JOURNAL OF BACTERIOLOGY, Feb. 1989, p. 636–642
0021-9193/89/020636-07$02.00/0

Vol. 171, No. 2

rhs Gene Family of *Escherichia coli* K-12

ALESIA B. SADOSKY, ANDREW DAVIDSON,† REN-JANG LIN,‡ AND CHARLES W. HILL*

Department of Biological Chemistry, The Milton S. Hershey Medical Center, The Pennsylvania State University, Hershey, Pennsylvania 17033

Received 29 August 1988/Accepted 26 October 1988

Two additional members of a novel *Escherichia coli* gene family, the *rhs* genes, have been cloned and characterized. The structures of these loci, *rhsC* and *rhsD*, have been compared with those of *rhsA* and *rhsB*. All four loci contain a homologous 3.7-kilobase-pair core. Sequence comparison of the first 300 nucleotides of the cores showed that *rhsA*, *rhsB*, and *rhsC* are closely related, with only 1 to 2% sequence divergence, whereas *rhsD* is 18% divergent from the others. The beginning of the core coincides with the initiation of an open reading frame that extends beyond the 300 nucleotides compared. Whether a protein product is produced from this open reading frame has not been established. However, nucleotide substitutions which differentiate the cores have highly conservative effects on the predicted protein products; this suggests that products are made from the open reading frame and are under severe selection. The four *rhs* loci have been placed on both the genetic and restriction maps of *E. coli* K-12. A fifth *rhs* locus remains to be characterized. In terms of size, number, and sequence conservation, the *rhs* genes make up one of the most significant repetitions in *E. coli*, comparable to the rRNA operons.

The first *rhs* genes to be discovered were involved in the *glyS* gene amplification that occurs in *Escherichia coli* K-12 (18). This amplification results from a *recA*-dependent, unequal crossover between *rhsA* and *rhsB*; these loci provide large, direct repeats flanking the amplified segment. This recurring phenomenon prompted us to name the repeated sequences *rhs*, for rearrangement hot spot. The *rhsA* and *rhsB* loci share highly conserved, 3.7-kilobase-pair (kb) core regions which are sufficiently similar to form an S1 endonuclease-resistant heteroduplex (18). Additional members of the *rhs* family were inferred from Southern analysis by probing genomic DNA with a portion of this core. At least one other gene cross-hybridized strongly with *rhsA* and *rhsB*, with additional genes hybridizing to a much lesser extent. This pattern suggested the presence of two *rhs* subfamilies: one consisting of *rhsA*, *rhsB*, and *rhsC*, and a second, more divergent one containing *rhsD* and possibly other genes.

The *rhs* gene family is particularly interesting because of the high degree of sequence conservation maintained among its members. Other examples of highly conserved families in *E. coli* include the rRNA operons, some tRNA genes, and certain multicopy insertion sequences (21). In addition, there are sets of protein genes which show homology within a set (reviewed in reference 20). However, with the exception of the *tufA-tufB* gene pair, these sets show much more sequence divergence. The major objective of this work was to investigate the apparent interrelatedness within the *rhs* family. We report the cloning and mapping of two new members of the *rhs* family, *rhsC* and *rhsD*. In addition, their structure is compared with that of *rhsA* and *rhsB*.

MATERIALS AND METHODS

Bacterial strains. The derivations of several *E. coli* K-12 mutants used have been described before: W3102 (2), AT1325 *lip-9* (14), ER (8), AX727 (10), CH931 (11), CH1480 (5), and CH1505 (17). CH1330 is a *polA1 argH* derivative of HfrC; it was prepared from PB153 (15) by a series of P1 cotransductions introducing successively the following sets of markers: TrpA$^+$ *tyrT*$^+$; Met$^+$ *argH*; *metE zig-1274*::Tn*10* *polA1*; Met$^+$ Tet^s. CH1592 is CH1330 transformed with pRL390. CH1555 is CH931 transformed with pJG1554. CH3113 is a *rhsD*::Kan^r *purE55* transductant of the *dnaZ2016*(Ts) strain, AX727. CH1698 is a *rhsC*::Kan^r transductant of the *asnA31 asnB32* strain, ER. CH1513 is a *polA1 mtl* derivative of W3102; the same strategy was used to introduce *polA1* into CH1513 as was used to prepare CH1330.

Plasmids. The vectors used for cloning were pBR322 (3), pBR325 (4) and pUC19 (19). The Kan^r determinant used for modifying the *rhsC* and *rhsD* loci was a *Kpn*I fragment derived from pIF11 and contains Tn*903* (7). Recombinant plasmids were isolated and maintained in the *recA56* strain CH1480 unless otherwise specified. Preparation of the *E. coli* *Hin*dIII-*Sal*I genomic library, derived from the *glyS-glyS* duplication mutant CH1505 and cloned into pBR325, has been described previously (18). pRL400 was prepared by subcloning the 3.2-kb *Pvu*II core fragment from *rhsA/B*. The *rhsA/B* locus is a hybrid gene produced by recombination between *rhsA* and *rhsB* and was taken from the plasmid pRL351 (18). The vector used to prepare pRL400 was the *Pvu*II fragment that contains the replication origin of pBR325.

Microbial genetic procedures. Growth media and procedures for conjugation, P1 transduction, and transformation have been specified previously (11). The following method was used to prepare a mutant which had the Kan^r determinant inserted into the *rhsC* gene. The *Kpn*I fragment containing Kan^r was isolated from pIF11 and ligated with a *Kpn*I digest of the *rhsC* plasmid pJG1626. The result was the deletion of the portion of *rhsC* between the internal *Kpn*I sites and its replacement with Kan^r; the disrupted gene was designated *rhsC*::Kan^r. This plasmid, pJG1672, was used to transform the *polA1* strain CH1330. Previous work has shown that successful transformation of a *polA1* recipient requires integration of the plasmid into the recipient chro-

* Corresponding author.
† Present address: Department of Internal Medicine, University of Florida, Gainesville, FL 32610.
‡ Present address: Department of Microbiology, University of Texas, Austin, TX 78712.

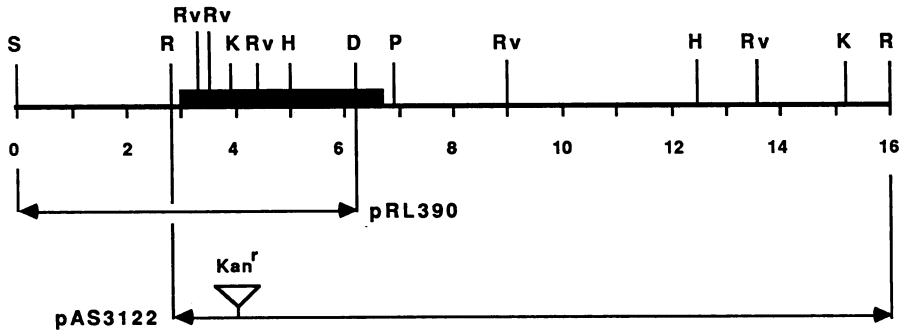

FIG. 1. Restriction enzyme map of the *rhsD* region, indicating the extent of recombinant plasmid inserts. Symbols: ▬, *rhsD* core; ▽, insertion position of the Kan^r gene into a *Kpn*I site. Coordinates are in kilobases. Restriction site designations: D, *Hin*dIII; H, *Hpa*I; K, *Kpn*I; P, *Pst*I; R, *Eco*RI; Rv, *Eco*RV; S, *Sal*I. There are no *Bam*HI sites in this region.

mosome (11). This integration depends on the homology between the plasmid insert and the chromosome. In this particular case, the gene sequence in the transformant would become *rhsC*$^+$ Amp^r *rhsC*::Kan^r. However, this structure is unstable, and Amp^s segregants are frequently produced. Some of these, such as CH1681, retain the *rhsC*::Kan^r allele, but lose *rhsC*$^+$ along with the Amp^r vector. An analogous approach was used to produce an *rhsD*::Kan^r mutant. In this case the Kan^r determinant was inserted into the single *Kpn*I site of pRL390, producing pAS3101. Exchange of *rhsD*:: Kan^r for the chromosomal *rhsD*$^+$ of CH1330 produced CH3102.

DNA isolation, restriction enzyme analysis, and sequencing. Procedures for the isolation of plasmid (6) and genomic DNA, as well as conditions for restriction endonuclease digestion, ligation, and gel electrophoresis (5) were specified previously. Procedures for subcloning fragments into pUC19, isolating overlapping deletions, and chemical sequencing were as specified previously (6). All nucleotide sequences reported were derived by independently determining the sequence of both DNA strands.

RESULTS

Cloning of *rhsD*. Lin et al. (18) prepared an *E. coli* DNA library of *Hin*dIII-*Sal*I fragments in pBR325 and screened it for clones sharing homology with *rhsA*. One recombinant plasmid, pRL390, contained a 6.3-kb insert identical in size to one from genomic DNA which cross-hybridizes weakly with *rhsA*. After preliminary characterization, it became clear that this fragment contained part but not all of *rhsD*, and so an additional, overlapping clone was sought. To facilitate the identification of this clone, we inserted a *Kpn*I fragment encoding Kan^r into the *Kpn*I site of pRL390. This modified *rhsD* locus was then transferred to the *E. coli* chromosome by recombination in a process that was dependent on the homology of the DNA flanking the Kan^r determinant (see Materials and Methods). The *rhsD*::Kan^r gene was then recloned as an *Eco*RI fragment in pBR322, producing pAS3122. A restriction map of the region and of the recombinant plasmids is shown in Fig. 1.

Cloning of *rhsC*. The results of Lin et al. (18) suggested that *rhsC*, the *rhs* locus showing the strongest cross-hybridization to *rhsA* and *rhsB*, was contained on a large (>30-kb) *Hin*dIII-*Sal*I fragment. However, we were unable to obtain this fragment when the collection of cloned *rhs*-homologous fragments was screened. Therefore, the *rhsC* locus was cloned by a different strategy (outlined in Fig. 2a). The first step involved transformation of CH931, a *polA1* derivative of HfrH, with plasmid pRL400. Since pBR325-derived plasmids cannot replicate in a *polA1* background, the only Amp^r transformants obtained are those in which the plasmid integrates into the host chromosome (11). This can happen by recombination between the plasmid insert and a homologous region of the chromosome. Since the insert of pRL400 consists of a 3.2-kb *Pvu*II fragment (Fig. 2a) from the *rhsA/B* core region, it was anticipated that the plasmid could integrate into *rhsC* as well as into *rhsA* or *rhsB*. A preliminary experiment suggested that *rhsC* mapped far from *rhsA* and *rhsB* on the *E. coli* chromosome. Since *rhsA* and *rhsB* were known to be transferred late by HfrH, we attempted to eliminate transformants in which the integration event had occurred at either *rhsA* or *rhsB* by screening the transformants for early transfer of the plasmid Amp^r marker to the *polA1* recipient CH1513. One recombinant, CH1530, was verified by Southern analysis as having pRL400 inserted into *rhsC* (data not shown). Next, a larger plasmid, containing chromosomal DNA flanking the integrated plasmid, was created by digesting CH1530 genomic DNA with *Hin*dIII and converting the fragments to circles by enzymatic ligation. The expectation was that only the fragment containing the inserted vector could become an autonomously replicating Amp^r plasmid. Although the resulting plasmid, pJG1544, contained *rhsC*, it was of limited immediate use because it also contained material derived from *rhsA/B*. Therefore, a 1.1-kb *Bam*HI-*Eco*RI fragment (Fig. 2a) was subcloned from pJG1544 into pBR325. This plasmid, pJG1554, was in turn used to transform a *polA1* recipient. In this case the only homology available for integration was the homologous segment adjacent to *rhsC*. In the next step, an *rhsC*-containing plasmid, pJG1568, was produced by digesting genomic DNA from a transformant, CH1555, with *Hin*dIII and circularizing the fragments by ligation. The *rhsC* subclone, pJG1626, consisting of a 15.1-kb *Ava*I-*Cla*I fragment spanning *rhsC* (Fig. 2b), was used for most of the characterization of *rhsC*.

Comparison of the four *rhs* loci. The *rhsC* and *rhsD* loci were characterized by restriction enzyme analysis and by partial sequence analysis. The restriction maps of these two loci were compared with those previously determined for *rhsA* and *rhsB*. The four loci have been aligned in Fig. 3 according to their major homologies. From an earlier study, we knew that *rhsA* and *rhsB* shared a 3.7-kb region of homology within which their restriction maps were identical (18). These core regions are indicated in Fig. 3. Inspection of the *rhsC* region revealed that this restriction pattern was

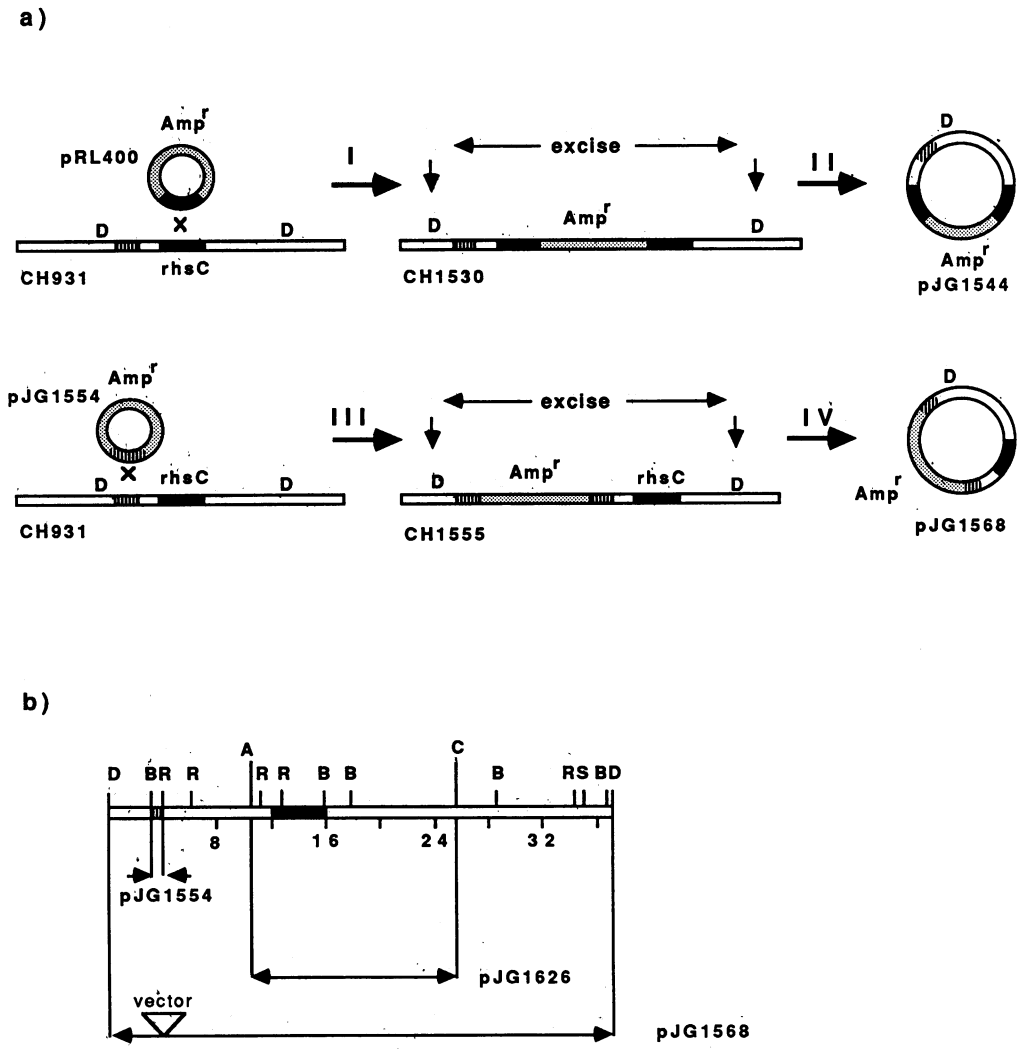

FIG. 2. Isolation and restriction mapping of the *rhsC* region. (a) Schematic representation of the strategy used to clone *rhsC*. Steps I and III involved the in vivo insertion of plasmids into the *polA1* recipient by recombination between homologous sequences. The plasmid used in step I, pRL400, contains a 3.2-kb insert derived from the *rhsA/B* core (see Materials and Methods). Steps II and IV involved the in vitro digestion of genomic DNA with *Hin*dIII and circularization of the fragments with T4 DNA ligase. The resulting plasmids were recovered by transformation of a *recA56* recipient, CH1480, selecting for Amp^r. Symbols: ■, portion of the *rhs* core region bounded by *Pvu*II sites (Fig. 3); ▥, segment bounded by *Bam*HI and *Eco*RI sites; ▩, pBR325 DNA; □, flanking chromosomal DNA. (b) Restriction map of the *rhsC* region. Symbol: ■, *rhsC* core. The extent of plasmid inserts is shown below the map. Both plasmids used pBR325 as vector. Restriction site designations: A, *Ava*I; B, *Bam*HI; C, *Cla*I; D, *Hin*dIII; R, *Eco*RI; S, *Sal*I. Only *Ava*I and *Cla*I sites used in constructing pJG1626 are shown; others may exist to the left of the *Ava*I sites or to the right of the *Cla*I sites. Coordinates are in kilobases.

present. This degree of conservation was consistent with the strong cross-hybridization observed between *rhsA*, *rhsB*, and *rhsC* (18). In contrast, none of the restriction sites were conserved in *rhsD*. This result was consistent with the weak cross-hybridization between *rhsD* and *rhsA* (18). Since no restriction sites were conserved, the alignment of *rhsD* with the other three loci required nucleotide sequence analysis (see below).

To determine more exactly the nature of the homology shared by the *rhs* family, we have determined the nucleotide sequence at the borders of the homology regions. The sequences at the left end of the homologies are shown in Fig. 4. Since a 3.7-kb, S1 endonuclease-resistant heteroduplex can be formed between *rhsA* and *rhsB* DNA, we expected to find a region of virtually identical sequence, at least for these two. Inspection of the sequences shown in Fig. 4 shows that the homology shared by all four *rhs* loci begins at nucleotide 101 and extends to the right. Interestingly, nucleotide 101 initiates an open reading frame that extends through the remaining 300 nucleotides of the depicted sequence. The sequence of *rhsA* and *rhsB* diverge at only 4 of these 300 nucleotides, whereas *rhsC* differs from each of them by 7 of 300 nucleotides. The homology between *rhsD* and the other three also begins with this potential start codon, but its sequence diverges from the others at 49 to 54 of the 300 positions. In addition, *rhsA* and *rhsC*, but not *rhsB* and *rhsD*, are highly homologous over the 118 nucleotides preceding the open reading frame (only 100 of these are shown in Fig. 4). We have defined nucleotide 101 in Fig. 4 as the beginning of the *rhs* core.

The other end of the core was located by sequence comparison of *rhsA*, *rhsC*, and *rhsD* (Fig. 5). The sequences

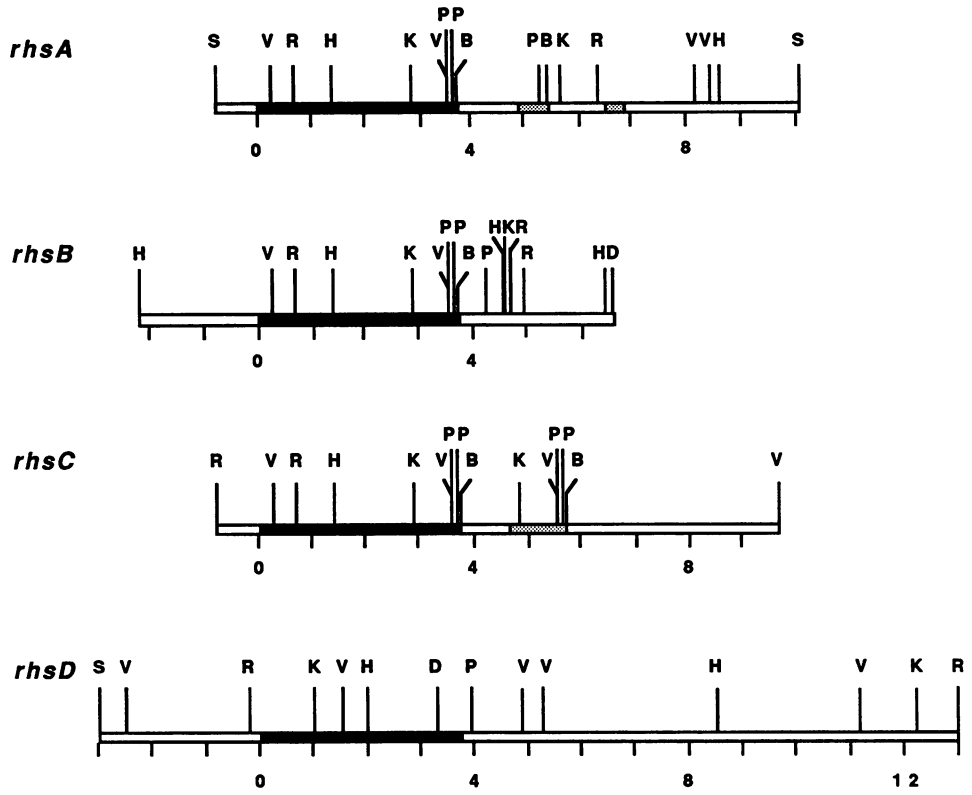

FIG. 3. Comparison of the restriction maps of the four *rhs* loci. Symbols: ■, *rhs* core homologies; ▦, partial repetitions of the core homologies. The restriction maps of *rhsA* and *rhsB* have been published previously (18). Restriction site designations: B, *Bam*HI; D, *Hin*dIII; H, *Hpa*I; K, *Kpn*I; P, *Pst*I; R, *Eco*RI; S, *Sal*I; V, *Pvu*II. Coordinates are in kilobases.

shown are the 140 nucleotides beginning with the conserved *Bam*HI site at the right end of the *rhsA* and *rhsC* cores (Fig. 3). Through the first 54 positions, the *rhsA* and *rhsC* sequences are identical except for 1 position. In keeping with the greater divergence expected for *rhsD*, it differs from *rhsA* at 7 of these 54 positions. We have chosen to designate position 55 as the end of the core.

Repetition of portions of the core sequence. In our previous report (18), we noted cross-hybridization between nonoverlapping restriction fragments from the *rhsA* region, suggesting tandem sequence homology within *rhsA*. A similar situation has been found in *rhsC*. For *rhsC*, the finding of five restriction sites (Fig. 3) identical in spacing to sites at the end of the core strongly suggested that the terminal portion of this core was repeated. Nucleotide sequencing has confirmed that the last 1,020 nucleotides of the core are repeated at the position indicated in *rhsC* (C. Hill and J. Gray, unpublished results). Sequence analysis has also shown that the terminal portion of the core is repeated twice in *rhsA* (Fig. 3). The first repeat contains the last 430 nucleotides, whereas the second contains the last 160 nucleotides. The nucleotide sequence at the end of the first repeat is compared with that of the *rhs* cores in Fig. 5, where it is designated *rhsA'*.

Mapping of the *rhs* loci. Preliminary mapping of the *rhs* loci was done by interrupted mating. Amp^r plasmids carrying inserts derived from either *rhsC* or *rhsD* were used to transform *polA1* Hfr strains. This procedure produces transformants in which the plasmid has integrated into the host chromosome through recombination between the plasmid insert and the homologous segment of the chromosome (Materials and Methods). Southern analysis verified that insertion had occurred at the appropriate chromosomal location. Since the *rhs* loci have no known phenotype, the Amp^r of the integrated plasmid was used as a selectable marker in conjugation. The time of entry during conjugal transfer by an HfrH derivative, CH1555, showed that *rhsC* was linked to the *lip* locus. *rhsD* was transferred very early by an HfrC derivative, CH1592, and was shown to be linked to *purE*.

To map these loci more precisely, we inserted a *Kpn*I fragment encoding Kan^r into both *rhsC* and *rhsD* by recombinant techniques. These modified genes were substituted for their respective chromosomal copies by recombinational exchange (Materials and Methods) and used in P1 cotransduction experiments (Fig. 6). The results placed *rhsC* near min 16, between *asnB* and *gal*, whereas *rhsD* mapped near min 12, between *dnaZ* and *purE*.

Comparison of plasmids containing *rhsC* with published restriction maps has allowed precise placement of *rhsC* with respect to nearby loci. pJG1568 overlaps the *phr* plasmid, pKY1 (22), and the *gltA* plasmid, pLC31-28 (12). This places the *phr* locus near coordinate 21 of the *rhsC* map in Fig. 2b and places *gltA* near coordinate 34. In addition, the 37-kb *Hin*dIII insert of pJG1568 is essentially identical to that of the *Hin*dIII interval found by Kohara et al. (16) at coordinates 733 to 769 of their *E. coli* restriction map. We could also place *rhsD* on the *E. coli* restriction map by taking into account our genetic and restriction mapping data along with those of Hadley et al. (13). The 13.2-kb *Eco*RI interval containing *rhsD* (Fig. 3) corresponds to coordinates 535 to 548 of the *E. coli* restriction map (16). Mapping results are summarized in Table 1.

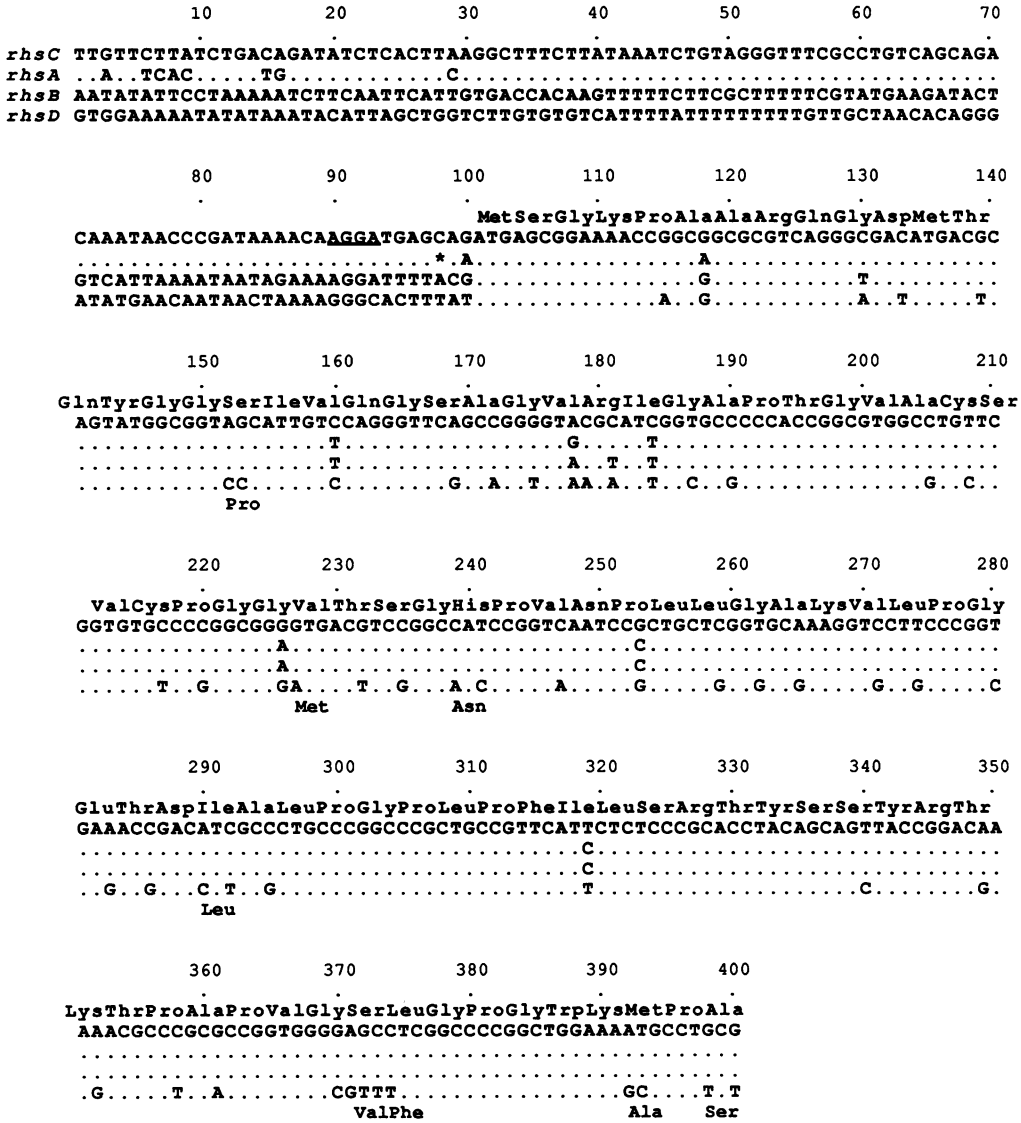

FIG. 4. Comparison of the nucleotide sequence at the beginning of the four *rhs* cores. Nucleotide 101 initiates an open reading frame that extends throughout the depicted sequence for all four loci. Translation of this open reading frame for *rhsA*, *rhsB*, and *rhsC* is shown above the DNA sequence, and the amino acid substitutions predicted from the *rhsD* sequence are shown below. Symbols: *, single nucleotide gap introduced into the *rhsA* sequence to permit its optimum alignment with the *rhsC* sequence; · , the base at this position is identical to the one shown above for *rhsC*. A potential ribosome-binding sequence is underlined.

DISCUSSION

One goal of this work was to extend our knowledge of the *rhs* gene family as regards the number and map positions of different *rhs* loci in *E. coli* K-12. Previous observations suggested that there was at least one (possibly several) *rhs* locus in addition to *rhsA* and *rhsB* (18). Reconsidering these earlier results in light of our knowledge of *rhsC* and *rhsD*, it is clear that at least one additional locus remains to be characterized. It is represented by a 1-kb *Hin*dIII-*Sal*I fragment which cross-hybridizes with a core-specific probe yet is clearly distinct from *rhsA*, *rhsB*, *rhsC*, or *rhsD*. This fragment has been isolated, and preliminary characterization suggests that its source is more closely related to *rhsD* than to *rhsA*, *rhsB*, or *rhsC* (G. Feulner, A. Sadosky, and C. Hill, unpublished observations).

From these and earlier results (18), we now know the positions of four *rhs* genes (Table 1). Although this mapping was accomplished by traditional microbial genetic techniques, all four loci can be further linked to reference loci by restriction analysis of the cloned genes. This was described for *rhsC* and *rhsD* in Results. The restriction pattern of the original clone containing *rhsB*, pRL389 (18), coincides with that of *Hin*dIII fragment 4087 to 4170 of the *E. coli* restriction map (16). In addition, the restriction map of pRL389 overlaps that of the *pit* plasmid, p5-2 (9). Similarly, the pattern of the original clone of *rhsA*, pRL276 (18), coincides with that of *Hin*dIII fragment 3939 to 3968 (16). The restriction map of pRL276 overlaps that of the *mtl* plasmid pLC15-48 (17). The positions of the *rhs* core regions are listed in Table 1. We should also note that all four *rhs* loci have the same orientation within the chromosome, the orientation shown in Fig. 3 corresponding to the clockwise direction of the *E. coli* genetic map. Since the section of the restriction map covering *rhsA* and *rhsB* is inverted with respect to the

```
        BamHI      10         20         30         40         50         60         70
                    .          .          .          .          .          .          .
rhsA    GGATCCGATTGGGCTGAAGGGGGGATGGAATTTTTATCAGTATCCGTTGAATCCAGTTACGAATACAGAT
rhsC    ..................A.....................................GA.CT.A....T....
rhsA'   ............A.....G.....................................G..AT....G.....
rhsD    ...C.....G...T....AG............A..........T..A.....AC.ACA.C.AA.T..C

                   80         90         100        110        120        130        140
                    .          .          .          .          .          .          .
        CCTCTGGGGTTAGAAGTTTTTCCTAGACCATTCCCCTTGCCAATTCCATGGCCCAAAAGCCCTGCACAGC
        ..AT.A..A......ACAC.AAAAT.CATTAAG..AC...ATTCAATGG.CGGA.CTG.TGAAAG.AGCG
        ..GCA...T..G.TTGATA....T.TATACCCCGA.AGTG..CTT..CCATTCTGTA.C...TGAG.TTA
        ..TATG..A..AT.GC.G.CTTGGGATG.TG..AG.TC..GAGCATGTACGGGG.G..T.TG..GTG..C
```

FIG. 5. Comparison of the nucleotide sequence at the end of the *rhsA*, *rhsC*, and *rhsD* core regions. The sequence designated *rhsA'* is the end of the first partial core repetition in *rhsA* (Fig. 3). The first six nucleotides constitute the conserved *Bam*HI site at the end of the *rhsA* and *rhsC* cores and of the *rhsA* repeat. The *Pst*I site in the *rhsD* region (Fig. 3) lies 37 nucleotides beyond the sequence shown. Symbol: ·, the base at this position is identical to the one immediately above.

genetic map (16), these two loci have a counterclockwise orientation on the restriction map.

Previous hybridization studies of cloned *rhsA* and *rhsB* suggested the existence of a 3.7-kb (originally estimated as 3.8-kb) core homology region shared by these loci (18). Sequence comparison of the four *rhs* genes allows a more precise definition of this core to be made. The homology shared by all four begins at nucleotide 101 in Fig. 4, and this position has been defined as the beginning of the core. The *Bam*HI site conserved in *rhsA*, *rhsB*, and *rhsC* (Fig. 3) lies 3,660 base pairs from the beginning of the core. Sequence comparison of *rhsA*, *rhsC*, and *rhsD* cores showed that the strong sequence similarity ends 55 nucleotides beyond this *Bam*HI site (Fig. 5). Therefore, the cores are 3,714 nucleotides in length. The sequence of the entire *rhsA* core has been completed (Hill and Gray, unpublished), confirming this size estimate. The sequence of the partial repetition of the core found in *rhsA*, designated *rhsA'* in Fig. 5, also diverges from the *rhsA* and *rhsC* cores at this same position. We have chosen to designate nucleotide 55 in Fig. 5 as the end of the core, although we note significant similarity in the sequences of *rhsA*, *rhsC*, *rhsD*, and *rhsA'* over the next 27 nucleotides. It is striking that within the region 68 to 81, four dinucleotides are conserved in all four sequences. Preliminary sequencing, based on data from one strand only, has shown that a similar divergence occurs at this position for *rhsB*, as well as at the end of the other partial repetitions found to the right of the *rhsA* and *rhsC* cores (Fig. 3). The existence of multiple occurrences of a common sequence diverging into various unique sequences poses interesting evolutionary questions about the establishment of these junctions. One possibility is that there is a site-specific recombination mechanism that joins the end of the cores to unique sequences.

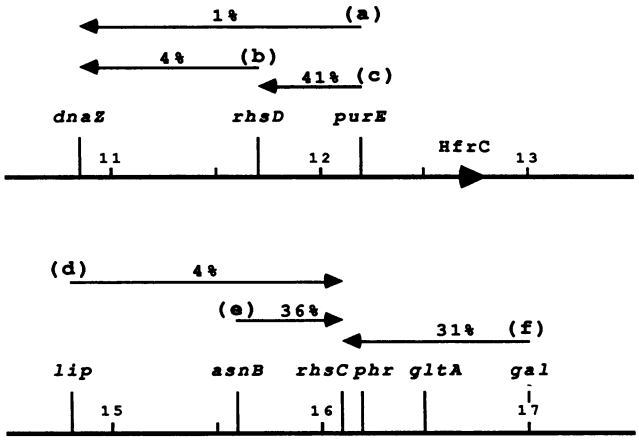

FIG. 6. Genetic mapping of *rhsC* and *rhsD*. Cotransduction frequencies of *rhs* genes with reference markers are shown above the arrows. In each cross, the gene situated at the butt of an arrow was initially used for selection, while the head represents the scored markers. The *rhs* loci were marked by the insertion of a Kan^r determinant into each. The scale in minutes and the positions of reference loci are from Bachmann (1); the location of the HfrC transfer origin is from Hadley et al. (13). The donor and recipient cultures for each cross along with relevant markers were as follows: (a) W3102 and CH3113 (*dnaZ purE*); (b) CH3113 (*dnaZ rhsD*::Kan^r) and W3102; (c) W3102 and CH3113 (*rhsD*::Kan^r *purE*); (d) CH1681 (*rhsC*::Kan^r) and AT1325 (*lip-9*); (e) CH1698 (*rhsC*::Kan^r) and ER (*asnB*); (f) CH1698 (*rhsC*::Kan^r) and W3102 (*gal*). The relative placements of *rhsC* and *phr* were based on the comparison of our restriction map with that published for pKY1 (22), and the placement of *gltA* was based on a comparison with the restriction map published for pLC31-28 (12).

The *rhs* loci are complex genetic elements and are clearly larger than the 3.7-kb core regions. Minimally, we consider them to include the repetitions found in *rhsA* and *rhsC* to the right of the cores (Fig. 3) and the 118 nucleotides preceding the *rhsA* and *rhsC* cores which differ at only 10 positions. Recent unpublished results lead us to believe that the *rhsA* locus is actually 8.3 kb in length (G. Feulner and C. Hill, unpublished observations).

An intriguing aspect shared by the four *rhs* genes is that a methionine codon is positioned at the junction of the core

TABLE 1. Locations of the *rhs* genes

Locus	Genetic map location (min)[a]	Restriction map location (kb)[b]
rhsA	*mtl* (80.6)	3,949
rhsB	*pit* (76.5)	4,093
rhsC	*phr* (16.1)	743
rhsD	*purE* (11.7)	535

[a] The closest reference locus is listed for each *rhs* locus, along with the genetic map position (1) for the *rhs* locus. Estimates of map positions are based on both genetic and restriction mapping as explained in the text.

[b] The coordinate on the *E. coli* restriction map (16) for the beginning of each *rhs* core is listed.

homology. This start codon initiates an open reading frame that in turn extends at least 300 nucleotides (Fig. 4). In addition, there is a good ribosome-binding site preceding the methionine codon (Fig. 4). The predicted peptide sequence is shown above the nucleotide sequence in Fig. 4. Whether this open reading frame is used to make a protein product has not been established. However, certain circumstances suggest that it does. Particularly, the nucleotide substitutions in the sequenced portion of the four cores are highly conservative. In fact, none of the eight nucleotide substitutions that differentiate *rhsA*, *rhsB*, and *rhsC* cause a change in the predicted amino acid sequence. The 49 to 54 nucleotide substitutions which differentiate the *rhsD* core from the other three would produce only eight amino acid changes (shown beneath the *rhsD* sequence in Fig. 4). The implication of this conservative divergence is that the protein product of each of the four genes has been under selection through evolution. When examined after Southern analysis, the *rhs* hybridization patterns of *E. coli* B/5 and C are very similar to that of K-12 (18), a fact we interpret as indicating that the *rhs* genes are relatively stable components of the *E. coli* chromosome. If this is true, it suggests that the functions of the four loci are not redundant but rather are simultaneously and independently under selection. It should be emphasized that whatever function the *rhs* cores may have, their individual integrity is not essential, since disruptions of *rhsC* and *rhsD*, constructed in the course of this work, had no observable effect on cell growth.

ACKNOWLEDGMENTS

We thank Jane Gray for competent technical assistance, Barbara Bachmann and James Walker for bacterial strains, and Mike Capage for a sample of pIF11.

This work was supported by Public Health Service grant GM 16329 from the National Institutes of Health.

ADDENDUM IN PROOF

The nucleotide sequence data reported in this paper have been submitted to the EMBL, GenBank, and DDBJ nucleotide sequence data bases and have been assigned the accession number J04224.

LITERATURE CITED

1. **Bachmann, B. J.** 1983. Linkage map of *Escherichia coli* K-12, edition 7. Microbiol. Rev. **47**:180–230.
2. **Bachmann, B. J.** 1987. Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12, p. 1190–1219. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
3. **Bolivar, F.** 1978. Construction and characterization of new cloning vehicles. III. Derivatives of plasmid pBR322 carrying unique *Eco* RI sites for selection of *Eco* RI generated recombinant DNA molecules. Gene **4**:121–136.
4. **Bolivar, F., R. L. Rodriquez, P. J. Greene, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow.** 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene **2**:95–113.
5. **Brody, H., A. Greener, and C. W. Hill.** 1985. Excision and reintegration of the *Escherichia coli* K-12 chromosomal element e14. J. Bacteriol. **161**:1112–1117.
6. **Brody, H., and C. W. Hill.** 1988. Attachment site of the genetic element e14. J. Bacteriol. **170**:2040–2044.
7. **Capage, M., and J. R. Scott.** 1983. SOS induction of P1 Km miniplasmids. J. Bacteriol. **155**:473–480.
8. **Cedar, H., and J. H. Schwartz.** 1969. The asparagine synthetase of *Escherichia coli*. J. Biol. Chem. **244**:4112–4121.
9. **Elvin, C. M., N. E. Dixon, and H. Rosenberg.** 1986. Molecular cloning of the phosphate (inorganic) transport (*pit*) gene of *Escherichia coli* K12. Mol. Gen. Genet. **204**:477–484.
10. **Filip, C. C., J. S. Allen, R. A. Gustafson, R. G. Allen, and J. R. Walker.** 1974. Bacterial cell division regulation: characterization of the *dnaH* locus of *Escherichia coli*. J. Bacteriol. **119**: 443–449.
11. **Greener, A., and C. W. Hill.** 1980. Identification of a novel genetic element in *Escherichia coli* K-12. J. Bacteriol. **144**: 312–321.
12. **Guest, J. R.** 1981. Hybrid plasmids containing the citrate synthase gene (*gltA*) of *Escherichia coli* K-12. J. Gen. Microbiol. **124**:17–23.
13. **Hadley, R. G., M. Hu, M. Timmons, K. Yun, and R. C. Deonier.** 1983. A partial restriction map of the *proA-purE* region of the *Escherichia coli* K-12 chromosome. Gene **22**:281–287.
14. **Herbert, A. A., and J. R. Guest.** 1968. Biochemical and genetic studies with lysine + methionine mutants of *Escherichia coli*: lipoic acid and α-ketoglutarate dehydrogenase-less mutants. J. Gen. Microbiol. **53**:363–381.
15. **Hill, C. W., J. Foulds, L. Soll, and P. Berg.** 1969. Instability of a missense suppressor resulting from a duplication of genetic material. J. Mol. Biol. **39**:563–581.
16. **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.
17. **Lee, C. A., and M. H. Saier, Jr.** 1983. Use of cloned *mtl* genes of *Escherichia coli* to introduce *mtl* deletion mutations into the chromosome. J. Bacteriol. **153**:685–692.
18. **Lin, R.-J., M. Capage, and C. W. Hill.** 1984. A repetitive DNA sequence, *rhs*, responsible for duplications within the *Escherichia coli* K-12 chromosome. J. Mol. Biol. **177**:1–18.
19. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene **26**:101–106.
20. **Riley, M.** 1984. Arrangement and rearrangement of bacterial genomes, p. 285–315. *In* R. P. Mortlock (ed.), Microorganisms as model systems for studying evolution. Plenum Publishing Corp., New York.
21. **Riley, M., and S. Krawiec.** 1987. Genome organization, p. 967–981. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
22. **Yamamoto, K., M. Satake, H. Shinagawa, and Y. Fujiwara.** 1983. Amelioration of the ultraviolet sensitivity of an *Escherichia coli recA* mutant in the dark by photoreactivating enzyme. Mol. Gen. Genet. **190**:511–515.