大腸菌rhlE遺伝子の構造解析

誌名	The Japanese journal of genetics
ISSN	0021504X
著者	大森, 治夫
巻/号	69巻1号
掲載ページ	p. 1-12
発行年月	1994年2月

Jpn. J. Genet. (1994) **69**, *pp.* 1–12

Structural analysis of the *rhlE* gene of *Escherichia coli*

Haruo OHMORI

Institute For Virus Research, Kyoto University
53 Shougoin, Kawahara-machi, Sakyo-ku, Kyoto 601-01, Japan

(*Received 10 October 1993*)

ABSTRACT

The *E. coli* chromosome is known to carry at least five genes, each of which codes for a "D-E-A-D" box protein that is presumed to possess an ATP-dependent RNA helicase activity. Four of such genes (*srmB*, *deaD*, *dbpA* and *rhlB*) were already mapped on the *E. coli* chromosome and their DNA sequences determined. We here report the complete nucleotide sequence of the remaining *rhlE* gene located at about 17.8 min on the *E. coli* genetic map. RhlE protein possesses all of the motifs (I to VI) conserved among prokaryotic and eukaryotic "D-E-A-D" proteins and has an arginine-rich carboxyl-terminal region. A null mutant of the *rhlE* gene was constructed by a new method with a ColE1 plasmid mutant that replicates in RNase HI-deficient bacterial strains, but not in the wild-type strains. The *ΔrhlE* mutant can grow normally, implying that the *rhlE* gene product is nonessential for bacterial cell growth.

1. INTRODUCTION

The "D-E-A-D" box proteins are a family of proteins found in a wide variety of organisms from bacteria to humans, which have a core region homologous to eIF-4A, a murine translation initiation factor (Schmid and Linder, 1992). This family of proteins share G-K-T (motif I) and D-E-A-D (motif II) sequences known to be a common A-motif and a special version of B-motif, respectively, for ATP binding sites (Walker et al., 1982). They also share other sequences such as P-T-R-E-L-A (motif IA) and H-R-I-G-R-T-G-R (motif VI) which are unique to this family (Gorbalenya et al., 1989; Linder et al., 1989). The "D-E-A-D" box proteins are assumed to have ATP-dependent RNA helicase or RNA-dependent ATPase activity, while such an enzyme activity has been demonstrated only with a few members of them, for example, the prototype eIF-4A (Ray et al., 1985) and the human p 68 protein (Hirling et al., 1989; Iggo and Lane, 1989). Individual members of the family possess specific amino- and carboxyl-terminal regions of variable lengths for respective functions, which are in most cases related to rearrangement of RNA structures during translation, ribosomal biogenesis, or RNA processing.

---

* The sequence of *rhlE* and its flanking regions has been submitted to GenBank and given the accession number L02123.

At least five D-E-A-D box genes have been found in *E. coli* (Kalman et al., 1991). First, the *srmB* gene was identified as a multiple copy suppressor for a temperature-sensitive (ts) mutation in the ribosomal protein L24 (Nishi et al., 1989). Purified SrmB protein was shown to have RNA-dependent ATPase activity. Another gene, *deaD*, was also identified as a multiple copy suppressor of the ts mutations in the ribosomal protein S2 (Toone et al., 1991). SrmB and DeaD proteins appear to have different functions or specificities since overexpression of DeaD protein does not suppress the ts mutation in L24 and overexpression of SrmB protein does not suppress the ts mutations in S2 (Toone et al., 1991). The third gene, designated *dbpA* for D-E-A-D box protein, was found by low stringency screening with a cDNA probe from the *S. pombe* homologue of the human p 68 protein (Iggo et al., 1990). The function of this gene is unknown. Furthermore, PCR screening with oligonucleotides corresponding to the common motifs (I, IA, II, and VI) identified two more genes, named *rhlB* and *rhlE* after RNA-helicase like protein (Kalman et al., 1991). The *rhlB* gene is necessary for viability only in some genetic backgrounds, and the conditional lethality is not complemented by the presence of a multiple copy plasmid carrying the *srmB* gene (Kalman et al., 1991). There are least information with regard to the *rhlE* gene; its partial sequence determined with the PCR fragment (from motif I to motif VI) and a rough mapping by hybridization of the fragment with the Kohara phage clone λ # 204 have been reported (Kalman et al., 1991). During the course of studying a gene, designated *rarB*, which is presumed to participate in the alternative replication mechanism of the *E. coli* chromosome independent of *oriC* and *dnaA* (to be published elsewhere), we noticed that the *rhlE* gene is located very closely to the *rarB* gene. This paper describes mapping on the *E. coli* chromosome, sequencing, and isolation of a null mutant, of the *rhlE* gene.

2. MATERIALS AND METHODS

Media and strains

LB [1% Bactotryptone (Difco), 0.5% yeast extracts (Difco), and 0.5% NaCl, pH 7.4] was used for cultivation of bacteria, and NZCYM (Sambrook et al., 1989) was used to grow lambda phages. L-agar contained 1.6% agar in LB. If necessary, ampicillin and kanamycin were added at a final concentration of 50 and 20μg/ml, respectively.

The *E. coli* K-12 strains used are JM101 [*supE*, *thi*, *Δ(lac-proAB)*/F', *traD36*, *proA*$^+$*B*$^+$, *lacI*^{q}*ΔM15*] (Yanish-Perron et al., 1985), its isogenic strains AK101 (*rnhA*::*cat*) (Kanaya et al., 1990) and MV1190 [*Δ(srl-recA)306*::Tn*10*] (Vieira and Messing, 1987). The phage λ#204 of the Kohara library (Kohara et al., 1987) was provided by Dr. Kohara (National Institute of Genetics, Japan). The λgt-λC phage (Thomas et al., 1974) was used for gene disruption experiment and provided by Dr. Inokuchi (Kyoto University, Japan). The phagemid vectors carrying an

ampicillin (Ap)-resistance gene, pTZ18U and pTZ19U were obtained from Bio-Rad (Richmond, CA, USA) and used for subcloning and sequencing. pUC4K (Pharmacia-LKB, Uppsala, Sweden) was used to prepare the DNA fragment containing the kanamycin (Km)-resistance gene.

DNA manipulations

Most procedures for DNA manipulation and bacterial transformation followed the standard protocols (Sambrook et al., 1989). *E. coli* chromosomal DNAs were prepared and subjected to Southern blotting analysis as described (Silhavy et al., 1984), using a nonradioactive DNA labeling and detection kit from Boehringer Mannheim (Mannheim, Germany).

Construction of pTZ19Urrh and other plasmids

An RNase H-sensitive replication mutant, pTZ19Urrh, was constructed from pTZ19U by sequence-directed mutagenesis of the Kunkel method (Kunkel, 1985). pTZ19Urrh has deleted the two bases at the unique *Hae*II site (from 5′-AGCGCC-3′ to 5′-AGCC-3′) in the region encoding the primer-precursor (RNA II) for initiation of the plasmid replication, and consequently its replication becomes sensitive to RNase H (Masukata and Tomizawa, 1984; Naito et al., 1984; Ohmori et al., 1987); that is, the plasmid can replicate in RNase HI-deficient (*rnhA*$^-$) strains of *E. coli*, but not in the wild-type strains containing the active enzyme. DNAs of pTZ19Urrh and its derivatives were prepared from the strain AK101 carrying such plasmids. Other plasmids were constructed in this experiment.

DNA sequence analysis

DNA sequencing by the dideoxy chain terminator method (Sanger et al., 1977) was done with single-stranded DNAs prepared by infection of the helper phage M13K07 (Vieira and Messing, 1987), using a Sequenase kit from US Biochemicals (Cleveland, OH, USA). Computer analyses of the nucleotide and amino-acid sequences were done with the SDC-GENETYX programs (Software Development, Tokyo, Japan).

3. RESULTS

The nucleotide sequence of the rhlE *gene*

The PCR-amplified DNA fragment of the RhlE core region was found to hybridize to the Kohara phage λ#204 (1B4) DNA (Kalman et al., 1991). The *E. coli* chromosomal DNA fragment carried by the phage was prepared after *Eco*RI digestion of the phage DNA and treated with *Sma*I and other restriction enzymes to locate their cleavage sites within the fragment. The fragment (Fig. 1) was found to contain some additional sites for *Eco*RV, *Bgl*I, and *Pvu*II to those shown in the original map (Kohara et al., 1987). The 858 base-pairs (bp) sequence of the

PCR-amplified *rhlE* fragment (GenBank database Accession Number X56037), contained single sites for *Eco*RV and *Sma*I, respectively, within a very short (109 bp) distance. This information allowed us to locate the *rhlE* fragment at around 2.0 kilo-bases (kb) away from the left end of the *E. coli* chromosomal segment. Figure 2 shows the 2600 bp sequence determined, which corresponds to the 842–845 kb region of the Kohara map and also to the 17.8 min region of the genetic map between *chlAM* and *glnQPH* (Bachmann, 1990). This location of the *rhlE* gene differs from the assignment by Rudd (1992), who mapped the gene on the adjacent *Bam*HI segment (the B1-B2 segment in Fig. 1), relying on the 858 bp sequence of the PCR-amplified *rhlE* fragment. Our sequence differs at 12 sites from the above-mentioned sequence of the PCR fragment, and generates three amino acid changes from the published RhlE core sequence (Kalman et al., 1991), all at the nonconserved residues. The coding region probably starts with ATG at 540–542, since it is preceded by a good candidate (GGAG at 531–534) for the Shine-Dalgarno sequence. This open reading frame (ORF) terminates with TGA at 2004–2006, which is followed by a repetitive extragenic palindromic element (REP) (Higgins et al., 1987). A short ORF in an opposite direction (f160, Fig. 1) terminates at 2138–2136, although there is no evidence that it is indeed translated. The REP element is also found in the 5.2 kb downstream, repeated twice, near the ends of two oppositely directed ORFs (Fig. 1). Another putative ORF

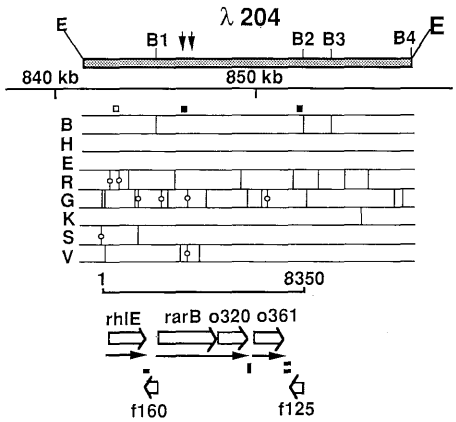

Fig. 1. Physical map of the *E. coli* chromosomal region carried by λ#204. The *E. coli* chromosomal region is flanked by an *Eco*RI site at each end, derived from the cloning vector λEMBL4. The second horizontal line from the top indicates the physical map position of the *E. coli* chromosome (Kohara et al., 1987). The open square indicates a *Sma*I site (another *Sma*I site present in this region is not precisely localized), and the filled squares indicate *Bgl*II sites. The symbols B, H, E, R, G, K, S, and V represent *Bam*HI, *Hin*dIII, *Eco*RI, *Eco*RV, *Bgl*I, *Kpn*I, *Pst*I, and *Pvu*II, respectively. The vertical lines with an open circle denote sites not present in the original Kohara map, but found from the DNA sequence. The rightmost *Bam*HI site (B4) was regenerated during cloning of the *Sau*3AI partial digest of *E. coli* DNA into the *Bam*HI site on λEMBL4. The two vertical arrows indicate the sites of mini-Tn*10* (Km) insertion in the *rarB11* and *rarB19* mutants (our unpublished data). Four rightward (clockwise direction in the genetic map) and two leftward ORFs as indicated were found in the determined DNA sequence from the left *Pst*I site to the second *Bam*HI (B2) site (8356 bp in total, our unpublished data). The thin horizontal arrows indicate the direction of transcription. The thick horizontal bar represents REP sequences and the vertical bar represents a putative Rho-independent transcription termination signal.

preceding *rhlE* terminates at 327–329, which is followed by a palindromic sequence that might function as a transcription terminator signal. These surrounding sequences suggest that the *rhlE* transcription is monocistronic.

```
        PstI
  1   CTGCAGGCGCGAAGGATCAGTTCACGGATGGCAGCCCGATCAGGCTGTGGTTGCGCGAAC
 61   AAGCGTTCGGCTTCCTCGGCATGCGGACGGAACTGCTCGCCAATAAAATCGGCAATCCAC
121   TGGGCGCAGGCGAGGTACAAATCTTCTTTCGAACCGAAGTAGTAGGTGATGGCAGCGATA
181   TTCTGCCCGGCCTGGGCGGCTATCTCGCGAGTGGTGGCGTTCATTCCATATTCACCAAAC
241   TGCGCCAGTGCGGCAGCAATCAGCTGTTTTTTCGCCTGTTCACCCTTGATTGTCATGGCA
                                                          -35
301   GGATTATTCATCGCACAGTCCATTCTTAATCAAATGATTGATTAAGATTATGACTCCATA
                  -10
361   GGGGAGTTGTCCAGTATGGCTAAGAATTTTAGCAACGCCAGTCACAGGGATAATTTATGC

421   GCTGCGTCACAAAAACTGCTACACTCCGCTCCCTCATGACATTGTGGTTTTTGTCATTTT
                                                          SD
481   CCTTTTCAGTATCTCCCTGAAAACTACACCGGTAACGGTCGGGGCGGTTCGGAGTAGTTA
                                                           M
541   TGTCTTTCGATTCTTTGGGTTTAAGCCCTGATATCCTGCGCGCCGTTGCCGAGCAGGGTT
  2     S  F  D  S  L  G  L  S  P  D  I  L  R  A  V  A  E  Q  G  Y
601   ACCGTGAACCCACCCCTATTCAGCAGCAGGCGATCCCTGCGGTGCTGGAAGGCCGCGACC
 22     R  E  P  T  P  I  Q  Q  Q  A  I  P  A  V  L  E  G  R  D  L
                *  *  *  Q        *  *  P        L     G  *  D  *
661   TGATGGCTAGCGCCCAGACCGGCACCGGCAAAACAGCGGGCTTTACGCTGCCGCTGTTGC
 42     M  A  S  A  Q  T  G  T  G  K  T  A  G  F  T  L  P  L  L  Q
                A     T  G  *  G  K  T  *  *
                                  motif I
721   AACACCTGATCACTCGCCAGCCGCACGCCAAAGGGCGTCGTCCGGTACGTGCGCTCATTC
 62     H  L  I  T  R  Q  P  H  A  K  G  R  R  P  V  R  A  L  I  L
                                                             L  *  *
781   TTACCCCGACCCGTGAACTGGCGGCGCAGATTGGCGAAAACGTCCGTGATTACAGCAAAT
 82     T  P  T  R  E  L  A  A  Q  I  G  E  N  V  R  D  Y  S  K  Y
          P  T  R  E  L  A     Q  *
                motif IA
841   ACCTGAACATTCGTTCGCTGGTGGTGTTTGGTGGTGTGAGTATTAACCCGCAGATGATGA
102    L  N  I  R  S  L  V  V  F  G  G  V  S  I  N  P  Q  M  M  K
                                  G  G
                                        SmaI
901   AACTGCGTGGCGGCGTTGATGTGCTGGTGGCAACCCCGGGACGTTTGCTGGACCTGGAAC
122    L  R  G  G  V  D  V  L  V  A  T  P  G  R  L  L  D  L  E  H
                         *  *  *  *  T  *  G  R  L
961   ATCAGAATGCAGTGAAGCTGGATCAGGTTGAAATCCTCGTCCTCGATGAAGCTGACCGCA
142    Q  N  A  V  K  L  D  Q  V  E  I  L  V  L  D  E  A  D  R  M
                               *        *  *  *  D  E  A  D     M
                                                        motif II
1021  TGCTCGACATGGGCTTTATCCACGATATCCGTCGCGTGTTAACAAAACTACCTGCGAAGC
162    L  D  M  G  F  I  H  D  I  R  R  V  L  T  K  L  P  A  K  R
             G  F           *        *
1081  GCCAGAACCTGTTATTCTCCGCGACCTTCTCTGACGATATTAAAGCCCTGGCGGAAAAAC
182    Q  N  L  L  F  S  A  T  F  S  D  D  I  K  A  L  A  E  K  L
       Q        L  F  S  A  T
                motif III
1141  TGTTGCACAACCCGCTGGAAATCGAAGTGGCACGCCGCAATACCGCGTCTGATCAGGTGA
202    L  H  N  P  L  E  I  E  V  A  R  R  N  T  A  S  D  Q  V  T
1201  CCCAGCACGTTCACTTTGTCGATAAGAAACGCAAACGCGAATTGCTGTCGCACATGATTG
222    Q  H  V  H  F  V  D  K  K  R  K  R  E  L  L  S  H  M  I  G
                                                 L
```

```
1261  GGAAAGGGAACTGGCAGCAGGTGCTGGTGTTTACCCGTACCAAACACGGCGCTAACCATC
 242    K  G  N  W  Q  Q  V  L  V  F  T  R  T  K  H  G  A  N  H  L
                          *  *  F        *                    *
                                motif IV
1321  TGGCTGAACAGCTCAATAAAGATGGCATCCGTAGTGCGGCGATCCACGGCAATAAATCGC
 262    A  E  Q  L  N  K  D  G  I  R  S  A  A  I  H  G  N  K  S  Q
                          G                    G              Q
1381  AAGGTGCGCGTACTCGTGCGCTGGCTGATTTTAAATCGGGCGATATTCGTGTACTGGTGG
 282    G  A  R  T  R  A  L  A  D  F  K  S  G  D  I  R  V  L  V  A
          R                                G           *  L  *  A
1441  CAACTGACATCGCTGCGCGCGGCCTGGATATTGAAGAGCTGCCGCACGTCGTCAACTATG
 302    T  D  I  A  A  R  G  L  D  I  E  E  L  P  H  V  V  N  Y  E
        T  D  *  A  A  R  G  *              *        V     N     *
                    motif V
1501  AACTGCCAAACGTACCTGAAGATTATGTCCACCGTATCGGGCGTACCGGTCGTGCGGCTG
 322    L  P  N  V  P  E  D  Y  V  H  R  I  G  R  T  G  R  A  A  A
           *                    *  H  R  I  G  R  T  *  R     *
                                          motif VI
1561  CTACCGGTGAAGCGTTGTCGCTGGTGTGTGTTGATGAACACAAACTGCTGCGTGATATCG
 342    T  G  E  A  L  S  L  V  C  V  D  E  H  K  L  L  R  D  I  E
           G        *  *
1621  AAAAACTGCTGAAAAAAGAGATCCCGCGCATTGCGATTCCGGGCTATGAGCCGGACCCGT
 364    K  L  L  K  K  E  I  P  R  I  A  I  P  G  Y  E  P  D  P  S
1681  CAATCAAAGCCGAACCGATCCAGAACGGTCGCCAGCAACGTGGCGGCGGCGGTCGTGGGC
 382    I  K  A  E  P  I  Q  N  G  R  Q  Q  R  G  G  G  G  R  G  Q
1741  AAGGTGGTGGTCGCGGTCAACAGCAACCACGCCGTGGGGAAGGTGGCGCAAAATCTGCAA
 402    G  G  G  R  G  Q  Q  Q  P  R  R  G  E  G  G  A  K  S  A  S
1801  GCGCGAAACCTGCAGAAAAACCGTCTCGCCGCCTCGGCGATGCCAAACCGGCAGGCGAAC
 422    A  K  P  A  E  K  P  S  R  R  L  G  D  A  K  P  A  G  E  Q
1861  AACAACGTCGCCGCCGTCCGCGTAAACCTGCCGCGTGCGCAGTAATCTTTTATGCGGGCT
 442    Q  R  R  R  R  P  R  K  P  A  A  C  A  V  I  F  Y  A  G  Y
                                                          StuI
1921  ATGCCCGGCATCAGGCTGATGAACAAACGCAAAACTGCCTGATGCGCTACGCTTATCAGG
 462    A  R  H  Q  A  D  E  Q  T  Q  N  C  L  M  R  Y  A  Y  Q  A
1981  CCTACGTGAACTCTGCAATATATTGAATTTGCATGCTTTTGTAGGCCGGATAAGGCGTTT
 482    Y  V  N  S  A  I  Y
2041  ACGCCGCATCCGGCATTTCACAACAAGCACTTGTCAGCAATTTGAGAACACGGGAAAAAA

2101  TTTATCTGTTTTACCGCCCATAAGCCACCGAGACGTTACTTCTCTATAGCCAATTGTTCG
2161  CGCAACTCCATTAAAAGGTAGCCCAGTCGATTCTTGCCCTTACCATGACCACCGTCTCCC
2221  CAGTAAGCATCGTTTTCCGTATGCTCAACCAGTTTTGCGGGCGCGGTTGCCAGCAAGAGC
2281  GCACGCAGTTCTGCATGCTGTTCGAATTTAGCGCGAAGTGCTTTTCGCATCACCTGTTCT
2341  TTGACCGACTCCCAGTTTTTACGCAGAGGCTTAGAACGATCGCGCCCCATGCGTGCGGCG
2401  ACCATAGGGGAAGAAACCCGACGGATCTCTTCGCGGTATTTTTCATCAAGGAATTTTTGT
2461  GCCTGAAAATAGTGTTCTGAGGTAGGCCAGGTTTTCCCGTCAACCTTGATGGGCCAGGCG
2521  GCAAAGTTAGAAAAATCACCGTAGTCATCGCTGGTGCTGTAGAAATTTATGATGGTGTCT
                BamHI
2581  TGCATGACGTGTTGGATCCT
```

Fig. 2. The nucleotide sequence around the *rhlE* gene. The 2600 bp sequence from the *Pst*I site to the first *Bam*HI (B1) site is shown. The initiation and termination codons of the *rhlE* gene and the preceding ORF, and the sequence complementary to the termination codon of f160 are indicated in boldface. The putative -35 and -10 sequences for the *rhlE* transcription and the Shine-Dalgarno sequence are also shown in boldface. The palindromic sequences probably acting as a transcriptional terminator and some restriction sites are underlined. Amino-acids are denoted by single letters. The identical residues among all of the five *E. coli* "D-E-A-D" box proteins are indicated below the RhlE sequence, and the conserved (not identical but similar) one are indicated by asterisks.

Expected properties of RhlE protein

From the deduced amino acid sequence, molecular weight of RhlE protein (488 amino-acid residues) is estimated to be 54 kilodalton (KD), similar to those of the four other *E. coli* D-E-A-D box proteins (50 KD of SrmB, 64 KD of DeaD, 46 KD of DbpA, and 47 KD of RhlB). RhlE protein has a core region with all of the well-conserved motifs and a relatively long carboxyl-terminal region, which is rich in basic amino acids, especially arginine (17 Arg and 10 Lys *vs* 5 Asp and 10 Glu residues among the total 150 residues in the carboxyl-terminal region beyond motif VI). The RRRRPRK sequence at 443–449 may correspond to the arginine-rich RNA binding motif as found in λN and HIV Tat proteins (Lazinski et al., 1989; Calnan et al., 1991). Another characteristic sequence in the RhlE carboxyl-terminal region is GRQQRGGGGRGQGGGRGQQQPRRGEGG at 390–416. Similar glycine-rich sequences containing arginines were also found in the carboxyl-terminal region of the other D-E-A-D box proteins, the yeast P 68 homologues DBP2 of *S. cerevisiae* and dbp2 of *S. pombe* (Iggo et al., 1991). The *Drosophila vasa* protein, which also belongs to the D-E-A-D family, contains a glycine-rich heptad repeat, F/SRGGE/QGG, five times in the amino-terminal region (Lasko and Ashburner, 1988; Hay et al., 1988). These glycine-rich sequences are believed to contribute for RNA binding since such sequences are found in some other RNA binding proteins (Jong et al., 1987). The presence of the above two characteristic sequences in the carboxyl-terminal region of RhlE supports that the region functions as a domain for RNA binding.

Disruption of the rhlE *gene*

To examine the *rhlE* gene function, isolation of a null mutant of the gene was undertaken. The method adopted here was to use a mutant of ColE1-type plasmids that can replicate in RNase HI-deficient (*rnhA*$^-$) strains of *E. coli*, but not in the wild-type strains (Ohmori, 1988). Derivatives of such a mutant plasmid carrying an *E. coli* chromosomal segment can be maintained as a high-copy-number plasmid in *rnhA*$^-$ hosts, but in the wild-type strains they can be maintained only in a state integrated into the host chromosome by homologous recombination. For such a purpose, we constructed a mutant of pTZ19U (pTZ19Urrh) by introducing a 2 bp deletion in a region essential for replication in the wild-type cells, as described in MATERIALS AND METHODS. This plasmid and its derivatives were propagated in an *rnhA*$^-$ strain, AK101. The 5 kb *Eco*RI-*Bgl*II DNA fragment containing the *rhlE* coding region (see Fig. 1) was inserted between the *Eco*RI and *Bam*HI sites of pTZ19Urrh to construct pKH 6001. Subsequently, about two-thirds of the *rhlE* gene, flanked by the *Sma*I site at the position 934 and the *Stu*I site at the position 1978, was replaced by the 1.2 kb *Hinc*II fragment carrying the kanamycin-resistance gene of pUC4K. The resultant ampicillin-resistant (Ap^R) and kanamycin-resistant (Km^R) plasmid (designated pKH6002, Fig. 3) generated Km^R transformants with JM101(the *rnhA*$^+$

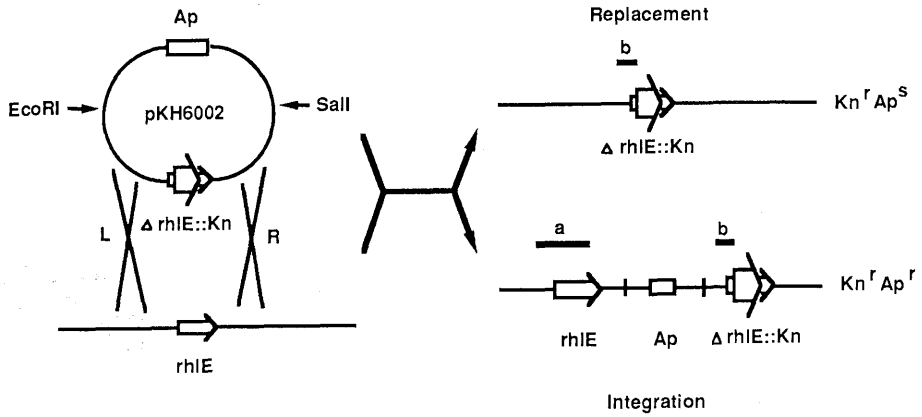

Fig. 3. A schematic representation of the events during transformation. One homologous recombination event at the right (R) side of the *rhlE* gene generates a Ap^S Km^R transformant shown in the lower right, and two such events, once at each side (L and R), generates a Ap^S Km^R transformant. The thick horizontal bar with the letter a or b indicates the DNA region that hybridizes with the probe used for Southern blotting analysis (see Fig. 4).

parental strain of AK101) at the frequency of about 10^{-4} of AK101, but it generated no Km^R transformants with MV1190 (a *ΔrecA* derivative of JM101) implying that *recA*-dependent homologous recombination underlay establishment of the transformants in RNase HI-containing cells. Most of the 38 Km^R transformants of JM101 thus obtained were resistant to ampicillin, but four of them were sensitive to the drug. It suggested that the entire sequence of the pKH 6002 genome was integrated in the chromosome in the major Ap^R Km^R transformants and the intact *rhlE* gene was replaced by the Km^R-inserted gene in the minor four Ap^S Km^R transformants. This result was as expected, since the replacement requires two homologous recombination events once at each side of the kanamycin-resistance gene, while the integration requires such an event just once at either side (Fig. 3). However, the appearance of the disruptants at about only one order less than that of the integrants suggested that the *rhlE* gene could be disrupted without an accompanied suppressor mutation (see below).

To confirm the DNA rearrangement occurred in the Ap^R Km^R and Ap^S Km^R transformants, the chromosomal DNAs were isolated from such transformants and also from the parental strain JM101, and subjected to Southern blotting analysis with a probe carrying a part of the *rhlE* gene (474 bp *Eco*RV fragment from 573 to 1046). The DNA sample from JM101 or the Ap^S Km^R transformants generated a single band, which had the size expected from the intact sequence or the Km-inserted one, respectively, (Fig. 3 and Fig. 4, lanes 1, 3, and 4). The DNA sample from the Ap^R Km^R transformant generated both of the two bands (Fig. 4, lane 2). This result is consistent with the conclusion that the *rhlE* gene is replaced by the corresponding region with Km^R-insertion in the Ap^S Km^R

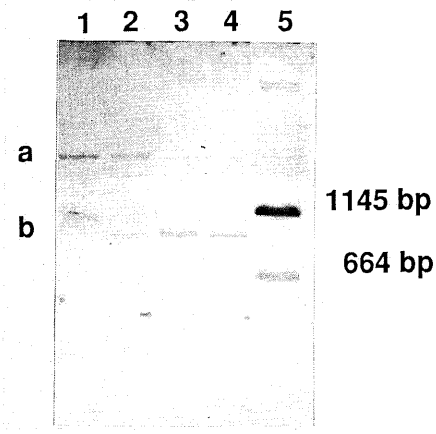

Fig. 4. Southern blotting analysis of two types of transformants. Chromosomal DNAs were prepared as described (Silhavy et al., 1984). Their *Pst*I digests were electrophoresed in an 1% agarose gel, transferred to a Zeta-Probe^{TM} blotting membrane (Bio-Rad), and hybridized with digoxigenin-labeled *Eco*RV fragment. The hybridized probes were detected with antidigoxigenin:alkaline phosphatase conjugate. Lane 1, JM101; lane 2, Ap^R Km^R transformant; lanes 3 and 4, two independent Ap^S Km^R transformants; lane 5, length markers (1145 and 664 bp fragment). The band a corresponds to the 1809 bp fragment from *Pst*I (at position 1) to *Pst*I (1810) and b corresponds to the 940 bp fragment from *Pst*I (1) to *Pst*I (the end of the kanamycin-resistance gene).

transformants, and the whole sequence of pKH6002 is integrated in the Ap^R Km^R transformant.

Since the above results do not necessarily exclude the possibility that the Ap^S Km^R transformants might carry a secondary mutation allowing the disruptants to survive, the *rhlE* gene disruption was examined by another method using a lambda phage vector with the temperature-sensitive repressor mutation *c*I857 (Silhavy et al., 1984). The whole sequence of pKH6002 linearized by *Eco*RI was inserted into λgt-λC (Thomas et al., 1974) to construct λ*ΔrhlE*::Km. JM101 strain lysogenized with this hybrid phage could be identified by the Ap^R Km^R and temperature-sensitive phenotype. Curing of the lysogenized phage by a short (1 min) heat pulse at 42°C generated temperature-resistant clones at approx. 10^{-2}. Nearly half of them were resistant to kanamycin and sensitive to ampicillin, implying that the *rhlE* gene can be disrupted without loss of viability, not accompanied by any suppressor mutation. Since the disruptants showed no difference from the original strain with regard to growth rate or colony size, the *rhlE* gene appears to be dispensable for cell growth under the normal conditions.

4. DISCUSSION

In this paper we have described mapping of the *rhlE* gene on the *E. coli* chromosome and determination of its complete sequence. We developed a new

method for gene disruption using a mutant of ColE1-type plasmid. The conventional methods for gene disruption in *E. coli* using a ColE1-type plasmid can be applied to only in special bacterial strains, because they take advantage of the fact that the vectors cannot be replicated in *polA*$^-$ strains (Gutterson and Koshland, 1983), or cannot be stably maintained in *recB*$^-$ *recC*$^-$ *sbcB*$^-$ (Winans et al., 1985) or in *recD*$^-$ strains (Russell et al., 1989). Contrarily, the plasmid constructed in this work cannot be replicated in the wild-type strains, although it can be maintained as a high-copy plasmid in *rnhA*$^-$ strains. The gene disruption method with this plasmid can be, therefore, applied to most of *E. coli* strains, even without separating the vector portion from the chromosomal region. If the gene in question were dispensable, transformation with the circular DNAs generates integration of the whole plasmid sequence into the chromosome predominantly and disruption/replacement less frequently. The relative frequency of the disruptants against the integrants can be increased by using the linearized DNAs for transformation, because integration becomes unfavorable unless the cleaved DNAs were religated within the cells. With this newly developed method, we were able to isolate a null mutant of the *rhlE* gene with no discernible growth defects. The *rhlE* gene might be just a remnant of an old gene which has lost its role during evolution, or its function might be exerted by one of the other D-E-A-D box genes in *E. coli*.

The *E. coli* chromosome carries many helicase genes. Most of the DNA helicase genes were first identified through isolating mutants defective in DNA replication, repair or recombination, and then their gene products were purified and shown to possess an activity to unwind duplex regions of DNA-DNA or DNA-RNA (Matson, 1991; and references therein). In contrast, studies on the putative RNA helicases encoded by the D-E-A-D box genes in *E. coli* are at present limited to isolating and sequencing of the genes. The *E. coli* strains in which *rhlB* or *rhlE* was disrupted were isolated, but they show no definite phenotype (Kalman et al., 1991; this work). The *srmB* and *deaD* genes, when overexpressed, suppress the temperature-sensitive mutation in the ribosomal protein L24 and S2, respectively, but their normal functions at a single copy level on the chromosome are obscure (Nishi et al., 1988; Toone et al., 1991). Since no functional complementation is observed between *srmB* and *deaD* or between *srmB* and *rhlB* (Kalman et al., 1991; Toone et al., 1991), these putative RNA helicases probably have specificity for RNA to act on, recognizing a primary sequence or secondary structure of substrate RNAs. Such specific interactions with RNA might be due to the carboxyl-terminal regions of the D-E-A-D box proteins, which differ in length and sequence but contain more basic amino-acids than acidic ones in general. Identification of such an RNA substrate and biochemical studies on the products of the D-E-A-D box genes will facilitate further understanding the functions of the *E. coli* RNA helicases.

I thank Dr. M. Cashel for informing me the DNA sequence of the *rhlE* PCR fragment, Drs. S. Kanaya, H. Inokuchi, and Y. Kohara for providing bacterial and phage strains, and also Dr. T. Nagata for improving the manuscript. This work was supported in part by Grants for the Priority Research Field 'Comprehensive Analysis of the *E. coli* Genome' from the Ministry of Education, Science and Culture, Japan.

REFERENCES

Bachmann, B. J. (1990). Linkage map of *Escherichia coli* K-12, Edition 8. *Microbiol. Rev.* **54**, 130–197.

Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D. and Frankel, D. (1991). Arginine-mediated RNA recognition. *Science* **252**, 1167–1171.

Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. and Blinov, V. M. (1989). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.* **17**, 4713–4730.

Gutterson, N. I. and Koshiland, D. E. Jr. (1983). Replacement and amplification of bacterial genes with sequences altered *in vitro*. *Proc. Natl. Acad. Sci. USA* **80**, 4894–4898.

Hay, B., Jan, L. Y. and Jan, Y. N. (1988). A protein component of Drosophila polar granules is encoded by *vasa* and has extensive sequence similarity to ATP-dependent helicases. *Cell* **55**, 577–587.

Higgins, D. G., McLaren, R. S. and Newbury, S. F. (1987). Repetitive extragenic palindromic sequences, mRNA stability and gene expression: evolution by gene conversion? A review. *Gene* **72**, 3–14.

Hirling, M., Scheffner, M., Restle, T. and Stahl, H. (1989). RNA helicase activity associated with the human p68 protein. *Nature* **339**, 562–564.

Iggo, R. D. and Lane, D. P. (1989). Nuclear protein p68 is an RNA-dependent ATPase. *EMBO J.* **8**, 1827–1831.

Iggo, R. D., Jamieson, D. J., MacNeill, S. A., Southgate, J., McPheat, J. and Lane, D. P. (1991). p68 RNA helicases: Identification of a nuclear form and cloning of related genes containing a conserved intron in yeasts. *Mol. Cell. Biol.* **11**, 1326–1333.

Iggo, R., Picksley, S., Souhgate, J., McPheat, J. and Lane, D. P. (1990). Identification of a putative RNA helicase in *E. coli*. *Nucleic Acids Res.* **18**, 5413–5417.

Jong, A. Y.-S., Clark, M. W., Gilbert, M., Oehm, A. and Campbell, J. L. (1987). *Saccharomyces cerevisiae* SSB1 protein and its relationship to nucleolar RNA-binding proteins. *Molec. Cell. Biol.* **7**, 2947–2955.

Kalman, M., Murphy, H. and Cashel, M. (1991). *rhlB*, a new *Escherichia coli* K-12 gene with an RNA helicase-like protein sequence motif, one of at least five such possible genes in a prokaryote. *New Biologist* **3**, 886–895.

Kanaya, S., Kohara, A., Miura, Y., Sekiguchi, A., Iwai, S., Inoue, H., Ohtsuka, E. and Ikehara, M. (1990). Identification of the aminoacid residues involved in an active site of *Escherichia coli* ribonuclease H by site-directed mutagenesis. *J. Biol. Chem.* **265**, 4615–4621.

Kohara, Y., Akiyama, K. and Isono, K. (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.

Kunkel, T. A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**, 488–492.

Lasko, P. F. and Ashburner, M. (1988). The product of the *Drosophila vasa* is very similar to eukaryotic initiation factor-4A. *Nature* **335**, 611–617.

Lazinski, D., Grzadzielska, E. and Das, A. (1989). Sequence-specific Recognition of RNA hairpins by

bacteriophage antiterminators requires a conserved arginine-rich motif. *Cell* **59**, 207–218.

Linder, P., Lasko, P. F., Ashburner, M., Leroy, P., Neilson, P. J., Nishi, K., Schnier, J. and Slominski, P. P. (1989). Birth of the D-E-A-D box. *Nature* **337**, 121–122.

Masukata, H. and Tomizawa, J. (1984). Effects of point mutations on formation and structure of RNA primer for ColE1 DNA replication. *Cell* **36**, 513–522.

Matson, S. W. (1991). DNA helicases of *Escherichia coli*. *Progress in Nucleic Acid Res. and Mol. Biol.* **40**, 289–326.

Naito, S., Kitani, T., Ogawa, T., Okazaki, T. and Uchida, H. (1984). *Escherichia coli* mutants suppressing replication-defective mutations of the ColE1 plasmid. *Proc. Natl. Acad. Sci. USA* **81**, 550–554.

Nishi, K., Morel-Deville, F., Hershey, J. W. B., Leighton, T. and Schnier, J. (1988). An eIF-4A-like protein is a suppressor of an *Escherichia coli* mutant defective in 50S ribosomal subunit assembly. *Nature* **336**, 496–498.

Ohmori, H. (1988). Structure and function of RNAII of ColE1. *Nucleic Acids Res.* **20A**, 131–132.

Ohmori, H., Murakami, Y. and Nagata, T. (1987). Nucleotide sequence required for a ColE1-type plasmid to replicate in *Escherichia coli* cells with or without RNase H. *J. Mol. Biol.* **198**, 223–234.

Ray, B. K., Lawson, T. G., Kramer, J. C., Cladaras, M. H., Grifo, J. A., Abramson, R. D., Merrick, W. C. and Thach, R. E. (1985). ATP-dependent unwinding of messenger RNA structure by eukaryotic initiation factors. *J. Biol. Chem.* **260**, 7651–7658.

Rudd., K. E. (1992). Alignment of *E. coli* DNA sequences to a revised, integrated genomic restriction map. In: *A Short Course in Bacterial Genetics* (ed.: J. H. Miller), Section 2A. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Russell., C. B., Thaler, D. S. and Dahlquist, F. W. (1989). Chromosomal transformation of *Escherichia coli recD* strains with linearized plasmids. *J. Bacteriol.* **171**, 2609–2613.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.

Schmid, S. R. and Linder, P. (1992). D-E-A-D protein family of putative RNA helicases. *Mol. Microbiol.* **6**, 283–292.

Silhavy, T. J., Berman, M. L. and Enquist, L. W. (1984). *Experiments with Gene Fusions.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Thomas, M., Cameron, J. R. and Davies, R. W. (1974). Viable molecular hybrids of bacteriophage lambda and eukaryotic DNA. *Proc. Natl. Acad. Sci. USA* **71**, 4579–4583.

Toone, W. M., Rudd, K. E. and Friesen, J. D. (1991). *deaD*, a new *Escherichia coli* gene encoding a presumed ATP-dependent RNA helicase, can suppress a mutation in *rpsB*, the gene encoding ribosomal protein S2. *J. Bacteriol.* **173**, 3291–3302.

Vieira, J. and Messing, J. (1987). Production of single-stranded plasmid DNA. *Methods in Enzymol.* **153**, 3–11.

Walker, J. E., Saraste, M., Runswick, M. J. and Gay, N. J. (1982). Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **8**, 945–951.

Winans, S. C., Elledge, S. J., Krueger, J. H. and Walker, G. C. (1985). Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*. *J. Bacteriol.* **161**, 1219–1221.

Yanish-Perron, C., Vieira, J. and Messing, J. (1985). Improved M13 phage vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.