Role of the C-terminal YG repeats of the primer-dependent streptococcal glucosyltransferase, GtfJ, in binding to dextran and mutan

Kim B. Kingston,[†] Donna M. Allen and Nicholas A. Jacques

Author for correspondence: Nicholas A. Jacques. Tel: +61 2 9845 8763. Fax: +61 2 9845 7599.
e-mail: njacques@dental.wsahs.nsw.gov.au

Institute of Dental Research,
Centre for Oral Health,
Westmead Hospital,
PO Box 533, Wentworthville,
NSW 2145, Australia

The recombinant primer-dependent glucosyltransferase GtfJ of *Streptococcus salivarius* possesses a C-terminal glucan-binding domain composed of eighteen 21 aa YG repeats. By engineering a series of C-terminal truncated proteins, the position at which truncation prevented further mutan synthesis was defined to a region of 43 aa, confirming that not all of the YG motifs were required for the formation of mutan by GtfJ. The role of the YG repeats in glucan binding was investigated in detail. Three proteins consisting of 3·8, 7·2 or 11·0 C-terminal YG repeats were expressed in *Escherichia coli*. Each of the three purified proteins bound to both the 1,6- α -linked glucose residues of dextran and the 1,3- α -linked glucose residues of mutan, indicating that a protein consisting of nothing but 3·8 YG repeats could attach to either substrate. Secondary structure predictions of the primary amino acid sequence suggested that 37% of the amino acids were capable of forming a structure such that five regions of β -sheet were separated by regions capable of forming β -turns and random coils. CD spectral analysis showed that the purified 3·8 YG protein possessed an unordered secondary structure with some evidence of possible β -sheet formation and that the protein maintained this relatively unordered structure on binding to dextran.

Keywords: *Streptococcus*, protein structure, circular dichroism

INTRODUCTION

The ability of various strains of oral streptococci to form biofilms on oral surfaces is enhanced by the formation of α -D-glucans (Hamada & Slade, 1980; Loesche, 1986), which are synthesized from dietary sucrose by a family of secreted enzymes known as the glucosyltransferases (Gtfs) (Monchois *et al.*, 1999b). Gtfs are broadly characterized as Gtf-S- or Gtf-I-type enzymes, depending upon whether the glucan they produce is water-soluble or -insoluble. Gtf-I-type enzymes were historically defined as those that synthesized insoluble, linear 1,3- α -linked glucans, called

mutans (Walker, 1978), though the definition ‘mutan’ no longer holds as mixed-linkage, insoluble 1,3- α -linked glucans containing 1,6- α -branch points and/or up to 50% 1,6- α -linked glucose residues have been described (Walker, 1978; Simpson *et al.*, 1995a; Monchois *et al.*, 1999b). GTF-S-type enzymes, on the other hand, produce water-soluble 1,6- α -linked dextrans, which may or may not possess 1,3- α -branch points (Walker, 1978; Simpson *et al.*, 1995a; Monchois *et al.*, 1999b). Gtfs may be further classified as primer-dependent or primer-independent, depending upon whether they require a preformed dextran for sucrose polymerization to occur at a near maximum rate (Walker, 1978; Simpson *et al.*, 1995a). The role of the primer remains controversial. It has been suggested that the addition of exogenous primer dextran may simply act as an activator by promoting a conformational change within the catalytic site of the enzyme (Monchois *et al.*, 1999a, b). However, this ignores the observation that the primer can act as an acceptor and become part of the polymer product

[†] **Present address:** Agriculture Victoria – Rutherglen, RMB 1145, Chiltern Valley Road, Rutherglen, Victoria 2685, Australia.

Abbreviations: CD, circular dichroism; Gtf, glucosyltransferase; Gtf-I, water-insoluble glucosyltransferase; Gtf-S, water-soluble glucosyltransferase; GBD, glucan-binding domain; GST, glutathione S-transferase.

(Walker, 1978). Primer-dependent enzymes are also able to hydrolyse sucrose in the absence of primer, reflecting their ability to use water as an alternative acceptor to dextran (Walker, 1978).

The mechanism of glucan-chain-elongation is also poorly understood. Evidence exists to support models for the addition of glucose at the reducing end of the growing polymer, while other data suggest that the addition of glucose occurs at the non-reducing end (Monchois *et al.*, 1999b).

An example of a primer-dependent Gtf-I-type enzyme is the GtfJ secreted by *Streptococcus salivarius*. This enzyme synthesizes chains of 1,3- α -linked glucosyl residues onto a 1,6- α -linked dextran primer. The resulting mixed linkage mutan is insoluble and co-precipitates with its enzyme *in vitro*, thus inactivating the enzyme (Simpson *et al.*, 1995a). Consistent with the predicted primary structures of all other streptococcal Gtfs, GtfJ possesses four domains: a signal sequence of 35 aa, a 144 aa variable domain, an 861 aa catalytic domain containing key catalytic amino acids, and a 478 aa C-terminal glucan-binding domain (GBD) (Giffard *et al.*, 1991; Monchois *et al.*, 1999b). As their name implies, the GBDs of Gtfs are able to bind to glucans of the dextran variety (Ferretti *et al.*, 1987; Mooser & Wong, 1988; Kobayashi *et al.*, 1989; Abo *et al.*, 1991). This has led to the assumption that where Gtfs adhere to mutan, binding occurs by way of the 1,6- α -linked glucosyl residues present within the predominantly 1,3- α -linked mutan (Haas & Banas, 2000). The recent observation that the glucan-binding protein GbpA does not bind to pseudonigeran, a polymer containing only 1,3- α -linked glucosyl residues, has been taken as evidence to support this conclusion (Haas & Banas, 2000).

The sequence identity and conservation across the C-terminal GBD regions of various Gtfs have been shown to vary considerably by multiple sequence alignment algorithms, even when gaps have been inserted within individual amino acid sequences. The C termini of these proteins, however, contain a number of amino acid repeat sequences, variously named A, B, C and D, that possess intra-repeat homology ranging from 40 to 100% (Ferretti *et al.*, 1987; Shiroza *et al.*, 1987; Giffard *et al.*, 1991; Giffard & Jacques, 1994). While different Gtfs possess different combinations of these repeats (Monchois *et al.*, 1999b), all GBDs have been shown to contain a characteristic pattern of amino acids that is also found in the GbpA protein of *Streptococcus mutans* (Haas *et al.*, 1998; Haas & Banas, 2000), the dextranase inhibitor from *Streptococcus sobrinus* (Sun *et al.*, 1994) and the repeat regions of virulence proteins from other Gram-positive bacteria (Wren, 1991; Wren *et al.*, 1991; von Eichel-Streiber *et al.*, 1992; Haas & Banas, 2000). This pattern of amino acids, i.e. the 21 aa YG repeat, transcends and extends beyond the N-terminal boundaries of all other repeat sequences found in the GBDs of the Gtfs (Giffard & Jacques, 1994). These YG repeats appear to have arisen as a consequence of a series of duplications and rearrangements that may be part of an

ongoing process (Giffard & Jacques, 1994; Simpson *et al.*, 1995b). The role of various lengths of YG motif in glucan binding by Gtfs has not been evaluated, whereas the need for specific numbers of higher order A, B and C repeats within the GBDs of Gtfs for both binding to glucans and for catalytic activity has been assessed (Ferretti *et al.*, 1987; Kato & Kuramitsu, 1990; Abo *et al.*, 1991; Lis *et al.*, 1995; Vickerman *et al.*, 1996; Monchois *et al.*, 1998). In this study we confirm the generally accepted role of the GBD in the catalysis of Gtfs by examining its function in the GtfJ of *S. salivarius*, while also further examining the ability of purified stretches of contiguous YG repeats to bind to dextran and to mutan.

METHODS

Chemicals and enzymes. All chemicals were of analytical grade or equivalent and were purchased from Ajax, BDH or Sigma. Radioactively labelled deoxynucleotides and [U-¹⁴C]-glucosyl-labelled sucrose were purchased from NEN. Restriction enzymes, T4 DNA ligase, T4 DNA polymerase, T4 polynucleotide kinase, DNase I, lysozyme, DNA molecular mass markers and antibiotics were obtained from Genesearch, Promega or Boehringer-Mannheim. The T7 DNA sequencing kit was purchased from Amrad Pharmacia, and the Coomassie Plus Protein Assay Reagent was purchased from Pierce. Broad-range and polypeptide molecular mass markers for SDS-PAGE were obtained from Bio-Rad and Gradipore electrophoresis gel stain was purchased from Gradipore. Oligonucleotides used for mutagenesis were synthesized using a Pharmacia Gene Assembler Plus and desalted using NAP-10 Columns (Pharmacia) according to the instructions supplied by the manufacturer. DNA was routinely purified using the Wizard Minipreps DNA Purification System (Promega). DNA sequencing (Sanger *et al.*, 1977) used the Pharmacia Biotech T7 sequencing kit. Epoxy-activated Sepharose 6B, the glutathione S-transferase (GST) Gene Fusion System and dextran T-10 were also obtained from Pharmacia.

Detection of Gtf activity. Hydrolysis of sucrose by GtfJ, or one of its mutated forms, was detected using a qualitative microtitre reducing sugar test for liberated fructose (Giffard *et al.*, 1991). Polymerizing activity was quantified with [U-¹⁴C]glucosyl-labelled sucrose (Jacques, 1983), for which one unit of glucan-forming activity (U) was defined as the amount of Gtf or its mutated form that catalysed the incorporation of 1 μ mol of the glucose moiety of sucrose into water-insoluble polysaccharide min⁻¹. The specific activity in U (mg dry wt *E. coli* host)⁻¹ was used throughout the study.

Phagemids, plasmids, bacteriophage and bacterial strains.

Table 1 lists the various *E. coli* strains and vectors used in this study. *E. coli* cells harbouring appropriate vectors were grown with shaking (150 r.p.m.) at 37 °C in conical flasks containing Luria-Bertani medium (Miller, 1972) supplemented with antibiotics where appropriate.

C-terminal deletions of GtfJ. Two C-terminal-deleted forms of GtfJ, designated C1J and C2J, were used in this study. The mutated enzyme C1J was encoded by phagemid pGSG103 which contained an exonuclease III 3' deletion of *gtfJ* obtained during the initial sequencing of the gene (Table 1; Giffard *et al.*, 1991). pGSG103 possessed a 0.9 kbp deletion of the *S. salivarius* chromosomal insert present in pGSG101, which resulted in the expression of a C-terminally truncated GtfJ

Table 1. Bacterial strains, plasmids, bacteriophages and phagemids

Bacteria, bacteriophage, phagemid or plasmid	Relevant characteristics	Source or reference
Strain		
<i>E. coli</i> LE392	F ⁻ e14 ⁻ (<i>mcrA</i> ⁻) <i>hsdR514</i> (<i>r</i> _k ⁻ <i>m</i> _k ⁻) <i>supE44 supF58 lacY1</i> or Δ (<i>lacIZY</i>)6 <i>galK2 galT22 metB1 trpR55</i>	Murray <i>et al.</i> (1977)
<i>E. coli</i> NM522	F' <i>lacI</i> ^q Δ (<i>lacZ</i>) M15 <i>proA</i> ⁺ <i>B</i> ⁺ - <i>supE thi</i> Δ (<i>lac-proAB</i>) Δ (<i>hsdMS-mcrB</i>)5 (<i>r</i> _k ⁻ <i>m</i> _k ⁻ <i>mcrBC</i> ⁻)	Gough & Murray (1983)
<i>E. coli</i> CJ236	<i>dut</i> ⁻ <i>ung</i> ⁻ F'	Kunkel <i>et al.</i> (1987)
<i>E. coli</i> BL21	F ⁻ <i>ompT bsdS</i> (<i>r</i> _B ⁻ <i>m</i> _B ⁻) <i>gal</i>	Studier & Moffat (1986)
Phagemid		
pIBI30	Ap ^r , f1 origin replication, β -galactosidase, T3 and T7 polymerase promoters	IBI
pIBI31	Ap ^r , f1 origin replication, β -galactosidase, T3 and T7 polymerase promoters	IBI
pGSG101 (pGS101)	pIBI30 with GtfJ-encoding 6.8 kbp <i>SacI</i> - <i>Bam</i> HI fragment of λ A-8	Giffard <i>et al.</i> (1991)
pGSG103	pGSG101 possessing a 576 bp 3' deletion within the coding region of the <i>gtfJ</i> gene	This study
pGSG105	pIBI31 containing <i>Bgl</i> II- <i>Sph</i> I fragment of pGSG101	This study
pGSG106	pGSG105 encoding W1100STOP of GtfJ	This study
pGSG107	pGSG101 encoding W1100STOP of GtfJ	This study
Plasmid		
pGEX-3X	Ap ^r , <i>lacI</i> ^q , GST utilizing P _{tac} and <i>lac</i> operator	Pharmacia
pGEX-5X-1	Ap ^r , <i>lacI</i> ^q , GST utilizing P _{tac} and <i>lac</i> operator	Pharmacia
pGSG108	pGEX-3X containing the <i>Dra</i> I- <i>Eco</i> RI fragment from the <i>gtfJ</i> gene	This study
pGSG109	pGEX-5X-1 containing the <i>Eco</i> RI $\rightarrow$ end fragment of the <i>gtfJ</i> gene	This study
pGSG110	pGEX-3X containing the <i>Dra</i> I $\rightarrow$ end fragment of the <i>gtfJ</i> gene	This study
Bacteriophage		
M13K07	M13 carrying a mutated version of gene II	Dotto & Zinder (1984)

Fig. 1. Position of YG repeats (□) in the GBDs of the parental GtfJ and C-terminal-truncated proteins C1J and C2J. The numbers refer to the position of the amino acids in the primary sequence of the proteins.

which had lost 191 aa, or 8.05 YG repeats (Fig. 1). The second C-terminal-deleted form of GtfJ, C2J, was a site-directed mutated GtfJ in which the TGG codon, encoding Trp-1300, was converted to a stop codon, TAA. C2J was therefore missing 219 aa (10.15 YG repeats) from its C terminus (Fig. 1). The site-directed mutated *gtf*, encoding C2J, was created by first subcloning the *Bgl*II-*Sph*I fragment of pGSG101 (encoding the GBD of GtfJ) into the phagemid vector pIBI31 cleaved with *Bam*HI-*Sph*I, to form pGSG105. Single-stranded uracylated DNA was obtained from this construct by transformation of *E. coli* CJ236 (Kunkel *et al.*, 1987) by electroporation (Bio-Rad Gene Pulser; 1.6 kV, 200 W, 25 μ F) and superinfection with M13K07 (Dotto & Zinder, 1984). The

mutagenic primer GGTATCAAGGATAGTTAACGAAAT-ATCAATGGT, containing the stop codon TAA (bold italic) and a unique *Hpa*I site (italic), was annealed to the single-stranded DNA and extended with Klenow enzyme at room temperature (18 °C) for 20 min, followed by T4 DNA polymerase at 37 °C for 2 h. The resultant double-stranded phagemids were electroporated into *E. coli* JM109 and amplified. A sample of the required site-directed mutated DNA (encoded by pGSG106) was detected by *Hpa*I digestion prior to excising the mutated *Sau*3A-*Sph*I fragment from the remaining DNA and using it to replace the *Bgl*II-*Sph*I fragment of pGSG101 to create pGSG107, encoding the Gtf W1300STOP, or C2J. The nature of the mutation in pGSG107 was confirmed by sequencing.

GtfJ and the C-terminal truncated Gtfs, C1J and C2J, were individually expressed in 100 ml cultures of *E. coli* NM522 following transformation of the host cells with the appropriate phagemid. Each culture was grown to OD₆₀₀ ~ 1.000 and harvested by centrifugation (10000 g, 10 min, 4 °C). The cell pellet was washed in 20 mM potassium phosphate buffer (pH 6.0) containing 1 mM PMSF before being resuspended in the same buffer to a final volume of 5 ml. The cells were immediately disrupted by sonication at 4 °C (Branson Sonicator; 4 $\times$ 50 W for 30 s with 30 s cooling periods). Following centrifugation (25000 g, 30 min, 4 °C), the supernatant was used to assay enzyme activity.

Expression of contiguous YG repeats. A *Dra*I restriction site, 3843 bp into the ORF of the *gtfJ* gene, occurs at the beginning of the coding region for the eighth YG repeat of the GBD of GtfJ (Giffard *et al.*, 1991). Three proteins containing 3.8 YG

Fig. 2. Diagrammatic representation of YG regions (□) of the GBD of GtfJ that were expressed by the chimeric pGEX vectors. 3-8 YG protein, the protein expressed by the *DraI*-*EcoRI* insert of the *gtfJ* gene in pGSG108; 7-2 YG protein, the protein expressed by the *EcoRI*-*End* insert of the *gtfJ* gene in pGSG109; 11-0 YG protein, the protein expressed by the *DraI*-*End* insert of the *gtfJ* gene in pGSG110. The numbers refer to the position of the amino acids in the primary sequence of GtfJ.

repeats (Lys-1282 to Phe-1364), 7-2 YG repeats (Glu-1363 to Asn-1518) or 11-0 YG repeats (Lys-1282 to Asn-1518) of molecular masses 9215, 17030 and 26000 Da, respectively, (Fig. 2) were expressed as GST fusion proteins using the GST Gene Fusion System. The first of these constructs was made by cloning the *DraI*-*EcoRI* 244 bp fragment of the *gtfJ* gene into *SmaI*-*EcoRI*-restricted pGEX-3X to form pGSG108 (Table 1). To construct the other two fusion proteins, DNA was first obtained by PCR amplification of pGSG101 from the *EcoRI* site to the end of the *gtfJ* gene using the forward primer TGACTAACGAATTCTTCACAACT, homologous to the DNA sequence from 4079 to 4101 bp within the coding region of the *gtfJ* gene, and the reverse primer CACATTAATAA-TAGATTAGAATTCTGTCATAAA, complementary to the DNA sequence 58-90 bp downstream of the stop codon of the ORF of the *gtfJ* gene except for the two altered nucleotides shown in bold italic (Giffard *et al.*, 1991). The primers were designed so that the resulting PCR products could be restricted at both ends with *EcoRI* (italic). pGSG109 (Table 1) expressing the GST::7-2 YG protein was obtained by cloning this *EcoRI*-restricted PCR product into *EcoRI*-restricted pGEX-5X-1 and determining the correct orientation of the insert by sequencing across the *EcoRI* site. In a similar manner, pGSG110 encoding the GST::11-0 YG protein was obtained by cloning the *EcoRI*-restricted PCR product into *EcoRI*-restricted pGSG108, which already contained the *DraI*-*EcoRI* 244 bp fragment of the *gtfJ* gene (Table 1).

The glutathione fusion proteins were expressed in *E. coli* BL21 and extracted by disruption of washed cell suspensions in PBS containing 1 mM PMSF, 2 mM EDTA and 0.02% (w/v) NaN_3 using a glass bead stirring method, as described by Song & Jacques (1997). The fusion proteins were purified according to the protocol obtained with the GST Gene Fusion System, except that the 10 ml glutathione/Sepharose 4B column was equilibrated with PBS containing 1 mM PMSF, 2 mM EDTA and 0.02% (w/v) NaN_3 before one of the fusion proteins was bound to the column and subsequently washed with 3 vols of the same buffer. The column was then eluted with 3 vols 50 mM tetraethylammonium chloride buffer (pH 7.5) containing 50 mM KCl and 20 mM MgCl_2 , followed by 3 vols 50 mM tetraethylammonium chloride buffer (pH 7.5) containing 50 mM KCl, 20 mM MgCl_2 and 5 mM ATP to remove unbound proteins (Thain *et al.*, 1996). The bound YG proteins were then cleaved from the bound GST::YG fusion protein

with Factor Xa and eluted from the column by following the protocol supplied with the GST Gene Fusion System.

Formation of mutan. Given sufficient time, primer-dependent Gtf-I-type enzymes produce an insoluble 1,3- α -linked mutan at high sucrose concentrations in the absence of primer (Hare *et al.*, 1978; Walker, 1978). A water-insoluble glucan was therefore synthesized by incubating GtfJ with 5% (w/v) sucrose for 5 d in the absence of dextran T-10 primer (Simpson *et al.*, 1995a). ^{13}C -NMR analysis of the water-insoluble glucan formed under these conditions confirmed that it was a mutan possessing > 99.9% 1,3- α -linked glucosyl residues.

Detection of glucan binding. The binding of GtfJ, mutated enzymes and variously expressed contiguous YG repeats to dextran or mutan were assessed by their ability to attach either to a Sepharose/dextran T-10 matrix prepared as described by Mooser & Wong (1988), or to the mutan synthesized by GtfJ in the absence of dextran primer. ^{13}C -NMR analysis showed that dextran T-10 possessed $\geq 95\%$ 1,6- α -linked glucosyl residues (Simpson *et al.*, 1995a), whereas the mutan possessed > 99.9% 1,3- α -linked glucosyl residues (see above). To determine the degree of binding, 400 ng of the appropriate protein or 100 μl *E. coli* lysate was incubated for 30 min at 37 °C, with occasional mixing with 500 μg mutan or 30 μg Sepharose/dextran T-10 in 100 μl 20 mM potassium phosphate buffer (pH 6.5). If dextran inhibition of protein binding to mutan was to be determined, 10 μg dextran T-10 was also added to the mutan slurry. Following incubation, the mixture was centrifuged (13000 g, 5 min, 20 °C) and the supernatant removed. The pellet was washed six times by resuspending with vigorous vortexing for 30 s in 200 μl 10 mM potassium phosphate buffer (pH 6.5) and recentrifuging (13000 g, 5 min, 20 °C) before subjecting the slurry to SDS-PAGE. Studies involving whole-enzyme-binding utilized 15% Tris/glycine gels (Russell, 1978), whereas those investigating YG binding motifs used 16% Tris/tricine gels (Schägger & von Jagow, 1987; Lesse *et al.*, 1990). After electrophoresis, any proteins that had remained bound to the glucan polymers were detected with Coomassie blue staining. The various proteins did not bind to Sepharose alone, while GtfJ and the mutated enzymes were concentrated out of the *E. coli* lysate by binding to mutan, leaving < 5% contaminating *E. coli* proteins after the subsequent washes (as determined by SDS-PAGE analysis).

Circular dichroism (CD) spectra. For CD spectral analysis, the 3-8 YG protein obtained from the glutathione/Sepharose 4B column was extensively dialysed against 18 M Ω H_2O to remove any contaminating ions, and further purified by HPLC (LKB Bromma 2248) using a Vydac C18 reverse-phase HPLC column. The protein was eluted from the column using a 5-85% CH_3CN gradient and extensively dialysed against 18 M Ω H_2O . Mass spectral analysis of purified protein samples was undertaken by the Australian Government Analytical Laboratories. CD spectra were subsequently accumulated at 22 °C in 18 M Ω H_2O within the range of 250 nm to 190 nm at 1.0 nm decrements, with a scan speed of 20 nm min $^{-1}$ and a 1 nm bandwidth using a JASCO J-710/720 Spectropolarimeter with a 1 mm light path. Five scans were averaged to improve the signal to noise ratio. To determine any conformational change following binding of the 3-8 YG protein to dextran, dextran T-10 was added to give a final concentration of 100 μM . A baseline was taken with 18 M Ω H_2O , or 18 M Ω H_2O containing dextran, under the same conditions used for the sample and subtracted from each protein spectrum.

Secondary structure prediction of YG repeats. The method of Garnier was used to predict the secondary structure of the primary amino acid sequence of contiguous YG repeats

(Pearson & Lipman, 1988). The program is considered to have a 75% prediction accuracy and was accessed through Entigen (<http://www.bionavigator.com>).

RESULTS

A role for the C-terminal GBD in mutan synthesis

Deletion studies of primer-dependent Gtfs have implied that the GBD plays a role in mutan synthesis by allowing the enzymes to bind to their primer dextrans (1,6- α -linked glucans). Consequently, the binding of GtfJ to Sepharose/dextran T-10 was not unexpected (Table 2; see footnote ‡). GtfJ also bound to insoluble mutan containing > 99.9% 1,3- α -linked glucose residues, indicating that it was not simply 1,6- α -linked glucosyl residues to which this primer-dependent enzyme adhered (Table 2; see footnote ‡).

To confirm a role for the GBD in the catalysis of GtfJ, a series of C-terminal deletion mutants were analysed for enzyme activity. Mutant C1J, encoded by pGSG103 (Table 1), possessed a 191 aa deletion, equivalent to a loss of 8.05 of the 18 YG repeats present at the C terminus of GtfJ (Fig. 1). C1J retained the ability to hydrolyse sucrose and to synthesize water-insoluble glucan in the presence of dextran T-10, albeit at a much reduced rate [range 30–100 mU (mg dry wt *E. coli* host)⁻¹; Table 2]. Other exonuclease III deletion mutants from the original sequencing of the *gtfJ* gene (Giffard *et al.*, 1991) that were truncated near the *DraI* restriction site, 3843 bp within the ORF of *gtfJ*, were enzymically inactive (data not shown). These mutated genes expressed C-terminal-truncated proteins missing at least 250 aa (11 YG repeats). It was therefore decided to construct the C-terminal-truncated mutant, C2J,

which was missing 219 aa (10.15 of the 18 YG repeats) (Fig. 1). This protein was expressed by pGSG107 (Table 1), in which Trp-1300 had been replaced with a stop codon by site-directed mutagenesis. C2J did not form glucan, while only four of seven preparations retained the ability to hydrolyse sucrose, reflecting the catalytic instability of the truncated enzyme during preparation for analysis (Table 2; see footnote §).

These observations implied that the position at which truncation within the C-terminal GBD prevented further glucan synthesis could be confined to a region of 43 aa. However, both mutated proteins, C1J and C2J, retained the ability to attach to Sepharose/dextran T-10 and mutan (Table 2), indicating that the removal of up to ten of the 18 C-terminal GBD YG repeats was insufficient to prevent attachment to either glucan. This observation did not preclude the possibility that the mutated Gtfs were so altered in their binding to dextran that this influenced the acceptor/activation function of the proteins (Monchois *et al.*, 1999a, b).

The role of YG repeats of the GBD in the binding of GtfJ to dextran and mutan

The 18 YG repeats within the GBD of GtfJ possess the consensus sequence IGGXXYYFGANGXQ(A,V)(T,K)-GXXVI, suggesting that it is unlikely that specific YG repeats or sequences of repeats play a unique role in glucan binding. However, to determine whether different lengths of YG repeats were necessary for binding to dextran or mutan, three proteins derived from the GBD region deleted from mutated proteins C1J and C2J were constructed and expressed in *E. coli*.

Table 2. Properties of C-terminal-deleted GtfJ proteins

Numbers in parentheses indicate the number of repeat experiments. ND, Not detected [less than the order of error of the assay method, ± 6 mU (mg dry wt *E. coli* host)⁻¹].

Form of enzyme	Mutation	Missing YG repeats	Sucrose hydrolysis*	Water-insoluble glucan [mU (mg dry wt <i>E. coli</i> host) ⁻¹]		Binding to glucan‡	
				– Dextran T-10	+ Dextran T-10	Dextran (1,6- α -glucan)	Mutan (1,3- α -glucan)
GtfJ	None	0	+	27 $\pm$ 13 (6)	1137 $\pm$ 448 (6)	Yes	Yes
Mutated C1J	C terminus missing 176 aa	8.05	+	ND (4)	44 $\pm$ 10 (7)	Yes	Yes
Mutated C2J	C terminus missing 219 aa	10.15	$\pm$ (7)§	ND (5)	ND (5)	Yes	Yes

* Determined by a positive red colour change (+) in the microtitre assay for reducing sugars on the addition of sucrose. Comparison was made with appropriate negative (no sucrose, no source of enzyme) and positive (fructose) controls.

† Mean $\pm$ SEM.

‡ Ability to bind to Sepharose/dextran or mutan as determined by detection of bound protein following SDS-PAGE. The amount of GtfJ bound by Sepharose/dextran and mutan was 77 and 83%, respectively, according to the amount of ¹⁴C-labelled glucan formed by the bound enzyme following its addition to glucan matrices in the range 4500–15 000 mU ml⁻¹ (*n* = 4). Similar data for C1J and C2J could not be acquired.

§ Four of the seven preparations gave a positive red colour. The other three remained colourless and were thus negative for reducing sugar.

Fig. 3. Representative SDS-PAGE analysis of 400 ng YG proteins binding to dextran and mutan. (a) 11.0 YG protein; (b) 7.2 YG protein; (c) 3.8 YG protein. Lanes: 1, molecular mass markers; 2, purified protein; 3, protein that bound to 500 μ g mutan; 4, protein that remained unbound to 500 μ g mutan; 5, protein that remained bound to 500 μ g mutan in the presence of 10 μ g dextran T-10; 6, protein that was eluted from 500 μ g mutan by 10 μ g dextran T-10; 7, protein that bound to 10 μ g dextran T-10; 8, protein that remained unbound to 10 μ g dextran T-10.

These three proteins, containing either 3.8, 7.2 or 11.0 YG repeats, were expressed by pGSG108, pGSG109 and pGSG110, respectively (Table 1), and formed a contiguous set of proteins with molecular masses of 9215, 17030 and 26000 Da, respectively (Fig. 2). In a series of at least three repeat experiments, each of the three proteins was observed to bind to dextran and mutan, indicating that a protein consisting of 3.8 YG repeats could attach to either substrate (Fig. 3). All three YG proteins were partially eluted from excess insoluble mutan by a fiftieth the amount of dextran T-10 (Fig. 3, lane 6). Higher ratios of dextran to mutan released more of each of the proteins (data not shown).

Secondary structure prediction and CD spectra of the 3.8 YG protein

Secondary structure prediction, based on the primary amino acid sequence of the 83 aa 3.8 YG protein, showed that the percentage composition of α -helix: β -sheet: β -turn:random coil was 16:37:29:18, compared with that of 16:39:24:21 for the complete 478 aa GBD of GtfJ. The predicted α -helix structure of the 3.8 YG protein related to two contiguous regions of four and nine amino acids, the latter being at the C terminus of the 3.8

Fig. 4. Secondary structure prediction (Pearson & Lipman, 1988) of the 83 aa 3.8 YG protein. The numbers refer to the amino acids of the GtfJ protein from which the 3.8 YG protein was derived.

YG protein (Fig. 4). Prediction of amino acids involved in β -sheet formation related to eight separate regions of either one, two, four, five or six contiguous amino acids separated by regions capable of forming β -turns and random coils (Fig. 4). Ignoring each run of single and double amino acids, it was conceivable that the remaining five regions could form β -sheet structures. Similar conclusions have been drawn for the 406 aa GBD of GbpA of *S. mutans*, following secondary structure prediction using the Chou, Fasman and Rose algorithm (Haas *et al.*, 1998).

Fig. 5. CD spectrum of the expressed 3-8 YG proteins from the GBD of GtfJ in the absence (solid line) and presence of 100 μ g dextran T-10 (dashed line).

Attempts to scale-up and purify the 3-8 YG protein for further structural analysis ran into difficulties, due to the production of inclusion bodies within the *E. coli* host. A number of modifications of the method of expression were attempted, including the use of alternative strains of *E. coli*, altering the growth medium, altering the time of induction, lowering the growth temperature to below 37 °C, altering the amount of IPTG used to induce the protein, and co-induction with chaperonins GroES and GroEL to stabilize the 3-8 YG protein. In all instances these alterations to the expression protocol resulted in a decrease in soluble protein expression whether carried out alone or, as was attempted in some cases, in concert with one another (data not shown).

The final protocol used resulted in a low mean yield of 700 μ g purified soluble 3-8 YG protein (l culture)⁻¹ that appeared as a single band following SDS-PAGE analysis (data not shown). Mass spectral analysis, however, indicated that there were two proteins present with molecular masses of 9913 and 9332 Da. Both of these proteins differed from the predicted mass of the cleaved GST::3-8 YG protein expressed by pGSG108. The differences could be accounted for by the removal of either the initial G, or the sequence GIPGKGK from the N terminus of the expressed protein. In the latter instance only the sequence KGK formed part of the original *S. salivarius* 3-8 YG protein (the remaining amino acids being derived from the C terminus of the GST protein following cleavage with factor Xa). This loss of N-terminal amino acids was not considered significant since the mixture readily bound to mutan and dextran (data not shown). This anomaly was therefore ignored and the mixture of the two forms of the 3-8 YG protein was used for CD spectral analysis.

The CD spectrum of the 3-8 YG expressed proteins revealed a single positive broad band centred at 230 nm and a negative band at 200 nm, with a broad shoulder between 203 and 216 nm (Fig. 5). The negative ellipticity in the far-UV spectrum below 220 nm was consistent

with the CD spectrum theoretically predicted for, and obtained with, proteins and peptides possessing a high level of unordered secondary structure (Woody, 1974; Venyaminov & Vassilenko, 1994). The CD spectrum in this region differed from that obtained with the 406 aa GBD of GbpA where a positive band was observed at 202 nm, effectively confirming that the GBD of GbpA was composed of a high percentage of β -sheets (Haas *et al.*, 1998). However, the broad shoulder between 203 and 216 nm was consistent with the presence of some β -sheet structures within the 3-8 YG protein (Venyaminov & Vassilenko, 1994). The presence of seven (8.4%) tyrosine, five (6.0%) phenylalanine and one (1.2%) tryptophan residues in the 3-8 YG protein may have obscured the secondary structure of the 3-8 YG protein (Woody, 1978; Manning & Woody, 1989; Fasman, 1996). This possibility has been commented on before, since the broad positive band centred at 230 nm in the CD spectrum of the 3-8 YG expressed proteins is also apparent in the CD spectrum of the GBD of GbpA. The positive broad band centred at 230 nm is most likely due to the high percentage of aromatic amino acids, especially tyrosine, within the 3-8 YG expressed proteins (Haas *et al.*, 1998) rather than to the folding of the polypeptide backbone.

As dextran T-10 was not optically active and therefore not detected by CD, the CD spectrum of the 3-8 YG expressed proteins bound to dextran T-10 was also obtained. The addition and binding to dextran T-10 at an equivalent concentration to that of the 3-8 YG expressed proteins caused an increase in the negative ellipticity at 200 nm while maintaining the shoulder between 203 and 216 nm. There was a corresponding decrease in the broad band centred at 230 nm (Fig. 5). The change in spectrum was consistent with a dilution effect, with little change in the predicted unordered structure of the 3-8 YG expressed proteins upon binding to dextran T-10.

DISCUSSION

The nature of the role for the GBD in Gtf catalysis remains unresolved. For instance, while our kinetic data show that the two primer-dependent Gtfs of *S. salivarius*, GtfJ and GtfK, are activated by and use dextrans as acceptors (Simpson *et al.*, 1995a), as do other Gtfs (Monchois *et al.*, 1999b), the nature and location of the acceptor-binding sites and whether they have anything to do with the GBDs remains unknown. Unfortunately, sequential removal of YG repeats from the C terminus of GtfJ does not in itself answer whether the GBD is the acceptor site in this enzyme. Sequential truncation of up to 10-15 of the 18 YG repeats in the GBD leads to the loss of glucan synthesis. Despite this, the C-terminal-truncated mutated enzymes retaining eight or more YG repeats were still capable of binding to dextran and mutan (Table 2). This finding is contrary to reports for other Gtfs (Kato & Kuramitsu, 1990; Abo *et al.*, 1991). Our finding that an intact GBD is required for the maximum rate of catalysis, however, confirms the findings of others (Ferretti *et al.*, 1987; Kato &

Kuramitsu, 1990; Abo *et al.*, 1991; Lis & Kuramitsu, 1995; Vickerman *et al.*, 1996; Monchois *et al.*, 1998), while resolving the point of loss of glucan synthesis to a region of 43 aa within the primary sequence of GtfJ. In all of these deletion studies, however, one observation appears to stand out: a specific minimal length of YG repeats at the end of the catalytic domain may be required for sucrose polymerization and further copies of these domains may be required for a maximum rate of catalysis either in the presence or absence of dextran. Thus it may not be the sequential removal of a specific number of YG motifs *per se*, or the removal of a unique combination of the higher order A, B and/or C repeats (Ferretti *et al.*, 1987; Kato & Kuramitsu, 1990; Lis & Kuramitsu, 1995; Vickerman *et al.*, 1996; Monchois *et al.*, 1998) that leads to the loss of polymerizing activity, but rather truncation to an overall minimum length, since the GBDs in Gtfs differ in length (Giffard & Jacques, 1994; Simpson *et al.*, 1995b; Monchois *et al.*, 1999b). The only exception to this is the recent finding that the catalytic domain of the primer-activated Gtf-I-type enzymes, the Gtfls of *Streptococcus downei* and *Streptococcus sobrinus*, devoid of their entire GBD, maintain polymer-forming activity at very high sucrose concentrations while losing their ability to be activated by dextran T-10 (Konishi *et al.*, 1999; Monchois *et al.*, 1999a). The Gtfl of *S. downei* differs from primer-dependent GtfJ in possessing a low affinity for sucrose and only being stimulated 1.5-fold by relatively high concentrations of dextran T-10. The authors concluded that deletions of C-terminal GBDs from different Gtfs clearly have different effects (Monchois *et al.*, 1999a). These differences may simply reflect variations or differences in catalytic mechanisms. Clearly some enzymes, like GtfJ, are highly dependent on dextrans for maximum activity while others, like the Gtfl of *S. salivarius*, are inhibited by dextran despite the obvious similarity in primary amino acid sequence (Simpson *et al.*, 1995a). YG repeats in the GBD may spatially orientate acceptor glucans towards the active site or, alternatively, may orientate and direct the glucan product away from the active site during polymer growth (Jacques, 1995; Monchois *et al.*, 1998). In a study of hybrid *S. mutans* enzymes, created by swapping regions of the GBDs of the primer-independent Gtf-I-type enzyme GtfB and the primer-dependent Gtf-S-type enzyme GtfD, less than 1% of the native activity was observed in the hybrid enzymes (Nakano & Kuramitsu, 1992). This may have been the consequence of the fact that the hybrid enzymes were constructed in a random manner by swapping different lengths of the GBDs, thus altering the spacial relationship between the YG repeats and the catalytic domains.

Given sufficient time, the primer-dependent GtfJ produces an insoluble 1,3- α -linked mutan at high sucrose concentrations in the absence of primer. As far as we are aware, the fact that the 3.8, 7.2 or 11.0 YG proteins were able to bind to this insoluble mutan represents the first report of portions of the GBD of a Gtf being able to bind to 1,3- α -linked glucosyl residues of mutan, rather than

the generally held assumption that Gtfs bind to stretches of 1,6- α -linked glucosyl residues present in a mixed-linkage mutan (Haas & Banas, 2000). The observation that YG proteins partially dissociated from mutan in the presence of a fiftieth the amount of added dextran T-10 (Fig. 3) is consistent with a related finding that GbpA, with its 16 C-terminal YG repeats, has a greater affinity for dextran than for an insoluble mixed-linkage mutan (Haas & Banas, 2000). *In vitro*, primer-dependent Gtfs, such as GtfJ, are inactivated by binding to, and precipitating out of solution with, their insoluble mutan product (Simpson *et al.*, 1995a). In dental plaque, it can be envisaged that soluble dextrans, synthesized by Gtf-S-type enzymes, would not only act as acceptors, but also as releasing agents for Gtf-I-type enzymes bound to mutan on the tooth surface, thus allowing a further round of mutan synthesis. This would result in insoluble and adhesive mutan polymers being formed with a high concentration of mixed 1,3- α - and 1,6- α -linkages while preventing pure 1,3- α -linked mutans from coalescing into insoluble ribbon-like structures that would be readily swallowed along with the soluble dextran products of the Gtf-S-type enzymes (Hare *et al.*, 1978; Walker, 1978; Jacques, 1994).

What is evident from our results is that arrays of contiguous YG repeats are able to bind to glucans with different three-dimensional structures, implying that a high degree of flexibility is most likely conferred by contiguous YG repeats. While it was not possible to determine the CD spectrum of the 3.8 YG protein binding to mutan containing only 1,3- α -linked glucosyl residues, due to the insolubility of the mutan, it was clear from the CD spectrum in the presence of dextran T-10 that the relatively unordered conformation was maintained on binding to dextran T-10. Binding by way of the clustered aromatic residues, tyrosine, tryptophan and phenylalanine, present in the YG repeats, would be expected to initially stabilize the protein-glucan complex, while the polar amino acids lysine, serine, threonine, asparagine and glutamine and the acidic amino acids, aspartic acid and glutamic acid, prevalent in the YG repeats, would allow the creation of hydrogen bonds with hydroxyl residues of the sugar (Quioco, 1986). The presence of high concentrations of glycine in the YG consensus sequence IGGXXYYFGANGXQ-(A,V)(T,K)GXXVI of the GBD of GtfJ would be expected to confer flexibility to the GBD allowing it to adjust to, and bind to, different glucan structures. While secondary structure predictions confirmed this (Fig. 4), it remains to be determined whether the 18 YG repeats present in the GBD would retain this unordered structure when associated with the larger catalytic domain in the native GtfJ protein. The flexibility of YG repeats, however, is perhaps best exemplified by the fact that contiguous YG repeats are not limited just to the GBDs of the Gtfs (Giffard & Jacques, 1994). They are also found in other Gram-positive bacterial proteins that bind to alternative substrates. For example, the binding domain containing YG repeats in the pneumococcal surface protein A of *Streptococcus pneumoniae* binds to

choline residues in lipoteichoic acid (Yother & Briles, 1992), a non-sugar residue, whereas toxin A from *Clostridium difficile* binds to the Type 2 polysaccharide core, Gal β 1-4GalNAc, present in epithelial glycoproteins (Wren *et al.*, 1991). While the structure of this latter disaccharide is significantly different to 1,6- α - and 1,3- α -glucans, it has been shown that antibodies raised against the YG repeats of toxin A react with GbpA from *S. mutans* (Wren *et al.*, 1991). With such a requirement for flexibility it is perhaps not surprising that the CD spectrum of the 3·8 YG protein shows a high degree of unordered structure.

ACKNOWLEDGEMENTS

This work was supported by the Australian National Health and Medical Research Council. We wish to thank Dr Philip Giffard for his help in producing the mutant C2J, Dr Joel Mackay for help with purifying the 3·8 YG protein and its subsequent CD analysis, and Associate Professor Norman Cheetham for ^{13}C -NMR analysis of mutant.

REFERENCES

- Abo, H., Matsumura, T., Kodama, T., Ohta, H., Fukui, K., Kato, K. & Kagawa, H. (1991). Peptide sequences for sucrose splitting and glucan binding within *Streptococcus sobrinus* glucosyltransferase (water-insoluble glucan synthetase). *J Bacteriol* **173**, 989–996.
- Dotto, G. P. & Zinder, N. D. (1984). Reduction of the minimal sequence for initiation of DNA synthesis by qualitative or quantitative changes of an initiator protein. *Nature* **311**, 279–280.
- von Eichel-Streiber, C., Sauerborn, M. & Kuramitsu, H. K. (1992). Evidence for a modular structure of the homologous repetitive C-terminal carbohydrate-binding sites of *Clostridium difficile* toxins and *Streptococcus mutans* glucosyltransferases. *J Bacteriol* **174**, 6707–6710.
- Fasman, G. D. (1996). *Circular Dichroism and the Conformational Analysis of Biomolecules*. New York: Plenum.
- Ferretti, J. J., Gilpin, M. L. & Russell, R. R. B. (1987). Nucleotide sequence of a glucosyltransferase gene from *Streptococcus sobrinus* Mfe28. *J Bacteriol* **169**, 4271–4278.
- Giffard, P. M. & Jacques, N. A. (1994). Definition of a fundamental repeating unit in streptococcal glucosyltransferase glucan-binding regions and related sequences. *J Dent Res* **73**, 1133–1141.
- 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. *J Gen Microbiol* **137**, 2577–2593.
- Gough, P. M. & Murray, N. L. (1983). Sequence diversity among related genes for recognition of specific targets in DNA molecules. *J Mol Biol* **166**, 1–19.
- Haas, W. & Banas, J. A. (2000). Ligand-binding properties of the carboxyl-terminal repeat domain of *Streptococcus mutans* glucan-binding protein A. *J Bacteriol* **182**, 728–733.
- Haas, W., MacColl, R. & Banas, J. A. (1998). Circular dichroism analysis of the glucan binding domain of *Streptococcus mutans* glucan binding protein-A. *Biochim Biophys Acta* **1384**, 112–120.
- Hamada, S. & Slade, H. D. (1980). Biology, immunology, and cariogenicity of *Streptococcus mutans*. *Biol Rev* **44**, 331–384.
- Hare, M. D., Svensson, S. & Walker, G. J. (1978). Characterization of the extracellular, water-insoluble α -D-glucans of oral streptococci by methylation analysis, and by enzyme synthesis and degradation. *Carbohydr Res* **66**, 254–264.
- Jacques, N. A. (1983). Membrane perturbation by cerulenin modulates glucosyltransferase secretion and acetate uptake by *Streptococcus salivarius*. *J Gen Microbiol* **129**, 3293–3302.
- Jacques, N. A. (1994). The glucosyltransferases of *Streptococcus salivarius*. *Aust Dent J* **39**, 111–114.
- Jacques, N. A. (1995). Extracellular sucrose metabolism by *Streptococcus salivarius*. In *Genetics of Streptococci, Enterococci and Lactococci*. Edited by J. J. Ferretti, M. S. Gilmore, T. R. Klaenhammer & F. Brown. *Dev Biol Stand* **85**, 315–322.
- Kato, C. & Kuramitsu, H. K. (1990). Carboxy-terminal deletion analysis of the *Streptococcus mutans* glucosyltransferase-I enzyme. *FEMS Microbiol Lett* **72**, 299–302.
- Kobayashi, S., Koga, K., Hayashida, O., Nakano, Y. & Hasegawa, Y. (1989). Glucan-binding domain of a glucosyltransferase from *Streptococcus sobrinus*: isolation of a 55-kilodalton peptide trypsin digest of glucosyltransferase prebound to insoluble glucan. *Infect Immun* **57**, 2210–2213.
- Konishi, N., Torii, Y., Yamamoto, T. & 8 other authors (1999). Structure and enzymatic properties of genetically truncated forms of the water-insoluble glucan-synthesizing glucosyltransferase from *Streptococcus sobrinus*. *J Biochem* **126**, 287–295.
- Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol* **154**, 367–382.
- Lesse, A. J., Campagnari, A. A., Bittner, W. E. & Apicella, M. A. (1990). Increased resolution of lipopolysaccharides and lipooligosaccharides utilizing tricine-sodium dodecyl sulphate-polyacrylamide gel electrophoresis. *J Immunol Methods* **126**, 109–117.
- Lis, M., Shiroza, T. & Kuramitsu, H. K. (1995). Role of the C-terminal direct repeating units of the *Streptococcus mutans* glucosyltransferase-S in glucan binding. *Appl Environ Microbiol* **61**, 2040–2042.
- Loesche, W. L. (1986). Role of *Streptococcus mutans* in human dental decay. *Microbiol Rev* **50**, 353–380.
- Manning, M. C. & Woody, R. W. (1989). Theoretical study of the contribution of aromatic side chains to the circular dichroism of basic bovine pancreatic trypsin inhibitor. *Biochemistry* **28**, 8609–8613.
- Miller, J. H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Monchois, V., Reverte, A., Remaud-Simeon, H., Monsan, P. & Willemot, R. M. (1998). Effect of *Leuconostoc mesenteroides* NRRL B-512F dextransucrase carboxy-terminal deletions on dextran and oligosaccharide synthesis. *Appl Environ Microbiol* **64**, 1644–1649.
- Monchois, V., Arguello-Morales, M. & Russell, R. R. B. (1999a). Isolation of an active catalytic core of *Streptococcus downei* Mfe28 GTF-I glucosyltransferase. *J Bacteriol* **181**, 2290–2292.
- Monchois, V., Willemot, R.-M. & Monsan, P. (1999b). Glucansucrases: mechanism of action and structure-function relationships. *FEMS Microbiol Rev* **23**, 131–151.
- Mooser, G. & Wong, C. (1988). Isolation of glucan-binding domain of glucosyltransferase (1,6- α -glucan synthase) from *Streptococcus sobrinus*. *Infect Immun* **56**, 880–884.
- Murray, N. E., Brammar, W. H. & Murray, K. (1977). Lambdaoid phages that simplify the recovery of *in vitro* recombinants. *Mol Gen Genet* **150**, 53–61.
- Nakano, Y. J. & Kuramitsu, H. K. (1992). Mechanism of *Streptococcus mutans* glucosyltransferases: hybrid-enzyme analysis. *J Bacteriol* **174**, 5639–5646.
- Pearson, W. R. & Lipman, D. J. (1988). Improved tools for

- biological sequence analysis. *Proc Natl Acad Sci USA* **85**, 2444–2448.
- Quijcho, F. A. (1986).** Carbohydrate-binding proteins: tertiary structures and protein–sugar interactions. *Annu Rev Biochem* **55**, 287–315.
- Russell, R. R. B. (1978).** Glycosyltransferases of *Streptococcus mutans* strain Ingbritt. *Microbios* **23**, 135–146.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977).** DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* **74**, 5463–5467.
- Schägger, H. & von Jagow, G. (1987).** Tricine-sodium dodecyl sulphate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem* **166**, 368–379.
- Shiroza, T., Ueda, S. & Kuramitsu, H. K. (1987).** Sequence analysis of the *gtfB* gene from *Streptococcus mutans*. *J Bacteriol* **169**, 4263–4270.
- Simpson, C. L., Cheetham, N. W. H., Giffard, P. M. & Jacques, N. A. (1995a).** Four glucosyltransferases, GtfJ, GtfK, GtfL and GtfM, from *Streptococcus salivarius* ATCC 25975. *Microbiology* **141**, 1451–1460.
- Simpson, C. L., Giffard, P. M. & Jacques, N. A. (1995b).** *Streptococcus salivarius* ATCC 25975 possesses at least two genes coding for primer-independent glucosyltransferases. *Infect Immun* **63**, 609–621.
- Song, D. D. & Jacques, N. A. (1997).** Cell disruption of *Escherichia coli* by glass bead stirring for the recovery of recombinant proteins. *Anal Biochem* **248**, 300–301.
- Studier, F. W. & Moffat, B. A. (1986).** Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113–130.
- Sun, J. W., Wanda, S. Y., Camilli, A. & Curtiss, R. (1994).** Cloning and DNA sequencing of the dextranase inhibitor gene (*dei*) from *Streptococcus sobrinus*. *J Bacteriol* **176**, 7213–7222.
- Thain, A., Gaston, K., Jenkins, O. & Clarke, A. R. (1996).** A method for the separation of GST fusion proteins from co-purifying GroEL. *Trends Genet* **12**, 209–210.
- Venyaminov, S. Y. & Vassilenko, K. S. (1994).** Determination of protein tertiary structure class from circular dichroism spectra. *Anal Biochem* **222**, 176–184.
- Vickerman, M. M., Sulavik, M. C., Minick, P. E. & Clewell, D. B. (1996).** Changes in the carboxy-terminal repeat region affect extracellular activity and glucan products of *Streptococcus gordonii* glucosyltransferase. *Infect Immun* **64**, 5117–5128.
- Walker, G. J. (1978).** Dextran. *Int Rev Biochem* **16**, 75–126.
- Woody, R. W. (1974).** Studies of theoretical circular dichroism of polypeptides: contributions of beta-turns. In *Peptides, Polypeptides and Proteins*, pp. 338–350. Edited by E. R. Blout, F. A. Bovey, N. Lotan & M. Goodman. New York: Wiley.
- Woody, R. W. (1978).** Aromatic side-chain contributions to the far ultraviolet circular dichroism of peptides and proteins. *Biopolymers* **17**, 1451–1467.
- Wren, B. W. (1991).** A family of clostridial and streptococcal ligand-binding proteins with conserved C-terminal repeat sequences. *Mol Microbiol* **5**, 797–803.
- Wren, B. W., Russell, R. R. B. & Tabaqchali, S. (1991).** Antigenic cross-reactivity and functional inhibition by antibodies to *Clostridium difficile* toxin A, *Streptococcus mutans* glucan-binding protein, and a synthetic peptide. *Infect Immun* **59**, 3151–3155.
- Yother, J. & Briles, D. E. (1992).** Structural properties and evolutionary relationships of PspA, a surface protein of *Streptococcus pneumoniae*, as revealed by sequence analysis. *J Bacteriol* **174**, 601–609.
- Received 28 June 2001; revised 6 September 2001; accepted 8 October 2001.