Limitations of the widely used GAM42a and BET42a probes targeting bacteria in the Gammaproteobacteria radiation

Christine Yeates, Aaron M. Saunders, Gregory R. Crocetti and Linda L. Blackall

Advanced Wastewater Management Centre, The University of Queensland, St Lucia, 4072, Australia

The 23S rRNA-targeted probes GAM42a and BET42a provided equivocal results with the uncultured gammaproteobacterium ‘Candidatus Competibacter phosphatis’ where some cells bound GAM42a and other cells bound BET42a in fluorescence in situ hybridization (FISH) experiments. Probes GAM42a and BET42a span positions 1027–1043 in the 23S rRNA and differ from each other by one nucleotide at position 1033. Clone libraries were prepared from PCR products spanning the 16S rRNA genes, intergenic spacer region and 23S rRNA genes from two mixed cultures enriched in ‘Candidatus C. phosphatis’. With individual clone inserts, the 16S rDNA portion was used to confirm the source organism as ‘Candidatus C. phosphatis’ and the 23S rDNA portion was used to determine the sequence of the GAM42a/BET42a probe target region. Of the 19 clones sequenced, 8 had the GAM42a probe target (T at position 1033) and 11 had G at position 1033, the only mismatch with GAM42a. However, none of the clones had the BET42a probe target (A at 1033). Non-canonical base-pairing between the 23S rRNA of ‘Candidatus C. phosphatis’ with G at position 1033 and GAM42a (G–A) or BET42a (G–T) is likely to explain the probing anomalies. A probe (GAM42_C1033) was optimized for use in FISH, targeting cells with G at position 1033 and was found to highlight not only some ‘Candidatus C. phosphatis’ cells, but also other bacteria. This demonstrates that there are bacteria in addition to ‘Candidatus C. phosphatis’ with the GAM42_C1033 probe target and not the BET42a or GAM42a probe target.

INTRODUCTION

An invaluable tool for the microbial ecologist has been the technique of fluorescence in situ hybridization (FISH), allowing the detection and in situ identification and quantification of bacteria within several different environmental samples (for an overview see Amann et al., 2001). FISH involves the mixing of a fixed microbial sample with fluorescently labelled oligonucleotide probes (Amann et al., 1990). Since the first report of FISH with bacteria (DeLong et al., 1989), FISH probes that specifically target the intracellular rRNAs of many different groups of microorganisms have been published. These phylogenetic group-specific (PGS) probes have been designed to target the whole bacterial or archaeal domains, several phyla, subphyla, genera and on occasion species (e.g. see Amann et al., 1995).

In addition, for uncultured environmental organisms where only 16S rDNA sequence information is available, PGS probes can provide a surrogate group definition (Crocetti et al., 2000, 2002; Hugenholtz et al., 2001).

The crux of reliable PGS probe design is the database of sequences from which probes can be designed. As more sequences become available, there is often a need to redesign probes to recognize cross-reacting groups or an altered phylogenetic group definition (e.g. Daims et al., 1999). Frequently, the information content of the 16S rRNA is a limiting factor. To overcome this, PGS probes have been designed to target other genes, most notably the 23S rRNA, such as the probes for the subphyla Gammaproteobacteria (GAM42a) and Betaproteobacteria (BET42a) in helix 42 of the 23S rRNA (Manz et al., 1992). However, there is a paucity of sequences for 23S rDNA in comparison to those for 16S rDNA and the phylogeny of the two genes may not be congruent. These issues are particularly relevant for investigations of uncultivated organisms.

Application of FISH probes requires optimization of the stringency of binding to avoid non-specific results. This is generally achieved with the use of controls such as cultured
organisms whose sequence in the rRNA probe target zone has zero or one mismatch to the designed oligonucleotide sequence. With PGS FISH probes that target uncultured bacterial groups, probe optimization is often carried out using environmental samples known to contain the organism and determined by reduction in fluorescent signal intensity of specific bacteria as observed by epifluorescence or confocal laser scanning microscopy. However, the specificity of any single probe to the target group is not guaranteed since the sample includes a large proportion of unknown organisms. A solution to this limitation is the design and application of multiple PGS probes that encompass the phylogenetic hierarchy of the uncultured target group. This has been described as the top-to-bottom approach (Amann et al., 1995). In any case, methodical re-evaluation and redesign of PGS probes should be practised (Daims et al., 1999; Loy et al., 2002) since the increasing and thus changing rDNA database affects the phylogenetic placement of sequences and consequent PGS FISH probe design (Amann et al., 1995). With the top-to-bottom approach, two of the most widely used PGS probes are GAM42a and BET42a, targeting helix 42 of the 23S rRNA from the subphyla Gammaproteobacteria and Betaproteobacteria, respectively (Manz et al., 1992). Use of these probes in FISH studies of wastewater communities has given unexpected results.

Microbial communities involved in wastewater treatment processes have been studied by FISH since 1993 (Wagner et al., 1993), and the procedure has led to the detection of several bacteria pivotal to success or failure of these processes (Crocetti et al., 2000, 2002; Juretschko et al., 1998; Rosselló-Mora et al., 1995; Schmid et al., 2000; Wagner et al., 1994a, b). Glycogen-accumulating organisms (GAOs) are considered important destabilizers of enhanced biological phosphorus removal (EBPR) due to their capacity to overgrow the organisms responsible for EBPR, the polyphosphate-accumulating organisms (PAOs). The taxon ‘Candidatus Competibacter phosphatis’ (henceforth called Competibacter) was recently identified and linked to the in situ GAO phenotype (Crocetti et al., 2002) and numerous PGS probes targeting Competibacter have been reported (Crocetti et al., 2002; Kong et al., 2002; Nielsen et al., 1999). Several researchers have noted that some cells hybridizing these Competibacter-specific FISH probes did not bind GAM42a, despite the phylogenetic placement of Competibacter in the Gammaproteobacteria subphylum (Crocetti et al., 2002; Kong et al., 2002; Liu et al., 2001; Nielsen et al., 1999). Nielsen et al. (1999) suggested this was related to the specificity of GAM42a which was originally designed based on a limited number of 235 rRNA sequences. Crocetti et al. (2002) and Kong et al. (2002) found some cells bound probes specific for Competibacter as well as the BET42a probe. Crocetti et al. (2002) quantified different bacterial groups and in two separate laboratory-scale processes, 88% and about 50% of the cells binding GAOQ431 (for Competibacter) also bound BET42a (for Betaproteobacteria). These results highlight the difficulties associated with the application of PGS probes for organisms without cultured representatives and the need for continual PGS probe re-evaluation. Research reported in this study investigated why the probing inconsistencies summarized above with Competibacter were occurring. This involved examining the sequence for helix 42 within the 23S rRNA gene from Competibacter. The study revealed that the probing inconsistencies were related to polymorphisms within the probe-binding site in helix 42 of the 23S rRNA.

### METHODS

**Clone library preparation, restriction fragment length polymorphism (RFLP) analysis and sequencing.** Separate bacterial clone libraries were generated from frozen-stored samples of both the Q and T sludge which have been described previously (Bond et al., 1999a, b; Crocetti et al., 2002). The Q sludge sample was collected on day 61 from a sequencing batch reactor (SBR) described by Bond et al. (1999a) and the T sludge sample was collected on day 225 from an SBR described by Bond et al. (1999b). DNA was extracted and purified (Yeates & Gillings, 1998) and primers were used to generate amplicons encompassing the 16S rDNA, intergenic spacer region (IGS) and the 23S rDNA. To selectively amplify DNA from Competibacter cells, the GAOQ989 probe (Table 1) was synthesized as a forward primer and a general bacteria-specific primer within the 23S rDNA was used as the reverse primer (1091r 5'-RGTAGCCTTTAAGGC-3'; Lane, 1991). The extracted purified Q or T sludge DNA solutions (10–100 ng) were PCR-amplified (Bond et al., 1995) except that 2 U Taq DNA Polymerase (Sigma) and 0.1 μg each primer were added, and a touchdown PCR program was used. The initial annealing temperature was 60 °C, which was reduced by 2 °C every 2 cycles with a final 29 cycles at 54 °C. The following thermal cycles were performed: 94 °C for 30 s; annealing temperature for 30 s; 72 °C for 2 min; with a final extension at 72 °C for 5 min.

Cloning of the purified PCR products was carried out as described by Bond et al. (1995). Clones were screened by PCR using plasmid primers (SP6 and T7) and conditions described above, except that the PCR thermal cycles were as follows: 94 °C for 30 s; 48 °C for 30 s; 72 °C for 2 min (30 cycles). The amplicons from each clone were digested with MspI (New England Biolabs) and grouped into operational taxonomic units (OTUs) based on the RFLP patterns. Representatives of some OTUs were fully sequenced using previously reported methods (Bond et al., 1995) except that the BigDye Terminator (Version 2; Applied Biosystems) sequencing kit was used. The primers used for sequencing included those described by Lane (1991) (1114f and 1492r for the 16S rDNA; 242r, 256f, 577f and 599r for the 23S rDNA) as well as primer 1492r (Lane, 1991) in the forward orientation.

**Analysis of sequence data.** Sequences for the 16S and 23S rDNAs were compiled and preliminarily analysed by BLAST (Altschul et al., 1997). The 23S rDNA sequence components of the clones (1060–1085 nt) were aligned in ARB (http://www.arb-home.de) and once missing or ambiguous characters were excluded, the phylogenetically analysed datasets contained 939 nt (23S rDNA). rDNA data analysis was performed by using previously reported methods (Björnsson et al., 2002). GenBank accession numbers for the sequences are presented in Table 2.

**FISH and microscopy.** FISH was performed on samples fixed in 4% paraformaldehyde (Amann et al., 1995). Probes (see Table 1) were commercially synthesized and 5’-labelled with either fluorescein isothiocyanate (FITC) or one of the sulfoindocyanine dyes, Cy3...
and Cy5 (Thermohybaid Interactiva). Samples were collected from a laboratory-scale anaerobic–aerobic cycled SBR enriched for the GAO phenotype and a full-scale EBPR wastewater treatment plant (Loganholme Wastewater Treatment Plant, South-East Queensland, Australia). Fixed samples from the laboratory-scale SBR were ground in a tissue homogenizer to break up large granules and flocs prior to probing. This pretreatment was not required for the fixed full-scale sludge samples. FISH-probed samples were observed and documented as described by Björnsson et al. (2002).

The GAM42a, BET42a and GAM42_C1033 probes all have only one base difference between them (Table 1). To increase the differentiation between probe and target due to the single base mismatch, unlabelled competitor probes were added (Manz et al., 1992). FISH using GAM42a, BET42a or GAM42_C1033 at all times included the labelled probe with the two others unlabelled and functioning as competitors at the same concentration.

The formamide concentration for optimum stringency of GAM42_C1033 was determined by performing FISH at formamide concentrations from 25 to 50% (5% increments) with paraformaldehyde-fixed pure cultures of a gammaproteobacterium and a betaproteobacterium isolated from activated sludge. The isolates were confirmed as members of these subphyla by analysis of their 16S rDNA sequences. Additionally, their 23S rDNA sequence at the GAM42a/BET42a probe target in helix 42 confirmed they possessed the correct probe target string and thus they were suitable as negative controls for probe GAM42_C1033. When the competitor probes were present, GAM42_C1033 did not bind to the control cultures at any formamide concentration. Therefore, the optimal stringency was determined to be at the formamide concentration where GAM42_C1033 no longer bound to the negative control cultures without the unlabelled competitor probes.

### RESULTS

**Clone library analysis**

The use of oligonucleotide GAOQ989 (Crocetti et al., 2002) as a forward primer in PCR should enrich the clone library in sequences originating from Competibacter, since GAOQ989 has a 100% match with most Competibacter sequences (Crocetti et al., 2002; Nielsen et al., 1999). Kong et al. (2002) reported Competibacter clone sequence data, but these were not available at the time our experiments were carried out. Although there is one mismatch between GAOQ989 and some of these sequences (subgroups 5 and 7 in Kong et al., 2002), they are centrally located by RFLP
analysis, 40 clones from the Q sludge fell into four OTUs. 70% of these (28 clones) were of a single type, 25% (10 clones) of a second type and the remaining two OTUs were each represented by a single clone. Two clones from OTU1, two clones from OTU2 and the representatives from OTUs 3 and 4 were selected for sequencing. RFLP analysis of the T sludge clones produced up to 22 OTUs. Resolution of individual OTUs was poor due to minor band length variations across the sample set. Thirteen clones that were considered broadly representative of the diversity in the T sludge were selected for sequencing.

The complete amplicon encompassing the 16S rDNA, IGS region and 23S rDNA was sequenced from two Q sludge clones and ten T sludge clones. Partial sequences only of the 16S rDNA portion and the 23S rDNA portion were generated from four Q sludge clones and three T sludge clones. Sequence analysis of the clones showed that for the T sludge the large number of OTUs did not indicate high diversity at the 16S rDNA or 23S rDNA level, because the clone sequences were 96–100% and 92–99% identical to each other, respectively. Examination of the 10 full-length sequences of the amplicons generated from the T sludge revealed that many of the restriction enzyme sites existed within the IGS (data not shown) and were likely to be responsible for the generation of the relatively large number of OTUs. The IGS regions from the clones of both sludges were found to contain genes for transfer RNAs (data not shown).

The DNA fragments sequenced here were derived from cloned PCR products that are assumed to come from the same cells. The 16S rDNA portion of each cloned fragment showed 96–100% identity to previously reported Competibacter sequences (Crocetti et al., 2002), identifying the source cells for the cloned fragments as Competibacter. BLAST searches using the 23S rDNA portions of the sequences showed the closest relatives of Competibacter to be Gammaproteobacteria at 88–93% and these sequences were included in the phylogenetic analysis shown in Fig. 1. Phylogenetic analysis of the 23S rDNA sequences placed Competibacter as a monophyletic group within the Gammaproteobacteria and outside the Betaproteobacteria (Fig. 1).

![Fig. 1. Evolutionary distance dendrogram of ‘Candidatus Competibacter phosphatis’ and some representatives from the Gammaproteobacteria and Betaproteobacteria based on phylogenetic analyses of 23S rDNA data in which 939 nt were compared. Only the longer generated sequences for Competibacter 23S rDNA were used in the analyses. The nucleotide at position 1033 for each sequence is indicated in square brackets. Branch points supported (bootstrap values > 74%) by all inference methods used are indicated by solid circles, and those supported by one inference method are indicated by open circles. Branch points without circles were not resolved (bootstrap values from all analyses < 75%). Clones sequenced in the present study are in bold type. The outgroup (not shown) was Rhodobacter capsulatus (X06485). The bar represents 10% estimated sequence divergence.](https://www.microbiologyreserach.org/FIG1.png)
Analysis of probe region for GAM42a and BET42a in the clones

The subphylum-specific probes GAM42a and BET42a were designed to target parts of helix 42 of the 23S rRNA of either the Gammaproteobacteria or Betaproteobacteria, respectively (Manz et al., 1992). These probes vary by one nucleotide at position 1033, where GAM42a has adenine (A) and BET42a has thymine (T) (see Table 1), with the complementary nucleotides in the 23S rRNA of cells being U or A, respectively. Sequence information was obtained for the region corresponding to the probe-binding site in 13 clones from the T sludge and 6 clones from the Q sludge. Position 1033 was the only variable nucleotide in the clones for this probe target, apart from clone Q19 which had a T at position 1043. Of the T sludge clones, eight had T at position 1033, while the other five clones had guanine (G) at position 1033. All six of the Q sludge clones had G at this position. Therefore, an additional probe called GAM42_C1033 was designed to bind to cells having G at position 1033 of the 23S rRNA (see Table 1). According to BLAST, there are no bacterial sequence matches in GenBank to the GAM42_C1033 probe.

Application of GAM42_C1033 probe

No binding of GAM42_C1033 to either of the negative control cultures was observed at any formamide concentration when unlabelled competitor probes were present. The sludge from a laboratory-scale anaerobic–aerobic cycling SBR demonstrated to have the GAO phenotype was examined by FISH with GAM42a, BET42a, GAM42_C1033, and GAOQMIX. Fig. 2(a–e) confirms the presence of numerous Competibacter cells (GAOQMIX binding) in this SBR. Most of these cells also bound either GAM42a or GAM42_C1033 (Fig. 2a, arrowed) which is consistent with the cloning results from the Q and T sludges (see above). However, there are some GAOQMIX-binding cells that also bound BET42a (Fig. 2b, arrowed) and some that bound none of the 23S rRNA-targeted probes (Fig. 2c, arrowed, and Fig. 2d). To examine the in situ arrangement of these cells, a non-homogenized sample from the laboratory-scale SBR was examined (Fig. 2e). Cells binding both GAM42_C1033 and GAOQMIX probes, or GAM42_C1033 alone or BET42a alone tended to exist as discrete clusters.

A full-scale treatment plant (Loganholme Sewage Treatment Plant) operating for EBPR had been previously shown to contain Competibacter cells that demonstrate the GAO phenotype (Crocetti et al., 2002). We found large clusters of Competibacter cells in Loganholme sludge that bound GAM42_C1033 (Fig. 2f) and additionally, some GAM42_C1033 binding cells in this sample did not bind GAOQMIX.

DISCUSSION

The ability to accurately assess microbial community structure in situ using FISH probes is reliant on the database of sequences available for probe design. As the amount of sequence information from uncultured bacteria increases, probe design has adapted to accommodate these sequences while maintaining probe specificity (e.g. Daims et al., 1999). Although the number of sequences described for 16S rDNA has been increasing rapidly, there has not been the same increase in 23S rDNA sequences. Only once before (Schmid et al., 2001) has a study attempted to obtain 16S rDNA, IGS and 23S rDNA sequences using PCR from the same uncultured environmental organism. The ability to link 16S rDNA and 23S rDNA sequence information will be likely to facilitate substantial improvements in FISH probe design.

Phylogenetic analysis of 16S and 23S rDNA sequences from Competibacter

The substantial 16S rDNA sequence database for Competibacter was initiated with partial sequences from denaturing gradient gel electrophoresis studies (Liu et al., 2000; Nielsen et al., 1999) which were complemented by near-complete sequences (Crocetti et al., 2002; Dabert et al., 2001) and additional partial sequences (Kong et al., 2002) from cloning studies. The 16S rDNA sequences described in our study were highly identical to other Competibacter sequences (Crocetti et al., 2002), confirming that Competibacter was the source for the linked 23S rDNA sequence information. The 23S rDNA sequence information therefore reflects both the GAM42a target site and phylogeny according to 23S rDNA for this organism.

The Proteobacteria, a much-studied and well-recognized phylum in the bacterial domain, is phylogenetically divided into several subphyla, including Betaproteobacteria and Gammaproteobacteria. Initially, the subdivisions were created with 16S rDNA data from pure-cultured representatives, but currently the 16S rDNA database is overwhelmingly composed of sequences obtained from cloning studies of 16S rDNA from diverse samples. In contemporary phylogenetic analyses, the Betaproteobacteria subphylum has been found to be an infrataxon of the Gammaproteobacteria, not a sister taxon (Ludwig & Klenk, 2001). The FISH probes used to identify these two subphyla are in the exact same part of the 23S rRNA and differ from each other by only one nucleotide.

The reason for probing inconsistencies

The research described in this paper was motivated by several reported probing inconsistencies where many Competibacter (a member of the Gammaproteobacteria) cells did not bind GAM42a (for Gammaproteobacteria) (Crocetti et al., 2002; Kong et al., 2002; Liu et al., 2001; Nielsen et al., 1999) and many Competibacter cells did bind BET42a (for Betaproteobacteria) (Crocetti et al., 2002; Kong et al., 2002). The 23S rDNA portion of the clones was used to determine the sequence of the probe region (positions 1027–1043) and specifically the only variable nucleotide in the probe region at position 1033. Of the 19 Competibacter clones, 8 had the GAM42a probe target, but 11 had G at
position 1033. Due to non-canonical base pairing, G-A or G-T hybrids can exist, and therefore GAM42a or BET42a could bind to Competibacter cells with G at position 1033. It is known that G-T and G-A hybrids are stronger than many other mismatches such as A-A, T-T, C-T or C-A (Lathe, 1990). The addition of an unlabelled competitor probe is required to differentiate the relatively weak A-A or T-T mismatch at position 1033 between GAM42a-Betaproteobacteria cells and BET42a-Gammaproteobacteria cells, respectively. So therefore, the absence of a competitor probe in FISH experiments studying organisms like Competibacter with the previously unknown G at position 1033 in their 23S rRNA, could readily lead to hybridization between them and GAM42a (G-A) or BET42a (G-T). The Competibacter cells with the perfect GAM42a probe target would bind GAM42a, but the Competibacter cells with G at position 1033 were likely to be responsible for the reported probing inconsistencies (Crocetti et al., 2002; Kong et al., 2002; Liu et al., 2001; Nielsen et al., 1999).

Helix 42 of the 23S rRNA – a good choice for probe design?

At the time GAM42a and BET42a were designed (Manz et al., 1992) their future comprehensiveness could not be guaranteed. Position 1027–1043 was the best region found by GAM42a in full-scale sludge samples, but their phylogenetic placement is not known. The sequence of the bacterial 23S rDNA from positions 1027–1043 should be comprehensively studied, especially in Betaproteobacteria and Gammaproteobacteria, and also in bacteria from other Proteobacteria lineages. The approach used in this study could be employed to address this issue with mixed microbial communities. It could be extended to the study of larger amplicons encompassing the complete 16S rDNA and more of the 23S rDNA. The ubiquity of Proteobacteria and the ecological significance of many of their members, has made GAM42a and BET42a two of the most widely used of all probes. Microbiologists should re-evaluate how these probes are used in the top-to-bottom approaches to community structure determination.

ACKNOWLEDGEMENTS

The authors would like to thank the operators of Loganholme Wastewater Treatment Plant for supplying sludge, Adrian Oehmen...

Fig. 2. Confocal laser scanning micrographs of FISH from a laboratory-scale SBR and from a full-scale EBPR process. Probes are indicated in Table 1; the bar on (a) is for all images and represents 10 µm. In all cases the colours of each probe are indicated in parentheses after the probe name. In superimposed images, the overlap between green and blue is cyan, between red and blue is magenta, and between red, green and blue is white. Images (a), (b) and (c) show different probe combinations in the same image of biomass from the laboratory-scale SBR. (a) GAM42a and GAM42_C1033 (both in red), and GAOQMIX (blue). The majority of the GAOQMIX-binding cells also bind either GAM42a or GAM42_C1033 (magenta, arrowed). (b). BET42a (green) and GAOQMIX (blue). Some of the GAOQMIX-positive cells also bind BET42a (cyan, arrowed). (c). GAM42a and GAM42_C1033 (both in red); BET42a (green) and GAOQMIX (blue). This is a combination of images (a) and (b). Some GAOQMIX-positive cells (blue, arrowed) did not bind GAM42a, BET42a or GAM42_C1033. (d). Laboratory-scale SBR with GAM42_C1033 (red), GAM42a (green) and GAOQMIX (blue). The majority of the GAOQMIX-binding cells also bind GAM42_C1033 (magenta, arrowed), but some GAOQMIX-binding cells also bind GAM42a (cyan) and others bind neither GAM42_C1033 nor GAM42a (blue). Red cells in this image bind GAM42_C1033 only. (e). Non-homogenized sample from laboratory-scale SBR showing the abundance of cells that bind GAM42_C1033 (red) or BET42a (green). There are few of the GAOQMIX cells binding BET42a (cyan) and clusters of GAOQMIX cells that bind GAM42_C1033 (magenta, arrowed). (f). Full-scale sample from Loganholme Wastewater Treatment Plant with GAM42_C1033 (red) GAOQMIX (green) and EUBMIX (blue). Large clusters of bacteria are seen that bind both the GAOQMIX and GAM42_C1033 probes (white). Also in the full-scale sample there are non-GAOQMIX-binding bacterial cells that bind GAM42_C1033 (magenta).
for supplying the laboratory-scale non-EBPR SBR sludge and Tim Beckenham for help with the 23S rDNA sequence alignment. Some of the work was supported by the CRC for Waste Management and Pollution Control Ltd, a centre established and supported under the Australian Government’s Cooperative Research Centres Program. Christine Yeates was supported by a University of Queensland Postdoctoral Research Fellowship, Greg Crocetti was supported by a University of Queensland Postgraduate Research Scholarship, and Greg Crocetti and Aaron Saunders were supported by a CRC-WMPC scholarship.

REFERENCES


