

## Genetic Diversity of the 28-Kilodalton Outer Membrane Protein Gene in Human Isolates of *Ehrlichia chaffeensis*

XUE-JIE YU, JERE W. McBRIDE, AND DAVID H. WALKER\*

Department of Pathology and WHO Collaborating Center for Tropical Diseases, The University of Texas Medical Branch at Galveston, Galveston, Texas 77555-0609

Received 3 August 1998/Returned for modification 14 December 1998/Accepted 4 January 1999

The *Ehrlichia chaffeensis* 28-kDa outer membrane protein (p28) gene was sequenced completely by genomic walking with adapter PCR. The DNA sequence of the p28 gene was nearly identical to the previously reported sequence (N. Ohashi, N. Zhi, Y. Zhang, and Y. Rikihisa, *Infect. Immun.* 66:132–139, 1998), but analysis of a further 75 bp on the 5' end of the gene revealed DNA that encoded a 25-amino-acid signal sequence. The leader sequence was removed from the N terminus of a 30-kDa precursor to generate the mature p28 protein. A monoclonal antibody (MAB), 1A9, recognizing four outer membrane proteins of *E. chaffeensis* (Arkansas strain) including the 25-, 26-, 27-, and 29-kDa proteins (X.-J. Yu, P. Brouqui, J. S. Dumler, and D. Raoult, *J. Clin. Microbiol.* 31:3284–3288, 1993) reacted with the recombinant p28 protein. This result indicated that the four proteins recognized by MAB 1A9 were encoded by the multiple genes of the 28-kDa protein family. DNA sequence alignment analysis revealed divergence of p28 among all five human isolates of *E. chaffeensis*. The *E. chaffeensis* strains could be divided into three genetic groups on the basis of the p28 gene. The first group consisted of the Sapulpa and St. Vincent strains. They had predicted amino acid sequences identical to each other. The second group contained strain 91HE17 and strain Jax, which only showed 0.4% divergence from each other. The third group contained the Arkansas strain only. The amino acid sequences of p28 differed by 11% between the first two groups, by 13.3% between the first and third groups, and by 13.1% between the second and third groups. The presence of antigenic variants of p28 among the strains of *E. chaffeensis* and the presence of multiple copies of heterogeneous genes suggest a possible mechanism by which *E. chaffeensis* might evade the host immune defenses. Whether or not immunization with the p28 of one strain of *E. chaffeensis* would confer cross-protection against other strains needs to be investigated.

*Ehrlichia chaffeensis* (1, 8) is a member of the family *Rickettsiaceae* and the etiologic agent of a newly emerging infectious disease, human monocytotropic ehrlichiosis. The disease is transmitted by ticks, and *Amblyomma americanum* is the predominant tick vector (2). *E. chaffeensis* is a small, obligately intracellular gram-negative bacterium which resides in an endosome in the host cell. *E. chaffeensis* possesses several immunodominant proteins, including the 120-, 66-, and 58-kDa proteins and a group of low-molecular-mass proteins in the range of 22 to 29 kDa (4, 5, 31). The low-molecular-mass proteins are encoded by a multiple gene family (20). The 28-kDa protein (p28) is a member of this protein family. The outer membrane proteins of *E. chaffeensis* may serve as adhesins for the organism or as targets for the host immune response. There is considerable interest in the 28-kDa protein because it is a surface-exposed outer membrane protein (6, 31), and immunization with recombinant p28 can prevent infection of mice with *E. chaffeensis* (20). However, the p28 protein is antigenically diverse (4), and heterogeneity of the reaction of convalescent-phase sera with *E. chaffeensis* antigen has been observed (7). Therefore, it is important to evaluate the genetic and antigenic diversity of the *E. chaffeensis* protein, which might be used for diagnostic purposes or vaccine production. In this study, we characterized the genetic divergence of the *E. chaffeensis* p28 gene.

### MATERIALS AND METHODS

**Ehrlichia spp.** *E. chaffeensis* strain Arkansas (1, 8) was obtained from Jacqueline Dawson (Centers for Disease Control and Prevention, Atlanta, Ga.). *E. chaffeensis* strains St. Vincent and Jax (22) were obtained from Christopher Paddock (Centers for Disease Control and Prevention, Atlanta, Ga.). *E. chaffeensis* strains 91HE17 and Sapulpa (7, 10) were isolated previously in our laboratory from patients with ehrlichiosis. Ehrlichiae were cultivated in DH82 cells, a canine macrophage-like cell line (9). DH82 cells were harvested with a cell scraper when 100% of the cells were infected with ehrlichiae. The cells were centrifuged at  $17,400 \times g$  for 20 min. The pellets were disrupted with a Braun-Sonic 2000 sonicator at 40 W for 30 s twice on ice. The cell lysate was loaded onto discontinuous gradients of 42%-36%-30% Renografin and then centrifuged at  $80,000 \times g$  for 60 min. Ehrlichiae in the heavy and light bands were collected (29) and washed by centrifugation with sucrose-phosphate-glutamate buffer (SPG; 218 mM sucrose, 3.8 mM  $\text{KH}_2\text{PO}_4$ , 7.2 mM  $\text{K}_2\text{HPO}_4$ , 4.9 mM glutamate [pH 7.0]).

**DNA preparation.** *E. chaffeensis* genomic DNA was prepared from Renografin density gradient-purified ehrlichiae by using an IsoQuick nucleic acid extraction kit (ORCA Research, Inc., Bothell, Wash.) according to the instructions of the manufacturer. Plasmid DNA was purified by using a High Pure plasmid isolation kit (Boehringer Mannheim Corp., Indianapolis, Ind.). The PCR product was purified by using a QIAquick PCR purification kit (Qiagen, Inc., Santa Clarita, Calif.).

**PCR amplification of the unknown DNA sequence of the p28 gene.** The DNA sequence of the p28 gene was amplified by using adapter PCR with the Genome-Walker kit (Clontech Laboratories, Inc., Palo Alto, Calif.). The *E. chaffeensis* (Arkansas strain) genomic DNA was digested completely with each of five restriction enzymes, including *Dra*I, *Eco*RV, *Pvu*II, *Sca*I, and *Stu*I. All five enzymes produced blunt end DNA fragments. Each batch of digested genomic DNA fragments was ligated with an adapter to create genomic libraries. Each end of an *E. chaffeensis* DNA fragment was ligated to an adapter. The genomic libraries were used as templates to amplify the unknown DNA sequence of the p28 gene by using adapter PCR. Primers were designed from the reported DNA sequence of the p28 gene (20). The unknown 5' end sequence of the p28 gene was amplified by nested PCR amplifications with primers complementary to the sense strand of the gene and the adapter primers. The downstream sequence of the p28 gene was amplified by nested PCR amplifications with primers complementary to the sense strand of the gene and the adapter primers. The down-

\* Corresponding author. Mailing address: Department of Pathology, WHO Collaborating Center for Tropical Diseases, 301 University Blvd., University of Texas Medical Branch, Galveston, TX 77555-0609. Phone: (409) 772-2856. Fax: (409) 772-2500. E-mail: dwalker@utmb.edu.

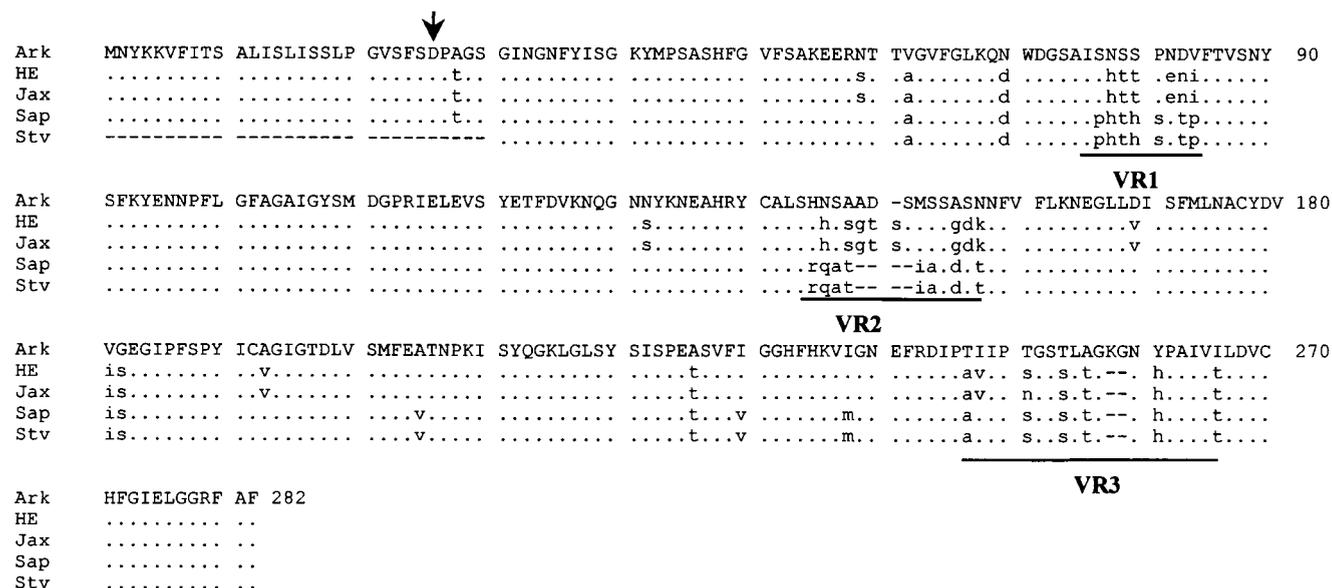


FIG. 1. Amino acid sequence alignment of the p28 proteins of five *E. chaffeensis* human isolates. The complete amino acid sequence of the Arkansas strain is presented as the consensus sequence. Differences from the consensus sequence are presented in lowercase. Dots represent the amino acids identical to those of the Arkansas strain, and dashes indicate gaps which were introduced for optimal alignment of the amino acid sequences. The variable regions (VR1, VR2, and VR3) are underlined. The arrow indicates the cleavage site of the signal peptide. The N-terminal portion of the St. Vincent strain sequence is incomplete. Ark, Arkansas; HE, 91HE17; Sap, Sapulpa; Stv, St. Vincent.

stream sequence of the p28 gene was amplified by nested PCR amplification with primers from the sense strand of the gene and the adapter primers.

**PCR amplification of the p28 gene from other *E. chaffeensis* human isolates.** Primers were designed from the p28 gene of Arkansas strain and used to amplify the p28 genes of other *E. chaffeensis* isolates by using PCR. The primer pair made up of p28f159 (ACT TCT ACT ATT GTT AAT TTA TTG TC) and p28r1336 (GCT GTT GTG TAA CTG TAG ACT GGT) was used to amplify the p28 genes of the 91HE17, Jax, and Sapulpa strains. The primer pair made up of p28f263 (AGT ATC ATT TTC CGA CCC AGC AGG TAG) and p28r1336 was used to amplify the p28 gene of the St. Vincent strain. PCR amplification was performed for 30 cycles at 94°C for 1 min, 55°C for 1 min, and 72°C for 2 min in a Perkin-Elmer thermal cycler (Perkin-Elmer Applied Biosystems, Foster City, Calif.).

**DNA sequencing.** The PCR products were sequenced directly by using PCR primers. DNA was sequenced with an ABI Prism 377 DNA sequencer (Perkin-Elmer Applied Biosystems).

**Gene analysis.** DNA sequences and deduced amino acid sequences were analyzed by using the Wisconsin GCG software package (Genetics Computer Group, Inc., Madison, Wis.) and DNASTAR software (DNASTAR, Inc., Madison, Wis.). The signal sequence of the deduced protein was analyzed by using the PSORT program (22a), which predicts the presence of signal sequences (17, 28) and detects potential transmembrane domains (15).

**Expression of the *E. chaffeensis* p28 gene in *Escherichia coli*.** pET29p28, a recombinant plasmid containing the gene coding for the mature p28 protein (20), was kindly provided by Yasuko Rikihisa (Ohio State University, Columbus, Ohio). The p28 gene in pET29p28 was removed by double digestion with *EcoRI* and *NotI* and then was directionally cloned into a pGEX expression vector (Amersham-Pharmacia Biotech, Piscataway, N.J.), which is routinely used in our laboratory, and was expressed as a glutathione *S*-transferase (GST) fusion protein in *E. coli* BL21. The GST fusion protein was affinity purified by using glutathione Sepharose 4B beads (Amersham Pharmacia Biotech). The recombinant p28 protein was cleaved from the GST fusion protein with thrombin.

**Western immunoblotting.** A murine monoclonal antibody (MAb), 1A9, to the *E. chaffeensis* (Arkansas strain) 28-kDa protein was reported previously (31). MAb 1A9 was used to react with the *E. chaffeensis* recombinant p28 protein by protein immunoblotting as described previously (31).

**Nucleotide sequence accession number.** The DNA sequences of the *E. chaffeensis* p28 gene were assigned the following GenBank accession numbers (by strain): Arkansas, AF068234; 91HE17, AF077732; Jax, AF077733; Sapulpa, AF077734; and St. Vincent, AF077735.

## RESULTS

**Precursor of the 28-kDa protein.** The p28 gene of *E. chaffeensis* (Arkansas strain) has previously been sequenced

partially (20). To sequence the p28 gene completely, both the upstream and downstream unknown sequences of the p28 gene were amplified from the genomic DNA of *E. chaffeensis* Arkansas by using adapter PCR. DNA sequencing of the PCR products demonstrated that the previously reported DNA sequence of the p28 gene is incomplete on its 5' end and the p28 gene sequence extends 75 nucleotides upstream from the previously reported sequence start site (20) (Fig. 1). The additional 75 nucleotides of the gene sequence encode a 25-amino-acid peptide that has characteristic features of the signal sequence for an exported protein in *E. coli* (20). Signal sequences for exported proteins in *E. coli* are characterized by a positively charged amino terminus (1 to 3 amino acids) followed by a hydrophobic core of 12 to 20 amino acids and a preferred signal peptidase cleavage site (21). Within the first 25 amino acids of the p28 of *E. chaffeensis*, positions 4 and 5 are strongly positively charged amino acids (both positions are lysine). Thirteen of 25 (52%) amino acids are hydrophobic. The three carboxyl-terminal amino acids of the signal sequence, serine, phenylalanine, and serine, at positions of 24, 25, and 26, respectively, are among the preferred amino acid sequences of signal peptidase at its processing site (21). The leader sequence was predicted by using a PSORT program (22a), which predicted the presence of a signal peptide (17, 28). The leader peptide of the 28-kDa protein has 72 and 68% homology with the N-terminal amino acid sequences of OMP-1E and OMP-1F of *E. chaffeensis*, respectively. The N-terminal sequences of OMP-1E and OMP-1F have been proposed as leader signal peptides (20).

The entire p28 gene of *E. chaffeensis* (Arkansas strain) was sequenced. The predicted amino acid sequence of the mature p28 was identical to the previously reported sequences (20), except for a single-amino-acid substitution in the C-terminal sequence. At position 255 of the mature 28-kDa protein, the amino acid is alanine instead of valine. The predicted molecular masses from the deduced amino acids are 30.3 kDa for the

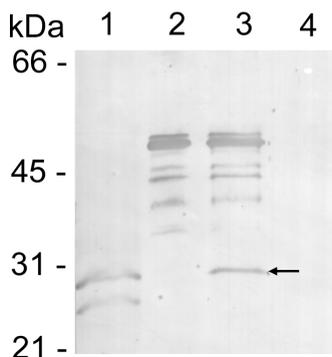


FIG. 2. Protein immunoblotting of MAb 1A9 reacted with the p28 recombinant protein. Lanes: 1, heat-denatured *E. chaffeensis* (Arkansas strain) antigen; 2, GST fusion protein with p28; 3, thrombin-cleaved GST fusion protein (the arrow indicates the thrombin-cleaved recombinant p28); 4, GST protein only. The multiple bands in lanes 2 and 3 were apparently degradation products of the GST fusion protein.

precursor, 2.7 kDa for the leader signal sequence, and 27.6 kDa for the mature protein.

**Recombinant p28 reactivity with MAb 1A9.** Protein immunoblotting demonstrated that MAb 1A9 reacted with both the thrombin-cleaved recombinant *E. chaffeensis* protein and the GST fusion protein, but not GST alone (Fig. 2). This result demonstrated that the p28 contained the antigenic epitope that reacts with MAb 1A9.

**PCR amplification of the p28 gene of *E. chaffeensis* isolates.** The p28 multiple gene family consists of at least six tandemly arranged genes, including the p28 gene. The DNA sequences of these genes are 70 to 80% homologous (20). The primers derived from one gene may amplify other genes in the multiple-gene family because of the substantial homology of particular segments of the genes. To specifically amplify the p28 gene, all primers were designed from the DNA sequences of the p28 gene which had no sequence homology with the DNA sequences of other genes. We aligned the complete DNA sequences of the p28 genes with the genes in the multiple-gene family, including DNA sequences of *omp-1c*, *omp-1d*, *omp-1e*, and *omp-1f* (20) and the related *Cowdria ruminantium* major antigenic protein-1 (MAP-1) gene (27). DNA alignment analysis revealed variable DNA sequence regions both inside the genes and in the intergenic sequences. The primer pair p28f159 and p28r1336 was designed from the consensus sequence of the p28 gene and the MAP-1 gene in the two variable DNA sequence regions. These sequences are noncoding sequences flanking the p28 gene (Fig. 3). The entire gene of the p28 precursor of the 91HE17, Sapulpa, and Jax strains was successfully amplified by primers p28f159 and p28r1336. However, this primer pair failed to amplify the St. Vincent strain p28

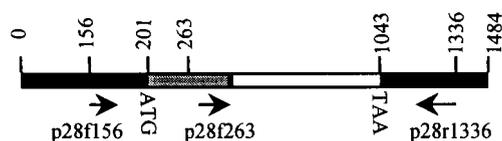


FIG. 3. Diagram of PCR amplification of the p28 gene. The dark boxes represent noncoding DNA sequences bordering the p28 gene. The shaded box and the open box represent the DNA sequence encoding the leader peptide and the sequence encoding the mature p28, respectively. The numbers on the top of the gene indicate the nucleotide positions in base pairs. Arrows indicate the directions of primers. The start point of each primer corresponds to the number at the end of the arrow and on the top of the p28 gene.

gene. To amplify the p28 gene of the St. Vincent strain, we reshuffled the six primers, including p28f159 and p28r1336, which did not amplify the p28 gene of the St. Vincent strain in the previous combinations to form three new primer pairs. The p28 gene of the St. Vincent strain was amplified by the primer pair p28f263 and p28r1336. Primer p28f263 was derived from the consensus DNA sequence of the DNA encoding the leader peptide of the p28 proteins of the Arkansas, 91HE17, Jax, and Sapulpa strains.

**Sequence homology of the 28-kDa protein gene of five *E. chaffeensis* isolates.** Comparison of the sequence data for all five human isolates of *E. chaffeensis* at the nucleotide and amino acid levels revealed divergence of the p28 gene. Nucleotide substitutions were present throughout the genes (Fig. 4). Notable nucleotide deletions or insertions occurred in two regions of the p28 genes, which correspond to the amino acid deletions or insertions in the variable regions 2 and 3 (Fig. 1 and 4). The overall DNA sequence homology of the five human isolates was 84.4 to 100%. There are three variable regions in the amino acid sequences which corresponded to the positions of Arkansas strain amino acids 78 to 85, 145 to 159, and 247 to 261 (Fig. 4). These regions are also highly variable in other genes of the multiple-gene family (20). However, the semivariable region observed at the N-terminal region of the amino acid sequences among the OMP-1 genes (20) was missing in the p28 amino acid sequences (Fig. 1). Based on the amino acid sequence similarities of the p28 proteins, the *E. chaffeensis* isolates were divided into three groups (Fig. 5). The first group included the Sapulpa and St. Vincent strains, which were identical. The second group consisted of the 91HE17 and Jax strains. These two strains were 99.6% homologous. The last group contained only the Arkansas strain. The divergence between the first two groups was 10.5 to 11%. The Arkansas strain differed by 13.1 to 13.3% from the four strains in the first two groups. We further analyzed the homology of the p28 with MAP-1 of *C. ruminantium* (27) and a major surface protein (MSP4) of *Anaplasma marginale* (19). *C. ruminantium* and *A. marginale* are genetically related to *Ehrlichia* spp. (26). The amino acid sequence homology was 57 to 67% between the p28 proteins of the five *E. chaffeensis* strains and MAP-1 and approximately 30% between the p28 proteins of the *E. chaffeensis* strains and MSP4. The amino acid sequence homology of the p28 genes of the *E. chaffeensis* strains was greater than the homology of these proteins with other members of the multiple gene family. The highest level of homology was observed between the p28 gene of 91HE17 and the *omp-1f* gene, 77.1% (Fig. 5). These results verified that the PCR-amplified products were derived from the p28 gene.

**Surface region of the 28-kDa protein.** The surface regions of the p28 of *E. chaffeensis* were analyzed by using the Protein program of Lasergene software (DNASTAR, Inc.). Nine surface-exposed regions containing three to 13 amino acids in the 28-kDa protein of Arkansas strain were predicted by using the Emimi method (12) (Fig. 6). The surface regions were predominantly located in the N-terminal half of the protein. Six of the nine surface regions are located between amino acids 39 and 137. The remaining three surface regions are located in the C-terminal half of the protein between amino acids 204 and 260. Analysis of the proteins by the Jameson-Wolf (13) method, which predicts potential antigenic determinants, indicated that p28 was very likely to be highly antigenic. Analysis of the proteins by the Rothbard-Taylor method (24), which locates potential T-lymphocyte antigenic determinants, predicted T-cell epitopes in the p28. The predicted surface probability, surface regions, antigenic index and T-cell epitopes for the Jax strain were similar to those of Arkansas strain, with

Ark	ATGAATTACA	AAAAAGTTTT	CATAACAAGT	GCATTGATAT	CATTAATATC	TTCTCTACCT	GGAGTATCAT	TTTCCGACCC	AGCAGGTAGT	90
HE	.....	.....	.....	.....	..c.....	.....	.....	.....t.....	.....a.....	
Jax	.....	.....	.....	.....	..c.....	.....	.....	.....t.....	.....a.....	
Sap	.....	.....	.....	.....	..c.....	.....	.....	.....t.....	.....a.....	
Stv	.....	.....	.....	.....	..c.....	.....	.....	.....t.....	.....a.....	
Ark	GGTATTAACG	GTAATTTCTA	CATCAGTGG	AAATACATGC	CAAGTGCTTC	GCATTTTGG	GTATTCTCTG	CTAAGGAAGA	AAGAAATACA	180
HE	.....	.....c.....	.....t.....	.....t.....	.....c.....	.....g.....	.....g..t.....	.....g.....	.....g.....	
Jax	.....	.....c.....	.....t.....	.....t.....	.....c.....	.....g.....	.....g..t.....	.....g.....	.....g.....	
Sap	.....	.....c.....	.....t.....	.....t.....	.....c.....	.....g.....	.....g..t.....	.....g.....	.....g.....	
Stv	.....	.....c.....	.....t.....	.....t.....	.....c.....	.....g.....	.....g..t.....	.....g.....	.....g.....	
Ark	ACAGTTGGAG	TGTTTGGACT	GAAGCAAAAT	TGGGACGGAA	GCGCAATATC	CAACTCCTCC	CCAAACGATG	TATTCACTGT	CTCAAATTAT	270
HE	.....c.....	.....a.....t.....	.....g.....	.....t.....c.....	.....t.....	.....tc..a..a..	.....g..aa..a.....	.....t.....	.....t.....	
Jax	.....c.....	.....a.....t.....	.....g.....	.....t.....c.....	.....t.....	.....tc..a..a..	.....g..aa..a.....	.....t.....	.....t.....	
Sap	.....c.....	.....a.....t.....	.....g.....	.....t.....c.....	.....t.....	.....tc..a..tcat	agt..tac.c.....	.....t.....	.....c..c.....	
Stv	.....c.....	.....a.....t.....	.....g.....	.....t.....c.....	.....t.....	.....tc..a..tcat	agt..tac.c.....	.....t.....	.....c..c.....	
Ark	TCATTTAAAT	ATGAAAACAA	CCCGTTTTTA	GGTTTTGCAG	GAGCTATTGG	TTACTCAATG	GATGGTCCAA	GAATAGAGCT	TGAAGTATCT	360
HE	.....g.....	.....t.....	.....a.....	.....a.....	.....	.....	.....c.....	.....	.....	
Jax	.....g.....	.....t.....	.....a.....	.....a.....	.....	.....	.....c.....	.....	.....	
Sap	.....g.....	.....t.....	.....a.....	.....a.....	.....	.....	.....c.....	.....	.....	
Stv	.....g.....	.....t.....	.....a.....	.....a.....	.....	.....	.....c.....	.....	.....	
Ark	TATGAAACAT	TTGATGTAAA	AAATCAAGGT	AACAATTATA	AGAATGAAGC	ACATAGATAT	TGTGCTCTAT	CCCAT--AA	CTCAGCAGCA	450
HE	.....g.....	.....c.....	.....	.....g.....c.....	.....c.....	.....g.....g.....	.....t.....	.....t...cac.g.....	.....t...g..a..	
Jax	.....g.....	.....c.....	.....	.....g.....c.....	.....c.....	.....g.....g.....	.....t.....	.....t...cac.g.....	.....t...g..a..	
Sap	.....g.....	.....t.....	.....	.....c.....	.....c.....	.....t.....	.....t.....	.....g..caggc.....	.....a..at.....	
Stv	.....g.....	.....t.....	.....	.....c.....	.....c.....	.....t.....	.....t.....	.....g..caggc.....	.....a..at.....	
Ark	GACATGAGTA	GTGCAAGTAA	TAATTTTGTG	TTTCTAAAAA	ATGAAGGATT	ACTTGACATA	TCATTTATGC	TGAACGCATG	CTATGACGTA	540
HE	ag...tcg.....	.....g.....g.....	.....a.....t.....	.....	.....	.....g.....	.....g..c.....	.....	.....t.....	
Jax	ag...tcg.....	.....g.....g.....	.....a.....t.....	.....	.....	.....g.....	.....g..c.....	.....	.....t.....	
Sap	-----	.....ga.....	.....ca.....t.....	.....	.....gc.....	.....	.....	.....	.....t.....	
Stv	-----	.....ga.....	.....ca.....t.....	.....	.....gc.....	.....	.....	.....	.....t.....	
Ark	GTAGCGGAAG	GCATACCTTT	TTCTCCTTAT	ATATGCGCAG	GTATCGGTAC	TGATTTAGTA	TCCATGTTTG	AAGCTACAAA	TCCTAAAATT	630
HE	a..a.t.....	.....a.....	.....c.....	.....t.....t.....	.....g.....	.....	.....	.....	.....	
Jax	a..a.t.....	.....a.....	.....c.....	.....t.....t.....	.....g.....	.....	.....	.....	.....	
Sap	a..a.t.....	.....a.....	.....c.....	.....t.....t.....	.....g.....	.....	.....	.....	.....	
Stv	a..a.t.....	.....a.....	.....c.....	.....t.....t.....	.....g.....	.....	.....	.....	.....	
Ark	TCTTACCAAG	GAAAGTTAGG	TTTAAGCTAC	TCTATAAGCC	CAGAAGCTTC	TGTGTTTATT	GGTGGGCACT	TTCATAAGGT	AATAGGGAAC	720
HE	.....	.....	.....	.....	.....a.....	.....	.....a..t.....	.....g.....a.....	.....g.....a.....	
Jax	.....	.....	.....	.....	.....a.....	.....	.....a..t.....	.....g.....a.....	.....g.....a.....	
Sap	.....	.....	.....	.....	.....a.....	.....c..g..c.....	.....c..a..t.....	.....c.....g.....a.....	.....g.....a.....	
Stv	.....	.....	.....	.....	.....a.....	.....c..g..c.....	.....c..a..t.....	.....c.....g.....a.....	.....g.....a.....	
Ark	GAATTTAGAG	ATATTCCTAC	TATAATACCT	ACTGGATCAA	CACTTGCAGG	AAAAGGAAAC	TACCCTGCAA	TAGTAATACT	GGATGTATGC	810
HE	.....	.....g.....g.....	.....g.....	.....g.....	gt..a.....	-----	.....c.....	.....c.....	.....c.....	
Jax	.....	.....g.....g.....	.....g.....	.....a.....	gt..a.....	-----	.....c.....	.....c.....	.....c.....	
Sap	.....	.....g.....g.....	.....g.....	.....g.....	gt..a.....	-----g..t.....	.....c.....	.....c.....	.....a.....	
Stv	.....	.....g.....g.....	.....g.....	.....g.....	gt..a.....	-----g..t.....	.....c.....	.....c.....	.....a.....	
Ark	CACTTTGGAA	TAGAACTTGG	AGGAAGGTTT	GCTTTCTAA						849
HE	.....t.....	.....g.....	.....	.....						
Jax	.....t.....	.....g.....	.....	.....						
Sap	.....t.....	.....g.....	.....a.....	.....						
Stv	.....t.....	.....g.....	.....a.....	.....						

FIG. 4. Alignment of the p28 gene coding DNA sequences. The complete DNA sequence of the Arkansas strain is presented as a consensus sequence. Differences from the consensus sequence are presented in lowercase. Dots represent the nucleotides of other strains of *E. chaffeensis* identical to those of the Arkansas strain, and dashes indicate gaps which were introduced for optimal alignment of the DNA sequences. The sequence of the St. Vincent strain is incomplete at the 5' end. Ark, Arkansas; HE, 91HE17; Sap, Sapulpa; Stv, St. Vincent.

minor differences in the number of surface regions and T-cell epitopes (Fig. 6).

DISCUSSION

The p28 of *E. chaffeensis* is posttranslationally modified. The previously reported sequence of the p28 gene of *E. chaffeensis* is not complete at the 5' end of the gene, since the first amino acid at the N-terminal portion of the amino acid sequence is aspartic acid. The first amino acid of any given protein is usually methionine or, less frequently, valine. The N-terminal portion of p28 had been determined by amino acid sequencing (20). Therefore, the true N terminus of the p28 was missing, and p28 must have been processed after translation with removal of a segment of amino acids from its N terminus. Our hypothesis was confirmed by finding an additional 75-nucleotide coding sequence on the 5' end of the p28 gene of *E. chaffeensis* (Arkansas strain). We believe the first 25-amino-acid peptide encoded by these 75 nucleotides is a leader signal

sequence of the p28 precursor and is removed following export to the surface of *E. chaffeensis*. This hypothesis is based on the facts that the N-terminal 25-amino-acid peptide has all the characteristics of a leader peptide and the peptide is missing from the N terminus of the mature 28-kDa protein when the amino acids were directly sequenced (20).

p28 is highly variable among *E. chaffeensis* strains. The pleomorphism of p28 of *E. chaffeensis* had been demonstrated previously by using MAbs (6, 7, 31). A group of lower-molecular-mass surface proteins of *E. chaffeensis* including the 29-, 27-, 26-, and 25-kDa proteins have been demonstrated to react with MAb 1A9 (31). These proteins could have represented the degradation products of a single protein or different proteins that shared an epitope. A recent study demonstrates that a multiple gene family of *E. chaffeensis* encodes six outer membrane proteins with molecular masses of approximately 28 kDa. The amino acid sequences of the proteins of the multiple-gene family of the Arkansas strain are 71 to 83% homologous. The protein immunoblotting pattern from the previous reports



FIG. 5. Phylogenetic tree constructed on the basis of the predicted amino acid sequences from the p28 genes of the *E. chaffeensis* strains. The predicted amino acid sequences of *E. chaffeensis* OMP-1F (19) and *C. ruminantium* MAP-1 (25) were included in the analysis to build the root of the tree. The Megalign program of Lasergene software was used to construct the tree. The length of each pair of branches represents the distance between sequence pairs. The scale beneath the tree measures the distance between the sequences.

showed the same patterns of *E. chaffeensis* antigens reacting with MAb 1A9 or with mouse antibodies to the recombinant p28 (20, 31). The mouse antisera to the recombinant p28 protein reacted with two *E. chaffeensis* proteins with molecular masses of 23 and 28 kDa (20). MAb 1A9 reacted with two heat-denatured *E. chaffeensis* proteins with molecular masses of 29 and 27 kDa, and it reacted with four proteins if *E. chaffeensis* was not heat denatured (31). Based on these obser-

vations, we hypothesize that the 28-kDa protein contains the epitope for MAb 1A9. Our results demonstrated that MAb 1A9 recognized the recombinant p28 protein. Therefore, the multiple protein bands recognized by MAb 1A9 are possibly member proteins encoded by the multiple-gene family. The slight differences between the molecular sizes of these proteins reported by different authors (7, 20, 31) were possibly merely differences in calculation. The genes of the p28 family are arranged in a contiguous series in the genome (20). However, it is unknown whether the genes are sequentially expressed or simultaneously expressed, or whether some of them are not expressed at all. At least four of these genes are apparently expressed, because MAb 1A9 recognizes four native proteins of *E. chaffeensis* (31). The antigenic divergence of p28 has been demonstrated among a limited number of *E. chaffeensis* strains (7). MAb 1A9 was originally described as an *E. chaffeensis* species-specific monoclonal antibody (31). Subsequently, when more strains of *E. chaffeensis* became available, it was demonstrated that MAb 1A9 reacted with polypeptides of the Arkansas and 91HE17 strains of different electrophoretic mobilities and affinities and that MAb 1A9 did not react with the Sapulpa strain (7). In this study, we demonstrated that the p28 gene is substantially divergent among the *E. chaffeensis* strains. The

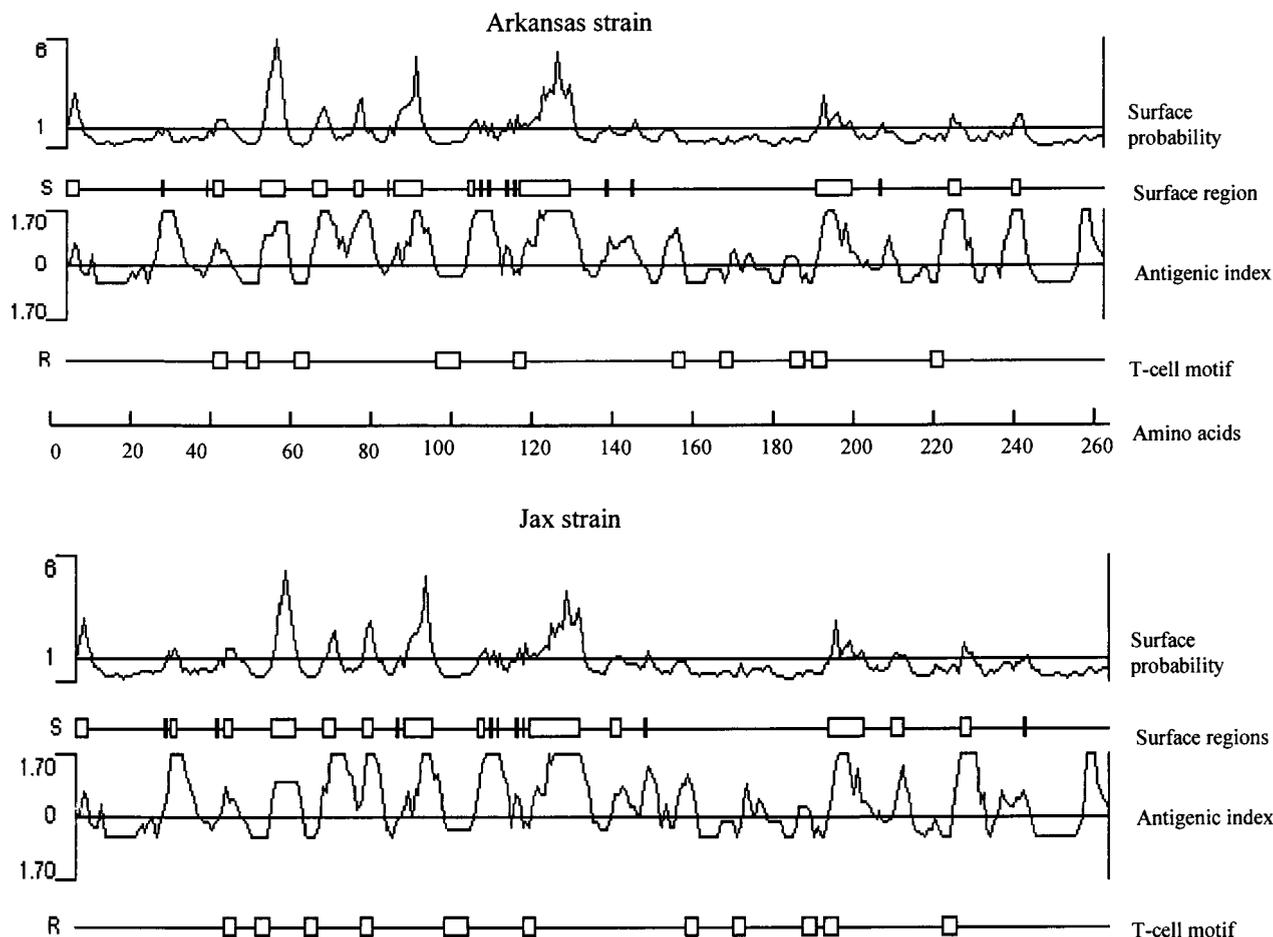


FIG. 6. Comparison of the predicted protein characteristics from the amino acids of the p28 proteins of the Arkansas and Jax strains. Surface probability predicts the surface residues by using a window consisting of a hexapeptide. A surface residue is any residue with  $>2.0 \text{ nm}^2$  of water-accessible surface area. A hexapeptide with a value of greater than 1 was considered as surface region. The antigenic index predicts potential antigenic determinants. The regions with a value above 0 are potential antigenic determinants. The T-cell motif locates the potential T-cell antigenic determinants by using a motif of five amino acids with residue 1 glycine or polar, residue 2 hydrophobic, residue 3 hydrophobic, residue 4 hydrophobic or proline, and residue 5 polar or glycine. The scale indicates the amino acid positions.

Arkansas strain, the prototype strain and the first isolate of *E. chaffeensis*, was most divergent from the other isolates. It is not surprising that the highest level of DNA sequence homology was observed between the Sapulpa and St. Vincent strains, because previous studies demonstrated that they have the same number of repeat units in the 120-kDa protein gene, but 1 repeat unit less than the other strains of *E. chaffeensis* that have been isolated (7, 22, 32). p28 may be a good genetic marker for classification of new isolates of *E. chaffeensis*.

No adequate animal model for *E. chaffeensis* infection is available currently. A mouse model has been used for evaluation of *E. chaffeensis* challenge (20). However, mice are not very susceptible to *E. chaffeensis* infection. No clinical signs are observed, and the organism is rarely reisolated after *E. chaffeensis* infection of mice. PCR is usually required for detection of *E. chaffeensis* DNA in mouse organs or blood to evaluate the response to immune challenge of the organisms. p28 was reported to prevent mouse infection with *E. chaffeensis* in a mouse model, because PCR amplification failed to detect *E. chaffeensis* DNA from the mouse organs (20). The protective immunity conferred by p28 needs to be evaluated in a more suitable animal model. The diversity of the p28 gene and the presence of the multiple copies of heterogeneous genes suggest that the p28 gene is under high selective pressure, possibly because of the host immune system, and that *E. chaffeensis* might use antigenic variation to evade the host immune surveillance. MAP-1 of *C. ruminantium* also shows diversity among strains. The MAP-1 amino acid sequences are divergent by 0.6 to 14% among strains of *C. ruminantium* (23). Persistent infection by *E. chaffeensis* in humans has been observed (11). It will be interesting to investigate whether persistent infection is caused by antigenic variation of the organism. The antigenic diversity of p28 may prevent it from serving as an effective vaccine, since immunization with p28 from one strain of *E. chaffeensis* may not prevent infection with another strain. On other hand, immunization with the conserved antigenic domains may confer cross-protection among strains of *E. chaffeensis*. Whether p28 immunization will protect animals from infection with heterogeneous strains of *E. chaffeensis* needs to be investigated. Similar issues are obvious in terms of the use of p28 as an antigen for serologic diagnosis. The surface regions of p28 are predominantly located on the N-terminal half. These data will facilitate epitope mapping of p28.

**p28 is a common antigen of genus *Ehrlichia*.** The classification of *Ehrlichia* is historical. Some organisms in the genus *Ehrlichia* may not be true *Ehrlichia* species, and some other organisms, such as *C. ruminantium*, which are not classified as *Ehrlichia* currently, are genetically and antigenically closely related to the genus *Ehrlichia*. Based on the DNA sequence homology of the 16S rRNA gene, *Ehrlichia* organisms can be classified into three groups (30). The first group includes *E. canis*, *E. chaffeensis*, *E. muris*, *E. ewingii* and *C. ruminantium*. The host cell tropism of the organisms in this group is for the monocyte, except for *C. ruminantium* and *E. ewingii*, which grow in the endothelial cell and granulocyte, respectively. The second group includes *E. equi*, *E. phagocytophila*, the human granulocytic ehrlichiosis agent, *E. platys*, *E. bovis*, and *A. marginale*. The tropism of the organisms in this group is very diverse. The first three organisms are granulocytotropic. They possibly represent a single species. *E. platys* infects platelets, and *E. bovis* infects monocytes and macrophages. *A. marginale* is an erythrocyte parasite. The p28 protein and the analog proteins had been detected in most species of *Ehrlichia*, except for *E. ewingii* and *E. platys*, in which the protein or its gene has yet to be studied to the best of our knowledge. The closer the organisms are genetically, the stronger the cross-reactions of

p28 are observed. Mouse antiserum to the recombinant p28 protein of *E. chaffeensis* cross-reacted with a 30-kDa protein of *E. canis* (20). Canine anti-*E. canis* serum reacted with the 28- and 30-kDa proteins of *E. chaffeensis*, *E. muris*, and *E. canis* (30). *C. ruminantium* MAP-1 cross-reacted with antibodies to *E. canis*, *E. ovina*, and *E. bovis* (14). *E. canis*, *E. sennetsu*, *E. equi*, and *E. risticii* have antigenic cross-reactive 25-kDa proteins (18). Anti-*E. sennetsu* or anti-*E. risticii* serum cross-reacted with the low-molecular-mass proteins in the range of 20 to 28 kDa of each other (3, 25). Although the low-molecular-mass proteins are not the immunodominant proteins of the human granulocytic ehrlichiosis agent, proteins with a molecular mass of approximately 30 kDa were detected in all strains, and the proteins are pleomorphic in different strains (33). The amino acid sequences deduced from p28 or its analogs from all *Ehrlichia* species evaluated are relatively conserved. The gene of the p28 of *E. chaffeensis* has high homology with the genes of the *C. ruminantium* MAP-1 and even has low homology with the *A. marginale* MSP4. Although not all of the investigations necessary to detect the existence of the p28 protein in *E. ewingii* and *E. platys* have been performed, we believe that the 28-kDa protein may be a characteristic of the genus *Ehrlichia*, since p28 is present in all *Ehrlichia* species characterized.

Because p28 and its homologs are surface-exposed proteins and are conserved among the members of the genus, they might play an important role in the structure of the ehrlichial outer membrane or in the physiology of the *Ehrlichia* organisms. The functions of these proteins have not been characterized. The cysteine content of the p28 is 1.6%, which is similar to the 1.6 and 1.9% cysteine contents of the 90-kDa envelope protein and the major outer membrane protein (MOMP), respectively, of *Chlamydia* (16). MOMP and the 90-kDa protein are involved in the disulfide-bonded cross-linking of the outer membrane of *Chlamydia*, which is responsible for the structural rigidity of the elementary body. Therefore, the 28-kDa protein may be involved in disulfide-bond cross-linking of the outer membrane of *Ehrlichia*, in which peptidoglycan, an important structural component of most bacterial cell walls, has yet to be identified.

#### ACKNOWLEDGMENTS

We thank Josie Ramirez-Kim for assistance in the preparation of the manuscript.

This study was supported by a grant from the National Institute of Allergy and Infectious Diseases (AI31431).

#### REFERENCES

- 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.
- Anderson, B. E., K. G. Sims, J. G. Olson, J. E. Childs, J. F. Piesman, C. M. Happ, G. O. Maupin, and B. J. Johnson. 1993. *Amblyomma americanum*: a potential vector of human ehrlichiosis. *Am. J. Trop. Med. Hyg.* **49**:239–244.
- Brouqui, P., J. S. Dumler, D. Raoult, and D. H. Walker. 1992. Antigenic characterization of ehrlichiae: protein immunoblotting of *Ehrlichia canis*, *Ehrlichia sennetsu*, and *Ehrlichia risticii*. *J. Clin. Microbiol.* **30**:1062–1066.
- Chen, S.-M., L. C. Cullman, and D. H. Walker. 1997. Western immunoblotting analysis of the antibody responses of patients with human monocytotropic ehrlichiosis to different strains of *Ehrlichia chaffeensis* and *Ehrlichia canis*. *Clin. Diagn. Lab. Immunol.* **4**:731–735.
- Chen, S. M., J. S. Dumler, H. M. Feng, and D. H. Walker. 1994. Identification of the antigenic constituents of *Ehrlichia chaffeensis*. *Am. J. Trop. Med. Hyg.* **50**:52–58.
- Chen, S. M., V. L. Popov, H. M. Feng, and D. H. Walker. 1996. Analysis and ultrastructural localization of *Ehrlichia chaffeensis* proteins with monoclonal antibodies. *Am. J. Trop. Med. Hyg.* **54**:405–412.
- Chen, S. M., X. J. Yu, V. L. Popov, E. L. Westerman, F. G. Hamilton, and D. H. Walker. 1997. Genetic and antigenic diversity of *Ehrlichia chaffeensis*: comparative analysis of a novel human strain from Oklahoma and previously isolated strains. *J. Infect. Dis.* **175**:856–863.

8. Dawson, J. E., B. E. Anderson, D. B. Fishbein, J. L. Sanchez, C. S. Goldsmith, K. H. Wilson, and C. W. Duntley. 1991. Isolation and characterization of an *Ehrlichia* sp. from a patient diagnosed with human ehrlichiosis. *J. Clin. Microbiol.* **29**:2741–2745.
9. Dawson, J. E., Y. Rikihisa, S. A. Ewing, and D. B. Fishbein. 1991. Serologic diagnosis of human ehrlichiosis using two *Ehrlichia canis* isolates. *J. Infect. Dis.* **163**:564–567.
10. Dumler, J. S., S.-M. Chen, K. Asanovich, E. Trigiani, V. L. Popov, and D. H. Walker. 1995. Isolation and characterization of a new strain of *Ehrlichia chaffeensis* from a patient with nearly fatal monocytic ehrlichiosis. *J. Clin. Microbiol.* **33**:1704–1711.
11. Dumler, J. S., W. L. Sutker, and D. H. Walker. 1993. Persistent infection with *Ehrlichia chaffeensis*. *Clin. Infect. Dis.* **17**:903–905.
12. Emini, E. A., J. Hughes, D. Perlow, and J. Boger. 1985. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J. Virol.* **55**:836–839.
13. Jameson, B. A., and H. Wolf. 1988. The antigenic index: a novel algorithm for predicting antigenic determinants. *CABIOS* **4**:181–186.
14. Jongejan, F., N. de Vries, J. Nieuwenhuijs, A. H. van Vliet, and L. A. Wassink. 1993. The immunodominant 32-kilodalton protein of *Cowdria ruminantium* is conserved within the genus *Ehrlichia*. *Rev. Elev. Med. Vet. Pays Trop.* **46**:145–152.
15. Klein, P., M. Kanehisa, and C. DeLisi. 1985. The detection and classification of membrane-spanning proteins. *Biochim. Biophys. Acta* **815**:468–476.
16. Longbottom, D., M. Russell, S. M. Dunbar, G. E. Jones, and A. J. Herring. 1998. Molecular cloning and characterization of the genes coding for the highly immunogenic cluster of 90-kilodalton envelope proteins from the *Chlamydia psittaci* subtype that causes abortion in sheep. *Infect. Immun.* **66**:1317–1324.
17. McGeoch, D. J. 1985. On the predictive recognition of signal peptide sequences. *Virus Res.* **3**:271–286.
18. Nyindo, M., I. Kakoma, and R. Hansen. 1989. Antigenic analysis of four species of the genus *Ehrlichia* by use of protein immunoblot. *Am. J. Vet. Res.* **52**:1225–1230.
19. Oberle, S. M., and A. F. Barbet. 1993. Derivation of the complete *msp4* gene sequence of *Anaplasma marginale* without molecular cloning. *Gene* **136**:291–294.
20. Ohashi, N., N. Zhi, Y. Zhang, and Y. Rikihisa. 1998. Immunodominant major outer membrane proteins of *Ehrlichia chaffeensis* are encoded by a polymorphic multigene family. *Infect. Immun.* **66**:132–139.
21. Oliver, D. 1985. Protein secretion in *Escherichia coli*. *Annu. Rev. Microbiol.* **39**:615–648.
22. Paddock, C. D., J. W. Sumner, G. M. Shore, D. C. Bartley, R. C. Elie, J. G. McQuade, C. R. Martin, C. S. Goldsmith, and J. E. Childs. 1997. Isolation and characterization of *Ehrlichia chaffeensis* strains from patients with fatal ehrlichiosis. *J. Clin. Microbiol.* **35**:2496–2502.
- 22a. PSORT. 5 June 1998, posting date. [Online.] <http://psort.nibb.ac.jp>. [16 July 1998, last date accessed.]
23. Reddy, G. R., C. R. Sulsona, R. H. Harrison, S. M. Mahan, M. J. Burrige, and A. F. Barbet. 1996. Sequence heterogeneity of the major antigenic protein 1 genes from *Cowdria ruminantium* isolates from different geographical areas. *Clin. Diagn. Lab. Immunol.* **3**:417–422.
24. Rothbard, J. B., and W. R. Taylor. 1988. A sequence pattern common to T cell epitopes. *EMBO J.* **7**:93–100.
25. Shankarappa, B., S. K. Dutta, and B. L. Mattingly-Napier. 1992. Antigenic and genomic relatedness among *Ehrlichia risticii*, *Ehrlichia sennetsu*, and *Ehrlichia canis*. *Int. J. Syst. Bacteriol.* **42**:127–132.
26. van Vliet, A. H. M., F. Jongejan, and B. A. M. van der Zeijst. 1992. Phylogenetic position of *Cowdria ruminantium* (*Rickettsiales*) determined by analysis of amplified 16S ribosomal DNA sequences. *Int. J. Syst. Bacteriol.* **42**:494–498.
27. van Vliet, A. H. M., F. Jongejan, M. van Kleef, and B. A. M. van der Zeijst. 1994. Molecular cloning, sequencing analysis, and expression of the gene encoding the immunodominant 32-kilodalton protein of *Cowdria ruminantium*. *Infect. Immun.* **62**:1451–1456.
28. von Heijne, G. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* **14**:4683–4690.
29. Weiss, E., J. C. Coolbaugh, and J. C. Williams. 1975. Separation of viable *Rickettsia typhi* from yolk sac and L cell host components by Renografin density gradient centrifugation. *Appl. Microbiol.* **30**:456–463.
30. Wen, B., Y. Rikihisa, J. Mott, P. A. Fuerst, M. Kawahara, and C. Suto. 1995. *Ehrlichia muris* sp. nov., identification on the basis of 16S rRNA base sequences and serological, morphological, and biological characteristics. *Int. J. Syst. Bacteriol.* **45**:250–254.
31. Yu, X.-J., P. Brouqui, J. S. Dumler, and D. Raoult. 1993. Detection of *Ehrlichia chaffeensis* in human tissue by using a species-specific monoclonal antibody. *J. Clin. Microbiol.* **31**:3284–3288.
32. Yu, X. J., P. A. Crocquet-Valdes, and D. H. Walker. 1996. Cloning and sequencing of the gene for a 120-kDa immunodominant protein of *Ehrlichia chaffeensis*. *Gene* **184**:149–154.
33. Zhi, N., Y. Rikihisa, H. Y. Kim, G. P. Wormser, and H. W. Horowitz. 1997. Comparison of major antigenic proteins of six strains of the human granulocytic ehrlichiosis agent by Western immunoblot analysis. *J. Clin. Microbiol.* **35**:2606–2611.