metal center. Could these differences be related to the conformational changes (Figure 6.1b, c) observed for both mutant structures? Indeed, the calculations show that despite a divalent ion being kept in the active site, different proton transfer behaviors are generated in the catalytic carboxylates (Figure 6.3), which may lead to different catalytic efficiencies of the enzymes, as observed experimentally.

Overall, the present studies show that a divalent metal cluster is needed in the native protein, close to the active site, to guarantee the correct protonation state of the catalytic residues and consequently the enzyme activity. Furthermore, the results, both for the native protein and for the mutants, suggest that a correct configuration of the binding pocket and of the active-site geometry is also needed in order to optimize enzyme efficiency.

Acknowledgments

José Inácio was responsible for the purification and concentration of the Abn2 H318A and H318Q mutants from *E. coli*. This work was supported by the Fundação para a Ciência e a Tecnologia [grant numbers PTDC/AGR-AAM/102345/2008 to I.S.N.; PEst-OE/BIA/UI0457/2011 to Centro de Recursos Microbiológicos; fellowship SFRH/BD/73039/2010 to M.J.F. and fellowship 069/BI-BI/2012 to S.L.].

Chapter VII

Concluding Remarks

Concluding Remarks

For the past 250 years, since the dawn of the industrial revolution in the 1760s, the demand for energy and fuel has increased exponentially, and those needs have been met mostly through the use of finite reserves of fossil fuels. First coal, to fuel steam-powered machines and transportation systems, and later oil to further supply transportation demands and also for the production of plastics, pharmaceuticals, fertilizers and various chemical reagents. The emergence of new economic superpowers, such as China, India or Brazil, in the second half of the 20th century, has contributed even further to increase this energy demand. Consequently, there have been a series of negative environmental impacts, namely oil spills and increased emissions of greenhouse gases, such as CO_2, and other air pollutants. In addition to environmental concerns, recent economy-damaging fluctuations in oil prices and the finite nature of fossil fuels have driven many researchers to develop and improve alternative and sustainable energy sources (Dellomonaco *et al.*, 2010).

Biofuels, such as bioethanol, are one of the short-term alternatives for the substitution of oil-based fuels. First generation biofuels were easily obtained from crop foods, such as maize and soybeans, and had attractive production costs, but led to direct competition between food and energy, ultimately increasing the price of food. Additionally, first generation biofuels could actually have a positive net contribution in CO_2 emissions, i.e., emissions from their production and combustion were higher than the total amount of CO_2 captured by the original crops (Yuan *et al.*, 2008). This led to the development of second generation biofuels, from the fermentation of sugars available from lignocellulosic biomass and agricultural waste products, which have an estimated negative net contribution in CO_2 emissions but are so far unprofitable due to high production costs (Yuan *et al.*, 2008). High production costs are mainly derived from the increased complexity of non-cellulosic polysaccharides composing cell walls of lignocellulosic plants – when compared to the less complex starch available from maize or sucrose from sugarcane – which complicates the sugar fermentation process.

In order to improve the bioethanol yield by ethanologenic bacteria and yeasts from lignocellulosic biomass, two main strategies are being applied: modification of plant cell wall biosynthesis to lower its lignin content or modify its structure, consequently reducing cell wall recalcitrance to enzymatic degradation (Pauly and Keegstra, 2010; Kalluri *et al.*, 2014; Loqué *et al.*, 2015); and improving the saccharification potential of ethanologenic microorganisms through the release of multiple enzymes capable of completely hydrolyzing plant cell wall polysaccharides (van Maris *et al.*, 2006; Edwards and Doran-Peterson, 2012). For the success of this second strategy, it is of critical importance to understand 1) how the concerted action of several glycosyl hydrolases contributes to the complete breakdown of plant cell wall polysaccharides, 2) how cells are able to uptake mono and oligosaccharides, and 3) how sugars are metabolized by different

organisms. Although research in polysaccharide utilization may be mainly driven by the biofuels industry, many of the obtained results concerning carbohydrate-degrading enzymes and sugar transport systems may be of high relevance for clinical purposes. For example, some pathogenic bacteria, such as *Streptococcus* spp., have highly evolved and concerted mechanisms for the uptake and utilization of carbohydrates that allow their proliferation in sugar-rich environments (Webb *et al.*, 2008; Marion *et al.*, 2011). The work presented in this thesis mainly addresses *Bacillus subtilis* oligosaccharide uptake via ABC-type importers (Chapter II, Chapter III and Chapter IV), and regulation and enzymatic activity of a *B. subtilis* glycosyl hydrolase involved in arabinan breakdown (Chapter V and Chapter VI).

It had been previously suggested by Quentin *et al.* (1999) that at least seven ABC-type importers encoded in the genome of *B. subtilis*, and which were lacking a NBD encoded in their vicinity, could be energized by a single shared ATPase, MsmX or YurJ, two orphan genes located in different regions of the genome. Additionally, a single ATPase, MsiK (with 55% identity to MsmX; Hurtubise *et al.*, 1995), was shown to be involved in the uptake of cellobiose, xylobiose and maltose by *Streptomyces lividans*, being the NBD responsible for energizing the three independent systems (Hurtubise *et al.*, 1995; Schlösser *et al.*, 1997). In *B. subtilis*, it was shown that a maltodextrin-specific ABC importer was energized by MsmX (Schönert *et al.*, 2006). The work presented in Chapter II – most of which published in Ferreira and Sá-Nogueira, 2010 – and Chapter III confirms that MsmX is the ATPase responsible for energizing the uptake of arabino- and galactooligosaccharides by *B. subtilis*. Additionally, MsmX is also implicated in the utilization of pectin, polygalacturonan and rhamnogalacturonan, which indicates that it is likely to be the energizing component of systems dedicated to the transport of galacturonic acid oligomers and/or rhamnose-galacturonic acid disaccharides. MsmX is thus confirmed as the multipurpose ATPase capable of energizing the uptake of most of the major oligosaccharides resulting from the breakdown of plant cell wall polysaccharides. Although we show that a second ATPase encoded in the *B. subtilis* genome, YurJ, is capable of functionally replacing MsmX when placed in a different locus, this protein is not present *in vivo* in any of the conditions tested in the wild-type strain of *B. subtilis*. The gene encoding this ATPase was shown to be transcribed but an unidentified post-transcriptional regulation mechanism must be responsible for the absence of protein.

Additionally, we also identified the AraNPQ, CycB-GanPQ, YesOPQ and YtcQP-YteQ transporters as the MsmX-dependent systems involved in the uptake of these sugars. An ABC system devoted to the uptake of xylooligosaccharides has thus far not been identified in *B. subtilis*, but such a system was characterized in *Geobacillus stearothermophilus* (Shulami *et al.*, 2014). It is thus possible that an ABC-type importer for xylose oligomers also exists in *B. subtilis*. More recently, in *Streptococcus pneumoniae*, MsmK was shown to be a shared ATPase for the sialic acid, raffinose, maltodextrins, and fructooligosaccharide transporters (Marion *et al.*, 2011; Linke *et al.*, 2013). Likewise, a shared ATPase has also been proposed in *Clostridium phytofermentans* (Petit *et al.*, 2015). The continued increase of reports showing that several species share a single

NBD for most ABC systems involved in sugar uptake suggests that this is a rather common feature in most polysaccharide-utilizing bacteria. Having a single ATPase shared between several systems that ultimately serve the same purpose, which is the uptake of a carbon and energy source, is an ingenious strategy that allows cells to rapidly respond to the simultaneous availability of multiple sugars. At the same time, this strategy avoids the burden of having to synthesize several ATP-hydrolyzing proteins in a short period of time, which could be very demanding for the cellular machinery responsible for transcription and translation, and could increase cytoplasmic ATP hydrolysis unrelated to substrate uptake. To better understand how some ABC ATPases are able to interact with multiple transporters while others are specific for a single system, important amino acids for the interactions between MsmX and TMDs from importers energized by it must be identified. Complementation assays with heterologous ATPases and the determination of the crystal structure of MsmX are currently underway and will contribute for the identification of important residues.

Here, in Chapter IV, we also show how arabinooligosaccharides are recognized by AraN, the SBP of the AraNPQ system. Our results suggest that recognition and binding of arabinose oligomers is initially made by only one of the two lobes of AraN and show that a tryptophan residue in that lobe is critical for substrate binding. In contrast with most SBPs from ABC-type sugar importers, which display a very high affinity to their substrates, the affinity of AraN towards arabinooligosaccharides is considerably low. The AraN-sugar interactions displayed a K_d about 100-fold higher than that reported for other ABC-type SBPs. This, however, is not shown here for the first time in *B. subtilis*. Similar affinities were determined for the interactions between the SBP CycB and cyclodextrins (Kamionka and Dahl, 2001), but it is not clear that CycB is actually involved in the uptake of these substrates. In fact, our results show that CycB is implicated in the transport of galactooligosaccharides. Also, in addition to arabinooligosaccharides, AraN was shown to bind cello-, xylo- and maltooligosaccharides with a similar affinity, although in a different site within the binding pocket, but uptake of these carbohydrates is not dependent on this SBP. Simultaneous binding of arabinose oligomers and cellodextrins was observed. We hypothesize that when cellodextrins are available they bind to a second site in the AraN binding pocket without blocking the binding site for arabinooligosaccharides. As a consequence, the binding of two molecules prevents AraN from changing to its closed conformation and triggering the mechanism for uptake of arabinose oligomers. Since AraN also binds maltodextrins, it is possible that a similar mechanism can also be mediated by these substrates, which are, like cellodextrins, rich in glucose. Therefore, the determination of the crystal structure of AraN, which is currently underway, would be of major significance to understand how this protein binds arabinooligosaccharides and also how the binding of other sugars might be involved in controlling the uptake of arabinose oligomers.

In order to understand why *B. subtilis* possesses two extracellular endo-arabinanases, AbnA and Abn2 presenting similar activity towards the same substrates, the mechanisms that regulate *abn2* expression were investigated. Preliminary results, presented in Chapter V, show

that *abn2* is apparently expressed constitutively at low levels during vegetative growth and that an increased expression is observed during post-exponential growth and in stationary phase. In addition, a decrease in *abn2* expression in the presence of arabinose was also observed. Since *abnA* is induced by the presence of this sugar, arabinose-mediated repression of *abn2* might work as a mechanism to prevent the parallel overexpression of two functionally similar enzymes. The major transition state regulator AbrB is directly implicated in repression of *abn2* during exponential growth phase through binding to a putative AbrB-binding site and preventing transcription elongation. The major sporulation regulator, Spo0A, is also indirectly involved in the regulation of *abn2* since it represses the *abrB* gene after being phosphorylated during the transition from vegetative growth to stationary phase. These results led us to hypothesize a scavenging role for Abn2, i.e., being constitutively expressed at low levels in the absence of a suitable carbon source this arabinanse would allow an immediate response to the availability of arabinose-containing polysaccharides. This way, arabinose and arabinose oligomers would be rapidly released and used by *B. subtilis* cells. At the same time cells would increase production of one arabinanase (AbnA) and repress the other (Abn2). Upon arabinose depletion, and if no other suitable carbon source is available to maintain exponential cell growth, *abn2* is again expressed at higher levels in an attempt to obtain a new carbon source before cells commit to the sporulation process. Further work is needed to better understand the role of Abn2, including the construction of *spo0H* mutants, and to identify the mechanisms by which arabinose is able to repress the expression of *abn2*.

In Chapter VI (published in McVey *et al.*, 2014), the enzymatic activity of Abn2 was shown to be dependent on a calcium cluster close to the catalytic site of the enzyme. This cluster is not only important for the overall enzyme stability, but also critical for its catalytic activity. Even though this cluster has been observed in crystal structures of Abn2 and is critical for enzyme activity, we also show that after removal of calcium ions, two divalent metallic ions, nickel and manganese, were capable of fully restoring endo-arabinanase activity. Further supporting the critical role of this cluster in enzyme activity, a mutant version of the protein which displays no detectable endo-arabinanase activity, was shown to lack a metal cluster.

List of References

Adams P. D., Afonine P. V., Bunkoczi G., Chen V. B., Davis I. W., Echols N., Headd J. J., Hung L. W., Kapral G. J., Grosse-Kunstleve R. W., McCoy A. J., Moriarty N. W., Oeffner R., Read R. J., Richardson D. C., Richardson J. S., Terwilliger T. C. and Zwart P. H. (2010). *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. Sect. D* **66**: 213–221.

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

Arbige M. V., Bulthuis B. A., Schultz J. and Crabb D. (1993). Fermentation of *Bacillus*. In: Sonenshein A. L., Hoch J. A. and Losick R. (ed.) *Bacillus subtilis* and other Gram-positive Bacteria. American Society for Microbiology, Washington DC, USA, 871–895.

Arnaud M., Chastanet A. and Débarbouille M. (2004). New Vector for Efficient Allelic Replacement in Naturally Nontransformable, Low-GC-Content, Gram-Positive Bacteria. *Appl. Environ. Microbiol.* **70**: 6887–6891.

Banse A. V., Chastanet A., Rahn-Lee L., Hobbs E. C. and Losick R. (2008). Parallel pathways of repression and antirepression governing the transition to stationary phase in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* **105**: 15547–15552.

Baptista A. M. and Soares C. M. (2001). Some theoretical and computational aspects of the inclusion of proton isomerism in the protonation equilibrium of proteins. *J. Phys. Chem. B*, **105**: 293–309.

Bashford D. and Gerwert K. (1992). Electrostatic calculations of the pKa values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.* **224**: 473–486.

Bastawde K. B. (1992). Xylan structure, microbial xylanases, and their mode of action. *World J. Microbiol. Biotechnol.* **8**: 353–368.

Battye T. G., Kontogiannis L., Johnson O., Powell H. R. and Leslie A. G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Cryst. Sect. D* **67**: 271–281.

Beldman G., Schols H. A., Pitson S. M., Searle-van Leeuwen M. J. F. and Voragen A. G. J. (1997). Arabinans and arabinan degrading enzymes. In: Sturgeon R. (ed.), Advances in Macromolecular Carbohydrate Research, vol. 1. JAI Press Inc., 1–64.

Berntsson R. P.-A., Smits S. H. J., Schmitt L., Slotboom D. J. and Poolman B. (2010). A structural classification of substrate-binding proteins. *FEBS Lett.* **584**: 2606–2617.

Blencke H. M., Homuth G., Ludwig H., Mäder U., Hecker M. and Stülke J. (2003). Transcriptional profiling of gene expression in response to glucose in *Bacillus subtilis*: regulation of the central metabolic pathways. *Metab. Eng.* **5**: 133–149.

Bordignon E., Grote M. and 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**: 1354–1366.

Bourgois T. M., van Craeyveld V., van Campenhout S., Courtin C. M., Delcour J. A., Robben J. and Volckaert G. (2007). Recombinant expression and characterization of XynD from *Bacillus subtilis* subsp. *subtilis* ATCC 6051: a GH 43 arabinoxylan arabinofuranohydrolase. *Appl. Microbiol. Biotechnol.* **75:** 1309–1317.

Bravman T., Zolotnitsky G., Belakhov V., Shoham G., Henrissat B., Baasov T. and Shoham Y. (2003). Detailed kinetic analysis of a family 52 glycoside hydrolase: a beta-xylosidase from *Geobacillus stearothermophilus*. *Biochemistry* **42**: 10528–10536.

Bravman T., Zolotnitsky G., Shulami S., Belakhov V., Solomon D., Baasov T., Shoham G., Shoham Y. (2001). Stereochemistry of family 52 glycosyl hydrolases: a L-xylosidase from *Bacillus stearothermophilus* T-6 is a retaining enzyme. *FEBS Lett.* **495**: 39–43.

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

Brüx C., Ben-David A., Shallom-Shezifi D., Leon M., Niefind K., Shoham G., Shoham Y. and Schomburg D. (2006). The structure of an inverting GH43 beta-xylosidase from *Geobacillus stearothermophilus* with its substrate reveals the role of the three catalytic residues. *J. Mol. Biol.* **359**: 97–109.

Caffall K. H. and Mohnen D. (2009). The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydr. Res.* **344**: 1879–1900.

Cai L., Cai S., Zhao D., Wu J., Wang L., Liu X., Li M., Hou J., Zhou J., Liu J., Han J. and Xiang H. (2014). Analysis of the transcriptional regulator GlpR, promoter elements, and posttranscriptional processing involved in fructose-induced activation of the phosphoenolpyruvate-dependent sugar phosphotransferase system in *Haloferax mediterranei*. *Appl. Environ. Microbiol.* **80**: 1430–1440.

Casula G. and Cutting S. M. (2002). *Bacillus* Probiotics: Spore Germination in the Gastrointestinal Tract. *Appl. Environ. Microbiol.* **68**: 2344–2352.

Chai Y., Beauregard P. B., Vlamakis H., Losick R. and Kolter R. (2012). Galactose Metabolism Plays a Crucial Role in Biofilm Formation by *Bacillus subtilis*. *mBio* **3**: e00184-12.

Chumsakul O., Nakamura K., Kurata T., Sakamoto T., Hobman J. L., Ogasawara N., Oshima T. and Ishikawa S. (2013). High-resolution mapping of *in vivo* genomic transcription factor binding sites using *in situ* DNase I footprinting and ChIP-seq. *DNA Res.* **20**: 325–338.

Chumsakul O., Takahashi H., Oshima T., Hishimoto T., Kanaya S., Ogasawara N. and Ishikawa S. (2011). Genome-wide binding profiles of the *Bacillus subtilis* transition state regulator AbrB and its homolog Abh reveals their interactive role in transcriptional regulation. *Nucleic Acids Res.* **39**: 414–428.

Correia I. L., Franco I. S. and Sá-Nogueira I. (2014). Towards Novel Amino Acid-Base Contacts in Gene Regulatory Proteins: AraR – A Case Study. *PLoS One* **9**(11): e111802.

Coutinho P. M. and Henrissat B. (1999). Life with no sugars? *J. Mol. Microbiol. Biotechnol.* **1**: 307–308.

Cuneo M. J., Beese L. S. and Hellinga H. W. (2009). Structural analysis of semi-specific oligosaccharide recognition by a cellulose-binding protein of *Thermotoga maritima* reveals adaptations for functional diversification of the oligopeptide periplasmic binding protein fold. *J. Biol. Chem.* **284**: 33217–33223.

Czjzek M., Ben David A., Bravman T., Shoham G., Henrissat B. and Shoham Y. (2005). Enzyme-substrate complex structures of a GH39 beta-xylosidase from *Geobacillus stearothermophilus*. *J. Mol. Biol.* **353**: 838–846.

Dahms A. S. (1974). 3-Deoxy-D-pentulosonic acid aldolase and its role in a new pathway of D-xylose degradation. *Biochem. Biophys. Res. Commun.* **60**: 1433–1439.

Daniel R. A., Haiech J., Denizot F. and Errington J. (1997). Isolation and Characterization of the *lacA* Gene Encoding β-Galactosidase in *Bacillus subtilis* and a Regulator Gene, *lacR*. *J. Bacteriol.* **179**: 5636–5638.

Daus M. L., Grote M., Müller P., Doebber M., Herrmann A., Steinhoff H.-J., Dassa E. and Schneider E. (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 ($MalFGK_2$). *J. Biol. Chem.* **282**: 22387–22396.

Davidson A. L., Shuman H. A. and Nikaido H. (1992). Mechanism of maltose transport in *Escherichia coli*: transmembrane signalling by periplasmic binding proteins. *Proc. Natl. Acad. Sci. USA* **89**: 2360–2364.

Davidson A. L., Dassa E., Orelle C. and Chen J. (2008). Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol. Mol. Biol. Rev.* **72**: 317–364.

Dawson R. J. and Locher K. P. (2006). Structure of a bacterial multidrug ABC transporter. *Nature* **443**: 180–185.

Débarbouillé M., Arnaud M., Fouet A., Klier A. and Rapoport G. (1990). The *sacT* gene regulating the *sacPA* operon in *Bacillus subtilis* shares strong homology with transcriptional antiterminators. *J. Bacteriol.* **172**: 3966–3973.

DeLano W. L. (2002) PyMOL. DeLano Scientific, San Carlos.

Dellomonaco C., Fava F. and Gonzalez R. (2010). The path to next generation biofuels: successes and challenges in the era of synthetic biology. *Microb. Cell Fact.* **9**: 3.

Desai T. A. and Rao C. V. (2010). Regulation of Arabinose and Xylose Metabolism in *Escherichia coli*. *Appl. Environ. Microbiol.* **76**: 1524–1532.

de Sanctis D., Inácio J. M., Lindley P. F., Sá-Nogueira I. and Bento I. (2010). New evidence for the role of calcium in the glycosidase reaction of GH43 arabinanases. *FEBS J*. **277**: 4562–4574.

Dippel R. and Boos W. (2005). The maltodextrin system of *Escherichia coli*: metabolism and transport. *J. Bacteriol.* **187**: 8322–8331.

Dodd D. and Cann I. K. O. (2009). Enzymatic deconstruction of xylan for biofuel production *Bioenergy* **1**: 2–17.

Duez C., Zervosen A., Teller N., Melkonian R., Banzubazé E., Bouillenne F., Luxen A. and Frère J. M. (2009). Characterization of the proteins encoded by the *Bacillus subtilis yoxA-dacC* operon. *FEMS Microbiol. Lett.* **300**: 42–47.

Earl A. M., Losick R. and Kolter R. (2008). Ecology and genomics of *Bacillus subtilis*. *Trends Microbiol.* **16**: 269–275.

Edwards M. C. and Doran-Peterson J. (2012). Pectin-rich biomass as feedstock for fuel ethanol production. *Appl. Microbiol. Biotechnol.* **95**: 565–575.

Ehrmann M., Ehrle R., Hofmann E., Boos W. and Schlösser A. (1998). The ABC maltose transporter. *Mol. Microbiol.* **29**: 685–694.

Eitinger T., Rodionov D. A., Grote M. and Schneider E. (2011). Canonical and ECF-type ATP-binding cassette importers in prokaryotes: diversity in modular organization and cellular functions. *FEMS Microbiol. Rev.* **35**: 3-67.

El Omari K., Iourin O., Kadlec J., Fearn R., Hall D. R., Harlos K., Grimesa J. M. and Stuart D. I. (2014). Pushing the limits of sulfur SAD phasing: *de novo* structure solution of the N-terminal domain of the ectodomain of HCV E1. *Acta Cryst. Sect. D* **70**: 2197–2203.

Emsley P. and Cowtan K. (2004). *Coot*: model-building tools for molecular graphics *Acta Cryst. Sect. D* **60**: 2126–2132.

Englesberg E., Anderson R., Weinberg R., Lee N., Hoffee P., Huttenhauer G. and Boyer H. (1962). L-arabinose-sensitive, L-ribulose 5-phosphate 4-epimerase-deficient mutants of *Escherichia coli*. *J. Bacteriol.* **84**: 137–146.

Englesberg E., Squires C. and Meronk F. (1969). The arabinose operon in *Escherichia coli* B/r: a genetic demonstration of two functional states of the product of a regulatory gene. *Proc. Natl. Acad. Sci. USA* **62**: 1100–1107.

Evans P. (2006). Scaling and assessment of data quality. *Acta Cryst. Sect. D* **62**: 72–82.

Evdokimov A. G., Anderson D. E., Routzahn K. M. and Waugh D. S. (2001). Structural basis for oligosaccharide recognition by *Pyrococcus furiosus* maltodextrin-binding protein. *J. Mol. Biol.* **305**: 891–904.

Farrell A. E., Plevin R. J., Turner B. T., Jones A. D., O'Hare M. and Kammen D. M. (2006). Ethanol can contribute to energy and environmental goals. *Science* **311**: 506–508.

Ferrari E., Jarnagin A. S. and Schmidt B. F. (1993). Commercial Production of Extracellular Enzymes. In: Sonenshein A. L., Hoch J. A. and Losick R. (ed.) *Bacillus subtilis* and other Gram-positive Bacteria. American Society for Microbiology, Washington DC, USA, 917–937.

Ferreira M. J. and Sá-Nogueira I. (2010). A Multitask ATPase Serving Different ABC-Type Sugar Importers in *Bacillus subtilis*. *J. Bacteriol.* **192**: 5312–5318.

Franco I. L. S. (2007). Regulation of arabinose metabolism in *Bacillus subtilis*: linking structure and function (PhD Thesis). ITQB-UNL, Lisbon, Portugal.

Franco I., Mota L. J., Soares C. and Sá-Nogueira I. (2006). Functional domains of the *Bacillus subtilis* transcription factor AraR and identification of amino acids important for nucleoprotein complex assembly and effector binding. *J. Bacteriol.* **188**: 3024–3036.

Franco I., Mota L. J., Soares C. and Sá-Nogueira I. (2007). Probing key DNA contacts in AraR-mediated transcriptional repression of the *Bacillus subtilis* arabinose regulon. *Nucleic Acids Res.* **35**: 4755–4766.

Fujita M., González-Pastor J. E. and Losick R. (2005). High- and Low-Threshold Genes in the Spo0A Regulon of *Bacillus subtilis*. *J. Bacteriol.* **187**: 1357–1368.

Galinier A., Deutscher J and Martin-Verstraete I. (1999). Phosphorylation of either Crh or HPr mediates binding of CcpA to the *Bacillus subtilis xyn cre* and catabolite repression of the *xyn* operon. *J. Mol. Biol.* **286**: 307–314.

Gärtner D., Geissendörfer M. and Hillen W. (1988). Expression of the *Bacillus subtilis xyl* Operon Is Repressed at the Level of Transcription and Is Induced by Xylose. *J. Bacteriol.* **170**: 3102–3109.

Gilson E., Nikaido H. and Hofnung M. (1982). Sequence of the *malK* gene in *E. coli* K12. *Nucleic Acids Res.* **10**: 7449–7458.

Godinho L. and Sá-Nogueira I. (2011). Characterization and regulation of a bacterial sugar phosphatase of the haloalkanoate dehalogenase superfamily, AraL, from *Bacillus subtilis*. *FEBS J.* **278**: 2511–2524.

Goubet F., Misrahi A., Park S. K., Zhang Z. N., Twell D. and Dupree P. (2003). AtCSLA7, a cellulose synthase-like putative glycosyltransferase, is important for pollen tube growth and embryogenesis in *Arabidopsis*. *Plant Physiol.* **131**: 547–557.

Grossman A. D. (1995). Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis*. *Annu. Rev. Genet.* **29**: 477–508.

Guérout-Fleury A. M., Frandsen N. and Stragier P. (1996). Plasmids for ectopic integration in *Bacillus subtilis*. *Gene* **180**: 57–61.

Guldan H., Sterner R. and Babinger, P. (2008). Identification and characterization of a bacterial glycerol-1-phosphate dehydrogenase: Ni^{2+}-dependent AraM from *Bacillus subtilis*. *Biochemistry* **47**: 7376–7384.

Hahn-Hägerdal B., Karhumaa K., Fonseca C., Spencer-Martins I. and Gorwa-Grauslund MF. (2008). Towards industrial pentose-fermenting yeast strains. *Appl. Microbiol. Biotechnol.* **74**: 937–953.

Hamon M. A. and Lazazzera B. A. (2001). The sporulation transcription factor Spo0A is required for biofilm development in *Bacillus subtilis*. *Mol. Microbiol.* **42**: 1199–1209.

Härtl B., Wehrl W., Wiegert T., Homuth G. and Schumann W. (2001). Development of a new integration site within the *Bacillus subtilis* chromosome and construction of compatible expression cassettes. *J Bacteriol.* **183**: 2696–2699.

Haydon D. J. and Guest J. R. (1991). A new family of bacterial regulatory proteins. *FEMS Microbiol. Lett.* **63**: 291–295.

Heath E., Hurwitz J., Horecker B. and Ginsburg A. (1958). Pentose fermentation by *Lactobacillus plantarum*. I. The cleavage of xylulose 5-phosphate by phosphoketolase. *J. Biol. Chem.* **231**: 1009–1013.

Henkin T. M., Grundy F. J., Nicholson W. L. and Chambliss G. H. (1991). Catabolite repression of alpha-amylase gene expression in *Bacillus subtilis* involves a trans-acting gene product homologous to the *Escherichia coli* LacL and GalR repressors. *Mol. Microbiol.* **5:** 575–584.

Higgins C. F. and Ames G. F. (1981). Two periplasmic transport proteins which interact with a common membrane receptor show extensive homology: complete nucleotide sequences. *Proc. Natl. Acad. Sci. USA* **78:** 6038–6042.

Higgins, C. F. (1992). ABC Transporters: From microorganisms to Man. *Annu. Rev. Cell Biol.* **8**: 67–113.

Higgins C. F. (2001). ABC transporters: physiology, structure and mechanism – an overview. *Res. Microbiol.* **152**: 205–210.

Hoch J. A. (1993). *spo0* Genes, the Phosphorelay, and the Initiation of Sporulation. In: Sonenshein A. L., Hoch J. A. and Losick R. (ed.) *Bacillus subtilis* and other Gram-positive Bacteria. American Society for Microbiology, Washington DC, USA, 747–755.

Holden H. M., Rayment I. and Thoden J. B. (2003). Structure and function of enzymes of the Leloir pathway for galactose metabolism. *J. Biol. Chem.* **278**: 43885–43888.

Hong H. A., Duc L. H. and Cutting S. M. (2005). The use of bacterial spore formers as probiotics. *FEMS Microbiol. Rev.* **29**: 813–835.

Hong H. A., To E., Fakhry S., Baccigalupi L., Ricca E. and Cutting S. M. (2009). Defining the natural habitat of *Bacillus* spore-formers. *Res. Microbiol.* **160**: 375–379.

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

Inácio J. M., Costa C. and Sá-Nogueira I. (2003). Distinct molecular mechanisms involved in carbon catabolite repression of the arabinose regulon in *Bacillus subtilis*. *Microbiology* **149**: 2345–2355.

Inácio J. M. and Sá-Nogueira I. (2007). *trans*-Acting and *cis* elements involved in glucose repression of arabinan degradation in *Bacillus subtilis*. *J. Bacteriol.* **189**: 8371–8376.

Inácio J. M., Correia I. L. and Sá-Nogueira I. (2008). Two distinct arabinofuranosidases contribute to arabino-oligosaccharide degradation in *Bacillus subtilis*. *Microbiology* **154**: 2719–2729.

Inácio J. M. and Sá-Nogueira I. (2008). Characterization of *abn2* (*yxiA*), Encoding a *Bacillus subtilis* GH43 Arabinanase, Abn2, and Its Role in Arabino-Polysaccharide Degradation. *J. Bacteriol.* **190**: 4272–4280.

Jordan D. B., Lee C. C., Wagschal K. and Braker J. D. (2013). Activation of a GH43 β-xylosidase by divalent metal cations: Slow binding of divalent metal and high substrate specificity. *Arch. Biochem. Biophys.* **533**: 79–87.

Joyet P., Bouraoui H., Aké F. M., Derkaoui M., Zébré A. C., Cao T. N., Ventroux M., Nessler S., Noirot-Gros M. F., Deutscher J. and Milohanic E. (2013). Transcription regulators controlled by interaction with enzyme IIB components of the phosphoenolpyruvate: sugar phosphotransferase system. *Biochim. Biophys. Acta.* **1834**: 1415–1424.

Kalluri U. C., Yin H., Yang X. and Davison B. H. (2014). Systems and synthetic biology approaches to alter plant cell walls and reduce biomass recalcitrance. *Plant Biotechnol. J.* **12**: 1207-1216.

Kamionka A. and Dahl M. K. (2001). *Bacillus subtilis* contains a cyclodextrin-binding protein which is part of a putative ABC-transporter. *FEMS Microbiol. Lett.* **204**: 55–60.

Kempf B. and Bremer E. (1995). OpuA, an osmotically regulated binding protein-dependent transport system for the osmoprotectant glycine betaine in *Bacillus subtilis*. *J. Biol. Chem.* **270**: 16701–16713.

Kim J.-H., Prabhu P., Jeya M., Tiwari M. K., Moon H.-J., Singh R. K. and Lee J.-K. (2010). Characterization of an L-arabinose isomerase from *Bacillus subtilis*. *Appl. Microbiol. Biotechnol.* **85**: 1839–1847.

Kobayashi K., Ehrlich S. D., Albertini A., Amati G., Andersen K. K., Arnaud M., *et al.* (2003). Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. USA* **100**: 4678–4683.

Kraus A., Hueck C., Gärtner D. and Hillen W. (1994). Catabolite Repression of the *Bacillus subtilis xyl* Operon Involves a *cis* Element Functional in the Context of an Unrelated

Sequence, and Glucose Exerts Additional *xylR*-Dependent Repression. *J. Bacteriol.* **176**: 1738–1745.

Kreuzer P., Gärtner D., Allmansberger R. and Hillen W. (1989). Identification and Sequence Analysis of the *Bacillus subtilis* W23 *xylR* Gene and *xyl* Operator. *J. Bacteriol.* **171**: 3840–3845.

Krispin O. and Allmansberger R. (1998). The *Bacillus subtilis* AraE protein displays a broad substrate specificity for several different sugars. *J. Bacteriol.* **180**: 3250–3252.

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**: 249–256.

Le Quéré C., Moriarty R., Andrew R. M., Canadell J. G., Sitch S., Korsbakken J. I., *et al.* (2015). Global Carbon Budget. *Earth Syst. Sci. Data* **7**: 349–396.

Leal T. F. and Sá-Nogueira I. (2004). Purification, characterization and functional analysis of an endo-arabinanase (AbnA) from *Bacillus subtilis*. *FEMS Microbiol. Lett.* **241**: 41–48.

Lee C. C., Braker J. D., Grigorescu A. A., Wagschal K. and Jordan D. B. (2013). Divalent metal activation of a GH43 β-xylosidase. *Enzyme Microb. Technol.* **52**: 84–90.

Lengeler J. W. and Jahreis K. (2009). Bacterial PEP-dependent carbohydrate: phosphotransferase systems couple sensing and global control mechanisms. *Contrib. Microbiol.* **16**: 65–87.

Lepesant J. A. and Dedonder R. (1967). Metabolisme du L-arabinose chez *Bacillus subtilis* Marburg Ind⁻ 168. *C. R. Acad. Sci. Ser. D.* 2683–2686.

Lindner C., Stülke J. and Hecker M. (1994). Regulation of xylanolytic enzymes in *Bacillus subtilis*. *Microbiology* **140**: 753–757.

Linke C. M., Woodiga S. A., Meyers D. J., Buckwalter C. M., Salhi H. E. and King S. J. (2013). The ABC transporter encoded at the pneumococcal fructooligosaccharide utilization locus determines the ability to utilize long- and short-chain fructooligosaccharides. *J. Bacteriol.* **195**: 1031–1041.

Liu Q., Dahmane T., Zhang Z., Assur Z., Brasch J., Shapiro L., Mancia F. and Hendrickson W. A. (2012). Structures from anomalous diffraction of native biological macromolecules. *Science* **336**: 1033–1037.

Locher K., Lee A. T. and Rees D. C. (2002). The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* **296**: 1091–1098.

Loqué D., Scheller H. V. and Pauly M. (2015). Engineering of plant cell walls for enhanced biofuel production. *Curr. Opin. Plant Biol.* **25**: 151–161.

Marion C., Aten A. E., Woodiga S. A. and King S. J. (2011). Identification of an ATPase, MsmK, which energizes multiple carbohydrate ABC transporters in *Streptococcus pneumoniae*. *Infect. Immun.* **79**: 4193–4200.

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

Mathews D. H. (2006). RNA Secondary Structure Analysis Using RNAstructure. *In Curr. Protoc. Bioinformatics*. Chapter 12, Unit 12.6. John Wiley & Sons, Inc.

Mayer M. and Meyer B. (1999). Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chem. Int.* **38**: 1784–1788.

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

McCoy J. G., Levin E. J. and Zhou M. (2015). Structural insight into the PTS sugar transporter EIIC. *Biochim. Biophys. Acta.* **1850**: 577–585.

McNeil M., Darvill A. G., Fry S. C. and Albersheim P. (1984). Structure and function of the primary cell walls of plants. *Annu. Rev. Biochem.* **53**: 625–663.

McVey C. E., Ferreira M. J., Correia B., Lahiri S., de Sanctis D., Carrondo M. A., Lindley P. F., Sá Nogueira I., Soares C. M. and Bento I. (2014) The importance of the Abn2 calcium cluster in the endo-1,5-arabinanase activity from *Bacillus subtilis*. *J. Biol. Inorg. Chem.* **19**: 505–513.

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

Miwa Y., Nakata A., Ogiwara A., Yamamoto M. and Fujita Y. (2000). Evaluation and characterization of catabolite-responsive elements (*cre*) of *Bacillus subtilis*. *Nucleic Acids Res.* **28**: 1206–1210.

Mohnen D. (2008). Pectin structure and biosynthesis. *Curr. Opin. Plant Biol.* **11**: 266–277.

Molle V., Fujita M., Jensen S. T., Eichenberger P., González-Pastor J. E., Liu J. S. and Losick R. (2003). The Spo0A regulon of *Bacillus subtilis*. *Mol. Microbiol.* **50**: 1683–1701.

Moraïs S., Salama-Alber O., Barak Y., Hadar Y., Wilson D. B., Lamed R., Shoham Y. and Bayer E. A. (2012). Functional Association of Catalytic and Ancillary Modules Dictates Enzymatic Activity in Glycoside Hydrolase Family 43 β-Xylosidase. *J. Biol. Chem.* **287**: 9213–9221.

Morris G. M., Huey R., Lindstrom W., Sanner M. F., Belew R. K., Goodsell D. S. and Olson A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**: 2785–2791.

Mota L. J., Tavares P. and Sá-Nogueira I. (1999). Mode of action of AraR, the key regulator of L-arabinose metabolism in *Bacillus subtilis*. *Mol. Microbiol.* **33**: 476–489.

Mota L. J., Sarmento L. M. and Sá-Nogueira I. (2001). Control of the arabinose regulon in *Bacillus subtilis* by AraR in vivo: crucial roles of operators, cooperativity, and DNA looping. *J. Bacteriol.* **183**: 4190–4201.

Mourez M., Hofnung M. and Dassa E. (1997). Subunit interactions in ABC transporters: a conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits. *EMBO J.* **16**: 3066–3077.

Münch R., Hiller K., Grote A., Scheer M., Klein J., Schobert M. and Jahn D. (2005). Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes. *Bioinformatics* **21**: 4187–4189.

Nataf Y., Yaron S., Stahl F., Lamed R., Bayer E. A., Scheper T.-H., Sonenshein A. L. and Shoham Y. (2009). Cellodextrin and Laminaribiose ABC Transporters in *Clostridium thermocellum*. *J. Bacteriol.* **191**: 203–209.

Nevoigt E. (2008). Progress in metabolic engineering of *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **72**: 379–412.

Nikaido H. and Hall J. A. (1998). Overview of bacterial ABC transporters. *Methods Enzymol.* **292**: 3–20.

Numan M. T. and Bhosle N. B. (2006). α-L-arabinofuranosidases: the potential applications in biotechnology. *J. Ind. Microbiol. Biotechnol.* **33**: 247–260.

Nurizzo D., Turkenburg J. P., Charnock S. J., Roberts S. M., Dodson E. J., McKie V. A., Taylor E. J., Gilbert H. J. and Davies G. J. (2002). *Cellvibrio japonicus* α-L-arabinanase 43A has a novel five-blade β-propeller fold. *Nat. Struct. Biol.* **9**: 665–668.

Ochiai A., Itoh T., Kawamata A., Hashimoto W. and Murata K. (2007). Plant Cell Wall Degradation by Saprophytic *Bacillus subtilis* Strains: Gene Clusters Responsible for Rhamnogalacturonan Depolymerization. *Appl. Environ. Microbiol.* **73**: 3803–3813.

Oldham M. L., Khare D., Quiocho F. A., Davidson A. L. and Chen J. (2007). Crystal structure of a catalytic intermediate of the maltose transporter. Nature **450**: 515–521.

Oldham M. L., Davidson A. L. and Chen J. (2008). Structural insights into ABC transporter mechanism. *Curr. Opin. Struct. Biol.* **18**: 726–733.

Oldham M. L. and Chen J. (2011) Crystal structure of the maltose transporter in a pretranslocation intermediate state. *Science* **332**: 1202–1205.

Oldham M. L., Chen S. and Chen J. (2013). Structural basis for substrate specificity in the *Escherichia coli* maltose transport system. *Proc. Natl. Acad. Sci. USA* **110**: 18132–18137.

Painter J. and Merritt E. A. (2005). A molecular viewer for the analysis of TLS rigid-body motion in macromolecules. *Acta Cryst. Sect. D* **61**: 465–471.

Pajatsch M., Gerhart M., Peist R., Horlacher R., Boos W. and Bock A. (1998). The periplasmic cyclodextrin binding protein CymE from *Klebsiella oxytoca* and its role in maltodextrin and cyclodextrin transport. *J. Bacteriol.* **180**: 2630–2635.

Panagiotidis C. H. and Shuman H. A. (1998) Maltose transport in *Escherichia coli*: Mutations that uncouple ATP hydrolysis from transport. *Methods Enzymol*. **292**: 30–39.

Pao S. S., Paulsen I. T. and Saier M. H., Jr. (1998). The major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* **62**: 1–32.

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

Pauly M., Gille S., Liu L., Mansoori N., de Souza A., Schultink A. and Xiong G. (2013). Hemicellulose biosynthesis. *Planta* **238**: 627–642.

Pauly M. and Keegstra K. (2010). Plant cell wall polymers as precursors for biofuels. *Curr. Opin. Plant Biol.* **13**: 305–312.

Petit E., Coppi M. V., Hayes J. C., Tolonen A. C., Warnick T., Latouf W. G., Amisano D., Biddle A., Mukherjee S., Ivanova N., Lykidis A., Land M., Hauser L., Kyrpides N., Henrissat B., Lau J., Schnell D. J., Church G. M., Leschine S. B. and Blanchard J. L. (2015). Genome and Transcriptome of *Clostridium phytofermentans*, Catalyst for the Direct Conversion of Plant Feedstocks to Fuels. *PLoS One* **10**(6): e0118285.

Petronilli V., and Ames G. F.-L. (1991). Binding protein-independent histidine permease mutants. Uncoupling of ATP hydrolysis from transmembrane signaling. *J. Biol. Chem.* **266**: 16293–16296.

Pettolino F. A., Walsh C., Fincher G. B. and Bacic A. (2012). Determining the polysaccharide composition of plant cell walls. *Nat. Protoc.* **7**: 1590–1607.

Pickl A., Johnsen U. and Schönheit P. (2012). Fructose degradation in the haloarchaeon *Haloferax volcanii* involves a bacterial type phosphoenolpyruvate-dependent phosphotransferase system, fructose-1-phosphate kinase, and class II fructose-1,6-bisphosphate aldolase. *J. Bacteriol.* **194**: 3088–3097.

Poelarends G. J. and Konings W. N. (2002). The transmembrane domains of the ABC multidrug transporter LmrA form a cytoplasmic exposed, aqueous chamber within the membrane, *J. Biol. Chem.* **277**: 42891–42898.

Pons T., Naumoff D. G., Martinez-Fleites C. and Hernandez L. (2004). Three acidic residues are at the active site of a β-propeller architecture in glycoside hydrolase families 32, 43, 62, and 68. *Proteins* **54**: 424–432.

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. American Society for Microbiology, Washington DC, USA, 3–16.

Pueyo C. and Lopez-Barea A. (1979). The L-arabinose-resistance test with *Salmonella typhimurium* strain SV3 selects forward mutations at several *ara* genes. *Mutat. Res.* **64**: 249–258.

Puls J. and Schuseil J. (1993). Chemistry of hemicelluloses: relationship between hemicellulose structure and enzyme required for hydrolysis. In: Coughlan M. P. and Hazlewood G. P. (ed.), Hemicellulose and Hemicellulases. Portland Press, London.

Quentin Y., Fichant G. and Denizot F. (1999). Inventory, assembly and analysis of *Bacillus subtilis* ABC transport systems. *J. Mol. Biol.* **287**: 467–484.

Raposo M. P., Inácio J. M., Mota L. J. and Sá-Nogueira I. (2004). Transcriptional regulation of arabinan-degrading genes in *Bacillus subtilis*. *J. Bacteriol.* **186**: 1287–1296.

Ren D., Navarro B., Xu H., Yue L., Shi Q. and Clapham D. E. (2001). A prokaryotic voltage-gated sodium channel. *Science* **294**: 2372–2375.

Rhee M. S., Wei L., Sawhney N., Rice J. D., St. John F. J., Hurlbert J. C. and Preston J. F. (2013). Engineering the xylan utilization system in *Bacillus subtilis* for production of acidic Xylooligosaccharides. *Appl. Environ. Microbiol.* **80**: 917–927.

Roncero M. I. G. (1983). Genes controlling xylan utilization by *Bacillus subtilis*. *J. Bacteriol.* **156**: 257–263.

Russell R. R. B., Aduse-Opoku J., Sutcliffe I. C., Tao L. and Ferretti J. J. (1992). A binding protein-dependent transport system in *Streptococcus mutans* responsible for multiple sugar metabolism. *J. Biol. Chem.* **267**: 4631–4637.

Rygus T., Scheler A., Allmansberger R. and Hillen W. (1991). Molecular cloning, structure, promoters and regulatory elements for transcription of the *Bacillus megaterium* encoded regulon for xylose utilization. *Arch. Microbiol.* **155**: 535–542.

Sadaie Y., Nakadate H., Fukui R., Yee L. M. and Asai K. (2008). Glucomannan utilization operon of *Bacillus subtilis*. *FEMS Microbiol. Lett.* **279**: 103–109.

Saha B. C. (2003). Hemicellulose bioconversion. *J. Ind. Microbiol. Biotechnol.* **30**: 279–291.

Saier M. H., Jr. (2000). Families of proteins forming transmembrane channels. *J. Membr. Biol.* **175**: 165–180.

Saier M. H., Jr. (2015). The Bacterial Phosphotransferase System: New Frontiers 50 Years after Its Discovery. *J. Mol. Microbiol. Biotechnol.* **25**: 73–78.

Saier M. H., Jr., Goldman S. R., Maile R. R., Moreno M. S., Weyler W., Yang N. and Paulsen I. T. (2002a). Overall transport capabilities of *Bacillus subtilis*. In: Sonenshein A. L., Losick R. and Hoch J. A. (ed.), *Bacillus subtilis* and its closest relatives – from genes to cells. American Society for Microbiology, Washington DC, USA.

Saier M. H., Jr., Goldman S. R., Maile R. R., Moreno M. S., Weyler W., Yang N. and Paulsen I. T. (2002b). Transport Capabilities Encoded Within the *Bacillus subtilis* Genome. *J. Mol. Microbiol. Biotechnol.* **4**: 37–67.

Saito A., Fujii T., Shinya T., Shibuya N., Ando A. and 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**: 3358–3365.

Sambrook J., Fritsch E. F. and Maniatis T. (1989). Molecular Cloning: a Laboratory Manual, 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sá-Nogueira I. and Lencastre H. (1989). Cloning and characterization of *araA*, *araB* and *araD*, the structural genes for L-arabinose utilization in *Bacillus subtilis*. *J. Bacteriol.* **171**: 4088–4091.

Sá-Nogueira I. and Mota L. J. (1997). Negative regulation of L-arabinose metabolism in *Bacillus subtilis*: characterization of the *araR* (*araC*) gene. *J. Bacteriol.* **179**: 1598–1608.

Sá-Nogueira I., Nogueira T. V., Soares S. and Lencastre H. (1997). The *Bacillus subtilis* L-arabinose (*ara*) operon: nucleotide sequence, genetic organization and expression. *Microbiology* **143**: 957–969.

Sá-Nogueira I. and Ramos S. S. (1997). Cloning, functional analysis, and transcriptional regulation of the *Bacillus subtilis araE* gene involved in L-arabinose utilization. *J. Bacteriol.* **179**: 7705–7711.

Saurin W., Köster W. and Dassa E. (1994) Bacterial binding protein-dependent permeases: characterization of distinctive signatures for functionally related integral cytoplasmic membrane proteins. *Mol. Microbiol*. **12**: 993–1004.

Sawhney N., Crooks C., St. John F. and Preston J. F. (2015). Transcriptomic Analysis of Xylan Utilization Systems in *Paenibacillus* sp. Strain JDR-2. *Appl. Environ. Microbiol.* **81**: 1490–1501.

Scheler A., Rygus T., Allmansberger R. and Hillen W. (1991). Molecular cloning, structure, promoters and regulatory elements for transcription of the *Bacillus licheniformis* encoded regulon for xylose utilization. *Arch. Microbiol.* **155**: 526–534.

Scheller H. V. and Ulvskov P. (2010). Hemicelluloses. *Annu. Rev. Plant. Biol.* **61**: 263–289.

Schlösser A. (2000). MsiK-dependent trehalose uptake in *Streptomyces reticuli*. *FEMS Microbiol. Lett.* **184**: 187–192.

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

Schönert S., Seitz S., Krafft H., Feuerbaum E., Andernach I., Witz G. and Dahl M. K. (2006). Maltose and maltodextrin utilization by *Bacillus subtilis*. *J. Bacteriol.* **188**: 3911–3922.

Serra C. R., Earl A. M., Barbosa T. M., Kolter R. and Henriques A. O. (2014). Sporulation during Growth in a Gut Isolate of *Bacillus subtilis*. *J. Bacteriol.* **196**: 4184–4196.

Shipkowski S. and Brenchley J. E. (2006). Bioinformatic, genetic, and biochemical evidence that some glycoside hydrolase family 42 β-galactosidases are arabinogalactan type I oligomer hydrolases. *Appl. Environ. Microbiol.* **72**: 7730–7738.

Shallom D. and Shoham Y. (2003). Microbial hemicellulases. *Curr. Opin. Microbiol.* **6**: 219–228.

Sharff A. J., Rodseth L. E., Spurlino J. E. and Quiocho F. A. (1992). Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the

maltodextrin binding protein involved in active transport and chemotaxis. Biochemistry **31**: 10657–10663.

Shitan N., Bazin I., Dan K., Obata K., Kigawa K., Ueda K., Sato F., Forestier C. and Yazaki K. (2003). Involvement of CjMDR1, a plant multidrug-resistance-type ATP-binding cassette protein, in alkaloid transport in *Coptis japonica*. *Proc. Natl. Acad. Sci. USA* **100**: 751–756.

Shulami S., Raz-Pasteur A., Tabachnikov O., Gilead-Gropper S., Shner I. and Shoham Y. (2011). The L-Arabinan Utilization System of *Geobacillus stearothermophilus*. *J. Bacteriol.* **193**: 2838–2850.

Shulami S., Shenker O., Langut Y., Lavid N., Gat O., Zaide G., Zehavi A., Sonenshein A. L. and Shoham Y. (2014). Multiple regulatory mechanisms control the expression of the *Geobacillus stearothermophilus* gene for extracellular xylanase. *J. Biol. Chem.* **289**: 25957–25975.

Shulami S., Zaide G., Zolotnitsky G., Langut Y., Feld G., Sonenshein A. L. and Shoham Y. (2007). A two-component system regulates the expression of an ABC transporter for xylo-oligosaccharides in *Geobacillus stearothermophilus*. *Appl. Environ. Microbiol.* **73**: 874–884.

Shuman H. A. (1982). Active transport of maltose in *Escherichia coli* K-12: role of the periplasmic maltose binding protein and evidence for a substrate recognition site in the cytoplasmic membrane. *J. Biol. Chem.* **257**: 5455–5461.

Simpson F., Wolin M. and Wood W. (1958). Degradation of L-arabinose by *Aerobacter aerogenes*. I. A pathway involving phosphorylated intermediates. *J. Biol. Chem.* **230**: 457–463.

Singh K. D., Schmalisch M. H., Stülke J. and Görke B. (2008). Carbon catabolite repression in *Bacillus subtilis*: quantitative analysis of repression exerted by different carbon sources. *J. Bacteriol.* **190**: 7275-7284.

Somogyi M. (1952). Notes on sugar determination. *J. Biol. Chem.* **195**: 19–23.

Song Y., Xue Y. and Ma Y. (2013). Global Microarray Analysis of Carbohydrate Use in Alkaliphilic Hemicellulolytic Bacterium *Bacillus* sp. N16-5. *PLoS One* **8**(1): e54090.

St. John F. J., Rice J. D. and Preston J. F. (2006). Characterization of XynC from *Bacillus subtilis* subsp. *subtilis* Strain 168 and Analysis of Its Role in Depolymerization of Glucuronoxylan. *J. Bacteriol.* **188**: 8617–8626.

Stephens C., Christen B., Fuchs T., Sundaram V., Watanabe K. and Jenal U. (2007). Genetic Analysis of a Novel Pathway for D-Xylose Metabolism in *Caulobacter crescentus*. *J. Bacteriol.* **189**: 2181–2185.

Strauch M. A., Webb V., Spiegelman G. and Hoch J. A. (1990). The SpoOA protein of *Bacillus subtilis* is a repressor of the *abrB* gene. *Proc. Natl. Acad. Sci. USA* **87**: 1801–1805.

Strauch M. A. (1993). AbrB, a transition state regulator. In: Sonenshein A. L., Hoch J. A. and Losick R. (ed.) *Bacillus subtilis* and other Gram-positive Bacteria. American Society for Microbiology, Washington DC, USA, 757–764.

Strauch M. A. (1995). Delineation of AbrB-binding sites on the Bacillus subtilis *spo0H*, *kinB*, *ftsAZ*, and *pbpE* promoters and use of a derived homology to identify a previously unsuspected binding site in the *bsuB1* methylase promoter. *J. Bacteriol.* **177**: 6999–7002.

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

Studier F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**: 207–234.

Stülke J. and Hillen W. (2000). Regulation of carbon catabolism in *Bacillus subtilis*. *Annu. Rev. Microbiol.* **54**: 849–880.

Sun T. and Altenbuchner J. (2010). Characterization of a mannose utilization system in *Bacillus subtilis*. *J. Bacteriol.* **192**: 2128–2139.

Tabachnikov O. and Shoham Y. (2013). Functional characterization of the galactan utilization system of *Geobacillus stearothermophilus*. *FEBS J.* **280**: 950–964.

Tam N. K. M, Uyen N. Q., Hong H. A., Duc L. H., Hoa T. T., Serra C. R., Henriques A. O. and Cutting S. M. (2006). The Intestinal Life Cycle of *Bacillus subtilis* and Close Relatives. *J. Bacteriol.* **188**: 2692–2700.

Tan M.-F., Gao T., Liu W.-Q., Zhang C.-Y., Yang X., Zhu J.-W., Teng M.-Y., Li L. and Zhou R. (2015). MsmK, an ATPase, Contributes to Utilization of Multiple Carbohydrates and Host Colonization of *Streptococcus suis*. *PLoS One* **10**(7): e0130792.

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

Teixeira V. H., Cunha C. A., Machuqueiro M., Oliveira A. S., Victor B. L., Soares C. M. and Baptista A. M. (2005). On the Use of Different Dielectric Constants for Computing Individual and Pairwise Terms in Poisson-Boltzmann Studies of Protein Ionization Equilibrium. *J. Phys. Chem. B* **109**: 14691–14706.

Teixeira V. H., Soares C. M. and Baptista A. M. (2002). Studies of the reduction and protonation behavior of tetraheme cytochromes using atomic detail. *J. Biol. Inorg. Chem.* **7**: 200–216.

Teplitsky A., Shulami S., Moryles S., Shoham Y. and Shoham G. (2000). Crystallization and preliminary X-ray analysis of an intracellular xylanase from *Bacillus stearothermophilus* T-6. *Acta Cryst. Sect. D* **56**: 181–184.

Terasaka K., Blakeslee J. J., Titapiwatanakun B., Peer W. A., Bandyopadhyay A., Makam S. N., Lee O. R., Richards E. L., Murphy A. S., Sato F. and Yazaki K. (2005). PGP4, an ATP binding cassette P-glycoprotein, catalyzes auxin transport in *Arabidopsis thaliana* roots. *Plant Cell* 17: 2922–2939.

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

Thomson J. A. (1993). Molecular biology of xylan degradation. *FEMS Microbiol. Rev.* **10**: 65-82.

Tobisch S., Glaser P., Krüger S. and Hecker M. (1997). Identification and characterization of a new beta-glucoside utilization system in *Bacillus subtilis*. *J. Bacteriol.* **179**: 496–506.

Tsujibo H., Kosaka M., Ikenishi S., Sato T., Miyamoto K. and Inamori Y. (2004). Molecular characterization of a high-affinity xylobiose transporter of *Streptomyces thermoviolaceus* OPC-520 and its transcriptional regulation. *J. Bacteriol.* **186**: 1029–1037.

van Maris A. J. A., Abbott D. A., Bellissimi E., van den Brink J., Kuyper M., Luttik M. A. H., Wisselink H. W., Scheffers W. A., van Dijken J. P. and Pronk J. T. (2006). Alcoholic fermentation of carbon sources in biomass hydrolysates by *Saccharomyces cerevisiae*: current status. *Anton. Leeuw. Int. J. G.* **90**: 391–418.

Velazquez-Campoy A. and Freire E. (2006). Isothermal titration calorimetry to determine association constants for high-affinity ligands. *Nat. Protoc.* **1**: 186–191.

Velázquez-Campoy A., Ohtaka H., Nezami A., Muzammil S. and Freire E. (2004). Isothermal titration calorimetry. *Curr. Protoc. Cell. Biol.* 23: 17.8.1–17.8.24.

Watanabe S., Kodaki T. and Makino K. (2006). Cloning, expression, and characterization of bacterial L-arabinose 1-dehydrogenase involved in an alternative pathway of L-arabinose metabolism. *J. Biol. Chem.* **281**: 2612–2623.

Webb A. J., Homer K. A. and Hosie A. H. F. (2008). Two closely related ABC transporters in *Streptococcus mutans* are involved in disaccharide and/or oligosaccharide uptake. *J. Bacteriol.* **190**: 168–178.

Weickert M. J. and Adhya S. (1992). A family of bacterial regulators homologous to Gal and Lac repressors. *J. Biol. Chem.* **267:** 15869–15874.

Weimberg R. (1961). Pentose oxidation by *Pseudomonas fragi*. *J. Biol. Chem.* **236**: 629–635.

Weinstein L. and Albersheim P. (1979). Structure of plant cell walls. IX. Purification and partial characterization of a wall-degrading endo-arabanase and an arabinosidase from *Bacillus subtilis. Plant Physiol.* **63**: 425–432.

Wenzel M. and Altenbuchner J. (2013). The *Bacillus subtilis* mannose regulator, ManR, a DNA-binding protein regulated by HPr and its cognate PTS transporter ManP. *Mol. Microbiol.* **88**: 562–576.

Winn M. D., Isupov M. N. and Murshudov G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Cryst. Sect. D* **57**: 122–133.

Xu K. and Strauch M. A. (1996). *In vitro* selection of optimal AbrB-binding sites: comparison to known *in vivo* sites indicates flexibility in AbrB binding and recognition of three-dimensional DNA structures. *Mol. Microbiol.* **19**: 145–158.

Yanisch-Perron C., Vieira J. and Messing J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103–119.

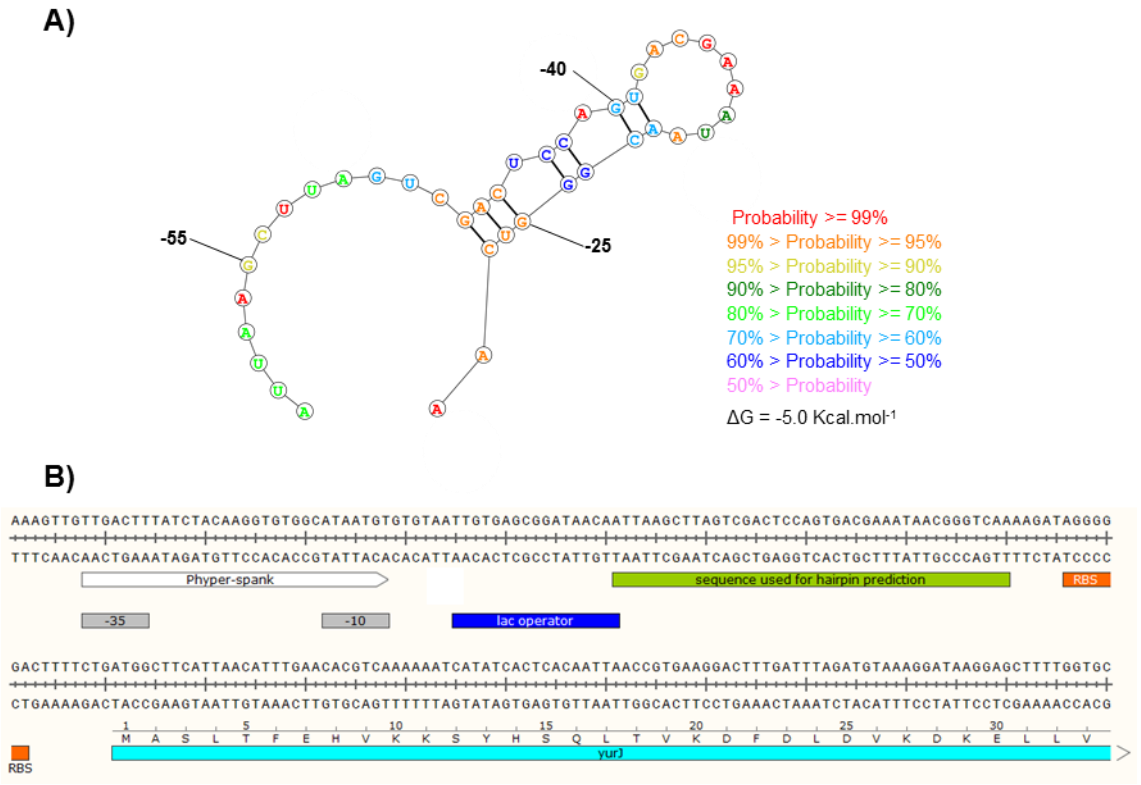
Yoshida K., Ishio I., Nagakawa E., Yamamoto Y., Yamamoto M. and Fujita Y. (2000). Systematic study of gene expression and transcription organization in the *gntZ–ywaA* region of the *Bacillus subtilis* genome. *Microbiology* **146**: 573–579.

Yuan J. S., Tiller K. H., Al-Ahmad H., Stewart N. R. and Stewart C. N., Jr. (2008). Plants to power: bioenergy to fuel the future. *Trends Plant Sci.* **13**: 421–429.

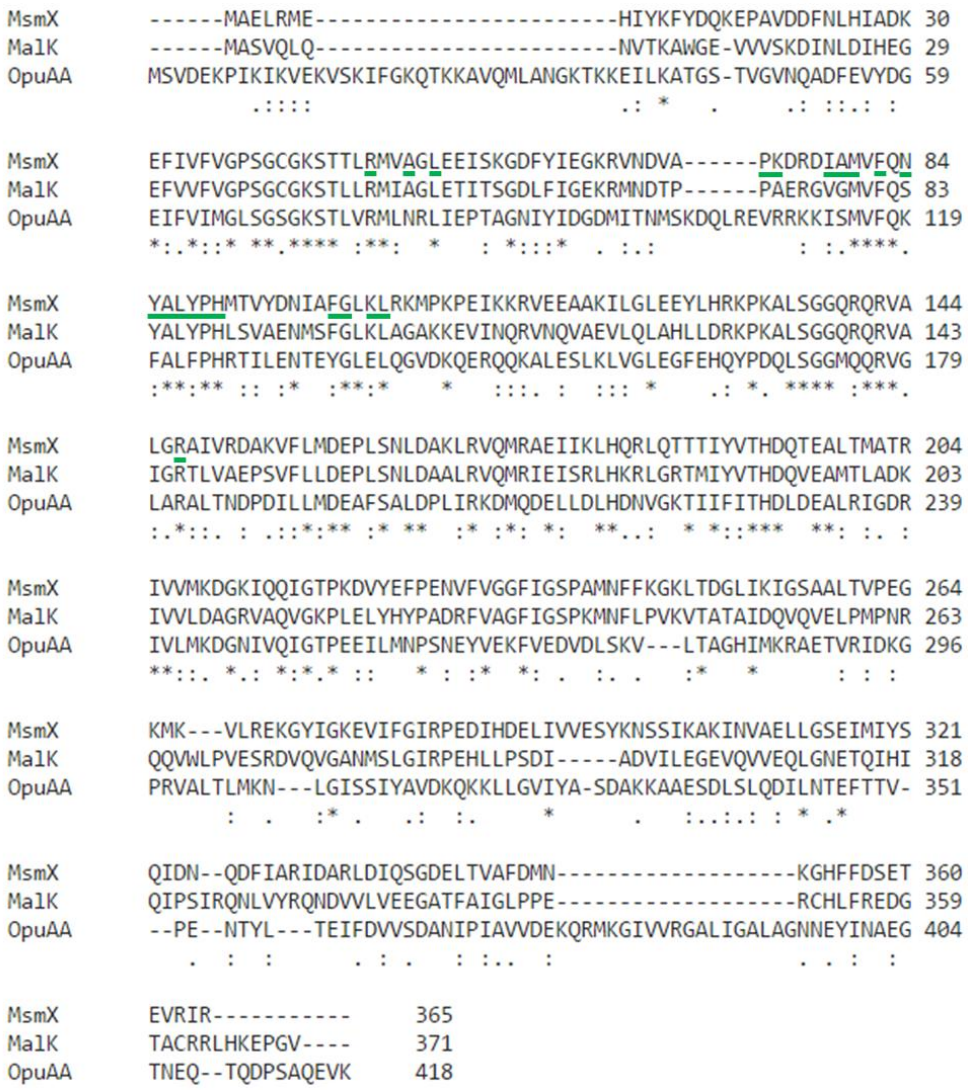
Zhang Y. (2007). Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins* **69**(Suppl 8): 108–117.

Zilhão R., Serrano M., Isticato R., Ricca E., Moran Jr. C. P. and Henriques A. O. (2004). Interactions among CotB, CotG, and CotH during assembly of the *Bacillus subtilis* spore coat. *J. Bacteriol.* **186**: 1110–1119.

Annexes

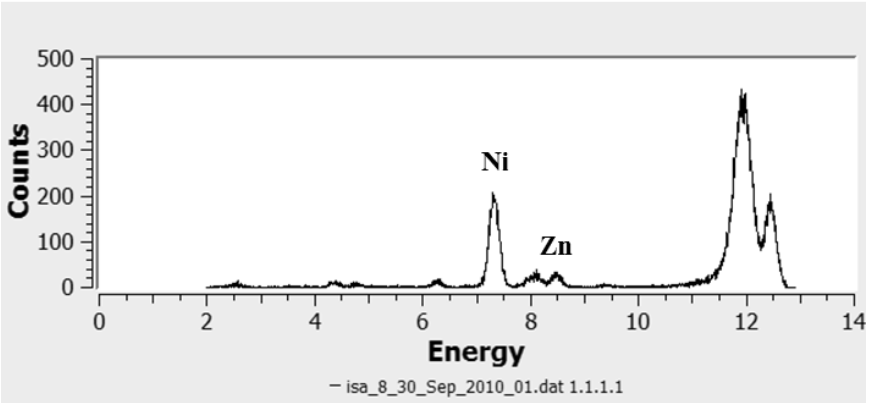


Supplementary Figure 1 – Secondary structure prediction of the *yurJ* transcript from the copy of the gene inserted in the *amyE* locus of the chromosome and under control of an IPTG-inducible promoter in strain IQB642. A) This putative structure has a higher free energy change (-5.0 Kcal.mol^{-1}) when compared to that observed for the wild-type strain and base-pairing probability is considerably low. The secondary structure prediction was obtained using the online tool RNAstructure (http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html; Mathews, 2006) and base-pair positions indicated are relative to translation initiation. The probability for base-pairing or unpairing indicated. **B)** Promoter region of the *yurJ* gene in strain IQB642. The *yurJ* ORF (*cyan*), the ribosome-binding site (RBS, *orange*), the Phyper-spank promoter (*white*, with the -10 and -35 boxes signaled in *gray*), the *lac* operator (*blue*) and the sequence used for hairpin prediction (*green*) are highlighted.

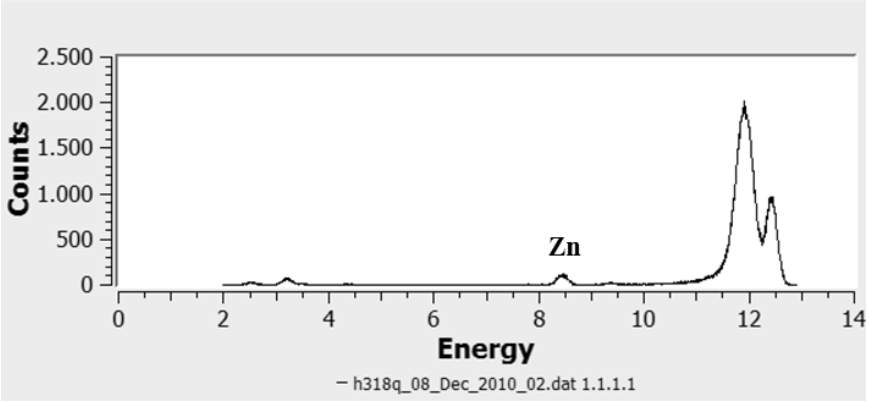


Supplementary Figure 2 – Sequence alignment of three ABC-type ATPases. The sequence alignments were obtained with the online tool Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). The amino acid sequences of MalK from *E. coli*, and MsmX and OpuAA from *B. subtilis* were used. *Green* lines identify residues of MalK involved in the interactions with the TMDs of the *E. coli* maltose transporter and that can also be involved in the interactions between MsmX and the systems it energizes in *B. subtilis*.

a)



b)



Supplementary Figure 3 – X-ray emission spectra of a) H318Q crystals and of b) H318Q protein solution.