The *ftf* gene encoding the cell-bound fructosyltransferase of *Streptococcus salivarius* ATCC 25975 is preceded by an insertion sequence and followed by *FUR1* and *clpP* homologues

PHILIP M. GIFFARD, CATHERINE RATHSAM, EDWARD KWAN, DEBORAH W. L. KWAN, KIM L. BUNNY, SU-PIN KOO and NICHOLAS A. JACQUES*

Institute of Dental Research, 2 Chalmers Street, Surry Hills, New South Wales 2010, Australia

(Received 2 November 1992; revised 18 December 1992; accepted 14 January 1993)

Analysis of the region downstream of the *ftf* gene of *Streptococcus salivarius* led to the detection of two open reading frames (ORFs). The deduced amino acid sequences of these ORFs were homologous to proteins encoded by genes not previously described and/or sequenced in Gram-positive bacteria. The deduced amino acid sequence of the first of these (*orf2*) showed strong homology to the product of the *FUR1* gene of *Saccharomyces cerevisiae*, which codes for a uracil phosphoribosyltransferase. The over-expression of the product of this gene appeared to be the source of the detrimental effect observed with phagemids carrying *orf2* in *Escherichia coli* hosts. The deduced amino acid sequence of the second ORF (*orf3*) was homologous to the ClpP family of proteases. Examination of the upstream region of the *ftf* gene led to the discovery of a new insertion sequence-like element which has been designated IS1161.

Introduction

Streptococcus salivarius is an abundant member of the human oral microflora. Most strains of this organism synthesize both glucans and fructans from sucrose by means of extracellular glucosyltransferases and extracellular cell-bound fructosyltransferases (FTFs) respectively (Walker & Jacques, 1987; Milward & Jacques, 1990). Recently, we isolated an FTF-encoding bacteriophage (λ A-39) from a genomic library of *S. salivarius* ATCC 25975; Southern hybridization and physiological evidence have indicated that this recombinant bacteriophage harbours the gene for the extracellular cell-bound FTF of *S. salivarius* (S. P. Koo, unpublished observations; Pitty *et al.*, 1989).

Here we report the physical mapping of the *S. salivarius* chromosomal DNA in λ A-39 and the sequencing of the DNA surrounding the *ftf* gene. Analysis of these sequences has revealed three open reading frames

(ORFs), one upstream and two downstream of the *ftf* gene, that show clear homology to genes of topical interest.

Methods

Chemicals and enzymes. All chemicals were purchased from Sigma, BDH or Ajax Chemicals and were of analytical reagent grade or equivalent. Radioactively labelled deoxynucleotides were purchased from NEN, restriction enzymes and T4 ligase from Pharmacia and/or Boehringer Mannheim and the T7 DNA sequencing kit from Pharmacia. Custom oligonucleotides were synthesized using a Pharmacia 'Gene Assembler Plus'.

Bacterial strains, bacteriophages, plasmids, and growth conditions. Phagemids were maintained either in *Escherichia coli* strain JM109 (Yanisch-Perron *et al.*, 1985) or in strain NM522 (Gough & Murray, 1983), while bacteriophage λ A-39 (Pitty *et al.*, 1989) was propagated on *E. coli* LE392 (Murray *et al.*, 1977). All subclones (Fig. 1) were based on one or the other of the phagemid vectors pIBI30 or pIBI31 (IBI Corporation) and production of single-stranded DNA from these subclones was induced by superinfection by bacteriophage M13KO7 (IBI Corporation). LB medium (Miller, 1972), supplemented where appropriate with ampicillin (100 μ g ml⁻¹), was used throughout these studies.

DNA manipulations. DNA from bacteriophage λ A-39 was purified from 1 litre lysates by the method of Silhavy *et al.* (1984). The construction and purification of phagemids followed the standard methods described by Sambrook *et al.* (1989), while induction and purification of single-stranded phagemid DNA using bacteriophage M13KO7 followed the method recommended by Pharmacia for their pTZ series of phagemids. DNA sequence determination (Sanger *et al.*, 1977) was carried out either on CsCl-purified double-stranded DNA or

* Author for correspondence. Tel. 61 2 2820353; fax 61 2 2820368.

Abbreviations: FTF, fructosyltransferase; UPRTase, uracil phosphoribosyltransferase; ANGIS, Australian National Genome Information Service.

The nucleotide sequence data reported in this paper have been submitted to GenBank and have been assigned the accession numbers L07793 (sequence containing *orf2* and *orf3*) and L07794 (sequence containing *orf4*).

on single-stranded DNA using the Pharmacia T7 sequencing kit according to the manufacturer's instructions. Custom-made oligonucleotide primers (17mers) were used where necessary and all sequencing was confirmed in both directions.

DNA sequence analysis. DNA sequences were assembled and ORFs detected using the IBI-Pustell sequence analysis software version 2.03. Database searches were carried out using the 'Fasta' program (Pearson & Lipman, 1988) while multiple sequence alignments and phylogenetic comparisons used the 'ClustalV' program (Higgins *et al.*, 1992). This program makes use of the neighbour-joining method of Saitou & Nei (1987) to deduce phylogenetic trees. Fasta and ClustalV were accessed through the ANGIS facility at the University of Sydney, Australia.

Detection of FTF activity. FTF activity was detected using a qualitative assay for released reducing sugars as described previously (Aduse-Opoku *et al.*, 1989; Giffard *et al.*, 1991).

Results and Discussion

Physical map of the *S. salivarius* chromosomal DNA in λ A-39

A restriction map of the insert of chromosomal DNA in λ A-39 was generated by means of double restriction digests and a series of phagemid subclones constructed based on this map (Fig. 1). *E. coli* strains carrying these phagemids had the phenotypes outlined in Table 1.

The first subclone to be constructed was pKRK102, which contained the central *Bam*HI fragment (Fig. 1).

Fig. 1. Physical map of the *ftf* locus of *S. salivarius* ATCC 25975. The fragments cloned into pIBI30/31 to form the pKRK series of plasmids are shown above the map. Coding regions are indicated by the filled regions. E, *Eco*RI; B, *Bam*HI; B2, *Bgl*II; S, *Sac*I.

Table 1. Phenotypes conferred on *E. coli* by phagemid subclones

FTF activity was detected as described in the text.

Phagemid	FTF Activity	Phagemid yield
pKRK102	+	($\pm$)*
pKRK104	+	+++
pKRK105	—	++
pKRK107	—	+++

*pKRK102 was almost impossible to maintain and had to be prepared from isolated colonies on solid growth medium.

This phagemid conferred an FTF⁺ phenotype on the host cell but was extremely difficult to maintain. In most instances the *E. coli* host died during attempts to propagate the recombinant phagemid in liquid culture. The difficulty in propagating pKRK102 was initially thought to be due to the product of the *ftf* gene carried on pKRK102. However, the stable FTF⁺ phenotype of strains carrying pKRK104 (the subclone containing the *Eco*RI–*Bgl*II insert; Fig. 1) as well as the stable phenotype of pKRK107 (the subclone containing the small *Bam*HI–*Eco*RI insert; Fig. 1), together with some slight difficulty encountered in maintaining pKRK105 (the subclone containing the small *Bgl*II–*Bam*HI insert; Fig. 1) raised the possibility that the cause of this phenomenon was not the product of the *ftf* gene itself. This was investigated further by DNA sequence determination.

The DNA sequence downstream of the *ftf* gene contains *FUR1* and *clpP* homologues

In order to investigate the observed detrimental effect of pKRK102 on *E. coli* hosts, the sequence of the 0.7 kbp *Bam*HI–*Bgl*II insert in pKRK105 (Fig. 1) was determined. Initial analysis of this sequence indicated that it contained fragments of two separate ORFs that were transcribed in the same direction. The sequences of these ORFs were extended until the complete sequences of both were obtained. This yielded a sequence of 1523 bp that contained the two ORFs starting with ATG codons and preceded by putative ribosome-binding sites (Fig. 2). The first of these, *orf2*, was 629 bp in length, while the other, *orf3*, separated by 173 bp from *orf2*, was 590 bp in length. Sequence analysis also showed that the *ftf* gene lies between the *Eco*RI and *Bgl*II sites, that it is transcribed in the same direction as *orf2* and *orf3*, and that it terminates 90 bp upstream of the beginning of *orf2* (Fig. 1 and unpublished results).

To seek potential homologues, the deduced amino acid sequences of *orf2* and *orf3* were compared with the GenBank translated ORF data base. The deduced amino acid sequences of *orf2* and *orf3* exhibited clear similarity to the *Saccharomyces cerevisiae* *FUR1* gene product and the ClpP family of polypeptides, respectively (Figs 3 and 4).

The *FUR1* gene encodes the enzyme uracil phosphoribosyltransferase (UPRTase) which catalyses the formation of uridine 5'-monophosphate (UMP) from uracil (Kern *et al.*, 1990). To our knowledge, *orf2* is the first sequenced *FUR1* homologue of prokaryotic origin. It is not known whether the putative product of *orf2* is in fact a UPRTase, although the 33% identity and 63% conservation of amino acids maintained across the prokaryote/eukaryote boundary not only suggests this

<i>S. salivarius</i>	MGKFQVISHPLI-----QHKLSILRR---EDTSTKDFRE--LVNEIAMLMG-YEVS
<i>Sacch. cerevisiae</i>	MNPLFFLASPFLLYLYIYPNKGFSVSKPRNLQKMSSEPFKNVLLPQTNQLGLYTII
	*. * * * *
<i>S. salivarius</i>	RD-----LPLEEVEIQTPITKTV-----QKQLSGK---KLAIVPILR
<i>Sacch. cerevisiae</i>	RKNNTTRPDFIFYSDRIRLLVEEGLNHLVPQKQIVETDTNENFEGVSFMGKICGVSLVR
	*. * * * *
<i>S. salivarius</i>	AGIGMVDGFLSLVPAKVGHIGMYRDEETLEPVEYLVKLPEDIDQRQIFVVDPLATGGS
<i>Sacch. cerevisiae</i>	AGESMEQGLRDCCRSVRIGKILQORDEETALPKLFYEKLPEDISERYVFLDPLATGGS
	** . * . * * * * *
<i>S. salivarius</i>	AILAVDSLKKRG--AANIKFVCLVAAPEGVKKLQDAHPDIDIYASLDEKLNENGYIVPG
<i>Sacch. cerevisiae</i>	AIMATEVLIKRGVKPERIYFLNLICSGEGIEKYHAAPFVRIVTGALDRGLDENKYLVP
	** * * * * *
<i>S. salivarius</i>	LGDAGDRLFGTK
<i>Sacch. cerevisiae</i>	LGDFGDRYYCV-
	*** ** . . .

Fig. 3. Alignment of the deduced amino acid sequence *orf2* of *S. salivarius* with the UPRTase of *Saccharomyces cerevisiae* (Kern *et al.*, 1990). Identical amino acids are denoted by asterisks, and functionally conserved amino acids by dots.

Fig. 4. (a) Multiple alignment of the deduced amino acid sequence of *orf3* of *S. salivarius* with the ClpPs of *E. coli* (Maurizi *et al.*, 1990a), liverwort (*Marchantia polymorpha*) (Ohya *et al.*, 1986), rice (*Oryza sativa*) (Hiratsuka *et al.*, 1989), tobacco (*Nicotiana tabacum*) (Shinozaki *et al.*, 1986) and wheat (*Triticum aestivum*) (Gray *et al.*, 1990); the first methionine is the initiator for each sequence. Sequence conservation is as shown in Fig. 3, except that identical amino acids are also shown above the six sequences. (b) Unrooted phylogenetic tree of the sequences in (a) constructed using the neighbour-joining method of Saitou & Nei (1987). Relative phylogenetic distances are indicated both numerically and by the length of the connecting lines.

but also implies a strong selective pressure for conservation of function.

The apparent lethality conferred on *E. coli* by pKRK102 would appear to be due to the presence of *orf2*

rather than the *ftf* gene, since interrupting *orf2* in pKRK104 yielded a phagemid that was no longer detrimental to host cells. Why the sequence of *orf2* should be detrimental to *E. coli* is unknown and is

somewhat surprising, since *E. coli* possesses its own UPRTase (Rasmussen *et al.*, 1986). It could, however, be speculated that unregulated overproduction of this enzyme would lead to a derangement of nucleotide synthesis. It should also be noted that as *S. salivarius* ATCC 25975 is auxotrophic for uracil (N. A. Jacques, unpublished observation), a UPRTase would be essential for the survival of the bacterium as UMP is the precursor for all pyrimidine nucleotide triphosphates.

The ClpP family of polypeptides comprises the proteolytic subunits of the Clp family of ATP-dependent proteases. The other subunit cleaves ATP and can belong to either the ClpA or ClpB family. These enzymes are highly conserved but their precise biological role has yet to be elucidated, although hypotheses have included degradation of damaged, degraded or abnormal proteins and chaperonin-like functions (Gottesman, 1989; Gottesman *et al.*, 1990; Maurizi *et al.*, 1990*a, b*; Squires & Squires, 1992). The clear demonstration that the putative product of *orf3* lies within this family (Fig. 4*a*) is of particular interest because in recent years it has become clear that *clp* homologues are found close to genes encoding surface or extracellular polypeptides in a range of bacteria (no simple explanation for this phenomenon has been forthcoming) and that in the case of *Streptococcus mutans* GS-5, its *ftf* gene is apparently preceded by a *clpB* homologue (Gottesman *et al.*, 1990). Thus, in the oral streptococci, there appears to be an association between *ftf* genes and *clp* genes. In this regard it may be pertinent that when the cell-bound FTF of *S. salivarius* is inactivated by oxygen radicals it is proteolytically removed from the cell surface (Jacques & Wittenberger, 1981). The nature of the proteinase(s) involved is unknown.

The phylogenetic tree shown in Fig. 4(*b*) emphasizes that the product of *orf3* may be assigned to the ClpP family. However, it also shows that the *E. coli* and *S. salivarius* sequences are more closely related to each other than to the chloroplast sequences. The topological separation of the bacterial sequences from the chloroplast sequences was tested by the 'bootstrap' sampling method (Felsenstein, 1985) using 1000 samples and gave a 100% confidence limit. The origin of higher-plant chloroplasts is currently a subject of debate (Lockart *et al.*, 1992, and references therein) but it is generally believed that chloroplasts arose from cyanobacterial ancestors. According to Woese (1987), there is some evidence that the Gram-positive bacteria and the cyanobacteria form a natural group that does not include the purple bacteria (which include the *Enterobacteriaceae*). Therefore, these results appear to be at variance with current phylogenetic hypotheses. One possible explanation for this is that the evolutionary pressures on ClpP protease sequences are different in

chloroplasts compared with free-living bacteria. Alternatively, the chloroplast *clpP* genes may have originated in the plant genome and so be phylogenetically somewhat remote from the bacterial sequences. It would be of interest to determine the sequence of *clpP* homologues in cyanobacteria and/or the allied genus *Prochloron* (cf. Lockart *et al.*, 1992).

DNA upstream of the ftf gene codes for an insertion sequence

In view of the association between *clp* homologues and the *ftf* genes in *S. salivarius* and *S. mutans*, it was considered worthwhile to determine whether the *S. salivarius* *ftf* gene was preceded by a *clpB* homologue as it is on the *S. mutans* chromosome.

A stretch of 1101 bp of DNA was sequenced in pKRK108 upstream of the *EcoRI* site in the *S. salivarius* chromosomal insert in λ A-39 (Fig. 1). Database searches failed to show any homology between this sequenced region and the *clp* gene families. However, the sequenced region did contain the beginning of an ORF that started with an ATG codon and was preceded by a putative ribosome-binding site (*orf4*; Fig. 5). Sequence determination was continued 240 bp downstream of the *EcoRI* site and the end of *orf4* was found 46 bp downstream of this *EcoRI* site. This region also contains the beginning of the *ftf* gene (Figs 1 and 5, and unpublished results).

The potential product of *orf4* is 326 amino acids in length (Fig. 5). A search of the GenBank translated ORF data base revealed clear similarity with two transposases – one encoded by the *E. coli* insertion sequence IS30 (Dalrymple *et al.*, 1984) and the other by the clindamycin-resistance transposon Tn4551 which is found in the *Bacteroides* plasmid pBI136 (Smith, 1987) (Fig. 6). Similarity between these two transposable elements has not previously been noted.

Transposable genetic elements are usually delineated by inverted repeats and flanked by direct repeats. The presence of such repeats is considered diagnostic for a sequence that is the product of a transposition event (Lewin, 1990). Such repeats were found in the expected positions with respect to *orf4*, and these are shown in Fig. 7. The structure within the boundary of these repeats can therefore be considered to be an insertion-sequence-like element and has been named IS1611. One interesting property common to both IS30 and the insertion sequences that delineate Tn4551 is that they can apparently increase the expression of downstream genes (Smith, 1987; Neuwald & Stauffer, 1990). It is not known whether IS1611 also possesses this property, but it is interesting to note that *S. salivarius* ATCC 25975 expresses three to four times the amount of FTF as two

```

IS1161 MNMSTNYSTTNQSYK-----HLSEAERGEIEAYL--SVGLKPAEIAARRLGRNR
IS30   MRRTITAEKASVFELWKNGTGFSEITNIGSKPGTIFTMLRDTGGIKPHERKRAVAHLT
Tn4551 MSKHITTEQRYAI-----SMM-----
      *
      . .

IS1161 STITREINRGSITQV--KKVNGAKGLLPTLLCRCCSITVIRHAREASYLLKLDVSDDFM
IS30   LSEREEIRAGLSAKMSIRAIATALNRSPSTISR--EVQRN---RGRYYKAVDANNRANR
Tn4551 -----LQIPMSKKAIAEAIGVDKSTVYR--EIKRNCDAISGSYSMELAQRKADRR
      . . . * . . . * . . . * . . .

IS1161 RA-----FTDAMREKPRVHSVDTFVHTYRLQHVDAVVPSTK-----TLNYNI
IS30   MAKRPKPCLLDQNLPLRKL--LEKLEMKWSPEQISGWLRRTKPRQKTLRISPETIYKTL
Tn4551 KQQKHKREVLTP--AMRKRI--IKLLKKGFSPEQIVGRSRL---EGIAMVSHETIYRWI
      . . . . . . . . . . * . * .

IS1161 ---HQGLLEIKVIDLPRVRIRKFKTKRPSTKKHLG-----KSIEERPEEINNRSRF
IS30   YFRSREALHHLNIQHLRRSHSLR--HGRRHTRKGERGTINIVNGTPIHERSRNIDNRRSL
Tn4551 WEDKRRG-GKLH-KYLR--QGR--YAKRGSKNAGRGI--PGRVDIDERPEIVELKERF
      . . . * * * . * . . * . * . . .

IS1161 GDWEIDSVLGKKTIGEPSILTVERQTRYAVTKKLVEKKAQYVNVQAVLE-CMKLYP--IK
IS30   GHWEGDLVSGTKN-SH--IATLVDRKSRYTIILRLRGKDSVSVNQALTDKFLSLPSELRK
Tn4551 GDLEIDTIIGKNHKA--ILTINDRATSRVWIRKLSGKEAIPVAK-IAVWALRKVKVNLH
      * . * . * . . * . * . . * . * . . .

IS1161 SITADNGNEFSS---LSKIEGLDVYFAHAYSSYERGNTNENFNGLLREFIPKGCSELKELNQ
IS30   SLTWDRGMELARHLEFTVSTGVKVYFCDPQSPWQGTNTNENGLIRQYFPKKTCLAQYTQ
Tn4551 TITADNGKEFAKHEEIAQKLEIKFYFCKPYHSWEGANENTNGLIRQYIPKCKDFSEVTN
      . * * * * . . . . * . . . * . * . * . * . . .

IS1161 NLEDYTKAINERPRRI-----
IS30   HELDLVAAQLNNRPRKTLKFKTPKE-----I IERGVALTD
Tn4551 KQIKWIENKLNRRPRKRLGYLTPNEKFKQIINQNSVAFAS
      . . . * . * . * .

```

Fig. 6. Multiple alignment of the deduced amino acid sequence of *orf4* of *S. salivarius* with the presumptive transposases from IS30 (Dalrymple *et al.*, 1984) and Tn4551 (Smith, 1987). Sequence conservation is shown as in Fig. 3.

(a)

```

      20      30      40      50      60
AGTTATTATTTTAAACAAAAAATAGTTTTTGCCTATTGTAAATGAA
: : : : : : : : : : : : : : : : : : : : : : : : : :
AATGATATTTTGTAAAAAATATTTTCCCTAATTGTCAAGTCAA
      1230      1220      1210      1200

      70      80
GTGCAACACAAAGACATGTTCTCTGAT
: : : : : : : : : : : : : : : : : : : : : : : :
GTGCAACAAAG-----ATGTTCTCTGAT
1190      1180      1170

```

(b)

```

      30      40
AAAAATAGTTTTT
: : : : : : : :
AAAAATAGTTTTT
      1210      1220

```

Fig. 7. Repeated sequences flanking *orf4*. The sequence numbering is taken from Fig. 5. (a) Inverted repeats; (b) direct repeats.

other strains of *S. salivarius* (Milward & Jacques, 1990). The relationship between the presence of IS1611 and FTF expression is currently being investigated.

Addendum

Since submission of this work, a paper by Dong *et al.* (1992) has come to our attention describing a transposase, IS1086, from *Alcaligenes eutrophus* that by pairwise alignment possesses 28% identical and 52%

identical plus conserved amino acids with the putative transposase of IS1611, bringing the total number of insertion sequences in this family to four.

This work was supported by a Project Grant awarded by the Australian National Health and Medical Research Council. Both E.K. and S.-P.K. were in receipt of Dr Fred Trebitsch Scholarships and D.W.L.K. was supported by a Summer Studentship awarded by the Australian Dental Research Fund Inc.

References

- ADUSE-OPUKU, J., GILPIN, M. L. & RUSSELL, R. R. B. (1989). Genetic and antigenic comparison of *Streptococcus mutans* fructosyltransferase and glucan-binding protein. *FEMS Microbiology Letters* **50**, 279–282.
- DALRYMPLE, B. P., CASPERS, P. & ARBER, W. (1984). Nucleotide sequence of the mobile genetic element IS30. *EMBO Journal* **3**, 2145–2149.
- DONG, Q., SADOUK, A., VAN DER LELIE, D., TAGHAVI, S., FERHAT, A., NUYTEN, J. M., BORREMANS, B., MERGEAY, M. & TOUSSAINT, A. (1992). Cloning and sequencing of IS1086, an *Alcaligenes eutrophus* insertion element related to IS30 and IS4351. *Journal of Bacteriology* **174**, 8133–8138.
- FELSENSTEIN, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.
- GIFFARD, P. M., SIMPSON, C. L., MILWARD, C. P. & JACQUES, N. A. (1991). Molecular characterization of a cluster of at least two glucosyltransferase genes in *Streptococcus salivarius* ATCC 25975. *Journal of General Microbiology* **137**, 2577–2593.
- GOTTESMAN, S. (1989). Genetics of proteolysis in *Escherichia coli*. *Annual Review of Genetics* **23**, 163–198.

- GOTTESMAN, S., SQUIRES, C., PICHERSKY, E., CARRINGTON, M. H., HOBBS, M., MATTICK, J. S., DALYRYMPLE, B., KURAMITSU, H., SHIROZA, T., FOSTER, T., CLARK, W. P., ROSS, B., SQUIRES, C. L. & MAURIZI, M. A. (1990). Conservation of the regulatory subunit for the Clp ATP-dependent protease in procaryotes and eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 3513–3517.
- GOUGH, J. A. & MURRAY, N. E. (1983). Sequence diversity among related genes for recognition of specific targets in DNA molecules. *Journal of Molecular Biology* **166**, 1–19.
- GRAY, J. C., HIRD, S. M. & DYER, T. A. (1990). Nucleotide sequence of the wheat chloroplast gene encoding the proteolytic subunit of an ATP-dependent protease. *Plant Molecular Biology* **15**, 947–950.
- HIGGINS, D. G., BLEASBY, A. J. & FUCHS, R. (1992). CLUSTAL V: improved software for multiple sequence alignment. *Computer Applications in the Biosciences (CABIOS)* **8**, 189–191.
- HIRATSUKA, J., SHIMADA, H., WHITTIER, R., ISHIBASHI, T., SAKAMOTO, M., MORI, M., KONDO, C., HONJI, Y., SUN, C.-R., MENG, B.-Y., LI, Y.-Q., KANNO, A., NISHIZAWA, Y., HIRAI, A., SHINOZAKI, K. & SUGIURA, M. (1989). The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Molecular and General Genetics* **217**, 185–194.
- JACQUES, N. A. & WITTENBERGER, C. L. (1981). Inactivation of cell-associated fructosyltransferase in *Streptococcus salivarius*. *Journal of Bacteriology* **148**, 912–918.
- KERN, L., DE MONTIGNY, J., JUND, R. & LACROUTE, F. (1990). The *FUR1* gene of *Saccharomyces cerevisiae*: cloning, structure and expression of wild-type and mutant alleles. *Gene* **88**, 149–157.
- LEWIN, B. (1990). *Genes IV*, pp. 649–662. Oxford: Oxford University Press.
- LOCKART, P. J., BEANLAND, T. J., HOWE, C. J. & LARKUM, A. W. D. (1992). Sequence of *Prochloron didemni atpDE* and the inference of chloroplast origins. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 2742–2746.
- MAURIZI, M. A., CLARK, W. P., KATAYAMA, Y., RUDIKOFF, S., PUMPHREY, J., BOWERS, B. & GOTTESMAN, S. (1990a). Sequence and structure of ClpP, the proteolytic component of the ATP-dependent Clp protease of *Escherichia coli*. *Journal of Biological Chemistry* **265**, 12536–12545.
- MAURIZI, M. A., CLARK, W. P., KIM, S.-H. & GOTTESMAN, S. (1990b). ClpP represents a unique family of serine proteases. *Journal of Biological Chemistry* **265**, 12546–12552.
- MILLER, J. H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- MILWARD, C. P. & JACQUES, N. A. (1990). Secretion of fructosyltransferase by *Streptococcus salivarius* involves the sucrose-dependent release of the cell-bound form. *Journal of General Microbiology* **136**, 165–169.
- MURRAY, N. E., BRAMMAR, W. J. & MURRAY, K. (1977). Lambdaoid phages that simplify the recovery of *in vitro* recombinants. *Molecular and General Genetics* **150**, 53–61.
- NEUWALD, A. F. & STAUFFER, G. V. (1990). IS30 activation of an *smp'-lacZ* gene fusion in *Escherichia coli*. *FEMS Microbiology Letters* **56**, 13–17.
- OHYAMA, K., FUKUZAWA, H., KOHCHI, T., SHIRAI, H., SANO, T., SANO, S., UMESONO, K., SHIKI, Y., TAKEUCHI, M., CHANG, Z., AOTA, S., INOKUCHI, H. & OZEKI, H. (1986). Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature, London* **322**, 572–574.
- PEARSON, W. R. & LIPMAN, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 2444–2448.
- PITTY, L. J., GIFFARD, P. M., GILPIN, M. L., RUSSELL, R. R. B. & JACQUES, N. A. (1989). Cloning and expression of glycosyltransferase activities from *Streptococcus salivarius*. *Journal of Dental Research* **68**, 1681–1682.
- RASMUSSEN, U. B., MYGIND, B. & NYGAARD, P. (1986). Purification and some properties of uracil phosphoribosyltransferase from *Escherichia coli* K12. *Biochimica et Biophysica Acta* **881**, 268–275.
- SAITOU, N. & NEI, M. (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.
- SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- SANGER, F., NICKLEN, S. & COULSEN, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463–5467.
- SHINOZAKI, K., OHME, M., TANAKA, M., WAKASUGI, T., HAYASHIDA, N., MATSUBAYASHI, T., ZAITA, N., CHUNWONGSE, J., OBOKATA, J., YAMAGUCHI-SHINOZAKI, K., OHTO, C., TORAZAWA, K., MENG, B. Y., SUGITA, M., DENO, H., KAMOGASHIRA, T., YAMADA, K., KUSUDA, J., TAKAIWA, F., KATO, A., TOHDOKI, N., SHIMADA, H. & SUGIURA, M. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO Journal* **5**, 2043–2049.
- SILHAVY, T. J., BERMAN, M. L. & ENQUIST, L. W. (1984). *Experiments with Gene Fusions*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- SMITH, C. J. (1987). Nucleotide sequence analysis of Tn4551: use of operon fusions to detect promoter activity in *Bacteroides fragilis*. *Journal of Bacteriology* **169**, 4589–4596.
- SQUIRES, C. & SQUIRES, C. L. (1992). The Clp proteins: proteolysis regulators or molecular chaperones. *Journal of Bacteriology* **174**, 1081–1085.
- WALKER, G. J. & JACQUES, N. A. (1987). Polysaccharides of oral streptococci. In *Sugar Transport and Metabolism in Gram-positive Bacteria*, pp. 39–68. Edited by J. Reiser & A. Peterkofsky. Chichester: Ellis Horwood.
- WOESE, C. R. (1987). Bacterial evolution. *Microbiological Reviews* **51**, 221–271.
- YANISCH-PERRON, C., VIEIRA, J. & MESSING, J. (1985). Improved M13 cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.