

# *Corynebacterium diphtheriae*: a PTS View to the Genome

Stephan Parche, Andreas W. Thomae,  
Maximilian Schlicht, and Fritz Titgemeyer\*

Lehrstuhl für Mikrobiologie, Friedrich-Alexander-  
Universität Erlangen-Nürnberg, Staudtstrasse 5, 91058  
Erlangen, Germany

## Abstract

We have surveyed the publicly available genome sequence of *Corynebacterium diphtheriae* ([www.sanger.ac.uk](http://www.sanger.ac.uk)) to identify components of the phosphotransferase system (PTS), which plays a central role in carbon metabolism in many bacteria. Three gene loci were found to contain putative *pts* genes. These comprise: (i) the genes of the general phosphotransferases enzyme I (*ptsI*) and HPr (*ptsH*), a fructose-specific enzyme IIABC permease (*fruA*), and a fructose 1-phosphate kinase (*fruK*); (ii) a gene that encodes an enzyme IIAB of the fructose/mannitol family, and a novel HPr-like gene, *ptsF*, that encodes an HPr domain fused to a domain of unknown function; (iii) and a gene for a glucose-specific enzyme IIBC (*ptsG*). A search for genes that may be putative PTS-targets or that may operate in general carbon regulation revealed a possible regulatory gene encoding an antiterminator protein downstream from *ptsG*. Furthermore, genes were detected encoding glycerol kinase, glucose kinase, and a homologue of the global activator of carbon catabolite repression in *Escherichia coli*, CAP. The possible significance of these observations in carbon metabolism and the novel features of the detected genes are discussed.

## Introduction

*Corynebacterium diphtheriae* is a non-differentiating aerobic actinomycete with a chinese character-like cell shape (Collins and Cummins, 1986). It is the causative agent of diphtheria, a highly infectious disease. Due to thorough vaccination, diphtheria is nowadays rare in the Western world. However in 1993/94, there were over 50,000 reported cases in the former USSR. The pathogenic properties of *C. diphtheriae* make it a difficult object for *in vivo* investigation and therefore, genome analysis is a powerful method to gain information about its biology (Saier, 2000).

*C. diphtheriae* can utilize numerous carbon sources including glucose, fructose, and maltose (Collins and Cummins, 1986). However, detailed data on carbon (C-)metabolism and C-regulation are not available.

Knowledge on C-metabolism in other actinomycetes is also fragmentary. By contrast, diverse mechanisms of C-regulation have been well-documented in species of Gram-negative and low-GC Gram-positive bacteria. In both bacterial groups the PTS plays a key role (for a review see Reizer *et al.*, 1988; Postma *et al.*, 1993). This system functions by protein phosphorylation as follows. The general phosphotransferase enzyme I (EI; gene *ptsI*) becomes autophosphorylated by PEP and transfers its phosphoryl group to the second general component HPr (*ptsH*). HPr in turn phosphorylates a number of sugar-specific permeases, the so-called enzyme IIABC(D)<sup>sugar</sup>-complexes (Saier and Reizer, 1992; Lengeler *et al.*, 1994). These complexes transport their substrates by concomitant phosphorylation. Consequently, PTS proteins are predominantly dephosphorylated as long as PTS substrates are available. The cell uses the fluctuations in the phosphorylation state of PTS proteins to exert C-regulation by mechanisms known as inducer exclusion and carbon catabolite repression. As an example, *E. coli* takes advantage of a non-phosphorylated IIA<sup>Glc</sup> to block the activities of certain sugar permeases and glycerol kinase (Saier, 1993). The phosphorylated form of IIA<sup>Glc</sup> stimulates adenylate cyclase to synthesize cAMP, which is the co-effector for the catabolite activator protein CAP that globally triggers carbon catabolite repression. In low-GC Gram-positive bacteria a major regulatory function is attributed to HPr. Via its phosphorylation state, HPr controls the activities of glycerol kinase, a set of antiterminator proteins that regulate genes required for utilization of glucose, sucrose, and  $\beta$ -glucosides, and the carbon catabolite repression protein, CcpA (Stülke *et al.*, 1998; Mahr *et al.*, 2000).

In actinomycetes, the presence of a PTS has been described for *Corynebacterium glutamicum* and for species of *Streptomyces*. The amino-acid producer *C. glutamicum* uses PTSs for the uptake of glucose, fructose, mannose, and sucrose (Shiio *et al.*, 1990; Lee *et al.*, 1993; Dominguez and Lindley, 1996). However, beside a gene encoding a glucose-specific enzyme II, none of the PTS components are described at the molecular level (Lee *et al.*, 1994). We have discovered the PTS in the genus *Streptomyces* (Titgemeyer *et al.*, 1994). These differentiating and antibiotic-producing bacteria appear to possess a fructose-specific PTS but lack a glucose PTS (Titgemeyer *et al.*, 1995; Parche *et al.*, 1999). Our analysis of the *S. coelicolor* genome sequence revealed that this microorganism possesses a set of four PTS permeases. A role of the PTS in C-regulation has so far not been demonstrated in *Streptomyces*, and this issue has not yet been addressed for species of *Corynebacteria* (Butler *et al.*, 1999; Parche *et al.*, 1999; Mahr *et al.*, 2001; Parche *et al.*, 2001).

In this communication we analyze genome data of the ongoing sequencing project of *C. diphtheriae*. We present a survey of *pts* gene homologues and genes that may be associated with PTS function and C-regulation.

\*For correspondence. Email [ftitgem@biologie.uni-erlangen.de](mailto:ftitgem@biologie.uni-erlangen.de);  
Tel. +49-9131-8528095; Fax. +49-9131-8528082.

## gene locus I

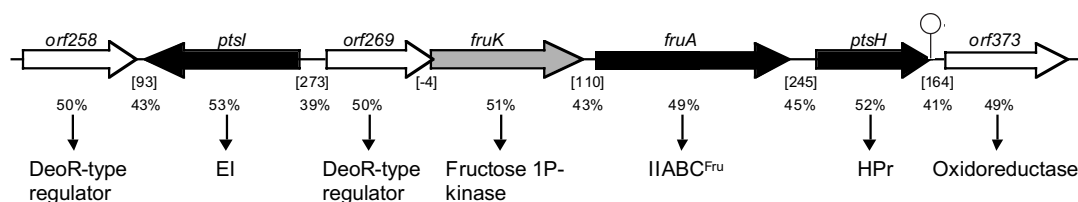


Figure 1. Gene locus I. Gene locus I includes the genes *ptsI*, *ptsH*, *fruA*, *fruK*, three *orfs* encoding two putative transcriptional regulators of the DeoR family (*orf258*, *orf269*), and a putative oxidoreductase (*orf373*). *orf* numbers denote the lengths of the gene products. Genes of the PTS are shaded black, and PTS-associated genes are labeled gray. The %GC content of coding and intergenic regions is specified. Numbers in brackets show the lengths of intergenic regions in bp. The LASERGENE WORKSTATION software (DNASTAR, Inc.) was used to process DNA sequence data. DNA databank and protein databank searches were performed using the BLAST server of the National Center of Biotechnology Information at the National Institutes of Health Bethesda, MD, USA ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)).

## Overview

The Sanger Centre (Cambridge, UK) has almost completed sequencing of the 2.8 Mb genome of strain *C. diphtheriae* NCTC13129 ([www.sanger.ac.uk](http://www.sanger.ac.uk)), which is a clinical representative of an epidemic clone circulating within Eastern Europe. The results of the survey presented in Table 1 denote six *pts* gene homologues. This set indicates that *C. diphtheriae* has a fructose- and glucose-specific PTS at its disposal. Hence, *C. diphtheriae* can be grouped into the class of bacteria that possess one to four PTSs like *Haemophilus influenzae* and *Mycoplasma* species (Reizer *et al.*, 1996b; Reizer *et al.*, 1996c). In comparison to that, bacteria such as *E. coli* or *Bacillus subtilis* can have more than fifteen PTSs, or can possess only the general phosphotransferases EI and HPr but lack PTS permeases (*Treponema pallidum*), or completely lack PTS gene

homologues (*Mycobacterium tuberculosis*) (Postma *et al.*, 1993; Cole *et al.*, 1998; Fraser *et al.*, 1998; Reizer *et al.*, 1999a).

Table 1 also provides a listing of genes (i) that may encode proteins required for utilization of glucose and fructose, (ii) that may encode PTS-interacting components, or (iii) that may be involved in C-regulation. Most of these genes have their closest homologue with genes of other actinomycetes, reflecting the phylogenetic grouping of *C. diphtheriae*. In this article we highlight novel features and discuss the physiological significance of the detected genes with respect to PTS-mediated carbon utilization and regulation.

Table 1. Genes encoding PTS homologues and PTS associated proteins of *C. diphtheriae*

| gene <sup>1</sup>                | function <sup>1</sup> | domain order <sup>2</sup> | no. of residues <sup>3</sup> | representative neighbor/ accession no. <sup>4</sup> |  |
|----------------------------------|-----------------------|---------------------------|------------------------------|---|--|
| general energy coupling proteins |                       |                           |                              |   |  |
| <i>ptsH</i>                      | encoded protein       |                           |                              |   |  |
|                                  | HPr                   | phosphotransferase        | H                            | 88  | HPr ( <i>S. coelicolor</i> ) / T35877                    |
| <i>ptsF</i>                      | PtsF                  | phosphotransferase        | FH                           | 222   | HPr ( <i>S. coelicolor</i> ) / T35877                    |
| <i>ptsI</i>                      | enzyme I              | phosphotransferase        | I                            | 560   | EI ( <i>S. coelicolor</i> ) / CAB88887                   |
| fructose / mannitol family       |                       |                           |                              |   |  |
| <i>fruA</i>                      | putative substrate    |                           |                              |   |  |
|                                  | fructose              | fructose permease         | IIABC                        | 670   | EII <sup>Fru</sup> ( <i>S. coelicolor</i> ) / CAB90980   |
| <i>ptx</i>                       | unknown               | phosphotransferase        | IIB                          | 264   | Ptx ( <i>S. coelicolor</i> ) / T37607                    |
| glucose / sucrose family         |                       |                           |                              |   |  |
| <i>ptsG</i>                      | glucose               | glucose permease          | IIBCA                        | 676   | EII <sup>Glc</sup> ( <i>C. ammoniagenes</i> ) / ACC27701 |
| PTS associated proteins          |                       |                           |                              |   |  |
| <i>orf275</i>                    | encoded protein       |                           |                              |   |  |
|                                  | GlcT                  | antiterminator            |                              | 275   | LicT ( <i>B. subtilis</i> ) / P39805                     |
| <i>fruK</i>                      | FruK                  | fructose 1P-kinase        |                              | 321   | FruK ( <i>S. coelicolor</i> ) / CAB90981                 |
| <i>sgaT</i>                      | SGAT                  | permease                  |                              | 520   | SGAT ( <i>S. coelicolor</i> ) / T37066                   |
| <i>orf224</i>                    | CAP-type              | regulator                 |                              | 224   | CAP-type ( <i>M. tuberculosis</i> ) / E70790             |
| <i>orf507</i>                    | GlpK                  | glycerol kinase           |                              | 507   | GlpK ( <i>M. tuberculosis</i> ) / O69664                 |
| <i>orf351</i>                    | Lacl-type             | regulator                 |                              | 351   | Lacl-type ( <i>S. coelicolor</i> ) / CAB66284            |

<sup>1</sup> Proposed designations of the genes and the proposed functional assignment of the gene products are based on sequence comparisons.

<sup>2</sup> Protein domains include: I, EI; H, HPr; IIA, the domain bearing the first permease-specific phosphorylation site; IIB, the domain bearing the second permease-specific phosphorylation site; IIC, the hydrophobic transmembrane permease domain; F, hypothetical domain of PtsF.

<sup>3</sup> Length of the encoded proteins: the most likely start codon was chosen. It appeared that beside ATG, GTG is also used as a frequent start codon. Since some of the gene sequences apparently contain sequencing errors, the actual sizes of *orfs* may be subject to change.

<sup>4</sup> Representative neighbors were chosen either as the closest homologues or, if available, as well-characterized homologues. Accession numbers from SWISSPROT have the prefix O or P while others are from GenBank.

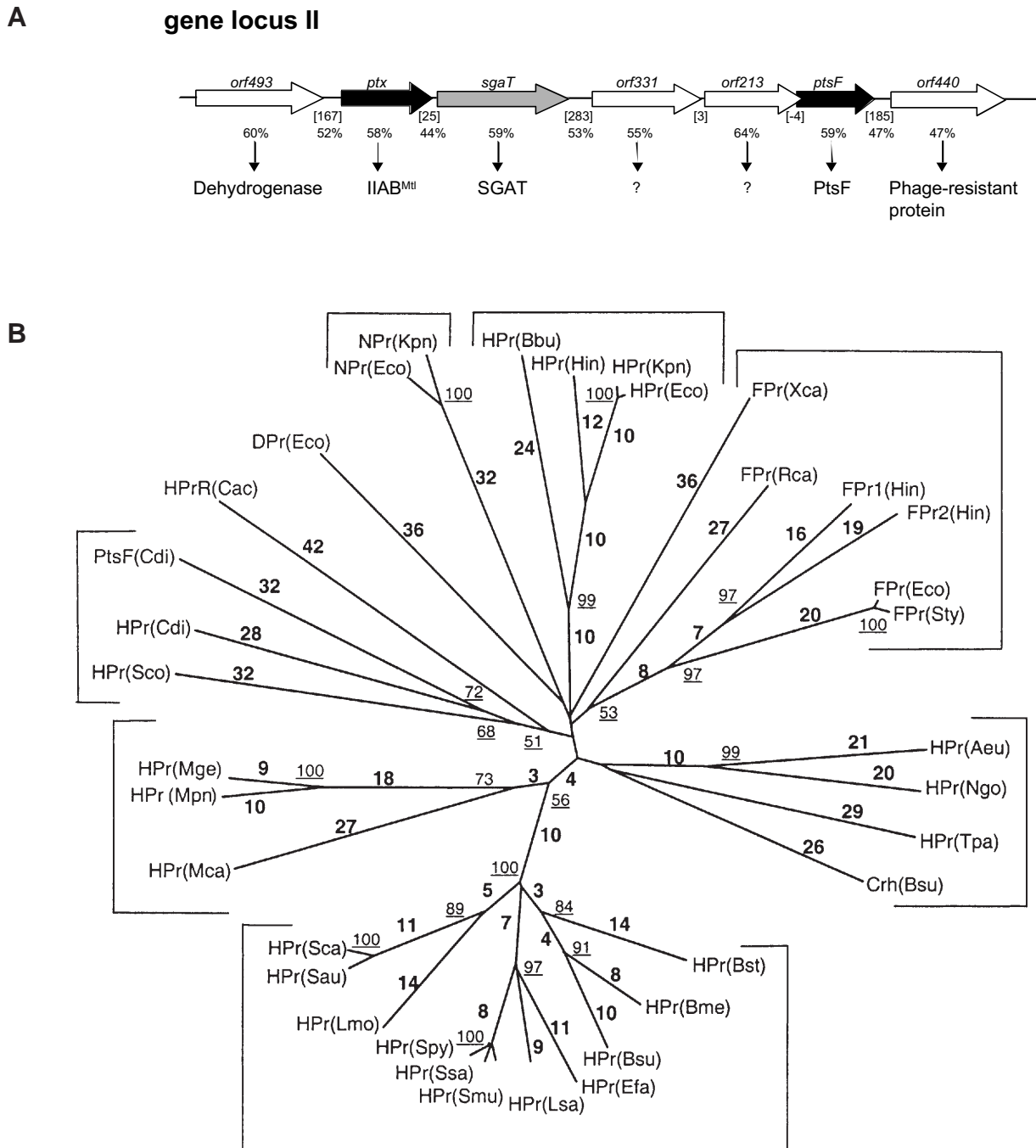


Figure 2. Gene locus II. A. Genetic map. The depicted region includes the genes *ptx*, *sgaT*, and *ptsF* together with the flanking genes encoding a putative dehydrogenase (493 aa), a phage resistant protein (440 aa), and two genes encoding proteins of unknown function (*orf331*, *orf213*). The convention of presentation is as described in the legend to Figure 1. B. Phylogenetic tree of the HPr family. An unrooted phylogenetic tree was computed with the CLUSTALW software using the implemented neighbor joining method. Branch lengths are labeled with arbitrary units (Higgins *et al* 1996). Bootstrap values (1000 sample runs) are expressed in percentages and underlined. A bootstrap value of 95% suggests a correct topology. Bootstrap values below 50% found for very short branches are not labeled. Abbreviations are as follows: HPr and PtsF proteins from Cdi, *C. diphtheriae* ([www.sanger.ac.uk](http://www.sanger.ac.uk)); Cac, *Clostridium acetobutylicum* ([www.cric.com/genesequences/clostridium/clospage.html](http://www.cric.com/genesequences/clostridium/clospage.html)); Sco, *S. coelicolor* (AL009204); Bsu, *Bacillus subtilis* (P08877); Bme, *Bacillus megaterium* (AJ005075); Bst, *Bacillus stearothermophilus* (P42013); Smu, *Streptococcus mutans* (P45596); Ssa, *Streptococcus salivarius* (P24366); Spy, *Streptococcus pyogenes* (contig 286); Lsa, *Lactobacillus sake* (P007125); Efa, *Enterococcus faecalis* (P07515); Sau, *Staphylococcus aureus* (P02907); Sca, *Staphylococcus carnosus* (P23534); Lmo, *Listeria monocytogenes* (O31148); Mca, *Mycoplasma capricolum* (P45611); Mpn, *Mycoplasma pneumoniae* (P75061); Mge, *Mycoplasma genitalium* (P47287); Psa, *Pseudomonas aeruginosa* (contig 299); Aeu, *Alcaligenes eutrophus* (P23537); Ngo, *Neisseria gonorrhoeae* (contig 253); Tpa, *Treponema pallidum* (AE001234); Eco, *Escherichia coli* (P07006); Kpn, *Klebsiella pneumoniae* (P16481); Hin, *Haemophilus influenzae* (P43921); Bbu, *Borrelia burgdorferi* (AE001157). FPr protein domains are from Eco, *E. coli* (P24217); Sty, *Salmonella typhimurium* (P17127); Hin, *H. influenzae* (P44715); Rca, *Rhodobacter capsulatus* (P23388); Xca, *Xanthomonas campestris* (P45597). NPR proteins are from Eco, *E. coli* (P33996) and Kpn, *K. pneumoniae* (P51185). Crh is from *B. subtilis* (Z94043). Accession numbers from SWISSPROT have the prefix 'O' or 'P', while others are from GenBank. Sequences from preliminary versions of the sequenced genomes have the contig number as identifier.

### Gene Locus I: a Complete Fructose-PTS

A screen for the general PTS components EI and HPr revealed one *ptsI*-like *orf* that was separated by 4.0 kb from a *ptsH*-like *orf* transcribed in opposite orientation (Figure 1A). The deduced protein sequences share the highest similarity with the respective genes of *S. coelicolor* (Parche *et al.*, 2001). The protein lengths of EI with 560 amino acids and HPr with 88 amino acids are typical, and the putative catalytic histidine sites (GGPTSHTAIAR in EI; GLHARPA in HPr) are well-conserved. HPr also carries a serine 49 that could represent a regulatory phosphorylation site.

Between *ptsI* and *ptsH*, three genes were identified that encode a fructose-specific enzyme II, a fructose 1-phosphate kinase, and a regulator of the DeoR family, giving the gene order *ptsI*, *orf269*, *fruK*, *fruA*, *ptsH*. This gene locus may comprise all genes required for PTS-dependent fructose utilization. The coding regions exhibit a GC content between 49-53%, which is in good agreement with the overall GC value of 52% of *C. diphtheriae* (www.sanger.ac.uk\Projects\Microbes\). The intergenic regions contain lower GC values ranging from 39-43%. This is consistent with observations in other high-GC bacteria, where promoter-containing regions exhibit a decreased GC content (Strohl, 1992).

Enzyme II<sup>Fructose</sup> has the domain order IIABC comprising putative HPr phosphorylation sites at a histidine within the IIA domain (IPHCRT) and at a cysteine within the IIB domain (ATTACPTGI). The domain sequence IIABC is a characteristic feature of enzyme II<sup>Fructose</sup> permeases of Gram-positive bacteria and *Mycoplasma* species (Parche *et al.*, 2001). The homologous II<sup>Fructose</sup> proteins of Gram-negative bacteria, by contrast, possess a IIB'BC domain arrangement. The gene order *fruKA* is similar to the putative *fruKA* operon found in *S. coelicolor* (Parche *et al.*, 2001). However, the genetic surrounding differs. *S. coelicolor* has a monocistronic *ptsH* gene and a *ptsI* gene that is adjacent to a gene encoding a IIA<sup>Glc</sup>-like protein at a distant gene locus.

This unique organization of the fructose-PTS may shed some light on the evolution of the PTS. The fructose-PTS has been considered as the archae-type of the system (Saier *et al.*, 1985), (i) because it is the most widespread PTS permease, and (ii) because Gram-negative bacteria such as *Xanthomonas campestris* or *Rhodobacter capsulatus* possess additional fructose-specific EI and HPr protein domains sometimes embedded in so-called three-

domain MTPs (multiphosphoryl protein; EI-HPr-IIA<sup>Fructose</sup>) (Wu *et al.*, 1990). Hence, gene locus I perhaps comprises a genetic arrangement that may represent a link between the evolution of a co-ordinately controlled *pts* operon with all five PTS domains (I, H, IIA, IIB, IIC) and a situation in which HPr and EI became independent of one PTS to also serve other enzyme II permeases. To substantiate this, the question whether *ptsH* and *ptsI* of *C. diphtheriae* are fructose-specific or of general specificity should be addressed. The single domain organization and the evolutionary relationship would favor that the latter is true (Figure 2B).

### Gene Locus II: IIAB-Type and HPr-Type Proteins

Screening of *ptsH* homologues led to the identification of an *orf222* encoding an HPr domain that is fused to a protein of about 140 amino acids. Interestingly, this two-domain protein is homologous to the ADI protein (*ptsD*) of *E. coli*, which in addition contains an EI-phosphorylation domain (Reizer and Saier, 1997; Paulsen *et al.*, 2000). We designated *orf222 ptsF*. Phylogenetic analysis revealed that the HPr domain falls into a loose cluster with HPrs of actinomycetes (Figure 2B). The N-terminal domain shares significant sequence identities with the N-terminal domain of PtsD and with a few hypothetical proteins of about 130 amino acids. PtsF belongs, like PtsD, DTP, and the recently described HPrR to a group of proteins that form hybrids with domains of yet unknown function (Reizer and Saier, 1997; Reizer *et al.*, 1999b). Therefore, it will be a great asset to elucidate the physiological significance of these proteins.

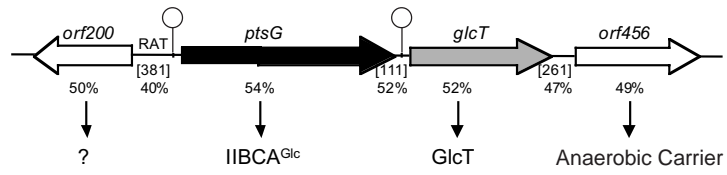
The fourth *orf* upstream of *ptsF* encodes an enzyme IIAB of the fructose/mannitol family that shows highest similarity to *ptx* of *S. coelicolor* (Figure 2A) (Parche *et al.*, 2001). As in *S. coelicolor*, *ptx* is followed by an *sgaT*-like gene that encodes a putative permease. Such a genetic organization may be a good hint as to a common physiological role and it may be conceivable that PtsF is involved in Ptx function. *ptx* has been discovered in *E. coli*, and a possible role in transport of pentitol-like substrates was suggested (Reizer *et al.*, 1996a).

Another interesting feature of gene locus II is that the genes from *orf493* to *ptsF* exhibit an increased GC content between 55 and 64%, which is significant higher than the genomic GC content of *C. diphtheriae* and could therefore be an indication that this region has been acquired by horizontal gene transfer.

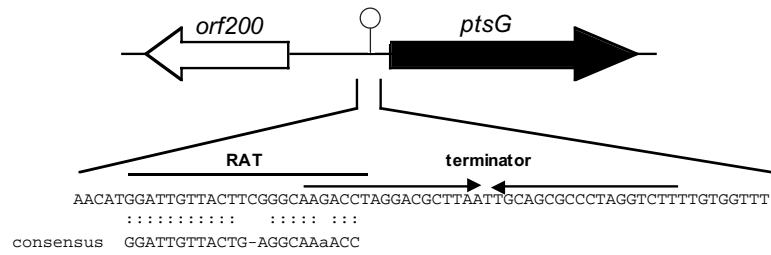
Figure 3. Gene locus III. A. Genetic map. The locus consisting of *ptsG* and *glcT* together with the flanking genes encoding a putative anaerobic carrier (*orf456*) and a gene encoding a protein of unknown function (*orf200*). The convention of presentation is as described in the legend to Figure 1. B. Putative terminator and antiterminator. Part of the sequence upstream of *ptsG* (-117 to -50) is depicted below the genetic map showing the intergenic region of *ptsG* and *orf200*. An inverted repeat forming a putative terminator is indicated by a stem loop sign and marked by two arrows. The terminator sequence is partially overlapping with a 23-bp RAT (ribonucleic antiterminator), indicated by a bar. A consensus sequence of putative RAT sequences found upstream of conditional p-independent terminators is given below the *C. diphtheriae* sequence. Identical nucleotides are shown by capital letters while highly conserved nucleotides are indicated by small letters. C. Domain structure of the putative antiterminator protein GlcT. A putative N-terminal RNA binding domain is shaded black while PTS regulation domains (PRD) are shaded gray. The regions around the four conserved histidine residues (\*) are shown together with the consensus sequence derived from an alignment of 12 homologues found in the current databank. Residues conserved in more than 90% of all proteins are shown in capital letters and similar residues in more than 50% of all proteins are shown in small letters. D. Phylogenetic tree of the antiterminator protein family. An unrooted phylogenetic tree was calculated as described in the legend to Figure 2B. Abbreviations are as follows: Antiterminator proteins from Cdi, *C. diphtheriae*; Lca, *Lactobacillus casei* (P24401); Bsu, *B. subtilis* (SacY, JU0294; SacT, S39702; LicT, S47216; GlcT, D69632); Bst, *Bacillus stearothermophilus* (AAB38977); Lla, *Lactococcus lactis* (AAA57135); Clo, *Clostridium longisporum* (AAC05712); Eco, *E. coli* (AAC76746); Ech, *Erwinia chrysanthemi* (A42603); and Sca, *Streptococcus carnosus* (CAA74358).

**A**

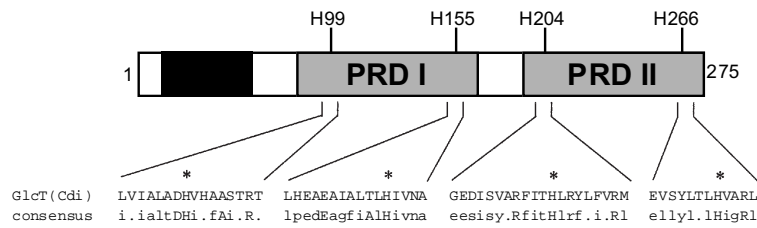
**gene locus III**



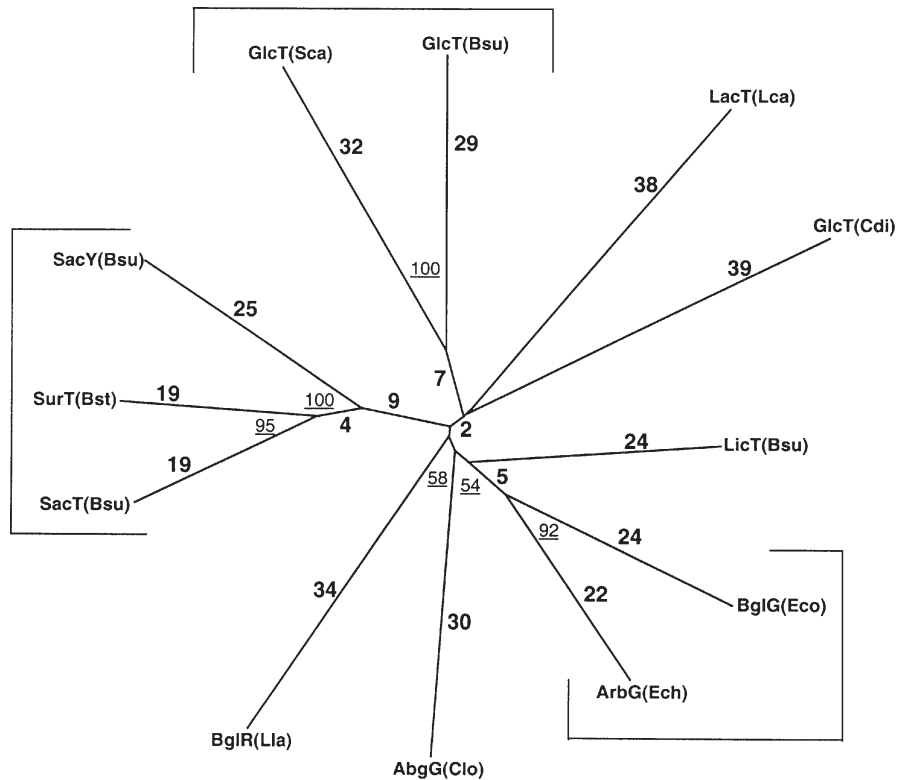
**B**



**C**



**D**



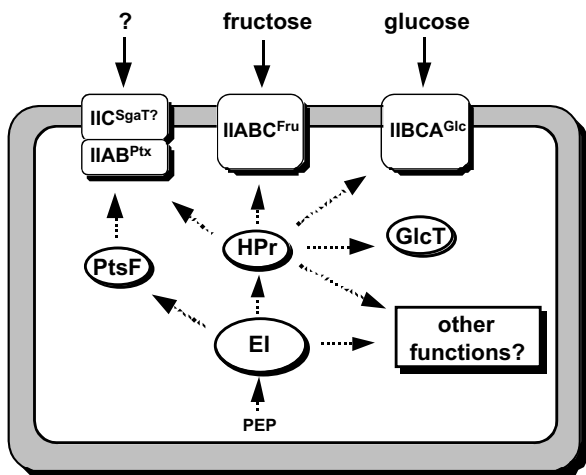


Figure 4. Proposed model of the *C. diphtheriae* PTS. The model illustrates recognized PTS proteins and phosphotransferase reactions (arrows). Transport of fructose and glucose is shown to occur via PEP, EI, HPr and enzyme IIBC<sup>Fru</sup> or enzyme IIBC<sup>Glc</sup>, respectively. An unknown substrate could be transported via IIB<sup>Ptx</sup> and IIC<sup>SgaT</sup>. GlcT could be phosphorylated by HPr, and PtsF could be phosphorylated by EI.

### Gene Locus III: a Glucose-Specific Enzyme II

A screen with the glucose-specific enzyme II<sup>Glc</sup> sequence of *C. glutamicum* led to a third locus including a gene of the glucose/sucrose transporter family followed by a gene encoding a putative antiterminator protein (Figure 3A). The flanking sequences encode two putative ORFs showing high similarity to an anaerobic carrier protein and a protein of unknown function. The percentages of GC content of coding and intergenic regions resemble those found in gene locus I (Figure 1A). The deduced primary sequence of the putative glucose transporter shares highest identity with the enzyme II<sup>Glc</sup> of *Corynebacterium ammoniagenes* and *C. glutamicum* (62 and 42%, respectively), all having the domain order IIBCA. We could demonstrate that the enzyme II of *C. glutamicum*, which was originally annotated as a mannose-specific enzyme II, was specific for glucose, having mannose and fructose as side substrates (Lee *et al.*, 1994; our unpublished results). For this reason the designation *ptsG* would be appropriate.

Surprisingly, these II<sup>Glc</sup> sequences possess a closer evolutionary relationship to the subcluster of the  $\beta$ -glucoside-, sucrose-, and trehalose permeases (Lengeler *et al.*, 1994; Reizer *et al.*, 1996b), which is distant from the subcluster comprising PTS enzymes II specific for glucose and N-acetylglucosamine. This suggests that *ptsG* of corynebacteria has evolved differently from *ptsG* genes present in other bacterial groups.

Inspection of the region upstream of *ptsG* led to the identification of a putative terminator structure partially overlapping a potential ribonucleic antiterminator (RAT) sequence (Figure 3B). This 23-bp RAT sequence shows high similarity to other RAT sequences found upstream of sucrose and  $\beta$ -glucoside operons of Gram-negative and Gram-positive bacteria (Aymerich and Steinmetz, 1992; Tangney and Mitchell, 2000). By contrast, length of possible

base pairings of this RAT sequence is shorter than of classical RAT sequences, which include about 29 nucleotides. Whether the nascent RNA is able to form a secondary structure to prevent termination of *ptsG* expression remains to be demonstrated. It should be noted that the region upstream of *C. glutamicum ptsG*, of which about 500 bp are reported, is lacking such a terminator/antiterminator sequence. Downstream of *ptsG* in *C. diphtheriae* a putative terminator of transcription was found, which suggests a monocistronic operon organization.

A regulatory protein, which possibly interacts with the identified RAT, could be encoded by the gene downstream of *ptsG*. This gene is the first of its kind within the branch of high-GC Gram-positive bacteria. Due to its close linkage to *ptsG*, we designated the gene *glcT*. The deduced amino acid sequence shows identities of 22 to 28% to other antiterminator proteins with the highest degree of identity to LicT and SacT of *B. subtilis*. Usually, antiterminator proteins contain three structural elements: an N-terminal RNA binding domain and two PTS regulation domains (PRDI and PRDII), which possess four conserved histidines (Figure 3C). LicT was demonstrated to become phosphorylated at three histidyl residues in the presence of HPr-His-P (Lindner *et al.*, 1999). These three histidines, which are highly conserved in most members of the BglG/SacY family of transcriptional antiterminators, correspond to His155, His204, and His266 of the *C. diphtheriae* GlcT. An unrooted tree was computed to analyze the phylogenetic relationship (Figure 3D). It showed that GlcT is closest to LicT of *B. subtilis*, and that it does not cluster with any of the other family members. Since almost all branches are similar in length, radiating from a central point, one might suggest that most of these proteins arose by early duplication events, and that further duplication events were rare. However, it remains to be demonstrated, if and how glucose utilization is regulated in *C. diphtheriae*.

### Genes Involved in Carbon Regulation?

The molecular mechanisms of C-regulation are still a mystery in high-GC Gram-positive bacteria (Angell *et al.*, 1994; Pope *et al.*, 1996; Parche *et al.*, 1999; Mahr *et al.*, 2001). It has been speculated that PTS components could fulfill C-regulatory function in actinomycetes as they do in other bacteria. This possibility prompted us to search for PTS-targets that are known from Gram-negative and low-GC Gram-positive bacteria.

Two such possible targets were found, one a putative glycerol kinase (*glpK*), and two an operon that encodes a maltose-like ABC transporter. Glycerol kinase shares 51%, 47%, and 46% identical residues with the respective kinases from *S. coelicolor*, *B. subtilis*, and *E. coli* throughout its entire sequence. Sequence alignment analyses revealed that the histidyl residue that is phosphorylated in low-GC Gram-positive bacteria by HPr is not present in the *C. diphtheriae* GlpK and also not in the homologous GylA gene of *S. coelicolor* (Wehtje *et al.*, 1995; Charrier *et al.*, 1997). However, since glycerol kinase is subject to PTS-mediated C-regulation in many bacteria, it remains to be seen whether a regulation of the glycerol utilization system requires the PTS (Hindle and Smith, 1994; Saier, 1989).

A *malEFG*, *orf410*, *malK* gene cluster was identified. MalK serves as the ATPase subunit of the maltose permease and is the target of inducer exclusion exerted through binding of IIA<sup>Glc</sup> in *E. coli* (Saier, 1989; Postma *et al.*, 1993). We have shown that a IIA<sup>Glc</sup>-like protein of *S. coelicolor* can inhibit maltose transport in *E. coli* (our unpublished results). It is possible that MalK is subject to PTS-mediated regulation in *C. diphtheriae*.

A search for homologues to known global bacterial regulators of carbon catabolite repression, CcpA (gene *ccpA*) and CAP (gene *crp*), revealed several *ccpA* homologues and one *crp* homologue. The CcpA regulators form a subfamily within the huge LacI/GalR family of bacterial regulators. CcpA proteins are characterized by 67 CcpA-specific residues that are distributed throughout the whole protein (Mahr *et al.*, 2000). The best CcpA-like gene product found contained only twelve CcpA-specific residues, which would argue against the presence of a CcpA protein. This conclusion may be supported by the fact that, despite extensive efforts, no CcpA-function could be detected in actinomycetes (van Wezel *et al.*, 1997). It further appears that actinomycetes do not possess an HPr kinase/phosphatase, an enzyme that is central to carbon catabolite control in low-GC Gram-positive bacteria (Titgemeyer *et al.*, 1995; Reizer *et al.*, 1998; Kravanja *et al.*, 1999; Parche *et al.*, 2001). This kinase phosphorylates HPr at a regulatory seryl residue and in that way generates the co-repressor HPr-Ser-P that is required for CcpA activity. A respective HPr kinase/phosphatase homologue could not be detected in the almost completed *C. diphtheriae* genome.

Analysis of a putative *crp* gene gave the following indicative data. The deduced protein shares 73%, 50%, and 30% protein sequence identity to homologues of *M. tuberculosis*, *S. coelicolor*, and to the well-studied CAP protein of *E. coli*. A putative adenylate cyclase, which in *E. coli* produces the CAP cofactor cAMP, was not found. Nevertheless, the *crp* homologue, which is not found in the low-GC Gram-positive bacteria, would be an interesting target for further investigation.

Finally, glucose kinase (*glkA*) was subject to our analysis, because it is considered to be a central component of C-regulation in streptomycetes (Angell *et al.*, 1994; Kwakman and Postma, 1994). A *glkA*-like gene was identified encoding a product of 317 amino acids. GlkA showed 62% and 40% identical amino acids to glucose kinase of *C. glutamicum* and *S. coelicolor*, respectively.

## Conclusions

The results summarized in this communication serve to characterize the complement of PTS enzymes encoded on the *C. diphtheriae* chromosome and therefore aid to define experimental investigations. As depicted in Figure 4, this organism possesses genes for the two energy-coupling proteins, EI and HPr, two complete permeases most probably specific for glucose and fructose, and two distinct genes encoding a IIA<sup>Ptx</sup>-like and a novel HPr-type protein of unknown function. The *C. diphtheriae* PTS exhibits several unique features that are worthy of note. First, it comprises a region, in which the fructose-specific genes are adjacent to the genes encoding the general

phosphotransferases. Second and so far unique among actinomycetes, it contains a regulator of the antiterminator class, which suggests that expression of *ptsG* is regulated by the PTS. Third, *C. diphtheriae* possesses a bipartite HPr hybrid protein, PtsF, which has its counterpart in the *E. coli* ADI protein that in addition contains an EI-like domain. Data derived from genes that may provide C-regulatory functions revealed that *C. diphtheriae* appears to lack components that are main actors in general carbon catabolite repression in low-GC Gram-positive bacteria. It remains to be determined whether similarities with mechanisms of C-regulation operative in Gram-negative bacteria exist, or whether *C. diphtheriae* and possibly other actinomycetes possess mechanisms that are dissimilar.

## Acknowledgements

We dedicate this work to Jonathan Reizer, who has pioneered PTS genome analysis. This work was carried out in the laboratories of Wolfgang Hillen. His support is greatly appreciated. We thank Milton Saier, Kerstin Mahr and Harald Nothhaft for critical reading of the manuscript. This work was supported through SFB473 of the Deutsche Forschungsgemeinschaft.

## References

- Angell, S., Lewis, C. G., Buttner, M. J., and Bibb, M. J. 1994. Glucose repression in *Streptomyces coelicolor* A3(2): a likely regulatory role for glucose kinase. *Mol. Gen. Genet.* 244: 135-143.
- 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.
- Butler, M. J., Deutscher, J., Postma, P. W., Wilson, T. J., Galinier, A., and Bibb, M. J. 1999. Analysis of a *ptsH* homologue from *Streptomyces coelicolor* A3(2). *FEMS Microbiol. Lett.* 177: 279-288.
- Charrier, V., Buckley, E., Parsonage, D., Galinier, A., Darbon, E., Jaquinod, M., Forest, E., Deutscher, J., and Claiborne, A. 1997. Cloning and sequencing of two enterococcal *glpK* genes and regulation of the encoded glycerol kinases by phosphoenolpyruvate-dependent, phosphotransferase system-catalyzed phosphorylation of a single histidyl residue. *J. Biol. Chem.* 272: 14166-14174.
- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltham, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Barrell, B. G., and *et al.*, 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393: 537-44.
- Collins, M. D., and Cummins, C. S. 1986. Genus *Corynebacterium*. *Bergey's Manual of Systematic Biology*, Vol. 2. Williams and Wilkins, Baltimore: 1266-1283.
- Dominguez, H., and Lindley, N. D. 1996. Complete sucrose metabolism requires fructose phosphotransferase activity in *Corynebacterium glutamicum* to ensure phosphorylation of liberated fructose. *Appl. Environ. Microbiol.* 62: 3878-3880.
- Fraser, C. M., Norris, S. J., Weinstock, G. M., White, O., Sutton, G. G., Dodson, R., Gwinn, M., Hickey, E. K., Clayton, R., Ketchum, K. A., Sodergren, E., Hardham, J. M., McLeod, M. P., Salzberg, S., Peterson, J., Khalak, H., Richardson, D., Howell, J. K., Chidambaram, M., Utterback, T., McDonald, L., Artiach, P., Bowman, C., Cotton, M. D., Venter, J. C., and *et al.*, 1998. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* 281: 375-388.
- Higgins, D. G., Thompson, J. D., and Gibson, T. J. 1996. Using CLUSTAL for multiple sequence alignments. *Methods. Enzymol.* 266: 383-402.
- Hindle, Z., and Smith, C. P. 1994. Substrate induction and catabolite repression of the *Streptomyces coelicolor* glycerol operon are mediated through the GylR protein. *Mol. Microbiol.* 12: 737-745.
- Kravanja, M., Engelmann, R., Dossonnet, V., Blüggel, M., Meyer, H. E., Frank, R., Galinier, A., Deutscher, J., Schnell, N., and Hengstenberg, W. 1999. The *hprK* gene of *Enterococcus faecalis* encodes a novel bifunctional enzyme: the HPr kinase/phosphatase. *Mol. Microbiol.* 31: 59-66.
- Kwakman, J. H., and Postma, P. W. 1994. Glucose kinase has a regulatory role in carbon catabolite repression in *Streptomyces coelicolor*. *J. Bacteriol.* 176: 2694-2498.
- Lee, J. K., Sung, M. H., Yoon, K. H., Pan, J.-G., Yu, J. H., and Oh, T. K.

1993. Cloning and expression of the gene encoding mannose enzyme II of the *Corynebacterium glutamicum* phosphoenolpyruvate-dependent phosphotransferase system system in *Escherichia coli*. *J. Microbiol. Biotechnol.* 3: 1-5.
- Lee, J. K., Sung, M. H., Yoon, K. H., Yu, J. H., and Oh, T. K. 1994. Nucleotide sequence of the gene encoding the *Corynebacterium glutamicum* mannose enzyme II and analyses of the deduced protein sequence. *FEMS Microbiol. Lett.* 119: 137-145.
- Lengeler, J. W., Jahreis, K., and Wehmeier, U. F. 1994. Enzymes II of the phosphoenolpyruvate-dependent phosphotransferase systems: their structure and function in carbohydrate transport. *Biochim. Biophys. Acta.* 1188: 1-28.
- Lindner, C., Galinier, A., Hecker, M., and Deutscher, J. 1999. Regulation of the activity of the *Bacillus subtilis* antiterminator LicT by multiple PEP-dependent, enzyme I- and HPr-catalysed phosphorylation. *Mol. Microbiol.* 31: 995-1006.
- Mahr, K., Hillen, W., and Titgemeyer, F. 2000. Carbon catabolite repression in *Lactobacillus pentosus*: analysis of the *ccpA* region. *Appl. Environ. Microbiol.* 66: 277-283.
- Mahr, K., van Wezel, G. P., Svensson, C., Kregel, U., Bibb, M. J., and Titgemeyer, F. 2001. Glucose kinase of *Streptomyces coelicolor* A3(2): large-scale purification and biochemical analysis. *Antonie van Leeuwenhoek* [in press].
- Parche, S., Nothaft, H., Kamionka, A., and Titgemeyer, F. 2001. Sugar uptake and utilization in *Streptomyces coelicolor*: a PTS view to the genome. *Antonie van Leeuwenhoek* in press.
- Parche, S., Schmid, R., and Titgemeyer, F. 1999. The phosphotransferase system (PTS) of *Streptomyces coelicolor*: identification and biochemical analysis of a histidine phosphocarrier protein HPr encoded by the gene *ptsH*. *Eur. J. Biochem.* 265: 308-317.
- Paulsen, I. T., Reizer, J., Jin, R. Z., Lin, E. C., and Saier, M. H., Jr. 2000. Functional genomic studies of dihydroxyacetone utilization in *Escherichia coli*. *Microbiology* 146: 2343-2344.
- Pope, M. K., Green, B. D., and Westpheling, J. 1996. The *bld* mutants of *Streptomyces coelicolor* are defective in the regulation of carbon utilization, morphogenesis and cell-cell signalling. *Mol. Microbiol.* 19: 747-756.
- Postma, P. W., Lengeler, J. W., and Jacobson, G. R. 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol. Rev.* 57: 543-594.
- Reizer, J., Bachem, S., Reizer, A., Arnaud, M., Saier, M. H., Jr., and Stülke, J. 1999a. Novel phosphotransferase system genes revealed by genome analysis - the complete complement of PTS proteins encoded within the genome of *Bacillus subtilis*. *Microbiology* 145: 3419-3429.
- Reizer, J., Schneider, B., Reizer, A., and Saier, M. H., Jr. 1999b. A hybrid response regulator possessing a PEP-dependent phosphorylation domain. *Microbiology* 145: 987-989.
- Reizer, J., Hoischen, C., Titgemeyer, F., Rivolta, C., Rabus, R., Stülke, J., Karamata, D., Saier, M. H., Jr., and Hillen, W. 1998. A novel protein kinase that controls carbon catabolite repression in bacteria. *Mol. Microbiol.* 27: 1157-1169.
- Reizer, J., and Saier, M. H., Jr. 1997. Modular multidomain phosphoryl transfer proteins of bacteria. *Curr. Opin. Struct. Biol.* 7: 407-415.
- Reizer, J., Charbit, A., Reizer, A., and Saier, M. H., Jr. 1996a. Novel phosphotransferase system genes revealed by bacterial genome analysis: operons encoding homologues of sugar-specific permease domains of the phosphotransferase system and pentose catabolite enzymes. *Genome Sci. Technol.* 1: 53-75.
- Reizer, J., Paulsen, I. T., Reizer, A., Titgemeyer, F., and Saier, M. H., Jr. 1996b. Novel phosphotransferase system genes revealed by bacterial genome analysis: the complete complement of *pts* genes in *Mycoplasma genitalium*. *Microb. Comp. Genomics.* 1: 151-164.
- Reizer, J., Reizer, A., and Saier, M.H., Jr. 1996c. Novel PTS proteins revealed by bacterial genome sequencing: a unique fructose-specific phosphoryl transfer protein with two HPr-like domains in *Haemophilus influenzae*. *Res. Microbiol.* 147: 209-215.
- Reizer, J., Saier, M.H., Jr., Deutscher, J., Grenier, F., Thompson, J., and Hengstenberg, W. 1988. The phosphoenolpyruvate:sugar phosphotransferase system in gram-positive bacteria: properties, mechanism, and regulation. *Crit. Rev. Microbiol.* 15: 297-338.
- Saier, M.H., Jr. 1993. Regulatory interactions involving the proteins of the phosphotransferase system in enteric bacteria. *J. Cell. Biochem.* 51: 62-68.
- Saier, M. H., Jr. 2000. Families of transmembrane sugar transport proteins. *Mol. Microbiol.* 35: 699-710.
- Saier, M.H., Jr., and Reizer, J. 1992. Proposed uniform nomenclature for the proteins and protein domains of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. *J. Bacteriol.* 174: 1433-1438.
- Saier, M.H., Jr. 1989. Protein phosphorylation and allosteric control of inducer exclusion and catabolite repression by the bacterial phosphoenolpyruvate:sugar phosphotransferase system. *Microbiol. Rev.* 53: 109-120.
- Saier, M.H., Jr., Grenier, F.C., Lee, C.A., and Waygood, E. B. 1985. Evidence for the evolutionary relatedness of the proteins of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. *J. Cell. Biochem.* 27: 43-56.
- Shiio, I., Sugimoto, S., and Kawamura, K. 1990. Effects of carbon source sugars on yield of amino acid production and sucrose metabolism in *Brevibacterium flavum*. *Agric. Biol. Chem.* 54: 1513-1519.
- Strohl, W. R. 1992. Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. *Nucleic. Acids. Res.* 20: 961-974.
- 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.
- Tangney, M., and Mitchell, W. J. 2000. Analysis of a catabolic operon for sucrose transport and metabolism in *Clostridium acetobutylicum* ATCC 824. *J. Mol. Microbiol. Biotechnol.* 2: 71-80.
- Titgemeyer, F., Walkenhorst, J., Reizer, J., Stuiver, M. H., Cui, X., and Saier, M. H., Jr. 1995. Identification and characterization of phosphoenolpyruvate:fructose phosphotransferase systems in three *Streptomyces* species. *Microbiology* 141: 51-58.
- Titgemeyer, F., Walkenhorst, J., Cui, X., Reizer, J., and Saier, M. H., Jr. 1994. Proteins of the phosphoenolpyruvate:sugar phosphotransferase system in *Streptomyces*: possible involvement in the regulation of antibiotic production. *Res. Microbiol.* 145: 89-92.
- van Wezel, G. P., White, J., Young, P., Postma, P. W., and Bibb, M. J. 1997. Substrate induction and glucose repression of maltose utilization by *Streptomyces coelicolor* A3(2) is controlled by *malR*, a member of the *lacI-galR* family of regulatory genes. *Mol. Microbiol.* 23: 537-549.
- Wehtje, C., Beijer, L., Nilsson, R. P., and Rutberg, B. 1995. Mutations in the glycerol kinase gene restore the ability of a *ptsGHI* mutant of *Bacillus subtilis* to grow on glycerol. *Microbiology* 141: 1193-8.
- Wu, L. F., Tomich, J. M., and Saier, M. H., Jr. 1990. Structure and evolution of a multidomain multiphosphoryl transfer protein. Nucleotide sequence of the *fruB(HI)* gene in *Rhodobacter capsulatus* and comparisons with homologous genes from other organisms. *J. Mol. Biol.* 213: 687-703.