Lia Raquel Marques Godinho

MSc. in Biotechnology

Characterization of genes of unknown function in *Bacillus subtilis*: gene regulation and functional analysis

Dissertation to obtain a PhD degree in Biology

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

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

May 2016

À minha Família,

Acknowledgements

I would like to present my acknowledgements to Fundação para a Ciência e Tecnologia (FCT) for financial support with PhD fellowship SFRH/BD/73109/2010 and UciBio @ Requimte (and previously Centro de Recursos Microbiológicos) for providing all the work conditions for the completion of this thesis.

I also want to thank and acknowledge Professor Isabel de Sá-Nogueira, my supervisor, without whom this thesis would not exist. Thank you for the opportunity to join your lab, your guidance and knowledge and for your enthusiasm for science. You have passed along valuable knowledge, science-related and otherwise and I am grateful for it.

I want to acknowledge my thesis committee members, Professor Helena Santos (Head of Cell Physiology and NMR group at ITQB António Xavier - NOVA) and Professor Adriano Henriques (Head of Microbial Development group at ITQB António Xavier - NOVA), for the scientific input in this thesis and helpful discussions.

A special acknowledgment to everyone at the Cell Physiology and NMR lab, who welcomed me as one of their own all the times I had to work there, particularly Dušica Radoš, who ran all the NMR experiments with me and provided major help with data analysis presented in this thesis. A word of appreciation to Ana Isabel Mingote and Ana Lúcia Carvalho, who participated in the early experiments.

To my lab mates from Microbial Genetics, from the first day I started at the lab, who were either permanent or transient: Joana Pedro and Joana Lima, Inês Martins, Sónia, Renato, Miguel, Maria João, Liliana and all other undergrads. I had the opportunity to get to know and help some of them along the way, but I also learned from it, scientifically and personally.

To Mário, for all the times we shared in the lab – work related discussions and flat out dorky and geeky discussions. It wouldn't have been the same without you dude! And sorry about all those times I spoiled your favorite TV shows! Maria Isabel, a thousand "thank yous" would not suffice to thank you for all you did for me, be it science-related or otherwise. Viviana and Tides, I am truly thankful you two boarded the 327 crazy train. I could not have asked for better co-workers, honestly – your sense of humor (even the lame jokes!), encouraging words and friendship made these times worth every minute.

To "330" members Damien and Bárbara, thanks for being a part of this ride, for always being there with an easy smile and words of encouragement. To Raquel, a true pillar for the 330 gang, beloved friend and PhD student next door – thank you for your friendship, helpful discussions and support. Thank you for always being on the other side of that wall. We're lucky to have "adopted" you.

Thank you to all my coworkers at the Departamento de Ciências da Vida at FCT-UNL, with special thanks to Márcia Palma, Carla Gonçalves, Miguel Larguinho, João Rosa, and at a later phase, the neighbors from Jaime Mota lab (Lia, Irina, Filipe, Sara, Nuno, Joana and Maria). Also a word of appreciation to the 427 lab from Requimte, especially to Professor Susana Barreiros and Alexandre Paiva, for allowing me to use laboratory equipment and material, and also for their kind words of encouragement.

Aos meus amigos de sempre, tão antigos que parece que estão comigo desde o início dos tempos. Ana, Hugo, Marta, PP e Rita, obrigada pelo vosso apoio, encorajamento, momentos de riso e de parvoíce. Obrigada por serem sempre constantes.

À Diana, cuja distância física de 1600 km foi sempre colmatada por chamadas pelo Skype que duravam até às 2h da manhã em dias de semana, conversas no Whatsapp ou vindas relâmpago a Portugal que tiveram sempre espaço para mim. Obrigada por me aturares, ouvires e aconselhares… e por me trazeres sempre chocolates.

Um agradecimento especial à família Pump, que entrou na minha vida já na fase final desta maratona: Lucy, Nat, Alvaro, Fábio, Di, Victória e tantos outros que não consigo nomear neste espaço. Obrigada pelo apoio, boa disposição e bons momentos passados no mundo do *fitness*, que foram muito importantes nestes últimos tempos. #tamosjuntos. Um obrigada especial à Catarina "Cacá Jam", que passou de sócia a instrutora e depois a amiga do coração: obrigada por estares sempre pronta para ouvir desabafos e reclamações e no fim me levares para fazer um Body Jam, fosse onde fosse.

Agradeço muito à minha família, em especial aos meus pais, Virgínia e Custódio, e à minha irmã, Irina, por todo o apoio incondicional e amor, por serem uma referência e um pilar e por estarem sempre a torcer por mim. Sem vocês não teria sido possível concretizar esta tese.

Finalmente, mas não menos importante, um obrigada muito especial ao Ângelo, por ser o poço de calma e motivação que me atura todos os dias há quase meia dúzia de anos. Fosse nos dias bons, nos dias maus, nos dias de sol ou nos dias de chuva houve sempre compreensão, carinho, amor e amizade ♥.

Abbreviations and Acronyms

3C –3 carbon

5C – 5 carbon

6C – 6 carbon

6PG – 6-phosphogluconate

$\mu g.mL^{-1}$ – micrograms per milliliter

µM – micromolar

µl – microliter

[ ] – concentration

αMG6P – α-methyl glucoside-6-phosphate

Abs – Absorbance

ADP – Adenosine diphosphate

Acetyl-CoA – acetyl Coenzyme A

ara – arabinose

ATP – Adenosine triphosphate

BCAA – branched- chain amino acids

bla – β-lactamase gene conferring resistance to ampicillin

bp – base pair

BSH – Bacillithiol

BSA – Bovine Serum Albumin

cAMP – cyclic adenosine monophosphate

cat – chloramphenicol acetyl transferase gene, conferring resistance to chloramphenicol

CcpA – catabolite control protein A

CCR – carbon catabolite repression

CO_2 – carbon dioxide

cre – catabolite responsive elements

Crh – catabolite repression HPr

CRP – cAMP receptor protein

DHAP – dihydroxyacetone phosphate

DDH – DDH phosphodiesterases family

DIP – D-myo-inositol

DNA – deoxyribonucleic acid

dNTPs – deoxynucleotides

E4P – erythrose 4-phosphate

ECL – electrochemiluminescence

EDTA – Ethylenediaminetetraacetic acid

EMP – Embden-Meyerhof-Parnas

ED – Entner–Doudoroff

erm – erythromycin resistance gene

F6P – fructose 6-phosphate

FAD – flavin adenine dinucleotide

FBP – fructose 1,6-bisphosphate

FDA – Food and Drug Administration

G6P – glucose 6-phosphate

GA3P – glyceraldehyde 3-phosphate

Glycerol-3P – glycerol 3-phosphate

GRAS – Generally Regarded As Safe

HPr – histidine-containing phosphocarrier protein

HADSF – Haloacid Dehalogenase Super Family

HD – HD phosphohydrolases family

HCl – hydrochloric acid

IPTG - Isopropyl β-D-1-thiogalactopyranoside

km – kanamycin resistance gene

kb – kilobase

k_{cat} – catalytic constant

KCl – potassium chloride

kDa – kiloDalton

K_M – Michaelis constant

LB – Luria Broth

M – molar

MG – methylglyoxal

mM – micromolar

mRNA – messenger RNA

NAD^+/NADH – nicotinamide adenine dinucleotide oxidized/reduced

$NADP^+$/NADPH – nicotinamide adenine dinucleotide phosphate oxidized/reduced

NTPases – nucleotide phosphatases

nm – nanometer

nM – nanomolar

NTA Ni^{2+} – nitrilotriacetic acid

NTC – no template control

OR – operator

ORF – open reading frame

Para – arabinose operon promoter

PEP – phosphoenol pyruvate

PCR – polymerase chain reaction

PHP – polymerase and histidinol phosphatase family of proteins

P*i* – inorganic phosphate

PMSF – phenylmethylsulfonyl fluoride

*p*NPP – *p*-nitrophenyl phosphate

ppm – parts per million

PPP – pentose phosphate pathway

PTS – phosphotransferase system

QPS – Qualified Presumption of Safe

qRT-PCR – quantitative real time polymerase chain reaction

R5P – ribulose 5-phosphate

RBS – ribosome binding site

rib – ribitol

RLU – relative luminescence units

RNA – ribonucleic acid

RNase – ribonuclease

S7P – sedoheptulose 7-phosphate

SDS-PAGE – sodium dodecyl sulfate polyacrylamide gel electrophoresis

TCA – tricarboxylic acid cycle

UDP-galactose – uracil-diphosphate galactose

UDP-glucose – uridine-diphosphate glucose

UV – ultraviolet

X5P – xylulose 5-phosphate

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

w/v – weight per volume

Amino acid	Three letter code	One letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Try	W
Tyrosine	Tyr	Y
Valine	Val	V

Base	Letter
Adenine	A
Cytosine	C
Guanine	G
Thymine	T

Abstract

Bacillus subtilis is involved in the enzymatic degradation of plant biomass, namely polysaccharides with a high content in arabinose, one of the most abundant pentose in nature. The operon *araABDLMNPQ-abfA* of *Bacillus subtilis*, is responsible for arabinan utilization and arabinose catabolism, however the role of two genes, *araL* and *araM*, in this context is still elusive. AraM is a dehydrogenase and AraL a phosphatase but they are not necessary for arabinose utilization. Transcription of the metabolic operon is induced by arabinose and negatively regulated by AraR. In an *araR*-null mutant, addition of arabinose to an exponentially growing culture results in immediate cessation of growth. In this study to investigate the role of both genes in the toxic effect caused by arabinose, in-frame deletions of *araL* and *araM* were constructed and their impact analyzed. The results strongly suggest that *araL* and *araM* do not participate in this phenomenon. In addition, AraL, which belongs to the haloacid dehalogenase HAD superfamily, was biochemically characterized and substrate screening showed AraL to have low specificity and catalytic activity towards several sugar phosphates. Thus, we propose a putative physiological role of AraL in detoxification of accidental accumulation of phosphorylated metabolites. AraL production is regulated by a structure in the translation initiation region of the mRNA, which probably blocks access to the ribosome-binding site, preventing protein synthesis.

Accumulation of sugar phosphate is known to be toxic for the majority of prokaryotic and eukaryotic cells, however the mechanisms that underlie toxicity are yet to be fully understood. Here, we investigated the growth arrest phenotype displayed in the presence of arabinose by *B. subtilis* strains lacking the regulator AraR. The current hypothesis is that the bacteriostatic effect observed could be due to an increased intracellular level of arabinose, which consequently arise the concentration of the metabolic sugar phosphates intermediates that are toxic to the cell. Analysis of both wild-type and mutant strains by quantification of mRNA levels, phosphorylated metabolites, accumulation of cytotoxic methylglyoxal, and ATP depletion, suggests distinct mechanisms underlying toxicity. This study highlights the importance of a secondary metabolic pathway regulator in the optimal growth of an industrial relevant species, *B. subtilis*, and how its deletion may negatively impact the overall central carbon metabolism.

Keywords: *Bacillus subtilis*; sugar phosphatase; Haloacid Dehalogenase; phosphosugar toxicity; arabinose metabolism; methylglyoxal

Resumo

Bacillus subtilis é uma bactéria Gram-positiva envolvida na degradação enzimática da biomassa vegetal nomeadamente polissacáridos com elevado conteúdo em arabinose, uma das pentoses mais abundantes na natureza. Em *B. subtilis* o operão *araABDLMNPQ-abfA* é responsável pela utilização de arabinano e arabinose, ainda que a função de dois dos seus genes, *araL* e *araM*, nesse contexto seja pouco clara. AraM é uma desidrogenase e AraL uma fosfatase, sendo ambas dispensáveis para a utilização de arabinose. A transcrição do operão metabólico é induzida pela arabinose e negativamente regulada pela proteína AraR. Num mutante nulo *araR*, a adição de arabinose a uma cultura em crescimento exponencial resulta numa paragem imediata no crescimento. Neste trabalho, para avaliar o papel dos dois genes no efeito tóxico causado pela arabinose foram construídas deleções em grelha de leitura dos genes *araL* e *araM* e o seu efeito analisado. Os resultados indicam que nem *araL* nem *araM* se encontram envolvidos neste fenómeno de toxicidade.

AraL, o produto do gene *araL*, pertencente à superfamília de proteínas HAD, foi bioquimicamente caracterizado e uma análise de substratos mostrou que AraL possui baixa especificidade e actividade catalítica em relação a vários açúcares fosforilados, podendo AraL ter uma função na destoxificação de metabolitos fosforilados que acumulem na célula. A produção de AraL é regulada por uma estrutura secundária ao nível do mRNA, situada na região de início de transcrição, bloqueando o acesso ao local de ligação ribossómico e impedindo a síntese proteica.

A acumulação de açúcares fosforilados é tóxica para a maioria dos organismos, no entanto os mecanismos associados a esta toxicidade não são ainda totalmente compreendidos. Aqui investigou-se o fenótipo de paragem de crescimento de uma cultura em fase exponencial após adição de arabinose num mutante de *B. subtilis* desprovido do regulador AraR. A hipótese actual assenta no pressuposto de o efeito bacteriostático ser causado por um aumento na concentração intracelular de arabinose, e consequente aumento na concentração intracelular de intermediários fosforilados, sendo estes tóxicos para a célula. A análise de uma estirpe selvagem e de estirpes mutantes através da quantificação de níveis de mRNA, determinação da acumulação do metabolito citotóxico metilglioxal e comparação dos níveis de ATP, aponta para mecanismos distintos na base desta toxicidade. Este trabalho realça a importância de um regulador de uma via metabólica secundária no crescimento de um microrganismo industrialmente relevante, como *B. subtilis*, e como a sua deleção pode ter um impacto negativo no metabolismo central de carbono.

Palavras-chave: *Bacillus subtilis*; fosfatase de açúcares; Desalogenase Haloácida; toxicidade de açúcares fosforilados; metabolismo da arabinose; metilglioxal

Thesis Outline

This thesis is organized in five chapters. Chapter I is an introduction to *Bacillus subtilis*, and carbohydrate uptake by this organism. Main carbohydrate utilization pathways, glycolysis, pentose phosphate pathway and tricarboxylic acid cycle are described, as well as general mechanisms of carbon catabolite control. This chapter also focuses on L-arabinose metabolism and regulation of arabinose utilization genes, followed by an overview of HAD phosphatases and sugar phosphate toxicity in bacteria.

Chapter II emphasizes the role of both *araL* and *araM* genes in the context of the arabinose operon, as they are not necessary for L-arabinose utilization. This chapter describes experiments to investigate their role in the phenomenon of toxicity caused by arabinose in a strain deregulated for arabinose utilization. The results strongly suggest that *araL* and *araM* do not play a role in the toxic effect of arabinose observed in an *araR*-null mutant.

Chapter III explains the cloning, overexpression, and biochemical characterization of AraL from *B. subtilis*. The enzyme displays phosphatase activity, low specificity and catalytic activity towards several sugar phosphates, which are metabolic intermediates of the glycolytic and pentoses phosphate pathways. Moreover, results evidence the existence a genetic regulatory mechanism controlling AraL production at the mRNA level, with formation of a secondary structure in the translation initiation region of the mRNA. The putative role of AraL in the context of arabinose utilization is discussed here.

Chapter IV focuses on sugar phosphate toxicity studies in bacteria, using *B. subtilis* as a model organism. Arabinose is toxic to a mutant lacking the negative regulator of the arabinose operon, AraR. By combining different techniques, it is shown that arabinose-sensitivity of the mutant strain is accompanied by an increase in gene expression concomitant with an increase in the levels of arabinose-degrading enzymes, as well as accumulation of several phosphorylated intermediates from the pentose phosphate and glycolytic pathways and the cytotoxic compound methylglyoxal, and a drop in ATP.

Finally, Chapter V includes concluding remarks and future perspectives on the work.

Table of Contents

Index of Figures

Index of Tables

Chapter I

General Introduction

Bacillus subtilis: a Gram-positive model organism

Identified in the early 19th century by German naturalist Christian Gottfried Ehrenberg and rediscovered in 1872 by Ferdinand Cohn (Skerman *et al.*, 1989), *Bacillus subtilis* rose to prominence after the description of the wild (non-domesticated) *B. subtilis* Marburg strain in 1930, and subsequent isolation of X-ray induced mutants of a domesticated strain, related to *B. subtilis* Marburg, in the late 1940's (Burkholder & Giles, 1947; Conn, 1930; Zeigler *et al.*, 2008). Domestication of several of those mutagenized strains and their adaptation to laboratory life allowed researchers, in the early stages of genetic engineering studies, to use them to study the development of a competent state for exogenous naked-DNA uptake (Anagnostopoulos & Spizizen, 1960; Spizizen, 1958) and sporulation. At the dawn of the molecular biology era, *B. subtilis* amenability for genetic engineering emerged as a powerful tool, turning this mesophilic, rod-shaped Gram-positive endospore-forming bacterium in one of the most studied microorganisms in the scientific community, along with its Gram-negative counterpart, *Escherichia coli*, and the eukaryotic microorganism *Saccharomyces cerevisiae*. Development of competence, production and secretion of hydrolytic enzymes, non-pathogenicity and endospore formation were driving forces in the study of biochemistry, genetics and physiology of *B. subtilis*, bringing this bacterium into the spotlight of academic research, being used for the development of recombinant DNA techniques, such as cell transformation or construction of cloning vectors (integrational, replicative and shuttle vectors) some of which were used to maximize expression of native and heterologous proteins (Harwood, 1992). Furthermore, *B. subtilis* was among the first microorganisms to have its genome fully sequenced by a consortium in 1997 (Kunst *et al.*, 1997).

B. subtilis and other *Bacillus* species left the exclusivity of wet labs in academia and entered the industrial setting (Harwood, 1992), which took advantage of their genetic manipulation amenability, coupled with their secretion capacity, remarkable fermentation properties, high product yields (van Dijl & Hecker, 2013; Westers *et al.*, 2004) and interesting properties of their spore (Cutting, 2011; Hong *et al.*, 2005). Considered a GRAS organism by the FDA or QPS by the European Commission (McNeil *et al.*, 2013), *Bacillus* species, such as *B. subtilis*, *B. amyloliquefaciens*, *B. licheniformis*, *B. thuringiensis* are used for the production of industrial enzymes (*e.g.* proteases, α-amylases, restriction enzymes), which are often improved by protein engineering, fine biochemicals (hypoxantine, riboflavin), antibiotics (bacilysin, subtilin) and insecticides.

As of 2004, over 60% of all commercially available enzymes originated from *Bacillus* species (Westers *et al.*, 2004). Commercial proteases from *Bacillus* are usually alkaline or neutral (Rao *et al.*, 1998; Schallmey *et al.*, 2004) , turning them suitable for use in food and detergent industry. Detergent proteases commonly found in household detergents are subtilisins, *Bacillus* spp. serine proteases.

Another type of enzymes of industrial interest are *B. subtilis* hemicellulases, namely for biotechnological purposes focused on cellulose hydrolysis from lignocellulosic substrates, which have potential as renewable energy source for bioethanol production (Araújo & Ward, 1990; Várnai *et al.*, 2011). Ethanol-producing processes require release of cellulose and hemicellulose from their complex with lignin, depolymerization of cellulose and hemicellulose to obtain free sugars, and fermentation of mixed hexose and pentose sugars to produce ethanol (Lee, 1997). *B. subtilis* hemicellulases, such as mannanases and galactanases can be used to synergistically breakdown lignin (Araújo & Ward, 1990) and its arabinofuranosidases degrade hemicellulosic homopolysaccharides (branched and debranched arabinans) and heteropolysaccharides (arabinoxylans, arabinogalactans). The release of monosaccharides such as arabinose, xylose and galactose, is accomplished together with β-xylosidases and β-galactosidases (Inácio *et al.*, 2008; Shallom & Shoham, 2003). Improvement of lignocellulose conversion to ethanol via metabolic and evolutionary engineering techniques can bypass the scarcity of microorganisms that can efficiently convert hexoses and pentoses to ethanol, although efficient sugar utilization coupled with high yield ethanol production is yet be achieved (Wiedemann & Boles, 2008). For example, the studies of Boles and co-workers attempt at enhancement of an efficient biomass-to-ethanol fermentation process using metabolic engineering of a high-yield ethanol producer *S. cerevisiae*. The substrate utilization range of the yeast was increased by establishing an L-arabinose utilization pathway from bacteria, namely *B. subtilis* (Becker & Boles, 2003; Subtil & Boles, 2011; Wiedemann & Boles, 2008). The genetic and physiological characteristics of *Bacillus* species and especially *B. subtilis* allowed it to be widely used in basic and applied research as a model organism for Gram-positive bacteria, as well as an industrial workhorse for over 60 years.

Carbohydrate metabolism of *Bacillus subtilis* in its natural habitat

Bacillus subtilis natural habitat. Gut or soil?

In order to delve into *B. subtilis* carbohydrate utilization, we first must establish the bacterium's natural habitat. This Gram-positive endospore-forming microorganism can be found either in terrestrial or in aquatic environments, and its spores have been isolated from the gastrointestinal tract of several animals, including humans (Hong *et al.*, 2009; Tam *et al.*, 2006). *B. subtilis* has been historically classified as a soil saprophyte involved in the common effort of degrading plant biomass, with the soil as primary reservoir, entering the gastrointestinal tract of animals by ingestion whilst associated to vegetal biomass. Recent work, however, has been evidencing its role not as a transient passenger of the gastrointestinal tract but as having adapted to carry out their entire life cycle, i.e. germinate and sporulate, within this environment (Tam *et al.*, 2006). As such, an intricate network of metabolic routes for biosynthesis of the building blocks of proteins, nucleic acids, lipids and carbohydrates, as well as different forms catabolism of various compounds to fuel cellular processes, emerged.

Whether in the gastrointestinal tract of animals or in the soil, *B. subtilis* is an active participant in plant cell wall degradation, possessing several intra and extracellular enzymes able to break down complex sugar molecules (i.e. pectin, cellulose, and hemicellulose). The genome encodes numerous genes involved in pathways for the utilization of plant-derived molecules (Barbe *et al.*, 2009; Earl *et al.*, 2008; Kunst *et al.*, 1997), and it secretes a vast number of polysaccharide-backbone-degrading enzymes, which yield oligosaccharides. The complete breakdown of these oligosaccharides is carried out by cell-associated or intracellular enzymes.

B. subtilis often encounters a mixture of different and transient carbon sources that can potentially be used. Mechanisms have evolved, enabling selective uptake and metabolism of carbon sources that allow rapid growth and yield best success whilst competing with other microorganisms for survival.

Central Carbon Metabolism in *Bacillus subtilis*

Glycolysis, or the Embden-Meyerhof-Parnas Pathway

The three major metabolic pathways for carbon oxidation in *B. subtilis* are glycolysis, the pentose phosphate pathway and the tricarboxylic acid cycle. Like most heterotrophic bacteria, *B. subtilis* preferred carbon source is glucose, which is metabolized via glycolysis. Glycolysis is defined as a sequence of reactions that metabolize one molecule of glucose to two molecules of pyruvate, producing two adenosine triphosphate (ATP) molecules along the way (Berg *et al.*, 2011).

In eukaryotes, glucose catabolism occurs through the Embden-Meyerhof-Parnas pathway (EMP), in order to oxidize glucose to pyruvate with the concomitant production of oxidized nicotinamide adenine dinucleotide (NADH). The majority of prokaryotes have similar mechanisms to oxidize glucose, although, due to microbial diversity, some of them metabolize glucose through unique pathways found only in Bacteria, like the Entner-Doudoroff (ED) pathway, found in *Zymomonas mobilis* and phosphoketolase (PK) pathway, found in lactic acid bacteria like *Lactobacillus* spp. or *Bifidobacterium* spp. (Kim & Gadd, 2008; Wolfe, 2015). Archaeal glycolysis presents major variations mainly in its upper branch, concerning the steps from glucose to 3-phosphoglycerate. Modified EMP and ED pathways are quite common in Archaea (Kim & Gadd, 2008; Verhees *et al.*, 2003).

B. subtilis presents the most common glycolytic route, the EMP pathway, where glucose is catabolized to pyruvate (reviewed in Deutscher *et. al.* 2002) . EMP starts with the phosphorylation of glucose, followed by the isomerization to fructose 6-phosphate and a second phosphorylation (to fructose 1,6-bisphosphate - FBP), the aldol cleavage of FBP and phosphorylation of glyceraldehyde 3-phosphate (GA3P), further converted into pyruvate (Figure 1.1)**.**

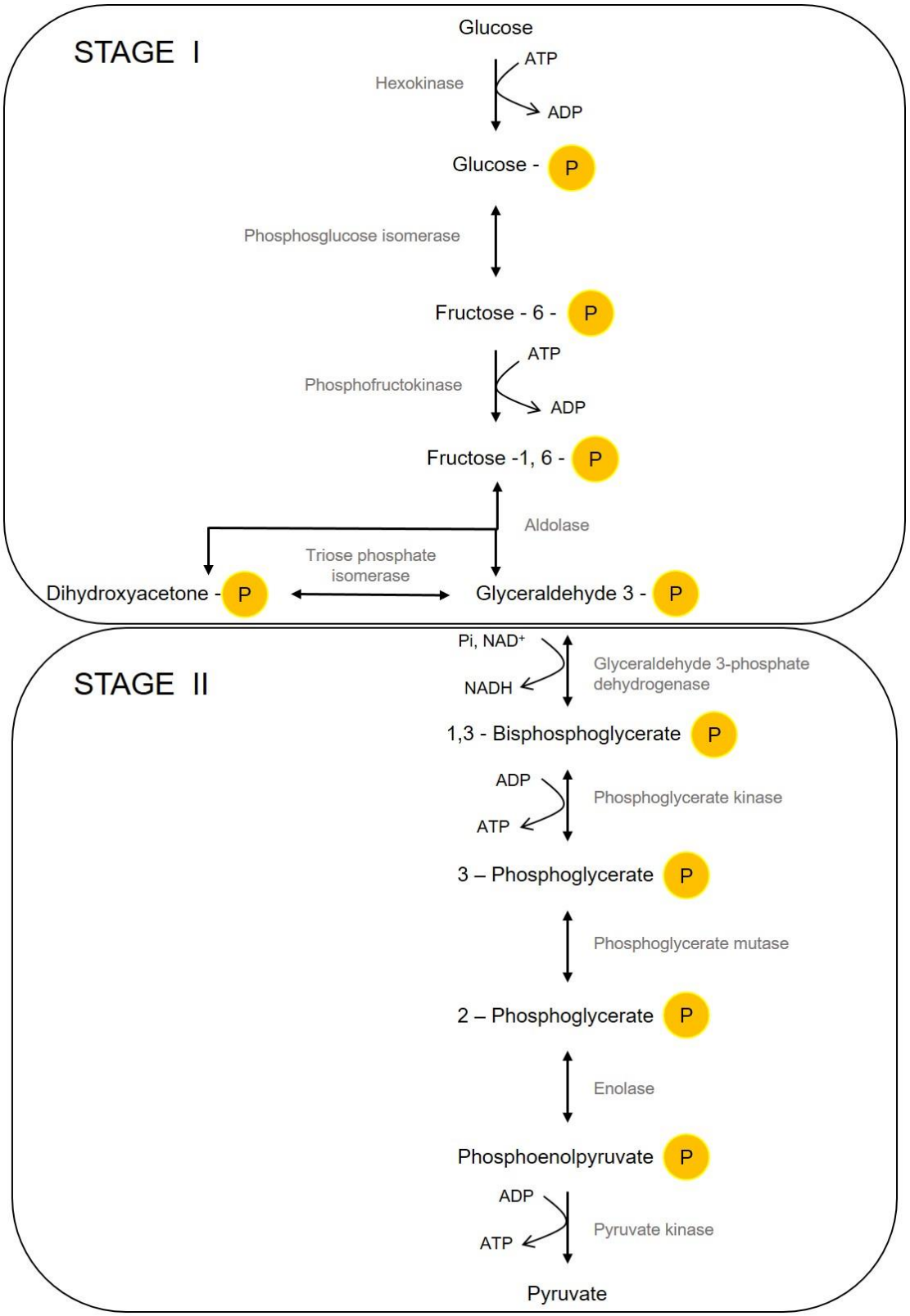


Figure 1.1. Embden–Meyerhof–Parnas pathway (EMP). Schematization of the conversion of glucose into pyruvate. Here represented are the two branches of glycolysis: Stage I, where 6C glucose is metabolized to triose phosphates (3C) and Stage II, where ATP and NADH are formed. Single-head arrows indicate irreversible enzymatic reactions and double-head arrows represent reversible enzymatic reactions.

Usually in the literature glycolysis is divided in two parts: the upper branch where glucose (6C) is converted to triose phosphates (3C), and where energy is consumed i.e. using ATP to energize the sugars, and the lower branch where there is a gain in energy through formation of reduction equivalents i.e. NADH (Commichau *et al.*, 2009).

In *B. subtilis* glucose enters the cell via group translocation, through the glucose-specific phosphotransferase system (PTS), which, in turn, is encoded by the *ptsGHI* operon. The phosphate donor in this transport system is phosphoenolpyruvate (PEP), a glycolytic intermediate and sugar translocation occurs via two PTS enzymes, enzyme I (EI) and a histidine-containing phosphocarrier protein (HPr), enzyme II (EII), the latter being sugar-specific (Stülke *et al.*, 1997). Glucose can also be phosphorylated to glucose 6-phosphate (G6P) by the cell glucokinase, GlcK, as unphosphorylated glucose may accumulate in the cytoplasm, resulting from disaccharide hydrolysis.

The following steps in EMP are the isomerization of G6P to fructose 6-phosphate (F6P) (reversible reaction) and the phosphorylation of F6P to FBP, which is the first non-reversible step in the pathway. The enzyme catalyzing this reaction is phosphofructokinase (Pfk), which is the most important control element in the glycolytic pathway, phosphorylating F6P at a second site irreversibly. This allosteric enzyme is inhibited by ATP and F6P, thus regulating the glycolytic flux (Berg *et al.*, 2011; Byrnes *et al.*, 1994).

The aldol cleavage of FBP by fructose-1,6-bisphosphate aldolase yields two triose phosphate molecules, dihydroxyacetone phosphate (DHAP) and GA3P. This cleavage is reversible, as this enzyme is also active during gluconeogenesis, when *B. subtilis* is using less preferred carbon sources such as malate or succinate. Because various degradation pathways feed into glycolysis at many different points, glycolysis or portions of it run in the forward or reverse direction, depending on the carbon source being utilized, in order to satisfy the cell need for precursor metabolites and energy.

The interconversion of DHAP together with the synthesis of phosphoenolpyruvate (lower branch of glycolysis) plays a central role in the metabolic network not only because it yields ATP, but also because it generates reducing power and important metabolic precursors for carbohydrate catabolism, linking metabolites from the upper branch of glycolysis and from the pentose phosphate pathway to the tricarboxylic acid cycle (Krebs cycle), namely G6P, Ribose 5-Phosphate (R5P) and Glycerol 3-Phosphate (glycerol-3P) (Doan & Aymerich, 2003).

Pentose Phosphate Pathway

The Pentose Phosphate Pathway (PPP) is another of the major metabolic pathways for carbon oxidation in Bacteria, and specifically in *B. subtilis*. The Pentose Phosphate Pathway becomes important as an entry point for several pentoses commonly found in nature, like xylose, arabinose or ribose. This

pathway is also comprised by two phases: oxidative generation of reduced nicotinamide adenine dinucleotide phosphate (NADPH) and non-oxidative interconversion of sugars.

The first phase, oxidative PPP is a major source of reductant (namely NADPH) for biosynthetic processes, such as fatty-acid synthesis and the assimilation of inorganic nitrogen, maintaining the redox potential necessary to protect against oxidative stress by oxidizing G6P to R5P. In this phase, the three steps necessary for the conversion of G6P to R5P are irreversible. R5P and its derivatives are the source of carbon skeletons for the synthesis of major cell components, like DNA and proteins, through synthesis of nucleotides, aromatic amino acids, phenylpropanoids and others (Berg *et al.*, 2011; Kim & Gadd, 2008; Moat *et al.*, 2002).

The second phase of PPP is a non-oxidative phase (Figure 1.2) comprising reversible reactions, which interconverts phosphorylated sugars. Combination of two C5 sugars, glycolytic intermediates that can be readily used in glycolysis or gluconeogenesis, depending on the cell metabolic state.

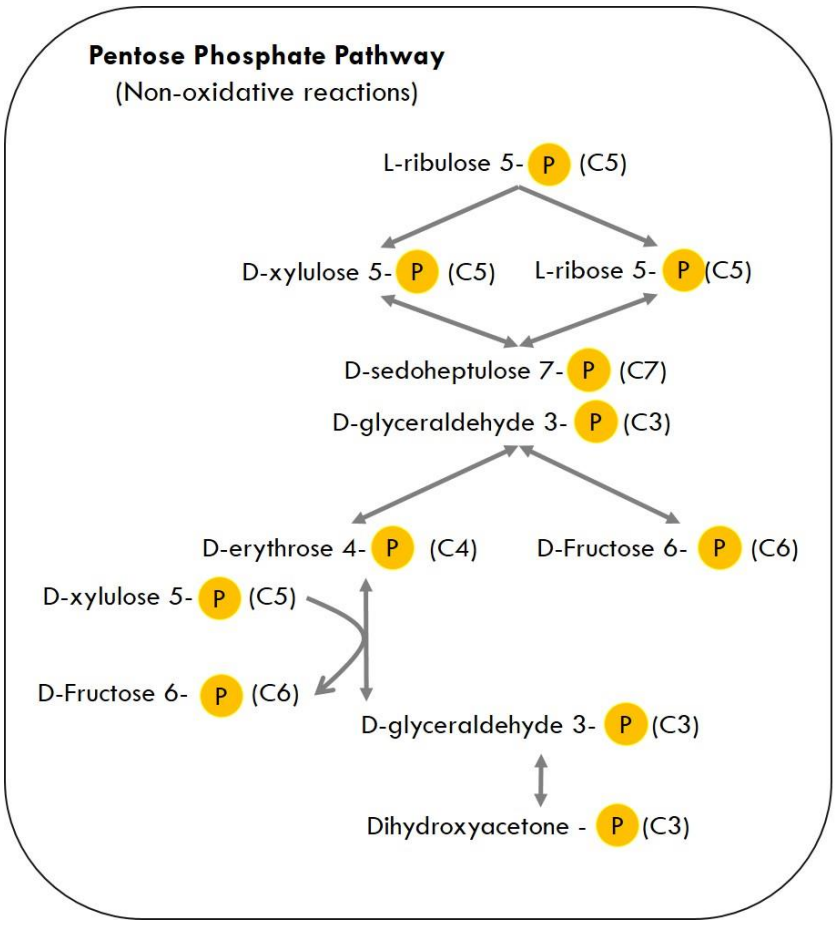


Figure 1.2. Non-oxidative reactions of the Pentose Phosphate Pathway (PPP). Single-head arrows indicate irreversible enzymatic reactions and double-head arrows represent reversible enzymatic reactions. Most of the reactions are reversible, meaning that the rearrangement of several sugar phosphates in the non-oxidative phase of the PPP allows intermediaries to be used for catabolism (F6P, DHAP, GA3P) and ATP production or to generate riboses from glycolytic intermediates for biosynthesis (R5P, E4P).

The transketolase accepts a 2C fragment from a 5C ketose (xylulose 5-phosphate, X5P) and then transfers this fragment to a 5C aldose (R5P), forming a 7C ketose (sedoheptulose 7-phosphate, S7P). Abstraction of 2 carbons from X5P yields a 3C-aldose, GA3P. Both molecules (S7P and GA3P) are then combined by transaldolase, yielding erythrose 4-phosphate (E4P) and F6P. E4P enters the chorismate biosynthesis pathway, which ultimately leads to aromatic amino acid formation (Light *et al.*, 2012). GA3P and F6P are phosphorylated sugars that can enter the lower part of glycolysis. PPP can either feed riboses into glycolysis for catabolism and ATP production or generate riboses from glycolytic intermediates for biosynthesis.

Tricarboxylic Acid Cycle

Under aerobic conditions, pyruvate resulting from glycolysis is converted into acetyl-CoA, which readily enters the Tricarboxylic Acid Cycle (TCA) (Figure 1.3), where ATP is generated and several building blocks for biosynthesis originate from. An example is α-ketoglutarate for glutamate synthesis or oxaloacetate, which can be converted to amino acids similar to aspartate (Bartholomae *et al.*, 2014). The oxidative decarboxylation of pyruvate (and other acetyl groups from different sources, such as succinate, fumarate or malate) to form acetyl-CoA is the link between glycolysis and the TCA; this irreversible channeling of the product of glycolysis into the TCA is catalyzed by the pyruvate dehydrogenase complex, encompassing a series of redox reactions that result in the oxidation of two molecules of acetyl-CoA to two molecules of CO_2 (Berg *et al.*, 2011; Cohen, 2011; Kim & Gadd, 2008). In summary, the TCA cycle starts with the irreversible oxidative decarboxylation of pyruvate to acetyl-CoA by the pyruvate dehydrogenase complex concomitant with NAD^+, $NADP^+$ and flavin adenine dinucleotide (FAD) reduction (Commichau *et al.*, 2009).

Citrate synthase (encoded by *citZ*) irreversibly condenses acetyl-CoA with oxaloacetate, yielding a 6C tricarboxylic acid (citric acid), which aconitase (*citB*) isomerizes to isocitrate. Isocitrate is oxidatively decarboxylated by isocitrate dehydrogenase (*icd*), which occurs simultaneously with the reduction of $NADP^+$ and the release of CO_2. The resulting five-carbon compound (α-ketoglutarate) is also oxidatively decarboxylated by enzymes of the α-ketoglutarate dehydrogenase complex (*odhA* and *odhB*) to yield a four-carbon thioester compound (succinyl-CoA). Succinyl-CoA synthetase (*sucC*) couples the cleavage of the high-energy succinyl-CoA into succinate to the synthesis of ATP from ADP + P*i*. In order to complete the cycle, succinate must be converted to oxaloacetate, which is accomplished by the three following reactions: dehydrogenation of succinate to fumarate by succinate dehydrogenase complex (*sdhABC*), one of which is a FAD-containing subunit, meaning FAD is reduced to $FADH_2$ (Sousa *et al.*, 2013); hydration of fumarate to malate by fumarase (*citG*); and finally, the closing reaction

of the cycle, regeneration of oxaloacetate from malate by malate dehydrogenase (*mdh*) and reduction of NAD^+ to NADH (Bartholomae *et al.*, 2014).

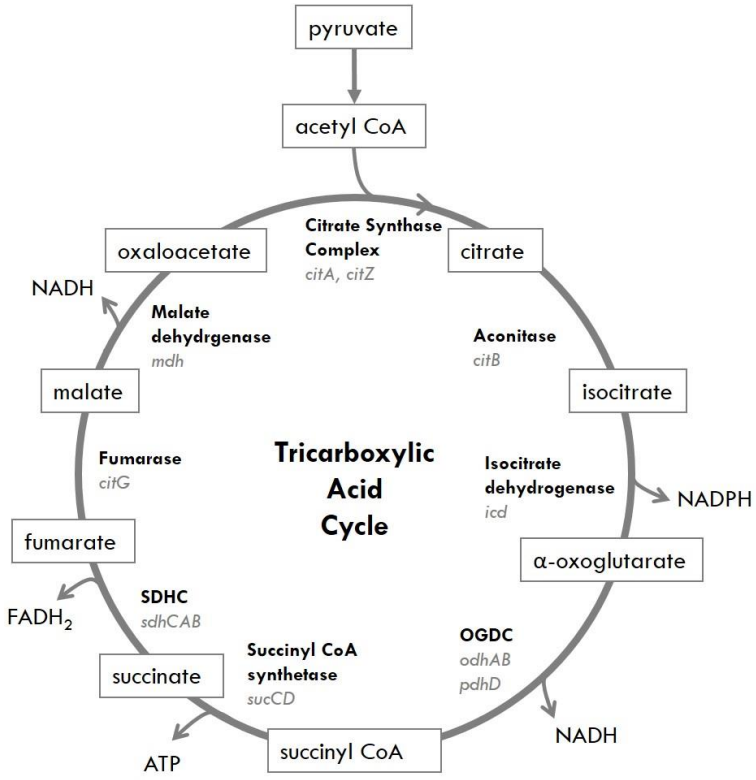


Figure 1.3. Tricaboxylic Acid Cycle. Also known as the Krebs cycle, this pathway yields ATP, reducing power and biosynthetic intermediaries, such as α-oxoglutarate, a precursor for glutamate and derivatives, oxaloacetate, a precursor of aspartate and succinyl-CoA. Relevant enzymes and enzymatic complexes are shown in bold, while genes encoding for those enzymes are italicized in grey. OGDC stands for oxoglutarate dehydrogenase complex and SDHC stands for succinate dehydrogenase complex.

Carbohydrate Uptake and Global Regulation of Carbohydrate Utilization

Sugar uptake is mediated by facilitator molecules, primary and secondary active transporters, and the PTS system (Postma *et al.*, 1993; Saier *et al.*, 2002; Simoni *et al.*, 1967).

Molecules like glucose or fructose, among others, are transported via a PTS system, a protein complex comprising the general proteins enzyme I (EI), the intermediate phosphoryl donor protein HPr and the substrate-specific complex protein enzyme II (EII). This complex translocates and simultaneously phosphorylates the substrate (reviewed in Stülke & Hillen 2000; Deutscher *et. al.* 2002).

Other sugar molecules are transported by ATP-binding cassette (ABC) transport systems. Their organization comprises two hydrophobic transmembrane domains (TMDs) coupled to two cytosolic nucleotide-binding domains (NBDs), or ATP-binding cassettes, responsible for ATP binding and hydrolysis-driven conformational changes for substrate translocation. The majority of NBDs is encoded in close proximity to their partner TMDs, although there are exceptions (Ferreira & Sá-Nogueira, 2010;

Quentin *et al.*, 1999). A functional ABC system associates itself to a solute binding protein (SBP), responsible for substrate recognition and internalization (Higgins, 2001; Quentin *et al.*, 1999). For example, oligosaccharides up to four L-arabinosyl units, such as α-1,5-arabinotetraose, resulting from arabinan degradation, are transported by the ABC-type importer AraNPQ (Ferreira & Sá-Nogueira, 2010).

Some molecules are translocated into the cell via a permease, like arabinose, transported by AraE (Krispin & Allmansberger, 1998a; Sá-Nogueira & Ramos, 1997), myo-inositol transported by IolT and IolF (major and minor transporter) (Morinaga *et al.*, 2010; Yoshida *et al.*, 2002) or gluconate, transported by GntP (Fujita & Fujita, 1989; Reizer *et al.*, 1991). So far, glycerol is the only carbohydrate known to enter the cell via the aquaporin-like carbohydrate facilitator, GlpF (Beijer *et al.*, 1993) (reviewed in Deutscher *et. al.* 2002).

Control of central carbon metabolism in bacteria has several control levels to regulate the use of carbon sources using global regulators. Some of these global regulators are also integrated into a larger regulatory network by which *B. subtilis* coordinates the metabolic flow through important metabolic intersections, namely carbon and nitrogen metabolism, in response to few signaling metabolites (reviewed in Sonenshein 2007). An example of such regulatory network is the carbon catabolite regulation.

B. subtilis thrives in many carbon sources, and as permanent expression of all transporters would consume valuable cellular resources and occupy limited membrane space, cells selectively express transport systems on the basis of extracellular and intracellular signals. The regulatory mechanism by which cells coordinate the metabolism of carbon and energy sources to maximize its efficiency is referred to as carbon catabolite control or regulation (CCR) (reviewed in Fujita 2009). Bacteria in general, and *B. subtilis* in particular, can preferentially use the carbon sources that are most easily accessible and allow fastest growth, which results in diauxic growth – immediate consumption of the most favored carbon source (i.e. glucose, hence the name glucose repression being usually interchanged with carbon catabolite repression) and only after its exhaustion the other carbon sources present in the medium will be used. CCR is defined as a regulatory phenomenon by which the expression genes encoding enzymes necessary for the use of secondary carbon sources are reduced in the presence of a preferred carbon source (Görke & Stülke, 2008; Singh *et al.*, 2008)

Carbon catabolite repression has been extensively studied in model organisms like *E. coli* and *B. subtilis*: although the outcome is similar, the mechanisms underlying CCR are quite different.

In *E. coli*, glucose is transported into the cell via a PTS system. Upon substrate translocation, Enzyme I (EI) is autophosphorylated and the phosphate group of phosphoenolpyruvate is sequentially transferred to the His15 residue in HPr. HPr then donates the phosphoryl group to a histidine residue in

the A domains of the various substrate-specific transporters or Enzyme II (EIIs). Finally, the phosphoryl group is transferred to a residue in the EIIB domain and from there to glucose during its uptake through the membrane domain. Because there is glucose in the medium, the concentration of phosphorylated EIIA decreases and it cannot activate membrane-bound enzyme adenylate cyclase. Low intracellular cyclic adenosine monophosphate (cAMP) concentration does not trigger the formation of a complex between cAMP and cAMP receptor protein (CRP), which is necessary for promoter activation of catabolic genes. RNA polymerase binding to the promoters subjected to CCR and formation of the open transcription complex required for transcription only occurs at the promoters if they are previously activated (reviewed in Deutscher *et. al.* 2006; Görke & Stülke 2008). In the absence of glucose, EII-P stimulates adenylate cyclase activity, increasing [cAMP] in the cell, which bind to CRP, thus activating promoters from catabolic operons responsible for the use of less favored carbon sources.

Carbon catabolite control in *B. subtilis* differs from the one used in *E. coli*, despite having the same outcome. Global regulation of carbon catabolite control occurs when the catabolite control protein A (CcpA) and histidine-containing phosphocarrier protein (HPr) phosphorylated at Ser46 (HPr-Ser46-P) complex bind to sequences known as catabolite responsive elements (*cre*), present in approximately 300 target operons (Fujita 2009 and references there in). *B. subtilis* HPr can be phosphorylated at two sites: the PEP-dependent phosphorylation of a histidine residue at position 15 (HPr-His15-P), which serves sugar translocation purposes, and the ATP-dependent phosphorylation of a serine at position 46 (HPr-Ser46-P), which serves regulatory purposes. Presence of glucose results in HPr-His15-P. A consequent increase in glycolytic metabolites from glucose metabolization, especially FBP, in the cell stimulates the ATP-dependent HPr kinase/phosphatase-catalyzed phosphorylation of HPr at serine 46 (Reizer *et al.*, 1998) (reviewed in Stülke & Hillen 2000; Deutscher *et. al.* 2002; Deutscher *et. al.* 2006; Fujita 2009). HPr-Ser46-P is able to form a complex with CcpA, binding to *cre* sites (Schumacher *et al.*, 2007) (Figure 1.4).In *B. subtilis* roughly 10% of the genome is under the control of CcpA mediated regulation, and the majority of those genes are repressed by CcpA (Blencke *et al.*, 2003; Moreno *et al.*, 2001). If the *cre* sites are located upstream of the -35 region, global regulation is achieved through activation of the promoters by interaction with RNA polymerase, thus originating carbon catabolite activation (CCA). Contrastingly, if the *cre* sites are located in the promoter region, binding of CcpA causes transcription repression, preventing RNA polymerase binding to the promoter, which results in carbon catabolite repression.

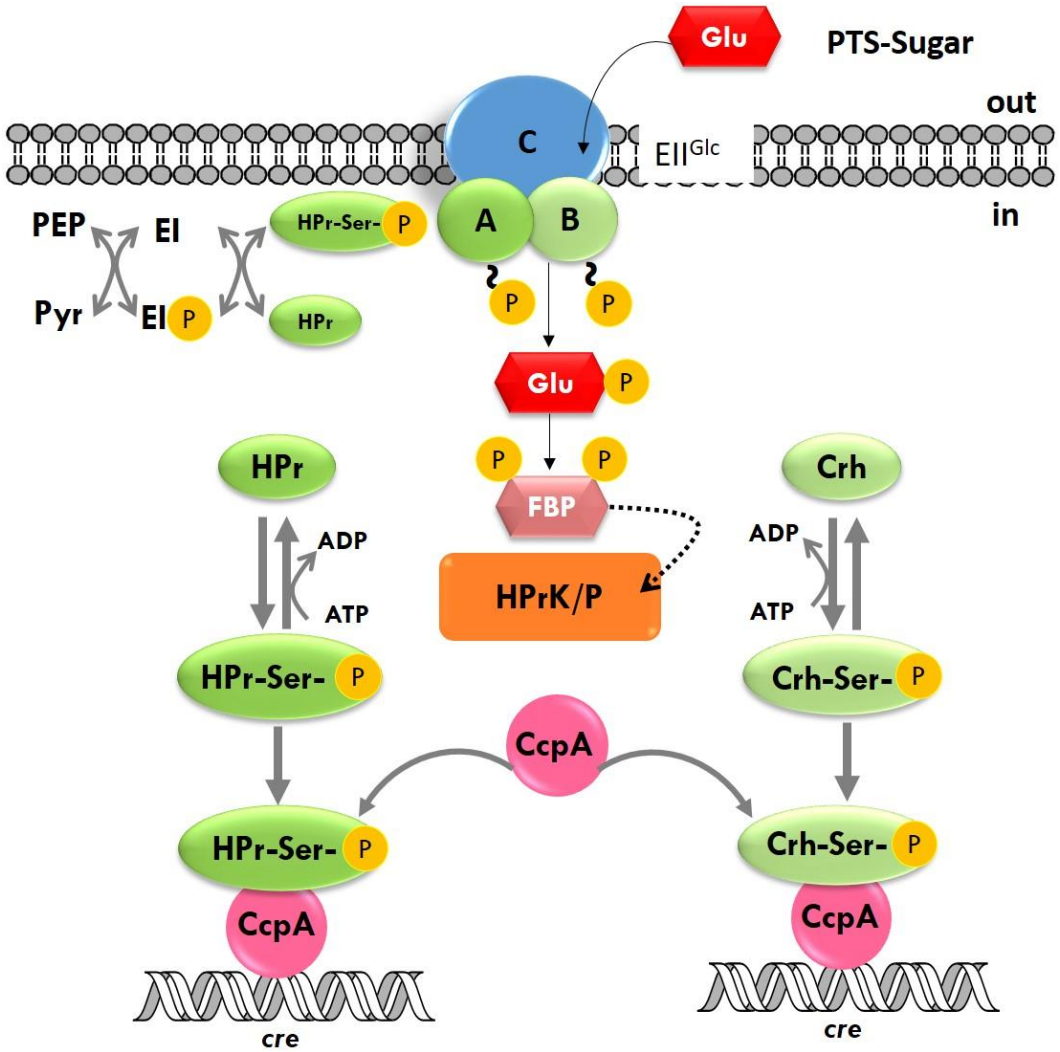


Figure 1.4. Carbon Catabolite Regulation in *Bacillus subtilis*. Uptake of a PTS-sugar (eg. glucose) leads to an increase in the intracellular [FBP], triggering ATP-dependent HPr kinase/phosphatase-catalyzed phosphorylation of HPr and Crh at Ser46. Only the Ser46-P forms of HPr and Crh bind to CcpA. The HPr-Ser46-P /CcpA and Crh-Ser46-P /CcpA complexes can bind to the catabolite responsive elements, *cre*, to cause Carbon Catabolite Repression or Carbon Catabolite Activation, depending on the position of the *cre*. Adapted from Deutscher *et al.* 2002 and Fujita 2009.

Similarly, if the *cre* site is located downstream of the transcription initiation site, CcpA binding causes transcription roadblock, prompting RNA polymerase release from DNA. The presence of two *cre* elements in an operon, synergistically contributing for CCR is not uncommon in *B. subtilis* – one of the elements prevents RNA polymerase binding whilst the other engages in a roadblocking mechanism – and can be found in the *gnt* (Miwa *et al.*, 1997), *iol* (Miwa & Fujita, 2001) and *ara* operons (Inácio *et al.*, 2003).

Global carbon catabolite control is also hierarchically achieved with substrates other than glucose, or non-PTS sugars. In *B. subtilis*, glycerol, fructose and mannitol also cause catabolite repression, as well as sucrose. Sugars like ribose or arabinose also contribute to CCR, although to a very little extent (Singh *et al.*, 2008). This phenomenon is not exclusive to *B. subtilis* and has been described in *Clostridium acetobutylicum* (Aristilde *et al.*, 2015) and in *E. coli* (Bettenbrock *et al.*, 2007).

CCR is also achieved in a CcpA-independent manner: catabolite control can be mediated by CcpB, a homolog protein of CcpA (Chauvaux *et al.*, 1998), CcpC, CcpN, or CggR (reviewed in Fujita, 2009). CcpC is known to be involved in TCA genes regulation, such as aconitase (*citB*) activation and repression in response to citrate levels (Mittal *et al.*, 2013), whereas CcpN has been described as an additional mediator of CCR in *B. subtilis*, acting on genes encoding gluconeogenic enzymes (Tännler *et al.*, 2008). CggR regulates the *gapA* operon (Ludwig *et al.*, 2001) CodY is also involved in global regulation, namely in the synthesis of leucine and isoleucine (BCAAs), histidine and arginine and transporters for amino acids, peptides and sugars, as well as regulation of transcription of carbon-overflow pathways and guanine nucleotide synthesis (Sonenshein 2007 and references therein).

Another mechanism involved in carbon catabolite regulation is inducer exclusion. In *E. coli*, for instance, dephosphorylation of enzyme IIA^{Glc} by G6P causes inhibition of uptake of a number of non-PTS carbon sources, namely lactose, melibiose or glycerol; dephosphorylated EII^{Glc} can bind to a number of enzymes involved in the metabolism of non-PTS carbon sources, inhibiting their utilization by the cell (Hogema *et al.*, 1998a, b). In *B. subtilis* cells metabolizing a PTS sugar, the glycerol kinase, GlpK, is not phosphorylated, causing a slow uptake of glycerol. Insufficient phosphorylation and low GlpK activity towards glycerol was hypothesized as the basis for the CcpA-independent CCR mechanism of inducer exclusion (Darbon *et al.*, 2002).

Additionally, *B. subtilis* and other *Bacilli* encode a paralogue of HPr, catabolite repression HPr protein (Crh), which bears over 40% sequence identity to HPr. Crh lacks the His15 site, but is phosphorylated at Ser46 (Galinier *et al.*, 1997; Landmann *et al.*, 2011; Singh *et al.*, 2008). Crh is known to repress *citM* when *B. subtilis* is grown on succinate and citrate as carbon sources (Warner & Lolkema, 2003) , but it is considered to have a weak contribution to global CCR, mainly due to its much lower levels in the cell and its lower binding affinity for CcpA as compared with HPr(Ser~P) (Görke *et al.*, 2004). In its phosphorylated form, this protein is also known to inhibit glyceraldehyde 3-phosphate dehydrogenase (GapA) activity in vitro (Pompeo *et al.*, 2007), and work by Inácio & De Sá-Nogueira 2007 indicated that Crh might be more important for CCR during the transition to stationary phase. An additional regulatory role for Crh was also proposed, as non-phosphorylated Crh binds to and inhibits activity of the metabolic enzyme methylglyoxal synthase, MgsA, in *B. subtilis*, initiating a glycolytic bypass (Landmann *et al.*, 2011). Despite not having a strong contribution for CCR, Crh seemingly regulates glycolytic flux through interaction with two metabolic enzymes, MgsA and GapA (Landmann *et al.*, 2012; Pompeo *et al.*, 2007).

Arabinose metabolism in *Bacillus subtilis*

In nature *B. subtilis* often encounters a mixture of different and transient carbon sources that can potentially be used. During its permanence either in the gastrointestinal tract of animals or in the soil, contributing to plant biomass degradation, *B. subtilis* encounters complex polymers comprising

lignocellulosic materials, namely hemicellulose and cellulose. Unlike cellulose, hemicellulose is a heteropolymer comprised of a wide variety of sugars and acids, like β-D-xylose, α-L-arabinose, β-D-mannose, β-D-glucose, α-D-galactose, α-D-glucuronic, α-D-4-*O*-methylgalacturonic and α-D-galacturonic acids and, to a lesser extent, α-L-rhamnose and α-L-fucose. Presence and prevalence of these carbon sources in the hemicellulose structure varies with the origin of plant biomass, yielding different types of structures: glucuronoxylans, galactoglucomannans, arabinoglucuronoxylans, xyloglucans and arabinoxylans. L-Arabinose, the second most abundant pentose in nature, is found in significant amounts not only in arabinoxylans and arabinogalactans, but also in homopolysaccharides, branched and debranched arabinans (Gírio *et al.*, 2010; Inácio & Sá-Nogueira, 2008).

The pathway for the utilization of arabinose in *B. subtilis* was described in 1967 (Lepesant & Dedonder, 1967), and studies with mutants unable to use arabinose led to the identification and mapping of three genes encoding arabinose catabolizing enzymes, organized in an operon (Lepesant & Dedonder, 1967; Paveia & Archer, 1992). *araA* encodes an arabinose isomerase, which converts arabinose into L-ribulose, *araB* encodes L-ribulose 5-phosphate kinase, which phosphorylates L-ribulose and, finally *araD* encodes an epimerase that catalyzes the conversion of R5P to X5P. D-xylulose 5-phosphate is then metabolized in the pentose phosphate pathway (Sá-Nogueira *et al.*, 1997) (Figure 1.5).

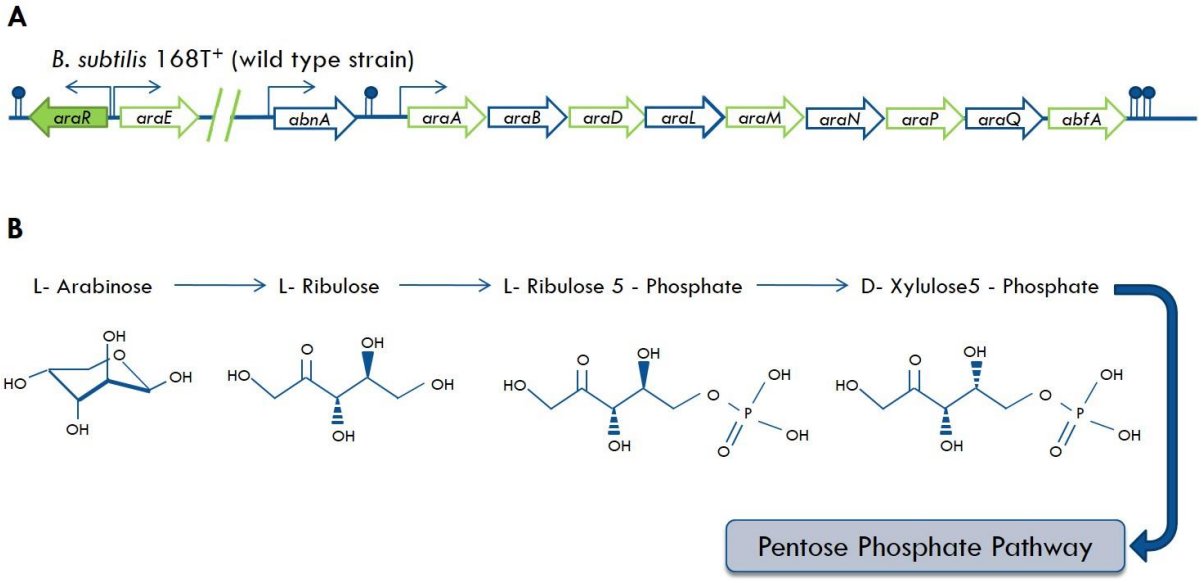


Figure 1.5. Genetic organization of the arabinose operon (A) and pathway for the utilization of arabinose in *Bacillus subtilis* (B). Operon genes, as well as the repressor and permease genes are represented by an arrow. Promoters (⤳) and terminators (⫯) are also represented.

The *ara loci* of *B. subtilis*

The complete genome sequence of *B. subtilis*, together with the arabinose operon sequence was released in 1997 (Kunst *et al.*, 1997; Sá-Nogueira *et al.*, 1997). Work by Sá-Nogueira and colleagues established *araABD* as part of a larger transcriptional unit of about 11kb (*ara* operon) located at about

256º in the *B. subtilis* chromosome, whose expression is directed by one σ^A-type promoter (Para). While *araABD* encode the enzymes required for intracellular conversion of arabinose into D-xylulose 5-phosphate, the remaining six genes are not necessary for arabinose utilization: *araL* encodes a phosphatase, *araM* encodes a glycerol 1-phosphate dehydrogenase, *araNPQ* encode components of an ABC-type transporter involved in the uptake of α-1,5-arabinooligosaccharides and *abfA* encodes an arabinofuranosidase (Ferreira & Sá-Nogueira, 2010; Godinho & Sá-Nogueira, 2011; Guldan *et al.*, 2008; Sá-Nogueira *et al.*, 1997).

In a different location of the *B. subtilis* chromosome, at about 294º, two open reading frames with divergently arranged promoters are found: *araR* encodes the negative regulatory protein of the operon, AraR, and the other, *araE*, encoding the AraE permease (Mota *et al.*, 1999; Sá-Nogueira & Ramos, 1997).

In addition to the arabinose degrading-enzymes, *B. subtilis* also encodes four enzymes devoted to the degradation of arabinan and arabinose-containing polysaccharides, which release arabinose and arabinosyl oligomers from these substrates. The *abfA* gene is part of the metabolic operon *araABDLMNPQ-abfA* and the *abf2* gene is located 23 kb downstream from the operon. Both genes encode α-L-arabinofuranosidases. The *abnA* gene, positioned immediately upstream from the metabolic operon, encodes an endo-α-1,5-arabinanase, an enzyme also encoded by the *abn2* gene, located in a different region of the chromosome. All genes, except for *abn2*, are under the negative regulation exerted by AraR (Inácio & Sá-Nogueira, 2008; Inácio *et al.*, 2008; Raposo *et al.*, 2004).

ara regulon

The arabinose regulon is comprised by thirteen genes under the control of the AraR transcription factor. In the absence of arabinose, AraR binds to palindromic sequences in the promoter regions of such genes, named *ara* boxes (operators - OR). AraR binds cooperatively to two in phase operators in the *araABDLMNPQ-abfA* (OR_{A1} and OR_{A2}), *araE* (OR_{E1}, OR_{E2}) and *abf2* (OR_{X1}, OR_{X2}) promoters, causing the formation of a small loop in the DNA strand, achieving a higher level of repression. Binding to only one operator is observed in self-regulation (*araR* gene - OR_{R3}) and in the promoter of the *abnA* gene (OR_{B1}), allowing a more flexible control of transcription.

The effector molecule arabinose causes a change in AraR conformation, preventing DNA binding. Thus, its mode of action is predominantly the result of protein-DNA interactions, although the cooperative binding involves protein-protein interactions. The AraR DNA contacting domain comprises a helix-turn-helix motif similar to that of the GntR family, while its C terminal domain presents homology to the LacI/GalR family (Correia *et al.*, 2014; Franco *et al.*, 2006, 2007; Mota *et al.*, 1999, 2001; Sá-Nogueira & Mota, 1997).

Negative regulation by AraR was first described by Sá-Nogueira & Mota 1997, as an insertion-deletion mutation in *araR* caused constitutive expression of the *araABDLMNPQ-abfA* operon. Moreover, sensitivity to the presence of arabinose was observed since the addition of this sugar to an early-exponentially growing culture of this mutant resulted in immediate cessation of growth (Sá-Nogueira & Mota, 1997).

Other than negative regulation by AraR, the arabinose regulon genes are also under CCR. Several *cre* are involved in the arabinan and arabinose utilization genes, namely two *cre* sites found in the region of the *araABDLMNPQ-abfA* operon (one located between the promoter region of the catabolic operon and the *araA* gene, and one located 2kb downstream within the *araB* gene), and one *cre* site found in the promoter region of *araE*, *abf2*, *abnA*, and *abn2* (Inácio & Sá-Nogueira, 2008; Inácio *et al.*, 2003; Raposo *et al.*, 2004). A roadblocking mechanism of CcpA was shown to operate in CCR of the operon (Inácio *et al.*, 2003).

Genes of unknown function in the context of the arabinose operon

The function of two genes present in the arabinose operon is still unclear in the context of arabinose utilization. AraM, which constitutes the first identified glycerol 1-phosphate dehydrogenase (G1PDH) from bacteria, has been proposed to generate glycerol 1-phosphate for the synthesis of phosphoglycerolipids in Gram-positive bacterial species (Guldan *et al.*, 2008). AraM catalyzes the NADPH-dependent reduction of DHAP to glycerol 1-phosphate, displaying a similar catalytic efficiency to its archaeal homologues, but its activity is dependent on the presence of Ni^{2+} instead of Zn^{2+}. The glycerol moiety of archaeal lipids is derived from *sn*-glycerol-1-phosphate, which is the enantiomer of *sn*-glycerol-3-phosphate, the precursor for bacterial/eukaryotic glycerolipids. The structural differences between archaeal and bacterial/eukaryotic lipids are believed to cause their distribution between these domains. These unique characteristics define what is called the "Lipid Divide", and no exception to this pattern has been observed to date (Koga, 2014; Yokoi *et al.*, 2012).

AraL was initially annotated as a *p*NPPase with sequence homology to sugar phosphatases members of he HAD superfamily (Sá-Nogueira *et al.*, 1997). Recent work (see Chapter III) has established AraL as a sugar phosphatase belonging to the HAD superfamily, displaying a wide range of substrate specificity (Godinho & Sá-Nogueira, 2011).

Haloacid Dehydrogenase Superfamily

Phosphoryl group transfer is a widely used process for energy transduction, regulation and for cell signaling in all organisms and phosphate transfer mechanisms are often part of strategies used to respond to different external and internal stimuli, having evolved many times in different circumstances. Cellular enzymes responsible for phosphoryl transfer and phosphate-ester hydrolysis evolved in several

superfamilies, developing successful strategies despite their distinct evolution and different structural topologies (Allen & Dunaway-Mariano, 2004), such as P-loop NTPases, the RNase H fold of ATPases the DDH, HD, PHP, calcineurin-like, synaptojanin-like, the Receiver domain and HAD superfamilies (Burroughs *et al.*, 2006).

One of the most successful and widely distributed enzyme superfamilies is the Haloacid Dehalogenase Superfamily (HADSF), originally named after the archetypal enzyme L-2-haloacid dehalogenase. Members of this family are related by their ability to form covalent enzyme-substrate intermediates via a conserved active site aspartic acid (D), which facilitates cleavage of C-Cl, P-C or P-O bonds. Interestingly, this superfamily is dominated by putative phosphatases (~79%) and ATPases (~20%), which are more prevalent, and also includes phosphonatases and phosphomutases, rather than by enzymes like the one it was named after (Allen & Dunaway-Mariano, 2009; Aravind *et al.*, 1998; Burroughs *et al.*, 2006).

HAD superfamily members are structurally characterized by the occurrence of a highly conserved α/β catalytic core domain. This α/β architecture adopts the topology of a typical Rossmanoid fold, a protein structural motif which alternates β strands with α helices (also called a βαβ fold). The scaffold of the core catalytic domain of the HADSF contains a three- layered α/β sandwich comprised of repeating β strand (S1 to S5) and α helix units, adopting the typical topology of a βαβ fold. This fold differs from related Rossmanoid folds in two key structural motifs: a nearly complete α-helical turn, the "squiggle", and a β-hairpin turn, the "flap" (Burroughs *et. al*. 2006; Lu *et. al*. 2008; Allen & Dunaway-Mariano 2009). Both structural motifs not only distinguish a HAD Rossmanoid fold from other Rossman folds, but also play an important role in HADSF catalysis (Figure 1.6).

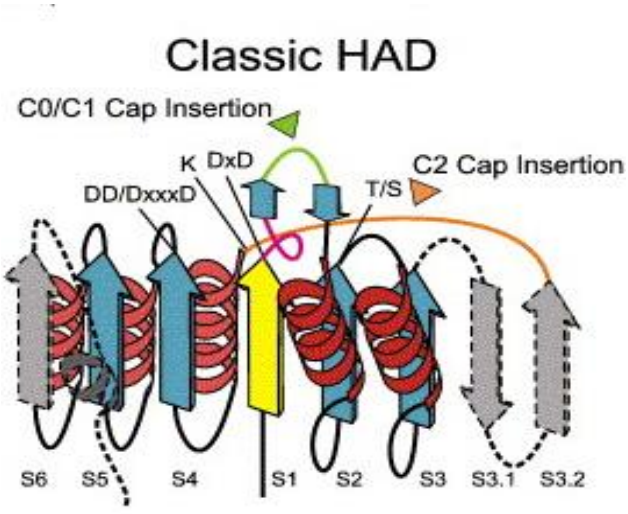


Figure 1.6. Topology diagram of the typical Rossmanoid-like fold from the HAD superfamily. Conserved core strands are in blue, non-conserved elements are depicted in grey. Broken lines indicate secondary structures that may not be present in all family members. The initial strand containing the conserved D residue is rendered in yellow, and C1 and C2 cap insertion points are depicted in green and orange, respectively. The α-helical turn (pink) and the β-hairpin turn (two blue strands projecting from the core of the domain, downstream from the "squiggle") (adapted from Burroughs *et al.* 2006).

In addition to the occurrence of the squiggle and the flap, HADSF members present four highly conserved sequence motifs, the loops constituting the active site (S1 to S4). These motifs constitute a

catalytic scaffold that positions the catalytic residues involved in metallic cofactor (Mg^{2+}) and substrate binding and acid-base and nucleophilic catalysis, using active site aspartate residues in nucleophilic catalysis. S1 to S5 position important residues that are conserved throughout the HADSF and that pinpoint HADSF members in amino-acid sequence alignments. Motif sequence I is located on S1 and it presents the D*x*D signature (where *x* is any amino acid). The two aspartates coordinate the Mg^{2+} cofactor and the first Asp residue is directly responsible for nucleophilic catalysis (Allen & Dunaway-Mariano, 2009). Motif II corresponds to the S2 strand, where a highly conserved serine or threonine is present, binding the substrate phosphoryl group. Motif III comprises a conserved arginine or lysine at the beginning of the helix located upstream of the S4 strand, and strand S4 itself, as well as its conserved acidic residues located on the end of the strand (Asp), account for motif IV. These residues display one of three basic signatures: DD, GD*xxx*D, or GD*xxxx*D, and together with the Asp residues from motif I they help coordinate the Mg^{2+} cofactor (Allen & Dunaway-Mariano, 2004, 2009; Burroughs *et al.*, 2006; Lu *et al.*, 2005). Most HAD phosphohydrolases present a highly conserved α/β core domain with a sequence insert which forms a mobile cap domain (cap). The cap accounts for substrate recognition and simultaneously divides the family into three generic subfamilies – cap insertion between motifs I and II typifies subfamily I (C1), whereas an α/β domain insertion between motifs II and III typifies subfamily II (C2). Subfamily III members do not possess a cap domain (C0) (Burroughs *et al.*, 2006; Peisach *et al.*, 2004; Selengut, 2001). There are two major evolutionary distinct, unrelated types of classical C2 caps (C2a and C2b) which further divide subfamily II in two types A and B. Both present an α/β domain, but while subfamily IIA presents an αβαββαβαβ topology, subfamily IIB displays an αββ(αβαβ)αββ topology (Burroughs *et al.*, 2006; Lu *et al.*, 2005; Peisach *et al.*, 2004).

Cap domains in the HADSF are intrinsically linked to substrate specificity. Both the modularity of the phosphoryl-transfer core domain and the specificity-conferring cap domain are considered to be the structural basis for the adaptation of the scaffold to a wide range of phosphate esters whilst maintaining catalytic efficiency - numerous C1 and C2 members show substrate ambiguity with a broad substrate range, while C0 members generally show a narrow substrate range.

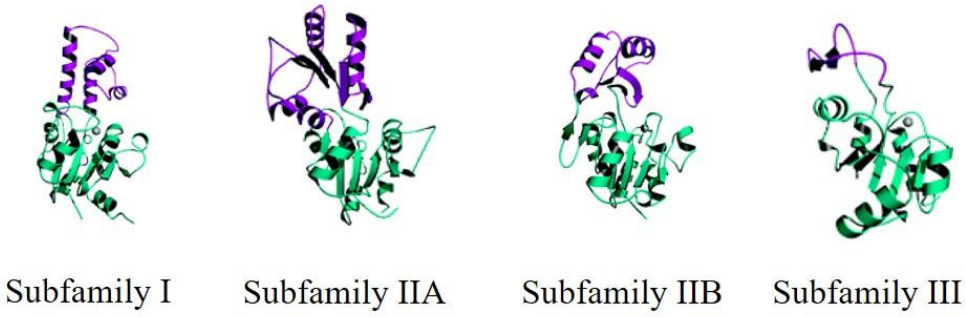


Figure 1.7. Schematic representation of the three HAD superfamily subfamilies. Ribbon diagrams, from left to right: Subfamily I, β-phosphoglucomutase; subfamily IIA, NagD; subfamily IIB, Phosphoglycolatephosphatase and subfamily III, magnesium dependent phosphatase 1. The common domain to all subfamilies is rendered in green, while the different cap domains are colored in purple. Adapted from Lu ***et al.*** **2005**.

This broad range in substrate specificity is consistent with the role HADSF members in diverse physiological functions in primary and secondary metabolism, membrane transport, signal transduction, metabolic homeostasis and nucleic-acid repair HADSF members have the ability to hydrolyze a wide range of phosphorylated metabolites, including carbohydrates, nucleotides, organic acids, coenzymes, and small phosphodonors. HAD family phosphatases play important roles in carbohydrate utilization and metabolic homeostasis. (Allen & Dunaway-Mariano, 2009; Godinho & Sá-Nogueira, 2011; Kuznetsova *et al.*, 2006; Lu *et al.*, 2005; Pandya *et al.*, 2014; Papenfort *et al.*, 2013; Park *et al.*, 2015; Tremblay *et al.*, 2006).

The Sugar Phosphate Toxicity Phenomenon

The sugar phosphate toxicity phenomenon is not new to scientists working with bacterial cells. Various studies starting in the late 1950's have described toxicity associated with sugar phosphates and its accumulation in bacteria, namely in *E. coli* (Englesberg *et al.*, 1962; Kadner *et al.*, 1992; Yarmolinsky *et al.*, 1959), with bacteriostatic (Englesberg *et al.*, 1962) or bactericidal effects (Irani & Maitra, 1977; Yarmolinsky *et al.*, 1959).

Although not new, this phenomenon remains largely uncharacterized, as well as its regulatory circuits or bypass mechanisms that help circumvent it. Most studies are focused on mutant strains that lack specific metabolic enzymes and, as a result, accumulate phosphorylated sugar intermediates upstream of the pathway (Mitchell *et al.*, 1987). At crossroads of metabolic pathways there are intermediary metabolites whose imbalances cause stress that send signals to genes to rectify the stress and achieve homeostasis (Lee *et al.*, 2009).

Sugar Phosphate Toxicity in *Escherichia coli*

Sugar phosphate toxicity phenomenon in *E. coli* has been extensively documented and was first described by researchers working with galactose-sensitive mutants. The authors reported abnormal sensitivity to galactose in *E. coli* strains due to a defective galactose 1-phosphate uridyl transferase (Kurahashi & Wahba, 1958; Yarmolinsky *et al.*, 1959). Authors stated that the presence of a functional galactokinase was a pre-requisite for growth inhibition, and that double mutants in the galactokinase and galactose-1-phosphate-transferease enzymes did not show this growth inhibition. At the time addition of glucose to the medium resulted in an almost resumption of growth, most likely due to what we now know to be catabolite repression exerted by glucose (Yarmolinsky *et al.*, 1959).

Similar studies were later published focusing on toxicity of UDP-glucose (Sundararajan *et al.*, 1962), α-glycerol-phosphate (Cozzarelli *et al.*, 1965) and arabinose (Englesberg *et al.*, 1962). The studies of Englesberg and collaborators in 1962 detected arabinose sensitive mutants, which were deficient in L-ribulose-5-phosphate epimerase, thus accumulating the phosphorylated metabolite

L-ribulose 5-phosphate. Growth of sensitive mutants in medium with L-arabinose originated L-arabinose-resistant mutants which were frequently deficient in both L- ribulose 5-phosphate 4-epimerase and L-ribulokinase (Englesberg *et al.*, 1962).

More recent studies on *E. coli*, however have focused on phosphosugar stress of PTS sugars, namely glucose. In this particular case, it is known that the mechanism involves a small non coding RNA (sRNA), SgrS, which appears to be dedicated to sugar phosphate stress and it is only induced by the accumulation of glucose-6-phosphate (G6P) or the non-metabolizable G6P analog, α-methyl glucoside 6-phosphate (αMG6P), and not by other phosphosugars (Vanderpool & Gottesman, 2004; Vanderpool, 2007). Like glucose, αMG (α-methyl glucoside) is transported and concomitantly phosphorylated by the PTS to αMG6P. Researchers observed that when G6P accumulated in strains lacking glucose 6-phosphate isomerase or in wild-type cells that were exposed to αMG and subsequently accumulate αMG6P intracellularly, the post-transcriptional regulation of *ptsG* (encoding the membrane-bound, glucose specific permease of the system) resulted in a marked destabilization of its mRNA under those stress conditions (Kimata *et al.*, 2001). This post-translational regulation exerted by SgrS directly promotes RNase E-dependent degradation of *ptsG* mRNA during glucose-phosphate stress. As phosphosugar levels rise, SgrS levels increase, full-length *ptsG* mRNA is depleted and RNA degradation products accumulate. SgrS then represses the *ptsG* mRNA at the level of translation, sequestering its ribosome binding site (RBS), helping relieve the phosphoglucose stress (Kawamoto *et al.*, 2006; Vanderpool & Gottesman, 2007). Authors presented evidence that glucose-phosphate stress results from depletion of glycolytic intermediates. Addition of glycolytic compounds like G6P and fructose-6-phosphate rescues the growth defect of an *sgrS* mutant caused by α-MG, making depletion of glycolytic intermediates the metabolic root of glucose-phosphate stress (Richards *et al.*, 2013). Regulatory circuits responsible for the control of disruptive glucose and PTS sugars influx have remained elusive. In recent work, the role of regulatory RNAs, like SgrS has become evident, highlighting a robust regulatory circuit that adjusts sugar influx to cell needs through repression of major sugar importers and upregulation of genes involved in phosphate-removal from phosphosugars. (Papenfort *et al.*, 2013).

Sugar Phosphate Toxicity in *Bacillus subtilis*

Although less documented in literature, sugar phosphate toxicity in *B. subtilis* was reported in the early 1970's. While working with glucose 6-phosphate dehydrogenase and phosphoglucose isomerase mutants of *B. subtilis*, Prasard & Freese in 1974 described cell lysis concomitant with intracellular glucose 1-phosphate accumulation, inhibiting N-acetyl glucosamine-6-phosphate synthesis (Prasad & Freese, 1974).

Subsequent studies also determined that addition of galactose is toxic for *galE*-negative *B. subtilis*, since it results in lethal intracellular concentration of UDP-galactose. This growth arrest

phenotype is reduced in cells lacking a phosphatase (encoded by *galK*) and/or a galactose 1-phosphate uridyltransferase (encoded by *galT*), which allow for the metabolization of galactose in UDP-galactose. Moreover, glucose is also toxic for *galE*-negative strains in the long run (late log phase). Lethal intracellular concentration of UDP-glucose was estimated to be 0.2 µM (Krispin & Allmansberger, 1998b). More recent studies have established a close relationship between the galactose metabolism and exopolysaccharide (EPS) formation (a key component of biofilm), as EPS production has been shown to relieve UDP-galactose toxicity in *galE*-negative strains (Chai *et al.*, 2012).

Sugar phosphate toxicity studies in *B. subtilis* focused on glycolytic intermediates (glucose and its epimer galactose) accumulating in strains with some deficiency in one of the metabolic genes encoding enzymes necessary for the downstream utilization of phosphorylated glycolytic sugars. Sugar phosphate toxicity of non-PTS sugars in *B. subtilis*, was described by Paveia & Archer, 1992 that characterized arabinose sensitive mutants, which were deficient in L-ribulose-5-phosphate epimerase, thus accumulating L- ribulose 5-phosphate (Paveia & Archer, 1992). Later, Sá-Nogueira *et. al*, reported pentose toxicity in an *B. subtilis* strain deficient in the regulator AraR (Sá-Nogueira *et al.*, 1997). Unlike glucose 1-phosphate and UDP-galactose toxicity, which was observed awhile after sugar addition in strains with defects in catabolic genes, arabinose toxicity was observed immediately after sugar addition to a repressor-null mutant in complex medium (Sá-Nogueira *et al.*, 1997).

Chapter II

Role of *araL* and *araM* in the context of the arabinose operon

The author of this Thesis performed all experiments reported in this chapter.

Abstract

The operon *araABDLMNPQ-abfA* of *Bacillus subtilis*, is responsible for arabinan utilization and arabinose catabolism, however the role of two genes, *araL* and *araM*, in this context is still elusive. In a strain deregulated for arabinose utilization, lacking the transcription factor AraR, addition of arabinose to an exponentially growing culture results in immediate cessation of growth. The toxic effect is suppressed by a large deletion of all genes downstream of *araD*. Since neither *araNPQ* nor *abfA* were considered to contribute to the toxic effect, due to its assigned function, the potential candidates are the genes *araL* and *araM*. To investigate their role in the phenomenon of toxicity caused by arabinose, we used two distinct and complementary approaches: in-frame deletions of *araL* and *araM* were constructed and analyzed in different genetic backgrounds, and in addition the effect of an ectopic and controlled expression of *araL* and *araM* was characterized. The results strongly suggest that *araL* and *araM* do not play a role in the toxic effect of arabinose observed in an *araR*-null mutant.

Introduction

Complete bacterial genome sequencing data obtained to date has revealed that close to half of all bacterial ORFs identified have no assigned functions. *Bacillus subtilis*, a model organism for Gram-positive bacteria, including the human pathogens *Bacillus anthracis*, *Listeria monocytogenes, Clostridium difficile, Streptococcus pneumoniae* and *Staphylococcus aureus*, has its genome sequenced since 1997, however about 870 genes, one-fifth of its genes, have unknown function (Arrieta-Ortiz *et al.*, 2015; Kunst *et al.*, 1997).

B. subtilis is able to utilize L-arabinose as sole carbon and energy source. The enzymes necessary to catabolize arabinose, AraA, AraB and AraD, which convert L-arabinose to D-xylulose 5-phosphate are encoded by the first three genes of the transcriptional unit *araABDLMNPQ-abfA* (Sá-Nogueira *et al.*, 1997). All the genes of the operon have an assigned function in the context of arabinose utilization except *araL* and *araM*. However it was shown that all open reading frames downstream from *araD* are not essential for L-arabinose utilization (Sá-Nogueira *et al.*, 1997), although *araNPQ* and *abfA* are involved in arabinosyl oligomers uptake and catabolism (Ferreira & Sá-Nogueira, 2010; Inácio *et al.*, 2008).

The purified product of the *araM* gene displays activity of a glycerol-1-phosphate dehydrogenase Ni^{2+} dependent. The exact function of this enzyme in *B. subtilis* is unknown, as it was the first glycerol-1P dehydrogenase to be identified in Gram-positive organisms (Guldan *et al.*, 2008). AraL is characterized as a phosphatase from the HAD superfamily (see Chater III; Godinho & Sá-Nogueira 2011). Both gene products remain elusive in their biological role in the context of the arabinose operon. The *araL* and *araM* genes are present across several *Bacillus* species - *B. subtilis* (*araL* and *araM*), *B. amyloliquefaciens* (*araL* and *araM*), *B. licheniformis* (*araM*), *B. halodurans* (*araM*), as well as in

Geobacillus stearothermophilus, where *araL* and *araM* are part of gene cluster involved in arabinan degradation. In the later, it is speculated that *araL* and *araM* might play a role in an alternative arabinose-utilization pathway (Shulami *et al.*, 2011), similar to that described first in *Azospirillum brasiliensis* (Watanabe *et al.*, 2006a, b). *A. brasiliensis* was the first bacterium in which an alternative pathway for L-arabinose degradation was described: arabinose is initially oxidized to L-arabino-ϒ-lactone, by a $NAD(P)^+$-dependent dehydrogenase, and is subsequently metabolized to α-ketoglutarate, which enters the TCA (Watanabe *et al.*, 2006b). Similarly, in Archaea, examples of L-arabinose pathways involving dehydrogenases have also been described (reviewed in Bräsen *et. al.* 2014), such as an oxidative pathway involving L-arabinose metabolization to α-ketoglutarate in *Haloferax volcanii* (Johnsen *et al.*, 2013). Fungal pathways for arabinose utilization are based on an unique oxidoreductive pathway, which converts L-arabinose via L-arabinitol (L-arabinose reductase), L-xylulose (L-arabinitol dehydrogenase) and xylitol (L-xylulose reductase) to D-xylulose (xylitol dehydrogenase) and finally to D-xylulose 5-phosphate (xylulokinase) (Seiboth & Metz, 2011).

Previous work by Sá-Nogueira and collaborators documented that addition of arabinose to an early-exponentially growing culture of an *araR*-null mutant strain resulted in immediate cessation of growth in a complex medium. The authors speculated that this effect could be due an increased intracellular level of arabinose which consequently caused an increase in the concentration of the toxic intermediates to the cell (Sá-Nogueira & Mota, 1997). Interestingly, the authors showed that this toxic effect of arabinose was suppressed by a deletion of all genes downstream of *araD* (Sá-Nogueira *et al.*, 1997). Since neither *araNPQ* nor *abfA* were considered to contribute to the toxic effect, due to nature of their function (transmembrane protein components and arabinofuranosidase, respectively), the potential candidates for this effect are the products of the genes *araL* and *araM*.

To ascertain the role of the *araL* and a*raM* gene products in the toxic effect of arabinose observed in the *araR*-null mutant, we generated in-frame deletions of *araL* and *araLM* in the *B. subtilis* chromosome. The effect of these deletions was analyzed in different genetic backgrounds. In addition, we expressed the *araL, araM* and *araLM* ectopically by placing the genes under the control of an inducible promoter derived from Pspac. The results obtained indicate that *araL* and *araM* do not contribute to the toxic effect of arabinose in *B. subtilis araR*-null mutant strains.

Materials and Methods

Substrates. All substrates were purchased from Sigma-Aldrich Co, St. Louis, MO, USA**.**

Bacterial strains and growth conditions. Strains used in this work are listed in Table 3.1. *B. subtilis* strains were grown on LB medium (Miller, 1972), SP medium (Anagnostopoulos & Spizizen, 1960) or minimal C medium supplemented with casein hydrolysate 1% (w/v) (Pascal *et al.*, 1971).

Chloramphenicol (5 μg.mL^{-1}), kanamycin (25 μg.mL^{-1}), spectinomycin (50 μg.mL^{-1}) or erythromycin (1 μg.mL^{-1}), L-arabinose (0.4% (w/v)), IPTG (1 mM) and X-Gal (40 μg.mL^{-1}) were added as appropriate. Solid medium was made with LB, or Sugar Free Agar Medium (LabM) containing 1.6% (w/v) agar (LabM). The *amyE* phenotype was tested in plates of solid LB medium containing 1% (w/v) potato starch; after overnight incubation, plates were flooded with a solution of 0.5% (w/v) I_2-5.0% (w/v) KI for detection of starch hydrolysis. *Escherichia coli* strain DH5α (Gibco BRL) was used for routine molecular cloning work. Ampicillin (100 μg.mL^{-1}), chloramphenicol (25 μg.mL^{-1}), kanamycin (30 μg.mL^{-1}), tetracycline (12 μg.mL^{-1}) and IPTG (100mM) were added as appropriate.

Growth kinetic parameters of the wild-type and mutant *B. subtilis* strains were determined in C medium (Pascal *et al.*, 1971) supplemented with 1% (w/v) casein hydrolysate (Sá-Nogueira *et al.*, 1997). Arabinose 0.4% (w/v) and IPTG (100mM) were added when appropriate. Cultures were grown on an Aquatron® Waterbath rotary shaker (Infors HT, Bottmingen, Switzerland), at 37 °C (unless stated otherwise) and 180 rpm, and OD_{600nm} was monitored in an Ultrospec™ 2100pro UV⁄ Visible Spectrophotometer (GE Healthcare Life Sciences, Uppsala, Sweden).

DNA manipulation, PCR amplification, and sequencing. DNA manipulations were carried out as described by (Joseph Sambrook & David W. Russel, 2001). Restriction enzymes were purchased from MBI Fermentas or New England Biolabs and were used according to the manufacturer's instructions. DNA ligations were performed using T4 DNA Ligase (MBI Fermentas). DNA was eluted from agarose gels with GFX Gel Band Purification kit (GE Healthcare), and plasmids were purified using the QIAGEN® Plasmid Midi kit (Qiagen) or QIAprep® Spin Miniprep kit (Qiagen). DNA sequencing was performed with ABI PRIS BigDye Terminator Ready Reaction Cycle Sequencing kit (Applied Biosystems). PCR amplifications were done using high-fidelity *Phusion*® DNA polymerase from Finnzymes.

Plasmid construction. Plasmid pLG1 was constructed by subcloning a 2240 bp NsiI-EcoRI DNA fragment from pDR111 (gift from David Rudner and Federico Gueiros Filho, Harvard University) between the NsiI and EcoRI sites of pPL82 (Quisel *et al.*, 2001). pLG2 was obtained by amplification of the *araL* rbs and coding sequence, from the wild-type strain *B. subtilis*168T$^+$, with oligonucleotides ARA253 and ARA424 (Table 3.2), which add unique restriction sites XbaI and SphI, and cloning this fragment between the NheI and SphI sites of pLG1. To construct pLG3, rbs and coding sequences of *araL* and *araM* were amplified with the ARA253 and ARA425 pair of primers, and then digested with HindIII, to yield only the *araM* rbs and coding sequence. This fragment was then cloned in the HindIII and SphI sites of pLG1. pLG4 was obtained with the insertion of the *araLM* sequence amplified with the primers ARA253 and ARA425, in the NheI and SphI sites of pLG1. To obtain pLG7, the arabinose promoter sequence from pLM32 (Mota *et al.*, 1999) was amplified using the primers ARA28 and ARA451, introducing EcoRI and HindIII sites, respectively, and it was subsequently cloned in the EcoRI

and HindIII sites of pLG1. pLG8 is a pLG7 derivative constructed by amplification or the *araLM* sequence with the pair of primers ARA253 and ARA452, which carry XbaI and BamHI sites, respectively, and inserting this fragment in pLG7 was digested with NheI and BamHI after removing a 1362bp that contained the *lacI* gene (schematic representation of transcription fusion in Figure 2.1).

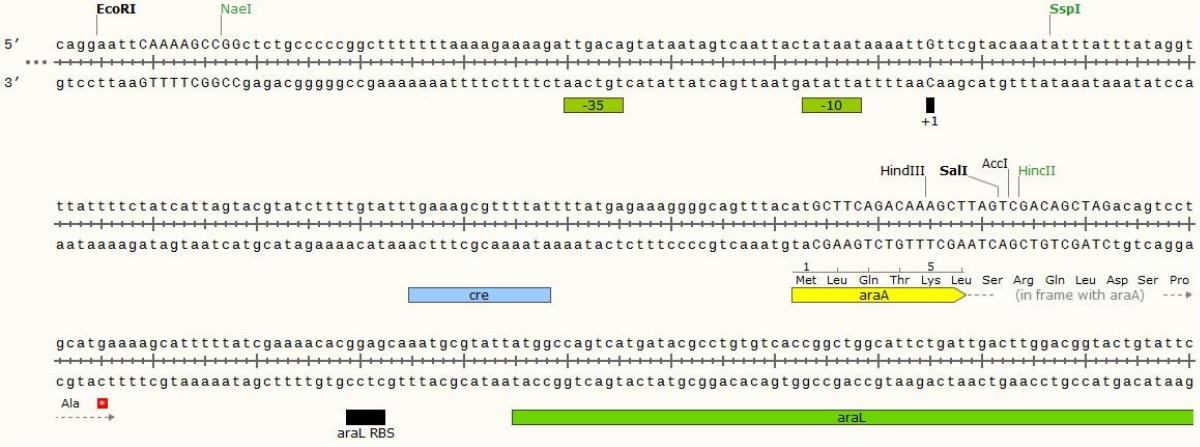


Figure 2.1. Schematic representation of transcription fusion in pLG8. Plasmid pLG8 harbors a transcriptional fusion of the arabinose promoter with the *araL* and *araM* genes. Promoter elements -10 and -35 regions are represented ( ), as well as the +1 site, *cre* and beginning of *araA* coding sequence ( ), *araL* rbs ( ) and beginning of the *araL* coding sequence ( ). Some restriction enzymes are indicated.

In-frame deletion of *araL* and *araLM* using pMAD. The pMAD vector (Arnaud *et al.*, 2004) was used to successfully delete the *araL* and *araLM* genes from the *B. subtilis* chromosome. To delete the *araL* gene, DNA fragments corresponding to the upstream region of *araL* (extending from positions -557 to +84 relatively to the *araL* start codon site) and downstream region of *araL* (extending from positions -72 and +586 relatively to the stop codon of *araL*) were amplified using two pairs of oligonucleotides, ARA444 and ARA458 (upstream) and ARA459 and ARA460 (downstream). The resulting PCR DNA fragments were purified by agarose electrophoresis gel, and equal amounts of both fragments were used as template for a new polymerase chain reaction, using the oligonucleotide pair ARA444 and ARA460. The resulting 1262bp fragment was purified, restricted with the enzymes BamHI and EcoRI, and cloned in the same sites of the pMAD vector. The recombinant pMADΔ*araL* vector was named pLG9. To delete both the *araL* and *araM* genes, DNA fragments corresponding to the upstream region of *araL* (extending from positions -557 to +84 relatively to the *araL* start codon site) and downstream region of *araM* (extending from positions -18 and +597 relatively to the stop codon of *araM*) were amplified using two pairs of oligonucleotides, ARA444 and ARA445 (upstream) and ARA446 and ARA447 (downstream). The PCR DNA fragments were purified on an agarose electrophoresis gel, and equal amounts of both fragments were used as template for a new polymerase chain reaction, using the oligonucleotide pair ARA444 and ARA447. The resulting 1275bp fragment was purified on an agarose electrophoresis gel, and was then restricted with the enzymes BamHI and

EcoRI and cloned in the same sites of the pMAD vector. The recombinant pMADΔ*araLM* vector was named pLG10.

Gene replacement in *B. subtilis*. *B. subtilis* strains 168T$^+$ and IQB215 were transformed with pLG9 and pLG10 derivatives of the pMAD vector (Figure 2.2). After addition of plasmid DNA to the cell culture, growth temperature was decreased from 37 ºC to 30 ºC, and transformation was carried out for 2 hours, at 180 rpm. 100 µL of cell suspension were plated on erythromycin plates, supplemented with X-Gal (80 µg.mL^{-1}), and incubated at 30 ºC for 48 hours. One blue colony was selected and inoculated in 20mL of LB medium supplemented with erythromycin for 1 hour at 30 ºC, 150 rpm. After 1 hour the temperature was gradually increased to 42 ºC, and incubation proceeded during 6 hours. Dilutions 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4} were made and 100 µL of each dilution was plated on solid LB medium supplemented with erythromycin and X-Gal (50 µg.mL^{-1}), and incubated at 42 ºC overnight. The integration of the plasmid into the chromosome by a single recombination event is achieved during growth at a non-permissive temperature (42 ºC) – co-integrate mutants. At a lower and permissive temperature (30 ºC) another single recombination event takes place, which allows the plasmid to be excised from the chromosome, deleting the target gene in the process The next day, 10 blue colonies were isolated and plated in the same conditions overnight. From these 10, 4 clones were selected and inoculated together in 20 mL of LB medium, without antibiotics. Growth was carried for 8 to 9 hours, at 30 ºC, 150 rpm, and then the culture was diluted 1:1000 in the same medium and incubated overnight at 37 ºC, 150 rpm. This procedure was repeated for two days, resulting in an overall of 6 cultures. The last culture overnight culture incubated at 37 ºC was diluted 10^{-5}, 10^{-6} and 10^{-7}, and 100 µL of each dilution plated in duplicate in plates supplemented with X-Gal (50 µg.mL^{-1}), and incubated overnight at 37 ºC. Ten white colonies were then re-isolated on LB solid medium plates supplemented with X-Gal (50 µg.mL^{-1}) and the appropriate antibiotics and incubated at 37 ºC overnight.

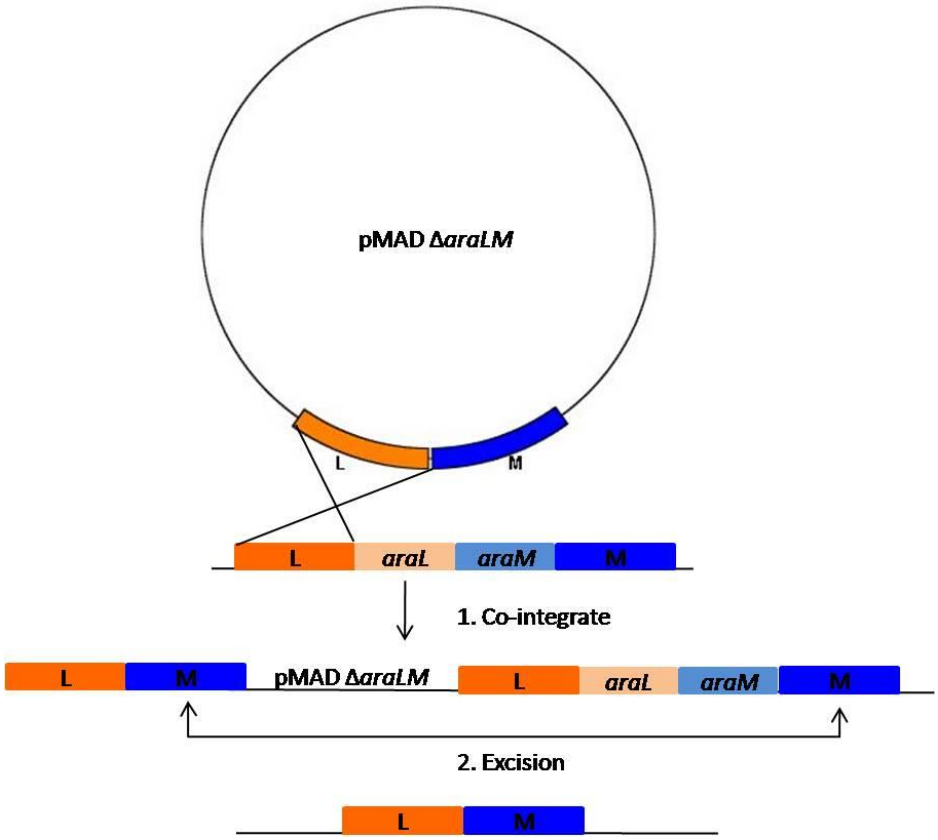


Figure 2.2. Schematic representation the two-step procedure used to obtain gene replacement recombination with pMAD. Areas labeled L and M represent DNA sequences located upstream and downstream from *araL* and *araM* genes. The crossed lines indicate crossover events. The integration of pMAD via homologous sequences can take place in area L or M. The co-integrate undergoes a second recombination event, regenerating the pMAD plasmid. Gene replacement occurs only if the second recombination event occurs in area M, as shown (adapted from Arnaud *et al.* 2004).

Extraction of *B. subtilis* chromosomal DNA. Clones re-isolated at the end of the gene replacement protocol were grown overnight in 3mL of LB medium, supplemented with the appropriate antibiotics, at 37 ºC and 150 rpm. A 2.2 mL volume of culture was subjected to centrifugation at 11 000 *g* for 5 minutes, and the supernatants were discarded. Cell pellets were resuspended in 2.2 mL of 50 mM Tris, 5 mM EDTA, pH 8, and centrifuged again at 11 000 *g* for 5 minutes. The resulting pellets were then resuspended in 175 µL of 50 mM Tris, 5mM EDTA, RNaseA 20 µg.mL^{-1}, lysozyme 1 mg/mL, pH 8, and incubated at 37 ºC for 30 minutes. The tubes were vigorously vortexed for 5 minutes and incubated again at 37 ºC for 15 minutes. Centrifugation was then carried out at 11 000 *g* for 10 minutes, and the supernatants were recovered to a new tube. 1 µL from each supernatant containing the chromosomal DNA was then used as template in a PCR reaction, to confirm the deletions. Oligonucleotide pairs ARA472 and ARA478 were used to confirm Δ*araL* and ARA472 and ARA473 were used to confirm Δ*araLM*.

Table 2.1. List of plasmids used in this study

Plasmids	Relevant construction	Source or reference
pPL82	contains the Pspac(hy) promoter	Quisel *et. al.* 2001
pDR111	derivative of the Pspac(hy) plasmid pJQ43; contains an additional *lacO* binding site	David Rudner
pLM32	plasmid containing the arabinose operon promoter region	Mota *et. al.* 1999
pLG1	contains the Pspank(hy) promoter and *lacI* gene from pPL82 and *cat* gene and *ori* from pDR111	This work
pLG2	pLG1 derivative, with *araL* under the control of Pspank(hy)	This work
pLG3	pLG1 derivative, with *araM* under the control of Pspank(hy)	This work
pLG4	pLG1 derivative, with *araLM* under the control of Pspank(hy)	This work
pLG8	pLG1 derivative, with *araLM* under the control of Para	This work

Table 2.2. List of oligonucleotides (Primers) used in this study

Primer	Sequence (5' → 3') (Restriction sites are underlined)
ARA28	CCTATTGAATTCAAAAGCCGG
ARA253	TAACCCCAATCTAGACAGTCC
ARA424	GATACGATGCATGCCAGAATCC
ARA425	ATTCGAGGTGCATGCTATTC
ARA431	TACGGATCCAAGCTTCAAGAGCATCAGCTG
ARA444	CGGGATCCACCGTGAAAAAGAAAGAATTGTC
ARA445	GTTGATCAGCGTTTTGTTTTCGTCCAAGTCAATCAGAATGCCAGCCGGTGACAC
ARA446	GTGTCACCGGCTGGCATTCTGATTGACTTGGACGAAAACAAAACGCTGATCAAC
ARA447	CTTCCCGAATTCTTTGCGCACTTTCTGTCC
ARA451	GAATTCATAAAGAAGCTTTGTCTGAAGC
ARA452	TTGGATCCGCGGGCTATTCATATAG
ARA458	CTCAGCCAATTTGGTTACATCCTTGTCCAAGTCAATCAGAATGCCAGCCGGTGCCAC
ARA459	GTGTCACCGGCTGGCATTCTGATTGACTTGGACAAGGATGTAACCAAATTGGCTGAG
ARA460	CGTGAATTCACCGAGCATGTCACCAAAGCC
ARA472	CCAACCTGAAGCTTCAAGAG
ARA473	GTTGCGGAATCATTTCTTTCC
ARA478	TGAACGATCTTAGCTCCTGC

Table 2.3. **List of strains used in this study**. Arrows indicate transformation and point from donor DNA to recipient strain.

Strain	Relevant genotype	Sources or Reference
E. coli strains		
DH5α	*fhuA*2 Δ(*argF-lacZ*) U169 *phoA gln*V44 Φ80 Δ(l*acZ*)M15 *gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	Gibco BRL
B. subtilis strains		
168T$^+$	Prototroph (wild-type)	F. E. Young
IQB215	Δ*araR*::*km*	Sá-Nogueira & Mota 1997
IQB206	Δ*araL-abfA*::*spc*	Sá-Nogueira *et. al.* 1997
IQB565	Δ*araL-abfA*::*spc* Δ*araR*::*km*	pLM8 → IQB206
IQB816	Δ*araL-abfA*::*spc* Δ*araR*::*km amyE*::Pspank(hy)-*araL*	pLG3 → IQB565
IQB817	Δ*araL-abfA*::*spc* Δ*araR*::*km amyE*::Pspank(hy)-*araM*	pLG2 → IQB565
IQB818	Δ*araL-abfA*::*spc* Δ*araR*::*km amyE*::Pspank(hy)-*araLM*	pLG4 → IQB565
IQB819	*amyE*::Pspank(hy)-*araL*	pLG2 → 168T$^+$
IQB820	*amyE*::Pspank(hy)-*araM*	pLG3 → 168T$^+$
IQB821	*amyE*::Pspank(hy)-*araLM*	pLG4 → 168T$^+$
IQB822	Δ*araR*::*km amyE*::Pspank(hy)-*araL*	pLG2 → IQB215
IQB823	Δ*araR*::*km amyE*::Pspank(hy)-*araM*	pLG3 → IQB215
IQB824	Δ*araR*::km *amyE*::Pspank(hy)-*araLM*	pLG4 → IQB215
IQB825	Δ*araL-abfA*::*spc*Δ*araR*::*km amyE*::Pspank(hy)	pLG1 → IQB565
IQB827	Δ*araL-abfA*::*spc* Δ*araR*::*km amyE*:: Para-*araLM*	pLG8 → IQB565
IQB829	Δ*araR*::*km*Δ*araLM*	pLG9 → IQB215
IQB830	Δ*araLM*	pLG9 → 168T$^+$
IQB831	Δ*araR*::*km*Δ*araL*	pLG10 → IQB215
IQB832	Δ*araL*	pLG10 → 168T$^+$

Results and Discussion

In-frame deletions of *araL* and *araLM*

The addition of arabinose to the *araR*-null mutant IQB215 results in growth arrest (Sá-Nogueira & Mota 1997), most probably due to the over-expression of the *araABDLMNPQ-abfA* operon. However, a deletion of all the genes downstream from *araD* in an *araR* null mutant, suppresses the toxic effect — strain IQB565 (Sá-Nogueira *et al.*, 1997). Since the deletion of genes *araL*, *araM*, *araN*, *araP*, *araQ* and *abfA* reverts the toxicity, it is reasonable to believe that it was the over-expression of one or more

than one of these genes may cause the toxic effect. The two more likely candidates are *araL* and *araM* encoding a phosphatase and a dehydrogenase, respectively. Due to their location in the middle of the transcriptional unit, we generated in-frame deletion to minimize the polar effect on the downstream genes. The deletion mutations were generated using the pMAD vector (Arnaud *et al.*, 2004), and followed a three-week protocol described in Materials and Methods. In the last step, ten clones were selected to confirm the in frame deletion, from which 50% were expected to bear the Δ*araL* or Δ*araLM*. The chromosomal DNA was extracted and the deletion was confirmed by PCR, using primers that hybridize outside of the deleted region, and by sequencing of the deletion regions (Figure 2.3). Strains IQB215 (*araR*::*km*) and 168T$^+$ were transformed with the pMAD derivatives containing the deletions Δ*araL* and Δ*araLM* (pLG9 and pLG10, respectively). The resulting strains are schematically represented in Figure 2.3. Growth kinetics parameters were determined for the different strains and are summarized in Table 2.4.

Table 2.4. Effect of distinct mutations in the growth kinetics of *B. subtilis* strains. Doubling times (minutes) for different strains in liquid minimal medium (C) supplemented with casein hydrolysate in the presence or absence of arabinose. Results are the averages of three independent assays and their respective standard deviations.

	Doubling Time (minutes)	
Strain (Relevant genotype)	**No sugar**	**arabinose 0.4%**
168T$^+$ (wild-type)	46.79±2.14	42.84±1.81
IQB215 (Δ*araR*::*km*)	46.44±3.93	No growth
IQB829 (Δ*araR*::*km* Δ*araLM*)	60.03±3.11	No growth
IQB830 (Δ*araLM*)	47.25±5.31	45.94±3.14
IQB831 (Δ*araR*::*km* Δ*araL*)	52.60±9.12	No growth
IQB832 (Δ*araL*)	46.44±2.93	43.63±3.82

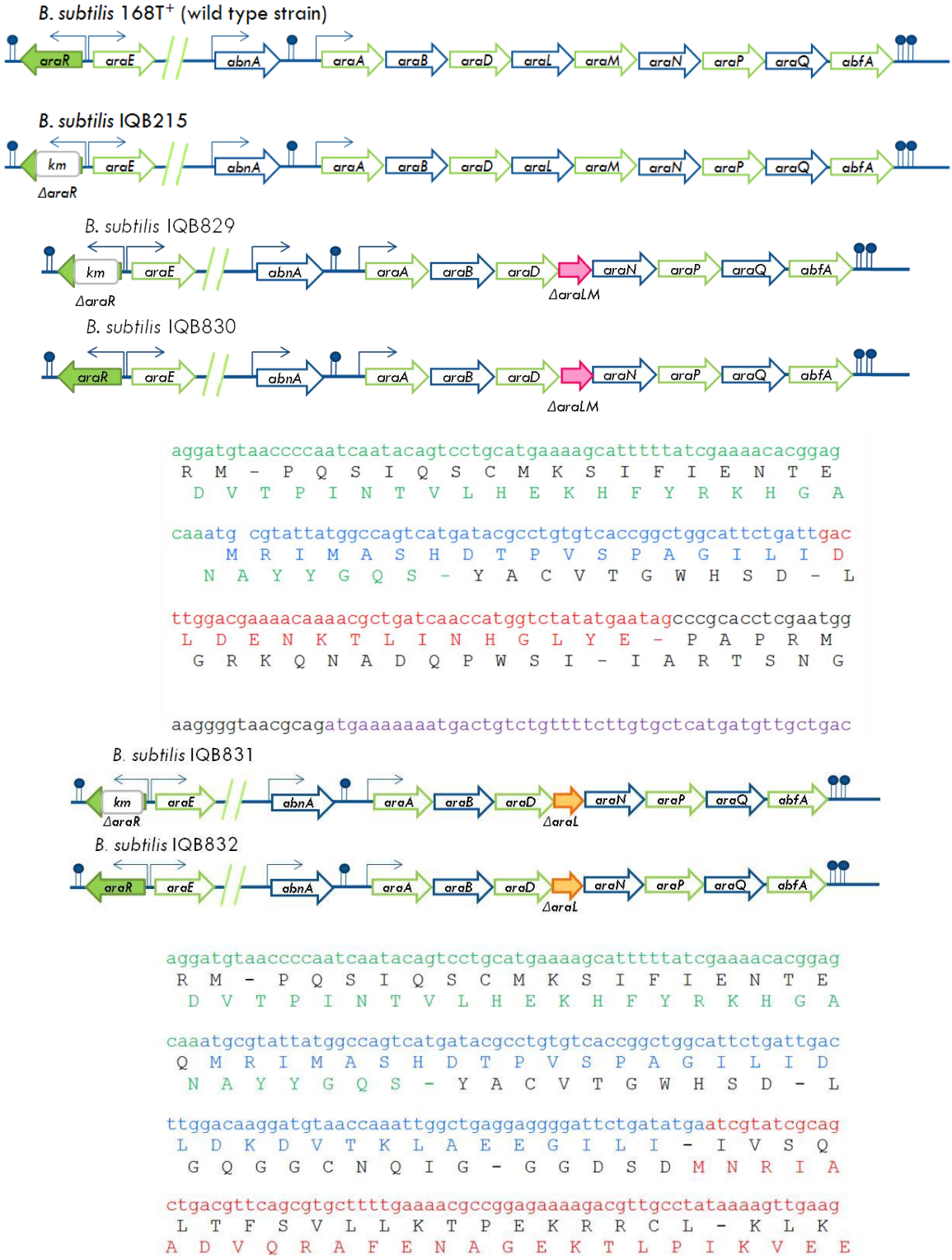


Figure 2.3. Schematic representation of the relevant genotype in *B. subtilis* strains 168T$^+$, IQB215 and IQB829, 830, 831 and 832 genotypes. The sequence of the deleted region originated by integration of the pMAD derivatives is shown. Nucleotide and amino acid sequence are represented according to the following color code: green (*araD* – AraD), blue (*araL* – AraL), red (*araM* – AraM) and purple (*araN* - AraN).

The results indicate that in frame deletion of *araL* and *araLM* in strains IQB831 (Δ*araR*::*km* Δ*araL*), and IQB829 (Δ*araR*::*km* Δ*araLM*) does not suppress the toxic effect of arabinose caused by absence of AraR, as observed with IQB565 (Δ*araR*::*km* Δ*araL-abfA*::*spc*) when compared to IQB215 (Δ*araR*::*km*). These observations indicate that the products of *araL* and *araM* do not play a role in the toxic phenomenon.

Ectopic expression of *araL* and *araLM* under the control of an inducible promoter

In parallel, to elucidate the role of *araL* and *araM* in the toxicity of arabinose, we constructed strains for ectopic expression of these genes. The genes were placed at the *amyE* locus under the control of an inducible promoter. If AraL, and/or AraM participate in the toxicity, one would expect that its ectopic expression in strain IQB565 (Δ*araR*::*km* Δ*araL-abfA*::*spc*) would re-establish growth arrest.

In previous experiments (Sá-Nogueira, unpublished results) *araL* and *araM* were placed under the control of the IPTG inducible Pspac promoter (Yansura & Henner, 1984), and integrated in the *amyE* locus of the *B. subtilis* chromosome of strain IQB565 (Δ*araR*::*km* Δ*araL-abfA*::*spc*), however, after IPTG induction in the presence of arabinose, the toxic effect was not re-established. Since the arabinose operon promoter is very strong, a reportedly stronger promoter was assayed. Pspac(hy) is a Pspac derivative carrying a point mutation, which raised the expression level of Pspac 10 to 20 fold, without affecting the fold induction by IPTG (Quisel *et al.*, 2001)**.** The Pspac promoter is known to be leaky (Jana *et al.*, 2000; Lindow *et al.*, 2002) thus, further improvement of Pspac(hy) was accomplished by inserting an additional *lac*O binding site, which allows better repression; the resulting promoter was named Pspank(hy) (D. Rudner and F. G. Filho, Harvard University).

For ectopic expression of *araL*, *araM*, and *araLM*, under the control of the Pspank(hy) promoter, plasmids pLG2, pLG3 and pLG4 were constructed and transformed into *B. subtilis* strains IQB565 (Δ*araR*::*km* Δ*araL-abfA*::*spc*) and into the wild-type strain. In addition a plasmid, pLG8, was constructed for ectopic expression of *araLM* under the control of the arabinose promoter, Para, which is a very strong promoter. Growth kinetics parameters of the *B. subtilis* strains were determined and the results are summarized in Table 2.5.

Table 2.5. Growth kinetics of distinct *B. subtilis* strains harboring ectopic expression of *araL, araM,* and *araLM*. Doubling times (minutes) for different strains in liquid minimal medium (C) supplemented with casein hydrolysate and IPTG in the presence and absence of arabinose. Results are the averages of three independent assays and their respective standard deviations.

	Doubling Time (min)	
Strain (Relevant Genotype)	**No sugar**	**Arabinose 0.4%**
168T$^+$ (prototroph)	46.79±2.14	42.84±1.81
IQB215 (Δ*araR*::*km*)	46.44±3.93	No growth
IQB565 (Δ*araR*::*km* Δ*araL-abfA*::*spc*)	50.08±3.66	56.52±2.65
IQB816 (Δ*araR*::*km* Δ*araL-abfA*::*spc amyE*::Pspank(hy)-*araL)*	56.22±3.97	65.06±2.75
IQB817 (Δ*araR*::*km* Δ*araL-abfA*::*spc amyE*::Pspank(hy)-*araM)*	51.85±5.33	66.27±5.06
IQB818 (Δ*araR*::*km* Δ*araL-abfA*::*spc amyE*::Pspank(hy)-*araLM*)	48.68±6.16	66.37±3.79
IQB819 (*amyE*::Pspank(hy)-*araL*)	49.31±1.59	43.25±1.21
IQB820 (*amyE*::Pspank(hy)-*araM*)	59.80±6.01	51.49±4.67
IQB821 (*amyE*::Pspank(hy)-*araLM*)	51.69±2.56	43.61±0.96
IQB825 (Δ*araR*::*km* Δ*araL-abfA*::*spc amyE*::Pspank(hy)	67.08±6.38	56.88±2.36
IQB827 (Δ*araR*::*km* Δ*araL-abfA*::*spc amyE*:: Para-*araLM*)	52.80±4.63	65.21±1.54

Analysis of growth kinetics of strains IQB816 (Δ*araR*::*km* Δ*araL-abfA*::*spc amyE*::Pspank(hy)-*araL*), IQB817 (Δ*araR*::*km* Δ*araL-abfA*::*spc amyE*::Pspank(hy)-*araM*), IQB18 (Δ*araR*::*km* Δ*araL-abfA*::*spc amyE*::Pspank(hy)-*araLM*), and IQB827 (Δ*araR*::*km* Δ*araL-abfA*::*spc amyE*:: Para-*araLM*) indicate that ectopic expression of *araL, araM* or *araLM* under the control of either the Spank(hy) promoter or the arabinose promoter fails to re-establish the growth arrest observed in strain IQB215 (Δ*araR*::*km*). These observations corroborate the results obtained in the characterization of the *in frame* deletion mutants, which concluded that the toxic effect of arabinose in strain IQB215 is not due to the deregulated expression of *araL* and/or *araM* genes.

By two complementary approaches we showed that the *araL* and *araM* are not involved in this phenomenon of toxicity. Firstly, in frame deletion mutations in both *araL and araM* genes were constructed in a *araR*-null mutant and failed to suppress the toxic effect of arabinose (Table 2.4). Secondly, ectopic expression of both *araL* and *araM* in a strain carrying an *araR*-null mutation and a deletion of all genes downstream of *araD* failed to re-establish the toxic effect (Table 2.5). Nevertheless

one cannot exclude the possibility that these constructs may result in low levels of AraL accumulation in the cell (see Chapter III). Neither *araL* or *araM* are responsible for the toxic effect, thus the hypothesis that the toxic effect could be due an increased intracellular level of arabinose in the *araR*- null mutant, caused by deregulation of all arabinose responsive genes, and consequently cause an increase in the concentration of the metabolic sugar phosphates intermediates that are toxic to the cell is still valid. Likewise, the suggestion that the large deletion in the genes downstream from *araD* may cause upstream mRNA destabilization which consequently causes a decrease in the concentration of the enzymes involved in the catabolism of arabinose leading to a lower the concentration of the sugar phosphates intermediates, thus suppressing the toxic effect (Chapter IV).

Acknowledgments

This work was partially funded by Fundação para a Ciência e a Tecnologia, grants no. [SFRH/BD/73109/2010 to L.G; PEst-OE/BIA/UI0457/2011 to Centro de Recursos Microbiológicos]

Chapter III

Characterization and regulation of a bacterial sugar phosphatase of the haloalkanoate dehalogenase superfamily, AraL, from *Bacillus subtilis*

Publication associated with this chapter:

Lia M. Godinho and Isabel de Sá-Nogueira (2011) Characterization and regulation of a bacterial sugar phosphatase of the haloalkanoate dehalogenase superfamily, AraL, from *Bacillus subtilis*. *FEBS Journal*, **278** 2511–2524.

The author of this Thesis performed all experiments described in this chapter.

Abstract

AraL from *Bacillus subtilis* is a member of the ubiquitous haloalkanoate dehalogenase, HAD, superfamily. The *araL* gene has been cloned, over-expressed in *Escherichia coli* and its product purified to homogeneity. The enzyme displays phosphatase activity, which is optimal at neutral pH (7.0) and 65°C. Substrate screening and kinetic analysis showed AraL to have low specificity and catalytic activity towards several sugar phosphates, which are metabolic intermediates of the glycolytic and pentose phosphate pathways. Based on substrate specificity and gene context within the arabinose metabolic operon, a putative physiological role of AraL in detoxification of accidental accumulation of phosphorylated metabolites has been proposed. The ability of AraL to catabolize several related secondary metabolites requires regulation at the genetic level. Here, by site-directed mutagenesis, we show that AraL production is regulated by a structure in the translation initiation region of the mRNA, which most probably blocks access to the ribosome-binding site, preventing protein synthesis. Members of HAD subfamily IIA and IIB are characterized by a broad-range and overlapping specificity that anticipated the need for regulation at the genetic level. In this study we provide evidence for the existence of a genetic regulatory mechanism controlling AraL production.

Introduction

Phosphoryl group transfer is a widely used signaling transfer mechanism in living organisms, from bacteria to animal cells. Phosphate transfer mechanisms are often part of strategies used to respond to different external and internal stimuli, and protein degradation (Dzeja, 2003). Phosphoryl-transfer reactions, catalyzed by phosphatases, remove phosphoryl groups from macromolecules and metabolites (Knowles, 1980). It is estimated that about 35-40% of the bacterial metabolome is composed of phosphorylated metabolites (Kuznetsova *et al.*, 2006). The majority of cellular enzymes responsible for phosphoryl transfer belong to a rather small set of superfamilies, all evolutionary distinct and with different structural topologies, but almost exclusively restricted to phosphoryl group transfer.

The HAD superfamily is one of the largest and most ubiquitous enzyme families identified to date (about 61435 sequences reported; http://pfam.xfam.org/clan/CL0137), and it is highly represent in individual cells. The family was named after the archetypal enzyme haloacid dehalogenase, which was the first family member structurally characterized (Aravind *et al.*, 1998; Koonin & Tatusov, 1994). This family, however, comprises a wide range of HAD-like hydrolases, like phosphatases (~79%), and ATPases (20%), the majority of which involved in phosphoryl group transfer to an active site aspartate residue (Allen & Dunaway-Mariano, 2004, 2009; Burroughs *et al.*, 2006). HAD phosphatases are involved in variety of essential biological functions, such as primary and secondary metabolism, maintenance of metabolic pools, housekeeping functions and nutrient uptake (Allen & Dunaway-Mariano, 2009). The highly conserved structural core of the HAD enzymes' consists on a α-β domain

that adopts the topology typical of the Rossmann α/β folds, housing the catalytic site, and is distinguished from all other Rossmanoid folds by two unique structural motifs: a nearly complete α-helical turn, named the "squiggle", and a β-hairpin turn, termed the "flap" (Allen & Dunaway-Mariano, 2009; Burroughs *et al.*, 2006; Lu *et al.*, 2008). The HAD superfamily can be divided into three generic subfamilies based on the existence and location of a cap domain involved in substrate recognition. Subfamily I possesses a small α-helical bundle cap between motifs I, and II, subfamily II displays a cap between the second and third motifs and subfamily III members present no cap domain (Lu *et al.*, 2005). Subfamily IIA, based on the topology of the cap domain can be further divided into two subclasses, subclass IIA and subclass IIB (Lu *et al.*, 2005).

Presently, about 19000 sequences are assigned to HAD subfamily IIA, which cover humans, other eukaryotes, Gram-positive and Gram-negative bacteria (http://www.ebi.ac.uk/interpro/IEntry?ac=IPR006357). The *Escherichia coli* NagD (Peri *et al.*, 1990) and the *Bacillus subtilis* putative product AraL (Sá-Nogueira *et al.*, 1997) typify this subfamily. NagD is a nucleotide phosphatase, encoded by the *nagD* gene, which is part of the N-acetylglucosamine operon (*nagBACD*). The purified enzyme hydrolyzes a number of phosphate containing substrates, and it has high specificity for nucleotide monophosphates, and in particular UMP and GMP. The structure of NagD has been determined and the occurrence of *NagD* in the context of the *nagBACD* operon suggested its involvement in the recycling of cell wall metabolites (Tremblay *et al.*, 2006). Although, this subfamily is widely distributed only few members have been characterized.

Here we report, the over production, purification and characterization of the AraL enzyme from *B. subtilis*. AraL is shown to be a phosphatase displaying activity towards different sugar phosphate substrates. Furthermore, we provide evidence that in both *E. coli* and *B. subtilis* production of AraL is regulated by the formation of an mRNA secondary structure, which sequesters the ribosome-binding site and consequently prevents translation. AraL is the first sugar phosphatase belonging to the family of NagD-like phosphatases characterized at the level of gene regulation.

Materials and Methods

Substrates. *p*-nitrophenyl phosphate (*p*NPP) was purchased from Apollo Scientific Ltd., and D-xylulose 5-phosphate, Glucose-6-phosphate, Fructose 6-phosphate, Fructose 1,6-bisphosphate, ribose 5-phosphate, D-arabinose 5-phosphate, Galactose 1-phosphate glycerol 3-phosphate, pyridoxal 5-phosphate, thiamine monophosphate, adenosine 5'-triphosphate (ATP), adenosine 5'-diphosphate (ADP), Adenosine monophosphate (AMP) from Sigma-Aldrich Co, St. Louis, MO, USA.

Bacterial strains and growth conditions. *E. coli* strains XL1Blue (Stratagene, La Jolla, CA, USA) or DH5α (Gibco-BRL, Carlsbad, CA, USA) were used for molecular cloning work and *E. coli* BL21 (DE3) pLysS used for overproduction of AraL (Table 2.3). *E. coli* strains were grown in LB medium (Miller, 1972) or in auto-induction medium (Studier, 2005). Ampicillin (100 $\mu g.mL^{-1}$), chloramphenicol (25 $\mu g.mL^{-1}$), kanamycin (30 $\mu g.mL^{-1}$), tetracycline (12 $\mu g.mL^{-1}$) and IPTG (100mM) were added as appropriate. *B. subtilis* was grown in liquid LB medium, LB medium solidified with 1.6% (w/v) agar with chloramphenicol (5 $\mu g.mL^{-1}$), erythromycin (1 $\mu g.mL^{-1}$), and X-Gal (50 $\mu g.mL^{-1}$) being added as appropriate. Growth kinetics parameters of the wild-type and mutant *B. subtilis* strains were determined in CSK liquid minimal medium (Debarbouillé *et al.*, 1990) as previously described (Inácio *et al.*, 2008). Cultures were grown on an Aquatron® Waterbath Rotary Shaker, at 37 ^{o}C (unless stated otherwise), 180 rpm, and OD was measured at 600 nm in an Ultrospec™ 2100 *pro* UV/Visible Spectrophotometer (GE Healthcare Life Sciences, Uppsala, Sweden).

DNA manipulation and sequencing. DNA manipulations were carried out as described by Sambrook *et. al.* (Joseph Sambrook & David W. Russel, 2001). Restriction enzymes were purchased from MBI Fermentas, Vilnius, Lithuania or New England Biolabs, United Kingdom, and used according to the manufacturer's instructions. DNA ligations were performed using T4 DNA Ligase (MBI Fermentas, Vilnius, Lithuania). DNA was eluted from agarose gels with GFX Gel Band Purification kit (GE Healthcare Life Sciences, Uppsala, Sweden), and plasmids purified using the QIAGEN® Plasmid Midi kit (Qiagen, Hilden, Germany) or QIAprep® Spin Miniprep kit (Qiagen, Hilden, Germany). DNA sequencing was performed with ABI PRIS BigDye Terminator Ready Reaction Cycle Sequencing kit (Applied Biosystems, Carlsbad, CA, USA). PCR amplifications were conducted using high-fidelity *Phusion*® DNA polymerase from Finnzymes, Espoo, Finland.

Plasmid constructions

Plasmids pLG5, pLG11 and pLG12 are pET30a (+) derivatives (Table 3.1), which harbor different versions of *araL* bearing a C-terminal His_6-tag, under the control of T7 inducible promoter. The coding sequence of *araL* was amplified by PCR using chromosomal DNA of the wild-type strain *B. subtilis* $168T^{+}$ as template and different sets of primers. To construct pLG5, oligonucleotides ARA439 and ARA440 (Table 3.2) were used and introduced unique NdeI and XhoI restriction sites at the 5' and 3' end, respectively, and the resulting PCR product inserted into pET30a (+) digested with the same restriction enzymes. Using the same procedure, primers ARA457 and ARA440 (Table 2.2) generated pLG11. ARA457 introduced mutation, which substitutes Val at position 8 to Met (Figure 2.1). Plasmid pLG12 was constructed with primers ARA456 and ARA440. Primer ARA456 inserted an NdeI restriction site in the *araL* sequence at the second putative start codon (Figure 3.1).

Site-Directed Mutagenesis. Vector pLG12 was used as template for site-directed mutagenesis experiments using the mutagenic oligonucleotides set ARA486 and ARA487 (Table 3.3). This pair of primers generated a G→A substitution in the 5' end of the *araL* coding region (Figure 3.1). This substitution originated a mutation in the residue at position 12 (Gly to Asp) in the resulting plasmid pLG13. A polymerase chain reaction was carried on using 1x Phusion® GC Buffer (Finnzymes, Espoo, Finland), 0.2 μM primers, 200 μM dNTPs, 3% DMSO, 0.4 ng.μL^{-1} pLG12 DNA and 0.02 U.μL^{-1} of Phusion® DNA polymerase in a total volume of 50 μL. The PCR product was digested with 10 U of DpnI, at 37 ºC, overnight. The mutation was confirmed by sequencing.

Overproduction and purification of recombinant AraL proteins in *E. coli.* Small scale growth of *E. coli* BL21 (DE3) pLysS cells harboring pLG5, pLG11, pLG12 and pLG13 was performed to assess the overproduction and solubility of the recombinant proteins. Cells were grown at 37 ºC, 180 rpm and 1 mM IPTG was added at OD_{600nm}=0.6. Cultures were then grown for an additional 3 hours at 37 ºC, 180 rpm. Whenever protein solubility was not observed, an auto-induction regime for the overproduction of AraL recombinant proteins was used (Studier, 2005). To prepare the cell-free extracts, the cells were resuspended in lysis buffer (20 mM sodium phosphate buffer, pH 7.4, 62.5 mM NaCl, 10 mM imidazole, glycerol 10%) and disrupted in the presence of lysozyme (1 mg.ml^{-1}) by three cycles of freezing in liquid N_2 and thawing for 5 min at 37 °C, followed by incubation with benzonase nuclease (Novagen®, Darmstadt, Germany). After 15 min centrifugation at 16 000*g* and 4°C the soluble and insoluble fractions of the crude extract were obtained.

For overproduction and purification of recombinant AraL-His_6, *E. coli* BL21(DE3) pLysS cells harboring pLG11 were grown in 100 mL of auto-induction medium (Studier, 2005). Cells were harvested by centrifugation at 4 °C, 6000*g*, 10 min. All subsequent steps were carried out at 4 °C. The harvested cells were resuspended in Start Buffer (Tris-HCl 100 mM buffer, pH 7.4, 62.5 mM NaCl, 10 mM imidazole, glycerol 10%) and lysed by passing three times through a French pressure cell. The lysate was centrifuged for 1 hour at 13 500 *g* and the proteins from the supernatant were loaded onto a 1 mL Histrap Ni^{2+}- NTA affinity column (GE Healthcare Life Sciences, Uppsala, Sweden). The bound proteins were eluted with a discontinuous imidazole gradient and the fractions containing AraL that were more than 95 % pure were dialyzed overnight against storage buffer (Tris-HCl 100 mM buffer, pH 7.4, 100 mM NaCl, glycerol 10%) and then frozen in liquid N_2 and kept at -80 °C until further use.

Protein analysis. Analysis of production, the homogeneity, and the molecular mass of the enzyme was determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), using broad-range molecular weight markers (Bio-Rad Laboratories, Hercules, CA, USA) as the standards. The degree of purification was determined by densitometric analysis of Coomassie blue-stained SDS-PAGE gels. The protein content was determined by using the Bradford reagent (Bio-Rad Laboratories, Hercules, CA, USA) with bovine serum albumin as the standard.

Enzyme assays. Phosphatase activity assays were performed using the general substrate *p*-nitrophenyl phosphate (*p*NPP). The reaction mixture containing 100 mM Tris-HCl buffer, pH 7 containing 15 mM $MgCl_2$ and appropriately diluted enzyme (20 µg) was incubated at 37 °C for 5 minutes. Addition of 20 mM of *p*NPP started the reaction and the mixture was further incubated for one hour. The reaction was stopped by adding 1 mL of 0.2 M NaOH, the tubes were centrifuged at 16 000 *g* for 1 min and 1 mL of the supernatant was recovered for measurement of OD_{405nm}. Calibration curve for phosphatase activity assays using *p*NPP as a substrate was performed using various concentrations ($mg.mL^{-1}$) of *p*-nitrophenol, within the measuring range of the method (Huggins & Smith, 1947). Negative controls were made using 20 µg of BSA, and blanks had no protein added. Enzymatic activity was also determined in the presence of 15 mM EDTA, using the same conditions. 1 unit of AraL hydrolyses 1 µmol of substrate per min. Both optimum temperature and pH for enzymatic activity of AraL-His_6 were determined as described above. The effect of temperature was tested in 100 mM Tris-HCl buffer, pH 7 containing 15 mM $MgCl_2$, at temperatures ranging from 25 to 70 °C. The effect of pH on the activity was assayed at 65°C in a series of Britton-Robinson buffers (0.1 M boric acid 0.1 M acetic acid and 0.1 M phosphoric acid (pH 3 to 6) and Tris-HCl buffers (pH 7.0 to 9.0).

Continuous Activity Assays. All continuous assays were carried out at 37 °C in 100 mM Tris-HCl buffer, pH 7 containing 15 mM $MgCl_2$, unless stated otherwise. Glucose production from G6P was monitored by measure of the glucose dehydrogenase catalyzed reduction of NADP. The initial velocity of glucose formation by dephosphorylation of G6P in reaction solutions initially containing 20 µg AraL, 0.7 unit of G6P dehydrogenase, 0.2 mM NADP, 1-15 mM α-G6P, and 15 mM $MgCl_2$ in 0.5 mL of 100 mM Tris-HCl (pH 7.5, 37 °C) was determined by monitoring the increase in the absorbance levels at 340 nm.

Discontinuous Assays. Initial phosphate hydrolysis for all substrates used in substrate screening was assessed to detect total phosphate release using the Malachite Green Phosphate Detection Kit (R&D Systems, Minneapolis, MN, USA) according to manufacturer's instructions. The 150 µL assay mixture, containing 100 mM Tris-HCl buffer (pH 7) containing 15 mM $MgCl_2$, was incubated for 1 h at 37 °C. Background phosphate levels were monitored in parallel using a control reaction without the AraL enzyme. The absorbance at 620 nm was measured. Steady-state kinetics was carried out using 20 µg AraL with varying concentrations of substrates. Kinetic parameters were determined using the enzyme kinetics software program GraphPad Prism version 5.00 for Windows (GraphPad Software).

In-frame deletion of *araL* in *B. subtilis*. To create *B. subtilis* mutant strains with an in-frame deletion of *araL* plasmid pLG10 was constructed using pMAD (Table 3.1). Regions immediately upstream and downstream of *araL* were amplified by two independent PCR experiments, from chromosomal DNA of *B. subtilis* 168T$^+$, using primers ARA444 and ARA458 (PCR1) and ARA459 and ARA460 (PCR2). The products were joined by overlapping PCR, with primers ARA444 and ARA460 (Table 3.2) and the resulting 1262 bp fragment digested with BamHI and EcoRI and cloned into pMAD BamHI-EcoRI yielding pLG10. This plasmid harboring an in-frame deletion of *araL* was used for integration and generation of a clean deletion in the *B. subtilis* chromosome following the published procedure described by Arnaud *et. al.* (Arnaud *et al.*, 2004). The in-frame deletion then confirmed by DNA sequencing and the resulting strain named IQB832. Transformation of *B. subtilis* was performed according to the method described by Anagnostopoulos & Spizizen (Anagnostopoulos & Spizizen, 1960).

Construction of an in-frame *araL'-'lacZ* fusion and integration at an ectopic site. To construct plasmid pLG25 the arabinose operon promoter region (- 81 to +129, relative to the transcriptional start site) was amplified from chromosomal DNA of the *B. subtilis* wild-type strain 168T$^+$ using oligonucleotides ARA28 and ARA451 (Table 3.2). The primers introduced unique EcoRI and HindIII restriction sites and the resulting fragment was sub-cloned into the same sites of the cloning vector pLG1 (see chapter II for details on this vector). Sequentially the *araL* 5'-end region comprising the rbs (position +3910 to +4020, relative to the transcriptional start site of the operon) was amplified from the wild-type strain with oligonucleotides ARA253 and ARA477 (Table 3.2), which carry unique XbaI and BamHI restriction sites and allow insertion of this fragment between the NheI and BamHI sites of pLG1. In the resulting plasmid, a deletion of the *araA* rbs and *araA* start site present in the arabinose promoter region (Para) was performed by overlapping PCR using two set of primers ARA358 and ARA514, and ARA515 and ARA516 (Table 3.2). The resulting fragment of 216 bp, comprises the arabinose promoter region (Para) from -81 to +80 fused to the *araL* 5'-end region from +3952 to + 4007, was inserted into the vector pAC5 (Table 3.1), yielding pLG25. Plasmid pLG25 carries a translational fusion between codon 10 of *araL* and codon 7 of *E. coli lacZ.* pLG25 was used as template for site-directed mutagenesis experiments using the mutagenic oligonucleotides set ARA509 and ARA510 (Table 3.2), as described above. This pair of primers generated a C→A substitution in the 5' end of the *araL* coding region (Figure 3.2 and Figure 3.6A). The substitution originated a mutation in the residue at position 9 (Thr to Lys) in the resulting plasmid pLG26. pLG26 was then used as template for site-directed mutagenesis using primers ARA549 and ARA550, which allowed a C → G substitution (Thr to Arg) in position 9, thus originating pLG28. Oligonucleotide set ARA551 and ARA552 introduced a double point mutation in the 5' end of *araL* coding sequence. Using pLG26 as template, the set of primers caused T → C (Arg to Pro) and C → G (Thr to Arg) mutations in the 2nd and 9th positions, respectively, yielding pLG27. Translation vector pLG29 was obtained by site-directed

mutagenesis from pLG26, using the oligonucleotide pair ARA553 and ARA554, originating a G → T substitution (Arg to Leu) in position 2. The constructions were confirmed by DNA sequencing.

DNA from plasmids pLG25 and pLG26, carrying the different *araL'*-'*lacZ* translational fusions, was used to transform *B. subtilis* strains (Table 3.3) and the fusions ectopically integrated into the chromosome via double recombination with the *amyE* gene back and front sequences. This event led to the disruption of the *amyE* locus and was confirmed as described previously (Sá-Nogueira & Mota, 1997).

β-Galactosidase activity assays. Strains of *B. subtilis* harboring the transcriptional *lacZ* fusions were grown were grown in liquid C minimal medium (Sá-Nogueira & Mota, 1997) supplemented with casein hydrolysate 1% (w/v) and arabinose was added to the cultures when necessary at a final concentration of 0.4% (w/v), as previously described (Sá-Nogueira & Mota, 1997). Samples of cell culture were collected two hours (exponential growth phase) after induction and the level of accumulated β-galactosidase activity determined as previously described (Sá-Nogueira & Mota, 1997).

In-frame point mutation in the *araL* locus of *B. subtilis*. The allelic replacement vector pMAD was again used to introduce a single point mutation already tested in the β-Galactosidase activity assays, using the in-frame *araL'-'lacZ* fusion integration at an ectopic site pLG26. The *araDLM* fragment from *B. subtilis* 168T^+ was amplified using oligos ARA472 and ARA461. This fragment was then cloned in the SmaI site of pBKS II (+), originating a vector with suitable size for site-directed mutagenesis, which was performed using mutagenic oligos ARA509 and ARA510, as described previously. The mutagenized *araDLM* region was then subcloned in pMAD in the EcoRI and BamHI sites, resulting in plasmid pLG30, and the allelic exchange procedure was carried on as mentioned above. Both plasmid and resulting strain IQB869 were sequenced to confirm the point mutation.

Immunoblotting of cell extracts. *B. subtilis* strains were grown as described for the β-galactosidase assays (see above), and 8 mL of cell culture was harvested 2 h after induction. After resuspension in lysis buffer (500 mM KCl, 20 mM HEPES- K^+, 10 mM EDTA, 1 mM dithioerythritol, 10% glycerol), lysozyme was added to the cell suspension at a final concentration of 1 $mg.mL^{-1}$, following an incubation at 37 °C for 10 min. Cells were subjected to 3 cycles of freezing in liquid N_2 and thawing at 37 °C. PMSF and benzonase were added, and incubation was continued for 10 min. Samples were rapidly frozen in liquid N_2 and then stored at -80 °C. Samples containing 10 µg of protein were resolved on 12.5% sodium dodecyl sulfate-polyacrylamide gel electrophoresis gels. Gels were transferred for 1h at 100 V to nitrocellulose membranes. Immunoblotting was performed with rabbit polyclonal anti-AraL serum (1:1000, Eurogentec, Liège, Belgium) and immunoblot detection was done with horseradish peroxidase-conjugated goat anti-rabbit antibody (1:10 000, Thermo Scientific, Pierce

Antibody Products, Rockford, IL, USA) by use of an ECL detection system (Western Lightning Plus ECL from Perkin Elmer, Waltham, MA, USA) as described by the manufacturer. Protein concentration was determined as described above**.**

Table 3.1. List of plasmids used in this study.

Plasmid	Relevant construction	Source
pET30a (+)	Expression vector allowing N- or C- terminal His_6 tag insertion; T7 promoter, *kan*	Novagen
pBKS II (+)	Standard cloning vector for *E. coli*, *bla*	Stratagene
pMAD	Plasmid used for allelic replacement in Gram-positive bacteria, *bla*, *erm*	(Arnaud *et al.*, 2004)
pAC5	Plasmid used for generation of *lacZ* translational fusions and integration at the *B. subtilis amyE* locus, *bla*, *cat*	Weinrauch *et. al.* 1991
pLG5	*araL* sequence with the first putative *araL* start codon cloned in the pET30a (+) vector	Present study
pLG10	pMAD derivative with an in-frame deletion Δ*araL*	Present study
pLG11	*araL* sequence with mutated GTG codon (valine at position 8) to ATG (methionine) cloned in the pET30a (+) vector	Present study
pLG12	*araL* sequence with the putative second *araL* start codon cloned in the pET30a (+) vector	Present study
pLG13	A pLG12 derivative with a mutation in the *araL* sequence GGC to GAC (Gly12 to Asp)	Present study
pLG25	A pAC5 derivative that contains a translational fusion of *araL* to the *lacZ* gene under the control of the Para	Present study
pLG26	A pLG25 derivative with a mutation in the *araL* sequence ACG to AAG (Thr9 to Lys)	Present study
pLG30	pMAD derivative with a point mutation in the *araL* sequence ACG to AAG (Thr9 to Lys)	Present study

Table 3.2. List of oligonucleotides (Primers) used in this study. Restriction sites are underlined, as are single-nucleotide point mutations

Primer	Sequence (5' → 3')
ARA28	CCTATTGAATTCAAAAGCCGG
ARA253	TAACCCCAATCTAGACAGTCC
ARA358	CTGCTGTAATAATGGGTAGAAGG
ARA439	GGAATTCCATATGCGTATTATGGCCAG
ARA440	TATTTACTCGAGAATCCCCTCCTCAGC
ARA444	CGGGATCCACCGTGAAAAAGAAAGAATTGTC
ARA451	GAATTCATAAAGAAGCTTTGTCTGAAGC
ARA456	CGGCGCGTCATATGGCCAGTCATGATA
ARA457	TGATACGCATATGTCACCGGCTGGC
ARA458	CTCAGCCAATTTGGTTACATCCTTGTCCAAGTCAATCAGAATGCCAGCCGGTGCCAC
ARA459	GTGTCACCGGCTGGCATTCTGATTGACTTGGACAAGGATGTAACCAAATTGGCTGAG
ARA460	CGTGAATTCACCGAGCATGTCACCAAAGCC
ARA461	CGAACCGACTGCGATCATGAC
ARA472	CCAACCTGAAGCTTCAAGAG
ARA477	AATCAGAATGGGATCCGGTGA
ARA486	CGGCTGACATTCTGATTGACTTGGACGG
ARA487	CAATCAGAATGTCAGCCGGTGACACAGG
ARA509	CCAGTCATGATAAGCCTGTGTCACCG
ARA510	CGGTGACACAGGCTTATCATGACTGG
ARA514	TAATACGCATTTGCTCCGTGTTTTCGTCATAAAATAAAACGCTTTCAAATAC
ARA515	GTATTTGAAAGCGTTTTATTTTATGACGAAAACACGGAGCAAATGCGTATTA
ARA516	CACCACGCTCATCGATAATTTCACC
ARA549	GGCCAGTCATGATAGGCCTGTGTCACC
ARA550	GGTGACACAGGCCTATCATGACTGGCC
ARA551	GCAAATGCCTATTATGGCCAGTCATGATAGGCCTGTGTC
ARA552	GACACAGGCCTATCATGACTGGCCATAATAGGCATTTGC
ARA553	CGGAGCAAATGCTTATTATGGCCAGTC
ARA554	GACTGGCCATAATAAGCATTTGCTCCG

Table 3.3. List of strains used in this study. Arrows indicate transformation and point from donor DNA to recipient strain.

Strain	Relevant Genotype	Source
***E. coli* Strains**		
DH5α	*fhuA*2 Δ(*argF-lacZ*) U169 *phoA gln*V44 Φ80 Δ(l*acZ*)M15 *gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	Gibco - BRL
XL1 Blue	(*recA1 endA1 gyrA96 thi-1 hsdr17 supE44 relA1 lac* [F' *proAB lacI*^{q} ZΔM15 Tn10 (*TetR*)]	Stratagene
BL21 (DE3) pLysS	F^{-} *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pLysS (Cm^R)	(Studier *et al.*, 1990)
***B. subtilis* strains**		
168T^{+}	Prototroph	(Sá-Nogueira *et al.*, 1997)
IQB832	Δ*araL*	Chapter II – Table 2.3
IQB831	Δ*araL* Δ*araR*::*km*	Chapter II – Table 2.3
IQB215	Δ*araR*::*km*	(Sá-Nogueira & Mota, 1997)
IQB847	*amyE*:: Para-*araL'*-*'lacZ cat*	pLG25 → 168T^{+}
IQB848	Δ*araR*::*km amyE*:: Para-*araL'*-*'lacZ cat*	pLG25 →IQB215
IQB849	*amyE*:: Para-*araL'* (C→A) -*'lacZ cat*	pLG26 → 168T^{+}
IQB851	*amyE*:: *'lacZ cat*	pAC5 → 168T^{+}
IQB853	*amyE*:: Para-*araL'* (T→C C→G) -*'lacZ cat*	pLG27 → 168T^{+}
IQB855	*amyE*:: Para-*araL'* (C→G) -*'lacZ cat*	pLG28 → 168T^{+}
IQB857	*amyE*:: Para -*araL'* (G→T) -*'lacZ cat*	pLG29 → 168T^{+}
IQB869	*araL* (C→A) T6K	pLG30 → 168T^{+}
IQB870	Δ*araR*::*km araL* (C→A) T6K	pLG30 → IQB215

Results and Discussion

The *araL* gene in the context of the *B. subtilis* genome and *in silico* analysis of AraL

The *araL* gene is the fourth cistron of the transcriptional unit *araABDLMNPQ-abfA* (Sá-Nogueira *et. al.* 1997). This operon is mainly regulated at the transcriptional level by induction in the presence of arabinose and repression by the regulator AraR (Mota *et al.*, 1999; Sá-Nogueira & Mota, 1997). To date, *araL* is the only uncharacterized open reading frame comprised in the operon (Figure 3.1). The putative product of *araL* displays some similarities to *p*-nitrophenyl phosphate-specific phosphatases from the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (Kaneko *et al.*, 1989; Yang *et al.*, 1991) and other phosphatases from the HAD superfamily, namely the NagD protein from *E. coli* (Tremblay *et al.*, 2006). Although the yeast enzymes were identified as phosphatases, no biologically relevant substrate could be determined, and both enzymes appeared to be dispensable for vegetative growth and sporulation. The purified NagD hydrolyzes a number of nucleotide and sugar phosphates.

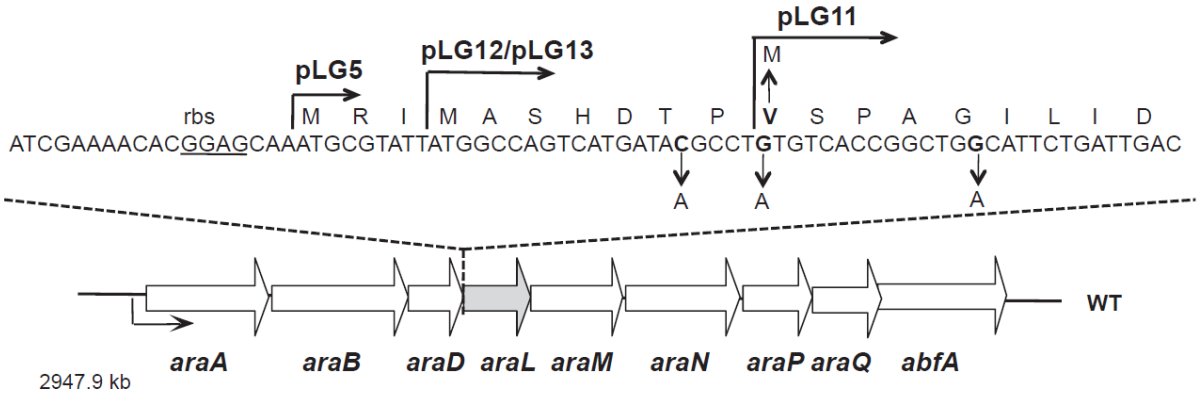


Figure 3.1. Schematic representation of the *araL* genomic context in *B. subtilis*. White arrows pointing in the direction of transcription represent the genes in the arabinose operon, *araABDLMNPQ-abfA*. The *araL* gene is highlighted in grey and the promoter of the transcriptional unit depicted by a black arrow. Above is displayed the coding sequence of the *araL* 5'-end. The putative ribosome-binding site, rbs, is underlined. The 5´-end of *araL* present in the different constructs pLG5, pLG11, pLG12 and pLG13, is indicated by an arrow above the sequence. Mutations introduced in the construction of pLG11, pLG13 and pLG26 are indicated below de DNA sequence and the corresponding modification in the primary sequence of AraL depicted above.

The *araL* gene contains two in-frame ATG codons in close proximity (within 6 bp; Figure 3.1). The sequence reported by Sá-Nogueira *et. al.* (Sá-Nogueira *et al.*, 1997), assumed as initiation codon the second ATG, placed further downstream (Figure 3.1), as the putative start codon for the *araL* gene because its distance relative to the ribosome-binding site is more similar to the average distance, 5 to 11 bp, observed in *Bacillus* (Rocha *et al.*, 1999). However, in numerous databases the upstream ATG is annotated as the initiation codon (Kunst *et al.*, 1997). Assuming the second ATG the *araL* gene encodes a protein of 269 amino acids with a molecular mass of 28.9 kDa.

HAD family members are identified in amino acid alignments by four active site loops which form the mechanistic gear for phosphoryl transfer (Allen & Dunaway-Mariano, 2009). The key residues are an aspartate in motif I (D), a serine or threonine motif II (S/T), an arginine or lysine motif III (K/Y) and an aspartate or glutamate motif IV (D/E). The NagD family members display a unique α/β cap domain involved in substrate recognition, located between the motifs II and III (Burroughs *et al.*, 2006). This family is universally spread, however only a few members have been characterized, such as NagD from *E. coli* (Burroughs *et al.*, 2006; Peri *et al.*, 1990). NagD members are divided in different subfamilies, like the AraL subfamily (Burroughs *et al.*, 2006), but all proteins present a GDxxxxD motif IV (Figure 3.2).

Homologues of the *B. subtilis* AraL protein are found in different species of Bacteria and Archea, genes encoding proteins with more than 50% amino acid identity to AraL are present in *Bacillus* and *Geobacillus* species clustered together with genes involved in arabinose catabolism. An alignment of the primary sequence of AraL with other members of the NagD family from different organisms, NagD from *E. coli* (27% identity), HdpA (formerly cgR_2128) from *Corynebacterium glutamicum* (28% identity) (Jojima *et al.*, 2012), the *p*-nitrophenyl phosphatases (*p*NPPases) from *S. cerevisiae* (24% identity), *Sz. pombe* (30% identity), and *Plasmodium falciparum* (31% identity), highlights similarities and divergence (Figure 2.2). AraL displays the conserved key catalytic residues that unify HAD members: the Asp at position 9 (motif I) that together with Asp 218 (motif IV) bind the cofactor Mg^{2+}, and Ser 52 (motif II) jointly with Lys 193 (motif III) bind the phosphoryl group (Figure 2.2). The cap domain is responsible for substrate binding/specificity, thus the uniqueness or similarity of the amino acid sequence in this domain may determine enzyme specificity or the lack of (Lahiri *et al.*, 2004; Lu *et al.*, 2005; Tremblay *et al.*, 2006). AraL like the other members of the NagD family share two Asp residues in the cap domain (Figure 3.2). To date the number of characterized members of this family is scarce. Here we showed that AraL possesses activity towards different sugar phosphates. The NagD enzyme was observed to have a nucleotide phosphohydrolase activity coupled with a sugar phosphohydrolase activity (Tremblay *et al.*, 2006). The *P. falciparum* enzyme displayed nucleotide and sugar phosphatase activity together with ability to dephosphorylate the vitamin B1 precursor thiamine monophosphate, TMP (Knöckel *et al.*, 2008). The yeast's enzymes are *p*-nitrophenyl phosphatases however, natural substrates were not found (Kaneko *et al.*, 1989; Yang *et al.*, 1991).

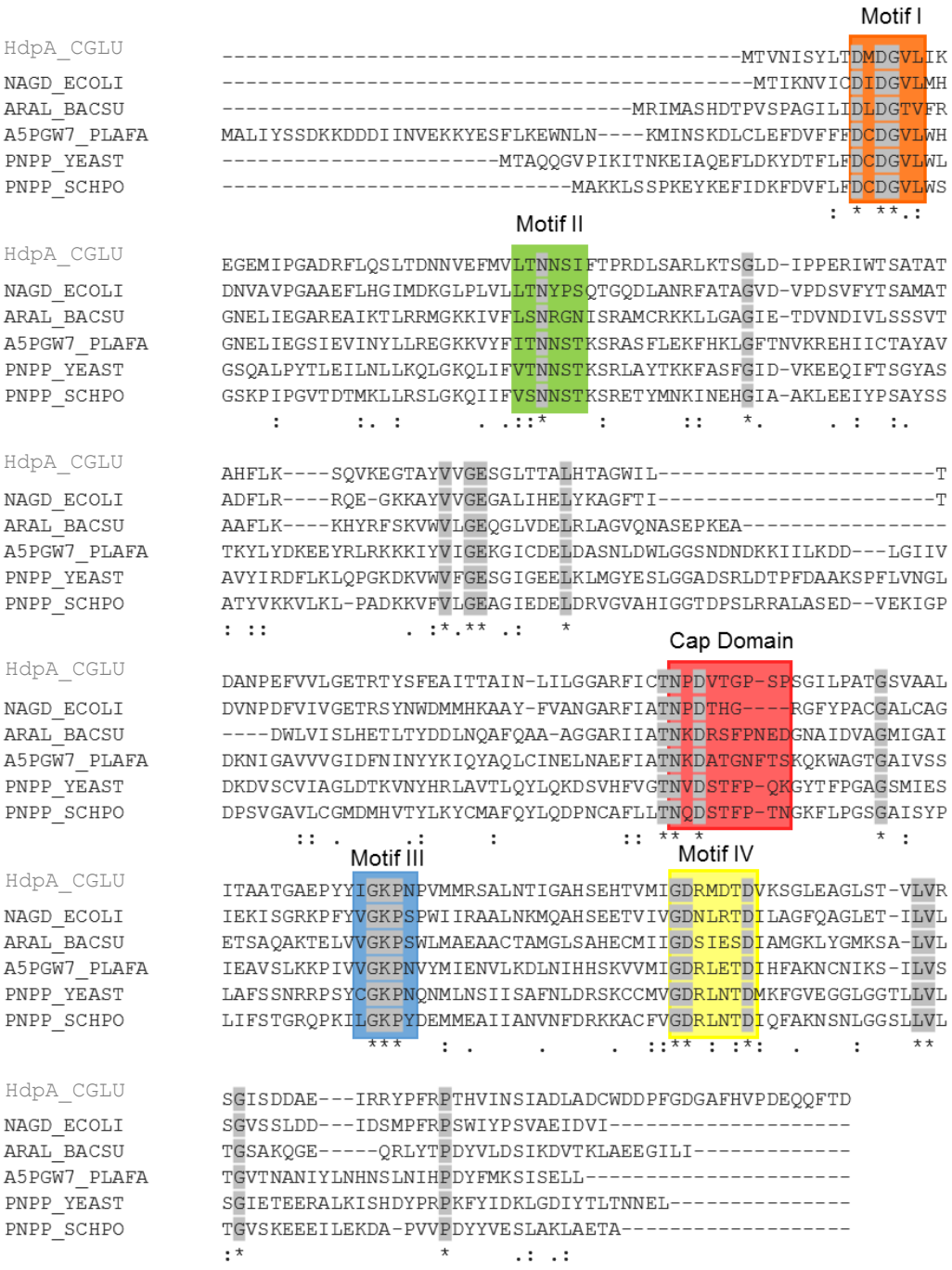


Figure 3.2. Alignment of AraL with other *p*NPPases members of the HAD superfamily (sub family IIA). The amino acid sequences of HdpA (formerly cgR_2128) from *Corynebacterium glutamicum*, AraL from *Bacillus subtilis*, NagD from *Escherichia coli* , the *p*-nitrophenyl phosphatases (*p*NPPases) from *Plasmodium falciparum* (A5PGW7,) *Saccharomyces cerevisiae*, and *Schizosaccharomyces pomb* were aligned using CLUSTAL W2 (Larkin *et al.*, 2007). Similar (.,:) and identical (*) amino acids are indicated. Gaps in the amino acid sequences inserted to optimize alignment are indicated by a '-'. The Motifs I, II, III and IV of the HAD superfamily and the cap domain C2 are boxed. Open arrowheads point to the catalytic residues in motifs I-IV. Identical residues in all five sequences are highlighted in dark grey.

The majority of the enzymes displayed in this alignment show activity to overlapping sugar phosphates (Papagianni *et al.*, 2007; Tremblay *et al.*, 2006) and it is tempting to speculate that is related to similarities in the cap domain. On the other hand, the variability and dissimilarity observed in this region may determine preference for certain substrates (Figure 3.2).

Over-production and purification of recombinant AraL

The full-length of *araL* coding region, starting at both the first and second putative initiation codon ATG and were separately cloned in the expression vector pET30a (+) (Table 3.1), which allows the insertion of a His_6-tag at the C-terminus. The resulting plasmids, pLG5 and pLG12 (Figure 3.1), bearing the different versions of recombinant AraL, respectively, under the control of a T7 promoter, were introduced into *E. coli* BL21(DE3) pLysS (Table 3.1) for the over-expression of the recombinant proteins. The cells were grown in the presence and absence of the inducer IPTG, soluble and insoluble fractions were prepared as described in Experimental Procedures and analyzed by SDS-PAGE. In both cases production of AraL was not detected although different methodologies for over-expression have been used (discussed below).

Based on the alignment of the primary sequence of AraL and NagD, we constructed in pET30a(+) a truncated version of AraL, with a small deletion at the N-terminus (pLG11; Figure 3.1). Production of this truncated version of AraL was achieved in *E. coli* BL21 (DE3) pLysS cells harboring pLG12, after IPTG induction, but the protein was obtained in the insoluble fraction. Thus, over-production was attempted using the auto-induction method described by Studier (Studier, 2005). In the soluble and insoluble fraction of IPTG-induced cells harboring pLG11 a protein of about 29 kDa was detected, which matched the predicted size for the recombinant AraL (Figure 3.3A). The protein was purified to more than 95% homogeneity by Ni-NTA agarose affinity chromatography (Figure 3.3B).

Characterization of AraL

AraL phosphatase activity was measured using the synthetic substrate 4-nitrophenyl phosphate (*p*NPP). AraL is characterized as a neutral phosphatase with optimal activity at pH 7 (Figure 3.4). Although, at pH 8 and 9 the activity was considerably lower than that observed at pH 7, the values are higher than that observed at pH 6, and no activity was measured below pH 4. The optimal temperature was analyzed over a range of temperatures from 25 to 70 °C. The enzyme was most active at 65 °C and at 25 °C no activity was detected (Figure 3.4). These biophysical AraL properties fall into the range found for other characterized phosphatases from *B. subtilis*: pH 7-10.5 and 55 to 65 ^{o}C (Hulett *et al.*, 1990; Ishikawa *et al.*, 2002; Mijakovic *et al.*, 2005; Sugahara *et al.*, 1991; Tye *et al.*, 2002).

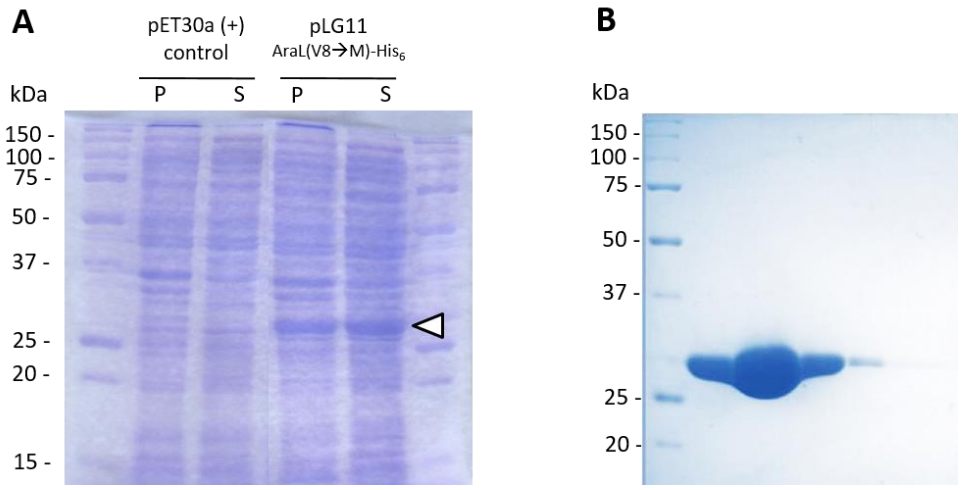


Figure 3.3. Over-production and purification of recombinant AraL-His$_6$. A. Analysis of the soluble (S) and insoluble (P) protein fraction (20 µg total protein) of induced cultures of *E. coli* BL21 (DE3) pLysS harboring pET30a (control) and pLG11 (AraL-His$_6$). **B**. Analysis of different fractions of purified recombinant AraL eluted with 300 mM of imidazole. The proteins were separated by SDS-PAGE 12.5% gels and stained with Coomassie blue. A white arrowhead indicates AraL-His$_6$. The size, in kDa, of the broad range molecular mass markers (Bio-Rad Laboratories, Hercules, CA, USA) are indicated.

HAD superfamily proteins typically employ a divalent metal cation in catalysis and phosphatases, particularly those belonging to the subclass IIA, frequently use Mg^{2+} as cofactor (Allen & Dunaway-Mariano, 2009; Burroughs *et al.*, 2006; Kuznetsova *et al.*, 2006; Tremblay *et al.*, 2006). The effect of divalent ions (Mg^{2+}, Zn^{2+}, Mn^{2+}, Ni^{2+}, Co^{2+}) in AraL activity was tested and the results indicated that catalysis absolutely requires the presence of Mg^{2+} (Figure 3.4). Addition of EDTA to a reaction containing $MgCl_2$, prevented AraL activity.

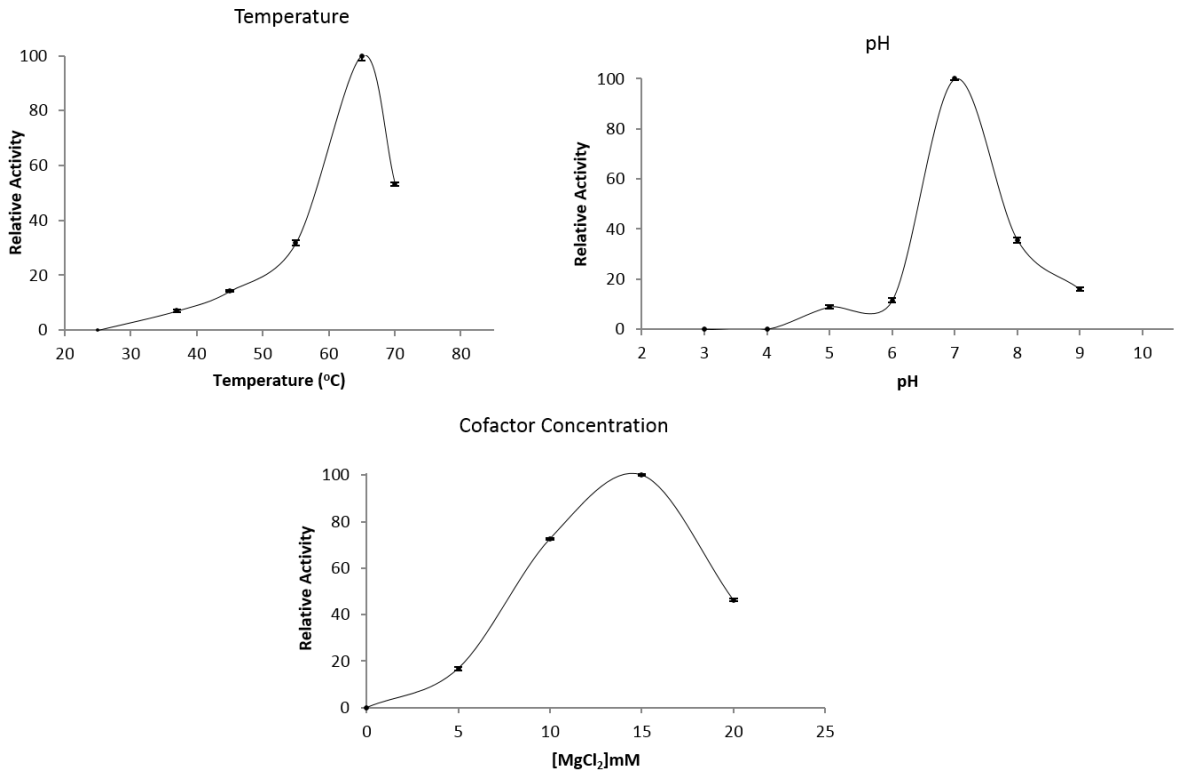


Figure 3.4. Effect of pH, temperature, and co-factor concentration on AraL activity. Enzyme activity was determined using *p*NPP as substrate, at 65 ˚C, pH 7 and 15 mM MgCl2 unless stated otherwise. The results represent the average of three independent experiments.

AraL is a sugar phosphatase

AraL is a phosphatase displaying activity towards the synthetic substrate *p*NPP but there is no evidence that *p*NPPase activity is physiologically relevant. The context of *araL* within the arabinose metabolic operon *araABDLMNPQ-abfA* involved in the transport of L-arabinose oligomers, further intracellular degradation, and catabolism of L-arabinose (Ferreira & Sá-Nogueira, 2010; Inácio *et al.*, 2008; Sá-Nogueira *et al.*, 1997) suggests a possible role as a phosphatase active towards sugar phosphate intermediates in L-arabinose catabolism, such as D-xylulose 5-phosphate. Based on both this clue and the observation that many HAD members display phosphatase activities against various intermediates of the central metabolic pathways, glycolysis and the pentose phosphate pathway (Kuznetsova *et al.*, 2006), we tested AraL activity towards glucose-6-P, fructose 6-P, fructose 1,6-bisphosphate, 3-phosphoglycerate, ribose 5-phosphate, D-xylulose 5-phosphate, and galactose-1-phosphate. Although, *B. subtilis* does not utilize D-arabinose the activity towards D-arabinose 5-phosphate was also assayed. In addition, the nucleotides AMP, ADP, ATP, pyridoxal 5-phosphate and thiamine monophosphate were

also screened (Table 3.4). Although the optimal temperature for enzyme activity is 65 °C the kinetics parameters were measured at 37 °C, the *B. subtilis* optimal growth temperature. It is noteworthy, that in these conditions the K_M determined for *p*NPP is 50 mM (Table 3.4) compared to 3 mM obtained at 65 °C.

Table 3.4. Kinetic constants form AraL against various substrates. Assays were performed at pH 7 and 37 °C as described in experimental procedures. The results are the mean value and standard deviation of triplicates. The following substrates were also tested, but no activity was detected: ATP, ADP, AMP, ribose-5-phosphate, glycerol-3-phosphate, pyridoxal-5-phosphate and thiamine monophosphate.

Substrate	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1} M^{-1})
D-xylulose 5-phosphate	29.14 ± 4.87	2.75 ± 0.26	0.943 x 10^2
Glucose 6-phosphate	24.96 ± 4.08	2.49 ± 0.26	0.998 x 10^2
D-arabinose 5-phosphate	27.36 ± 1.8	2.92 ± 0.10	1.06 x 10^2
Fructose 6-phosphate	34.89 ± 4.51	2.817 ± 0.22	0.807 x 10^2
Fructose 1,6-bisphosphate	40.78 ± 11.40	1.49 ± 0.26	0.365 x 10^2
Galactose 1-phosphate	40.74 ± 6.03	4.28 ± 0.40	1.02 x 10^2
***p*NPP**	50.00 ± 23.32	0.012 ± 0.0006	0.24

The AraL enzyme showed reactivity with D-xylulose-5-phosphate, D-arabinose 5-phosphate, galactose-1-phosphate, glucose-6-phosphate, fructose-6-phosphate and fructose 1,6-bisphosphate (Table 3.4).

The K_M values are high (~30 mM) and above the range of the known bacterial physiological concentrations. In *E. coli* the intracellular concentration of ribose 5-phosphate, glucose-6-phosphate, fructose 6-phosphate and fructose 1,6-bisphosphate, is 0.18 - 6 mM (Kuznetsova *et al.*, 2006) , and in *B. subtilis* the measured concentration of fructose 1,6-bisphosphate when cells were grown in the presence of different carbon sources, including arabinose, varies within the range 1.8 - 14.1 mM (Singh *et al.*, 2008). However, we cannot rule them out as feasible physiological substrates because in certain conditions the intracellular concentration of glucose-6-phosphate, fructose 6-phosphate and fructose 1,6-bisphosphate, may reach concentrations of 20 - 50 mM, as reported for *Lactococcus lactis* (Papagianni *et al.*, 2007). Nevertheless, the mean value of the substrate specificity constant k_{cat}/K_M is

low (1 x 10^2 M^{-1} s^{-1}), thus the ability of AraL to distinguish between these sugar phosphate substrates will be limited. The results obtained for AraL are comparable to those obtained for other members of HAD from subfamilies IIA and IIB, which have in common a low substrate specificity and catalytic efficiencies (k_{cat}/K_M < 1 x 10^5 M^{-1} s^{-1}) and lack defined boundaries of physiological substrates (Lu *et al.*, 2005; Tremblay *et al.*, 2006). These features point to enzymes functioning in secondary metabolic pathways.

Production of AraL in *E. coli* is subjected to regulation

In silico DNA sequence analysis of pLG12 and pLG5 detected the possible formation, in both plasmids, of an mRNA secondary structure, which sequesters the ribosome-binding site. Both, hairpin structures, display low free energy -17.5 kcal mol^{-1} (Figure 3.5A) and -22.7 kcal mol^{-1}, respectively, and could impair translation preventing the production of AraL observed in these constructs (see above). In plasmid pLG11, carrying the truncated version of AraL, overproduction was successful (Figure 3.3). The deletion of the *araL* gene 5'-end caused an increase of the free energy of the putative mRNA secondary structure (-11.8 kcal mol^{-1}). To test the potential involvement of the mRNA secondary structure in the lack of production of the recombinant AraL versions constructed in plasmids pLG12 and pLG5, site directed mutagenesis was performed using pLG12 as template. A single-base substitution G→A introduced at the 5´end of the gene (Figure 3.1) was designed in order to increase the free energy of the mRNA secondary structure in the resulting plasmid pLG13. This point mutation increased the free energy from -17.5 kcal mol^{-1} to -13.1 kcal mol^{-1} (Figure 3.5A). In addition, this modification caused the substitution of a glycine to an aspartate at position 12 in AraL (G12→D; Figure 3.1) however, based on the structure of NagD from *E. coli* (Tremblay *et al.*, 2006) this amino acid substitution close to the N-terminus it is not expected to cause major interference in the overall protein folding. Cell extracts of induced *E. coli* Bl21 (DE3) pLysS cells carrying pLG13 were tested for the presence of AraL. A strong band with an estimated size around 29 kDa, was detected (Figure 3.5B) strongly suggesting that recombinant AraL is produced in *E. coli* when the mRNA secondary structure is destabilized. This observation indicates that production of AraL is modulated by a secondary mRNA structure placed at the 5'-end of the *araL* gene.

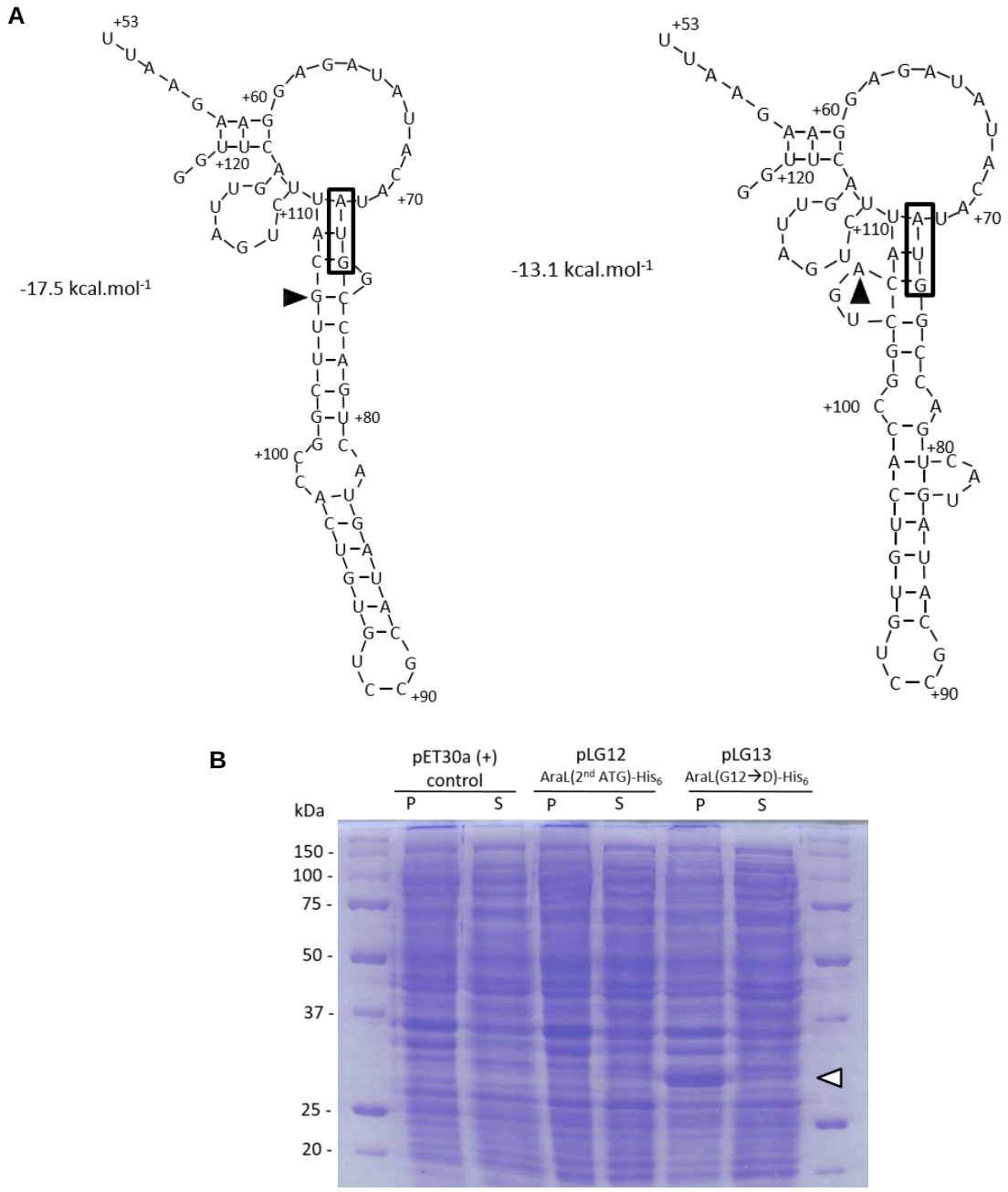
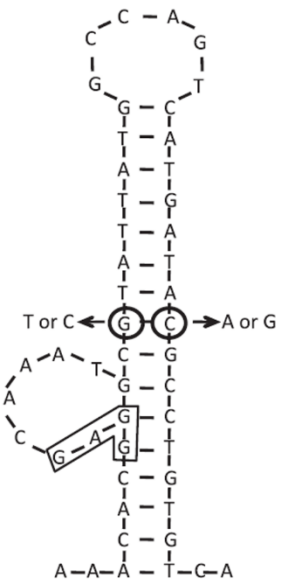


Figure 3.5. Site-directed mutagenesis in the 5'-end *araL* and over-production of recombinant AraL-His$_6$. **A**. The secondary structure of the *araL* mRNA in pLG12 (left) and pLG13 (right), which bears a single nucleotide change. An arrowhead highlights the mutated nucleotide located at the beginning of the *araL* coding region. The ribosome-binding site, rbs, and the initiation codon (ATG) are boxed. Position relative to the transcription start site is indicated. The free energy of the two secondary structures, calculated by DNAsis v 3.7 (©Hitachi Software Engineering Co. Ltd), is shown. **B**. Over-production of recombinant AraL-His$_6$. Analysis of the soluble (S) and insoluble (P) protein fraction (20 µg total protein) of induced cultures of *E. coli* BL21 (DE3) pLysS harboring pLG12 (AraL-His$_6$) and pLG113 (AraL-His$_6$ G→A). The proteins were separated by SDS-PAGE 12.5% gels and stained with Coomassie blue. A white arrowhead indicates AraL-His$_6$. The sizes, in kDa, of the broad range molecular mass markers (Bio-Rad Laboratories, Hercules, CA, USA) are indicated.

Regulation and putative role of AraL in *B. subtilis*

In *B. subtilis* the formation of a similar hairpin structure at the same location is possible and displays a free energy of -21.4 kcal mol^{-1} (Figure 3.6A). To determine its role in the regulation of *araL* expression a translational fusion of the *araL* 5'-end to the *lacZ* reporter gene from *E. coli* was constructed and integrated into the *B. subtilis* chromosome, as single copy, at an ectopic site. The construct comprises the *araL* ribosome-binding site, initiation codon, and a fusion between codon 10 of *araL* and codon 7 of *E. coli lacZ*. The *araL'-'lacZ* translational fusion is under the control of the strong promoter (Para) of the *araABDLMNPQ-abfA* operon (Figure 3.6A). However, expression from the *araL'-'lacZ* fusion in the presence of arabinose (inducer) is very low as determined by measuring the levels of accumulated β-galactosidase activity in strain IQB847 (Figure 3.6). In contrast, strain IQB849 carrying a single-base substitution C→A introduced in the hairpin region displayed an augment in *araL'-'lacZ* expression of about 30-fold in the presence of inducer (Figure 3.6B). This point mutation increased the free energy of the mRNA secondary structure from -21.4 kcal mol^{-1} to -15.4 kcal mol^{-1} (Figure 3.6A).

A



B

Strain (Relevant genotype)		△G° (kcal.mol^{-1})	β-galactosidase activity (Miller Units) -Ara	β-galactosidase activity (Miller Units) +Ara
IQB847	hairpin WT	-21.4	2.67 ± 0.20	6.81 ± 0.30
IQB849	hairpin mut (C→A)	-15.4	3.01 ± 0.09	69.17 ± 5.48
IQB857	hairpin mut (C→A and G→T)	-19.7	1.55 ± 0.30	3.35 ± 0.26
IQB855	hairpin mut (C→G)	-18.2	2.80 ± 0.15	3.32 ± 0.23
IQB853	hairpin mut (C→G and G→C)	-23.3	1.17 ± 0.13	1.68 ± 0.20

Figure 3.6. Regulation of *araL* in *B. subtilis*. A. Site-directed mutagenesis in the 5'-end *araL*. The secondary structure of the *araABDLMNPQ-abfA* mRNA in the 5'-end *araL* region is depicted. An arrow highlights the mutated nucleotide (circled) located at the beginning of the *araL* coding region. The ribosome-binding site, rbs, is boxed. The free energy of the wild-type (WT) and mutated (mut C→A) secondary structures, calculated by DNAsis v 3.7 (©Hitachi Software Engineering Co. Ltd), are shown. **B. Expression from the wild-type and mutant *araL'-lacZ* translational fusion.** The *B. subtilis* strains IQB847 (Para-*araL'-'lacZ*) and IQB849 (Para-*araL'* (C→A) -*'lacZ*) were grown on C minimal medium supplemented with casein hydrolysate in the absence (non-induced) or presence (induced) of arabinose. Samples were analyzed 2 h after induction. The levels of accumulated β-galactosidase activity represent the average ± standard deviation of three independent experiments each performed with triplicate measurements.

In order to determine if the mutation introduced ectopically in strain IQB849 also influenced the protein production of AraL *in vivo* we introduced that same mutation in the *araL* locus, obtaining strain IQB869 (mut C→A, AraL Thr6Lys). Both the wild-type strain and IQB869 were grown in the same conditions and we were able to detect differences in AraL production in the presence and absence of inducer. Additionally, we analyzed the effect of the single point mutation in the absence of the regulator (AraR-null mutant background – strain IQB870) and the results show in increase in intracellular accumulation of AraL, which is not dependent of AraR (Figure 3.7):

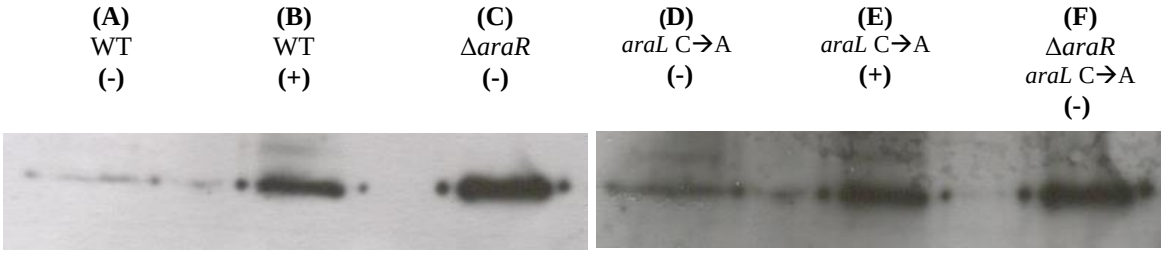


Figure 3.7. AraL accumulation in the cell determined by Western Immunoblot analysis. Equal amounts of the soluble fractions of cell extracts (10 µg of protein) obtained from *B. subtilis* cultures harboring a wild-type or mutant araL allele and grown in the absence (-) or presence (+) of inducer (arabinose) were prepared as described in Materials and Methods. (**A**) Wild-type strain $168T^+$ (**B**) Wild-type strain $168T^+$ in the presence of arabinose (**C**) *araR*-null mutant strain IQB215 (**D**) Strain IQB869, bearing the C→A, AraL Thr6Lys mutation (**E**) Strain IQB869, bearing the C→A, AraL Thr6Lys mutation in the presence of arabinose (**F**) Strain IQB870, bearing the C→A, AraL Thr6Lys mutation in an *araR*-null mutant background.

These results clearly show that the hairpin structure play an active role in the control of *araL* expression *in vivo*. The regulatory mechanism operating in this situation is most probably sequestration of the ribosome binding by the mRNA secondary structure, which consequently prevents translation, although premature transcription termination by early RNA polymerase release cannot be excluded. Translational attenuation by mRNA secondary structure comprising the initiation region is present in many systems of Bacteria, including *B. subtilis* (Grundy & Henkin, 2006). Due to the nature of the NagD family members displaying low specificity and catalytic activities and lacking clear boundaries defining physiological substrates, regulation at the genetic level was anticipated (Tremblay *et al.*, 2006).

The AraL enzyme encoded by the arabinose metabolic operon *araABDLMNPQ-abfA*, was previously shown dispensable for arabinose utilization in a strain bearing a large deletion comprising all genes downstream from *araD*. However this strain displayed some growth defects (Sá-Nogueira *et al.*, 1997). To confirm this hypothesis an in-frame deletion mutation in the *araL* gene was generated by allelic replacement, to minimize the polar effect on the genes of the *araABDLMNPQ-abfA* operon located downstream of *araL* (Figure 1.5, Figure 2.3. See Chapter II). The physiological effect of this knock-out mutation in *B. subtilis* (strain IQB832 Δ*araL*; Table 3.3) was assessed by determination of the growth kinetics parameters using glucose and arabinose as the sole carbon and energy source. In the

presence of glucose and arabinose the doubling time of the mutant, 49.7 ± 0.3 and 52.4 ± 0.1 min, respectively, is comparable to that of the wild-type strain, 46.6 ± 0.4 and 52.2 ± 0.5 min, respectively, indicating both the stability of the strain bearing the in-frame deletion, and that the AraL enzyme is not involved in L-arabinose utilization. The substrate specificity of AraL points to a biological function within the context of carbohydrate metabolism. The location of the *araL* gene in the arabinose metabolic operon together with the observation that AraL is active towards D-xylulose 5-phosphate, a metabolite resultant from arabinose catabolism, suggests that AraL has a wide substrate utilization profile and low substrate specificity, which is consistent with other members of HAD phosphatase superfamily, possibly helping the cell to get rid of phosphorylated metabolites that accidentally accumulate via stalled pathways (Guggisberg *et al.*, 2014; Kim *et al.*, 2015; Kuznetsova *et al.*, 2006; Pandya *et al.*, 2014). The arabinose operon is under the negative control of the transcription factor AraR and in an *araR*-null mutant expression of the operon is constitutive. In previous work, it was observed that addition of arabinose to an early-exponentially growing culture of this mutant resulted in immediate cessation of growth. It is believed that this effect could be due an increased intracellular level of arabinose, which would consequently cause an increase in the concentration of the metabolic sugar phosphates intermediates that are toxic to the cell (Sá-Nogueira & Mota, 1997). One hypothesis is that there is a possibility of the AraL playing a role in the dephosphorylation of substrates related with the arabinose metabolism, namely L-ribulose phosphate and/or D-xylulose phosphate. In addition, due to its capacity to catabolize other related secondary metabolites this enzyme needs to be regulated. Moreover, the *araL* gene is under the control of the operon promoter, which is a very strong promoter and basal expression in the absence of inducer is always present (Sá-Nogueira & Mota, 1997). The second level of regulation within the operon that operates in *araL* expression acts to drastically reduce the production of AraL.

Acknowledgments

We would like to thank Jörg Stülke for helpful discussions.

This work was partially funded by Fundação para a Ciência e a Tecnologia, grant no. [SFRH/BD/73109/2010 to L.G] and Fundação para a Ciência e a Tecnologia, POCI and FEDER grant no. [PPCDT/BIA-MIC/61140/2004 to I.S-N].

Chapter IV

Sugar Phosphates Toxicity in *Bacillus subtilis*

This chapter contains data to be published in:

"Toxicity of sugar phosphates in *Bacillus subtilis*" manuscript in preparation.

All experiments were performed by the author of this thesis except for the acquisition of NMR spectroscopy data, as well as identification of phosphorylated metabolites by NMR, which was performed with the collaboration of Dušica Radoš, PhD from the Cell Physiology and NMR group at ITQB António Xavier - NOVA.

Abstract

Accumulation of sugar phosphate is known to be toxic for the majority of prokaryotic and eukaryotic cells, however the mechanisms that underlie toxicity are yet to be fully understood. Conversely, little is known concerning the mechanisms utilized by the cells to adapt to sugar phosphate toxicity, or sugar phosphate stress.

Arabinose is one of the most abundant pentose in plant biomass. In *Bacillus subtilis* the ability to metabolize arabinose is dependent on three intracellular enzymes encoded by the *araA*, *araB* and *araD* genes that convert L-arabinose to D-xylulose 5-phosphate, which enters the pentose phosphate pathway. The transcription of these genes is induced by arabinose and negatively controlled by AraR, a repressor allosterically regulated by arabinose. In addition to the catabolism AraR also controls the uptake of arabinose. In an *araR*-null mutant, addition of arabinose to an exponentially growing culture results in immediate cessation of growth. The current hypothesis is that the observed bacteriostatic effect could be due to an increased intracellular level of arabinose, which consequently increases the concentration of the metabolic sugar phosphates intermediates that are toxic to the cell.

Here we used this model to address unanswered questions regarding the physiology of arabinose-induced stress, including the cellular signals and targets involved. Analysis of both wild-type and mutant strains by quantification of mRNA levels, phosphorylated metabolites, accumulation of cytotoxic methylglyoxal, and ATP depletion, suggests distinct mechanisms underlying toxicity. Furthermore, our study highlights the importance of a secondary metabolic pathway regulator in the growth of an industrial relevant species, like *B. subtilis*, and how its deletion negatively impacts the overall central carbon metabolism.

Introduction

The use of the *Bacillus* genus members is frequent in the industrial setting, either for the production of recombinant proteins, diverse secondary metabolites or as probiotic agent. As of 2004 about 60% of the commercially available enzymes are produced by *Bacillus* species (Westers *et al.*, 2004). *Bacillus subtilis* has been one of the Gram-positive pioneering hosts for recombinant protein production and it is still one of the preferred hosts for industrial bioprocesses due to both accumulated know-how on *B. subtilis* genetics and physiology and the increasing number of tools for genetic engineering adapted to this organism. The screening of different *B. subtilis* mutant strains is a promising route for the development of better production systems assisted by expanding systems metabolic engineering and synthetic biology tools (Küppers *et al.*, 2014).

The occurrence of metabolic bottlenecks either for protein expression or secondary metabolite production is well known (Breitling *et al.*, 2013; Westers *et al.*, 2004). Metabolomics is one of the tools researchers have been using to minimize and circumvent issues such as the accumulation of unwanted or toxic side products and intermediates, or the depletion of precursors necessary for certain reactions (Breitling *et al.*, 2013). Metabolic engineering and synthetic biology tools applied to strain enhancement must be used carefully, as bottlenecks causing either depletion of precursors or accumulation of toxic products can be a result of said enhancement, such as accumulation of sugar phosphate in cells.

The toxicity caused by the intracellular accumulation of phosphorylated sugars intermediates has been observed in bacteria, yeasts, insects, animal cells, and humans. Although not new to scientists working with bacterial cells, the mechanisms underlying this phenomenon are apparently different in each case-study or remain uncharacterized. Research conducted decades ago described toxicity associated with sugar phosphates and its accumulation in bacteria, namely in *Escherichia coli* (Englesberg *et al.*, 1962; Kadner *et al.*, 1992; Yarmolinsky *et al.*, 1959). However, the majority of these studies are focused on mutant strains that lack specific catabolic enzymes and as consequence accumulate phosphorylated sugar intermediates of the metabolic pathway. Methylglyoxal (MG) is a toxic byproduct known to accumulate in bacterial cells, when there is a sudden increase in carbohydrate uptake, causing an imbalance between flux through the upper branch of glycolysis and the capacity of the lower branch of the Embden-Meyerhof-Parnas pathway. Synthesis of MG is thought to function as an overflow mechanism that prevents accumulation of phosphorylated intermediates (Cooper & Anderson, 1970; Landmann *et al.*, 2011) Methylglyoxal, however, is highly cytotoxic, reacting with DNA, lipids and proteins and was shown to inhibit growth in bacteria including *B. subtilis* (Nguyen *et al.*, 2009). In the case of *E.coli* sensitivity to accumulation of phosphoglucose, researchers presented evidence that glucose-phosphate stress is the result of glycolytic intermediates depletion rather than accumulation of the known toxic byproduct methylglyoxal (Kadner *et al.*, 1992; Richards & Vanderpool, 2011; Richards *et al.*, 2013)

Fermentation of lignocellulosic material is industrially relevant and requires microorganisms that have the ability to use pentoses, such as xylose or arabinose (Garcia Sanchez *et al.*, 2010). Arabinose is one of the most abundant pentoses found in lignocellulosic feedstocks, like agricultural wastes and hardwoods, in the form of hemicellulose and pectin. The AraR protein is a negative regulator involved in arabinose-inducible expression of the *B. subtilis araABDLMNPQ-abfA* metabolic operon and of the *araR/araE* transcriptional unit. The presence of arabinose induces a conformational change in AraR such that recognition and binding to DNA is no longer possible and the *ara* genes are expressed (Mota *et al.*, 1999; Sá-Nogueira & Mota, 1997; Sá-Nogueira & Ramos, 1997). Addition of arabinose to a culture of a *B. subtilis araR*-null mutant causes immediate cessation of growth. A direct correlation between the cessation of growth of the *araR* mutant cells and the minimal L-arabinose concentration, which corresponds to full induction of the *araABDLMNPQ-abfA* operon promoter in the wild-type

strain, was established (Sá-Nogueira & Mota, 1997). Since the arabinose uptake system is under the negative control of AraR and the deficiency of repressor induces a two-fold increase in the expression from the metabolic operon promoter, it was hypothesized that the absence of AraR could lead to an intracellular increase of arabinose, consequently causing an increase in the concentration of the metabolic sugar phosphates intermediates which are toxic to the cell (Sá-Nogueira & Mota, 1997). In this study we examined this bacteriostatic effect to gain insight into the mechanisms that underlie bacterial cell toxicity of sugar-phosphates. By using quantification of mRNA levels, Nuclear Magnetic Resonance (NMR) analysis of phosphorylated metabolites, accumulation of cytotoxic methylglyoxal, and ATP depletion, we established a correlation between growth arrest of the arabinose-sensitive *araR*-null mutant strain and increased level of mRNA encoding the arabinose catabolic enzymes, as well as of several phosphorylated intermediates of the pentose phosphate pathway. Furthermore, associated with the bacteriostatic effect, an increased level of the cytotoxic compound methylglyoxal is detected together with intracellular depletion of ATP.

Materials and Methods

Substrates. All reagents were purchased from Sigma-Aldrich (St. Louis, MO, USA), with the exception of X-Gal, purchased from Apollo Scientific (Stockport, UK).

Bacterial strains and growth conditions. *Escherichia coli* strains XL1Blue (Stratagene, La Jolla, CA, USA) or DH5α (Gibco-BRL, Carlsbad, CA, USA) were used for molecular cloning work. *E. coli* strains were grown in LB medium (Miller, 1972). Ampicillin (100 $\mu g.mL^{-1}$), and tetracycline (12 $\mu g.mL^{-1}$) were added as appropriate. *B. subtilis* was grown in liquid LB medium, LB medium solidified with 1.6% (w/v) agar, with chloramphenicol (5 $\mu g.mL^{-1}$), kanamycin (10 $\mu g.mL^{-1}$), erythromycin (1 $\mu g.mL^{-1}$), spectinomycin (100 $\mu g.mL^{-1}$) and X-Gal (50 $\mu g.mL^{-1}$ or 80 $\mu g.mL^{-1}$) being added as appropriate. Specific growth rates of the wild-type and mutant *B. subtilis* strains were determined in C medium [70 mM K_2HPO_4 30 mM KH_2PO_4 5.25 mM $(NH_4)_2SO_4$ 0.5 mM $MgSO_4$ 0.1 mM ferric ammonium citrate 22 µg/mL (Pascal *et al.*, 1971)] supplemented with 1% (w/v) casein hydrolysate (Sá-Nogueira *et. al.* 1997). Arabinose 0.4% (w/v) or ribitol 0.4% (w/v) were added when appropriate. Cultures were grown on an Aquatron® Waterbath rotary shaker (Infors HT, Bottmingen, Switzerland), at 37 °C (unless stated otherwise) and 180 rpm, and OD_{600nm} was monitored in an Ultrospec™ 2100pro UV/Visible Spectrophotometer (GE Healthcare Life Sciences, Uppsala, Sweden).

DNA manipulation and sequencing. DNA manipulations were carried out as described previously by Sambrook *et. al.* (Joseph Sambrook & David W. Russel, 2001). Restriction enzymes were purchased from Thermo Fisher Scientific (Waltham, MA, USA) and used in accordance with the

manufacturer's instructions. DNA ligations were performed using T4 DNA Ligase from Thermo Fisher Scientific (Waltham, MA, USA). DNA was eluted from agarose gels with GFX Gel Band Purification kit (GE Healthcare Life Sciences, Uppsala, Sweden) and plasmids were purified using the Qiagen® Plasmid Midi kit (Qiagen, Hilden, Germany) or NZYMiniprep from NZYTech, Lda. - Genes and Enzymes (Lisboa, Portugal). DNA sequencing was performed with ABI PRIS Big-Dye Terminator Ready Reaction Cycle Sequencing kit (Applied Biosystems, Carlsbad, CA, USA). PCR amplifications were conducted using high-fidelity Phusion® DNA polymerase from Thermo Fisher Scientific (Waltham, MA, USA).

Plasmid constructions and in-frame deletions in *B. subtilis*. Plasmids pMO1 and pMO2 were used to introduce a small in-frame deletion of selected residues in the *araB* and *araD*, genes of *B. subtilis*, respectively. To construct pMO1, regions immediately upstream and downstream of *araB* were amplified by two independent PCR experiments, from chromosomal DNA of *B. subtilis* 168T$^+$, using primers ARA570 and ARA571 (PCR1) and ARA572 and ARA573 (PCR2). The products were joined by overlapping PCR, with primers ARA570 and ARA573 and the resulting fragment was cloned into pMAD (Arnaud *et al.*, 2004) digested with SmaI, yielding pMO1. pMO2 was obtained in the same way as pMO1, but using oligonucleotides ARA574 and ARA575 to amplify the upstream region of *araD* and oligonucleotides ARA576 and ARA577 to amplify the downstream region of *araD*. ARA574 and ARA577 were used to obtain the overlap PCR. The resulting fragment was cloned in pMAD using the SmaI site. Both plasmids were sequenced. pLG32 was obtained for the construction of an in-frame deletion of the *araLMNPQ* genes (Δ*araLMNPQ*). Regions immediately upstream and downstream of *araL* and *araQ* were amplified by two independent PCR experiments, from chromosomal DNA of *B. subtilis* IQB830 (Δ*araLM*, Chapter II). Primers used for PCR1 were ARA665 (introduction of a BglII site) and ARA448. Primers for PCR2 were ARA449 and ARA666 (introduction of a SalI site). This fragment was cloned in pMAD between the BglII and SalI sites. The in-frame deletion was sequenced. Plasmid pLG39 was used to insert an in-frame deletion of the methylglyoxal synthase gene (*mgsA*) in the *B. subtilis* chromosome. Primers ARA750 (BglII site) and ARA751 were used for PCR1 and oligos ARA752 and ARA753 (SalI site) were used for PCR2. ARA750 and ARA753 were used for the overlap PCR. The overlap PCR was then digested with BglII and SalI and the insert was cloned in the same sites of pMAD. Insertion of the in-frame deletions was sequenced. *B. subtilis* transformation with plasmid DNA was performed as previously described (Anagnostopoulos & Spizizen, 1960) and all strains were sequenced for the in-frame deletion.

Total RNA extraction *B. subtilis* strains were grown in C minimal medium supplemented with 1% (w/v) of casein hydrolysate in the presence and absence of 0.4% (w/v) arabinose, as previously reported (Sá-Nogueira & Mota, 1997). Cells were collected 2 h after the addition of arabinose and cell pellets were frozen at -80°C. Total RNA extraction was performed using the Absolutely RNA Miniprep

Kit (Agilent Technologies – Stratagene, La Jolla, CA, USA), according to the manufacturer's instructions. RNA quantification was determined on a Thermo Scientific NanoDrop™ ND-1000 Spectrophotometer (NanoDrop Technologies, LLC, USA).

Real-time PCR Experiments. One set of primers targeting *araB*, was used in the experiments (ARA581 and ARA582). An additional pair of primers targeting the housekeeping gene *16S rRNA* was used as internal control (ARA583 and ARA584). Primers were designed with the help of an internet-based interface, Primer3 (Rozen and Skaletsky, 2000), and are listed in Table 4.1. Real time amplification was performed in a Corbett Research Rotor-Gene RG6000 (Cambridgeshire, UK) using the SensiFAST™ SYBR® No-ROX One-Step Kit (Bioline Reagents Ltd., London, UK). Each reaction contained: 6.25 µl of the SensiFAST™ SYBR® No-ROX One-Step Kit, 0.5 µl of forward and reverse primers (100nM each), 0.125 µl of reverse transcriptase, 0.25 µl of RiboSafe RNase inhibitor and 40 ng of total RNA, to a final volume of 12.5 µl, adjusted with DEPC-treated H_2O. The reaction conditions were as follows: 10 min at 45 °C, followed by 2 min at 95 °C and then 40 cycles of 5 s at 95 °C and 10 s at 55 °C. PCR reaction for each sample was carried out in triplicate. NTC, no-template control (reagent alone without template) was included in each assay to detect any possible contamination of the PCR reagents. Following the final cycle, melting curve analysis was performed to assess the specificity in each reaction tube (absence of primer dimers and other nonspecific products). Results were analyzed using the Pfaffl mathematical model (Hellemans *et al.*, 2007; Pfaffl, 2001). Statistical analyses were performed with GraphPad Prism version 5.00 for Windows (GraphPad Software) using *Ct* values obtained from three independent assays. *p* values were determined using an unpaired two-tailed *t* test.

Methylglyoxal assay. To determine accumulation of methylglyoxal in the medium, *B. subtilis* cells were grown as described above. The methylglyoxal concentration was determined by the reaction of methylglyoxal with 2,4-dinitrophenylhydrazine as described previously (Cooper, 1974; Huynh *et al.*, 2000; Landmann *et al.*, 2011). The cells were removed by centrifugation and various volumes (120 µl or 320 µl) of the supernatant were incubated with 120 µl of 2, 4- dinitrophenyl hydrazine (10 $mg.ml^{-1}$ in 2 M HCl) for 15 min at 30 °C. Then, 560 µl of 10% (w/v) NaOH was added and the mixture was further incubated for 10 min at room temperature. Subsequently, the Abs_{550nm} was determined. An absorbance of 1.64 corresponds to 0.1 mmol of methylglyoxal present in the reaction mixture.

Cold Ethanolic Extracts of *B. subtilis* cultures. Cell were grown in C minimal medium supplemented with 1% (w/v) of casein hydrolysate in the presence and absence of 0.4% (w/v) L-arabinose, as described above. Cells were collected 2 h after the addition of arabinose and cell pellets were frozen at -80 °C. Each cell pellet as re-suspended in 70 mL of cold ethanol 70% (previously chilled at -20 °C), and then stirred vigorously in an ice bath for 30 minutes. The re-suspended cell pellets were centrifuged at 27000 *g* for 45 min at 4 °C using the SS34 rotor in a Sorvall RC-5C Plus centrifuge (Thermo Scientific®, Waltham, MA, USA). The supernatants (cell extracts) were transferred to round

bottom flasks and ethanol was evaporated using a rotary evaporator. The cell extracts were then lyophilized from 6 hours to overnight and re-suspended in MilliQ H_2O.

Identification of phosphorylated metabolites by NMR spectroscopy. Freeze-dried extracts were dissolved in MilliQH_2O and analyzed by ^{31}P- nuclear magnetic resonance (NMR). NMR spectra were acquired in a Bruker Avance II 500-MHz spectrometer (Bruker BioSpin GmbH) using a 5-mm ^{31}P-selective probe head at 25 °C and a standard Bruker program. The presence of each metabolite was confirmed by adding a small amount of standard solution to the extract. Calibration and quantification were performed using di-*myo*-inositol-1,3′-phosphate (DIP) of known concentration, with a resonance at -0.57 ppm (Rodrigues *et al.*, 2009).

Quantification of intracellular ATP. Intracellular ATP was quantified by using the BacTiter-Glo™ Microbial Cell Viability Assay Kit (Promega, Madison, WI, USA). *B. subtilis* strains were grown in C minimal medium supplemented with 1% (w/v) of casein hydrolysate in the presence and absence of 0.4% (w/v) arabinose, as described above. Cells were collected 2 h after the addition of arabinose and the OD_{600} was normalized to the lowest OD_{600} obtained. 100 µl of cells were mixed with an equal volume of the BacTiter-Glo™ reagent and incubated for 5 min, according to the manufacturer's instructions. The emitted luminescence was detected by using a *GloMax*® 96 Microplate Luminometer from Promega (Madison, WI, USA*)* and was expressed as relative luminescence units (RLU).

Table 4.1. List of plasmids used in this study

Plasmid	Relevant Construction	Source
pMAD	Plasmid used for allelic replacement in Gram-positive bacteria, *bla*, *erm*	Arnaud *et. al.* 2004
pMO1	In-frame deletion of 3 residues directly involved in the catalytic mechanism of the L-ribulokinase, AraB	Present study
pMO2	In-frame deletion of 5 residues directly involved in the catalytic mechanism of the L-ribulose 5 phosphate epimerase, AraD	Present study
pLG32	In-frame deletion of the *araLMNPQ* genes (Δa*raLMNPQ*)	Present study
pLG39	In-frame deletion of the methylglyoxal synthase gene (Δ*mgsA*)	Present study

Table 4.2. List of oligonucleotides (Primers) used in this study. Restriction sites are underlined.

Primers	Sequence (5' → 3')
ARA448	GATAAAGTACTTTTCGAAAAAAGTCATTTTTTTCATCTGCGTTACCCCTTC
ARA449	GAAGGGGTAACGCAGATGAAAAAAATGACTTTTTTCCAAAAGTACTTTATC
ARA570	CCTTCCTATTGACAGCAGGG
ARA571	GACGCGCGGTACCGAAACTACATTGGCAACAACCGCAAC
ARA572	GTTGCGGTTGCCAATGTAGTTTCGGTACCGGCGGTC
ARA573	CATACGCATAAATCTGCATG
ARA574	GAAAACCATGTCATGAAGCG
ARA575	AAATTCCGTCGGAAGCTGCTGGTTTAAGCGAGCCTTCGACGACC
ARA576	GGTCGTCGAAGGCTGGCTTAAACCTACCCATGTTTATCTATAAA
ARA577	GTGATTATTCACGAGCACAC
ARA581	TGGAACACACCATTCCGTCT
ARA582	TGCGGCTCCTCTTCATATTC
ARA583	TCGCGGTTTCGCTGCCCTTT
ARA584	AAGTCCCGCAACGCGCGCAA
ARA665	GGAGATCTCGGCATTCACCGT
ARA666	GGCAGTCGACTGTTTGAGCTG
ARA750	TTGCATTAGAGATCTGAGTCCGTC
ARA751	TTATACATTCGGCTCTTCTCCCCGAGCAATTTTCATTGTTTATCCCCC
ARA752	GGGGGATAAACAATGAAAATTGCTCGGGGAGAAGAGCCGAATGTATAA
ARA753	GTAGGCGTCGACGCTTTGTTTCTT

Table 4.3. List of strains used in this study. Arrows indicate transformation and point from donor DNA to recipient strain

Strain	Relevant Genotype	Source
***E. coli* Strains**		
DH5α	*fhuA*2 Δ(*argF-lacZ*) U169 *phoA glnV44* Φ80 Δ(*lacZ*)*M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	Gibco - BRL
XL1 Blue	(*recA1 endA1 gyrA96 thi-1 hsdr17 supE44 relA1 lac* [F' *proAB lacI*^{q} ZΔM15 Tn10 (*TetR*)]	Stratagene
***B. subtilis* strains**		
168T^{+}	Prototroph (wild-type)	F.E. Young
IQB206	Δ*araLMNPQ-abfA*::*spc*	Sá-Nogueira *et. al.* 1997
IQB565	Δ*araLMNPQ-abfA*::*spc* Δ*araR*::*km*	Sá-Nogueira, unpublished results
IQB861	*araB*$^{-}$	pMO1 → 168T^{+}
IQB862	Δ*araR*::*km araB*$^{-}$	pMO1 → IQB215
IQB865	*araD*$^{-}$	pMO2 → 168T^{+}
IQB866	Δ*araR*::*km araD*$^{-}$	pMO2 → IQB215
IQB871	Δ*araLMNPQ*-in-frame	pLG32 → 168T^{+}
IQB872	Δ*araLMNPQ*-in-frame Δ*araR*::*km*	pLG32 → IQB215
IQB875	Δ*mgsA*	pLG39 → 168T^{+}
IQB876	Δ*mgsA* Δ*araR*::*km*	pLG39 → IQB215

Results and Discussion

Analysis of arabinose sensitivity in different *Bacillus subtilis* mutant strains

In previous studies, to characterize the role of the *araR* gene in the catabolism of L-arabinose, an *araR*-null mutant strain was constructed (IQB215 Δ*araR*::*km*). This strain was unable to grow on minimal medium with arabinose as the sole carbon and energy source and presented constitutive expression of the arabinose metabolic genes, *araA*, *araB*, and *araD* on rich medium (Sá-Nogueira & Mota, 1997). In liquid minimal C medium supplemented with casein hydrolysate, in which *B. subtilis* cell are utilizing casamino acids as source of carbon and energy, addition of arabinose during early-exponential-phase growth causes immediate cessation of growth. A direct correlation between the growth arrest of the Δ*araR* mutant cells and the minimal arabinose concentration in the medium was established, which corresponds to full induction of the *araABDLMNPQ-abfA* operon promoter in the wild-type strain (Sá-Nogueira & Mota, 1997). Moreover, a strain bearing a large deletion downstream

of *araD*, which replaced the sequences of the *araLMNPQ* genes and the majority of the *abfA* gene with the spectinomycin resistance gene (strain IQB206, Δ*araLMNPQ-abfA*::*spc)*, was able to grow on arabinose as the sole carbon an energy source. However, it was noticed that the specific growth rate of this strain, when fully induced by arabinose, was slightly affected by the large deletion when compared to the wild-type strain (Sá-Nogueira *et. al.*, 1997; Table 4.4).

Interestingly, it was shown that the toxic effect of arabinose was suppressed by introduction of the insertion-deletion mutation in an *araR*-null mutant background (strain IQB565, Δ*araR*::*km* Δ*araLMNPQ-abfA*::*spc*) (Sá-Nogueira *et. al.* unpublished results; Table 4.4). These observations, together with the data obtained in Chapter II, led us to speculate that the growth cessation phenotype observed in the Δ*araR* strain was most likely due to accumulation of phosphorylated sugars derived from arabinose metabolism, and that recovery from growth arrest in the double mutant strain was probably caused by destabilization of mRNA of the operon, which was also previously hypothesized to explain the decrease in specific growth rate in the insertion-deletion mutant strain IQB206 (Δ*araLMNPQ-abfA*::*spc*), when compared to the wild-type strain in the presence of arabinose.

To test this hypothesis we constructed new strains by generating a marker-free in-frame deletion of genes *araLMNPQ* in the wild-type and *araR*-null mutant background, strains IQB871 (Δ*araLMNPQ* in-frame) and IQB872 (Δ*araR*::*km* Δ*araLMNPQ* in-frame), respectively. The growth kinetic parameters of the strains were determined in liquid minimal C medium supplemented with casein hydrolysate in the presence and absence of arabinose and the results are summarized in Table 4.4. The strain bearing the in-frame deletion IQB871 (Δ*araLMNPQ* in-frame) displayed a doubling time very similar to the wild-type. Furthermore, contrary to that observed in the strain bearing the insertion-deletion mutation (strain IQB565, Δ*araR*::*km* Δ*araLMNPQ-abfA*::*spc*), in strain IQB872 (Δ*araR*::*km* Δ*araLMNPQ* in-frame) the in-frame deletion did not revert the growth arrest phenotype observed in an *araR*-null mutant background (Table 4.4). These results are consistent with instability of the message of the operon, namely of the *araA*, *araB* and *araD* mRNA, caused by the insertion of the spectinomycin resistance gene. Thus to confirm this results the mRNA levels of these strains were determined in the presence and absence of arabinose.

Table 4.4. **Growth kinetics of B. *subtilis* wild-type and mutant strains in complex medium.** Cells were grown in C minimal medium supplemented with casein hydrolysate in the presence of arabinose or ribitol and in the absence of sugar. Results are the averages of three independent assays and their respective standard deviations.

	Doubling Time (min)		
Strain	**0.4% arabinose**	**0.4% ribitol**	**No sugar**
168T$^+$ (wild-type)	42.84 ± 1.81	53.08 ± 6.00	46.79 ± 2.14
IQB215 (Δ*araR*::*km*)	No growth	No growth	46.44 ± 3.93
IQB565 (Δ*araR*::*km* Δ*araLMNPQ*::*spc*)	56.52 ± 2.65	No growth	50.08 ± 3.66
IQB861 (*araB* $^-$)	47.23 ± 1.04	46.81 ± 2.33	47.27 ± 4.01
IQB862 (Δ*araR*::*km araB* $^-$)	48.99 ± 0.24	47.10 ± 1.84	49.45 ± 2.13
IQB866 (Δ*araR*::*km araD* $^-$)	No growth	No growth	48.84 ± 5.29
IQB865 (*araD* $^-$)	No growth after 30 min	46.86 ±3.00	46.87 ± 3.00
IQB872 (Δ*araR*::*km* Δ*araLMNPQ* in-frame)	No growth	No growth	51.05 ± 5.01
IQB871 (Δ*araLMNPQ* in-frame)	43.33 ± 5.94	44.87±1.93	48.05 ± 5.64
IQB875 (Δ*mgsA*)	51.49 ± 3.43	61.35 ± 4.23	63.81 ± 1.37
IQB876 (Δ*araR*::*km* Δ*mgsA*)	No growth	No growth	67.23 ± 4.83

mRNA levels of the operon were relatively quantified by real-time PCR using the *araB* gene. The analysis showed that in the presence of arabinose the levels of messenger RNA in the strains bearing the insertion-deletion mutation decrease considerably when compared to the wild-type strain fully induced (Table 4.5).

Table 4.5. **Measurement of *araB* mRNA levels in different *B. subtilis* strains by qRT-PCR.** The results represent the fold-change of the expression in the target conditions versus the control conditions. Cells were grown in minimal C medium supplemented with 1% (w/v) of casein hydrolysate in the presence (+) and absence (-) of arabinose. Statistical analyses were performed with GraphPad Prism version 5.00 for Windows (GraphPad Software) using C*t* values obtained from three independent assays. *p* values were determined using an unpaired two-tailed *t* test (ns, non-significant difference; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

	Target Situation				
Control Situation	IQB215 Δ*araR*::*km* (-)	IQB206 Δ*araLMNPQ-abf*A::*spc* (+)	IQB565 Δ*araR*::*km* Δ*araLMNPQ-abf*A::*spc* (+)	IQB871 Δ*araLMNPQ* (+)	IQB872 Δ*araR*::*km* Δ*araLMNPQ* (-)
$168T^+$ (+)	$3.78 \pm 0.19^{***}$	$0.65 \pm 0.127^{*}$	$0.025 \pm 0.01^{**}$	$0.90 \pm 0.21^{*}$	$4.84 \pm 1.26^{**}$

Growth kinetic experiments in the presence of arabinose together with real-time PCR analysis of the mRNA levels allowed us to conclude that the recovery from growth arrest displayed by strain IQB565 (Δ*araR*::*km* Δ*araLMNPQ-abfA*::*spc*) in the presence of this sugar was due to a decrease in the concentration of arabinose catabolic enzymes, hence decreasing the levels of phosphorylated sugar intermediates of arabinose catabolism.

The wild-type *B. subtilis* strain $168T^+$ is not able to utilize the pentose alcohol ribitol as sole carbon an energy source; however, it can be phosphorylated by the ribulokinase AraB, originating ribitol phosphate (ribitol-P). We determined the growth kinetics parameters of the mutant strains in the presence of ribitol in the same conditions as described above. Addition of 0.4% ribitol resulted in immediate growth arrest in the *araR*-null mutant strains IQB565 (Δ*araR*::*km* Δ*araLMNPQ-abfA*::*spc*) and IQB872 (Δ*araR*::*km* Δ*araLMNPQ*-in-frame) (Table 4.4), similarly to that observed in strain IQB215 (Δ*araR*::*km*), thus sustaining the hypothesis of toxicity caused by sugar-phosphate accumulation.

In addition, we constructed and tested strains with small in-frame deletions in the catabolic genes of the operon, namely *araB* and *araD*, targeting residues directly involved in the catalytic mechanism of the enzymes encoded by these genes, preventing their ribulokinase and epimerase activity, respectively. The mutations were generated by in-frame marker-free deletions in order to minimize polar

effects in the downstream genes yielding the following strains, IQB861 (*araB*$^-$), IQB862 (Δ*araR*::*km araB*$^-$), IQB865 (a*raD*$^-$) and IQB866 (Δ*araR*::*km araD*$^-$). For the *araB*$^-$ strains we introduced a small mutation in the *araB* gene, which deleted 3 residues directly involved in the catalytic mechanism of the L-ribulose kinase. In the resulting strain, this mutation reverted the toxic effect of arabinose and ribitol addition in the *araR*-null mutant. Both strains, *araB* $^-$ and Δ*araR*::*km araB*$^-$ were able to grow in the presence of both sugars (Table 4.4). In parallel, the introduction of a small mutation in the *araD* gene that deleted 5 residues directly involved in the catalytic mechanism of the L-ribulose 5-phosphate epimerase, resulted in growth arrest in both strains in the presence of arabinose. However, addition of arabinose to strain IQB865 (*araD*$^-$) did not cause immediate growth arrest, as observed in the *araR*-null mutant strain IQB866 (Δ*araR*::*km araD*$^-$) – cessation of growth occurs approximately 30 minutes after sugar addition to the medium. The presence of ribitol does not affect growth, since it is not an inducer of the arabinose-responsive promoter. In the *araR*-null strain, Δ*araR*::*km araD*$^-$, addition of both ribitol and arabinose result in immediate growth arrest, as a result of an increase in AraA, AraB and AraD concentration, caused by the de-repression of the Para promoter in the absence of the AraR (Table 4.4). These results further support that the bacteriostatic effect is due to the intracellular accumulation of sugar-phosphates.

Production of methylglyoxal by mutant *B. subtilis* strains

Methylglyoxal (also known as 2-oxopropanal or as pyruvaldehyde) is a typical cellular 2-oxoaldehyde and a known by-product of metabolic pathways in living organisms. Toxicity of methylglyoxal accumulation is a well-established phenomenon (Chandrangsu *et al.*, 2014; Cooper & Anderson, 1970; Inoue & Kimura, 1995; Kadner *et al.*, 1992; Landmann *et al.*, 2011; Nguyen *et al.*, 2009; Subedi *et al.*, 2008). When there is an increase in the uptake of carbohydrates into the cell, an asymmetry is caused between flux through the upper branch of the Embden-Meyerhof-Parnas pathway (glycolysis) and the capacity of its lower branch (tricarboxylic acid cycle), and synthesis of methylglyoxal is thought to function as an overflow mechanism that prevents accumulation of phosphorylated intermediates (Cooper & Anderson, 1970; Inoue & Kimura, 1995; Kadner *et al.*, 1992). The way methylglyoxal (MG) is further processed in the cell, resulting in D-lactate production or being excreted to the medium, was recently investigated in *B. subtilis* (Chandrangsu *et al.*, 2014).

Since the results discussed above pointed towards an increase in phosphorylated metabolites as a consequence of the upregulation of the *araABD* genes in the absence of the repressor AraR, most probably the accumulation of dihydroxyacetone phosphate (DHAP) originating from the pentose phosphate pathway is directed to MG production by action of the methylglyoxal synthase (MgsA) enzyme. Consequently, accumulation of MG might be playing a role in growth arrest of strains lacking AraR. To test this hypothesis we measured MG production in the wild-type and mutant strains grown as described above in the presence of arabinose, or ribitol, and in the absence of sugars (Fig. 4.1).

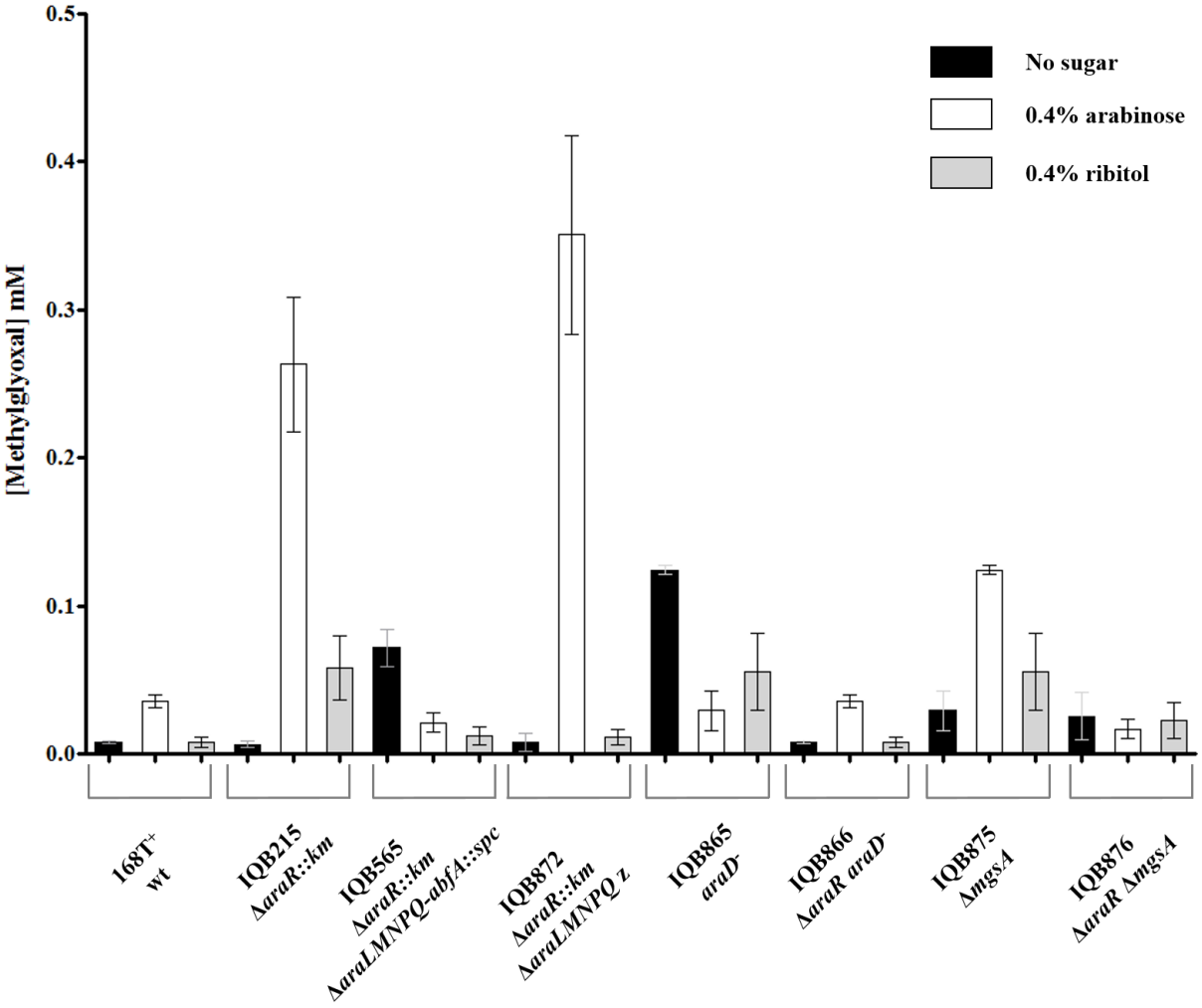


Figure 4.1. Methylglyoxal production in *B. subtilis* strains. Methylglyoxal presence in the medium was measured in the absence of sugar (white bars), in the presence of arabinose (black bars) and in the presence of ribitol (grey). Error bars represent the standard deviation of at least three independent experiments and differences were considered statistically significant. Unpaired Two-tailed *t* test and GraphPad Prism version 5.00 for Windows (GraphPad Software) were used for statistical analysis.

The data show that accumulation of MG in the culture medium is higher in strains lacking AraR, thus strains producing more MG cease growth upon arabinose addition to the medium and display higher levels of mRNA for the arabinose catabolic genes in the same conditions (2 h after arabinose addition). Moreover the amount of MG measured in these cases correlates to results obtained by other authors reporting growth arrest of *B. subtilis* cells at similar concentrations of MG (Landmann *et al.*, 2011). As expected, strains unable to synthesize DHAP, such as IQB865 (a*raD*$^-$) and IQB866 (Δ*araR*::*km araD*$^-$) do not present significant accumulation of MG, as well as strains lacking the MgsA gene, IQB875 (Δ*mgsA*) and IQB876 (Δ*araR*::*km* Δ*mgsA*). The results obtained in the determination of the growth kinetics parameters (Table 4.4) showed that strains IQB865 (a*raD*$^-$) IQB866 (Δ*araR*::*km araD*$^-$) and

IQB876 (Δ*araR*::*km* Δ*mgsA*) are unable to grow in the presence of arabinose, however no significant accumulation of MG in the growth medium is detected when compared to the Δ*araR*::*km* strains bearing an intact arabinose catabolic pathway (IQB215 and IQB872). Most interesting is to verify the statistical significance of the decrease of accumulated MG in strain IQB565 vs IQB872 (insertion-deletion and in-frame deletion).

Accumulation of MG in the presence of ribitol is not significant, which is in agreement with the non-catabolization of ribitol phosphate by the PPP in all strains derived from the wild-type 168T$^+$. Thus, in strains unable to grow in the presence of ribitol, namely IQB215 (Δ*araR*::*km*), (IQB565 (Δ*araR*::*km* Δ*araLMNPQ-abfA*::*spc*), IQB872 (Δ*araR*::*km* Δ*araLMNPQ*-in-frame) and IQB876 (Δ*araR*::*km* Δ*mgsA*) growth arrest is independent from MG accumulation, and may be due to depletion of cellular phosphate upon accumulation of phosphosugars, like L-ribulose 5-phosphate (in strains a*raD*$^-$ and Δ*araR*::*km araD*$^-$) or ribitol phosphate.

Accumulation of phosphosugars is a major driving force of arabinose toxicity in the absence of AraR

Then we analyzed cell extracts of *B. subtilis* by ^{31}P-nuclear magnetic resonance (NMR), to identify which metabolites were accumulating in the *araR*-null mutant strains when compared to the wild-type strain. Cold ethanolic cell extracts of cultures grown in identical conditions as described above in the previous analysis were used in the study to detect if those metabolites correlate, in any way, with metabolites from the arabinose degrading pathway, the pentose phosphate pathway or even from glycolysis.

The acquired ^{31}P-NMR spectra display a significant difference in peak number and typology in the wild-type and the mutant strain IQB215 Δ*araR*::*km* when arabinose is present (Panels A and C, respectively, Figure 4.2). The occurrence of phosphate monoesters (phosphosugars) in the wild-type strain in the presence and absence of arabinose (panels A and B, respectively, Figure 4.2) indicated that full induction of the arabinose catabolizing genes does not generate accumulation of phosphorylated metabolites. Yet, spectra analysis of the deregulated *araR*-null mutant strain in the presence and absence of arabinose (panels C and D, respectively, Figure 4.2) showed a substantial difference, as several peaks corresponding to phosphorylated metabolites are identifiable in the presence of arabinose.

The location of the peaks in the spectra (in ppm) allowed us to conclude that phosphorylated sugars are indeed accumulating in the mutant strain. Furthermore, we were able to identify the majority of the peaks present in the phosphate monoester region as phosphosugars belonging to both the pentose phosphate pathway and glycolysis (Figure 4.3). It was possible to assign the existence of L-ribulose 5-phosphate (R5P), D-xylulose 5-Phosphate (X5P), fructose 1,6-bisphosphate (FBP), fructose

6-phosphate (F6P), glucose 6-phosphate (G6P), 6-phosphogluconate (6PG) in the Δ*araR*::*km* mutant strain sample obtained in the presence of arabinose (Figure 4.3).

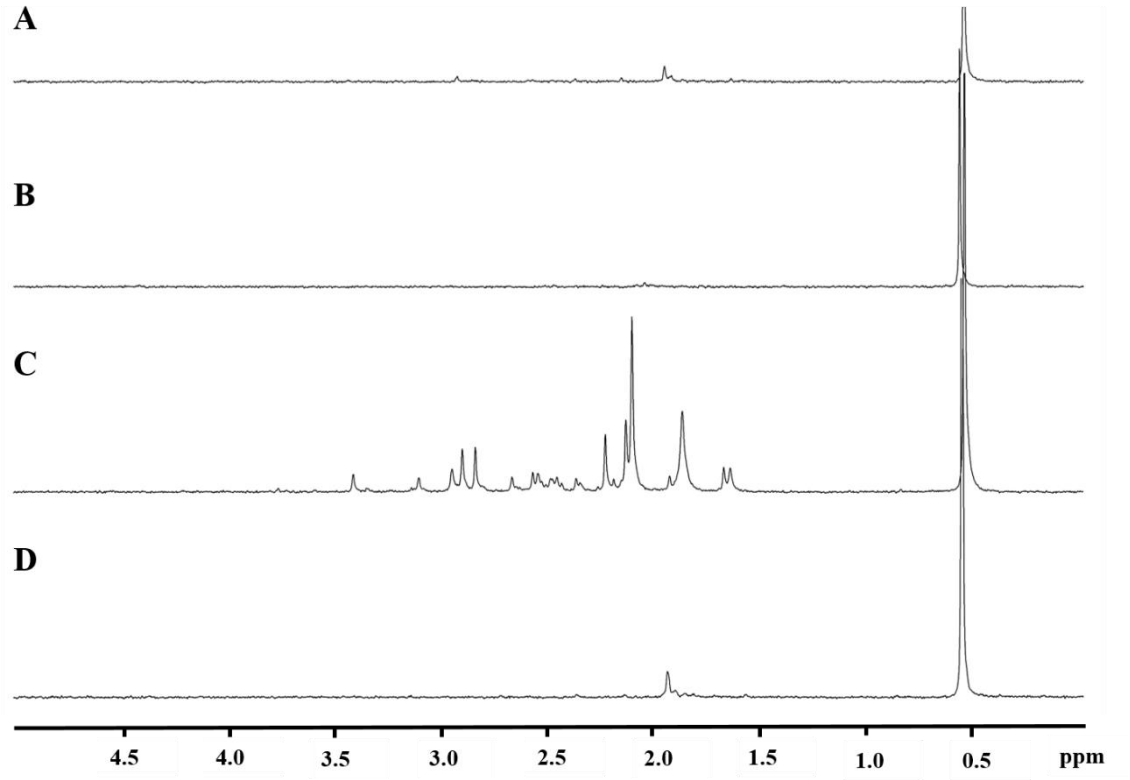


Figure 4.2. ^{31}P-NMR analysis of *B. subtilis* cell extracts. Freeze-dried extracts were dissolved in MilliQH_2O and analyzed by ^{31}P- nuclear magnetic resonance (NMR). NMR spectra were acquired in a Bruker Avance II 500-MHz spectrometer. On the left, NMR spectra of the wild-type strain 168T$^+$ acquired in the presence (**A**) and in the absence (**B**) of arabinose. On the right, NMR spectra of the mutant *araR*-null strain IQB215 acquired in the presence (**C**) and in the absence (**D**) of arabinose. Accumulation of several phosphate monoesters, between 1.5 and 3.5 ppm, corresponding to phosphorylated sugars can be seen in **C**, when compared to **A**, **B** or **D**).

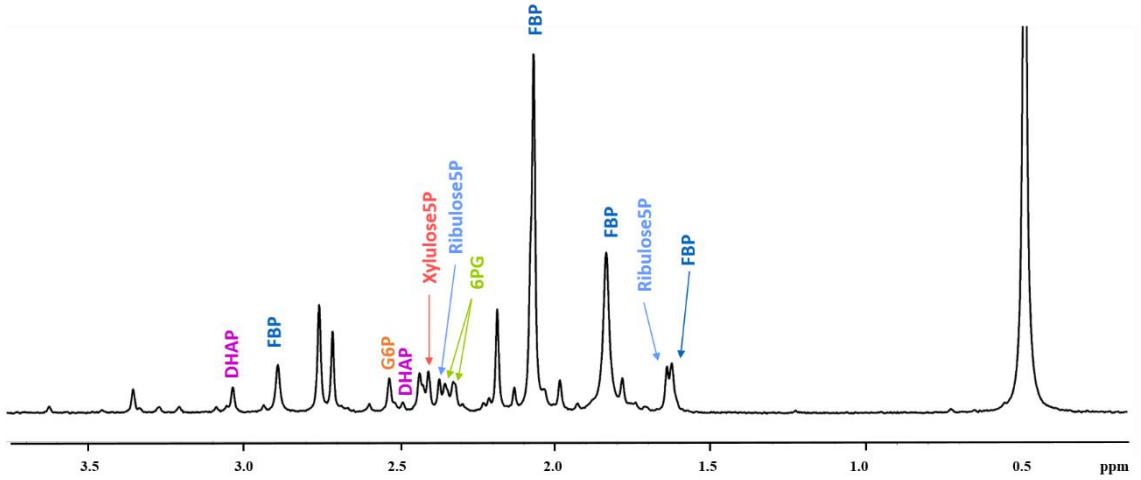


Figure 4.3. Metabolite Identification of Phosphorylated Sugars by ^{31}P-NMR. ^{31}P-NMR spectrum of freeze-dried extract of the mutant *araR*-null strain IQB215 acquired in the presence of arabinose. P*i* is visible at 0.5 ppm, while the identified metabolites all fall in the phosphate monoester region, between 1.5 and 3.5 ppm. Arrows point towards the phosphosugars, namely L-ribulose 5-phosphate (Ribulose5P), D-xylulose 5-Phosphate (Xylulose5P), fructose 1,6-bisphosphate (FBP), glucose 6-phosphate (Glucose6P), 6-phosphogluconate (6PG). Although identified through spiking, fructose 6-phosphate (F6P) is not shown here, as its signal is masked by the stronger FBP signal.

Other substrates tested, namely glyceraldehyde 3-phosphate, 2-phosphoglycerate, ribose 5-phosphate, phosphoenol pyruvate and acetyl phosphate, were not found in the mutant strain.

The largest peak found in sample corresponds to accumulation of FBP and could be a result of the interconversion of several metabolites found in both the PPP and glycolysis (Figure 4.4). This observation correlates with the results we obtained for methylglyoxal accumulation in strains depleted of AraR (Figure 4.1), as high FBP levels lead to an increase in MG production (Landmann *et al.*, 2011).

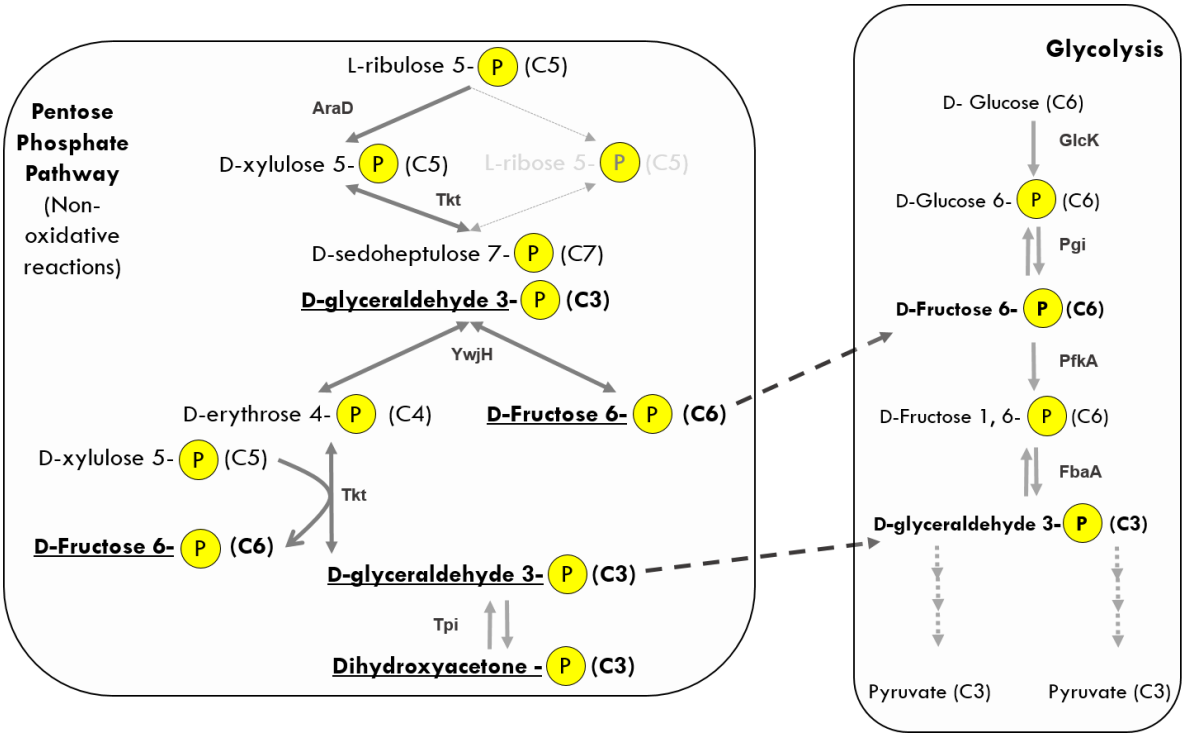


Figure 4.4. Interconnection between the pentose phosphate pathway and glycolysis. Dashed arrows indicate entry points of PPP metabolites in the glycolytic pathway, namely F6P and GA3P. Also indicated are the enzymes that catalyze PPP and glycolysis reactions: AraD - L-ribulose 5-phosphate epimerase; Tkt – transketolase; YwjH – transaldolase; Tpi – triose phosphate isomerase; GlcK – glucose kinase; Pgi - glucose-6-phosphate isomerase; PfkA – phosphofructokinase; FbaA – fructose 1,6-bisphosphate aldolase.

Imbalance of ATP in the cell plays a role in arabinose toxicity in *B. subtilis*

ATP is the universal currency of free energy in biological systems, as it is the free-energy-donor in most energy-requiring processes, due to its two phosphoanydre bonds, which release a large amount of free energy when ATP is hydrolyzed. ATP is generated in the breakdown of sugars as the oxidation energy of carbon atoms is transformed into phosphoryl transfer potential.

However, phosphorylated metabolites arising from the PPP are phosphorylated using ATP (namely L-ribulose 5-phosphate), thus ATP molecules are being invested in the phosphorylation of

sugars while less ATP molecules are generated via the lower branch of glycolysis, because metabolism is stalled at the production of FBP in strains that cease growth upon arabinose addition. So, the uncontrolled phosphorylation of metabolites resulting from arabinose catabolism may function as an ATP sink depleting the available ATP pool. To evaluate intracellular ATP depletion, we measured the relative intracellular ATP using the BacTiter-Glo™ Microbial Cell Viability Assay Kit (Promega), which allows generation of a luminescent signal proportional to the amount of ATP present in the sample. The results obtained in the wild-type and mutant strains grown in the presence of arabinose or ribitol, and absence of sugar are summarized in Figure 4.5.

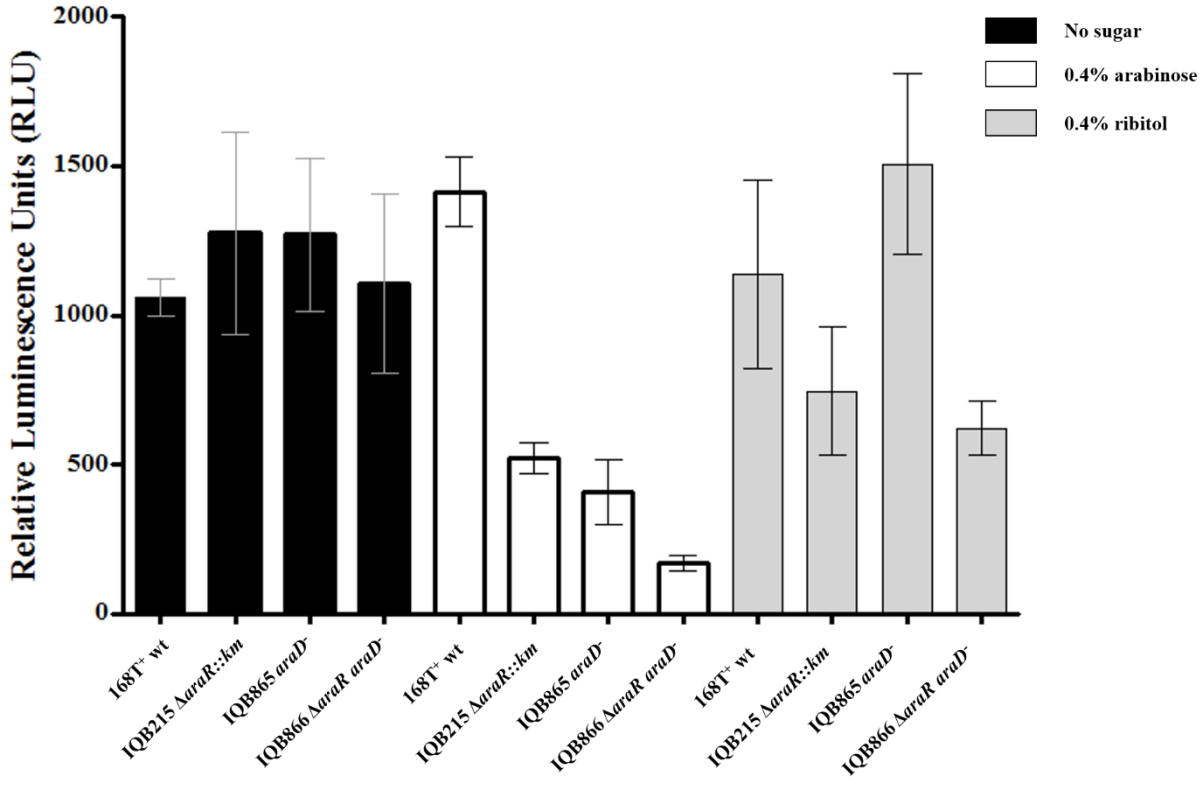


Figure 4.5. Relative ATP quantification in *B. subtilis* cell cultures. The bars represent the relative luminescence units (RLU) and are means of data obtained from at least three independent experiments each conducted in triplicate. Cell number was normalized previous to assay. Error bars represent the standard deviation. Unpaired two-tailed *t* test and GraphPad Prism version 5.00 for Windows (GraphPad Software) was used for statistical analysis.

Our results show that in the presence of arabinose the ATP levels of the *araR*-null mutant strains (IQB215 and IQB866) and the *araD*- strain (IQB865) decrease when compared to the values obtained in the absence of sugar. Moreover, there is a significant drop in intracellular ATP detected in arabinose-sensitive strains when compared to the wild-type strain. This could be caused, as we hypothesize above, by a stalled metabolism in FBP, which originates from DHAP and G3P by action of fructose-1,6-bisphosphate aldolase (FbaA, Figure 4.4), a constitutively expressed enzyme from glycolysis (Ludwig *et al.*, 2001). FBP is produced simultaneously from phosphorylation of F6P by the

phosphofructokinase enzyme (Pfk, Figure 4.4) that has an increased activity when the ATP/AMP ratio decreases (Berg *et al.*, 2011; Byrnes *et al.*, 1994). When the concentration of ATP lowers, Pfk detects an energy deficit in the cell and phosphorylates FBP to proceed with ATP and NADH generation via the lower branch of glycolysis and TCA cycle allowing growth. In addition, growth cessation in the *araR*-null mutant strains caused by ribitol is also associated with a decrease in ATP concentration, although intracellular concentration is not as low as detected in the presence of arabinose. Although ribitol is not catabolized by the *B. subtilis* strains used in this study, ribitol phosphate is formed by the action of ribulokinase constitutively expressed in the *araR*-null mutant strains, therefore contributing to ATP depletion.

Growth arrest phenotype in an *araR*-null mutant strain of *B. subtilis* is caused by multiple factors

The data indicate that there is no one single or simple cause that leads to growth arrest in an *araR*-null mutant strain in the presence of arabinose. The mutant strain IQB215 (Δ*araR*::*km*) displays arabinose-sensitivity and accumulates not only MG but also several phosphorylated sugars, namely FBP, and displays a decrease in intracellular ATP when compared to cells grown in the absence of sugar, or compared to the wild-type strain grown in the presence of arabinose. Decrease on intracellular ATP content is concomitant with an increase in phosphofructokinase activity in the cell (Berg *et al.*, 2011; Byrnes *et al.*, 1994), which phosphorylates F6P to FBP. Additionally, an increase in FBP levels in the cell leads to phosphorylation of Crh, a paralogue protein of HPr, responsible for controlling the methylglyoxal bypass in the MG pathway (Landmann *et al.*, 2011). The enzyme MgsA converts DHAP to MG initiating a glycolytic bypass that prevents the deleterious accumulation of phospho-sugars under carbon overflow conditions. The non-phosphorylated form of Crh interacts with MgsA and inhibits its activity. Phosphorylation of Crh impairs its binding to MgsA, thus coupling high activity of the MG pathway to high intracellular levels of FBP, reflecting overflow of carbon sources. Thus, bacteria produce a toxic substance to get rid of phospho-sugar stress and, on the other hand, MgsA activity restores inorganic phosphate levels.

Strains with an impaired arabinose catabolic pathway, namely *araD*$^{-}$ strains, IQB865 (*araD*$^{-}$) and IQB866 (Δ*araR*::*km araD* $^{-}$) do not accumulate MG in the growth medium but still display a decrease in intracellular ATP content in the presence of arabinose. This result could be due to ATP molecules being invested in the phosphorylation L-ribulose and no ATP being generated via the lower branch of glycolysis, mainly from 1,3-bisphosphoglycerate, because the metabolism has come to a stop at L-ribulose 5-phosphate. As such, growth arrest could be attributed mostly to the intracellular decrease of ATP, which can have major implications in the redox balance of the cell disrupting the NADH/NAD^{+} ratio.

Similarly, the presence of ribitol causes immediate cessation of growth of strains bearing a deletion in *araR* but does not result in MG accumulation in the medium, although an intracellular ATP decrease is observed (Table 4.4, Figure 4.1 and Figure 4.4, respectively). Since ribitol-phosphate accumulates in the cell (data not shown) and is not further metabolized by the cells, these observations might be explained as discussed above in the case of strains unable to catabolize L-ribulose phosphate.

Additionally, we observed that the effect of arabinose-sensitivity, growth arrest phenotype, is bacteriostatic rather than bactericidal implying that the cells eventually find mechanisms to cope with an overflow of carbon source that originates a variety of toxic stimuli. We performed growth experiments in the same conditions described above but incubation proceeded for an extended period and the results obtained with the *araR*-null strain showed that the toxic effect of arabinose and ribitol was, in fact, transient rather than permanent (Figure 4.6).

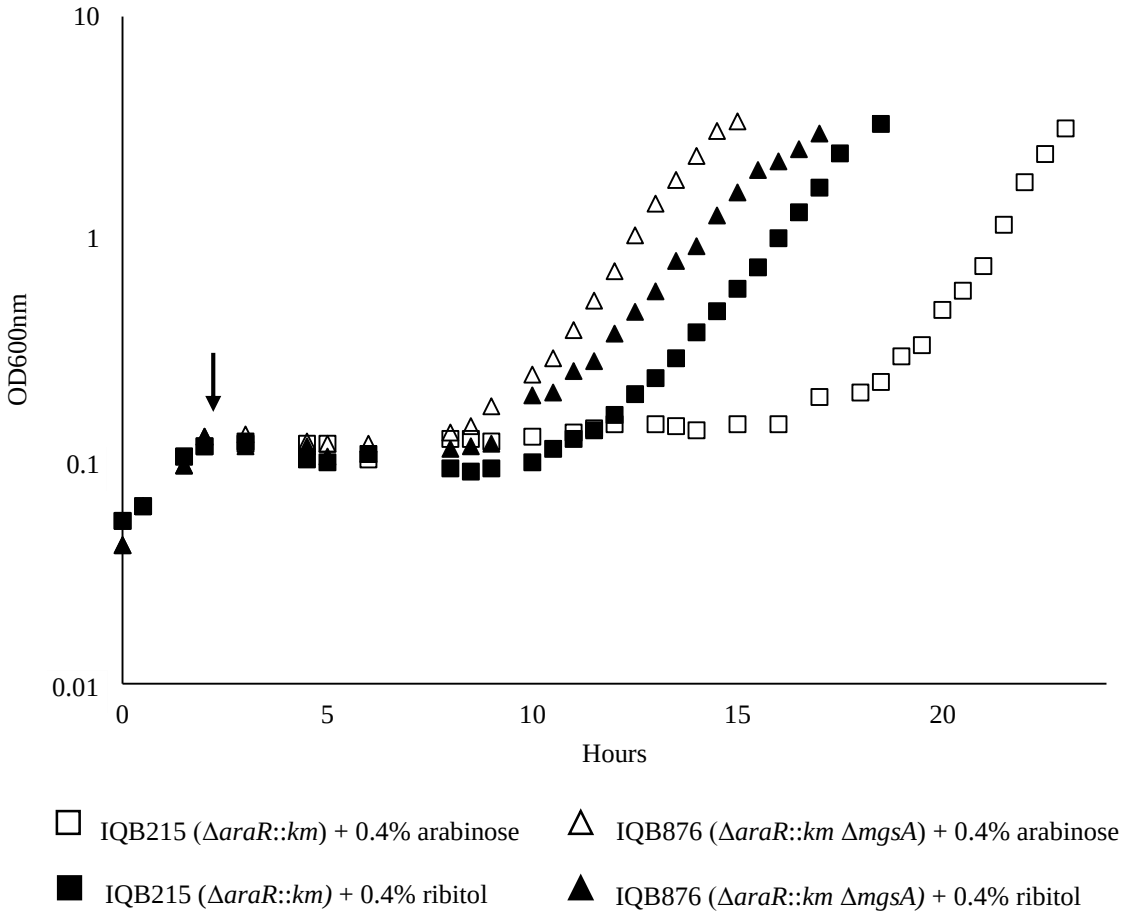


Figure 4.6. Recovery of growth of *B. subtilis araR*-null mutant strains in complex medium. Black arrow indicates time of arabinose or ribitol addition to an early exponential growing culture of *B. subtilis*. Strains IQB215 (Δ*araR*::*km*) and IQB 876 (Δ*araR*::*km* Δ*mgsA*) were tested to ascertain recovery from growth arrest upon arabinose or ribitol addition.

In the presence of arabinose, the *araR*-null mutant took longer to resume growth than in the presence of ribitol. We hypothesize that this could be due to the known cell toxicity effect of MG, since no significant production of MG was detected when the strains were grown in the presence of ribitol (Figure 4.1). So, we examined the behavior of the Δ*araR*::*km* Δ*mgsA* double mutant strain in identical conditions (presence of arabinose and ribitol) and the data revealed that this strain restarted growth phenotype much earlier than the single *araR*-null mutant (Figure 4.6). This observation is most probably due to the lack of MgsA, and consequent absence of MG in the cell, thus allowing the double mutant strain to recover growth more rapidly than the single mutant strain. The mechanisms involved in the capacity of the strains to recovery from growth arrest is unknown, however one possible explanation is that the selective pressure applied by sugar-phosphates stress favors the proliferation of cells carrying spontaneous mutations in the *araB* gene encoding the L-arabinose ribulokinase, which is able to phosphorylate both substrates arabinose and ribitol.

Also, addition of ribitol to IQB215 (Δ*araR*::*km*) and IQB866 (Δ*araR*::*km araD*$^-$) causes immediate cessation of growth, but does not result in MG accumulation in the medium nor in intracellular ATP decrease. We know that ribitol-phosphate accumulates, at least, in IQB215 (data not shown). Thus, if ribitol phosphorylation does not result in significant intracellular ATP decrease as we expected, we can hypothesize that accumulation of ribitol-P in our *araR*-null mutant strains could be impairing cell wall biosynthesis, due to action of enzymes involved in that process that may have affinity to ribitol-P, disrupting cell wall biosynthesis processes.

Acknowledgments

We would like to thank Prof. Helena Santos Head of the Cell Physiology and NMR group at ITQB António Xavier – NOVA for helpful discussions.

This work was partially funded by Fundação para a Ciência e a Tecnologia, grants no. [SFRH/BD/73109/2010 to L.G; 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]

Chapter V

Concluding Remarks and Future Perspectives

Concluding Remarks and Future Perspectives

In this work, investigated the role of *araL* and *araM* in the toxic effect observed upon addition of arabinose to an early-exponentially growing culture of an *araR*-null mutant. Previous results obtained in our laboratory showed that the toxic effect of arabinose verified in an *araR*-null mutant is suppressed by a deletion of all genes downstream of *araD*. Neither *araNPQ* nor *abfA* gene products are considered to contribute to the toxic effect, due to the nature of their function, leaving as potential candidates for this effect the genes *araL* and *araM*. By two complementary approaches we showed that *araL* and *araM* are not involved in this phenomenon of toxicity. First, in-frame deletion mutations in both *araL and araM* genes constructed in an *araR*-null mutant failed to suppress the toxic effect of arabinose addition. Then, ectopic expression of both *araL* and *araM* under the control of an inducible promoter in a strain carrying an *araR*-null mutation and a deletion of all genes downstream of *araD* failed to re-establish the toxic effect. Neither *araL* or *araM* are responsible for the toxic effect, so the most plausible explanation is that the toxic effect could be due an increased intracellular level of arabinose in the *araR*- null mutant, caused by deregulation of all arabinose responsive genes, consequently leading to an increase in the concentration of the metabolic sugar phosphates intermediates that are toxic to the cell.

Furthermore, we demonstrate that the large deletion in the genes downstream from *araD*, constructed by an insertion-deletion mutation leads to destabilization of the upstream mRNA. The generation of markerless in-frame deletions of *araLMNPQ* combined with quantitative analysis of mRNA showed that in the strain carrying the insertion deletion mutation a lower level of *araABD* mRNA is present resulting in a decrease in the concentration of the enzymes involved in the catabolism of arabinose and consequently to a minor the concentration of the sugar phosphates intermediates.

We prove that the *araL* gene encodes a phosphatase, with activity towards several phosphorylated sugars. Over production, purification and biochemical characterization of a recombinant version of AraL was successful. Experimental data indicates that AraL is a magnesium dependent alkaline phosphatase functioning at high temperatures, with optimal temperature at 65 ºC. Substrate specificity of AraL points to a biological function within the context of carbohydrate metabolism – like other members of the HAD superfamily, it displays a wide substrate utilization profile and low substrate specificity. Because nonspecific HAD enzymes have broad and low substrate specificities and do not participate in specific metabolic pathways, we proposed that these enzymes detoxify sugar phosphates that accidentally accumulate at metabolic bottlenecks.

We have also shown, for the first time, that a genetic regulatory mechanism controls expression/production of a member of the HADSF, NagD family. The *araL* gene is under the control of the operon promoter, which is very strong promoter and basal expression in the absence of inducer is always present. A second level of regulation that involves sequestering of the *araL* ribosome binding site and consequently drastically reduces production of AraL is observed, as demonstrated by gene reporter fusion assays and immunoblotting experiments.

In an *araR*-null mutant growth arrest caused by accumulation of phosphosugars was tested by inserting small in-frame deletions of specific codons in the arabinose catabolic genes *araB* and *araD*. Targeting residues directly involved in the catalytic mechanism of the enzymes, their activity was abolished whilst limiting a polar effect at the mRNA level. Strains lacking AraB (kinase) activity were successful in the reestablishment of the growth phenotype in the presence of arabinose, whereas strains deficient in AraD (epimerase) activity were still unable to grow in the presence of arabinose due to accumulation of ribulose 5-phosphate. In order to identify the metabolites building-up in the *araR*-null mutant strain, as opposed to the wild-type strain, we used ^{31}P-NMR and were able to identify the majority of the peaks present in the phosphate monoester region as phosphosugars from both the pentose phosphate pathway and glycolysis. Enhanced levels of ribulose 5-phosphate, xylulose 5-phosphate, fructose 1,6-bisphosphate, fructose 6-phosphate, glucose 6-phosphate and 6-phosphogluconate were found.

Our studies led to the conclusion that the major driving forces for arabinose toxicity in a strain lacking *araR* are accumulation of phosphorylated sugars, together with accumulation of methylglyoxal and ATP depletion. Synthesis of methylglyoxal is widely accepted as an overflow mechanism that prevents accumulation of phosphorylated intermediates, despite its cytotoxicity. Accumulation of methylglyoxal in the growth medium was higher in the *araR*-null mutant, which draw a parallel with the growth kinetics results: strains accumulating more methylglyoxal stop growth after arabinose addition and display higher levels of mRNA for the arabinose catabolic genes in the same conditions. These observations together with relative ATP quantification, enable the establishment of a correlation between growth arrest phenotype (arabinose-sensitivity), methylglyoxal accumulation and drop in ATP level in the *araR*-null mutant when compared to the wild-type strain. Metabolites arising from the PPP are phosphorylated using ATP, while less ATP molecules are generated via the lower branch of glycolysis, because metabolism is stalled at the production of FBP, as shown by the ^{31}P-NMR experiments.

Decrease on intracellular ATP content is concomitant with an increase in phosphofructokinase activity in the cell, phosphorylating F6P and yielding FBP. An increase in FBP levels in the cell leads to phosphorylation of Crh which impairs binding to MgsA, coupling high activity of the methylglyoxal pathway to high intracellular levels of FBP. Strains with an impaired arabinose catabolic pathway, *araD*$^{-}$ strains, do not accumulate MG in the growth medium but still display a decrease in intracellular ATP content in the presence of arabinose. This could be due to ATP molecules being invested in the phosphorylation L-ribulose and no ATP being generated via the lower branch of glycolysis, because the pathway is stalled at L-ribulose 5-phosphate. Similarly, the presence of ribitol causes immediate cessation of growth of strains bearing a deletion in *araR* but does not result in methylglyoxal accumulation in the medium, although an intracellular ATP decrease is observed. Since ribitol-phosphate accumulates in the cell and is not further metabolized by the cells there is an investment of ATP molecules in the phosphorylation of ribitol. In the case of these strains, growth arrest

could be attributed to the decrease in intracellular ATP, which ultimately may have implications in the redox balance of the cell, disrupting the $NADH/NAD^+$ ratio.

Our study clearly shows that inactivation of a regulatory protein responsible for controlling the utilization of a secondary carbon source can have a major impact in central carbon metabolism. Understanding how secondary carbon sources can be used in major pathways (like glycolysis) is key to further understand fermentation of these secondary carbon sources like pentoses, which are industrially relevant, as strain enhancement can lead to bottlenecks causing either depletion of precursors or accumulation of toxic products.

While AraL apparently does not play a role in the toxic effect described, it is worth mentioning a study focusing on the role of a HAD phosphatase (YigL) in carbon efflux for glucose homeostasis in *E. coli* (Papenfort *et al.*, 2013). The phosphatase is regulated post-transcriptionally by a small RNA SgrS and was shown to possibly remove phosphorylated 2-deoxyglucose and the glucose analog α-methyl-glucoside as an immediate response to reduce intracellular stress. AraL production is also post-transcriptionally regulated although by a different mechanisms involving a secondary structure in the mRNA which sequesters the ribosome binding site. No signal for the disruption of this hairpin-like structure has yet been identified and no definite physiological role for AraL has been assigned thus is plausible to speculate that AraL may also play a role in the detoxification and carbon efflux, in a way similar to YigL.

References

Allen, K. N. & Dunaway-Mariano, D. (2004). Phosphoryl group transfer: evolution of a catalytic scaffold. *Trends Biochem Sci* **29**, 495–503.

Allen, K. N. & Dunaway-Mariano, D. (2009). Markers of fitness in a successful enzyme superfamily. *Curr Opin Struct Biol* **19**, 658–65.

Anagnostopoulos, C. & Spizizen, B. Y. J. (1960). Requirements for transformation in *Bacillus subtilis*. *J Bacteriol* **81**, 741–746.

Araújo, A. & Ward, O. P. (1990). Hemicellulases of *Bacillus species*: Preliminary comparative studies on production and properties of mannanases and galactanases. *J Appl Bacteriol* **68**, 253–261.

Aravind, L., Galperin, M. Y. & Koonin, E. V. (1998). The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. *Trends Biochem Sci* **23**, 127–9.

Aristilde, L., Lewis, I. a., Park, J. O. & Rabinowitz, J. D. (2015). Hierarchy in Pentose Sugar Metabolism in *Clostridium acetobutylicum*. *Appl Environ Microbiol* **81**, 1452–1462.

Arnaud, M., Chastanet, A. & Débarbouillé, M. (2004). New Vector for Efficient Allelic Replacement in Naturally Gram-Positive Bacteria. *J Bacteriol* **70**, 6887–6891.

Arrieta-Ortiz, M. L., Hafemeister, C., Bate, A. R., Chu, T., Greenfield, A., Shuster, B., Barry, S. N., Gallitto, M., Liu, B. & other authors. (2015). An experimentally supported model of the *Bacillus subtilis* global transcriptional regulatory network. *Mol Syst Biol* **11**, 839.

Barbe, V., Cruveiller, S., Kunst, F., Lenoble, P., Meurice, G., Sekowska, A., Vallenet, D., Wang, T., Moszer, I. & other authors. (2009). From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. *Microbiology* **155**, 1758–75.

Bartholomae, M., Meyer, F. M., Commichau, F. M., Burkovski, A., Hillen, W. & Seidel, G. (2014). Complex formation between malate dehydrogenase and isocitrate dehydrogenase from *Bacillus subtilis* is regulated by tricarboxylic acid cycle metabolites. *FEBS J* **281**, 1132–43.

Becker, J. & Boles, E. (2003). A Modified *Saccharomyces cerevisiae* Strain That Consumes L -Arabinose and Produces Ethanol. *Appl Environ Microbiol* **69**, 4144–4150.

Beijer, L., Nilsson, R. P., Holmberg, C. & Rutberg, L. (1993). The *glpP* and *glpF* genes of the glycerol regulon in *Bacillus subtilis*. *J Gen Microbiol* **139**, 349–59.

Berg, J. M., Tymoczko, J. L. & Stryer, L. (2011). *Biochemistry: International Edition*. W. H. Freeman; 7th edition edition.

Bettenbrock, K., Sauter, T., Jahreis, K., Kremling, A., Lengeler, J. W. & Gilles, E.-D. (2007). Correlation between growth rates, EIIACrr phosphorylation, and intracellular cyclic AMP levels in *Escherichia coli* K-12. *J Bacteriol* **189**, 6891–900.

Blencke, H.-M., Homuth, G., Ludwig, H., Mäder, U., Hecker, M. & 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.

Bräsen, C., Esser, D., Rauch, B. & Siebers, B. (2014). Carbohydrate metabolism in Archaea: current insights into unusual enzymes and pathways and their regulation. *Microbiol Mol Biol Rev* **78**, 89–

175.

Breitling, R., Ceniceros, A., Jankevics, A. & Takano, E. (2013). Metabolomics for secondary metabolite research. *Metabolites* **3**, 1076–83.

Burkholder, P. & Giles, N. H. (1947). Induced Biochemical Mutations in *Bacillus subtilis*. *Am J Bot* **34**, 345–348.

Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D. & Aravind, L. (2006). Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. *J Mol Biol* **361**, 1003–34.

Byrnes, M., Zhu, X., Younathan, E. S. & Chang, S. H. (1994). Kinetic characteristics of phosphofructokinase from *Bacillus stearothermophilus*: MgATP nonallosterically inhibits the enzyme. *Biochemistry* **33**, 3424–31.

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

Chandrangsu, P., Dusi, R., Hamilton, C. J. & Helmann, J. D. (2014). Methylglyoxal resistance in *Bacillus subtilis*: contributions of bacillithiol-dependent and independent pathways. *Mol Microbiol* **91**, 706–15.

Chauvaux, S., Paulsen, I. T. & Saier, M. H. (1998). CcpB, a novel transcription factor implicated in catabolite repression in *Bacillus subtilis*. *J Bacteriol* **180**, 491–7.

Cohen, G. N. (2011). *Microbial Biochemistry*. *Experientia*. New York: Springer.

Commichau, F. M., Rothe, F. M., Herzberg, C., Wagner, E., Hellwig, D., Lehnik-Habrink, M., Hammer, E., Völker, U. & Stülke, J. (2009). Novel activities of glycolytic enzymes in *Bacillus subtilis*: interactions with essential proteins involved in mRNA processing. *Mol Cell Proteomics* **8**, 1350–60.

Conn, H. (1930). The Identity of *Bacillus subtilis*. *J Infect Dis* **46**, 341 – 350.

Cooper, R. A. (1974). Methylglyoxal Synthase. In *Methods in Enymology*, pp. 502–508.

Cooper, R. A. & Anderson, A. (1970). The Formation and Catabolism of Methylglyoxal during glycolysis in *Escherichia coli*. *FEBS J* **11**, 273–276.

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

Cozzarelli, N. R., Koch, J. P., Hayashi, S. & Lin, E. C. (1965). Growth stasis by accumulated L-alpha-glycerophosphate in *Escherichia coli*. *J Bacteriol* **90**, 1325–9.

Cutting, S. M. (2011). *Bacillus* probiotics. *Food Microbiol* **28**, 214–20.

Darbon, E., Servant, P., Poncet, S. & Deutscher, J. (2002). Antitermination by GlpP, catabolite repression via CcpA and inducer exclusion triggered by P~GlpK dephosphorylation control *Bacillus subtilis glpFK* expression. *Mol Microbiol* **43**, 1039–1052.

Debarbouillé, M., Arnaud, M., Fouet, A., Klier, A. & Rapoport, G. (1990). The *sacT* gene regulating the *sacPA* operon in *Bacillus subtilis* shares strong homology with transcriptional antiterminators.

J Bacteriol **172**, 3966–73.

Deutscher, J., Galinier, A. & Martin-Verstraete, I. (2002). *Bacillus subtilis and Its Closest Relatives* (A. L. Sonenshein, J. A. Hoch & R. Losick, Eds.). American Society of Microbiology.

Deutscher, J., Francke, C., Postma, P. W., Deutscher, J., Francke, C. & Postma, P. W. (2006). How Phosphotransferase System-Related Protein Phosphorylation Regulates Carbohydrate Metabolism in Bacteria. *Microbiol Mol Biol Rev* **70**.

van Dijl, J. M. & Hecker, M. (2013). *Bacillus subtilis*: from soil bacterium to super-secreting cell factory. *Microb Cell Fact* **12**, 3.

Doan, T. & Aymerich, S. (2003). Regulation of the central glycolytic genes in *Bacillus subtilis*: Binding of the repressor CggR to its single DNA target sequence is modulated by fructose 1,6-bisphosphate. *Mol Microbiol* **47**, 1709–1721.

Dzeja, P. P. (2003). Phosphotransfer networks and cellular energetics. *J Exp Biol* **206**, 2039–2047.

Earl, A. M., Losick, R., Kolter, R., Wang, J., Dye, B. T., Rajashankar, K. R., Kurinov, I. & Brenda, A. (2008). Ecology and Genomics of *Bacillus subtilis*. *Trends Microbiol* **16**, 1–11.

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

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

Franco, I. S., Mota, L. J., Soares, C. M. & 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. S., Mota, L. J., Soares, C. M. & Sá-Nogueira, I. (2007). Probing key DNA contacts in AraR-mediated transcriptional repression of the *Bacillus subtilis* arabinose regulon. *Nucleic Acids Res* **35**, 4755–66.

Fujita, Y. & Fujita, T. (1989). Effect of mutations causing gluconate kinase or gluconate permease deficiency on expression of the *Bacillus subtilis gnt* operon. *J Bacteriol* **171**, 1751–4.

Fujita, Y. (2009). Carbon catabolite control of the metabolic network in *Bacillus subtilis*. *Biosci Biotechnol Biochem* **73**, 245–59.

Galinier, A., Haiech, J., Kilhoffer, M.-C., Jaquinod, M., Stülke, J., Deutscher, J. & Martin-Verstraete, I. (1997). The *Bacillus subtilis crh* gene encodes a HPr-like protein involved in carbon catabolite repression. *Proc Natl Acad Sci U S A* **94**, 8439–44.

Garcia Sanchez, R., Karhumaa, K., Fonseca, C., Sànchez Nogué, V., Almeida, J. R., Larsson, C. U., Bengtsson, O., Bettiga, M., Hahn-Hägerdal, B. & Gorwa-Grauslund, M. F. (2010). Improved xylose and arabinose utilization by an industrial recombinant *Saccharomyces cerevisiae* strain using evolutionary engineering. *Biotechnol Biofuels* **3**, 13.

Gírio, F. M., Fonseca, C., Carvalheiro, F., Duarte, L. C., Marques, S. & Bogel-Łukasik, R. (2010).

Hemicelluloses for fuel ethanol: A review. *Bioresour Technol* **101**, 4775–4800. Elsevier Ltd.

Godinho, L. M. & 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.

Görke, B. & Stülke, J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol* **6**, 613–624.

Görke, B., Fraysse, L. & Galinier, A. (2004). Drastic differences in Crh and HPr synthesis levels reflect their different impacts on catabolite repression in *Bacillus subtilis*. *J Bacteriol* **186**, 2992–5.

Grundy, F. J. & Henkin, T. M. (2006). From ribosome to riboswitch: control of gene expression in bacteria by RNA structural rearrangements. *Crit Rev Biochem Mol Biol* **41**, 329–38.

Guggisberg, A. M., Park, J., Edwards, R. L., Kelly, M. L., Hodge, D. M., Tolia, N. H. & Odom, A. R. (2014). A sugar phosphatase regulates the methylerythritol phosphate (MEP) pathway in malaria parasites. *Nat Commun* **5**, 4467. Nature Publishing Group.

Guldan, H., Sterner, R. & Babinger, P. (2008). Identification and characterization of a bacterial glycerol-1-phosphate dehydrogenase: $Ni^{(2+)}$-dependent AraM from *Bacillus subtilis*. *Biochemistry* **47**, 7376–84.

Harwood, C. R. (1992). *Bacillus subtilis* and its relatives: molecular biological and industrial workhorses. *Trends Biotechnol* **10**, 247–256.

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* **8**, R19.

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

Hogema, B. M., Arents, J. C., Bader, R., Eijkemans, K., Inada, T., Aiba, H. & Postma, P. W. (1998a). Inducer exclusion by glucose 6-phosphate in *Escherichia coli*. *Mol Microbiol* **28**, 755–765.

Hogema, B. M., Arents, J. C., Bader, R., Eijkemans, K., Yoshida, H., Takahashi, H., Aiba, H. & Postma, P. W. (1998b). Inducer exclusion in *Escherichia coli* by non-PTS substrates: the role of the PEP to pyruvate ratio in determining the phosphorylation state of enzyme IIAGlc. *Mol Microbiol* **30**, 487–498.

Hong, H. A., Khaneja, R., Tam, N. M. K., Cazzato, A., Tan, S., Urdaci, M., Brisson, A., Gasbarrini, A., Barnes, I. & Cutting, S. M. (2009). *Bacillus subtilis* isolated from the human gastrointestinal tract. *Res Microbiol* **160**, 134–43.

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

Huggins, C. & Smith, D. R. (1947). Chromogenic Substrates III. *p*-nitrophenyl sulfate as a substrate

for the assay of phenolsulfatase activity. *J Biol Chem* **170**, 391 – 398.

Hulett, F. Ma., Bookstein, C. & Karren Jensen. **(1990).** Evidence for Two Structural Genes for Alkaline Phosphatase in *Bacillus subtilis*. *J Bacteriol* **172**, 735–740.

Huynh, P. L., Jankovic, I., Schnell, N. F. & Brückner, R. (2000). Characterization of an HPr kinase mutant of *Staphylococcus xylosus. J Bacteriol* **182**, 1895–902.

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

Inácio, J. M. & 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**, 4272–80.

Inácio, J. M., Costa, C. & Sá-Nogueira, I. (2003). Distinct molecular mechanisms involved in carbon catabolite repression of the arabinose regulon in *Bacillus subtilis*. *Microbiology* **149**, 2345–55.

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

Inoue, Y. & Kimura, A. (1995). Methylglyoxal and regulation of its metabolism in microorganisms. *Adv Microb Physiol* **37**, 177–227.

Irani, M. H. & Maitra, P. K. (1977). Properties of *Escherichia coli* mutants deficient in enzymes of glycolysis. *J Bacteriol* **132**, 398 – 410.

Ishikawa, S., Core, L. & Perego, M. (2002). Biochemical characterization of aspartyl phosphate phosphatase interaction with a phosphorylated response regulator and its inhibition by a pentapeptide. *J Biol Chem* **277**, 20483–9.

Jana, M., Luong, T. T., Komatsuzawa, H., Shigeta, M. & Lee, C. Y. (2000). A method for demonstrating gene essentiality in *Staphylococcus aureus*. *Plasmid* **44**, 100–4.

Johnsen, U., Sutter, J.-M., Zaiß, H. & Schönheit, P. (2013). L-Arabinose degradation pathway in the haloarchaeon *Haloferax volcanii* involves a novel type of L-arabinose dehydrogenase. *Extremophiles* **17**, 897–909.

Jojima, T., Igari, T., Gunji, W., Suda, M., Inui, M. & Yukawa, H. (2012). Identification of a HAD superfamily phosphatase, HdpA, involved in 1,3-dihydroxyacetone production during sugar catabolism in *Corynebacterium glutamicum*. *FEBS Lett* **586**, 4228–32.

Joseph Sambrook & David W. Russel. **(2001).** *Molecular Cloning - A Laboratory Manual 3rd edition*, 3rd edn. New York: Cold Spring Harbor Laboratory Press.

Kadner, R. J., Murphy, G. P. & Stephens, C. M. (1992). Two mechanisms for growth inhibition by elevated transport of sugar phosphates in *Escherichia coli*. *J Gen Microbiol* **138**, 2007–14.

Kaneko, Y., Toh-e, A. & Oshima, Y. (1989). Molecular characterization of a specific *p*-nitrophenylphosphatase gene, PHO13, and its mapping by chromosome fragmentation in *Saccharomyces cerevisiae*. *Mol Gen Genet* **220**, 133–139.

Kawamoto, H., Koide, Y., Morita, T. & Aiba, H. (2006). Base-pairing requirement for RNA silencing

by a bacterial small RNA and acceleration of duplex formation by Hfq. *Mol Microbiol* **61**, 1013–22.

Kim, B. H. & Gadd, G. M. (2008). *Bacterial Physiology and Metabolism*. New York: Cambridge Univeristy Press.

Kim, S. R., Xu, H., Lesmana, A., Kuzmanovic, U., Au, M., Florencia, C., Oh, E. J., Zhang, G., Kim, K. H. & Jin, Y.-S. (2015). Deletion of PHO13 , Encoding Haloacid Dehalogenase Type IIA Phosphatase, Results in Upregulation of the Pentose Phosphate Pathway in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* **81**, 1601–1609.

Kimata, K., Tanaka, Y., Inada, T. & Aiba, H. (2001). Expression of the glucose transporter gene , *ptsG* , is regulated at the mRNA degradation step in response to glycolytic flux in *Escherichia coli*. *EMBO J* **20**, 3587–95.

Knöckel, J., Bergmann, B., Müller, I. B., Rathaur, S., Walter, R. D. & Wrenger, C. (2008). Filling the gap of intracellular dephosphorylation in the *Plasmodium falciparum* vitamin B1 biosynthesis. *Mol Biochem Parasitol* **157**, 241–3.

Knowles, J. R. (1980). Enzyme-catalyzed phosphoryl transfer reactions. *Annu Rev Biochem* **49**, 877–919.

Koga, Y. (2014). From promiscuity to the lipid divide: On the evolution of distinct membranes in Archaea and Bacteria. *J Mol Evol* **78**, 234–242.

Koonin, E. V & Tatusov, R. L. (1994). Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity. Application of an iterative approach to database search. *J Mol Biol*.

Krispin, O. & Allmansberger, R. (1998a). The *Bacillus subtilis* AraE protein displays a broad substrate specificity for several different sugars. *J Bacteriol* **180**, 3250–2.

Krispin, O. & Allmansberger, R. (1998b). The *Bacillus subtilis galE* Gene Is Essential in the Presence of Glucose and Galactose. *J Bacteriol* **180**, 2265–2270.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, a M., Alloni, G., Azevedo, V., Bertero, M. G., Bessières, P., Bolotin, A. & other authors. **(1997).** The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249–56.

Küppers, T., Steffen, V., Hellmuth, H., O'Connell, T., Bongaerts, J., Maurer, K.-H. & Wiechert, W. (2014). Developing a new production host from a blueprint: *Bacillus pumilus* as an industrial enzyme producer. *Microb Cell Fact* **13**, 46.

Kurahashi, K. & Wahba, A. J. (1958). Interference with growth of certain *Escherichia coli* mutants by galactose. *Biochim Biophys Acta* **30**, 298–302.

Kuznetsova, E., Proudfoot, M., Gonzalez, C. F., Brown, G., Omelchenko, M. V, Borozan, I., Carmel, L., Wolf, Y. I., Mori, H. & other authors. **(2006).** Genome-wide analysis of substrate specificities of the *Escherichia coli* haloacid dehalogenase-like phosphatase family. *J Biol Chem* **281**, 36149–61.

Lahiri, S. D., Zhang, G., Dai, J., Dunaway-Mariano, D. & Allen, K. N. (2004). Analysis of the substrate specificity loop of the HAD superfamily cap domain. *Biochemistry* **43**, 2812–20.

Landmann, J. J., Busse, R. a, Latz, J.-H., Singh, K. D., Stülke, J. & Görke, B. (2011). Crh, the paralogue of the phosphocarrier protein HPr, controls the methylglyoxal bypass of glycolysis in *Bacillus subtilis*. *Mol Microbiol* **82**, 770–87.

Landmann, J. J., Werner, S., Hillen, W., Stülke, J. & Görke, B. (2012). Carbon source control of the phosphorylation state of the *Bacillus subtilis* carbon-flux regulator Crh in vivo. *FEMS Microbiol Lett* **327**, 47–53. The Oxford University Press.

Larkin, M. a, Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. a, McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. & other authors. **(2007).** Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–8.

Lee, J. (1997). Biological conversion of lignocellulosic biomass to ethanol. *J Biotechnol* **56**, 1–24.

Lee, S. J., Trostel, A., Le, P., Harinarayanan, R., FitzGerald, P. C. & Adhya, S. (2009). Cellular stress created by intermediary metabolite imbalances. *Proc Natl Acad Sci U S A* **106**, 19515–19520.

Lepesant, J. A. & Dedonder, R. (1967). Metabolism of L-arabinose in *Bacillus subtilis* Marburg Ind-168. *C R Acad Sci Hebd Seances Acad Sci D* **264**, 2683–6.

Light, S. H., Halavaty, A. S., Minasov, G., Shuvalova, L. & Anderson, W. F. (2012). Structural analysis of a 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase with an N-terminal chorismate mutase-like regulatory domain. *Protein Sci* **21**, 887–895.

Lindow, J. C., Britton, R. A. & Grossman, A. D. (2002). Structural maintenance of chromosomes protein of *Bacillus subtilis* affects supercoiling in vivo. *J Bacteriol* **184**, 5317–22.

Lu, Z., Dunaway-Mariano, D. & Allen, K. N. (2005). HAD superfamily phosphotransferase substrate diversification: structure and function analysis of HAD subclass IIB sugar phosphatase BT4131. *Biochemistry* **44**, 8684–96.

Lu, Z., Dunaway-Mariano, D. & Allen, K. N. (2008). The catalytic scaffold of the haloalkanoic acid dehalogenase enzyme superfamily acts as a mold for the trigonal bipyramidal transition state. *Proc Natl Acad Sci U S A* **105**, 5687–92.

Ludwig, H., Homuth, G., Schmalisch, M., Dyka, F. M., Hecker, M. & Stülke, J. (2001). Transcription of glycolytic genes and operons in *Bacillus subtilis*: evidence for the presence of multiple levels of control of the *gapA* operon. *Mol Microbiol* **41**, 409–22.

McNeil, B., Archer, D., Giavasis, I. & Harvey, L. (Eds.). **(2013).** *Microbial Production of Food Ingredients, Enzymes and Nutraceuticals*. Cambridge: Woodhead Publishing Limited.

Mijakovic, I., Musumeci, L., Tautz, L., Petranovic, D., Edwards, R. a, Jensen, P. R., Mustelin, T., Deutscher, J. & Bottini, N. (2005). In vitro characterization of the *Bacillus subtilis* protein tyrosine phosphatase YwqE. *J Bacteriol* **187**, 3384–90.

Miller, J. H. (1972). **(1972).** *Experiments in Molecular Genetics*. *Exp Mol Genet*. New York.

Mitchell, W. J., Saffen, D. W. & Roseman, S. (1987). Sugar Transport by the Bacterial Phosphotransferase System. *J Biol Chem* **262**, 16254 – 16260.

Mittal, M., Pechter, K. B., Picossi, S., Kim, H.-J., Kerstein, K. O. & Sonenshein, A. L. (2013). Dual role of CcpC protein in regulation of aconitase gene expression in *Listeria monocytogenes* and *Bacillus subtilis*. *Microbiology* **159**, 68–76.

Miwa, Y. & Fujita, Y. (2001). Involvement of two distinct catabolite-responsive elements in catabolite repression of the *Bacillus subtilis* myo-inositol (*iol*) operon. *J Bacteriol* **183**, 5877–84.

Miwa, Y., Nagura, K., Eguchi, S., Fukuda, H., Deutscher, J. & Fujita, Y. (1997). Catabolite repression of the *Bacillus subtilis gnt* operon exerted by two catabolite-responsive elements. *Mol Microbiol* **23**, 1203–1213.

Moat, A. G., Foster, J. W. & Spector, M. P. (2002). *Microbial Physiology*, 4th edn. New York: John Wiley & Sons.

Moreno, M. S., Schneider, B. L., Maile, R. R., Weyler, W. & Saier, M. H. (2001). Catabolite repression mediated by the CcpA protein in *Bacillus subtilis*: novel modes of regulation revealed by whole-genome analyses. *Mol Microbiol* **39**, 1366–81.

Morinaga, T., Matsuse, T., Ashida, H. & Yoshida, K. (2010). Differential substrate specificity of two inositol transporters of *Bacillus subtilis*. *Biosci Biotechnol Biochem* **74**, 1312–4.

Mota, L. J., Tavares, P. & Sá-Nogueira, I. (1999). Mode of action of AraR, the key regulator of L-arabinose metabolism in *Bacillus subtilis*. *Mol Microbiol* **33**, 476–89.

Mota, L. J., Morais Sarmento, L. & 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.

Nguyen, T. T. H., Eiamphungporn, W., Mäder, U., Liebeke, M., Lalk, M., Hecker, M., Helmann, J. D. & Antelmann, H. (2009). Genome-wide responses to carbonyl electrophiles in *Bacillus subtilis*: control of the thiol-dependent formaldehyde dehydrogenase AdhA and cysteine proteinase YraA by the MerR-family regulator YraB (AdhR). *Mol Microbiol* **71**, 876–94.

Pandya, C., Farelli, J. D. & Allen, K. N. (2014). Enzyme Promiscuity : Engine of Evolutionary Innovation. *J Biol Chem* **289**, 30229 – 30236.

Papagianni, M., Avramidis, N. & Filiousis, G. (2007). Glycolysis and the regulation of glucose transport in *Lactococcus lactis* spp. *lactis* in batch and fed-batch culture. *Microb Cell Fact* **6**, 16.

Papenfort, K., Sun, Y., Miyakoshi, M., Vanderpool, C. K. & Vogel, J. (2013). Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. *Cell* **153**, 426–37.

Park, J., Guggisberg, A. M., Odom, A. R. & Tolia, N. H. (2015). Cap-domain closure enables diverse substrate recognition by the C2-type haloacid dehalogenase-like sugar phosphatase *Plasmodium falciparum* HAD1. *Acta Crystallogr Sect D Biol Crystallogr* **71**, 1824–1834.

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

Paveia, H. & Archer, L. (1992). Genes for L-arabinose utilization in *Bacillus subtilis*. *Brotéria Genética* **13**, 149–159.

Peisach, E., Selengut, J. D., Dunaway-Mariano, D. & Allen, K. N. (2004). X-ray crystal structure of the hypothetical phosphotyrosine phosphatase MDP-1 of the haloacid dehalogenase superfamily. *Biochemistry* **43**, 12770–9. American Chemical Society.

Peri, K. G., Goldie, H. & Waygood, E. B. (1990). Cloning and characterization of the N-acetylglucosamine operon of *Escherichia coli*. *Biochem Cell Biol* **68**, 123–137.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**, e45.

Pompeo, F., Luciano, J. & Galinier, A. (2007). Interaction of GapA with HPr and its homologue, Crh: Novel levels of regulation of a key step of glycolysis in *Bacillus subtilis*? *J Bacteriol* **189**, 1154–7.

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

Prasad, C. & Freese, E. (1974). Cell lysis of *Bacillus subtilis* caused by intracellular accumulation of glucose-1-phosphate. *J Bacteriol* **118**, 1111–1122.

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

Quisel, J. D., Burkholder, W. F. & Grossman, A. D. (2001). In vivo effects of sporulation kinases on mutant Spo0A proteins in *Bacillus subtilis*. *J Bacteriol* **183**, 6573–8.

Rao, M. B., Tanksale, A. M., Ghatge, M. S. & Deshpande, V. V. (1998). Molecular and biotechnological aspects of microbial proteases. *Microbiol Mol Biol Rev* **62**, 597–635.

Raposo, M. P., Inácio, J. M., Mota, L. J. & Sá-Nogueira, I. (2004). Transcriptional Regulation of Genes Encoding Arabinan-Degrading Enzymes in *Bacillus subtilis*. *J Bacteriol* **186**, 1287–1296.

Reizer, A., Deutscher, J., Saier, M. H. & Reizer, J. (1991). Analysis of the gluconate (*gnt*) operon of *Bacillus subtilis*. *Mol Microbiol* **5**, 1081–9.

Reizer, J., Hoischen, C., Titgemeyer, F., Rivolta, C., Rabus, R., Stülke, J., Karamata, D., Saier Jr, M. H. & Hillen, W. (1998). A novel protein kinase that controls carbon catabolite repression in bacteria. *Mol Microbiol* **27**, 1157–1169.

Richards, G. R. & Vanderpool, C. K. (2011). Molecular call and response: the physiology of bacterial small RNAs. *Biochim Biophys Acta* **1809**, 525–31.

Richards, G. R., Patel, M. V, Lloyd, C. R. & Vanderpool, C. K. (2013). Depletion of glycolytic intermediates plays a key role in glucose-phosphate stress in *Escherichia coli*. *J Bacteriol* **195**, 4816–25.

Rocha, E. P., Danchin, A. & Viari, A. (1999). Translation in *Bacillus subtilis*: roles and trends of initiation and termination, insights from a genome analysis. *Nucleic Acids Res* **27**, 3567–76.

Rodrigues, M. V, Borges, N., Almeida, C. P., Lamosa, P. & Santos, H. (2009). A unique beta-1,2-

mannosyltransferase of *Thermotoga maritima* that uses di-myo-inositol phosphate as the mannosyl acceptor. *J Bacteriol* **191**, 6105–15.

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

Sá-Nogueira, I. & Mota, L. J. (1997). Negative regulation of L-arabinose metabolism in *Bacillus subtilis*: characterization of the *araR* (*araC*) gene. *J Bacteriol* **179**, 1598–608.

Sá-Nogueira, I. & Ramos, S. S. (1997). Cloning, functional analysis, and transcriptional regulation of the *Bacillus subtilis araE* gene involved in L-arabinose utilization. *J Bacteriol* **179**, 7705–11.

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

Schallmey, M., Singh, A. & Ward, O. P. (2004). Developments in the use of *Bacillus* species for industrial production. *Can J Microbiol* **50**, 1–17.

Schumacher, M. A., Seidel, G., Hillen, W. & Brennan, R. G. (2007). Structural mechanism for the fine-tuning of CcpA function by the small molecule effectors glucose 6-phosphate and fructose 1,6-bisphosphate. *J Mol Biol* **368**, 1042–50.

Seiboth, B. & Metz, B. (2011). Fungal arabinan and L-arabinose metabolism. *Appl Microbiol Biotechnol* **89**, 1665–73.

Selengut, J. D. (2001). MDP-1 Is a New and Distinct Member of the Haloacid Dehalogenase Family of Aspartate-Dependent Phosphohydrolases. *Biochemistry* **40**, 12704–12711. American Chemical Society.

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

Shulami, S., Raz-Pasteur, A., Tabachnikov, O., Gilead-Gropper, S., Shner, I. & Shoham, Y. (2011). The L-Arabinan utilization system of *Geobacillus stearothermophilus*. *J Bacteriol* **193**, 2838–50.

Simoni, R. D., Levinthal, M., Kundig, F. D., Kundig, W., Anderson, B., Hartman, P. E. & Roseman, S. (1967). Genetic evidence for the role of a bacterial phosphotransferase system in sugar transport. *Proc Natl Acad Sci U S A* **58**, 1963–70.

Singh, K. D., Schmalisch, M., Stülke, J. & Görke, B. (2008). Carbon catabolite repression in *Bacillus subtilis*: quantitative analysis of repression exerted by different carbon sources. *J Bacteriol* **190**, 7275–84.

Skerman, V., McGowan, V. & Sneath, P. (1989). Approved Lists of Bacterial Names (Amended). Washington (DC): ASM Press.

Sonenshein, A. L. (2007). Control of key metabolic intersections in *Bacillus subtilis*. *Nat Rev Microbiol* **5**, 917–27.

Sousa, P. M. F., Videira, M. A. M., Santos, F. A. S., Hood, B. L., Conrads, T. P. & Melo, A. M. P.

(2013). The bc:caa3 supercomplexes from the Gram positive bacterium *Bacillus subtilis* respiratory chain: A megacomplex organization? *Arch Biochem Biophys* **537**, 153–160.

Spizizen, B. Y. J. (1958). Transformation of Biochemically Deficient Strains of *Bacillus subtilis* by Deoxyribonucleate. *Proc Natl Acad Sci U S A* **44**, 1072–1078.

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

Studier, F. W., Rosenberg, a H., Dunn, J. J. & Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol*.

Stülke, J. & Hillen, W. (2000). Regulation of Carbon Catabolism in *Bacillus* Species. *Annu Rev Microbiol* **54**, 849 – 880.

Stülke, J., Martin-Verstraete, I., Zagorec, M., Rose, M., Klier, A. & Rapoport, G. (1997). Induction of the *Bacillus subtilis ptsGHI* operon by glucose is controlled by a novel antiterminator, GlcT. *Mol Microbiol* **25**, 65–78.

Subedi, K. P., Kim, I., Kim, J., Min, B. & Park, C. (2008). Role of GldA in dihydroxyacetone and methylglyoxal metabolism of *Escherichia coli* K-12. *FEMS Microbiol Lett* **279**, 180–7.

Subtil, T. & Boles, E. (2011). Improving L-arabinose utilization of pentose fermenting *Saccharomyces cerevisiae* cells by heterologous expression of L-arabinose transporting sugar transporters. *Biotechnol Biofuels* **4**, 38.

Sugahara, T., Konno, Y., Ohta, H., Ito, Z., Kaneko, J., Kamio, Y. & Kazuo Izaki. **(1991).** Purification and Properties of Two Membrane Alkaline Phosphatases from *Bacillus subtilis* 168. *J Bacteriol* **173**, 1824–1826.

Sundararajan, T. A., Rapin, A. M. C. & Kalckar, H. M. (1962). Biochemical Observations on *E.coli* Mutants Defective in Uridine Diphosphoglucose. *Proc Natl Acad Sci U S A* **48**, 2187–2193.

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

Tännler, S., Fischer, E., Le Coq, D., Doan, T., Jamet, E., Sauer, U. & Aymerich, S. (2008). CcpN controls central carbon fluxes in *Bacillus subtilis*. *J Bacteriol* **190**, 6178–6187.

Tremblay, L. W., Dunaway-Mariano, D. & Allen, K. N. (2006). Structure and activity analyses of *Escherichia coli* K-12 NagD provide insight into the evolution of biochemical function in the haloalkanoic acid dehalogenase superfamily. *Biochemistry* **45**, 1183–93.

Tye, a J., Siu, F. K. Y., Leung, T. Y. C. & Lim, B. L. (2002). Molecular cloning and the biochemical characterization of two novel phytases from *B. subtilis* 168 and *B. licheniformis*. *Appl Microbiol Biotechnol* **59**, 190–7.

Vanderpool, C. K. (2007). Physiological consequences of small RNA-mediated regulation of glucose-phosphate stress. *Curr Opin Microbiol* **10**, 146–51.

Vanderpool, C. K. & Gottesman, S. (2004). Involvement of a novel transcriptional activator and small

RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Mol Microbiol* **54**, 1076–89.

Vanderpool, C. K. & Gottesman, S. (2007). The novel transcription factor SgrR coordinates the response to glucose-phosphate stress. *J Bacteriol* **189**, 2238–48.

Várnai, A., Huikko, L., Pere, J., Siika-Aho, M. & Viikari, L. (2011). Synergistic action of xylanase and mannanase improves the total hydrolysis of softwood. *Bioresour Technol* **102**, 9096–104.

Verhees, C. H., Kengen, S. W. M., Tuininga, J. E., Schut, G. J., Adams, M. W. W., De Vos, W. M. & Van Der Oost, J. (2003). The unique features of glycolytic pathways in Archaea. *Biochem J* **375**, 231–246.

Warner, J. B. & Lolkema, J. S. (2003). A Crh-specific function in carbon catabolite repression in *Bacillus subtilis*. *FEMS Microbiol Lett* **220**, 277–80.

Watanabe, S., Kodaki, T., Kodak, T. & Makino, K. (2006a). Cloning, expression, and characterization of bacterial L-arabinose 1-dehydrogenase involved in an alternative pathway of L-arabinose metabolism. *J Biol Chem* **281**, 2612–23.

Watanabe, S., Shimada, N., Tajima, K., Kodaki, T. & Makino, K. (2006b). Identification and characterization of L-arabonate dehydratase, L-2-keto-3-deoxyarabonate dehydratase, and L-arabinolactonase involved in an alternative pathway of L-arabinose metabolism. Novel evolutionary insight into sugar metabolism. *J Biol Chem* **281**, 33521–36.

Weinrauch, Y., Msadek, T., Kunst, F. & Dubnau, D. (1991). Sequence and properties of *comQ*, a new competence regulatory gene of *Bacillus subtilis*. *J Bacteriol* **173**, 5685–93.

Westers, L., Westers, H. & Quax, W. J. (2004). *Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. *Biochim Biophys Acta* **1694**, 299–310.

Wiedemann, B. & Boles, E. (2008). Codon-optimized bacterial genes improve L-Arabinose fermentation in recombinant *Saccharomyces cerevisiae*. *Appl Environ Microbiol* **74**, 2043–50.

Wolfe, A. J. (2015). Glycolysis for the Microbiome Generation. *Microbiol Spectr* **3**, 1–20.

Yang, J., Dhamija, S. S. & M. Ernst Schweingruber. (1991). Characterisation of the specific *p*-nitrophoshatase gene and protein of *Schizosaccharomyces pombe*. *Eur J Biochem* **198**, 493–497.

Yansura, D. G. & Henner, D. J. (1984). Use of the *Escherichia coli lac* repressor and operator to control gene expression in *Bacillus subtilis*. *Proc Natl Acad Sci U S A* **81**, 439–43.

Yarmolinsky, M. B., Wiesmeyer, H., Kalckar, H. M. & Jordon, E. (1959). Hereditary defects in galactose metabolism in *Escherichia coli* mutants. II. Galactose-induced sensitivity. *Proc Natl Acad Sci U S A* **45**, 1786–1791.

Yokoi, T., Isobe, K., Yoshimura, T. & Hemmi, H. (2012). Archaeal Phospholipid Biosynthetic Pathway Reconstructed in *Escherichia coli*. *Archaea* **2012**, 1–9.

Yoshida, K., Yamamoto, Y., Omae, K., Yamamoto, M. & Fujita, Y. (2002). Identification of two myo-inositol transporter genes of *Bacillus subtilis*. *J Bacteriol* **184**, 983–91.

Zeigler, D. R., Prágai, Z., Rodriguez, S., Chevreux, B., Muffler, A., Albert, T., Bai, R., Wyss, M. & Perkins, J. B. (2008). The origins of 168, W23, and other *Bacillus subtilis* legacy strains. *J Bacteriol* **190**, 6983–95.