JOURNAL OF BACTERIOLOGY, Mar. 1995, p. 1393–1398
0021-9193/95/$04.00+0

Vol. 177, No. 5

Reshuffling of *Rhs* Components To Create a New Element

SHENG ZHAO AND CHARLES W. HILL*

Department of Biochemistry and Molecular Biology, Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033

Received 23 September 1994/Accepted 24 December 1994

***RhsF* has been identified as the fourth member of the *RhsABCF* subfamily of genetic elements. This new element is found in *Escherichia coli* ECOR-50 and several other strains but not in strain K-12. A novel feature of *RhsF* is that it represents a new arrangement of components previously uniquely associated with *RhsA* and *RhsC* of strain K-12.**

The *Rhs* elements comprise a family of large, composite accessory sequences. These elements are found at several chromosomal locations in *Escherichia coli* K-12 (9) (Fig. 1). Each element contains several distinct components, some of which are highly conserved with respect to other members of the *Rhs* family and some of which are highly divergent or even unique. These components are identified in Fig. 2A, which depicts an idealized *Rhs* element. The *Rhs* elements are considered accessory genetic elements because not all natural *E. coli* cells have them, and their components have G+C contents that are very atypical of *E. coli* genes. The most prominent *Rhs* component is a conserved 3.7-kb core that maintains a single open reading frame (ORF) throughout its length. This large ORF does not terminate at the boundary of the core homology but rather continues into a region of high sequence divergence for another 139 to 177 codons. The transition from core to core extension is marked by an abrupt change in G+C content (Fig. 2A). The *Rhs* core sequence predicts a large protein containing 28 repetitions of a peptide motif that can be written xxGxxRYxYDxxGRL(I or T)xxxx (3). The *Rhs* core peptide motif is similar to a motif that appears 31 times in a wall-associated protein of *Bacillus subtilis* (4). This and other considerations led to the conjecture that the *Rhs* core protein is a component of the cell surface (9), but this has not been established experimentally. *E. coli* K-12 contains five *Rhs* elements that have been placed in two subfamilies according to the sequence divergence of their cores. The *RhsABC* and the *RhsDE* subfamilies diverge about 22% at the nucleotide level, while divergence within the subfamilies is limited to between 1 and 4% (18, 23). The two K-12 subfamilies are readily distinguished by the strength of the hybridization signal produced with a core-specific probe (11): *RhsA*, *RhsB*, and *RhsC* cores strongly cross-hybridize, but they weakly hybridize with *RhsD* and *RhsE* and vice versa. The *Rhs* elements commonly contain putative insertion sequences such as H-rpt (see below). Our concept of the *Rhs* elements is that they are composite structures assembled from discrete components that have had independent evolutionary origins outside of the *E. coli* species.

An ongoing project of this laboratory concerns the *Rhs* profiles of natural *E. coli* isolates, particularly the ECOR collection (15). Many of these independent strains have been found to possess one or more of the same elements as found in strain K-12. (The identity of an *Rhs* element is determined by its chromosomal location, not necessarily by its structure.) We have observed that some natural strains have an additional band that hybridizes strongly with a probe specific for the *RhsABC* subfamily. The source of this hybridization clearly does not correspond to one of the five elements from strain K-12. This suggested that natural populations harbor at least one additional member of the *RhsABC* subfamily. In this paper, we identify *RhsF* as the source of this signal in strain ECOR-50, and we describe some novel features of this new element that reflect additional complexities of *Rhs* biology. In work to be described elsewhere, we have found *RhsF* in 24 of the 72 ECOR strains.

Isolation and mapping of *RhsF*. The characterization of *RhsF* is based primarily on the isolation of two large restriction fragments from the ECOR-50 chromosome (Fig. 3). One was a 4.1-kb *Sma*I-*Bgl*II fragment (pSZ3958), while the other was a 5.6-kb *Bgl*II fragment (pSZ3828). Preliminary considerations suggested that the pSZ3828 insert contained part of a new *Rhs* element related to the *RhsABC* subfamily. In an attempt to delineate the structure and location of this element, a hybridization probe was prepared from the 0.7-kb *Hpa*I-*Bgl*II interval from the insert's right end and applied to restriction digests of strain K-12 DNA on Southern blots. From the results, a restriction map of the homologous K-12 region was deduced.

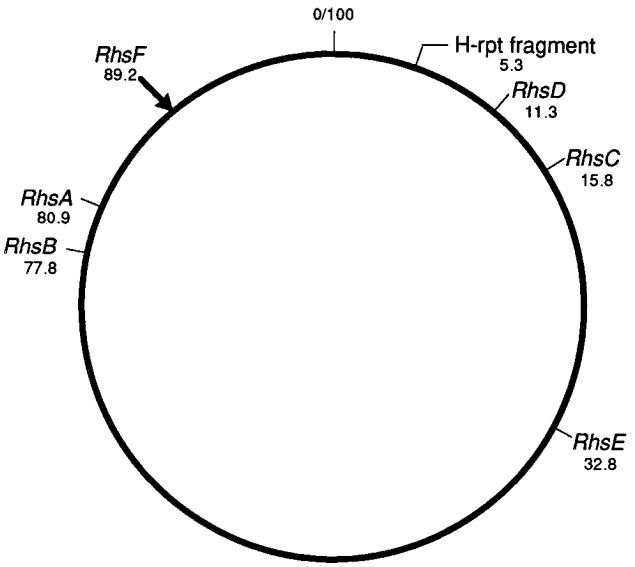

FIG. 1. *RhsF* insertion relative to the *E. coli* K-12 map. The locations of the five *Rhs* elements of K-12 are shown, as is the position of an additional segment of the H-rpt (23).

* Corresponding author. Phone: (717) 531-8592. Electronic mail address: chill@cor-mail.biochem.hmc.psu.edu.

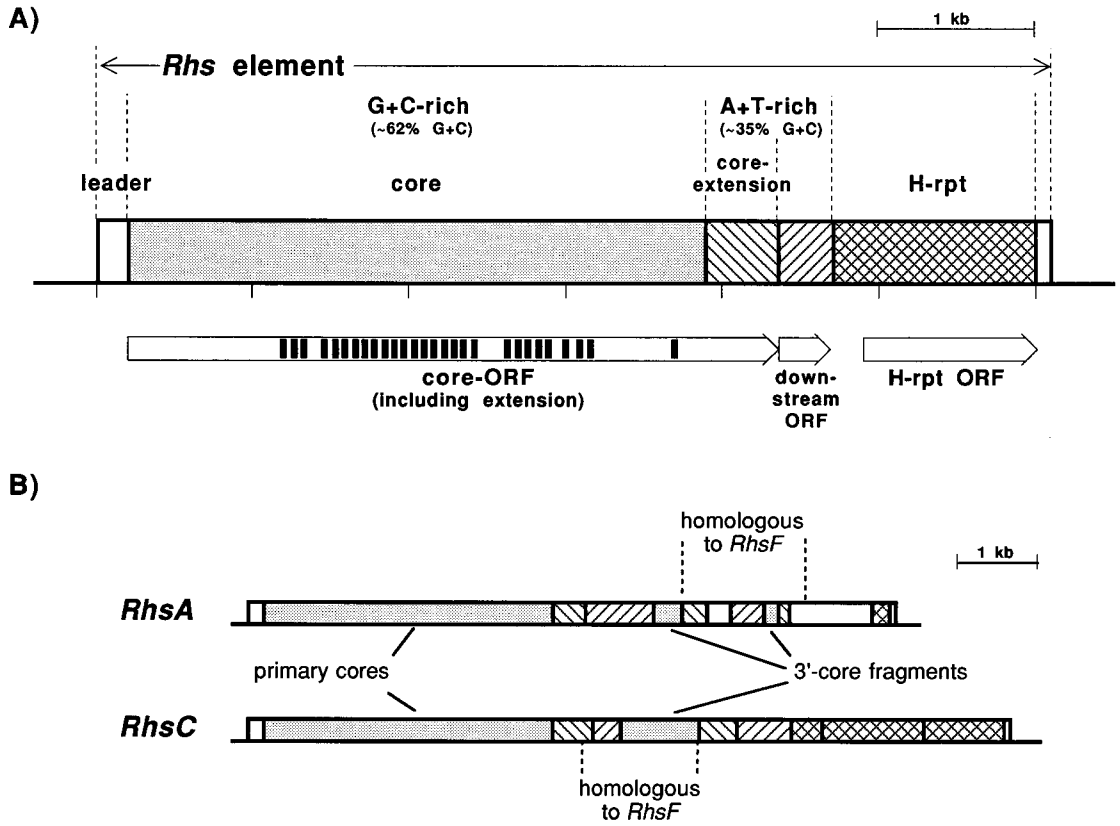

FIG. 2. *Rhs* elements. (A) Structure of a typical *Rhs* element. No individual element has precisely the structure shown, but all have most of the features depicted. Boxes denoting different components are identified in the diagram, and the thin line denotes flanking genomic DNA. ORFs are indicated by the arrows, and the repetitions of the peptide motif are marked by bars within the core ORF. (B) Structures of *E. coli* K-12 *RhsA* and *RhsC*. Various components are shown by patterned boxes with the same conventions as in panel A. The partial core repetitions of *RhsA* and *RhsC* are emphasized, and the portions shared with *RhsF* are indicated.

This map was in turn compared with the Kohara restriction map of the K-12 genome (10) and found to closely match the pattern around minute 89. The location of the homology was confirmed when it was found that the *Hpa*I-*Bgl*II probe hybridized strongly to λ 538 DNA, a clone from the Kohara collection corresponding to minute 89 of the *E. coli* K-12 map (10). This position is distant from the location of any of the K-12 *Rhs* elements (Fig. 1), confirming that a new *Rhs* element was involved. The precise relationship of *RhsF* to the K-12 genome was shown by DNA sequencing commencing at the left *Sma*I site of pSZ3958 and the right *Bgl*II site of pSZ3958. Both of these sites are conserved between K-12 and ECOR-50, and in

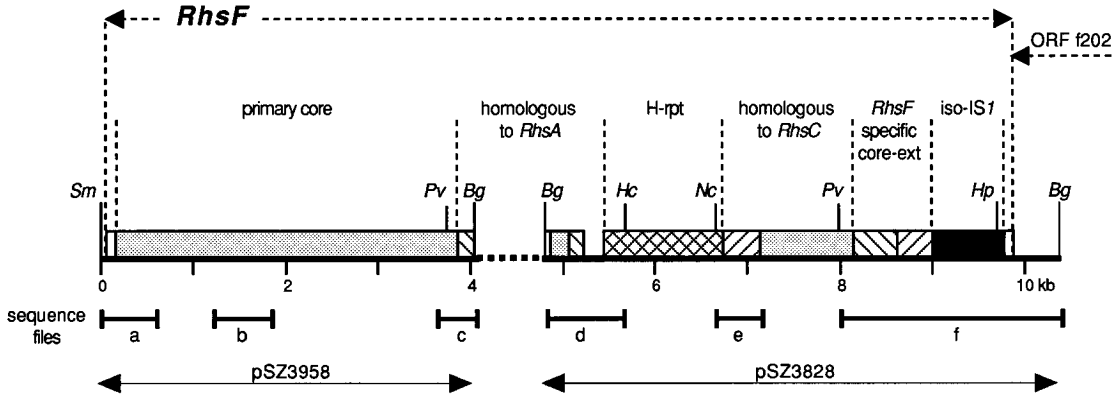

FIG. 3. Structure of *RhsF* from ECOR-50. The *Rhs* components are shown by patterned boxes as in Fig. 2. The first coordinate is the first base of a *Sma*I site and corresponds to base 88843 of the GenBank file ECOUW87, while the *Bgl*II site on the right includes base 89442 (17). ORF f202 is a 609-bp ORF that is common to both K-12 and ECOR-50. The inserts of two clones essential for *RhsF* characterization are shown; pSZ3958 used pOK12 (20) as a vector, while pSZ3828 used pUC19 (14). Details of their isolation are described elsewhere (22). The portions of *RhsF* and adjacent regions that were sequenced are indicated as sequence files a to f. Only selected, illustrative restriction sites are shown for *Sma*I (*Sm*), *Pvu*II (*Pv*), *Bgl*II (*Bg*), *Hin*cII (*Hc*), *Nco*I (*Nc*), and *Hpa*I (*Hp*). The dotted line indicates DNA that has not been cloned but whose identity was deduced by Southern analysis.

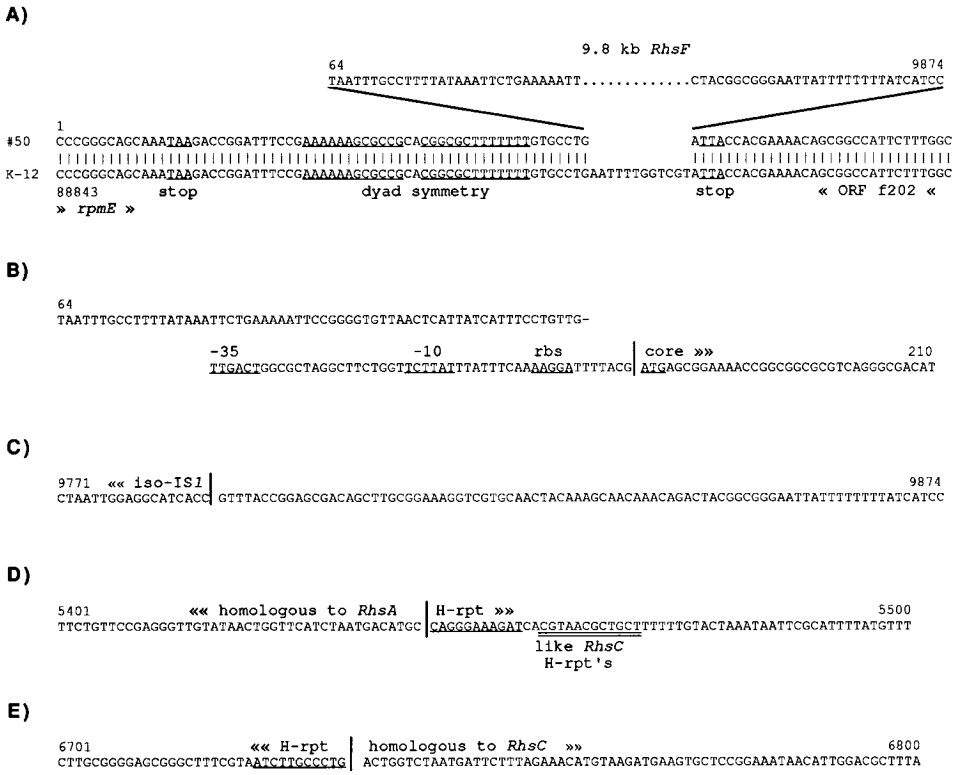

FIG. 4. Sequence of selected segments of *RhsF*. The *RhsF* coordinates refer to a hypothetical file assembled from the sequences determined as specified in Fig. 3 and then connected by the sequences derived from the homologous portions of K-12 *RhsA* and *RhsC*. (A) Comparison of sequences at the divergence between the ECOR-50 and K-12 chromosomes. The 12-bp sequence of K-12 DNA that is replaced by *RhsF* corresponds to nucleotides 88906 to 88917 of ECOUW87 (17). The stop codons for converging ORFs derived from the K-12 sequence and a large dyad symmetry are identified by underlining. (B) Sequence of the leader between the left end of the *RhsF* element and its core. Proposed ribosome binding (rbs) and promoter −35 and −10 sites are underlined. (C) Sequence of the interval between the iso-IS*1* homology and the right end of the *RhsF* element. (D) Sequence at the junction of the K-12 *RhsF-RhsA* homology and the left end of the H-rpt. The 11-bp inverted terminal repeat is underlined, and the 12-bp sequence that distinguishes the H-rpts of *RhsB* and *RhsE* from those of *RhsC* is double underlined. (E) Sequence of the junction between the right end of the H-rpt and the K-12 *RhsF-RhsC* homology. The 11-bp inverted terminal repeat is underlined.

K-12, they embrace a 600-bp segment (17). The K-12 and ECOR-50 sequences were identical for 63 bp downstream from the *Sma*I site but then diverged (Fig. 4A). The sequences again converged 525 bp upstream from the *Bgl*II site. Consequently, 12 bp of the K-12 chromosome is replaced by a 9.8-kb sequence in ECOR-50. In keeping with our operational definition of an *Rhs* element as the entire DNA segment that is absent from a reference *E. coli* strain but which contains an *Rhs* core homolog (9), this 9.8-kb sequence was named *RhsF*. It lies between two converging ORFs of K-12, *rpmE* for ribosomal protein L31 and ORF f202 of unknown function (17). The stop codon for ORF f202 is adjacent to the divergence (Fig. 4A). Such close proximity of the point of an *Rhs* element insertion to the terminus of conserved ORFs has precedents in *RhsA* and *RhsB* (3, 23).

Features of *RhsF*. The *RhsF* element of ECOR-50 was subcloned and characterized by a combination of cloning, restriction site mapping, Southern analysis, and DNA sequencing. Procedures used have been described previously (23), except that CircumVent Thermal Cycle (New England Biolabs) sequencing was used in most cases. The six regions sequenced are shown in Fig. 3 as sequence files a to f, and they cover critical transitions at the ends of the element and between components of the element.

A schematic summary of *RhsF* structure is presented in Fig. 3. The coordinates of Fig. 3 refer to kilobases commencing with the *Sma*I site upstream from *RhsF*. The pSZ3958 and pSZ3828 inserts did not actually overlap but could be related by Southern analysis of digests of ECOR-50 genomic DNA (see below). *RhsF* was found to have many of the features observed for other *Rhs* elements. The ends of its core were sequenced (Fig. 3 sequence files a and c) along with a 609-bp internal segment (sequence file b). These sequences were compared with the homologous segments of *RhsA*, *RhsC*, and *RhsD* (Table 1). The *RhsF* core diverged from *RhsA* and *RhsC* by about 4%, but from *RhsD* by about 24%, placing *RhsF* in the *RhsABC* subfamily. Additional evidence for this assignment came from restriction mapping of pSZ3959 with *Eco*RV, *Pvu*II, *Eco*RI, *Hpa*I, *Kpn*I, *Pst*I, and *Bam*HI: 11 sites found in all members of the *RhsABC* subfamily (23) were present in

TABLE 1. Sequence similarities of *Rhs* core segments[a]

Rhs element	Nucleotide similarity (%)								
	RhsA core segment			*RhsC* core segment			*RhsD* core segment		
	a	b	c	a	b	c	a	b	c
RhsF	97.1	95.9	96.0	97.4	95.7	95.0	82.1	70.9	80.0
RhsA				98.3	97.0	98.0	80.7	72.1	83.5
RhsC							81.9	71.4	82.5

[a] Core segments a, b, and c contain, respectively, the first 420 bp of the core, 609 bp from the interior of the core, and the last 200 bp of the core.

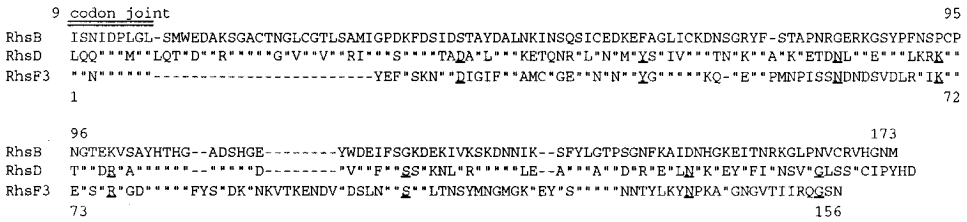

FIG. 5. Similarity of the *RhsF* core extension to those of K-12 *RhsB* and *RhsD*. The *RhsF* sequence is a conceptual translation of the extension of the larger 3′ core fragment (coordinates 8.17 to 8.64 kb in Fig. 3). The *RhsB* and *RhsD* sequences were described previously (3). Alignment utilized the MacMolly Tetra program. Matches of either *RhsD* or *RhsF* to *RhsB* are indicated by quotation marks; positions where *RhsF* matches *RhsD* but not *RhsB* are underlined. Dashes are introduced to aid alignment. The 9-amino-acid sequence that joins the core and extension is marked by a double overline.

RhsF. In addition, one new *Pst*I and one new *Bam*HI site were seen in the *RhsF* core. We conclude that the *RhsF* core belongs to the *RhsABC* subfamily (hereafter called *RhsABCF*) and that the core does not contain insertions or deletions large enough for us to detect by restriction mapping.

In addition to the primary core described above, the *RhsF* element contained two other core homologs. These were located at 4.89 to 5.05 kb and 7.15 to 8.17 kb (shaded boxes in Fig. 3), and in both cases, they were fragments from the 3′ end of the core. Partial core fragments have been found previously in K-12 *RhsA*, *RhsC* (Fig. 2B), and *RhsD*.

RhsF possesses an H-rpt (cross-hatched in Fig. 3). The H-rpt is a 1,291-bp sequence found in *RhsB* and *RhsE* (23). Defective copies are present in *RhsC*, and a highly divergent and degenerate copy (relative to that of *RhsB*) is present in *RhsA*. The H-rpt is a homolog of the *Aeromonas salmonicida* insertion sequence, ISAS2 (5), as well as a related sequence in *Salmonella enterica* (21). The H-rpt is considered to be an insertion sequence although the movement of those found in the *Rhs* elements has not been observed through direct experimentation. The first 213 bp and last 63 bp of the *RhsF* H-rpt were sequenced (Fig. 3 sequence files e and f). This partial sequence along with the observation of a 1.0-kb *Hin*cII-*Nco*I spacing led us to conclude that a complete edition of the H-rpt, very similar to those found in *RhsB* and *RhsE*, was located at coordinates 5.44 to 6.73 kb (Fig. 3). In K-12, the *RhsB* and *RhsE* H-rpts are distinguished from two of the *RhsC* H-rpts by a dissimilar 12-base sequence near the left end; *RhsF* was like the *RhsC* H-rpts in this region (Fig. 4D). Otherwise, it differed from the *RhsB* and *RhsE* H-rpts by less than 4%. A unique aspect of the *RhsF* H-rpt is that it is located in the middle of the element rather than at the extreme right end as in other elements.

A particularly novel feature of *Rhs* elements is the extension of the conserved core ORF into a highly divergent adjacent region of radically different G+C content (Fig. 2A). Nine different core extensions are found among the five K-12 *Rhs* elements; five are associated with primary cores, and four are associated with 3′ core fragments. Only one pair of the nine core extensions has shown any discernible sequence similarity (9). None of the extensions are found in more than one copy in the same genome. Therefore, it was of immediate interest to determine whether the *RhsF* core extension is related to one known for K-12. We sequenced the segment between the core and the *Bgl*II site that terminated the pSZ3958 insert (Fig. 3 sequence file c) where a core extension would be expected. Comparison of this 173-bp segment with the others revealed that it was identical to the core extension associated with a 3′ core fragment in K-12 *RhsA*. This suggested that additional portions of K-12 *RhsA* might be represented in ECOR-50 *RhsF*. Unfortunately, the immediately adjacent material was not cloned. However, sequencing the left end of the pSZ3828 insert (Fig. 3 sequence file d) revealed that its first 620 bp were also homologous to K-12 *RhsA*, differing by only four substitutions and a 6-bp deletion. Included in this sequence was a 160-bp core fragment with just a single mismatch. One substitution generated the *Bgl*II site that determined the left end of the pSZ3828 insert. In *RhsA*, the sequences homologous to the right end of pSZ3958 and the left end of pSZ3828 are separated by 761 bp. The following observations showed that ECOR-50 *RhsF* has a similar sequence and that the *RhsA-RhsF* shared sequence is a single continuous block in both elements. First, *Bgl*II digestion of ECOR-50 genomic DNA produced a 0.7-kb fragment that specifically hybridized with a probe specific for the K-12 *RhsA* sequence in question. Second, *Pvu*II digestion of ECOR-50 genomic DNA produced a 4.3-kb fragment that hybridized both with this probe and with an H-rpt-specific probe. This was exactly the prediction from the structure proposed in Fig. 3 if the core *Pvu*II sites at 3.73 and 8.01 kb had been cut.

As mentioned above, another, larger 3′ core fragment was found to the right of the H-rpt (coordinates 7.15 to 8.17 kb in Fig. 3). This region (Fig. 3 sequence file e) was also similar to sequences from K-12, in this case to *RhsC*. The core fragment was precisely the same size as the core repetition of *RhsC*, and the preceding 419-bp segment was similar to the corresponding sequence of *RhsC*, differing by three substitutions and a deletion of 47 bp. In K-12, this 419-bp sequence is found only in *RhsC*.

The fact that these 3′ core fragments of ECOR-50 *RhsF* and K-12 *RhsC* are identical in extent, not to mention the similarity of the upstream sequence, is strong presumptive evidence that they are related by descent. Therefore, the nature of the core extension following the second core repetition was of interest. In particular, we wanted to know whether it was the same as the extension in K-12 *RhsC*. In fact, the core extension in *RhsF* was quite different. The *RhsF* core extension (coordinates 8.17 to 8.64 kb in Fig. 3) did retain properties generally observed for these *Rhs* components: it had a very low G+C content, 31.2%, and it extended the remnant core ORF by 156 codons, a value within the range generally observed. The first nine codons of all core extensions retain significant sequence homology (9), but beyond these codons, we generally cannot find homology even at the amino acid level. The only exception has been a clear relationship between the primary extensions of *RhsB* and *RhsD*. This new extension in *RhsF* was compared with the others, and similarity to the *RhsB-RhsD* pair could be detected. However, it was considerably less similar to either than they were to each other (Fig. 5). Specifically, the *RhsF* sequence matched *RhsB* at 45 amino acids and *RhsD* at 38 amino acids, whereas *RhsB* and *RhsD* matched at 84 amino acids.

A general feature of *Rhs* elements is the presence of a small downstream ORF in the A+T-rich region that follows a core extension (Fig. 2A). This was true for this new *RhsF* core

extension, in which a 107-codon ORF occurred at coordinates 8.65 to 8.97. The G+C content of this ORF was only 29.9%. As is generally true for these downstream ORFs (9), the sequence of its amino terminus (MKVIQCCYFLALVSSHLMAGV MSD. . .) suggested that it might serve as a signal peptide for export from the cytoplasm. No specific homology with other *Rhs* downstream ORFs was discernible.

The sequences preceding the *Rhs* cores presumably contain promoter and other sequences necessary for core expression. In contrast to the high conservation of the cores, these leader sequences generally show limited similarity (23). The leader of *RhsF* was only 112 bp long, which makes it the smallest leader in the *Rhs* family. It had a very good match to the promoter consensus (Fig. 4B). The start of the leader showed no particular relationship to the distal end of the element (Fig. 4C), nor was either end related to the ends of other *Rhs* elements. This lack of relationship between *Rhs* ends sets them apart from most other accessory genetic elements.

The final component of *RhsF* was a segment that appeared to be an insertion sequence of the IS*1* family (Fig. 3). Of the various IS*1* homologs, it was most similar to the iso-IS*1* element, νξ, which occurs in 150 copies in a strain of *Shigella dysenteriae* (16). The 764-bp *RhsF* sequence was 80.0% similar to the 766-bp *S. dysenteriae* sequence. Like IS*1*, it contains two overlapping ORFs, and the stop codon of the first is preceded by the sequence A-A-A-A-A-A-C, known to cause translational frameshifting in IS*1* (19). While IS*1* homologs have not been previously associated with *Rhs* elements, insertion sequences are common components. As discussed above, the H-rpt is considered to be an insertion sequence, and the most distal portion of *RhsD* is an apparently defective homology (65% identity over 122 amino acids) of an insertion sequence from *Actinobacillus pleuropneumoniae* (1).

Inferences from the structure of *RhsF*. These results have important implications for the evolution of *Rhs* elements. The significance of the 3′ core fragments in various *Rhs* elements has been a matter of speculation. A plausible hypothesis is that they, and the extensions of their remnant ORFs, were once part of a primary core unit but subsequently a new extension became attached to the primary core by a mechanism that left a piece of the core and the old extension nearby. Finding the same core extension at both primary (in ECOR-50 *RhsF*) and secondary (in K-12 *RhsA*) positions supports this idea. However, the fact that the extension at issue resides in different *Rhs* elements in the two strains adds an additional layer of complexity. Although ECOR-50 has several other *Rhs* elements, it does not have *RhsA*. Like ECOR-55, a strain that has no *Rhs* elements at all, ECOR-50 has a 32-bp alternative sequence at the position occupied by *RhsA* (2, 3). If the secondary extension in *RhsA* is merely evolutionary debris, it is either recent or remarkably conserved. It is identical to the copy in *RhsF*, at least over the 173 bp of available sequence. This contrasts with other ECOR-50 and K-12 sequence comparisons in which at least some divergence is seen. For example, we found that ORF f202 (Fig. 3) diverged at 15 of 519 positions (2.9%), while divergences of 3.9% for a segment of the *trp* operon, 1.5% in the region distal to *trp*, 1.7% in *celC*, and 1.0% in *crr*, have been reported previously (6, 12, 13). Combination of these comparisons of ECOR-50 and K-12 alleles indicates that synonymous substitutions exceed nonsynonymous substitutions by a factor of 7. The conservation of the core extension becomes more significant if it is assumed that once one copy became secondary, all substitutions were neutral. The implication is that the rearrangement(s) that dispersed this extension to different *Rhs* elements occurred recently relative to the divergence of the general chromosomal frame of these two strains. ECOR-50 and K-12 are rather distantly related according to multilocus enzyme electrophoresis, restriction analysis, and sequence comparison (7, 12). In our investigation of the *Rhs* profiles of the ECOR strain collection, we have found this particular extension at the *RhsF* location only in the distinct clonal group of ECOR-46, ECOR-49, and ECOR-50 (8). It is present in 42 other ECOR strains besides ECOR-46, -49, and -50 but always at the *RhsA* location.

As mentioned above, the larger of the *RhsF* 3′ core fragments was identical in size to the one in K-12 *RhsC*. This, along with the homology of the respective upstream sequences, strongly indicates that these two core fragments descend from a single rearrangement event that defined the core fragment's left end. However, the respective core extensions were found to be dissimilar. This suggests that the event that caused the core repetitions to diverge on the right was more recent than the event that defined the left end.

As stated at the outset, our concept of the *Rhs* elements is that they are composite structures assembled from discrete components that had independent evolutionary origins outside of the *E. coli* species. This concept was developed from observation of the elements present in strain K-12 (9). We now see that *Rhs* elements can be assembled by reshuffling portions of other *Rhs* elements. Specifically, ECOR-50 *RhsF* consists of components that are associated with *RhsA* and *RhsC* in K-12. How this came to be is obscure, but it has been proposed that the H-rpt causes illegitimate recombination that aids *Rhs* element assembly (9). The fact that the H-rpt precisely connects the *RhsA* homology to the *RhsC* homology in *RhsF* (Fig. 4D and E) supports such a role. The iso-IS*1* element of *RhsF* might play a similar role. An H-rpt homolog has been implicated in mediating genetic diversity of the *rfb* locus of *S. enterica* (21).

Nucleotide sequence accession number. Nucleotide sequences of the six files, a to f, have been submitted to the GenBank database under accession numbers U15122 to U15127.

We thank Robert Selander and Ken Rudd for bacterial and phage stocks.

This work was supported by Public Health Service grant GM16329 from the National Institutes of Health.

REFERENCES

1. **Anderson, C., A. A. Potter, and G.-F. Gerlach.** 1991. Isolation and molecular characterization of spontaneously occurring cytolysin-negative mutants of *Actinobacillus pleuropneumoniae* serotype 7. Infect. Immun. **59:**4110–4116.
2. **Feulner, G.** 1990. Ph.D. thesis. The Pennsylvania State University, University Park.
3. **Feulner, G., J. A. Gray, J. A. Kirschman, A. F. Lehner, A. B. Sadosky, D. A. Vlazny, J. Zhang, S. Zhao, and C. W. Hill.** 1990. Structure of the *rhsA* locus from *Escherichia coli* K-12 and comparison of *rhsA* with other members of the *rhs* multigene family. J. Bacteriol. **172:**446–456.
4. **Foster, S. J.** 1993. Molecular analysis of three major wall-associated proteins of *Bacillus subtilis* 168: evidence for processing of the product of a gene encoding a 258 kDa precursor two-domain ligand-binding protein. Mol. Microbiol. **8:**299–310.
5. **Gustafson, C. E., S. Chu, and T. J. Trust.** 1994. Mutagenesis of the paracrystalline surface protein array of *Aeromonas salmonicida* by endogenous insertion elements. J. Mol. Biol. **237:**452–463.
6. **Hall, B. G., and P. M. Sharp.** 1992. Molecular population genetics of *Escherichia coli*: DNA sequence diversity at the *celC*, *crr*, and *gutB* loci of natural isolates. Mol. Biol. Evol. **9:**654–665.
7. **Herzer, P. J., S. Inouye, M. Inouye, and T. S. Whittam.** 1990. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. J. Bacteriol. **172:**6175–6181.
8. **Hill, C. W.** Unpublished data.
9. **Hill, C. W., C. H. Sandt, and D. A. Vlazny.** 1994. *Rhs* elements of *Escherichia coli*: a family of genetic composites each encoding a large mosaic protein. Mol. Microbiol. **12:**865–871.
10. **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.
11. **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.
12. **Milkman, R., and M. M. Bridges.** 1993. Molecular evolution of the *Escherichia coli* chromosome. IV. Sequence comparisons. Genetics **133:**455–468.
13. **Milkman, R., and A. Stoltzfus.** 1988. Molecular evolution of the *Escherichia coli* chromosome. II. Clonal segments. Genetics **120:**359–366.
14. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene **26:** 101–106.
15. **Ochman, H., and R. K. Selander.** 1984. Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. **157:**690–693.
16. **Ohtsubo, H., K. Nyman, W. Doroszkiewicz, and E. Ohtsubo.** 1981. Multiple copies of iso-insertion sequences of IS*1* in *Shigella dysenteriae* chromosome. Nature (London) **292:**640–643.
17. **Plunkett, G., III, V. Burland, D. L. Daniels, and F. R. Blattner.** 1993. Analysis of the *Escherichia coli* genome. III. DNA sequence of the region from 87.2 to 89.2 minutes. Nucleic Acids Res. **21:**3391–3398.
18. **Sadosky, A. B., J. A. Gray, and C. W. Hill.** 1991. The *RhsD-E* subfamily of *Escherichia coli* K-12. Nucleic Acids Res. **19:**7177–7183.
19. **Sekine, Y., and E. Ohtsubo.** 1989. Frameshifting is required for production of the transposase encoded by insertion sequence *1*. Proc. Natl. Acad. Sci. USA **86:**4609–4613.
20. **Vieira, J., and J. Messing.** 1991. New pUC-derived cloning vectors with different selectable markers and DNA replication origins. Gene **100:**189–194.
21. **Xiang, S.-H., M. Hobbs, and P. R. Reeves.** 1994. Molecular analysis of the *rfb* gene cluster of a group D2 *Salmonella enterica* strain: evidence for intraspecific gene transfer in O antigen variation. J. Bacteriol. **176:**4357–4365.
22. **Zhao, S.** 1992. Ph.D. thesis. The Pennsylvania State University, University Park.
23. **Zhao, S., C. H. Sandt, G. Feulner, D. A. Vlazny, J. A. Gray, and C. W. Hill.** 1993. *Rhs* elements of *Escherichia coli* K-12: complex composites of shared and unique components that have different evolutionary histories. J. Bacteriol. **175:**2799–2808.