Oral Microbiology Immunology 2006: 21: 21–27
Printed in Singapore. All rights reserved

ORAL MICROBIOLOGY
AND IMMUNOLOGY

Genomic variation in *Streptococcus mutans*: deletions affecting the multiple pathways of β-glucoside metabolism

L. A. Old, S. Lowes, R. R. B. Russell
Oral Biology, School of Dental Sciences, University of Newcastle, Newcastle upon Tyne, UK

Old LA, Lowes S, Russell RRB. Genomic variation in Streptococcus mutans*: deletions affecting the multiple pathways of β-glucoside metabolism.*
Oral Microbiol Immunol 2006: 21: 21–27. © 2006 The Authors. Journal compilation © 2006 Blackwell Munksgaard.

The genome of *Streptococcus mutans* UA159 contains two phospho-β-glucosidase genes, *bglA* and *celA*, which occur in operon-like arrangements along with genes for components of phosphotransferase transport systems and a third phospho-β-glucosidase encoded by the *arb* gene, which does not have its own associated transport system but relies on uptake by the *bgl* or *cel* systems. Targeted inactivation of each of the phospho-β-glucosidase genes revealed that *bglA* is involved in aesculin hydrolysis, *celA* is essential for utilisation of cellobiose, amygdalin, gentobiose and salicin, and *arb* is required for utilisation of arbutin. Inactivation of genes for the phosphotransferase systems revealed an overlap of specificity for transport of β-glucosides and also indicated that further, unidentified transport systems exist. The *cel* and *arb* genes are subject to catabolite repression by glucose, but the *regM* gene is not essential for catabolite repression. Screening a collection of isolates of *S. mutans* revealed strains with deletions affecting the *msm*, *bgl* and/or *cel* operons. The phenotypes of these strains could largely be explained on the basis of the results obtained from the knockout mutants of *S. mutans* UA159 but also indicated the existence of other pathways apparently absent from UA159. The extensive genetic and phenotypic variation found in β-glucoside metabolism indicates that there may be extensive heterogeneity in the species.

Key words: genome; β-glucoside; *Streptococcus mutans*

Prof. R. R. B. Russell, Oral Biology, School of Dental Sciences, University of Newcastle, Newcastle upon Tyne, NE2 4BW, UK
Tel.: +44 191 2228240;
fax: +44 191 2226137;
e-mail: r.r.russell@ncl.ac.uk
Accepted for publication July 6, 2005

Streptococcus mutans is capable of utilising a broad range of carbohydrates and it is thought that this versatility is one of the characteristics contributing to its survival in plaque and involvement in the caries process. The carbohydrates may originate from dietary sources or from host macromolecules. Among the substrates utilised are β-glucosides such as aesculin, arbutin, cellobiose and salicin. All these substrates contain a glucose moiety that can enter glycolysis, coupled to another glucose or some other nonmetabolisable component. Different species of oral streptococci vary in their capacity to attack these substrates and this variation has been of value in the development of identification schemes (33). The description of *S. mutans* includes its ability to hydrolyse aesculin and ferment amygdalin, arbutin and salicin but it has been reported that many isolates of *S. mutans* do not possess one or more of these properties (2, 33). While such variation can confuse identification schemes, variation within the species may also reflect variation in the competitiveness of different strains, and possibly in their virulence. Furthermore, if *S. mutans* as a species demonstrates variability in sugar utilisation patterns, it seems probable that other important properties, for which the phenotype is not so readily observable, may also be subject to variability. Such variation clearly is of considerable relevance to our understanding of *S. mutans* virulence, and in the selection of potential targets for preventive measures. However, little is known as yet about the extent of genomic variation within *S. mutans* or the mechanisms for generating diversity. We have therefore used analysis of genes involved in carbohydrate metabolism as a model system to explore genetic variation, and this work was undertaken to unravel the complexities of β-glucoside metabolism in *S. mutans*.

In a number of species of bacteria, it has been demonstrated that multiple systems exist for the transport and utilisation of β-glucosides; for example, *Bacillus subtilis* and various species of enterobacteria have four (13, 26). The systems in *Escherichia coli* have been the most thoroughly studied and it is known that in this species there are several sets of cryptic (silent) genes that can be activated by mutation or insertion of a mobile genetic element (10). To understand such cases, it is necessary to apply a genetic approach in parallel with studies of the phenotypic capabilities of the species. β-glucosides are generally transported by β-glucoside-specific phosphotransferase systems (PTS) whereby the substrate is vectorially phosphorylated as it is taken up, and subsequently cleaved by a phospho-β-glucosidase (31). In *S. mutans* NG8, Cote et al. (5) have described the *bgl* regulon that includes *bglA*, encoding a phospho-β-glucosidase, *bglP,* that encodes an Enzyme II PTS component and a regulatory gene, *licT*. They also provided evidence from gene-inactivation experiments that one or more additional systems for β-glucoside utilisation must be present in *S. mutans*. A similar conclusion was obtained from a study of *S. mutans* isolates naturally defective in *bgl* genes (18). Annotation of the *S. mutans* UA159 genome sequence has subsequently identified genes potentially involved in β-glucoside metabolism, on the basis of homology to genes from other bacteria (1).

We have previously reported that the inability of certain strains of *S. mutans* to utilise melibiose and galactose is due to an 18 kb chromosomal deletion involving the *msm* and *gal* operons and that a number of such strains also had a 4 kb deletion affecting the *bglP* and *bglC* genes (18). In this work, we exploited the availability of the genome sequence of *S. mutans* strain UA159 to investigate the function of genes associated with β-glucosidase metabolism and examined a collection of clinical isolates for the presence or absence of these genes.

Material and methods

Bacterial strains

S. mutans UA159 was the strain selected for the genome sequencing project (1); it is 'wild-type' and capable of utilising all the sugars normally expected for the species. *S. mutans* NG8 was used in a previous study of the *bgl* genes (5, 6). Strains At10, B2, MT4653, MT4863, 13.1, 12, 29, AS1, L18, C4 and G8 are independent melibiose-negative isolates used in previous studies (4, 18, 30). A further 20 *S. mutans* isolates from our local collection were screened as described below, including strains 34, LML1 and LML5. Conditions of growth, fermentation and physiological tests and assay for β-glucosidase using the fluorescent substrate 4-methylumbelliferyl-β-D-glucoside were as described previously (4). Hydrolysis of aesculin on solid media with and without glucose was assayed as described by Cote et al. (5).

Bioinformatic analysis

The complete 2,032,327 bp *S. mutans* UA159 genome sequence (GenBank Accession NC004350) was searched with the BLAST suite of programmes at http://www.ncbi.nlm.nih.gov/BLAST/ to identify open reading frames (ORFs) with homology to known β-glucosidases and phospho-β-glucosidases. Genes were located on the *S. mutans* UA159 genome map visualised with the ARTEMIS genome viewer programme (23) from http://www.sanger.ac.uk/. ORFs are numbered according to the GenBank genome annotation (1).

Extraction of DNA

Bacteria were grown overnight in Todd-Hewitt broth containing 0.5% yeast extract and 20 mM threonine. After harvesting (3000 *g* for 5 min) the cell pellet was washed in 2 ml of sterile distilled water and resuspended in 2 ml of 50 mM Tris HCl pH 8.0, 10 mM EDTA (Tris-EDTA) supplemented with 0.45 mg/ml lysozyme and 12 μg/ml mutanolysin, and incubated at 50°C for 2 h. Complete lysis was achieved by adding 2 ml of hot (65°C) Tris-EDTA buffer containing 2% SDS. Proteinase K was added to a final concentration of 1.4 mg/ml and the tubes were incubated at 50°C for a further 2 h. Cell debris was removed by centrifugation and the supernatant was extracted six times with equal volumes of phenol chloroform. Traces of phenol were removed by two extractions with an equal volume of water-saturated ether. Residual ether was boiled off by incubation at 37°C for 15 min. The DNA was precipitated by the addition of 2 volumes of ethanol and holding at −20°C for 30 min. After centrifugation at 12,000 *g* for 5 min, the DNA pellet was washed once with prechilled 70% ethanol at −20°C, dried, and resuspended in 0.5 ml of sterile distilled water. Dilutions of this stock were used as polymerase chain reaction templates.

Polymerase chain reaction (PCR)

The confirmation of strains as *S. mutans* was done using primers for 16S ribosomal RNA genes described by Rupf et al. (22). Primers for amplification of regions of chromosomal DNA were designed with the aid of the Primer3 programme of Rozen & Skaletsky (21) available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html using published sequences or data from the genome of *S. mutans* UA159. The ARTEMIS programme was used to locate open reading frames and select regions for primer design. PCR experiments were carried out with the high fidelity, premixed Extensor Long PCR Master Mix (ABgene, Epsom, Surrey, UK) without oil overlays, on a GeneAmp9700 thermal cycle essentially as described previously, with appropriate variation in extension time for different primer pairs (18). Forward and reverse primers used to screen for the *cel* operon were TCTCTGCCTTTTGAGGCATT and AACTTCTTGGCGCGGTTTG; those for the *arb* gene were AACTGACGCCCTAATTTCTC and TGGAACTTTTGAAGGGCAAC, respectively.

DNA sequencing

PCR products separated by agarose gel electrophoresis were purified from the gel using the Qiaex II kit (Qiagen, Crawley, West Sussex, UK). DNA sequencing was carried out at the Molecular Biology Unit, University of Newcastle, using Thermosequenase and dye terminator chemistry on an ABI377 sequencer. Primers used for sequencing were the same as primers used to generate the PCR products.

Gene inactivation

The three phospho-β-glucosidase genes were inactivated by gene replacement mutagenesis in which regions flanking the gene to be inactivated were amplified from *S. mutans* UA159 chromosomal DNA by PCR, using primers designed using the genome sequence and ligated to an erythromycin-resistance determinant (16). The construct was then used to transform *S. mutans* UA159. Transport genes were knocked out by insertion-duplication mutagenesis in which an internal fragment of *ca*. 200 bp from the 5′ end of the gene to be inactivated was amplified by PCR. The primers incorporated an *Xma*I site at either end. Amplicons were purified and cloned into the *Sma*I site of the suicide vector

pVA8912, which carries erythromycin resistance (27) or pSF151 (29), which carries kanamycin resistance. Recombinant plasmids were used to transform *S. mutans* UA159 by the biofilm transformation method of Li et al. (17) using competence stimulating peptide kindly provided by D. Cvitkovitch, Toronto. Colonies in which the resistance marker had integrated into the chromosome were selected on erythromycin or kanamycin as required. Chromosomal insertion of the resistance markers at the correct location was confirmed using PCR with primers within the insert and flanking regions. A *regM* mutant of *S. mutans* UA159 was constructed by transformation with chromosomal DNA from the mutant previously constructed in *S. mutans* LT11 (28).

Results

Search for genes involved in β-glucoside metabolism

The genome of *S. mutans* UA159 was searched by the BLAST suite of programs for ORFs potentially encoding β-glucosidases or phospho-β-glucosidases having homology to the *bglA* gene sequence described by Cote et al. (5) or to related genes from other gram-positive bacteria available in public databases. No β-glucosidases were detected but two ORFs paralogous to *bglA* and predicted to encode phospho-β-glucosidase were discovered (Table 1). The three ORFs are located in different parts of the genome and all are of similar length (478–479 residues). SMU.1102 shares 54% identical amino acids with *bglA* and SMU.1601 has 52% identity. In addition, ORF SMU.1490 has 32% overall identity with *bglA*. This ORF can be identified as a phospho-β-galactosidase because of its possession of a tryptophan residue in position 429 instead of the alanine found in phospho-β-glucosidases (25), and corresponds to *lacG* in the lactose operon (20).

The occurrence of phospho-β-glucosidases indicates that β-glucosides are likely to be phosphorylated as they enter the cell, as occurs when uptake is via the phosphotransferase system. Cote et al. (5) described the presence of β-glucoside specific enzyme II component of the PTS system, encoded by *bglP*, in the same regulon as *bglA*. Enzyme II components were also found in an operon-like arrangement with SMU.1601 (Fig. 1). In contrast to *bglP*, the enzyme II in this region is constructed from three separate peptide chains encoded by different ORFs. Also within this operon-like arrangement is SMU.1597, predicted to encode a transmembrane protein that is of unknown function and has no homologues in the sequence databases. The arrangement of the ORFs SMU.1595–SMU.1601 is shown in Fig. 1. In their annotation of the *S. mutans* UA159 genome, Ajdic et al. (1) identified these genes as possibly being involved in cellobiose metabolism on the basis of their homology with genes in other bacterial species. Since experimental evidence (below) supports such a function, it is proposed that the genes be renamed *cel*, as indicated in Table 1 and Fig. 1.

The third phospho-β-glucosidase, SMU.1102, is not linked to any PTS components and its separation from adjacent genes suggests that it is independently transcribed. Ajdic et al. (1) found that SMU.1102 showed homology to the *ascB* gene of *E. coli*, which encodes an enzyme that acts on *a*esculin, *s*alicin and *c*ellobiose (10). However, results presented below show that in *S. mutans* this gene is clearly required for the degradation of arbutin, so it is proposed that the name *arb* be applied. *Lactobacillus delbrueckii* has also been shown to possess an arbutin-specific enzyme (32).

Table 1. Summary of open reading frames (ORFs) and names of genes associated with β-glucoside metabolism in *S. mutans* UA159

ORF	Current annotation (1)	Proposed gene name	Function
bgl region			
SMU.977	*licT*	*licT*	transcriptional antiterminator
SMU.980	*bglP*	*bglP*	PTS system, enzyme EII component
SMU.981, 982	*bglB*	*bglB*	putative lipase, paralogue of *gbpD*
SMU.983	*bglC*	*bglC*	putative transcriptional regulator
SMU.985	*bglA*	*bglA*	6-phospho-β-glucosidase
arb region			
SMU.1102	*ascB*	*arb*	6-phospho-β-glucosidase
cel region			
SMU.1601	*bgl*	*celA*	6-phospho-β-glucosidase
SMU.1600	*ptcB*	*celA*	PTS system IIB component
SMU.1599	*celR*	*celR*	Transcriptional regulator
SMU.1598	*ptcA*	*celC*	PTS system IIA component
SMU.1596	*ptcC*	*celD*	PTS system IIC component
SMU.1595	*cah*	*cah*	putative carbonic anhydrase precursor
SMU.1597	–	–	conserved hypothetical protein (transmembrane protein)

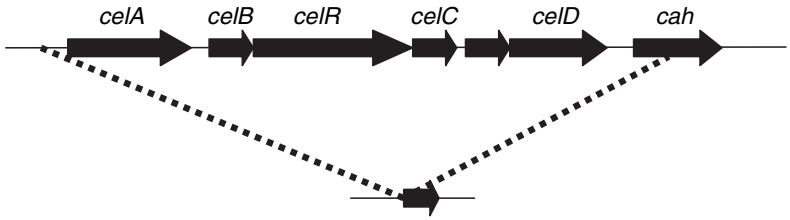

Fig. 1. Gene arrangement of the *cel* region of the *S. mutans* UA159 chromosome, derived from the published genome sequence (1) with proposed new gene names. The upper line shows the arrangement in *S. mutans* UA159, the lower line the arrangement in strains with a chromosomal deletion of the *cel* region. *celA*, phospho-β-glucosidase. *celB*, PTS system IIB component. *celR*, transcription regulator. *celC*, PTS system IIA component. *celD*, PTS system IIC component. *cah*, carbonic anhydrase. Between *celC* and *celD* lies SMU.1597, encoding a hypothetical transmembrane protein.

Knockout of phospho-β-glucosidase genes

To explore their contribution to the metabolism of the various β-glucoside substrates, the genes encoding each of the three phospho-β-glucosidases were inactivated. Cote et al. (5) have previously reported that inactivation of *bglA* in *S. mutans* NG8 did not affect fermentation of arbutin, cellobiose or salicin. However, they could demonstrate that aesculin is a substrate for BglA by incorporating glucose into the test medium. This resulted in catabolite repression of other phospho-β-glucosidases (5) and the *bglA* knockout removed the residual activity. In contrast, we found that glucose completely repressed aesculin hydrolysis in *S. mutans* UA159 and all its derivative mutants (Fig. 2a, compare ± glucose) so that this strain displays the same phenotype as NG8 mutants in which *bglA* has been inactivated. Inactivation of *arb* results in inability to ferment arbutin but other substrates are unaffected. In contrast, *celA* appears to encode an enzyme with broad specificity because its inactivation destroyed the ability to ferment cellobiose, amygdalin, gentobiose and salicin. From this observation,

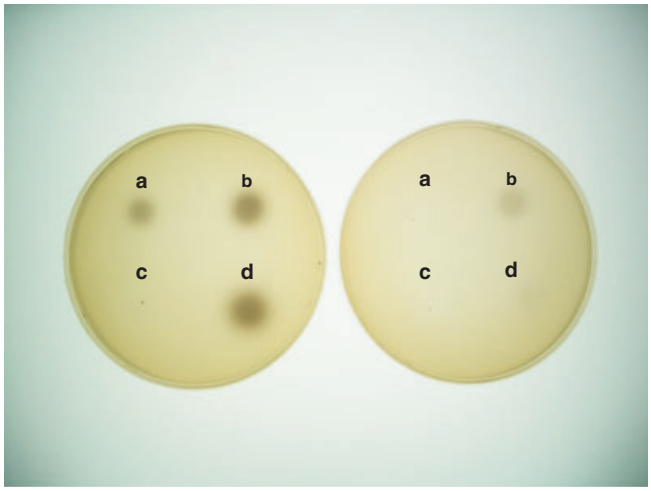

Fig. 2. Phenotype of *S. mutans* strains on medium containing aesculin (left) or aesculin + glucose (right). a) *S. mutans* UA159; b) *S. mutans* NG8; c) *S. mutans* UA159 *celA*$^{-}$ *arb*$^{-}$; d) *S. mutans* NG8 *regM*.

it can also be concluded that neither BglA nor Arb attacks these substrates. Construction of a *bglA*/*arb* double mutant by insertional inactivation with kanamycin and erythromycin-resistance determinants confirmed that cellobiose, amygdalin, gentobiose and salicin were fermented even when CelA was the only enzyme present. The *celA*/*arb* double mutant could not attack aesculin (Fig. 2c) or any of the other substrates tested, indicating that BglA is nonfunctional.

Knockout of PTS genes

A similar set of observations was made with a mutant in which the *bglP* uptake system was inactivated, confirming the observation of Cote et al. (5) that loss of *bglP* alone did not alter fermentation patterns but that *bglP* is required for uptake and subsequent hydrolysis of aesculin when glucose is present. When the *cel* PTS system is inactivated by disrupting the sugar-specific *celD* enzyme II component, the resultant mutant can no longer ferment cellobiose, amygdalin or gentobiose, suggesting that the *cel* genes encode the only uptake system for these substrates. However, it seems that both the *bgl* and *cel* PTS systems are capable of uptake of aesculin and salicin and the fluorescent substrate methylumbelliferyl-β-D-glucoside, since it is only when both genes are inactivated that there is an effect on the utilisation of these. Finally, even when both *bglP* and *celD* are inactivated in a double mutant, arbutin can still be taken up by some other transport system as fermentation activity is still present (Table 2).

Catabolite repression

The results of Cote et al. (5) showed that uptake and breakdown of aesculin is partially repressed by glucose in *S. mutans* NG8, and that the residual activity is abolished by inactivation of *bglA*. That observation is confirmed here but note that in strain UA159, activity against aesculin is completely repressed by glucose (Fig. 2, compare a and b ± glucose). Amongst the collection of 40 strains, 50% showed the same pattern of complete repression as UA159. In order to examine the involvement of *regM* gene in catabolite repression, the effect of inactivating *regM* on hydrolysis of aesculin in the presence and absence of glucose was examined. The result clearly shows that repression by glucose in strain NG8 is even more effective when *regM* is inactivated (Fig. 2b,d, compare ± glucose).

Genetic variation among isolates of *S. mutans*

We have previously reported that amongst a collection of isolates with deletions of the *msm* and *gal* operons, some also had deletions affecting the *bglP* and *bglC* genes (18). An additional 20 clinical isolates were screened by PCR using primers flanking the *msm*, *bgl* and *cel* regions. Strains with deletions give smaller PCR products than UA159 (18). The screening led to the identification of one (strain LML1) that resembled strains such as At10 described earlier with regard to fermentation pattern and deletion of *msm* and *bgl* regions and two others (strains LML5 and 34) that had sugar fermentation patterns the same as wild-type but with a 4 kb deletion affecting *bglP* and *bglC*, of the type described previously (18). These strains showed the same phenotype as the *bglP* knockout mutant of UA159.

As we found previously that deletions were a frequent cause of the loss of sugar fermentation ability, we explored whether the *cel* operon and *arb* gene might also be subject to deletion events. Primers were therefore designed to amplify these genes and c.1 kb of flanking sequence on either side. For the *cel* operon, these primers yielded a PCR product of 10 kb from UA159 and other strains with wild-type fermentation pattern. However, certain strains (Table 3) gave a product of only 3 kb, indicating that a deletion event had occurred. The precise site of the deletion was determined by sequencing from either end of the 3 kb amplicon and aligning the sequence with the UA159 genome sequence. This revealed that the deletion affected the entire *cel* operon as well as the adjacent *cah* gene, encoding carbonic anhydrase (Fig. 1). The precise endpoints corresponded to genome sequence nucleotide positions 152,8067–153,2890. Examination of the sequences at the two endpoints revealed no distinctive feature such as inverted repeats or repetitive sequences.

Similar experiments were carried out to determine whether *arb* was affected by deletions, but no deletions were found among the strains examined.

Correlations between genotype and phenotype

An attempt was made to explain the sugar fermentation patterns of the clinical isolates (Table 3) with reference to the insights gained from targeted gene knockout in UA159 (Table 2). As previously described, strains such as L18 with only a deletion of *msm* and *gal* genes are affected in their utilisation of these sugars but still utilise β-glucosides (18). Similarly, strains 34 and LML5, in which the only detected change was in *bgl* genes, were only altered with respect to their hydrolysis of aesculin (i.e were aesculin-negative in the presence of glucose, as observed for NG8). The group of strains MT4863, C4 and G8 that had deletions of the *cel* genes could no longer ferment amygdalin. This was expected because *celA* and *celD* were shown to be essential for amygdalin fermentation in UA159. Unexpectedly, they could still use cellobiose and gentobiose even though *celD* is needed for fermentation of these sugars in

Table 2. Biochemical reactions of *S. mutans* UA159 and mutants of UA159 in which specific genes had been inactivated

Mutant	Aesc	Arb	Cell	Amyg	Gent	Sal	F-β-Glu
UA159	+	+	+	+	+	+	+
bglA	+	+	+	+	+	+	+
arb	+	–	+	+	+	+	+
celA	+	+	–	–	–	–	+
bglA/arb	+	–	+	+	+	+	+
celA/arb	–	–	–	–	–	–	–
bglP	+	+	+	+	+	+	+
celD	+	+	–	–	–	+	+
bglP/celD	–	+	–	–	–	–	–

Mel, melibiose. Aesc, aesculin. Glu, glucose. Arb, arbutin. Cell, cellobiose. Amyg, amygdalin. Gent, gentobiose. Sal, salicin. F-β-Glu, 4-methylumbelliferyl-β-D-glucoside.

Table 3. Biochemical reactions of *S. mutans* UA159 and other isolates of *S. mutans* in which chromosomal deletions (Δ) have been detected

Strain	Genotype	Mel	Aesc	Aesc + Glu	Arb	Cell	Amyg	Gent	Sal	F-β-Glu
UA159		+	+	–	+	+	+	+	+	+
L18	Δ*msm*	–	+	+	+	+	+	+	+	+
LML5	Δ*bgl*	+	+	–	+	+	+	+	+	+
34	Δ*bgl*	+	+	–	+	+	+	+	+	+
MT4863	Δ*msm*, Δ*cel*	–	+	+	+	+	–	+	+	+
C4	Δ*msm*, Δ*cel*	–	+	+	+	+	–	+	+	+
G8	Δ*msm*, Δ*cel*	–	+	+	+	+	–	+	+	+
AT10	Δ*msm*, Δ*bgl*, Δ*cel*	–	–	–	+	–	–	–	+	+
B2	Δ*msm*, Δ*bgl*, Δ*cel*	–	–	–	+	–	–	–	+	+
MT4653	Δ*msm*, Δ*bgl*, Δ*cel*	–	–	–	+	–	–	–	+	+
13.1	Δ*msm*, Δ*bgl*, Δ*cel*	–	–	–	+	–	–	–	+	+
12	Δ*msm*, Δ*bgl*, Δ*cel*	–	–	–	+	–	–	–	+	+
29	Δ*msm*, Δ*bgl*, Δ*cel*	–	–	–	+	–	–	–	+	+
AS1	Δ*msm*, Δ*bgl*, Δ*cel*	–	–	–	+	–	–	–	+	+

Mel, melibiose. Aesc, aesculin. Glu, glucose. Arb, arbutin. Cell, cellobiose. Amyg, amygdalin. Gent, gentobiose. Sal, salicin. F-β-Glu, 4-methylumbelliferyl-β-D-glucoside.

UA159. The final group, in which deletions affect *msm*, *bgl* and *cel* regions, behave as predicted from the results with UA159, with the exception that they still ferment salicin.

Discussion

The use of a representative strain of a bacterial species for detailed laboratory investigation, particularly if its genome sequence is available, can give valuable insights into its properties. However, recent developments in our understanding of bacterial populations have led to the realisation that no single strain is likely ever to be entirely representative of its species. The extent of diversity varies in different species and in some instances even the current concept of species borderlines may prove to be inappropriate. A full understanding of the properties of a species such as *S. mutans* will therefore require study of a range of strains to determine the spectrum of metabolic capabilities encoded by the 'supragenome' (8). There is an extensive literature on the variation in phenotypic properties of *S. mutans* and studies of genetic polymorphisms indicate extensive variation at the nucleotide sequence level, though how this relates to phenotypic variation is unknown. From the limited information available on specific chromosomal regions, we know that, in comparison with UA159, some strains have lost some genetic information through chromosomal deletions, while others have genes present that are not represented in the UA159 genome (18, 24). Besides deletions of the size discussed in this report, there will of course also be small-scale deletions, insertions, inversions and point mutations not detected by the PCR method used but which could be detected by sequencing. One example pertinent to this paper is *bglB*, which exists as a single ORF in strain NG8 but is interrupted by a frameshift in UA159 (1, 5).

Results obtained from laboratory studies of knockout mutations in the *bgl* region and from study of β-glucoside-negative clinical isolates led to the conclusion that *S. mutans* possesses other systems for metabolism of β-glucosides beside the *bgl* regulon (5, 18). In the present study, the *arb* gene and *cel* operons were identified from the genome sequence, and results from targeted knockout experiments were consistent with the proposed function of their constituent genes in metabolism of various β-glucosides. As in other species such as *B. subtilis* and *E. coli*, there is redundancy of transport systems and overlapping specificity for β-glucosidases (Table 2). However, it is not known what the preferred natural substrates encountered in nature might be. β-glucosides are predominantly found in plants, but chitin, from the exoskeleton of insects and marine animals, is the second most abundant polysaccharide in nature after cellulose and is composed of β1,4-linked N-acetyl glucosamine. In *E. coli*, the enzyme originally thought to be required for cellobiose degradation is now considered to primarily be concerned with chitobiose, a disaccharide derived from chitin, and the *cel* genes have been redesignated as *chb* (11, 12). The genes identified in *S. mutans* may be needed for metabolism of substrates as yet untested. However, we propose the *cel* nomenclature because cellobiose is a readily available substrate and widely used in biochemical identification schemes. The *bgl* and *cel* PTS systems account for uptake of most of the substrates tested, but even when both are inactivated, arbutin is still broken down (by Arb) and fermented. It must therefore be transported by another, unidentified PTS system. A total of 14 PTSs were identified in the genome of UA159, though the specificity of most of them remains unknown, (1) and it is possible that one or more may be capable of recognising β-glucosides.

Previous study of isolates of *S. mutans* that fail to ferment melibiose led to the discovery that this is commonly due to an 18 kb chromosomal deletion in which the entire *msm* and *gal* operons are replaced by an incomplete insertion element, named ISSmu3 (18). Some of the strains studied also had deletions affecting *bglP* and *bglC* and here we show that deletions of the *cel* operon are also common. The origin of the deletion is not obvious, as there is no indication of a history of IS element insertion/excision or of slipped-strand repair as proposed for the *msm* and *bgl* deletions (18). The deletion affects the entire *cel* operon and also inactivates the adjacent *cah* gene that encodes carbonic anhydrase. The physiological consequences of loss of *cah* are unknown. However, it is interesting to note that carbonic anhydrase in *E. coli* is required to detoxify cyanate (15), and amygdalin is a diglucoside with a cyanide radical attached. It will thus be interesting to explore whether carbonic anhydrase in *S. mutans* serves a function to detoxify intracellular products generated from breakdown of amygdalin.

Information from the knockout of specific genes in UA159 provides a basis for interpretation of the different pattern of fermentation found amongst clinical isolates. Some strains such as L13 studied previously only have a deletion of *msm* and *gal* genes; as a consequence they fail to ferment melibiose or grow on galactose as sole carbon source (18). The two newly discovered strains LML5 and 34 have the *bglP, bglC* deletion but their *msm* and *gal* regions are intact so they appear wild-type in the regular sugar fermentation tests. Their mutant phenotype can only be observed on aesculin plates in the presence of glucose so they mimic the behaviour of the *bglA* and *bglP* knockout mutants. The Δ*cel* strains MTc, C4 and G8 are unable to ferment amygdalin, like the *celA* and *celD* knockout mutants of UA159. However, they still ferment cellobiose, gentobiose and salicin and so differ from the *celA* and *celD* knockout mutants. The pathways by which these three sugars are taken up and metabolised in MTc, C4 and G8 are presently unknown. One explanation may be that these strains possess transport systems and/or phospho-β-glucosidases that are not present in UA159, or it could be that the specificity of their *bgl* and *arb* gene products differ from those of UA159. Similar arguments may explain why the final group of isolates, with three identified chromosomal deletions, can still utilise salicin, though their inability to utilise arbutin is in agreement with predictions from the knockout experiments in UA159. Another explanation for the differences observed between the natural isolates with deletions and the laboratory-constructed mutants might be that some genes are cryptic in UA159 but expressed in other strains, in a way similarly observed for cryptic *bgl* and *cel* genes in *E. coli* K12. However, bioinformatic analysis of the UA159 genome revealed no likely candidate genes beyond those described here.

Catabolite repression is clearly an important influence on the expression of the various genes concerned with β-glucoside metabolism. It was previously concluded that the *bglA* gene in *S. mutans* NG8 is not subject to catabolite repression by glucose (6), whereas the other two phospho-β-glucosidases are repressed. In this study we could detect no hydrolysis of aesculin in the presence of glucose in *S. mutans* UA159. To distinguish between the two possibilities that BglA is subject to catabolite repression in UA159 but not in NG8, or that BglA is inactive in UA159, we constructed the *celA*/*arb* double mutant. In this mutant, *bglA* is the only phospho-β-glucosidase gene present, but there is no hydrolysis of aesculin or any other substrate. There thus appears to be a clear difference between expression of the *bglA* genes of strains UA159 and NG8. Comparison of the *bglA* sequences from the two strains reveals 12 nucleotide differences, 10 of which result in a different amino acid. In the absence of structural information, it is not possible to conclude how these might impact on the function of BglA in strain UA159 but the data presented here indicate that *bglA* is a pseudogene, its functional loss being covered by the redundancy of phospho-β-glucosidase activity provided by the other two enzymes. However, it is also possible that the lack of *bglA* expression in UA159 is due to defects in *licT* or *bglB* genes, which are both involved in the regulation of *bglA* (6, 7). The mechanism of catabolite repression in *S. mutans* is incompletely understood but in many other gram-positive bacteria the protein CcpA is required for catabolite repression. The homologue of *ccpA* in *S. mutans* has been named *regM* and construction of a knockout mutant of *regM* showed that it was not required for catabolite repression of α-galactosidase, mannitol-1-phosphate dehydrogenase or phospho-β-galactosidase (28). In contrast, there is increased catabolite repression of these enzymes in a *regM* mutant (28). Results presented here show that *regM* is not required for catabolite repression of the pathways associated with metabolism of β-glucosides either, and that knockout of *regM* increases the repression in strain NG8 (Fig. 2). *S. mutans, Streptococcus gordonii* (19) and *Streptococcus pneumoniae* (9) thus differ from the CcpA paradigm established for other bacteria.

Recently, Kilic et al. (14) have investigated the β-glucoside metabolism of *S. gordonii* and concluded that this organism possesses four metabolic pathways for these substrates, though the arrangement of genes seems to be quite different from that found in *S. mutans*. There are, however, a number of similarities, including the susceptibility to catabolite repression. These authors also proposed that the β-glucoside pathways may influence adhesion, biofilm formation and *in vivo* colonisation (14), possibly acting as environmental sensors in a manner similar to that observed in *Listeria monocytogenes* (3). We do not have any evidence for or against such a function for the pathways in *S. mutans* but it is an intriguing possibility that changes caused by the deletion of *bgl* or *cel* genes might have far-reaching consequences for the physiological responsiveness of a strain.

Our original collection of strains for study was made on the basis of their easily recognised melibiose-negative phenotype. In this study, a further 20 recent clinical isolates were screened by PCR, as well as by phenotypic tests. This resulted in the discovery of strains with genetic and phenotypic properties similar to strains already known, but also to the two Δ*bgl* strains, which are phenotypically the same as wild-type in routine biochemical assays. This illustrates the advantage of screening at the nucleotide sequence level to explore the extent of variation within *S. mutans*.

The work on β-glucoside metabolism has drawn attention to three regions of the *S. mutans* genome (*msm*, *bgl* and *cel*) where some strains differ from UA159, apparently because deletions have occurred. In comparison with UA159, some strains may also have additional genetic information inserted, such as the *cnm* gene for a collagen-binding adhesin (24) or an additional sequence replacing the *gbpA* gene (our unpublished observations). Preliminary results from intragenome comparisons by DNA–DNA hybridisation on microarrays representing ORFs of UA159 indicate that besides the *msm*, *gal*, *bgl* and *cel* genes missing in strain At10, other ORFs are missing (J. Waterhouse, personal communication). Whether there are also insertions of genes not found in UA159 is unknown. While UA159 thus offers an enormously valuable reference point for exploration of other strains, it is clear that much remains to be learnt about the diversity within the species. Many important questions are also raised concerning the mechanism for generating deletions and their ecological advantage or disadvantage. We have found deletions in both recent isolates and long-established laboratory strains, so there is no evidence for a selective pressure imposed by subculturing. To date, the only isolates found with a deletion of the *cel* genes are ones that also lack *msm* but it is not clear whether this observation is significant: Which occurred first? Does one deletion create a selective pressure for the other? Since chromosomal deletions are irreversible events, it is probable that they will be transmitted clonally and that these strains share a common ancestor in which the two deletions originally arose independently. It will be of considerable interest to study a wider collection of strains in order to address the question of whether deletions have occurred as independent events on a number of occasions, or in

some ancestral *S. mutans* that has subsequently become widely disseminated.

Acknowledgements

We are grateful for the help of project students Lisa Lister, Teniola Oyeleye, Arnie Dickin, Shaheda Ahmed, Susanna Sherwin and Nicola Griffiths in screening of the collection of *S. mutans* isolates.

References

1. Ajdic D, McShan WM, McLaughlin RE, Savic G, Chang J, Carson MB, et al. Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. Proc Natl Acad Sci U S A 2002: **99**: 14434–14439.
2. Beighton D, Russell RRB, Whiley RA. A simple biochemical scheme for the differentiation of *Streptococcus mutans* and *Streptococcus sobrinus*. Caries Res 1991: **25**: 174–178.
3. Brehm K, Ripio MT, Kreft J, Vazquez-Boland JA. The *bvr* locus of *Listeria monocytogenes* mediates virulence gene repression by α-glucosides. J Bacteriol 1999: **181**: 5024–5032.
4. Colby SM, Harrington DJ, Russell RRB. Identification and genetic characterisation of melibiose-negative isolates of *Streptococcus mutans*. Caries Res 1995: **29**: 407–412.
5. Cote CK, Cvitkovitch D, Bleiweis AS, Honeyman AL. A novel β-glucoside-specific PTS locus from *Streptococcus mutans* that is not inhibited by glucose. Microbiology 2000: **146**: 1555–1563.
6. Cote CK, Honeyman AL. The transcriptional regulation of the *Streptococcus mutans bgl* regulon. Oral Microbiol Immunol 2002: **17**: 1–8.
7. Cote CK, Honeyman AL. The LicT protein acts as both a positive and a negative regulator of loci within the *bgl* regulon of *Streptococcus mutans*. Microbiology 2003: **149**: 1333–1340.
8. Fux CA, Shirtliff M, Stoodley P, Costerton JW. Can laboratory reference strains mirror 'real-world' pathogenesis? Trends Microbiol 2005: **13**: 58–63.
9. Giammarinaro P, Paton JC. Role of RegM, a homologue of the catabolite repressor protein CcpA, in the virulence of *Streptococcus pneumoniae*. Infect Immun 2002: **70**: 5454–5461.
10. Hall BG, Xu L. Nucleotide sequence, function, activation, and evolution of the cryptic *asc* operon of *Escherichia coli* K12. Mol Biol Evol 1992: **9**: 688–706.
11. Keyhani NO, Roseman S. Wild-type *Escherichia coli* grows on the chitin disaccharide, N,N′-diacetylchitobiose, by expressing the *cel* operon. Proc Natl Acad Sci U S A 1997: **94**: 14367–14371.
12. Keyhani NO, Wang LX, Lee YC, Roseman S. The chitin disaccharide, N,N′-diacetylchitobiose, is catabolized by *Escherichia coli* and is transported/phosphorylated by the Phosphoenolpyruvate:Glycose phosphotransferase system. J Biol Chem 2000: **275**: 33084–33090.
13. Kharat AS. Phenotypic variability of α-glucoside utilization and its correlation to pathogenesis process in a few enteric bacteria. FEMS Microbiol Lett 2001: **199**: 241–246.
14. Kilic AO, Tao L, Zhang Y, Lei Y, Khammanivong A, Herzberg MC. Involvement of *Streptococcus gordonii* β-glucoside metabolism systems in adhesion, biofilm formation, and *in vivo* gene expression. J Bacteriol 2004: **186**: 4246–4253.
15. Kozliak EI, Fuchs JA, Guilloton MB, Anderson PM. Role of bicarbonate/CO_2 in the inhibition of *Escherichia coli* growth by cyanate. J Bacteriol 1995: **177**: 3213–3219.
16. Lau PC, Sung CK, Lee JH, Morrison DA, Cvitkovitch DG. PCR ligation mutagenesis in transformable streptococci: application and efficiency. J Microbiol Methods 2002: **49**: 193–205.
17. Li YH, Lau PC, Lee JH, Ellen RP, Cvitkovitch DG. Natural genetic transformation of *Streptococcus mutans* growing in biofilms. J Bacteriol 2001: **183**: 897–908.
18. Robinson WG, Old LA, Shah DSH, Russell RRB. Chromosomal insertions and deletions in *Streptococcus mutans*. Caries Res 2003: **37**: 148–156.
19. Rogers JD, Scannapieco FA. RegG, a CcpA homolog, participates in regulation of amylase-binding protein A gene (abpA) expression in *Streptococcus gordonii*. J Bacteriol 2001: **183**: 3521–3525.
20. Rosey EL, Stewart GC. Nucleotide and deduced amino acid sequences of the *lacR*, *lacABCD*, and *lacFE* genes encoding the repressor, tagatose 6-phosphate gene cluster, and sugar-specific phosphotransferase system components of the lactose operon of *Streptococcus mutans*. J Bacteriol 1992: **174**: 6159–6170.
21. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 2000: **132**: 365–386.
22. Rupf S, Merte K, Eschrich K, Stosser L, Kneist S. Peroxidase reaction as a parameter for discrimination of *Streptococcus mutans* and *Streptococcus sobrinus*. Caries Res 2001: **35**: 258–264.
23. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. Bioinformatics 2000: **16**: 944–945.
24. Sato Y, Okamoto K, Kagami A, Yamamoto Y, Igarashi T, Kizaki H. *Streptococcus mutans* strains harboring collagen-binding adhesin. J Dent Res 2004: **83**: 534–539.
25. Schulte D, Hengstenberg W. Engineering the active center of the 6-phospho-β-galactosidase from *Lactococcus lactis*. Protein Eng 2000: **13**: 515–518.
26. Setlow B, Cabrera-Hernandez A, Cabrera-Martinez RM, Setlow P. Identification of aryl-phospho-β-D-glucosidases in *Bacillus subtilis*. Arch Microbiol 2004: **181**: 60–67.
27. Simpson CL, Russell RR. Intracellular alpha-amylase of *Streptococcus mutans*. J Bacteriol 1998: **180**: 4711–4717.
28. Simpson CL, Russell RR. Identification of a homolog of CcpA catabolite repressor protein in *Streptococcus mutans*. Infect Immun 1998: **66**: 2085–2092.
29. Tao L, LeBlanc DJ, Ferretti JJ. Novel streptococcal-integration shuttle vectors for gene cloning and inactivation. Gene 1992: **120**: 105–110.
30. Ushiro I, Lumb SM, Aduse-Opoku J, Ferretti JJ, Russell RRB. Chromosomal deletions in melibiose-negative isolates of *Streptococcus mutans*. J Dent Res 1991: **70**: 1422–1426.
31. Vadeboncoeur C, Pelletier M. The phosphoenolpyruvate:sugar phosphotransferase system of oral streptococci and its role in the control of sugar metabolism. FEMS Microbiol Rev 1997: **19**: 187–207.
32. Weber BA, Klein JR, Henrich B. The *arbZ* gene from *Lactobacillus delbrueckii* subsp. *lactis* confers to *Escherichia coli* the ability to utilize the β-glucoside arbutin. Gene 1998: **212**: 203–211.
33. Whiley RA, Beighton D. Current classification of the oral streptococci. Oral Microbiol Immunol 1998: **13**: 195–216.