Defining the Bacteroides Ribosomal Binding Site

Udo Wegmann,* Nikki Horn, Nikki R. Carding* a,b
Gut Health and Food Safety Programme, Institute of Food Research, a, and Norwich Medical School, University of East Anglia, b, Norwich Research Park, Norwich, United Kingdom

The human gastrointestinal tract, in particular the colon, hosts a vast number of commensal microorganisms. Representatives of the genus Bacteroides are among the most abundant bacterial species in the human colon. Bacteroidetes diverged from the common line of eubacterial descent before other eubacterial groups. As a result, they employ unique transcription initiation signals and, because of this uniqueness, they require specific genetic tools. Although some tools exist, they are not optimal for studying the roles and functions of these bacteria in the human gastrointestinal tract. Focusing on translation initiation signals in Bacteroides, we created a series of expression vectors allowing for different levels of protein expression in this genus, and we describe the use of pepI from Lactobacillus delbrueckii subsp. lactis as a novel reporter gene for Bacteroides. Furthermore, we report the identification of the 3' end of the 16S rRNA of Bacteroides ovatus and analyze in detail its ribosomal binding site, thus defining a core region necessary for efficient translation, which we have incorporated into the design of our expression vectors. Based on the sequence logo information from the 5' untranslated region of other Bacteroidales ribosomal protein genes, we conclude that our findings are relevant to all members of this order.

The human gastrointestinal tract hosts a plethora of commensal microorganisms, the majority of which are located in the colon, where they reach cell densities of 10^{11} per gram of content (1). It has been shown that these bacteria contain representatives of hundreds of different phylotypes (2, 3). Although the composition of the microbiota of individuals is unique and variable, a dominant phylogenetic core set has been described (4), with members of the phyla Firmicutes and Bacteroidetes constituting up to 90% of the microbiota resident in the colon of all human populations (2, 5).

The intestinal microbiome contributes considerably to the health of the human host by, for example, providing essential vitamins and breaking down and fermenting dietary fiber into short-chain fatty acids, which, besides supplying energy, have a wider physiological impact on the host (6). The microbiome also plays a key role in the development of the immune system (7), and its dysbiosis has been implicated in the development of chronic inflammatory disorders, such as Crohn's disease and ulcerative colitis (8), and may contribute to obesity (9). Among members of the Bacteroidetes phylum, representatives of the genus Bacteroides are some of the most abundant bacterial species in the human colon (10). Consequently, there is considerable interest in studying these organisms that diverged from the common line of eubacterial descent before the major eubacterial groups, including Gram-positive bacteria, and are distinct from the other major Gram-negative phylum, the Proteobacteria (11). Their membranes contain sphingolipids (12), and the structure of their promoters is different from those of proteobacteria (13), which is reflected in the fact that Bacteroides species possess a unique primary sigma factor (14) and which explains the failure to express genes from other Gram-negative bacteria (e.g., Escherichia coli) in this genus (15). Functional analyses of these bacteria require, therefore, tools specific to this order of bacteria.

To address this need, we have focused on translation initiation, the next step in protein expression, and the sequence involved in translation initiation, the ribosomal binding site (RBS) or the Shine-Dalgarno (SD) sequence (16). The available information on Bacteroides Shine-Dalgarno sequences is inconclusive. Mas-tropaolo et al. (17) used the SD sequence (5'-AGAAAGGAG-3') suggested by Tribble et al. (18) in their study of 16S rRNA expression signals. By exchanging transcription and translation initiation signals between E. coli and Bacteroides thetaiotaomicron, they concluded that the RBS sequence and spacing recognition by the ribosome in Bacteroides are more selective than those in E. coli (17). This is seemingly akin to Gram-positive bacteria such as Bacillus subtilis, which have more-stringent criteria for their translation initiation signals and require a higher degree of complementarity to the 3' end of the 16S rRNA (19). Recently, Accetto and Avgustin (20) noted from the genome sequences of several representatives of the phylum Bacteroidetes that there is enrichment in adenine and thymine in the 5' untranslated region (UTR) of genes. Furthermore, they reported an inability of Prevotella bryantii (a member of the order Bacteroidales) to form a functional Shine-Dalgarino interaction, although it was somewhat surprising that the best-performing 5' UTR contained a perfect core SD sequence (5'¬GGAGG-3').

Proline iminopeptidases (EC 3.4.11.5) cleave N-terminal proline residues from peptides, thereby making them more accessible to the enzymatic activities of other aminopeptidases with broader substrate specificity (21). General aminopeptidases cannot cleave the Pro-X peptide bond efficiently due to structural constraints on the peptide chain conferred by the proline residue (22). Here, we report a novel reporter gene, pepI from Lactobacillus delbrueckii subsp. lactis (21), for Bacteroides and describe in detail the Bacteroides ribosomal binding site. Furthermore, we present a set of

Received 9 October 2012 Accepted 9 January 2013
Published ahead of print 18 January 2013
Address correspondence to Udo Wegmann, udo.wegmann@ifr.ac.uk.
U.W. and N.H. contributed equally to this article.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.03086-12.
Copyright © 2013, American Society for Microbiology. All Rights Reserved.
doi:10.1128/AEM.03086-12
expression vectors allowing for high, medium, and low levels of constitutive protein expression.

MATERIALS AND METHODS

Media, growth conditions, and transformations. Escherichia coli strains were grown in Luria-Bertani medium at 37°C. Bacteroides ovatus V975 and derivative strains were grown under anaerobic conditions at 37°C in brain heart infusion (BHI) medium (Oxoid, United Kingdom) supplemented with 0.001% hemin. Lactococcus lactis UKLc10 and derivative strains were grown at 30°C in M17 medium (Oxoid) supplemented with 5 g/liter glucose. Antibiotics, ampicillin at 200 μg/ml and tetracycline, erythromycin, and chloramphenicol at 5 μg/ml, were added as selective agents when appropriate. E. coli JM109 and L. lactis UKLc10 were transformed by electroporation using a Gene Pulser II (Bio-Rad, United Kingdom). The construction of most plasmids described below was performed using E. coli JM109 as the host strain. For constructs relating to pUK2001, the host strain L. lactis UKLc10 was used. The E. coli strain J53/R751 was supplemented with 200 μg/ml trimethoprim when grown for 18 h. Plasmids were mobilized from E. coli strain J53/R751 into B. ovatus V975 using a triparental mobilization filter mating protocol (23).

Calculation of codon adaptation indices. Codon adaptation indices (CAIs) were calculated using the EMBOSS software package (24). First, codon usage tables were created on the basis of all coding sequences (excluding pseudogenes) of the organism using the EMBOSS ccap algorithm, and then the codon adaptation index for each ribosomal gene was computed using the respective codon usage table and the EMBOSS cai algorithm.

Construction of plasmids. (i) Expression vector pFI2716. Initially, a 1,883-bp region carrying a β-d-glucuronidase gene (gusA) was removed from the Bacteroides plasmid pMF1 (25) through restriction digestion with BamHI and Sacl, T4 polymerase treatment, and subsequent religation of the larger fragment to create pFI2708. Next, the primer pair tetQ_5' and tetQ_3' (general primers are listed in Table S1 in the supplemental material) was used to amplify a 1,500-bp region carrying gusA from the Bacteroides plasmid pBT-2 (26). The tetQ fragment was cloned into the blunt EcorI site of pFI2708 to create pFI2714. The P1 promoter from the 16S rRNA gene of Bacteroides thetaiotaomicron VPI 5482 was cloned using the primer pair BAC_P1_f and BAC_P1_r to generate a 228-bp PCR product that was digested with BamHI and PstI and ligated into the same sites in pFI2714 to create pFI2715. Finally, the primer pair BglII_PUK200-linker_5' and BglII_PUK200-linker_3' was used to amplify a 239-bp polylinker region from plasmid pUK200 (27) comprising an SD sequence, unique cloning sites for NcoI (to be used for translational linker fragment was digested with BglII and cloned into the BamHI site of pFI2715 to create pFI2716.

(ii) Cloning of the pepI reporter gene. An 882-bp fragment was amplified from pUK2001 plasmid DNA (27) using the primer pair pepL_5' and pepL_3'. The pepL fragment was cloned into the Ncol (blunted) and Smal sites of pFI2716 to create pGH001. In pGH001, the distance between the SD sequence (5'-AAGGAGG-3') and the ATG start codon is 8 nucleotides. To move the SD sequence closer to the ATG start codon, the spacing was adjusted by either 2 or 6 nucleotides. To this end, primers phoshorylated at the 5' end were employed to amplify the 8,876-bp pGH001 plasmid. The primer pair RBS-F and RBS_R_dis_6 was used to generate an 8,874-bp product (∆2nt), and RBS-F and RBS_R_dis_2 were used to generate an 8,870-bp product (∆6nt); each was then religated, creating plasmids pGH004 and pGH006, respectively.

For the generation of splice overlap extension PCR products, the details of the primers and templates are listed in Table S2 in the supplemental material. In each case, two PCR products (A and B) were amplified, and the amplified fragments from each of these reactions were then used as templates for the final splice reaction involving the forward primer from reaction A and the reverse primer from reaction B. (iii) Cloning of the 5' UTR upstream of orf1. To exchange the SD sequence in pFI2716 with the 5' UTR of orf1 in the B. ovatus V975 xylose utilization operon, splice overlap extension PCR was employed. A 365-bp Splhl- and Ncol-digested fragment of the splice PCR product was then used to replace the corresponding 257-bp fragment in the SD sequence-related Bacteroides expression vector pFI2716, creating pGH090. To replace the upstream sequences between the transcription initiation site (TIS) and the ATG start codon in pGH090 to either 20 or 30 nucleotides (nt), splice overlap extension PCR was used. A 169-bp or 179-bp Splhl- and Ncol-digested fragment of the splice PCR products was then used to replace the corresponding 257-bp fragment of pFI2716, creating pGH020 and pGH021, respectively.

(iv) Removal of RNase E site from pFI2716. To remove a region containing the RNase E site upstream of the SD sequence in pFI2716, splice overlap extension PCR was employed. A 329-bp Splhl- and Ncol-digested fragment of the splice PCR product was used to replace the corresponding 257-bp fragment of pGH020 to create pGH078.

(v) Replacing tetQ with ermF. The primer pair ccr_amont2 and ccr_a-val2 was used to amplify a 1,500-bp region carrying ermF from the plasmid pDF15 (28). The ermF fragment was digested with Ndel and cloned into Ndel-digested (blunted) and Nsil-digested pFI2716, pGH020, and pGH021 to replace the existing 1,839-bp fragment of each, creating pGH043, pGH044, and pGH090, respectively.

(vi) Altered 5' UTR lengths. To reduce the length of the 5' UTR in pGH024 to 24 nt, splice overlap extension PCR was used. A 333-bp Splhl- and Ncol-digested fragment of the splice PCR product was then used to replace the corresponding 329-bp fragment of pGH024 to create pGH079. To remove the entire 5' UTR between the TIS and the start codon or to reduce the length of the 5' UTR in pGH022 to 10, 16, or 18 nt, splice overlap extension PCR was employed. A 309-bp (0-nt 5' UTR), 319-bp (10-nt 5' UTR), 325-bp (16-nt 5' UTR), or 327-bp (18-nt 5' UTR) Splhl- and Ncol-digested fragment of each splice PCR product was then used to replace the corresponding 329-bp fragment of pGH024 to create pGH039, pGH040, pGH041, and pGH042, respectively.

(vii) Exchanging individual bases within the 5' UTR. Splice overlap extension PCR was used to exchange bases in pGH022 at positions –11 (replacing a guanosine with a cytidine), –10 (replacing an adenosine with a cytidine), and –9 (replacing an adenosine with a cytidine). A 329-bp Splhl- and Ncol-digested fragment of each splice PCR product was then used to replace the corresponding 329-bp fragment of pGH024 to create pGH045, pGH046, and pGH047, respectively.

To move the RBS closer to the ATG start codon in pGH022, the spacing was adjusted by either 2 or 4 nucleotides using splice overlap extension PCR. A 327-bp (∆2nt) or 325-bp (∆4nt) Splhl- and Ncol-digested fragment of each splice PCR product was then used to replace the corresponding 329-bp fragment of pGH024 to create pGH048 and pGH049, respectively.

(viii) Comparing proline iminopeptidase I expression levels in a Gram-positive host. The 20-nucleotide lactococcal SD sequence upstream of pepl in pUK2001 was exchanged for the 20 nt of the 5' UTR from pGH022 using splice overlap extension PCR. A 362-bp BglII- and Ncol-digested fragment of the splice PCR product was then used to replace the corresponding restriction fragment of pUK2001 to create pGH068.

(ix) Cloning of the gusA reporter gene. A 1,806-bp fragment was amplified using the primer pair GUS_5' and GUS_3', with pBH101 plasmid DNA (Clontech/TaKaRa Bio Europe, France) as the template. The resulting PCR fragment was then cloned into Ncol-digested (blunted) and SmaI-digested pFI2716 to create pGH075. The SmaI region present in pGH075 was replaced with the 5' UTR from pGH020 by way of splice overlap extension PCR. A 555-bp Splhl- and SnaBl-digested fragment of the splice PCR product was then used to replace the corresponding 643-bp fragment of pGH075 to create pGH076.
The insert regions for all plasmids constructed as described above were sequenced in full using the BigDye Terminator v3.1 cycle sequencing kit (Life Technologies Ltd., United Kingdom). Table S3 in the supplementary material summarizes all plasmid constructs.

Preparation of cell extracts. A 20-ml culture of B. ovatus or E. coli was grown to an optical density at 600 nm (OD$_{600}$) of 1.0. Cells from two independent cultures were harvested by centrifugation at 10,000 $\times g$ for 25 min at 4°C. Cell pellets were washed once in 0.2 M Tris–HCl (pH 7.2) before freezing at −20°C. Each thawed cell pellet was resuspended in 250 μl of 0.2 M Tris–HCl (pH 7.2), and cell disruption via sonication was performed using eight 10-s pulses (amplitude, 6 micrometers), with 30-s pauses on ice between each pulse. Cell extracts were obtained in the form of supernatants after centrifugation at 14,000 $\times g$ for 30 min at 4°C. For L. lactis, cultures were grown to an OD$_{600}$ of 0.5 before induction with 1 ng/ml nisin A (Aplin & Barret Ltd., Beaminster, United Kingdom) in order to switch on the mva promoter. Following a 2-h induction period, cells from two independent cultures were collected by centrifugation at 10,000 $\times g$ for 15 min at 4°C using sample volumes equivalent to 37 OD$_{600}$ units. Cell extracts were prepared principally as described by Wegmann et al. (27).

Enzyme assays and protein determination. Peptidase I activity was assayed using 50 μl of cell-free samples or dilutions thereof in 0.2 M Tris–HCl (pH 7.2) and 0.1% bovine serum albumin fraction V (Sigma, United Kingdom), with 0.7 mM L-proline p-nitroanilide trifluoroacetate salt (Sigma) as the substrate. The release of p-nitrophenol was measured at 405 nm at 32°C. β-Glucuronidase (GUS) activity was assayed using 25 μl of cell-free samples or dilutions thereof in 50 mM Na$_2$HPO$_4$ (pH 7.0), 10 mM β-mercaptoethanol, 1 mM EDTA, and 0.1% Triton X-100, with 1.25 mM 4-nitrophenol-β-D-glucuronide (PNPG) as the substrate. The release of p-nitrophenol was measured at A$_{405}$ at 37°C. The protein concentrations of samples were measured at A$_{280}$ using the Bio-Rad protein assay (Bio-Rad, United Kingdom) with bovine serum albumin (New England BioLabs UK Ltd.) as the standard. Specific enzyme activity is expressed as nanomoles of nitrophenol released from the chromogenic substrate per milligram of protein per minute. The means and standard deviations presented are based upon two biological replicates with three technical replicates each.

RNA purification. To isolate RNA for use in quantitative reverse transcription–PCR (qRT-PCR) experiments, B. ovatus cells were grown to an OD$_{600}$ of 1.0. Subsequently, two independent 1-ml culture samples were removed and added to 200 μl of phenol/ethanol (5:95), and the cell pellet was collected by centrifugation at 16,000 $\times g$ for 30 min at 4°C, after a 30-min incubation period on ice. RNA purification was performed using the SV total RNA isolation system (Promega UK Ltd.) and following the manufacturer’s instructions, with an additional DNase treatment step using a Turbo DNA-free kit (Life Technologies Ltd., United Kingdom). The RNA was subsequently chloroformed, a ThermoScript RT-PCR system (Life Technologies Ltd., Paisley, United Kingdom) was used according to the manufacturer’s recommendation to reverse transcribe the polyadenylated RNA and subsequently amplify the 3′ ends of the cDNA product using the primers 5′-GCAATGTTGTGCAGCTCG-3′ and 5′-(TTTTT)$_x$-3′. The resulting amplicons were gel purified on a 5% acrylamide gel, electroeluted, and subsequently used as the template for a second round of PCR using the primers 5′-ACCCGGAGGTCGCCGTCG-3′ and 5′-(TTTTT)$_x$-3′. Following gel purification, the fragments were cloned into pJE1.2 (Fermentas, St. Leon-Rot, Germany) and subsequently sequenced using pJE forward and reverse sequencing primers.

RESULTS

Comparative analysis of the 3′ ends of 16S rRNAs. To identify optimal SD sequences for our expression vectors, we used the Bacteroidales genome sequences available in public databases. Figure 1A shows alignment of the 3′ ends of the 16S rRNA gene sequences extracted from the genome sequences of Bacteroides thetaiotaomicron VPI 5482 (29), Bacteroides vulgatus ATCC 8442 (30), Bacteroides ovatus V975 (U. Wegmann, unpublished data), Bacteroides fragilis NCTC 9543 (30), Prevotella denticola F0289 (CP002589), and Prevotella bryantii B14 (31) with the sequenced 3′ ends of 16S rRNA sequences from Escherichia coli (32, 33), Bacillus subtilis (34), Bacillus steaerotherophilus (35) (now Geobacillus steaerotherophilus), and Lactococcus lactis (36), which demonstrates the perfect sequence conservation with regard to the 3′ ends of the 16S rRNAs. In the cases of B. thetaiotaomicron, B. vulgatus, P. denticola, and P. bryantii, the 16S rRNA genes annotated in the respective genomes were extended with additional downstream sequence data on the basis that all RNA gene operons present in the genome were followed by the same stretch of sequence that would be transcribed into the same 3′ ends as those reported for Gram-positive 16S rRNA sequences.

Analysis of 5′ UTRs of ribosomal protein genes. Subsequently, we extracted the nucleotide sequences of genes encoding ribosomal proteins from positions −30 to +5 (the first base of the start codon is position 0), from the Bacteroides genome sequences mentioned above, and the genome sequences of Bacillus subtilis (37), Escherichia coli (38), and Lactococcus lactis (39). Based on codon bias, the ribosomal proteins of most bacteria are predicted to be highly expressed and, aside from optimized codon usage, this is achieved through the presence of an SD sequence that has higher affinity for the 3′ end of the 16S rRNA and occurs at closer-to-optimal spacing than the SD sequences of predicted moderately
expressed genes (40). First, we inspected the *Bacteroidales* 5′ UTR with respect to the presence of Shine-Dalgarno-like sequences, the results of which are summarized in Table 1. In contrast to the well-documented prevalence of these sequences in the 5′ UTRs of ribosomal genes from many bacterial species, which often reaches 80% (40), the respective 5′ UTRs of the examined *Bacteroidales* genes contained Shine-Dalgarno-like sequences to a far lesser extent (18 to 28%). Interestingly, in comparison to the group of genes that lack such sequences, this group of genes contained an equal or greater proportion of members with a codon adaptation index (CAI) above the average calculated from the CAIs of all ribosomal genes of the organism in question, which would indicate high expression levels, also requiring efficient translation initiation.

Due to the small number of canonical SD sequences in the 5′ UTRs of *Bacteroidales* ribosomal genes, WebLogo 3.3 (41) was used to identify the optimal ribosomal binding site consensus sequence.

![Table 1](https://example.com/table1.png)

**Table 1** Distribution of Shine-Dalgarno sequences in *Bacteroides* ribosomal protein genes

<table>
<thead>
<tr>
<th>Species</th>
<th>No. of ribosomal protein genes</th>
<th>% of ribosomal protein genes with an SD sequence in the 5′ UTR</th>
<th>% of genes with a CAI above the mean</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Bacteroides ovatus</em></td>
<td>45</td>
<td>20.0</td>
<td>80</td>
</tr>
<tr>
<td><em>Bacteroides thetaiotaomicron</em></td>
<td>33</td>
<td>18.8</td>
<td>80</td>
</tr>
<tr>
<td><em>Bacteroides fragilis</em></td>
<td>52</td>
<td>18.8</td>
<td>50</td>
</tr>
<tr>
<td><em>Bacteroides vulgatus</em></td>
<td>51</td>
<td>17.6</td>
<td>50</td>
</tr>
<tr>
<td><em>Prevotella denticola</em></td>
<td>53</td>
<td>28.3</td>
<td>47</td>
</tr>
<tr>
<td><em>Prevotella bryantii</em></td>
<td>48</td>
<td>22.9</td>
<td>50</td>
</tr>
</tbody>
</table>

* SD sequences are GGAGG, GGAG, and AGG. The 5′ UTR comprises positions −1 to −20.
sequence for the respective organism by creating sequence logos of these regions without any prior alignment. The resulting sequence logos corresponded well with the 3’ ends predicted from the genome sequence of the organism as well as from the total lack of any PepI activity in the vector-only control. They fall into four groups. Group 1 is based on the Gram-positive bacterial 16S rRNA through 3’ ends predicted from their genome sequences (Fig. 1B). Instead, they are enriched in adenosine and, to a far lesser extent, thymidine. Positions 12 to 18 with respect to the start codon stand out because of the strong conservation of adenosine and thymidine peaks in this region. Furthermore, there is a clear dip in the sequence conservation leading up to this region, albeit at position 8 for B. vulgatus instead of position 10/11 for the other Bacteroides species. There are other positions with less striking conservation, for example, positions –1, –3, and –7, and there is a conspicuous stretch of adenosine and thymidine enrichment from positions –19 to –28, which might be important.

Determining the 3’ ends of the 16S rRNA. Based on their unique sequence logos, we considered the possibility that the 3’ ends of the Bacteroides 16S rRNAs might be different from their counterparts in other eubacteria due to, for example, different RNase processing or posttranscriptional modification. We decided therefore to determine the 3’ end of the B. ovatus 16S rRNA through 3’ rapid amplification of cDNA ends (RACE). The corresponding sequencing results confirmed the 3’ ends predicted from the genome sequence of the organism as 5’-ACCTCCCTTCT-3’, which is identical to the sequences determined for B. subtilis and L. lactis (Fig. 1).

Proline iminopeptidase I reporter and expression vectors for Bacteroides. We constructed a series of different expression vectors using pepI as a novel marker gene to define the Bacteroides-specific translation initiation signal in detail, to create high- and low-level protein expression vectors for Bacteroides, and to demonstrate the dynamic range of the pepI reporter. Figure 2 shows a schematic representation of the expression vectors generated. They fall into four groups. Group 1 is based on the Gram-positive expression cassette from the lactococcal expression vector pUK200 (27), allowing for translational fusion of the gene of interest with the strong SD sequence from the lactococcal misA gene (43) under the control of the P1 promoter from the 16S rRNA gene of B. thetaiotaomicron. Group 2 is identical to group 1 apart from the region between the transcription initiation site and the start codon, which carries various derivatives of the 5’ UTR of pepI from the xylanase operon of B. ovatus V975. Alignment of the orf1 5’ UTR with the B. ovatus ribosomal protein gene 5’ UTR logo is presented in Fig. S2 in the supplemental material. Groups 3 and 4 represent vectors from groups 1 and 2, respectively, carrying the gusA reporter gene instead of pepI (i.e., pGH075);*, pGH001 carrying the gusA reporter gene instead of pepI (i.e., pGH076).

![Figure 2](https://example.com/figure2.png)

**FIG 2** Schematic representation of the different expression vectors with the region stretching from the promoter to the start codon expanded to the nucleotide level. Tbrnq represents the terminator of the Lactobacillus delbrueckii subsp. lactis brnQ gene. Positions –33 and –7 and the transcription initiation site (position +1) of the P1 promoter (17) are in bold capital letters, ribosomal binding sites are underlined and in capital letters, and nucleotide changes are underlined and in bold lowercase letters. The 3’ end of the 16S rRNA (3’-UCUUUCUCCA-5’) is shown for comparison. Groups 3 and 4 represent vectors from groups 1 and 2, respectively; *, pGH001 carrying the gusA reporter gene instead of pepI (i.e., pGH075);†, pGH022 carrying the gusA reporter gene instead of pepI (i.e., pGH076).
the xylanase operon \textit{orf}1 5’ UTR (positions −1 to −20) for the \textit{pepI} gene (pGH022) is high enough to visualize the corresponding protein by SDS-polyacrylamide gel electrophoresis (data not shown), something that, as far as we are aware, has not been achieved in \textit{Bacteroides} with any other protein and expression system.

**Transcript levels.** Although all constructs carry the P1 promoter from the 16S rRNA gene of \textit{B. thetaiotaomicron}, there is scope for the differences in transcript levels produced by these constructs due to the sequence variations in their 5’ UTRs. Therefore, we determined transcript levels through quantitative PCR, the results of which are shown in Fig. 3B. Most constructs produced similar levels of transcript, with the noticeable exceptions of pGH001, pGH004, and pGH006. Plasmid pGH006 showed the lowest transcript level and served as the calibrator in our qRT-PCR experiments. We identified a potential RNase E site at the 5’ end of the transcripts of these 3 vectors, explaining the stark difference in mRNA levels. Removal of this part of the sequence in plasmid pGH078 stabilized the respective mRNA (Fig. 3B). In comparison to pGH022, significantly lower levels of transcript were also determined for pGH039 and pGH040 (two-sample two-tailed \( t \) test, \( P < 0.05 \)).

**Defining the \textit{Bacteroides} ribosomal binding site.** To define the \textit{Bacteroides} ribosomal binding site, we needed to look at the translation initiation efficiency represented by PepI activity while minimizing the effects of different transcript levels. To this end, PepI expression data were normalized with respect to the qRT-PCR results (Fig. 3C). The normalized PepI expression levels, representing the translation initiation efficiency, were 198 nmol mg\(^{-1}\) min\(^{-1}\) with the lactococcal SD sequence (pGH001) and decreased markedly as the distances between the start codon and the SD sequence were progressively reduced (see pGH004 and pGH006 in Fig. 3C). The shortening of the 5’ UTR to 22 nt in pGH078 resulted in a 45% increase in activity, and exchanging the core SD sequence-containing region for the 5’ UTR of \textit{orf}1 from the xylanase operon of \textit{B. ovatus} V975 (pGH10) led to a slight decrease in PepI activity to 232 nmol mg\(^{-1}\) min\(^{-1}\). This was increased 3.2-fold when the stretch of nucleotides between positions −2 (transcription) and −30 (translation) was deleted (pGH023), which, with a normalized PepI activity level of 752 nmol mg\(^{-1}\) min\(^{-1}\), represents the most efficiently initiating RBS in the series. The progressive shortening of the RBS region to positions −24 (pGH079), −20 (pGH022), −18 (pGH042), and −16 (pGH041) resulted in successive decreases in PepI activity levels to 74, 67, 66, and 61%,
respectively, of the peak activity. Deletion to position −10 (pGH040) reduced the peptidase activity level to 25% of the peak activity obtained with pGH023, but the activity still was twice that of the negative control (pGH039), which had only two cytidine residues in the pepI 5′ UTR. A two-sample two-tailed t test analysis was performed on the above-mentioned comparisons and, apart from the comparison of pGH006 with pGH004, all the differences were significant (P < 0.05). To define the RBS further, nucleotide exchanges were made at positions −11, −10, and −9, none of which had a substantial effect on reporter gene expression levels in comparison to the levels obtained with the equivalent plasmid pGH022. With the core RBS defined as ranging from positions −12 to −18, the spacing between the RBS and the start codon in construct pGH022 was changed. Moving the RBS closer by 2 nt (pGH048) reduced expression levels by one-third (P < 0.05), and a move of 4 nt (pGH049) resulted in the halving of expression levels (P < 0.05), demonstrating the importance of the spacing between the Bacteroides ribosomal binding site and the start codon.

To ensure that our results were not dependent on this particular reporter gene, we constructed gusA (uidA) equivalents for some of the key pepI-carrying constructs (pGH075* and pGH076 in Fig. 2). In general, the specific enzyme activity observed with these constructs was more than 1 order of magnitude lower than that seen with the pepI reporter (Fig. 3D), demonstrating the superior performance of the pepI reporter gene. With regard to the expression levels obtained with the classical SD sequence or the Bacteroides ribosomal binding site, the findings with the pepI reporter gene were confirmed, in that employment of the latter was far more efficient and resulted in 23-fold-higher specific GusA activity.

Use of Bacteroides ribosomal binding site in other bacteria.

In light of the high expression levels achieved with the RBS originating from the B. ovatus xylanase operon orf1, we determined how it would perform in bacteria that normally employ the SD sequence for translation initiation. Two organisms were investigated: E. coli, which has less stringent SD sequence requirements and employs the ribosomal protein S1 to aid translation initiation (44), and Gram-positive L. lactis, which has more stringent requirements. The L. lactis cultures were induced with nisin. In E. coli, some expression might be derived from the Bacteroides P1 promoter, although previous research has shown that this promoter does not work very well in E. coli (17). We analyzed the upstream region of the pepI gene of the constructs in question with the SoftBerry promoter prediction program (http://linux1.softberry.com/berry.phtml), which returned 2 plausible E. coli promoters that, in our opinion, are the more likely sources of gene expression. As expected, the classical SD sequence (5′-GGAGG-3′) present in pGH078 and the lactococcal expression vector pUK200 performed best in both of these bacteria, but interestingly, the ribosomal binding site originating from Bacteroides delivered 19% of the PepI activity achieved with the SD sequence in E. coli (pGH022 in Fig. 3E) and the respective value was 30% in the case of L. lactis (pGH68 in Fig. 3F).

DISCUSSION

We have analyzed the upstream regions of ribosomal genes in Bacteroides to identify optimized ribosomal binding sites for the construction of expression vectors using a sequence logo approach. Our results, though more detailed, are in keeping with the findings by Accetto and Avgustin (20), who used the total gene contents of the organisms for their logos, containing alien as well as highly and moderately expressed genes. This explains why our logos are more extensive and detailed, as ribosomal genes are predicted to be among the most highly expressed genes and show high levels of occurrence of strong SD sequences (40), therefore providing a better starting point for identifying strong ribosomal binding sites through a sequence logo approach.

Some authors have suggested that translation initiation in the Bacteroidales phylum is facilitated solely through ribosomal protein S1 (RPS1) (20, 45). It has been demonstrated that translational enhancers derived from the tobacco mosaic virus, whose sequence fits the binding criteria for RPS1 (no guanosine) (46), can replace the SD sequence in a variety of Gram-negative bacteria (47). However, the ability of RPS1 to initiate translation by itself has not been demonstrated (48).

In examining regions upstream of the ribosomal genes of B. thetaiotaomicron VPI 5482, B. ovatus V975, B. fragilis NCTC 9343, B. vulgatus ATCC 8482, P. bryantii B14, and P. denticola F0289, we identified the core SD sequence in 31 of a total of 309 genes from position −1 to position −20 but not at all from position −21 to position −40. Interestingly, the presence in front of certain genes is preserved (see Table S4 in the supplemental material), for example, the genes for ribosomal proteins S18 and L6 in all 6 strains containing the core SD sequence upstream of the respective start codon. In contrast, the occurrence of 5′-GGAGG-3′ is evenly distributed between these two regions with regard to the total gene content and occurs at a rate of 2.8% in region −1 to −40 (B. ovatus not included). Strengthening the notion that these particular ribosomal genes are highly expressed is the fact that in the five members of Bacteroidales examined, an equal or greater proportion of the genes (Table 1) that possess the core SD sequence or remnants thereof in the position −1 to −20 region also display a codon adaptation index that is above the average for the ribosomal genes of the bacterial species in question. Together with the results for the reporter gene expression plasmid pGH078, these findings provide clear evidence that the core SD sequence enables efficient translation initiation in Bacteroides, although at rates somewhat lower than those for the Bacteroides ribosomal binding site defined here. It is clear from our data that B. ovatus is able to initiate translation in the absence of the SD sequence or the ribosomal binding site defined here (pGH040 in Fig. 3). Indeed, the presence of 2 nt in front of the ATG start codon (pGH039 in Fig. 3) is sufficient. This so-called leaderless translation initiation is not uncommon in bacteria (49).

The mRNA produced from constructs pGH001, pGH004, and pGH006 is very labile, most probably due to an RNase E site near the 5′ end of the molecule. Normalization of the PepI expression data with respect to the corresponding transcript data reveals that spacing is a factor affecting translation initiation efficiency for the canonical SD sequence, as described for other bacteria (50–52). A spacing effect can also be seen with regard to the Bacteroides ribosomal binding site (comparing pGH022 and pGH049). Assuming that RPS1 binds to this sequence, this is somewhat different from the situation for the translation-enhancing effect that was described by Boni et al. for RPS1 (46), where the U-rich binding sites are described as a “drifting” element of the primary mRNA structure. In a review by Kozak (53), the minimization of secondary structure in the 5′ UTR is put forward as an alternative explanation for the enhancing effects of these A/U-rich sequences. With
this in mind, we compared the results of expression vector pairs that differed only in the extent of their 5’ UTRs (pGH001 with pGH078 and pGH100 with pGH023). In both cases, the efficiency of translation initiation is negatively affected by the presence of an extended 5’ UTR, leading to losses of 31 and 69% of PepI activity, respectively (Fig. 3). Both 5’ UTRs contain secondary structure elements as determined with mFold (54), with combined ΔG values of −20 and −38 kcal mol−1, respectively. Whether these predicted secondary structure elements or the sheer presence of a longer 5’ UTR is responsible for the knockdown in translation initiation cannot be elucidated from our data.

Gram-positive bacteria such as *B. subtilis* and *L. lactis* were thought to lack RPS1 (55–57). It is now known that *B. subtilis* has a counterpart, YpfD (58), that is not involved in translation. *L. lactis* also carries an *rpsA* gene on its chromosome (59), although Salah et al. showed that the corresponding gene product lacks the typical ribosome binding domain (domain 1), supporting the idea that the protein cannot bind to the ribosome (60). This protein is therefore unlikely to be involved in translation initiation in *L. lactis*. Salah et al. (60) also identified the full complement of domains in the *B. fragilis* RPS1 protein supporting its role in translation initiation. Together, these findings make it difficult to see how translation initiation at the *Bacteroides* RBS in *L. lactis* might also be achieved through the *rpsA* gene product, especially as the expression level achieved amounts to 30% of the level seen with the strong SD sequence of the *nisA* gene. This raises the questions of how translation initiation is achieved in *L. lactis* under these circumstances and whether RPS1 is the sole entry point for mRNAs of this type in *Bacteroidetes* unless RPS1 plays a role in *L. lactis* or there are two distinct systems present in these organisms.

Reporter genes are useful tools for studying the expression of genes of interest in response to extrinsic or intrinsic factors. Ideally, the gene products are easily detectable, and the measured activity is absent in the organism of interest to avoid background problems at low levels of expression. A number of reporter genes have been described for *Bacteroides*, although they all have limitations and suffer from high costs and complexity (chloramphenicol acetyltransferase assay), high background levels under certain conditions or in certain organisms (β-glucuronidase and xyllosidase) (25,61), relatively low levels of enzyme activity, or the need for specialized equipment to detect fluorescence or luminescence. The pepI reporter gene described here offers a simpler, more sensitive, and cheaper alternative. Measured PepI activities ranging from 102 to 104 nmol mg−1 min−1 demonstrate a greater dynamic range than the *E. coli gusA* gene (10 to 102 nmol mg−1 min−1) or the *B. ovatus xa* gene (38 to 75 nmol mg−1 min−1), measured in *Bacteroides uniformis* and *B. thetaiotaomicron*, respectively (61).

Also, due to the absence of any significant constitutive PepI activity in *B. thetaiotaomicron*, *B. ovatus*, and *B. fragilis*, it is possible to measure low levels of expression in these organisms. The relatively small size of the pepI gene and its product (294 amino acids [aa]), in comparison with the reporter gene products GusA (602 aa) and LuxAB (683 aa), puts a smaller burden on the host cells, which can be a problem if strong promoters are examined, and is an advantage in terms of plasmid stability.

In summary, this work gives for the first time a detailed view of the unique *Bacteroides* ribosomal binding site. While we do not demonstrate direct binding to the ribosome, our findings are supported by extensive expression data. Based on the respective sequence logos (see Fig. S1 in the supplemental material), we predict that our findings apply to a great number of, if not all, members of the *Bacteroidetes* phylum. Furthermore, we show that canonical SD sequences initiate translation in *B. ovatus* and most probably in other *Bacteroidales* species as well. We have generated a set of expression vectors that allow the expression of a gene of interest at appropriate levels in these organisms. They are presented in Fig. 4 and offer different combinations of expression levels and antibiotic resistance genes.

**ACKNOWLEDGMENTS**

This work was supported in part by the Biotechnology and Biological Sciences Research Council (BBSRC) in support of the Gut Health and Food Safety research program (grant BB/J004529/1). We thank Bernhard Henrich (University of Kaiserslautern, Germany) for providing plasmid pUK200 and the pepI gene from *Lactobacillus delbrueckii* subsp. *lactis*. We are grateful to Terry Whitehead for the gift of the plasmid pMJF1.

**REFERENCES**


