JOURNAL OF BACTERIOLOGY, June 1989, p. 3553–3556
0021-9193/89/063553-04$02.00/0

Vol. 171, No. 6

Unusual Nucleotide Arrangement with Repeated Sequences in the *Escherichia coli* K-12 Chromosome

ATSUO NAKATA,* MITSUKO AMEMURA, AND KOZO MAKINO

Department of Experimental Chemotherapy, Research Institute for Microbial Diseases, Osaka University, 3-1, Yamadaoka, Suita, Osaka, Japan 565

Received 19 December 1988/Accepted 13 March 1989

Between 59 and 60 min on the *Escherichia coli* genetic map, there is a highly conserved sequence of 29 base pairs, containing an inverted repeat of seven base pairs that appears 14 times, 32 or 33 base pairs apart, downstream of the *iap* gene coding region. About 24 kilobase pairs downstream of the 14 repeats, a similar 29-base-pair sequence with a spacing of 32 base pairs appears seven times. Nucleotide sequences hybridizing with the 29-base-pair fragment were also detected in *Shigella dysenteriae* and *Salmonella typhimurium* but not in *Klebsiella pneumoniae* or *Pseudomonas aeruginosa*.

In *Escherichia coli*, there are repeated nucleotide sequences such as duplicated genes, genes presumably derived from a common ancestral gene, short duplicated segments within genes sharing a common function within a protein product, large directly repeated or inverted sequences, and small repeated sequences both within and between coding segments (10). Among short duplications in noncoding regions, the repetitive extracistronic palindromic sequences or palindromic units of 20 to 40 nucleotides with dyad symmetry occupy almost 1% of the whole genome of *E. coli* (4).

We described an unusual nucleotide arrangement in which a highly conserved sequence of 29 base pairs (bp) is tandemly repeated five times with a spacing of 32 bp downstream of the *iap* gene (6). Although the 29-bp sequence contains an inverted repeat of 7 bp, it has no similarity to the repetitive extracistronic palindromic sequences or *hin* and *gin* group sequences (4, 10).

Detection of the 29-bp repeats in the *E. coli* genome. The DNA fragment containing the *iap* gene was cloned from a strain carrying an F′ factor (9), so we could not rule out the possibility that the repeated sequences were derived from the F factor. To detect the 29-bp sequence in the genome of *E. coli*, the chromosomal DNA was isolated from strain K10 (3), as described previously (1), and digested with *Bam*HI, *Eco*RI, or *Hin*dIII. The DNA fragments were analyzed by Southern blot hybridization (13) after electrophoresis in a 0.8% agarose gel with a mixture of two ^{32}P-end-labeled synthetic oligonucleotides, CGGTTTATCCCCGCTGGCGC GGGGAACTC and CGGTTTATCCCCGCTAACGCGGGG AACTC, both of which are the conserved sequences for the repeats (6).

Two *Bam*HI fragments, about 8.0 and 3.9 kilobase pairs (kb), two *Hin*dIII fragments, about 9.7 and 1.5 kb, and one *Eco*RI fragment of about 30 kb hybridized with the probe oligonucleotides (Fig. 1). These results indicate that the two *Bam*HI or *Hin*dIII fragments are arranged contiguously and that the repeats similar to the 29-bp sequence exist in one region of the *E. coli* chromosome or that these fragments are in separate regions within the *Eco*RI fragment. Based on the intensity of these bands on the autoradiogram, we assumed that more repeats similar to the 29-bp sequence existed in the 3.9-kb *Bam*HI and the 9.7-kb *Hin*dIII fragments than in the 8.0-kb *Bam*HI or 1.5-kb *Hin*dIII fragments.

Cloning and sequencing of the DNA fragments containing the repeats. Chromosomal DNA isolated from strain K10 was digested with *Hin*dIII and ligated to the double-stranded DNA (replicative form DNA) of bacteriophage M13mp18. Strain JM103 (8) was transfected with the recombinant DNA and screened for plaques containing the repeated sequences by hybridization with the end-labeled synthetic 29-mer oligonucleotides used as probes. Two phage clones containing the 9.7- and the 1.5-kb chromosomal DNA fragments that hybridized with the probe DNAs were obtained. The *iap* gene and the five repeats of the 29-bp sequence are contained in the 9.7-kb *Hin*dIII fragment (6, 9). The DNA fragment was

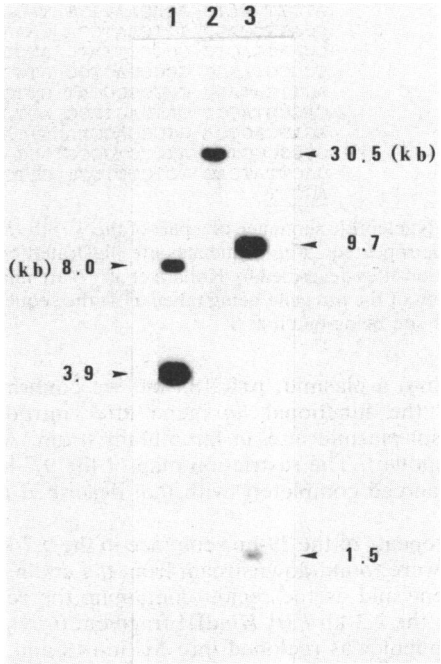

FIG. 1. Southern blot hybridization of chromosomal DNA isolated from *E. coli* K10 after digestion with *Bam*HI (lane 1), *Eco*RI (lane 2), and *Hin*dIII (lane 3). A mixture of two end-labeled synthetic oligonucleotides, CGGTTTATCCCCGCTAACGCGGGG AACTC and CGGTTTATCCCCGCTGGCGCGGGGAACTC, was used as a probe.

* Corresponding author.

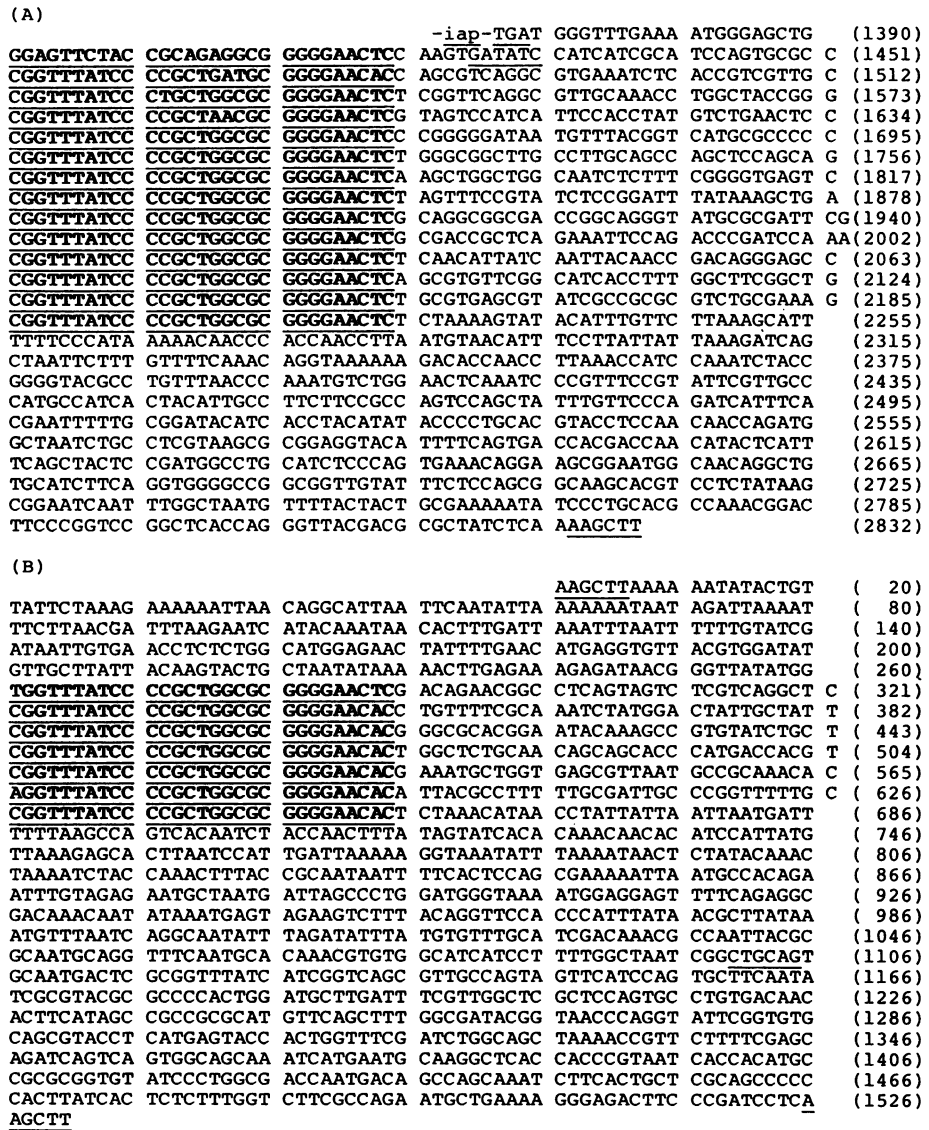

FIG. 2. Nucleotide sequence of a part of the 9.7-kb *Hin*dIII fragment (A) and of the entire 1.5-kb *Hin*dIII fragment (B) containing the 29-bp repeats. The repeated 29-bp sequences are illustrated by underlined boldface letters, and the translation termination codon of the *iap* gene and restriction sites described by Kohara et al. (7) are underlined. Nucleotide numbers in parentheses in panel A were counted with the *Eco*RI site upstream of the *iap* gene being taken as 1; the sequence up to 1664 was as described previously (6). Those in panel B were counted with the *Hin*dIII site being taken as 1.

recloned into a plasmid, pUC18, and we confirmed that it contained the functional *iap* gene after introducing the recombinant plasmid into an Iap mutant strain, ANJ426 (9; data not shown). The restriction map of the 9.7-kb *Hin*dIII fragment agreed completely with that described previously (9).

As the repeats of the 29-bp sequence in the 9.7-kb *Hin*dIII fragment were found downstream from the coding region of the *iap* gene and as the region containing the repeats was carried on the 2.3-kb *Pst*I-*Hin*dIII fragment (6, 9), the latter DNA fragment was recloned into M13mp18 and M13mp19 phages for sequencing. Nucleotide sequences of the 2.3-kb DNA fragment and the 1.5-kb *Hin*dIII fragment that hybridized with the probe DNA were analyzed by the dideoxy-chain termination method (8, 11). A total of 14 repeats of the conserved 29-bp sequence were found in the 2.3-kb *Pst*I-*Hin*dIII fragment, and seven repeats were found in the 1.5-kb *Hin*dIII fragment (Fig. 2A and B). Nucleotide C at 1664 in Fig. 2A was mistakenly identified as G in a previous paper (6); this nucleotide was derived from the vector plasmid. The repeated 29-bp sequences are highly conserved. The nucleotide sequences of 15 repeats among the 21 were exactly the same, except for the 28th nucleotides, where A or T was found. The homology of the first repeat to the consensus sequence was weaker (56%) than the remaining repeats, which deviated from the consensus only by 1 or 2 bp. Of the 29 nucleotides, 20 were either G or C (71%), and therefore the G+C content in this region was higher than the mean for the chromosome DNA, 51% (12). Spacing between each of the 29-bp repeats was exactly 32 bp, except for the two spacer regions that were made of 33 nucleotides.

Possible open reading frames were sought in both DNA segments, with consideration of the possible role of the repeats in regulation of the neighboring genes. Two open

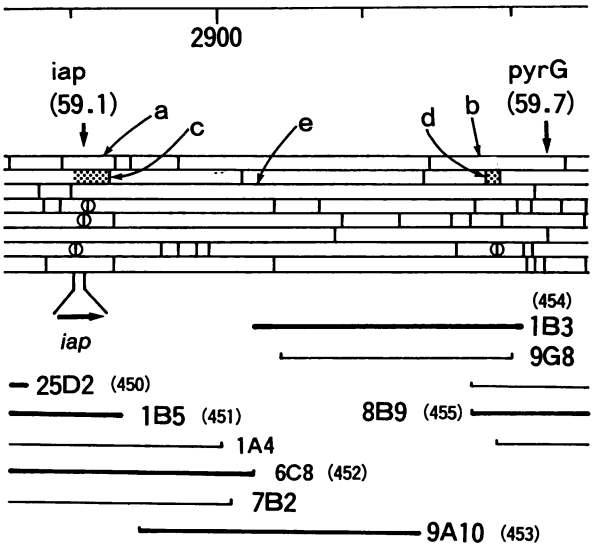

FIG. 3. Restriction map of the *iap-pyrG* gene region of *E. coli* chromosomal DNA (7) and the location of the repeated sequences. The restriction cleavage sites are *Bam*HI (top line), *Hin*dIII (line 2), *Eco*RI (line 3), *Eco*RV (line 4), *Bgl*I (line 5), *Kpn*I (line 6), *Pst*I (line 7), and *Pvu*II (bottom line). The regions sequenced in our previous study and this study are shown by stippled boxes, and the restriction cleavage sites found in the sequenced regions are circled. Letters: a, 3.9-kb *Bam*HI fragment; b, 8.0-kb *Bam*HI fragment; c, 9.7-kb *Hin*dIII fragment; d, 1.5-kb *Hin*dIII fragment; e, 30-kb *Eco*RI fragment.

reading frames that could code for proteins with molecular weights of 10,000 or more were found in the region covering the repeats in each DNA segment, one downstream of the *iap* gene in the same orientation (nucleotides 1929 to 2216 [Fig. 2A]) and the other in the 1.5-kb segment (nucleotides 257 to 550 [Fig. 2B]). However, no consensus sequence for ribosome-binding sites was found at an appropriate position upstream from these open reading frames. An open reading frame coded for by the other DNA strand was found in each of the DNA segments. The one continuing from the 3′ end of the sequenced region and ending at nucleotide 998 in the 1.5-kb segment, which coded for 174 amino acids, may be a distal part of an unknown gene. No repeats of the 29-bp sequence were in this region. Thus, these repeated sequences were probably in noncoding regions. The biological function of the repeated sequences, such as chromosomal rearrangement or modulation of expression of the gene immediately upstream or downstream, is not known.

Location of the DNA fragment containing the repeats. The *iap* gene and one set of the repeats are carried on the 9.7-kb *Hin*dIII fragment, the other set is on the 1.5-kb *Hin*dIII fragment, and both sets are on the single 30-kb *Eco*RI fragment, so two sets of the repeat should be on a 30-kb *Eco*RI fragment in the vicinity of the *iap* gene. In addition, each set of the repeats is carried on the 8.0- and 3.9-kb *Bam*HI fragments, respectively. Therefore, both *Bam*HI fragments should be contained in or overlap the 30-kb *Eco*RI fragment. To locate these two sets, we examined the restriction map of the *E. coli* chromosome constructed by Kohara et al. (7) in the vicinity of the *iap* locus. A region that fulfills the above conditions was found between *iap* (59.1 min) and *pyrG* (59.7 min); one set was near the *iap* gene and the other was near the *pyrG* gene (Fig. 3).

The chromosomal segments in this region have been

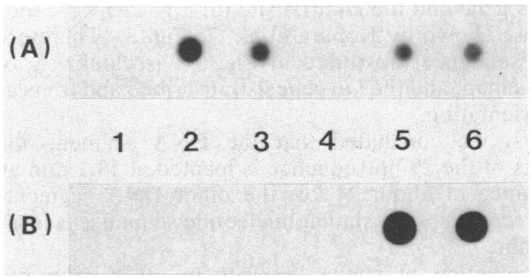

FIG. 4. Spot hybridization of lambda phage DNA isolated from phage clones. Lanes: 1, 25D2(450); 2, 1B5(451); 3, 6C8(452); 4, 9A10(453); 5, 1B3(454); and 6, 8B9(455). End-labeled probe DNAs used were synthetic oligonucleotides (A) and the 1.5-kb *Hin*dIII fragment (B).

cloned into the lambda clones 1B5, 6C8, 9A10, 1B3, and 8B9 (7; Fig. 3). To locate each set of repeats, we hybridized the DNAs isolated from these phages and the ^{32}P-labeled synthetic oligonucleotides or the labeled 1.5-kb *Hin*dIII fragment (Fig. 4). DNAs of phage clones 1B5 and 6C8 hybridized with the 29-mer probe oligonucleotides more than those of 1B3 and 8B9 did, and clones 1B3 and 8B9 hybridized with the 1.5-kb *Hin*dIII fragment much more than the other two did. DNAs of 25D2 and 9A10 clones, neither covering the 3.9-kb *Bam*HI fragment or the 1.5-kb *Hin*dIII fragment, did not hybridize with either of these DNA probes. Chromosomal DNA segments cloned into these phage clones did not complement the *iap* gene mutation (data not shown). Thus, the two DNA segments containing the repeated sequences were on the *Eco*RI fragment, one in the end at 59.1 min and the other in the other end near 59.7 min (Fig. 3).

The order and relative positions of the restriction sites of *Pst*I, *Bgl*I, and *Eco*RV found in the DNA segment between

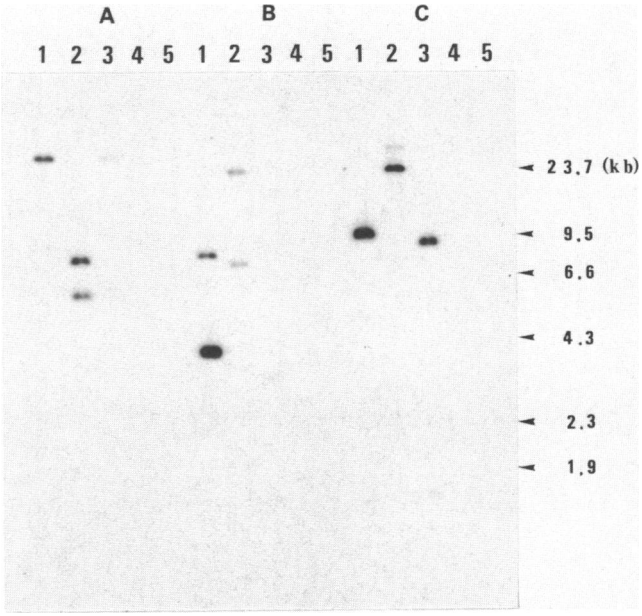

FIG. 5. Southern blot hybridization of chromosomal DNA isolated from *E. coli* K10 (lanes 1), *Salmonella typhimurium* TA1535 (lanes 2), *Shigella dysenteriae* Sh (lanes 3), *K. pneumoniae* (lanes 4), and *P. aeruginosa* (lanes 5) after digestion with *Eco*RI (A), *Bam*HI (B), or *Hin*dIII (C). A mixture of end-labeled synthetic oligonucleotides was used as a probe.

the *iap* gene and the *Hin*dIII site (6; Fig. 2A) were the same as those shown by Kohara et al. (7; Fig. 3). Therefore, the DNA sequence illustrated in Fig. 2A is clockwise on the genetic map, and the *iap* gene is transcribed and translated in this orientation.

Thus, we concluded that the DNA segment with 14 repeats of the 29-bp sequence is located at 59.1 min and, at a distance of about 24 kb, the other DNA segment with seven repeats of the similar nucleotide sequence is located at 59.6 min.

Examination of similar repeats in other gram-negative bacteria. To see whether sequences similar to the 29-bp sequence exist in other *E. coli* strains and in other species of gram-negative bacteria, Southern blotting analysis was done with chromosomal DNA from *E. coli* C600 and Ymel (3), *Salmonella typhimurium* TA1535 (2), *Shigella dysenteriae* Sh (5), *Klebsiella pneumoniae*, and *Pseudomonas aeruginosa* (obtained from the Institute for Fermentation, Osaka), with a mixture of synthetic oligonucleotides as a probe (Fig. 5). Chromosomal DNAs of strains of *Salmonella typhimurium* and *Shigella dysenteriae*, as well as of other *E. coli* strains (data not shown), hybridized with the probe, but no hybridizable DNA sequence was detected in the genome of *K. pneumoniae* or *P. aeruginosa*.

We thank M. Yamamoto and T. Takeda of the Laboratory of Molecular Genetics, Institute for Medical Sciences, University of Tokyo, for supplying synthetic oligonucleotides, Y. Kohara for lambda phage clones, T. Fukasawa for the *Shigella dysenteriae* strain, and H. Shinagawa for critical reading of the manuscript.

This study was supported by a grant-in-aid for Scientific Research from the Ministry of Education, Science, and Culture, Japan.

LITERATURE CITED

1. **Amemura, M., H. Shinagawa, K. Makino, N. Otsuji, and A. Nakata.** 1982. Cloning of and complementation tests with alkaline phosphatase regulatory genes (*phoS* and *phoT*) of *Escherichia coli*. J. Bacteriol. **152**:692–701.
2. **Ames, B. N., F. D. Lee, and W. E. Durston.** 1973. An improved bacterial test system for the detection and classification of mutagens and carcinogens. Proc. Natl. Acad. Sci. USA **70:** 782–786.
3. **Bachmann, B. J.** 1972. Pedigrees of some mutant strains of *Escherichia coli* K-12. Bacteriol. Rev. **36**:525–557.
4. **Gilson, E., J. M. Clément, D. Perrin, and M. Hofnung.** 1987. Palindromic units: a case of highly repetitive DNA sequences in bacteria. Trends Genet. **3**:226–230.
5. **Hattman, S., and T. Fukasawa.** 1963. Modification of T-even phages due to defective glucosylation of their DNA. Proc. Natl. Acad. Sci. USA **50**:297–300.
6. **Ishino, Y., H. Shinagawa, K. Makino, M. Amemura, and A. Nakata.** 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. J. Bacteriol. **169:** 5429–5433.
7. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole E. coli chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50**:495–508.
8. **Messing, J., R. Crea, and P. H. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
9. **Nakata, A., H. Shinagawa, and M. Amemura.** 1982. Cloning of alkaline phosphatase isozyme gene (*iap*) of *Escherichia coli*. Gene **19**:313–319.
10. **Riley, M., and S. Krawiec.** 1987. Genome organization, p. 967–981. *In* J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umberger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
11. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
12. **Schildkraut, C. L., J. Marmur, and P. Doty.** 1962. Determination of the base composition of deoxyribonucleic acid from its buoyant density in CsCl. J. Mol. Biol. **4**:430–443.
13. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.