JOURNAL OF BACTERIOLOGY, June 1994, p. 3559–3567
0021-9193/94/$04.00+0

Vol. 176, No. 12

Characterization of Genetic Determinants for R Body Synthesis and Assembly in *Caedibacter taeniospiralis* 47 and 116

DANIEL P. HERUTH,* FINN R. POND,† JUDITH A. DILTS,‡ AND ROBERT L. QUACKENBUSH§

Department of Microbiology, School of Medicine, University of South Dakota, Vermillion, South Dakota 57069

Received 27 December 1993/Accepted 6 April 1994

***Caedibacter taeniospiralis*, an obligate bacterial endosymbiont of *Paramecium tetraurelia*, confers a killing trait upon its host paramecium. Type 51 R bodies (refractile inclusion bodies) are synthesized by these endosymbionts and are required for expression of the killing trait. The nucleotide sequence of the genetic determinants for type 51 R body synthesis and assembly was determined for *C. taeniospiralis* 47 and 116. Three independently transcribed genes (*rebA*, *rebB*, and *rebC*) were characterized. To date these are the only genes from *C. taeniospiralis* to be sequenced and characterized. DNA regulatory regions are recognized by *Escherichia coli*, and codon usage appears similar to that in *E. coli*. A fourth open reading frame with appropriate regulatory sequences was found within the *reb* locus, but no evidence was obtained to suggest that this putative gene is expressed in *E. coli*. The R body-encoding sequences from both strains are identical. Two-dimensional gel electrophoresis of deletion derivatives shows that two polymerization events are involved in R body assembly. One polymerization event requires only RebB and RebC; the other requires all three proteins. Expression of RebC is necessary for the posttranslational modification of RebA and RebB into species with three and two different molecular weights, respectively. In the presence of RebC, each species of RebB with a different molecular weight has six different isoelectric points.**

Refractile inclusion bodies, known as R bodies, are produced by only a few species of bacteria. These inclusion bodies are highly insoluble protein ribbons, typically seen coiled into cylindrical structures within the cell. Five classes or types of R bodies have been described and are distinguished on the basis of physical dimensions, morphology, and behavior in response to certain physical and chemical treatments (16, 21, 22).

Several classes of R bodies are produced by *Caedibacter* species, more commonly known as kappa particles. *Caedibacter* species are obligate bacterial endosymbionts of *Paramecium* species and are characterized by their ability to produce both R bodies and a toxin. R body-containing forms of *Caedibacter* species are toxic to sensitive strains of paramecia. Thus, paramecia that carry any *Caedibacter* strain are referred to as killers. Killing occurs when a sensitive paramecium ingests an R body-containing kappa particle that has been released into the environment by a killer paramecium (6).

Five *Caedibacter* species, producing three of the five known types of R bodies (types 7, 51, and Cc [18, 19]), have been identified. R body synthesis in four of the species is determined by extrachromosomal elements thought to be defective phages (17). In the fifth species, *Caedibacter taeniospiralis*, R body synthesis is directed by plasmid DNA (23). The R bodies of *C. taeniospiralis* are type 51. They are 0.4 μm wide, have a maximum length of 20 μm, possess acute angles at each end, and unroll in a telescopic fashion when exposed to a pH of 6.5 or lower.

The R body-encoding region from the native 43-kbp plasmid (pKAP47) in *C. taeniospiralis* 47 has been cloned and expressed in *Escherichia coli* (23). The recombinant plasmid (pBQ65) contains a 2,700-bp fragment encoding three small polypeptides. These polypeptides were estimated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) to be 18, 13, and 10 kDa (8). An analysis of deletion derivatives of pBQ65 showed that synthesis of each of these polypeptides is necessary for R body assembly and indicated that each polypeptide species is controlled by a separate gene which is expressed as an independent transcriptional unit possessing regulatory signals recognized by *E. coli* (8). Earlier work also suggested that the 18- and 13-kDa polypeptides are posttranslationally modified into 3 and 12 species, respectively, and that R body assembly involves polymerization of the 13-kDa species into high-molecular-mass complexes (8).

In this paper we present the DNA sequence for the R body locus (*reb*) from two *C. taeniospiralis* strains, 47 and 116. Three genes (*rebA*, *rebB*, and *rebC*), which code for the 18-kDa (RebA), 13-kDa (RebB), and 10-kDa (RebC) products, respectively, have been identified. Analysis of pBQ65 deletion derivatives which further elucidates the functional relationships among the three polypeptides is presented.

MATERIALS AND METHODS

Bacterial strains and plasmids. The bacterial strains and plasmids used in this study are listed in Table 1.

Media and antibiotics. *E. coli* JM101 and JM107 were grown in YT broth (0.8% Bacto Peptone, 0.5% yeast extract, 0.5% NaCl [pH 7.5]). *E. coli* JM105 was grown in L broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl). The minicell-producing *E. coli* P678-54 was grown in M9 minimal medium (0.1 mM $CaCl_2$, 1 mM $MgSO_4$, 0.6% $NaHPO_4$, 0.3% KH_2PO_4, 0.05% NaCl, 0.1% NH_4Cl) supplemented with 0.4% glucose, 1% Casamino Acids (Difco Laboratories, Detroit, Mich.), 4 mg of thiamine hydrochloride per liter, 80 mg of threonine per liter, and 40 mg of leucine per liter. Solid media contained 1.5% agar. Top agar used for plating M13 clones was made by

* Corresponding author. Present address: School of Biological Sciences, University of Missouri-Kansas City, Kansas City, MO 64108. Phone: (816) 931-8080. Fax: (816) 753-1307.

† Present address: Department of Biology, Whitworth College, Spokane, WA 99251.

‡ Present address: Department of Biology, William Jewell College, Liberty, MO 64108.

§ Present address: National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892.

TABLE 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Genotype or phenotype[a]	Source or reference
E. coli strains		
JM101	*supE thi* Δ(*lac-proAB*) (F′ *traD36 proAB lacI*^{q}*ZΔM15*)	32
JM105	*thi rpsL endA sbcB15 hspR4* Δ(*lac-proAB*) F′ *traD36 proAB lacI*^{q}*ZΔM15*)	32
JM107	*endA1 gyrA96 thi hsdR17 supE44 relA1* λ$^-$ Δ(*lac-proAB*) (F′ *traD36 proAB lacI*^{q}*ZΔM15*)	32
P678-54	*thr-1 leuB6 minA1 minB2 thi-1 ara-13 lacY1 gal-6 malA1 xyl-7 mtl-2 azi-8 strA135 tonA2*	12
Plasmids		
pUC18	Amp^r	32
pBQ59	Amp^r *C. taeniospiralis* 47 *rebA rebB*	8
pBQ65	Amp^r *C. taeniospiralis* 47 *reb*	8
pBQ69	Amp^r *C. taeniospiralis* 47 Δ*rebA rebB rebC*	8
pBQ70	Amp^r *C. taeniospiralis* 47 *rebB rebC*	8
pBQ80	Amp^r *C. taeniospiralis* 116 *reb*	This study
pBQ88	Amp^r *C. taeniospiralis* 116 *reb*	This study
pBQ89	Amp^r *C. taeniospiralis* 47 *reb*	This study
pKAP47	*reb*	20
pKAP116	*reb*	20

[a] Amp, ampicillin.

the addition of 0.6% agar to B medium (1% Bacto Peptone, 0.8% NaCl [pH 7.5]). Antibiotic concentrations were 100 μg of ampicillin per ml, 25 μg of chloramphenicol per ml, and 30 μg of streptomycin per ml. All cultures were grown at 37°C with shaking.

Plasmid purification and manipulations. *C. taeniospiralis* 116 cells were purified from lysates of host paramecia by chromatography on ECTEOLA columns (13). Plasmid DNA (pKAP116) was isolated by using ethidium bromide-cesium chloride gradients (3). Plasmid DNA from bacterial transformants was isolated by the cleared-lysate method of Clewell and Helinski (2). The procedures for restriction endonuclease analysis, cloning, and transformation have been described previously (23). Recombinant plasmids were detected by the miniprep protocol of Birnboim and Doly (1). Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs, Beverly, Mass., and Bethesda Research Laboratories, Gaithersburg, Md., respectively.

Nucleotide sequence analysis. The 5′ ends of DNA fragments from pBQ65 were labeled with T4 polynucleotide kinase (Bethesda Research Laboratories) and [γ-^{32}P]ATP (ICN Pharmaceuticals, Inc., Irvine, Calif.) and then reduced by restriction enzyme cleavage to smaller fragments labeled at one end (11). Labeled fragments electroeluted from polyacrylamide gels were sequenced by the Maxam and Gilbert procedure (11).

Following the construction of pBQ88 and pBQ89, restriction fragments from these plasmids were subcloned into M13mp18 and M13mp19 (14) and sequenced by the dideoxy-chain termination method (27) as recommended in the Bethesda Research Laboratories M13 sequencing manual. All restriction sites were bridged, and complementary sequences were obtained from both strands. Nucleotides, Klenow fragment, and the universal primer were purchased from both Bethesda Research Laboratories and Pharmacia Inc., Piscataway, N.J. Reaction mixtures were resolved by 8% polyacrylamide gel electrophoresis by the method of Sanger and Coulson (28).

Assembly and analysis of the nucleotide sequence were performed on a VAX computer (Digital Equipment Corp.) with the Sequence Analysis Software Package (version 6.1) of the Genetics Computer Group (University of Wisconsin, Madison) and the program MacVector 3.5, International Biotechnologies, Inc., New Haven, Conn. The NBRF Protein Identification Resource data bank (release 23.0) and GenBank (release 71.0) were searched for matches to the Reb proteins. Hydrophobicity plots were constructed by using the program DNA Strider (10) and the parameters of Kyte and Doolittle (9). Codon usage analysis was aided by the program DNA Strider (10).

Identification of polypeptides encoded by *rebA*, *rebB*, and *rebC*. Minicells produced by *E. coli* P678-54 and its derivatives were purified on linear sucrose gradients by the procedure of Roozen et al. (26). Plasmid-encoded polypeptides in minicells were labeled as described previously (8) except that 50 μCi of [^{35}S]methionine (ICN Pharmaceuticals, Inc.) per ml was added and the minicells were incubated for 45 min at 37°C.

Labeled protein samples were analyzed by both one- and two-dimensional SDS-PAGE. Isoelectric focusing tube gels (1 mm by 14 cm; 8 M urea, 3.5% polyacrylamide [5% cross-linking], 20% Triton X-100, 0.5% Ampholine pH 3 to 10, 4.5% Ampholine pH 4 to 6.5 [Ampholines were purchased from Sigma Chemical Co., St. Louis, Mo.]) were run for 16 h at 400 V followed by 2 h at 800 V. The catholyte buffer was 0.1 M NaOH, and the anolyte buffer was 0.06% phosphoric acid. Extruded isoelectric focusing gels were covered with SDS reducing buffer (62.5 mM Tris-HCl [pH 6.8], 10% glycerol, 2% SDS, and 0.715 M β-mercaptoethanol with bromphenol blue as a tracking dye) and either immediately loaded onto a slab gel or frozen in an acetone-dry ice bath and stored at −70°C. SDS-PAGE was carried out as previously described (8). Autoradiography was enhanced by treating the gels with Enlightning (New England Nuclear Corp., Boston, Mass.).

Nucleotide sequence accession numbers. The nucleotide sequences reported in this paper have been deposited in the GenBank/EMBL data library (accession numbers U04523 [for pBQ88]) and U04524 [for pBQ89]).

RESULTS

Cloning and sequencing the *reb* locus from *C. taeniospiralis* 47 and 116. The plasmid pBQ65, which encodes the R body proteins, was used for earlier studies on R body expression in *E. coli* (8, 23). In addition to the pKAP47 *reb* locus, pBQ65 contains a fragment of the transposon Tn*4503* located upstream to *rebA*. The recombinant plasmid pBQ89 contains the pKAP47 *reb* locus and was constructed by inserting a 2,000-bp

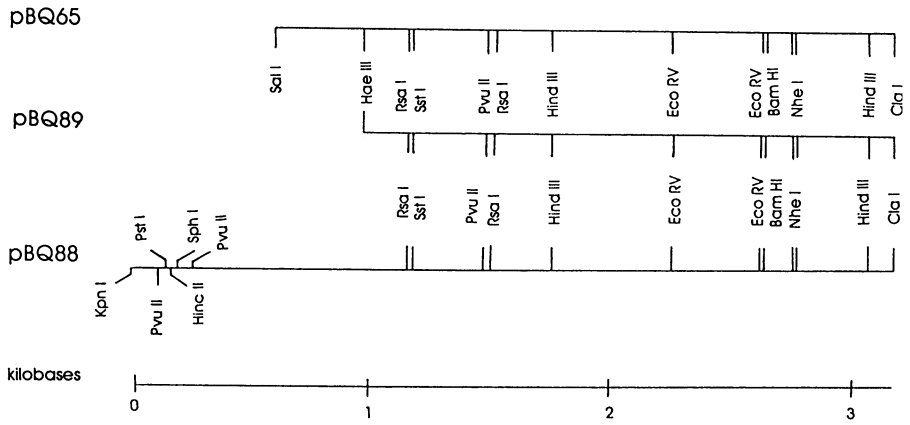

FIG. 1. Comparison of restriction maps of cloned R body-encoding sequences derived from pKAP47 (pBQ65 and pBQ89) and pKAP116 (pBQ88). The inserted sequences in pBQ65 and pBQ89 contain a portion of Tn*4503* upstream of the R body-encoding sequence. This transposon is not present in pBQ88. Modified from reference 16 and printed with permission.

*Cla*I-*Hae*III fragment from pBQ65 into a pUC18 vector digested with *Acc*I and *Sma*I.

The recombinant plasmid pBQ80 was constructed by inserting a 13.2-kbp fragment encoding the R body proteins from pKAP116 into pUC18. pBQ88 was then constructed by inserting a 3,200-bp *Cla*I-*Hae*III fragment from pBQ80 into pUC18. pBQ88 contains the *reb* locus plus 1,000 bp of upstream DNA. Unlike pBQ65, this large upstream region does not contain any transposon sequence.

Figure 1 compares the cloned fragments from pBQ88, pBQ89, and pBQ65. The region containing the *reb* locus from *C. taeniospiralis* 116 (pBQ88) has a restriction endonuclease profile identical to that of the equivalent region in *C. taeniospiralis* 47 (pBQ65 and pBQ89).

Identification of *rebA*, *rebB*, and *rebC*. The nucleotide sequence of the *reb* locus from both pBQ88 and pBQ89 appears in Fig. 2. The sequence presented for pBQ88 is 1,931 bp and runs from 110 bp upstream of the *Rsa*I site to the *Hin*dIII site (Fig. 1). The sequence presented for pBQ89 is a 1,893-bp *Hae*III-*Hin*dIII fragment. The 5′ end of this sequence contains a portion of the long terminal repeat from Tn*4503* (5). Three open reading frames (ORFs) corresponding to the regions of DNA sequence predicted by deletion derivative analysis of pBQ65 (8) have been identified. The names *rebA*, *rebB*, and *rebC* were assigned to these genes. The sizes of the three predicted protein products are 114 (11.6 kDa), 74 (7.7 kDa), and 56 (6.0 kDa) amino acid residues, respectively. The predicted molecular weight of each polypeptide product is approximately 60% of that calculated by comparison with molecular weight standards on an SDS-polyacrylamide gel (8). Since these are the first genes sequenced and identified from *C. taeniospiralis*, no consensus sequences for *Caedibacter* regulatory elements are available. However, because these genes are expressed in *E. coli* by using their own regulatory sequences, analysis was performed by using the consensus sequences for *E. coli* regulatory elements.

The first ORF, designated *rebA*, extends from base 148 through 492. *rebA* is preceded by the potential ribosome-binding site (RBS) (30, 31) AGGAG, located 6 bp before the initiation codon. In *E. coli*, the consensus promoter sequences for the −10 and −35 regions are TTATA and TTGACA, respectively (4). The upstream region of *rebA* was searched for comparable regulatory sequences. The probable −10 region for the *rebA* gene in pKAP116 (pBQ88) is TTATAT (nucleotides 50 to 55 in Fig. 2). The probable −35 region, TTGAGA, lies 18 bp upstream at nucleotides 26 to 31. The *rebA* gene in pKAP47 (pBQ89), however, does not use the same regulatory sequences. As described below, the different promoter usage in pBQ89 is apparently a consequence of a transpositional event.

The transposon Tn*4503* was known previously to be close to the *reb* locus of pKAP47 (24, 25). This study and the work of Hodel-Christian (5) indicate that the insertion of Tn*4503* disrupted the normal *rebA* regulatory sequences. Remarkably, however, a functional promoter region was generated by the insertional event. Situated 21 bp downstream from the −10 region of pBQ88 (nucleotides 77 to 82 in Fig. 2) is a sequence that matches the −10 consensus sequence for *E. coli* (TATAAT). This sequence, present in both pBQ88 and pBQ89, has no obvious function in regulating transcription of the *rebA* gene in pBQ88 but appears to be an active component of the *rebA* promoter sequence in pBQ89. Within the long terminal repeat of Tn*4503* is a potential −35 sequence, TGGATA. Transposon insertion brought this −35 region into proximity with the potential −10 region. These sequences, separated by 18 bp, form a new *rebA* promoter.

The *rebA* gene uses the TAA stop codon. Termination appears to be rho dependent since there are no strong inverted repeats followed by runs of thymidine.

The second ORF, *rebB*, extends from base 573 through 890. Interestingly, there are four possible ATG initiation codons in frame at nucleotides 573, 639, 666, and 671. These initiation sites are all preceded by potential regulatory sequences. Depending on which initiation codon is used, the resulting gene product would be 105, 82, 74, or 72 amino acids long. The potential regulatory sequences for the initiation codon located at position 573 are the RBS AGGA (bases 562 to 565), the −10 sequence TTATAA (bases 511 to 516), and the −35 sequence TTGATA (bases 479 to 484). The ATG codon at position 639 is preceded by the RBS AGGT (bases 632 to 635) and the −10 and −35 sequences AGTAAT (bases 576 to 581) and CA GACA (bases 551 to 556), respectively. The initiation codons at positions 666 and 671 are preceded by the potential RBS GGAAG (bases 657 to 661), −10 sequence TATTAG (bases 628 to 633), and −35 sequence TTCAGT (bases 602 to 607). The *rebB* gene utilizes a TAA stop codon, and transcriptional termination appears to be rho dependent.

The third ORF, *rebC*, extends from base 1431 through 1601. The initiation codon for *rebC* is preceded by a possible RBS

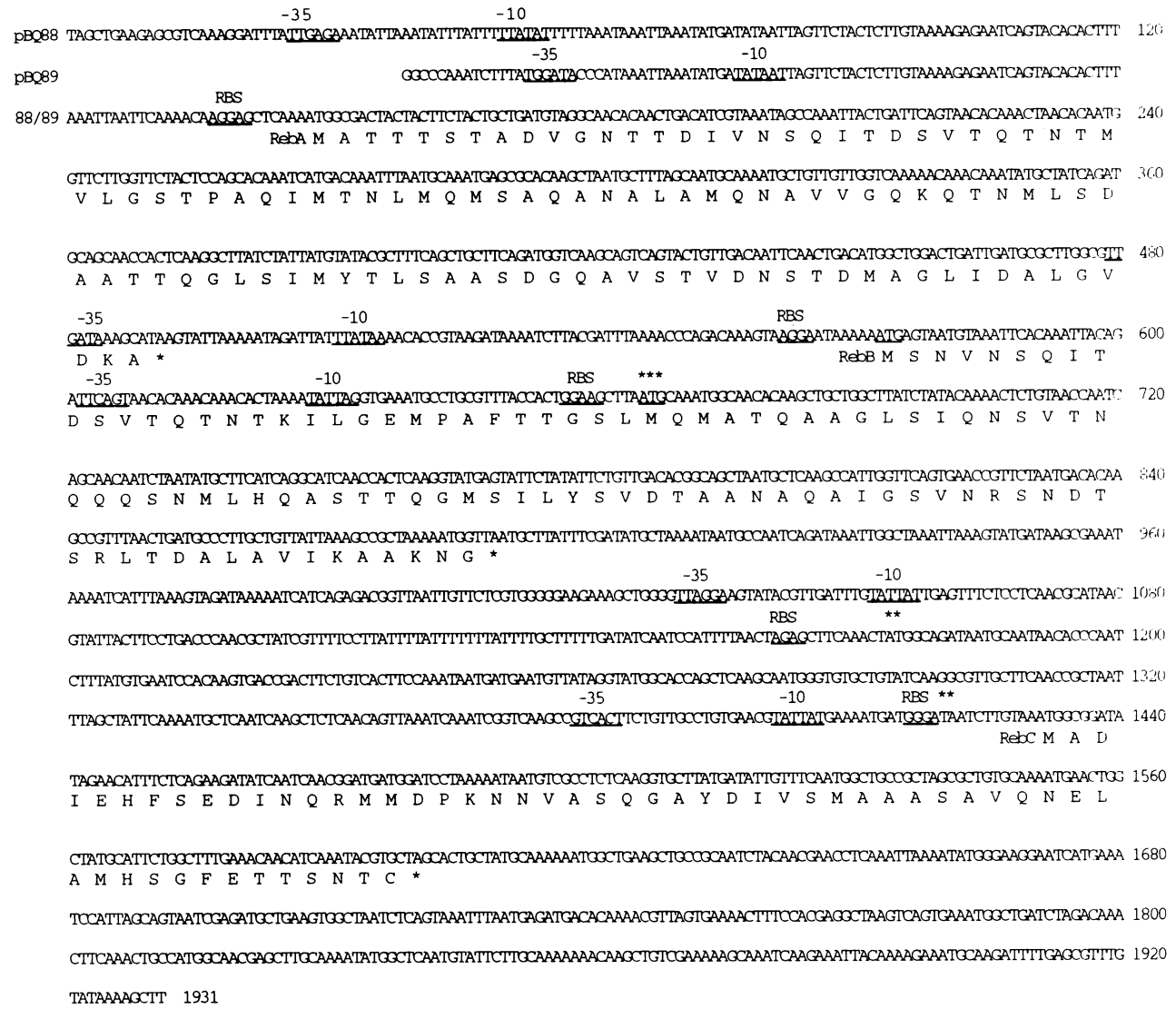

FIG. 2. Nucleotide sequence of the *reb* locus from *C. taeniospiralis* 47 and 116 and deduced primary structure of the protein products of *rebA*, *rebB*, and *rebC*. Probable −35 and −10 promoter regions are underlined and numbered. Possible RBSs are also underlined. The initiation codon which would encode the short version of RebB is denoted by ***, and the corresponding regulatory sequences are underlined. The initiation and termination codons for a fourth ORF between *rebB* and *rebC* are denoted by **. The putative regulatory sequences for this ORF are underlined.

GGGA (bases 1416 to 1419). The probable −10 and −35 promoter sequences, TATTAT (bases 1401 to 1406) and GTCACT (bases 1378 to 1383), respectively, are separated by 17 bp. The *rebC* gene utilizes the TAG stop codon, and transcriptional termination appears to be rho dependent.

A fourth ORF, extending from base 1174 through 1422, is present between *rebB* and *rebC*. The initiation codon and the possible RBS AGAG (bases 1161 to 1164) are separated by 9 bp. The probable −10 TATTAT (bases 1052 to 1057) and −35 TTAGGA (bases 1030 to 1035) regulatory sequences are separated by 16 bp. The stop codon, TAA, is immediately downstream of the RBS for *rebC*, and therefore termination signals for transcription would overlap with the ORF of *rebC*. There are no inverted repeats or runs of thymidine, so termination would be rho dependent. This ORF would code for a protein of 82 amino acids, with a molecular mass of 8.8 kDa. The protein would contain eight Met residues and two Cys residues and therefore should have been radiolabeled in our assay (see Materials and Methods) if it was synthesized. However, we obtained no evidence for the use of this ORF in *E. coli*.

Codon usage. The distribution of codons in each of the four ORFs is depicted in Table 2. The codon distribution is similar for all four ORFs and is not unusual when compared with codon usage in *E. coli*. The codon adaptive index (CAI) is a statistical analysis that measures the relative level at which natural selection has influenced the pattern of codon usage in a gene (29). Highly expressed genes have been selected to have optimal codon usage, whereas genes that are expressed at a low level are not subjected to such selection. Therefore, high CAI values are representative of high levels of gene expression. The CAI values of the three genes of the *reb* locus are 0.298 for *rebA*, 0.237 for *rebB*, and 0.249 for *rebC*; for comparison, values for *E. coli* genes are 0.77 for genes expressed at a high level, 0.25 for genes expressed at a low level, and 0.17 for a random sequence of nucleotides (29). The CAI values for the *rebB*

TABLE 2. Distribution of codons in each of the four ORFs[a]

First base	Second Base T A	B	C	4th	C A	B	C	4th	A A	B	C	4th	G A	B	C	4th	third base
	0	1	1	0	3	5	2	2	1	1	1	2	0	0	0	2	T
	phe				ser				tyr				cys				
	0	0	1	2	0	0	0	0	0	0	0	0	0	0	1	0	C
T	3	4	0	2	5	4	3	2	1	1	0	1	0	0	0	0	A
									stop								
	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	G
													trp				
	3	2	0	1	0	1	1	0	0	1	2	0	0	2	0	1	T
	leu				pro				his				arg				
C	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	C
	1	1	0	0	1	0	0	2	10	9	3	9	0	0	0	0	A
									gln								
	1	0	1	1	0	0	0	0	0	2	0	1	0	0	1	0	G
	3	4	1	2	11	4	0	1	6	7	5	7	1	2	0	0	T
	ile				thr				asn				ser				
	2	0	1	1	1	3	0	2	2	3	0	0	2	2	1	1	C
A	0	2	1	3	5	5	2	1	2	3	1	1	0	0	0	0	A
									lys				arg				
	9	6	5	8	1	1	1	0	0	0	0	0	0	0	0	0	G
	met																
	5	2	1	3	7	6	5	5	6	2	4	1	3	4	1	3	T
	val				ala				asp				gly				
G	1	0	1	1	0	3	2	2	3	2	0	1	3	1	1	1	C
	3	3	0	0	7	3	0	4	0	1	4	1	1	1	0	1	A
									glu								
	0	1	1	3	2	1	1	0	0	0	0	0	0	0	0	0	G

[a] A, *rebA* ORF; B, *rebB* ORF; C, *rebC* ORF; 4th, fourth ORF.

ORF that would encode smaller polypeptides (74 residues, 0.279; 82 residues, 0.265) are actually higher than those for the complete ORF (105 residues). All three genes have CAI values similar to those of *E. coli* genes that are expressed at a moderate or low level. The CAI value for the possible fourth ORF (0.220) is above the standard for a random sequence and only slightly lower than the value for the three known genes of the *reb* locus.

In bacteria with high G+C content, a high proportion of codons ending in G or C is characteristic of moderately or highly expressed genes. Conversely, if the G+C content is low, a low proportion of the codons are expected to end in G or C. In *C. taeniospiralis* the G+C content is between 40 and 44%, and therefore a low percentage of G's and C's in the third position of the codons is expected. This holds for all four ORFs of the *reb* locus. The distribution of bases in the four ORFs is presented in Table 3.

Amino acid composition and hydropathy. RebA consists of 114 amino acid residues with a predicted molecular mass of 11.6 kDa. Of the amino acid residues, 56% are Ala, Gln, Met, Ser, or Thr. There are no Cys, Glu, Phe, Pro, or Trp residues, and 33% of the residues have hydrophobic characteristics. A plot depicting the hydrophobic character of RebA is presented in Fig. 3. This profile suggests a protein with alternating hydrophobic and hydrophilic regions. The calculated isoelectric point (pI) for the protein is 3.393.

TABLE 3. Distribution of bases in the ORFs of the *reb* locus

Codon position	% G+C in: *rebA*	*rebB*	*rebC*	4th ORF
1st	49.7	45.3	50.8	51.8
2nd	46.0	45.3	38.5	37.3
3rd	23.5	23.6	33.3	31.4
Total	40.0	38.0	41.0	40.2

The *rebB* ORF would encode a polypeptide of 105 amino acid residues with a predicted molecular mass of 11.0 kDa. Of the amino acid residues, 53% are Ala, Gln, Met, Ser, or Thr. There are no Cys or Trp residues, and 38% of the residues possess hydrophobic characteristics. The hydropathy plot of RebB reveals a peptide, similar to *rebA*, with alternating hydrophobic and hydrophilic regions (Fig. 3). The calculated pI is 6.998.

Of the 56 residues in RebC, 45% are Ala, Gln, Met, Ser, or Thr. There are no Trp residues, and 30% of the residues are hydrophobic. The hydropathy plot predicts a hydrophilic N terminus, a hydrophobic core, and a hydrophilic C terminus (Fig. 3). The calculated pI is 4.068.

If the fourth ORF situated between *rebB* and *rebC* is translated, the protein would have an amino acid composition similar to that of the other three proteins. Of the residues, 46% would be Ala, Gln, Met, Thr, or Ser. No Trp residues would be encoded, and 34% of the residues would be hydrophobic. The calculated pI of this predicted polypeptide is 5.283.

Sequence homology. Both the nucleic acid and amino acid sequences were compared with GenBank and NBRF data files for similarities with known sequences. No strong similarities exist. However, when the amino acid sequences for the three proteins are compared with one another, several interesting similarities are noted. The alignment of RebA with RebB reveals 55% similarity over a region of 98 amino acids (Fig. 4); 32% of the residues are conserved amino acid changes. This finding suggests that a gene duplication occurred to form *rebB*. Analysis of the DNA sequences for *rebA* and *rebB* shows a remarkable 63.1% similarity over a region of 306 nucleotides (Fig. 4). The 113 nucleotides that differ between the two sequences consist of 47 transitions and 66 transversions. The

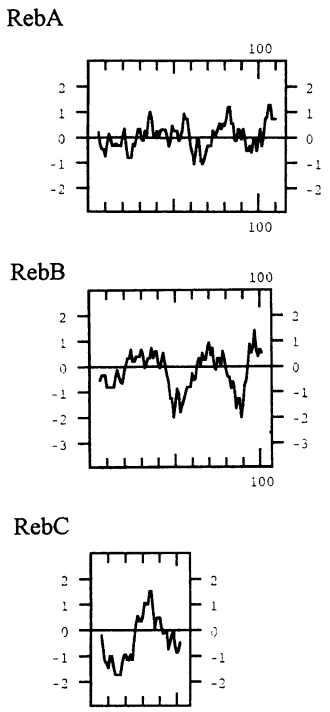

FIG. 3. Hydrophobicity of the three proteins encoded by the *reb* locus. The program DNA Strider was used to plot hydrophobicity by using the Kyte-Doolittle algorithm. Hydrophobic sequences are above the axis, and hydrophilic sequences are below.

```
rebA: ATG GCG ACT ACT ACT TCT ACT GCT GAT GTA GGC AAC ACA ACT GAC
       m   a   t   t   t   s   t   a   d   v   g   n   t   t   d
rebB:                                                   ATG AGT
                                                          m   s

rebA: ATC GTA AAT AGC CAA ATT ACT GAT TCA GTA ACA CAA ACT AAC ACA
       i   V   N   S   Q   I   T   D   S   V   T   Q   T   N   T
rebB: AAT GTA AAT TCA CAA ATT ACA GAT TCA GTA ACA CAA ACA AAC ACT
       n   V   N   S   Q   I   T   D   S   V   T   Q   T   N   T

rebA: ATG GTT CTT GGT TCT ACT CCA GCA CAA ATC ATG ACA AAT TTA ATG
       m   v   L   G   s   t   P   A   q   i   m   t   n   L   M
rebB: AAA ATA TTA GGT GAA ATG CCT GCG TTT ACC ACT GGA AGC TTA ATG
       k   i   L   G   e   m   P   A   f   t   t   g   s   L   M

rebA: CAA ATG AGC GCA CAA GCT AAT GCT TTA GCA ATG CAA AAT GCT GTT
       Q   M   s   a   Q   A   n   a   L   a   m   Q   N   a   V
rebB: CAA ATG GCA ACA CAA GCT GCT GGC TTA TCT ATA CAA AAC TCT GTA
       Q   M   a   t   Q   A   a   g   L   s   i   Q   N   s   V

rebA: GTT GGT CAA AAA CAA ACA AAT ATG CTA TCA GAT GCA GCA ACC ACT
       v   g   Q   k   Q   t   N   M   L   s   d   A   a   T   T
rebB: ACC AAT CAG CAA CAA TCT AAT ATG CTT CAT CAG GCA TCA ACC ACT
       t   n   Q   q   Q   s   N   M   L   h   q   A   s   T   T

rebA: CAA GGC TTA TCT ATT ATG TAT ACG CTT TCA GCT GCT TCA GAT GGT
       Q   G   l   S   I   m   Y   t   l   s   a   A   s   d   g
rebB: CAA GGT ATG AGT ATT CTA TAT TCT GTT GAC ACG GCA GCT AAT GCT
       Q   G   m   S   I   l   Y   s   v   d   t   A   a   n   a

rebA: CAA GCA GTC AGT ACT GTT GAC AAT TCA ACT GAC ATG GCT GGA CTG
       Q   A   v   s   t   V   d   n   S   t   D   m   a   g   L
rebB: CAA GCC ATT GGT TCA GTG AAC CGT TCT AAT GAC ACA AGC CGT TTA
       Q   A   i   g   s   V   n   r   S   n   D   t   s   r   L

rebA: ATT GAT GCG CTT GGC GTT GAT AAA GCA TAA
       i   D   A   L   g   V   d   K   A   *
rebB: ACT GAT GCC CTT GCT GTT ATT AAA GCC GCT AAA AAT GGT TAA
       t   D   A   L   a   V   d   K   A   a   k   n   g   *
```

FIG. 4. Alignment of the DNA sequence and corresponding amino acid sequences for *rebA* and *rebB*. Identical nucleotide sequences are underlined. Identical amino acid residues are capitalized.

hydropathy plots predict similar secondary structures for these regions of the polypeptides.

Expression of the complete *reb* locus. To determine the events of the polymerization processes, the proteins encoded by pBQ89 were radiolabeled in minicells. The labeled proteins were analyzed by two-dimensional gel electrophoresis (Fig. 5A). As previously reported, 3 species of RebA and 12 species of RebB can be identified (8). RebA exists as three species with different molecular weights. RebB has two molecular weights, each with six isoelectric points. Evidence for two polymerization patterns was obtained through analysis of pulse-chase experiments with two-dimensional gels. One polymerization event manifests itself as a ladder of polypeptides in the basic portion of the gel. The other polymerization event results in a ladder of polypeptides that begins in the acidic portions of the gel and becomes more basic as the molecular weights of the polymerization products increase. Several species of proteins with different pIs exist for each stage of the polymerization process. The basic polymerization products are the faster-migrating "minor proteins" observed by Kanabrocki et al. (8), whereas the major products are those whose pIs shift from acidic to basic as polymerization proceeds (8).

Expression of *rebB* and *rebC* in the absence of *rebA*. The major structural subunit of R bodies appears to be RebB (8). To determine the role of RebA and RebC in the polymerization of RebB, the proteins encoded by pBQ70 were labeled in minicells. pBQ70 is a deletion derivative of pBQ65 that contains only *rebB* and *rebC* (8). In the absence of RebA, only the basic polymerization event occurs (Fig. 5B). At each stage of the basic polymerization event, several species of proteins with different pIs can be identified. Twelve species of RebB and one species of RebC are present.

Pulse-chase labeling of deletion derivatives. To further analyze the role of each individual gene product in the polymerization event, proteins encoded by deletion derivatives of pBQ65 were labeled in minicells for 1 min and then chased for various lengths of time. pBQ69 codes for a truncated version of RebA (*rebB*, *rebC*, and the 3′ end of *rebA* were deleted). Pulse-chase analysis of pBQ69 shows that the truncated version of RebA migrates at 15 kDa (Fig. 6A). Analysis of pBQ59, which encodes RebA and RebB, shows each protein migrating as a single-molecular-mass species (Fig. 6B). Pulse-chase analysis of pBQ70 reveals that RebB migrates at two molecular masses when RebC is present (Fig. 6C).

DISCUSSION

DNA hybridization, restriction endonuclease mapping, and immunological studies had previously demonstrated strong similarities among pKAP plasmids (20, 25) and among type 51 R bodies (7, 15) from various strains of *C. taeniospiralis*. This study was undertaken with the expectation that some differences may exist between the DNA sequences of the *reb* locus from the two strains studied. Regions of conserved DNA sequence might then suggest corresponding regions of amino acid sequence that are critical to R body assembly. However, this study revealed identical nucleotide sequences for the refractile body (*reb*) locus in two strains of *C. taeniospiralis* (strains 47 and 116).

The remarkable conservation of these genes is quite interesting when one considers that strain 47 was isolated in California and strain 116 was isolated in Indiana. Even more puzzling is the observation that the R body-coding plasmid (pKAP169) from *C. taeniospiralis* 169 (isolated in Japan) has the same restriction enzyme map as pKAP116 (25). The

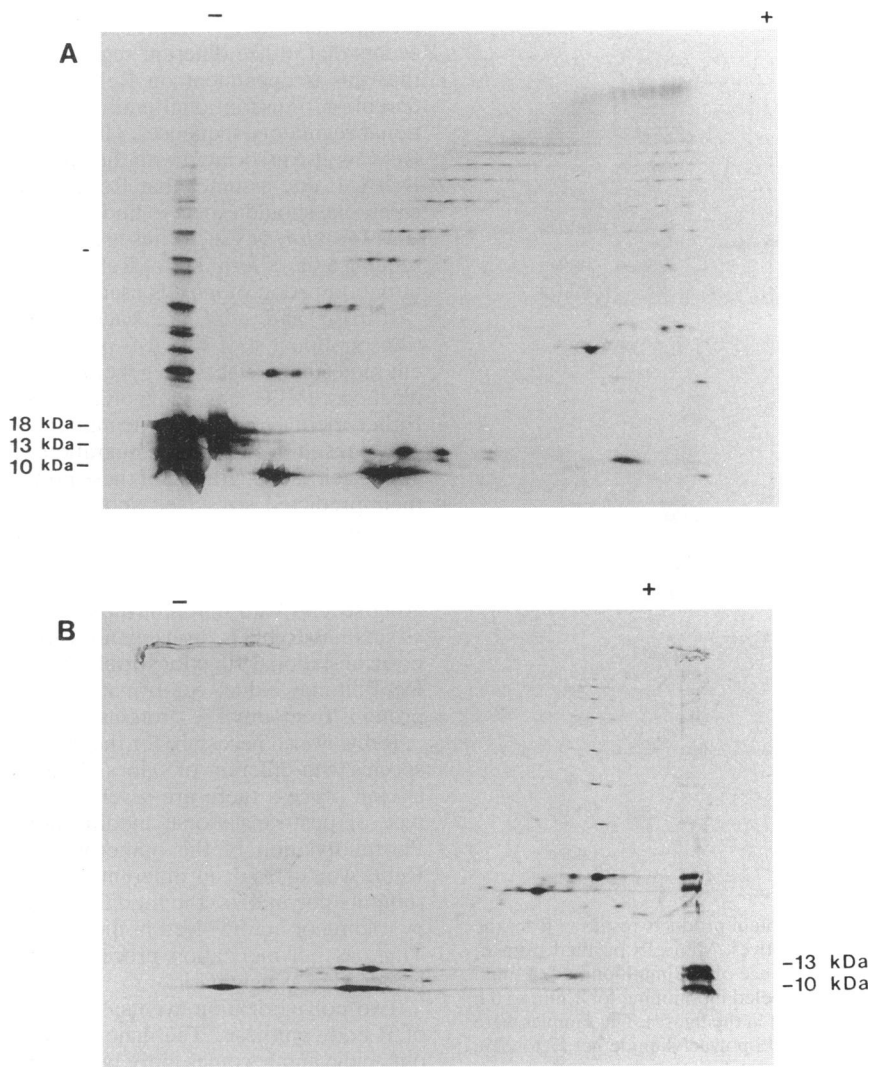

FIG. 5. Two-dimensional electrophoresis of proteins encoded by the recombinant plasmids pBQ89 and pBQ70. Minicells containing the plasmids were labeled with [^{35}S]methionine for 45 min. The proteins were separated by isoelectric focusing and then separated in a second direction on an SDS-polyacrylamide gel (5 to 22% gradient). The basic pIs (+) are to the right, and the acidic pIs (−) are to the left. The cloning vector encodes β-lactamase, which can be detected as two species (molecular masses of 28 and 30 kDa) in the basic edge of each gel. (A) Two-dimensional analysis of proteins encoded by pBQ89 (encodes the three proteins necessary for complete assembly of type 51 R bodies: RebA, RebB, and RebC). The standard lane (left) contains a one-dimensional profile of pBQ89-encoded proteins radiolabeled in minicells. Note the characteristic ladderlike profile. Three species of RebA exist in the acid region of the gel. RebB, which exists as species with two different molecular weights and at least six pIs, can be seen in the bottom center region of the gel. Two polymerization events are occurring. One polymerization event is basic, and the other starts out acidic and becomes progressively more basic as the molecular weight of the complex increases. Several species of proteins differing in pI can be identified at each level of polymerization. (B) Two-dimensional electrophoresis of proteins encoded by pBQ70 (encodes RebB and RebC). The standard lane (right) contains a one-dimensional profile of pBQ70-encoded proteins radiolabeled in minicells. Only the basic polymerization event can be detected. RebB exists as species with two different molecular weights and with several pIs. Modified from reference 16 and reprinted with permission.

geographic distribution of these strictly freshwater, non-cyst-forming organisms suggests that the various strains have been isolated for countless generations. Identical nucleotide sequences suggest that it is possible that R body structure and function requires an extremely high level of sequence conservation. This is not unreasonable considering that the proteins are small and the polymerization event is complex. It is also possible, however, that the geographic distribution of these symbiont-containing paramecia is of more recent origin, perhaps as a result of human activity. That is, *C. taeniospiralis* 47 and 116 may in fact be only recently separated (see reference 16 for further discussion).

In *C. taeniospiralis* 47 and 116, three genes at the *reb* locus, *rebA*, *rebB*, and *rebC*, have been identified and sequenced. Associated with each gene are promoter and RBS regions which correspond to *E. coli* consensus sequences. These findings confirm the prediction that each gene acts as an independent transcriptional unit (8). Codon usage is consistent among all three genes and is not unusual when compared with that of *E. coli*. The DNA sequence-based prediction that RebA and RebC are acidic and that RebB is neutral was confirmed by observed migration of the polypeptides in isoelectric focusing gels.

DNA sequence analysis predicts that no Cys residues are

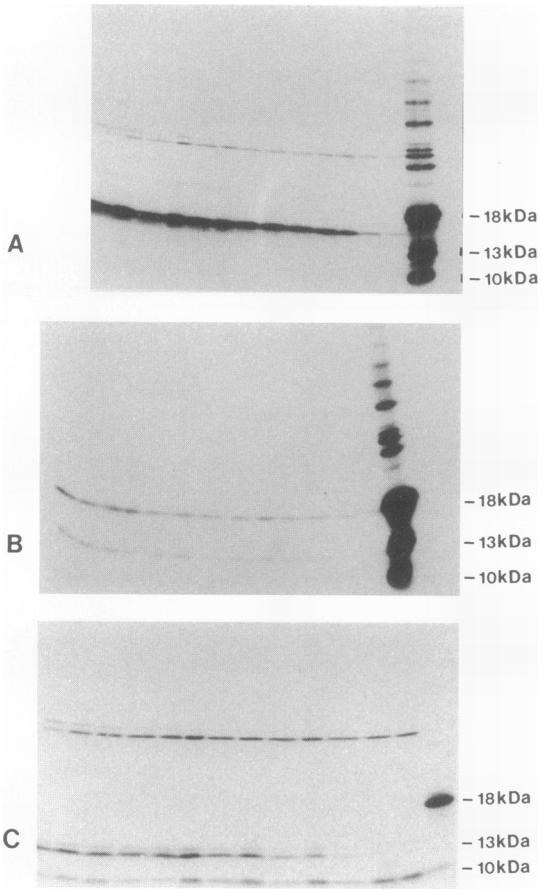

FIG. 6. Pulse-chase analysis of protein products resulting from the expression of pBQ65 deletion derivatives. Minicells purified from *E. coli* P678-54 were labeled in the presence of [^{35}S]methionine for 1 min (leftmost lane) and chased with unlabeled methionine for 1 min to 6 h (chase time increases from left to right in the figure). The samples were resolved by electrophoresis on an SDS-polyacrylamide gel (7 to 20% gradient). The standard lane (rightmost lane) contains proteins encoded by pBQ65 (which contains the complete *reb* locus) labeled with [^{35}S]methionine. (A) Pulse-chase of pBQ69 which codes for a truncated version of RebA. The only proteins labeled are the truncated version of RebA, which migrates at 15 kDa, and the vector-encoded β-lactamase, migrating at 28 and 30 kDa. (B) Pulse-chase of pBQ59 which encodes RebA (18 kDa) and RebB (13 kDa). No polymerization or modification events have occurred. Both proteins migrate at only one molecular weight. No β-lactamase can be detected in this exposure. (C) Pulse-chase of pBQ70 which encodes RebB (13 kDa) and RebC (10 kDa). RebB exists as two species with different molecular weights, while RebC is present at one molecular weight. Both species of β-lactamase (28 and 30 kDa) can be identified.

encoded by *rebA* and *rebB* and that *rebC* encodes only one Cys residue. This analysis, plus the inability of standard denaturing treatments to completely dissociate type 51 R bodies (8), indicates that covalent linkages other than disulfide bridges are involved. The hydropathy profiles of the three proteins do not predict that any regions will strongly associate with lipids.

Three possible promoter regions for *rebB* were identified. They precede regions of the *rebB* ORF that would encode 105 amino acids, 82 amino acids, and either 74 or 72 amino acids (depending on which of two possible start codons is used). This is interesting because in the presence of RebC, two species of RebB with different molecular weights are present. It is possible that the two RebB species result from transcriptional events that utilize different regulatory sequences for *rebB* and that this is dependent on RebC. That is, RebC may act to derepress transcriptional inhibition of one of these transcriptional regulatory sequences. However, it should be noted that RebC is also associated with the appearance of three species of RebA. If one assumes that RebC would act similarly in both cases, one should expect to find alternative transcriptional start sites for *rebA*. No such sites are found. It will be necessary to sequence the N terminus of RebB to precisely determine which initiation codon of *rebB* is used.

If *rebB* and *rebA* are each transcribed from only one transcriptional start site, it is possible that species with different molecular weights are generated by posttranslational modifications; RebC may be involved in the modification process. Either methylation or the trimming of terminal amino acids would result in a change of migration within an SDS-polyacrylamide gel. The migration of these polypeptides does not reflect their predicted sizes (see above), and anomalous migration patterns are seen when these polypeptides are electrophoresed in gels with different percentages of acrylamide (8), making it difficult to assess which processes might be involved in generating species with different molecular weights. A polypeptide as small as RebC is probably not capable of carrying out such posttranslational functions itself, but, by binding to RebA or RebB, it may induce conformational changes which allow host proteins to modify the proteins.

RebC is also necessary for the modification of RebB into six species with different pI values. At each stage of the polymerization process there are several species differing in pI. The type of posttranslational modification involved is unknown. Phosphorylation of the numerous Ser and Thr residues in RebB would result in different pIs. It is possible that as the subunits polymerize, the modified acidic residues are blocked by folding or are involved in the covalent linkages themselves. Thus, as polymerization proceeds, the R body complex becomes increasingly basic.

Two polymerization events, "major" and "minor," are part of R body synthesis. The major polymerization process starts out acidic and becomes more basic as the sizes of the polymerized products increase. The complexes of the minor polymerization process are basic and migrate slightly faster than those of the major polymerization event. All three proteins are required for the major polymerization event, but only RebB and RebC are necessary for the basic polymerization process. To form the ribbon structure, there must be polymerization in both length and width. The two polymerization events identified may represent polymerization events in each dimension.

Width is uniform (0.5 μm) in type 51 R bodies, while length varies. How this ordered assembly of the R body occurs is not known. It has been suggested that RebA may act as a scaffolding protein to facilitate the major polymerization process (16). The identity in amino acid sequence between RebA and RebB suggests a similar structure and function. Like RebB, RebA is modified into two or more species with different molecular weights before the major polymerization event occurs. The polymerization of RebB may involve an array of scaffolding proteins. Comparison of the hydropathy plots for RebB with those for RebA suggests a similar secondary structure for these regions. Therefore, RebA should be capable of entering into either a temporary or permanent association with the polymerized complexes of RebB. If the acidic RebA proteins are linked to the growing R body complex (as scaffolding or at the site of polymerization), they may be responsible for the increased pIs during the major polymerization event. That is, the proportional contribution of RebA may decline as polymerization

proceeds, resulting in a shift of the pIs. Pulse-chase analysis of the protein products encoded by pBQ65 reveals that, like RebB, the modified faster-migrating species of RebA decrease in concentration over time as the higher-molecular-weight polymerization complexes are formed (8). This is evidence that RebA proteins may associate directly with the polymerization complexes.

The role of a fourth protein in the synthesis and assembly of R bodies has been postulated (8). This study has identified a fourth ORF within the *reb* locus. The regulatory sequences, codon usage, and amino acid composition are similar to those of the three identified genes. However, radiolabeling of proteins encoded by plasmids in minicells has not revealed any translation product from this region. Analysis of the pBQ65 deletion derivative pBQ72, which deletes *rebB*, shows that only *rebA* and *rebC* protein products are translated (8). Therefore, the product of the fourth ORF is not masked by comigration with RebB. It is possible that this ORF, which is either not transcribed or not translated in *E. coli*, is functional in *C. taeniospiralis*. Another possibility is that this ORF (which has a relatively low CAI value) is expressed at a low frequency or for only a short period and was therefore not identified by our assay. The presence of the stop codon for the fourth ORF immediately downstream of the RBS for *rebC* raises the possibility of translational coupling and suggests a means of regulating levels of the two gene products. This is an interesting possibility since RebC appears to trigger synthesis and assembly of R bodies. However, since R bodies can be synthesized in *E. coli* in the apparent absence of the putative protein, it may not be necessary for R body production.

Our present understanding of R body synthesis is still too limited to allow us to present a model for polymerization. However, now that the nucleotide sequence for the *reb* locus has been determined, it should be possible, by using the technique of site-directed mutagenesis, to undertake a detailed analysis of the polymerization event. In addition, delineating the posttranslational modifications of the gene products should aid in further characterization of this unique protein structure.

ACKNOWLEDGMENTS

We thank Jean Pond, Paul Rothberg, and Joe Kanabrocki for helpful discussions.

This research was supported by Public Health Service grant GM36293 awarded to R.L.Q. from the National Institutes of Health and Public Health Service grant 1R15 GM46045-01 awarded to F.R.P. from the National Institutes of Health.

REFERENCES

1. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:**1513–1523.
2. **Clewell, D. B., and D. R. Helinski.** 1969. Supercoiled circular DNA-protein complex in *Escherichia coli*: purification and induced conversion to an open circular DNA form. Proc. Natl. Acad. Sci. USA **62:**1159–1166.
3. **Dilts, J. A.** 1976. Covalently closed circular DNA in kappa endosymbionts of *Paramecium*. Genet. Res. **27:**161–170.
4. **Hawley, D. K., and W. R. McClure.** 1983. Comparison and analysis of *Escherichia coli* promoter sequences. Nucleic Acids Res. **11:**2237–2249.
5. **Hodel-Christian, S. L.** 1988. Ph.D. thesis. University of South Dakota, Vermillion.
6. **Jurand, A., B. M. Rudman, and J. R. Preer, Jr.** 1971. Prelethal effects of killing action by stock 7 of *Paramecium aurelia*. J. Exp. Zool. **177:**365–388.
7. **Kanabrocki, J. A., J. Lalucat, B. J. Cox, and R. L. Quackenbush.** 1986. Comparative study of refractile (R) bodies and their genetic determinants: relationship of type 51 R bodies to R bodies produced by *Pseudomonas taeniospiralis*. J. Bacteriol. **168:**1019–1022.
8. **Kanabrocki, J. A., R. L. Quackenbush, and F. R. Pond.** 1986. Organization and expression of genetic determinants for synthesis and assembly of type 51 R bodies. J. Bacteriol. **168:**40–48.
9. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
10. **Marck, C.** 1988. 'DNA Strider': A 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. **16:**1829–1836.
11. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.
12. **Meagher, R. R., R. C. Tait, M. Betlach, and H. W. Boyer.** 1977. Protein expression in *E. coli* minicells by recombinant plasmids. Cell **10:**521–536.
13. **Mueller, J. A.** 1963. Separation of kappa particles with infective activity from those with killing activity and identification of the infective particles in *Paramecium aurelia*. Exp. Cell Res. **30:**492–508.
14. **Norrander, J., T. Dempe, and J. Messing.** 1983. Improved M13 vectors using oligonucleotide-directed mutagenesis. Gene **26:**101–106.
15. **Pond, F. R., J. A. Burbach, and B. Wells.** 1987. Relationships among the R bodies of caedibacteria. Micron Microsc. Acta **18:**153–158.
16. **Pond, F. R., I. Gibson, J. Lalucat, and R. L. Quackenbush.** 1989. R body-producing bacteria. Microbiol. Rev. **53:**25–67.
17. **Preer, J. R., Jr., L. B. Preer, and A. Jurand.** 1974. Kappa and other endosymbionts in *Paramecium aurelia*. Bacteriol. Rev. **38:**113–163.
18. **Preer, L. B., A. Jurand, J. R. Preer, Jr., and B. M. Rudman.** 1972. The classes of kappa in *Paramecium aurelia*. J. Cell Sci. **11:**581–600.
19. **Quackenbush, R. L.** 1978. Genetic relationships among bacterial endosymbionts of *Paramecium aurelia*: deoxyribonucleotide sequence relationships among members of *Caedibacter*. J. Gen. Microbiol. **108:**181–187.
20. **Quackenbush, R. L.** 1983. Plasmids from bacterial endosymbionts of hump-killer paramecia. Plasmid **9:**298–306.
21. **Quackenbush, R. L.** 1987. R body-producing bacteria. Micron Microsc. Acta **18:**37–40.
22. **Quackenbush, R. L.** 1987. The killer trait, p. 406–418. *In* H. D. Gortz (ed.), Current topics in *Paramecium* research. Springer-Verlag KG, Berlin.
23. **Quackenbush, R. L., and J. A. Burbach.** 1983. Cloning and expression of DNA sequences associated with the killer trait of *Paramecium tetraurelia* stock 47. Proc. Natl. Acad. Sci. USA **80:**250–254.
24. **Quackenbush, R. L., B. J. Cox, and J. A. Kanabrocki.** 1986. Extrachromosomal elements of extrachromosomal elements of *Paramecium* and their extrachromosomal elements, p. 265–278. *In* A. Hollaender et al. (ed.), Extrachromosomal elements of lower eukaryotes. Plenum Publishing Corp., New York.
25. **Quackenbush, R. L., J. A. Dilts, and B. J. Cox.** 1986. Transposonlike elements in *Caedibacter taeniospiralis*. J. Bacteriol. **166:**349–352.
26. **Roozen, K. J., R. G. Fenwick, Jr., and R. Curtiss III.** 1971. Synthesis of ribonucleic acid and protein in plasmid-containing minicells of *Escherichia coli* K-12. J. Bacteriol. **107:**21–33.
27. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
28. **Sanger, F. S., and A. R. Coulson.** 1978. The use of thin acrylamide gels for DNA sequencing. FEBS Lett. **87:**107–110.
29. **Sharp, P. M., and W. H. Li.** 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. Nucleic Acids Res. **14:**7737–7749.
30. **Shine, J., and L. Dalgarno.** 1977. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **76:**575–579.
31. **Stormo, G. D., T. D. Schneider, and L. M. Gold.** 1982. Characterization of translational initiation sites in *E. coli*. Nucleic Acids Res. **10:**2971–2996.
32. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.