INFECTION AND IMMUNITY, Jan. 1991, p. 390–397
0019-9567/91/010390-08$02.00/0

Vol. 59, No. 1

Tandem Insertion Sequence-Like Elements Define the Expression Site for Variable Antigen Genes of *Borrelia hermsii*

ALAN G. BARBOUR,[1]* CAROL J. CARTER,[1] NILS BURMAN,[2] CYNTHIA S. FREITAG,[1] CLAUDE F. GARON,[3] AND SVEN BERGSTRÖM[2]

Departments of Microbiology and Medicine, University of Texas Health Science Center, San Antonio, Texas 78284[1]; Department of Microbiology, University of Umeå, S-901 87 Umeå, Sweden[2]; and Laboratory of Vectors and Pathogens, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, Hamilton, Montana 59840[3]

Received 8 August 1990/Accepted 18 October 1990

The spirochete *Borrelia hermsii* avoids the immune response of its mammalian host through multiphasic antigenic variation. Serotype specificity is determined by variable antigens, Vmp proteins, in the outer membrane. Through nonreciprocal recombination between linear plasmids, a formerly silent *vmp* gene replaces another *vmp* gene downstream from a common expression site. To further characterize this activating site, we determined the nucleotide sequence of 6.9 kb of the common upstream expression region of strain HS1 of *B. hermsii*. Preceding the *vmp* gene promoter and a poly(dT · dA) run were three imperfectly repeated segments of 2 kb. Each of the 2-kb segments contained 1-kb elements with inverted repeats of approximately 0.2 kb each at their termini. The potential of the 1-kb elements to form stem-and-loop structures was demonstrated by heteroduplex analysis. There was no evidence of the presence of the elements elsewhere in the genome of *B. hermsii*. One or more of these elements may confer the unidirectionality that characterizes *vmp* gene switches.

The spirochete *Borrelia hermsii* causes relapsing fever, a serious infection of humans that features episodes of fever interspersed with periods of well-being (5). Borrelias survive in humans and other mammalian hosts by avoiding the immune response. This is accomplished by changing from one surface antigen, called a Vmp protein, to another variable antigen (2, 7, 8). A variant bacterium displaying a new Vmp appears spontaneously in the population at the frequency of 10^{-4} to 10^{-3} per cell per generation (38). The variant and its progeny flourish while the host produces antibodies to the strain that initiated the infection. There are at least 25 different serotypes of the HS1 strain of *B. hermsii* (6).

Serotypes 7 and 21 of strain HS1 are the best-characterized variants of *B. hermsii* (6). Separate genes, *vmp7* and *vmp21*, encode the serotype-specific proteins Vmp7 and Vmp21, respectively (10). In a borrelia cell of serotype 7, *vmp7* is found in two environments, silent and active (29, 35). In the same serotype 7 cell, there is only one environment for the *vmp21* gene. In serotype 21 cells, the situation is reversed; there are two environments for the *vmp21* gene and only one for the *vmp7* gene. Both types of *vmp* genes are located on novel procaryotic replicons, linear plasmids of 24 to 28 kb with covalently closed ends (4, 26, 35). Whereas active *vmp7* and *vmp21* are usually near the telomeres of the expression-linked linear plasmids, the silent versions of these genes are located farther from the termini of their silent linear plasmids (26). The linear plasmids bearing active *vmp7* and active *vmp21* genes were designated bp7E and bp21E, respectively; the silent versions of these genes are on plasmids bp7S and bp21S, respectively (26).

A recombination between an expression plasmid and a silent plasmid activates a new *vmp* gene in a borrelia through promoter addition (3). The recombination is effectively nonreciprocal; the second possible recombination product appears to be either lost or never created (26, 35). The switch is also unidirectional; a silent gene displaces a gene at the expression site and not vice versa (29, 35). The unidirectionality of the genetic switch appears to be determined by sequences upstream of the *vmp* gene on an expression plasmid (26, 35). To further characterize the upstream expression region, we examined DNA 5′ to an active *vmp* gene. We were specifically interested in DNA sequences that would possibly provoke recombination and would provide for the unidirectionality that characterizes the recombination.

MATERIALS AND METHODS

Bacterial strains, phage, and plasmids. *B. hermsii* HS1 (ATCC 35209) serotype 7 was grown in BSK II broth medium and harvested as described previously (8). Recombinant plasmid p7.16 contains the entire expressed *vmp7* gene and the upstream flanking sequence; plasmids p7.1 and p21.8 contain the 5′ end of expressed *vmp* genes and upstream flanking regions of serotypes 7 and 21, respectively (35). Subclones were created by using the vectors pBR322 and pUC19 and were maintained in *Escherichia coli* JM101 or DH5α (Bethesda Research Laboratories, Gaithersburg, Md.). M13mp18 and mp19 phages were propagated in JM101 or JM109 (44).

DNA techniques. Restriction endonucleases, T4 DNA ligase (Boehringer Mannheim, Indianapolis, Ind.), reverse transcriptase (Life Sciences, St. Petersburg, Fla.), Sequenase (U.S. Biochemical, Cleveland, Ohio), and Klenow fragment of DNA polymerase I (Pharmacia, Piscataway, N.J.) were used as recommended by the manufacturers. The isolation of plasmid DNA from *B. hermsii* and *E. coli* and agarose gel electrophoresis were performed essentially as described previously (4, 9). Recovery of DNA fragments from gels was carried out with an analytical electroelutor (International Biotechnologies) or by agarose extraction with GeneClean (Bio 101, La Jolla, Calif.). Ligation of insert

* Corresponding author.

DNA into plasmid or M13 vectors was performed by standard techniques (30). Transformation, transfection, and production of competent cells were performed as described by Hanahan (23). Custom oligonucleotides for probes, specific-primer-directed DNA sequencing, and polymerase chain reaction were synthesized on an Applied Biosystems DNA synthesizer. Oligonucleotides were 5′ end labeled with T4 kinase and were used in Southern blot analyses as described previously (29).

Heteroduplex formation and electron microscopy. Purified DNA of plasmids p7.1 and p21.8, which contain *Pst*I fragments (35), was digested with *Eco*RI. Heteroduplexes were formed by mixing equal quantities (1 to 2 μg) of each plasmid, denaturing with alkali, and renaturing for 2 h at 37°C in a solution containing 50% formamide. Samples were mounted for electron microscopy as previously described (20). Grids were examined in a JEOL 100B electron microscope at 40 kV of accelerating voltage. Electron micrographs were taken on Kodak Electron Image plates at a magnification of ×7,000. The magnification was calibrated for each set of plates with a grating replica (E. F. Fullam; catalog no. 10000). Contour lengths were measured with a Numonics Graphics calculator interfaced with a Tektronix 4052A computer. The known lengths of plasmid pBR322 *Eco*RI-*Pst*I arms were used as internal calibration standards. Measurements of various segments are expressed (in kilobases) as mean lengths for a total of 10 molecules.

DNA sequence analysis. DNA sequences of recombinant DNA clones were determined by dideoxy procedures by using the Klenow fragment of DNA polymerase I (30), reverse transcriptase (45), or Sequenase (41). The complete sequences of both strands were determined. Three general strategies were used: (i) sequencing small, overlapping fragments with vector primers, (ii) sequencing larger fragments with vector primers for the 5′ ends and custom oligonucleotide primers for internal sequences (39), and (iii) sequencing single-stranded, amplified *B. hermsii* genomic DNA. Overlapping subclones for sequencing were produced by forced cloning with known restriction sites, by shotgun cloning and subsequent fragment assembly (25), or by making sequential deletions with exonuclease III (13). Templates were obtained from single-stranded M13 phage (30) or from double-stranded pBR322, pUC18, or pUC19 plasmids (13); oligonucleotide primers were synthesized as described above or were obtained from Bethesda Research Laboratories. Genomic *B. hermsii* DNA was sequenced by using amplified fragments and specific oligonucleotide primers in a modification of the method of Gyllensten and Erhlich (3, 22). Initial amplifications in the polymerase chain reaction were performed in a solution containing 10 mM Tris (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% gelatin, 0.2 mM deoxynucleoside triphosphate, 2.5 U of *Thermus aquaticus* DNA polymerase (Perkin-Elmer-Cetus), 25 ng of *B. hermsii* DNA, and 50 pmol of each of the two oligonucleotide primers. Samples were heated to 94°C for 1.5 min in a thermal cycler (MJ Research, Cambridge, Mass.) for initial denaturation and then subjected to 30 cycles, each consisting of heating at 94°C for 1 min, 55°C for 1 min, and 72°C for 3 min. During the last cycle, the samples were kept at 72°C for 10 min. The amplification products were recovered by electrophoresis and electroelution. Single-stranded DNA was generated from the first amplification products by the same procedure, with the exception that only one primer was used. Samples were then precipitated with ammonium acetate and isopropanol. The single-stranded DNA was in turn sequenced with Sequenase and specific primers as described above.

The software developed by Harr et al. (24) for VAX computers (Digital Equipment Corporation) was used to assemble the DNA sequences. This and additional software from the University of Wisconsin Genetics Computer Group (GCG; Sequence Analysis Software version 6.0) for the VAX were used for subsequent sequence analysis (15). GenBank data base release 64.0, EMBL data base release 23.0, and NBRF protein data base release 25.0 were available for searches. The codon frequency table of Burman et al. (10) was used for codon preference analysis, and the algorithm of Fickett (18) was used for identifying protein coding regions. Predicted amino acid sequences of open reading frames (ORFs) were compared by GCG's Gap algorithm, which is based on the method of Needleman and Wunsch (31) and which assesses amino acid similarity by using the measurements of Gribskov and Burgess (21). Derived amino acid sequences were also compared by the jumbling test of Doolittle (16); the sequence orders were randomized for this analysis by using the Shuffle algorithm of GCG.

RESULTS

DNA sequence of the upstream expression region. The physical maps of Plasterk et al. (35) and of Kitten and Barbour (26) indicated that at least 10 kb of sequence upstream of their active *vmp* genes were the same in bp7E and bp21E. The contour lengths of bp7E and bp21E are each 24 kb (20a). In the present study, 6.9 kb of DNA upstream of the start codon (position +1) for *vmp7* on bp7E was sequenced (Fig. 1). Figure 2 schematically summarizes the sequence findings. The active *vmp7* gene (3, 10), the downstream homology sequence, and the right telomere (26) of expression plasmid bp7E are 3′ to the sequence in Fig. 1. The sequence begins with the left *Pst*I site of the 4.7-kb *Pst*I fragment of p7.16 and extends rightward into the adjacent 2.9-kb *Pst*I fragment (29).

The upstream expression sequence of bp7E has several notable features, one of which is the repetitive character of the DNA. The sequence is divided into a series of three 2-kb segments, each containing the same motif (Fig. 1 and 2). The first segment spans positions −6870 to −4753. For ease of description, the basic pattern of the segment was divided into three areas or domains: the first, indicated by dots, from −6870 to −6484; the second, designated by bars, from −6179 to −5518; and the third, indicated by stars, from −5347 to −4753.

In the middle segment (−4752 to −2172) and last segment (−2171 to −85), sequences similar to the "dot," "bar," and "star" domains were found (Fig. 1). The greatest relatedness was between the domains of the first and third 2-kb segments; the dot, bar, and star domains of the third segment were 95, 82, and 97% identical, respectively, to the corresponding domains of the first segment. In the interdomain regions, there was little sequence similarity within or between the three segments, a finding represented by the different fill-in patterns in Fig. 2.

From positions −1033 to −84 of the third segment are 950 bp bounded by imperfect inverted repeats of 201 bp (left inverted repeat [LIR]) and 224 bp (right inverted repeat [RIR]) (Fig. 1). Sequences highly similar to these inverted repeats were also found in the first and second segments (Fig. 1 and 2). The alignment of the LIRs and the inverse sequences of the RIRs of the three elements is shown in Fig. 3. When one base replaced another in the sequence, it was commonly one pyrimidine replacing another pryimidine; 28

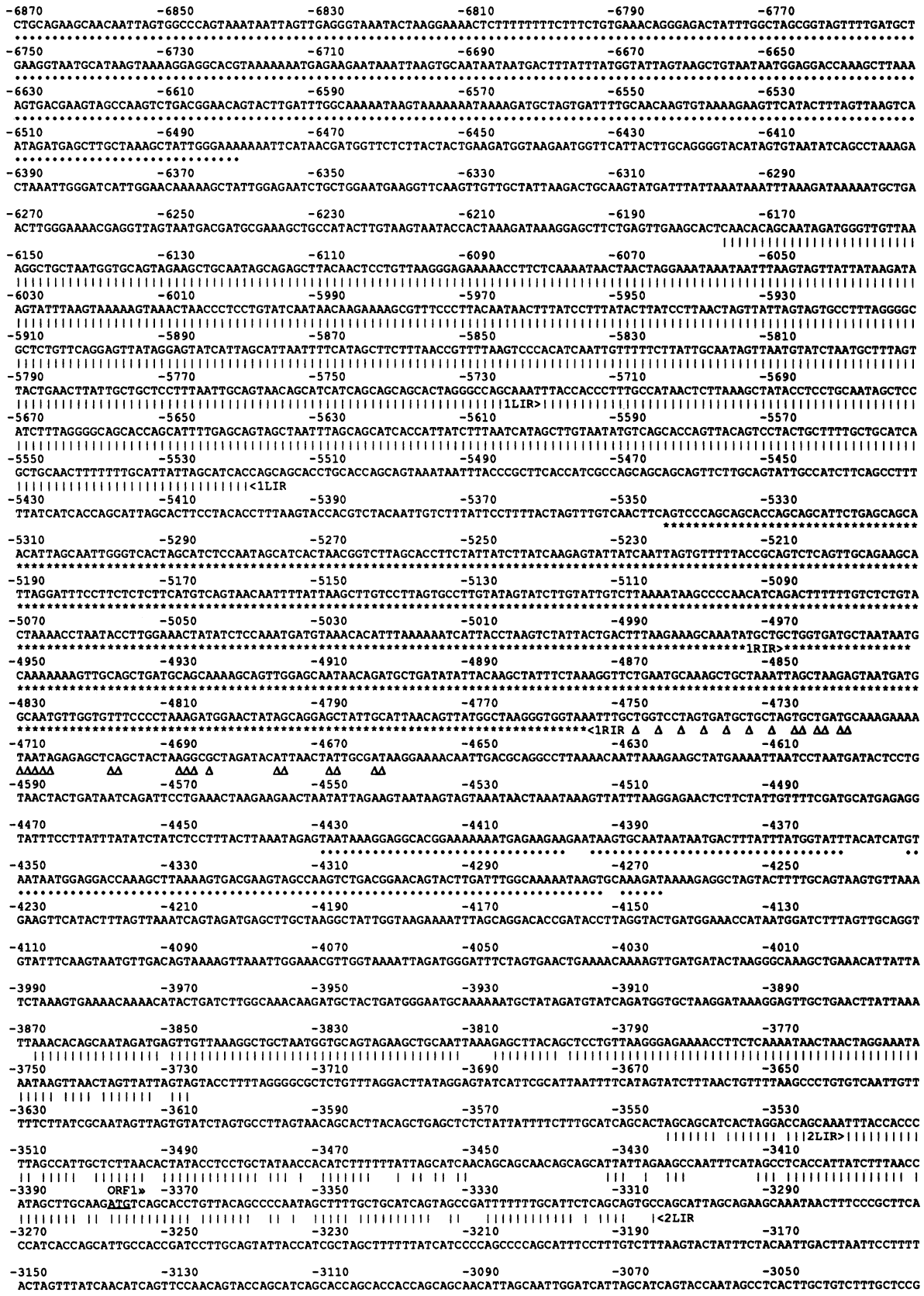
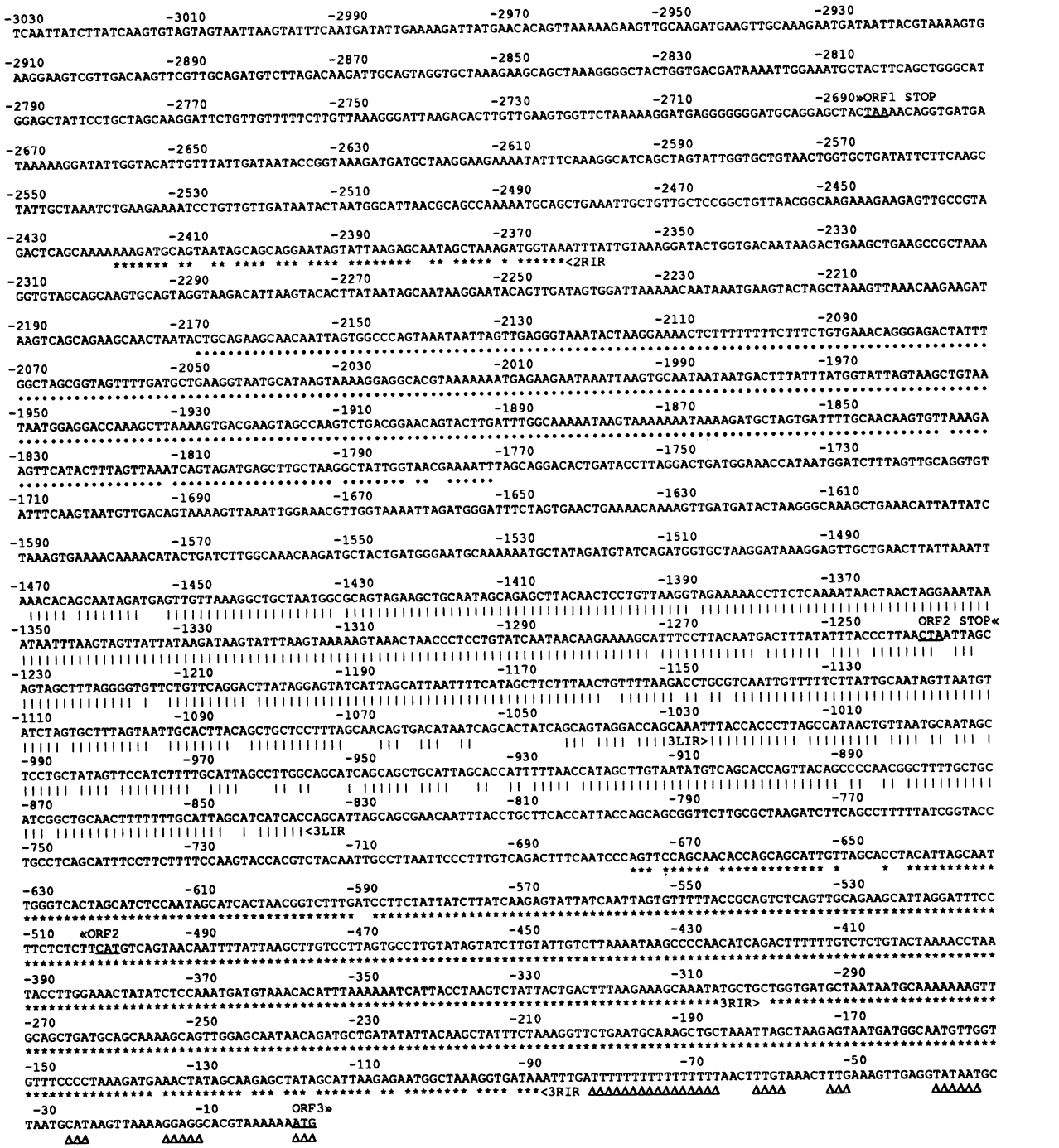


FIG. 1. Sequence of DNA upstream of expressed *vmp7* of *B. hermsii*. The numbers above each line refer to nucleotide positions. The +1 position is the first base of the start codon for *vmp7*. Similarities between different regions of the sequence are indicated by dots (·), bars (|), or stars (*) below the nucleotides. The starts and stops (< or >) of LIRs and RIRs are shown underneath the sequence. The starts and stops of ORFs are indicated above the sequence. Selected regulatory sequences or their analogs, which are described in the text, are indicated by triangles below the nucleotides.



C-T or T-C changes were found instead of the expected 9. The sequence similarity between repeats was greatest at their 5′ and 3′ ends. The RIR of the second segment (2RIR) showed the greatest divergence from the first set of inverted repeats. Sequences similar to the inverted repeats were not found in the GenBank or EMBL data bases.

Previous hybridization studies had suggested that the DNA region shown in Fig. 1 was not substantially duplicated elsewhere in the *B. hermsii* genome (26, 29, 35). To determine whether the sequence homologous to the LIRs and RIRs was present elsewhere, the oligonucleotide 5′-CAAAT TTACCACCCTT[A/T]GCCAT-3′, which is highly conserved in the inverted repeats, was labeled and used as a probe in a Southern blot under conditions allowing for hybrid formation with up to two or three mismatches. There was no evidence of this sequence in other regions of bp7E, on any other plasmid, or on the chromosome (results not shown). Thus, these elements appear to be restricted to the expression region shown in Fig. 1.

Starting at position −82, 1 base from the end of the third RIR, is an uninterrupted string of 16 Ts, as noted previously (3). When the oligonucleotide T_{16} was end labeled and used

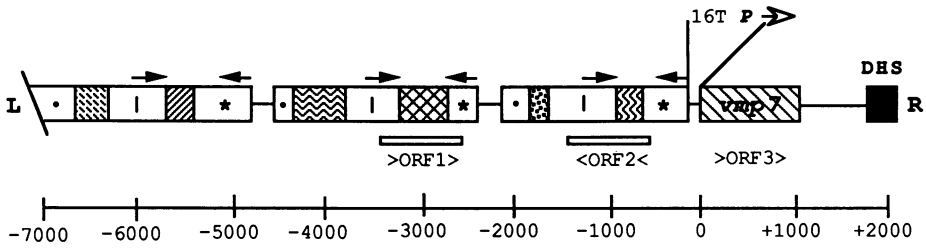


FIG. 2. Schematic diagram of the right arm of the expression plasmid bp7E of *B. hermsii*. The three 2-kb imperfectly repeated segments are shown with the dot (·), bar (|), and star (*) domains of Fig. 1. The different fill-in patterns in the segments represent interdomain sequence heterologies. The oppositely oriented short arrows indicate the approximate locations and sizes of pairs of inverted repeats in each segment. The locations of ORF1 and ORF2 are shown below. The active *vmp7* (ORF3) is shown in relation to the plasmid's right telomere and the downstream homologous sequence (DHS) (3, 10, 26). The site of 16 Ts, the *vmp* gene promoter (*P*), and the direction of transcription are indicated. The scale at the bottom (in base pairs) uses the start of the *vmp7* gene as the 0 position.

as a probe, there was no evidence by Southern blot analysis of 16 or more Ts elsewhere in the *B. hermsii* genome (results not shown). Like the inverted repeats, the run of A-T pairs appeared to be unique to the expression site.

Following the run of Ts and beginning at position −67 are the pair of imperfect 9-mer direct repeats (AACTTTG[T/A]A) in tandem and the 6-mer inverted repeats (AACTT T....AAAGTT) (3). A few bases after these short repeats are

```
1LIR: ( AAATTTACCACCCTTtGCCATaaCTCTTAAaGCTATACCTCCTGCaATA
1RIR: CAAATTTACCACCCTTAGCCATaaCTgTTAATGCaATACCTCCTGCTATA
2LIR: CAAATTTACCACCCTTAGCCATtgCTCTTAAcaCTATACCTCCTGCTATA
2RIR: tAAATTTACCAtCtTTAGCtATtgCTCTTAATaCTATtCCTgCTGCTATt
3LIR: CAAATTTACCACCCTTAGCCATaaCTgTTAATGCaATAgCTCCTGCTATA
3RIR: CAAATTTAtCACCtTTAGCCAT.tCTCTTAATGCTATAgCTCtTGCTATA
      Y       Y Y Y  W  Y  Wn  S    nR W  WS  SY   W  W
                .         .         .         .         .
1LIR: GcTcCATCTTTaGGGGcAgCACCAgCATT.............TTAGCtaA
1RIR: GtTcCATCTTTaGGGGaAaCACCAACATTgcCAtCATTActCTTAGCtaA
2LIR: accaCATCTTTttattAgCAtCAACAgcagCAaCAgcAgcaTTAttagA
2RIR: acTgCATCTTTttttGc.....................................
3LIR: GtTcCATCTTTtG.....................CATTAgcCTTgGCagc
3RIR: GtTtCATCTTTaGGGGaAaCACCAACATTgcCAtCATTActCTTAGCtaA
      RYYn       WKKnKn R  Y  R  KYRS  W  KY SYM  RKYWRM
                .         .         .         .         .
1LIR: tTtAGCAGC..aGCATcacCAtt.AtCTTTAatcATAGCTTGTAATATgT
1RIR: tTtAGCAGCttTGCATT..CAga.AcCTTTAgaaATAGCTTGTAATATaT
2LIR: agccaatttcaTagccTcaCcattAtCTTTAaccATAGCTTGcAAgATgT
2RIR: .....................................................
3LIR: aTcAGCAGC..TGCATTagCAcc.AttTTTAaccATAGCTTGTAATATgT
3RIR: tTtAGCAGCttTGCATT..CAga.AcCTTTAgaaATAGCTTGTAATATaT
      WKYMRMWKYYWWRSMYYMn Mnn  YY    RnM        Y  K  R
                .         .         .         .         .
1LIR: CAGCAcCaGTTAcaG.tCCtACtGCTTTTGCTGCATCAGCTGCAACTTTT
1RIR: CAGCAtCtGTTAttGCtCCAACtGCTTTTGCTGCATCAGCTGCAACTTTT
2LIR: CAGCAcCtGTTAcaGCcCCAAtaGCTTT.GCTGCATCAGtaGCcgaTTTT
2RIR: .....................................................
3LIR: CAGCAcCaGTTAcaGCcCCAACgGCTTTTGCTGCATCgGCTGCAACTTTT
3RIR: CAGCAtCtGTTAttGCtCCAACtGCTTTTGCTGCATCAGCTGCAACTTTT
           Y W    YW  Y  W Yn              R YW  MRM
                .         .
1LIR: TTTGCATTATtAGCATCACCAGCA
1RIR: TTTGCATTATtAGCATCACCAGCA
2LIR: TTTGCATTcTcAGCAgtgCCAGCA
2RIR: ........................
3LIR: TTTGCATTAgcAtCATCACCAGCA
3RIR: TTTGCATTATtAGCATCAgCAGCA
            M Y K  KYRS
```

FIG. 3. Sequence similarities between three sets of LIRs and the inverse of the RIRs of the upstream expression sequence of *B. hermsii*. At each position, nucleotides common to four or more repeats are represented by capital letters; other nucleotides are represented by lowercase letters. Gaps were introduced for optimum alignment. Below the aligned sequences, the type of base substitution is indicated by one of the following abbreviations: Y, pyrimidine-pyrimidine; R, purine-purine; W, A-T; S, C-G; K, T-G; M, A-C; and n, any of three or four bases.

the *vmp* promoter with its consensus −35 and −10 sequences, the transcriptional start sequence, the ribosomal binding sequence, and the start codon (3, 10). These sites, as well as the poly(dT) tract, are indicated with triangles in Fig. 1.

Remnants or antecedents of these regulatory sequences are present in an analogous region 3′ to the first RIR (positions −4751 to −4665 in Fig. 2). These sequences are also indicated by triangles. A tract with a T every third base is followed by sequences identical, except for one or two bases, to the promoter, transcriptional start, ribosomal binding site, and start codon.

The two ORFs that exceed 600 bases and begin with an ATG are indicated in Fig. 1 and 2. The first (ORF1) starts at position −3381, ends at position −2688, and would encode a 231-residue polypeptide. A sequence resembling a σ^{70} procaryotic promoter does not precede ORF1. Analysis of ORF1 with the codon frequency table for *Borrelia* genes (10) showed that the ORF's codon preference substantially differed from that of known *Borrelia* genes. For these reasons, we concluded that ORF1 probably does not encode a polypeptide.

ORF2 runs oppositely, beginning at nucleotide −502 and ending at position −1240 (Fig. 1 and 2). ORF2 has a coding capacity for a 246-residue polypeptide (Fig. 4). ORF2, like ORF1, is not preceded by a discernible procaryotic promoter. However, by the algorithm of Fickett (18), the measure of randomness at every third base in ORF2 was above the threshold for a coding sequence. In addition, an analysis using the codon table of Burman et al. (10) showed that codon preference and the pattern of rare codon use in ORF2 were similar to those of known *Borrelia* genes (results not shown). The deduced ORF2 polypeptide has 30.5% identity with Vmp7 (10) when gaps are allowed (Fig. 4). Because this amount of identity might be attributable to chance, 21 different pairs of randomly jumbled ORF2 and Vmp7 amino acid sequences were similarly compared. The mean percent identity between the pairs of jumbled sequences was 22.4 ± 1.2 (mean ± standard error), indicating that the observed similarity between ORF2 and Vmp7 is significant (16). ORF2 differs from sequenced outer membrane lipoproteins of *Borrelia* spp. in lacking a cysteine residue (9, 10). A search of the GenBank, EMBL, and NBRF DNA and protein data bases did not reveal a sequence identity greater than 24% between ORF2 and any other protein or ORF.

```
ORF2>   1 .........................................MKREGNPN 8
                                                    : |::
Vmp7>  51 MELGRSAENAFYAFIELVSDVLGFTAKSDTTKQEVGGYFNSLGAKLGEAS 100
ORF2>   9 ASATETAVKTLIDNTLDKIIEGSKTVSDAIGDASDPIANVG..... 49
          : |||   :  :||   :   : :|: :| : :: :
Vmp7> 101 NDLEQVAVKA..ETGVDKSDSSKNPIREAVNEAKEVLGTLKGYVESLGTI 148
ORF2>  50 ........ANNAAGVAGTGI.ESLTKGIKAIVDVVLGKEGNAEAGTDKKA 90
                  |||||| : |:  : | |::||: ::| :  : :  :    |
Vmp7> 149 GDSNPVGYANNAAGSGTTAADDELRKAFKALQEIVKAATDAGVKALKIGA 198
ORF2>  91 EDLSARTAAGNGEAGKLFAANAGDDANAKKVAADAAKAVGAVTGADILQA 140
          | |  :|:| |::|::|  :|: | |    | ||  :: | | ::|
Vmp7> 199 TTLQANEGADNKEGAKILATSGGNPAAAD..VAKAAAILSSVSGEEMLSS 246
ORF2> 141 MVKNGANAAADAAKANA.....................KDGTIAGAIAL 168
          :|| | | |  || |::                    | : :||:|||
Vmp7> 247 IVKSGENDAQLAAAADGNTSAISFAKGGSDAHLAGANTPKAAAVAGGIAL 296
ORF2> 169 TVMAKGGKFAGPTADSA.DYVTVAKGAAVSAITKALDTLTIAIRKTIDAG 217
            : | ||:|:  || | :      | :| |  | |  :   |:||:
Vmp7> 297 RSLVKTGKLAAGAADNATGGGKEVQGVGVAAANKLLRAVEDVIKKTVKNV 346
ORF2> 218 LKTVKEAM.KINANDTPISPEQNTPKATAN* 247
          |   || : |  : : |:|
Vmp7> 347 LEKAKEKIDKARGSQEPVSESSK*..... 370
```

FIG. 4. Alignment of the deduced amino acid sequences of ORF2 and the Vmp7 protein of *B. hermsii*. The Vmp7 sequence is from the work of Burman et al. (10). The amino acids are given in single-letter code. Aligned amino acids that are identical are indicated by vertical bars (|). Aligned amino acids that are chemically similar (21) are indicated by colons (:).

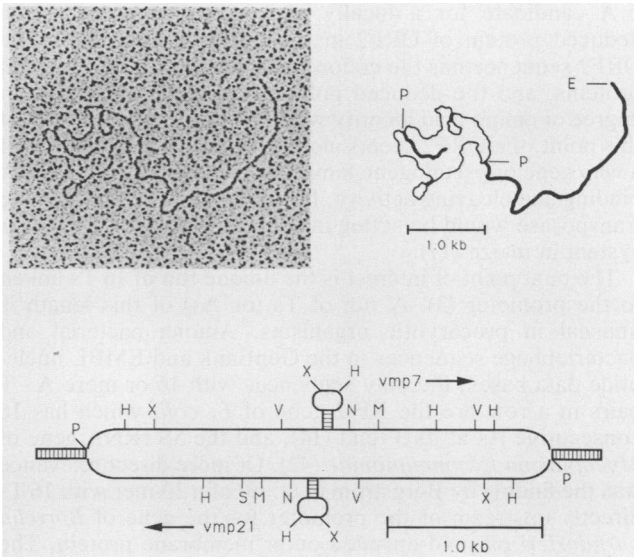


FIG. 5. Heteroduplex analysis of sequences upstream of expressed *vmp7* and *vmp21* of *B. hermsii* in recombinant plasmids p7.1 and p21.8, respectively. (Top) Heteroduplex of p7.1 and p21.8. The plasmids were cut with *Eco*RI (E). The oppositely oriented *Pst*I (P) inserts in pBR322 remain single stranded, and each contains stem-loop structures. The *Eco*RI-*Pst*I arms of pBR322 form double-stranded DNA. (Bottom) Schematic representation of the heteroduplex. Plasmid p7.1 is above and plasmid p21.8 is below. The direction of transcription of *vmp7* and *vmp21* is shown with arrows. Other restriction enzymes are *Hin*dIII (H), *Xho*II (X), *Pvu*II (V), *Nsi*I (N), and *Msp*I (M). Only a portion of each duplex arm of pBR322 is shown.

Heteroduplex analysis. The long inverted repeats and intervening sequences of the expression site might form stem-and-loop structures after denaturation. We examined the sequence immediately upstream of expressed *vmp7* and *vmp21* for stem-loop potential by heteroduplex analysis. For this study, we used the recombinant plasmids p7.1 and p21.8 (29, 35). Plasmid p7.1 contains approximately two-thirds of active *vmp7* and 2 kb of flanking upstream sequence from bp7E in a 2.9-kb *Pst*I fragment. Plasmid p21.8 has a 2.8-kb *Pst*I fragment containing the analogous region of active *vmp21* and bp21E. The inserts of p7.1 and p21.8 are oppositely oriented within the vector pBR322.

The plasmids were first cut with *Eco*RI, producing pBR322 arms of unequal length with respect to the *Pst*I insert. The linearized plasmids were then denatured, allowed to reanneal, and examined by electron microscopy. A typical heteroduplex molecule is shown in Fig. 5. As expected, the complementary single strands of the pBR322 arms of p7.1 and p21.8 annealed to one another. The oppositely oriented inserts could not align in this state and remained single stranded. Stem-loop structures formed in the borrelia DNA inserts of both p7.1 and p21.8. The lengths of the duplex stems were estimated to be 0.15 kb; the single-stranded loops were approximately 0.5 kb. Under the conditions of hybridization, the duplex that formed the stem structures required at least 80% sequence homology for annealing (20).

The location of the stem-loops in both inserts was the same with respect to the common upstream *Pst*I site and the *Nsi*I site. The stem-loops occupied the same position within the expression site as the third pair of inverted repeats (Fig. 1 and 2). This experiment also confirmed that the DNA upstream of the expressed *vmp21* gene on bp21E was highly similar, if not identical, to the sequence 5′ of the active *vmp7* gene on bp7E.

DISCUSSION

Little is known of the genetics of spirochetes, many of which cannot be cultivated in vitro. Several members of the genus *Borrelia* are cultivable, but under conditions more typical of mammalian cell cultures than of cultivation of other types of bacteria (5). As yet, there is neither a transformation nor a transduction system for spirochetes, and transposonlike elements had not previously been noted in this division of bacteria. Countering these impediments, and perhaps providing an advantage for investigators, is the siting of genes encoding major surface proteins on linear plasmids in the microorganisms (35). In previous studies, we analyzed genes for the variable proteins and the telomeric regions of these plasmids (10, 26). In the present study, we were particularly interested in the sequence 5′ to the active *vmp* gene and its regulatory sequences because of the site-specific character of the gene activation process (3, 29, 35). A single- or double-stranded break near or in the expression region has been proposed as the initiating event in the recombination between linear plasmids (1).

A 2-kb sequence segment was repeated with varying fidelity three times in the space of 7 kb. Each segment in turn has a 1-kb element that resembles an insertion sequence or a transposable element in having terminal inverted repeats (19). The heteroduplex analysis demonstrated the ability of the borrelia elements to form structures resembling insertion sequence elements under similar conditions (20). Although the pairs of inverted repeats lack target duplications typical of insertion sequences (19), the confinement of these inverted repeats to the expression plasmid suggests a role for the repeats in expression or recombination. The borrelia inverted-repeat pairs may stimulate recombination at the expression site by creating chromosomal breaks in the manner of transposable elements of maize (17). If this were the case, one might suspect the action of a "transposase" in the breakage.

A candidate for a locally encoded transposase is the deduced protein of ORF2 in the third 2-kb segment. The ORF2 sequence has the codon preferences of other *Borrelia* proteins, and the deduced polypeptide shows a significant degree of amino acid identity with the Vmp7 protein (10). At this point, the ORF2 seems more likely to be the member of a *vmp* gene or pseudogene family than an enzyme with DNA binding and cleaving activity. If this is the case, the putative transposase would be acting in *trans*, similarly to the *Ac-Ds* system in maize (17).

The next point of interest is the unique run of 16 Ts linked to the promoter (3). A run of Ts (or As) of this length is unusual in procaryotic organisms. Among bacterial and bacteriophage sequences in the GenBank and EMBL nucleotide data bases, the only sequences with 16 or more A · T pairs in a row are the RF-2 gene of *E. coli*, which has 16 consecutive As at its 3′ end (14), and the 5S rRNA gene of *Mycoplasma hyopneumoniae* (42). Of more direct relevance was the finding by Bergström et al. (9) of a 20-mer with 16 Ts directly upstream of the promoter for the gene of *Borrelia burgdorferi* plasmid-encoded outer membrane protein. The upstream T-rich tracts of *B. hermsii* and *B. burgdorferi* may enhance transcription, similarly to yeast positive regulatory sequences that are characterized by runs of Ts (40).

Apart from possibly enhancing expression, could the run of A·T pairs serve to stimulate recombination? Would, for instance, strand disruption be more likely here than in a sequence more closely approximating random nucleotide order? A consideration is the structure of this DNA. Homopurine-homopyridine tracts of DNA of this length may not be B DNA; poly(dT · dA) exhibits a helical periodicity that is distinctly different from that of random sequence DNA (34, 36). There may be sufficient strain on the homopyrimidine strand for deformation or "cracking" of the strand to occur (43). The non-B DNA inherent in this homopolymeric stretch could be the target of nuclease attack (12); a bubble in this region would be a substrate for a single-strand-specific endonuclease. The major cutting site in the nuclease-hypersensitive region upstream from the chicken adult β-globin gene is 16 consecutive deoxyguanosine residues (32). The hypothetical bubble may form at temperatures typically found in mammals; the calculated T_m of the run of 16 A-T pairs is 29°C.

In making these considerations, we have assumed that the *vmp* gene switch could be managed in a borrelia either by activities of a more general housekeeping nature or by enzymes associated with the insertion sequence-like elements. Nevertheless, we cannot rule out the mediation of recombination by an endonuclease acting specifically on telltale DNA sequences at the expression site, such as occurs in the mating-type switch of *Saccharomyces cerevisiae*. In that organism, a site-specific endonuclease makes a double-strand break that initiates recombination (27). The mating-type switching in the fission yeast *Schizosaccharomyces pombe* is also preceded by a double-stranded break, and the cutting site is located in the midst of a run of A · T pairs similar to the poly(dT · dA) stretch 5′ to the *vmp* gene (33).

Further comparison of the relapsing fever borrelias with African trypanosomes should also be made. The many similarities in the biology of antigenic variation between the procaryote and the eucaryote have been pointed out (1). Moreover, both the borrelias and the trypanosomes feature telomeric siting of their expressed genes for surface antigens and activate these genes through telomere conversion and promoter addition (3, 26). Can we now identify similarities between the 5′ regions of expressed variable antigen genes in these two groups of organisms? Here the analogy weakens. Sequences similar to the 5′ region of expressed *vmp* genes have not been noted in a similar location in trypanosomes. Upstream of the transcription unit for the variable antigen gene in *Trypanosoma brucei* is a 5- to 40-kb region characterized by a paucity of restriction enzyme sites and tandem repeats of 72 to 76 bp (11, 28, 37). Insertion sequence-like elements have not been noted in these regions.

In this study, we have demonstrated features of the upstream expression region that are unique to this part of the *B. hermsii* genome. Whether these different elements and unusual DNA sequences have a functional role in *vmp* gene switching remains to be determined. The present findings allow future studies to be directed to certain sites of the expression plasmids that have a high potential for strand breakage.

ACKNOWLEDGMENTS

We thank Vojo Deretic, Todd Kitten, and Mel Simon for advice and assistance and Connie Stahl for help in preparing the manuscript.

This work was supported by grants from the National Institutes of Health (AI24424) and National Science Foundation (DMB-8806112) to A.G.B. and from the Swedish Medical Research Council (Dnr07922) to S.B.

REFERENCES

1. **Barbour, A. G.** 1989. Antigenic variation in relapsing fever *Borrelia* species: genetic aspects, p. 783–789. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
2. **Barbour, A. G., O. Barrera, and R. Judd.** 1983. Structural analysis of the variable major proteins of *Borrelia hermsii*. J. Exp. Med. **158**:2127–2140.
3. **Barbour, A. G., N. Burman, C. J. Carter, T. Kitten, and S. Bergström.** Mol. Microbiol., in press.
4. **Barbour, A. G., and C. F. Garon.** 1987. Linear plasmids of the bacterium *Borrelia burgdorferi* have covalently closed ends. Science **237**:409–411.
5. **Barbour, A. G., and S. F. Hayes.** 1986. Biology of *Borrelia* species. Microbiol. Rev. **50**:381–400.
6. **Barbour, A. G., and H. G. Stoenner.** 1984. Antigenic variation of *Borrelia hermsii*. UCLA Symp. Mol. Cell. Biol. New Ser. **20**:123–125.
7. **Barbour, A. G., and S. L. Tessier, and H. G. Stoenner.** 1982. Variable major proteins of *Borrelia hermsii*. J. Exp. Med. **156**:1312–1324.
8. **Barstad, P. A., J. E. Coligan, M. G. Raum, and A. G. Barbour.** 1985. Variable major proteins of *Borrelia hermsii*. Epitope mapping and partial sequence analysis of CNBr peptides. J. Exp. Med. **161**:1302–1314.
9. **Bergström, S., V. G. Bundoc, and A. G. Barbour.** 1989. Molecular analysis of linear plasmid-encoded major surface proteins, OspA and OspB, of the Lyme disease spirochete *Borrelia burgdorferi*. Mol. Microbiol. **3**:479–486.
10. **Burman, N., S. Bergström, B. I. Restrepo, and A. G. Barbour.** 1990. The variable antigens Vmp7 and Vmp21 of the relapsins fever bacterium *Borrelia hermsii* are structurally analogous to the VSG proteins of the African trypanosome. Mol. Microbiol. **4**:1715–1726.
11. **Campbell, D., D. Thornton, and J. Boothroyd.** 1984. Apparent discontinuous transcription of *Trypanosoma brucei* variant surface antigen genes. Nature (London) **311**:350–355.
12. **Cantor, C. R., and A. Efstratiadis.** 1984. Possible structures of homopurine-homopyrimidine S1-hypersensitive sites. Nucleic Acids Res. **12**:8059–8072.
13. **Chen, E. J., and P. H. Seeburg.** 1985. Alkaline denaturation for double strand plasmid sequencing. DNA **4**:165–170.
14. **Craigen, W. J., R. G. Cook, W. P. Tate, and C. T. Caskey.** 1985. Bacterial peptide chain release factors: conserved primary

structure and possible frameshift regulation of release factor 2. Proc. Natl. Acad. Sci. USA **82**:3616–3620.
15. **Devereux, J., P. Haeverli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
16. **Doolittle, R. F.** 1986. Of URFs and ORFs. A primer on how to analyze derived amino acid sequences. University Science Books, Mill Valley, Calif.
17. **Federoff, N. V.** 1989. Maize transposable elements, p. 375–411. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
18. **Fickett, J.** 1982. Recognition of protein coding regions in DNA sequences. Nucleic Acids Res. **10**:5303–5318.
19. **Galas, D. J., and M. Chandler.** 1989. Bacterial insertion sequences, p. 109–162. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
20. **Garon, C. F.** 1986. Electron microscopy of nucleic acids, p. 161–179. *In* H. Aldrich and W. Todd (ed.), Ultrastructure techniques for microorganisms. Plenum Publishing Corp., New York.
20a. **Garon, C. F., and A. G. Barbour.** Unpublished data.
21. **Gribskov, M., and R. R. Burgess.** 1986. Sigma factors from *E. coli*, *B. subtilis*, phage SP01, and phage T4 are homologous proteins. Nucleic Acids Res. **14**:6745–6763.
22. **Gyllensten, U. B., and H. A. Erhlich.** 1988. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the *HLA-DQA* locus. Proc. Natl. Acad. Sci. USA **85**:7652–7656.
23. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
24. **Harr, R., P. Fallman, M. Haggström, L. Wahlström, and P. Gustafsson.** 1986. GENEUS, a computer system for DNA and protein sequence analysis containing an information retrieval system for the EMBL data library. Nucleic Acids Res. **14**:273–284.
25. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
26. **Kitten, T., and A. G. Barbour.** 1990. Juxtaposition of expressed variable antigen genes with a conserved telomere in the bacterium *Borrelia hermsii*. Proc. Natl. Acad. Sci. USA **87**:6077–6081.
27. **Kostriken, R., J. N. Strathern, A. J. S. Klar, J. B. Hicks, and F. Heffron.** 1983. A site-specific endonuclease essential for mating-type switching in *Saccharomyces cerevisiae*. Cell **33**:167–174.
28. **Liu, A. L., L. Van der Ploeg, F. Rijsewijk, and P. Borst.** 1983. The transposition unit of variant surface glycoprotein gene 118 of *Trypanosoma brucei*. Presence of repeated elements at its border and absence of promoter-associated sequences. J. Mol. Biol. **167**:57–75.
29. **Meier, J. T., M. I. Simon, and A. G. Barbour.** 1985. Antigenic variation is associated with DNA rearrangements in a relapsing fever borrelia. Cell **41**:403–409.
30. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
31. **Needleman, S. B., and C. D. Wunsch.** 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48**:443–453.
32. **Nickol, J. M., and G. Felsenfeld.** 1982. DNA conformation at the 5′ end of the chicken adult β-globin gene. Cell **35**:467–477.
33. **Nielsen, O., and R. Egel.** 1989. Mapping the double-strand breaks at the mating-type locus in fission yeast by genomic sequencing. EMBO J. **8**:269–276.
34. **Peck, L. J., and J. C. Wang.** 1981. Sequence dependence of the helical repeat of DNA in solution. Nature (London) **292**:375–378.
35. **Plasterk, R. H. A., M. I. Simon, and A. G. Barbour.** 1985. Transposition of structural genes to an expression sequence on a linear plasmid causes antigenic variation in the bacterium *Borrelia hermsii*. Nature (London) **318**:257–263.
36. **Rhodes, D., and A. Klug.** 1981. Sequence-dependent helical periodicity of DNA. Nature (London) **292**:378–380.
37. **Shah, J., J. Young, B. Kimmel, K. Iams, and R. Williams.** 1987. The 5′ flanking sequence of a *Trypanosoma brucei* variable surface glycoprotein gene. Mol. Biochem. Parasitol. **24**:163–174.
38. **Stoenner, H. G., T. Dodd, and C. Larsen.** 1982. Antigenic variation of *Borrelia hermsii*. J. Exp. Med. **156**:1297–1311.
39. **Strauss, E. C., J. A. Kobori, G. Siu, and L. E. Hood.** 1986. Specific-primer-directed DNA sequencing. Anal. Biochem. **154**:353–360.
40. **Struhl, K., W. Chen, D. E. Hill, I. A. Hope, and M. A. Oettinger.** 1985. Constitutive and coordinately regulated transcription of yeast genes: promoter elements, positive and negative regulatory sites, and DNA binding proteins. Cold Spring Harbor Symp. Quant. Biol. **50**:489–503.
41. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84**:4767–4771.
42. **Taschke, C., M. Q. Klinkert, J. Wolters, and R. Herrmann.** 1986. Organization of the ribosomal RNA genes in Mycoplasma hyopneumoniae: the 5S rRNA gene is separated from the 16S and 23S rRNA genes. Mol. Gen. Genet. **205**:428–433.
43. **Wohlrab, F., M. J. McClean, and R. D. Wells.** 1987. The segment inversion site of Herpes simplex virus type 1 adopts a novel DNA structure. J. Biol. Chem. **262**:6407–6416.
44. **Yanisch-Perron, C., J. Viera, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
45. **Zagursky, R. J., K. Baumeister, N. Lomax, and M. L. Berman.** 1985. Rapid and easy sequencing of large linear double-stranded DNA and supercoiled plasmid DNA. Gene Anal. Tech. **2**:89–94.