

Characterization of the gene encoding the protective paracrystalline-surface-layer protein of *Rickettsia prowazekii*: Presence of a truncated identical homolog in *Rickettsia typhi*

(protein processing/Rickettsiales/bacterial gene/DNA sequence)

M. CARL*, M. E. DOBSON, W.-M. CHING, AND G. A. DASCH

Rickettsial Diseases Division, Infectious Diseases Department, Naval Medical Research Institute, Bethesda, MD 20814-5055

Communicated by J. H. S. Gear, July 23, 1990

ABSTRACT The DNA sequence of the gene encoding the protective surface protein antigen (SPA) of *Rickettsia prowazekii* has been determined. The open reading frame of 4836 nucleotides with promoter and ribosome-binding site is present on a 10.1-kilobase *EcoRI* fragment. The encoded carboxyl terminus of the 169-kDa protein contains a potential transmembrane region and hydrophilic regions with many lysine and arginine residues potentially accessible to proteolytic cleavage. Because the rickettsia-derived SPA has an estimated molecular mass of only 120 kDa and does not contain several predicted large carboxyl-region CNBr fragments, the SPA product appears to be processed by the rickettsiae. Eight other CNBr fragments were identical in sequence to those predicted from the encoded gene. A complementary 8.7-kilobase *EcoRI* fragment of *Rickettsia typhi* DNA was cloned. This fragment lacked a 1433-base-pair region that included the promoter, ribosome-binding site, and the initial 1162 base pairs of the open reading frame encoding the *R. prowazekii* SPA but had a 3674-base-pair region identical with the remainder of the *R. prowazekii* SPA gene sequence.

Many bacteria possess paracrystalline surface layers (S-layers) as the outermost component of their cell envelopes (1-3). The S-layers of many bacteria are composed of protein or glycoprotein subunits arranged on the cell surface in a regularly repeating hexagonal, tetragonal, or linear pattern. The functions of these S-layers are poorly understood. However, an S-layer may function as a protective barrier against environmental hazards, as a molecular sieve and ion trap, as a promoter for cell adhesion and surface recognition, or as a mold for determining cell shape and envelope rigidity. The S-layer proteins of several pathogenic bacteria appear to be important virulence determinants (3, 4).

Rickettsia prowazekii also contains a tetragonally arranged S-layer (5, 6). S-layers have been isolated and characterized as the species-specific surface protein antigens (SPAs) of *R. prowazekii* and *Rickettsia typhi* (7). The evidence that supports the classification of the SPAs as S-layer proteins has been extensively reviewed (8). The SPAs of both *R. prowazekii* and *R. typhi* have apparent molecular masses of 120 kDa. SPAs are readily released by shaking the typhus rickettsiae in hypotonic solution and are partially purified by filtration and by pelleting contaminants with ultracentrifugation (9). Extraction of the SPA in hypotonic solution results in the loss of the repetitive subunits from the rickettsial outer membrane and in a smooth outer-membrane appearance (8). Although this ease of extraction is relatively unusual for an S-layer because strong chaotropic ions or detergents are often used for their release (10), the large amount and high degree of purity of this high-molecular-mass rickettsial pro-

tein is exactly that expected for an S-layer because these layers form a monomolecular layer around the entire microorganism (11, 12). The SPAs of the typhus rickettsiae also contain a high proportion of acidic and hydrophobic amino acids and are low in sulfur-containing amino acids, as are other described S-layer proteins (1-3, 8). In addition, SPAs of the typhus rickettsiae, like other S-layer proteins, are rich in β -sheet structure but have little α -helical structure (8).

The SPAs of the typhus rickettsiae are highly immunogenic in humans and animals (13). In addition, the SPAs have been shown to stimulate different classes of human lymphocytes, including lymphokine-activated killers (14), suppressor T cells (15), and γ -interferon-producing T helper cells (16) that participate in the immunological defense against these organisms. More important however, purified preparations of SPA from the typhus rickettsiae have been shown to be effective protective antigens in an animal model. Guinea pigs immunized with different doses of the SPA appear to be protected against challenge with live organisms as compared with nonimmunized controls (13).

The gene *spaP* encoding the SPA of *R. prowazekii* has been previously cloned (M.E.D., M.C., and G.A.D., unpublished work). We describe here the complete sequence of the 4836 nucleotide (nt) open reading frame (ORF) that encodes a 169-kDa SPA protein as well as its flanking regions.[†] We provide evidence that the 120-kDa SPA obtained from rickettsiae is derived by truncating the carboxyl end of the encoded gene product. We have also cloned and sequenced a fragment of *R. typhi* DNA that contains an exact 3674-base-pair (bp) copy (*spaPt*) of the 3' end of *spaP*, but lacks the promoter, ribosome-binding site, and 1162 bp of the 5' end of the ORF of *spaP*.

MATERIALS AND METHODS

Construction of Recombinant Bacteriophages, Plasmids, and DNA Sequencing. The construction and identification of recombinant λ gt11 and recombinant λ gtWES bacteriophages containing 3.7-kilobase (kb) and 10.1-kb *R. prowazekii* Breinl-derived DNA fragments, respectively, have been described (M.E.D., M.C., and G.A.D., unpublished work). When plated onto the *Escherichia coli* host strain ED8654 (*supE supF hsdR metB lacY gal trpR*), recombinant bacteriophages containing all or part of the SPA gene were identified by screening plaques with monoclonal antibodies (mAbs) specific for the SPA of *R. prowazekii* (17). Similarly, whole-cell DNA was prepared from purified *R. typhi* (Wilmington strain), digested with a mixture of *Alu I*, *Hae III*, and

Abbreviations: S-layers, paracrystalline surface layers; PCR, polymerase chain reaction; SPA, surface protein antigen; nt, nucleotide(s); mAb, monoclonal antibody; ORF, open reading frame.

*To whom reprint requests should be addressed.

[†]The sequence reported in this paper has been deposited in the GenBank data base (accession no. M37647).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Rsa I, and ligated into *λgt11* using *Eco*RI linkers (18). The library was titered on *E. coli hsdR* and screened for antigen expression with *R. typhi*-specific mAbs (17). A total *Eco*RI digest of *R. typhi*-derived DNA was ligated into *λgtWES* (19). Recombinant phage were screened using mAbs as above.

Recombinant bacteriophage DNA was prepared (20), and purified DNA fragments were subcloned into pUC8 and pUC19 plasmids. Recombinant plasmids were sequenced directly from plasmid DNA by the dideoxynucleotide chain-termination method (21) with Sequenase (United States Biochemical) as the polymerase. Sequencing was initiated on both strands with M13 primers (Promega and United States Biochemical). Subsequently, sequencing was continued on both strands by using custom synthetic oligonucleotides as primers.

Restriction Map and Southern Blot Analysis. DNA from the desired plasmids were digested with the following endonucleases: *Bgl* II, *Eco*RI, *Hind*III, and *Pst* I, according to the supplier's recommendations (Bethesda Research Laboratories). Plasmid DNA digested with single or double enzymes was then electrophoresed in 0.8% agarose gel.

For Southern blot analysis, DNA was digested with the appropriate restriction enzymes, electrophoresed in 0.8% agarose gels, and transferred to nitrocellulose (Schleicher & Schuell). Hybridizations were carried out largely as described by Southern (22). Probes were derived from recombinant plasmid DNA labeled with deoxyadenosine 5-[γ -³⁵S]thio]triphosphate by nick-translation (23). Alternatively, probes were generated using polymerase chain reaction (PCR)-derived products labeled with digoxigenin (24), and detection was accomplished by using the stringent conditions recommended by the manufacturer (Boehringer Mannheim).

Preparation and Sequencing of CNBr Fragments from SPA. Cleavage of purified SPA protein at methionine residues with CNBr was done according to the method of Gross and Witkop (25). Fragments were then separated by SDS/PAGE (26), electroblotted onto polyvinyl difluoride paper (Millipore), and stained directly with Coomassie blue. Amino acid sequence analysis was performed on a model 477A sequencer with model 120A automated on-line phenylthiohydantoin analyzer (Applied Biosystems).

PCR. DNA was amplified using the PCR, as described (27). Genomic DNA derived from *R. typhi* and *R. prowazekii*, as well as recombinant plasmids containing inserts derived from these organisms, served as templates. Oligonucleotide primers were selected from those used for sequence determinations. Amplified DNA was detected by agarose gel electrophoresis and ethidium bromide staining.

RESULTS

Recombinant plasmids pMD306 and pMDL7 were created by subcloning 3.7-kb and 10.1-kb inserts (originally derived from *R. prowazekii* Breinl), from recombinant *λgt11* and *λgtWES* bacteriophages, respectively (M.E.D., M.C., and G.A.D., unpublished work). These phages had been detected by screening libraries with SPA-specific mAbs (17). Recombinant phage *λgt11*-RTW9 contains a 986-bp insert derived from *R. typhi* that is recognized by mAbs specific for the SPA. pMAC1 was created by subcloning this 986-bp insert into pUC19. pMD306 was used as a probe to identify a homologous 8.7-kb *Eco*RI fragment in *R. typhi* genomic DNA. A recombinant *λgtWES* clone containing this 8.7-kb fragment was identified with SPA-specific mAbs. pMAC2 is a recombinant pUC19 plasmid that contains this 8.7-kb *Eco*RI fragment. The restriction maps of all four of these plasmid inserts are shown in Fig. 1. This data suggested that the pMD306 insert is identical with a region contained within pMDL7 and that the pMAC1 insert is identical with a region contained within pMAC2. Maps for plasmids pMDL7 (de-

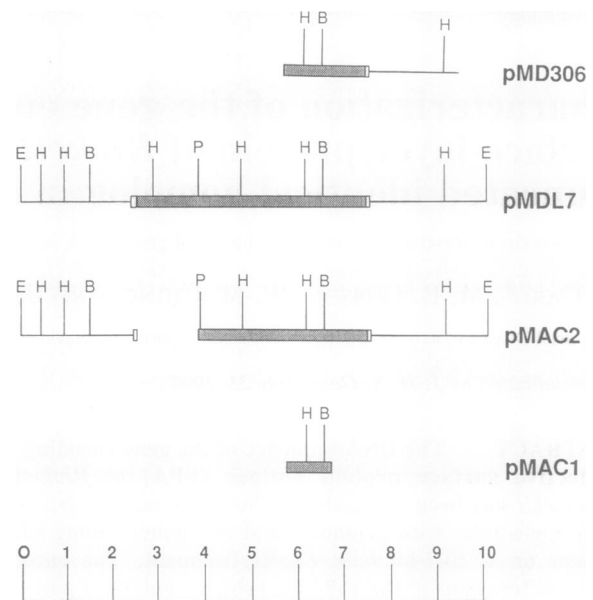


FIG. 1. Restriction maps of the rickettsial inserts of four recombinant pUC plasmids. The inserts from pMD306 and pMDL7 are derived from *R. prowazekii*, and the inserts from pMAC1 and pMAC2 are derived from *R. typhi*. Thickened lines represent sequenced regions; a shaded area represents a continuous ORF. Restriction endonucleases used include *Bgl* II (B), *Eco*RI (E), *Hind*III (H), and *Pst* I (P). Scale is in kb.

rived from *R. prowazekii*) and pMAC2 (derived from *R. typhi*) were identical except for a 1.4-kb fragment of DNA present in pMDL7 but not in pMAC2.

The complete nucleotide sequences of plasmids pMD306 and pMAC1 were determined for both strands of DNA. The entire sequences for both of these plasmids were contained within pMDL7 and pMAC2. Further sequencing of both strands of pMDL7 revealed an ORF of 4836 nt between the ATG triplet at positions 394–396 and the TAA stop codon at positions 5230–5232 (Fig. 2). Presumptive ribosome-binding site and –10 and –35 regions are underlined. A large inverted repeat forming a stem loop structure consisting of a 14-bp stem and an 11-bp loop is present downstream of the translational stop codons (nt 5271–5309) and might function as a transcriptional termination signal. Amino-terminal amino acid sequences of eight CNBr fragments of purified *R. prowazekii* SPA were found within the ORF and are underlined in Fig. 2. The DNA sequence predicts the existence of an additional seven CNBr fragments at the carboxyl end of the SPA with molecular masses of 0.87–8.14 kDa. However, despite the fact that all other major CNBr fragments throughout the rest of the molecule were identified, no fragments corresponding to these seven predicted carboxyl terminal fragments could be found. A 5.5-kDa CNBr fragment beginning at amino acid 1255 (corresponding to nt 4216–4218) was closest to the putative SPA carboxyl terminus.

The sequence of pMAC2 contained an ORF of 3674 bp identical to pMDL7 from nt 1556–5229 (Fig. 2). A region of DNA just 5' to this homologous ORF is shown in Fig. 3 (nt 1–87) and is identical to the nucleotide sequence of pMDL7 from nt 36–122. pMAC2, therefore, lacks a 1433-bp segment of DNA (nt 123–1555), present in pMDL7. The DNA sequence of pMAC2 found 3' to the ORF is also identical to that found in pMDL7 through nt 5322 (Fig. 2) and, therefore, includes an inverted repeat downstream of the translational stop codons. No presumptive ribosome-binding site or –10 and –35 regions could be discerned.

To confirm that this homologous sequence in *spaPt* was not a laboratory artifact, we selected an oligonucleotide primer (primer 1, GGTGTGATAGTATTGTAATC) based on

1 TTAAGTACGTAATGACATATGTAAAMTTGATATACAAACTCATCGCAATAACT
 58 GATTATATTTCATAACAACCAATGTGATAAATAAACAATTATTAACAACTTTATAAAAACTACTGT
 128 GATCGTCTGTAATTATGTACAATATATAGTACACTAGCOOCTAGTGTGAAACACTTGTAAACAAATATT
 198 AGGTTATTTCCCTATCAAGTGTGGGATATCTGACTGCTATTGATTAAATTTGTTTAACTAGATACT
 268 AAATTTTAACTTAAATATAGGAAAAAATATGGCTCAAAAACCTTATTTCTAAAAAATAATTTCCGCG
 -35 -10 S/D
 337 AGGATGGTAAGTACTGCTTCCAGGCTACTATAGTACGCTGTTTCCTGGTGTAGCA ATG GGT GCT GCT
 M G A A
 406 ATG CAA TAT AAT AGG ACA ACA AAT GCA GCA GCT ACA ACC TTT GAT GGT ATA GCG TTT GAT
 M Q Y N R T T N A A A A A T T E D G I G E D
 466 CAA GCT GCT GGT AAT ATT OCT GTC GCT CCA AAT TCA GTT ATT ACT GCT AAT GCT AAT
 O A A G A A N I P V A P N S V I T A N A N
 526 AAT OCT AAT ACT TTT AAT ACT CCA AAC GGT CAT TTA AAT AGT TTA TTT TTG GAT AGC GCA
 N P I T F N T P N G H L N S L F L D A T T
 586 AAT GAT TTA GCA GTA ACA ATT AAT GAG GAT ACT ACC TTA GGA TTT ATA ACA AAT ATT GCT
 N D L A V T I N E D T T L G F I T N I A
 646 CAG CAG GCT AAG TTC TTT AAT TTT ACT GCT GCT GGT AAA ATT CTT AAC ATA ACA GGG
 O O A K F F N F T V A A G K I L N I T G
 706 CAG GGT AAT ACT GTT CAA GAA GCT TCT AAT ACA ATA AAT GCT CAA AAT GCT CTT ACA AAA
 O G I T V O E A S A T I N A Q N A L T K
 766 GTG CAT GGT GGC GCT GCT ATT AAC GCT AAT GAT CTT AGC GGG CTA GGA TCA ATA ACC TTT
 V H G G A A I N A N D L S G L G S I T F
 826 GCT GTG TGT CCT TCT GTA TTA GAA TTT AAT TTA ATA AAT CCT ATC AAC TCA AGA AGC TCC
 A V C P S V L E F N L I N P I N S R S S
 886 TCT TAT CAC TTG GCT AAT TCT AAA ATA GTT AAT GGT GGT AAT GGG ATA TTA AAT ATT
 S Y H L V S N S K I V N G G N G I L N I
 946 ACT AAT GGA TTT ATT CAG GTT TCA GAT AAC ACT TTT GCT GGT ATT AAG ACC AAT ATT ATC
 T N G F I Q O V S D N T F A G I K T I N I
 1006 GAT GAT TGT CAA GGT TTA AAT TCT ACT OCT GAT GOC GCT AAT ACT TTA AAT TTA
 D C O G L M F N S T P D A A N T I N L
 1066 CAA GCA GGT GGT AAT CTT AAT GGA ATA GAC GGT ACT GGT AAA TTA GTA TTA
 O A G G N T I N F N G I D G T G K I V L
 1126 GTC AGT AAG AAT GGT GCT ACC GAA TTT AAT GTT ACA GCA ACT TTA GGT GGT AAT CTA
 V S K N G A T A E F N V T G T L G N N L
 1186 AAA GGT AAT ATT GAA TTT AAC ACT GCA GCA GTA GCT GGT AAA CTT ATC TCT CTT GGA GGT
 K G I E L N T A A V A G K L I S L G G
 1246 GCT GCT AAT GCA GTA ATA GGT ACA GAT AAT GGA GCA GGT AGA GCT GCA GGA TTT ATT GTT
 A N A V A T T D N G A G R A A G F I V
 1306 AGT GGT AAT GGT AAT GCA GCA ACA ATT TCT GCA CAA GGT TAT GCT AAA AAG ATG GTG
 S V D N G N A A T I S G O V Y A K N M Y
 1366 ATA CAA AGT GCT AAT GCA GGT GGA CAA GTC ACT TTT GAA CAC ATA GTT GAT GGT GGT TTA
 O S A N A G G O V T F E H I V D V G I
 1426 GGC GGT ACC ACC AAC TTT AAA ACT GCA GAT TCT AAA GTT ATA ATA GAA AAC TCA AAC
 G G T I N F K T A D S K V I I T E N S N
 1486 TTT GGT TCT ACT AAT TTT GGT AAT CTT GCA ACA CAG AAT GTA GCT OCT GAT ACT AAG ATT
 F G S T N F N G N L D T O I V P D T K I
 1546 CTT AAA GGT AAC TTC ATA GGT GAT GTA AAA AAT AAC GGT AAT ACT GCA GGT GGT ATT ACT
 L K N F I G D V K N N G N T A G V I T
 1606 TTT AAT GCT AAT GGT GCT TTA GTA AGT GCT AGT ACT GAT CCA AAT ATT GCA GTA ACA AAT
 F N A N G A L V S A S T D P N I A V T N
 1666 ATT AAT GCA ATT GAA CAA GAA GGG GCG GGT GTA GAA TTA TCA GAA ATA CAT ATT GCA
 I N A I E A E G A G V V E L S G I H I A
 1726 GAA TTA CGT TTA GGG AAT GGT GGT TTT AAA CTT GCT GAT GGC ACA GTA ATT AAT
 E L E L G N G G S I F K L A D G T V I N
 1786 GGT CCA GTT AAC CAA AAT GCT CTT ATG AAT AAT AAT GCT CTT GCA GCT GGT TCT ATT CAG
 G P V N O N S L M N N S L A A G S I O
 1846 TTA GAT GGG AGT GCT ATA AAT ACC GGT GAT ATA GGT AAC GGT GGT ATT AAT GCT GCG TTA
 L D G S A I I T G D I G N G V N A A L
 1906 CAA CAC ATT ACT TTA GCT AAC GAT GCT TCA AAA ATA TTA GCA CTC GAT GGC GCA AAT ATT
 O H I T L A N D A S K I L A L D G A N I
 1966 ATC GGG GCT AAT GGT GGT GGT GAT TTT CAA GCT AAC GGT GGT AAT AAT AAT AAT
 I G A N V G G A I H F O A N G G T I K L
 2026 ACA AAT ACT CAA AAT AAT ATT GTA GTT AAT TTT GAT TTA GAT ATA ACT ACT GAT AAA ACA
 T N T O N N I V V N F D L D I T D K I
 2086 GGT GGT GAT GCA AGT AGT TTA ACA AAT AAT CAA ACT TTA ACT AAT AAT GGT AGT ATC
 G V V D A S S L T N N O T L T I N G S I
 2146 GGT ACT GGT GAT GCT AAT ACT AAA ACA CTT GCA CAA TTA AAC ATC GCG TCA AAT AAA ACA
 G T V V A N T K T L A O L N I G S S K T
 2206 ATA TTA AAT GCT GGC GAT GTC GCT ATT AAC GAG TTA GTT ATA GAA AAT AAT GCT TCA GTA
 I L N A G D V A I N E L V I N G S V
 2286 CAA CTT AAT CAC AAT ACT TAC TTA ATA ACA AAA ACT ATC AAT GCT GCA AAC CAA GGT CAA
 O L N H N T Y L I T K T I N A A N O G O
 2326 ATA ATC GTT GCG GCT GAT OCT CTT AAT ACT AAT ACT ACT CTT GCT GAT GGT ACA AAT TTA
 I V A A D P L N T N T L A D G C T N L
 2386 GGT AGT GCA GAA AAT CCA CTT TCT ACT ATT CAT TTT GGC ACT AAA GCT GCT AAT GCT GAC
 G S A E N P L S T I H F A T K A A N D
 2446 TCT ATA TTA AAT GTA GGT AAA GGA GTA AAT TTA TAT GCT AAT AAT ATT ACT ACT AAC GAT
 S L N V G K G V N L Y A N N I T T N D
 2506 GCT AAT GTA GGT TCT TTA CAC TTT AGS TCT GGT GGT ACA AGT ATA GTA AGT GGT ACA GTT
 A N V G S L H F R S G G T S I V S G T V
 2566 GGT GGA CAG CAA GGT CAA GGT CTT AAT AAT TTA ATA TTA GAT AAT GGT ACT ACT GTT AAG
 G G O O G H K L N N L I L D N G T V K
 2626 TTT TTA GGT GAT ACA ACA TTT AAT GGT GGT ACT AAA ATT GAA GGT AAA TCC ACT TTG CAA
 F L G D T T F N G G T K I E G K S I L O

2686 ATT AGC AAT AAT TAT ACT ACT GAT CAT GTT GAA TCT GCT GAT AAT ACT GGT ACA TTA GAA
 I S N N Y T T D H V E S A D N T G T L E
 2746 TTT GTT AAC ACT GAT OCT ATA ACC GTA ACA TTA AAT AAA CAA GGT GCT TAT TTT GGT GTT
 F V N T D P I T V T L N K O G A Y F G V
 2806 TTA AAA CAA GTA ATT ATT TCT GGT CCA GGT AAC ATA GTA TTT AAT GAG ATA GGT AAT GTA
 L K O V I I S G P G N I V F N E I G N V
 2866 GGA ATT GTA CAT GGT ATA GCA GCT AAT TCA ATT TCT TTT GAA AAT GCA AGT TTA GGT ACA
 G I V H G I A A N S I S F E N A S L G T
 2926 TCT TTA TTA OCT AGT GGT ACT CCA TTA GAT GTT TTA ACA ATT AAA AGT ACC GTA GGT
 S L F L P S G T P L D V L T I K S T V G
 2986 AAT GCA ACA GTA GAT AAT TTT AAT GCT OCT ATT GTA GTT GTA TCA GGT ATT GAT AGT ATG
 N G T V D N F N A P I V V S G I D S M
 3046 ATC AAT AAC GGT CAA ATC ATC GGT GAT AAA AAG AAT ATT ATA GCT CTA TCG CTT GGA AGT
 I N N G O I I G D K A N N I A I S T G S
 3106 GAT AAC AGT ATT ACT GGT AAT GCT AAT ACA TTA TAT TCA GGT ATC AGA ACT ACA AAA AAT
 D N S I T V N A N T L Y S G I R T T K N
 3166 AAT CAA GGT ACT GTG ACA CTT AGT GGT GGT ATG OCT AAT AAT OCT GGT ACA ATT TAT GGT
 N O G T V T L S G G M P N N P G T I Y G
 3226 TTA GGT TTA GAG AAT GGT AGT CCA AAG TTA AAA CAA GTG ACA TTT ACT ACA GAT TAT AAC
 G L E N G S P K L K O C V T T F T D Y N
 3286 AAC TTA GGT AGT ATT GCA AAT AAT GTA ACA AAT AAT GAT GAT ACT CTT ACT ACA
 I N A I E A E G A G V V E L S G I H I A
 3346 GGA GGT ATA GCA GGG ACA GAT TTT GAG GCT AAA ATT ACT CTT GCA AGT GTT AAC GGT AAC
 G I A G T D F D A K I V T L G S V N G N
 3406 GCT AAC GTA AGG TTT GAT GAT ACT TTT TCT GAT OCT GAT AAT ATG ATT GTT GCT ACT
 A N V R F V D S T F S D P R S M L V A I
 3466 CAA GCT AAT AAG GGT ACT GTA ACT TAT TTA GGT AAT GCA TTA GGT AAT GGT GGT AGT
 Q A N K G T V T Y I G N A L V S N I G S
 3526 TTA GAT ACT OCT GTA GCT TCT GTT AGA TTA ACA GGT AAT GAT AGT GGS GCA GGA TTA CAA
 L D T P V A S V R F T G N D S G A L O
 3586 GGC AAT AAT TAT TCA CAA AAT ATA GAT TTT GGT ACT TAT AAT TCA ACT ATT CTA AAT TCT
 G N I Y S O N I D F G T Y N L T I L N S
 3646 AAT GTC ATT TTA GGT GGT GGT ACT ACT GAT AAT GGT GAA ATC GAT CTT CTG ACA AAT
 N V I L G G G T T A I N G E I D L T N
 3706 AAT TTA ATA TTT GCA AAT GGT ACT TCA ACA TGG GGT GAT AAT ACT TCT ATT AGT ACA AGG
 N L I F A N G T S T W G D N T S I S T
 3766 TTA AAT GTA TCA AGC GGT AAT ATA GGT CAA GTA GTC ATT GOC GAA GAT GCT CAA GTT AAC
 L N V S S G N I G O V V I A E D A O V N
 3826 GCA ACA ACT ACA GGA ACT ACA ACC ATT AAA ATA CAA GAT AAT GCT AAT GCA AAT TTC AGT
 A T T T G T T T I K I O D N A N A N F S
 3886 GGC ACA CAA GCT TAT ACT TTA ATT CAA GGT GCT GCT AAT GCT AAT GGT ACT TTA GGA GCT
 G T O A Y T L I O G G A R F N G T L G A
 3946 CCT AAC TTT GCT GTA ACA GGA AGT AAT ATT TTC GTA AAA TAT GAA CTA ATA GCT GAT TCT
 P N F A V T G S N I F V K Y E L I R D S
 4006 AAC CAG GAT TAT GTA TTA ACA CGT ACT AAC GAT GTA TTA AAC GTA GTT ACA ACA GCT GTT
 N O D Y V L T R T N D V L N V V T T A V
 4066 GGA AAT AGT GCA ATT GCA AAT GCA OCT GGT GTA AGT CAG AAC ATT TCT AGA TCC TTA GAA
 G N S A I A N A P G V S O N I S R C L E
 4126 TCA ACA AAT ACA GCA GCT TAT AAT AAT ATG CTT TTA GCT AAA GAT OCT TCT GAT GTT GCA
 S T N T A A Y N N M L L A K D P S D V A
 4186 ACA TTT GTA GGA GCT ATT GCT ACA GAT ACA AGT GCG GCT AAT ACT ACA GTA AAG TTA AAT
 T F V G A I A T D T S A V T T V N L N
 4246 GAT ACA CAA AAA ACT CAA GAT CTA CTT AGT AAT AGG CTA GGT ACA CTT AGA TAT CTA AGT
 D T O K T O D L L S R N L G T L Y R L S
 4306 AAT GCT GAA ACT TCT GAT GTT GCT GCA TCT GCA ACA GGT GCA GTG TCT TCA GGT GAT GAA
 N A E T S D V A G S A T G A V S S G D E
 4366 GCG GAA GTA TCT TAT GGT GTA TGG WAK AAA OCT TCT TAC AAT ATT GCA GAA CAA GAA AAA
 A E V S Y G V W A K P F Y N I A E O D K
 4426 AAA GGT GGT ATA GGT TAT AAA GCA AACT ACT GGG GTT GTA GTT GGT TTA GAT ACT
 K G I Y A G Y K A K T T G V V V G L D T
 4486 CTC GCT AGC GAT AAC CTA ATG ATT GGG GCA GCT ATT GGG ATC ACT AAA GAT GAT ATA AAA
 L A S D N L M I G A I T I G I T K T D I K
 4546 CAC CAA GAT TAT AAG AAA GGT GAT AAA ACT GAT AAT AAT GGT TTA TCA TCT CTA TAT
 H O D Y K G D K T D I N G L S F S L Y
 4606 GGT TCC CAA CAG CTT GTT AAG AAT TTT TTT GCT CAA GGT AAT TCA ATC TTT ACT TTA AAC
 G S O Q L V K N F F A O G N S I F T L N
 4666 AAA GTC AAA AGT AAA AGT CAG OCT TAC TTC GAG TCT AAT GGT AAG ATG AGC AAG CAA
 K V K S K S O R Y F F E S N G K M S K O
 4726 ATT GCT GCT GGT AAT TAC GAT AAC ATG ACA TTT GGT GGT AAT TTA ATA TTT GGT TAT GAT
 I A A G N Y D N M T F G G N L I F G Y D
 4786 TAT AAT GCA ATG CCA AAT GTA TTA GTA ACT CCA ATG GCA GGA CTT AGC TAC TTA GAA TCT
 Y N A M P N V L V T P M A G L S L Y K S
 4846 TCT AAT GAA AAT TAT AAA GAA ACC GGT ACA ACA GTT GCA AAT AAG GCG AAT AAT AGC AAA
 S N E N Y K E T E T G T V A N K R I N S K
 4906 TTT AGT GAT AGA GTC GAT TTA ATA GTA GGG GCT AAA GTA GCT GGT AGT ACT GTG AAT ATA
 F S D R V D L I V G A K V A G S T V N I
 4966 ACT GAT ATT GTG ATA TAT CCG GAA ATT CAT TCT TTT GTG GTG CAC AAA GTA AAT GGT AAA
 T D I V I Y P E I H S F V V H K V N G K
 5026 TTA TCT AAC TCT CAG TCT ATG TTA GAT GGA CAA ACT GCT CCA TTA ATG AAG CAA OCT GAT
 L S N S O S M L D G O T A P F I S O P D
 5086 ASA ACT GCT AAA AGC TCT TAT AAT ATA GGC TTA AGT GCA AAC ATA AAA TCT GAT GCT AAG
 R T A K T S Y N I G L S A N I K S D A K
 5146 ATG GAG TAT GGT ATC GGT TAT GAT TTT AAT TCT GCA AGT AAA TAT ACT GCA CAT CAA GGT
 M E Y G I G Y D F N S A S K Y T A H O G
 5206 ACT TTA AAA GTA CGT GTA AAC TTC TAA TTA TTT GGT GAT TTT AAG TTT TAA AACTTGATTA
 T L K V R V N F
 5271 AGAAAAAGCCCACTTTGAAAAATGGGCTTTTTTCTAGTATGTAATAA

FIG. 2. Nucleotide sequence and deduced amino acid sequence (in one-letter code) of the *spaP* gene. Potential -35 and -10 regions, a potential ribosome-binding site, and a terminal 14-bp inverted repeat are underlined. Deduced amino acid sequences corresponding to sequenced CNBr fragments are also underlined.

additional partial sequence of a region 5' to the 1.4-kb deletion in *R. typhi* (not shown in Fig. 2) and used it in PCR reactions paired with each of two primers 3' to the deletion (primer 2, CAGTAGTTATATCTAAATCA and primer 3, ABAGCCACCATTCCCTAAAC, reverse complements of bp 2058-2077 and 1733-1752, respectively, Fig. 2). The DNA templates used in these PCR reactions were genomic DNA isolated from *R. typhi* or from *R. prowazekii*. DNA amplified from these two different templates differed by ≈1.4 kb for

each of the two primer pairs examined, thus confirming the origin of the DNA inserts in pMDL7 and pMAC2 from *R.*

1 CA ACAC TCA TCA TCG CAA TAA CTG ATT ATT ATT TCA TAA CAC CAA TGT GAT AAA AAT ATA ACA
 N S S S O - L I I I S - H O C D K N I T
 63 TTT ATT AAA ACT TTA TAA AAA AAC TAC TTT ATA GGT GAT GTA AAA
 F I K T L - K N Y F I G D V K

FIG. 3. Nucleotide sequence adjacent to the 5' end of *spaPt* in *R. typhi*. The underlined nucleotides correspond to nt 1556-1575 in the nucleotide sequence of *spaP* shown in Fig. 2. Amino acids are shown in one-letter code.

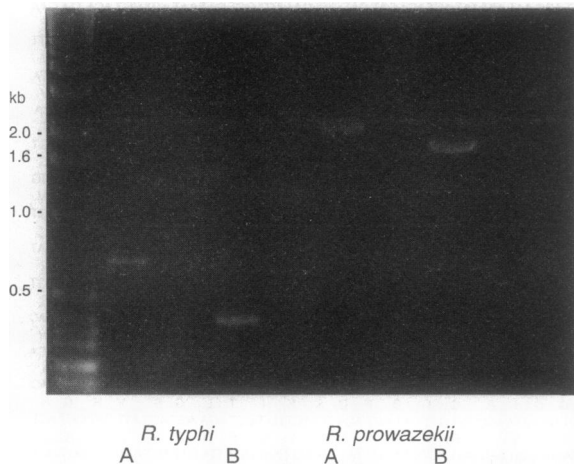


FIG. 4. Polymerase chain reaction-amplified DNA using *R. typhi* or *R. prowazekii* genomic DNA as the template. Primer pair A (primers 1 and 2) was used to amplify a 700-bp fragment from *R. typhi* and a 2.1-kb pair fragment from *R. prowazekii*. Primer pair B (primers 1 and 3) was used to amplify a 400-bp fragment from *R. typhi* and a 1.8 kb-fragment from *R. prowazekii*.

prowazekii and *R. typhi*, respectively (Fig. 4). Therefore, the observed 1.4-kb deletion seen is present in *R. typhi* and did not occur as a deletional event within the transformed *E. coli* host. Although large enough to encode a protein of 120 kDa, we are certain that *spaPt* does not encode the *R. typhi* SPA because some sequenced CNBr fragments of *R. typhi*-derived SPA are not present in the deduced amino acid sequence of *spaPt* (W.-M.C., unpublished work). Therefore, we hypothesized that there exists another gene (*spaT*) highly homologous to *spaPt* that encodes the *R. typhi* SPA. To prove this, Southern blot analysis of DNA derived from *R. typhi* was done. A digoxigenin-labeled DNA probe (generated using PCR) from the 5' end of the *spaP* gene (nt 394–1641) hybridized to a single 3.4-kb *Pst* I fragment (presumably containing part of *spaT*), whereas a second digoxigenin-labeled DNA probe (nt 3238–4437) hybridized to the same 3.4-kb *Pst* I fragment and also to a 12-kb fragment (presumably containing part of *spaPt*, Fig. 5). Because the second

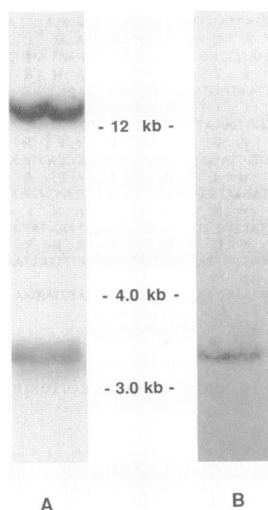


FIG. 5. Genomic DNA purified from *R. typhi* was digested with *Pst* I, electrophoresed in 0.8% agarose gel, and transferred to nitrocellulose before Southern blot analysis. (A) Probe used for hybridization was a PCR-generated product, and its nucleotide sequence corresponded to nt 3238–4437 of *spaP* (Fig. 2). (B) Probe used for hybridization was generated using the PCR, and its nucleotide sequence corresponded to nt 394–1641 of *spaP* (Fig. 2).

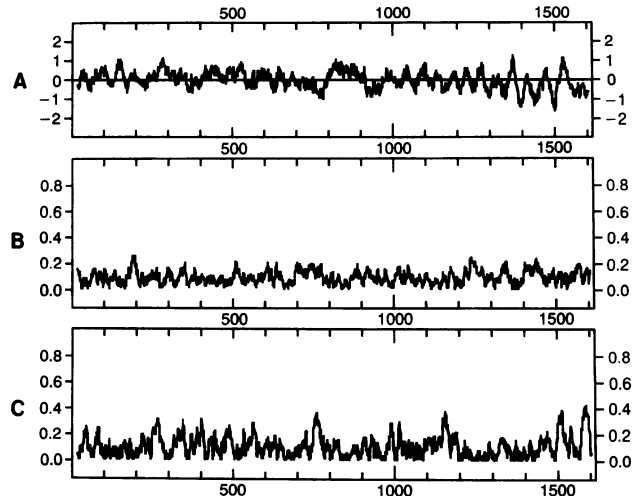


FIG. 6. Theoretical structure analysis of the *R. prowazekii* SPA. Hydrophathy (A), α -helical structure (B), and β -sheet structure (C) are based on the deduced amino acid sequence.

probe consisted of DNA the sequence of which was located entirely within a single *Pst* I fragment derived from *R. prowazekii*, this data strongly suggests the presence of two highly homologous genes, only one of which (*spaT*) contains sequence homology to the 5' end of *spaP*.

The ORF of pMDL7 encoded a protein with a calculated molecular mass of 169,874 and with a theoretical pI of 5.81. Codon usage in the *spaP* gene is similar to that described for the *R. prowazekii* citrate synthetase gene (28). The hydrophobicity profile, α -helical structure, and β -sheet structure (29) are plotted in Fig. 6 (DNA Strider, Institut de Recherche Fondamentale). Although most of the protein appears to be hydrophobic, the carboxyl portion of the molecule appears more hydrophilic.

DISCUSSION

We describe the cloning and sequencing of the *spaP* gene. Although *R. prowazekii*-derived SPA has an estimated molecular mass of 120 kDa, the gene described here encodes a protein with a molecular mass of 169.87 kDa. Although widely disparate molecular masses for the SPA have been obtained by different methods and the SPA migrates anomalously by PAGE (8, 30, 31), this discrepancy is rather large. However, seven CNBr fragments predicted at the carboxyl end of the molecule from the deduced amino acid sequence were absent in the rickettsia-derived SPA. It is striking that the carboxyl end of the SPA for *R. prowazekii* contains both a hydrophobic region with α -helical configuration consisting of 24 amino acids (amino acids 1356–1379, corresponding to nt 4519–4590, Fig. 5), which could span the bacterial cell membrane and an adjacent 16 hydrophilic amino acids in a β -turn region (amino acids 1382–1396, corresponding to nt 4597–4641, Fig. 5) could serve, respectively, as a hydrophobic anchor and translocation stop. This hypothesis is consistent with what is presently known about hydrophobic anchor sequences that have been identified in transmembrane proteins (32, 33). Cleavage of the SPA protein near this anchor region would then result in a soluble surface protein with approximately the same molecular mass (130 kDa) and almost the identical amino acid composition as determined for *R. prowazekii*-derived SPA (9). Although the calculated pI of 5.4 for the protein that results from this hypothesized cleavage would differ from the pI of 4.1 already determined for the SPA (34), various modifications of the protein might account for this difference—such as phosphorylation, deamidation, and glycosylation.

We have compared the deduced amino acid sequence of the *spaP* gene for *R. prowazekii* with previously published sequences for the S-layer proteins of several prokaryotes (35–39) and are unable to identify any major regions of homology despite the fact that most S-layer proteins typically have a high proportion of acidic amino acids and usually a small amount of sulfur-containing amino acids (8). This result seems to suggest that similar characteristics of S-layer proteins of various prokaryotes did not arise via common ancestors or via the transfer of genes among bacteria. Signal sequences that are cleaved from mature S-layer proteins have been demonstrated for *Halobacterium*, *Bacillus brevis*, and *Deinococcus* (35–39). No cleavage of the relatively hydrophobic amino terminus was detected in the S-layer of *Caulobacter* (40). The amino terminus of *R. prowazekii* SPA is not very hydrophobic, does not exhibit homology to other signal sequences, and is not cleaved in the mature protein.

In the present study we have also cloned and sequenced *spaPt*, which is identical to the *spaP* gene but lacks the promoter region, the ribosomal-binding site, and 1162 bp of coding sequence located at the 5' end of the *spaP* gene. Because the SPA of *R. typhi* appears highly conserved among different strains and its gene appears highly homologous to *spaPt* (17), it is not readily apparent why or how *spaPt* is maintained stably in *R. typhi*. Deletion of the 1.4-kb fragment of DNA from an ancestor of *R. typhi* and *R. prowazekii* may be one of many such events that occurred during divergence of *R. typhi* as a species. The fact that the ORF of *spaPt* is identical to that of *spaP* suggests that some selective pressure exists to maintain *spaPt*, perhaps even an unknown functional role for the protein encoded by this truncated gene.

We thank Laura Perlinn, Cindy Jerominski, Kevin Swinson, Walter Sewell, and Benjamin Allen for expert technical assistance and Drs. Ken Stover, Patricia Guerry, Al Churilla, and Emilio Weiss for helpful discussion. This paper is dedicated to the memory of Dr. Richard Wistar, Jr., whose gentle and kindly nature and keen intellect have been an inspiration to those of us who were privileged to have worked with him. This investigation was supported by the Naval Medical Research and Development Command, Department of the Navy, Research Task No. 3M162770A870.AN.121. The opinions and assertions contained herein are the private ones of the authors and are not to be construed as official or reflecting the views of the Navy Department or the naval service at large.

- Sleytr, U. B. & Messner, P. (1988) *J. Bacteriol.* **170**, 2891–2897.
- Hovmoller, S., Sjogren, A. & Wang, D. N. (1988) *Prog. Biophys. Mol. Biol.* **51**, 131–163.
- Sleytr, U. B. & Messner, P. (1983) *Annu. Rev. Microbiol.* **37**, 311–339.
- Blaser, M. J., Smith, P. F., Repine, J. E. & Joiner, K. A. (1988) *J. Clin. Invest.* **81**, 1434–1444.
- Palmer, E. L., Mallavia, L. P., Tzianabos, T. & Obijeski, J. F. (1974) *J. Bacteriol.* **118**, 1158–1166.
- Silverman, D. J. & Wisseman, C. L., Jr. (1978) *Infect. Immun.* **22**, 233–246.
- Dasch, G. A., Samms, J. R. & Williams, J. C. (1981) *Infect. Immun.* **31**, 276–288.
- Ching, W.-M., Dasch, G. A., Carl, M. & Dobson, M. E. (1990) *Ann. N.Y. Acad. Sci.* **590**, 334–351.
- Dasch, G. A. (1981) *J. Clin. Microbiol.* **14**, 333–341.
- Koval, S. F. & Murray, G. E. (1984) *Can. J. Biochem. Cell Biol.* **62**, 1181–1189.
- Lepault, J. & Pitt, T. (1984) *EMBO J.* **3**, 101–105.
- Baumeister, W., Karrenberg, F., Rachel, R., Engel, A., ten Heggeler, B. & Saxton, W. O. (1982) *Eur. J. Biochem.* **125**, 535–544.
- Bourgeois, A. L. & Dasch, G. A. (1981) in *Rickettsiae and Rickettsial Diseases*, eds. Burgdorfer, W. & Anacker, R. L. (Academic, New York), pp. 71–80.
- Carl, M. & Dasch, G. A. (1986) *J. Immunol.* **136**, 2654–2659.
- Misiti, J. & Dasch, G. A. (1985) *J. Immunol.* **134**, 2689–2694.
- Carl, M., Vaida, S., Robbins, F. M., Hartzman, R. J. & Dasch, G. A. (1989) *Infect. Immun.* **57**, 1276–1280.
- Raoult, D. & Dasch, G. A. (1989) *J. Immunol. Methods* **125**, 57–65.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
- Leder, P., Tiemeier, D. & Enquist, L. (1977) *Science* **196**, 175–179.
- Young, R. A. & Davis, R. W. (1983) *Science* **222**, 778–782.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
- Rigby, P., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–242.
- Carl, M., Tibbs, C. W., Paparello, S. F., Dobson, M. E. & Dasch, G. A. (1990) *J. Infect. Dis.* **161**, 791–794.
- Gross, E. & Witkop, B. (1962) *J. Biol. Chem.* **237**, 1856–1860.
- Schagger, H. & von Jagow, G. (1987) *Anal. Biochem.* **166**, 368–379.
- Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. (1985) *Science* **230**, 1350–1354.
- Wood, D. O., Williamson, L. R., Winkler, H. H. & Krause, D. C. (1987) *J. Bacteriol.* **169**, 3564–3572.
- Marck, C. (1988) *Nucleic Acids Res.* **16**, 1829–1836.
- Smith, D. K. & Winkler, H. H. (1979) *J. Bacteriol.* **137**, 963–971.
- Ching, W.-M., Dobson, M. E., Falk, M., Weaver, J., Dasch, G. A., Carl, M. & Williams, R. (1989) *J. Cell Biochem.* **13**, 51.
- Eisenberg, D. A. (1984) *Annu. Rev. Biochem.* **53**, 595–623.
- Davis, G. D., Boeke, J. D. & Model, P. (1985) *J. Mol. Biol.* **181**, 111–121.
- Oaks, E. V., Wisseman, C. L., Jr., & Smith, J. F. (1981) in *Rickettsiae and Rickettsial Diseases*, eds. Burgdorfer, W. & Anacker, R. L. (Academic, New York), pp. 461–472.
- Peters, J., Peters, M., Lottspeich, F., Schafer, W. & Baumeister, W. (1987) *J. Bacteriol.* **169**, 5216–5223.
- Tsuboi, A., Uchihi, R., Tabata, R., Takahashi, Y., Hashiba, H., Sasaki, T., Yamagata, H., Tsukagoshi, N. & Udaka, S. (1986) *J. Bacteriol.* **168**, 365–373.
- Yamagata, H., Adachi, T., Tsuboi, A., Takao, M., Sasaki, T., Tsukagoshi, N. & Udaka, S. (1987) *J. Bacteriol.* **169**, 1239–1245.
- Tsuboi, A., Uchihi, R., Adachi, T., Sasaki, T., Hayakawa, S., Yamagata, H., Tsukagoshi, N. & Udaka, S. (1988) *J. Bacteriol.* **170**, 935–945.
- Lechner, J. & Sumper, M. (1987) *J. Biol. Chem.* **262**, 9724–9729.
- Fisher, J. A., Smit, J. & Agabian, N. (1988) *J. Bacteriol.* **170**, 4706–4713.