6102–6115 *Nucleic Acids Research, 2006, Vol. 34, No. 21* *Published online 29 October 2006*
doi:10.1093/nar/gkl733

Keeping signals straight in transcription regulation: specificity determinants for the interaction of a family of conserved bacterial RNA–protein couples

Oliver Schilling, Christina Herzberg, Tina Hertrich[1], Hanna Vörsmann, Dirk Jessen, Sebastian Hübner, Fritz Titgemeyer[1] and Jörg Stülke*

Department of General Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen, Germany and [1]Lehrstuhl für Mikrobiologie, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Received August 3, 2006; Revised September 20, 2006; Accepted September 21, 2006

ABSTRACT

Regulatory systems often evolve by duplication of ancestral systems and subsequent specialization of the components of the novel signal transduction systems. In the Gram-positive soil bacterium *Bacillus subtilis*, four homologous antitermination systems control the expression of genes involved in the metabolism of glucose, sucrose and β-glucosides. Each of these systems is made up of a sensory sugar permease that does also act as phosphotransferase, an antitermination protein, and a RNA switch that is composed of two mutually exclusive structures, a RNA antiterminator (RAT) and a transcriptional terminator. We have studied the contributions of sugar specificity of the permeases, carbon catabolite repression, and protein–RAT recognition for the straightness of the signalling chains. We found that the β-glucoside permease BglP does also have a minor activity in glucose transport. However, this activity is irrelevant under physiological conditions since carbon catabolite repression in the presence of glucose prevents the synthesis of the β-glucoside permease. Reporter gene studies, *in vitro* RNA–protein interaction analyzes and northern blot transcript analyzes revealed that the interactions between the antiterminator proteins and their RNA targets are the major factor contributing to regulatory specificity. Both structural features in the RATs and individual bases are important specificity determinants. Our study revealed that the specificity of protein–RNA interactions, substrate specificity of the permeases as well as the general mechanism of carbon catabolite repression together allow to keep the signalling chains straight and to avoid excessive cross-talk between the systems.

INTRODUCTION

To sense their environment and to adapt to changing conditions, all organisms possess signal transduction systems which are composed of a sensor that perceives the signal, a regulator that can modify its activity in response to the signal, and a target of regulation. This general scheme can be modified in many ways: The sensor and the regulator are often combined in the same molecule as in the Lac repressor. The sensor and the regulator are usually proteins, but regulatory RNAs continue to be uncovered. The target of the regulation may be a protein, i.e. an enzyme, but for the control of gene expression, specific DNA or RNA sequences are the most common targets.

In bacteria, the number of environmental or internal signals that need to be sensed is much higher than the number of non-related regulatory systems. Thus, large families of regulation systems are present in bacteria. Among the most common families are the two-component regulatory systems, sigma factors with their anti-sigma factors as well as several families of repressor and activator proteins (1–5). All these families can be divided to sub-families that do often respond to similar signals. The evolution of signalling families is still in progress and can be observed in the transcriptional regulation of biodegradation pathways. Even more, new regulatory systems can be generated artificially (6,7). The similarity of the components of many families of signal transduction systems raises the question how the bacteria avoid excessive cross-talk, i.e. the activation of a regulatory protein by gratuitous inducers or the induction of a gene by a non-cognate regulator protein that recognizes a similar DNA sequence. This problem was the subject of extensive analyzes for the

*To whom correspondence should be addressed. Tel: +49 551 393781; Fax: +49 551 393808; Email: jstuelk@gwdg.de

two-component regulatory systems in the gram-positive soil bacterium *Bacillus subtilis* (8).

We are interested in the control of glucose utilization in *B.subtilis*. This sugar is transported by a specific permease of the phosphotransferase system (PTS) encoded by *ptsG* and is subsequently catabolized via the glycolytic pathway (9). The expression of the *ptsG* gene and of several glycolytic genes is inducible by glucose, however, the mechanisms differ. While *ptsG* expression is induced by transcriptional antitermination, the glycolytic *gapA* operon is controlled by the repressor CggR (10–13). Induction of *ptsG* expression involves a RNA switch which is the target of the antitermination protein GlcT, and the sensory glucose permease, PtsG. As part of the PTS, the glucose permease possesses two soluble domains that are involved in the phosphate transfer from phosphoenolpyruvate to the incoming sugar, the domains IIA and IIB (14). If glucose is present, the phosphate groups are immediately transferred to the sugar, whereas they accumulate on the glucose permease as well as on the two general proteins of the PTS, enzyme I and HPr, in the absence of glucose. Under these conditions, the glucose permease can transfer a phosphate residue to GlcT thereby inactivating the antitermination protein (15,16). GlcT is made up of three domains, an N-terminal RNA-binding domain, and two homologous PTS-regulation domains called PRD-I and PRD-II (15,17,18). Phosphorylation of a conserved histidine residue in PRD-I by the glucose permease results in GlcT inactivation in the absence of glucose. Biochemical studies revealed that PRD-II of GlcT can also be phosphorylated on a conserved histidine residue, however this phosphorylation is catalyzed by the HPr protein of the PTS and has only a very minor impact on the activity of GlcT (16). If in the right phosphorylation state, i.e. if non-phosphorylated in PRD-I, GlcT can bind its target site on the *ptsG* mRNA called RNA antiterminator (RAT, 15,19). The RAT overlaps a transcriptional terminator located in the leader region of the *ptsG* mRNA and the two structures form a RNA switch since they are mutually exclusive. Binding of GlcT to the RAT is thought to prevent the formation of the terminator and to allow transcription elongation into the *ptsG* structural gene. This regulatory system couples the availability of the inducer glucose to the phosphorylation state of the sensor permease and the antitermination protein GlcT resulting in either of two states of the *ptsG* RNA switch and subsequently in *ptsG* gene expression.

The regulatory system controlling *ptsG* expression is part of a family made up of highly conserved components, i.e. sensor permeases, antitermination proteins and RAT targets for the regulatory proteins (see Figure 1A). The additional permeases transport sucrose and the β-glucoside salicin. Two antitermination proteins, SacT and SacY, regulate expression of sucrose catabolic genes. While SacT is thought to be active at low sucrose concentrations, high concentrations of sucrose are required to activate SacY since its cognate permease SacX has a very weak transport activity (9). LicT controls the expression of the *bglPH* operon and the *licS* gene in the presence of salicin (20). SacT and LicT are only active if (i) their inducers are present and if (ii) no glucose is present in the medium. This allows their phosphorylation by HPr in the PRD-II leading to activation of the antiterminator proteins. In contrast, SacY and GlcT,

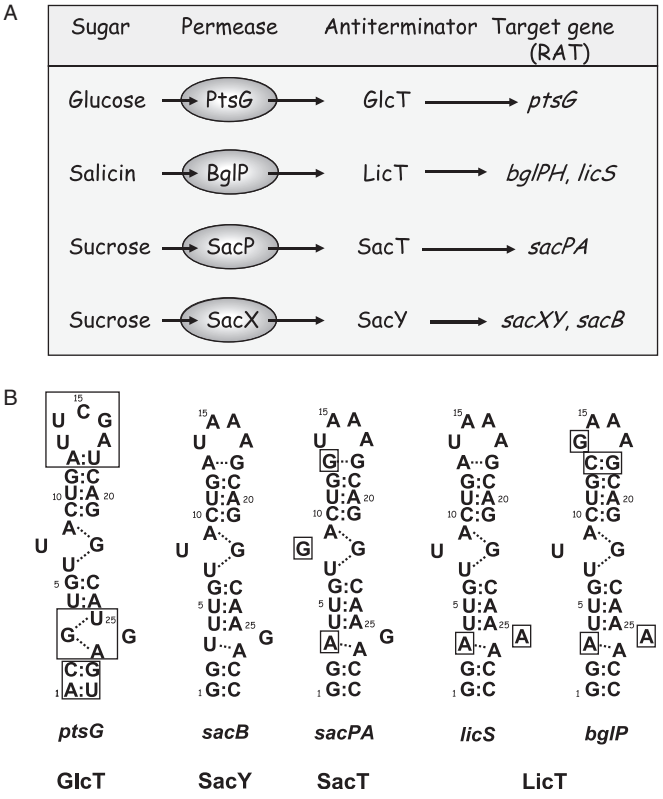

Figure 1. A family of antitermination systems controls sugar uptake and metabolism in *B.subtilis*. (**A**) The design of the four signalling systems. The target genes encode the following proteins: *ptsG*, glucose permease of the PTS; *bglP*, β-glucoside permease of the PTS; *bglH*, phospho-β-glucoside hydrolase; *licS*, β-1,3-1,4-glucanase; *sacP*, sucrose permease of the PTS; *sacA*, sucrase; *sacX*, sucrose permease of the PTS (low affinity); *sacY*, antitermination protein; *sacB*, extracellular levansucrase. (**B**) Comparison of the secondary structures of the related RAT-RNAs of *B.subtilis* (24,43). The relevant antiterminator proteins are indicated below their cognate RAT structures. Boxes indicate nucleotides that differ from the *sacB* RAT. Dashed lines indicate bases that are proposed to be in direct contact to each other.

which are active in the presence of high sugar concentrations (or with the preferred sugar), are independent on a HPr-dependent activation even though HPr can phosphorylate these proteins. The glucose-dependent control of the antiterminator proteins' activity by HPr is part of the phenomenon of carbon catabolite repression which results in the preferential utilization of easily metabolizable carbon sources (18,21,22).

The RAT targets of the four antiterminator proteins are all similar to each other (Figure 1B). Some determinants causing specificity of protein–RNA interaction have been identified in a pioneering work (19). The determination of the structure of the complex between the RNA-binding domain of LicT and its cognate RAT-RNA suggested that the essential contacts between the protein and the RNA are made in the loop regions of the RAT (23). The *ptsG* RAT recognized by GlcT is most different from all other RAT structures, and neither this RAT nor GlcT are involved in any cross-talk (24). In this work, we identified determinants that result in the regulatory specificity of the four distinct antitermination systems.

MATERIALS AND METHODS

Bacterial strains and growth conditions

The *B.subtilis* strains used in this study are shown in Table 1. All *B.subtilis* strains are derivatives of the wild type strain 168. Strains used in the cause of site-directed mutagenesis studies are listed in Tables 2 and 3. These strains were all derived by transformation from the basal mutant strains listed in Table 1. *Escherichia coli* DH5α and BL21(DE3) (25) were used for cloning experiments and for expression of recombinant proteins, respectively.

B.subtilis was grown in SP medium or in CSE minimal medium (26). The media were supplemented with auxotrophic requirements (at 50 mg/l), carbon sources and inducers as indicated. *E.coli* was grown in Luria–Bertani medium (LB medium) and transformants were selected on plates containing ampicillin (100 μg/ml). LB and SP plates were prepared by the addition of 17 g Bacto agar/l (Difco) to LB or SP medium, respectively.

Table 1. *B.subtilis* strains used in this study

Strain	Genotype	Source[a]
168	*trpC2*	Laboratory collection
BGW10	*trpC2 lys-3 ΔlicTS*::*erm*	45
GM1112	*sacXYΔ3 sacBΔ23 sacTΔ4 bglP*::Tn*10 erm amyE*::(*sacB-lacZ phl*)	46
QB5435	*trpC2 ΔptsG*::*cat*	10
QB5448	*trpC2 amyE*::(DLA *ptsG'-'lacZ aphA3*)	10
GP109	*trpC2* D*glcT8 amyE*:: (ΔLA *ptsG'-'lacZ aphA3*)	15
GP150	*trpC2 ΔglcT8 amyE*::(*'lacZ cat*)	36
GP385	*trpC2 amyE*:: (ΔLA *ptsG-R1'-'lacZ aphA3*)	24
GP425	*trpC2 ΔsacY*::*cat*	see Materials and Methods
GP427	*trpC2 ΔlicTS*::*erm*	BGW10→168
GP429	*trpC2 ΔsacT*::*spc*	see Materials and Methods
GP430	*trpC2 ΔsacY*::*cat ΔsacT*::*spc*	GP429→GP425
GP431	*trpC2 ΔlicTS*::*ermΔ sacT*::*spc*	GP429→GP427
GP432	*trpC2 ΔlicTS*::*erm ΔsacY*::*cat*	GP425→GP427
GP433	*trpC2 ΔlicTS*::*erm ΔsacY*::*cat sacT*::*spc*	GP429→GP432
GP437	*trpC2 amyE*::(*sacB-lacZ aphA3*)	pGP564→168
GP470	*trpC2 ΔptsG*::*cat bglP*::Tn*10 erm*	GM1112→QB5435

[a]Arrows indicate construction by transformation.

Transformation and characterization of the phenotype

B.subtilis was transformed with plasmid DNA according to the two-step protocol described previously (27). Transformants were selected on SP plates containing kanamycin (Km 5 μg/ml), chloramphenicol (Cm 5 μg/ml), spectinomycin (Spc 100 μg/ml), or erythromycin plus lincomycin (Em 1 μg/ml and Lin 10 μg/ml).

In *B.subtilis*, amylase activity was detected after growth on SP medium supplemented with 10 g hydrolyzed starch/l (Connaught). Starch degradation was detected by sublimating iodine onto the plates.

Quantitative studies of *lacZ* expression in *B.subtilis* in liquid medium were performed as follows: cells were grown in CSE medium supplemented with the carbon

Table 2. Effect of mutations in the *ptsG* RAT on recognition by the different antiterminator proteins

Strain	RAT	Relevant genotype	β-Galactosidase activity (U/mg protein)[a]				
			CSE	CSE Glc	CSE Suc (0.1%)	CSE Suc (2%)	CSE Sal
QB5448	*ptsG*	wild type	9	548	231	329	267
GP109	*ptsG*	*glcT*	4	5	2	8	9
GP387	*ptsG*	*ΔsacT ΔsacY*	10	327	85	166	335
GP389	*ptsG*	*ΔlicT*	12	412	222	422	312
GP385	*ptsG*-R1	wild type	6	8	34	37	102
GP386	*ptsG*-R1	*glcT*	10	18	48	49	121
GP390	*ptsG*-R1	*ΔlicT*	16	11	17	14	13
GP413	*ptsG*-R2	wild type	3	2	2	3	4
GP415	*ptsG*-R3	wild type	14	11	12	7	14
GP416	*ptsG*-R4	wild type	11	3	8	6	94
GP396	*ptsG*-R4	*glcT*	7	9	34	20	90
GP417	*ptsG*-R4	*ΔlicT*	10	10	11	10	9
GP404	*ptsG*-R5	wild type	34	33	174	121	877
GP400	*ptsG*-R5	*glcT*	24	100	457	178	633
GP402	*ptsG*-R5	*ΔlicT*	12	18	66	80	16
GP455	*ptsG*-R5	*ΔlicT ΔsacY*	7	9	42	40	8
GP454	*ptsG*-R5	*ΔlicT ΔsacT*	6	10	10	9	9
GP456	*ptsG*-R5	*ΔlicT ΔsacT ΔsacY*	10	9	10	11	8
GP434	*ptsG*-R5	*glcT ΔlicT*	9	8	—	—	6
GP436	*ptsG*-R5	*glcT ΔsacT*	32	70	—	—	—
GP435	*ptsG*-R5	*glcT ΔsacY*	33	73	—	—	—
GP408	*ptsG*-R6	wild type	34	16	45	50	702
GP399	*ptsG*-R6	*glcT*	19	50	239	156	514
GP409	*ptsG*-R6	*ΔlicT*	14	16	20	20	12
GP464	*ptsG*-R7	wild type	37	32	36	37	60
GP419	*ptsG*-R8	wild type	147	188	157	238	1038
GP420	*ptsG*-R8	*glcT*	160	174	660	798	786
GP421	*ptsG*-R8	*ΔlicT*	209	150	300	287	217

[a]Representative values of *lacZ* expression. All measurements were performed at least twice.

Table 3. Conversion analysis of the *sacB* RAT

Strain	RAT	Relevant genotype	β-Galactosidase activity (U/mg protein)[a]				
			CSE	CSE-Glc	CSE Suc (0.1%)	CSE Suc (2%)	CSE Sal
GP437	*sacB*	wild type	9	6	54	78	5
GP440	*sacB*	Δ*sacT*	5	4	5	28	4
GP438	*sacB*	Δ*sacY*	6	5	50	49	6
GP441	*sacB*	Δ*sacT* Δ*sacY*	6	3	5	4	3
GP461	*sacB*-R1	wild type	7	8	132	104	35
GP465	*sacB*-R1	Δ*sacT* Δ*licT*	5	4	3	4	2
GP466	*sacB*-R1	Δ*sacY* Δ*licT*	5	5	80	74	3
GP463	*sacB*-R1	Δ*sacT* Δ*sacY*	11	8	10	12	28
GP462	*sacB*-R1	Δ*licT*	7	9	85	82	5
GP460	*sacB*-R2	wild type	7	6	93	145	8
GP472	*sacB*-R2	Δ*sacY* Δ*licT*	3	5	97	112	5
GP471	*sacB*-R2	Δ*sacT* Δ*licT*	4	3	4	7	3
GP519	*sacB*-R3	wild type	5	2	96	88	4
GP521	*sacB*-R3	Δ*sacT*	4	—	4	9	4
GP540	*sacB*-R4	wild type	8	8	70	117	7
GP541	*sacB*-R4	Δ*sacT*	7	8	9	35	7
GP542	*sacB*-R4	Δ*sacT* Δ*sacY*	4	3	3	4	2
GP520	*sacB*-R5	wild type	4	3	216	208	6
GP522	*sacB*-R5	Δ*sacT*	4	2	5	7	2
GP537	*sacB*-R6	wild type	4	6	349	463	11
GP538	*sacB*-R6	Δ*sacT*	6	5	6	22	9
GP539	*sacB*-R6	Δ*sacT* Δ*sacY*	3	4	4	4	3
GP476	*sacB*-R7	wild type	3	6	17 6	145	92
GP544	*sacB*-R7	Δ*sacT*	5	4	7	8	136
GP484	*sacB*-R7	Δ*licT*	3	2	195	185	2
GP536	*sacB*-R7	Δ*sacT* Δ*licT*	2	2	3	5	2
GP477	*sacB*-R8	wild type	6	7	248	294	218
GP486	*sacB*-R8	Δ*licT*	2	1	305	188	2
GP487	*sacB*-R8	Δ*sacT*	5	4	9	8	376
GP480	*sacB*-R9	wild type	7	11	280	253	510
GP492	*sacB*-R9	Δ*licT*	3	4	291	189	3
GP493	*sacB*-R9	Δ*licT* Δ*sacT*	4	2	3	4	2
GP494	*sacB*-R9	Δ*licT* Δ*sacY*	3	3	213	193	3
GP444	*sacB*-R10	wild type	6	3	24	18	10
GP543	*sacB*-R10	Δ*sacT*	5	-	5	8	12
GP446	*sacB*-R11	wild type	48	32	43	34	23
GP448	*sacB*-R12	wild type	5	4	4	6	4
GP450	*sacB*-R13	wild type	28	95	112	129	33
GP453	*sacB*-R13	Δ*licT* Δ*sacT* Δ*sacY*	48	86	104	80	28
GP451	*sacB*-R13	*glcT*	31	33	20	36	18

[a]Representative values of *lacZ* expression. All measurements were performed at least twice.

sources indicated. Cells were harvested at OD_{600} 0.6–0.8. Cell extracts were obtained by treatment with lysozyme and DNase. β-Galactosidase activities were determined as previously described using *o*-nitrophenyl-galactoside as a substrate (27). One unit is defined as the amount of enzyme which produces 1 nmol of *o*-nitrophenol per min at 28°C.

DNA manipulation

Transformation of *E.coli* and plasmid DNA extraction were performed using standard procedures (25). Restriction enzymes, T4 DNA ligase and DNA polymerases were used as recommended by the manufacturers. DNA fragments were purified from agarose gels using the QIAquick gel extraction kit (Qiagen®, Hilden, Germany). *Pfu* DNA polymerase was used for the PCR as recommended by the manufacturer. The combined chain reaction and the multiple mutation reaction were performed with *Pfu* DNA polymerase and thermostable DNA ligase (Ampligase®, Epicentre, Wisconsin, USA). DNA sequences were determined using the dideoxy chain termination method (25). Chromosomal DNA of *B.subtilis* was isolated as described (27).

Construction of *sacT* and *sacY* mutant strains by allelic replacement

To construct *sacT* and *sacY* mutant strains, the long flanking homology PCR (LFH-PCR) technique was used (28). Briefly, cassettes carrying the *cat* and *spc* resistance genes were amplified from the plasmids pGEM-*cat* and pDG1726, respectively (29,30). DNA fragments of ∼1000 bp flanking the regions to be deleted at their 5′ and 3′ ends were amplified. The 3′ end of the upstream fragment as well as the 5′ end of the downstream fragment extended into the gene(s) to be deleted in a way that all expression signals of genes up- and downstream of the targeted genes remained intact. The joining of the two fragments to the resistance cassette was performed in a second PCR as described previously (31). In these reaction we used the primer pairs cat-fwd (5′-CGGCAATAGTTACCCTTATTATCAAG)/cat-rev (5′-CCAGCGTGGACCGGCGAGGCTAGTTACCC) and spec-fwd/spec-rev (31) for the

amplification and joining of the cat and spc cassettes, respectively. The PCR products were directly used to transform *B.subtilis*. The integrity of the regions flanking the integrated resistance cassettes was verified by sequencing PCR products of ~1000 bp amplified from chromosomal DNA of the resulting mutants. The resulting strains were GP425 (Δ*sacY::cat*) and GP429 (Δ*sacT::spc*).

Site-directed mutagenesis

Translational fusions of variants of the *ptsG* and *sacB* regulatory regions with the *lacZ* gene were constructed using the vector pAC7 (32) containing the kanamycin resistance gene *aphA3*. The plasmid harbours a *lacZ* gene without a promoter located between two fragments of the *B.subtilis amyE* gene. To construct a translational *sacB-lacZ* fusion the DNA upstream from the *sacB* gene [−464 to +15 nt relative to the translational start point of *sacB* (33)] was amplified by PCR using the primers OS49 (5′- AAAGAATTCGATCCTTTTTAACCCATCACATATAC) and OS50 (5′-TTTGGATCCTTTTTGATGTTCATCGTTCATGTC). The primers introduced BamHI and EcoRI cloning sites at the ends of the amplified fragment and created an in-frame translational fusion of the *lacZ* gene with the 5th codon of *sacB*. The PCR product was inserted into pAC7, both linearized with the same enzymes producing plasmid pGP437.

To study the effect of point mutations in the RAT sequences the following strategy was applied: a DNA fragment carrying the mutant form of the RAT was constructed by site-directed mutagenesis using either the combined chain reaction or the multiple mutation reaction (to introduce three or more mutations simultaneously) as outlined previously (34,35). Plasmids pGP66 (10) and pGP437 containing the *ptsG* and *sacB* promoter regions, respectively, served as templates. The mutagenic primers and the resulting plasmids are available upon request. The oligonucleotides JS11 (10)/IL5 (36) and OS49/OS50 (see above) were used as outer primers for *ptsG* and *sacB*, respectively. The final PCR products were purified and cut by BamHI and MfeI (for *ptsG*) or BamHI and EcoRI (for *sacB*) sites introduced by the PCR primers. To introduce the constructed *lacZ* fusions into the chromosome of *B.subtilis*, competent cells of the wild type strain 168 were transformed with the plasmids carrying the respective mutations linearized with ScaI.

Construction of expression vectors for the RNA-binding domains of antiterminator proteins

A plasmid allowing the fusion of any protein to a Strep tag at the C-terminus was constructed as follows: First, the expression vector pET3C (Novagen) was digested with NdeI and BamHI. The insert containing a small multiple cloning site and the Strep tag was prepared by annealing the complementary oligonucleotides OS91 (5′-TATGGAGCTCGGATCCTGGAGCCACCCGCAGTTCGAAAAATGATAGT) and OS92 (5′-GATCACTATCATTTTTCGAACTGCGGGTGGCTCCAGGATCCGAGCTCCA). The resulting DNA fragment carries ends compatible with NdeI and BamHI. Upon ligation, the NdeI site was conserved whereas the BamHI site was lost. The resulting plasmid, pGP574, carries an IPTG-inducible promoter, a small cloning site (NdeI–SacI–BamHI) for the insertion of the coding sequences, and the sequence encoding the Strep tag followed by two stop codons.

To fuse the RNA-binding domains of GlcT, LicT and SacT to a Strep tag at their C-termini, plasmids pGP575, pGP576 and pGP577 were constructed: DNA fragments corresponding to amino acids 1–60 of GlcT, and 1–57 of LicT and SacT were amplified by PCR using chromosomal DNA of *B.subtilis* QB5448 and the primer pairs OS93/OS94, OS95/OS96, and OS97/OS98, respectively (the primer sequences are available upon request). The PCR products were digested with NdeI and BamHI, and the resulting fragments were cloned into the expression vector pGP574 cut with the same enzymes.

Protein purification

E.coli BL21(DE3)/pLysS was used as host for the overexpression of recombinant proteins. Expression was induced by the addition of IPTG (final concentration 1 mM) to exponentially growing cultures (OD_{600} of 0.8). Cells were lysed using a french press. After lysis the crude extracts were centrifuged at 15 000 *g* for 30 min and then passed over a Streptactin column (IBA, Göttingen, Germany). The recombinant protein was eluted with desthiobiotin (Sigma, final concentration 2.5 mM). After elution the fractions were tested for the desired protein using 12.5% SDS–PAGE gels. The relevant fractions were combined and dialysed overnight. Purified proteins were concentrated using Microsep™ Microconcentrators with a molecular weight cut-off of 3 kDa (Pall Filtron, Northborough, MA). The protein concentration was determined according to the method of Bradford using the Bio-rad dye-binding assay and BSA as the standard.

Assay of interaction between the RNA-binding domains and RAT-RNA

To obtain templates for the *in vitro* synthesis of the *ptsG* RAT-RNA, the primers OS25/OS26 (24) were used to amplify a 99 bp PCR product using pGP66 or the plasmid carrying the desired mutation as template. Similarly, a 99 bp DNA fragment encompassing the *sacB* RAT was amplified using the oligonucleotides OS86 (5′-CCAAGTAATACGACTCACTATAGGCGAAAAGTAAATCGCGCG) and OS87 (5′-GTATACACTTTGCCCTTTACAC) and pGP437 or a mutant variant as template. The presence of a T7 RNA polymerase recognition site on primers OS25 and OS86 (underlined) allowed the use of the PCR product as a template for *in vitro* transcription with T7 RNA polymerase (Roche Diagnostics). The integrity of the RNA transcripts was analyzed by denaturating agarose gel electrophoresis (12).

Binding of the RNA-binding domains to RAT-RNA was analyzed by gel retardation experiments. The RAT-RNA (in water) was denatured by incubation at 90°C for 2 min and renatured by dilution 1:1 with ice cold water and subsequent incubation on ice. Purified protein was added to the RAT-RNA and the samples were incubated for 10 min at room temperature in TAE buffer in the presence of 300 mM NaCl. After this incubation, glycerol was added to a final concentration of 10% (w/v). The samples were then analyzed on 10% Tris–acetate PAA gels.

Northern blot analysis

RNA was prepared by the modified 'mechanical disruption protocol' described previously (12). Briefly, 20 ml of cells were harvested at the exponential phase. After mechanical cell disruption, the frozen powder was instantly resuspended in 3 ml lysis buffer [4 M guanidine isothiocyanate; 0.025 M sodium acetate, pH 5.3; 0.5% *N*-laurylsarcosine (w/v)]. Subsequently, total RNA was extracted using the RNeasy Mini kit (Quiagen, Germany). Digoxigenin RNA probes specific for the *E.coli lacZ* gene were obtained by *in vitro* transcription with T7 RNA polymerase (Roche Diagnostics) using a PCR-generated fragment as templates. The primers used for PCR were SHU55 (5′-GTTTTACAACGTCGTGACTGG) and SHU56 (5′-CTAATACGACTCACTATAGGGAGGTGTGC-AGTTCAACCACCG). The reverse primers contained a T7 RNA polymerase recognition sequence. *In vitro* RNA labelling, hybridization and signal detection were carried out according to the manufacturer's instructions according to the instructions of the manufacturer (DIG RNA labelling kit and detection chemicals; Roche Diagnostics).

Uptake of radioactive glucose *in vivo*

B.subtilis strains were grown in CSE medium with glucose (10 g/l). Sugar uptake assays were performed as described previously (10). Exponentially growing cells were harvested at an OD_{600} of 0.6–0.8 and washed once with the incorporation medium. Labelled [^{14}C] glucose (184 mCi mmol^{-1}) and non-labelled glucose (final concentration 0.4 mM) were added. Samples were taken and treated as described (10).

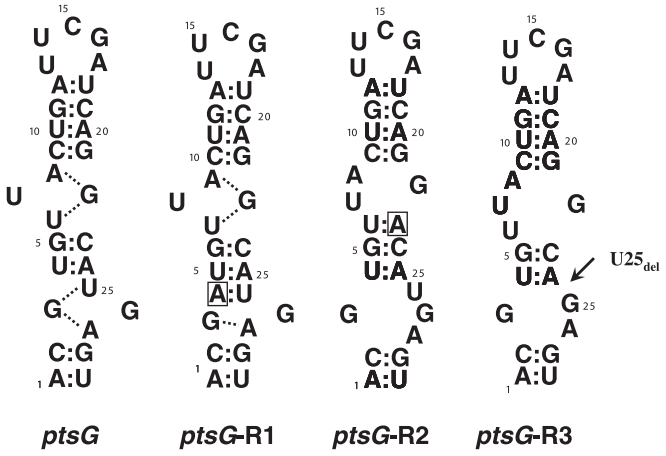

Figure 2. Predicted secondary structures of *ptsG* RAT and the *ptsG* RAT mutants *ptsG*-R1, *ptsG*-R2, and *ptsG*-R3. The insertions of an adenine at position 4 in the *ptsG*-R1 and at position 23 in the *ptsG*-R2 RAT mutants are boxed. The deletion of the base U25 in *ptsG*-R3 mutant is indicated by an arrow.

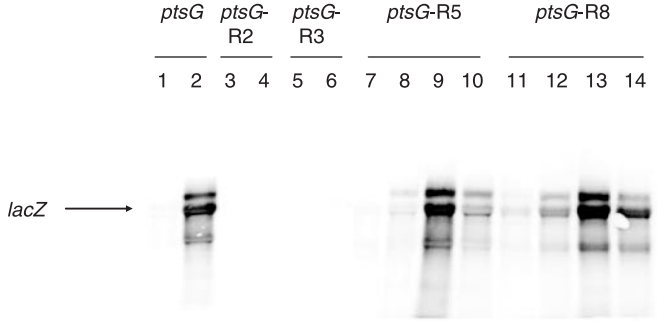

Figure 3. Northern Blot analysis of the expression of the *lacZ* gene under control of wild type and mutant RATs. Total RNA was separated by electrophoresis in 1.0% agarose gels and, after blotting, nylon membranes were hybridized to a riboprobe specific for *lacZ*. The mRNA corresponding to the *lacZ* gene is marked by an arrow. Note that a larger transcript was detected due to imperfect termination of *lacZ* transcription. 5µg RNA per lane were applied. The RNAs were isolated from the wild type strain QB5448 (lanes 1, 2), and the RAT mutant strains GP413 (*ptsG*-R2, lanes 3, 4), GP415 (*ptsG*-R3, lanes 5, 6), GP404 (*ptsG*-R5, lanes 7 to 10), and GP419 (*ptsG*-R8, lanes 11 to 14). The cultures for RNA isolation were grown in CSE minimal medium (lanes 1, 3, 5, 7, 11), in CSE medium supplemented with glucose (lanes 2, 4, 6, 8, 12), salicin (lanes 9, 13), or sucrose (lanes 10, 14).

RESULTS

Analysis of the loop structures in the *ptsG* RAT

The *ptsG* RAT differs from all other RAT sequences recognized by antiterminator proteins of the BglG/SacY family in the structure of the lower loop (Figure 1B). In a previous work, we have demonstrated that the insertion of one base into the lower loop of the *ptsG* RAT (the *ptsG*-R1 mutation, see Figure 2) makes its structure similar to that recognized by the other antiterminator proteins and results in exclusive binding of LicT to this structure whereas it is not bound by GlcT (24). From this result it was concluded that structure rather than the nucleotide sequence is important for antiterminator protein-RAT recognition.

Since the antitermination proteins bind as dimers to the RAT, and LicT contacts different structures of the lower and the upper loop, we asked whether a RAT with an 'inversion' of the lower and upper RAT structures might be recognized by any of the antitermination proteins (see Figure 2). The activity of this mutant RAT, *ptsG*-R2, was assayed by analyzing the expression of a translational fusion of the mutated *ptsG* control region to a promoterless *lacZ* gene (see Table 2). While the presence of glucose in the growth medium resulted in a strong GlcT-dependent induction of the *ptsG* promoter in the wild type, salicin induced the *ptsG*-R1 promoter region in a LicT-dependent manner. In contrast, the *ptsG*-R2 promoter region did not allow expression of the *lacZ* fusion irrespective of the potential inducing carbohydrate present in the medium. Thus, this RAT is not bound by any of the antitermination proteins in *B.subtilis* (Table 2).

The *ptsG*-R1 structure was obtained by inserting an A after position 3 of the RAT sequence. This did not only create a lower loop strucure similar to those present in RAT structures bound by LicT, SacT, and SacY, but did also generate an additional base pair between the lower and upper loops (see Figures 1 and 2). To rule out any effect of this extra base pair we constructed the *ptsG*-R3 RAT mutant by deleting the U at position 25. This results in a lower loop identical to that in the *ptsG*-R1 RAT, but separated from the upper loop by only 2 bp (Figure 2). The biological activity of this RAT mutant was determined by studying its effect on the expression of a *ptsG-lacZ* fusion. As shown in Table 2, the *ptsG*-R3 RAT did not confer induction under any of the conditions tested

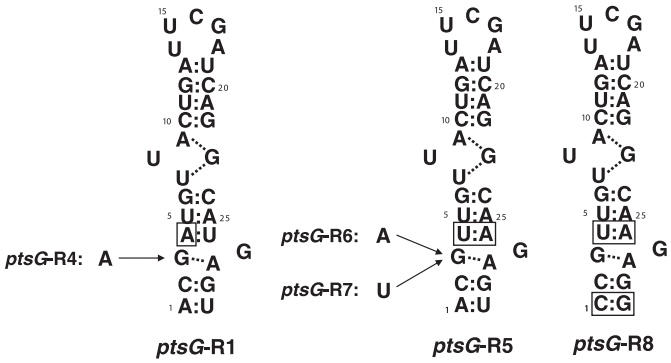

Figure 4. Secondary structures of the *ptsG*-R1, *ptsG*-R5, and *ptsG*-R8 RAT mutants. Bases that differ from *ptsG* RAT are boxed.

suggesting that this structure is not recognized by any of the antitermination proteins (see Discussion).

The inability of GlcT to bind to the *ptsG*-R2 and *ptsG*-R3 RATs was verified by a northern blot analysis. The amounts of *lacZ* mRNA were compared in the wild type strain QB5448 and the two mutant strains GP413 and GP415. As can be seen in Figure 3, the *lacZ* mRNA was strongly induced in cells grown in the presence of glucose whereas no induction was observed in the two mutant strains. This result is in perfect agreement with those obtained by the reporter gene assays (Table 2).

Contribution of individual bases to the recognition of the RAT sequence by antitermination proteins

The cognate RATs bound by LicT, SacT, and SacY are very similar to each other both in terms of structure and sequence (Figure 1B). However, the *ptsG*-R1 RAT is recognized by LicT only and not by SacY or SacT. Therefore, we decided to introduce further mutations into the *ptsG*-R1 RAT that allow the evaluation of the contribution of individual nucleotides to protein–RNA recognition.

The RATs recognized by LicT and SacT contain an A at position 3 in the lower loop rather than a G as in the *ptsG*-R1 RAT. Therefore, we exchanged the G3 for an A. The effect of this mutation, present in the *ptsG*-R4 RAT (see Figure 4), was tested by the analysis of a *ptsG*-R4-*lacZ* fusion. This mutation resulted in a *lacZ* expression comparable to that observed with the *ptsG*-R1 RAT. As determined for *ptsG*-R1, the expression driven by the *ptsG*-R4 promoter region was completely dependent on a functional *licT* gene (Table 2). Thus, LicT is the only antiterminator protein binding to both the *ptsG*-R1 and *ptsG*-R4 RATs.

Another important difference between the *ptsG*-R1 RAT and all other RATs recognized by LicT, SacT, or SacY is the U:A base pair above the lower loop, which is A:U in *ptsG*-R1 (see Figures 1B and 2). Previous results suggested that inversions of base pairs in the stems of the RAT are tolerated as long as the general structure is conserved (24). However, due to the strict conservation of the U:A pair in this position in all RATs except *ptsG*-R1, we addressed the effect of such a base pair inversion. As can be seen in Table 2, this inversion, present in the *ptsG*-R5 mutation (Figure 4), resulted in an increased expression of the fusion under all conditions tested. However, the *ptsG*-R5 RAT

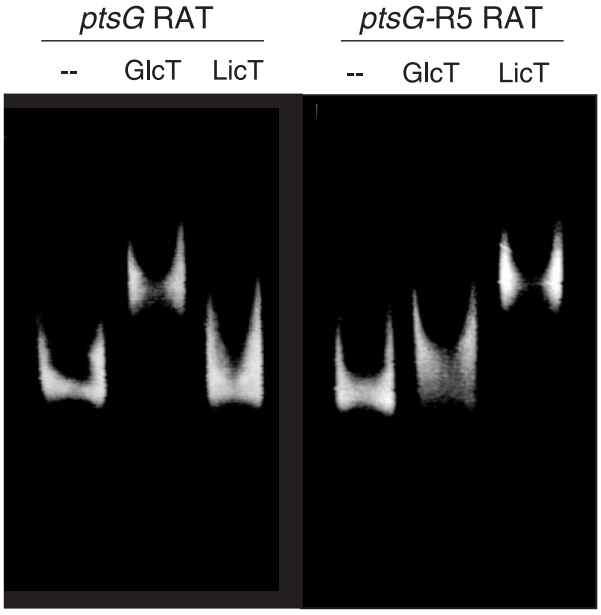

Figure 5. Electrophoretic mobility shift analysis of the interaction between the wild type *ptsG* and *ptsG*-R5 RATs, and the RNA-binding domains of GlcT or LicT. 100 pmol of the *ptsG* and *ptsG*-R5 RAT-RNAs were used. GlcT or LicT (250 pmol) were added to the RNA as indicated prior to electrophoresis.

conferred a strong induction in the presence of salicin, and this induction was dependent on the presence of the LicT antiterminator protein. In contrast, glucose did not induce this fusion suggesting that GlcT is unable to bind this RAT (Table 2). These observations were verified by an electrophoretic mobility shift analysis using the purified RNA-binding domains of GlcT and LicT (see Figure 5). As reported previously (24), GlcT efficiently bound the *ptsG* RAT-RNA. In contrast, LicT was unable to bind this RNA. In good agreement with the reporter gene analysis, LicT but not GlcT was capable of binding the *ptsG*-R5 RAT *in vitro* (Figure 5). For the *ptsG*-R1 RAT, induction in the presence of salicin and, to a lesser extent, sucrose, is strictly LicT-dependent (24, Table 2). We tested therefore whether the induction of the *ptsG*-R5 RAT by sucrose was also due to binding of LicT. As mentioned above, a deletion of the *licT* gene resulted in loss of *ptsG*-R5 induction by salicin. However, the *licT* mutation did not abolish the induction by sucrose at the *ptsG*-R5 RAT (Table 2) suggesting that either SacT or SacY (or both) bind this RNA and cause antitermination. To test this possibility, we assayed the expression of the *ptsG*-R5-*lacZ* fusion in strains containing combinations of mutations of the three antiterminator genes. In a *licT sacY* double mutant, a slight reduction of sucrose induction of the *ptsG*-R5-*lacZ* fusion was observed as compared to the *licT* mutant strain (Table 2). In contrast, induction was completely lost in the *licT sacT* and the *licT sacT sacY* double and triple mutant strains carrying deletions of two and of three antiterminator protein-encoding genes, respectively. From this result we may conclude that SacT can recognize the *ptsG*-R5 RAT in addition to LicT. A northern blot analysis of the *lacZ* mRNA confirmed the strong induction by salicin and, to a lesser extent, by sucrose, conferred by the *ptsG*-R5 RAT (Figure 3). Taken together, these data

demonstrate that the U:A base pair just above the lower loop is important to facilitate binding of LicT to this structure (compare the high β-galactosidase activity to that driven by the *ptsG*-R1 fusion, Table 2) and to allow binding of SacT.

The role of the base at position 3 in the lower loop was also analyzed in the context of the *ptsG*-R5 RAT. However, as observed with *ptsG*-R4, only minor effects of a substitution of G3 by A were observed (Table 2, see *ptsG*-R6, see Figure 4). Induction by salicin was slightly decreased, and the induction with sucrose was also completely dependent on LicT indicating that SacT did not bind the *ptsG*-R6 RAT. A substitution of G3 by U (*ptsG*-R7, as present in the *sacB* RAT recognized by SacY, Figure 4) resulted in loss of induction by sucrose and only weak induction upon the addition of salicin (15-fold reduction as compared to *ptsG*-R5, see Table 2). Taken together, these results indicate that the G at position 3 facilitates binding of the antitermination proteins. In contrast, an U at this position strongly diminishes binding by LicT. These conclusions are validated by an analysis of the *sacB* RAT (see below).

Binding of the antitermination proteins to their RAT targets allows the formation of otherwise non-favoured RAT structures and prevents concomitantly the formation of the transcription terminators. The relative stability of the RAT structures may therefore be important for the level of gene expression. The RATs recognized by LicT, SacT, and SacY contain two G:C base pairs in the bottom stem whereas the *ptsG* RAT contains a A:U and a C:G base pair at this position. It seemed therefore possible that the replacement of the A:U base pair by a C:G base pair would result in a more stable RAT structure and thus affect transcription. To test this idea, the *ptsG*-R8 RAT was constructed based on *ptsG*-R5 and analyzed (see Figure 4, Table 2). While the *ptsG*-R5 RAT allowed only a weak basal expression in the absence of any inducer (CSE medium), a strongly increased basal expression was found for *ptsG*-R8 (34 versus 147 U of β-galactosidase). This was also reflected in a northern blot analysis of *lacZ* mRNA if expressed under the control of the *ptsG*-R5 and *ptsG*-R8 RAT (compare Figure 3, lanes 7 and 11). The *ptsG*-R8-*lacZ* fusion was also induced by salicin, and the induced expression was the sum of read through (~150 U) and real induction (~900 U, see *ptsG*-R5, Table 2). However, the increased read through might also result from a destabilization of the terminator even though an extra mutation was introduced in the terminator to restore base pairing.

Carbon catabolite repression interferes with the transport of glucose by BglP

The analysis of the *ptsG*-R5 RAT revealed that this structure is efficiently bound by LicT but not by GlcT. The disruption of the *glcT* gene in a strain carrying the *ptsG*-R5-*lacZ* fusion resulted in induction of β-galactosidase by salicin and sucrose (see Table 2) as expected due to the binding of LicT and SacT, respectively (see above). Surprisingly, glucose did also activate expression of this fusion in a *glcT* mutant strain. Since GlcT is not available in this mutant, LicT or SacT must be activated in the presence of glucose in the *glcT* mutant. To test this idea, we studied the activity of the *ptsG*-R5 control region in *glcT licT* or *glcT sacT* double mutants. As shown in Table 2, only a minor effect of the *sacT* deletion was observed, whereas the deletion of the *licT* gene resulted in complete loss of induction by glucose. Thus, glucose can activate LicT in a *glcT* mutant strain.

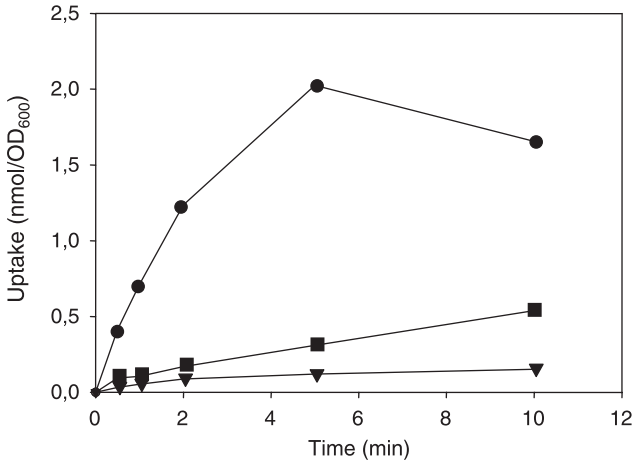

Figure 6. The implication of BglP in glucose uptake. For uptake measurements of radioactively labelled glucose the *B.subtilis* strains 168 (wild type, circles), QB5435 (*ptsG*, squares), and GP470 (*ptsG bglP*, triangles) were grown in CSE minimal medium supplemented with 0.5% of glucose.

Two scenarios for the activation of LicT by glucose can be envisaged. First, there might be some non-specificity in BglP that results in the transport of glucose by this permease and the subsequent dephosphorylation and activation of the cognate antiterminator LicT. Second, there might be some cross-talk between the glucose permease PtsG and the LicT antiterminator that results in LicT activation upon glucose transport. Several lines of evidence demonstrate that the former possibility reflects the truth: (i) The glucose permease PtsG is not expressed in a *glcT* mutant strain and is therefore unable to activate LicT in a *glcT* mutant (10,15). (ii) BglP phosphorylates and thereby inactivates LicT in the absence of the substrate salicin, and this regulation would be dominant over any minor PtsG-dependent dephosphorylation of LicT (21,22). (iii) To provide direct evidence for glucose uptake by BglG we measured the glucose transport of glucose-grown cells of a wild type strain (*B.subtilis* 168), a *ptsG* mutant (QB5435) and a *ptsG bglP* double mutant (GP470). As shown in Figure 6, glucose was efficiently transported by the wild type strain (initial uptake rate 620 ± 110 pmol glucose per minute and OD_{600}), whereas a significant reduction was observed in the *ptsG* mutant (initial uptake rate 62 ± 5 pmol glucose per minute and OD_{600}). These results are in good agreements with previous studies of glucose transport in *ptsG* mutants (10,14). In the *ptsG bglP* double mutant GP470, the transport of glucose was further reduced (see Figure 6, initial uptake rate 24 ± 1 pmol glucose per minute and OD_{600}), confirming that BglP has some minor glucose transport activity which may explain glucose-dependent activation of LicT in the *glcT* mutant background (see Discussion).

Conversion analysis of the *sacB* RAT sequence towards new recognition specificities

The similarity of the *ptsG*-R6 RAT to that of the *sacPA* operon (Figures 1 and 4) suggests that both RNA structures

might be recognized by the same proteins. However, as shown in Table 2, *ptsG*-R6 is bound exclusively by LicT whereas the *sacPA* RAT is the target of SacT and is not recognized by LicT (17). Similarly, the *ptsG*-R7 RAT which is poorly recognized by LicT but by none of the other antiterminator proteins resembles strongly the *sacB* RAT which is the target of SacY (see Figures 1 and 4). Thus, additional components seem to play a role in RAT-antiterminator protein recognition. To unravel these factors, we decided to perform an in-depth conversion analysis of the *sacB* RAT to mutate it and shift it gradually to sequences that are not longer recognized by SacB but rather by one of the three other family members. We chose the *sacB* RAT for this purpose since *sacB* lacks any additional regulation by carbon catabolite repression (37). This analysis was aimed at the identification of bases that are responsible for the specificity for one or the other antiterminator protein.

Discrimination between SacY and SacT

First, we determined the regulation mediated by the wild type *sacB* RAT. If the *lacZ* gene was expressed under the control of this RAT, induction was observed only in the presence of sucrose confirming that neither GlcT nor LicT bind the *sacB* RAT. Induction by sucrose occurred both at low and high sucrose concentrations which activate SacT and SacY, respectively. Indeed, induction at a low sucrose concentration was lost in the *sacT* mutant. In the *sacY* mutant strain, induction was still visible at both concentrations suggesting that SacT is active under both conditions. In the *sacT sacY* double mutant, the *sacB* RAT-terminator couple did not allow induction under all the condition tested (Table 3).

The *ptsG*-R5 RAT which is recognized by LicT and SacT, closely resembles the *sacB* RAT but contains a G at position 3 rather than a U as in the *sacB* RAT. We constructed therefore the *sacB*-R1 RAT by replacing U3 by a G (see Figure 7). This single mutation resulted in a significant specificity shift.

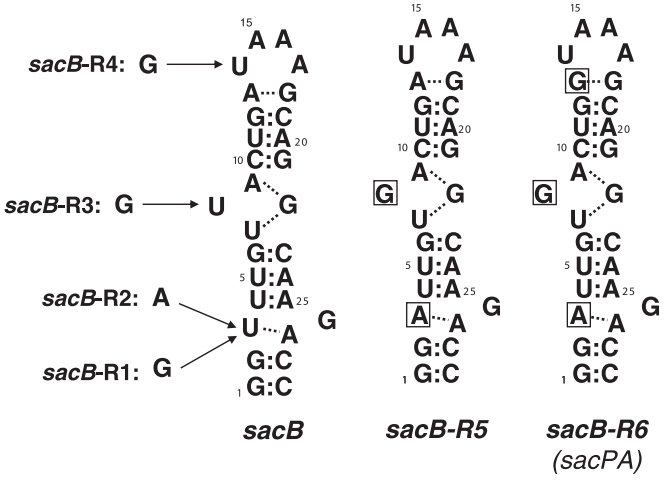

Figure 7. Secondary structures of the wild type *sacB*, *sacB*-R5, and *sacB*-R6 mutant RATs. Bases that differ from *sacB* RAT are boxed in the *sacB*-R5 double- and in the *sacB*-R6 triple mutant. The *sacB*-R1 to R4 mutant RATs are indicated as single base exchanges into *sacB* RAT. The *sacB*-R6 triple mutant RAT is identical to the *sacPA* wild type RAT (see Figure 1).

The *sacB*-R1 RAT was not longer a target for SacY, whereas the activation of SacT allowed a higher β-galactosidase expression as compared to the wild type *sacB* RAT (Table 3). Moreover, the *sacB*-R1 RAT allowed LicT-dependent induction by salicin. These results are in good general agreement with the observed affinity of the similar *ptsG*-R5 RAT for SacT and LicT, however, the preference for the two antitermination proteins was inverse. The *sacPA* RAT, which is the cognate target of SacT, contains also a purine base at position 3, i.e. an A (see Figure 1B). Therefore, the *sacB*-R2 containing an A at position 3 was constructed (Figure 7). The presence of this RAT conferred induction by sucrose but neither by salicin nor glucose (Table 3). Thus, this RAT is not bound by LicT. To distinguish whether it is recognized by SacY or SacT we analyzed the expression driven by the *sacB*-R2-*lacZ* fusion in *licT sacT* and *licT sacY* double mutant strains. As shown in Table 3, the *sacB*-R2 RAT is efficiently bound by SacT at both low and high sucrose concentrations whereas it is not recognized by SacY. The results obtained with the *sacB*-R1 and -R2 mutants suggest that the U at position 3 of the RAT is important for recognition by SacY. In contrast, SacT tolerates all three tested bases at this position.

The data presented above demonstrate that U3 is important for SacY binding in the context of the *sacB* RAT. However, since SacY is capable of recognizing the *sacPA* RAT (17), it seems to be able to accept bases different from U at position 3. To address this question we exchanged the three bases in the *sacB* RAT that are different from the *sacPA* RAT. A replacement of U8 in the middle loop by a G as in *sacPA* (*sacB*-R3, Figure 7) resulted in loss of binding by SacY whereas SacT bound this RAT as judged from loss of sucrose induction in the *sacT* mutant (see Table 3). Both the *sacB* and *sacPA* RATs contain a UAAA tetraloop at the top. This loop is flanked by A-G and G-G pairs in *sacB* and *sacPA*, respectively (see Figure 1B). Therefore, we constructed the *sacB*-R4 mutant RAT with a G-G pair at the bottom of the top-loop (A13G exchange, Figure 7). This mutation did not affect binding by SacY and SacT as compared to the wild type *sacB* RAT. Moreover, it did not confer induction by salicin indicating that it is no target for LicT (see Table 3). A combination of the two mutations of *sacB*-R2 and *sacB*-R3 (U3A and U8G) present in the *sacB*-R5 RAT (Figure 7) resulted in enhanced induction by sucrose which was exclusively dependent on SacT as determined using a *sacT* mutant strain (Table 2). Indeed, the *sacB*-R5 RAT was efficiently bound by the RNA-binding domain of SacT (see Figure 8). In contrast, the wild type *sacB* RAT was only weakly bound by SacT. These observations are in very good agreement with the high SacT-dependent induction of gene expression mediated by *sacB*-R5 as compared to induction conferred by the wild type *sacB* RAT. An additional mutation of the base pair at the bottom of the top loop (A13G) made the resulting *sacB*-R6 RAT identical to that of *sacPA*, but in a sequence context of *sacB* (see Figures 1B and 7). As expected, this RAT is most efficiently recognized by SacT. In the *sacT* mutant, only a very weak induction by sucrose was observed which was lost in the *sacT sacY* double mutant strain (Table 3). Thus, the A3 and G8 do both discriminate against binding by SacY. However, as demonstrated using the *sacB*-R6 RAT, the G-G base pair at

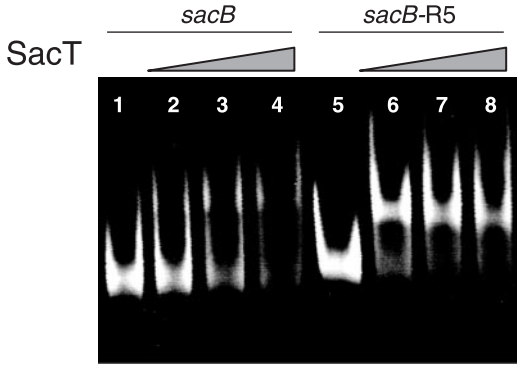

Figure 8. Electrophoretic mobility shift analysis of the interaction between the *sacB* and *sacB*-R5 RATs, and the RNA-binding domain of SacT. Lanes 1–4 and 5–8 contain 100 pmol of *sacB* and *sacB*-R5 RAT-RNAs, respectively. Increasing concentrations of SacT were added to the RNA in lanes 2–4 and 6–8 prior to electrophoresis. Aliquots of 75, 150 and 300 pmol SacT were used.

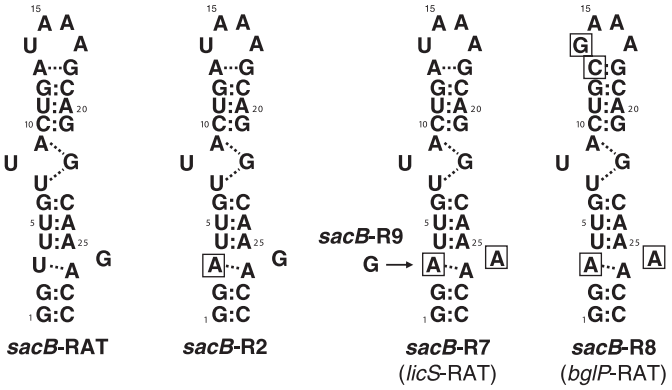

Figure 9. Secondary structures of the wild type *sacB*, *sacB*-R2, *sacB*-R7 and *sacB*-R8 RATs. Bases that differ from the *sacB* RAT are boxed. The mutations introduced into the *sacB* RAT convert it gradually to the *licS* (*sacB*-R7) and the *bglP* RATs (*sacB*-R8).

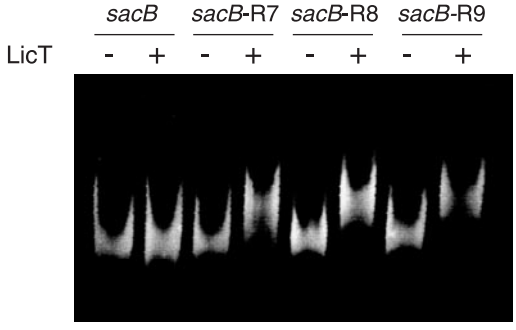

Figure 10. Electrophoretic mobility shift analysis of the interaction between the wild type *sacB* and several mutant RAT-RNAs (see Figure 8) with the RNA-bindig domain of LicT. In all lanes, 100 pmol of RNA were used. In the lanes labelled with '+', 250 pmol of LicT were added prior to electrophoresis.

the bottom of the top loop seems to weaken this discrimination and does thus allow weak binding by SacY.

Discrimination between SacY and LicT

The *sacB* RAT differs from the *licS* and *bglPH* RATs that are the cognate targets of LicT by two bases in the lower loop. Additionally, the *bglP* RAT contains a C-G base pair at the bottom of the top loop and a GAAA tetraloop at the top (see Figure 1B). The first step in the conversion of the *sacB* RAT to a structure expected to be recognized by LicT was the *sacB*-R2 mutation (U3A) described above. This RAT was bound by SacT but not by LicT (see Table 3). With the introduction of a second mutation in the lower loop (G26A) the resulting *sacB*-R7 RAT was identical to that of *licS* (see Figures 1B and 9). The determination of β-galactosidase regulation conferred by this RAT demonstrated induction not only by sucrose but also by salicin (see Table 3). Induction by salicin was completely lost in a LicT mutant whereas sucrose induction was lost in the *sacT* mutant strain. Binding of LicT to the *sacB*-R7 RAT was verified by an electrophoretic mobility shift assay. While the RNA-binding domain of LicT was unable to retard the wild type RAT of *sacB*, the *sacB*-R7 RAT was bound by this protein (see Figure 10). Thus, the A at position 26 is an important feature that makes the RAT a target for LicT. A mutation of 2 bp affecting the top loop converts the *licS*-RAT (*sacB*-R7) to the *bglP* RAT (*sacB*-R8). These mutations increase the affinity of both LicT and SacT as inferred from β-galactosidase activities of the *sacB*-R8-*lacZ* fusion strains (see Table 3). This finding is in good agreement with the observed stronger salicin-dependent induction of the *bglPH* operon as compared to the *licS* gene by the antiterminator LicT (20). As observed for the *sacB*-R7 RAT, the *sacB*-R8 RAT was bound by the RNA-binding domain of LicT *in vitro* (see Figure 10). The importance of A26 for recognition by LicT is underlined by the analysis of the *sacB*-R9 RAT in which the A at position 3 (present in *sacB*-R7 and the cognate targets of LicT) is replaced by a G (see Figure 9). As shown in Table 3, the presence of this RAT allows even higher LicT-dependent induction by salicin. Again, this RAT was recognized by LicT *in vitro* (see Figure 10). However, this mutation did not affect binding specificity since *sacB*-R9 was also a target for SacT. Thus, for LicT and SacT, position 3 seems to be most important to maintain the proper RAT structure, whereas the opposing A at position 26 is important for allowing efficient binding by LicT.

Discrimination between SacY and GlcT

The *ptsG* RAT is most different from all other RAT structures in *B.subtilis* due to the triple base pairing in the lower loop region (24, see above, Figure 1B). It has been proposed that this distinct structure rather than the details of the actual nucleotide sequence is important for recognition by GlcT. To verify this assumption we introduced mutations into the *sacB* RAT that made its structure gradually more similar to that of the *ptsG* RAT. In a first step, the U at position 4 was deleted (Figure 11). This mutation is a reversal of the conversion from the *ptsG* to the *ptsG*-R1 RAT (insertion of one base at position 4, see Figure 2), but in the context of the *sacB* RAT. The resulting *sacB*-R10 RAT allowed a very weak SacT-dependent induction by sucrose. In contrast, this RAT was not at all recognized by GlcT as concluded from the absence of induction by glucose. Here, the U3 might form a base pair with either A24 or A26 thus forming a structure weakly recognized and sufficiently stabilized by SacT

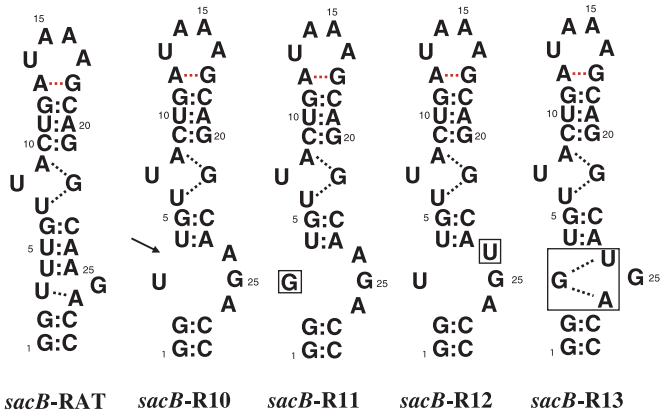

Figure 11. Gradual conversion of the lower loop region of the wild type *sacB* RAT to a stucture similar to that found in the *ptsG*-RAT. A deletion of an U in *sacB*-R10 is indicated by an arrow. Bases that differ from *sacB* RAT are boxed. The lower loop region in the *sacB*-R13 mutant is identical to that of the *ptsG*-RAT (see Figure 1).

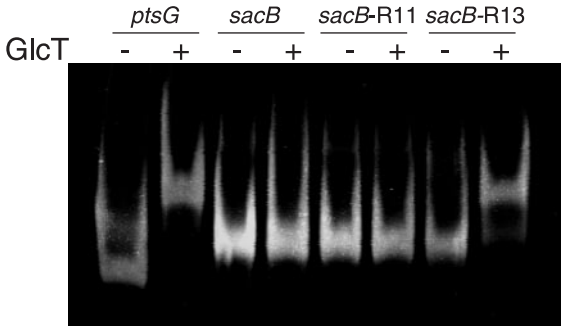

Figure 12. Electrophoretic mobility shift analysis of the interaction between the *ptsG* and *sacB* RAT-RNAs, several *sacB* mutant RAT-RNAs and the RNA-binding domain of GlcT. In all lanes, 100 pmol of RAT-RNA were used. In the lanes labelled with '+', 250 pmol of GlcT were added prior to electrophoresis.

to allow antitermination. The *sacB*-R10 mutant RAT was then parent to two further variants. In the *sacB*-R11-RAT, the U3 was replaced by a G (Figure 11). In the *ptsG* RAT, a G at this position contacts the nucleotides at positions 24 (U) and 26 (A). However, this does not seem to be the case with the two adenines in *sacB*-R11. This RAT does not confer induction to the *lacZ* gene, however, the read through was somewhat increased even in the absence of any inducer (see Table 3). A replacement of A24 present in *sacB*-R10 by a U (as in the *ptsG* RAT at this position) resulted in complete loss of expression of the reporter gene (see Table 3, *sacB*-R12, Figure 11). By replacing the U3 of *sacB*-R12 by a G, we obtained a lower loop that is identical to that found in the *ptsG* RAT (*sacB*-R13, Figure 11). Indeed, the *sacB*-R13-*lacZ* fusion was induced by glucose and sucrose (see Table 3). Since salicin and sucrose are known to activate GlcT (36), we tested the expression of this fusion in a *licT sacT sacY* triple mutant as well as in a *glcT* mutant strain. As expected, the combined deletion of *licT*, *sacT*, and *sacY* did not affect the induction by any of the sugars whereas no induction was observed in the *glcT* mutant strain (see Table 3). Thus, this RAT is exclusively recognized by GlcT. To verify this observation we performed electrophoretic mobility shift assays in the presence of the RNA-binding domain of GlcT using the RATs of *ptsG* and *sacB* as controls as well as the *sacB*-R11- and *sacB*-R13-RATs (see Figure 12). As previously observed, GlcT is capable of binding its cognate *ptsG* RAT. In contrast, the *sacB* RAT was not recognized by GlcT. Similarly, the *sacB*-R11-RAT was not retarded. As expected from the transcription regulation conferred by GlcT and the *sacB*-R13 RAT, an RNA-fragment containing this RAT was bound by the RNA-binding domain of GlcT. These findings confirm the important role of G3, U24, and A26 for the formation of the structure in the lower loop and in GlcT binding.

DISCUSSION

Several distinct mechanisms contribute to the specificity of the four antitermination systems present in *B.subtilis*. These include, first, the sugar permeases and their interactions with their substrates and with the cognate antitermination proteins. Second, carbon catabolite repression limits the conditions under which certain systems are expressed and the antiterminator proteins active. Finally, the interaction between the antiterminator proteins and the RAT-RNAs makes a major contribution to regulatory specificity.

The sugar permeases of the PTS can transport and phosphorylate only one substrate, or they can exhibit a relaxed specificity, i.e. they may transport more than one sugar. The glucose permease PtsG is known to transport sucrose and salicin in addition to glucose thus explaining the induction of *ptsG* expression by these sugars (36, see Table 2). A relaxed specificity has also been observed for the GlcB permease from *Staphylococcus carnosus* which is also capable of transporting salicin in addition to glucose (38). Similarly, the β-glucoside permease BglP is able to transport glucose, although with a low efficiency (see Figure 6). In wild type strains, the *bglP* gene is strongly repressed in the presence of glucose, thus, this relaxed specificity has no biological consequence. In contrast, the two sucrose permeases seem to be highly specific for sucrose, and the SacX permease is regarded as being inactive since it does not contribute to sucrose transport (39). All experiments with the different antitermination systems published so far did not provide any indication that a permease might interact with a non-cognate antiterminator protein. This might reflect the parallel evolution of the permeases and their targets, the PRD-I domains of the antitermination proteins (40). Indeed, the control of the antitermination proteins by the corresponding sugar-specific permeases works beyond the species barrier as shown for *B.subtilis* LicT in *E.coli* or *S.carnosus* GlcT in *B.subtilis* (20,41).

Bacteria use carbon sources in a hierarchical order, i.e. those that are most easily metabolized with a maximum yield of energy are preferred. In *B.subtilis*, glucose is the preferred carbon source, and the presence of glucose prevents the activity of many enzymes as well as the expression of genes and operons that are required for the utilization of alternative carbon sources. Among the genes studied here, only *ptsG* is induced by glucose (via antitermination) whereas *sacPA*, *bglPH* and *licS* are repressed. This repression is achieved by two independent mechanisms: First, the CcpA

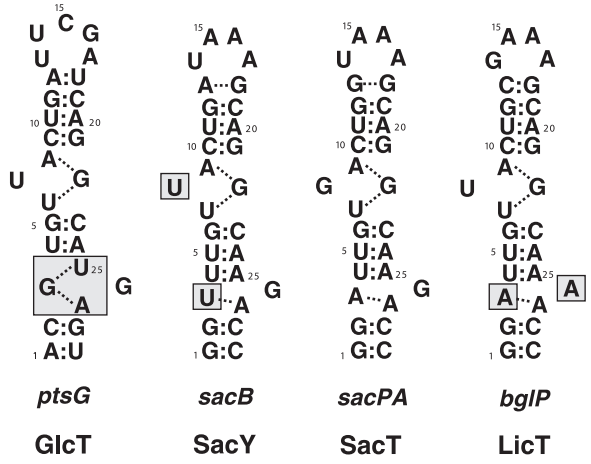

Figure 13. Summary of the relevant features that cause protein–RNA recognition specificity of the RAT-RNAs of *B.subtilis*. Boxes indicate nucleotides that switch specificity towards the appropriate antiterminator proteins. For *ptsG*, the structure of the lower loop region discriminates it from all the other RAT-stuructures and facilitates exclusive GlcT binding. SacY binding depends on the Us found at positions 3 and 8 in the *sacB* RAT. No specificity determinant could be found for SacT, as it binds to all RAT structures except for that of *ptsG*. SacT signalling specificity is achieved by the control of the protein's activity by dual PTS-dependent phosphorylation. The LicT targets (*bglP*, *licS*) are characterized by the essential As found at positions 3 and 26.

repressor protein binds target sites in the promoter regions of these genes and prevents their expression if glucose is present. Second, the antiterminator proteins SacT and LicT which are required for the expression of these genes, are inactive as long as glucose is present. In the absence of glucose they are phosphorylated at their PRD-II and thereby activated by HPr (18,22,42). These two mechanisms result in the absence of the BglP permease if glucose is available. Only in the *glcT* mutant strain, if glucose is unable to exert carbon catabolite repression since it can not efficiently be transported into the cell, the *bglP* gene can escape carbon catabolite repression and the BglP protein may exert a weak glucose transport activity which is sufficient for the activation of LicT (see Figure 6, Table 2).

A major specificity determinant in transcription regulation by the four antiterminator proteins is the RNA-protein interaction. As shown previously, the loop structures of the RATs are crucial for the specific recognition (23,24). In this work, we have identified all the factors that determine the specificity for any of the four antiterminator proteins (see Figure 13).

GlcT is unique in that it requires two identical and nearly symmetrical triple base pairings in the RAT. In contrast, LicT, SacT, and SacY bind RAT structures that resemble the upper triple base pair in the *ptsG* RAT, but differ significantly in the lower loop. A mutation that changes the lower loop of the *ptsG* RAT towards that found in the RAT recognized by the other antiterminator proteins (*ptsG*-R1, *ptsG*-R5, see Table 2), prevented recognition by GlcT and allowed binding by SacT and LicT. Similarly, a mutation of the *sacB* RAT which affected the lower loop and allowed the formation of a triple base pair, resulted in loss of SacY and SacT binding whereas GlcT recognized such a structure (*sacB*-R13, see Figures 11 and 12; Table 3).

LicT recognizes structures that are highly similar to the targets of SacT and SacY. An inspection of the RAT structures reveals that the LicT targets are unique in having an A at position 26 (see Figure 2B). The importance of this position is underlined by our mutation analysis of the *sacB* RAT. The importance of A26 for recognition by LicT is also supported by the structure of the LicT–RAT complex. There are several contacts of LicT with A26 and the sugar phosphate backbone in its immediate neighbourhood (23). The differential role of guanine and adenine residues for recognition of nucleic acids by proteins is well established (43). Moreover, the data indicate that the A at position 26 is necessary but not sufficient for LicT binding. In addition, a purine base is required at the opposing position 3 of the RAT. This is in good agreement with a previous study (19). Interestingly, a G at position 3 (*sacB*-R9) allows much higher LicT-, but also SacT-dependent antitermination as compared to a similar RAT containing an A at this position (*sacB*-R7).

There are conflicting reports on the recognition of the *sacB* RAT by SacT (17,19). We observed that the *sacB* RAT is recognized by both SacT and SacY. Interestingly, the SacT-dependent induction of *sacB* is stronger than the induction mediated by the cognate antiterminator, SacY. Thus, SacT induces both the *sacPA* operon and the *sacB* gene encoding levansucrase. In contrast, SacY exerts only a very minor effect at the *sacPA* RAT (identical to *sacB*-R6). Since SacT is active at both high and low sucrose concentrations whereas SacY is active only in the presence of large amounts of sucrose, SacT may be regarded as the major antiterminator protein controlling sucrose utilization. The minor role of SacY is also illustrated by the weak affinity of this protein to the *sacB* RAT which is two orders of magnitude lower than the affinities observed for LicT and GlcT with their respective targets (44, I. Langbein and J. Stülke, unpublished data). A step-wise conversion of the *sacB* RAT into a *sacPA* RAT like structure revealed the following observations: Single base mutations (Figure 7, *sacB*-R2, *sacB*-R3, *sacB*-R4, Table 3) all enhanced the binding of SacT, whereas the double and triple mutations (Figure 7, *sacB*-R5, *sacB*-R6) had additive effects. Most single mutations and the double mutation prevented SacY binding. However, the triple mutation which did also affect the top loop neutralized the negative effect of the two other nucleotides and restored binding of SacY.

Taken together, our data indicate that SacT is the most promiscuous of the antitermination proteins whereas GlcT at the other end of the spectrum is strictly confined to its cognate *ptsG* RAT due to its specific structural demands. In the living cell, glucose plays a special role as the by far most preferred carbon source. Therefore it is advantageous for the bacteria to have a regulatory system for glucose utilization that avoids any risk of cross-talk. On the other hand, SacT induces both sucrose catabolic systems, but it does not mediate antitermination at the *bglP* RAT in wild type bacteria (45). As shown here and in previous publications, SacT can bind *bglP*-like RAT structures (17, Table 3). It is so far unknown why SacT does not induce the *bglPH* operon in wild type bacteria. More factors such as the sequence context surrounding the RAT, the top loop, and the overall stability of the different RAT/terminator couples may provide additional levels for controlling the effective interaction with the antiterminator

proteins. A careful analysis of the data presented here indicates that this is indeed the case. More work will be required to study the contributions of these factors.

ACKNOWLEDGEMENTS

We are grateful to Matthias H. Schmalisch for helpful discussions, and Timo Hupfeld for the help with some experiments. Thorsten Mascher is acknowledged for the gift of primers for LFH-PCR. This work was supported by grants from the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie to J.S. Funding to pay the Open Access publication charges for this article was provided by the DFG.

Conflict of interest statement. None declared.

REFERENCES

1. Helmann,J.D. (1999) Anti-sigma factors. *Curr. Opin. Microbiol*., **2**, 135–141.
2. Stock,A.M., Robinson,V.L. and Goudreau,P.N. (2000) Two-component signal transduction. *Annu. Rev. Biochem*., **69**, 183–215.
3. Pané-Farré,J., Lewis,R.J. and Stülke,J. (2005) The RsbRST stress module in bacteria: a signalling system that may interact with different output modules. *J. Mol. Microbiol. Biotechnol*., **9**, 65–76.
4. Huffman,J.L. and Brennan,R.G. (2002) Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Curr. Opin. Struct. Biol*., **12**, 98–106.
5. Ramos,J.L., Martinez-Bueno,M., Molina-Henares,A.J., Teran,W., Watanabe,K., Zhang,X., Gallegos,M.T., Brennan,R. and Tobes,R. (2005) The TetR family of transcriptional repressors. *Microbiol. Mol. Biol. Rev*., **69**, 326–356.
6. Galvao,T.C. and de Lorenzo,V. (2006) Transcriptional regulators *à la carte*: engineering new effector specificities in bacterial regulatory proteins. *Curr. Opin. Biotechnol*., **17**, 34–42.
7. Garmendia,J., Devos,D., Valencia,A. and de Lorenzo,V. (2001) *À la carte* transcriptional regulators: unlocking responses of the prokaryotic enhancer-binding protein XylR to non-natural effectors. *Mol. Microbiol*., **42**, 47–59.
8. Hoch,J.A. and Varughese,K.I. (2001) Keeping signals straight in phosphorelay signal transduction. *J. Bacteriol*., **183**, 4941–4949.
9. Stülke,J. and Hillen,W. (2000) Regulation of carbon catabolism in *Bacillus* species. *Annu. Rev. Microbiol*., **54**, 849–880.
10. Stülke,J., Martin-Verstraete,I., Zagorec,M., Rose,M., Klier,A. and Rapoport,G. (1997) Induction of the *Bacillus subtilis ptsGHI* operon by glucose is controlled by a novel antiterminator, GlcT. *Mol. Microbiol*., **25**, 65–78.
11. Fillinger,S., Boschi-Muller,S., Azza,S., Dervyn,E., Branlant,G. and Aymerich,S. (2000) Two glyceraldehyde-3-phosphate dehydrogenases with opposite physiological roles in a nonphotosynthetic bacterium. *J. Biol. Chem*., **275**, 14031–14037.
12. Ludwig,H., Homuth,G., Schmalisch,M., Dyka,F.M., Hecker,M. and 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.
13. Doan,T. and Aymerich,S. (2003) Regulation of the central glycolytic genes in *Bacillus subtilis*: binding of the repressor CggR to its single DNA target sequence is modulated by fructose-1,6-bisphosphate. *Mol. Microbiol*., **47**, 1709–1721.
14. Bachem,S., Faires,N. and Stülke,J. (1997) Characterization of the presumptive phosphorylation sites of the *Bacillus subtilis* glucose permease by site-directed mutagenesis: Implication in glucose transport and catabolite repression. *FEMS Microbiol. Lett*., **156**, 233–238.
15. Bachem,S. and Stülke,J. (1998) Regulation of the *Bacillus subtilis* GlcT antiterminator protein by components of the phosphotransferase system. *J. Bacteriol*., **180**, 5319–5326.
16. Schmalisch,M., Bachem,S. and 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.
17. Manival,X., Yang,Y., Strub,M.P., Kochoyan,M., Steinmetz,M. and Aymerich,S. (1997) From genetic to structural characterization of a new class of RNA-binding domain within the SacY/BglG family of antiterminator proteins. *EMBO J*., **16**, 5019–5029.
18. Stülke,J., Arnaud,M., Rapoport,G. and Martin-Verstraete,I. (1998) PRD—a protein domain involved in PTS-dependent induction and carbon catabolite repression of catabolic operons in bacteria. *Mol. Microbiol*., **28**, 865–874.
19. Aymerich,S. and Steinmetz,M. (1992) Specificity determinants and structural features in the RNA target of the bacterial antiterminator proteins of the BglG/SacY family. *Proc. Natl Acad. Sci. USA*, **89**, 10410–10414.
20. Schnetz,K., Stülke,J., Gertz,S., Krüger,S., Krieg,M., Hecker,M. and Rak,B. (1996) LicT, a *Bacillus subtilis* transcriptional antiterminator of the BglG family. *J. Bacteriol*., **178**, 1971–1979.
21. Tortosa,P., Aymerich,S., Lindner,C., Saier,M.H., Jr, Reizer,J. and Le Coq,D. (1997) Multiple phosphorylation of SacY, a *Bacillus subtilis* antiterminator negatively controlled by the phosphotransferase system. *J. Biol. Chem*., **272**, 17230–17237.
22. Lindner,C., Galinier,A., Hecker,M. and Deutscher,J. (1999) Regulation of the activity of the *Bacillus subtilis* antiterminator LicT by multiple, enzyme I- and HPr-catalysed phosphorylation. *Mol. Microbiol*., **31**, 995–1006.
23. Yang,Y., Declerck,N., Manival,X., Aymerich,S. and Kochoyan,M. (2002) Solution structure of the LicT-RNA antitermination complex: CAT clamping RAT. *EMBO J*., **21**, 1987–1997.
24. Schilling,O., Langbein,I., Müller,M., Schmalisch,M.H. and Stülke,J. (2004) A protein dependent riboswitch controlling *ptsGHI* operon expression in *Bacillus subtilis*: RNA structure rather than sequence provides interaction specificity. *Nucleic Acids Res*., **32**, 2853–2864.
25. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
26. Faires,N., Tobisch,S., Bachem,S., Martin-Verstraete,I., Hecker,M. and Stülke,J. (1999) The catabolite control protein CcpA controls ammonium assimilation in *Bacillus subtilis*. *J. Mol. Microbiol. Biotechnol*., **1**, 141–148.
27. Kunst,F. and Rapoport,G. (1995) Salt stress is an environmental signal affecting degradative enzyme synthesis in *Bacillus subtilis*. *J. Bacteriol*., **177**, 2403–2407.
28. Wach,A. (1996) PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in *S. cerevisiae*. *Yeast*, **12**, 259–265.
29. Youngman,P. (1990) Use of transposons and integrational vectors for mutagenesis and construction of gene fusions in *Bacillus subtilis*. In Harwood,C.R. and Cutting,S.M. (eds), *Molecular Biological Methods for Bacillus*. John Wiley & sons, Chichester, UK, pp. 221–266.
30. Guérout-Fleury,A.M., Shazand,K., Frandsen,N. and Stragier,P. (1995) Antibiotic resistance cassettes for *Bacillus subtilis*. *Gene*, **167**, 335–336.
31. Jordan,S., Junker,A., Helmann,J.D. and Mascher,T. (2006) Regulation of LiaRS-dependent gene expression in *Bacillus subtilis*: identification of inhibitor proteins, regulator binding sites, and target genes of a conserved cell envelope stress-sensing two-component system. *J. Bacteriol*., **188**, 5153–5166.
32. Weinrauch,Y., Msadek,T., Kunst,F. and Dubnau,D. (1991) Sequence and properties of *comQ*, a new competence regulatory gene of *Bacillus subtilis*. *J. Bacteriol*., **173**, 5685–5693.
33. Steinmetz,M., Le Coq,D., Aymerich,S., Gonzy-Tréboul,G. and Gay,P. (1985) The DNA sequence of the gene for the secreted *Bacillus subtilis* enzyme levansucrase and its genetic control sites. *Mol. Gen. Genet*., **200**, 220–228.
34. Bi,W. and Stambrook,P.J. (1998) Site-directed mutagenesis by combined chain reaction. *Anal. Biochem*., **256**, 137–140.
35. Hames,C., Halbedel,S., Schilling,O. and Stülke,J. (2005) Multiple-mutation reaction: a method for simultaneous introduction of multiple mutations into the *glpK* gene of *Mycoplasma pneumoniae*. *Appl. Environ. Microbiol*., **71**, 4097–4100.
36. Langbein,I., Bachem,S. and Stülke,J. (1999) Specific interaction of the RNA-binding domain of the *Bacillus subtilis* transcriptional

antiterminator GlcT with its RNA target, RAT. *J. Mol. Biol*., **293**, 795–805.
37. Steinmetz,M., Le Coq,D. and Aymerich,S. (1989) Induction of saccharolytic enzmyes by sucrose in *Bacillus subtilis*: evidence for two partially interchangeable regulatory pathways. *J. Bacteriol*., **171**, 1519–1523.
38. Christiansen,I. and Hengstenberg,W. (1999) Staphylococcal phosphoenolpyruvate-dependent phosphotransferase system—two highly similar glucose permeases in *Staphylococcus carnosus* with different glucoside specificity: protein engineering *in vivo*? *Microbiology*, **145**, 2881–2889.
39. Kunst,F., Pascal,M., Lepesant,J.A., Walle,J. and Dedonder,R. (1974) Purification and some properties of an endocellular sucrase from a constitutive mutant of *Bacillus subtilis* 168. *Eur. J. Biochem*., **42**, 611–620.
40. Greenberg,D.G., Stülke,J. and Saier,M.H. (2002) Domain analysis of transcriptional regulators bearing PTS regulatory domains. *Res. Microbiol*., **153**, 519–526.
41. Knezevic,I., Bachem,S., Sickmann,A., Meyer,H.E., Stülke,J. and Hengstenberg,W. (2000) Regulation of the glucose-specific phosphotransferase system (PTS) of *Staphylococcus carnosus* by the antiterminator protein GlcT. *Microbiology*, **146**, 2333–2342.
42. Krüger,S., Gertz,S. and Hecker,M. (1996) Transcriptional analysis of *bglPH* expression in *Bacillus subtilis*: evidence for two distinct pathways mediating carbon catabolite repression. *J. Bacteriol*., **178**, 2637–2644.
43. Nobeli,I., Laskowski,R.A., Valdar,W.S.J. and Thornton,J.M. (2001) On the molecular discrimination between adenine and guanine by proteins. *Nucleic Acids Res*., **29**, 4294–4309.
44. Declerck,N., Vincent,F., Hoh,F., Aymerich,S. and van Tilbeurgh,H. RNA recognition by transcriptional antiterminators of the BglG/SacY family: functional and structural comparison of the CAT domain from SacY and LicT. *J. Mol. Biol*., **294**, 389–402.
45. Krüger,S. and Hecker,M. (1995) Regulation of the putative *bglPH* operon for aryl-β-glycoside utilization in *Bacillus subtilis*. *J. Bacteriol*., **177**, 5590–5597.
46. Le Coq,D., Lindner,C., Krüger,S., Steinmetz,M. and Stülke,J. (1995) New β-glucoside (*bgl*) genes in *Bacillus subtilis*: the *bglP* gene product has both transport and regulatory functions similar to those of BglF, its *Escherichia coli* homolog. *J. Bacteriol*., **177**, 1527–1535.