Past Departments Faculty Publications (CELS)   College of the Environment and Life Sciences

1998

Molecular Cloning and Sequencing of Three Granulocytic Ehrlichia Genes Encoding High-Molecular-Weight Immunoreactive Proteins

James R. Storey

Linda A. Doros-Richert

Cindy Gingrich-Baker

Kenneth Munroe

Thomas N. Mather
University of Rhode Island, tmather@uri.edu

See next page for additional authors

Follow this and additional works at: https://digitalcommons.uri.edu/cels_past_depts_facpubs

Molecular Cloning and Sequencing of Three Granulocytic Ehrlichia Genes Encoding High-Molecular-Weight Immunoreactive Proteins

Authors

James R. Storey, Linda A. Doros-Richert, Cindy Gingrich-Baker, Kenneth Munroe, Thomas N. Mather, Richard T. Coughlin, Gerald A. Beltz, and Cheryl I. Murphy

Terms of Use

Molecular Cloning and Sequencing of Three Granulocytic *Ehrlichia* Genes Encoding High-Molecular-Weight Immunoreactive Proteins

JAMES R. STOREY,[1] LINDA A. DOROS-RICHERT,[1] CINDY GINGRICH-BAKER,[1]
KENNETH MUNROE,[1] THOMAS N. MATHER,[2] RICHARD T. COUGHLIN,[1]
GERALD A. BELTZ,[1] AND CHERYL I. MURPHY[1]*

*Aquila Biopharmaceuticals, Inc., Worcester, Massachusetts 01605,[1] and Center for Vector-Borne Disease,
University of Rhode Island, Kingston, Rhode Island 02881[2]*

Granulocytic *Ehrlichia* was isolated from canine blood obtained from animals challenged with field-collected *Ixodes scapularis* and propagated in HL60 cells. PCR primers specific for the 16S ribosomal DNA (rDNA) of the *Ehrlichia* genogroup comprising *E. equi*, *E. phagocytophila*, and the agent of human granulocytic ehrlichiosis (HGE) amplified DNA from extracts of these cells. Sequence analysis of this amplified DNA revealed that it is identical to the 16S rDNA sequence of the HGE agent. A genomic library was constructed with DNA from granulocytic *Ehrlichia* and screened with pooled sera from tick-challenged, granulocytic *Ehrlichia*-infected dogs. Several clones were isolated and sequenced. Three complete genes encoding proteins with apparent molecular masses of 100, 130, and 160 kDa were found. The recombinant proteins reacted with convalescent-phase sera from dogs and human patients recovering from HGE. This approach will be useful for identifying candidate diagnostic and vaccine antigens for granulocytic ehrlichiosis and aid in the classification of genogroup members.

Members of the genus *Ehrlichia* include species which have a tropism for mononuclear phagocytes (*E. canis*, *E. chaffeensis*, *E. muris*, *E. sennetsu*, and *E. risticii*) (33) and those which infect granulocytes (*E. ewingii* [33], *E. phagocytophila* [10, 33], *E. equi* [14, 26], and the recently discovered agent of human granulocytic ehrlichiosis [HGE] [3, 7]). Disease caused by granulocytic *Ehrlichia* (GE) is manifested by fever, lethargy, thrombocytopenia, and death, and many species from diverse geographical locations have shown evidence of natural infection, including horses (25, 27, 29, 38), dogs (24, 34, 35), small mammals (39, 40), and humans (4, 15).

The similar host range and near identity of the 16S rRNA genes of *E. phagocytophila*, *E. equi*, and the HGE agent (7) have raised the possibility that these organisms represent a single species (2, 28). In addition, Dumler et al. (9) have shown that they share significant antigenicity by immunofluorescence and immunoblot assays. Objective methods for species classification, e.g., molecular genetic analysis, have not been readily available, primarily because of an inability to culture these ehrlichiae in vitro. However, we (reference 43 and unpublished data) and others (13) have recently demonstrated successful cultivation of GE isolates from dogs and humans, respectively.

In this paper, we describe the use of purified GE obtained from in vitro culture of infected HL60 cells, a promyelocytic human cell line, to generate a genomic DNA library for expression screening with sera obtained from dogs experimentally infected with GE. The screening resulted in the isolation of recombinant clones containing complete genes encoding three putative proteins of GE, GE 160, GE 130, and GE 100 (named for apparent molecular mass in kilodaltons). One of these proteins, the 100-kDa protein, is similar in both glutamic

acid content and repeated amino acid structure to an immunodominant 120-kDa *E. chaffeensis* protein (45). Both the 100- and 130-kDa granulocytic *Ehrlichia* proteins share some amino acid sequence homology to the 120-kDa *E. chaffeensis* protein.

MATERIALS AND METHODS

Isolation of GE in cell culture. The GE agent (strain USG3) was isolated as described previously (43). Briefly, blood from a dog experimentally infected by adult *Ixodes scapularis* ticks collected from both Westchester County, N.Y., and Montgomery County, Pa., was inoculated into a suspension culture of the human promyelocytic cell line HL60 (ATCC CCL 240). The cells were cultured in RPMI 1640 medium supplemented with 20% heat-inactivated fetal bovine serum, 2 mM L-glutamine, 1 mM sodium pyruvate, and 0.1 mM nonessential amino acids and maintained at 37°C and 6% CO_2 in a humidified chamber. The fetal bovine serum was reduced to 10% once the culture was established. Ehrlichial infection was indicated by lysis of the HL60 cells. Cultures of USG3 (passages 7 to 21) were grown for 5 days prior to GE purification.

Purification of GE. USG3 cultures at approximately 80% cell lysis (monitored microscopically) were centrifuged at 840 × g for 15 min at 4°C to remove host HL60 cell debris. The supernatant was filtered through a Poretics (Livermore, Calif.) 5-µm-pore-diameter polycarbonate membrane, 47 mm in diameter, followed by a Poretics 3-µm-pore-diameter filter under negative pressure. The USG3 filtrate was centrifuged at 9460 × g in a Sorvall centrifuge for 30 min at 4°C. Following centrifugation, the GE pellet was resuspended in 5 ml of 25 mM Tris (pH 8.0)–10 mM $MgCl_2$–0.9% NaCl. DNase I (Life Technologies, Gaithersburg, Md.) was added to a final concentration of 9 µg per ml, and the solution was incubated for 15 min at 37°C. Following incubation, the DNase was inactivated by the addition of 0.5 ml of 0.5 M EDTA, and the GE was pelleted at 14,000 × g in a Sorvall centrifuge for 30 min at 4°C.

PCR amplification and cloning of GE 16S ribosomal DNA. Universal eubacterial primers for the 16S rRNA gene (41) were modified to include restriction enzyme recognition sites as follows: forward primer, 5′ CTGCAGGTTTGATC CTGG 3′ (*Pst*I site); reverse primer, 5′ GGATCCTACCTTGTTACGACTT 3′ (*Bam*HI site). These primers (0.5 µM) were added to a 100-µl reaction mixture containing 1× PCR buffer II (Perkin-Elmer Corp.); 1.5 mM $MgCl_2$ (Perkin-Elmer Corp.); 200 µM (each) dATP, dGTP, dCTP, and dTTP; 2.5 U of Amplitaq DNA polymerase; and 20 µl of USG3 DNA. Amplification was performed under the following conditions: 35 cycles at 94°C for 1 min, 53°C for 1 min, and 72°C for 2 min, followed by a 72°C incubation for 10 min. The amplified 1,500-bp fragment was digested with *Pst*I and *Bam*HI and ligated to pUC19 linearized with the same enzymes. The resulting clone, pUCHGE16S, was sequenced as described below.

* Corresponding author. Mailing address: Aquila Biopharmaceuticals, 365 Plantation St., Worcester, MA 01605. Phone: (508) 797-5777, ext. 188. Fax: (508) 797-4014. E-mail: molbio@splusnet.com.
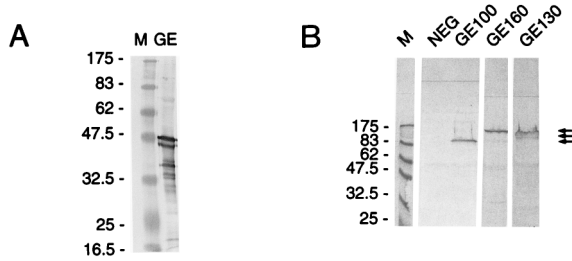
FIG. 1. Expression of GE proteins by Western blotting. (A) Purified USG3 disrupted in SDS (lane GE). (B) Individual recombinant clones containing the genes coding for GE 100, GE 160, GE 130, and a negative control (NEG [no insert]) were grown and incubated with IPTG to induce protein expression as described in Materials and Methods. Samples of each were electrophoresed on SDS-PAGE gels and transferred to nitrocellulose for Western blotting. Blots were probed with convalescent-phase dog sera. Molecular mass markers (M) (in kilodaltons) are shown to the left of each panel.

Construction of a GE genomic library. Genomic DNA was isolated from purified GE with the QIAamp Genomic DNA kit (Qiagen, Chatsworth, Calif.) for library preparation (Stratagene, La Jolla, Calif.). The DNA was mechanically sheared to a 4- to 10-kb size range and ligated to *Eco*RI linkers. Linkered fragments were ligated into the *Eco*RI site of Lambda Zap II, and the library was amplified in *Escherichia coli* XL1-Blue MRF′ to a titer of 10^{10} PFU/ml.

Expression screening of the genomic library. Bacteriophage were plated with XL1-Blue MRF′ and induced to express protein with 10 mM isopropyl-β-D-thiogalactopyranoside (IPTG) (Sigma, St. Louis, Mo.). Proteins were transferred to nitrocellulose filters, and the filters were washed with Tris-buffered saline (TBS; 25 mM Tris HCl [pH 7.5], 0.5 M NaCl). Washed filters were blocked in TBS containing 0.1% polyoxyethylene 20 cetyl ether (Brij 58) and incubated with a 1:50 dilution of pooled sera (depleted of anti-*E. coli* antibodies) taken from four dogs experimentally infected by exposure to field-collected adult *I. scapularis* ticks (8). The filters were washed and incubated with rabbit anti-dog horseradish peroxidase-conjugated immunoglobulin (Ig) antibody, rewashed, and developed with 4-chloronaphthol. Positive plaques were isolated, replated, and screened again. Plasmid DNA containing the putative recombinant clones was obtained by plasmid rescue (Stratagene).

DNA sequencing and sequence analysis. DNA sequencing of recombinant clones was performed by the primer walking method and with an ABI 373A DNA sequencer (ACGT, Northbrook, Ill.; Lark Technologies, Houston, Tex.; and Sequegen, Shrewsbury, Mass.). Sequences were analyzed by the MacVector (Oxford Molecular Group) sequence analysis program, version 6.0. The BLAST algorithm, D version 1.4 (19, 20), was used to search for homologous nucleic acid and protein sequences available on the National Center for Biotechnology Information (NCBI) server.

PCR analysis of USG3 and HL60 DNA. PCR primer sets were designed based on the sequences of each of the three GE clones and are as follows: S2 (forward, 5′-GCGTCTCCAGAACCAG-3′; reverse, 5′-CCTATATAGCTTACCG-3′), S7 (forward, 5′-GATGTTGCTTCGGGTATGC-3′; reverse, 5′-CAGAGATTACT TCTTTTTTGCGG-3′), and S22 (forward, 5′-CACGCCTTCTTCTAC-3′; reverse, 5′-CTCTGTTGCTATAGGGGC-3′). Each 50-μl reaction mixture contained 0.5 μM each primer, 1× PCR Supermix (Life Technologies, Gaithersburg, Md.) and either 100 ng of USG3 DNA, 100 ng of HL60 DNA, or 200 ng of plasmid DNA. PCR amplification was performed under the following conditions: 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min. After 30 cycles, a single 10-min extension at 72°C was done. PCR products were analyzed on 4% Nusieve 3:1 agarose gels (FMC Bioproducts, Rockland, Me.).

Western blot analysis. Individual recombinant plasmid-containing cultures were induced to express protein with 5 mM IPTG. Bacterial cells were pelleted by centrifugation and resuspended in 5× Laemmli buffer (12% glycerol, 0.2 M Tris-HCl [pH 6.8], 5% sodium dodecyl sulfate [SDS], 5% β-mercaptoethanol) at 200 μl per optical density unit of culture. Samples were boiled for 5 min, and 10 μl of each was analyzed by SDS-polyacrylamide gel electrophoresis (PAGE) (for dog sera) or NuPage (Novex, San Diego, Calif.) electrophoresis (for human sera). Proteins were transferred to nitrocellulose filters, the filters were blocked in TBS-Brij 58, and the blots were probed with either the pooled dog sera referenced above or 1:1,000 dilutions of human sera. The sera used in this study were two convalescent-phase serum samples from patients (no. 2 and 3 from New York, kindly provided by M. Aguero-Rosenfeld) and one sample from an individual in Wisconsin who was part of a seroprevalence study (no. 1, kindly provided by J. Bakken). Blots were washed and incubated with biotin-labeled goat anti-dog IgG (Kirkegaard & Perry Laboratories, Inc., Gaithersburg, Md.) followed by peroxidase-labeled streptavidin (Kirkegaard & Perry Laboratories, Inc.) or horseradish peroxidase-conjugated antihuman IgG (Bio-Rad, Hercules, Calif.). After several additional washes, the dog serum blots were developed with 4-chloronaphthol (Bio-Rad), and the human serum blots were detected by en-

hanced chemiluminescence (Pierce, Rockford, Ill.) and viewed by autoradiography.

Nucleotide sequence accession number. The nucleotide sequences of the GE genes described here have been assigned the following GenBank accession numbers: *ank* (GE 160), AF020521; *rea* (GE 130), AF020522; and *gra* (GE 100), AF020523.

RESULTS

Cloning and sequencing of the USG3 16S rRNA gene. Sources of GE used to generate a cell culture isolate have included blood from either infected humans (13) or dogs (32, 43). To assess the relatedness of the USG3 isolate to other GE isolates, including the HGE agent, the 16S rRNA gene was amplified by PCR with universal eubacterial primers (41). A 1,500-bp DNA fragment was isolated and cloned into pUC19 (see Materials and Methods), and the insert DNA was sequenced. The USG3 16S rRNA gene sequence was found to be identical to the GenBank sequence of the HGE agent (accession no. U02521). The sequences of other 16S ribosomal DNA PCR fragments generated with USG3 DNA as the template were also identical to the HGE agent sequence (reference 43 and data not shown).

Isolation of recombinant clones identified with GE-positive dog sera. Using pooled sera from adult *I. scapularis*-challenged, GE-infected dogs, positive clones were identified and purified as single plaques. pBluescript plasmids were rescued according to the Stratagene protocol, and DNA was purified with Qiagen plasmid purification kits. A number of restriction digests were performed with each clone to assess their relatedness (data not shown). Based on these digests, restriction maps were generated which showed that all of the clones represented just three different GE genes. One clone from each of the three gene groups (clones S2, S7, and S22) was chosen for further analysis.

Analysis of proteins encoded by GE clones. Because recombinant clones were selected based on immunoreactivity, bacterial lysates of each clone were analyzed by SDS-PAGE and Western blotting to identify the immunoreactive proteins. The blots were probed with the same pooled canine sera used to screen the library. Figure 1A shows the reactivity of the antisera against purified USG3. Several immunodominant proteins with sizes of 36, 43, and 45 kDa were observed, and other less-immunoreactive proteins of various molecular masses are
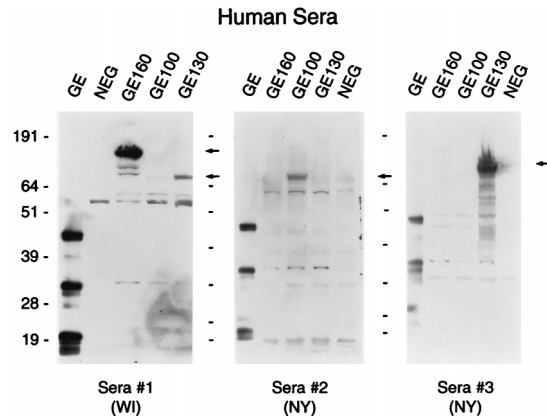
FIG. 2. Expression of GE proteins by Western blotting. Three different human serum samples were used to probe Western blots containing SDS-disrupted USG3 (GE lanes), GE 160, GE 100, and GE 130. A pBluescript library clone containing no insert was used as a negative control (NEG). The origin of the sera is indicated at the bottom of each panel (WI, Wis.; NY, N.Y.). Molecular mass markers (in kilodaltons) are shown to the left of each panel.

```
   1  TGTACCATCCCCAGTGACAGGTGAAACGCTGCGGTAGCTGCGTATAAAGGGGGCAACACT
  61  GAGGATGTTAAGACTATGTTACGCTGTAATAGCATGGACGTAGATGCTCGGTCACATGAT
                     M  L  R  C  N  S  M  D  V  D  A  R  S  H  D
 121  GGTAGAACTATAATACATTACGCAGCAAAGGATGGAAATTTAGAGATATTGCAGCAGGCT
         G  R  T  I  I  H  Y  A  A  K  D  G  N  L  E  I  L  Q  Q  A
 181  CTTGGAAGGAAGAGTAGTTATTCTAAGTTTCCTGTAAAGGATGGTGTTCCTACTCCAGGT
         L  G  R  K  S  S  Y  S  K  F  P  V  K  D  G  V  P  T  P  G
 241  GTATATGCGATTCGTGAAGCAAGTGGTGGAAAAGTATCGCTACAAGCACTTGACATGTTA
         V  Y  A  I  R  E  A  S  G  G  K  V  S  L  Q  A  L  D  M  L
 301  ATGAGATATGAGCCTCACCCGCAGCATGTTGCTGTGCGAGGCAGTAAGAACAGGTGCAGTA
         M  R  Y  E  P  H  P  Q  H  V  A  V  E  A  V  R  T  G  A  V
 361  GGTGTATTGGAGCACCTTATTACCACTGAAGTGATTAGTGTAAATGAAGAAATTACAACT
         G  V  L  E  H  L  I  T  T  E  V  I  S  V  N  E  E  I  T  T
 421  CCTGAAGGAAAAAAGACAACTTTGACCGCTGAAGCACTAACTAGTGGTAAATATGGTGTA
         P  E  G  K  K  T  T  L  T  A  E  A  L  T  S  G  K  Y  G  V
 481  GTGAAGGCGTTAATTAAAAACAGTGCTGATGTAAATGCGTCTCCAGAACCAGCTATTACT
         V  K  A  L  I  K  N  S  A  D  V  N  A  S  P  E  P  A  I  T
 541  TTGGGTATACAAGGAAGGTGCTTTCAGGGGAGTAAAGCTATAAAGCATTTAAAGCGTGTT
         L  G  I  Q  G  R  C  F  Q  G  S  K  A  I  K  H  L  K  R  V
 601  GTAGAAGCTGGGGCACATATAAATACTCCTACCGGATCTATGAGCCCTTTAGCTGCTGCA
         V  E  A  G  A  H  I  N  T  P  T  G  S  M  S  P  L  A  A  A
 661  GTTCAAGCGGCAAATGAGGCAAGTAACCTTAAAGAGGCTAATAAGATTGTAAATTTCCTT
         V  Q  A  A  N  E  A  S  N  L  K  E  A  N  K  I  V  N  F  L
 721  TTACATAGGGGTGCAGATCTTTCGTCTACGGAACACACTGGAACTCCAGCCTTGCATTTA
         L  H  R  G  A  D  L  S  S  T  E  H  T  G  T  P  A  L  H  L
 781  GCAACAGCTGCTGGCAACCATAGGACTGCTATGTTGCTCTTGGATAAAGGGGCTCCAGCA
         A  T  A  A  G  N  H  R  T  A  M  L  L  L  D  K  G  A  P  A
 841  ACGCAGAGAGATGCTAGGGGTAGGACGGCTTTACATATAGCAGCTGCTAATGGTGACGGT
         T  Q  R  D  A  R  G  R  T  A  L  H  I  A  A  A  N  G  D  G
 901  AAGCTATATAGGATGATTGCGAAAAAATGCCCAGATAGCTGTCAACCACTCTGTTCTGAT
         K  L  Y  R  M  I  A  K  K  C  P  D  S  C  Q  P  L  C  S  D
 961  ATGGGAGATACAGCGTTACATGAGGCTTTATATTCTGATAATGTTACAGAAAAATGCTTT
         M  G  D  T  A  L  H  E  A  L  Y  S  D  N  V  T  E  K  C  F
1021  TTAAAGATGCTTAAAGAGTCTCGAAAGCATTTGTCAAACTCATCTTTTTCGGAGACTTG
         L  K  M  L  K  E  S  R  K  H  L  S  N  S  S  F  F  G  D  L
1081  CTTAATACTCCTCAAGAAGCAAATGGTGACACGTTACTGCATCTGGCTGCATCGCGTGGT
         L  N  T  P  Q  E  A  N  G  D  T  L  L  H  L  A  A  S  R  G
1141  TTCGGTAAAGCATGTAAAATACTACTAAAGGCTGGGGCGTCAGTATCAGTCGTGAATGTA
         F  G  K  A  C  K  I  L  L  K  A  G  A  S  V  S  V  V  N  V
1201  GAGGGCAAAAACACCGGTAGATGTTGCGGATCCATCATTGAAAACTCGTCCGTGGTTTTTT
         E  G  K  T  P  V  D  V  A  D  P  S  L  K  T  R  P  W  F  F
1261  GGAAAGTCCGTTGTCACAATGATGGCTGAACGTGTTCAAGTTCCTGAAGGGGGATTCCCA
         G  K  S  V  V  T  M  M  A  E  R  V  Q  V  P  E  G  G  F  P
1321  CCATATCTGCCGCCTGAAAGTCCAACTCCTTCTTTAGGATCTATTTCAAGTTTTGAGAGT
         P  Y  L  P  P  E  S  P  T  P  S  L  G  S  I  S  S  F  E  S
1381  GTCTCTGCGCTATCATCCTTGGGTAGTGGCCTAGATACTGCAGGAGCTGAGGAGTCTATC
         V  S  A  L  S  S  L  G  S  G  L  D  T  A  G  A  E  E  S  I
1441  TACGAAGAAATTAAGGATACAGCAAAAGGTACAACGGAAGTTGAAAGCACATATACAACT
         Y  E  E  I  K  D  T  A  K  G  T  T  E  V  E  S  T  Y  T  T
1501  GTAGGAGCTGAGGAGTCTATCTACGAAGAAATTAAGGATACAGCAAAAGGTACAACGGAA
         V  G  A  E  E  S  I  Y  E  E  I  K  D  T  A  K  G  T  T  E
1561  GTTGAAAGCACATATACAACTGTAGGAGCTGAAGGTCCGGAGAACACCAGAAGGTGAAGAT
         V  E  S  T  Y  T  T  V  G  A  E  G  P  R  T  P  E  G  E  D
1621  CTGTATGCCTACTGTGGGAGCTGCAATTACTTCCGAGGCGCAAGCATCAGATGCGGCGTCA
         L  Y  A  T  V  G  A  A  I  T  S  E  A  Q  A  S  D  A  A  S
1681  TCTAAGGGAGAAAGGCCGGAATCCATTTATGCTGATCCATTTGATATAGTGAAACCTAGG
         S  K  G  E  R  P  E  S  I  Y  A  D  P  F  D  I  V  K  P  R
1741  CAGGAAAGGCCTGAATCTATCTATGCTGACCCATTTGCTGCGGAACGAACATCTTCTGGA
         Q  E  R  P  E  S  I  Y  A  D  P  F  A  A  E  R  T  S  S  G
1801  GTAACGACATTTGGCCCTAAGGAAGAGCCGATTTATGCAACAGTGAAAAAGGGTCCTAAG
         V  T  T  F  G  P  K  E  E  P  I  Y  A  T  V  K  K  G  P  K
1861  AAGAGTGATACTTCTCAAAAAGAAGGAACAGCTTCTGAAAAAGTCTGCTCAACAATAACT
         K  S  D  T  S  Q  K  E  G  T  A  S  E  K  V  C  S  T  I  T
1921  GTGATTAAGAAGAAAGTGAAACCTCAGGTTCCAGCTAGGACAAGTAGTTTGCCTACTAAA
         V  I  K  K  K  V  K  P  Q  V  P  A  R  T  S  S  L  P  T  K
1981  GAAGGTATAGGTTCTGATAAAGACCTGAGTTCAGGAACTAGTAGCTCTTTTGCAGCTGAG
         E  G  I  G  S  D  K  D  L  S  S  G  T  S  S  F  A  A  E
2041  CTGCAAGCACAAAGGGGTAAATTGCGTCCTGTGAAGGGAGGTGCTCCGGATTCTACCAAA
         L  Q  A  Q  R  G  K  L  R  P  V  K  G  G  A  P  D  S  T  K
2101  GACAAAACAGCTACTTCTATATTCTCCAGTAAAGAGTTCAAAAAGGAACTAACAAAAGCT
         D  K  T  A  T  S  I  F  S  S  K  E  F  K  K  E  L  T  K  A
2161  GCCGAAGGATTACAGGGAGCAGTTGAAGAAGCTCAGAAGGGTGATGGAGGAGCTGCAAAG
         A  E  G  L  Q  G  A  V  E  E  A  Q  K  G  D  G  G  A  A  K
2221  GCAAAGCAAGATCTTGGCATGGAATCTGGTGCCCCAGGATCTCAACCAGAAGCTCCTCAA
         A  K  Q  D  L  G  M  E  S  G  A  P  G  S  Q  P  E  A  P  Q
2281  AGTGAAGGCCCTAAGTCTGTAAAAGGAGGTCGCGGTAGGTAGAATTATACCGAAAAATCG
         S  E  G  P  K  S  V  K  G  G  R  G  R
2341  CTGAGGTACT
```

FIG. 3. Sequence of the 160-kDa protein gene. Nucleotide numbers are indicated to the left. The ATG start codon and TAA stop codon are shown in boldface type. The translated amino acid sequence for the open reading frame is displayed underneath the DNA sequence according to the single-letter amino acid code. Upstream regulatory sequences are underlined and represent the −35, −10, and ribosome binding sites.

shown. Figure 1B shows the Western blot of the library clones, S2, S7, and S22. A single high-molecular-mass protein (indicated by arrows) was detected by the sera for each isolated clone. These proteins were not expressed in the absence of

IPTG induction and were not detected with preimmune dog sera (data not shown). The approximate molecular masses of each protein are as follows: 160 (S2), 100 (S7), and 130 (S22) kDa. These proteins are not the major immunodominant antigens of the purified cell culture isolate USG3, two of which migrate at about 43 to 45 kDa (Fig. 1A), but they could be present among the higher-molecular-mass immunoreactive proteins.

The three GE proteins were also analyzed by Western blotting for their reactivity with human sera. Figure 2 shows three different human serum samples tested against purified USG3 antigen and the three GE recombinant proteins. Each serum sample reacted strongly with one of the three GE proteins: no. 1 with GE 160 and less strongly with GE 130, no. 2 with GE 100, and no. 3 with GE 130. All three serum samples reacted with several proteins in the purified USG3 lanes (GE lanes). A

```
   1  ATCGTGTAGCACTTAAACAAATGCTGGTGAACGTAGAGGGATTAAAGGAGGATTTGCGTA
  61  TATGTATGGTATAGATATAGAGCTAAGTGATTACAGAATTGGTAGTGAAACCATTTCCAG
         M  Y  G  I  D  I  E  L  S  D  Y  R  I  G  S  E  T  I  S  S
 121  TGGAGATGATGGCTACTACGAAGGATGTGCTTGTGACAAAGATGCCAGCACTAATGCGTA
         G  D  D  G  Y  Y  E  G  C  A  C  D  K  D  A  S  T  N  A  Y
 181  CTCGTATGACAAGTGTAGGGTAGTACGGGGAACGTGGAGACCGAGCGAACGTTGGTTTTATA
         S  Y  D  K  C  R  V  V  R  G  T  W  R  P  S  E  L  V  L  Y
 241  TGTTGGTGATGAGCATGTGGCATGTAGAGATGTTGCTTCGGGTATGCATCATGGTAATTT
         V  G  D  E  H  V  A  C  R  D  V  A  S  G  M  H  H  G  N  L
 301  GCCAGGGAAGGTGTATTTTATAGAGGCAGAAGCGGGCAGAGCTGCTACTGCTGAAGGTGG
         P  G  K  V  Y  F  I  E  A  E  A  G  R  A  A  T  A  E  G  G
 361  TGTTTATACTACCGTTGTGGAGGCCATTATCGCTGGTGCAAGAGGAAGAGGGTACAGGTAT
         V  Y  T  T  V  V  E  A  L  S  L  V  Q  E  E  E  G  T  G  M
 421  GTACTTGATAAACGCACCAGAAAAAGCGGTCGTAAGGTTTTTCAAGATAGAAAAGAGTGC
         Y  L  I  N  A  P  E  K  A  V  V  R  F  F  K  I  E  K  S  A
 481  AGCAGAGGAACCTCAAACAGTAGATCCTAGTGTAGTTGAGTCAGCAACAGGGTCGGGTGT
         A  E  E  P  Q  T  V  D  P  S  V  V  E  S  A  T  G  S  G  V
 541  AGATACGCAAGAAGAACAAGAAATAGATCAAGAAGCACCAGCAATTGAAGAAGTTGAGAC
         D  T  Q  E  E  Q  E  I  D  Q  E  A  P  A  I  E  E  V  E  T
 601  AGAAGAGCAAGAAGTTATTCTGGAAGAAGGTACTTTGATAGATCTTGAGCAACCTGTAGC
         E  E  Q  E  V  I  L  E  E  G  T  L  I  D  L  E  Q  P  V  A
 661  GCAAGTACCTGTAGTAGCTGAAGCAGAATTACCTGGTGTTGAAGCTGCAGAAGCGGATTGT
         Q  V  P  V  V  A  E  A  E  L  P  G  V  E  A  A  E  A  I  V
 721  ACCATCACTAGAAGAAAATAAGCTTCAAGAAGTGGTAGTTGCTCCAGAAGCGCCAACAACT
         P  S  L  E  E  N  K  L  Q  E  V  V  V  A  P  E  A  Q  Q  L
 781  AGAATCAGCTCCTGAAGTTTCTGCGCCAGCACAACCTGAGTCTACAGTTCTTGGTGTTGC
         E  S  A  P  E  V  S  A  P  A  Q  P  E  S  T  V  L  G  V  A
 841  TGAAGGTGATCTAAAGTCTGAAGTATCTGTAGAAGCTAATGCTGATGTACCGCAAAAAGA
         E  G  D  L  K  S  E  V  S  V  E  A  N  A  D  V  P  Q  K  E
 901  AGTAATCTCTGGTCAACAAGAGCAAGAAATTGCAGAAGCACTAGAGGGAACTGAAGCTCC
         V  I  S  G  Q  Q  E  Q  E  I  A  E  A  L  E  G  T  E  A  P
 961  TGTAGAAGTAAAAGAAGAAACAGAAGTTCTTCTAAAGGAAGATACTTTGATAGATCTTGA
         V  E  V  K  E  E  T  E  V  L  L  K  E  D  T  L  I  D  L  E
1021  GCAACCTGTAGCACAAGTACCTGTAGTAGCTGAAGCAGAATTACCTGGTGTTGAAGCTGC
         Q  P  V  A  Q  V  P  V  V  A  E  A  E  L  P  G  V  E  A  A
1081  AGAAGCGGATTGTACCATCACTAGAAGAAAATAAGCTTCAAGAAGTGGTAGTTGCTCCAGA
         E  A  I  V  P  S  L  E  E  N  K  L  Q  E  V  V  V  A  P  E
1141  AGCGCAACAACTAGAATCAGCTCCTGAAGTTTCTGCGCCAGCACAACCTGAGTCTACAGT
         A  Q  Q  L  E  S  A  P  E  V  S  A  P  A  Q  P  E  S  T  V
1201  TCTTGGTGTTACTGAAGGTGATCTGAAGTCTGAAGTATCTGTAGAAGCTGATGCTGGTAT
         L  G  V  T  E  G  D  L  K  S  E  V  S  V  E  A  D  A  G  M
1261  GCAGCAAGAAGCAGGAATCTCTGATCAAGAGACACAAGCAACTGAAGAAGTTGAAAAGGT
         Q  Q  E  A  G  I  S  D  Q  E  T  Q  A  T  E  E  V
1321  TGAAGTATCTGTAGAAACAAAAACGGAAGAGCCAGAAGTTATTCTAGAAGAAGGTACTTT
         E  V  S  V  E  T  K  T  E  E  P  E  V  I  L  E  E  G  T  L
1381  GATAGATCTTGAGCAACCTGTAGCGCAAGTACCTGTAGTAGCTGAAGCAGAATTACCTGG
         I  D  L  E  Q  P  V  A  Q  V  P  V  V  A  E  A  E  L  P  G
1441  TGTTGAAGCTGCAGAAGCGATTGTACCATCACTAGAAGAAAATAAGCTTCAAGAAGTGGT
         V  E  A  A  E  A  I  V  P  S  L  E  E  N  K  L  Q  E  V  V
1501  AGTTGCTCCAGAAGCGCAACAACTAGAATCAGCTCCTGAAGTTTCTGCGCCAGTACAACC
         V  A  P  E  A  Q  Q  L  E  S  A  P  E  V  S  A  P  V  Q  P
1561  TGAGTCTACAGTTCTTGGTGTTACTGAAGGTGATCTGAAGTCTGAAGTATCTGTAGAAGC
         E  S  T  V  L  G  V  T  E  G  D  L  K  S  E  V  S  V  E  A
1621  TGATGCTGGTATGCAGCAAGAAGCAGGAATCTCTGATCAAGAGACACAAGCAACTGAAGA
         D  A  G  M  Q  Q  E  A  G  I  S  D  Q  E  T  Q  A  T  E  E
1681  AGTTGAGAAGGTTGAAGTATCTGTAGAAGCTGATGCTGGTATGCAGCAAGAGTTAGTAGA
         V  E  K  V  E  V  S  V  E  A  D  A  G  M  Q  Q  E  L  V  D
1741  TGTTCCGACTGCTTTGCCGTTAAAGGATCCTGACGATGAAGATGTTCTAAGTTATTAGGA
         V  P  T  A  L  P  L  K  D  P  D  D  E  D  V  L  S  Y
1801  TATCTTTCTCGTGAAAAGTATGGGGAAGGT
```

FIG. 4. Sequence of the 100-kDa protein gene. Nucleotide numbers are indicated to the left. The ATG start codon and TAA stop codon are shown in boldface type. The translated amino acid sequence for the open reading frame is displayed underneath the DNA sequence according to the single-letter amino acid code. A possible ribosome binding site is underlined.

```
   1  TCGACGCAGGCGGATTTTGTACTATGTTTCACTTAACAAGGTATTATAAATGTTTGAACA
                                                           M  F  E  H
  61  CAATATTCCTGATACATACACAGGAACAACTGCAGAAGGTTCTCCTGGCTTAGCAGGCGG
       N  I  P  D  T  Y  T  G  T  T  A  E  G  S  P  G  L  A  G  G
 121  GGATTTTAGCTTAAGTTCTATTGACTTTACAAGGGACTTTACAATTGAATCACATAGAGG
       D  F  S  L  S  S  I  D  F  T  R  D  F  T  I  E  S  H  R  G
 181  AAGCTCAGCTGATGACCCAGGTTACATCAGCTTTAGGGATCAAGACGGAAACGTCATGTC
       S  S  A  D  D  P  G  Y  I  S  F  R  D  Q  D  G  N  V  M  S
 241  ACGTTTTCTTGATGTGTACGTAGCTAATTTCAGCTTGCGATGCAAGCATTCTCCCTATAA
       R  F  L  D  V  Y  V  A  N  F  S  L  R  C  K  H  S  P  Y  N
 301  CAACGACAGAATGGAAACAGCTGCGTTCTCTCTAACTCCCGACATAATAGAGCCTTCTGC
       N  D  R  M  E  T  A  A  F  S  L  T  P  D  I  I  E  P  S  A
 361  TTTATTGCAAGAATCACATAGTACACAAAACAATGTAGAAGAGGCAGTACAAGTTACAGC
       L  L  Q  E  S  H  S  T  Q  N  N  V  E  E  A  V  Q  V  T  A
 421  TCTTGAGTGCCCTCCATGTAATCCAGTCCCTGCCGAGGAAGTAGCTCCTCAACCGTCTTT
       L  E  C  P  P  C  N  P  V  P  A  E  E  V  A  P  Q  P  S  F
 481  TCTAAGCAGAATAATTCAGGCGTTCTTGTGGTTATTCACGCCTTCTTCTACTACCGACAC
       L  S  R  I  I  Q  A  F  L  W  L  F  T  P  S  S  T  T  D  T
 541  TGCTGAAGACAGCAAGTGTAATAGTAGCGATACTTCAAAATGTACCTCTGCTAGCAGTGA
       A  E  D  S  K  C  N  S  S  D  T  S  K  C  T  S  A  S  S  E
 601  GTCATTAGAGCAGCAACAAGAATCAGTGGAAGTGCAACCATCTGTACTTATGTCTACTGC
       S  L  E  Q  Q  Q  E  S  V  E  V  Q  P  S  V  L  M  S  T  A
 661  CCCTATAGCAACAGAGCCTCAAAATGCGGTTGTTAACCAAGTAAACACTACTGCAGTACA
       P  I  A  T  E  P  Q  N  A  V  V  N  Q  V  N  T  T  A  V  Q
 721  AGTAGAATCATCCATTATTGTGCCAGAATCGCAACACACTGACGTTACCGTGCTCGAAGA
       V  E  S  S  I  I  V  P  E  S  Q  H  T  D  V  T  V  L  E  D
 781  TACTACTGAGACGATAACTGTTGATGGGGAATATGGACATTTTAGTGACATTGCTTCAGG
       T  T  E  T  I  T  V  D  G  E  Y  G  H  F  S  D  I  A  S  G
 841  TGAACACAATAACGATCTGCCTGCCATGTTGTTAGATGAAGCAGACTTCACTATGTTATT
       E  H  N  N  D  L  P  A  M  L  L  D  E  A  D  F  T  M  L  L
 901  AGCGAACGAGGAGTCAAAGACCCTGGAGTCTATGCCTTCTGATAGCCTAGAAGACAATGT
       A  N  E  E  S  K  T  L  E  S  M  P  S  D  S  L  E  D  N  V
 961  TCAGGAACTAGGTACATTGCCTTTACAAGAAGGTGAAACAGTTTCTGAGGGCAACACACG
       Q  E  L  G  T  L  P  L  Q  E  G  E  T  V  S  E  G  N  T  R
1021  AGAGTCACTACCCACTGACGTTTCACAAGACTCAGTTGGTGTAAGTACAGATCTTGAAGC
       E  S  L  P  T  D  V  S  Q  D  S  V  G  V  S  T  D  L  E  A
1081  TCATTCTCAAGAAGTTGAAACAGTTTCTGAGGTCAGCACACAAGATTCACTATCCACTAA
       H  S  Q  E  V  E  T  V  S  E  V  S  T  Q  D  S  L  S  T  N
1141  CATTTCACAAGACTCAGTTGGTGTAAGTACAGATCTTGAAGCTCATTCTAAAGGAGTTGA
       I  S  Q  D  S  V  G  V  S  T  D  L  E  A  H  S  K  G  V  E
1201  AATAGTTTCTGAGGGCGGCACACAAGATTCACTATCCGCTGATTTTCCAATAAACACAGT
       I  V  S  E  G  G  T  Q  D  S  L  S  A  D  F  P  I  N  T  V
1261  TGAAAGTGAAAGTACAGATCTTGAAGCTCATTCTCAAGAAGTTGAAACTGTTTCTGAATT
       E  S  E  S  T  D  L  E  A  H  S  Q  E  V  E  T  V  S  E  F
1321  CACACAAGATTCACTATCCACTAACATTTCACAAGACTCAGTTGGTGTAAGTACAGATCT
       T  Q  D  S  L  S  T  N  I  S  Q  D  S  V  G  V  S  T  D  L
1381  TGAAGTTCATTCTCAAGAAGTTGAAATAGTTTCTGAGGGCGGCACACAAGATTCACTATC
       E  V  H  S  Q  E  V  E  I  V  S  E  G  G  T  Q  D  S  L  S
1441  CACTAACATTTCACAAGACTCAGTTGGTGTAAGTACAGATCTTGAAGCTCATTCTCAAGA
       T  N  I  S  Q  D  S  V  G  V  S  T  D  L  E  A  H  S  Q  E
1501  AGTTGAAACTGTTTCTGAATTCACACAAGATTCACTATCCACTAACATTTCACAAGACTC
       V  E  T  V  S  E  F  T  Q  D  S  L  S  T  N  I  S  Q  D  S
1561  AGTTGGTGTAAGTACAGATCTTGAAGTTCATTCTCAAGAAGTTGAAATAGTTTCTGAGGG
       V  G  V  S  T  D  L  E  V  H  S  Q  E  V  E  I  V  S  E  G
1621  CGGCACACAAGATTCACTATCCACTAACATTTCACAAGACTCAGTTGGTGTAAGTACAGA
       G  T  Q  D  S  L  S  T  N  I  S  Q  D  S  V  G  V  S  T  D
1681  TCTTGAAGCTCATTCTAAAGGAGTTGAAATAGTTTCTGAGGGCGGCACACAAGATTCACT
       L  E  A  H  S  K  G  V  E  I  V  S  E  G  G  T  Q  D  S  L
1741  ATCCGCTGATTTTCCAATAAACACAGTTGAAAGTGAAAGTACAGATCTTGAAGCTCATTC
       S  A  D  F  P  I  N  T  V  E  S  E  S  T  D  L  E  A  H  S
1801  CCCAGAAGGTGAAATAGTTTCTGAGGTCAGCACACAAGATGCGCCATCCACTGGAGTAGA
       P  E  G  E  I  V  S  E  V  S  T  Q  D  A  P  S  T  G  V  E
1861  GATCAGATTTATGGATCGTGATTCTGATGATGACGTGCTCGCGTTGTGAAGTGATCATGG
       I  R  F  M  D  R  D  S  D  D  D  V  L  A  L
1921  TAGGGGAAA
```

FIG. 5. Sequence of the 130-kDa protein gene. Nucleotide numbers are indicated to the left. The ATG start codon and TAA stop codon are shown in boldface type. The translated amino acid sequence for the open reading frame is displayed underneath the DNA sequence according to the single-letter amino acid code. A possible ribosome binding site is underlined.

serum sample from a Rhode Island HGE patient recognized two of the three GE proteins, and other HGE patient sera tested also recognized one or two of the proteins (data not shown). None of the samples we have tested to date react strongly with all three recombinant GE proteins. Negative control human sera did not react with the GE proteins on Western blots (data not shown).

It is interesting to note that these human sera do not strongly recognize proteins with sizes of 100, 130, and 160 kDa with USG3 lysates as the antigen source. This may be due to underrepresentation of these proteins in the purified cell culture antigen, either because they are secreted or because they are expressed only after infection of a host species.

DNA sequencing and database analysis of recombinant clones. The inserts from each of the recombinant plasmids S2, S7, and S22 were completely sequenced by the primer walking method. Sequence analysis (MacVector 6.0; Oxford Molecular Group) showed that each clone contained a single large open reading frame encoded by the plus strand of the insert and that each one appeared to be a complete gene. The portions of the DNA sequence containing the open reading frame and corresponding amino acid sequence of each of the clones are shown as follows: Fig. 3, S2, GE 160; Fig. 4, S7, GE 100; and Fig. 5, S22, GE 130. A possible ribosome binding site sequence upstream from the initiating ATG codon is underlined in each figure.

The numbers of nucleotides sequenced upstream of each reading frame were 1,500 bp for the S2 insert, 500 bp for the S22 insert, and 200 bp for the S7 insert. These sequences were examined for likely ehrlichial promoter sequences based on homology to the *E. coli* −35 and −10 consensus sequences (16). Of the three clones, S2 had the closest homology to the *E. coli* promoter sequences, and these are underlined in Fig. 3. A number of possible promoter sequences were found for the S22 and S7 genes, but in the absence of direct experimental evidence, it is difficult to determine which sequence elements are necessary for promoter function. Protein expression in all three clones was IPTG dependent, and thus it is likely that the vector *lacZ* promoter is used for transcription of the *Ehrlichia* RNA in *E. coli*.

Amino acid sequence analysis of the proteins encoded by the three gene clones showed that all contain regions of repeated amino acids. A schematic version of these repeat structures is shown in Fig. 6 and 7. The 160-kDa protein (Fig. 6) has three groups of repeats. The first set consists of a number of ankyrin-like repeat units of 33 amino acids, the second consists of repeat units of 27 amino acids, and the third consists of repeat units of 11 amino acids. The ankyrin repeats were revealed by a BLAST database search for protein homologies. Ankyrin repeats occur in at least four consecutive copies and are present in yeast, plants, bacteria, and mammals (5, 6). Figure 6 shows a multiple alignment of the 160-kDa protein ankyrin repeats under a consensus sequence derived from the analysis


```
                 160 kDa Protein

Amino acid    100  200  300  400  500  600  700

      ■   Ankyrin-like repeat units (1-8)
      ▨   27 amino acid repeats
      □   11 amino acid repeats
```

```
              160 kDa protein ankyrin repeats

Consensus:      t  S/TP L H   A h    t t  t h h t  L L t    t    t
Repeat 1    D G R T I I H Y A A K D G N L E I L Q Q A L G R K S S Y S K F P V
       2    K K T T L T A E A L T S G K Y G V V K A L I K N S A D V N A S P E
       3    A V Q A A N E A S N L K E A N K I V N F L L H R G A D L S S T E H
       4    T G T P A L H L A T A A G N H R T A M L L L D K G A P A T Q R D A
       5    R G R T A L H I A A A N G D G K L Y R M I A K K C P D S C Q P L C
       6    M G D T A L H E A L Y S D N V T E K C F L K M L K E S R K R L S N
       7    N G D T L L H L A A S R G F G K A C K I L L K A G A S V S V V N V
       8    E G K T P V D V A D P S L K T R P W F F G K S V V T M M A E R V Q
```
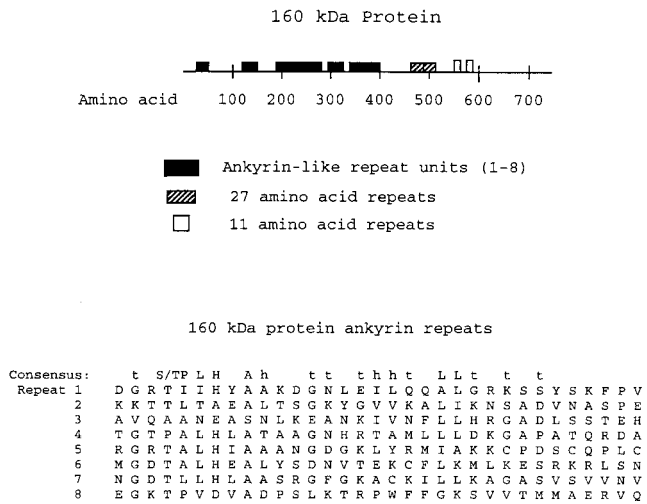
FIG. 6. Schematic diagram of the GE 160-kDa protein. Repeat regions are indicated by the boxes. Sequences of proposed ankyrin repeats, numbered 1 to 8, are aligned according to the consensus sequence at the top (6). h, hydrophobic; t, turn-like or polar; S/T, serine or threonine; capital letters, conserved amino acids.
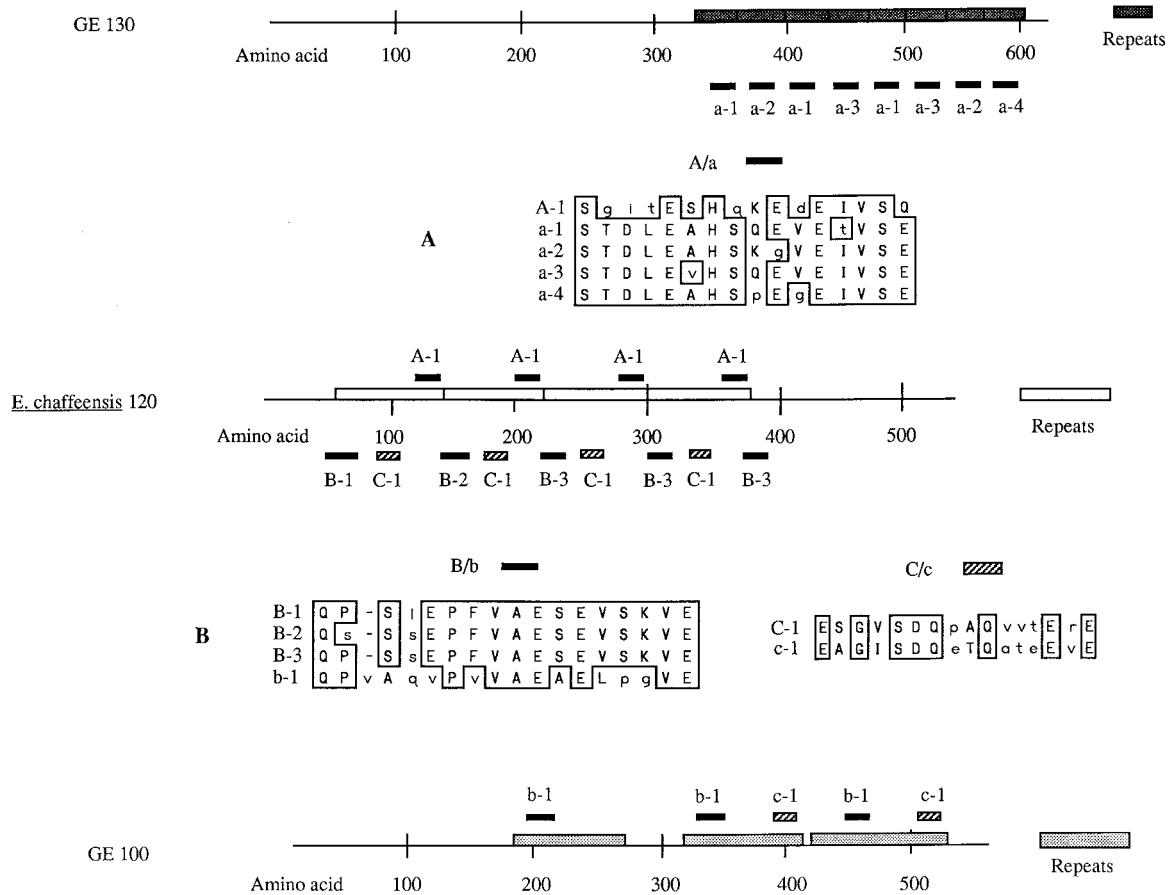
FIG. 7. Amino acid sequence alignments of selected regions of GE 130-kDa and *E. chaffeensis* 120-kDa proteins (A) and GE 100-kDa and *E. chaffeensis* 120-kDa proteins (B). Each protein is shown as a linear amino acid sequence, and amino acids are numbered in hundreds. Boxed regions on the linear sequence represent repeated amino acids. (A) Amino acid alignments of a sequence which occurs four times in the *E. chaffeensis* protein (45) (top line of alignment, A-1) and eight times in the GE 130-kDa protein (a-1 to a-4). Sequence a-1 is repeated three times, related sequences a-2 and a-3 are each repeated twice, and related sequence a-4 is found once. The positions of these sequences in the proteins are indicated by the small bold lines. (B) Alignments of two different sequence motifs which occur in the *E. chaffeensis* 120-kDa protein (B-1 to B-3 and C-1) and the GE 100-kDa protein (b-1 and c-1). Bold and cross-hatched boxes indicate the positions of these sequences in the proteins. Identical amino acids are surrounded by boxes, and conserved amino acids are in capital letters.

of several hundred similar ankyrin-like motifs (6). The eighth repeat sequence holds to the consensus only through the first half of the repeat unit and may not represent a full ankyrin-like repeat.

The 130-kDa protein (Fig. 7) has a repeat unit of 26 to 34 amino acids which occurs eight times in the carboxy-terminal half of the protein. The sequence varies somewhat from repeat to repeat. A database homology search with the NCBI BLAST algorithm revealed that the 130-kDa protein has limited homology to the *E. chaffeensis* 120-kDa protein (45). An amino acid sequence alignment of a motif common to both proteins is shown in Fig. 7A. This motif is represented by a bold line and occurs four times in an identical fashion in the *E. chaffeensis* protein (designated A-1) and eight times with four variations in the GE 130-kDa protein (a-1, a-2, a-3, and a-4).

The 100-kDa protein (Fig. 7) has three large repeat units, which differ somewhat in length. A database search revealed that it is similar to the 120-kDa *E. chaffeensis* protein, which contains four repeats of 80 amino acids each (45). Both proteins contain large amounts of glutamic acid: 18% for the 100-kDa protein and 17% for the 120-kDa protein. When the two protein sequences are aligned, most of the homology occurs in the repeat regions. Figure 7B shows alignments for two homologous groups of amino acid motifs from the two proteins

(designated B/b and C/c) found with the BLAST algorithm. These are not the only possible alignments of the two proteins but do provide an example of their similarities. The locations of the homologous sequences are indicated by bold or hatched lines above (GE 100) or below (*E. chaffeensis* 120) the respective proteins. The B sequence represented by the bold line varies slightly in the *E. chaffeensis* protein (shown as B-1, B-2, and B-3) and occurs a total of five times. The GE 100 protein equivalent, b-1, is invariant and occurs three times. The sequence represented by the hatched line occurs four times in *E. chaffeensis* 120 (C-1) and two times in GE 100 (c-1).

All of the GE proteins migrated at higher apparent molecular masses under reducing conditions than calculations would predict. The GE 160 protein consists of 748 amino acids with a calculated molecular mass of 78.8 kDa. The GE 100 protein has 578 amino acids, and its calculated molecular mass is 61.4 kDa. The GE 130 protein contains 619 amino acids and has a calculated molecular mass of 66.1 kDa.

Verification that clones are GE derived with PCR analysis. Based on the DNA sequences of each clone, PCR primers were designed to amplify specific regions of each open reading frame (see Materials and Methods). Primer pairs specific for S22, S2, and S7 were used in separate PCRs to amplify three different templates: GE DNA (from USG3), HL60 DNA, or
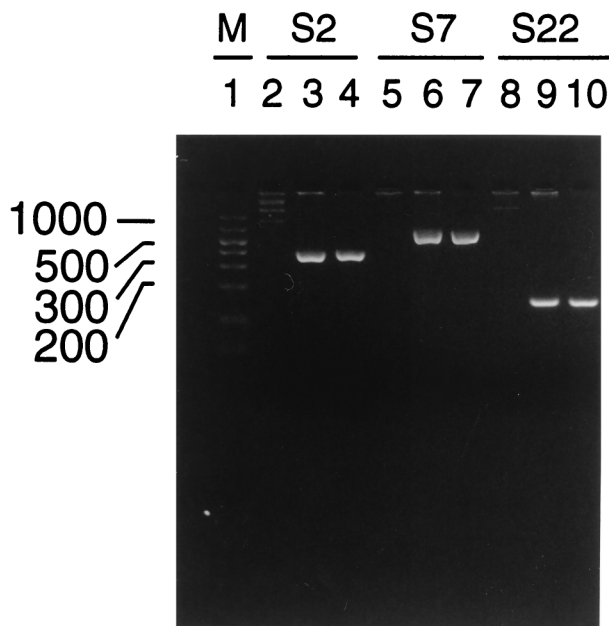
FIG. 8. PCR analysis of GE genes. PCRs were performed as described in Materials and Methods, and the products were analyzed with 4% Nusieve gels. S2 primers were used to amplify a 395-bp region of S2 DNA with HL60 DNA (lane 2), S2 plasmid DNA (lane 3), and USG3 DNA (lane 4) used as templates. S7 primers were used to amplify a 643-bp region of S7 DNA with HL60 DNA (lane 5), S7 plasmid DNA (lane 6), and USG3 DNA (lane 7) used as templates. S22 primers were used to amplify a 159-bp region of S22 DNA with HL60 DNA (lane 8), S22 plasmid DNA (lane 9), and USG3 DNA (lane 10) used as templates. DNA molecular size markers (M) (50 to 1,000 bp [FMC, Rockland, Maine]) are present in lane 1.

the purified plasmid DNA of each clone. Figure 8 shows the results obtained for primers of S2 (lanes 2 to 4), S7 (lanes 5 to 7), and S22 (lanes 8 to 10). All three clones were specific to GE and were not present in HL60 DNA (lanes 2, 5, and 8). In each case, the size of the PCR product determined with genomic DNA as the template was the same as that generated by the purified plasmid DNA.

DISCUSSION

Ehrlichiae which have a tropism for peripheral blood granulocytes and are closely related based on the sequence of the 16S rRNA gene have been classified as either *E. phagocytophila*, *E. equi*, or the agent of HGE, depending on the species and geographical location of the infected host. *E. phagocytophila* is responsible for a tick-borne fever of cattle, sheep, and goats and occurs primarily in Europe (10, 33), and *E. equi* is found in horses in North and South America (14, 26). The agent of HGE has been identified as the organism responsible for disease in patients in the upper midwestern and northeastern United States (3, 7). However, classification of GE has been difficult because of a lack of genetic information.

The GE isolate we describe here is indistinguishable from the HGE agent based on the sequence of the 16S rRNA gene. We also describe three previously unidentified proteins of GE, the 160-, 130-, and 100-kDa proteins. Aside from the 16S rRNA gene and the relatively conserved *groEL* homolog (22), the genes coding for these proteins are the first sequenced genes reported for this organism. Using PCR primers based on the sequences of the three genes, we have confirmed that the genes are specific for GE and that they are not derived from the HL60 cell line used to culture the organism. Preliminary

results also show that these genes are present in other members of the GE genogroup, including *E. phagocytophila* (31a).

The three GE proteins described in this study exhibit several interesting features. All three of the proteins contain repeated regions of amino acids. In the 100-kDa protein, the repeats comprise the majority of the protein, with the exception of the amino terminus. The direct repeats in both the 160- and 130-kDa proteins occur in the carboxyl half of the molecules. Various biological functions have been demonstrated or implicated for repeat regions found in other proteins. These include antigenic variation (17, 31) and ligand interaction (11, 42), both of which are thought to occur in several species of pathogenic bacteria (42). In several proteins of the malaria parasite *Plasmodium falciparum*, repeated regions are often the major targets for the host antibody response (21). Proteins with several repeated domains are also often found associated with the cell surface (12, 42). In addition to the direct repeats, the 160-kDa protein contains seven to eight ankyrin-like repeats. These domains have been found in over 90 different proteins from nearly all phyla, including both *E. coli* and humans (6). They are present in functionally diverse proteins, such as enzymes, toxins, and transcription factors, and are thought to play a role in protein-protein interactions (5). Human erythrocyte ankyrin contains 22 repeats, some of which may be involved in binding tubulin, spectrin, or vimentin. The repeats could thus enable the protein to act like a bridge between cytoskeleton and membrane components (23). The myriad functions ascribed to proteins containing repeated regions may imply that the GE proteins described here will be found to have roles in host cell binding, antigenic variation, or some other cell surface function.

The presence of repeated regions in the 100-, 130-, and 160-kDa proteins may in part explain the differences between the expected molecular masses and the observed molecular masses on SDS-PAGE, which are much higher. This phenomenon has been reported for other proteins containing repeated regions, including rickettsial proteins (1, 36), malaria antigens (21), the fibronectin-binding protein from *Staphylococcus aureus* (18, 37), and the *E. chaffeensis* 120-kDa protein (45). The difference between calculated and observed molecular mass varies from 10 to 30% for the malaria antigens to almost 100% for the *S. aureus* and *E. chaffeensis* proteins. Investigators have speculated that high percentages of certain amino acids in these proteins or posttranslational modifications may contribute to the aberrant migration in an SDS-polyacrylamide gel system (18, 37, 45).

A search of the major protein and nucleotide databases with the NCBI BLAST algorithm did not reveal any significant similarities of the GE proteins to any known proteins, other than the ankyrin repeats of the 160-kDa protein and the homology between the 100-kDa protein or 130-kDa protein and the 120-kDa protein of *E. chaffeensis*. The function of the *E. chaffeensis* 120-kDa protein is not known, but the hydrophilic nature of its repeat domain suggests that it may be located on the surface of *E. chaffeensis* (45). The GE 100-kDa protein is also hydrophilic in the repeat region (data not shown) and may also be surface associated. Although both the 100- and 130-kDa proteins have some homology to the *E. chaffeensis* 120-kDa protein, they are not homologous to one another. It is possible that whatever the function(s) of the 120-kDa protein is for *E. chaffeensis*, similar functions may be carried out in GE through the proteins GE 100 and GE 130.

It is unclear whether the three GE proteins described in this report may be useful as potential diagnostic reagents or vaccine candidates. Although they are not well represented in the purified-culture-grown organism, either because they are se-

creted into the culture medium or their expression is tied more closely to natural infection, they are clearly important immunologically. All three react with convalescent-phase sera from dogs, sheep (data not shown), and humans diagnosed with HGE. The *E. chaffeensis* 120-kDa protein, the counterpart of the 100-kDa GE protein, has shown sensitivity and specificity as a diagnostic reagent with a panel of *E. chaffeensis* patient sera (44). However, Western blots with convalescent-phase sera from HGE patients do not show a consistent response to the GE 100-kDa protein. It is not known whether there is any immunologic cross-reactivity between the GE 100-kDa protein and the *E. chaffeensis* 120-kDa protein or whether the two could be used to distinguish monocytic from granulocytic ehrlichiosis. Some human antisera have been reported to react with both *E. chaffeensis* and HGE antigens, but it is not known whether this represents cross-reactive epitopes or separate exposures to the two ehrlichial agents (30, 32).

In summary, the expression library approach for the isolation and characterization of GE proteins has resulted in the discovery of three previously unidentified genes which express high-molecular-weight, immunoreactive proteins. Further study of these proteins will aid in the classification of the members of the GE genogroup as corresponding genes are sequenced and compared and will be important in elucidating the role of these proteins in immunity and pathogenesis.

ACKNOWLEDGMENTS

REFERENCES

1. **Allred, D. R., T. C. McGuire, G. H. Palmer, S. R. Leib, T. M. Harkins, T. F. McElwain, and A. F. Barbet.** 1990. Molecular basis for surface antigen size polymorphisms and conservation of a neutralization-sensitive epitope in *Anaplasma marginale*. Proc. Natl. Acad. Sci. USA **87:**3220–3224.
2. **Anderson, B. E., J. E. Dawson, D. C. Jones, and K. H. Wilson.** 1991. *Ehrlichia chaffeensis*, a new species associated with human ehrlichiosis. J. Clin. Microbiol. **29:**2838–2842.
3. **Bakken, J. S., J. S. Dumler, S.-M. Chen, M. R. Eckman, L. L. Van Etta, and D. H. Walker.** 1994. Human granulocytic ehrlichiosis in the upper Midwest. A new species emerging? JAMA **272:**212–218.
4. **Bakken, J. S., J. Krueth, C. Wison-Nordskog, R. L. Tilden, K. Asanovich, and J. S. Dumler.** 1996. Clinical and laboratory characteristics of human granulocytic ehrlichiosis. JAMA **275:**199–205.
5. **Blank, V., P. Kourilsky, and A. Israel.** 1992. NF-kB and related proteins: Rel/dorsal homologies meet ankyrin-like repeats. Trends Biochem. Sci. **17:**135–140.
6. **Bork, P.** 1993. Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? Proteins Struct. Funct. Genet. **17:**363–374.
7. **Chen, S.-M., J. S. Dumler, J. S. Bakken, and D. H. Walker.** 1994. Identification of a granulocytotropic *Ehrlichia* species as the etiologic agent of human disease. J. Clin. Microbiol. **32:**589–595.
8. **Coughlin, R. T., D. Fish, T. N. Mather, J. Ma, C. Pavia, and P. Bulger.** 1995. Protection of dogs from Lyme disease with a vaccine containing outer surface protein (Osp) A, Osp B, and the saponin adjuvant QS-21. J. Infect. Dis. **171:**1049–1052.
9. **Dumler, J. S., K. M. Asanovich, J. S. Bakken, P. Richter, R. Kimsey, and J. E. Madigan.** 1995. Serologic cross-reactions among *Ehrlichia equi*, *Ehrlichia phagocytophila*, and human granulocytic ehrlichia. J. Clin. Microbiol. **33:**1098–1103.
10. **Foggie, A.** 1951. Studies on the infectious agent of tick-borne fever in sheep. J. Pathol. Bacteriol. **63:**1–15.
11. **Gaillard, J.-L., P. Berche, C. Frehel, E. Gouin, and P. Cossart.** 1991. Entry of *L. monocytogenes* into cells is mediated by internalin, a repeat protein reminiscent of surface antigens from Gram-positive cocci. Cell **65:**1127–1141.
12. **Gilmore, R. D.** 1993. Comparison of the *rompA* gene repeat regions of *Rickettsiae* reveals species-specific arrangements of individual repeating units. Gene **125:**97–102.
13. **Goodman, J. L., C. Nelson, B. Vitale, J. E. Madigan, J. S. Dumler, T. J. Kurtti, and U. G. Munderloh.** 1996. Direct cultivation of the causative agent of human granulocytic ehrlichiosis. N. Engl. J. Med. **334:**209–215.
14. **Gribble, D. H.** 1969. Equine ehrlichiosis. J. Am. Vet. Med. Assoc. **155:**462–469.
15. **Hardalo, C. J., V. Quagliarello, and J. S. Dumler.** 1995. Human granulocytic ehrlichiosis in Connecticut: report of a fatal case. Clin. Infect. Dis. **21:**910–914.
16. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11:**2237–2255.
17. **Hoft, D. F., K. S. Kim, K. Otsu, D. R. Moser, W. J. Yost, J. H. Blumin, J. E. Donelson, and L. V. Kirchhoff.** 1989. *Trypanosoma cruzi* expresses diverse repetitive protein antigens. Infect. Immun. **57:**1959–1967.
18. **Hollingshead, S. K., V. A. Fischetti, and J. R. Scott.** 1986. Complete nucleotide sequence of type 6M protein of the group A *Streptococcus*. J. Biol. Chem. **261:**1677–1686.
19. **Karlin, S., and S. F. Altschul.** 1990. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proc. Natl. Acad. Sci. USA **87:**2264–2268.
20. **Karlin, S., and S. F. Altschul.** 1993. Applications and statistics for multiple high scoring segments in molecular sequences. Proc. Natl. Acad. Sci. USA **90:**5873–5877.
21. **Kemp, D. J., R. L. Coppel, and R. F. Anders.** 1987. Repetitive proteins and genes of malaria. Annu. Rev. Microbiol. **41:**181–208.
22. **Kolbert, C. P., E. S. Bruinsma, A. S. Abdulkarim, E. K. Hofmeister, R. B. Tompkins, S. R. Telford III, P. D. Mitchell, J. Adams-Stich, and D. H. Persing.** 1997. Characterization of an immunoreactive protein from the agent of human granulocytic ehrlichiosis. J. Clin. Microbiol. **35:**1172–1178.
23. **Lux, S. E., K. M. John, and V. Bennett.** 1990. Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. Nature (London) **344:**36–42.
24. **Madewell, B. R., and D. H. Gribble.** 1982. Infection in two dogs with an agent resembling *Ehrlichia equi*. J. Am. Vet. Med. Assoc. **180:**512–514.
25. **Madigan, J. E., and D. H. Gribble.** 1987. Equine ehrlichiosis in northern California: 49 cases (1968–1981). J. Am. Vet. Med. Assoc. **190:**445–448.
26. **Madigan, J. E., S. Hietala, S. Chalmers, and E. DeRock.** 1990. Seroepidemiologic survey of antibodies to *Ehrlichia equi* in horses of northern California. J. Am. Vet. Med. Assoc. **196:**1962–1964.
27. **Madigan, J. E.** 1993. Equine ehrlichiosis, p. 209–214. *In* Z. Woldehiwet and M. Ristic (ed.), Rickettsial and chlamydial diseases of domestic animals. Pergamon Press, Oxford, United Kingdom.
28. **Madigan, J. E., P. J. Richter, R. B. Kimsey, J. E. Barlough, J. S. Bakken, and J. S. Dumler.** 1995. Transmission and passage in horses of the agent of human granulocytic ehrlichiosis. J. Infect. Dis. **172:**1141–1144.
29. **Madigan, J. E., J. E. Barlough, J. S. Dumler, N. S. Schankman, and E. DeRock.** 1996. Equine granulocytic ehrlichiosis in Connecticut caused by an agent resembling the human granulocytotropic ehrlichia. J. Clin. Microbiol. **34:**434–435.
30. **Magnarelli, L. A., J. S. Dumler, J. F. Anderson, R. C. Johnson, and E. Fikrig.** 1995. Coexistence of antibodies to tick-borne pathogens of babesiosis, ehrlichiosis, and Lyme borreliosis in human sera. J. Clin. Microbiol. **33:**3054–3057.
31. **Matsumoto, M., Y. Tange, T. Okada, Y. Inoue, T. Horiuchi, Y. Kobayashi, and S. Fujita.** 1996. Deletion in the 190 kDa antigen repeat region of *Rickettsia rickettsii*. Microb. Pathog. **20:**57–62.
31a.**Murphy, C. I., et al.** Unpublished observations.
32. **Nicholson, W. L., J. A. Comer, J. W. Sumner, C. Gingrich-Baker, R. T. Coughlin, L. A. Magnarelli, J. G. Olson, and J. E. Childs.** 1997. An indirect immunofluorescence assay using a cell culture-derived antigen for detection of antibodies to the agent of human granulocytic ehrlichiosis. J. Clin. Microbiol. **35:**1510–1516.
33. **Rikihisa, Y.** 1991. The tribe *Ehrlichiae* and ehrlichial diseases. Clin. Microbiol. Rev. **4:**286–308.
34. **Rikihisa, Y., S. A. Ewing, J. C. Fox, A. G. Siregar, F. H. Pasaribu, and M. B. Malole.** 1992. Analyses of *Ehrlichia canis* and a canine granulocytic *Ehrlichia* infection. J. Clin. Microbiol. **30:**143–148.
35. **Rodgers, S. J., R. J. Morton, and C. A. Baldwin.** 1989. A serological survey of *Ehrlichia canis*, *Ehrlichia equi*, *Rickettsia rickettsii*, and *Borrelia burgdorferi* in dogs in Oklahoma. J. Vet. Diagn. Invest. **1:**154–159.
36. **Schuenke, K. W., and D. H. Walker.** 1994. Cloning, sequencing, and expression of the gene coding for an antigenic 120-kilodalton protein of *Rickettsia conorii*. Infect. Immun. **62:**904–909.
37. **Signas, C., G. Raucci, K. Jonsson, P. Lindgren, G. M. Anantharamaiah, M. Hook, and M. Lindberg.** 1989. Nucleotide sequence of the gene for a fibronectin-binding protein from *Staphylococcus aureus*: use of this peptide sequence in the synthesis of biologically active peptides. Proc. Natl. Acad. Sci. USA **86:**699–703.
38. **Stannard, A. A., D. H. Gribble, and R. S. Smith.** 1969. Equine ehrlichiosis: a disease with similarities to tick-borne fever and bovine petechial fever. Vet. Rec. **84:**149–150.
39. **Telford, S. R., III, J. E. Dawson, P. Katavolos, C. K. Warner, C. P. Kolbert, and D. H. Persing.** 1996. Perpetuation of the agent of human granulocytic ehrlichiosis in a deer tick-rodent cycle. Proc. Natl. Acad. Sci. USA **93:**6209–6214.

40. **Walls, J. J., B. Greig, D. F. Neitzel, and J. S. Dumler.** 1997. Natural infection of small mammal species in Minnesota with the agent of human granulocytic ehrlichiosis. J. Clin. Microbiol. **35:**853–855.
41. **Wilson, K. H., R. B. Blitchington, and R. C. Greene.** 1990. Amplification of bacterial 16S ribosomal DNA with polymerase chain reaction. J. Clin. Microbiol. **28:**1942–1946.
42. **Wren, B. W.** 1991. A family of clostridial and streptococcal ligand-binding proteins with conserved C-terminal repeat sequences. Mol. Microbiol. **5:**797–803.

43. **Yeh, M.-T., T. N. Mather, R. T. Coughlin, C. Gingrich-Baker, J. W. Sumner, and R. F. Massung.** 1997. Serologic and molecular detection of granulocytic ehrlichiosis in Rhode Island. J. Clin. Microbiol. **35:**944–947.
44. **Yu, X.-J., P. Crocquet-Valdes, L. C. Cullman, and D. H. Walker.** 1996. The recombinant 120-kilodalton protein of *Ehrlichia chaffeensis*, a potential diagnostic tool. J. Clin. Microbiol. **34:**2853–2855.
45. **Yu, X.-J., P. Crocquet-Valdes, and D. H. Walker.** 1997. Cloning and sequencing of the gene for a 120-kDa immunodominant protein of *Ehrlichia chaffeensis*. Gene **184:**149–154.

Editor: P. E. Orndorff