JOURNAL OF BACTERIOLOGY, June 1999, p. 3857–3859 Vol. 181, No. 12
0021-9193/99/$04.00+0

Fused and Overlapping *rpoB* and *rpoC* Genes in Helicobacters, Campylobacters, and Related Bacteria

NATALYA ZAKHAROVA,[1,2] BRUCE J. PASTER,[3] IRENE WESLEY,[4] FLOYD E. DEWHIRST,[3] DOUGLAS E. BERG,[5,6] AND KONSTANTIN V. SEVERINOV[1,2]*

Waksman Institute[1] and Department of Genetics, Rutgers, The State University of New Jersey[2], Piscataway, New Jersey 08854; Forsyth Dental Center, Boston, Massachusetts 02115[3]; National Animal Disease Center, USDA-ARS, Ames, Iowa 50010[4]; and Departments of Molecular Microbiology[5] and Genetics,[6] Washington University School of Medicine, St. Louis, Missouri 63110

Received 3 March 1999/Accepted 14 April 1999

The genes coding for the β (*rpoB*) and β′ (*rpoC*) subunits of RNA polymerase are fused in the gastric pathogen *Helicobacter pylori* but separate in other taxonomic groups. To better understand how the unique fused structure evolved, we determined DNA sequences at and around the *rpoB-rpoC* junction in 10 gastric and nongastric species of *Helicobacter* and in members of the related genera *Wolinella*, *Arcobacter*, *Sulfurospirillum*, and *Campylobacter*. We found the fusion to be specific to *Helicobacter* and *Wolinella* genera; *rpoB* and *rpoC* overlap in the other genera. The fusion may have arisen by a frameshift mutation at the site of *rpoB* and *rpoC* overlap. Loss of good Shine-Dalgarno sequences might then have fixed the fusion in the *Helicobacteraceae*, even if fusion itself did not confer a selective advantage.

A most unexpected natural fusion of *rpoB* and *rpoC*, the genes coding for the β and β′ subunits of DNA-dependent RNA polymerase (RNAP), respectively, was discovered while sequencing the genome of *Helicobacter pylori* 26695 (7), an ε-group proteobacterium that is the primary cause of peptic ulcer disease and an early risk factor for gastric cancer (4). We found that this extraordinary *rpoB-rpoC* fusion is typical of *H. pylori* as a species and also of one other gastric *Helicobacter* tested (*H. felis*) and that it results in a stable fused β-β′ subunit of RNAP (8). In contrast, *Campylobacter jejuni* and *Campylobacter fetus*, species related to *H. pylori* but which colonize intestinal and not gastric sites and often cause diarrheal disease, have separate *rpoB* and *rpoC* genes (8), as do all other eubacterial species studied to date.

If the *rpoB-rpoC* fusion were a characteristic and specific feature of all gastric helicobacters, it might contribute to the special ability of these bacteria to colonize their unique gastric niche. For example, one can speculate that the tethered structure of RNAP β and β′ is useful for *H. pylori* and other gastric helicobacters in facilitating the multisubunit RNAP assembly in the hostile, urea-rich or low-pH gastric environment. A simple prediction of such a model is that *rpoB* and *rpoC* genes might be separate in helicobacters that colonize nongastric sites. To test this prediction and to better understand how this unusual gene structure evolved, we studied the distribution of translational fusion of *rpoB* and *rpoC*.

Our collection of ε-group proteobacteria included 10 species of *Helicobacter*. Two of these species colonize gastric sites, and the rest colonize intestinal sites. In addition, three species of *Arcobacter*, which are significant animal and occasional human pathogens (1); *Wolinella succinogenes*; *Campylobacter rectus* (formerly known as *Wolinella recta* [5]); and one species of *Sulfurospirillum*, *S. barnesii*, all of which are *Helicobacter* related, were included in the analysis. Two primers that target sequences flanking the *rpoB-rpoC* junction and that are complementary to highly conserved sequences in the *H. pylori* 26695 *rpoB-rpoC* gene (8) were used for PCR amplification with genomic DNA from these organisms. In all cases, a single major PCR fragment ca. 500 bp in length was amplified. Each fragment was cloned and sequenced, and the DNA sequences were used to derive amino acid sequences of the RNAP subunit(s). The translational fusion of *rpoB* and *rpoC* was maintained in all helicobacters, gastric and nongastric, as well as in *W. succinogenes*. Based on our previous results with *H. pylori* (6, 8), we conclude that these organisms use natural fusion polypeptide as the sole source of the largest RNAP subunits, β and β′. In contrast, equivalent DNA sequence analysis showed that the *rpoB* and *rpoC* reading frames are separate in all three *Arcobacter* species, as well as in *C. rectus* and *Sulfurospirillum*. Hence, the β and β′ subunits are separate in these organisms. The likely position of an ATG codon coding for the β′ subunit Met^1 was inferred based on sequence comparisons with β′ sequences from other organisms and on the presence of an appropriately spaced A and G rich sequence that could serve as a ribosome-binding site. The analysis suggests that in arcobacters and in *C. rectus*, the appropriate ATG codon is found overlapping with two last codons of the *rpoB* gene, as is also the case in *C. jejuni* and *C. fetus* (8). It is worth noting that the inferred initiating ATG is found in the same reading frame, −1, relative to the *rpoB* reading frame in all these organisms with separate *rpoB* and *rpoC* genes (Fig. 1). In *Sulfurospirillum*, the *rpoB-rpoC* gene overlap is more extensive, as the two genes overlap by 7 codons (Fig. 1).

The resultant collection of deduced amino acid sequences was aligned by using the Clustal method, and the phylogenetic tree shown in Fig. 2A was built with the DNASTAR program. The previously determined sequences of the *rpoB-rpoC* junction of *C. jejuni*, *C. fetus*, *H. felis* (8), and *H. pylori* (7) were also included in this analysis. As can be seen, the phylogenetic tree reveals two major clusters: members of the family *Helicobacteraceae* (the helicobacters and *W. succinogenes*) and members of the family *Campylobacteraceae* (arcobacters, campylobacters, and *S. barnesii*). The campylobacters and arcobacters form their own coherent groups within their cluster, with *Sulfurospirillum* just "outside" the campylobacters. Helicobacters also

* Corresponding author. Mailing address: Waksman Institute, 190 Frelinghuysen Rd., Piscataway, NJ 08854. Phone: (732) 445-6095. Fax: (732) 445-5735. E-mail: severik@waksman.rutgers.edu.

```
H. n.  GCT TTG GAT ATT AAT ATT TTT GGG GAC GAT GTG GAT GAG GAT GGA GCG CCT AAA CCC ATT GTC ATT AAA GAA GAT GAC AGG
       ala leu asp ile asn ile phe gly asp asp val asp glu asp gly ala pro lys pro ile val ile lys glu asp asp arg

H. c.  GCA CTT GAT GTG AAT ATT TAT GGC GAA GAA GTT GAT GAA AAT GGA ATG CCT GTG CCT ATT ACC ATT AAA GAA GAT GAT CGT
       ala leu asp val asn ile tyr gly glu glu val asp glu asn gly met pro val pro ile thr ile lys glu asp asp arg

W. s.  GCG TTG GAT GTC ACC GTC TAT GGC GAG ACC GAA GAG GAT TCT TTT GTT CCT ATG CCC ATC AAA GAA GAC GAT CGA CCC TCT
       ala leu asp val thr val tyr gly glu thr glu glu asp ser phe val pro met pro ile lys glu asp asp arg pro ser

C. r.  GCA CTA GAT GTC GAG ATA TAC GAT GAG GAT GAA AAT AAT GAG TGA
       ala leu asp val glu ile tyr asp glu asp glu asn asn glu OPA
                                  AGG ATG AAA ATA ATG AGT GAG TTA AAA CCT ATT GAG ATA AAA GAA GAA CGC AGA CCG
                                              met ser glu leu lys pro ile glu ile lys glu glu arg arg pro

A. b.  GCA CTA GAT GTA GAG ATT TTT GGA GAG GTA GAA AAC AAT GAG CAA TAA
       ala leu asp val glu ile phe gly glu val glu asn asn glu gln OCH
                                  AGG TAG AAA ACA ATG AGC AAT AAT GAA AAA GTA TTG TCA CCA ATT GAG ATA AAA GAG
                                              met ser asn asn glu lys val leu ser pro ile glu ile lys glu

S. b.  GCT TTA GAT GTT GAG ATT TAT GAT GAG GTG GAA GAA GAT GAC ACG ACT AGA ACC AAT TGA
       ala leu asp val glu ile tyr asp glu val glu glu asp asp thr thr arg thr asn OPA
                                  AGG TGG AAG AAG ATG ACA CGA CTA GAA CCA ATT GAG ATT CAC GAA GAG AGC CGT CCT
                                              met thr arg leu glu pro ile glu ile his glu glu ser arg pro

A. a.  GAG AAG CCT TGT GAC GAG GTT GAA GTT AAA GAG GAG GAA GAA AAA TGA
       glu lys pro cys asp glu val glu val lys glu glu glu glu lys OPA
                                                GGA GGA AGA AAA ATG AGT GAA GCA AGA AGG GGT ATC TTC CCC TTC TCA AAA
                                                        Met ser glu ala arg arg gly ile phe pro phe ser lys
```

FIG. 1. The structure of the *rpoB-rpoC* junction in selected ε-proteobacteria and *A. aeolicus*. The DNA sequence of the *rpoB-rpoC* junction in gastric *H. nemestineae* (*H. n.*) corresponding to *H. pylori* 26695 codons 1366 to 1392 is aligned to the corresponding sequences from intestinal *H. cinaedi* (*H. c.*), *W. succinogenes* (*W. s.*), *C. rectum* (*C. r.*), *A. butzleri* (*A. b.*), and *S. barnesii* (*S. b.*). The sequence of the *rpoB-rpoC* junction in *A. aeolicus* (*A. a.*) does not align well with the helicobacter sequence and is shown only to illustrate the structure of the *rpoB-rpoC* overlap in this organism. The deduced amino acid sequences are also shown. The deduced initiating codons are shown in bold, and the likely ribosome-binding sites are underlined.

form a coherent group with *W. succinogenes* just outside. Overall, these data correlate very well (with some minor differences) with 16S rRNA data (Fig. 2B) as well as with an *rpoB-rpoC* tree built using DNA sequences (data not shown). Previous rRNA sequence analysis had suggested the classification of *C. rectus* from its initial placement as a *Wolinella* (5). Our analysis firmly supports this placement, based on (i) sequence analysis per se and (ii) the fact that *rpoB* and *rpoC* in this organism are separate genes.

The most striking feature of the phylogenetic tree presented in Fig. 2 is that all members of the *Helicobacteraceae*, both gastric and nongastric, have fused *rpoB* and *rpoC* genes (shown

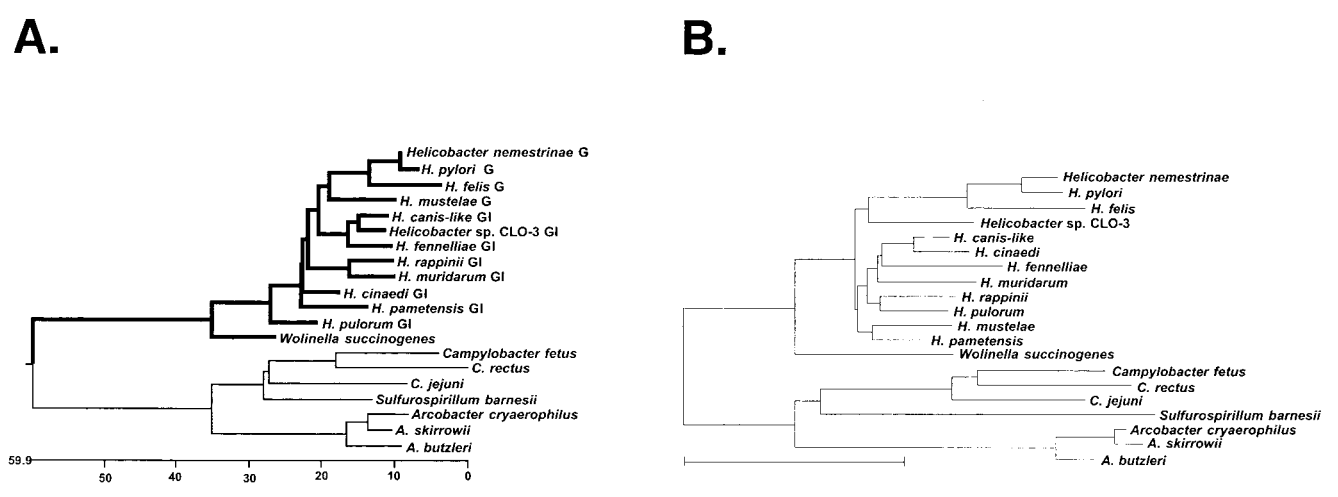

FIG. 2. (A) Phylogenetic tree of derived amino acid sequences from around the *rpoB-rpoC* genes junction in ε-proteobacteria. The following helicobacters were used in the analysis: *H. pylori* 26695 (7), *H. felis* (8), *H. nemestrinae* (NADC 6904), and *H. mustelae* (ATCC 43772T) (all gastric; labeled "G" in the figure); *H. cinaedi* (CCUG 18818T), *H. fennelliae* (NADC 3235), *H. muridarum* (NADC 6895), *H. canis-like* (NADC 29176), *H. pametensis* (NADC 6900), *Helicobacter* sp. CLO3 (CCUG 14564), *H. rappini* (NADC 9615), and *H. pulorum* (all gastrointestinal; labeled "GI" in the figure). Other organisms used were *W. succinogenes* (ATCC 29543), *C. rectus* (NADC 3165), *A. skirrowii* (NADC 3700), *A. cryaerophilus* (NADC 3544), *A. butzleri* (3492), and *S. barnesii* (SES-3). The deduced amino acid sequences were aligned and the tree was built by using the Clustal method with the PAM250 residue weight table (Windows 32 MegAlign program, version 3.1.7 of DNASTAR). (B) Phylogenetic tree derived from 16S rRNA sequences.

by bold lines in Fig. 2A). In contrast, in members of the *Campylobacteraceae*, these two genes are separate. Although the *rpoB* and *rpoC* genes are separated by an untranslated linker of 20 to 100 bp in most bacterial species, in the *Campylobacteraceae*, these two genes partially overlap. Interestingly, the only other eubacterial species with an *rpoB-rpoC* overlap is *Aquifex aeolicus*, apparently the most deeply branching member of the eubacteria, based on 16S rRNA sequence analysis (2). However, the *A. aeolicus rpoB* and *rpoC* sequences strongly resemble proteobacterial sequences based on their primary sequence and the presence of long dispensable regions typical of proteobacteria (data not shown). In addition, phylogenetic analysis of primary sigma factors also places *A. aeolicus* and *H. pylori* together (3). These results strongly suggest the horizontal transfer of *rpo* genes during the evolution of *A. aeolicus*.

The fused *rpoB-rpoC* structure of the *Helicobacteraceae* probably originated in the common ancestor of present-day helicobacters and wolinellas by a simple frameshift mutation (either an insertion of 1 nucleotide base pair or the deletion of 2 bp) at the site of the *rpoB* and *rpoC* overlap. This original frameshift mutation might have been an evolutionary accident which was not specifically selected against nor required (at least initially) for gastric colonization. Indeed, an engineered *H. pylori* strain with separated *rpoB* and *rpoC* genes is viable and can colonize conventional mice at least for the short run (6). Additional experiments are needed, however, to examine more closely the possible contribution of the *rpoB-rpoC* fusion to *H. pylori* fitness during chronic infection and severe inflammatory responses.

The sequences reported in this paper have been deposited in the GenBank (accession no. AF136503 to AF136518).

This work was supported by the Burroughs Wellcome Career Award, a Charles and Johanna Busch Biomedical grant, and NIH grant RO1 GM 59295 (to K.S.); NIH grants DK48029 and AI138166 (to D.E.B.); and NIDCR grants DE-10374 (to F.E.D.) and DE-11443 (to B.J.P.).

We are grateful to John F. Stolz for providing *S. barnesii* DNA. N.Z. is a recipient of a Charles and Johanna Busch Postdoctoral Fellowship.

ADDENDUM IN PROOF

Recent phylogenetic analysis of *rpoB* and *rpoC* genes of *Aquifex pyrophilus* confirms the placement of the genus *Aquifex* within or close to the ϵ group of proteobacteria (H.-P. Klenk, T.-D. Meier, P. Durovic, V. Schwass, F. Lottspeich, D. P. Dennis, and W. Zillig, J. Mol. Evol. **48:**528–541, 1999).

REFERENCES

1. **Anderson, K. F., J. A. Kiehlbauch, D. C. Anderson, H. M. McClure, and I. K. Wachsmuth.** 1993. *Arcobacter* (*Campylobacter*) *butzleri*-associated diarrheal illness in a nonhuman primate population. Infect. Immun. **61:**2220–2223.
2. **Deckert, G., P. V. Warren, T. Gaasterland, W. G. Young, A. L. Lenox, D. E. Graham, R. Overbeek, M. A. Snead, M. Keller, M. Aujay, R. Huber, R. A. Feldman, J. M. Short, G. J. Olsen, and R. V. Swanson.** 1998. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. Nature **392:** 353–358.
3. **Gruber, T. M., and D. A. Bryant.** 1998. Characterization of the group 1 and group 2 sigma factors of the green sulfur bacterium *Chlorobium tepidum* and the green non-sulfur bacterium *Chloroflexus aurantiacus*. Arch. Microbiol. **170:**285–296.
4. **Parsonnet, J., R. A. Harris, H. M. Hack, and D. K. Owens.** 1996. Modeling cost-effectiveness of *Helicobacter pylori* screening to prevent gastric cancer: a mandate for clinical trials. Lancet **348:**150–154.
5. **Paster, B. J., and F. E. Dewhirst.** 1988. Phylogeny of campylobacters, wolinellas, *Bacteroides gracilis*, and *Bacteroides ureolyticus* by 16S rRNA sequencing. Int. J. Syst. Bacteriol. **38:**56–62.
6. **Raudonikiene, A., N. Zakharova, W. W. Su, J.-Y. Jeong, L. Bryden, P. S. Hoffman, D. E. Berg, and K. Severinov.** 1999. *Helicobacter pylori* with separate β and β′ subunits of RNA polymerase is viable and can colonize conventional mice. Mol. Microbiol. **32:**131–138.
7. **Tomb, J. F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, B. D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzgerald, N. Lee, M. D. Adams, J. C. Venter, et al.** 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature **388:**539–547.
8. **Zakharova, N., P. S. Hoffman, D. E. Berg, and K. Severinov.** 1998. The largest subunits of RNA polymerase from gastric helicobacters are tethered. J. Biol. Chem. **273:**19371–19374.