JOURNAL OF BACTERIOLOGY, Sept. 1992, p. 5540–5548 Vol. 174, No. 17
0021-9193/92/175540-09$02.00/0

A *Coxiella burnetii* Repeated DNA Element Resembling a Bacterial Insertion Sequence

TIMOTHY A. HOOVER,[1]* MICHAEL H. VODKIN,[2,3] AND JIM C. WILLIAMS[4]

Bacteriology Division, U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Frederick, Maryland 21702-5011[1]; Department of Veterinary Pathobiology, School of Veterinary Medicine, Purdue University, West Lafayette, Indiana 47907-1027[2]; Department of Veterinary Pathobiology, University of Illinois, Urbana, Illinois 61801[3]; and Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, Maryland 20892[4]

Received 2 April 1992/Accepted 20 June 1992

A DNA fragment located on the 3′ side of the *Coxiella burnetii htpAB* operon was determined by Southern blotting to exist in approximately 19 copies in the Nine Mile I genome. The DNA sequences of this *htpAB*-associated repetitive element and two other independent copies were analyzed to determine the size and nature of the element. The three copies of the element were 1,450, 1,452, and 1,458 bp long, with less than 2% divergence among the three sequences. Several features characteristic of bacterial insertion sequences were discovered. These included a single significant open reading frame that would encode a 367-amino-acid polypeptide which was predicted to be highly basic, to have a DNA-binding helix-turn-helix motif, to have a leucine zipper motif, and to have homology to polypeptides found in several other bacterial insertion sequences. Identical 7-bp inverted repeats were found at the ends of all three copies of the element. However, duplications generated by many bacterial mobile elements in the recipient DNA during insertion events did not flank the inverted repeats of any of the three *C. burnetii* elements examined. A second pair of inverted repeats that flanked the open reading frame was also found in all three copies of the element. Most of the divergence among the three copies of the element occurred in the region between the two inverted repeat sequences in the 3′ end of the element. Despite the sequence changes, all three copies of the element have retained significant dyad symmetry in this region.

Coxiella burnetii is the bacterial agent that causes Q fever, a widely disseminated illness with a broad range of susceptible hosts. The microorganism is an obligate intracellular parasite that replicates only in the phagolysosomal compartments of nucleated cells (1, 6, 7, 23). Although virulence factors such as lipopolysaccharide (22, 54), acid activation (21, 23–25, 30), and immunomodulatory complex (52) are identified, little is known concerning the genetic basis of virulence of *C. burnetii* because of the inability to grow the microorganism axenically. Therefore, application of classical genetic techniques, such as mutagenesis and transformation, to the study of genes and gene products involved in virulence has not been easily pursued. However, the availability of specific antibodies, enzyme assays, and certain *Escherichia coli* mutants makes possible the isolation and characterization of several *C. burnetii* genes whose gene products are antigens or are involved in intermediary metabolism (27, 28, 32, 42, 51).

Two of the known *C. burnetii* genes comprise the heat shock-inducible *htpAB* operon, which encodes protein homologs of *E. coli groES* and *groEL* (51). The latter gene encodes the so-called common antigen polypeptide, a ubiquitous protein with a high degree of similarity among a number of bacterial *groEL* gene products (13, 48, 51). In the cloning of *C. burnetii htpAB*, multiple bands of genomic DNA, in addition to the appropriately sized ones, were detected in Southern blots when subcloned DNA (pCS26C1) containing the operon and approximately 2 kbp of flanking sequence was used as a probe (50). This result suggested that a repetitive DNA element was linked to *C. burnetii htpAB*.

Repeated DNA elements have been described for many bacterial and eucaryotic organisms and are of a variety of sizes and functions. These elements can have various effects, including modulation of gene expression, induction of chromosomal deletions and inversions, and generation of antigenic variation. A repetitive element associated with *groEL* of *Mycobacterium leprae*, another obligate intracellular parasite, has also been described (11, 20). It has likewise been shown to reside downstream of the common antigen gene, is present in at least 15 copies per genome, and has features of bacterial insertion sequences (16, 19). Insertion sequences, which are but one of several types of repetitive elements found in bacteria, are transposable elements found in many different bacteria, capable of mobilizing themselves as discrete units or often as integral parts of transposons. Ranging in size from 800 to 2,500 bp, they are present in from one to hundreds of copies per genome. Recent reports indicate the utility of the insertion sequence as a target for the detection of *M. leprae* by means of the polymerase chain reaction (PCR) (57). In the current study, the region of the 3′ side of *C. burnetii htpB* was sequenced to determine the nature of this repetitive DNA element.

MATERIALS AND METHODS

Bacterial strains and plasmids. Total genomic *C. burnetii* DNA was purified from the phase I Nine Mile clone 7 strain. Growth and purification of *C. burnetii* cultures were as previously described (55). Subcloning of the *htpAB* operon from a cosmid (pHC79) library of *C. burnetii* DNA, resulting in pCS26C1, was described previously (51). *E. coli* strain DH5αF′ (Bethesda Research Laboratories Life Technologies, Gaithersburg, Md.) was used in cloning and DNA

* Corresponding author.

template preparation. Plasmid vector pBluescript KS (Stratagene Cloning Systems, La Jolla, Calif.) was used to clone the insertion sequence copies. DNA fragments were sequenced in plasmid pBluescript KS and bacteriophage vectors M13mp18 and M13mp19 (Boehringer Mannheim Corp., Indianapolis, Ind.). *E. coli* was grown in LB medium supplemented with 100 μg of ampicillin per ml to select plasmids used in this study.

DNA isolation, purification, and analysis. Total genomic DNA was prepared and purified in anionic chromatography columns supplied with A.S.A.P. kits (Boehringer Mannheim) to obtain high-molecular-weight DNA. Plasmids were purified by an alkaline lysis procedure (3). Southern blots were performed by transferring DNA fragments onto 0.45-μm Magnagraph Nylon 66 (Micron Separations, Inc., Westboro, Mass.) membrane filters in a Stratagene Cloning Systems model Posiblot Pressure Blotter with 0.4 N NaOH as transfer buffer. DNA probes were labeled with the Boehringer Mannheim Genius system, which incorporates digoxigenin-labeled nucleotides by DNA polymerase (Klenow fragment) into random primer-directed DNA fragments. DNA-DNA hybridizations, antidigoxigenin antibody and enzyme-linked secondary antibody binding, and color developments were carried out according to the manufacturer's specifications. DNA was sequenced by the Applied Biosystems, Inc., model 381 automated sequencer. Universal primers labeled with fluorescent dyes were used with *Taq* DNA polymerase in the Sanger method for DNA sequencing (45). Double-stranded templates were primarily used because of the instability of these fragments in M13. To verify two single-base uncertainties in IS*1111a*, fragments were generated from genomic DNA by PCR amplification with oligonucleotides specific for IS*1111* and *htpB* in a Perkin-Elmer Cetus DNA Thermal Cycler and sequenced by the Applied Biosystems, Inc., *Taq* cycle sequencing procedure. The sequences were analyzed for open reading frames (ORFs), secondary structures, direct and inverted repeats, predicted secondary structures of the encoded polypeptide, and DNA or amino acid homology with other sequences in the GenBank and EMBL data bases with programs from the University of Wisconsin Genetics Computer Group software package (14). Putative transposases were aligned with the CLUSTAL program of PCGene from Intelligenetics, Inc. (31). Assignment of the designation IS*1111* was made by Esther Lederberg at the Plasmid Reference Center, Stanford University School of Medicine, Stanford, Calif.

Nucleotide sequence accession number. The DNA sequence data reported here have been assigned GenBank accession number M80806.

RESULTS

Location and boundaries of the *htpAB*-associated repetitive DNA element and its chromosomal distribution. Plasmid pCS26C1 contains a 4.3-kbp segment of the *C. burnetii* chromosome, which includes the *htpAB* operon (51) plus 500 bp of upstream and 2,000 bp of downstream flanking sequences. DNA probes were made from various pCS26C1 restriction endonuclease fragments isolated from low-melting-point agarose gels. Probes containing at least part of the repetitive element were identified by Southern blot analysis on the basis of the appearance of multiple bands on the hybridized, color-developed filter (data not shown). The repetitive element was determined to be downstream of *htpB*, extending for at least 1,200 bp. Probes strictly from within or upstream of the *htpAB* operon, in contrast, hybridized to single genomic restriction fragments.

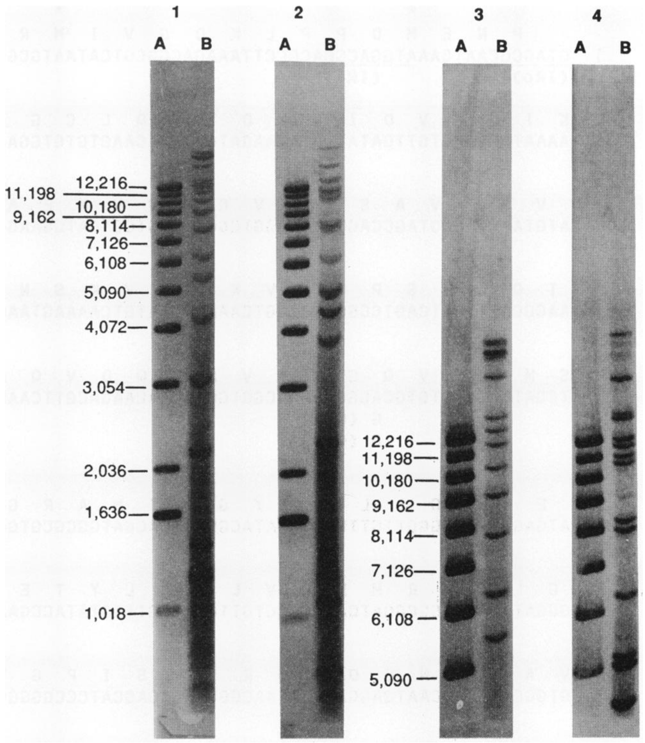

FIG. 1. Southern blot analysis of the distribution of IS*1111* in the *C. burnetii* Nine Mile I genome. Total DNA was digested with *Sma*I and electrophoresed through a 0.8% agarose gel (panels 1 and 2) or a 0.55% agarose gel (panels 3 and 4) and blotted onto nylon membranes. Panels 1 and 3 were hybridized with digoxygenin-labeled molecular weight markers and an IS*1111a* fragment 5' to its internal *Sma*I site. Panels 2 and 4 are the same blots as those in panels 1 and 3, respectively, with previous dye and probes removed by filter washings. For panels 2 and 4, hybridization was with labeled molecular weight markers and an IS*1111a* fragment 3' to its internal *Sma*I site. In all panels, lanes A are molecular weight markers and lanes B are *Sma*I-digested *C. burnetii* total DNA.

To obtain an accurate estimate of the number of copies of the repetitive element, we performed Southern blots with various digests of *C. burnetii* genomic DNA in an attempt to obtain a well-separated banding pattern. Restriction enzymes with single sites in the repetitive element were chosen, as these would yield genomic DNA fragments with only one partial copy of the repetitive element at each end. Probes from either side of a particular restriction site within the repetitive element could, therefore, be used to count the number of genomic copies of the repetitive element. Because anomalies could arise from elements that had lost the restriction site, or if DNA fragments of nearly identical size were generated, several sites were used in the enumeration. Southern blots of *C. burnetii* genomic DNA digested with *Sma*I (Fig. 1), *Sst*I, *Xho*I, and *Sst*I plus *Xho*I (data not shown) revealed about 19 fragments that hybridized with the repetitive element probes.

Nucleotide sequences of three independent copies of the repetitive element. The nucleotide sequences of the *htpAB*-associated repetitive element (IS*1111a*) and two independent copies that mapped elsewhere are presented in Fig. 2. When we analyzed the DNA sequence of IS*1111a*, we predicted

```
                   *                             *                                            *
          P  N  E  M  D  P  P  L  K  D  G  V  I  M  R  Q  H  R  I  S  I  F  K  K  R  R  R  S  M  K  D  I
   1  GTAGCCCAATGAAATGGACCCACCCCTTAAAGACGGCGTCATAATGCGCCAACATAGAATTTCTATTTTCAAAAAAAGGAGAAGGTCCATGAAAGATATT   100
      (IRo)>>         (IRi)>>

      K  I  L  G  V  D  I  A  K  D  V  F  Q  L  C  G  I  D  E  W  G  K  V  I  Y  T  R  R  V  K  R  A  Q  Y
 101  AAAATACTGGGTGTTGATATTGCAAAAGATGTTTTTCAACTGTGTGGAATTGATGAGTGGGGTAAAGTGATCTACACGAGACGGGTTAAGCGTGCTCAGT   200

        V  S  T  V  A  S  L  K  V  G  C  V  V  M  E  A  C  G  G  A  N  H  W  Y  R  T  F  M  G  M  G  I  P
 201  ATGTATCCACCGTAGCCAGTCTTAAGGTGGGCTGCGTGGTGATGGAAGCGTGTGGAGGAGCGAACCATTGGTATCGGACGTTTATGGGGATGGGTATCCC   300

       T  Q  L  I  S  P  Q  H  V  K  P  Y  V  K  S  N  K  N  D  R  N  D  A  Q  A  I  A  E  A  A  S  R  A
 301  AACGCAGTTGATCAGTCCGCAGCACGTCAAACCGTATGTCAAAAGTAACAAGAATGATCGTAACGATGCGCAGGCGATAGCTGAAGCGGCTTCCCGCGCC   400

      S  M  R  F  V  Q  G  K  T  V  E  Q  Q  D  V  Q  A  L  L  K  I  R  D  R  L  V  K  S  R  T  A  L  I  N
 401  TCGATGCGGTTTGTGCAGGGTAAAACGGTGGAACAACAAGACGTTCAAGCGCTGTTAAAGATACGCGATCGTTTAGTCAAAAGCCGCACGGCGCTGATCA   500
                      G (Q-R)
                      G (Q-R)

        E  I  R  G  L  L  Q  E  Y  G  L  T  M  A  R  G  A  K  R  F  Y  E  E  L  P  L  I  L  A  S  E  A  V
 501  ATGAGATTCGGGGGTTGTTGCAAGAATACGGACTCACGATGGCGCGTGGTGCCAAGCGATTTTATGAAGAGCTCCCGTTGATTTTAGCGAGCGAAGCGGT   600

       G  L  T  P  R  M  K  R  V  L  N  C  L  Y  T  E  L  L  N  R  D  E  A  I  G  D  Y  E  E  E  L  K  A
 601  GGGATTAACACCGCGGATGAAACGGGTGTTGAATTGTTTGTATACCGAATTGTTGAACCGGGACGAAGCGATTGGTGATTACGAGGAGGAATTAAAAGCG   700

      V  A  K  A  N  E  D  C  Q  R  V  Q  S  I  P  G  V  G  Y  L  T  A  L  S  V  Y  A  S  V  G  D  I  H  Q
 701  GTGGCAAAAGCCAATGAGGATTGTCAACGGGTACAGAGCATCCCGGGGGTGGGTTATTTAACGGCGCTCTCGGTTTATGCGAGCGTGGGTGACATTCATC   800

        F  H  R  S  R  Q  L  S  A  F  I  G  L  V  P  R  Q  H  S  S  G  N  K  E  V  L  L  G  I  S  K  R  G
 801  AATTTCATCGTTCCCGGCAGTTGTCGGCGTTTATTGGGTTGGTCCCTCGACAACATTCGAGTGGGAATAAGGAGGTGTTGTTGGGGATTAGTAAACGCGG   900

       N  V  M  L  R  T  L  L  I  H  G  A  R  A  L  L  R  H  V  K  N  K  T  D  K  K  S  L  W  L  K  A  L
 901  CAATGTGATGTTAAGGACGTTATTGATTCATGGCGCCCGTGCGCTATTGCGTCATGTAAAAAATAAAACGGATAAAAAGAGTCTGTGGTTAAAAGCACTC   1000

      I  E  R  R  G  M  N  R  A  C  V  A  L  A  N  K  N  A  P  I  I  W  A  L  L  T  R  Q  E  T  Y  R  C  G
1001  ATTGAGCGCCGCGGAATGAATCGCGCTTGTGTGGCGTTAGCGAATAAAAATGCGCCGATCATTTGGGCGCTTTTAACACGCCAAGAAACGTATCGCTGTG   1100
                                                             G (P-R)
                                                             G (P-R)
        A
1101  GCGCCTAAACACCGCCGTGGGTAAAAAAAAGAATTAACAAAAGGAGACACACCAACCGAGTTCGAAACAATGAGGGCTGATGAAAGTAAGGTAAAACCTG   1200

1201  AGGTTGATTAAGCTGATTCATACGGTGGCTCTGTGAAGCCGATAGCCCGATAAGCATCAACCTTGCATAATTCATCAAGGCACCAATGGTGGCCAATTTA   1300

1301  AATCGTGATGCCGGATATACGAATGCAACCGCTTTCTTTTCATCAAAACTGATTGTTGACAACAAGGGTGGGTCCATATACGTATGCA.CGAAGCGGAAT   1400
         T                                                               <<(IRi)            . G       A
         T                                                                                    G

1401  ACGAGGATGAAATCATTTTCT.........CCGTATTCCGCTTCGCTGCATACGGGCTAC  1460
         G   AAC   G  G CA CTATTGTCG       T         .     <<(IRo)
         G   AC   AGCACGGA T........
```

FIG. 2. Nucleotide sequence of *C. burnetii* IS*1111a*, IS*1111b*, and IS*1111c*. Only differences in the sequences of IS*1111b* and IS*1111c* (second and third sequence lines, respectively) are included. The deduced amino acid sequence of the putative transposase is presented above the IS*1111a* sequence in single-letter symbols. (Note that two amino acids in the predicted IS*1111a*-encoded transposase differ for the putative IS*1111b* and IS*1111c* transposases, with arginine replacing glutamine at codon 138 and arginine replacing proline at codon 351 of the ORF in the latter two sequences.) Three amino-terminal methionine residues, potential translation initiation codons, are noted with asterisks. The 7-bp outer inverted repeats are underlined and designated (IRo), and the 12-bp inner inverted repeats are underlined and designated (IRi).

and later verified the restriction endonuclease *Dde*I to cleave a 1,000-bp fragment from the interior of the element. In a previous study, *Apa*I, *Cla*I, and *Eco*RI *C. burnetii* genomic fragments inserted into pBluescript KS were combined into pools of 10 clones each. Plasmid DNA was prepared from these pools, and aliquots were digested with *Dde*I. DNA bands of approximately 1,000 bp were detected in agarose gels after electrophoresis from several of the pools. Two independent isolates from these pools which retained the 1,000-bp *Dde*I fragment, had single *Sst*I and *Sma*I sites (as did IS*1111a*), and hybridized to an IS*1111a* probe (data not shown) were obtained. The similarity of restriction fragment profiles generated with several other endonucleases further confirmed that these two clones were also members of the family, and they were therefore subjected to DNA sequence analysis. IS*1111b* was isolated on a 2.8-kb *Apa*I fragment, and IS*1111c* was isolated on a 3.0-kb *Cla*I fragment. Restriction endonuclease fragment and nucleotide sequence differences in the flanking regions (data not shown) verified that the three copies of the repetitive element were distinct and independent.

The first indication that the repetitive element was actually an insertion sequence was the discovery of the 12-bp, perfect inverted repeats (ATGGACCCACCC; Fig. 2). Separating the repeats were 1,340 bp, which included a 364-codon ORF that began with a methionine codon in the 5′ 12-bp inverted repeat. A purine-rich region, which preceded the third methionine codon in the major ORF, was similar to typical sites of procaryotic translational initiation. Synthesis of a polypeptide from the third methionine of the major ORF would produce a protein of 339 amino acids before a termination codon was reached. Proximal to the major ORF was a site with similarity to the *E. coli* consensus promoter sequence. Four of six bases in the -35 hexanucleotide sequence (TGGACC) and five of six bases in the -10 sequence (CATAAT) agreed with the *E. coli* consensus σ^{72} promoter sequence (TTGACA and TATAAT, respectively), with a near-optimal spacing of 19 bp. This promoter would be just inside the 5′ end of the insertion sequence and, in fact, would be partly contained in the 12-bp inverted repeat.

Boundaries of the repetitive element were determined by comparing the three nucleotide sequences. Surprisingly, homology between the three clones extended well past the 12-bp inverted repeats. The 5′ end, as defined by the orientation of the major ORF, extended an additional 13 bp, and the 3′ end extended an additional 73 to 81 bp. The additional 5′ sequences lengthened slightly the major ORF from 364 codons to a maximum possible 367 codons for all three IS*1111* copies. Furthermore, the 7 bp at each end of these extended sequences of homology (GTAGCCC) were inverted repeats. Thus, 7-bp inverted repeats flanked the 12-bp inverted repeats, but at unequal intervals within each insertion sequence. In the case of the insertion sequence on the *Apa*I fragment (IS*1111b*), the outer inverted repeats extended past the sequences homologous with IS*1111a* and IS*1111c* an additional 7 bp, forming a 14-bp inverted repeat (GAGCTAAGTAGCCC). All three insertion sequences had 6 bp between the two 5′ inverted repeats, but the lengths between the 7- and 12-bp inverted repeats located in the 3′ end of the element differed slightly among the three copies of the insertion sequence (66, 68, and 74 bp). These 66-, 68-, and 74-bp regions contained the majority of sequence divergence for the three copies of the insertion sequence. Duplications of the sequences at the apparent sites of insertion were not present.

Inspection of the three sequences for secondary structures revealed extensive dyad symmetries that spanned the entire region between the 3′ inverted repeats, despite sequence differences there (Fig. 3). Although the DNA sequences diverged, the secondary structures were conserved. A stretch of uridines, typically found immediately after the stem-loop structures of rho-independent terminators, was not present on either side of the insertion sequence stem loops.

The 3′ 7-bp inverted repeat of IS*1111a* abutted the terminator at the 3′ end of the *htpAB* operon (Fig. 3), juxtaposing the dyad symmetry in the 3′ end of IS*1111a* with the putative terminator of the *htpAB* operon. Secondary structures were not found in the DNA flanking the 3′ end of IS*1111b* or IS*1111c*. Homology searches with IS*1111b* and IS*1111c* flanking DNA did not reveal related sequences in the GenBank or EMBL data base.

There were two base changes in the ORF of IS*1111a* (ORF-a) that altered the amino acid sequence of the putative transposases encoded by IS*1111b* (ORF-b) and IS*1111c* (ORF-c). Residue 110 (of the polypeptide initiated at the third methionine) changed from glutamine to arginine (CAG to CGG), and residue 323 changed from proline to arginine (CCG to CGG). These changes were verified by sequence analysis of PCR-amplified fragments generated by oligonucleotides specific for IS*1111a* (one internal to IS*1111a* and the other from *htpB*). These changes rendered ORF-b and ORF-c slightly more basic (predicted pI of 10.74 versus 10.64) and extended the second helix of the helix turn helix (see below) by an additional four residues.

Features of the major ORF-encoded polypeptide consistent with bacterial transposases. The deduced amino acid sequence of the major ORF indicated that the polypeptide had several characteristics of DNA-binding proteins, suggesting its role as a transposase. From sequence analysis, we predicted the *C. burnetii* IS*1111a*-encoded polypeptide to be quite basic, with calculated pIs of 10.64 for the 339-residue protein and 10.85 for the extended 364-residue protein, the predicted polypeptide of which began with the ATG codon found within the 5′ 12-bp inverted repeat.

Another characteristic of some DNA-binding proteins is a helix-turn-helix (HTH) motif. Computer analysis predicted an HTH structure near the carboxy terminus of the putative *C. burnetii* IS*1111* transposase (Fig. 4). The analysis further predicted that a glycine residue breaks the alpha helix and begins the turn, similar to the structure present in the InsA protein, a DNA-binding protein encoded by and involved in transposition of *E. coli*-derived IS*1*, as well as in several other bacterial DNA-binding proteins (16).

A third feature of DNA-binding proteins, albeit described almost exclusively for eucaryotic proteins heretofore, is the leucine zipper (8, 36). The pattern in these cases consists of four leucine residues spaced seven amino acids apart in the polypeptide chain, in a proposed alpha-helical conformation. We found this motif near the middle of the deduced amino acid sequence, from residues 122 through 143. Computer analysis predicted portions of this region to be helical in conformation (Fig. 4).

Similarity of the putative *C. burnetii* transposase to other bacterial transposases. Figure 5 shows an alignment of the deduced amino acid sequences of ORFs suggested to encode transposases from several reported bacterial insertion sequences. Leskiw et al. (37) found significant conservation of residues by position in pairwise rather than global homology comparisons of putative transposases. Inspection of the alignment showed strict conservation of 20 residues for the six polypeptides, with conservative replacements at other locations. In contrast, randomization and realignment of the six sequences resulted in no conserved residues by the same analysis (not shown). While the four *Streptomyces* and *Mycobacterium* transposases are around 400 amino acids long, a computer-generated gap near the middle brought both the *C. burnetii* and *Thermus thermophilus* sequences into reasonable pairwise alignment with them. Comparing the *C. burnetii* protein with the individual proteins revealed 19 to 20% identity for each with an average of 12 gaps introduced. Although the *C. burnetii* protein was not more closely related to any one of these five on the basis of sequence homology, the *T. thermophilus* protein was closer in size to it than the others. A leucine zipper pattern appeared in only the *Streptomyces clavuligerus* and *C. burnetii* polypeptides, and the two patterns were in the same relative positions. Computer analysis, however, did not predict helix-turn-helix motifs for the other aligned proteins.

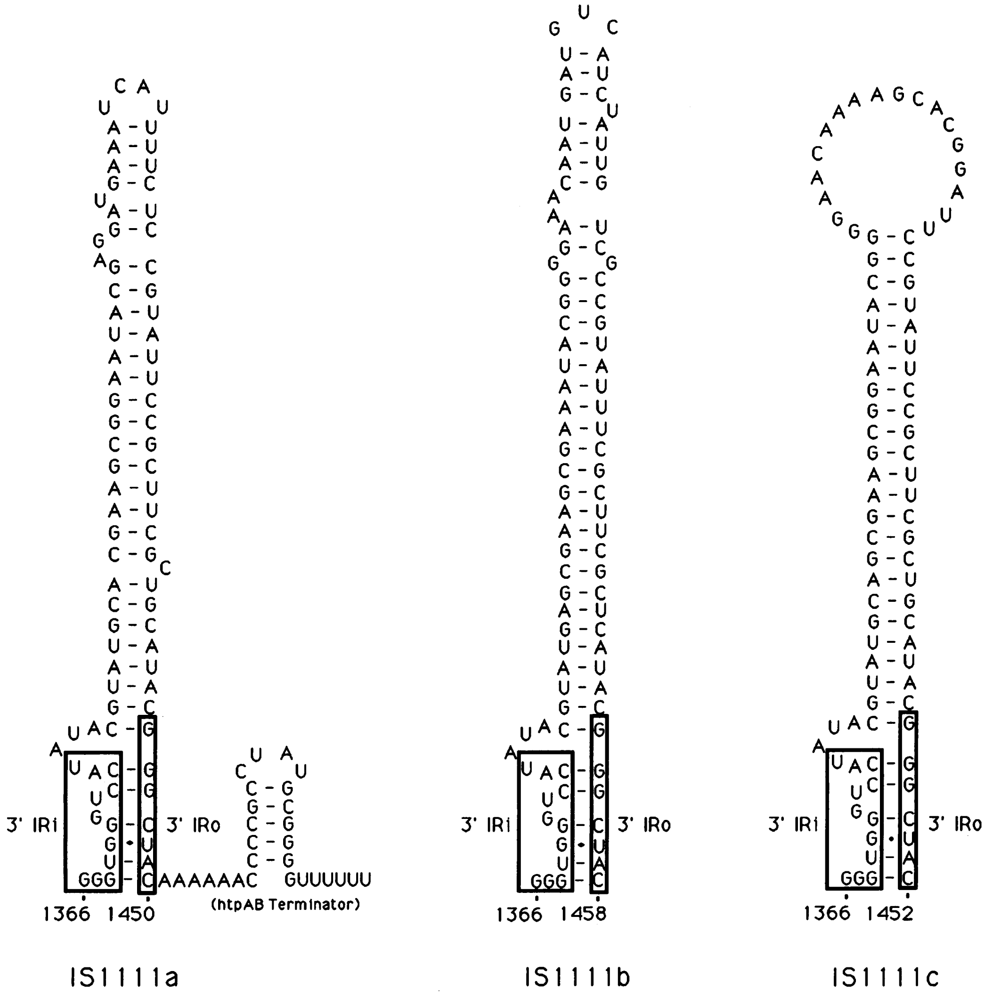

FIG. 3. Potential secondary structures of the sequences between the 3′ 7- and 12-bp inverted repeats of IS*1111a*, IS*1111b*, and IS*1111c*. The 12- and 7-bp inverted repeat sequences are boxed and designated IRi and IRo, respectively, as in Fig. 2.

DISCUSSION

A 1.45-kb DNA fragment was shown by Southern blot analysis to be present at least 19 times in the type strain *C. burnetii* Nine Mile I clone 7 genome. Several features of the nucleotide sequence of this repetitive DNA element suggested that it is an insertion sequence. These included an ORF encoding a deduced polypeptide with DNA-binding characteristics flanked by inverted repeats. The characteristics of the deduced polypeptide suggestive of transposases are (i) a possible carboxyl terminus HTH motif, (ii) a basic pI, (iii) a leucine zipper motif, and (iv) positional homology with several reported transposases. IS*1111* had a novel structure in that there was a second, outer pair of inverted repeats flanking the inner pair. Furthermore, the sequence between the 3′ inner and outer inverted repeats was composed of an extensive region of dyad symmetry, in contrast to the short 6-bp sequence between the 5′ inner and outer inverted repeats. There was a suggestion that the cloned transposase could function in vivo: clones were unstable and underwent rearrangements during the original subcloning of *htpAB* (50).

HTH motifs have been described for a number of eucaryotic and procaryotic DNA-binding proteins, and the structure and specificity with which these interactions occur have been well documented (for reviews, see references 4, 26, and 44). These secondary structures bind DNA in the major groove of the recognition site for several DNA-binding regulatory proteins, such as lambda cI repressor and Cro (40, 47, 49). Secondary structures of the IS*1111a* and IS*111b/c* ORF-encoded polypeptides, as calculated by the Chou-Fasman (10) and Robson- (17) methods, included an HTH predicted to lie near the carboxyl terminus. Examining the ORF for HTH motifs by the method of Dodd and Egan, which compares a 22-residue window from any amino acid sequence with HTH motifs from 91 reference sequences (15), showed that only two sequences of the protein scored above 1 standard deviation (SD). Both of these sequences, though, were below their 2.5-SD threshold and by this analysis considered unlikely to be HTH motifs. However, as Dodd and Egan point out, functional information can be used to adjust scores upward or downward. Thus, known DNA-binding proteins that score low by this method should not necessarily be eliminated as having the HTH motif. Amino acid residues 299 to 320, the higher-scoring sequence of the two HTH candidates in the IS*1111a* ORF, comprised the HTH in the carboxyl terminus, as predicted by Chou-Fasman and Robson-Garnier analyses. The best Dodd and Egan scores for the apparently related transposases in IS*110*,

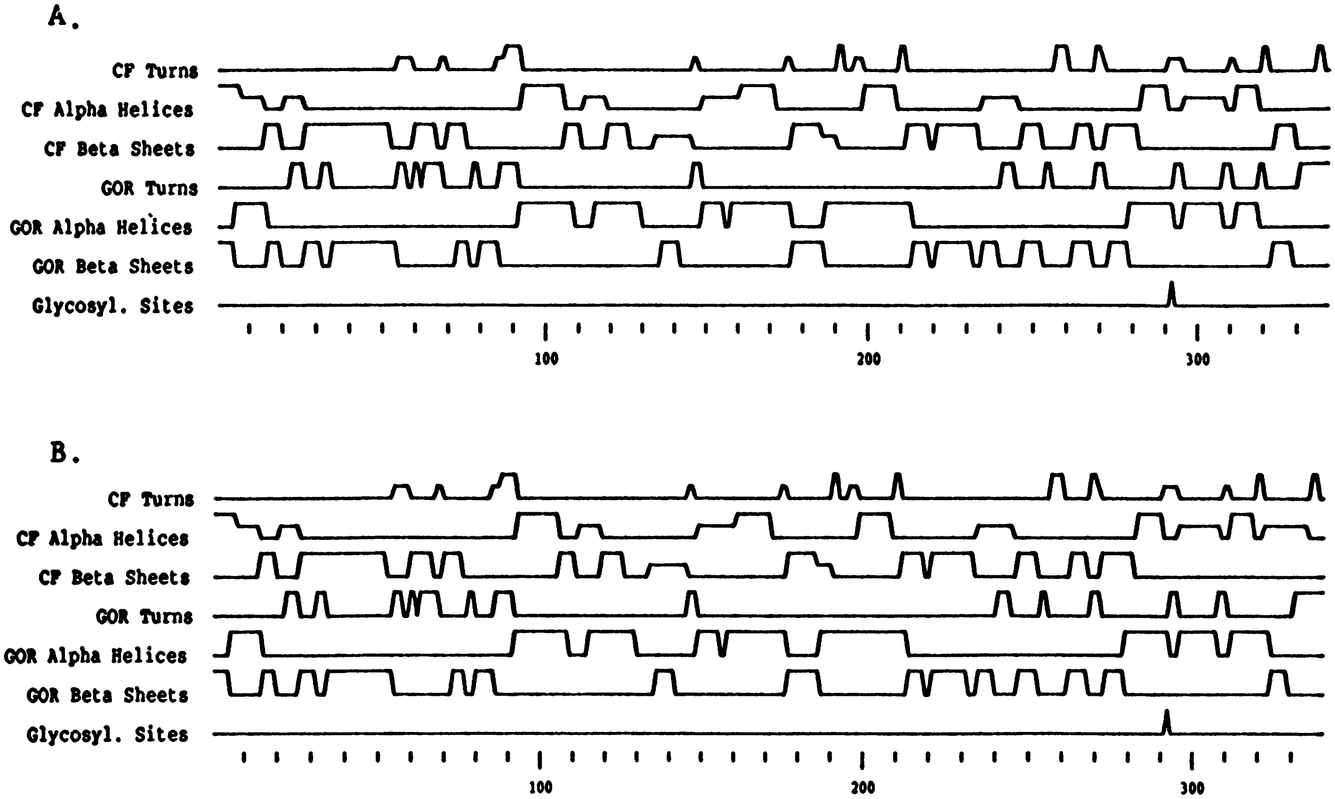

FIG. 4. Predicted secondary structures of the polypeptide encoded by ORF-a (A) and by ORF-b and ORF-c (B). The programs PEPTIDESTRUCTURE and PLOTSTRUCTURE of the Genetics Computer Group software package were used to predict and plot several parameters of protein secondary structures. The entire 339-residue ORF was analyzed. The potential HLH structures can be seen as peaks in the graphs of both Chou-Fasman (CF) and Garnier-Robson (GOR) predicted alpha-helical and turn regions near the carboxyl terminus of the sequence. As discussed in the text, while residues 296 through 323 fit the HLH motif, the structure may actually be a helix-turn-helix-turn-helix from residues 279 through 323. Furthermore, the proline-to-arginine replacement in IS*1111b* and IS*1111c* appears to extend the helical region which ends at residue 320 of IS*1111a*.

IS*116*, IS*900*, IS*1000*, and the *Streptomyces coelicolor* minicircle were likewise below the 2.5-SD threshold. In contrast, the score for the IS*1* InsA transposase was 4.36 SD, making it highly likely that it contains an HTH motif. HTH motifs were detected in several resolvases and invertases and subsequently included in the reference set by Dodd and Egan. HTH motifs are thus likely to be found in proteins involved in transposition and may occur in the putative IS*1111* transposase.

Another feature of many transposase proteins is a high content of basic residues. InsA and InsB from IS*1* have estimated pIs of 10.9 and 11.0, respectively (35, 39, 43). Presumably the high positive charge of these polypeptides ensures their activity in the vicinity of cognate mRNA translation. Nonspecific binding affinity to nearby negatively charged DNA with subsequent diffusion to and tight binding at specific sites is one explanation for extremely low transposase activity in trans (16, 56). The calculated pI of the IS*1111a* ORF from the third methionine was 10.64, and that from the extreme 5′ Met codon was 10.85. The predicted pIs of the five related transposases ranged from 10.0 to 10.6.

While the leucine zipper motif was first described for and appears to be more characteristic of eucaryotic DNA-binding proteins, a study of MetR from *E. coli* demonstrated a requirement for the integrity of the leucine zipper region of that protein for activity (41). MetR is a positive effector of *metE* and *metH* expression and functions as a homodimer. Other examples of this structure in procaryotes are the two leucine zippers found in *E. coli* σ^{54} (46), which promote intramolecular associations for the positioning of DNA-melting domains rather than for intermolecular dimerization. This motif is thus not restricted to eucaryotic DNA-binding proteins and may direct associations between the IS*1111* transposase and a host factor(s) required for transposition or may facilitate IS*1111* transposase homodimerization.

Inspection of the six similar sequences in the region of the *C. burnetii* leucine zipper revealed a leucine zipper motif for *S. clavuligerus* and possible leucine zipper motifs for the other four. In fact, only the *S. coelicolor* sequence has an incompatible residue in any of the four heptad repeats (33), a threonine in the fourth leucine position. However, it does have a leucine residue in what would be a fifth heptad leucine position of the leucine zipper. Immediately amino terminal to the characterized eucaryotic leucine zippers are highly basic regions that bind DNA. Homodimeric DNA-binding proteins with binding domains held in close proximity by a leucine zipper thus recognize sequences with dyad symmetry or hyphenated dyad symmetry, which can be considered inverted repeats. The six procaryotic transposases that we compared in this study, while highly basic overall, were not enriched in basic residues near the putative leucine zippers. This finding suggested that the DNA-binding domain(s) of these procaryotic transposases would not be contiguous with the leucine zippers, in contrast with known eucaryotic

```
CBURNETII    mKdiKil----gvdiaKdvfqlcgidewgKviytRRvKRaq-----yvstvasl----Kvgcvvmeacgganhwy Rtfmgmgiptqlisp   77
MPARATUBER   va---qp-vwagvdagKadhycmvinddaqRllsqRvandeaalleliaavttladggevtwaidlnaggaallialliaagqRllyi-p   86
SCLAVULIG    lstRhdR-iwvgidagKghhwavavdadgetlfstKvindeaqvltlietaR---eReevRwavdisgRastlllallvahgqnvvyv-p   86
SCOELICOL    mfdtedvgvflgldvgKtahhghgltpagKKvldKqlpnsepRlRavfdKla--aKfgtvlvivdqpasigalpltvaRdagcKvayl-p   88
MINICIRCLE   mwe-dsltvfcgidwaeRhhdvaivddtgtllaKaRitddvagynKlldllaehgdssatpipvaietshg-llvaalRtgsRKvfainp   88
TTHERMOPH    m-------tfagidvsKt-hldlalvsnspKptRlRfpnspegRqalla---alahhnpawvaleptsayhlpllKllaenRlqvalvnp   79
             .          *.* ..       .   .      .   .        .          .   .               .       . *

                                                        ------ LEUCINE ZIPPER -------
CBURNETII    qhvKpyvKs-----nKndRndaqaiaeaas--RasmRfvqgKtveqqdvqallKiRdRlvKsRtalineiRgllqey-------------   147
MPARATUBER   gRtvhhaagsyRgegKtdaKdaaiiadqaR-mRhdlqplRagddiavelRiltsRRsdlvadRtRaiepnaRpaagilsaleRafdyn--   172
SCLAVULIG    gRtvnRmsgayKgegKtdaKdaRviadqaR-mRRdfapldRppelvttlRlltnhRadliadRvRlinRlRdlltgicpaleRafdys--   172
SCOELICOL    glamRRiadlypgeaKtdaKdaaviadaaRtmahtlRsleltdeitaelsvlvgfdqdlaaeatRtsnRiRglltqfhpsleRvlgpR--   175
MINICIRCLE   laaaRyRdRhgvsRKKsdpgdalvlanilRtdmhahRplpadselaqaitvlaRaqqdavwnRqqvanqvRsllReyypaalhafqsKdg   178
TTHERMOPH    yhlaafRKaKg-eRqKtdRqdalllaRyaqvyhedlRaytlppetlRelKalvgyRedlagReRtilnqmeaaew---------------   148
                            *.*  **  .*        .            ..  *               .

CBURNETII    glt-maRgaKRfyeelplilase-------------------------------------avgltpRmKRvlnclytellnRdeaigdye   199
MPARATUBER   --KsRaalilltgyqtpdalRsaggaRvaaflRKRKaRnadtvaat----alqaanaqhsivpgqqlaatvvaRlaKevmaldteigdtd   256
SCLAVULIG    --aaKgpvvmlteyqtpaalRRtgvKRlttwlgRRKvRdadtvaaK----aieaaRtqqvvlpgeKRatKlvcdlahqllaldeRiKdnd   256
SCOELICOL    -ldhqavtwlleRygspaalRKagRRRlvelvRpKapRmaqRlidd----ifdaldeqtvvvpgtgtldivvpslassltavheqRRale   260
MINICIRCLE   gltRpdaRviltmaptpaKaaKltlaqlRaglKRsgRtRafnteieRlRgifRseyaRqlpavedafghqllall-Rqldatclaaddla   267
TTHERMOPH    -----------------------------------------------------------------agsKevlallqKelacvKgllgeve   178
                            (*) .       .   .      (*)                             ..  *   .

CBURNETII    eelKavaKanedcqRvqsipgvg-yltalsvyasvgdihqfhRsRqlsafiglvpRqhssgnKevllgisKRgnvmlRtllihgaRallR   288
MPARATUBER   amieeRfRRhRhaeiilsmpgfgvilgaeflaatggdmaafasadRlagvaglapvpRdsgRisgnlKRpRRydRRllRacylsa---lv   343
SCLAVULIG    ReiRetfRtddRaeiiesmpgmgpvlgaefva-ivgdlsgyKdagRlashaglapvpRdsgRRtgnyhRpqRynRRlRwlfymsa---qt   342
SCOELICOL    aqinalleahplspvltsmpgvgvRtaavllv-tvgdgtsfptaahlasyaglapttKssgtsihgehapRggnRqlKRamflsa---fa   346
MINICIRCLE   KavedafRehadseillsfpglgpllgaRvlaeigddRsRftdaRalKsyagsapitRasgRKhfvgRRfvKnnR-lmnagflwa---fa   353
TTHERMOPH    aRiqallatlpeaevlmalpgvgpqvaaavlallppel--wgRaKRaasyaglipeReesgKsveRsRlsKKgpplRRKlymga---lv   263
               . .          . ..** *   .*  .     .   .  .    .  *  *   .**               *       *

CBURNETII    hvKnKtdKKslwlKalieRRgmnRacvalanKnapiiwalltRqetyRc----------ga   339
MPARATUBER   siRtdpssRtyydRKRtegKRhtqavlalaRRRlnvlwamlRdhavyh---p-att-taaa   399
SCLAVULIG    ammRpgpsRdyylKKRgegllhtqallslaRRRvdvlwamlRdKRlft---p-appvtqta   399
SCOELICOL    cmnadpasRtyydRqRaRgKthtqallRlaRqRisvlfamlRdgtfyesRmp-agve-laa   406
MINICIRCLE   alqaspganahyRRRRehgdwhaaaqRhllnRflgqlhhclqtRqhfdeqRafapllqaaa   414
TTHERMOPH    avRhdpemRafyhRllsRgKRKKqalvavahKllRRmmgRlRe--yyatqld----qg-va   318
              .      .. . .          *   . ..    .   *     .            .*
```

FIG. 5. CLUSTAL (PCGene) alignment of six bacterial insertion sequence-encoded putative transposases. Asterisks indicate amino acid identity for all sequences (asterisks in parentheses indicate identity in regions where gaps exist), and dots indicate conservative replacements. The region in which the leucine zipper motifs occurred is indicated above residues 122 through 143 of the IS*1111a* major ORF, with the leucines (or functional replacements [33]) underlined. Arginine and lysine residues, often found in DNA-binding domains, are shown in uppercase bold type. Strains and insertion sequences represented are as follows: CBURNETII, IS*1111a* of *C. burnetii*; MPARATUBER, IS*900* of *Mycobacterium paratuberculosis* (18); SCLAVULIG, IS*116* of *S. clavuligerus* (37); SCOELICOL, IS*110* of *S. coelicolor* A3(2) (5, 9); MINICIRCLE, the minicircle of *S. coelicolor* A3(2) (29, 38); and TTHERMOPH, IS*1000* of *T. thermophilus* (2).

leucine zipper-containing DNA-binding proteins. The transposases and potential transposases from IS*110*, IS*116*, IS*900*, and the minicircle all had a highly basic region approximately 50 residues carboxy terminal from the leucine zippers, while the IS*1000* and IS*1111* ORFs had highly basic regions on the amino-terminal side of the zipper. All six polypeptides also had highly basic regions in the carboxyl terminus. Thus, while the eucaryotic leucine zipper serves to align two identical DNA-binding domains along a contiguous region of DNA with dyad symmetry, the IS*1111* transposase may bind and possibly bring together (as do some transposases) the remote ends of the insertion sequence by means of a DNA-binding domain at some distance from the leucine zipper. The insertion sequence ends were inverted repeats and were therefore merely dyad symmetric with extensive intervening DNA. The fact that highly basic regions were not adjacent to the leucine zippers in these transposases would therefore not be incompatible with the noncontiguous, distal binding sites (the inverted repeats) in these bacterial insertion sequences.

The arrangement of the ORF with the two sets of inverted repeats and the 3′-end dyad symmetry represents a novel structure for an insertion sequence. Secondary structures have been reported in the 5′ end of insertion sequences, with speculation that they function as modulators of expression by masking the translation initiation site (16). The absence of a run of adjacent thymidines implied that the secondary structures were not rho-independent terminators; nevertheless, transcriptional termination is a reasonable conjecture of their function. Juxtaposition of IS*1111a* immediately downstream of and in opposite orientation to the *htpAB* operon positioned the IS*1111a* secondary structure near that operon's terminator, which coincidentally has a run of thymidines on both sides of the stem-loop and may thus be functional as a bidirectional terminator. Divergent transcription through the 3′ end of and into the *htpAB* operon is, therefore, highly unlikely with these structures in place and may confer protection from interfering RNA polymerases to these required genes.

Studies of several insertion sequences and transposons show that the inverted repeats generally consist of two or more transposase-binding and, in some cases, host factor-binding domains (12, 34, 53). The first is found toward the interior side of the repeats and is the site of binding of the transposase and/or host factors; and a second domain, located near the outer ends of the repeats, is for directing scission of the DNA backbone just beyond the inverted repeat. It appears that IS*1111* may have two domains as well: the inner and outer inverted repeats. In this case, a separation of the domains may have occurred. The 5′ inverted repeats may well have once been a typical insertion sequence end, possibly including the 6-bp gap, while the 3′ end may have diverged to include the secondary structures between the two domains. Elucidating the mechanism of transposition of IS*1111* will be extremely interesting. Does the IS*1111* transposase bind to the inner, 12-bp inverted repeat and then direct strand scission at the 7-bp inverted repeat? How does this occur at the 3′ end with the interrupted sequence? One hypothesis is that a break in the DNA at the 5′ end may allow factor-directed cruciform formation to occur in the 3′ secondary structure, allowing the 12- and the 7-bp 3′-end inverted repeats that flank it to be brought into closer proximity to one another, with binding of the transposase at the 12-bp inverted repeat and strand scission now possible at the 7-bp inverted repeat.

While interesting questions remain concerning the mechanism of transposition, the presence of multiple copies of this sequence in the genome provides a tempting target to detect this microorganism through DNA hybridization and/or PCR technologies. Rapid diagnostic methods not requiring *C. burnetii* cultures would be valuable for detecting this obligate intracellular parasite. We found no significant homology to the IS1111 DNA sequence in searches of the major sequence data bases, and there was only limited, positional similarity for the amino acid sequence of the ORF. Thus, either hybridization or PCR-based diagnostic assays may be feasible and distinctive with IS*1111*. Furthermore, a means for distinguishing between *C. burnetii* strains may be possible should differences in the sites of insertion in various strains of *C. burnetii* be determined. Southern blotting and PCR analysis have demonstrated the presence of IS*1111*, with some variations in copy number (ranging from 19 to 37 copies) and restriction patterns, in all 16 *C. burnetii* strains examined thus far (32a).

Any role for the insertion sequence in the attenuation of virulence, as seen in various nonvirulent *C. burnetii* strains, or in other strain-to-strain differences remains to be determined. However, IS*1111* occupies a significant (1 to 2%) portion of the *C. burnetii* genome. Experiments designed to determine the flanking sequences of the various sites of IS*1111* insertion are just beginning. This information may implicate IS*1111* in deletions or other rearrangements of the chromosome, one or more of which may have been involved in relegating *C. burnetii* to its status as an obligate intracellular parasite, or perhaps in the smooth-to-rough phase transitions that occur in serial egg passages of *C. burnetii*.

ACKNOWLEDGMENTS

We thank Dale Seburn for technical assistance and Jeff Chulay and David Waag for helpful comments on the manuscript.

REFERENCES

1. **Akporiaye, E. T., J. D. Rowatt, A. A. Aragon, and O. G. Baca.** 1983. Lysosomal response of a murine macrophage-like cell line persistently infected with *Coxiella burnetii*. Infect. Immun. **40:**1155–1162.
2. **Ashby, M. K., and P. L. Bergquist.** 1990. Cloning and sequence of IS1000, a putative insertion sequence from *Thermus thermophilus* HB8. Plasmid **24:**1–11.
3. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:**1513–1523.
4. **Brennan, R. G., and B. W. Matthews.** 1989. Structural basis of DNA-protein recognition. Trends Biochem. Sci. **14:**286–290.
5. **Bruton, C. J., and K. F. Chater.** 1987. Nucleotide sequence of IS110, an insertion sequence of Streptomyces coelicolor A3 (2). Nucleic Acids Res. **15:**7053–7065.
6. **Burton, P. R., N. Kordova, and D. Paretsky.** 1971. Electron microscopic studies of the rickettsia *Coxiella burnetii*: entry, lysosomal response, and fate of the rickettsial DNA in L-cells. Can. J. Microbiol. **17:**143–150.
7. **Burton, P. R., J. Stueckemann, R. M. Welsh, and D. Paretsky.** 1978. Some ultrastructural effects of persistent infections by the rickettsia *Coxiella burnetii* in mouse L cells and green monkey kidney (Vero) cells. Infect. Immun. **21:**556–566.
8. **Busch, S. J., and P. Sassone-Corsi.** 1990. Dimers, leucine zippers and DNA-binding domains. Trends Genet. **6:**36–40.
9. **Chater, K. F., C. J. Bruton, S. G. Foster, and I. Tobek.** 1990. Physical and genetic analysis of IS110, a transposable element of Streptomyces coelicolor A3(2). Mol. Gen. Genet. **200:**235–239.
10. **Chou, P. Y., and G. D. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. Relat. Areas Mol. Biol. **47:**45–148.
11. **Clark-Curtiss, J. E., and M. A. Docherty.** 1989. A species-specific repetitive sequence in *Mycobacterium leprae* DNA. J. Infect. Dis. **159:**7–15.
12. **Craigie, R., M. Mizuuchi, and K. Mizuuchi.** 1984. Site-specific recognition of the bacteriophage Mu ends by the Mu A protein. Cell **39:**387–394.
13. **Dasch, G. A., W. Ching, P. Y. Kim, H. Pham, C. K. Stover, E. V. Oaks, M. E. Dobson, and E. Weiss.** 1990. A structural and immunological comparison of rickettsial HSP60 antigens with those of other species. Ann. N.Y. Acad. Sci. **590:**352–369.
14. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
15. **Dodd, I. B., and J. B. Egan.** 1990. Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res. **18:**5019–5026.
16. **Galas, D. J., and M. Chandler.** 1989. Bacterial insertion sequences, p. 109–164. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
17. **Garnier, J., D. J. Osguthorpe, and B. Robson.** 1978. Analysis of

the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120**:97–120.

18. **Green, E. P., M. L. V. Tizard, M. T. Moss, J. Thompson, D. J. Winterbourne, J. J. McFadden, and J. Hermon-Taylor.** 1989. Sequence and characteristics of IS900, an insertion element identified in a human Crohn's disease isolate of Mycobacterium paratuberculosis. Nucleic Acids Res. **17**:9063–9073.
19. **Grindley, N. D. F., and R. R. Reed.** 1985. Transpositional recombination in prokaryotes. Annu. Rev. Biochem. **54**:863–896.
20. **Grosskinsky, C. M., W. R. Jacobs, Jr., J. E. Clark-Curtiss, and B. R. Bloom.** 1989. Genetic relationships among *Mycobacterium leprae*, *Mycobacterium tuberculosis*, and candidate leprosy vaccine strains determined by DNA hybridization: identification of an *M. leprae*-specific repetitive sequence. Infect. Immun. **57**:1535–1541.
21. **Hackstadt, T.** 1983. Estimation of the cytoplasmic pH of *Coxiella burnetii* and effect of substrate oxidation on proton motive force. J. Bacteriol. **154**:591–597.
22. **Hackstadt, T.** 1990. The role of lipopolysaccharides in the virulence of *Coxiella burnetii*. Ann. N.Y. Acad. Sci. **590**:27–32.
23. **Hackstadt, T., and J. C. Williams.** 1981. Biochemical stratagem for obligate parasitism of eukaryotic cells by *Coxiella burnetii*. Proc. Natl. Acad. Sci. USA **78**:3240–3244.
24. **Hackstadt, T., and J. C. Williams.** 1981. Stability of the adenosine 5′-triphosphate pool in *Coxiella burnetii*: influence of pH and substrate. J. Bacteriol. **148**:419–425.
25. **Hackstadt, T., and J. C. Williams.** 1983. pH dependence of the *Coxiella burnetii* glutamate transport system. J. Bacteriol. **154**:598–603.
26. **Harrison, S. C., and A. K. Aggarwal.** 1990. DNA recognition by proteins with the helix-turn-helix motif. Annu. Rev. Biochem. **59**:933–969.
27. **Heinzen, R. A., M. E. Frazier, and L. P. Mallavia.** 1990. Nucleotide sequence of *Coxiella burnetii* superoxide dismutase. Nucleic Acids Res. **18**:6437.
28. **Heinzen, R. A., and L. P. Mallavia.** 1987. Cloning and functional expression of the *Coxiella burnetii* citrate synthase gene in *Escherichia coli*. Infect. Immun. **55**:848–855.
29. **Henderson, D. J., D. J. Lydiate, and D. A. Hopwood.** 1989. Structural and functional analysis of the mini-circle, a transposable element of *Streptomyces coelicolor* A3(2). Mol. Microbiol. **3**:1307–1318.
30. **Hendrix, L., and L. P. Mallavia.** 1984. Active transport of proline by *Coxiella burnetii*. J. Gen. Microbiol. **130**:2857–2863.
31. **Higgins, D. G., and P. M. Sharp.** 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene **73**:237–244.
32. **Hoover, T. A., and J. C. Williams.** 1990. Characterization of *Coxiella burnetii pyrB*. Ann. N.Y. Acad. Sci. **590**:485–490.

32a. **Hoover, T. A., et al.** Unpublished data.

33. **Hu, J. C., E. K. O'Shea, P. S. Kim, and R. T. Sauer.** 1990. Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions. Science **250**:1400–1403.
34. **Huisman, O., P. R. Errada, L. Signon, and N. Kleckner.** 1989. Mutational analysis of IS10's outside end. EMBO J. **8**:2101–2109.
35. **Johnsrud, L.** 1979. DNA sequence of the transposable element IS1. Mol. Gen. Genet. **169**:213–218.
36. **Landschulz, W. H., P. F. Johnson, and S. L. McKnight.** 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science **240**:1759–1764.
37. **Leskiw, B. K., M. Mevarech, L. S. Barritt, S. E. Jensen, D. J. Henderson, D. A. Hopwood, C. J. Bruton, and K. F. Chater.** 1990. Discovery of an insertion sequence, IS116, from *Streptomyces clavuligerus* and its relatedness to other transposable elements from actinomycetes. J. Gen. Microbiol. **136**:1251–1258.
38. **Lydiate, D. J., H. Ikeda, and D. A. Hopwood.** 1986. A 2.6 kb DNA sequence of *Streptomyces coelicolor* A3(2) which functions as a transposable element. Mol. Gen. Genet. **203**:79–88.
39. **Machida, Y., C. Machida, and E. Ohtsubo.** 1984. Insertion element IS1 encodes two structural genes required for its transposition. J. Mol. Biol. **177**:229–245.
40. **Matthews, B. W., D. H. Ohlendorf, W. F. Anderson, and Y. Takeda.** 1982. Structure of the DNA-binding region of lac repressor inferred from its homology with cro repressor. Proc. Natl. Acad. Sci. USA **79**:1428–1432.
41. **Maxon, M. E., J. Wigboldus, N. Brot, and H. Weissbach.** 1990. Structure-function studies on *Escherichia coli* MetR protein, a putative prokaryotic leucine zipper protein. Proc. Natl. Acad. Sci. USA **87**:7076–7079.
42. **Minnick, M. F., R. A. Heinzen, M. E. Frazier, and L. P. Mallavia.** 1990. Characterization and expression of the cbbE′ gene of *Coxiella burnetii*. J. Gen. Microbiol. **136**:1099–1107.
43. **Ohtsubo, H., and E. Ohtsubo.** 1978. Nucleotide sequence of an insertion element, IS1. Proc. Natl. Acad. Sci. USA **75**:615–619.
44. **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu. Rev. Biochem. **53**:293–321.
45. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
46. **Sasse-Dwight, S., and J. D. Gralla.** 1991. Role of eukaryotic-type functional domains found in the prokaryotic enhancer receptor factor sigma^{54}. Cell **62**:945–954.
47. **Sauer, R. T., R. R. Yocum, R. F. Doolittle, M. Lewis, and C. O. Pabo.** 1982. Homology among DNA-binding proteins suggests use of a conserved super-secondary structure. Nature (London) **298**:447–451.
48. **Shinnick, T. M., M. H. Vodkin, and J. C. Williams.** 1988. The *Mycobacterium tuberculosis* 65-kilodalton antigen is a heat shock protein which corresponds to common antigen and to the *Escherichia coli* GroEL protein. Infect. Immun. **56**:446–451.
49. **Steitz, T. A., D. H. Ohlendorf, D. B. McKay, W. F. Anderson, and B. W. Matthews.** 1982. Structural similarity in the DNA-binding domains of catabolite gene activator and cro repressor proteins. Proc. Natl. Acad. Sci. USA **79**:3097–3100.
50. **Vodkin, M., and J. Williams.** 1988. Unpublished data.
51. **Vodkin, M. H., and J. C. Williams.** 1988. A heat shock operon in *Coxiella burnetii* produces a major antigen homologous to a protein in both mycobacteria and *Escherichia coli*. J. Bacteriol. **170**:1227–1234.
52. **Waag, D. M., and J. C. Williams.** 1988. Immune modulation by *Coxiella burnetii*: characterization of a phase I immunosuppressive complex differentially expressed among strains. Immunopharmacol. Immunotoxicol. **10**:231–260.
53. **Wiater, L. A., and N. D. Grindley.** 1988. Gamma delta transposase and integration host factor bind cooperatively at both ends of gamma delta. EMBO J. **7**:1907–1911.
54. **Williams, J. C., T. A. Hoover, D. M. Waag, N. Banerjee-Bhatnagar, C. R. Bolt, and G. H. Scott.** 1990. Antigenic structure of *Coxiella burnetii*. A comparison of lipopolysaccharide and protein antigens as vaccines against Q fever. Annu. N.Y. Acad. Sci. **590**:370–380.
55. **Williams, J. C., and S. Stewart.** 1984. Identification of immunogenic proteins of *Coxiella burnetii* phase variants, p. 257–262. *In* L. Lieve and D. Schlessinger (ed.), Microbiology—1984. American Society for Microbiology, Washington, D.C.
56. **Winter, R. B., O. G. Berg, and P. H. von Hippel.** 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli lac* repressor-operator interaction: kinetic measurements and conclusions. Biochemistry **20**:6961–6977.
57. **Woods, S. A., and S. T. Cole.** 1989. A rapid method for the detection of potentially viable *Mycobacterium leprae* in human biopsies: a novel application of PCR. FEMS. Microbiol. Lett. **53**:305–309.