Der Einfluss der Glutamatdehydrogenasen auf die Verknüpfung des Kohlenstoff- und Stickstoffstoffwechsels in *Bacillus subtilis*

Dissertation zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Katrin Gunka
aus Cuxhaven

Göttingen 2010

"Ohne Spekulation gibt es keine neue Beobachtung."

(Charles Darwin)

Danksagung

Table of Contents

List of Abbreviations

% (vol/vol)	% (volume/volume)
% (wt/vol)	% (weight/volume)
3-m-2-o-b-ate	3-methyl-2-oxo-butanoate
4-m-2-o-p-ate	4-methyl-2-oxo-pentanoate
ADP	adenosine diphosphate
Amp	ampicillin
AP	alkaline phospatase
ATP	adenosine triphosphate
B.	*Bacillus*/*Bordetella*
B2H	bacterial two-hybrid
C.	*Clostridium*/*Corynebacterium*
cAMP	cyclic AMP
CcpA	catabolite control protein A
CDP*	disodium 2-chloro-5-(4-methoxyspiro {1,2-dioxetane-3,2'-(5'-chloro) tricyclo[3.3.1.13,7]decan}-4-yl) phenyl phosphate
Cm	chloramphenicol
cre	catabolite responsive element
DNA	deoxyribonucleic acid
E.	*Escherichia*
EDTA	ethylenediaminetetraacetic acid
Erm	erythromycin
EMSA	electrophoretic mobility shift assay
FA	formaldehyde
FAD	flavin adenine dinucleotide (oxidized)
FADH	flavin adenine dinucleotide (reduced)
FBA	Flux Balance Analysis
Fig.	figure
fwd	forward
GDH	glutamate dehydrogenase
Glc	glucose

GOGAT	glutamine-oxoglutarate-aminotransferase
GS	glutamine synthetase
IgG	Immunglobulin G
IPTG	isopropyl-β-D-thiogalactopyranoside
k_m	Michaelis constant
LB	Luria Bertani (medium)
LFH-PCR	Long Flanking Homology PCR
Lin	Lincomycin
MILP	mixed-integer linear programming
MOMA	minimization of metabolic adjustment
mRNA	messenger RNA
NAD	nicotinamide adenine dinucleotide (oxidized)
NADH	nicotinamide adenine dinucleotide (reduced)
$NADP^+$	nicotinamide adenine dinucleotide phosphate (oxidized)
NADPH	nicotinamide adenine dinucleotide phosphate (reduced)
NER	nucleotide excision repair
OD_x	optical density , measured at the wavelength $\lambda = x$ nm
ONPG	ortho-Nitrophenyl-β-galactoside
ori	origin of replication
P	promoter
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
pH	power of hydrogen
PVDF	polyvinylidene difluoride
r.m.s.d.	root mean square deviation
rev	reverse
RNA	ribonucleic acid
ROOM	regulatory on/off minimization
S	Succinate
S.	*Saccharomyces*
SASP	small acid soluble proteins
SAT	Boolean satisfiability

SD   Shine-Dalgarno

SDS   sodium dodecyl sulfate

SigL   sigma factor L

SP   sporulation medium

SPABBATS   shortest pathway between the basal and target sets

Spec   spectinomycin

SPINE   Strep-protein interaction experiment

SPR   surface plasmon resonance

Tab.   table

TCA   tricarboxylic acid

TCR   transcription-coupled nucleotide excision repair

TCRF   transcription repair coupling factors

Tet   tetracycline

Tris   tris(hydroxymethyl)aminomethane

U   units

WT   wild type

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

YFP   yellow fluorescent protein

List of Publications

Gunka, K., Newman, J. A., Commichau, F. M., Herzberg, C., Rodrigues, C., Hewitt, L., Lewis, R. J. & Stülke, J. (2010) Functional dissection of a trigger enzyme: Mutations of the *Bacillus subtilis* glutamate dehydrogenase RocG that affect differentially its catalytic activity and regulatory properties. J. Mol. Biol. **400**: 815-827.

Flórez, L. A., Gunka, K., Polanía, R., Tholen, S. & Stülke, J. SPABBATS: A pathway-discovery method based on Boolean satisfiability that facilitates the characterization of suppressor mutants. BMC Systems Biology submitted

Gunka, K., Tholen, S., Neme, R., Freytag, B., Commichau, F. M., Herzberg, C. & Stülke, J. A high-frequency mutation in *Bacillus subtilis*: Requirements for the decryptification of the *gudB* glutamate dehydrogenase gene. Mol. Microbiol. submitted

Summary

Glutamate is the central amino group donor for all nitrogen containing compounds in the cell. In the Gram-positive soil-dwelling bacterium *Bacillus subtilis* glutamate is exclusively synthesized by the combined reactions of the glutamine synthetase and the glutamate synthase. Since the synthesis of glutamate requires 2-oxoglutarate that is derived from the tricarboxylic acid cycle, glutamate metabolism is an important intersection between the carbon and the nitrogen metabolism. Thus, the expression of the biosynthetic enzymes of the glutamate metabolism is highly regulated. The opposite reaction, the degradation of glutamate to 2-oxoglutarate is catalyzed by the glutamate dehydrogenase RocG. This enzyme is not capable of glutamate synthesis probably due to its low affinity for ammonium. In addition to its enzymatic function, RocG triggers the activity of the transcriptional regulator GltC that is essential for the expression of the *gltAB* operon encoding the glutamate synthase.

In this work, mutant variants of the RocG protein were isolated and analyzed. Single amino acid exchanges uncoupled the two functions of the RocG protein. One class of mutants is severely impaired in its catalytic activity but strongly inhibits the GltC protein, thus preventing the expression of the *gltAB* operon. The second class completely lost the ability to inhibit GltC but retained full enzymatic activity. The data provide an insight into the regulatory mechanism of the RocG-GltC interaction.

B. subtilis encodes a second glutamate dehydrogenase, GudB. In the laboratory strain 168 the *gudB* gene is cryptic due to a direct repeat of nine base pairs leading to a duplication of three amino acids in the active center of the enzyme. In a *rocG* mutant strain the *gudB* allele is readily decryptified upon growth on complex medium by the precise deletion of one half of the direct repeat.

This work shows that the *gudB* mutation occurs at an extremely high rate of 10^{-4}. Evidence was provided that a perfect direct repeat is crucial for the rapid decryptification of *gudB*. Moreover, by using an artificial mutagenesis system it turned out that transcription is involved in the high mutation rate of the *gudB* gene. Indeed, the transcription repair coupling factor Mfd is required for the decryptification of the *gudB* gene.

The results of this work emphasize the importance of glutamate homeostasis in *B. subtilis.*

Zusammenfassung

Glutamat ist der zentrale Aminogruppendonor für alle stickstoffhaltigen Verbindungen in der Zelle. In dem Gram-positiven Bodenbakterium *Bacillus subtilis* wird Glutamat ausschließlich in den gekoppelten Reaktionen der Glutaminsynthetase und der Glutamatsynthase synthetisiert. Da für den Aufbau von Glutamat 2-Oxoglutarat aus dem Citrat Zyklus benötigt wird, stellt die Glutamatsynthese einen wichtigen Knotenpunkt zwischen dem Kohlenstoff- und Stickstoffstoffwechsel dar. Deshalb ist die Expression der beteiligten Enzyme streng reguliert. Der Abbau von Glutamat wird von der Glutamatdehydrogenase RocG katalysiert. Durch die geringe Affinität zu Ammonium kann dieses Enzym nicht die Synthese von Glutamat katalysieren. Neben der enzymatischen Funktion reguliert RocG die Aktivität des Transkriptionsregulators GltC, der bedeutend für die Expression der Glutamatsynthase ist.

In dieser Arbeit wurden mutierte RocG-Varianten isoliert und charakterisiert. Durch einen Aminosäureaustausch wurden die beiden Funktion von RocG voneinander getrennt. Eine Klasse von Varianten inaktiviert GltC sehr stark und ist in ihrer zeigt enzymatischen Aktivität schwer beeinträchtigt. Die zweite Klasse kann GltC nicht mehr inhibieren, zeigt aber enzymatische Aktivität. Diese Ergebnisse geben einen Einblick in den regulatorischen Mechanismus der RocG-GltC Interaktion.

B. subtilis 168 kodiert für eine zweite Glutamatdehydrogenase, GudB. Im Laborstamm 168 ist das *gudB*-Gen durch eine direkte Wiederholung von neun Basenpaaren kryptisch. Beim Wachstum einer *rocG*-Mutante auf Komplexmedium wird das *gudB*-Allel schnell dekryptifiziert, wobei immer eine Hälfte der Sequenzwiederholung deletiert wird.

In dieser Arbeit wurde gezeigt, dass die *gudB*-Mutation mit einer extrem hohen Rate von 10^{-4} auftritt. Außerdem wurde bewiesen, dass eine perfekte Basen-wiederholung für die hohe Mutationsrate notwendig ist. Darüber hinaus zeigte sich in einem artifiziellen Mutagenese-System, dass die Transkription für die hohe Mutationsrate des *gudB*-Gens notwendig ist. Das Mfd Protein, welches die Transkription mit der DNA Reparatur koppelt, spielt eine entscheidende Rolle bei der Dekryptifizierung des *gudB* Gens.

Die Ergebnisse dieser Arbeit unterstreichen die herausragende Bedeutung der Glutamathomöostase in *B. subtilis*.

1.   **Introduction**

Prologue

Bacillus subtilis is a rod-shaped bacterium that was first described as *Vibrio subtilis* by C. G. Ehrenberg in 1835 but renamed in *Bacillus subtilis* by J. F. Cohen in 1872 (Gordon *et al.*, 1981). *B. subtilis* belongs to the phylum of the Firmicutes with low GC content and has become famous as a model organism for Gram-positive spore-forming bacteria. Naturally, the bacterium is common in the soil. In this environment the bacterium has to cope with fluctuating conditions which probably have led to its ability to adapt rapidly to stress conditions (Boylan *et al.*, 1993; Helmann *et al.*, 2001; Hecker *et al.*, 2007). These survival strategies as well as the close relation to pathogens such as *B. anthracis* directed the interest of research on *B. subtilis*. The secretion of proteases, attractive for the industry, the production of antibiotics and the development of competence are only a few aspects worth mentioning of the complex reactions to unfavorable growth conditions of *B. subtilis* (Hamoen *et al.*, 2003). The mechanism of DNA uptake and integration of this DNA in its genome, referred to as competence, allows genetic manipulation of *B. subtilis* (Harwood, 1990).The whole genome (4.2 Mbp) of *B. subtilis* was sequenced in 1997 (Kunst *et al.*, 1997) and a resequencing approach in 2009 provides even more reliable sequence information now (Barbe *et al.*, 2009). As the knowledge on *B. subtilis* has rapidly grown, new platforms were developed to manage all present and upcoming information. Recently, two complementary online databases namely SubtiWiki and SubtiPathways were published (Flórez *et al.*, 2009; Lammers *et al.*, 2010). These tools represent a compact accessibility of data which will benefit to the further research on *B. subtilis*.

1.1.   **The general role of glutamate in life**

The amino acid glutamate plays an important role in all domains of life. Glutamate fulfills a great variety of functions in all organisms. It serves as the major amino group donor for all nitrogen containing compounds of the cell. Glutamate is the most important amino group shuttle for the synthesis of amino acids, amino sugars and nucleotides (Ikeda *et al*, 1996; Reitzer, 2003). In *Escherichia coli*, 88% of the nitrogen in the biomass is derived from glutamate mainly via transamination reactions and only 12% is supplied by glutamine (Reitzer, 1996; Goss *et al*., 2001). Also in other microorganisms e.g. the yeast *Saccharomyces cerevisiae*, glutamate is the source of about 80% of the cellular nitrogen (Magasanik, 2003). The importance of glutamate in physiological processes is reflected by its high abundance. In *E. coli*, glutamate is the most abundant metabolite with over 40% of the total detectable intracellular metabolome (Bennett *et al*., 2009). Furthermore, glutamate is also present at the highest concentration of all metabolites in the opportunistic human pathogen *Pseudomonas aeruginosa* (Frimmersdorf *et al*., 2010). Although glutamate seems to be needed in high amounts in the cell, its synthesis has to be tightly regulated as it provides the link between carbon and nitrogen metabolism. For *de novo* glutamate synthesis 2-oxoglutrate is required. This carbon backbone is derived from the tricarboxylic acid (TCA) cycle, which is a part of the carbon metabolism. Therefore, the need for glutamate and the energy state of the cell have to be concerted.

From bacteria to archaea, glutamate is also of importance if the cell has to cope with external osmolarity changes. During high external osmolarity the outflow of water is prevented by the influx of potassium ions. Glutamate counteracts as an anion for the positively charged potassium. To mention only a few, some Archaea or Enterobacteria like *E. coli* are known to accumulate glutamate as a so called compatible solute in response to sudden hypersalinity of the environment (Kempf & Bremer, 1998; da Costa *et al*., 1998). In *B. subtilis*, glutamate is not the main osmoprotectant, but it is needed in high amounts during osmotic stress. *B. subtilis* accumulates proline in molar concentration as main compatible solute and proline is synthesized with glutamate as precursor (Kempf & Bremer, 1998).

For food industry, glutamate has to be produced in large scale to serve as an additive in a variety of products. Besides other organisms like *Brevibacteria*, especially the Gram-positive bacterium *Corynebacterium glutamicum* is used for L-glutamate

production, with an output of about 1.5 million tons per year (Sano, 2009; Rehm & Burkovski, 2010). Glutamate enhances the flavor of certain products e.g. potato chips or sauces. The effect on the human gustative nerve is called "umami" (Japanese) also referred to as savoriness (Kondoh & Torii, 2008). This term means that glutamate brings the flavor of a savory product such as meat, cheese or sea food to mind, foods that contain a lot of proteins. Nowadays, the application of glutamate is controversial as it is supposed to trigger a variety of symptoms e.g. the "Chinese restaurant syndrome", but this has not been directly supported by clinical data yet (Williams & Woessner, 2009; Jinap & Hajeb, 2010).

1.1.1. Glutamate metabolism in *Bacillus subtilis*

Bacteria prefer different nitrogen sources for their cellular processes. *E. coli* favors the inorganic molecule ammonium as nitrogen source (Reitzer, 2003). In contrast, *B. subtilis* utilizes the organic compound glutamine as the preferred source of nitrogen (Detsch & Stülke, 2003).

In the absence of glutamine, *B. subtilis* can also assimilate ammonium that is together with 2-oxoglutarate converted to glutamate (Fisher, 1999). At high external concentrations of ammonium, a small fraction of ammonia is present due to the chemical equilibrium. As an uncharged molecule, ammonia can freely diffuse into the cell. At low ammonia concentration or if the pH value drops, the positively charged ion ammonium has to be actively taken up by the cell. Under this condition the ammonium transporter NrgA is expressed that allows ammonium uptake (Wray *et al.*, 1994; Detsch & Stülke, 2003). Furthermore, within the cell ammonium is formed by the degradation of amino acids and nucleotides (Fisher & Débarbouillé, 2002).

In *B. subtilis*, glutamate is exclusively synthesized via two reactions in the GOGAT cycle (Deshpande & Kane, 1980; Fisher 1999). The two reactions in the GOGAT cycle are performed by the glutamine synthetase (GS) and the glutamate synthase [GOGAT (glutamine-oxoglutarate-aminotransferase)] which is a heterodimeric enzyme encoded in the *gltAB* operon. In the first step of the cycle, ammonium is transferred to glutamate catalyzed by the GS in an ATP-dependent manner (see Fig. 1.1). This reaction leads to the formation of glutamine. In the following reaction, the GOGAT catalyzes the transfer of the δ-amino group of glutamine to 2-oxoglutarate that

is derived from the TCA cycle. In this reaction, the co-factor NADPH is oxidized and two molecules of glutamate are produced. One molecule of glutamate is needed to balance the cycle and the other one can be used as nitrogen donor for the anabolism of the cell. As described above, the biosynthesis of glutamate consumes 2-oxoglutarate as carbon backbone. That is why glutamate synthesis is a very important link between the carbon and the nitrogen metabolism in the cell.

The opposite reaction, the deamination of glutamate to 2-oxoglutarate and ammonium allows the cell to utilize glutamate as a source of carbon. This reaction is catalyzed by the glutamate dehydrogenase (GDH) that is RocG in *B. subtilis* (see Fig.1.1). This enzyme is only capable of catalyzing the degradation of glutamate and not of its synthesis (Belitsky & Sonenshein, 1998). Probably, RocG lacks the ability to perform the reverse reaction *in vivo* due to its low affinity for ammonium (k_m value for ammonium of 18 mM).
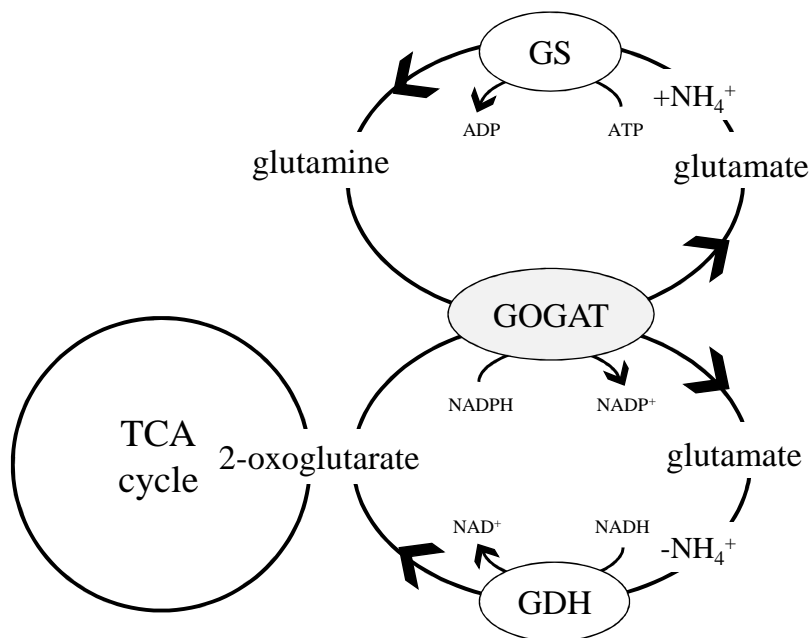
Fig. 1.1 – The GOGAT cycle in *B. subtilis.*

In *B. subtilis* glutamate is exclusively synthesized by the combined reactions of glutamine synthetase (GS) and the glutamate synthase (GOGAT). In contrast to other organisms, e.g. *E. coli* or *C. glutamicum*, the glutamate dehydrogenase (GDH) is only capable of glutamate degradation and not of its synthesis.

It was shown that the *B. subtilis* GDH can synthesize glutamate with the excess of ammonium *in vitro*, thus it is hypothesized that the intracellular ammonium concentration in *B. subtilis* is not sufficient to facilitate this reaction.

In *E. coli,* the GDH can catalyze the anabolic reaction and forms glutamate in a reductive amination of 2-oxoglutarate with ammonium (Reitzer, 2003). This enzyme exhibits a higher affinity for ammonium than the GDH of B. subtilis (k_m for ammonium of 2.5 mM and 18 mM, for *E. coli* GDH and *B. subtilis* RocG, respectively; Sharkey & Engel, 2008; Gunka *et al*, 2010). Therefore, *E. coli* GDH is only active at high cellular concentrations of ammonium. Under ammonium excess the GS is poorly expressed and furthermore inhibited by a covalent modification and glutamate is produced by the GDH (Arcondéguy *et al.*, 2001; Reitzer, 2003; Ninfa & Jiang, 2005). If the ammonium levels drops in the cell, glutamate is synthesized in the GOGAT cycle. Switching between these two pathways, the GOGAT cycle and the GDH reaction, depending on the ammonium supply is very common in microorganisms. Also *C. glutamicum* can circumvent the energy-consuming reaction of the GS at high ammonium concentrations and forms glutamate catalyzed by the GDH (Rehm & Burkovski, 2010).

1.1.2. **The regulation of glutamate metabolism in *B. subtilis***

As mentioned above, the glutamate metabolism is a very important intersection between the carbon and the nitrogen metabolism. In order to assure that the carbon and nitrogen metabolisms are balanced, glutamate synthesis has to be tightly regulated.

In *B. subtilis,* this regulation is achieved by a complex network including the action of transcriptional regulators, protein-protein interactions and the impact of metabolites on enzyme activities and protein complex formation.

The GS is encoded in the *glnA* gene that is part of the *glnAR* operon (Gardner & Aronson 1984; Strauch *et al.*, 1988). The expression of the GS is regulated in response to nitrogen availability by the two transcriptional regulators TnrA and GlnR (see Tab. 1.1 for an overview on regulation; Schreier *et al.*, 1989; Wray *et al.*, 1996; Fisher, 1999; Fisher & Débarbouillé, 2002).

Tab. 1.1 - Regulation of the key enzymes of glutamate metabolism in *B. subtilis.*

Target genes	Regulator	Regulation and effectors
glnA	GlnR	Negative, in complex with feedback inhibited GS by glutamine
	TnrA	Positive in the absence of glutamine; in the presence of glutamine inactivated by interaction with feedback inhibited GS
gltAB	TnrA	Activation in the presence of ammoinum
	GltC	Activation by 2-oxoglutarate; inhibition by glutamate
	RocG	Inhibition of GltC by protein-protein interaction in the presence of glutamate
rocG	CcpA	Inhibition by CcpA in complex with HPr phosporylated at serine 46
	RocR	Activation in the presence of arginine, ornithine, citrulline or proline
	SigL	Activation dependent on RocR
	AhrC	Activation, requires arginine
	AbrB	Inhibtion

TnrA is a global regulator of nitrogen metabolism. During nitrogen limitation, TnrA activates among other genes that are required for the utilization of alternative nitrogen sources, the expression of the *glnA* gene, whereas under nitrogen excess, the GS expression is repressed by GlnR (Wray *et al.*, 1996; Wray *et al.*, 2001). The GS influences both, TnrA dependent activation und the transcriptional repression exerted by GlnR. At high concentration of glutamine, the GS is subject to feedback inhibition (Deuel & Prusiner, 1974). The feedback inhibited GS prevents TnrA from DNA binding by a protein-protein interaction thus inhibiting its own expression if glutamine is available (Wray *et al.*, 2001). Moreover, the feedback inhibited GS activates the DNA binding activity of GlnR and stabilizes GlnR-DNA interaction which represses the *glnA* expression (Fisher & Wray, 2008). Due to this regulation, the GS is only expressed at

high level, if nitrogen is limiting and glutamate has to be synthesized. In this regulation mechanism the GS acts as a trigger enzyme that controls its own synthesis.

The *gltAB* operon codes for the GOGAT in *B. subtilis* (Deshpande & Kane, 1980). The big subunit (168 kDa) encoded in the *gltA* gene is involved in glutamine deamination. The *gltB* gene codes for the small subunit (55 kDa) that is responsible for NADPH binding and catalyzes the transfer of the amino group to 2-oxoglutarate (Matsuoka & Kimura, 1986). The expression of the GOGAT depends on both, the carbon and the nitrogen sources (see Tab. 1.1). If glucose and ammonium are present, the *gltAB* operon is highly transcribed (see Fig. 1.2A). The expression is low in the presence of a poor carbon source or if glutamate or its precursor arginine, ornithine or proline is available (Bohannon *et al.*, 1985; Wacker *et al.*, 2003; Belitsky & Sonenshein, 2004). In the absence of ammonium, repression of the *gltAB* operon is mediated by the global regulator TnrA (Belitsky *et al.*, 2000). Thus, the GOGAT is only expressed, if ammonium is present and can be utilized in the GOGAT cycle. Furthermore, for *gltAB* expression the activity of the transcriptional regulator GltC is required (Bohannon & Sonenshein, 1989; Belitsky & Sonenshein, 1995). The GltC protein is encoded directly upstream of the *gltAB* operon in divergent orientation, whereas the promoter regions of *gltC* and *gltAB* are overlapping (Belitsky *et al.*, 2000). GltC is a member of the LysR transcription regulator family that are characterized by a conserved structure with an N-terminal helix-turn-helix motif, necessary for DNA binding, and a C-terminal co-inducer-binding domain (Bohannon & Sonenshein, 1989; Maddocks & Oyston, 2008). In addition to the activation of the *gltAB* operon, GltC is a negative autoregulator of its own synthesis. The *gltCAB* locus exhibits three GltC binding boxes referred to as Box I, II and III (Belitsky *et al.*, 1995; Belitsky & Sonenshein, 1995; Picossi *et al.*, 2007). Binding of GltC to Box I is responsible for the negative autoregulation, whereas binding to Box II affects both the autoregulation as well as the activation of the *gltAB* transcription. *In vitro* analyses showed the impact of low-molecular weight effectors on GltC binding. At a high glutamate concentration GltC binds to the Boxes I and III leading to the repression of the GOGAT expression (see Fig. 1.2B). In the presence of 2-oxoglutarate, GltC binds to Box I and Box II resulting in a high transcription level of the *gltAB* operon (Picossi *et al.*, 2007*)*. This regulation by low-molecular weight effectors might contribute to a coordinated expression of the GOGAT in dependence of glutamate availability.
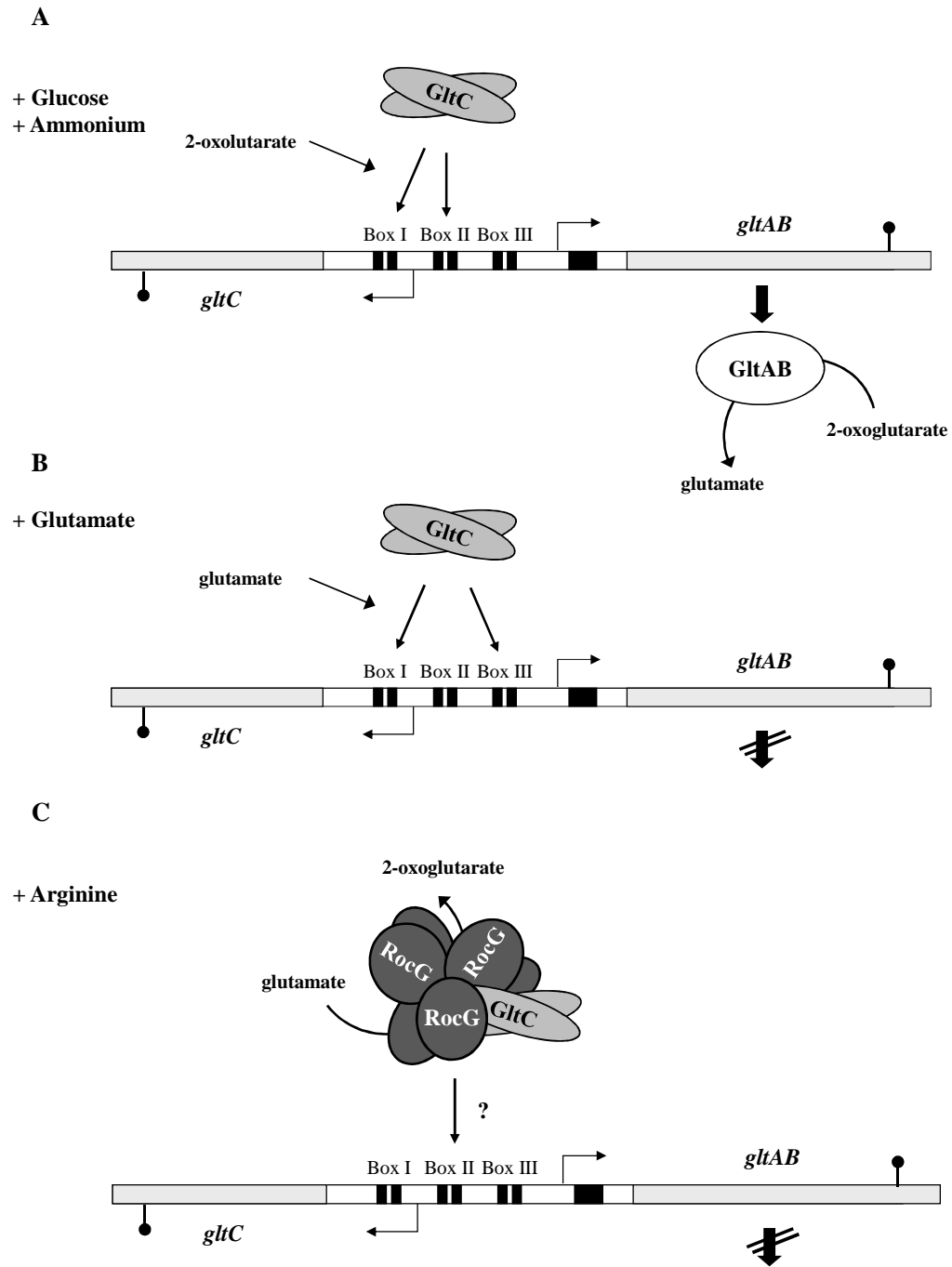
Fig. 1.2 – GOGAT expression in response to the carbon and nitrogen source.

(A) In the presence of glucose and ammonium GltC binds to the Boxes I and II which results in GOGAT expression. Moreover, TnrA can induce the transcription of the *gltAB* operon. (B) In the absence of glucose and the presence of glutamate, GltC binds to the Boxes I and III leading to an inhibition of the transcription of the *gltAB* operon. (C) In the presence of Arginine RocG is expressed and prevents GltC by protein-protein interaction from the activation of the GOGAT expression.

The expression pattern of the GOGAT and the GDH RocG is mutually exclusive. This is physiologically important in order to avoid a futile cycle of glutamate production and subsequent glutamate degradation. Furthermore, the GOGAT is constitutively expressed in a *rocG* mutant strain implying that the RocG protein prevents *gltAB* transcription (Wacker *et al.*, 2003; Commichau *et al.*, 2007b). By an *in vivo* cross-linking approach it was demonstrated that the RocG protein directly interacts with GltC thus inhibiting the capacity of GltC to activate *gltAB* transcription (Commichau *et al.*, 2007a). The inhibition of GltC by RocG seems to be furthermore dependent on the presence of the low-molecular weight effector glutamate (Gunka *et al.*, 2010). It has not been analyzed yet, how RocG inhibits the binding capacity of GltC to the promoter regions.

The expression of the *rocG* gene coding for the catabolically active glutamate dehydrogenase is also subject to a complex regulation exerted by both the carbon and the nitrogen sources (see Tab. 1.1). The *rocG* gene is induced by arginine, ornithine, citrulline or proline (Belitsky & Sonenshein, 1998). Therefore, *B. subtilis* can utilize arginine as the sole source of carbon. Arginine is degraded in several reactions to glutamate that can be consumed by RocG to form 2-oxoglutarate (Calogero *et al.*, 1994). The transcription of the *rocG* gene depends on the alternative sigma factor L (SigL) as well as the two transcription regulators RocR and AhrC (Débarbouillé *et al.*, 1991; Calogero *et al.*, 1994; Gardan *et al.*, 1995). Besides the *rocG* gene the two operons *rocABC* and *rocDEF* are part of the RocR regulon that is responsible for the uptake and utilization of arginine. The RocR protein acts together with SigL in the activation of RocG expression (Belitsky & Sonenshein, 1999; Ali *et al.*, 2003). The inducers of RocR were identified to be ornithine or citrulline (Gardan *et al.*, 1997). AhrC is a global regulator of the arginine metabolism that induces the r*ocABC* and *rocDEF* genes for arginine utilization but represses the gene products for arginine synthesis (Klingel *et al.*, 1995; Miller *et al.*, 1997). It was shown that also the *rocG* gene is positively regulated by AhrC but the binding site of AhrC to the promoter region has not been analyzed yet (Commichau *et al.*, 2007b). Recently, the *rocG* gene was identified as a target of AbrB that is a repressor that is active during exponential growth phase (Chumsakul *et al.*, 2010).

In addition to this complex regulation exerted by nitrogen sources the *rocG* gene is subject to carbon catabolite repression (Belitsky & Sonenshein, 2004). In the

presence of a preferred carbon source e.g. glucose, the CcpA protein (catabolite control protein A) interacts with the HPr protein phosphorylated at a serine 46 residue. This complex represses the transcription of the *rocG* gene (Belitsky *et al.*, 2004).

As mentioned above, the RocG protein is involved itself in the complex regulation of the glutamate metabolism in *B. subtilis.* Additionally to its function as the glutamate dehydrogenase it controls GltC activity in the presence of glutamate or a precursor of it (see Fig. 1.2C) (Commichau *et al.*, 2007a). Therefore, RocG is classified as a trigger enzyme that controls gene expression in response to the nitrogen level of the cell to avoid a futile cycle of glutamate synthase and degradation (Commichau & Stülke, 2008).

In addition to the *rocG* gene*, B. subtilis* possesses a second gene, *gudB,* that codes for a glutamate dehydrogenase. Both proteins share an identity of 74% and were crystallized as homohexamers (Gunka *et al.*, 2010). In the laboratory strain 168, only the product of the *rocG* gene is enzymatically active (Belitsky & Sonenshein, 1998). The *gudB* gene is characterized by a direct duplication of nine base pairs in the open reading frame. This direct repeat of nine base pair leads to a direct duplication of three amino acids directly located in the active site of the enzyme, which results in an inactive protein. In contrast to the RocG expression that is highly regulated, the *gudB* gene is constitutively transcribed at a high level (Belitsky & Sonenshein, 1998; Gunka *et al.*, subm.).

The cultivation of a *rocG* mutant strain on rich medium or the presence of glutamate as single carbon source provokes the reactivation of the *gudB* gene (Belitsky & Sonenshein, 1998; Commichau *et al.*, 2008). The gain of function process is characterized by the precise excision of nine base pairs of the direct repeat leading to the expression of an active glutamate dehydrogenase designated GudB1. The GudB1 protein can fully replace the trigger enzyme RocG. GudB1 is not only able to take over the function of RocG in glutamate degradation but is also capable of controlling GltC activity (Commichau *et al.*, 2007b).

1.1.3.   **Impact of glutamate metabolism on phenotypical features of *B. subtilis***

The enzymes involved in glutamate metabolism are not only important for the appropriate supply with glutamate of the cell during growth. The loss of the GOGAT

has a severe impact on cell physiology in terms of differentiation processes, emphasizing the pivotal role of glutamate for the cell.

A survival strategy of *B. subtilis* is the formation of biofilms. Biofilms are multicellular communities that are surface associated (López *et al*., 2010). Wild strain isolates of *B. subtilis* form biofilms that are highly differentiated in architecture whereas domesticated strains can only form relatively undifferentiated biofilms (Branda *et al*., 2004).

In the laboratory strain 168, the loss of the GOGAT results in a strong defect on biofilms formation. This defect cannot be suppressed by the addition of glutamate to the medium, indicating that not the capacity of the GOGAT to produce glutamate is the factor needed for normal biofilms assembly (Chagneau & Saier, 2004). However, the detailed influence of the GOGAT on biofilms formation has not been characterized yet.

Interestingly, it was shown that the GOGAT affects even more differentiation processes in *B. subtilis*. Sporulation is a survival strategy that is induced under starvation and high cell density. The dormant cell can overcome nutrient limitation allowing the survival even under the most adverse conditions (Claverys & Håvarstein, 2007). A GOGAT deficient *B. subtilis* strain is impaired in sporulation and germination. Moreover, the resistance of the spores in a GOGAT mutant strain is reduced (Ruzal & Sanchez-Rivas, 2003). The addition of glutamate restores the spore resistance partially in a GOGAT mutant strain but the outgrow of the spores is still strongly impaired (Ruzal & Sanchez-Rivas, 2003). Spores contain small acid soluble proteins (SASP) in high abundance. These SASPs mainly contribute to spore resistance to UV, heat and osmotic strength (Hackett & Setlow, 1988). The defect of a *gltAB* mutant in spore resistance properties linked to SASPs implies that the GOGAT activity is very important for the biosynthesis of SASPs (Ruzal & Sanchez-Rivas, 2003). Moreover, the GOGAT is needed for the return to a vegetative lifestyle of *B. subtilis* (Ruzal & Sanchez-Rivas, 2003).

The loss of the GDH RocG results in a severe growth defect of *B. subtilis* on rich medium (see Fig. 1.3). This is probably caused by the accumulation of toxic metabolites that are formed during the degradation of arginine (Belitsky & Sonenshein, 1998). This growth defect is suppressed by the activation of the cryptic glutamate dehydrogenase gene *gudB* (Belitsky & Sonenshein, 1998, Gunka *et al*., subm.).
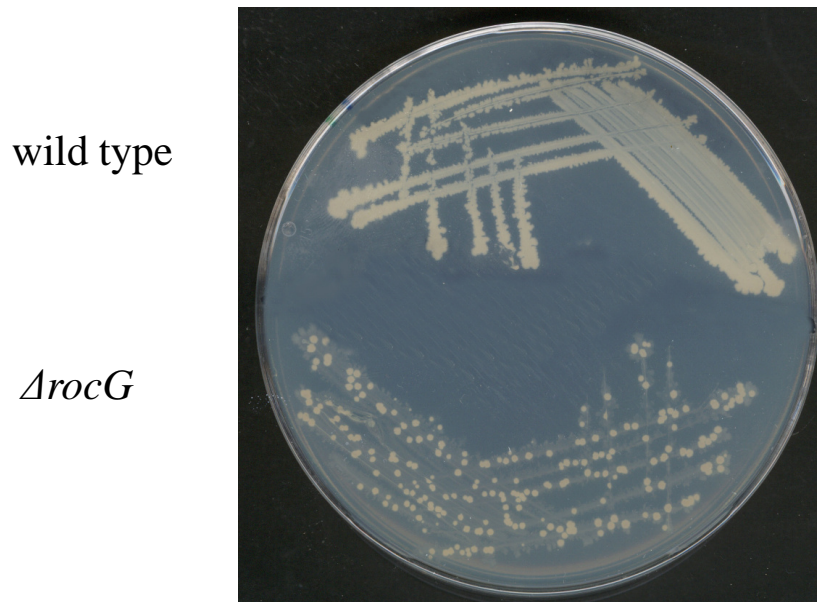
Fig. 1.3 – The growth defect of a *B. subtilis rocG* mutant on rich medium.

B. subtilis wild type grows well on rich medium (upper part of the agar plate). In contrast a *rocG* mutant forms translucent colonies and shows an impaired growth (lower part in the background). By the decryptification of the *gudB* allele, coding for the second glutamate dehydrogenase, this defect is suppressed. The *gudB1* mutants grow as well as the wild type (papillae in the lower part).

1.1.4. **The importance of glutamate homeostasis**

The rapid decryptification of the *gudB* allele in a *rocG* mutant is probably the most extensively studied suppressor mutation observed after perturbating the glutamate homeostasis in *B. subtilis*. In addition to the *gudB1* mutation a set of other mutations concerning other genes involved in glutamate metabolism were observed in *B. subtilis*.

A *rocG gudB* double mutant strain that is obviously not able to activate the *gudB* gene upon cultivation on rich medium accumulates other mutations that restore growth. A mutant was isolated that had acquired a frameshift mutation in the *gltB* gene resulting in the expression of a truncated protein and glutamate auxotrophy of this strain (Commichau *et al*., 2008). The loss of a functional GOGAT seems to balance the glutamate homeostasis of a GDH deficient strain on rich medium. Additionally to mutations affecting the catabolic part of glutamate metabolism, also mutation events occur concerning the anabolic branch of glutamate metabolism in *B. subtilis*. A mutant defective in the transcription activator GltC is auxotrophic for glutamate, as the

expression of the GOGAT cannot be induced. However, mutants were found that show an elevated expression of *gltA* in the absence of the GltC protein. These mutants had acquired a gain of function mutation in the LysR-type regulator GltR (Belitsky & Sonenshein, 1997). This mutation allows GltR to replace GltC in the activation of *gltAB* transcription.

As described above, a *rocG gudB* suppressor mutant was isolated that had accumulated a frameshift mutation namely a two base pair deletion in the *gltB* gene. The cultivation of this glutamate auxotrophic mutant on minimal medium with ammonium as nitrogen source resulted in the reactivation of a functional *gltB* gene by a third base pair deletion that restores the *gltB* reading frame (Commichau *et al.*, 2008). These examples clearly demonstrate the robustness of glutamate metabolism according to the availability of the nitrogen source.

Not only in *B. subtilis* but also in Enterobacteria, suppressor mutations balancing the glutamate homeostasis are known. In 2007, Yan carried out an intensive study on the maintenance of the glutamate pool in *Salmonella typhimurium* and *E. coli.* By using mutant strains defective in the GOGAT enzyme a series of suppressor mutants were isolated. In the absence of the GOGAT the glutamine produced by the GS cannot be converted to glutamate. In this strain glutamate can only be synthesized by the GDH, which has a lower affinity for ammonium compared to the GS. This results in an increase of the cellular glutamine and a drastic depletion of the glutamate pool of the cell. The disproportion is balanced by several suppressor mutations. One kind of suppressor mutation affected the GDH activity. For one mutant it was shown that the mutation was located in the *gdhA* promoter region leading to a high level of GDH expression. Another mutation affected the *glnA* gene and reduced the activity of the GS. Both suppressor mutations compensate the absence of the GOGAT and restore the glutamate/glutamine ratio to wild type level (Yan, 2007).

1.2.   **The controversy of adaptive mutations**

In order to maintain genomic stability, deleterious mutations have to be avoided by the cell. Mutations are changes in the nucleotide sequence including base pair substitutions, deletions or insertions. Nevertheless, without the occurrence of beneficial mutations the process of evolution would not have been possible. Natural selection as a driving force in evolution was first described by Darwin in 1859 (Darwin, 1859). The classic experiment of Luria and Delbrück, the fluctuation test, showed that mutations are even spontaneously acquired without selection pressure (Luria & Delbrück, 1943). They incubated bacteria cells over a certain period and plated the culture on media treated with bacteriophages. If the presence of the phages would induce mutations in terms of resistance of the bacteria, it would be expected to find almost the same number of resistant bacteria on each plate. Strikingly, the number of resistant bacteria varied extremely on each plate. This experimental setup demonstrated that mutations occur permanently. Before the contact with the phages, the bacteria in the cultures have already randomly acquired mutations that led to resistance. If the mutation event occurs early in the culture, resistant bacteria are highly representative, whereas a late mutation results in only a small number of resistant bacteria. Based on this experiment Lea and Coulson derived a distribution, named Luria–Delbrück distribution that considers the observed variance and allows the determination of the mutation rate (Lea & Coulson, 1949).

However, experiments with *E. coli lac⁻* mutants challenged the model proposed by Luria and Delbrück (Cairns *et al*., 1988). In contrast to Luria and Delbrück who applied lethal selection, the Cairns system imposes nonlethal stress on cells. An *E. coli* strain deficient in the *lacZ* gene cannot utilize lactose. *E. coli* was shown to possess a cryptic gene that is capable of the hydrolysis of lactose upon decryptification. The experiment demonstrated that activation of this cryptic gene occurs much faster in the presence of lactose than in its absence. Based on this result, it is tempting to speculate, that bacteria might control which mutation occurs in response to selective pressure (Cairns *et al*., 1988). In 1991, Cairns and Foster modified this system and used an *E. coli* strain that is chromosomally *lacZ* negative but harbors a plasmid which encodes a *lacZ* gene with a frameshift mutation. The gene product exhibits a remaining hydrolysis activity of 2% of the wild type enzyme and allows poor growth of lactose. Under unrestricted growth, a gain of function mutation occurred at a rate of 10^{-8} whereas with lactose as carbon

source the mutation rate increased hundred fold (Cairns & Foster 1991; Roth *et al.*, 2006). This experiment suggests the generation of adaptive mutations during selection and argues the statement given by Luria and Delbrück.

1.2.1.   **The frequency of mutations**

The impact of mutations on the organism can be extremely different. The mutation can be deleterious, silent or can even increase the fitness. In *E. coli,* the chance of acquiring a beneficial mutation is about 10^{-9}, whereas the rate of deleterious mutations is in range of 10^{-4} to 10^{-5} (Denamur & Matic, 2006). Most of the deleterious mutations are purged from the population whereas beneficial mutations increase the overall fitness (Imhof & Schlötterer, 2001). Studies on *E. coli* revealed a mutation rate in the range of 10^{-7} to 10^{-8} during replication (Kunkel *et al*, 2004). Most of the errors are eliminated by the DNA mismatch repair system (MMR) and are repaired in favor of the base on the template strand. The MMR is evolutionarily highly conserved. The process involves recognition of the mismatch, incision, nucleotide removal and strand resynthesis (Polosina & Cupples, 2010). In *E. coli*, recognition of mismatches in the hemimethylated DNA is carried out by the MutS protein. Subsequently, the MutL protein binds to the MutS:DNA complex in an ATP-dependent manner but without ATP hydrolysis. The UvrD helicase is recruited to separate the strands. By complex formation with MutSL the endonucleolytic activity of the MutH protein is activated and MutH excises the mismatch (Modrich & Lahue, 1996). The resulting gap is filled in by DNA polymerase III, and the DNA ligase carries out the ligation (Polosina & Cupples, 2010). *B. subtilis* possesses a functional equivalent of the *E. coli* MMR system with some striking differences. *B. subtilis* lacks MutH and UvrD homologues and discrimination between the parental and the newly synthesized strand is not mediated by the methylation state but by the presence of single-strand breaks (Rossolillo & Albertini, 2001).

Another cellular pathway, transcription, turned out to be important for the emergence of mutations.

1.3.   Transcription-associated DNA repair

Besides DNA replication, the process of transcription is important for mutagenesis. During transcription, the nontranscribed strand is subject to mutations. The phenomenon is called transcription-associated mutagenesis. When the RNA polymerase forms a complex with the transcribed strand, the nascent nontranscribed strand is not protected by base pairing anymore and becomes vulnerable to mutagens. The mutations that arise are not expressed until replication (Kim *et al.*, 2010). A study on *E. coli* suggests that actively transcribed genes acquire more cytosine to thymine mutations due to deamination in the nontranscribed than in the transcribed strand (Beletskii & Bhagwat, 1996). Moreover, an increase in mutations that changed a TGA stop codon in the non-transcribed strand to a sense codon was observed at high transcription (Klapacz & Bhagwat, 2005).

Another mechanism that is supposed to generate mutations during transcription is the transcription-coupled nucleotide excision repair (TCR) that acts selectively on the transcribed strand. TCR is a subpathway of nucleotide excision repair (NER) in which DNA lesions are recognized that interfere with the proceeding of the RNA polymerase in a transcribed gene (Hanawalt & Spivak, 2008). There is evidence that the mechanism of TCR is widespread among bacteria and eukaryotes but it has not been demonstrated in archaea yet (Eisen & Hanawalt, 1999; Fousteri & Mullenders, 2008). The process of TCR was first described for eukaryotes by demonstrating that pyrimidine dimers were preferentially removed by nucleotide excision repair in transcribed regions (Bohr *et al.*, 1985).

1.3.1.   The role of Mfd in cellular physiology

The Mfd protein is the driving force in TCR and is highly conserved in microbial genomes (Selby & Sancar, 1993; 1994). Mfd stands for mutation frequency decline due to its originally identified effect on mutagenesis (Witkin, 1956; 1966). Later on, Mfd was shown to mediate TCR and has been called transcription repair coupling factor (TCRF) that provides the continuation of transcription after DNA lesions (Selby & Sancar, 1993; 1994; 1995a; 1995b). Mfd recognizes stalled RNA polymerase caused by a DNA lesion. After displacing the RNA polymerase and its transcripts, Mfd recruits the DNA excision repair machinery that resolves the lesion and allows the restart of

transcription (Roberts & Park, 2004; Borukhov *et al*., 2005; Deaconescu *et al*., 2006). Evidence was provided that Mfd operates with the nucleotide excision repair system (NER). *In vitro* analysis showed that Mfd stimulates the UvrABC system (Selby & Sancar, 1993). The UvrABC system cleaves damaged DNA in an ATP-dependent manner. The UvrA protein initially recognizes a DNA lesion and forms an ATP-dependent heterodimer with the UvrB protein. The ATP hydrolysis is performed by the ATPase activity of the UvrB protein that is stimulated upon contact with UvrA (Truglio *et al*., 2006). In the TCR pathway, Mfd is supposed to recruit UvrA to the damaged DNA (see Fig. 1.4). For the recruitment, the N-terminal part of the Mfd protein seems to be necessary as it has a high homology to the region in the UvrB protein that is essential for the UvrA-UvrB interaction (Selby & Sancar, 1993; Assenmacher *et al*. 2006). After loading UvrB onto the damaged DNA, UvrA dissociates and UvrB forms a stable complex with the DNA. Subsequently, UvrC binds to the UvrB:DNA complex and initiates the cleavage of the 3' as well as the 5' end of the DNA (Truglio *et al*., 2006). In *E. coli,* the UvrD protein, a helicase II, removes the excised nucleotides and displaces UvrC. Then, the gap is filled in by the DNA polymerase I and ligated by a DNA ligase (Sancar, 1994). The genome of *B subtilis*, encodes also the UvrABC proteins but lacks an UvrD homologue (Smith *et al*., 2001; 2002). Besides the Mfd protein, other proteins are supposed to modify RNA polymerase activity. The transcription elongation factors GreA ad GreB suppress RNA polymerase arrest. While GreA can only prevent RNA polymerase from pausing, GreB can even reactivate transcription after RNA polymerase arresting at a roadblock (Borukhov *et al*., 2005). NusA is an essential transcription termination factor and fulfills a variety of roles. NusA can facilitate transcription termination caused by hairpin-structures. Moreover, it is necessary to synchronize the process of transcription and translation by inducing RNA polymerase pausing (Borukhov *et al*., 2005). With the exception of Mfd, the other transcription factors have not been proven to be involved in the TCR.
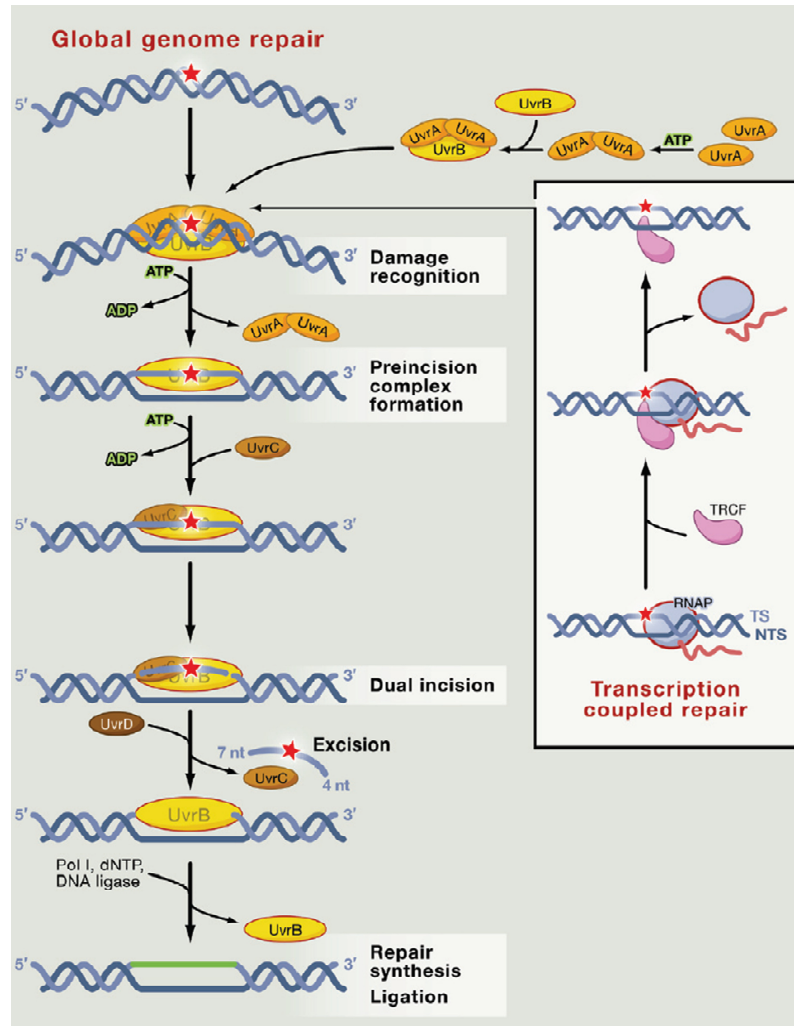
Fig. 1.4 – The role of Mfd in NER (Guo *et al.*, 2010).

The TCRF, Mfd, recognizes a stalled RNA polymerase at a DNA lesion and recruits the NER system to solve the lesion. (The figure is reprinted with kind permission of Elsevier.)

In *B. subtilis* the role of Mfd in mutagenesis was subject to intensive investigations in the last two decades. First evidence for the role of Mfd in TCR was provided by showing that *B. subtilis* Mfd can displace the RNA polymerase *in vitro* (Ayora *et al.*, 1996). Furthermore, Mfd was shown to be involved in the carbon catabolite repression by displacing RNA polymerase stalled at downstream *cre* sites (Zalieckas, 1998). Additionally, the role of Mfd in the phenomenon called adaptive or stationary phase mutagenesis in *B. subtilis* was discovered. The stationary phase mutagenesis means the occurrence of a hypermutable subpopulation, when a nondividing population has to cope with e.g. amino acid starvation pressure or as in the

Cairns system with carbon end energy starvation (Sung & Yasbin, 2002). Mfd turned out to be involved in the generation of mutants during amino acid starvation. *B. subtilis* strains that are auxotrophic for certain amino acids were incubated on medium lacking one of the required amino acids. In a short period of time mutants prototrophic for the previously required amino acid were isolated, indicating the occurrence of stationary phase mutagenesis. A *mfd* mutant strain was strikingly diminished in its capacity to generate prototrophic revertants in stationary phase implying an important role of the Mfd protein in the process of mutagenesis (Ross *et al.*, 2006). Recently, the importance of transcription in the stationary phase mutagenesis was pointed out in *B. subtilis*. In this study, a strain that bore a missense mutation in an allele for the synthesis of a particular amino acid was used. Under selection a high number of prototrophic revertants were found, but only if the gene was transcribed (Pybus *et al.* 2010). This is in agreement with studies on the yeast *S. cerevisiae* that showed a high frequency of mutations according to high transcription level (Kim *et al.*, 2007). In an isogenic *mfd* mutant strain not only the transcription level of the respective gene was reduced but moreover, the rate of revertants was significantly decreased (Pybus *et al.* 2010). This result strongly suggests an important role of Mfd in the stationary phase mutagenesis and the accumulation of adaptive mutations.

Recently, another role of the Mfd protein in *E. coli* was demonstrated. *In vitro* analysis showed that Mfd is necessary to clear the conflict between the DNA and the RNA polymerase in a head-on collision. The Mfd protein arranges to solve this conflict in favor of replication by pausing transcription (Pomerantz & O'Donnell, 2010).

1.4.   **Aims of this work**

In *B. subtilis*, the glutamate dehydrogenase RocG is a bifunctional trigger enzyme (Commichau *et al*., 2007a). On the one hand, RocG catalyzes the degradation of glutamate. On the other hand, RocG directly interacts with the transcriptional regulator GltC, thus controlling the ability of GltC to induce the expression of the glutamate synthase. The region in the RocG protein that is necessary for interaction with GltC and the direct mechanism of regulation have not been clarified yet. In order to study the mechanism of RocG-GltC interaction it was intended to isolate mutant RocG proteins that lack the ability to control GltC activity. These mutant proteins would be an appropriate subject for analyzing the mode of interaction between RocG and GltC.

The inactive glutamate dehydrogenase GudB is readily decryptified in a *rocG* mutant strain upon growth on complex medium. This suppressor mutation always occurs as the precise deletion of nine base pairs of the direct repeat in the *gudB* gene and complements the severe growth defect of the *rocG* mutant on complex medium. To understand whether this mutation is acquired with an increased rate compared to other mutations, the occurrence of this mutation has to be quantified. In the case of an increased rate it would be of interest to analyze the mechanism and the requirements for the decryptification of the *gudB* allele.

2.   **Chapter 2**

Functional dissection of a trigger enzyme: Mutations of the *Bacillus subtilis* glutamate dehydrogenase RocG that affect differentially its catalytic activity and regulatory properties

Abstract

Any signal transduction requires communication between a sensory component and an effector. Some enzymes engage in signal perception and transduction, as well as in catalysis, and these proteins are known as "trigger" enzymes. In this report, we detail the trigger properties of RocG, the glutamate dehydrogenase of *Bacillus subtilis*. RocG not only deaminates the key metabolite glutamate to form α-ketoglutarate but also interacts directly with GltC, a LysR-type transcription factor that regulates glutamate biosynthesis from α-ketoglutarate, thus linking the two metabolic pathways.

We have isolated mutants of RocG that separate the two functions. Several mutations resulted in permanent inactivation of GltC as long as a source of glutamate was present. These RocG proteins have lost their ability to catabolize glutamate due to a strongly reduced affinity for glutamate. The second class of mutants is exemplified by the replacement of aspartate residue 122 by asparagine. This mutant protein has retained enzymatic activity but has lost the ability to control the activity of GltC. Crystal structures of glutamate dehydrogenases that permit a molecular explanation of the properties of the various mutants are presented. Specifically, we may propose that D122N replacement affects the surface of RocG. Our data provide evidence for a correlation between the enzymatic activity of RocG and its ability to inactivate GltC, and thus give insights into the mechanism that couples the enzymatic activity of a trigger enzyme to its regulatory function.

Introduction

All signal transduction processes require a component that senses environmental change, and this task can be fulfilled by regulatory proteins that bind directly to specific effector molecules. Signal perception and regulatory output, however, are frequently determined by individual proteins. This mode of signal transduction is arguably best exemplified by two-component signaling systems. By contrast, metabolism is regulated ultimately by the availability of metabolites, and the players that have direct access to this information are the metabolic enzymes. A subset of these enzymes is able to engage in signal transduction and to perform in catalysis, and these are collectively termed "trigger" enzymes (Commichau & Stülke, 2008).

We are interested in the regulation of nitrogen metabolism in the Gram-positive soil-dwelling bacterium *Bacillus subtilis*, whose preferred sources of nitrogen are glutamine and ammonium. Ammonium is assimilated by the synthesis of glutamine by glutamine synthetase, encoded by the *glnA* gene. Glutamine is subsequently used for the reductive amination of 2-oxoglutarate by glutamate synthase, encoded by the genes of the *gltAB* operon (see Fig. 2.1A) (Fisher, 1999). Expression of this operon is controlled by two transcription factors. First, TnrA protein represses the *gltAB* operon in the absence of ammonium (Wray *et al*., 1996; Belitsky *et al*., 2000; Yoshida *et al*., 2003). Second, the LysR-type regulator GltC activates transcription of the operon in the presence of sugars and in the absence of arginine (Bohannon & Sonenshein, 1989; Commichau *et al*., 2007b). The activity of these two transcriptional regulators can be controlled by trigger enzymes. GlnA interacts with TnrA, and the catabolic glutamate dehydrogenase RocG interacts with GltC (Wray *et al*., 2001; Belitsky & Sonenshein, 2004; Commichau *et al*., 2007a; Commichau *et al*., 2007b) (see Fig. 2.1B). In addition, GltC is modulated by metabolites of the glutamate dehydrogenase reaction: 2-oxoglutarate and glutamate stimulate and inhibit the activity of GltC, respectively (Picossi, *et al*., 2007).

Thus, the information available to the enzymes of glutamine synthesis and glutamate degradation is shared with transcription factors in the modulation of their activities. The use of trigger enzymes in the transcriptional regulation of metabolic genes allows direct feedback between metabolite availability and gene expression, thus

maintaining a pool of key metabolic intermediates in the cell at a constant concentration.

(A)

(B)
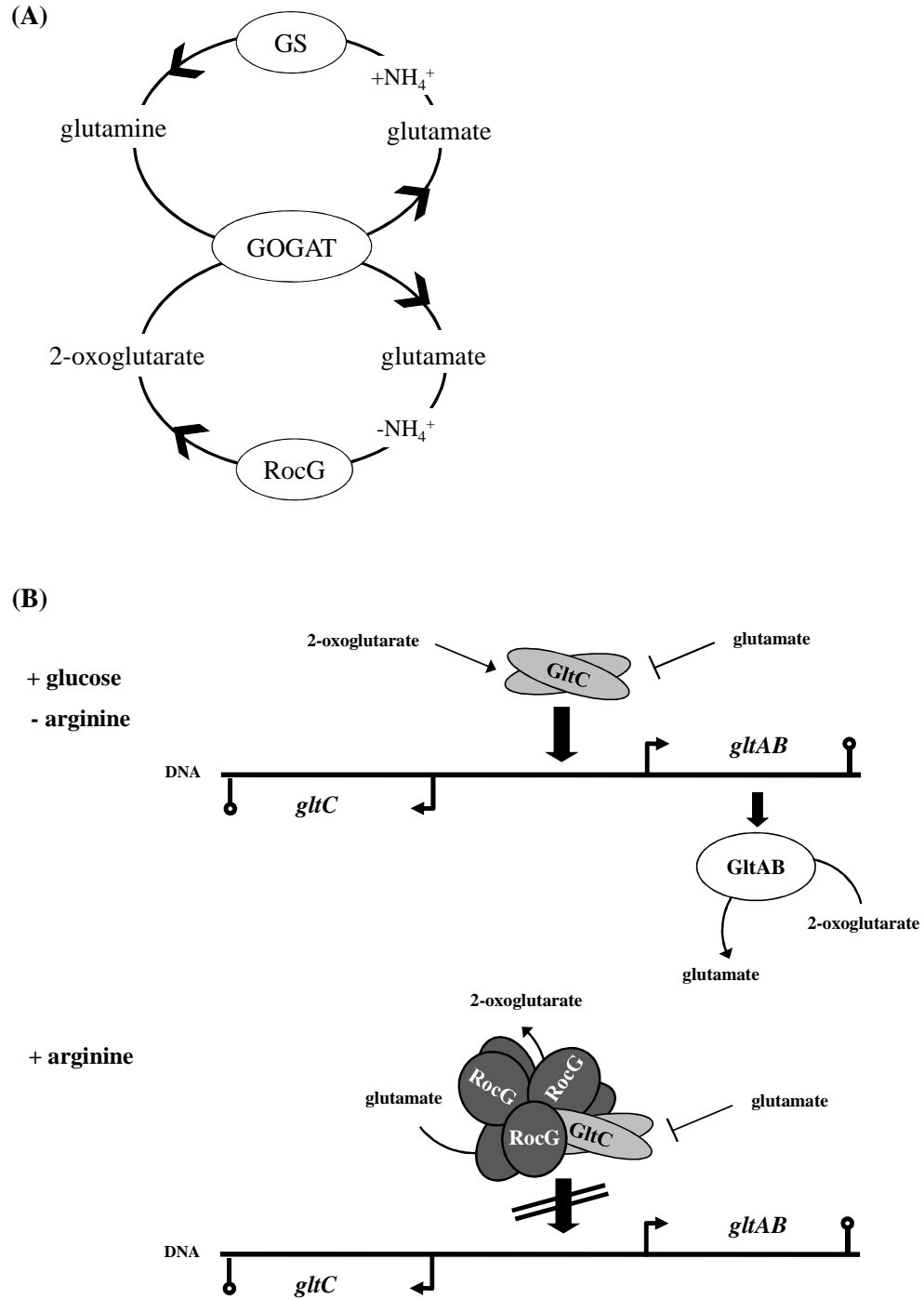
Fig. 2.1 - Glutamate metabolism in *B. subtilis* and its regulation.

Fig. legend see next page.

Fig. 2.1 - Glutamate metabolism in *B. subtilis* and its regulation.

(A) Glutamate is synthesized in the GOGAT cycle by the combined reactions of glutamine synthetase (GS) and glutamate synthase (GOGAT). The degradation of glutamate to 2-oxoglutarate and ammonium is catalyzed by the strictly catabolic glutamate dehydrogenase RocG. (B) In the absence of arginine, the transcriptional activator GltC (light-gray ovals) stimulates the expression of the *gltAB* operon. In the presence of arginine, RocG (dark-gray ovals) directly interacts with GltC, leading to inactivation of the transcription factor. Moreover, the activity of GltC is affected by 2-oxoglutarate and glutamate.

Glutamate is one of the major metabolites in any living cell and is either taken up directly from the environment, generated from peptides or amino acids such as arginine and proline, or synthesized from 2-oxoglutarate. In *Escherichia coli*, glutamate accounts for as much as 40% of the cellular metabolite pool (Yuan *et al.*, 2009). In the absence of organic nitrogen sources, glutamate is synthesized by glutamate synthase, as summarized above. In many organisms, including *E. coli*, glutamate dehydrogenase is capable of directly generating glutamate by the amination of 2-oxoglutarate. However, glutamate dehydrogenase has a low affinity for ammonium; therefore, this reaction can take place only at high ammonium concentrations (Yuan *et al.*, 2009). Glutamate synthase is the only enzyme in *B. subtilis* that can produce glutamate; by contrast, *rocG*-encoded glutamate dehydrogenase has an exclusively catabolic activity in this organism (see Fig. 2.1A) (Belitsky & Sonenshein*,* 1998).

Glutamate would thus appear to be a particularly important cellular metabolite in *B. subtilis* and other organisms. For instance, glutamate metabolism is very tightly controlled by at least five different transcription factors: TnrA and GltC control the biosynthetic operon *gltAB*; the expression of *rocG* is subject to carbon catabolite repression exerted by CcpA; and synthesis of *rocG* in the presence of arginine is mediated by the transcription activators RocR and AhrC (Klingel *et al.*, 1995; Belitsky & Sonenshein, 1999; Belitsky *et al.*, 2004). Moreover, both *B. subtilis* and *E. coli* respond to any perturbation in glutamate homeostasis with spontaneous suppressor mutations that bring the glutamate pools back into balance (Yan, 2007; Commichau *et al.*, 2008).

Glutamate metabolism in *B. subtilis* is, unusually, characterized by the presence of a second glutamate dehydrogenase, *gudB* (Belitsky & Sonenshein, 1998). The enzyme encoded by this gene is active in wild-type isolates of *B. subtilis*, but is cryptic in laboratory strain 168 due to an insertion of three amino acids close to the active site of

the enzyme (Zeigler *et al*., 2008). Indeed, *rocG* mutant strains readily recover glutamate homeostasis by decryptification of *gudB* (Belitsky & Sonenshein, 1998; Commichau *et al*., 2008).

Hence, RocG appears to have a dual role in glutamate metabolism. It is involved in glutamate catabolism but, if glutamate is available, interacts with GltC, the transcription activator of the *gltAB* operon, and this interaction is inhibitory to GltC. However, in the absence of arginine or in the presence of glucose, either of which leads to a reduction in RocG expression, free GltC activates transcription of the *gltAB* operon, resulting in the biosynthesis of glutamate (Commichau *et al*., 2007b). Some variants of GltC are constitutively active as a transcription activator, suggesting that these GltC proteins cannot be inactivated by RocG (Belitsky & Sonenshein, 1995). Indeed, a reduction in the interaction between RocG and one of these GltC variants was previously observed *in vivo* (Commichau *et al*., 2007b).

In this study, we have isolated and characterized RocG variants whose ability to control GltC activity is affected. Several amino acid substitutions in RocG cause permanent inactivation of GltC in media containing glutamate or its precursor, arginine. These enzymes are, on the other hand, strongly impaired in catalysis, and the affinity for glutamate is severely reduced. One of the RocG variants that we have isolated has retained its enzymatic activity but has lost the ability to inactivate GltC. We have also determined the crystal structures of both the major glutamate dehydrogenase and the secondary glutamate dehydrogenase from *B. subtilis* (RocG and GudB1, respectively). These structures have allowed us to analyze the effect of RocG variations at the molecular level and to gain insights into the nature of the interaction between RocG and GltC.

Materials and Methods

Bacterial strains and growth conditions

The *E. coli* strains DH5α, (Sambrook *et al*., 1989) XL1-Red (Stratagene, Amsterdam, Netherlands), and BL21/DE3 (pLysS) (Novagen, Darmstadt, Germany) were used for cloning experiments, random mutagenesis, and protein expression, respectively. The *B. subtilis* strain GP28 (*trpC2 rocG*::*Tn10 spc* Δ*gudB*::*cat amyE*::*gltA'–'lacZ aphA3*) (Commichau *et al*., 2007a) was used for the screening and characterization of *rocG* mutants. *B. subtilis* GP27 (*trpC2* Δ*gudB*::*cat amyE*::*gltA'–' lacZ aphA3*) (Commichau *et al*., 2007a) was used to estimate the level of overexpression of RocG in the screening system. *B. subtilis* was grown in C minimal medium containing ammonium as the nitrogen source, and 5 g/l glucose (C-Glc medium), 5 g/l arginine (CR medium), or 8 g/l glutamate (CE medium) as the carbon source (Wacker *et al*., 2003). CR-Glc medium is C minimal medium containing 5 g/l glucose and 5 g/l arginine. CS-Glc contains 6 g/l sodium succinate and 5 g/l glucose. CSE-Glc medium is CS-Glc with 8 g/l glutamate. The media were supplemented with tryptophan (50 mg/l). *E. coli* was grown in LB medium, and transformants were selected on plates containing ampicillin (100 μg/ml). LB and SP plates were prepared by addition of 17 g/l Bacto agar (Difco, Lawrence, USA) to the medium.

DNA manipulation

Transformation of *E. coli* was performed using standard procedures (Sambrook *et al*., 1989). The NucleoSpin Plasmid kit (Macherey and Nagel, Düren, Germany) was used to isolate plasmid DNA from *E. coli* and *B. subtilis*. Prior to plasmid purification, *B. subtilis* cells were incubated with 3 mg/ml lysozyme (Serva, Heidelberg, Germany) dissolved in buffer 1 of the NucleoSpin Plasmid kit to increase the yield of plasmid DNA recovery. Chromosomal DNA of *B. subtilis* was isolated as described previously (Kunst & Rapoport, 1995). Restriction enzymes, T4 DNA ligase, and, for PCR, Phusion™ DNA polymerase (New England Biolabs, Frankfurt am Main, Germany) were all used as recommended by the manufacturers. DNA fragments were purified using the QIAquick PCR Purification kit (Qiagen, Hilden, Germany), and all DNA sequences were determined using the dideoxy chain termination method (Sambrook *et al*., 1989).

Random mutagenesis of *rocG*

To study the effect of point mutations in RocG that affect GltC activity, we generated plasmid pGP529, which allows the expression of *rocG* in *B. subtilis* under the control of the strong *degQ*36 promoter (Martin-Verstraete *et al*., 1994) as follows: *rocG* was amplified using the primers IW18 (5′AAAAGGATCCGCGCTTACATTACAG) and IW19 (5′AAACTGCAGTCATTAGACCCATCCGCGGAAACG), and the resulting PCR product was digested with BamHI and PstI before cloning into similarly restricted pBQ200 to yield pGP529. Plasmid pGP529 was subjected to random mutagenesis using the *E. coli* mutator strain XL1-Red. For this purpose, pGP529 was used to transform *E. coli* XL1-Red, and the cells were plated on 40 LB plates resulting in approximately 80 colonies per plate. The colonies from each plate were resuspended in 1 ml of LB medium, and 100 μl of each suspension was used to inoculate 100-ml flasks containing 10 ml of LB medium. The cultures were grown for 48 h at 37°C to allow the emergence of mutations. Plasmid DNA from each culture was isolated individually and used to transform the indicator strain *B. subtilis* GP28, which contains a translational *gltA–lacZ* fusion to monitor the effect of mutant RocG proteins on the activity of GltC. Transformants were selected on SP plates containing 2 μg/ml erythromycin plus 25 μg/ml lincomycin.

Plasmids for the expression of *rocG* and *gudB* alleles

Plasmid pGP902 was used for the overexpression of Strep-tagged wild-type RocG (Commichau *et al*., 2007a). All plasmids used in this study are listed in Tab. S2. For the expression of GudB1, the active variant of *B. subtilis* GudB, we constructed plasmid pGP864 by cloning a PCR fragment obtained with primers KG20 (5′ AAAGAGCTCGATGGCAGCCGATCGAAACACCG) and KG21 (5′ TTTGGA TCCTCATTATATCCAGCCTCTAAAACGCGAAGCTT), which have used the chromosomal DNA of the *gudB1* mutant strain GP801 (Commichau *et al*., 2008) as template. For overexpression of Strep-tagged RocG mutant proteins, the plasmids pGP856, pGP857, pGP858, pGP865, and pGP866 were constructed. The corresponding *rocG* mutant alleles were amplified with the primer pair PT5/PT6, (Commichau *et al*., 2007a) using plasmids pGP852, pGP853, pGP854, pGP932, and pGP933 as template DNAs (see Tab. 2.1). All PCR products were digested with SacI and BamHI, and

cloned into the vector pGP172 (Merzbacher *et al*., 2004) cut with the same enzymes. The identity of the cloned inserts was verified by DNA sequencing.

Protein purification

E. coli BL21/DE3 (pLysS), transformed by plasmids encoding the appropriate glutamate dehydrogenases, was used as host for the overexpression of recombinant proteins. Cultures were grown in 1 l of LB medium at 37°C; expression was induced by the addition of IPTG (final concentration, 1 mM) to logarithmically growing cultures (OD_{600} = 0.8), and the cells were harvested 2 h after induction. The pellets were resuspended in 30 ml of disruption buffer (100 mM Tris–HCl, 150 mM NaCl, and 10 mM ethylenediaminetetraacetic acid, pH 8.0). The cells were disrupted by using a French Press (20,000 psi, 138,000 kPa; Spectronic Instruments, Garforth, UK). The extracts were passed over a 1-ml Strep-Tactin® column (IBA, Göttingen, Germany), and the bound material was washed with 6 ml of disruption buffer before being eluted with 3.5 ml of disruption buffer containing 2.5 mM desthiobiotin (IBA, Göttingen, Germany). The Bio-Rad dye-binding assay was used to determine protein concentrations, with bovine serum albumin as standard.

Western blot analysis

For Western blot analyses of RocG, *B. subtilis* crude extracts were separated by SDS-PAGE and transferred to a polyvinylidene difluoride membrane (Bio-Rad, Munich, Germany) by electroblotting. RocG was detected with rabbit polyclonal antiserum raised against *B. subtilis* RocG (Commichau et al., 2007a). The anti-RocG antibodies were visualized with anti-rabbit IgG-AP secondary antibodies (Chemikon International, Temecula, USA) with the CDP* detection system (Roche Diagnostics, Basel, Switzerland).

Enzyme assays

The amination reaction catalyzed by RocG was assayed at room temperature by measuring the decrease in $NADH_2$ absorption at 340 nm (Ultrospec 2100 pro; GE Healthcare, Munich, Germany) under the following conditions: 10 mM 2-oxoglutarate, 0.1 mM $NADH_2$, and 0.1 M ammonium chloride in 100 mM Tris–HCl buffer (pH 7.3).

The deamination reaction was assayed by measuring the increase in $NADH_2$ absorption at 340 nm in a mixture of 0.1 mM NAD^+ and 20 mM monosodium L-glutamate in 50 mM potassium phosphate buffer (pH 7.7). For the determination of K_m values, a wide concentration range of the relevant substrates (glutamate and ammonium for catabolic and anabolic reactions, respectively) was used.

Crystallization and structure determination

RocG(E93K) and GudB1 proteins were purified as described above, with the exception of an additional purification step of gel filtration using a Superdex 200 column (GE Healthcare, Buckinghamshire, UK) and with the elution profile of both proteins being consistent with a hexameric assembly. The proteins were buffer exchanged into 10 mM Tris–HCl (pH 8.0) and 100 mM NaCl, and concentrated for crystallization at 10 mg/ml using a Millipore 30-kDa cutoff centrifugal concentrator. RocG(E93K) crystals were grown by hanging-drop vapor diffusion using a mother liquor of 20% (wt/vol) PEG 3350 and 20% (vol/vol) glycerol. Single crystals were loop mounted and flash cooled in liquid nitrogen. GudB1 crystals were grown from solutions containing 20% (wt/vol) PEG 3350 and 0.15 M sodium malonate. Crystals were transferred for approximately 1 min to a cryoprotectant of crystallization buffer supplemented with 25% (vol/vol) glycerol before being loop mounted and flash cooled in liquid nitrogen. Data collection for both RocG(E93K) and GudB1 was performed on beamline I02 at the Diamond light source, with diffraction extending to maximum resolutions of 2.3 and 2.4 Å, respectively. Data sets were processed using XDS (Leslie, 1992) then scaled and merged using SCALA (Evans, 1993) in the CCP4 suite of programs (CCP4, 1994). Both structures were solved by molecular replacement using the program MOLREP (Vagin & Teplyakov, 1997) and the *Pyrococcus furiosus* glutamate dehydrogenase (1GTM) as search model (Yip *et al*., 1995). Manual model building in Coot (Emsley & Cowtan, 2004) was interspersed with rounds of refinement until convergence. The RocG(E93K) structure was refined using the program REFMAC (Murshudov *et al*., 1997) to *R*-factor/R_{free} values of 0.19/0.24, whereas GudB1 was refined with PHENIX REFINE (Adams *et al*., 2002) to *R*-factor/R_{free} values of 0.25/0.28.

Results

Isolation of *rocG* mutant alleles

Plasmid pGP529 carrying the *rocG* gene under the control of a strong constitutive promoter was subjected to random mutagenesis using the *E. coli* mutator strain XL1-Red, and the plasmid pools obtained were used to transform the indicator strain *B. subtilis* GP28 as described in Materials and Methods (Commichau *et al.*, 2007a). To identify *rocG* mutant alleles, we transferred the transformants onto CR and CR-Glc minimal media containing 40 µg/ml 5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside (X-Gal). Cells expressing wild-type RocG form pale blue colonies on CR-Glc-X-Gal plates. In our system, RocG is constitutively expressed under the control of a glucose-insensitive promoter. The RocG amounts present in the cell in this system were about 20-fold increased as compared to a wild-type strain expressing *rocG* from its own promoter (*B. subtilis* GP27; data not shown). The arginine present in the medium can be converted into glutamate, the substrate of the glutamate dehydrogenase RocG. The catalytically active RocG inhibits GltC, thus reducing the expression of *gltA–lacZ* fusion. A reasonable explanation for the residual activity of GltC is that glucose is converted into 2-oxoglutarate, which is a positive effector of GltC (Picossi *et al.*, 2007). As a result, GltC cannot be fully inhibited by RocG if the bacteria grow on CR-Glc-X-Gal plates due to sources of α-ketoglutarate. By contrast, wild-type RocG completely inhibits GltC activity on CR-X-Gal plates, resulting in the formation of white colonies. In the absence of glucose, no α-ketoglutarate is formed by the tricarboxylic acid cycle, and GltC cannot be stimulated by 2-oxoglutarate and cannot activate the *gltA–lacZ* fusion.

CR-Glc-X-Gal medium was used to screen for RocG variants that permanently inhibited GltC, resulting in white colonies. About 60 white colonies were obtained, and permanent inhibition of GltC due to the *rocG* mutant allele was verified by retransformation for 10 candidates. These plasmidborne *rocG* alleles encode so-called "superrepressor" variants of RocG.

Since catalytically competent glutamate dehydrogenase is required for *B. subtilis* to utilize arginine as carbon and nitrogen source, we used CR-X-Gal plates to isolate monofunctional RocG variants that are catalytically active but are unable to inhibit GltC. Strain GP28 expressing wild-type RocG forms white colonies on CR-X-Gal

plates due to the complete inactivation of GltC by RocG. By contrast, cells expressing monofunctional RocG proteins were expected to form dark-blue colonies on CR-X-Gal plates. This expectation was further supported by an *in silico* analysis of the regulatory network of glutamate metabolism, which highlighted the possibility of using CR-X-Gal medium to screen for monofunctional RocG variants. Indeed, two darkblue colonies on CR-X-Gal plates were obtained after the transformation of *B. subtilis* GP28 with the pool of mutant plasmids. The plasmids of these two candidates were isolated and used to retransform *B. subtilis* GP28. These transformants were able to grow with arginine as the sole carbon and nitrogen source, and formed dark-blue colonies on CR-X-Gal plates. Thus, these *rocG* alleles encode active glutamate dehydrogenases that appear to have lost the ability to inhibit the transcription activator GltC.

Mutations of the superrepressor and monofunctional *rocG* alleles were identified by sequencing. In all cases, changes in RocG properties were caused by single base-pair exchanges resulting in single amino acid substitutions. Superrepressor mutations are scattered throughout the entire *rocG* sequence, whereas the two monofunctional *rocG* alleles carried the same mutation. The position of the mutations in the amino acid sequence of RocG is shown schematically in Fig. 2.2, and the plasmids containing the *rocG* mutant alleles and the corresponding mutations are listed in Tab. 2.1.
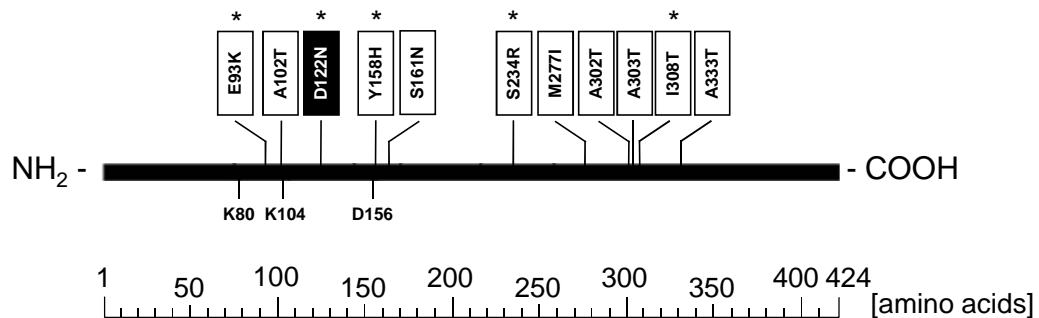
Fig. 2.2 - Mutations found in RocG that affect the regulation of GltC activity by RocG.

Amino acid exchanges leading to a superrepressor phenotype are shown in empty boxes. The mutation of the monofunctional RocG variant is depicted by a filled box. RocG mutant variants that were further analyzed are indicated by asterisks. Amino acids involved in glutamate binding at the active site of RocG are indicated below the primary sequence.

Tab. 2.1 - Mutations affecting the inhibition of GltC by RocG

Plasmid	Allele	Mutation	Amino acid substitution	Corresponding expression vector
pGP852	*rocG*-SR1	T923C	I308T	pGP856
pGP853	*rocG*-SR2	C702A	S234R	pGP857
pGP854	*rocG*-SR3	G277A	E93K	pGP858
pGP855	*rocG*-SR4	G997A	A333T	
pGP933	*rocG*-SR5	T472C	Y158H	pGP866
pGP867	*rocG*-SR6	G907A	A303T	
pGP868	*rocG*-SR7	G904A	A302T	
pGP869	*rocG*-SR8	A304G	A102T	
pGP870	*rocG*-SR9	A479G	S161N	
pGP871	*rocG*-SR10	G831A	M277I	
pGP932	*rocG*-MF	G364A	D122N	pGP865

For further analyses, we selected four superrepressor mutants and one monofunctional RocG variant. To exclude the possibility that the observed effects result from altered cellular levels of the mutant RocG proteins in comparison to the wild-type protein, we determined the RocG concentration by Western blot analysis. Strains carrying plasmid pGP529 (wild-type RocG) and pBQ200 (empty vector) were used as controls. The strain carrying pBQ200 did not synthesize any protein that was recognized by the anti-RocG antibodies, indicating that any immunodetection is specific for RocG (Fig. 3.2). The cellular concentrations of all RocG variants were similar to that of the wild-type protein (see Fig. 2.3). Thus, the *rocG* phenotypes are directly attributable to changes in the biochemical properties of the mutant enzymes.
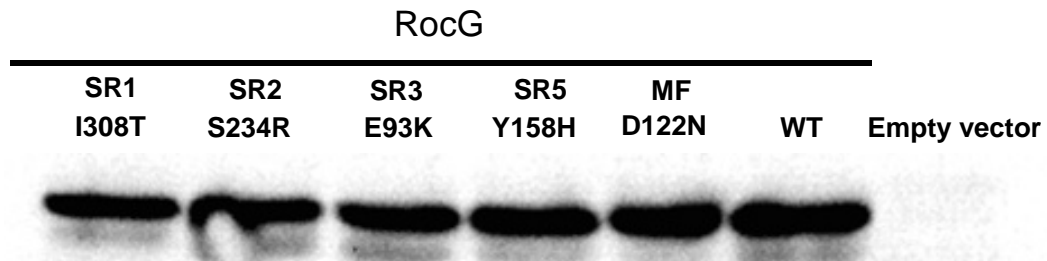
Fig. 2.3 - Expression of RocG mutant proteins.

Crude extracts were isolated from *B. subtilis* strain GP28 (*rocGΔgudB*) expressing the RocG superrepressor (SR) proteins SR1, SR2, SR3, and SR5, and the monofunctional (MF) RocG protein grown in CSE minimal medium. Strain GP28 either expressing the RocG wild-type (WT) protein or carrying the empty vector pBQ200 served as control. The proteins were detected using rabbit polyclonal antibodies raised against *B. subtilis* RocG. Samples (15 μg) of crude extract were applied per lane.

Impact of RocG mutant proteins on GltC activity

The impact of RocG mutations on the activity of GltC was quantified after the growth of *B. subtilis* GP28 carrying relevant plasmids in minimal media. As shown in Tab. 2.2, *B. subtilis* GP28 carrying the empty vector pBQ200 exhibited constitutive *gltA* expression because of the absence of a functional glutamate dehydrogenase to inhibit GltC. The strain carrying pGP529, encoding wild-type *rocG*, showed little promoter activity on media containing arginine. By contrast, *gltA–lacZ* activity was high in the absence of arginine (presumably reflecting the activity of GltC). This suggests that the inhibition of GltC by RocG did not occur under these conditions. This is in good agreement with previous reports indicating that RocG does not inhibit GltC activity in the absence of arginine even if *rocG* is overexpressed. Moreover, the effect of arginine cannot be due to any effect of RocG expression, but some other role of arginine, such as acting as a source of glutamate, must be involved (Belistky & Sonenshein, 2004; Commichau *et al*., 2007b).

The plasmids encoding the superrepressor *rocG* alleles all caused very low GltC activity in CR medium containing glucose. These RocG variants did not support growth with arginine as the single carbon source, suggesting that their enzymatic activity was affected by the mutations (see the text below). The GltC activity of the strains carrying the superrepressor RocG variants was very low in CSE medium containing glucose. By contrast, these RocG variants had only a minor inhibitory impact on GltC activity in a

medium that did not contain glutamate, or a source of it (i.e., CS medium containing glucose). This observation suggests that the presence of glutamate facilitates the inhibition of GltC by RocG. Moreover, the mutant RocG proteins are still capable of responding to the presence of glutamate. The monofunctional RocG protein did not inhibit GltC activity under any condition tested in this study, and the *B. subtilis* strain expressing this mutant protein was able to utilize arginine as the single carbon source, suggesting that RocG-D122N was enzymatically active in glutamate degradation. Thus, the D122N mutation appears to have uncoupled the two functions of RocG and converted it from a trigger enzyme into a conventional enzyme with just a metabolic role.

Tab. 2.2 - Effect of the different RocG variants on GltC activity

Plasmid	RocG variant[a]	*gltA-lacZ* expression[b]			
		CR	CR-Glc	CS-Glc	CSE-Glc
pBQ200	none	NG[c]	211 ± 11	509 ± 44	321 ± 80
pGP529	wild type	7 ± 4	26 ± 8	562 ± 85	423 ± 100
pGP852	RocG-SR1, I308T	NG	9 ± 2	207 ± 24	15 ± 5
pGP853	RocG-SR2, S234R	NG	5 ± 2	164 ± 31	9 ± 6
pGP854	RocG-SR3, E93K	NG	9 ± 2	140 ± 42	18 ± 11
pGP933	RocG-SR5, Y158H	NG	8 ± 4	187 ± 25	17 ± 2
pGP932	RocG-MF, D122N	277 ± 23	505 ± 58	650 ± 62	840 ± 134

[a] SR and MF denote the superrepressor and monofunctional RocG variants, respectively.

[b] *gltA* expression is expressed in units per milligram of protein. All measurements were performed at least three times. Average values with standard deviations are shown.

[c] NG, no growth.

Enzymatic activities of mutant glutamate dehydrogenases

When we determined the activity of GltC in the strains expressing the mutant alleles of *rocG*, we observed that strains expressing the superrepressor variants were unable to grow with arginine as the single carbon source. This might result from a reduced or absent enzymatic activity of these RocG variants; consequently, we studied the enzymatic activity of RocG and its mutant variants in more detail.

First, we analyzed the growth of *B. subtilis* GP28 carrying the different *rocG* alleles on CR minimal medium in the presence or in the absence of glucose. Strains bearing the empty vector pBQ200 or pGP529 (wild-type RocG) were used as controls. Both strains grew well on a medium containing both glucose and arginine. Similarly, all strains expressing the mutant RocG proteins grew well on this medium, suggesting that these strains also have full biosynthetic capacity required for growth on minimal medium (see Tab. 2.3). The strain expressing wild-type RocG was also able to grow with arginine as the only source of carbon and energy. By contrast, the strain carrying the empty vector did not grow on this medium because it lacks a functional glutamate dehydrogenase required for the last step of arginine catabolism, the conversion of glutamate into 2-oxoglutarate (Commichau *et al.*, 2008). The strains expressing the superrepressor RocG variants were unable to utilize arginine as the only carbon source, suggesting that these RocG variants are severely impaired in their enzymatic activity. The monofunctional RocG protein that was originally isolated on CR plates allowed arginine utilization, suggesting that this protein is enzymatically active.

Next, we purified recombinant RocG variants, and their enzymatic activities were determined. Wild-type RocG exhibited catabolic and anabolic activities of 3.9 and 39.7 U/mg protein, respectively. The catabolic activities of all superrepressor proteins were severely reduced (see Tab. 2.3), in agreement with the inability of the corresponding *B. subtilis* strains to grow with arginine as the only carbon source. In addition to their reduced activities, these enzymes have a drastically increased K_m for glutamate (between 23 and 75 mM, compared to 2.9 mM for the wild-type protein), indicating a reduced affinity for the substrate. Similarly, with the exception of the RocG(E93K) variant, the anabolic activities and the K_m values for ammonium of the superrepressors were reduced and increased, respectively, albeit to a lesser extent (see Tab. 2.3). The monofunctional RocG variant had a 4-fold higher K_m for glutamate than for the wild-type enzyme; however, this variant still enabled the parent strain GP28 to grow with

arginine as the single carbon source. Similarly, the anabolic properties of this RocG mutant are not significantly altered in comparison to the wild-type enzyme, and confirm that the D122N substitution in RocG predominantly affects the inhibitory interaction of RocG with GltC, but not its enzymatic properties.

Tab. 2.3 - Enzymatic activities of the RocG variants.

Plasmid[a]	RocG variant	Growth on plates CR-Glc	CR	catabolic activity (U/mg of protein)[b]	K_M (glutamate) (mM)[b]	anabolic activity (U/mg of protein)[b]	K_M (ammonium) (mM)[b]
pBQ200	none	+++	-	NA[c]	NA	NA	NA
pGP529	wild type	+++	+	3.9 ± 0.67	2.9 ± 1.25	39.7 ± 2.52	18 ± 5
pGP852	RocG-SR1, I308T	+++	-	ND[d]	ND	ND	ND
pGP853	RocG-SR2, S234R	+++	-	0.12 ± 0.03	74.7 ± 7	4.3 ± 1.28	58.7 ± 1.53
pGP854	RocG-SR3, E93K	+++	-	0.4 ± 0.06	46.3 ± 4.16	36 ± 7.54	40 ± 3
pGP933	RocG-SR5, Y158H	+++	-	0.69 ± 0.17	23 ± 3.61	14.3 ± 3.51	38.7 ± 3.60
pGP932	RocG-MF, D122N	+++	+	2.27 ± 0.32	12 ± 2.65	37 ± 3.61	10.3 ± 2.52

[a] These plasmids were used for the growth experiments.

[b] The experiments were performed in triplicate. Average values with standard deviations are shown.

[c] NA, not applicable.

[d] ND, not determined due to low stability of this protein.

Note that the assays for catabolic and anabolic activities were performed with an excess of ammonium and glutamate, respectively. *In vivo*, RocG is unable to perform the anabolic reaction most likely due to its low affinity for ammonium.

Crystal structures of RocG(E93K) and GudB1

In an attempt to gain a better molecular understanding of structure–function relationships among the different mutant RocG alleles, we have determined the crystal structures of the superrepressor RocG(E93K) mutant and the decryptified secondary glutamate dehydrogenase GudB1. This protein lacks the three-amino-acid insertion in the active site that renders the GudB protein completely inactive (Belitsky & Sonenshein, 1998; Commichau *et al.*, 2007b). During the crystallographic aspect of this study, a number of RocG variants were purified and subjected to crystallization trials. The wild-type enzyme and most of the mutants studied produced comparatively small crystals that diffracted to 4 Å at best, whereas the RocG(E93K) crystals diffracted to a resolution of 2.3 Å and belonged to space group $P2_12_12_1$. GudB1 crystals diffracted to 2.4 Å and belonged to space group $P2_1$. Both structures were solved by molecular replacement with six subunits in the asymmetric unit and were refined until convergence. The final RocG(E93K) model contains 2409 out of a possible 2544 residues with 5 polyethylene glycol (PEG) ions and 829 water molecules, whereas the GudB1 model contains 2354 out of a possible 2538 residues and 414 water molecules. Both structures have been restrained to standard bond lengths and angles, with over 99% of residues lying in the most favored and additionally allowed regions of the Ramachandran plot, (Ramachandran *et al.*, 1963) as calculated by the program PROCHECK (Laskowski *et al.*, 1993). A full summary of the data collection, processing, and refinement statistics can be found in Tab. 2.4.

RocG(E93K)

Overall, the electron density for RocG(E93K) is of high quality throughout all six molecules, with the exception of the N-terminal 15 residues that were not visible in the electron density, have been omitted from the model in chains B, C, D, E, and F, and are only partially included in chain A. Similarly, the entire length of chains A and B could be traced in the electron density with high confidence, whereas in the regions corresponding to residues 215–342 in chains C–F, the electron density is of poorer overall quality, and residues 274–276 in chain D, residues 270–286 in chain E, and residues 270–285 in chain F have been omitted from the final model.

Tab. 2.4 - **Data collection and refinement statistics.**

	RocG(E93K)	GudB1
Data collection		
Space group	P $2_12_12_1$	P 2_1
Unit cell parameters		
$a, b, c,$ (Å)	137.6, 143.1, 162.6	86.1, 158.0, 193.8
α, β, γ (°)	90, 90, 90	90, 118.7, 90
Wavelength (Å)	0.98	0.98
Resolution (Å)	20 – 2.3 (2.42 – 2.30)	50.0 – 2.4 (2.46 – 2.40)
R_{merge} (%)[a]	6.3 (44.5)	7.0 (38.0)
$I/\sigma I$	16.0 (3.1)	11.6 (3.2)
Completeness (%)	99.8 (100.0)	97.8 (97.8)
Redundancy	4.2 (4.3)	3.6 (3.7)
Refinement		
Resolution (Å)	20.0 – 2.3	50.0 – 2.4
No. reflections	142264	98133
R_{work}/R_{free} (%)	19.1/24.3	24.7/27.8
No. Atoms		
Protein	19609	15594
Water	824	414
Ligand/Ion	35	0
B factors		
Protein	56.0	45.5
Water	48.1	40.2
Ligand/Ion	56.0	-
R.m.s. deviations		

	RocG(E93K)	GudB1
Bond lengths (Å)	0.014	0.003
Bond angles(°)	1.44	0.82
PDBid	3K92	3K8Z

Values in parentheses are for the highest-resolution shell.

[a] $R_{\mathrm{merge}}=\Sigma_{hkl}\Sigma_i|I_i-I_{\mathrm{m}}|/\Sigma_{hkl}\Sigma_iI_i$, where I_i and I_{m} are the observed intensity and mean intensity of related reflections, respectively.

The structure of RocG(E93K) is very similar to those of a number of glutamate dehydrogenases studied to date, being hexameric with a subunit comprising two domains separated by a cleft, which forms the active site (Yip *et al*., 1995; Baker *et al*., 1997; Knapp *et al*., 1997; Britton *et al*., 1999; Nakasako *et al*., 2001). Domain I comprises residues 1–193 at the N-terminus and the final 17 residues at the C-terminus, and forms a typical amino acid dehydrogenase fold with a central six-stranded mixed β-sheet (strand order a↑, b↓, d↑, c↑, e↑, f↑) flanked by two α-helices on one side and by one α- helix on the other side. Domain II is composed of a central seven-stranded parallel β-sheet (strand order m, l, k, g, h, i, j) flanked by four α-helices on one side and by three α-helices on the other side, and forms a variant of the common NAD(P)-binding Rossmann fold (Rossmann *et al*., 1974). The two domains are connected by a number of long α-helices that converge upon a single hinge region, allowing for relatively large domain motions to occur (up to 14 Å). Indeed, these domain motions are critical to the catalytic mechanism of this enzyme (Stillman *et al*., 1999) and are believed to allow the nicotinamide ring of the nucleotide cofactor to come into contact with the C^{α} of the glutamate substrate in order for hydride transfer to occur.

The six subunits in the asymmetric unit form a single compact homohexamer with 32-point group symmetry, and the vast majority of intersubunit interactions around the 2-fold and 3-fold axes are provided by domain I. The six chains in the asymmetric unit represent, to varying degrees, the open and closed forms of glutamate dehydrogenase. Chain A represents a more closed conformation, and chains B, C, D, E, and F are more open, with maximal displacements of 14 Å in comparison to chain A . These differences in domain architecture can be explained in part by crystal contacts within the lattice,

with domain II of chain A making extensive contacts with a symmetry-related copy of chain C. Furthermore, the absence of crystal contacts in chains C, D, E, and F may account for the disordered residues at the periphery of domain II. While the closed form of the enzyme is usually associated with glutamate binding (Stillman *et al*., 1999), this kind of intersubunit variation has already been observed in the apo form of the homologous enzyme from *Thermococcus profundus* (Nakasako *et al*., 2001

Neither glutamate nor any other substrate was added to the RocG crystals, and examination of the $2F_{obs}-F_{calc}$ and $F_{obs}-F_{calc}$ electron density maps in the active site did not reveal any significant unexplained density.

The superrepressor E93K variant of RocG described above is, we believe, virtually identical in structure to the wild-type enzyme. The mutation, which changes a glutamate into a lysine, is located within the first turn of the third helix of domain I, within an acidic stretch of amino acids. The local structure reveals no significant differences when compared to other bacterial glutamate dehydrogenases or to GudB1. Indeed, the high overall level of structural similarity between RocG(E93K) and GudB1 (r.m.s.d., 0.95 Å) indicates that the RocGE93K structure is an appropriate template with which to investigate the other mutants described above. Wild-type RocG could be crystallized under the same conditions as RocG(E93K), and although the diffraction obtained from these morphologically identical crystals was limited to 4 Å, their growth indicates that the structural differences arising from the mutation are minor.
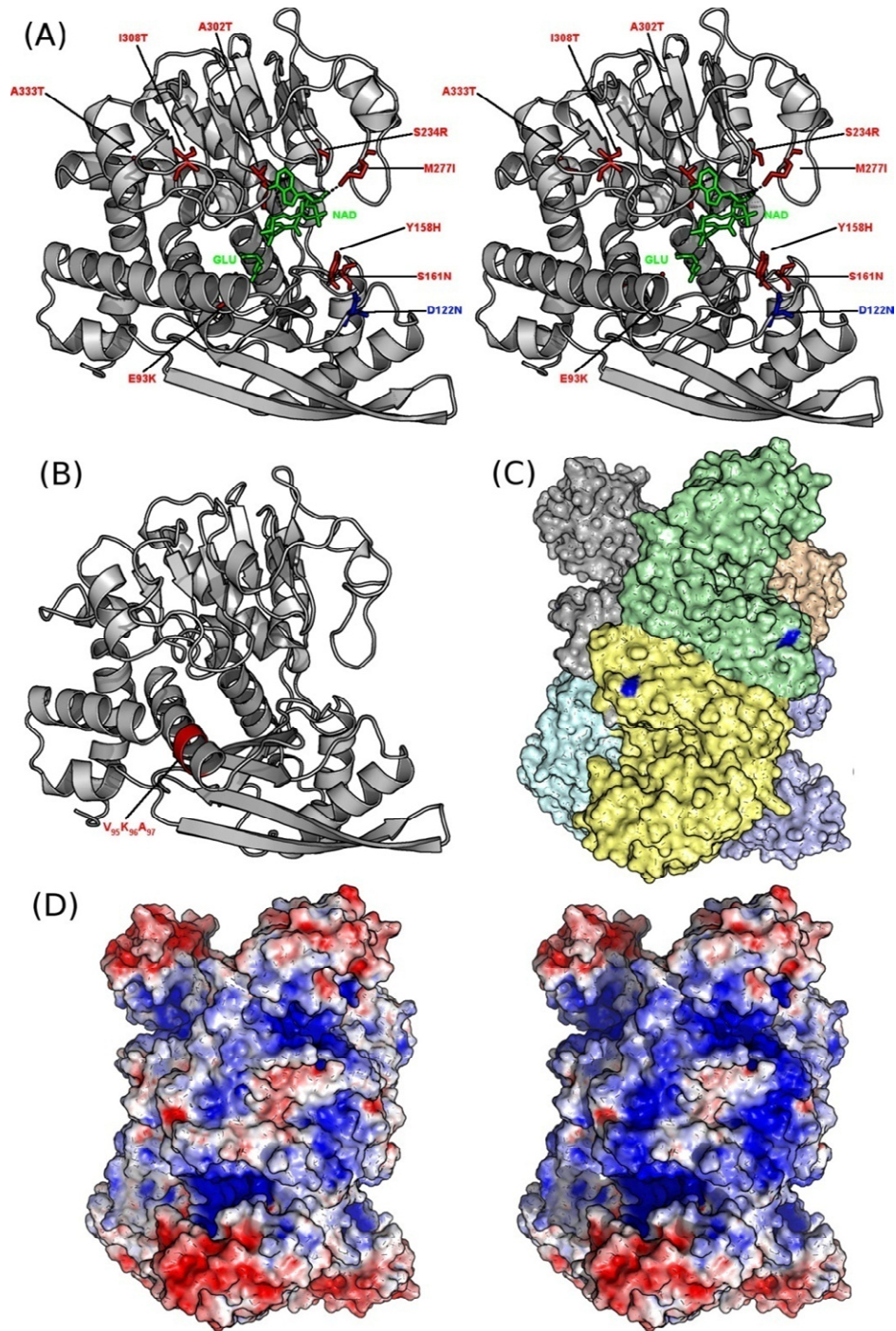
Fig. 2.4 - Structural organization of RocG/GudB1 variants.

(A) Stereo view of the RocG monomer shown in cartoon format, with superrepressor mutations highlighted in red and with the single monofunctional variant highlighted in blue. For reference, the nucleotide cofactor and the glutamate substrate are shown in green in their expected positions in the active site based on structural superpositions with the *C. symbiosum* glutamate dehydrogenase–NAD

complex (Stillman *et al*., 1993) and the *P. islandicum* glutamate dehydrogenase–glutamate complex, (Bhuiya *et al*., 2005) respectively. Direct interactions between the superrepressor mutants and the substrate are shown as dotted lines. (B) The GudB1 monomer, viewed in the same orientation as RocG in (A). The three residues that are directly repeated in the *B. subtilis* 168 strain but are present only once in the "wild type" are highlighted in red. (C) The surface of the RocG hexamer viewed along the 2-fold axis, with each chain colored individually. The monofunctional D122N mutation is shown in blue and locates to a possible interaction surface centered on the molecular 2-fold axis. (D) Analysis of the effect of the D122N substitution on the surface electrostatics of RocG. The left panel shows wild-type RocG, while the right panel shows the D122N-substituted variant; both variants are shown in the same orientation as in Fig. 4c, with electrostatic calculations performed by the program APBS (Baker *et al*. 2001) and contoured at $\pm 5 K_b T / e_c$.

The majority of the mutated residues of the other RocG variants isolated in this study, are surface exposed and cluster to the immediate surroundings of the active-site cleft. Superposition of the structures of glutamate dehydrogenase from *Clostridium symbiosum* (Stillman *et al*., 1993) and *Pyrobaculum islandicum* (Bhuiya *et al*., 2005) has allowed us to model the expected position of the substrate and the nucleotide cofactor, respectively. Six out of 10 of these mutants (Glu93, Tyr158, Ser234, Met277, Ala302, and Ala303) are within 8 Å of the expected position of the nucleotide cofactor (with respect to the closed form of the enzyme), and three of these (Ala302, Ala303, and Met277) are in positions where they are able to make direct van der Waals interactions (Fig. 2.4A) On the other hand, the D122 residue, at which a mutation to asparagine increases the K_m for glutamate, is not within the active-site groove; instead, it is over 12 Å away from either nucleotide or substrate and solvent exposed, facing towards a shallow depression that surrounds the molecular 2-fold axis (Fig. 2.4C).

GudB1

Similarly, the GudB1 variant used in this study is a mutant with respect to the laboratory strain *B. subtilis* 168, but this allele is found in environmental isolates and can thus be regarded as "wild type" (Zeigler *et al*., 2008). The laboratory strain contains an insertion of the three-residue repeat $V_{95}K_{96}A_{97}$, which is located within the third helix of domain I (Fig. 4b). This insertion presumably causes severe destabilization of the fold of the protein, leading to an inactive enzyme that is very quickly degraded; consistent with this premise, recombinant GudB from *B. subtilis* 168 cannot be purified

in our hands. This is supported by the extremely rapid proteolytic degradation of GudB in *B. subtilis* (Gerth *et al*., 2008). By contrast, GudB1 is an active glutamate dehydrogenase with enzymatic activities of 6.7 and 24.5 U/mg of protein for catabolic and anabolic reactions, respectively. The K_m values were found to be 17.9 mM for glutamate and 41 mM for ammonium.

Both the open form and the closed form of GudB1 were evident in the six molecules in the asymmetric unit (two closed subunits and four open subunits), which associate to form the same compact homohexamer as RocG. This variation indicates that the closed conformation is quite readily adopted under different crystal forms. As would be expected given the 74% sequence identity between RocG and GudB1, their structures are very similar: the 'open' conformers can be superimposed on 409 aligned residues to yield an r.m.s.d. of 0.95 Å. A comparison of the 'closed' conformers of RocGE93K and GudB1 yields an r.m.s.d. of 1.2 Å, whereas that between the 'open' form and the 'closed' form yields an r.m.s.d. of 2.5 Å.

Discussion

In this study, we have isolated and characterized a set of RocG variants that uncouple the enzymatic and regulatory activities of this trigger enzyme. Several mutations result in loss of enzymatic activity and are accompanied by permanent substrate-independent inactivation of GltC. A second class, exemplified by one mutant protein, has retained enzymatic activity but has lost the ability to inactivate GltC. It has been shown that the catalytically active glutamate dehydrogenases RocG and GudB1 are able to inactivate GltC and thereby prevent transcription activation of glutamate biosynthetic genes (Belitsky & Sonenshein, 2004; Commichau *et al*., 2007a; Commichau *et al*., 2007b). Moreover, the cofactor glutamate is required to elicit the formation of the inhibitory forms of RocG or GudB1 (Commichau *et al*., 2007a). Similarly, the *B. subtilis* glutamine synthetase binds the transcription factors TnrA and GlnR if the enzyme is feedback inhibited by glutamine (Wray *et al*., 2001; Wray & Fisher, 2005).

The separation of enzymatic and regulatory activities has already been observed for other trigger enzymes such as the *B. subtilis* glutamine synthetase. As shown in this

study for RocG, many of these mutations cluster to the region of the active site of the glutamine synthetase (Wray & Fisher, 2005; Fisher & Wray, 2009). We showed that the mutations causing the superrepressor phenotype reduce the catalytic activity of RocG (see Tab. 2.3). Thus, one could assume that the conformation of these RocG variants required for GltC inhibition is more stable than that of the wild-type RocG enzyme. These variants might require less glutamate to elicit the "inhibitory" conformation of glutamate dehydrogenase. By contrast, the monofunctional RocG enzyme is enzymatically active, but glutamate does not induce the inhibitory conformation required to inactivate GltC activity.

Dual effects on regulatory interactions have also been observed for sugar permeases of the *B. subtilis* phosphotransferase system, which phosphorylate and thereby inactivate transcriptional anti-terminators and activators that control the expression of the permeases. Mutations that result in complete inactivation of the permeases (such as deletion of corresponding genes) result in the constitutive activity of the corresponding transcription factors. Similarly, mutations affecting the phosphorylation of the anti-terminators or activators by the sugar permeases (such as mutations of the phosphorylation sites) lead to permanent activity of the transcription factors and concomitant loss of sugar transport activity. By contrast, mutations that affect the membrane-bound transport domain lock the permeases in the phosphorylated state, since the phosphoryl group can never be transferred to the incoming sugar. These permease variants phosphorylate and thereby inactivate their cognate regulator proteins even in the presence of the normally inducing sugar in the medium. Such mutations have been intensively studied for the *B. subtilis* fructose permease (LevDEFG) and the glucose permease PtsG, which control the activities of the transcription activator LevR and the transcriptional anti-terminator GlcT, respectively (Stülke *et al.*, 1997; Bachem & Stülke, 1998; Martin-Verstraete *et al.*, 1998; Schmalisch *et al.*, 2003). In *E. coli*, proline dehydrogenase acts as the repressor for its own transcription in the absence of the substrate proline. In this case, the link between the two activities of the protein is more intricate, and mutations that lead to permanent repression of *putA* expression result from loss of proline binding and, thus, enzymatic activity (Muro-Pastor & Maloy, 1995).

Mutations affecting the control of gene expression by trigger enzymes have not only been found in the genes encoding the trigger enzymes. Similarly, mutations

affecting the controlled transcription factors may interfere with their productive interaction with the cognate trigger enzyme. Such mutations have been isolated for GltC, and reduced interaction with RocG has been demonstrated for one of these variants (Belitsky & Sonenshein, 1995; Commichau *et al*., 2007b).

An interesting result of this study is the identification of 10 different superrepressor variants of RocG, while only one monofunctional protein incapable of inhibiting GltC was found. Originally, we would have expected to find more monofunctional proteins than superrepressors. The small number of monofunctional RocG variants isolated might indicate that the GltC interaction surface is relatively large and involves multiple side chains, and that the interaction between the two proteins and the inhibition of GltC by RocG cannot be easily disrupted by single amino acid substitutions in RocG. The location of Asp122 on the surface of the enzyme, close to a shallow depression that surrounds the 2-fold axis of RocG (Fig. 2.4C), suggests that this region may be involved in the interaction of RocG with GltC. Indeed, it could be envisaged that this 2-fold axis is coincident with one of the 2-fold axes of GltC, which, based on structural studies of other LysR-type transcriptional regulators, is believed to be a tetramer with two 2-fold axes (Maddocks & Oyston, 2008). Furthermore, the fact that the substitution of this residue with an isosteric asparagine is capable of disrupting the inactivation of GltC by RocG(E93K) indicates that the charge properties of this residue are more important than its shape and size. This may be explained by the fact that in the RocG(E93K) structure, Asp122 forms an ion pair with Arg124; this interaction would most likely be disrupted in the D122N mutation, perhaps causing a larger alteration in the surface properties of the enzyme than would otherwise be expected. Analysis of the impact of this substitution on the surface electrostatics of RocG using the program APBS (Baker *et al*., 2001) reveals a significant increase in the basic character of the region surrounding the shallow depression, perhaps indicating how an isosteric substitution could cause a large disturbance in an interaction surface (Fig. 2.4D). On the other hand, any mutation that interferes with enzymatic activity without drastic changes in the overall structure of the protein might lead to permanent inhibition of GltC. This idea is supported by the clustering of the superrepressor mutations around the position of the nucleotide cofactor (Fig. 2.4A) and is also in good agreement with the large number of such mutations.

In most bacteria, glutamate dehydrogenase is required for ammonium assimilation. In *E. coli*, the enzyme can synthesize glutamate at high ammonium concentrations due to the low affinity for this substrate (Reitzer, 2003). In *B. subtilis*, the glutamate synthase encoded by the *gltAB* operon is the only enzyme for ammonium assimilation. Even the decryptified GudB1 protein is unable to catalyze glutamate formation *in vivo* (Commichau *et al*., 2008). The reason for this difference between the glutamate dehydrogenase of *E. coli* and the glutamate dehydrogenase of *B. subtilis* remains opaque. The data presented herein do, however, provide an explanation for the exclusive catabolic activity of the *B. subtilis* enzyme: glutamate dehydrogenase of *E. coli*, with a K_m of 2.5 mM for ammonium (Sharkey & Engel, 2008), is only active at high ammonium concentrations (Reitzer, 2003). However, the K_m values of the *B. subtilis* enzymes are even higher (18 and 41 mM for RocG and GudB1, respectively). The ammonium concentration in the cell is restricted by an active transport system. At high extracellular ammonium concentrations, the very small fraction that is present as ammonia diffuses freely into the cell, whereas at low ammonium concentrations, the ammonium transporter NrgA is expressed, and ammonium is taken up from external sources (Wray *et al*., 1994; Detsch & Stülke, 2003).

Thus, even high external ammonium concentrations do not result in high internal concentrations that would be sufficient for significant glutamate biosynthesis by the glutamate dehydrogenase.

The results presented in this study support the idea of an inhibitory interaction between the glutamate dehydrogenase RocG and the transcription factor GltC. Moreover, they provide new insights into the relation between the enzymatic activities and the regulatory activities of the trigger enzyme RocG. The identification of the monofunctional RocG variant that has lost its regulatory function but has retained enzymatic activity is in excellent agreement with the previous conclusion that the enzymatic activity of RocG is important but not sufficient for the control that it exerts on GltC. The availability of mutations in both partners that interfere with the inhibition of GltC by RocG will be of great value in the elucidation of the molecular details behind this regulatory interaction.

Acknowledgements

3. **Chapter 3**

Analysis of the RocG-GltC interaction

Abstract

In the glutamate metabolism of *B. subtilis* the glutamate dehydrogenase RocG fulfills two functions. First, RocG catalyzes the deamination of glutamate to 2-oxoglutarate that serves as a source of carbon. Second, RocG acts as a trigger enzyme by controlling the activity of the transcriptional regulator GltC. The activity of GltC is essential for the expression of the *gltAB* operon, encoding the glutamate synthase (GOGAT). In the presence of glutamate or its precursor arginine RocG prevents GltC from the induction of GOGAT expression, thus a futile cycle of glutamate synthesis and degradation is avoided. Two classes RocG variants were characterized. The first class is drastically diminished in enzymatic activity but strongly inhibits GltC. The second class consists of a mutant RocG protein that lost the capacity to control GltC activity but remained enzymatically active.

This study provides information about the interaction between the glutamate dehydrogenase RocG and the transcriptional activator GltC. It is suggested that the mode of interaction involves more factors. Moreover, the capacity of the mutant RocG variants to physically interact with GltC was tested in *B. subtilis.* The monofunctional RocG protein that was shown to fail in inhibiting the GltC protein still interacts with GltC. This result leads to the assumption that interaction is not the equivalent of inhibition and the mechanism of RocG to trigger the GltC activity might be more complex.

Introduction

In *B. subtilis* glutamate metabolism is one of the most tightly regulated pathways. Briefly, glutamate is exclusively synthesized by the combined reactions of the glutamine synthetase and the glutamate synthase (GltAB) in the GOGAT cycle (see Fig. 1.1A) (Fisher, 1999). The catabolic reaction, the oxidative deamination of glutamate to 2-oxoglutarate and ammonium is performed by the glutamate dehydrogenase (GDH), RocG (Belitsky & Sonenshein, 1998). This enzyme is only capable of glutamate degradation (Belitsky & Sonenshein, 1998, Commichau *et al.*, 2008). The expression of the *gltAB* operon and the *rocG* gene is mutually exclusive (Commichau et al., 2007b). In the absence of arginine and in the presence of glucose, the *rocG* gene is subject to

catabolite repression. Under this condition the *gltAB* operon is expressed due to the activation by the transcriptional regulator GltC. In the absence of glucose and the presence of arginine, the *rocG* gene is strongly transcribed. Then, the RocG protein sequesters the transcriptional regulator GltC in the presence of its substrate glutamate which prevents the expression of the *gltAB* operon (Commichau et al., 2007a) (see Fig. 2.1B). Besides the positive effect of GltC on the *gltAB* expression, GltC exerts autorepression on the *gltC* gene that is divergently transcribed with respect to the *gltAB* operon (Belitsky *et al.*, 1995; Belitsky & Sonenshein, 1995). This autoregulation has not been shown to be dependent on the nitrogen source (Belitsky & Sonenshein, 1995).

In this work mutant variants of the trigger enzyme RocG were isolated, that show a different inhibition pattern concerning the GltC protein (Chapter 2). Two classes of variants were isolated. Firstly, the so called superrepressor RocG proteins (RocG-SR) were characterized. These RocG-SR proteins strongly inhibit GltC if glutamate or its precursor arginine is available but show lower enzymatic activity than the wild type protein. Secondly, a monofunctional RocG protein (RocG-MF) was isolated. This protein is fully active as a glutamate dehydrogenase but lost the capacity to inhibit GltC. All mutant proteins are based on a single amino acid exchange that alters the trigger enzyme RocG in a protein with only one function. In the RocG-SR proteins the enzymatic activity is abolished but they still regulate GltC activity, whereas the RocG-MF protein completely lost the ability to control GltC but this protein is still active as a glutamate dehydrogenase.

In this study, mainly two questions are addressed. On the one hand it was analyzed whether the RocG-GltC interaction can be detected in the bacterial two-hybrid system. This approach might give information about the mode of interaction. The result can provide hints whether the interaction is binary or whether other factors are involved. On the other hand the ability of the mutant RocG proteins to interact with GltC *in vivo* was investigated in a SPINE approach.

Materials and Methods

Bacterial strains and growth conditions

E. coli DH5α and BTH101 (Sambrook *et al*., 1989; Karimova *et al*., 1998) were used for cloning experiments and bacterial two-hybrid (B2H) analyses, respectively. The *E. coli* strains were grown in LB medium. The *B. subtilis* strains GP717 (*trpC2 ΔgudB::cat rocG::*Tn*10 spc amyE::(gltA-lacZ aphA3) gltB1 ansR-C107A*) (Commichau *et al*., 2008) and GP28 (*trpC2 rocG*::Tn*10 spc ΔgudB*::*cat amyE*::*gltA'–'lacZ aphA3*) (Commichau *et al*., 2007b) were grown in SP medium, LB medium or in C minimal medium supplemented with tryptophan (at 50 mg/l) (Wacker *et al*., 2003). CSE medium is C minimal medium supplemented with sodium succinate (6 g/l) and potassium glutamate (8 g/l). C-Glc is C minimal medium supplemented with glucose (5 g/l), and CS is supplemented with sodium succinate (6 g/l) (Wacker *et al*., 2003). CR medium is supplemented with 5 g/l arginine. LB and SP and were prepared by the addition of 17 g Bacto agar/l (Difco, Lawrence, USA) to LB or SP medium, respectively.

In vivo detection of protein–protein interactions

The isolation of protein complexes from *B. subtilis* cells was performed by the SPINE technology (Herzberg *et al*., 2007). To express the RocG protein and its derivatives fused to an N-terminal Strep-tag, the different *rocG* alleles were amplified (for primers see Tab. S1) and the resulting PCR products were cloned into the expression vector pGP380. The PCR products were digested with BamHI and HindIII and ligated to vector pGP380 (Herzberg *et al*., 2007). The resulting plasmids are pGP1708 (RocG), pGP1709 (RocG-MFD122N). These plasmids and pGP652 (RocG-SR3 E93K) (Tholen, 2008) were used to transform *B. subtilis* GP717 (Commichau *et al*., 2008). For cultivation one liter culture was inoculated to an OD_{600} of 0.1 with an overnight culture. This culture was grown at 37°C until OD_{600} 0.9–1.0 and divided. One half and aliquots of 1.5 ml for the *lacZ* assay were harvested immediately, and the other was treated with formaldehyde (6 g/l, 20 min) to facilitate the cross-linking (Herzberg *et al*., 2007). After cross-linking, the cells were also harvested and washed with a buffer containing 50 mM Tris–HCl (pH 7.5) and 200 mM NaCl. The pellets were lysed using a French press (20,000 p.s.i., 138,000 kPa; Spectronic Instruments, Garforth UK). After

lysis the crude extracts were centrifuged at 100,000 g for 1 h. For purification of the Strep-tagged proteins the resulting supernatants were passed over a Streptactin column (IBA, Göttingen, Germany; 0.5 ml bed volume). The recombinant proteins were eluted with desthiobiotin (IBA, Göttingen, Germany, final concentration 2.5 mM). Aliquots of the different fractions were subjected to SDS-PAGE. Prior to electrophoresis, the protein samples were boiled for 20 min in Laemmli buffer to reverse the cross-links. As a control, the *B. subtilis* strain GP717 carrying the empty vector pGP380 was used.

B2H assay

The bacterial two-hybrid (B2H) assay was used to analyze the protein–protein interactions between RocG and GltC (Karimova *et al*., 1998). The B2H system is based on the interaction-mediated reconstruction of adenylate cyclase (CyaA) activity from *Bordetella pertussis* in *E. coli*. The CyaA enzyme consists of two complementary fragments T18 and T25 that are not active when physically separated. Fusion of these fragments to interacting proteins results in functional complementation between the T18 and T25 fragments and the synthesis of cAMP. cAMP production can be monitored by measuring the β-galactosidase activity of the cAMP-CAP-dependent promoter of the *E. coli lac* operon. Thus, a high β-galactosidase activity reflects the interaction between the hybrid proteins. Plasmids pUT18 and p25-N allow the expression of proteins fused to the N-terminus of the T18 and T25 fragments of the CyaA protein, respectively, and the plasmids pUT18C and pKT25 allow the expression of proteins fused to the C-terminus of the T18 and T25 fragments of the CyaA protein, respectively (Karimova *et al*., 1998; Claessen *et al*., 2008). The plasmids pKT25-zip and pUT18C-zip served as positive controls for complementation. These plasmids express T18-zip and T25-zip fusion proteins that can associate due to the leucine zipper motifs resulting in an active CyaA enzyme and a high β-galactosidase activity. DNA fragments of the *rocG* and *gltC* genes were obtained by PCR and digested with the enzymes KpnI and XbaI (for primers, see Tab. S1). The PCR products were cloned into the four vectors of the two-hybrid system digested with the same enzymes, respectively. The resulting plasmids (see Tab. S2) were used for cotransformations of *E. coli* BTH101 and the protein–protein interactions were then analyzed by plating the cells on LB plates containing ampicillin (100 mg/ml), kanamycin (50 mg/ml), X-Gal (40 mg/ml) (5-bromo-4-chloro-

3-indolyl-β-D-galactopyranoside) and IPTG (0.5 mM) (isopropyl-β-D-thiogalacto-pyranoside), respectively. The plates were incubated for a maximum of 48 h at 30°C.

Materials and Methods Summary

Western blot analysis

Western blot analysis of GltC was performed as described previously (Gunka *et al.*, 2010). GltC were detected with rabbit polyclonal antiserum raised against *B. subtilis* GltC (Commichau *et al.*, 2007a).

Activity of the *gltA-lacZ* fusion

Quantitative assays of *lacZ* expression in *B. subtilis* were performed as described in Gunka *et al.*, subm.

DNA manipulation

All primer sequences are provided in Tab. S1. Methods used for DNA manipulation were performed as described previously (Gunka *et al.*, 2010).

Results

RocG-GltC interaction in the bacterial two-hybrid assay: Limits of this method

A variety of experiments were performed to characterize the Roc-GltC interaction in more detail. But none of the methods used (e.g. surface plasmon resonance (SPR), electrophoretic mobility shift assay (EMSA) data not shown) was successful in giving new insights into the mechanism of interaction (J. A. Newman, R. J. Lewis, F. M. Commichau, C. Herzberg, pers. comm.).

To test whether the protein-protein interaction between RocG and GltC can be reconstructed heterologously in *E. coli,* both protein were fused to the C- and N-terminus of the T18 and T25 domain of the adenylate cyclase of *B. pertussis*, respectively. Different combinations of the GltC and the RocG fusions were co-expressed in the *E. coli* strain BTH101. If a direct interaction between RocG and GltC occurred, the domains of the adenylate cyclase would get in close proximity leading to the production of cAMP (Karimova *et al*., 1998). The cAMP production was qualitatively monitored on plates containing IPTG and X-Gal (see Fig. 3.1). Thus, a strong interaction between the respective proteins is reflected by the formation of dark blue colonies on plates. The leucine zipper of the yeast GCN4 served as the positive control in this experiment.

As shown in Fig. 3.1, the GltC as well as the RocG protein tended to perform self-interaction indicated by blue colored colonies. This result is in good agreement with previous data for the RocG protein. Like many other glutamate dehydrogenases RocG assemblies as a hexamer consisting of two trimers (Gunka *et al*., 2010). The GltC protein has not been successfully subjected to crystallization until now. But it was shown for other members of the LysR family that they act as tetramers consisting of two dimers (Picossi *et al*., 2007; Maddocks & Oyston, 2008). The observed self-interaction of both proteins implies that all fusions were expressed and active in the bacterial two-hybrid assay. However, a clear interaction between the RocG and GltC protein could not be observed in this assay. *E. coli* cells expressing T18-GltC and RocG-T25 were only light-blue colored. As the co-expression of T18-GltC and leucine zipper control T25-zip resulted in an even more intensive blue color, the interaction between T18-GltC and RocG-T25 was defined as non specific.

Although all used fusions of the RocG and GltC protein with domains of the adenylate cyclase seemed to be expressed and active in this assay, no specific interaction between RocG and GltC was detected. Thus, this method is not appropriate to analyze this protein-protein interaction.

pUT18/pUT18C derivatives
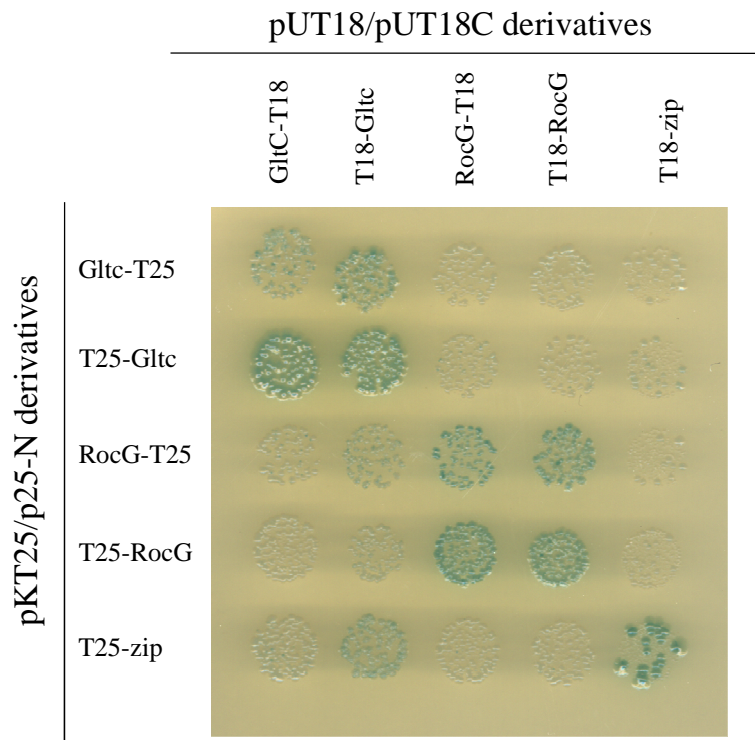
Fig. 3.1 – Bacterial two-hybrid assay of the RocG-GltC interaction.

Both genes were cloned into the plasmids pUT18, pUT18C, p25-N and pKT25. As illustrated, different combinations of the plasmids were co-expressed in *E. coli*. The transformants were incubated for 48 h at 30°C. Degradation of X-Gal (blue color) indicates the formation of a functional adenylate cyclase resulting from the interaction of the fused proteins.

SPINE approach with RocG variants

The expression of superrepressor RocG (RocG-SR) proteins in *B. subtilis* results in a strongly abolished *gltA-lacZ* expression in medium containing glutamate or its precursor arginine. Furthermore, the presence of the monofunctional RocG (RocG-MF) does not repress GltC activity indicated by a high *gltA-lacZ* expression (Gunka *et al*., 2010). The *gltA-lacZ* expression reflects the capacity of the RocG variants to control GltC. However, these data do not provide evidence whether the RocG variants,

especially the RocG-MF protein, have an altered interaction pattern with the GltC protein in comparison to the wild type RocG.
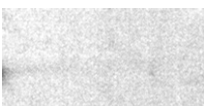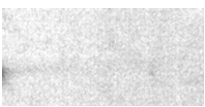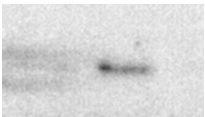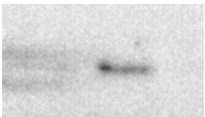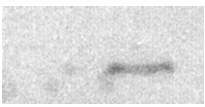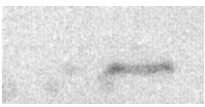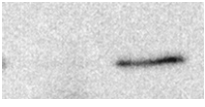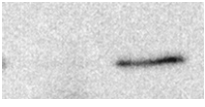
In order to get a snapshot of the GltC interaction exerted by the mutant RocG variants, a SPINE experiment was performed. In this approach the RocG protein was expressed with a Strep-tag in *B. subtilis*. By using formaldehyde, the status of the cell was fixed and the RocG protein could be purified via the Strep-tag with its *in vivo* interaction partners.

The *B. subtilis* strain GP717 was used in this study. This strain lacks both glutamate dehydrogenase genes and bears a *gltA-lacZ* fusion to monitor GltC activity. GP717 was transformed with plasmids containing the *rocG* wild type, the *rocG*-SR3 or the *rocG*-MF allele (Gunka *et al.*, 2010). All alleles were constitutively expressed from the plasmids. The empty vector pGP380 served as the negative control. The strains were grown in C minimal medium, containing 5 g/l glucose and 5 g/l arginine, to an OD_{600} of 1.0. One half was harvested and the other half was treated with 6 g/l (final concentration) formaldehyde and incubated for 20 min under agitation and harvested. Prior to the formaldehyde treatment small aliquots were taken for measuring *gltA-lacZ* activity just in time of the cross-linking. After the Strep-tag purification, the elution fractions were boiled to resolve the cross-linking and separated by SDS-PAGE. Additionally, a Western blot experiment with antibodies raised against the GltC protein was performed (see Fig. 3.2A).

As shown in Fig. 3.2A there was no GltC protein detected in the elution fraction of the empty vector pGP380. After cross-linking with formaldehyde, GltC was detected in the elution fraction of the RocG wild type protein, the superrepressor protein (RocG-SR3) and even the monofunctional RocG protein (RocG-MF). This result demonstrates that all used RocG proteins interact with the GltC protein under the condition tested.

Fig. 3.2B shows the *gltA-lacZ* activity of the strain GP717 transformed with the plasmids as indicated just prior to cross-linking with formaldehyde. As expected, in absence of any RocG protein GltC was fully active (empty vector pGP380) resulting in a high activity of the *gltA-lacZ* fusion with 443 Miller units. The strain expressing the RocG wild type protein showed a lower *gltA-lacZ* activity with 56 Miller units.

(A)			(B) *gltA-lacZ* expression (Miller Units)
	-FA	+FA	
pGP380 Empty vector			443
pGP1708 RocG WT			56
pGP652 RocG-SR3 (E93K)			13
pGP1709 RocG-MF (D122N)			561

Fig. 3.2 – Detection of the RocG-GltC interaction.

(A) 15 µl of each elution fraction were separated by a SDS-PAGE. The SDS-PAGE was blotted on a PVDF membrane. For detection of GltC polyclonal antibodies raised against the GltC protein were used. (B) *gltA-lacZ* expression in crude extracts of the *B. subtilis* strain GP717 transformed with the respective *rocG* alleles cultivated in C minimal medium with glucose (5 g/l) and arginine (5 g/l).

The presence of the RocG SR3 protein decreased the GltC activity even stronger than the wild type RocG resulting in only 13 Miller units of the *gltA-lacZ* fusion. This is in line with previously results (Gunka *et al.*, 2010). Due to these low *gltA-lacZ* activities it was tempting to speculate that the RocG wild type and SR3 protein interact with GltC under this condition. The presence of the RocG MF protein does not lead to an inhibition of the GltC activity. The *gltA-lacZ* fusion is highly expressed with 561 Miller units. However, this abolished inhibition cannot be explained by an eliminated interaction of RocG MF with GltC as shown in Fig. 3.2A. This result suggests that the pure physical interaction between RocG and GltC is not sufficient for GltC inactivation. Actually, the RocG MF protein still interacts with GltC but is not capable of exerting a negative effect on the GltC activity anymore.

Mutant RocG variants have no impact on the GltC protein level

In order to exclude that the effect of the mutant RocG variants on *gltA-lacZ* expression is based on an altered GltC protein level, the GltC expression was analyzed in a Western blot experiment (see Fig. 3.3).

GltC

SR1	SR2	SR3	SR5	MF	WT	
I308T	S234R	E93K	Y158H	D122N		Empty vector
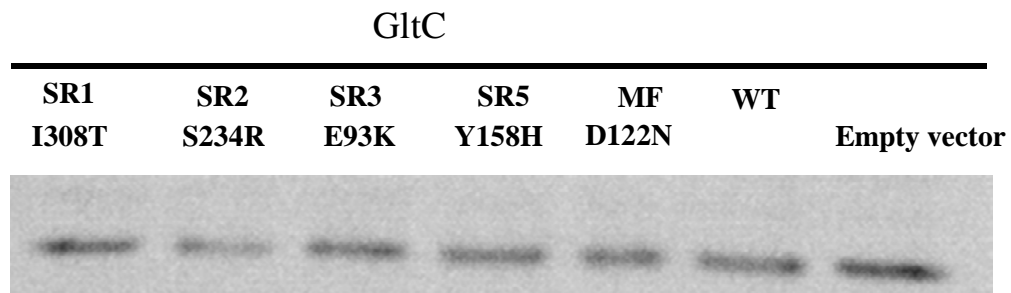
Fig. 3.3 – Expression of GltC in the context of RocG mutant proteins.

Crude extracts were isolated from *B. subtilis* strain GP28 (*rocG ΔgudB*) expressing the RocG superrepressor (SR) proteins SR1, SR2, SR3, and SR5, and the monofunctional (MF) RocG protein grown in CSE minimal medium. Strain GP28 either expressing the RocG wild-type (WT) protein or carrying the empty vector pBQ200 served as control to compare GltC protein level. GltC was detected using rabbit polyclonal antibodies raised against *B. subtilis* GltC. Samples (15 μg) of crude extract were applied per lane.

As shown in Fig. 3.3 the expression of the mutant RocG variants had no impact on the GltC expression. Therefore, it can be assumed that the effect of these RocG variants on *gltAB* expression is not linked to an altered GltC protein level.

Discussion

To the best of our knowledge, every method to reconstruct the RocG-GltC interaction *in vitro* has failed until now. Although the fusions of RocG and GltC protein seemed to be expressed and active in the bacterial two-hybrid assay as indicated by self-interactions, no specific interaction between RocG and GltC was detected. Even though it is accepted that every individual method to study protein–protein interactions can detect only about one third of the actual interactions (Braun *et al.*, 2009), the question remains why also approaches like SPR have failed in showing the interaction (J. A. Newman., pers. comm.) The most obvious explanation why the bacterial two-hybrid assay was not successful in showing the RocG-GltC interaction is that a factor is missing in *E. coli* that is present in *B. subtilis*. In principle, there are different classes of factors possible that facilitate protein-protein interactions.

First, low-molecular weight factors are known to be involved in a particular interaction. A well studied example is the glutamine synthetase in *B. subtilis*. In the presence of glutamine the glutamine synthetase interacts with the transcription regulator TnrA inhibiting the DNA binding activity of TnrA (Wray *et al.*, 2001). Moreover, the glutamine synthetase also regulates the activity of another important transcription regulator of nitrogen metabolism, GlnR, in a glutamine-dependent manner (Fisher & Wray, 2008). In analogy to this regulation it can be hypothesized that the interaction between RocG and GltC is triggered by glutamate as this amino acid is the substrate of the glutamate dehydrogenase RocG and can therefore by easily sensed by RocG. This idea is supported by the observation that RocG can only repress GltC activity in the presence of glutamate or its precursor arginine (Gunka *et al.*, 2010). But as it was demonstrated that glutamate is one of the most abundant metabolites also in *E. coli* (Bennett *et al.*, 2009) it is unlikely that glutamate is limiting in the bacterial two-hybrid assay. Thus, the availability of glutamate cannot be the only requirement for the RocG-GltC interaction.

The second possibility is that a third protein might be involved in the formation of the RocG-GltC complex. But if the interaction of RocG with GltC is not direct but mediated by another protein we would have expected to find an additional interaction partner in a SPINE approach in relevant amounts (Commichau *et al.*, 2007a). However, this possibility cannot be completely excluded.

Third, an obvious difference between the bacterial two-hybrid assay compared to the situation in *B. subtilis* is the presence of the *gltAB* promoter region that is the target of GltC binding. In this promoter region GltC can bind to multiple locations (Bohannon & Sonenshein, 1989; Belitsky *et al.*, 1995; Picossi *et al.*, 2007). The ratio of 2-oxoglutarate to glutamate is supposed to regulate GltC binding capacity to the different locations. At high 2-oxoglutarate concentration GltC activates the *gltAB*-expression whereas at higher glutamate concentrations the GltC conformation is altered and the expression of the *gltAB* genes is not activated (Picossi *et al.*, 2007). In this complex model of regulation it is tempting to speculate that RocG modulates the GltC activity by binding to GltC which has to be associated with its target promoter region. Perhaps GltC does not assemble in such a conformation that can be bound by RocG until it is located at its promoter region. This hypothesis would explain why the RocG-GltC interaction cannot be detected in the bacterial two-hybrid assay.

Moreover, the cross-linking experiment with the mutant RocG protein implies that interaction is not the equivalent of inhibition. It was demonstrated that all RocG proteins interact with GltC, although the RocG-MF lacks the ability to repress GltC activity. This result supports the idea that the RocG-GltC interaction is a very complex regulation mechanism. The interaction between RocG and GltC seems to have different effects on the GltC activity. Perhaps the effect is influenced by the presence of metabolites e.g. glutamate and 2-oxoglutarate. Furthermore, it was demonstrated that the RocG mutant variants specifically influence the GltC activity in terms of *gltAB* transcription activation and not the autoregulative properties of GltC. To get a more precise insight into the interaction mechanism, this topic needs to be subject or further investigation.

Future perspectives

As discussed above, alternative evidence to the SPINE experiment for the RocG-GltC interaction is still missing. To test the hypothesis, whether the protein-protein interaction between RocG and GltC depends on the presence of the target promoter region of the transcription activator GltC, the defined DNA sequence can be introduced in the background of the bacterial two-hybrid assay. This can be simply achieved by cloning the promoter region into one of the vectors for the RocG and GltC co-expression. If the *gltAB* promoter is necessary to facilitate the right conformation of

GltC and enables RocG to interact with GltC, this interaction should be detected in the bacterial two-hybrid assay in the presence of the promoter sequence.

Furthermore, a SPINE experiment in the absence of glutamate or a source of it can provide more insights into the mechanism of GltC regulation by RocG. If RocG also interacts with GltC in the absence of glutamate this would support the idea that the impact of the interaction on GltC activity depends on other factors e.g. the ratio glutamate to 2-oxoglutarate. A barrier in performing this experiment is that the RocG protein is naturally not expressed under this condition (Belitsky & Sonenshein 1999). In order to circumvent this problem RocG could be constitutively expressed from a plasmid and serve as the bait to catch GltC. Another opportunity to test the interaction in the absence of glutamate is to express GltC with a Strep-tag in a *B. subtilis gudB1* mutant strain. The active glutamate dehydrogenase GudB1 was shown to interact with GltC and is expressed under this condition (Commichau *et al.*, 2007a, Gunka *et al.*, 2010 subm.).

4.   **Chapter 4**

A high-frequency mutation in *Bacillus subtilis*: Requirements for the decryptification of the *gudB* glutamate dehydrogenase gene

The work described in this chapter was submitted for publication in:

Gunka, K., Tholen, S. Neme, R., Freytag, B., Commichau, F. M., Herzberg, C. & Stülke, J. A high-frequency mutation in *Bacillus subtilis*: Requirements for the decryptification of the gudB glutamate dehydrogenase gene. Mol. Microbiol. submitted

Author contributions: This study was designed and interpreted by KG, ST, FMC and JS. KG and ST performed most of the experiments. RNG constructed several mutant strains and performed Northern Blot experiments. BF contributed to this study by generating some mutant strains. FMC helped to develop the artificial mutation system and CH was essential to improve the method part. KG and JS wrote the paper.

Abstract

Bacillus subtilis encodes two glutamate dehydrogenases, the enzymatically active protein RocG and the cryptic enzyme GudB that is inactive due to a duplication of three amino acids in its active centre. The inactivation of the *rocG* gene results in poor growth of the bacteria on complex media likely due to the accumulation of toxic intermediates. Therefore, *rocG* mutants readily acquire suppressor mutations that decryptify the *gudB* gene. This decryptification occurs by a precise deletion of one part of the nine base pair direct repeat that causes the amino acid duplication. This mutation occurs at the extremely high rate of 10^{-4}. Mutations affecting the integrity of the direct repeat result in a strong reduction of the mutation rate; however, the actual sequence of the repeat is not essential. The mutation rate of *gudB* was not affected by the position of the gene on the chromosome. When the direct repeat was placed in the completely different context of an artificial promoter, the precise deletion of one part of the repeat was also observed, but the mutation rate was reduced by three orders of magnitude. Thus, transcription of the *gudB* gene seems to be essential for the high rate of the appearance of the *gudB1* mutation. This idea is supported by the finding that the transcription-repair coupling factor Mfd is required for the decryptification of *gudB*. The Mfd-mediated coupling of transcription to mutagenesis can be regarded as a built-in precaution that facilitates the accumulation of mutations preferentially in transcribed genes.

Introduction

As the central amino group donor for nearly all biosynthetic pathways in any living cell, glutamate plays a key role in the biochemistry and physiology of all organisms (Commichau *et al.*, 2006). Investigations with *Escherichia coli* demonstrate that glutamate is by far the most abundant metabolite in these bacteria accounting for about 40% of the internal metabolite pool (Yuan *et al.*, 2009). Moreover, glutamate is one of the most highly embedded metabolites: In the Gram-positive soil bacterium *Bacillus subtilis*, at least 37 reactions make use of this amino acid (Oh *et al.*, 2007).

In *B. subtilis*, glutamate is exclusively synthesized from 2-oxoglutarate and glutamine by the activity of glutamate synthase. 2-Oxoglutarate is replenished in the

citric acid cycle, whereas glutamine can be synthesized with ammonium as the nitrogen source and one of the two molecules of glutamate that are generated by glutamate synthase as the acceptor. Glutamate does also serve as a precursor for proline biosynthesis, and under conditions of osmotic stress, molar concentrations of proline have to be produced (Kempf & Bremer, 1998). Thus, it is not surprising that glutamate synthesis has to be a highly efficient process, and indeed, interactions between enzymes of the branch of the citric acid cycle that generates 2-oxoglutarate and glutamate synthase have been reported (Meyer *et al.*, 2011). Glutamate can also serve as source of carbon and nitrogen. Its utilization is initiated by an oxidative deamination catalyzed by the glutamate dehydrogenase. The expression of the genes encoding glutamate biosynthetic and catabolic enzymes is subject to complex control mechanisms that allow the adjustment of the intracellular glutamate concentration to the actual requirement (Belitsky & Sonenshein, 1999; Belitsky *et al.*, 2004; Commichau *et al.*, 2007b; Picossi *et al.*, 2007; Sonenshein, 2007).

   B. subtilis encodes two glutamate dehydrogenases, GudB and RocG (Belitsky & Sonenshein, 1998). However, the *gudB* gene experienced an inactivating mutation during domestication, resulting in an inactive pseudogene in the laboratory strain *B. subtilis* 168. In contrast, the *gudB* gene encodes an active enzyme in wild isolates and in non-domesticated strains such as NCIB3610 (Zeigler *et al.*, 2008). The inactivation of *gudB* is caused by a duplication of nine base pairs of the coding sequence resulting in a duplication of three amino acids in the active centre of the protein. The glutamate dehydrogenase RocG catalyzes the final step of the catabolic pathway for arginine, ornithine and citrulline. Accordingly, its expression is strongly induced in the presence of arginine (Belitsky & Sonenshein, 1998). Interestingly, the glutamate dehydrogenases are not only required for glutamate utilization, but they are also involved in the control of glutamate biosynthesis: In the presence of glutamate they inhibit the transcription activator GltC that is necessary for the expression of the glutamate synthase operon, *gltAB* (Commichau *et al.*, 2007a; Commichau & Stülke, 2008; Gunka *et al.*, 2010). In the active state, the two glutamate dehydrogenases are very similar to each other, both at the level of the amino acid sequence and also concerning their structures. In contrast, the inactive GudB protein seems to misfold and is subject to rapid degradation (Gerth *et al.*, 2008; Gunka *et al.*, 2010).

The importance of glutamate for the cellular physiology is underlined by the observation that any mutation that disturbs the glutamate homoeostasis results in the accumulation of suppressing mutations. This is true for both *E. coli* and *B. subtilis* (Yan, 2007; Commichau *et al*., 2008). In the laboratory strain of *B. subtilis*, the inactivation of the *rocG* gene encoding the only active glutamate dehydrogenase results in the appearance of mutants with an active GudB enzyme (these alleles are designated GudB1) (Belitsky & Sonenshein, 1998). Moreover, *rocG gudB* double mutants easily acquire suppressive mutations affecting the glutamate synthase (Commichau *et al*., 2008). The *rocG gudB* double mutants are unable to utilize glutamate as the source of carbon and nitrogen. However, cultivation of such mutants in the presence of glutamate or its precursors such as arginine results in the selection of suppressor mutants that catabolize glutamate by a pathway that is not operative in wild type bacteria. The analysis of one such mutant revealed constitutive expression of the aspartase pathway due to the inactivation of the repressor of the corresponding *ansAB* operon, AnsR (Flórez *et al*., subm.).

The accumulation of mutations that restore growth of mutants or that allow faster growth is a common phenomenon in bacteria. Several studies suggest that mutations that overcome the specific limitation are preferentially acquired (Cairns *et al*., 1988; Barrick *et al*., 2009); however, the underlying mechanisms have not yet been elucidated.

Mutations can be acquired during replication. Most of the errors are eliminated by DNA mismatch repair, including the MutSL system which contributes to genome stability (Modrich & Lahue, 1996; Fukui, 2010). Some errors can escape from repair and may be beneficial for the organism. Many bacteria including *B. subtilis* possess systems for the induction of mutations in the stationary phase (Sung & Yasbin, 2002). The emergence of these mutations is associated with transcription rather than with DNA replication and plays an important role in the generation of diversity in nondividing populations of *B. subtilis*. The process of transcription-coupled DNA repair is crucial for the accumulation of mutations in the stationary phase, and this involves the transcription repair coupling factor Mfd (Ayora *et al*., 1996). The Mfd protein targets DNA lesions during transcription that provoked a roadblock of transcription. Subsequently, Mfd may displace the RNA polymerase and recruit the nucleotide excision repair system to resolve the lesion (Borukhov *et al*., 2005; Truglio *et al*., 2006).

It was suggested that this process favors the acquisition of beneficial mutations of highly transcribed genes (Ross *et al.*, 2006; Pybus *et al.*, 2010).

We are interested in the mechanism by which the decryptification of the *gudB* gene occurs in *rocG* mutants. The *gudB1* mutation appears during growth and requires a deletion of nine base pairs. Therefore, *gudB* provides a unique system to study the emergence of mutations. Our results suggest that the decryptification of *gudB* requires the presence of a perfect direct repeat. Moreover, a part of this repeat is only deleted in the context of a transcribed gene, and this deletion requires the Mfd transcription repair coupling factor.

Materials and Methods

Bacterial strains and growth conditions

All *B. subtilis* strains used in this work are derived from the laboratory wild type strain 168. They are listed in Tab. S3. *E. coli* DH5α (Sambrook *et al*., 1989) was used for cloning experiments. *B. subtilis* was grown in SP medium, LB medium or in C minimal medium supplemented with tryptophan (at 50 mg/l) (Wacker *et al*., 2003). CSE medium is C minimal medium supplemented with sodium succinate (6 g/l) and potassium glutamate (8 g/l). C-Glc is C minimal medium supplemented with glucose (5 g/l), and CS is supplemented with sodium succinate (6 g/l) (Wacker *et al*., 2003). Additional sources of carbon and nitrogen were added as indicated. *E. coli* was grown in LB medium and transformants were selected on plates containing ampicillin (100 µg/ml). LB, SP and CS plates were prepared by the addition of 17 g Bacto agar/l (Difco, Lawrence, USA) to LB, SP or CS medium, respectively.

DNA manipulation, transformation and phenotypic analysis

Transformation of *E. coli* and plasmid DNA extraction were performed using standard procedures (Sambrook *et al*., 1989). Restriction enzymes, T4 DNA ligase and DNA polymerases were used as recommended by the manufacturers. DNA fragments were purified using the QIAquick PCR Purification kit (Qiagen, Hilden, Germany) Phusion™ DNA polymerase (New England Biolabs, Frankfurt am Main, Germany) was used for the polymerase chain reaction as recommended by the manufacturer. All primer sequences are provided as supplementary material (Tab. S1). DNA sequences were determined using the dideoxy chain termination method (Sambrook *et al*., 1989). All plasmid inserts derived from PCR products were verified by DNA sequencing. Chromosomal DNA of *B. subtilis* was isolated as described (Kunst & Rapoport, 1995).

E. coli transformants were selected on LB plates containing ampicillin (100 µg/ml). *B. subtilis* was transformed with plasmid or chromosomal DNA according to the two-step protocol described previously (Kunst & Rapoport, 1995). Transformants were selected on SP plates containing kanamycin (Kan 10 µg/ml), chloramphenicol (Cm 5 µg/ml) spectinomycin (Spec 150 µg/ml) or erythromycin plus lincomycin (Erm 2 µg/ml and Lin 25 µg/ml).

In *B. subtilis*, amylase activity was detected after growth on plates containing nutrient broth (7.5 g/l), 17 g Bacto agar/l (Difco, Lawrence, USA) and 5 g hydrolyzed starch/l (Merck, Darmstadt, Germany). Starch degradation was detected by sublimating iodine onto the plates.

Quantitative studies of *lacZ* expression in *B. subtilis* were performed as follows: cells were grown in CSE medium supplemented with different carbon and nitrogen sources as indicated. Cells were harvested at OD_{600} of 0.6 to 0.8 for cultures in CSE medium and 0.8 to 1.0 for cultures in CSE medium with sugar. β-Galactosidase specific activities were determined with cell extracts obtained by lysozyme treatment as described previously (Kunst & Rapoport, 1995). One unit of β-galactosidase is defined as the amount of enzyme which produces 1 nmol of o-nitrophenol per min at 28° C.

Ectopic expression of *gudB* variants

To express the *gudB* gene at an ectopic site, we used plasmid pAC5 (Martin-Verstraete *et al.*, 1992). This plasmid allows integration of the cloned fragments into the *amyE* site of the *B. subtilis* chromosome. Briefly, the *gudB* gene was amplified with its natural promoter using the oligonucleotides ST1 and KG92 using chromosomal DNA of *B. subtilis* 168 as the template. The PCR product was digested with EcoRI and BamHI and cloned into pAC5 linearized with the same enzymes. The resulting plasmid pGP900 was used to introduce the *gudB* allele into the chromosome.

The direct repeat of *gudB* present in pGP900 was subjected to site-directed mutagenesis by a modified PCR protocol, the combined chain reaction (Bi & Stambrook, 1998). Primers ST1 and KG92 were used as outer primers. The primers KG119, KG120, and KG133 were used to introduce point mutations into the *gudB* coding region. These primers were phosphorylated at their 5' end and allowed ligation of the nascent elongation product initiated from ST1. The resulting products carrying the mutations were cut with EcoRI and BamHI and cloned into pAC5 digested with the same enzymes. The resulting plasmids were pGP1714 (G3T G9T), pGP1715 (G3T G9T G12T G18T), and pGP1721 (G12T G18T). The plasmids were linearized with PstI and used to transform *B. subtilis* (see Tab. S2).

Design and construction of a mutagenesis reporter system

In order to analyze the occurrence of the deletion of the repeat in a non-related sequence context, we developed a reporter system which is based on a promoter that is only active upon deletion of one part of the *gudB*-derived direct repeat. This artificial *alf* promoter controls the expression of genes coding for a kanamycin resistant determinant (*aphA3*) and *E. coli* β-galactosidase (see Fig. 4.2). To obtain the reporter strain, we first constructed plasmid pGP655 as follows: The promoterless *aphA3* gene was amplified from pDG780 (Guérot-Fleury *et al*., 1995) using the primer pair ST4/ST9. These oligonucleotides attached restriction sites for EcoRI and BamHI (ST9) and for BglII (ST4) to the PCR product. The fragment was digested with EcoRI and BglII and cloned into the integration vector pAC6 (Stülke *et al*., 1997) linearized with EcoRI and BamHI. The resulting plasmid pGP653 contained a promoterless *aphA3-lacZ* operon. The *alf* promoter fragment was obtained by hybridization of the complementary oligonucleotides ST7 and ST8. It was cloned between the EcoRI and BamHI sites of pGP653, resulting in plasmid pGP655.

Construction of mutant strains

The Δ*gudB* and Δ*mfd* mutant strains were obtained by applying the long flanking homology PCR (LFH-PCR) technique (Wach *et al*., 1996). For the construction of the Δ*gudB* mutant, an *aphA3* resistance cassette was first amplified from plasmid pDG780 using the primer pair kan-fwd/kan-rev (Guérot-Fleury *et al*., 1995). DNA fragments of about 1,000 bp flanking the *gudB* region at its 5' and 3' end were amplified using the primer pairs KG100/KG101 and KG102/KG103. Chromosomal DNA of *B. subtilis* 168 served as the template. The 3' end of the upstream fragment as well as the 5' end of the downstream fragments extended into the respective gene region, in a way that all expression signals of genes up- and downstream remained intact. The joining of the two fragments to the resistance cassette was performed in a second PCR. In this reaction we used the primer pair KG100/KG103 for the deletion of *gudB*. The PCR product was directly used to transform *B. subtilis* 168. The integrity of the regions flanking the integrated resistance cassette was verified by sequencing PCR products of about 1,000 bp amplified from chromosomal DNA of the resulting mutant GP1160.

The Δ*mfd* mutant GP1167 was obtained in a similar way. Briefly, the *ermC* resistance gene was amplified from pDG647 with the primers mls-fwd(kan)/mls-rev(kan) (Guérot-Fleury *et al*., 1995), and the regions flanking *mfd* were amplified using the primer pairs KG86/KG87 and KG88/KG89. The joining PCR was performed with the primer pairs KG86/KG89. The PCR product was used to transform *B. subtilis* 168. As described above, the integrity of the DNA fragments flanking the resistance cassette in GP1167 was verified by sequencing.

Construction of a *gudB-lacZ* fusion

To determine the activity of the *gudB* promoter, a translational fusion of the *gudB* promoter to a promoterless *lacZ* gene encoding β-galactosidase was constructed as follows. A DNA fragment containing the *gudB* promoter region was generated by PCR using the primers ST1 and ST2, digested with BamHI and EcoRI and cloned into the plasmid pAC5. The plasmid pAC5 contains a promoterless *lacZ* gene and allows the introduction of translational fusions into the *amyE* locus of *B. subtilis* (Martin-Verstraete *et al*., 1992). The resulting plasmid pGP651 was used to introduce the fusion into different *B. subtilis* mutants (see Tab. S3).

Determination of mutation rates

All *rocG* mutant strains were plated on CS medium to verify the presence of the cryptic *gudB* wild type allele in the respective strains. The *rocG* mutants were able to grow on this medium as long as the *gudB* gene had remained inactive (Commichau *et al*., 2007b). Mutation rates were determined by the method of the median (Lea & Coulson, 1949). Briefly, eleven cultures in CSE-Glc were inoculated to a density of 100 cells/ml with an overnight culture grown in the same medium. The cultures were incubated at 37°C to an OD_{600} of 2.0. For the analysis of culture titers appropriate dilutions of four cultures were plated on SP medium containing glucose to allow growth of the *rocG* mutant strains. To screen for *gudB1* mutations, appropriate dilutions of each culture were plated on SP medium. After 24 h colonies showing the *gudB1* phenotype (wild type-like colonies on SP plates) were counted. To be sure of the identity of the mutations, the *gudB* allele was sequenced for at least three independent suppressor mutants in each experiment. In every single case, the correct excision of one part of the

repeat (i.e. the *gudB1* mutation) was observed. For the determination of mutation rates of the *alf* promoter present in the strains GP1123 and GP1168, the bacteria were plated on SP medium containing kanamycin (60 µg/ml) and X-Gal (80 µg/ml).

Northern blot analysis

Preparation of total RNA and Northern blot analysis were carried out as described previously (Ludwig *et al*., 2001). Digoxigenin (DIG) RNA probes were obtained by *in vitro* transcription with T7 RNA polymerase (Roche Diagnostics, Basel, Switzerland) using PCR-generated DNA fragments as templates. The primer pairs used to amplify DNA fragments specific for *gudB* and *gapA* are listed in Tab. S1. The reverse primers contained a T7 RNA polymerase recognition sequence. *In vitro* RNA labelling, hybridization and signal detection were carried out according to the instructions of the manufacturer (DIG RNA labelling kit and detection chemicals; Roche Diagnostics, Basel, Switzerland). To determine the size of the *gudB* mRNA, we used the transcripts observed with a *gapA* probe as the standard. RNA stability was analyzed as described previously (Meinken *et al*., 2003). Briefly, rifampicin was added to logarithmically growing cultures (final concentration 100 µg/ml) and samples were taken at the time points indicated. The quantification was performed using the ImageJ software v1.42 (Abramoff *et al*., 2004).

Western blotting

For Western blot analysis, proteins were separated by 12.5% SDS-PAGE and transferred onto polyvinylidene difluoride (PVDF) membranes (Bio-Rad, Munich, Germany) by electroblotting. Rabbit anti-RocG (1:15,000) (Commichau *et al*., 2007b) served as the primary antibody. The antibodies were visualised by using anti-rabbit immunoglobulin G-alkaline phosphatase secondary antibodies (Promega, Madison, USA) and the CDP* detection system (Roche Diagnostics, Basel, Switzerland), as described previously (Commichau *et al*., 2007a).

Results

The *gudB1* suppressor mutation appears at an extremely high rate

In our previous studies we observed that the *gudB* gene readily acquired the *gudB1* suppressor mutation if the *rocG* gene was inactivated (Commichau *et al*., 2007b; Commichau *et al*., 2008). In order to describe this mutation event in a quantitative way, the rate of the *gudB* reversion was determined. For this purpose, eleven independent cultures of the *rocG* mutant strain *B. subtilis* GP747 were inoculated with approximately 100 cells/ml to reduce the likelyhood of very early mutants. The cultures were grown under non-selective conditions (in CSE medium supplemented with glucose) for 20 generations and plated on complex medium (SP medium) which is toxic for the *rocG* mutant but not for emerging *rocG gudB1* suppressor strains. Suppressor mutants were recognized since they grew as solid colonies (like the wild type strain *B. subtilis* 168) whereas the *rocG* mutant strain GP747 formed only very small opaque colonies on complex medium. To ascertain that the mutation had appeared during the cultivation and not as a result of selection on the plates, only suppressor mutants that were present after 24 hours were taken into consideration. The mutation rate was 1 x 10^{-4}. To the best of our knowledge, such a high mutation rate has never been observed before in *B. subtilis*.

The role of chromosomal location and the direct repeat for the high reversion rate of *gudB*

The extremely high rate of reversion of *gudB* might result from the presence of a direct repeat of nine base pairs. However, the chromosomal arrangement might play a role as well. To distinguish between these possibilities, we decided to address the role of the chromosomal location of the *gudB* gene first. For this purpose, we used the *B. subtilis* strain GP1163. In this strain, the chromosomal *gudB* gene was deleted and another copy of *gudB* under the control of its own promoter was inserted ectopically at the *amyE* site of the chromosome. The mutation rate of this strain was 0.49 x 10^{-4}. Sequence analysis of three randomly selected suppressor mutants revealed that all contained the *gudB1* mutation. This observation suggests that the chromosomal location has no major impact on the occurrence of the *gudB1* suppressor mutation (see Fig. 4.1A).

The direct repeat in *gudB* is a common feature irrespective of the chromosomal location of the *gudB* allele. Therefore, we attempted to address the role of this repeat by a mutation analysis. Since the direct repeat is located within the coding sequence of *gudB*, any mutation to be introduced into the direct repeat had to conserve the *gudB* open reading frame. In order to destroy the direct repeat, we replaced two G residues by T (position 3 and 9 of the repeat, corresponding to wobble bases of the codons for valine and alanine). This mutation was introduced into both the first and the second part of the direct repeat; the corresponding strains are *B. subtilis* GP1179 and GP1197, respectively. Moreover, we restored a direct repeat; albeit with a sequence that deviates from the original repeat by introducing the same mutations in both parts of the repeat. This strain was GP1180 (see Fig. 4.1A).

A

168 (*gudB*)	...	GTG	AAG	GCG	GTG	AAG	GCG...
		Val	Lys	Ala	Val	Lys	Ala
168 (*gudB1*)	...	---	---	---	GTG	AAG	GCG...
		-	-	-	Val	Lys	Ala
GP1179 (*gudB_mut1*)	...	GT**T**	AAG	GC**T**	GTG	AAG	GCG...
GP1180 (*gudB_mut2*)	...	GT**T**	AAG	GC**T**	GT**T**	AAG	GC**T**...
GP1197 (*gudB_mut3*)	...	GTG	AAG	GCG	GT**T**	AAG	GC**T**...

B

GP1179 (*gudB_mut1*)	...	---	---	---	GTG	AAG	GCG...	11 x
		-	-	-	Val	Lys	Ala	
GP1197 (*gudB1_mut3*)	...	---	---	---	GT**T**	AAG	GC**T**...	9 x
		-	-	-	Val	Lys	Ala	
GP1197 (*gudB1_mut3*)	...	GT**G**	---	---	---	AAG	GC**T**...	5 x
		Val	-	-	-	Lys	Ala	

Fig. 4.1 - The crucial role of the direct repeat for the decryptification of *gudB*.

(A) The wild type *gudB* sequence was mutated without changing the amino acid sequence. In GP1179 two G residues were replaced by T in the first half of the repeat (position 3 and 9 of the repeat). In GP1197 theses mutations were introduced in the second part of the direct repeat (position 12 and 18). The perfect direct repeat was restored in the strain GP1180. This study served to analyze the role of a perfect direct repeat in the rapid decryptification of the *gudB* allele. (B) Destroying of the direct repeat leads to a 15-fold decrease of mutation frequency. In all *gudB1* mutants derived from the strain GP1179, the first part of the imperfect repeat was excised. In nine out of 14 *gudB1* mutants derived from the strain GP1197 also the first half of the imperfect repeat was deleted whereas in five *gudB1* mutants an internal excision had occurred.

A comparison of the mutation rates revealed that the perfect repeat was a prerequisite for efficient accumulation of *gudB1* suppressor mutants. In the absence of a perfect direct repeat, the mutation rate was reduced by a factor of about 15 (0.036 x 10^{-4} and 0.034 x 10^{-4} for GP1179 and GP1197, respectively vs. 0.49 x 10^{-4} strain GP1163 carrying the wild type repeat). The introduction of compensatory mutations that restore the direct repeat did also restore the high rate of the appearance of the *gudB1* mutation (0.89 x 10^{-4} for GP1180). These results clearly demonstrate that the presence of the direct repeat is the decisive factor for the high *gudB1* mutation rate.

Selective excision of the first part of the direct repeat

In all experiments to determine mutation rates, we analyzed the nucleotide sequence of the *gudB* suppressor mutations. As stated above, a precise deletion of the direct repeat was observed in all cases. However, the presence of a perfect repeat precluded the identification of the nucleotides that had actually been excised. This question became tractable with the availability of the suppressor mutants of *B. subtilis* GP1179 and GP1197 in which the repeat is not perfect. The sequence analysis of the *gudB1* alleles of eleven suppressor mutants derived from GP1179 (mutated in the first part of the repeat) revealed that the first half of the repeat was deleted in all cases (see Fig. 4.1B). This strong bias might indicate that either the first part of the repeat is preferentially excised or that the naturally occurring sequence is retained with preference. This question was addressed by the analysis of suppressor mutants derived from GP1197 (mutations in the second part of the repeat). In this case, of 14 analyzed mutants, nine had a deletion of the first part of the repeat. Moreover, five mutants exhibited internal deletions of the repeat that restored a sequence coding for the active GudB protein (see Fig. 4.1B). Thus, none of the mutants derived from GP1197 restored the original nucleotide sequence of the remainder of the repeat. Instead, we observed again a strong bias towards deletion of the first part of the repeat suggesting that this selective deletion is inherent to the mutagenesis process that decryptifies the *gudB* gene.

Construction and analysis of a deletion reporter system

The results presented above demonstrate that the deletion of one part of the *gudB* repeat occurs at a very high frequency both in the native and in a non-related genomic context as long as the repeat is intact. These findings prompted us to ask whether the deletion would also take place as efficiently in a completely different sequence context as it does in the *gudB* gene. For this purpose, we constructed a reporter system consisting of an *aphA3-lacZ* operon encoding a resistance to kanamycin and β-galactosidase under the control of an artificial (*alf*) promoter. This promoter was designed to have perfect recognition sequences for the housekeeping sigma factor of the RNA polymerase (-10 and -35), however the spacing between the two boxes was 26 rather than the canonical 17 to 18 base pairs. The perfect repeat of the *gudB* gene should be a part of this spacer. This promoter is not likely to be recognized by the RNA polymerase unless one part of the repeat is deleted and the optimal 17 bp spacing is restored (see Fig. 4.2).
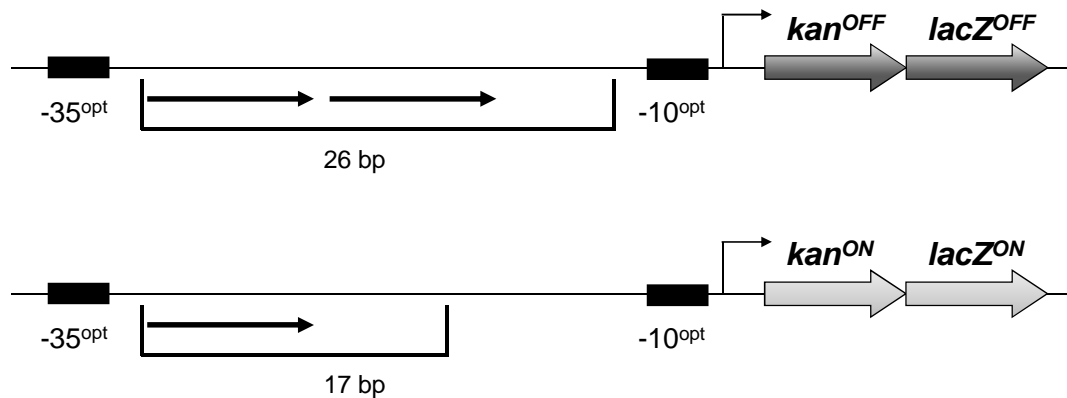

Fig. 4.2 - The mutagenesis test system.

The direct repeat originating from the *gudB* allele of *B. subtilis* was placed as the spacer between an optimal -10 and -35 region (upper part). An operon consisting of a kanamycin resistance gene (*aphA3*) and the β-galactosidase gene (*lacZ*) was placed under the control of the artificial promoter. Due to the long spacer, the promoter is not active. By the precise deletion of nine bp in the spacer region the promoter gains function and the kanamycin resistance und the β-galactosidase are highly expressed (lower part).

Such a reporter system was constructed as described in Material and Methods and introduced into the genome of *B. subtilis*, resulting in strain GP1123 (see Tab. S3). *B subtilis* GP1123 was unable to grow in the presence of kanamycin and formed white

colonies on plates containing X-Gal suggesting that neither kanamycin resistance nor β-galactosidase was expressed by these bacteria. These findings demonstrate that the *alf* promoter was inactive, as expected. However, we observed the sporadic appearance of kanamycin-resistant blue colonies that might result from the activation of the *alf* promoter. Indeed, a sequence analysis of the promoter for several colonies revealed the deletion of one part of the repeat resulting in a promoter (*alf1*) with perfect -10 and -35 regions separated by the preferred distance of 17 bp. Thus, the deletion of one part of the repeat occurs also in a non-related sequence context.

Next, we asked whether the deletion of the repeat in the *alf* promoter took place with a similar high frequency as observed for the decryptification of *gudB*. For this purpose, the rate of appearance of kanamycin-resistant suppressor mutants of GP1123 was determined. It was found to be 1.3×10^{-7}. This mutation rate is in the range typically observed in bacteria (Kunkel, 2004), but three orders of magnitude lower than the rate found for the deletion event in the *gudB* gene context. Thus, there seems to be a relevant difference between the sequence contexts of the *gudB* gene and the *alf* promoter that results in drastically changed mutation rates.

Expression of the *gudB* gene and stability of the cryptic and active glutamate dehydrogenases

While the direct repeat is part of a putatively expressed coding region in the *gudB* gene, it is present in the non-transcribed spacer in the artificial *alf* promoter. This difference might contribute to the different mutation rates observed in the two sequence contexts. Therefore, we decided to study first the expression of the *gudB* gene to some detail. Previous studies have shown that *gudB* expression is not modulated by the source of nitrogen present in the medium (Belitsky & Sonenshein, 1998). We have studied the activity of the *gudB* promoter by determining the expression of a *gudB-lacZ* fusion in wild type, Δ*rocG* and *gudB1* genetic backgrounds. As shown in Tab. 4.1, the fusion was expressed constitutively irrespective of the genetic background or medium analyzed. The expression level of about 500 units/mg of protein is quite high as compared to other *lacZ* fusions based on the same reporter system (Schilling *et al*., 2007). Thus, even the cryptic *gudB* gene coding for an inactive protein is constantly expressed at high levels in *B. subtilis*.

To allow the action of selective pressure on the decryptification of *gudB*, the accumulation of the active protein is required. However, the inactive GudB protein was reported to be one of the most unstable proteins of *B. subtilis* (Gerth *et al.*, 2008). In contrast, preliminary evidence suggested that the active GudB1 protein is much more stable (Gunka *et al.*, 2010). The issue of stability might apply not only at the level of the protein, but may also be relevant for the *gudB* mRNA. To address these problems, we first determined the stability of the *gudB* mRNA of *B. subtilis* GP747 and the isogenic *gudB1* mutant GP753 by a Northern blot analysis. As shown in Fig. 4.3A, we detected a single transcript of about 1,300 bp for *gudB*. This corresponds to a monocistronic transcript and is in good agreement with previous suggestions based on genome analysis (Belitsky & Sonenshein, 1998). The quantitative evaluation of the mRNA stability revealed a half-life of about 4 min. The stability of the mRNA was similar in both strains, demonstrating that it is not affected by the presence of the direct repeat.

Tab.4.1 - Analysis of *gudB* expression.

Strain	Relevant genotype	β-Galactosidase activity [a]				
		C-Glc	CE	CE-Glc	CR	CR-Glc
GP1101	wild type	357	NG	504	573	415
GP1102	*gudB1*	182	384	268	415	242
GP1104	*rocG*::Tn*10*	422	NG	557	NG	394
GP1105	*rocG*::Tn*10 gudB1*	225	658	478	410	295

[a] Bacteria were grown in C minimal medium. Glucose (Glc), glutamate (E), and arginine (R) were added to final concentrations of 5 g/l (Glc and R) or 8 g/l (E). β-Galactosidase activities are given as units/mg of protein. Experiments were carried out at least threefold. Representative results from one series are shown.

NG = no growth.

The accumulation of the glutamate dehydrogenase GudB was studied by Western blot analysis. For this purpose, we used the cell extracts of the *rocG* mutant GP747 and its isogenic *gudB1* derivative GP753 that were prepared for the

determination of the mRNA stability (just prior to rifampicin addition, t_0 in Fig. 4.3A). To detect the GudB protein, we used an antibody raised against RocG. Both proteins are very similar and the antibody recognizes GudB as well. Since both strains used for this experiment are *rocG* mutants, the only signal is obtained with GudB (Commichau *et al*., 2007b).
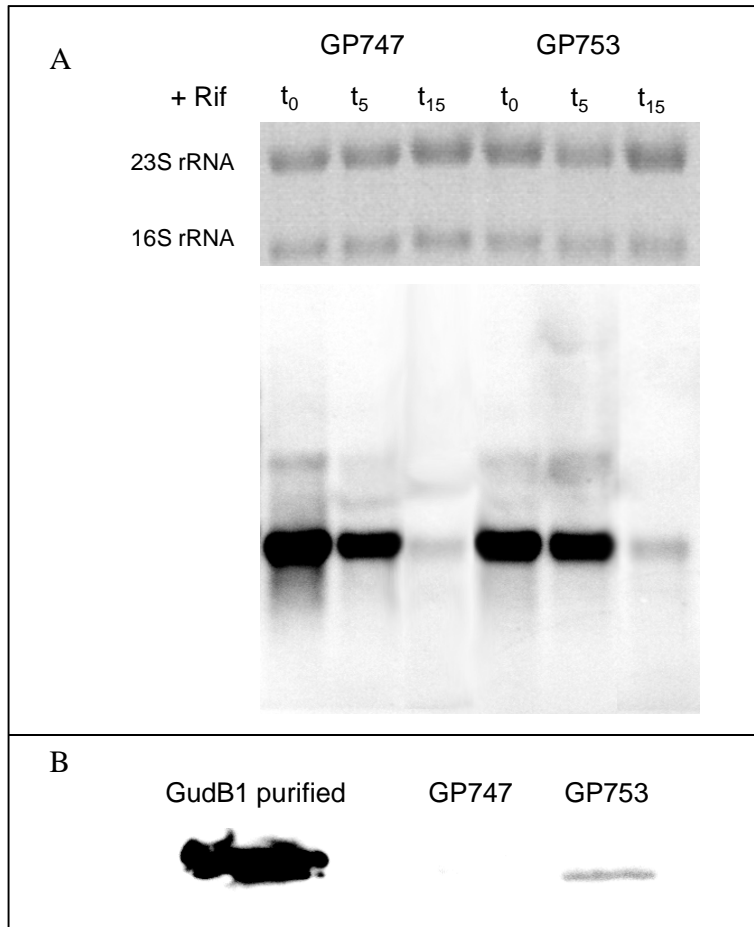

Fig. 4.3 - Expression of the *gudB* gene and the GudB protein level.

(A) Northern blot analysis was performed to determine the stability of the *gudB* mRNA of *B. subtilis* GP747 and the isogenic *gudB1* mutant GP753. Both mRNAs do not differ in stability and half-life, implying that the direct repeat does not influence mRNA levels. (B) A Western blot analysis was performed to compare the protein levels of GudB with the level of GudB1. Crude extracts of *B. subtilis* GP747 and the isogenic *gudB1* mutant GP753 were used and the GudB protein was detected by using antibodies raised against RocG that cross-react with the GudB protein.

As shown in Fig. 4.3B, the active enzyme GudB1 could be detected in the extract of GP753. In contrast, no signal was observed for the cryptic GudB protein. Since the mRNA amounts are similar for both strains (see Fig. 4.3A), we may conclude that the inactive GudB protein is highly unstable, as suggested by a previous study (Gerth *et al*., 2008). In contrast, the active glutamate dehydrogenase GudB1 is a stable protein that accumulates in the cell. Thus, the decryptification of *gudB* is sufficient for the cell to obtain immediately an active glutamate dehydrogenase that may help to overcome the metabolic imbalance of the *rocG* mutant.

Implication of repair and recombination proteins in the decryptification of *gudB*

The extremely high frequency at which the *gudB* decryptification occurs and the fact that the rate is much higher in the *gudB* locus as compared to the *alf* mutagenesis reporter system suggest the involvement of proteins in the mutagenesis process. The sequence of the direct repeat in *gudB* is somewhat similar to the *chi* sequence that is recognized and bound by the AddAB helicase/nuclease, a component of the recombination machinery of the cell. Since the recombination protein RecA is also involved in the generation of mutations, we determined the *gudB* mutation rate of the *addAB* and *recA* mutant strains GP1107 and GP1103, respectively. The rates were similar to those observed with the isogenic *rocG* mutants (0.9×10^{-4} and 1.1×10^{-4} for the *addAB* mutant and the wild type GP754; 0.3×10^{-4} vs. 1.3×10^{-4} for the *recA* mutant and the wild type GP747). Therefore, AddAB and RecA do not seem to play a major role in the deletion of the direct repeat in the *gudB* gene.

If AddAB and RecA would have played a role in the deletion of the direct repeat in *gudB*, we would have expected that they are not selective for the *gudB* gene context as compared to the context of the *alf* promoter. Thus, the genetic context plays a decisive role in the decryptification of *gudB*. As shown above, the *gudB* gene is constitutively expressed. In contrast, the core promoter of the mutagenesis reporter system is a non-transcribed region. The transcription-repair coupling factor Mfd might therefore participate in the deletion of the *gudB* repeat. To test this idea, we constructed the *mfd* deletion mutant GP1169 and compared the mutation rate in this strain to that of the isogenic *rocG* mutant GP747. In this case, we detected a hundred-fold reduction of the frequency of *gudB1* mutants (1.25×10^{-6} vs. 1.3×10^{-4}). Next, we investigated the impact of the *mfd* mutation on the deletion of the repeat in the *alf* promoter. In this case,

the mutation rates of the wild type (GP1123) and the isogenic *mfd* mutant GP1168 were very similar (1.3×10^{-7} vs. 2.3×10^{-7}). Thus, the *mfd* mutation affects the deletion of the direct repeat only in the context of the transcribed gene. This observation strongly supports the idea that transcription of the *gudB* gene is essential for obtaining the high rate of decryptification.

Discussion

In bacteria, mutations occur with a frequency of about 10^{-7} and the frequency of beneficial mutations is estimated to be even two orders of magnitudes less frequent (Imhof & Schlötterer, 2001; Kunkel, 2004). The *gudB1* mutation studied in this work appeared with a rate of about 10^{-4}. To the best of our knowledge, this is the highest mutation rate for a specific allele that has ever been observed in bacteria.

Since the pioneering work of Darwin it was thought that the occurrence of a mutation is a stochastic event that may affect each nucleotide with the same probability and that natural selection provides the choice between different variants (Darwin, 1859). However, research in the last few years suggests that selective pressure may somehow favor the appearance of beneficial mutations (Cairns *et al.*, 1988; Roth *et al.*, 2006). A recent long-term study with *E. coli* suggested that early during starvation the cells acquire the most beneficial mutations (*i.e.* those mutations that have the highest positive impact on fitness) and that independent bacterial cultures are likely to accumulate the same beneficial mutations (Barrick *et al.*, 2009). The high frequency and the high precision of the *gudB* decryptification have two implications: First, a specific selective pressure seems to stimulate the occurrence of the *gudB1* mutation. Second, the molecular tools to generate the mutation must be present in *B. subtilis*.

Previous work has shown that glutamate homoeostasis is important for both *E. coli* and *B. subtilis* (Yan, 2007; Commichau *et al.*, 2008). To any mutation that affects glutamate metabolism, the bacteria respond readily with the emergence of suppressor mutations. In the lab strain of *B. subtilis*, the inactivation of *gudB* seems to be very stable under laboratory growth conditions. Indeed, a derivative of the strain 168 with an active GudB glutamate dehydrogenase could only be selected on minimal medium with glutamate as the single carbon source (Belitsky & Sonenshein, 1998; Commichau *et al.*, 2008). The situation is completely different when the *rocG* gene that

encodes the final enzyme of the arginine degradation pathway is inactivated. These bacteria form only small translucent colonies on complex medium and rapidly acquire the *gudB1* mutation. The occurrence of the *gudB1* mutation at such a high frequency suggests the existence of a strong selective pressure exerted on the *rocG* mutant. The *rocG* gene product, the glutamate dehydrogenase, converts glutamate to 2-oxoglutarate. This suggests that glutamate or one of its precursors in the arginine degradation pathway might accumulate in the *rocG* mutant and this might be problematic for the cell. We have tested the growth of mutants affected in the different steps of arginine degradation on complex medium; however, the strong growth defect was unique to the *rocG* mutant (our unpublished results). Thus, the accumulation of glutamate may be toxic for the cell. This idea is in good agreement with the observation that a strain with a constitutive high-level expression of the glutamate synthesizing enzyme glutamate synthase acquired a mutation that inactivates this enzyme when grown in the presence of glutamate (Commichau *et al.*, 2008). This leaves us with the question why glutamate should be toxic for the cell when it is the most abundant metabolite anyway. The enzyme glutamate racemase (encoded by the essential gene *racE* in *B. subtilis*) catalyzes the conversion of L-glutamate to D-glutamate that is a building block for peptidoglycan biosynthesis (Kimura *et al.*, 2004; Spies *et al.*, 2009). Indeed, the accumulation D-glutamate was shown to be toxic for *B. subtilis* (Kimura *et al.*, 2004). In the presence of very high intracellular amounts of L-glutamate due to the strong induction of the enzymes of the arginine degradation pathway, RacE probably generates higher concentrations of D-glutamate than tolerated by the cell. The activation of the normally cryptic glutamate dehydrogenase GudB might then bring the glutamate concentration to a level that does not longer result in the accumulation of harmful D-glutamate.

A mutation can only be beneficial for a bacterium if it confers the cell with an immediate selective advantage. This was observed in the case of long-term evolution of *E. coli*: most of the mutations that were fixed after 5,000 generations resulted in an increased fitness of the cells (Barrrick *et al.*, 2009). In the case of GudB, a mutation has two solve two problems at the same time: first, the extremely unstable protein needs to be stabilized and, second, the stable enzyme must bring an advantage for the cell. As shown in this work, the decryptification of GudB is accompanied by drastic increase of the stability of the protein as judged from the increased amounts of the protein while the

expression of the gene is not affected by the mutation. The active GudB protein may than participate in reducing the cellular glutamate pool as outlined above. Thus, the decryptification of *gudB* meets both requirements.

A particular feature of the *gudB* gene is the presence of a perfect direct repeat. To our knowledge, this is a unique situation in the genome of *B. subtilis*. As shown in this work, the repeat is essential for the high rate of *gudB* decryptification. Tandem repeats in bacteria are usually instable and multiple mechanisms are involved in their contraction or expansion (Bichara *et al*., 2006). However, the well-studied tandem repeats such as the *lgtC* repeat in *Haemophilus influenzae* or the *nadA* repeat in *Neisseria meningitidis* consist of five to 36 repeats of tetranucleotides (Bayliss *et al*., 2001; Martin *et al*., 2005). In those cases, the instability is not linked to transcription (Bichara *et al*., 2006). The *gudB* repeat is unique in possessing a large repeat (unit repeat of nine nucleotides) that is present in only two copies. In contrast to the well-studied tandem repeats the *gudB* repeat is very stable in a non-related, non-transcribed genomic context. Our work provides first insights into the mechanism of the deletion of one part of the *gudB* repeat: The Mfd protein is essential for the high-frequency decryptification of *gudB* if the repeat is located in a transcribed region, and then the first part of the repeat is preferentially deleted. The identification of the particular enzymes that are required for the decryptification of *gudB* will be the subject of further analyses.

The Mfd-mediated coupling of transcription to DNA repair and mutagenesis can be regarded as a built-in precaution that facilitates the accumulation of mutations preferentially in transcribed genes. This has several implications: (i) The coupling allows that the mutations occur in genes that are expressed at the given timepoint; therefore the mutant variants of the encoded proteins might help to overcome the actual limitation. (ii) Non-transcribed genes that may be required under different conditions are in this way protected from potentially harmful mutations. Both effects facilitate the adaptation of bacteria to all kind of challenges that limit their growth and are therefore crucial for bacterial evolution.

Acknowledgements

5.   **Chapter 5**

Further factors involved in the Mfd-dependent decryptification of the *gudB* allele in *Bacillus subtilis*

Abstract

In the laboratory *Bacillus subtilis* strain 168 the *gudB* gene encoding an inactive glutamate dehydrogenase is rapidly decryptified upon deletion of the *rocG* gene coding for the active glutamate dehydrogenase. This decryptification occurs at a high mutation rate of about 10^{-4}. In the decryptified *gudB* allele, designated *gudB1*, nine base pairs of a direct repeat in the coding region are precisely excised, leading to the production of an active enzyme. A recent study showed that the high rate of *gudB1* mutation in a *rocG* mutant background depends on the perfect direct repeat and the presence of the Mfd protein. The involvement of Mfd, which is a transcription repair coupling factor, in the *gudB1* mutations implies the contribution of other proteins that are part of the DNA repair system to the mutation event. In order to identify these putative factors, mutants that are deficient in the nucleotide excision repair and the DNA repair mismatch systems were constructed. Both, the loss of the UvrA and UvrB proteins and the MutS and MutL proteins do not reduce the *gudB1* mutation rate as much as the Mfd deficiency. These results support the idea that other factors may operate together with Mfd in the *gudB1* decryptification.

Introduction

The genome of *B. subtilis* codes for two glutamate dehydrogenases, RocG and GudB. In the laboratory strain 168 only the RocG protein is active in degrading glutamate to 2-oxoglutarate and ammonium (Belitsky & Sonenshein, 1998), whereas the *gudB* gene encodes a cryptic glutamate dehydrogenase that cannot utilize glutamate as a source carbon (Belitsky & Sonenshein, 1998; Zeigler *et al.*, 2008). The inactivity of the GudB protein is caused by a duplication of nine base pairs in the coding region of the *gudB* gene leading to the duplication of the three amino acids in the active site of the protein (Belitsky & Sonenshein, 1998; Commichau *et al*, 2007b; Gunka *et al.*, subm.).

In the absence of the active glutamate dehydrogenase RocG, *B. subtilis* is not capable of utilizing glutamate anymore which results in a severe growth defect of a *rocG* mutant on medium containing glutamate and a variety of its precursors. This growth defect is rapidly suppressed by a mutation which activates the cryptic *gudB* allele. It was demonstrated that the suppressor mutation always occurs as the precise

deletion of nine base pairs of the direct repeat in the *gudB* gene (Belitsky & Sonenshein, 1998, Commichau *et al*., 2007a). The decryptification of the *gudB* allele occurs at an extremely high rate (1.3 x 10^{-4}) (Gunka *et al*., subm.). By this mutation the *gudB* allele, designated as *gudB1*, gains function and codes for a stable and active glutamate dehydrogenase (Belitsky & Sonenshein, 1998).

As shown, there are at least two requirements for the rapid decryptification of the *gudB* allele. (i) The presence of a perfect direct repeat and (ii) and the transcription repair coupling factors Mfd are crucial for the high mutation rate (Gunka *et al*., subm.). The lack of one of these requirements reduces the mutation rate about hundred-fold, respectively.

The Mfd protein is evolutionarily highly conserved and belongs to the group of transcription repair coupling factors (TCRF). These factors recognize stalled RNA polymerase caused by a DNA lesion. TRCFs displace the RNA polymerase and its transcripts and recruit the DNA excision repair machinery that resolves the lesion and allows the restart of transcription (Roberts & Park, 2004; Borukhov *et al*., 2005; Deaconescu *et al*., 2006). Until now, little is known about the molecular mechanism of the detailed recruitment of the DNA excision repair machinery by the TCRF Mfd. But evidence was provided that Mfd directly interacts with the UvrA subunit of the UvrABC nucleotide excision repair system (Selby & Sancar, 1993; Selby & Sancar 1995a).

Furthermore, it was shown that the Mfd protein is involved in the phenomenon of stationary phase mutagenesis in *B. subtilis* (Ross *et al*., 2006; Robleto *et al*., 2007; Pybus *et al*., 2010). Indeed, in a *B. subtilis* strain lacking the Mfd protein the accumulation of beneficial mutations is diminished. This is in good agreement with our findings that the frequency of *gudB1* mutations is decreased in a *mfd* mutant strain (Gunka *et al*., subm.).

Recently, it was demonstrated that another role of the Mfd protein is to clear the conflict between the DNA and the RNA polymerase in a head-on collision in *E. coli*. The Mfd protein arranges to solve this conflict in favor of replication by pausing transcription (Pomerantz & O'Donnell, 2010).

To further elucidate the molecular mechanism of the Mfd-dependent rapid *gudB* decryptification, it was studied which factors also contribute to the generation of *gudB* suppressors. Therefore, mutants of components of the UvrABC DNA excision repair

system and the MutSL proteins that participate in the DNA repair mismatch were generated. Experimental evidence cannot be provided that any of the proteins analyzed is involved in the case of the *gudB1* mutation. Thus, it is still an open question with which DNA repair machinery Mfd operates in the *gudB* decryptification.

Materials and Methods

Bacterial strains and growth conditions

All *B. subtilis* strains used in this work are derived from the laboratory wild type strain 168. They are listed in Tab. S3. *B. subtilis* was grown in SP medium, LB medium or in C minimal medium supplemented with tryptophan (at 50 mg/l) (Wacker *et al*., 2003). CSE medium is C minimal medium supplemented with sodium succinate (6 g/l) and potassium glutamate (8 g/l). C-Glc is C minimal medium supplemented with glucose (5 g/l), and CS is supplemented with sodium succinate (6 g/l) (Wacker *et al*., 2003). LB, SP and CS plates were prepared by the addition of 17 g Bacto agar/l (Difco, Lawrence, USA) to LB, SP or CS medium, respectively.

Construction of mutant strains

The Δ*mutSL* and Δ*uvrAB* mutant strains were obtained by applying the long flanking homology PCR (LFH-PCR) technique (Wach *et al*., 1996). For the construction of the *uvrAB* mutant, *ermC* resistance gene was amplified from pDG647 with the primers mls-fwd (kan)/mls-rev (kan) (Guérot-Fleury *et al*., 1995). Internal DNA fragments of about 1,000 bp flanking the *uvrAB* region at its 5' and 3' end were amplified using the primer pairs KG114/KG115 and KG116/KG117. All primer sequences are provided in Tab. S1. Chromosomal DNA of *B. subtilis* 168 served as the template. The 3' end of the upstream fragment as well as the 5' end of the downstream fragments extended into the *uvrAB* gene region, in a way that all expression signals of genes up- and downstream remained intact. The joining of the two fragments to the resistance cassette was performed in a second PCR. In this reaction we used the primer pair KG114/KG117 for the deletion of the *uvrAB* region. The PCR product was directly used to transform *B. subtilis* 168 resulting in the mutant GP1175.

The Δ*mutSl* mutant GP1190 was obtained in a similar way. Briefly, the *aphA3* resistance gene was amplified from pDG780 with the primers mls-fwd (kan)/mls-rev (kan) (Guérot-Fleury *et al.*, 1995), and the regions flanking the *mutSL* genes were amplified using the primer pairs KG127/KG128 and KG129/KG130. The joining PCR was performed with the primer pairs KG127/KG130. The PCR product was used to transform *B. subtilis* 168. The integrity of the regions flanking the integrated resistance cassette was verified by sequencing. PCR products of about 1,000 bp were amplified from chromosomal DNA of the resulting strain GP1190 by using the primer pairs KG131/kan-check rev and KG132/kan-check fwd.

Determination of mutation rates

The *gudB* mutation rates of the *B. subtilis* strains GP1176, GP1191 and GP1192 were determined as previously described in Gunka *et al.*, subm.

DNA manipulation

B. subtilis was transformed with chromosomal DNA or PCR products according to the two-step protocol described previously (Kunst & Rapoport, 1995). Transformants were selected on SP plates containing kanamycin (Kan 10 µg/ml), spectinomycin (Spec 150 µg/ml) or erythromycin plus lincomycin (Erm 2 µg/ml and Lin 25 µg/ml).

Chromosomal DNA of *B. subtilis* was isolated as described (Kunst & Rapoport, 1995). DNA sequences were determined using the dideoxy chain termination method (Sambrook *et al.*, 1989).

Results

Involvement of the UvrAB proteins in the *gudB* mutation

We observed a hundred-fold decrease in the *gudB* decryptification in the *mfd* mutant strain GP1169 (Gunka *et al.*, subm.). This suggests that the nucleotide excision repair NER system might also be involved in the precise excision of the nine base pairs in the *gudB* allele. The Mfd protein does not possess nuclease activity but is required for recruiting the NER machinery to a certain DNA lesion in the process of the transcription coupled nucleotide excision repair (TCR) (Selby & Sancar, 1993; 1994; Deaconescu *et al.*, 2006).

As a potential candidate for the excision system, operating with the Mfd protein in the *gudB* decryptification, the UvrABC system was chosen. In *E. coli,* the UvrA protein is able to interact with transcription repair coupling factors (TCRFs) like Mfd leading to the repair of the damaged DNA (Truglio *et al.*, 2006). Thus, the *uvrAB* operon was deleted in a *rocG* mutant strain (GP1176) and the *gudB* mutation rate was determined (Gunka *et al.*, subm.). The mutation rate of the *uvrAB* mutant strain was similar to those observed with the isogenic *rocG* mutant GP747 (0.1×10^{-4} and 1.3×10^{-4}, respectively).

If UvrABC would be involved in the excision of the nine base pairs in the *gudB* allele, we would expect a more striking effect of an *uvrAB* deletion on the mutation rate. In conclusion, the UvrABC system does not seem to operate with Mfd in the generation of *gudB* suppressors.

Effect of a *mutSL* gene deletion on the *gudB* mutation

In order to gain further insights into the molecular mechanism of the *gudB* decryptification, we looked for other interaction partners of the Mfd protein that could have the capacity to repair DNA damage. For *E. coli* it was shown, that the deletion of the DNA repair mismatch genes *mutS* and *mutL* selectively abolishes DNA repair in the transcribed DNA strand to the same extent as a *mfd* deletion does (Mellon *et al.*, 1996). Therefore, both the *mutS* and the *mutL* gene were deleted and it was tested whether these deletions have an impact on the mutation rate of the *gudB* allele. We determined the mutation rate of the *rocG mutSL* double mutant (GP1191) in comparison to its isogenic *rocG* parent GP747. The mutation rate was 0.2×10^{-4} for the *mutSL* mutant strain GP1191 which is similar to the rate of the *rocG* mutant GP747 (1.3×10^{-4}). Thus,

the deletion of *mutS* and *mutL* has no significant effect on the precise excision of the nine base pairs in the *gudB* allele.

To exclude the possibility that the UvrAB and the MutSL proteins can replace each other in the transcription coupled nucleotide excision repair, we constructed a *B. subtilis* strain lacking the *uvrAB* and the *mutSL* genes (GP1192). The mutation rate of the *gudB* gene in this strain was 0.13 x 10^{-4} which is similar to its isogenic *rocG* mutant strain GP747 (1.3 x 10^{-4}). This result leads to two conclusions: First, it can be assumed that the UvrA and the UvrB proteins cannot be displaced by the proteins MutS and MutL and vice versa in the process of *gudB* decryptification. Secondly, we propose that neither the nucleotide excision repair proteins UvrA and UvrB nor the DNA repair mismatch protein MutS and MutL contribute to the Mfd mediated *gudB* mutation.

Discussion

The data presented herein, imply that in the case of the *gudB1* mutation the Mfd protein does not act in concert with the UvrABC or the MutSL systems. Even the simultaneous inactivation of both systems does not lead to a significant decrease in the *gudB1* mutation rate.

Mfd possess structural characteristics which are highly conserved (Deaconescu *et al.*, 2006). Allowing six gaps, the amino acids 15 to 473 of the Mfd protein of *B. subtilis* share a 19% identity with the N-terminal region of UvrB protein of *B. subtilis* see (Fig. 5.1). This region in the UvrB protein is supposed to contain residues necessary for the interaction with the UvrA protein. Therefore, the N-terminal part of the Mfd protein seems to be necessary for the recruitment of the NER system by interacting with UvrA (Selby & Sancar, 1993). As the very last C-terminal part of UvrB was shown to interact with the UvrC protein and this part is not homologous to any part of the Mfd protein, it is unlikely that Mfd itself can interact with UvrC. It was demonstrated that the Mfd protein interacts with the RNA polymerase (see Fig. 5.1) and is therefore sufficient to recognize a stalled transcription complex (Ayora *et al.*, 1996). The C-terminal domain of Mfd is supposed to block the UvrA interaction site until a stalled RNA polymerase is recognized by the RNA polymerase interaction domain and displaced by the translocation domain (Deaconescu *et al.*, 2006).
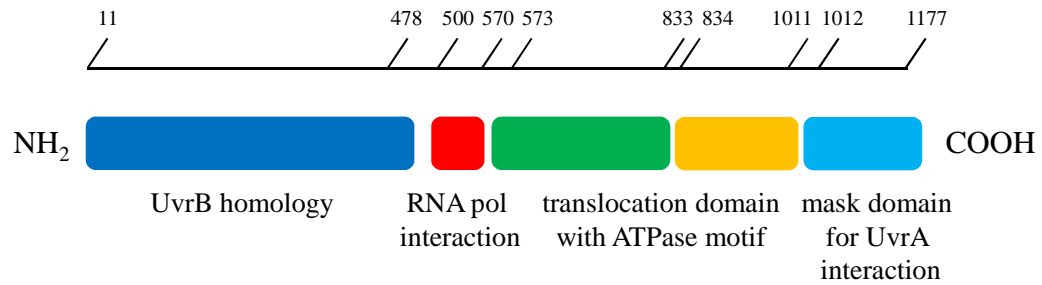
Fig. 5.1 – Domain structure of the *B. subtilis* Mfd protein.

Characteristic and conserved domains of TCRFs are mapped on the primary structure of the *B. subtilis* Mfd protein (Illustration adapted to Deaconescu *et al*., 2006). Each domain is highlighted in a different color. The domains were determined by using a multiple alignment of Mfd homologues to (Deaconescu *et al*., 2006) (RNA pol = RNA polymerase).

It is obvious that the Mfd protein can neither recruit UvrC nor can it perform nucleotide excision itself. Therefore, it is tempting to speculate that the Mfd mediated generation of the *gudB1* mutation occurs in yet unknown mechanism and involves proteins that have not been linked with the transcription coupled nucleotide excision repair until now.

Future perspectives

In order to gain insights into the Mfd-mediated generation of *gudB1* mutations a SPINE experiment using a *rocG* mutant strain can be performed (Herzberg *et al*., 2007). By this *in vivo* cross linking approach potential interaction partners of the Mfd protein can be identified. To increase the chance of cross-linking Mfd in complex with its potential target, the *gudB* gene, and additional interaction partners, the *gudB* gene has to be enriched. Besides the analysis of the elution fraction by mass spectrometry to identify Mfd interaction partners, a Northern blot with a *gudB* probe can be performed. This experiment might demonstrate whether the *gudB* transcript can be specifically cross-linked with the Mfd protein.

6.   Chapter 6

SPABBATS: A pathway-discovery method based on Boolean satisfiability that facilitates the characterization of suppressor mutants

The work described in this chapter was submitted for publication in:

Flórez, L. A., Gunka, K., Polanía, R., Tholen, S. & Stülke, J. SPABBATS: A pathway-discovery method based on Boolean satisfiability that facilitates the characterization of suppressor mutants. BMC Systems Biology submitted

Author contributions:

LAF, KG, and JS planned the project. LAF and RP conceived and implemented the SPABBATS algorithm. KG and JS developed the strategies for the experimental validation of the method. ST contributed to the performance of the experiments. LAF, KG, and JS wrote the manuscript and created the Table and Figures. All authors read and approved the final manuscript.

Abstract

Background

Several computational methods exist to suggest rational genetic interventions that improve the productivity of industrial strains. Nonetheless, these methods are less effective to predict possible genetic responses of the strain after the intervention. This problem requires a better understanding of potential alternative metabolic and regulatory pathways able to counteract the targeted intervention.

Results

Here we present SPABBATS, an algorithm based on Boolean satisfiability (SAT) that computes alternative metabolic pathways between input and output species in a reconstructed network. The pathways can be constructed iteratively in order of increasing complexity. SPABBATS allows the accumulation of intermediates in the pathways, which permits discovering pathways missed by most traditional pathway analysis methods. In addition, we provide a proof of concept experiment for the validity of the algorithm. We deleted the genes for the glutamate dehydrogenases of the Gram-positive bacterium *Bacillus subtilis* and isolated suppressor mutant strains able to grow on glutamate as single carbon source. Our SAT approach proposed candidate alternative pathways which were decisive to pinpoint the exact mutation of the suppressor strain.

Conclusion

SPABBATS is the first application of SAT techniques to metabolic problems. It is particularly useful for the characterization of metabolic suppressor mutants and can be used in a synthetic biology setting to design new pathways with specific input-output requirements.

Background

A holistic understanding of cellular metabolism is central to systems biology and metabolic engineering: In order to amplify the flux through production pathways in industrial strains we have to understand how the metabolic network responds to our interventions.

Several methods can suggest rational interventions that may lead to favorable industrial phenotypes (Feist & Palsson, 2008). Their goal is to optimize the distribution of metabolic fluxes towards the product of interest, either directly (e.g. FBA, MOMA or ROOM) or indirectly by coupling it to another characteristic (e.g. OptKnock) that facilitates further strain improvements via mutation and screening.

While these methods can predict a final flux distribution, they do not predict the range of genetic and metabolic responses of the organism after the targeted mutation. At the same time, it would be highly desirable to have tools that may predict these responses, since they can suggest ways to generate more stable strains, or accelerate the adaptation to an intended optimal flux. The challenge of the question is the need to understand why particular genetic responses make sense in an evolutionary setting. Thus, the ultimate question is: Which parallel pathways - that were not active previously - result in an adaptive advantage under the screening conditions?

Pathway analysis has received increased attention due to the reconstruction of genome scale metabolic networks for many organisms. These methods can be divided into two categories: stoichiometric and path oriented (Planes & Beasley, 2008). The first approach generates all pathways that conform to the pseudo-steady-state assumption for internal metabolites. However, it presents two problems: the number of predicted pathways is in the order of millions for genome scale models, making the approach totally intractable for the question at hand (Klamt & Stelling, 2002). Its second shortcoming is the constraint imposed by the pseudo-steady-state assumption for internal metabolites. This assumption may rule out feasible pathways or (in case we include a large number of "freely available" metabolites) result again in a combinatorial explosion of pathways. The alternative approach - path oriented pathway reconstructions - is advantageous since it usually generates a small (and thus tractable) set of possible pathways. This is due to the choice of starting and ending metabolites and heuristics on the characteristics of the "optimal" pathway. However, the path-

oriented approach may result in unrealistic pathways that consume internal metabolites not present in sufficient quantities inside the cell.

What is needed is an algorithm that reconstructs stoichiometrically balanced pathways in increasing order of complexity, with relaxed mass-balance constraints in comparison to the traditional pseudo-steady-state restriction.

A solution based on mixed-integer linear programming (MILP) has been suggested by de Figueiredo *et al.*, but it has not been used in an evolutionary context so far (de Figueiredo *et al.*, 2009). Here we describe the use of Boolean satisfiability (SAT) for the reconstruction of alternative pathways in metabolic networks (Claessen *et al.*, 2009). Given a set of basal metabolites (that are considered freely available) and a set of target metabolites (whose concentration *must* increase), our SAT method constructs the shortest pathway between the basal and target sets (SPABBATS) of metabolites that is stoichiometrically balanced, while allowing the concentration of the intermediate metabolites to increase, if needed. The constraints are more relaxed than the ones for e.g. flux balance analysis, thus retaining the metabolically significant pathways. Using the algorithm iteratively, we obtain a prioritized list of pathways, whose elements can be tested individually by common molecular biology techniques.

To demonstrate the power of this concept, we applied the SPABBATS algorithm to a complex physiological problem, which is a result of an evolutionary experiment. We have elucidated a novel pathway of glutamate degradation present in the metabolic network of *B. subtilis* that had been decryptified upon inactivation of the normal glutamate catabolic genes. By using our SAT approach, we proposed four different new pathways that could be present in the mutant to utilize glutamate as single carbon source. These predictions were experimentally tested and revealed that one of these pathways was indeed active in the mutant strain and that this novel "suppressor" pathway is required and sufficient for glutamate utilization. This proves that the results of our approach correspond to valid metabolic alternatives for living cells.

Materials and Methods

Algorithm for finding short pathways between a basis and a target set of metabolites (SPABBATS)

Our approach draws inspiration from flux-balance analysis (FBA) (Orth *et al.*, 2010) in the sense that it searches the flux space of a metabolic network for fluxes that comply with a set of stoichiometric constraints. The major difference to FBA lies in the optimality criterion; in FBA the value to optimize is the target flux. In our case we change from optimization to satisfiability: we search for a flux that satisfies all the constraints, including a maximum number of allowed reactions.

Another important difference, that is a consequence of satisfiability approach, is that we use two variables for each flux instead of one. The first variable is a positive integer, which is a relative measure of the contribution of that particular flux to the total pathway. The second variable is Boolean and defines whether or not the particular flux takes part in the solution.

As in FBA, we define $\mathbf{S}$ as the stoichiometric matrix of the network with n reactions and m compounds. Reversible reactions are split into two unidirectional reactions. We divide the set of compounds into three disjoint sets:

i) $\mathbf{B}$ is the set of basis compounds that are considered freely available, either because they are provided in the medium, or because they are "currency metabolites", whose concentration is buffered by the whole system (e.g. ATP, ADP, NADH, etc.)

ii) $\mathbf{T}$ is the set of target compounds, the ones constrained to be produced in the pathways of interest

iii) $\mathbf{I}$ is the set containing all other compounds, that can be intermediates of the resulting pathway

We use different constraints for each of these sets. The compounds in the set $\mathbf{B}$ are left unconstrained. For each compound in the set $\mathbf{T}$, we write a constraint in the form:

$$\sum_{i=1}^{n} s_{ij} a_i b_i > 0, \qquad (1)$$

where s_{ij} is the stoichiometric coefficient of compound j in reaction i, and a_i and b_i are the integer and Boolean valued variables of reaction i, respectively. These

constraints mean that in the solution pathway the overall flux to these metabolites should be positive.

For the compounds in the set **I** we use a constraint similar to (1), with the difference that we use a "greater than or equal to" ($\geq$) sign. In FBA, an equality sign is used here, to constraint the fluxes to the steady-state space. We purposely do not constrain the pathway to the steady-state space, since the candidate solutions to the problem will not be the only pathway active in the cell and the intermediates that are accumulated in our pathway can be used by other pathways operating in parallel in the system. We require the total flux to these compounds to be non-negative, since the supposition is that they are not present in sufficiently high amounts to allow sustained growth on their consumption.

Next, we add constraints that limit the directionality of reversible reactions. This is done with constraints in the form:

$$b_i + b_j < 2, \qquad (2)$$

where b_i and b_j are the Boolean variables of two reactions that together characterize a reversible reaction. These constraints mean that no two directions of a reversible reaction can appear in the final pathway at the same time.

Last, we add a constraint for the total length of the solution. This constraint is:

$$\sum_{i=1}^{n} b_i \leq k, \qquad (3)$$

where k is a positive integer value that determines the maximum number of reactions that can appear in the pathway. This constraint does not immediately find the best solution, but it puts successively stricter upper-bounds to the maximum number of reactions that are allowed. Thus, it is able to find the shortest solution after some iterations by choosing successively smaller numbers for k.

The constraints for the compounds in **T** and **I** are not linear, since each term in the sum is composed of two variables instead of one. For this reason, a linear optimization strategy cannot be used directly. This limitation is not present when we use the SAT-solver HySAT (Fränzle *et al*., 2007). It is able to find assignments to the variables that satisfy all the constraints in the system, even when these are non-linear. It is also able to detect if no such assignment exists.

If the shortest solution has been found, the best sub-optimal solution can be found by adding an additional constraint in the form:

$$\sum_{i \in K} b_i < k_{op},\qquad (4)$$

where k_{op} is the number of reactions in the shortest solution and K is the set of indices for the reactions in the shortest solution. In other words, we constrain the sum of all the Boolean variables of the optimal solution to be less than k_{op}, thus leaving out the shortest solution from the solution space. By iterating this process with the Boolean variables of the sub-optimal pathway, we can find solutions with successively higher number of reactions.

The particular implementation of this algorithm for the problem mentioned in the Results section is as follows: we used the genome-scale reconstruction of *B. subtilis* (Oh *et al.*, 2007). We removed the biomass "reaction"; it is useful for FBA, since it describes the target flux to cellular growth, but is meaningless in our context. In addition, we removed the reaction "glutamate dehydrogenase" (R_GLUDxi) to simulate the conditions of the strain GP717. We also scaled the non-integer stoichiometric coefficients of the model to integer values (and divided by the greatest common denominator). In our case, the set **B** contained the metabolites ATP, ADP, NAD^+, NADH, FAD, $FADH_2$, H_2O, H^+, NH_4^+, and glutamate. These "currency metabolites" were chosen due to their participation in most catabolic pathways in the cell. The set **T** contained just 2-oxoglutarate. The remaining compounds were assigned to the set **I**. We set the interval for the a_i to [1, 1000]. The calculations were done using an Intel Core2 Duo processor at 2.66GHz, with 3.25GB of RAM. The first pathway (the one involving leucine as intermediate) was found after 28 seconds. All other pathways took less than 8 minutes each to calculate.

Bacterial strains and growth conditions

All *B. subtilis* strains used in this work are derived from the laboratory wild type strain 168. They are listed in Tab. S3. *E. coli* DH5α (Sambrook *et al.*, 1989) was used for cloning experiments. *B. subtilis* was grown in C minimal medium containing ammonium as basic source of nitrogen (Wacker *et al.*, 2003). Glutamate and/ or glucose were added as carbon source as indicated. The medium was supplemented with

auxotrophic requirements (at 50 mg/l). *E. coli* was grown in LB medium and transformants were selected on plates containing ampicillin (100 µg/ml). LB, SP and CSE plates were prepared by the addition of 17 g Bacto agar/l (Difco, Lawrence, USA) to LB, SP or CSE medium, respectively.

DNA manipulation and transformation

Transformation of *E. coli* and plasmid DNA extraction were performed using standard procedures (Sambrook *et al*., 1989). Restriction enzymes, T4 DNA ligase and DNA polymerases were used as recommended by the manufacturers. DNA fragments were purified from agarose gels using the Nucleospin Extract kit (Macherey and Nagel, Düren, Germany). Phusion™ DNA polymerase was used for the polymerase chain reaction as recommended by the manufacturer (New England Biolabs, Frankfurt am Main, Germany). All primer sequences are provided as supplementary material (Tab. S2). DNA sequences were determined using the dideoxy chain termination method (Sambrook *et al*., 1989). All plasmid inserts derived from PCR products were verified by DNA sequencing. Chromosomal DNA of *B. subtilis* was isolated as described (Kunst & Rapoport, 1995).

E. coli transformants were selected on LB plates containing ampicillin (100 µg/ml). *B. subtilis* was transformed with plasmid DNA or PCR products according to the two-step protocol described previously (Kunst & Rapoport, 1995). Transformants were selected on SP plates containing tetracycline (Tet 10 µg/ml), or erythromycin plus lincomycin (Erm 2 µg/ml and Lin 25 µg/ml).

Plasmid and mutant strain construction

To express a plasmid-borne *ansR* gene in *B. subtilis*, we constructed plasmid pGP873. For this purpose the *ansR* gene was amplified with the primers KG18 and KG19 using chromosomal DNA of *B. subtilis* as a template (all primer sequences are provided in Tab. S1). The PCR product was digested with BamHI and SalI and cloned into the overexpression vector pBQ200 (Martin-Verstraete *et al*., 1994).

Deletion of the *ansAB* and *ansR* genes was achieved by transformation with PCR products constructed using oligonucleotides to amplify DNA fragments flanking the target genes and an intervening erythromycin and tetracycline resistance cassettes from

plasmids pDG647 and pDG1514, respectively (Guérout-Fleury *et al.*, 1995), as described previously (Wach, 1996). The PCR products were used to transform GP717 and GP28 for the deletion of the *ansAB* and *ansR*, respectively.

Reverse transcription-real-time quantitative PCR

For RNA isolation, the cells were grown to an OD_{600} of $0.5 - 0.8$ and harvested. Preparation of total RNA was carried out as described previously (Ludwig *et al.*, 2001). cDNAs were synthesized using the One-Step RT-PCR kit (BioRad, Munich, Germany) as described (Rietkötter *et al.*, 2008). Real time quantitative PCR was carried out on the iCycler instrument (BioRad, Munich, Germany) following the manufacturer's recommended protocol by using the primers KG26/KG27 for the *ansA* gene, KG38/KG39 for the *ald* gene and KG40/KG41 for the *bcd* gene, respectively. Their recommended data analysis procedure was also used. The *rpsE* and *rpsJ* genes encoding constitutively expressed ribosomal proteins were used as internal controls and were amplified with the primers *rpsE*-RT-*fwd/ rpsE*-RT-*rev* and *rpsJ*-RT-*fwd/ rpsJ*-RT-*rev,* respectively. The expression ratios were calculated as fold changes as described (Rietkötter *et al.*, 2008). RT-PCR experiments were performed in duplicate.

Results

Isolation of a mutation that allows a bypass of the glutamate dehydrogenase for the utilization of glutamate

Glutamate is the most abundant metabolite in a bacterial cell. Although its exact concentration in *B. subtilis* is unknown, it is known to account for about 40% of the internal metabolite pool of an *Escherichia coli* cell (Yuan *et al*., 2009). Glutamate serves as an osmotic regulator (Whatmore *et al*., 1990), as well as universal amino group donor in anabolism thus linking carbon and nitrogen metabolism (Commichau *et al*., 2006). In *B. subtilis*, at least 37 reactions make use of glutamate as cofactor for transamination (Oh *et al*., 2007).

The key reactions of glutamate biosynthesis and degradation in *B. subtilis* are summarized in Fig. 6.1 2-oxoglutarate, an intermediate of the citric acid cycle, is aminated by the glutamate synthase, encoded by the *gltA* and *gltB* genes. Glutamate degradation to 2-oxoglutarate requires the glutamate dehydrogenase RocG. Additionally, the laboratory strain *B. subtilis* 168 harbors a cryptic gene, *gudB*, coding for an inactive glutamate dehydrogenase. This gene is readily decryptified in *rocG* mutants (Belitsky & Sonenshein, 1998; Commichau *et al*., 2008). In addition, RocG controls the expression of the *gltAB* operon and therefore prevents glutamate biosynthesis in the presence of arginine (Belitsky & Sonenshein, 2004; Commichau *et al*., 2007b).
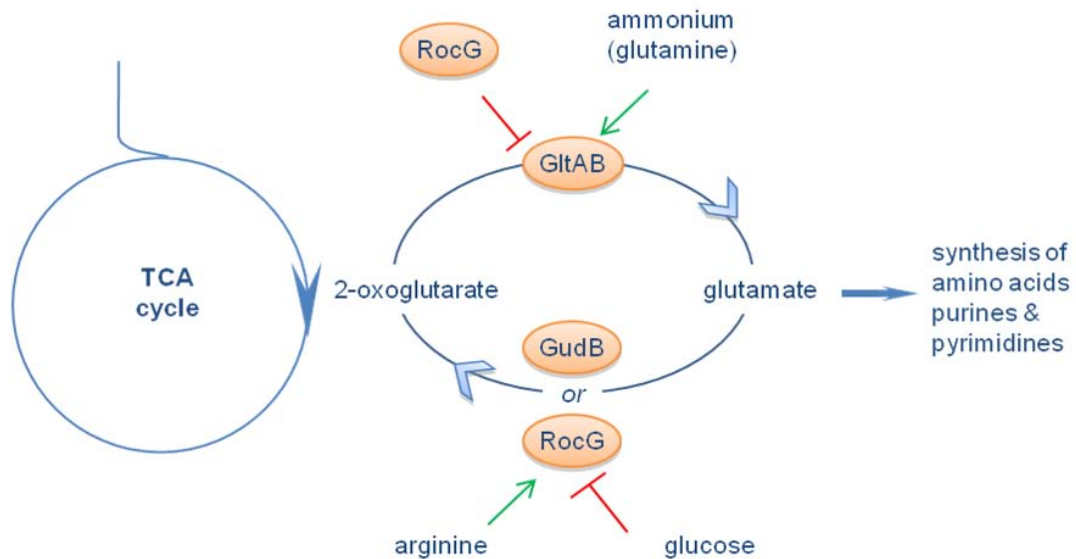
Fig. 6.1 – Key reactions for glutamate biosynthesis and degradation in *B. subtilis*.
Glutamate is the universal amino group donor in all living cells and in that way links the carbon and nitrogen metabolisms. In *B. subtilis* the synthesis of glutamate depends on the glutamate synthase GltAB. In addition, the genome encodes two glutamate dehydrogenases, RocG and GudB, although the latter is inactive in the laboratory *B. subtilis* strain 168 (see text). The synthesis and degradation of glutamate are tightly regulated in response to the availability of carbon and nitrogen sources.

Inactivation of both the *rocG* and the *gudB* gene results in loss of any glutamate dehydrogenase activity and concomitant inability of the bacteria to utilize glutamate (Belitsky & Sonenshein, 1998; Commichau *et al*., 2008). The *rocG gudB* double mutant strain GP28 grows poorly on SP medium (an amino acid-rich medium) due to the accumulation of degradation products of arginine metabolism (Commichau *et al*., 2007a). However, cultivation of GP28 on SP plates eventually resulted in the isolation of a mutant (GP717) that carries a mutation inactivating the *gltB* gene, encoding a subunit of the glutamate synthase (Commichau *et al*., 2008). This *gltB1* mutation leads to glutamate auxotrophy and might therefore prevent the accumulation of intermediates of arginine degradation. We have observed that toxic intermediates of arginine degradation result in poor growth of mutants lacking a functional glutamate dehydrogenase (our unpublished results). If intrinsic glutamate synthesis is blocked by a mutation, such an accumulation of toxic intermediates might be reduced. A careful analysis of the mutant strain revealed that it had acquired the ability to utilize glutamate

as the only source of carbon and energy. This might have resulted from a re-activation of the *rocG* or *gudB* genes or from the establishment of a novel pathway for glutamate utilization. We tested therefore the *rocG* and *gudB* alleles by PCR analysis. Both the transposon insertion in *rocG* and the replacement of the *gudB* gene by a chloramphenicol resistance gene were identical to the parent strain GP28. Clearly, a new pathway of glutamate degradation was activated in this suppressor mutant that was not active in the wild type and *rocG gudB* mutant cells.

Development of a pathway-finding algorithm

The most reasonable hypothesis to explain the suppression was that the mutation had activated a redundant pathway that is inactive in the wild type strain in a medium with glutamate as single carbon source. Since glutamate is a highly abundant metabolite and is involved as a substrate in 20 reactions in *B. subtilis*, it was not obvious which mutation could have lead to glutamate utilization proficiency in *B. subtilis* GP717.

To address this problem by use of the power of bioinformatics, we developed an approach that harnesses the strengths of Boolean satisfiability (SAT) to find valid pathways (see Materials and Methods). It is able to find short pathways between a basis and a target set (SPABBATS) of metabolites that can operate in a sustained way. It is convenient for its focus on short pathways and the fact that it can calculate pathways that comply with the steady-state constraint. It also allows the relaxation of this constraint, by allowing some metabolites to accumulate if necessary.
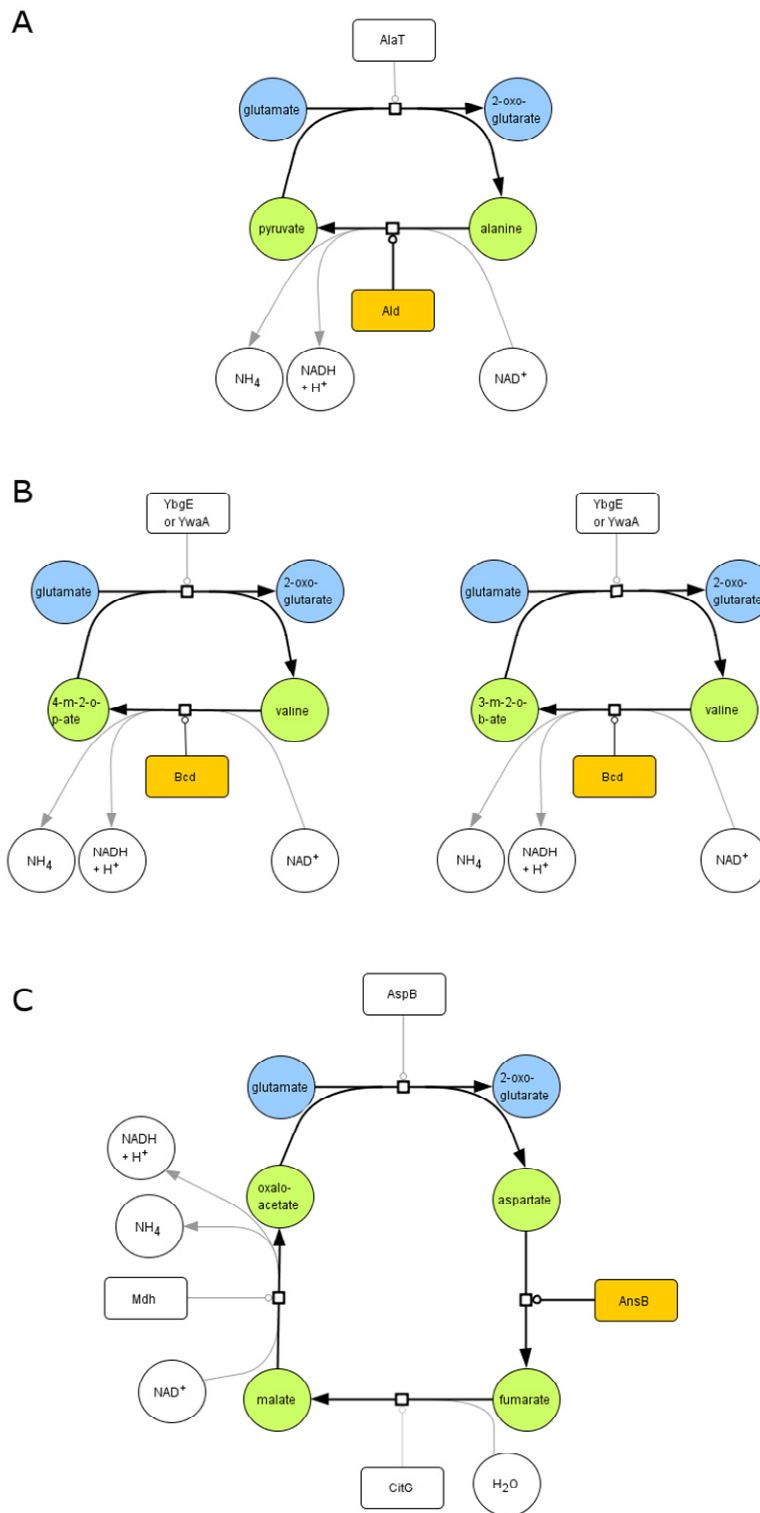
Figure 6.2 - Predictions of alternative pathways for glutamate utilization based on SAT techniques.

Fig. legend see next page.

Figure 6.2 - Predictions of alternative pathways for glutamate utilization based on SAT techniques.

A *B. subtilis* strain (GP28) was constructed that lacks the glutamate dehydrogenases. An evolutionary adaptation resulted in a strain (GP717) that acquired the capacity to grow on glutamate as single carbon source. Using a SAT based search algorithm (see Materials and Methods) we predicted four alternative pathways that could be activated in the GP717. The genes coding for the enzymes in orange were analyzed further (see text). 4-m-2-o-p-ate = 4-methyl-2-oxo-pentanoate; 3-m-2-o-b-ate = 3-methyl-2-oxo-butanoate.

The first four pathways suggested by our algorithm are presented in Fig. 6.2 In each case, the first step is a transamination reaction that leads to the production of 2-oxoglutarate. The substrate for transamination is then replenished via the remaining reaction(s) of the pathway. The first pathway (Fig. 6.2A) involves transamination to form alanine and subsequent oxidative deamination of alanine by the alanine dehydrogenase Ald resulting in the net formation of 2-oxoglutarate. The next two pathways (Fig. 6.2B) are very similar and involve enzymes of branched amino acid metabolism. In the transamination step, both pathways use the transaminases YbgE and YwaA. The branched chain amino acid dehydrogenase Bcd is then used for the oxidative deamination of the transamination products valine or leucine. Again, the net result of this pathway is the production of 2-oxoglutarate from glutamate. The last pathway (Fig. 6.2C) requires four steps, (i) the reaction of the aspartate amino-transferase AspB, (ii) the deamination of asparate to fumarate by the aspartase AnsB, (iii) the fumarase reaction (CitG) of the citric acid cycle, and finally (iv) the oxidation of malate by the malate dehydrogenase Mdh. As described for the other pathways, this reaction sequence results in the net formation of 2-oxoglutarate from glutamate. Since the original mutant GP28 did not grow with glutamate as the single carbon source, it is obviously not able to use any of these proposed pathways suggesting that they were activated by a suppressor mutation in GP717.

Experimental validation of the predictions

Our experiments were performed in minimal medium suggesting that the activity of transaminases was not limiting. Similarly, the two enzymes of the citric acid cycle (CitG and Mdh) are constitutively expressed (Jin & Sonenshein, 1994; Feavers *et al*., 1998; Blencke *et al*., 2003). Thus, the mutation may have affected the expression of one

of the deaminases Ald, Bcd or AnsB. This hypothesis was tested by reverse transcription-real-time quantitative PCR. As shown in Fig. 6.3, the levels of *ald* and *bcd* mRNA are comparable for the original mutant GP28 and the suppressor strain GP717. In contrast, a strong increase of the expression of the *ansAB* operon encoding the asparaginase and aspartase was observed for the suppressor mutant that was able to utilize glutamate. This observation suggests that it is the high-level expression of AnsB that allows glutamate utilization in GP717.
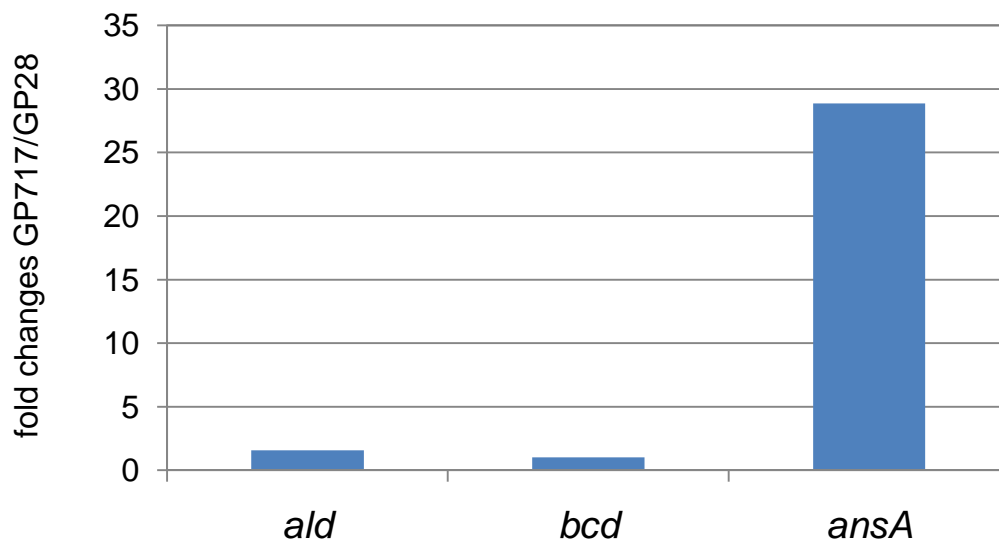
Fig. 6.3 – Comparison of gene expression patterns between mutant and parental strains, based on the predictions of the SPABBATS algorithm for pathway analysis.

The predictions of the SPABBATS algorithm (see Fig. 6.2) were further characterized by transcription analysis. The expression of the *ald* and *bcd* genes remains constant between the mutant (GP717) and parental (GP28) strains, suggesting that these genes are not involved in the newly activated catabolic pathway. In contrast, the expression of the *ansAB* operon is strongly increased in the mutant. This hints to a gain of function in the mutant strain that was analyzed further.
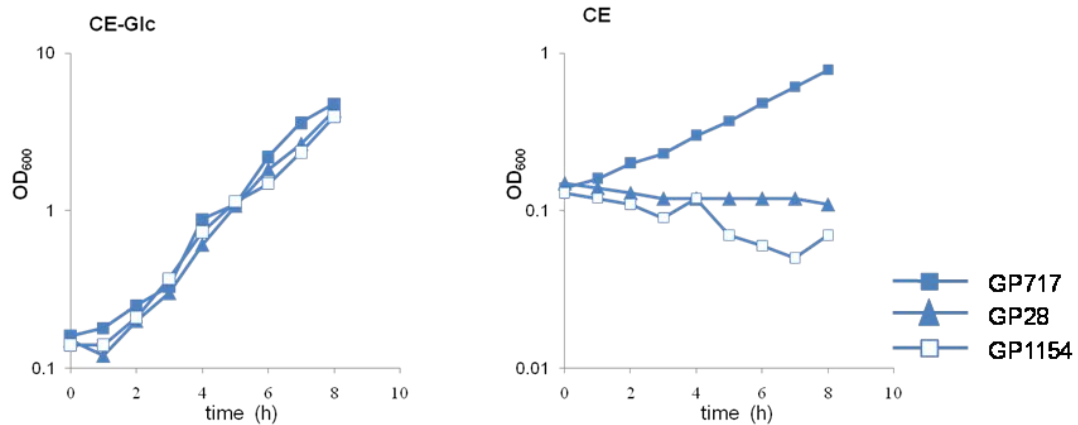
Fig. 6.4 - Requirement of the aspartase gene in the alternative pathway for glutamate utilization.

The SPABBATS algorithm (see Fig. 6.2) and the transcription analysis (see Fig. 6.3) suggested that the overexpression of the asparaginase and aspartase genes (*ansAB*) is the cause for the metabolic gain of function of the mutant strain GP717. To prove this, the *ansAB* operon was deleted in the GP717 strain. The resulting strain GP1154 lost the capacity to utilize glutamate as single carbon source. This strongly indicates that the induction of the aspartase gene is required and sufficient for the newly activated catabolic pathway. CE = Minimal medium containing 8 g/l glutamate, CE-Glc = CE medium with an addition 5 g/l glucose.

The involvement of the aspartase AnsB in the novel glutamate utilization pathway was verified by analyzing the effect of a deletion of the *ansAB* operon. Growth of the original strain GP28, the suppressor mutant GP717 and its isogenic Δ*ansAB* mutant derivative GP1154 in minimal medium with glutamate or with glutamate and glucose was recorded. As shown in Fig. 6.4, all three strains were able to grow with glutamate and glucose. In contrast, the deletion of the *ansAB* operon reverted the capability of the suppressor strain of using glutamate as the single carbon source, and the Δ*ansAB* mutant GP1154 was unable to grow with glutamate as was the original strain GP28. This finding strongly supports the idea that the activity of the aspartase AnsB is the reason for the ability of the suppressor strain GP717 to utilize glutamate.

The *ansAB* operon is induced in the presence of asparagine due to inactivation of the AnsR repressor (Sun & Setlow, 1991; Sun & Setlow, 1993; Fisher & Wray, 2002). A comparative analysis of *ansAB* expression revealed about 30-fold induction by asparagine in GP28, whereas the expression levels were unaffected by the availability of asparagine in the suppressor mutant GP717 (data not shown). The observed induction in

the wild type strain is in good agreement with previous reports. The loss of regulation in GP717 and the high expression of the operon as compared to GP28 suggest constitutive *ansAB* expression that might be the result of an inactivation of the *ansR* repressor gene.

To test the hypothesis that inactivation of the AnsR repressor allowed glutamate utilization by GP717, we performed two tests: First, we deleted the *ansR* gene of the parental strain GP28 and tested the ability of the resulting strain GP811 to grow with glutamate as the single carbon source. Unlike GP28, this strain GP811 (Δ*ansR*) grew in CE minimal medium. Thus, inactivation of the *ansR* gene is sufficient to open a new pathway for glutamate catabolism. In a complementary approach, we complemented *B. subtilis* GP717 with a plasmid-borne copy of the *ansR* gene (present on pGP873) and tested the ability of the transformants to use glutamate. While the control strain (GP717 transformed with the empty vector pBQ200) grew well on CE medium, expression of AnsR from the plasmid completely blocked growth in this medium, i.e. the utilization of glutamate. This result confirms that a mutation in the *ansR* gene must be present in GP717 and that it is this mutation, which confers the bacteria with the ability to utilize glutamate via the new aspartase pathway.

To identify the mutation in *ansR*, we sequenced the *ansR* alleles of the parental strain GP28 and the glutamate-utilizing suppressor mutant GP717. While the wild type allele of *ansR* was present in GP28, a C-to-A substitution at position 107 of the *ansR* open reading frame was found in GP717. This mutation changes codon 36 from UCA (Ser) to UAA (stop) and results in premature translation termination and the formation of an incomplete and non-functional AnsR repressor protein.

Taken together, these experiments confirmed that the metabolic pathway predicted by the SPABBATS algorithm corresponds to a valid metabolic state of the *rocG gudB ansR* mutant strain GP717.

Discussion

Comparison of SPABBATS with other methods for metabolic analysis

Flux balance analysis and the majority of methods derived from it are based on constraining the admissible intracellular flux space to steady-state and choosing an adequate optimality criterion to calculate intracellular fluxes (Orth *et al*., 2010). Commonly used optimization criteria are biomass production and the maximization of energy output.

Although these methods predict the essentiality of genes with high accuracy (Oh *et al*., 2007), they are less suited for the characterization of alternative metabolic pathways in viable mutants. On the one hand, by restricting the admissible intracellular flux to steady-state, they discard pathways where a by-product accumulates. Nonetheless, the cell is still viable if this by-product is consumed by other pathways in the cell, not directly related to the process that is studied. SPABBATS solves this problem by allowing a larger flux-space, where intermediate products can accumulate, if necessary.

On the other hand, the optimality criterion can be artificial. For instance, maximizing cellular growth might lead to a theoretical maximum growth rate, or a flux distribution that is as close to the wild-type flux as possible, but it is hard to argue that the regulatory network of the strain is directed to the same target. The pathways discovered by SPABBATS are a structural property of the network and do not depend on an extrinsic optimality criterion (beyond the number of reactions of the resulting pathway). For this reason, the resulting pathways can be interpreted objectively.

Other methods for structural decomposition (e.g. extreme pathways and elementary flux modes, [Planes & Beasley, 2008]) rely on the same steady-state restriction of FBA related methods and for this reason share some of their disadvantages. Moreover, SPABBATS does not require the calculation of all possible pathways. Instead, it can be used iteratively to calculate pathways of increasing length, which results in a dramatic improvement in performance for finding relevant pathways in large networks.

An advantage over the method of de Figueiredo *et al*. (de Figueiredo *et al*.) is that we do not make use of an optimization framework, but select for satisfiability instead. Similar problems in other areas of computational biology (e.g. Graça *et al*., 2007) show

a performance improvement of SAT methods over traditional mixed-integer linear programming methods.

Future perspectives

So far, our analysis of networks using SAT has been restricted to metabolic networks. Nonetheless, since SAT is especially suited for problems that involve Boolean constraints, it is possible to expand the analysis to regulatory networks. For *B. subtilis*, this implies the reconstruction of the metabolic network together with its regulatory complement. This reconstruction is in progress (Goelzer *et al.*, 2008; Lammers *et al.*, 2010).

In parallel, we envision the development of novel SAT solvers that are optimized for the solution of metabolic constraints. This will result in the adoption of SAT based methods for metabolic engineering as well as for the design of synthetic circuits that are able to perform computations in the same way as their silicon-made counterparts (Lou *et al.*, 2010).

Conclusions

In this contribution we have shown the use of SAT techniques to discover alternative pathways that connect sets of starting and target species. In addition, we provided a proof of concept for the applicability of the algorithm. We started with a complex physiological problem in *B. subtilis*: the need to characterize a suppressor mutation that allowed growth on glutamate without glutamate dehydrogenases. SPABBATS predicted four potential pathways for glutamate utilization that were decisive to suggest target genes for experimentation. These experiments confirmed the validity of the SPABBATS' prediction, closing the cycle between modelling and wet lab experimentation.

SPABBATS relies on Boolean satisfiability (SAT) to construct the metabolic pathways. SAT has been used for the determination of haplotypes from sequenced genotypes (Graça *et al.*, 2007), the analysis of genome biology networks (Chin *et al.*, 2008), the understanding of myogenic differentiation (Piran *et al.*, 2009), and the characterization of steady states of regulatory circuits (Tiwari *et al.*, 2007; de Jong &

Page, 2008). Here we report the first application of SAT techniques to metabolic problems.

The SPABBATS algorithm was applied here to a specific problem, the analysis of glutamate metabolism in *B. subtilis*. However, the solution strategies are applicable to a broad spectrum of metabolic problems. For instance, SPABBATS can be particularly useful in the characterization of suppressor mutants. Moreover, SPABBATS can also be useful in synthetic biology. Although used here to find pathways in a reconstruction of the metabolism of *B. subtilis*, it is also possible to use a database of enzymes as the starting model. In this way, it can be used to construct synthetic pathways that satisfy specific input-output and mass-balance requirements.

Acknowledgements

7.   **Discussion**

7.1.   **The emergence of suppressor mutations in response to glutamate imbalance**

In the absence of the active GDH RocG in *B. subtilis*, the cryptic GDH GudB1 is rapidly decryptified upon growth on complex medium. The rate of this *gudB1* mutation is 10^{-4} (Gunka *et al*., subm.). Compared to rates of beneficial mutations in bacteria that are in a range of 10^{-9} the *gudB* decryptification occurs with an extremely high rate (Imhof & Schlötterer, 2001). The gain of function mutation in the *gudB* gene allows a *rocG* mutant to degrade glutamate and to overcome the severe growth defect of this mutant (see Fig. 1.3).

Previous results revealed that a *rocG gudB* double mutant can also accumulate suppressor mutations that facilitate good growth of this mutant when incubated on complex medium (Commichau *et al*., 2008). One mutation was found to inactivate the GOGAT leading to the loss of glutamate synthesis in this mutant. Furthermore, this strain had acquired a second suppressor mutation that allows the utilization of glutamate as carbon source even in the absence of a glutamate dehydrogenase (Flórez *et al*., subm.). This mutation is located in the gene for the repressor AnsR of the aspartase/asparaginase operon *ansAB* (Sun & Setlow, 1991). A nonesense mutation inactivated the repressor AnsR resulting in a constitutive expression of the *ansAB* operon. This alternative pathway was discovered by using a Boolean satisfiability (SAT) method that constructs the shortest pathway between the basal and target sets (SPABBATS) of metabolites that is stoichiometrically balanced (Flórez *et al*., subm.). Both mutations, the inactivation of the GOGAT and the activation of an alternative pathway for the glutamate degradation, probably result in a decrease of the glutamate level of the cell. It was also reported for Enterobacteria that the perturbation of glutamate homeostasis results in the accumulation of suppressor mutations that balance the glutamine to glutamate ratio to wild type level (Yan, 2007). Taken together with the *gudB1* suppressor mutation, it is tempting to speculate that the accumulation of glutamate might be highly problematic for *B. subtilis*, although glutamate is the most abundant molecule in the cell anyway. The complex medium contains beef extract and is therefore composed of a variety of nitrogen containing compounds such as arginine

and glutamate but the medium lacks sugars. A *rocG* mutant is not capable of converting glutamate to 2-oxoglutarate. Perhaps the lack of glutamate degradation is disadvantageous upon cultivation on complex medium and might lead to a decrease in TCA intermediates. By the activation of GudB this bottleneck might be resolved. Another possibility that could explain the growth defect of a *rocG* mutant is that toxic metabolites accumulate in the cell in the absence of an active GDH.

The enzyme glutamate racemase (encoded by the essential gene *racE* in *B. subtilis*) catalyzes the conversion of L-glutamate to D-glutamate that is a building block for peptidoglycan biosynthesis (Kimura *et al*., 2004; Spies *et al*., 2009). Indeed, the accumulation D-glutamate was shown to be toxic for *B. subtilis* (Kimura *et al*., 2004). In the presence of very high intracellular amounts of L-glutamate due to the strong induction of the enzymes of the arginine degradation pathway, RacE probably generates higher concentrations of D-glutamate than tolerated by the cell. The activation of the cryptic glutamate dehydrogenase GudB, the initialization of an alternative pathway to degrade glutamate and the inactivation of the GOGAT might then bring the glutamate concentration to a level that does not longer result in the accumulation of harmful D-glutamate. This hypothesis has to be subject of further investigation.

Additionally to the suppression of the growth defect on complex medium, the *ansR* mutation is also sufficient to provide growth of this mutant strain with glutamate as single carbon source (Flórez *et al*., subm.). The *gudB1* mutation also facilitates the growth of *B. subtilis* with glutamate as sole source of carbon (Commichau *et al*., 2008). Interestingly, the *ansR* mutant strain was isolated on complex medium as described. Thus, the activation of the alternative pathway for glutamate degradation is not only beneficial for the growth of this mutant on rich medium but is also essential for a condition to which the mutant was not exposed during isolation.

7.2.   The cryptification of the *gudB* gene

It is well established that cellular processes are highly regulated. The expression of genes required for the utilization of a nutritional source is induced as a response to the availability of this particular source. Moreover, enzymes for the biosynthesis of a certain metabolite are only expressed when the metabolite is limiting. Actually, the accumulation of a metabolite often represses the transcription of genes, needed for its

synthesis (Sonenshein, 2007). This regulation guarantees the utilization of a particular compound and the supply with a particular metabolite at a defined point of time. This can be regarded as efficiency, since a waste of energy and resources is prevented.

In this context the presence and most notably the expression of the cryptic glutamate dehydrogenase seem paradoxical in *B. subtilis*. This inactive enzyme is highly expressed even in the absence of its substrate glutamate and moreover it is rapidly degraded (Chapter 4; Gerth *et al*., 2008). An in depth analysis of a variety of laboratory *B. subtilis* strains and wild type isolates revealed that a functional *gudB* allele naturally occurs in the two wild Marburg strains as well as in Burkholder and Giles strain 122 (Zeigler *et al*., 2008). The presence of a functional *gudB* allele in the ancestors of the laboratory *B. subtilis* strains has two implications. First, the domestication of *B. subtilis* might have caused the *gudB* cryptification and second, the active GudB enzyme might be the main glutamate dehydrogenase in wild type isolates.

It is well established that the cultivation in the laboratory can have a striking impact on bacteria. *B. subtilis* wild type isolates form highly differentiated multicellular communities whereas the laboratory strain can only form thin and relatively undifferentiated biofilms (Branda *et al.*, 2001; Kearns & Losick, 2003). In the *B. subtilis* strain 168, not only biofilm formation is affected but also the swarming motility. The strain 168 is unable to swarm due to a frameshift mutation in the *swrA* gene (Patrick & Kearns, 2009).

Interestingly, cultivation of bacteria in the laboratory does not always lead to loss of function mutations but also to gain of function mutations. In naturally occurring *E. coli* strains the *bgl* operon is cryptic. Under selective conditions in the laboratory, the transcription of the *bgl* operon is enhanced by the acquisition of a mutation and this enables *E. coli* cells to utilize β-glucosides (Hall & Betts, 1987). It is tempting to speculate that cryptic genes might encode unusual functions that can be activated upon selective pressure (Tamburini & Mastromei, 2000).

The hypothesis that GudB is the original enzyme for glutamate degradation is not only supported by the fact that GudB is an active glutamate dehydrogenase in wild type isolates of *B. subtilis*. Moreover, the expression of the glutamate dehydrogenase RocG, the active enzyme in *B. subtilis* 168, is not induced by its substrate glutamate but by arginine, ornithine and citrulline as indicated by its name, **Roc**G for the abbreviations of arginine (**R**), ornithine (**o**), citrulline (**c**) (Belitsky & Sonenshein, 1998). Moreover, the

rocG gene is chromosomally clustered with other genes which are necessary for arginine utilization (RocABC). These observations might support the idea that GudB is responsible for glutamate utilization and RocG is needed when precursors of glutamate are available. In biochemical analyses the enzymatic parameters of RocG and GudB1 were determined (see Chapter 2). The GudB1 protein exhibits a higher enzymatic activity in carrying out glutamate degradation than the RocG protein (6.7 U/mg of protein and 3.9 U/mg of protein, respectively). In contrast, the GudB1 protein shows a lower affinity for its substrate glutamate compared to RocG (k_m values of 17.9 mM and 2.9 mM, respectively). These data might support the following idea: During glutamate excess, GudB1 is relatively active in degrading glutamate and therefore supplies the cell with 2-oxoglutarate. Subsequently, 2-oxoglutarate can be used in the TCA cycle for energy supply and other biosynthetic steps. But if glutamate is limiting in the cell, the high k_m value for glutamate might prevent that the low amount of glutamate is readily catabolized by GudB1, which is constitutively expressed irrespective of glutamate availability. The higher affinity of the RocG protein for glutamate might support growth with precursors of glutamate e. g. arginine. Due to its biochemical properties RocG might carry out glutamate degradation already at low glutamate concentrations.

The question remains why the cryptic *gudB* allele is highly stable in *B. subtilis* 168 upon cultivation in the laboratory but is so readily decryptified upon growth on complex medium in a *rocG* mutant. It is possible that both glutamate dehydrogenases are redundant upon cultivation in the laboratory, as the medium contains all nutrients required for growth, implying that the presence of RocG is sufficient to support growth on the medium *B. subtilis* is faced in the lab. But the absence of RocG is supposed to provoke a fast reactivation of GudB on complex medium as well as upon growth with glutamate as single carbon source (Belitsky & Sonenshein, 1998; Commichau *et al.*, 2008).

Based on the fact that the *gudB* allele is not cryptic in the ancestors of *B. subtilis* 168, it should be taken into account to rename the allele in the 168 strain. It would be easier to understand to use *gudB* for the functional gene and *gudB1* for the cryptic gene in *B. subtilis.*

7.3.   **The Mfd protein is required for the *gudB* decryptification**

The work revealed that the *gudB*1 mutation occurs with a rate of 10^{-4}. To the best of our knowledge, this is an extremely high mutation rate not only in *B. subtilis*. A striking feature of the *gudB* allele is the direct repeat of nine base pairs. Mutational analysis of the nine bp direct repeat showed that a perfect direct repeat is essential for the high decryptification rate of the *gudB* allele (see Chapter 4). It is well established that repeated sequences are common in a variety of genomes. This feature is highly abundant in particular in eukaryotic genomes but was also found to be a major factor for genomic rearrangement in bacteria (Viguera *et al.*, 2001; Michel, 2000). The instability of tandem repeats is supposed to be linked with replication. If the DNA polymerase is arrested at a repeated DNA sequence a deletion or an expansion of the repeated sequence can occur. The arrest of the DNA polymerase at the first part of the repeat can lead to a deletion whereas the arrest to the second part might result in an expansion (Michel, 2000). This model was termed as replication slippage. Especially tandem repeats are highly unstable and often undergo a contraction as well as an expansion (Bichara *et al.*, 2006). Well-studied tandem repeats such as the *lgtC* repeat in *Haemophilus influenzae* or the *nadA* repeat in *Neisseria meningitidis* consist of five to 36 repeats of tetranucleotides (Bayliss *et al.*, 2001; Martin *et al*, 2005). In those cases, the instability is linked to replication rather than to transcription (Bichara *et al.*, 2006). The *gudB* repeat is unique in possessing a large repeat (unit repeat of nine nucleotides) but this sequence is only repeated in two units. Compared to the length of well characterized repeats, two units are relatively short.

If a replication slippage mechanism would be the driving force for the rapid *gudB* decryptification, we would have expected to obtain the same high mutation rate of this repeat in a non-related genomic context. However, the use of the artificial mutagenesis promoter system revealed that the deletion of the nine bp of the direct repeat can occur in a different context but the mutation rate was reduced by three orders of magnitude. The artificial promoter system is located in a nontranscribed genomic region. Therefore, only replication can be involved in the emergence of mutations. This implies that the high mutation rate of the *gudB* allele needs another pathway than replication that can generate mutations. Indeed, the transcription-repair coupling factor Mfd turned out to be required for the high mutation rate of the *gudB* allele (see Chapter 4). In a *mfd* deficient background the mutation rate of the *gudB* allele was reduced hundred fold. Mfd

mediates the transcription-coupled DNA repair that is induced by the arrest of the RNA polymerase at a DNA lesion during transcription. Mfd displaces the RNA polymerase and recruits the nucleotide excision repair system to solve the lesion (Roberts & Park, 2004). Other studies have suggested that Mfd is necessary for the accumulation of mutations in transcribed genes during the stationary phase in *B. subtilis* (Ross *et al*., 2006; Pybus *et al*., 2010). The Mfd-mediated coupling of transcription to DNA repair and might facilitate the accumulation of mutations preferentially in transcribed genes. This has several implications: (i) The coupling allows that the mutations occur in genes that are expressed at the given time point; therefore the mutant variants of the encoded proteins might help to overcome the actual limitation. (ii) Non-transcribed genes that may be required under different conditions are in this way protected from potentially harmful mutations. Both effects facilitate the adaptation of bacteria to all kind of challenges that limit their growth and are therefore crucial for bacterial evolution. As shown, the *gudB* allele is highly expressed. Thus, the *gudB* allele fulfills the prerequisite to be subject of transcription-coupled DNA repair.

Furthermore, this work addressed the question which DNA repair system acts in concert with the Mfd protein in the *gudB* decryptification. The loss of the UvrABC as well as the MutSL system does not decrease the *gudB* mutation rate (see Chapter 5). Even the simultaneous deletion of the *uvrAB* and the *mutSL* genes does not influence the *gudB* decryptification. Therefore, the repair system that operates with the Mfd protein in the *gudB1* mutation has not been identified yet and might be subject to further investigation.

7.4.   **A perfect direct repeat is required for the efficient *gudB* decryptification**

Evidence was provided that a perfect direct repeat is crucial for the high mutation rate of the *gudB* allele. Destroying the direct repeat by introducing mutations in the first half as well as in the second half led to a fifteen fold decrease in the mutation rate and also the occurrence of *gudB1* revertants upon long time incubation on complex medium was significantly diminished (our unpublished data) These results emphasize the importance of the nine bp direct repeat for the rapid decryptification of the *gudB* gene. Since compensatory mutations in both parts of the direct repeat did not decrease the

mutation frequency, the wild type sequence of the direct repeat is not necessary for the high mutation rate.

As shown, the process of transcription coupled DNA repair is involved in the decryptification of *gudB*. Therefore, it is important whether the *gudB* sequence exhibits a particular feature that could cause the arrest of the RNA polymerase during transcription. One feature could be the formation of secondary structures in the coding strand.

An analysis of the direct repeat DNA sequence in the genomic context of the *gudB* allele in the transcribed strand (30 bases) with the mfold web server revealed a putative secondary structure (see Fig. 7.1) (Zucker, 2003). The predicted structure has a free energy of -2.85 kcal/mol (see Fig.7.1A). This secondary structure might be involved in the Mfd-mediated mutation in the *gudB* allele. During transcription the DNA is single stranded which might favor the formation of secondary structures. The replacement of two G residues by T in the first and the second half of the repeat (position 3 and 9 of the repeat) leads to a fifteen fold decrease of mutation frequency. This might correlate to a weaker secondary structure formation of these sequences. The free energy of putative secondary structures of these sequences is -0.88 kcal/mol for the mutations in the first half and -0.65 kcal/mol for the mutations in the second part of the direct repeat (see Fig. 7.1BC). A weaker secondary structure might reduce the chance of a roadblock during transcription. Moreover, the use of a non perfect direct repeat allowed us to address the question which half of the direct repeat has been excised. It turned out that the first part of the direct repeat is preferentially excised. However, in the minority of the sequenced mutants an internal deletion occurred (see chapter 4). Restoring the perfect direct repeat with T substitutions in both parts of the direct repeat leads to a mutation rate that is in the range of the wild type rate. Indeed, the calculated energy for a secondary structure of this DNA sequence is roughly comparable to the wild type structure (see Fig. 7.1.A and Fig. 7.1.D; -2.43 kcal/mol for the mutant perfect repeat and -2.85 kcal/mol for the wild type perfect repeat, respectively).

gudB WT

5' ...aagag**gtgaaggcggtgaaggcg**ctttcaa...
3' ...ttctc**cacttccgccacttccgc**gaaagtt...

gudB_mut1 (G3T G9T)

5'...aagag**gttaaggctgtgaaggcg**ctttcaa...
3'...ttctc**caattccgacacttccgc**gaaagtt...

gudB_mut3 (G12T G18T)

5'...aagag**gtgaaggcggttaaggct**ctttcaa...
3'...ttctc**cacttccgccaattccga**gaaagtt...

gudB_mut2 (G3T G9T G12T G18T)

5'...aagag**gtgaaggcggttaaggct**ctttcaa...
3'...ttctc**cacttccgccaattccga**gaaagtt...

Fig. 7.1 – Putative DNA secondary structures of the *gudB* direct repeat and its mutant derivatives.

Fig. legend see next page.

Fig. 7.1 – Predicted DNA secondary structures of the *gudB* direct repeat and its mutant derivatives.

For the secondary structure prediction the mfold web server was used (Zucker, 2003). As the template for prediction 30 bases of the coding strand were used containing the direct repeat. The first and the last base of the direct repeat are marked with an asterisk. The DNA sequence used for prediction is shown above the structure and the repeat is indicated by bold letters (A). The calculated free energy of the structure for the wild type sequence is -2.85 kcal/mol. Predictions of the secondary structures of the imperfect direct repeats are (B) -0.88 kcal/mol and (C) -0.65 kcal /mol. (D) Recovery ofthe direct repeat with the identical substitutions in the first and the second half results in a predicted free energy of -2.43 kcal/mol.

This result implies that the deletion might depend on a stable secondary structure rather than on the wild type sequence. The secondary structure prediction might support the idea that the rapid decryptification of the *gudB* allele is induced by an RNA polymerase arrest caused by a DNA lesion in the coding strand and requires the Mfd protein.

The formation of secondary structures might be also possible in the nontranscribed strand. It was demonstrated that the nontranscribed strand is also subject to mutations during transcription that are not mediated by Mfd (Kim *et al*., 2010). In the case of the *gudB* mutation, this mechanism of mutagenesis is entirely conceivable. Possibly, further mechanisms of mutagenesis might contribute to the *gudB* mutation. In agreement with this hypothesis is that after the loss of the Mfd protein, the mutation rate is in fact hundred fold decreased but the acquisition of the mutation is not completely abolished and upon long term incubation of the *mfd rocG* double mutant on complex medium *gudB1* revertants still occur. Although the mutation rate of the *mfd* mutant is drastically reduced, a mutation rate in the range of 10^{-6} is still striking compared to other observed mutation rates that are in the range of 10^{-8} (Kunkel, 2004). This result implies that further factors are involved in the precise deletion of the 9 bp in the direct repeat of the *gudB* gene. This topic has to be subject of further investigation.

7.5.   **The separation of enzymatic and regulatory activities of RocG**

In this work, two classes of mutant variants of the GDH RocG were isolated and characterized. Both classes uncouple the enzymatic and regulatory activities of this trigger enzyme RocG. Several single amino acid exchanges result in loss of enzymatic activity and are accompanied by permanent substrate-independent inactivation of GltC. This class of RocG variants was designated superrepressor RocG proteins (RocG-SR). A second class, exemplified by one mutant protein (designated monofunctional RocG, RocG-MF) has retained enzymatic activity but has lost the ability to inactivate GltC.

The mutations leading to the RocG SR phenotype are clustered in the active site of the enzyme. The separation of enzymatic and regulatory activities has already been observed for other trigger enzymes such as the *B. subtilis* glutamine synthetase (GS). The GS is regulated by glutamine feedback inhibition. The feedback inhibited GS triggers the DNA binding activity of two nitrogen transcriptional factors TnrA and GlnR (Wray *et al.*, 2001; Fisher & Wray, 2005). GS mutants that lost the capacity to regulate the activity of TnrA and GlnR also show mutations clustered to the region of the active site of the glutamine synthetase and are impaired in their enzymatic activity (Wray & Fisher, 2005; Fisher & Wray, 2009; Wray & Fisher, 2010). Mutations causing the RocG superrepressor phenotype reduce the catalytic activity and the affinity for glutamate (see Tab. 2.3). Thus, it can be assumed that the conformation of these RocG variants required for GltC inhibition is more stable than that of the wild-type RocG enzyme. These variants might require less glutamate to elicit the "inhibitory" conformation of glutamate dehydrogenase. By contrast, the monofunctional RocG enzyme is enzymatically active, but glutamate does not induce the inhibitory conformation required to inactivate GltC activity.

Mutations affecting the control of gene expression by trigger enzymes have not only been found in the genes encoding the trigger enzymes. Similarly, mutations affecting the controlled transcription factors may interfere with their productive interaction with the cognate trigger enzyme. Such mutations have been isolated for GltC, and reduced interaction with RocG has been demonstrated for one of these variants (Belitsky & Sonenshein, 1995; Commichau *et al.*, 2007a).

An interesting result of this work is the identification of 10 different superrepressor variants of RocG, while only one monofunctional protein incapable of inhibiting GltC was found. The small number of monofunctional RocG variants isolated

might indicate that the GltC interaction surface is relatively large and involves multiple side chains, and that the interaction between the two proteins and the inhibition of GltC by RocG cannot be easily disrupted by single amino acid substitutions in RocG.

The results presented support the idea of an inhibitory interaction between the glutamate dehydrogenase RocG and the transcription factor GltC. Moreover, they provide new insights into the relation between the enzymatic activities and the regulatory activities of the trigger enzyme RocG. The identification of the monofunctional RocG variant that has lost its regulatory function but has retained enzymatic activity is in excellent agreement with the previous conclusion: The enzymatic activity of RocG is important but not sufficient for the control exerted on GltC.

7.6.   **A new model of RocG-GltC interaction**

An *in vivo* cross-linking approach revealed that the RocG-MF protein still interacts with the transcriptional regulator GltC, even though the inhibitory effect of this interaction on GltC is completely abolished. Thus, it can be assumed that the interaction itself is not sufficient to regulate GltC activity. Moreover, it is tempting to speculate that the mode of inhibition is based on a particular conformation of the RocG-GltC complex that might be excluded due to the mutation in the RocG-MF protein. The amino acid exchange of an aspartate to an asparagine at the position 122 is surface located. The surface exposed location of the mutated residue could have two implications. First, the aspartate residue could be essential for binding a yet unknown factor that mediates GltC regulation by RocG. Second, the aspartate residue 122 of the RocG protein is in the interface of the interaction and is essential to transmit an inhibitory signal to GltC. The hypothesis that a third protein or an unknown low molecular weight factor is involved in the regulation is supported by the failure to reconstruct the RocG-GltC interaction *in vitro* (our unpublished results). Other examples showed that the reconstructs of a particular regulatory protein-protein interaction can be a challenge *in vitro*. The concept of carbon catabolite repression (CCR) has been an intensively studied subject in *E. coli*. There is strong evidence that the phosphorylated subunit of the glucose permease exerts a regulatory role in the CCR by stimulating the activity of the adenylate cyclase (Görke

& Stülke, 2008). Nevertheless, this model has never been proven *in vitro* (Park *et al*., 2006).

    In addition to the involvement of a third factor in the regulation of GltC by RocG it should be taken into account that the conformation of GltC might be important for regulation. GltC binds to different promoter sites in response to the presence of the low-molecular weight effectors, glutamate and 2-oxoglutarate (Picossi *et al*., 2007). High 2-oxoglutarate level favors the binding to the Boxes I and II *in vitro* which induces expression of the GOGAT. In the presence of glutamate GltC preferentially binds to the Boxes I and III leading to a loss of transcription activation. *In vivo*, this model is entirely conceivable to elucidate the regulation of GltC activity in the absence of RocG. But when RocG expression is induced, the RocG protein might be additionally responsible for sensing the glutamate to 2-oxoglutarate ratio and for transmitting this information to the GltC protein. This hypothesis might be explained with the following model (see Fig. 7.2). RocG interacts with GltC that has to bind previously to its target promoter region. It cannot be excluded that an additional protein is involved in the RocG-GltC interaction. In the model it is hypothesized that GltC has to be available in a particular quaternary structure that is formed upon DNA binding. Previous results suggest that GltC binds the DNA in a dimer-of-dimer fashion (Picossi *et al*., 2007). This assumption is supported by the observation that LysR regulators are functional as tetramer consisting of two dimers (Maddocks & Oyston, 2008) If the glutamate level is high reflected by catalytic performance of RocG, the binding capacity of GltC might be altered leading to a lack of induction of the GOGAT expression (see Fig. 7.2A). The depletion of glutamate might facilitate a conformational change in the interaction that allows GltC activity (see Fig. 7.2B). This model hypothesizes that RocG triggers a switch in GltC activity dependent on the glutamate concentration *in vivo*. This idea is in agreement with the previous observations concerning the impact of RocG on GltC activity (Commichau *et al*., 2007a; 2007b).
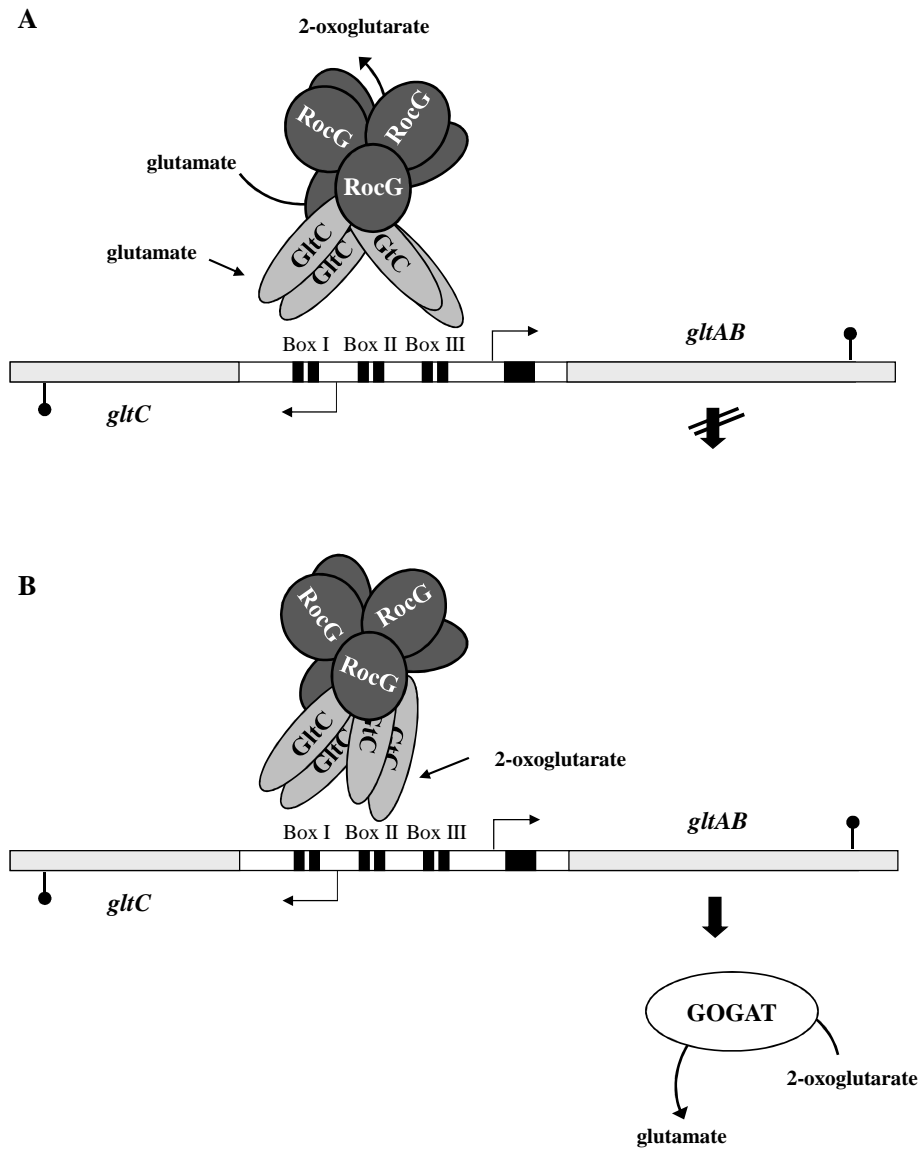
Fig. 7.2 – Model for the regulatory impact of RocG on GltC.

(A) RocG is capable of interacting with the transcriptional regulator GltC. In the presence of glutamate RocG carries out the degradation of glutamate to 2-oxoglutarate. The catalysis might favor a conformational change of GltC that does not allow the expression of the *gltAB* operon. (B) If the glutamate level drops, RocG might be no longer able to force GltC in an inhibitory conformation. This leads to the transcription of the *gltAB* operon.

The model is also supported by the result that the RocG-SR proteins exert a strong inhibition on GltC in the presence of glutamate but do not influence GltC activity in the absence of glutamate. The activity in terms of glutamate degradation of the RocG-SR proteins is severely impaired and not sufficient to allow growth with arginine as sole carbon source. Nevertheless, they are still capable of catalyzing the reaction very slowly. This result leads to the hypothesis that the RocG-SR proteins are frozen in a glutamate bound state and therefore mediating a permanent strong inhibition of GltC activity. The RocG-MF protein is still capable of efficient glutamate utilization but lost the capacity to switch GltC activity. It is possible that RocG-MF cannot process an impact of an additional factor that might influence the effect of the RocG-GltC interaction. Another explanation might be that RocG-MF cannot transmit the information to GltC that leads to a conformational change of GltC multimers. These hypotheses have to be subject of further investigation.

7.7.   **Perspectives**

The results revealed the requirement of the Mfd protein in the rapid *gudB* decryptification. However, the DNA repair system that operates with Mfd in the mutagenesis has not been identified yet. In order to gain further insights into the Mfd-mediated generation of *gudB1* mutations a SPINE experiment using a *rocG* mutant strain can be performed (Herzberg *et al*., 2007). By this *in vivo* cross linking approach potential interaction partners of the Mfd protein might be identified. To increase the chance of cross-linking Mfd in complex with its potential target the *gudB* gene and additional interaction partners, the *gudB* gene has to be enriched and transcribed at high level. This might be of importance as a high transcription rate is supposed to enhance the transcription coupled DNA repair mediated by Mfd (Gunka *et al*., subm.). Besides the analysis of the elution fraction by mass spectrometry to identify interaction partners of Mfd, a Northern blot with a *gudB* probe can be performed. This experiment might demonstrate whether the *gudB* transcript can be specifically cross-linked with the Mfd protein.

To test the hypothesis that the protein-protein interaction between RocG and GltC depends on the presence of the target promoter region of the transcription activator GltC, the defined DNA sequence can be introduced in the background of the bacterial two-hybrid assay. This can be simply achieved by cloning the promoter region into one of the vectors for the RocG and GltC co-expression. If the *gltAB* promoter is necessary to facilitate the right conformation of GltC and enables RocG to interact with GltC, this interaction might be detected in the bacterial two-hybrid assay in the presence of the promoter sequence.

A new SPINE approach might be of interest to identify the putative interaction partner that mediates the RocG-GltC interaction. With a newly developed mass spectrometry method the complete elution fractions of the co-purification can be analyzed very sensitively (E. Hammer, pers. comm.).

Furthermore, *in vitro* studies on the RocG-GltC interaction would be of great value to understand the regulatory mechanism. In order to check whether the RocG-GltC interaction influences the binding properties of GltC to its target promoter region, DNA footprint analysis would be an appropriate approach. For this approach RocG-SR variants could be used in comparison to the RocG-MF protein, that was shown to lack GltC inhibition.

8.   **References**

Abramoff, M. D., Magelhaes, P. J., & Ram, S. J. (2004) Image processing with ImageJ. Biophotonics. Internat. **11**: 36-42.

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A., *et al***.** (2002) PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D Biol. Crystallogr. **58:** 1948–1954.

Ali, N. O., Jeusset, J., Larquet, E., Le Cam, E., Belitsky, B., *et al***.** (2003) Specificity of the interaction of RocR with the *rocG-rocA* intergenic region in *Bacillus subtilis*. Microbiology **149:** 739-750.

Arcondéguy, T., Jack, R. & Merrick, M. (2001) P(II) signal transduction proteins, pivotal players in microbial nitrogen control. Microbiol. Mol. Biol. Rev. **65:** 80-105.

Assenmacher, N. Wenig, K. Lammens, A. & Hopfner, K. H. (2006) Structural basis for transcription-coupled repair: the N terminus of Mfd resembles UvrB with degenerate ATPase motifs. J. Mol. Biol. **355:** 675-83.

Ayora, S., Rojo, F., Ogasawara, N., Nakai, S. & Alonso. J. C. (1996) The Mfd protein of *Bacillus subtilis* 168 is involved in both transcription-coupled DNA repair and DNA recombination. J. Mol. Biol. **256:** 301-318.

Bachem, S. & Stülke, J. (1998) Regulation of the *Bacillus subtilis* GlcT antiterminator protein by components of the phosphotransferase system. J. Bacteriol. **180:** 5319–5326.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. Proc. Natl Acad. Sci. U.S.A. **98:** 10037–10041.

Baker, P. J., Waugh, M. L., Wang, X. G., Stillman, T. J., Turnbull, A. P., *et al***.** (1997) Determinants of substrate specificity in the superfamily of amino acid dehydrogenases. Biochemistry **36:** 16109–16115.

Barbe, V., Cruveiller, S., Kunst, F., Lenoble, P., Meurice, G., *et al.* (2009) From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. Microbiology **155:** 1758-1775.

Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., *et al.* (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli.* Nature **461:** 1243-1247.

Bayliss, C. D., Field, D. & Moxon, E. R. (2001) The simple sequence contingency loci of *Haemophilus influenzae* and *Neisseria meningitidis*, J. Clin. Invest. **107:** 657–662.

Beletskii, A., & Bhagwat, A.S. (1996) Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. **93:** 13919–13924.

Belitsky, B. R., Janssen, P. J. & Sonenshein, A. L. (1995) Sites required for GltC-dependent regulation of *Bacillus subtilis* glutamate synthase expression. J. Bacteriol. **177:** 5686-5695.

Belitsky, B. R. & Sonenshein, A. L. (1995) Mutations in GltC that increase *Bacillus subtilis gltA* expression. J. Bacteriol. **177:** 5696-5700.

Belitsky, B. R. & Sonenshein, A. L. (1997) Altered transcription activation specificity of a mutant form of *Bacillus subtilis* GltR, a LysR family member. J. Bacteriol. **179:** 1035-1043.

Belitsky, B. R. & Sonenshein, A. L. (1998) Role and regulation of *Bacillus subtilis* glutamate dehydrogenase genes. J. Bacteriol. **180:** 6298–6305.

Belitsky, B. R. & Sonenshein, A. L. (1999) An enhancer element located downstream of the major glutamate dehydrogenase gene of *Bacillus subtilis*. Proc. Natl. Acad. Sci. U.S.A. **96:** 10290–10295.

Belitsky, B. R., Wray, L. V., Jr., Fisher, S. H., Bohannon, D. E. & Sonenshein, A. L. (2000) Role of TnrA in nitrogen source-dependent repression of *Bacillus subtilis* glutamate synthase gene expression. J. Bacteriol. **182:** 5939–5947.

Belitsky, B. R., Kim, H. J. & Sonenshein, A. L. (2004) CcpA-dependent regulation of *Bacillus subtilis* glutamate dehydrogenase gene expression. J. Bacteriol. **186:** 3392–3398.

Belitsky, B. R. & Sonenshein, A. L. (2004) Modulation of activity of *Bacillus subtilis* regulatory proteins GltC and TnrA by glutamate dehydrogenase. J. Bacteriol. **186:** 3399–3407.

Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J. & Rabinowitz, J. D. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli.* Nat. Chem. Biol. **5:** 593-599.

Bhuiya, M. W., Sakuraba, H., Ohshima, T., Imagawa, T., Katunuma, N. & Tsuge, H. (2005) The first crystal structure of hyperthermostable NAD-dependent glutamate dehydrogenase from *Pyrobaculum islandicum.* J. Mol. Biol. **345:** 325–337.

Bi, W. & Stambrook, P. J. (1998) Site-directed mutagenesis by combined chain reaction. Anal. Biochem. **256:** 137-140

Bichara, M., Wagner, J. & Lambert, I. B. (2006) Mechanisms of tandem repeat instability in bacteria. Mutat. Res. **598:** 144-163.

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.

Bohannon, D. E., Rosenkrantz, M. S. & Sonenshein, A. L. (1985) Regulation of *Bacillus subtilis* glutamate synthase genes by the nitrogen source. J. Bacteriol. **163:** 957-964.

Bohannon, D. E. & Sonenshein, A. L. (1989) Positive regulation of glutamate biosynthesis in *Bacillus subtilis*. J. Bacteriol. **171:** 4718–4727.

Bohr, V. A., Smith, C. A., Okumoto, D. S. & Hanawalt, P. C. (1985) DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. Cell **40:** 359-369.

Boylan, S. A., Redfield, A. R., Brody, M. S. & Price, C. W. (1993) Stress-induced activation of the sigma B transcription factor of *Bacillus subtilis*. J. Bacteriol. **175:** 7931-7937.

Branda, S. S., Gonzalez-Pastor, J. E., Ben-Yehuda, S., Losick, R. & Kolter, R. (2001) Fruiting body formation by *Bacillus subtilis*. Proc. Natl Acad. Sci. U.S.A. **98:** 11621-11626.

Branda, S. S., González-Pastor, J. E., Dervyn, E., Ehrlich, S. D., Losick, R. & Kolter R. (2004) Genes involved in formation of structured multicellular communities by *Bacillus subtilis.* J. Bacteriol. **186:** 3970-3979.

Braun, P., Tasan, M., Dreze, M., Barrios-Rodiles, M., Lemmens, I., *et al***.** (2009) An experimentally derived confidence score for binary protein–protein interactions. Nat. Methods **6:** 91–97.

Britton, K. L., Yip, K. S., Sedelnikova, S. E., Stillman, T. J., Adams, *et al***.** (1999) Structure determination of the glutamate dehydrogenase from the hyperthermophile *Thermococcus litoralis* and its comparison with that from *Pyrococcus furiosus*. J. Mol. Biol. **293:** 1121–1132.

Borukhov, S., Lee, J. & Laptenko, O. (2005) Bacterial transcription elogantion factors: new insights into molecular mechanism of action. Mol. Microbiol. **55:** 1315-1342.

Cairns, J. & Foster, P. L. (1991) Adaptive reversion of a frameshift mutation in *Escherichia coli*. Genetics **128:** 695–701.

Cairns, J., Overbaugh, J. & Miller, S. (1988) The origin of mutants. Nature **335:** 142-145.

Calogero, S., Gardan, R., Glaser, P., Schweizer, J., Rapoport, G. & Débarbouillé, M. (1994) RocR, a novel regulatory protein controlling arginine utilization in *Bacillus subtilis*, belongs to the NtrC/NifA family of transcriptional activators. J. Bacteriol. **176:** 1234-1241.

CCP4 (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. **50:** 760–763.

Chagneau, C. & Saier, M. H., Jr. (2004) Biofilm-defective mutants of *Bacillus subtilis*. J. Mol. Microbiol. Biotechnol. **8:** 177-188.

Chédin, F., Noirot, P., Biaudet, V., & Ehrlich, S. D. (1998) A five-nucleotide sequence protects DNA from exonucleolytic degradation by AddAB, the RecBCD analogue of *Bacillus subtilis*. Mol. Microbiol. **29:** 1369-1377.

Chin, G., Jr., Chavarria, D. G., Nakamura, G. C. & Sofia, H. J. (2008) BioGraphE: high-performance bionetwork analysis using the Biological Graph Environment. BMC Bioinformatics **9:** S6.

Chumsakul, O., Takahashi, H., Oshima, T., Hishimoto, T., Kanaya, S., *et al*. (2010) Genome-wide binding profiles of the *Bacillus subtilis* transition state regulator AbrB and its homolog Abh reveals their interactive role in transcriptional regulation. Nucleic Acids Res. in press (PMID: 20817675).

Claessen, D., Emmins, R., Hamoen, L. W., Daniel, R. A., Errington, J. & Edwards, D. H. (2008) Control of the cell elongation-division cycle by shuttling of PBP1 protein in *Bacillus subtilis*. Mol. Microbiol. **68:** 1029–1046.

Claessen, K., Een, N., Sheeran, M., Sörensson, N., Voronov, A. & Åkesson, K. (2009) SAT-solving in practice, with a tutorial example from supervisory control. Discrete Event Dyn. Syst. **19:** 495-524.

Claverys, J. P. & Håvarstein, L. S. (2007) Cannibalism and fratricide: mechanisms and raisons d'être. Nat. Rev. Microbiol. **5:** 219-229.

Commichau, F. M. (2006) Regulation der Glutamatsynthese in *Bacillus subtilis* durch die Glutamatdehydrogenase RocG und das Aktivatorprotein GltC. PhD thesis. Georg-August-Universität Göttingen; Germany

Commichau, F. M., Forchhammer, K. & Stülke J. (2006) Regulatory links between carbon and nitrogen metabolism. Curr. Opin. Microbiol. **9:** 167-172.

Commichau, F. M., Herzberg, C., Tripal, P., Valerius, O. & Stülke, J. (2007a) A regulatory protein–protein interaction governs glutamate biosynthesis in *Bacillus subtilis*: the glutamate dehydrogenase RocG moonlights in controlling the transcription factor GltC. Mol. Microbiol. **65:** 642–654.

Commichau, F. M., Wacker, I., Schleider, J., Blencke, H. M., Reif, I., *et al.* (2007b) Characterization of *Bacillus subtilis* mutants with carbon-source independent glutamate biosynthesis. J. Mol. Microbiol. Biotechnol. **12:** 106–113.

Commichau, F M., Gunka, K., Landmann, J. J. & Stülke, J. (2008) Glutamate metabolism in *Bacillus subtilis*: gene expression and enzyme activities evolved to avoid futile cycles and to allow rapid responses to perturbations of the system. J. Bacteriol. **190:** 3557-3564.

Commichau, F. M. & Stülke, J. (2008) Trigger enzymes: bifunctional proteins active in metabolism and in controlling gene expression. Mol. Microbiol. **67:** 692–702.

da Costa, M. S., Santos, H. & Galinski, E. A. (1998) An overview of the role and diversity of compatible solutes in Bacteria and Archaea. Adv. Biochem. Eng. Biotechnol. **61:** 117-153.

Darwin, C. R. (1859) The origin of species. Ed. Avenel Books, New York 1979.

de Figueiredo, L. F., Podhorski, A., Rubio, A., Kaleta, C., Beasley, J. E., *et al.* (2009) Computing the shortest elementary flux modes in genome-scale metabolic networks. Bioinformatics **25:** 3158-3165.

de Jong, H. & Page, M. (2008) Search for steady states of piecewise-linear differential equation models of genetic regulatory networks. IEEE/ACM Trans Comput Biol Bioinform **5:** 208-22.

Deaconescu, A. M., Chambers, A. L., Smith, A. J., Nickels, B. E., Hochschild, A., *et al.* (2006) Structural basis for bacterial transcription-coupled DNA repair. Cell **124:** 507-520.

Débarbouillé, M., Martin-Verstraete, I., Kunst, F. & Rapoport, G. (1991) The *Bacillus subtilis sigL* gene encodes an equivalent of sigma 54 from gram-negative bacteria. Proc. Natl .Acad. Sci. U.S.A. **88:** 9092-9096.

Denamur, E. & Matic, I. (2006) Evolution of mutation rates in bacteria. Mol. Microbiol. **60:** 820-7.

Deshpande, K. L. & Kane, J. F. (1980) Glutamate synthase from *Bacillus subtilis*: *in vitro* reconstitution of an active amidotransferase. Biochem. Biophys. Res. Commun. **93:** 308-314.

Detsch, C. & Stülke, J. (2003) Ammonium utilization in *Bacillus subtilis*: transport and regulatory functions of NrgA and NrgB. Microbiology **149:** 3289–3297.

Deuel, T. F. & Prusiner, S. (1974) Regulation of glutamine synthetase from *Bacillus subtilis* by divalent cations, feedback inhibitors, and L-glutamine. J. Biol. Chem. **249:** 257-264.

Eisen, J. A. & Hanawalt, P. C. (1999) A phylogenomic study of DNA repair genes, proteins, and processes. Mutat. Res. **43:** 171-213.

Emsley, P. & Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. **60:** 2126–2132.

Evans, P. R. (1993) Data collection and processing. Proc. CCP4 Study Weekend, 114–122.

Feavers, I. M., Price, V. & Moir, A. (1998) The regulation of the fumarase (*citG*) gene of *Bacillus subtilis* 168. Mol. Gen. Genet. **211:** 465-471.

Feist, A. M. & Palsson, B. Ø. (2008) The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. Nat. Biotechnol. **26:** 659-667.

Fisher, S. H. (1999) Regulation of nitrogen metabolism in *Bacillus subtilis*: vive la différence! Mol. Microbiol. **32:** 223–232.

Fisher, S. H. & Débarbouillé, M. (2002) Nitrogen Source Utilization and Its Regulation. A. L. Sonenshein, J. A. Hoch and R. Losick. *Bacillus subtilis* and Its Closest Realtives:from Genes to Cells. ASM Press, Washington, D. C. **pp. 181-191**.

Fisher, S. H. & Wray, L. V., Jr. (2002) *Bacillus subtilis* 168 contains two differentially regulated genes encoding L-asparaginase. J. Bacteriol. **184:** 2148-54.

Fisher, S. H. & Wray, L. V., Jr. (2008) *Bacillus subtilis* glutamine synthetase regulates its own synthesis by acting as a chaperone to stabilize GlnR–DNA complexes. Proc. Natl. Acad. Sci. U.S.A. **105:** 1014-1019.

Fisher, S. H. & Wray, L. V., Jr. (2009) Novel transacting *Bacillus subtilis glnA* mutations that derepress *glnRA* expression. J. Bacteriol. **191:** 2485–2492.

Flórez, L. A., Roppel, S. F., Schmeisky, A. G., Lammers, C. R. & Stülke J. (2009) A community-curated consensual annotation that is continuously updated: the *Bacillus subtilis* centred wiki SubtiWiki. Database 2009:bap012.

Flórez, L. A., Gunka, K., Polanía, R., Tholen, S. & Stülke J. SPABBATS: A pathway-discovery method based on Boolean satisfiability that facilitates the characterization of suppressor mutants BMC Systems Biology submitted

Fousteri, M. & Mullenders, L. F. H. (2008) Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. Cell Res. **8:** 73-84.

Fränzle, M., Herde, C. & Teige, T. (2007) Efficient solving of large non-linear arithmetic constraint systems with complex Boolean structure. Journal on Satisfiability **1:** 209-236.

Frimmersdorf, E., Horatzek, S., Pelnikevich, A., Wiehlmann, L. & Schomburg, D. (2010) How *Pseudomonas aeruginosa* adapts to various environments: a metabolomic approach. Environ. Microbiol. **12:** 1734-1747.

Fukui, K. (2010) DNA mismatch repair in eukaryotes and bacteria. J. Nucleic. Acids. pii: 260512.

Gardan, R., Rapoport, G. & Débarbouillé, M. (1995) Expression of the *rocDEF* operon involved in arginine catabolism in *Bacillus subtilis*. J. Mol .Biol. **249:** 843-856.

Gardan, R., Rapoport, G. & Débarbouillé, M. (1997) Role of the transcriptional activator RocR in the arginine-degradation pathway of *Bacillus subtilis*. Mol. Microbiol. **24:** 825-837.

Gardner, A. L. & Aronson, A. I. (1984) Expression of the *Bacillus subtilis* glutamine synthetase gene in *Escherichia coli*. J. Bacteriol. **158:** 967-971.

Gerth, U., Kock, H., Küsters, I., Michalik, S., Switzer, R. L. & Hecker, M. (2008) Clp-dependent proteolysis down-regulates central metabolic pathways in glucose-starved *Bacillus subtilis*. J. Bacteriol. **190:** 321–331.

Goelzer, A., Bekkal Brikci, F., Martin-Verstraete, I., Noirot, P., Bessières, P., *et al.* (2008) Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*. BMC Syst Biol **2:** 20.

Gordon, R. E. (1981) One hundred and seven years of the genus *Bacillus*. R. C. Berkeley, and M. Goodfellow (ed.) The aerobic endosporeforming bacteria. Published for the Society for General Microbiology by Academic Press, London ; New York.

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.

Goss, T. J., Perez-Matos, A. & Bender, R. A. (2001) Roles of glutamate synthase, *gltBD*, and *gltF* in nitrogen metabolism of *Escherichia coli* and *Klebsiella aerogenes*. J Bacteriol. **183:** 6607-6619.

Graça, A., Marques-Silva, J., Lyncem I. & Oliveira, A. L. (2007) Efficient haplotype inference with pseudo-Boolean optimization. In Algebraic Biology Springer Verlag Berlin/Heidelberg: **2007:**125-139.

Greener, A. & Callahan, M. (1994) XL1-Red: a highly efficient random mutagenesis strain. Strategies **7:** 32-34.

Guérout-Fleury, A. M., Shazand, K., Frandsen, N. & Stragier, P. (1995) Antibiotic resistance cassettes for *Bacillus subtilis*. Gene **167:** 335-336.

Gunasekaran, K. & Nussinov, R. (2007) How different are structurally flexible and rigid binding sites? Sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. J. Mol. Biol. **365:** 257-273

Gunka, K., Newman, J. A., Commichau, F. M., Herzberg, C., Rodrigues, C., *et al.* (2010) Functional dissection of a trigger enzyme: mutations of the *Bacillus subtilis* glutamate dehydrogenase RocG that affect differentially its catalytic activity and regulatory properties. J. Mol. Biol. **400:** 815-827.

Gunka, K., Tholen, S., Neme, R., Freytag, B., Commichau, F. M., *et al.* A high-frequency mutation in *Bacillus subtilis*: Requirements for the decryptification of the gudB glutamate dehydrogenase gene. Mol. Microbiol. submitted

Guo, C., Tang, T. S. & Friedberg, E. C. (2010) SnapShot: nucleotide excision repair. Cell **140:** 754-4.e1.

Hackett, R. H. & Setlow, P. (1988) Properties of spores of *Bacillus subtil*is strain which lack the major small, acid-soluble protein. J. Bacteriol. **170:** 1403–1404

Hall, B. G. & Betts, P. W. (1987) Cryptic genes for cellobiose utilization in natural isolates of *Escherichia coli*, Genetics **115:** 431–439.

Hamoen, L. W., Venema, G. & Kuipers, O. P. (2003) Controlling competence in *Bacillus subtilis*: shared use of regulators. Microbiology **149:** 9-17.

Hanawalt, P. C. & Spivak, G. (2008) Transcription-coupled DNA repair: two decades of progress and surprises Nat. Rev. Mol. Cell Biol. **9:** 958-970.

Harwood, C. (1990) Molecular biological methods for *Bacillus*. Cutting, S. M. (ed.) John Wiley & Sons, Ltd., Chichester, United Kingdom.

Hecker, M., Pané-Farré, J. & Völker, U. (2007) SigB-dependent general stress response in *Bacillus subtilis* and related Gram-positive bacteria. Annu. Rev. Microbiol. **61:** 215-236.

Helmann, J. D., Wu, M. F., Kobel, P. A., Gamo, F. J., Wilson, M., *et al*. (2001) Global transcriptional response of *Bacillus subtilis* to heat shock. J. Bacteriol. **183:** 7318-7328.

Herzberg, C., Weidinger, L. A., Dörrbecker, B., Hübner, S., Stülke, J. & Commichau, F. M. (2007) SPINE: a method for the rapid detection and analysis of protein–protein interactions *in vivo*. Proteomics **7:** 4032–4035.

Ikeda, T. P., Shauger, A. E. & Kustu, S. (1996) *Salmonella typhimurium* apparently perceives external nitrogen limitation as internal glutamine limitation. J. Mol. Biol. **259:** 589-607.

Imhof, M. & Schlötterer, C. (2001) Fitness effects of advantageous mutations in evolving *Escherichia coli* populations. Proc. Natl. Acad. Sci. U.S.A. **98:** 1113-1117.

Jeffery, C. J. (1999)**.** Moonlighting proteins. Trends Biochem. Sci. **24:** 8-11.

Jin, S. & Sonenshein, A. L. (1994): Transcriptional regulation of *Bacillus subtilis* citrate synthase genes. J. Bacteriol. **176:** 4680-4690.

Jinap, S. & Hajeb, P. (2010) Glutamate. Its applications in food and contribution to health. Appetite **55:** 1-10.

Karimova, G., Pidoux, J., Ullmann, A. & Ladant, D., (1998) A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proc. Natl. Acad. Sci. U.S.A. **95:** 5752–5756.

Kearns, D. B. & Losick, R. (2003) Swarming motility in undomesticated *Bacillus subtilis*. Mol. Microbiol. **49:** 581-590.

Kempf, B. & Bremer, E. (1998) Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. Arch. Microbiol. **170:** 319-330.

Kim, H., Lee, B. S., Tomita, M. & Kanai, A. (2010). Transcription-associated mutagenesis increases protein sequence diversity more effectively than does random mutagenesis in *Escherichia coli*. PLoS ONE **5:** e10567.

Kim, N., Abdulovic, A. L., Gealy, R., Lippert, M. J. & Jinks-Robertson, S. (2007) Transcription-associated mutagenesis in yeast is directly proportional to the level of gene expression and influenced by the direction of DNA replication. DNA Repair **6:** 1285-1296.

Kimura, K., Tran, L.-S. P. & Itoh, Y. (2004) Roles and regulation of the glutamate racemase isogenes, *racE* and *yrpC*, in *Bacillus subtilis*. Microbiology **150:** 2911-2920.

Klamt, S. & Stelling, J. (2002) Combinatorial complexity of pathway analysis in metabolic networks. Mol. Biol. Rep. **29:** 233-236.

Klapacz, J. & Bhagwat, A. S. (2005) Transcription promotes guanine to thymine mutations in the non-transcribed strand of an *Escherichia coli* gene. DNA Repair **4:** 806-813.

Klingel, U., Miller, C. M., North, A. K., Stockley, P. G. & Baumberg, S. (1995) A binding site for activation by the *Bacillus subtilis* AhrC protein, a repressor/activator of arginine metabolism. Mol. Gen. Genet. **248:** 329–340.

Knapp, S., de Vos, W. M., Rice, D. W. & Ladenstein, R. (1997) Crystal structure of glutamate dehydrogenase from the hyperthermophilic eubacterium *Thermotoga maritima* at 3.0 Å resolution. J. Mol. Biol. **267:** 916–932.

Kondoh, T. & Torii, K. (2008) Brain activation by Umami substances via gustatory and visceral signaling pathways, and physiological significance. Biol. Pharm. Bull. **31:** 1827-1832.

Kunkel, T. A. (2004) DNA replication fidelity. J. Biol. Chem. **279:** 16895-16898.

Kunst, F. & Rapoport, G. (1995) Salt stress is an environmental signal affecting degradative enzyme synthesis in *Bacillus subtilis*. J. Bacteriol. **177:** 2403–2407.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., *et al***.** (1997) The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. Nature **390:** 249-256.

Lammers, C. R., Flórez, L. A., Schmeisky, A. G., Roppel, S. F., Mäder, U., *et al.* (2010) Connecting parts with processes: SubtiWiki and SubtiPathways integrate gene and pathway annotation for *Bacillus subtilis*. Microbiology **156:** 849-859.

Laskowski, R. A., MacAurthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. **26:** 283–291.

Lea, D. E., & Coulson, C. A. (1949) The distribution of mutants in bacterial populations. J. Genet. **49:** 248-264.

Leigh, J. A. & Dodsworth, J. A. (2007) Nitrogen regulation in Bacteria and Archaea. Annu. Rev. Microbiol. **61:** 349-377.

Lemon, K. P., Kurtser, I. & Grossman, A. D. (2001) Effects of replication termination mutants on chromosome partitioning in *Bacillus subtilis.* Proc. Natl. Acad. Sci. U.S.A. **98:** 212-217.

Lehnik-Habrink, M., Pförtner, H., Rempeters, L., Pietack, N., Herzberg, C. & Stülke, J. (2010) The RNA degradosome in *Bacillus subtilis*: identification of CshA as the major RNA helicase in the multiprotein complex. Mol. Microbiol. **77**: 958–971.

Leslie, A. G. W. (1992) Joint CCP4 + ESF-EAMCB. Newsl. Protein Crystallogr. **26**.

López, D., Vlamakis, H. & Kolter, R. (2010) Biofilms. Cold Spring Harb Perspect Biol **2:** a000398.

Lou, C., Liu, X., Ni, M., Huang, Y., Huang, Q., *et al.* (2010) Synthesizing a novel genetic sequential logic circuit: a push-on push-off switch. Mol. Syst. Biol. **6:** 350.

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:** 409-422.

Luria, S. E. & Delbrück, M. (1943) Mutations of bacteria from virus sensitivity to virus resistance. Genetics **28:** 491-511.

Maddocks, S. E. & Oyston, P. C. (2008) Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. Microbiology **154:** 3609-3623.

Magasanik, B. (2003) Ammonia assimilation by *Saccharomyces cerevisiae*. Eukaryotic Cell **2:** 827-829.

Martin, P., Makepeace, K., Hill, S. A., Hood, D. W. & Moxon, E. R. (2005) Microsatellite instability regulates transcription factor binding and gene expression. Proc. Natl. Acad. Sci. U.S.A **102:** 3800–3804.

Martin-Verstraete, I., Débarbouillé, M., Klier, A. & Rapoport, G. (1992) Mutagenesis of the *Bacillus subtilis* "-12,-24" promoter of the levanase operon and evidence for the existence of an upstream activating sequence. J. Mol. Biol. **226:** 85-99.

Martin-Verstraete, I., Débarbouillé, M., Klier, A. & Rapoport, G. (1994) Interactions of wild-type and truncated LevR of *Bacillus subtilis* with the upstream activating sequence of the levanase operon. J. Mol. Biol. **241:** 178–192.

Martin-Verstraete, I., Charrier, V., Stülke, J., Galinier, A., Erni, B., *et al.* (1998) Antagonistic effects of dual PTS catalyzed phosphorylation on the *Bacillus subtilis* transcriptional activator LevR. Mol. Microbiol. **28:** 293–303.

Matsuoka, K. & Kimura, K. (1986) Glutamate synthase from *Bacillus subtilis* PCI 219. J. Biochem. (Tokyo) **99:** 1087-1100.

Meinken, C., Blencke, H. M., Ludwig, H. & Stülke, J. (2003) Expression of the glycolytic *gapA* operon in *Bacillus subtilis*: differential syntheses of proteins encoded by the operon. Microbiology **149:** 751-761.

Mellon, I. & Champe, G. N. (1996) Products of DNA mismatch repair genes *mutS* and *mutL* are required for transcription-coupled nucleotide-excision repair of the lactose operon in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **93:** 1292-1297.

Merzbacher, M., Detsch, C., Hillen, W. & Stülke, J. (2004) *Mycoplasma pneumoniae* HPr kinase/phosphorylase: assigning functional roles to the P-loop and the HPrK/P signature sequence motif. Eur. J. Biochem. **271:** 367–374.

Meyer, F. M., Gerwig, J., Hammer, E., Herzberg, C., Commichau, F. M., *et al.* (2011) Physical interactions between tricarboxylic acid cycle enzymes in *Bacillus subtilis*: evidence for a metabolon. Metab, Engn. in press (PMID: 20933603).

Michel, B. (2000) Replication fork arrest and DNA recombination. Trends Biochem. Sci. **25:** 173-178.

Miller, C. M., Baumberg, S. & Stockley, P. G. (1997) Operator interactions by the *Bacillus subtilis* arginine repressor/activator, AhrC: novel positioning and DNA-mediated assembly of a transcriptional activator at catabolic sites. Mol. Microbiol. **26:** 37-48.

Modrich, P. & Lahue, R. (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu. Rev. Biochem. **65:** 101-133.

Morett, E. & Segovia, L. (1993) The sigma 54 bacterial enhancer-binding protein family: mechanism of action and phylogenetic relationship of their functional domains. J. Bacteriol. **175:** 6067-6074.

Muro-Pastor, A. M. & Maloy, S. (1995) Proline dehydrogenase activity of the transcriptional repressor PutA is required for induction of the *put* operon by proline. J. Biol. Chem. **270:** 9819–9827.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. **53:** 240–255.

Nakasako, M., Fujisawa, T., Adachi, S., Kudo, T. & Higuchi, S. (2001) Large-scale domain movements and hydration structure changes in the active-site cleft of unligated glutamate dehydrogenase from *Thermococcus profundus* studied by cryogenic X-ray crystal structure analysis and small-angle X-ray scattering. Biochemistry **40:** 3069–3079.

Ninfa, A. J. & Jiang, P. (2005) PII signal transduction proteins: sensors of alpha-ketoglutarate that regulate nitrogen metabolism. Curr. Opin. Microbiol. **8:** 168-173.

Oh, Y. K., Palsson, B. Ø., Park, S. M., Schilling , C. H. & Mahadevan, R. (2007) Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. J. Biol. Chem. **282:** 28791-28799.

Orth, J. D., Thiele, I. & Palsson, B. Ø. (2010) What is flux balance analysis? Nat. Biotechnol. **28:** 245-248.

Park, Y. H., Lee, B. R., Seok, Y.J. & Peterkofsky, A. (2006) *In vitro* reconstitution of catabolite repression in *Escherichia coli*. J Biol Chem. **281:**6448-6454.

Patrick, J. E. & Kearns, D. B. (2009) Laboratory strains of *Bacillus subtilis* do not exhibit swarming motility. J. Bacteriol. **191:** 7129-7133.

Persuh, M., Mandic-Mulec, I. & Dubnau, D. (2002) A MecA paralog, YpbH, binds ClpC, affecting both competence and sporulation. J Bacteriol. **184:** 2310-2313.

Picossi, S., Belitsky, B. R. & Sonenshein, A. L. (2007) Molecular mechanism of the regulation of *Bacillus subtilis gltAB* expression by GltC. J. Mol. Biol. **365:** 1298–1313.

Piran, R., Halperin, E., Guttmann-Raviv, N., Keinan, E. & Reshef, R. (2009) Algorithm of myogenic differentiation in higher-order organisms. Development **136:** 3831-3840.

Planes, F. J. & Beasley, J. E. (2008) A critical examination of stoichiometric and path-finding approaches to metabolic pathways. Brief. Bioinformatics **9:** 422-436.

Polosina, Y. Y. & Cupples, C. G. (2010) MutL: conducting the cell's response to mismatched and misaligned DNA. Bioessays **32:** 51-59.

Pomerantz, R. T., O'Donnell, M. (2010) Direct restart of a replication fork stalled by a head-on RNA polymerase. Science **327:** 590-592.

Pybus, C., Pedraza-Reyes, M., Ross, C. A., Martin, H., Ona, K., *et al.* (2010) Transcription-associated mutation in *Bacillus subtilis* cells under stress. J. Bacteriol. **192:** 3321-3328.

Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. J. Mol. Biol. **7:** 95–99.

Rehm, N. & Burkovski A. (2010) Engineering of nitrogen metabolism and its regulation in *Corynebacterium glutamicum*: influence on amino acid pools and production. Appl. Microbiol. Biotechnol. in press (PMID: 20922371).

Reitzer, L. J. (1996) Ammonia assimilation and the biosynthesis of glutamine, glutamate, aspartate, asparagine, L-alanine, and D-alanine, p. 391-407. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. American Society for Microbiology, Washington, D.C.

Reitzer, L. (2003) Nitrogen assimilation and global regulation in Escherichia coli. Annu. Rev. Microbiol. **57:** 155–176.

Rietkötter, E., Hoyer, D. & Mascher, T. (2008) Bacitracin sensing in *Bacillus subtilis*. Mol. Microbiol. **68:** 768-785.

Roberts, J. & Park J. S. (2004) Mfd, the bacterial repair coupling factor: translocation, repair and termination. Curr. Opin. Microbiol. **7:** 120-125

Robleto, E. A., Yasbin, R., Ross, C. & Pedraza-Reyes, M. (2007) Stationary phase mutagenesis in *B. subtilis*: a paradigm to study genetic diversity programs in cells under stress. Crit. Rev. Biochem. Mol. Biol. **42:** 327-339

Ross. C., Pybus, C., Pedraza-Reyes, M., Sung, H. M., Yasbin, R. E. & Robleto, E. (2006) Novel role of mfd: effects on stationary-phase mutagenesis in *Bacillus subtilis*. J. Bacteriol. **188:** 7512-7520.

Rossmann, M. G., Moras, D. & Olsen, K. W. (1974) Chemical and biological evolution of nucleotide-binding protein. Nature **250:** 194–199.

Rossolillo, P. & Albertini, A. M. (2001) Functional analysis of the *Bacillus subtilis yshD* gene, a *mutS* paralogue. Mol. Gen. Genet. **264:** 809-818.

Roth, J. R., Kugelberg, E., Reams, A. B., Kofoid, E. & Andersson, D. I. (2006) Origin of mutations under selection: the adaptive mutation controversy. Annu. Rev. Microbiol. **60:** 477-501.

Ruzal, S. M. & Sanchez-Rivas, C. (2003) Effect of glutamate synthase (GOGAT) activity on *Bacillus subtilis* spore properties. Curr. Microbiol. **47:** 208-213.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) Molecular cloning: a laboratory manual, 2nd edit. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sancar, A. (1994) Mechanisms of DNA excision repair. Science **266:** 1954-1956.

Sano, C. (2009) History of glutamate production. Am. J. Clin. Nutr. **90:** 728S-732S

Schilling, O., Frick, O., Herzberg, C., Ehrenreich, A., Heinzle, E., *et al.* (2007) Transcriptional and metabolic responses of *Bacillus subtilis* to the availability of organic acids: transcription regulation is important but not sufficient to account for metabolic adaptation. Appl. Environ. Microbiol. **73:** 499-507.

Schirmer, F., Ehrt, S. & Hillen, W (1997) Expression, inducer spectrum domain structure, and function of MopR, the regulator of phenol degradation in *Acinetobacter calcoaceticus* NCIB8250. J. Bacteriol. **179**: 1329-1336.

Schmalisch, M. H., Bachem, S. & Stülke, J. (2003) Control of the *Bacillus subtilis* antiterminator protein GlcT by phosphorylation: Elucidation of the phosphorylation chain leading to inactivation of GlcT. J. Biol. Chem. **278:** 51108–51115.

Schreier, H. J., Brown, S. W., Hirschi, K. D., Nomellini, J. F. & Sonenshein, A. L. (1989) Regulation of *Bacillus subtilis* glutamine synthetase gene expression by the product of the *glnR* gene. J. Mol. Biol. **210:** 51-63.

Selby, C. P. & Sancar, A. (1993) Molecular mechanism of transcription-repair coupling. Science **260:** 53-58

Selby, C. P. & Sancar, A. (1994) Mechanisms of transcription-repair coupling and mutation frequency decline. Microbiol. Rev. **58:** 317-329.

Selby, C. P. & Sancar, A. (1995a) Structure and function of transcription-repair coupling factor. I. Structural domains and binding properties. J. Biol. Chem. **270:** 4882-4889.

Selby, C. P. & Sancar, A. (1995b) Structure and function of transcription-repair coupling factor. II. Catalytic properties. J. Biol. Chem. **270:** 4890-4895.

Sharkey, M. A. & Engel, P. C. (2008) Apparent negative co-operativity and substrate inhibition in overexpressed glutamate dehydrogenase from *Escherichia coli*. FEMS Microbiol. Lett. **281:** 132–139.

Smith, B. T., Grossman, A. D. & Walker, G. C. (2001) Visualization of mismatch repair in bacterial cells. Mol. Cell **8:** 1197-1206.

Smith, B. T., Grossman, A. D. & Walker, G. C. (2002) Localization of UvrA and effect of DNA damage on the chromosome of *Bacillus subtilis*. J. Bacteriol. **184:** 488-493.

Sonenshein, A. L. (2007) Control of key metabolic intersections in *Bacillus subtilis*. Nat. Rev. Microbiol. **5:** 917-927.

Spies, M. A., Reese, J. G., Dodd, D., Pankow, K. L., Blanke, S. R. & Baudry, J. (2009) Determinants of catalytic power and ligand binding in glutamate racemase. J. Am. Chem. Soc. **131:** 5274-5284.

Stillman, T. J., Baker, P. J., Britton, K. L. & Rice, D. W. (1993) Conformational flexibility in glutamate dehydrogenase. Role of water in substrate recognition and catalysis. J. Mol. Biol. **234:** 1131–1139.

Stillman, T. J., Migueis, A. M., Wang, X. G., Baker, P. J., Britton, K. L., *et al.* (1999) Insights into the mechanism of domain closure and substrate specificity of glutamate dehydrogenase from *Clostridium symbiosum*. J. Mol. Biol. **285:** 875–885.

Strauch, M. A., Aronson, A. I., Brown, S. W., Schreier, H. J. & Sonenshein, A. L. (1988) Sequence of the *Bacillus subtilis* glutamine synthetase gene region. Gene **71:** 257-265.

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.

Sun, D. X. & Setlow, P. (1991) Cloning, nucleotide sequence, and expression of the *Bacillus subtilis ans* operon, which codes for L-asparaginase and L-aspartase. J. Bacteriol. **173:** 3831-3845.

Sun, D. X. & Setlow, P. (1993) Cloning and nucleotide sequence of the *Bacillus subtilis ansR* gene, which encodes a repressor for the *ans* operon coding for L-asparaginase and L-aspartase. J. Bacteriol. **175:** 2501-2506

Sung, H. M. & Yasbin, R. E. (2002) Adaptive, or stationary-phase, mutagenesis, a component of bacterial differentiation in *Bacillus subtilis*. J. Bacteriol. **184:** 5641-5653.

Tamburini, E. & Mastromei, G. (2000) Do bacterial cryptic genes really exist? Res. Microbiol. **151:** 179-182.

Tholen, S. (2008) Charakterisierung des *gudB*-Gens aus *Bacillus subtilis.* Diploma thesis. Georg-August-Universität Göttingen; Germany

Tiwari, A., Talcott, C., Knapp, M., Lincoln, P. & Laderoute, K. (2007) Analyzing Pathways Using SAT-Based Approaches. In Algebraic Biology Springer Verlag Berlin/Heidelberg: 155-169.

Truglio, J. J., Croteau, D. L., Van Houten, B. & Kisker, C. (2006) Prokaryotic nucleotide excision repair: the UvrABC system. Chem. Rev. **106:** 233-252.

Vagin, A. A. & Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. J. Appl. Crystallogr. **30:** 1022–1025.

Veening, J-W., Smits, W. K., Hamoen, L. W., Jongbloed, J. D. & Kuipers, O. P. (2004) Visualization of differential gene expression by improved cyan fluorescent protein and yellow fluorescent protein production in *Bacillus subtilis.* Appl. Environ. Microbiol. **70:** 6809-6815.

Verstee, S., Mogk, A. & Schumann, W. (1999) The Bacillus subtilis *htpG* gene is not involved in thermal stress management. Mol. Gen. Genet. **1261:** 582-588.

Viguera, E., Canceill, D. & Ehrlich, S. D. (2001) Replication slippage involves DNA polymerase pausing and dissociation. EMBO J. **20:** 2587-2595.

Wach, A. (1996) PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in *Saccharomyces cerevisiae*. Yeast **12:** 259-265.

Wacker, I., Ludwig, H., Reif, I., Blencke, H. M., Detsch, C. & Stülke, J. (2003) The regulatory link between carbon and nitrogen metabolism in *Bacillus subtilis*: regulation of the *gltAB* operon by the catabolite control protein CcpA. Microbiology **149:** 3001-3009.

Whatmore, A. M., Chudek, J. A. & Reed, R. H. (1990) The effects of osmotic upshock on the intracellular solute pools of *Bacillus subtilis*. J. Gen. Microbiol. **136:** 2527-2535.

Williams, A. N. & Woessner, K. M. (2009) Monosodium glutamate 'allergy': menace or myth? Clin. Exp. Allergy **39:** 640-646.

Witkin, E. M. (1956) Time, temperature, and protein synthesis: a study of ultraviolet-induced mutation in bacteria. Cold Spring Harb. Symp. Quant. Biol. **21:** 123-140.

Witkin, E. M. (1966) Radiation-induced mutations and their repair. Science **152:** 1345-1353.

Wray, L. V., Jr., Atkinson, M. R. & Fisher, S. H. (1994) The nitrogen-regulated *Bacillus subtilis nrgAB* operon encodes a membrane protein and a protein highly similar to the *Escherichia coli glnB*-encoded PII protein. J. Bacteriol. **176:** 108–114.

Wray, L. V., Jr., Ferson, A. E., Rohrer, K. & Fisher, S. H. (1996) TnrA, a transcription factor required for global nitrogen regulation in *Bacillus subtilis*. Proc. Natl Acad. Sci. U.S.A. **93:** 8841–8845.

Wray, L. V., Jr., Zalieckas, J. M. & Fisher, S. H. (2001) *Bacillus subtilis* glutamine synthetase controls gene expression through protein–protein interaction with transcription factor TnrA. Cell **107:** 427–435.

Wray, L. V., Jr. & Fisher, S. H. (2005) A feedback-resistant mutant of *Bacillus subtilis* glutamine synthetase with pleiotropic defects in nitrogen-regulated gene expression. J. Biol. Chem. **280:** 33298–33304.

Wray, L. V., Jr. & Fisher, S. H. (2010) Functional roles of the conserved Glu304 loop of *Bacillus subtilis* Glutamine Synthetase. . J. Bacteriol. **92:** 5018-5025.

Yan, D. (2007) Protection of the glutamate pool concentration in enteric bacteria. Proc. Natl Acad. Sci. U.S.A. **104:** 9475–9480.

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

Yip, K. S., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., *et al.* (1995) The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. Structure **3:** 1147–1158.

Yoshida, K., Yamaguchi, H., Kinehara, M., Ohki, Y. H., Nakaura, Y. & Fujita, Y. (2003) Identification of additional TnrA-regulated genes of *Bacillus subtilis* associated with a TnrA box. Mol. Microbiol. **49:** 157–165.

Yuan, J., Doucette, C. D., Fowler, W. U., Feng, X. J., Piazza, M., *et al.* (2009) Metabolomics-driven quantitative analysis of ammonia assimilation in *E. coli*. Mol. Syst. Biol. **5:** 302.

Zalieckas, J. M., Wray, L. V., Jr., Ferson, A. E. & Fisher, S. H. (1998) Transcription–repair coupling factor is involved in carbon catabolite repression of the *Bacillus subtilis hut* and *gnt* operons. Mol. Microbiol. **27:** 1031-1038.

Zeigler, D. R., Prágai, Z., Rodriguez, S., Chevreux, B., Muffler, A., *et al*. (2008) The origins of 168, W23 and other *Bacillus subtilis* legacy strains. J. Bacteriol. **190:** 6983–6995.

Zucker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31:** 3406-3415.

9.   **Appendix**

9.1.   **Oligonucleotides**

Oligonucleotides were purchased from Eurofins MWG Operon (Ebersberg, Germany) and Sigma Aldrich (Munich, Germany).

Tab. S1 - Oligonucleotides

Name	Sequence (5' → 3')[a-d]	Description
CD13	5' AAACATATGGCTAGCTGGAGCCACCCGCAGTTC	sequencing of pGP172 constructs, fwd
FC146	5' CGATGCGTTCGCGATCCAGGC	sequencing of pUT18 constructs
FC147	5' CCAGCCTGATGCGATTGCTGCAT	sequencing of p25-N constructs
FC148	5' GTCACCCGGATTGCGGCGG	sequencing of pUT18C constructs
FC149	5' GCTGGCTTAACTATGCGGCATCAGA	sequencing of pUT18C constructs,
FC150	5' GATTCGGTGACCGATTACCTGGC	sequencing of pKT25 constructs
FC151	5'CGCCAGGGTTTTCCCAGTCACG	sequencing of pKT25 constructs
HL47	5' *CTAATACGACTCACTATAGGGAGA*GTCATAGCGGGACATATAATG	*gapA* probe, rev, contains T7 promoter
HL48	5' TATTCGAAAGAACCAAGTCAG	*gapA* probe, fwd
HMB 74	5' CGCGGATCCAGAAGATTCTAGGAGGTTAAC	sequencing of *gudB*
IR3	5' CGCGGATCCACGCCGATTCAGAAGACGA	sequencing of *gudB*
IR6	5' AAACGCGTCGACCTCCGATTTCGAGCTTTTC A	sequencing of *gudB*
JL7	5' TTT<u>GGATCC</u>GTTCCCCCAGTTGTCAACATCAATAATTTTT	*ansA-lacZ* fusion, rev (BamHI)

Name	Sequence (5' → 3')[a-d]	Description
JS39	5' TCTATCAACAGGAGTCCAAGC	Sequencing of pWH844 constructs, rev
kan-fwd	5' CAGCGAACCATTTGAGGTGATAGG	LFH-PCR, amplification of *aphA3* from pDG780, fwd
kan-rev	5' CGATACAAATTCCTCGTAGGCGCTCGG	LFH-PCR, amplification of *aphA3* from pDG780, rev
kan-check rev	5' CTGCCTCCTCATCCTCTTCATCC	LFH-PCR, sequencing of the up-fragment
kan-check fwd	5' CATCCGCAACTGTCCATACTCTG	LFH-PCR, sequencing of the down-fragment
KG3	5'AAAGGATCCCTGGTTCCGCGTGGTTCCATGTCAGCAAAGCAAGTCTCGAAAGATGAAG	cloning of *rocG* into pWH844, fwd (BamHI) with thrombin cleavage site
KG4	5' TTTAAGCTTTCATTAGACCCATCCGCGGAAACGCGAT	cloning of *rocG* into pWH844, rev (HindII)
KG5	5' ATCAATCGAAGAGCAGAGGCATCTTCG	LFH-PCR *rocG* (fwd up-fragment)
KG6	5'CCTATCACCTCAAATGGTTCGCTGGCTTCCTTAATGATTGTTTGGGTAGACAG	LFH-PCR *rocG* (rev up-fragment)
KG7	5' CCGAGCGCCTACGAGGAATTTGTATCGGCGTTTGGCGGCTTACATGACGG	LFH-PCR *rocG* (fwd down-fragment)
KG8	5' GGATGCAGCAAGGTCAAGATCAGCG	LFH-PCR *rocG* (rev down-fragment)
KG9	5' CCGTGGGATGATGAGACGATCGG	LFH-PCR *rocG* (fwd up-fragment, sequencing)

Name	Sequence (5' → 3')[a-d]	Description
KG10	5' GCGGGAGCCCGCAGAACACT	LFH-PCR *rocG* (rev down-fragment, sequencing)
KG11	5' CGCGAGCTTGACGGGGGAG	amplification of *ansR* region
KG12	5' CCTATCACCTCAAATGGTTCGCTGGATGG CC AGCCGCTGAGTGAAG	LFH-PCR *ansR* (rev up-fragment)
KG13	5'CCGAGCGCCTACGAGGAATTTGTATCGCC GAGAAGGTCAGCTGTATATTGAAGC	LFH-PCR *ansR* (fwd down-fragment)
KG14	5' ACCTCGTAAATGCTCATGTCTTCGCC	LFH-PCR *ansR* (fwd up-fragment)
KG15	5' CCGGAAGTCATTCTAGAGCTTGAGGA	LFH-PCR *ansR* (rev down-fragment)
KG16	5' CTAGCGCCCACATCAATTTTGGCAC	LFH-PCR *ansR* (fwd up-fragment, sequencing)
KG17	5' GAAAGCCGGAGGAGGAGGAACC	LFH-PCR *ansR* (rev down-fragment, sequencing)
KG18	5'AAA<u>GGATCC</u>CAGCTCAAGGTGAAAAAGG AGCGGAA	cloning of *ansR* into pBQ200, fwd (BamHI)
KG19	5'TTT<u>GTCGAC</u>TCATTAACTCAGTTCCTCCTG TACTTTTCTTTTGTG	cloning of *ansR* into pBQ200, rev (SalI)
KG20	5'AAA<u>GAGCTC</u>GATGGCAGCCGATCGAAAC ACCG	cloning of *gudB* into pGP172, fwd (SacI)
KG21	5'TTT<u>GGATCC</u>TCATTATATCCAGCCTCTAAA ACGCGAAGCTT	cloning of *gudB* into pGP172, rev (BamHI)

Name	Sequence (5' → 3')[a-d]	Description
KG22	5' AGAAAAGGTTCGCGCTCGGCAC	sequencing of *fur*
KG23	5' CCTTATTCTTATTTATATGAGCAGGACG	sequencing of *fsrA*, fwd
KG24	5' TTTGCAGCCGTGTGATGATACCC	sequencing of *fsrA*, rev
KG25	5' TTGAAGGGGAAAATGGGCTG	RT PCR *ansA*, fwd
KG26	5' CTATTTCCACCCAGTATTCAGG	RT PCR *ansA*, rev
KG27	5' AAAGAATTCGCTGGCCATCAACAGAAATG	translational *ansA-lacZ* fusion, fwd (EcoRI)
KG28	5' ATGGCTTGGACCCGTTATTGGGG	LFH-PCR *ansAB* (fwd up-fragment)
KG29	5' CCTATCACCTCAAATGGTTCGCTGGAGCCA GCCCATTTTCCCCTTC	LFH-PCR *ansAB* (rev up-fragment)
KG30	5' CCGAGCGCCTACGAGGAATTTGTATCGCGGCGCTGATCATCTTGTTGATG	LFH-PCR *ansAB* (fwd down-fragment)
KG31	5' AAGTCGGCACAACGCCTCCGG	LFH-PCR *ansAB* (rev down-fragment)
KG32	5' GCGCATTCGTTGGGGAAAATCGG	LFH-PCR *ansAB* (fwd up-fragment, sequencing)
KG33	5' GCCGGCAGTTCCGACTGTTCC	LFH-PCR *ansAB* (rev down-fragment, sequencing)
KG34	5'AAATCTAGAGATGGAGGTGCGCCAACTGCGTTAT	cloning of *gltC* into B2H vectors, fwd (XbaI)
KG35	5'TTTGGTACCCGTTGATACTGCTCCAGCTTAGAGAAAAATTG	cloning of *gltC* into B2H vectors, rev (KpnI)

Name	Sequence (5' → 3')[a-d]	Description
KG36	5' AAA<u>TCTAGA</u>GATGTCAGCAAAGCAAGTC TCGAAAGATGAAG	cloning of *rocG* into B2H vectors, fwd (XbaI)
KG37	5' TTT<u>GGTACC</u>CGGACCCATCCGCGGAAAC GCGAT	cloning of *rocG* into B2H vectors, rev (KpnI)
KG38	5' CCGTGTCGCATTAACACC	RT PCR *ald*, fwd
KG39	5' ACCTGCTTCGGATCAGCA	RT PCR *ald*, rev
KG40	5' TAAACCTTGGCGGCGGAA	RT PCR *bcd*, fwd
KG41	5' CCATATCCTCGACCGTTG	RT PCR *bcd*, rev
KG42	5` GAAACGGCAAAACGTTCTGG	RT PCR *rpsJ*, fwd
KG43	5` GTGTTGGGTTCACAATGTCG	RT PCR *rpsJ*, rev
KG44	5`GCGTCGTATTGACCCAAGC	RT PCR *rpsE*, fwd
KG45	5` TACCAGTACCGAATCCTACG	RT PCR *rpsE*, rev
KG46	5' AA<u>GAATTC</u>GATAAACCCAGCGAACCATT TG	*aphA3* from pDG780, fwd (EcoRI)
KG47	5' TTT<u>CCCGGG</u>ATCGATACAAATTCCTCGTA GGC	*aphA3* from pDG780, rev (SmaI)
KG48	5' AAA<u>GAATTC</u>GATCCTTTAACTCTGGCAAC CC	*ermC* from pDG647, fwd (EcoRI)
KG49	5' TTT<u>CCCGGG</u>GCCGACTGCGCAAAAGACA TAAT	*ermC* from pDG647, rev (SmaI)
KG50	5' TTT<u>ATCGATGCGGCCGC</u>AATGGTTTCTTA GACGTCAGGTG	amplification of *bla* and *ori* from pUC18, rev (ClaI, NotI)
KG51	5' AAA<u>GAATTC</u>GCTGTTTCCTGTGTGAAATT GTTAT	amplification of *bla* and *ori* from pUC18, fwd (EcoRI)
KG52	5' GGTTATTGTCTCATGAGCGGAT	sequencing of pGP882, fwd in *bla* gene
KG53	5' GCGCAACGCAATTAATGTGAG	sequencing of pGP882, rev in *ori*

Name	Sequence (5' → 3')[a-d]	Description
KG54	5' AAA<u>ATCGAT</u>CGTCGGTCTATTCAATTTAG TGAAT	amplification *lacA*5', fwd (ClaI)
KG55	5' TTT<u>GTCGACGGATCC</u>AGCTGATGGCATC GACATGCTT	amplification *lacA*5', rev (BamHI, SalI)
KG56	5' AAA<u>GTCGACCCCGGG</u>CACCATATAAACT GCTGATCGTC	amplification *lacA*3', fwd (SalI, SmaI)
KG57	5' TTT<u>CAATTG</u>ATCATACGGGTGATTCCAGA TG	amplification *lacA*3' rev (MfeI)
KG58	5' AAA<u>AGATCT</u>ATCTTACATTGTAATCATGT CCAGAAAATGATC	amplification *xylR,* fwd (BglII)
KG59	5' TTT<u>CAATTGCCCGGGGTCGACGGATCC</u>A T GATTGTTTCCTCCTTTCAGATGCATTTTATTC ATATAGTAAGTAC	amplification P*xylA* contains *gapA* SD-sequence and ATG, rev (BamHI, SalI, SmaI, MfeI)
KG60	5' AAA<u>AGATCTTCTAGAGGATCCGGTACCG AATTC</u>AGGTGGATCAGGCTCGGGATCTGGT TCAATGGCCGACAAGGAGAAGAACG	amplification *yfP* C-terminal fragment, fwd (BglII, XbaI, BamHI, KpnI, EcoRI)
KG61	5' TTT<u>GTCGAC</u>TTATCACTTGTACAGCTCGT CCATGCC	amplification *yfP* C-terminal fragment, rev (SalI)
KG62	5' AAA<u>AGATCT</u>CATGGCAGTAAAAGTCGGT ATTAACGGTTTTGGTCGTATGGCCGACAAG GAGAAGAACG	amplification *yfP* C-terminal fragment, contains *gapA* SD-sequence, fwd (BglII)
KG63	5' TTT<u>GTCGAC</u>TCATTAT<u>GAATTCGGTACCG GATCCTCTAGA</u>GAACCAGATCCCGAGCCTG ATCCACCTGACTTGTACAGCTCGTCCATGCC	amplification *yfP* C-terminal fragment, rev, (SalI, XbaI, BamHI, KpnI, EcoRI)

Name	Sequence (5' → 3')[a-d]	Description
KG64	5' TATCAGGGCCTCGACTACA	sequencing of pGP882 derivatives, fwd in *lacA5'*
KG65	5' CGCTGATTAAATACAGCATCGG	sequencing of pGP882 derivatives, rev *lacA3'*
KG66	5' GAAAATACTGACGAGGTTATATAAGATGA	sequencing of pGP884, pGP885 derivatives, fwd in *PxylA*
KG67	5' CTCGATGTTGTGGCGGATCTT	sequencing of pGP888, rev in *yfp* C-terminal fragment
KG68	5' ATCACATGGTCCTGCTGGAGT	sequencing of pGP889, fwd in *yfp*-C- terminal fragment
KG69	5'AAA<u>AGATCT</u>AGCGATATCCACTTCATCCACT	amplification P*xylA*, fwd (BglII)
KG70	5' AAA<u>ATCGAT</u>AGCTGTACGGAGAGGACATTAT	amplification *xkdE5'*, fwd (ClaI)
KG71	5' TTT<u>GTCGACGGATCC</u>TTTCTAGCAGTATCCGCTGTC	amplification *xkdE5'*, rev (BamHI, SalI)
KG72	5' AAA<u>GTCGACCCCGGG</u>TCAGGCTGACATTAAAAGGACC	amplification *xkdE3'*, fwd (SalI, SmaI)
KG73	5' TTT<u>CAATTG</u>CAAGCATGTCTTCAACGAGCTT	amplification *xkdE3'*, rev (MfeI)
KG74	5' AAA<u>AGATCTTCTAGAGGATCCGGTACCGAATTC</u>AGGTGGATCAGGCTCGGGATCTGGTTCAATGGTGAGCAAGGGCGAGGA	amplification *yfP* N-terminal fragment, fwd (BglII, XbaI, BamHI, KpnI, EcoRI)
KG75	5' TTT<u>GTCGAC</u>TTATCACATGATATAGACGTTGTGGCTGTTG	amplification *yfP* N-terminal fragment, rev (SalI)

Name	Sequence (5' → 3')[a-d]	Description
KG76	5' AAA<u>AGATCT</u>CATGGCAGTAAAAGTCGGTATTAACGGTTTTGGTCGTATGGTGAGCAAGGGCGAGGA	amplification *yfP* C-terminal fragment, contains *gapA* SD-sequence, fwd (BglII)
KG77	5' TTT<u>GTCGAC</u>TCATTAT<u>GAATTCGGTACCGGATCCTCTAGA</u>GAACCAGATCCCGAGCCTGATCCACCTGACATGATATAGACGTTGTGGCTGTTG	amplification *yfP* N-terminal fragment, rev, (SalI, XbaI, BamHI, KpnI, EcoRI)
KG78	5' GCAGAAATAAGCTGCGTTCC	sequencing of pGP883 derivatives, fwd in *xkdE5'*
KG79	5' GAGCACCTTCGCAATCTCAA	sequencing of pGP883 derivatives, rev *xkdE3'*
KG80	5' TCCAGCTCGACCAGGATGG	sequencing of pGP886, rev in *yfp* N- terminal fragment
KG81	5' GGGCACAAGCTGGAGTACAA	sequencing of pGP887, fwd in *yfp* N-terminal fragment
KG82	5' CCTTGGCGTCAAAGACCACG	sequencing of *taq*, rev
KG83	5' ACGACGAGCTGGTCCTCGAG	sequencing of *taq*, fwd
KG84	5' AAA<u>AGATCT</u>ATGGAATTCGGGATGCTGCCC	cloning of *taq* into pWH844, fwd (BglII)
KG85	5' TTT<u>GTCGAC</u>CTATCACTCCTTGGCGGAGAG	cloning of *taq* into pWH844, rev (SalI)
KG86	5' AGGATTCGCCATGCTTGTGA	LFH-PCR *mfd* (fwd up-fragment)
KG87	5' CCTATCACCTCAAATGGTTCGCTGCCTCGTGTAAACCGTTGATG	LFH-PCR *mfd* (rev up-fragment)
KG88	5' CCGAGCGCCTACGAGGAATTTGTATCGGGCATGCTGAAGGGCTTAAA	LFH-PCR *mfd* (fwd down-fragment)

Name	Sequence (5' → 3')[a-d]	Description
KG89	5' ATCCAACATCGCCGACAATG	LFH-PCR *mfd* (rev down-fragment)
KG90	5' GCGGTTACGTCTTTTGTGCT	LFH-PCR *mfd* (fwd up-fragment, sequencing)
KG91	5' GAAATCCTGATGTGGACAGCA	LFH-PCR *mfd* (rev down-fragment, sequencing)
KG92	5' TTT<u>GGATCC</u>TCATTATATCCAGCCTCTAA AACGCG	cloning of *gudB* into pAC5 (BamHI)
KG93	5'AAA<u>GGATCC</u>CTCGCAGGAAACAGAGAGG A	cloning of *mfd* into pBQ200, fwd (BamHI)
KG94	5' TTT<u>GTCGAC</u>TCATTACGTTGATGAAATG GTTTGCTTTTTC	cloning of *mfd* into pBQ200, rev (SalI)
KG95	5' GTGGATTCAATCCGCAGCTT	sequencing of *mfd*, fwd (base 600)
KG96	5' GGAACGTACGCAGAAATTGTC	sequencing of *mfd*, fwd (base 1251)
KG97	5' GCTCCGTTCTATTCACGAAATC	sequencing of *mfd*, fwd (base 1884)
KG98	5' GTAGAGGACATTGAGCGGAA	sequencing of *mfd*, fwd (base 2524)
KG99	5' TTCGCTGTGTCTCCTTTACG	sequencing of *mfd*, rev (base 3045)
KG100	5' GCAGCAATAACACCGGCAATAA	LFH-PCR *gudB* (fwd up-fragment)
KG101	5' CCTATCACCTCAAATGGTTCGCTGGCTG GATATAAGTTGATGATTTGCAT	LFH-PCR *gudB* (rev up-fragment)
KG102	5' CCGAGCGCCTACGAGGAATTTGTATCGG CTGAAATCAGCAGTTTTCCGT	LFH-PCR *gudB* (fwd down-fragment)
KG103	5' GCCATAATCCGGAGATTCATG	LFH-PCR *gudB* (rev down-fragment)

Name	Sequence (5' → 3')[a-d]	Description
KG104	5' CGATTTCCGCTGCGATATGC	LFH-PCR *gudB* (fwd up-fragment, sequencing)
KG105	5' GGTTGATGATATCAGGATGGAG	LFH-PCR *gudB* (rev down-fragment, sequencing)
KG106	5' AAA<u>GGATCC</u>ATGTGGGAAAGTAAATTTT CA AAAGAAGG	cloning of *guaB* into pGP380, fwd (BamHI)
KG107	5' TTT<u>CTGCAG</u>TCATTATGAAATTGTATAGT TAGGTGATTCTTTTG	cloning of *guaB* into pGP380, rev (PstI)
KG108	5' AAA<u>AGATCT</u>GCGGTAGTACAGTAAGGAA GGGGA	cloning of *gltB* into pGP1460 fwd, contains SD-sequence (BglII)
KG109	5' TTT<u>GTCGACGGATCC</u>TCATTACGGAAGA ACTGAACTCCCCATCAA	cloning of *gltB* into pGP1460, rev, contains stop codon (BamHI, SalI)
KG110	5' AAA<u>GGATCC</u>ATGGTTACATAAAAAGGGA GGCTGAGAG	cloning of *yerD* into pGP1460, fwd, contains SD-sequence (BamHI)
KG111	5' TTT<u>CTGCAG</u>TCATTATGAAACAAATTGAT GGATCAGGTTGTCG	cloning of *yerD* into pGP1460, rev, contains stop codon (PstlI)
KG112	5' TTT<u>GTCGAC</u>TGAAACAAATTGATGGATC AGGTTGTCGATA	cloning of *yerD* into pGP1331 (SalI)
KG113	5' GCGAGAAGCCATTTTATGTGAAGG	sequencing of *yerD*
KG114	5' AGTCTCGAAATATCAGCCCCAG	LFH-PCR *uvrAB* (fwd up-fragment)
KG115	5' CCTATCACCTCAAATGGTTCGCTGAAGC GTATACGGCGTTGAACCT	LFH-PCR *uvrAB* (rev up-fragment)

Name	Sequence (5' → 3')[a-d]	Description
KG116	5' CCGAGCGCCTACGAGGAATTTGTATCGA AG TGAATGCCAAGTTCCCGCT	LFH-PCR *uvrAB* (fwd down-fragment)
KG117	5' CAAGAGCGATTTCATGCGTGTT	LFH-PCR *uvrAB* (rev down-fragment)
KG118	5' GCGGCTAAGAAGAGAGGCATCGATA	*gudB* probe, fwd
KG119	5' P-GAACGTAACAGAAAAGAGGT**T**AAGG C**T**GTGAAGGCGCTTTCAATTTGG	mutagenesis: *gudB_{mut1}*
KG120	5' P-GAACGTAACAGAAAAGAGGT**T**AAGG C**T**GT**T**AAGGC**T**CTTTCAATTTGGATGAGTTT AAAATGC	mutagenesis: *gudB_{mut2}*
KG121	5' TATCCAGCCTCTAAAACGCGAAGCTTC	amplification of *gudB*, rev, without stop codon for C-terminal fusion to *yfp*
KG122	5' GAAGCTTCGCGTTTTAGAGGCTGGATA TGGTGAGCAAGGGCGAGGAGCT	amplification of *yfp*, fwd, for N-terminal fusion to *gudB* (contains complementary sequence to *gudB*)
KG123	5' *CTAATACGACTCACTATAGGGAGA*AACGCC G ACCATATATGCAGCGAG	*gudB* probe, rev, contains T7 promoter
KG124	5' AAA<u>GGATCC</u>GGGCATCCTGAAGAATTGC ATGAGC	cloning of *yerD* into pGP1331, fwd (BamHI)
KG125	5' <u>AAAGGATCC</u>GCAAAAGTTGTCGGCATCT CAGATGC	cloning of *gudB* into pGP1331, fwd (BamHI)
KG126	5' TTT<u>GTCGAC</u>TATCCAGCCTCTAAAACGC GAAGCTTC	cloning of *gudB* into pGP1331, rev (SalI)
KG127	5' TCTAAACACGGTGCCTTTACAGGCC	LFH-PCR *mutSL* (fwd up-fragment)

Name	Sequence (5' → 3')[a-d]	Description
KG128	5' CCTATCACCTCAAATGGTTCGCTGGGCAT CCTGGTGCTCTGCCTTT	LFH-PCR *mutSL* (rev up-fragment)
KG129	5' CCGAGCGCCTACGAGGAATTTGTATCGG CACATCAGACCCATTTACATGCCC	LFH-PCR *mutSL* (fwd down-fragment)
KG130	5' GGCTGATTCAAGGGTGCTGTTTGTAT	LFH-PCR *mutSL* (rev down-fragment)
KG131	5' CATCACATCAACCGGAGGCGAC	LFH-PCR *mutSL* (fwd up-fragment, sequencing)
KG132	5' TCCCCAGTGGAAAGGGCCTTTTTTG	LFH-PCR *mutSL* (rev down-fragment, sequencing)
KG133	5' P-CAGAAAAAGAGGTGAAGGCGGT**T**AAG GC**T**CTTTCAATTTGGATGAGTTT	mutagenesis: *gudB**mut3*
M13 pUC fwd	5'GTAAAACGACGGCCAGTG-	sequencing of pUC derivatives, fwd
M13 pUC rev	5' GGAAACAGCTATGACCATG	sequencing of pUC derivatives, rev
mls fwd (kan)	5' CAGCGAACCATTTGAGGTGATAGGGATC CTTTAACTCTGGCAACCCTC	LFH-PCR, amplification of *ermC* from pDG647, fwd
mls rev (kan)	5' CGATACAAATTCCTCGTAGGCGCTCGGG CCGACTGCGCAAAAGACATAATCG	LFH-PCR, amplification of *ermC* from pDG647, rev
mls-check rev	5'GTTTTGGTCGTAGAGCACACGG	LFH-PCR, sequencing of the up-fragment
mls-check fwd	5'CCTTAAAACATGCAGGAATTGACG	LFH-PCR, sequencing of the down-fragment

Name	Sequence (5' → 3')[a-d]	Description
NP20	5' GCAGCAGCCAACTCAGCTTCCTTTCGGGC	sequencing of pGP172 constructs, rev
PAC 5F	5' GCGTAGCGAAAAATCCTTTTC	sequencing of pAC constructs, fwd
PAC 5R	5' CTGCAAGCGATAAGTTGG	sequencing of pAC constructs, fwd
PT05	5' AAACGAGCTCGATGTCAGCAAAGCAAGT CTCG	cloning of *rocG* into pGP172, fwd (SacI)
PT06	5' CGCGGATCCTCATTAGACCCATCCGCGGA AAC	cloning of *rocG* into pGP172, rev (BamHI)
PT12	5' CCCAAGCTTTCATTAGACCCATCCGCGGA AAC	cloning of *rocG* into pGP380, fwd (HindIII)
pWH 844 fwd	5' TATGAGAGGATCGCATCACCAT	sequencing of pWH844 constructs, fwd
rpsJ-RT-fwd	5' GAAACGGCAAAACGTTCTGG	RT PCR *rpsJ* fwd
rpsJ-RT-rev	5' GTGTTGGGTTCACAATGTCG	RT PCR *rpsJ* rev
rpsE-RT-fwd	5' GCGTCGTATTGACCCAAGC	RT PCR *rpsE* fwd
rpsE-RT-rev	5' TACCAGTACCGAATCCTACG	RT PCR *rpsE* rev
ST1	5' AAAGAATTCCATTCAGCTTTCAGAAAGC TTACAGCGAATC	*gudB-lacZ* furion, fwd (EcoRI)
ST2	5' AAAGGATCCCCCAATTTTTCCAGAGCCT TATGTATTACG	*gudB-lacZ* furion, rev (BamHI)
ST3	5' AAAGGATCCATGTCAGCAAAGCAAGTCT CGAA	cloning of *rocG* into pGP380, fwd (BamHI)

Name	Sequence (5' → 3')[a-d]	Description
ST4	5' AAA<u>AGATCT</u>CTACTAAAACAATTCATCCAGTAA	construction of the *alf* promoter fusion, amplification of *aphA3* from pDG780, rev (BglII)
ST7	5' <u>AATTC</u>TTGACAAGTGAAGGCGGTGAAGGCGCGCTATGCTATAATACAGCTTGGAAATG	construction of the *alf* promoter fusion, fwd (EcoRI)
ST8	5' <u>GATCC</u>ATTTCCAAGCTGTATTATAGCATAGCGCGCCTTCACCGCCTTCACTTGTCAAG	construction of the *alf* promoter fusion, rev (BamHI)
ST9	5'AAA<u>GAATTCCCGGGGATCC</u>TAATGTTAGAAAAGAGGAAGGAAATAA	construction of the *alf* promoter fusion, amplification of *aphA3* from pDG780, (EcoRI, BamHI)
Tc fwd1 (kan)	5'CAGCGAACCATTTGAGGTGATAGGGCTTAT C AACGTAGTAAGCGTGG	LFH-PCR, amplification of *tet* from pDG1514, fwd
Tc rev (kan)	5' CGATACAAATTCCTCGTAGGCGCTCGGGAACTCTCTCCCAAAGTTGATCCC	LFH-PCR, amplification of *tet* from pDG1514, rev
Tc-check rev	5' CGGCTACATTGGTGGGATACTTGTTG	LFH-PCR, sequencing of the up-fragment
Tc-check fwd	5'CATCGGTCATAAAATCCGTAATGC	LFH-PCR, sequencing of the down-fragment

[a] Restriction sites are underlined.

[b] The "P" at the 5' end of primer sequences indicates phosphorylation.

[c] Base pair exchanges in bold.

[d] T7 promoter sequence in italic.

9.2.   **Plasmids**

Tab. S2 - Plasmids

Name	Relevant characteristics[a]	Used restriction sites	Reference
pAC5	allows translational *laZ* promoter fusions, homologous recombination into the *amyE* site of *B. subtilis* (Cm^R); *E. coli* Amp^R		Martin-Verstraete *et al.*, 1992
pAC6	allows transcriptional *laZ* promoter fusions, contains SD-sequence and ATG of *sacB*, homologous recombination into the *amyE* site of *B. subtilis* (Cm^R); *E. coli* Amp^R		Stülke *et al.*, 1997
pBlueskript SK(-)	cloning vector		Stratagene, Amsterdam, Netherlands
pBQ200	allows the overexpression of proteins in *B. subtilis*; *E. coli* (Amp^R), *B. subtilis* (Erm^R)		Martin-Verstraete *et al.*, 1994
pDG647	template for erythromycin resistance (*erm^C*), *E. coli* (Amp^R)		Guérout-Fleury *et al.*, 1995
pDG780	template for kanamycin resistance (*aphA3*), *E. coli* (Amp^R)		Guérout-Fleury *et al.*, 1995
pDG1514	template for tetracycline resistance (*tet^C*), *E. coli* (Amp^R)		Guérout-Fleury *et al.*, 1995
pGP172	allows expression of proteins carrying a Strep-tag at their N-terminus in *E. coli* BL21(DE3)/pLysS (Amp^R)		Merzbacher *et al.*, 2004

Name	Relevant characteristics[a]	Used restriction sites	Reference
pGP380	Allows overexpression of N-terminal Strep-tag fusion proteins in *B. subtilis* (Erm[R]); *E. coli* (Amp[R]);		Herzberg *et al.*, 2007
pGP526	translational *gltA-lacZ* fusion in pAC7		Wacker *et al.*, 2003
pGP529	pBQ200-*rocG*		Commichau *et al.*, 2008
pGP651	translational *gudB-lacZ* fusion in pAC5 (ST1/ST2)	BamHI/EcoRI	Gunka *et al.* subm.
pGP652	pGP380-*rocG*-SR3 (*G277A*) (ST3/PT12)	BamHI/HindIII	Tholen, 2008
pGP653	promoterless *aphA3* gene in pAC6 (ST9/ST4), allows promoter fusion to a *aphA3-lacZ* operon	BglII/EcoRI	Gunka *et al.* subm.
pGP655	artificial *alf* promoter system in pGP653 (ST7/ST8)	EcoRI/BamHI	Gunka *et al.* subm.
pGP852	pBQ200- *rocG*-SR1 (*T923C*), mutagenesis of pGP529 in XL1-Red		Gunka *et al.*, 2010
pGP853	pBQ200-*rocG*-SR2 (*C702A*) mutagenesis of pGP529 in XL1-Red		Gunka *et al.*, 2010
pGP854	pBQ200-*rocG*-SR3 (*G277A*) mutagenesis of pGP529 in XL1-Red		Gunka *et al.*, 2010
pGP855	pBQ200-*rocG*-SR4 (*G997A*) mutagenesis of pGP529 in XL1-Red		Gunka *et al.*, 2010
pGP856	pGP172-*rocG*-SR1 (*T923C*)	SacI/BamHI	Gunka *et al.*, 2010
pGP857	pGP172-*rocG*-SR2 (*C702A*)	SacI/BamHI	Gunka *et al.*, 2010
pGP858	pGP172-*rocG*-SR3 (*G277A*)	SacI/BamHI	Gunka *et al.*, 2010

Name	Relevant characteristics[a]	Used restriction sites	Reference
pGP859	pGP172-*rocG*-SR4 (*G997A*)	SacI/BamHI	Gunka *et al.*, 2010
pGP860	pWH844-*rocG* (KG3/KG4), contains thrombin cleavage site	BamHI/HindIII	This work
pGP861	pWH844-*rocG*-SR3 (*G277A*) (KG3/KG4), contains thrombin cleavage site	BamHI/HindIII	This work
pGP862	pWH844-*rocG*-MF (*G364A*) (KG3/KG4), contains thrombin cleavage site	BamHI/HindIII	This work
pGP863	pGP172-*gudB* (KG20/KG21)	SacI/BamHI	Gunka *et al.*, 2010
pGP864	pGP172-*gudB1* (KG20/KG21)	SacI/BamHI	Gunka *et al.*, 2010
pGP865	pGP172-*rocG*-MF (*G364A*) (PT5/PT6)	SacI/BamHI	Gunka *et al.*, 2010
pGP866	pGP172-*rocG*-SR5 (*T472C*) (PT5/PT6)	SacI/BamHI	Gunka *et al.*, 2010
pGP867	pBQ200-*rocG*-SR6 (*G907A*) mutagenesis of pGP529 in XL1-Red		Gunka *et al.*, 2010
pGP868	pBQ200-*rocG*-SR7 (*G904A*) mutagenesis of pGP529 in XL1-Red		Gunka *et al.*, 2010
pGP869	pBQ200-*rocG*-SR8 (*A304G*) mutagenesis of pGP529 in XL1-Red		Gunka *et al.*, 2010
pGP870	pBQ200-*rocG*-SR9 (*A479G*) mutagenesis of pGP529 in XL1-Red		Gunka *et al.*, 2010
pGP871	pBQ200-*rocG*-SR10 (*G831AA*) mutagenesis of pGP529 in XL1-Red		Gunka *et al.*, 2010
pGP872	*ansA-lacZ* fusion in pAC5 (KG27/JL7)	BamHI/EcoRI	This work

Name	Relevant characteristics[a]	Used restriction sites	Reference
pGP873	pBQ200-*ansR* (KG18/19)	BamHI/SalI	This work
pGP874	pUT18-*gltC* (KG34/KG35)	XbaI/KpnI	This work
pGP875	pUT18C-*gltC* (KG34/KG35)	XbaI/KpnI	This work
pGP876	p25-N-*gltC* (KG34/KG35)	XbaI/KpnI	This work
pGP877	pKT25-*gltC* (KG34/KG35)	XbaI/KpnI	This work
pGP878	pUT18-*rocG* (KG36/KG37)	XbaI/KpnI	This work
pGP879	pUT18C-*rocG* (KG36/KG37)	XbaI/KpnI	This work
pGP880	p25-N-*rocG* (KG36/KG37)	XbaI/KpnI	This work
pGP881	pKT25-*rocG* (KG36/KG37)	XbaI/KpnI	This work
pGP882	*ori-bla* from pUC18 (KG50/51), *lacA5'* (KG54/KG55), *lacA3'*(KG56/KG56); *E. coli* Amp^R	*	This work
pGP883	*ori-bla* from pUC18 (KG50/51), *xkdE5'*(KG70/71), *xkdE3'*(KG72/KG73); *E. coli* Amp^R	*	This work
pGP884	pGP882-*xylR*+P*xylA* (KG58/KG59) *aphA3* (KG46/47); *E. coli* Amp^R	*	This work
pGP885	pGP883- P*xylA* (KG69/59)-*ermC* (KG48/49); *E. coli* Amp^R	*	This work
pGP886	pGP883-*yfp* N-terminus for C-terminal fusions (KG74/KG75), allows homologous recombination into the *xkdE* site of *B. subtilis* (Erm^R)	BglII, SalI	This work
pGP887	pGP883- *yfp* N-terminus for N-terminal fusions (KG76/KG77), allows homologous recombination into the *xkdE* site of *B. subtilis* (Erm^R)	BglII, SalI	This work

Name	Relevant characteristics[a]	Used restriction sites	Reference
pGP888	pGP882-*yfp* C-terminus for C-terminal fusions (KG60/KG61), allows homologous recombination into the *lacA* site of *B. subtilis* (Kan[R])	BglII, SalI	This work
pGP889	pGP882- *yfp* C-terminus for N-terminal fusions (KG62/KG63), allows homologous recombination into the *lacA* site of *B. subtilis* (Kan[R])	BglII, SalI	This work
pGP890	pGP886-*gltC*	XbaI/KpnI	This work
pGP891	pGP888-*gltC*	XbaI/KpnI	This work
pGP892	pGP887-*gltC*	XbaI/KpnI	This work
pGP893	pGP889-*gltC*	XbaI/KpnI	This work
pGP894	pGP886-*rocG*	XbaI/KpnI	This work
pGP895	pGP888-*rocG*	XbaI/KpnI	This work
pGP896	pGP887-*rocG*	XbaI/KpnI	This work
pGP897	pGP889-*rocG*	XbaI/KpnI	This work
pGP898	pGP884-*yfp* for C-terminal (KG61/KG74) fusions allows homologous recombination into the *lacA* site of *B. subtilis* (Kan[R])	SalI/MfeI	This work
pGP900	*gudB* in pAC5 (ST01/KG92)	EcoRI/BamHI	Gunka *et al.*, subm.
pGP901	pGP380-*guaB* (KG106/107)	(BamHI/PstI)	This work
pGP919	*gltA-lacZ* fusion in pAC5		Commichau *et al.*, 2007b
pGP932	pBQ200-*rocG*-MF (*G364A*) mutagenesis of pGP529 in XL1-Red		Gunka *et al.*, 2010
pGP933	pBQ200-*rocG*-SR5 (*T472C*) mutagenesis of pGP529 in XL1-Red		Gunka *et al.*, 2010

Name	Relevant characteristics[a]	Used restriction sites	Reference
pGP1331	pUS19-3xFLAG, integrate by single crossing-over into *B. subtili*s chromosome; allows fusion to a 3xFLAG at the C-terminus; *E. coli* (Amp[R]), *B. subtilis* (Spec[R])		Lehnik-Habrink *et al.*, 2010
pGP1702	pBluescript (SK-)-rocG-SR3 (*G277A*)	XbaI/KpnI	This work
pGP1703	pBluescript (SK-)-rocG-MF (*G364A*)	XbaI/KpnI	This work
pGP1708	pGP380-*rocG*-MF (*G364A*) (ST3/PT12)	BamHI/HindIII	This work
pGP1709	pGP380-*rocG* (ST3/PT12)	BamHI/HindIII	This work
pGP1714	*gudB* mut1 in pAC5 (ST01/KG119/KG92)	EcoRI/BamHI	Gunka *et al.*, subm.
pGP1715	*gudB* mut2 in pAC5 (ST01/KG120/KG92)	EcoRI/BamHI	Gunka *et al.*, subm.
pGP1720	pGP1331-*gudB* (3xFLAG) (KG125/KG126)	BamHI/SalI	This work
pGP1721	*gudB* mut3 in pAC5 (ST01/KG133/KG92)	EcoRI/BamHI	Gunka *et al.*, subm.
pIYFP	template for *yfp* gene, *B. subtilis* (Cm[R]), *E. coli* (Amp[R])		Veening *et al.*, 2004
p25-N	*Plac*-mcs-*cyaA* (Kan[R])		Claessen *et al.*, 2008
pKT25	*Plac*-*cyaA*-mcs (Kan[R])		Karimova *et al.*, 1998
pKT25::zip	*Plac*-*cyaA*-*zip* (Kan[R])		Karimova *et al.*, 1998
pUT18	*Plac*-mcs-*cyaA* (Amp[R])		Karimova *et al.*, 1998
pUT18C	*Plac*-*cyaA*-mcs (Amp[R])		Karimova *et al.*, 1998

Name	Relevant characteristics[a]	Used restriction sites	Reference
pUT18C::zip	P_{lac}-*cyaA*-*zip* (Amp^R)		Karimova *et al.,* 1998
pUC18	cloning vector		Yanisch-Perron, *et al.*, 1985
pWH844	allows expression of proteins carrying a His tag at their N-terminus in *E. coli* (Amp^R)		Schirmer *et al.*, 1997

[a] Resistance abbreviations as follows: Amp, ampicillin; Cm, chloramphenicol; Erm, erythromycin; Kan, kanamycin; Spec, spectinomycin

* constructed by three fragment ligation

9.3. **Strains**

Tab. S3 - Strains

Strain	Genotype [a]	Reference/Construction [b]
Bacillus subtilis		
168	*trpC2*	Laboratory collection
BD3349	BD3349 trpC2 ypbH::spc	Persuh *et al.* 2002
GP28	*trpC2 ΔgudB::cat rocG::*Tn*10 spc amyE::(gltA-lacZ aphA3)*	Commichau *et al.*, 2007a
GP655	*trpC2 ΔrocF::aphA3*	Commichau, 2006
GP656	*trpC2 ΔrocD::aphA3*	Commichau, 2006
GP669	*trpC2 amyE::(gltA´-´lacZ cat)*	Commichau *et al.*, 2006
GP717	*trpC2 ΔgudB::cat rocG::*Tn*10 spc amyE::(gltA-lacZ aphA3) gltB1 ansR(C107A)*	Commichau *et al.*, 2008
GP738	*trpC2 gltC::*Tn*10 spc*	Commichau, 2006
GP747	*trpC2 rocG::*Tn*10 spc*	Commichau *et al.*, 2007b
GP753	*trpC2 rocG::*Tn*10 spc gudB1*	Gunka *et al.*, subm.; spontaneous mutation of GP747 on SP
GP754	*trpC2 rocG::cat amyE::(gltA-lacZ aphA3)*	Commichau *et al.*, 2007a
GP801	*trpC2 rocG::cat gudB1 amyE::(gltA-lacZ aphA3)*	Commichau *et al.*, 2008
GP804	*trpC2 gudB1 amyE::(gltA-lacZ aphA3)*	Commichau *et al.*, 2008
GP807	*trpC2 ΔgltAB::tet*	LFH-PCR → 168
GP808	*trpC2 amyE::(gltA-lacZ aphA3) ΔgltAB::tet*	pGP526 → GP807
GP809	*trpC2ΔgudB::cat ΔrocG::tet ansR(C107A) gltB(ΔT1010Δ1011) amyE::(gltA-lacZ aphA3)*	LFH-PCR → GP717
GP810	*trpC2 ΔrocG::tet*	LFH-PCR→ 168
GP811	*trpC2 ΔgudB::cat rocG::*Tn*10 spc amyE::(gltA-lacZ aphA3) ansR::tet*	Flórez *et al.*, subm.; LFH→ GP28
GP812	*trpC2 ypbH::spec amyE::(gltA-lacZ cat)*	This work; BD3349→ GP669

Strain	Genotype [a]	Reference/Construction[b]
GP815	*trpC2 amyE::(gltA-lacZ aphA3) lacR::Tn10 spc gudB1*	This work; transposon mutagenesis pIC333 → GP804
GP816	*trpC2 gudB1 gltC(C578T) amyE::(gltA-lacZ aphA3)*	This work; spontaneous mutation of GP804 on C-Glc
GP817	*trpC2 gudB1 gltC(C578T) proJ::Tn10 spc amyE::(gltA-lacZ aphA3)*	This work; transposon mutagenesis pIC333 → GP816
GP819	*trpC2 gudB1 amyE::(gltA-lacZ cat)*	This work, pGP919→GP804
GP1101	*trpC2 amyE::(gudB-lacZ cat)*	Gunka *et al*., subm.; pGP651 → 168
GP1102	*trpC2 gudB1 amyE::(gudB-lacZ cat)*	Gunka *et al*., subm.; pGP651 → GP804
GP1103	*trpC2 rocG::Tn10 spc recA::erm cat*	Gunka *et al*., subm.; IRN444 → GP747
GP1104	*trpC2 rocG::Tn10 spc amyE::(gudB-lacZ cat)*	Gunka *et al*., subm.; pGP651→ GP747
GP1105	*trpC2 rocG::Tn10 spc amyE::(gudB-lacZ cat) gudB1*	Gunka *et al*., subm.; spontaneous mutation of GP1104 on SP
GP1106	*trpC2 ΔaddAB::spc*	Gunka *et al*., subm.; HVS666 → 168
GP1107	*trpC2 ΔaddAB::spc rocG::cat amyE::(gltA-lacZ aphA3)*	Gunka *et al*., subm.; GP1106 → GP754
GP1123	*trpC amyE::(alf-aphA3 lacZ cat)*	Gunka *et al*., subm.; pGP655 → 168
GP1127	*trpC2 amyE::(alf1-aphA3 lacZ cat)*	Gunka *et al*., subm.; spontaneous mutation of GP1123 on SP-Kan

Strain	Genotype [a]	Reference/Construction[b]
GP1152	*trpC2 ΔansR::tet*	This work; GP811→168
GP1153	*trpC2 ΔansAB::erm*[C]	This work; LFH→168
GP1154	*trpC2 ΔgudB::cat rocG::*Tn*10 spc amyE::(gltA-lacZ aphA3) ansR (C107A)gltB (ΔT1010Δ1011) ΔansAB::erm*[C]	Flórez *et al.*, subm.; LFH→GP717
GP1155	*trpC2 amyE::(ansA-lacZ cat)*	This work; pGP872→168
GP1156	*trpC2 ΔansR::tet amyE:(ansA-lacZ cat)*	This work; pGP872→GP1152
GP1157	*trpC2 rocG::cat*	This work; GP754→168
GP1158	*trpC2 gltC::*Tn*10 spc rocG::cat*	This work; GP738→GP1157
GP1159	*trpC2 ΔhtpG::cat*	This work; SV01→GP168
GP1160	*trpC2 ΔgudB::aphA3*	Gunka *et al.*, subm.; LFH→ 168
GP1161	*trpC2 ΔgudB::aphA3 rocG::*Tn*10 spc*	Gunka *et al.*, subm.; GP1160→ GP747
GP1162	*trpC2 ΔgudB::cat rocG::*Tn*10 ΔrocF::aphA3*	This work; GP655 →GP1199
GP1163	*trpC2 ΔgudB::aphA3 rocG::*Tn*10 spc amyE::(gudB lacZ cat)*	Gunka *et al.*, subm.; pGP900→GP1161
GP1164	*trpC2 ΔgudB::cat rocG::*Tn*10 ΔrocD::aphA3*	This work; GP656 →GP1199
GP1165	*trpC2 ΔgudB::aphA3 rocG::*Tn*10 spc amyE::(gudB1-lacZ cat)*	This work; spontaneous mutation of GP1163 on SP

Strain	Genotype [a]	Reference/Construction[b]
GP166	*trpC2 rocG*::Tn*10 spc ΔhtpG::cat*	This work; GP747 →GP1159
GP1167	*trpC2 Δmfd::ermC*	Gunka *et al*., 2010 subm.; LFH→ 168
GP1168	*trpC2 Δmfd::ermC amyE::(alf-aphA3 lacZ cat)*	Gunka *et al*., 2010 subm.; GP1167→GP1123
GP1169	*trpC2 rocG*::Tn*10 spc Δmfd::ermC*	Gunka *et al*., 2010 subm.; GP1167→GP747
GP1171	*trpC2 xkdE::*(N-*yfp ermC)*	This work; pGP886 → 168
GP1172	*trpC2 xkdE::*(N-*yfp emrC)*	This work; pGP887 → 168
GP1173	*trpC2 lacA::(C-yfp aphA3)*	This work; pGP888 → 168
GP1174	*trpC2 lacA::(C-yfp aphA3)*	This work; pGP889 → 168
GP1175	*trpC2 ΔuvrAB::ermC*	This work; LFH→ 168
GP1176	*trpC2 rocG*::Tn*10 spc ΔuvrAB::ermC*	This work; GP1175→GP747
GP1177	*trpC2 ΔgudB::aphA3 amyE::(gudB$_{mut1}$ lacZ cat)*	Gunka *et al*., subm.; pGP1714→ GP1160
GP1178	*trpC2 ΔgudB::aphA3 amyE::(gudB$_{mut2}$ lacZ cat)*	Gunka *et al*., subm.; pGP1715→ GP1160
GP1179	*trpC2 ΔgudB::aphA3 rocG*::Tn*10 spc amyE:: (gudB$_{mut1}$ lacZ cat)*	Gunka *et al*., subm.; GP747→ GP1177
GP1180	*trpC2 ΔgudB::aphA3 rocG*::Tn*10 spc amyE:: (gudB$_{mut2}$ lacZ cat)*	Gunka *et al*., subm.; GP747→ GP1178
GP1188	*trpC2 ΔgudB::aphA3 rocG*::Tn*10 spc amyE:: (gudB1$_{mut2}$ lacZ cat)*	This work; spontaneous mutation of GP1180 on SP

Strain	Genotype [a]	Reference/Construction[b]
GP1189	*trpC2 ΔuvrAB::ermC rocG*::Tn*10 spc gudB1*	This work; spontaneous mutation of GP1176 on SP
GP1190	*trpC2 ΔmutSL::aphA3*	This work; LFH→ 168
GP1191	*trpC2 rocG::*Tn*10 spc ΔmutSL::aphA3*	This work; GP1191 → GP747
GP1192	*trpC2 rocG::*Tn*10 spc ΔmutSL::aphA3ΔuvrAB::ermC*	This work; GP1176 → GP1191
GP1193	*trpC2 rocG::cat gudB1*	This work; spontaneous mutation of GP1157 on SP
GP1194	*trpC2 gudB -3xFlag spc*	This work; pGP1720 → 168
GP1195	*trpC2 gudB1 -3xFlag spc*	This work; pGP1720 → GP804
GP1196	*trpC2 rocG::Tn10 spc ΔrocF::aphA3*	This work; GP747 → GP655
GP1197	*trpC2 ΔgudB::aphA3 rocG*::Tn*10 spc amyE:: (gudB$_{mut3}$ lacZ cat)*	Gunka *et al*., subm.; pGP1721→ GP1161
GP1198	*trpC2 gudB::aphA3 rocG*::Tn*10 spc amyE:: (gudB1$_{mut3}$ lacZ cat)*	Gunka *et al*., subm.; spontaneous mutation of GP1197 on SP
GP1199	*trpC2 ΔgudB::cat rocG::*Tn*10*	This work; GP28→ 168
GP1200	*trpC2 rocG::Tn10 spc ΔrocD ::aphA3*	This work; GP747 → GP656
HVS666	*trpC2 ΔaddAB::spc*	Gunka *et al*., subm.; Chédin *et al.*, 1998
IRN444	*trpC2 recA::ermC cat*	Lemon *et al.*, 2001
SV01	*leuA8 metB5 trpC2 hsrM1 ΔhtpG::cat*	Versteeg *et al.*, 1999

Strain	Genotype [a]	Reference/Construction[b]
Escherichia coli		
BL21 (DE3)/ pLysS	*F- lon ompT rBmB hsdS gal (cIts857ind1 Sam7 nin5 lacUV5- T7 gene1)*	Sammbrock *et al.*, 1989
BTH101	*F- cya-99 araD139 galE15 galK16 rpsL1 (StrR) hsdR2 mcrA1 mcrB1*	Karimova *et al.,* 2005
DH5α	*recA1 endA1 gyrA96 thi hsdR17*r_K^- m_K^+*relA1 supE44* Φ80Δ*lacZ*Δ*M15* Δ(*lacZYA-argF*)U169	Sammbrock *et al.*, 1989
XL1-Red	*endA1 gyrA96 thi- 1 hsdR17 supE44 relA1 lac mutD5 mutS mutT* Tn*10* (*tet*)	Greener & Callahan, 1994

[a] Resistance gene abbreviations as follows: *aphA3*, kanamycin; *cat*, chloramphenicol; *ermC*, erythromycin; *sp*c, spectinomycin; *tet*, tetracycline.

[b] Arrows indicate construction by transformation.

Curriculum vitae

Persönliche Angaben

Name	Katrin Gunka
Geburtsdatum, -ort	08.09.1981 in Cuxhaven
Staatsangehörigkeit	deutsch

Schulbildung

1988 – 1992	Grundschule Sahlenburg
1992 – 1994	Orientierungstufe Franzenburg
1994 - 2001	Abitur am Lichtenberg-Gymnasium Cuxhaven

Studium

04/2002 - 09/2002	Lehramtsstudium für die Sekundarstufe II, Georg-August-Universität Göttingen
10/2002 - 01/2007	Diplomstudium Biologie, Georg-August-Universität Göttingen
02/2007 - 11/2007	Anfertigung der Diplomarbeit „Isolation von Mutationen, die die regulatorische Wechselwirkung von RocG und GltC in *Bacillus subtilis* beeinträchtigen" Arbeitsgruppe von Prof. Dr. Jörg Stülke, Institut für Mikrobiologie und Genetik, Georg-August-Universität Göttingen

Promotion

01/2008 – 02/2011	Promotionsthema: „Der Einfluss der Glutamat-dehydrogenasen auf die Verknüpfung des Kohlenstoff- und Stickstoffstoffwechsels in *Bacillus subtilis*" Arbeitsgruppe von Prof. Dr. Jörg Stülke