JOURNAL OF BACTERIOLOGY, Sept. 1989, p. 5199–5201
0021-9193/89/095199-03$02.00/0

Vol. 171, No. 9

Comparative Sequence Analysis of a Genus-Common Rickettsial Antigen Gene

BURT E. ANDERSON* AND THEODORE TZIANABOS

Viral and Rickettsial Zoonoses Branch, Division of Viral and Rickettsial Diseases, Center for Infectious Diseases, Centers for Disease Control, Atlanta, Georgia 30333

Received 23 March 1989/Accepted 10 June 1989

The genes encoding the 17-kilodalton genus-common antigen have been cloned and sequenced from *Rickettsia conorii*, *Rickettsia prowazekii*, and *Rickettsia typhi*. Compared with the *R. rickettsii* sequence, this sequence had a high degree of homology within the coding and control regions (*R. conorii*, 99.8%; *R. prowazekii*, 88.1%; *R. typhi*, 88.7%). The 5′ flanking regions, including the promoter and the transcription initiation sites, were extremely well conserved for all four species, suggesting that control and expression of this locus are important to the survival of the rickettsiae.

The 17,000-molecular-weight-antigen (17K antigen) gene of *Rickettsia rickettsii* is one of the most thoroughly characterized loci from any rickettsial genome (2, 3; B. E. Anderson, Ph.D. dissertation, Georgia State University, Atlanta, 1988). Nucleotide sequences for transcription initiation and putative translation initiation, signal peptide, and lipid modification sites have been identified (2). To determine the degree of genetic conservation among these regions, as well as other areas within the coding region for the mature protein, we cloned and sequenced this genus-common antigen from three additional pathogenic species of rickettsiae: *Rickettsia conorii*, *Rickettsia prowazekii*, and *Rickettsia typhi*. In addition, nucleotide sequences were examined for regions of variability that may serve as group- or species-specific priming sites for polymerase chain reaction amplification of rickettsial DNA from clinical samples.

To determine the level of conservation of the 17K antigen throughout the genus *Rickettsia*, immunoblot analysis was performed. *R. rickettsii* Sheila Smith, *R. conorii* Moroccan, *R. prowazekii* Breinl, *R. typhi* Wilmington, *Rickettsia bellii* RML 369-C, *Rickettsia akari* Hartford, and *Rickettsia canada* McKiel were grown in Vero cells as previously described (3) and purified by Renografin density gradient centrifugation (14). Purified rickettsiae were solubilized at 100°C and subjected to electrophoresis by the method of Laemmli (8). Resolved proteins in the gel were electroblotted to nitrocellulose filters by the method of Towbin et al. (11). The resulting filters were reacted with rabbit antiserum raised to the synthetic peptide NH_2-Asp-Asn-Gly-Asn-Tyr-Gly-Tyr-Val-Thr-Pro-Asn-Lys-Thr-Tyr-Arg-COOH, which is derived from residues 106 to 120 of the deduced amino acid sequence of the 17K antigen of *R. rickettsii* (3; Anderson, Ph.D. dissertation). When the filters were reacted with a horseradish peroxidase-conjugated anti-rabbit serum, all species of rickettsiae tested exhibited an antigenic protein with an approximate molecular weight of 17,000 (Fig. 1). This antigenic protein has not been observed in members of other genera, including *Bacillus*, *Proteus*, *Neisseria*, *Escherichia*, and *Chlamydia* (data not shown). Thus, the 17K antigen described for *R. rickettsii* appears to be a genus-common antigen found in members of both the spotted fever and the typhus group rickettsiae.

DNA from *R. rickettsii*, *R. conorii*, *R. typhi*, and *R. prowazekii* was isolated by lysis of purified rickettsiae with proteinase K and Sarkosyl, followed by phenol-chloroform extractions and subsequent ethanol precipitation. A number of primers homologous to different regions of the coding and flanking sequences of the 17K-antigen gene from *R. rickettsii* were chosen for polymerase chain reaction amplification of 1 ng of genomic rickettsial DNA. Twenty-five cycles of amplification were performed with a thermal cycler and a Gene-Amp kit according to the directions of the manufacturer (Perkin-Elmer/Cetus, Norwalk, Conn.). Various combinations of primer pairs were used for amplifying fragments of the gene from each of the four rickettsial species until the entire structural genes and flanking regions of each were obtained. The amplified fragments were cloned into plasmid vector pUC19 and transformed into *Escherichia coli* DH5α by the method of Hanahan (5), and the DNA sequence was determined for both strands by the Sanger chain termination method (10). The nucleotide sequences from overlapping fragments were connected to provide the sequence for the full-length coding region and promoter (Fig. 2). To ensure

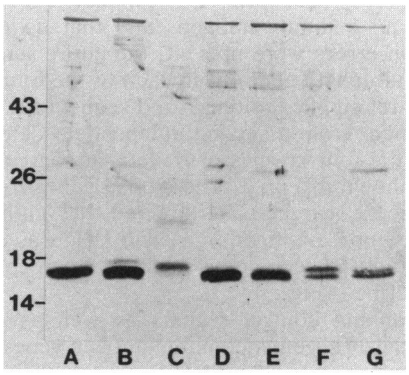

FIG. 1. Immunoblot of solubilized rickettsial proteins reacted with rabbit antiserum specific for the 17K genus-common antigen. Antiserum was raised to the synthetic peptide corresponding to residues 106 to 120 of the *R. rickettsii*-deduced amino acid sequence. Lane A, *R. akari*; lane B, *R. conorii*; lane C, *R. bellii*; lane D, *R. rickettsii*; lane E, *R. canada*; lane F, *R. typhi*; lane G, *R. prowazekii*. Electrophoresis was performed on a 12.5% acrylamide gel by the method of Laemmli (8). Molecular mass standards (in kilodaltons) are shown at left.

* Corresponding author.

```
          -35                      -10          +1 ---------->
Rr   1 TTTACAAAAT TCTAAAAACC ATATACTTAT T.AATTATATA ..TTAATTTAGA 50
Rc   1                               .           --             50
Rt   1                               -  A  -    GA     ---     48
Rp   1                               T  --  -  -- -     -       47
                fMET ------>
Rr  51 GAGAATTATA TGAAACTATT ATCTAAAATT ATGATTATAG CTCTTGCAAC   100
Rc  51                                                         100
Rt  49  -  -                      G A       T           G      96
Rp  48    G                         A                   G      97

Rr 101 TTCTATGTTA CAAGCCTGTA ACGGTCCGGG CGGTATGAAT AAACAAGGTA   150
Rc 101                                  C                      150
Rt  97                A      T         T       C         G    146
Rp  98                T      T    A A  T       C              147

Rr 151 CAGGAACACT TCTTGGCGGT GCTGGCGGCG CATTACTTGG TTCTCAATTC   200
Rc 151                                                         200
Rt 147  T             T  C   C     T                          196
Rp 148  C  T          T  C   A  A  T                          197

Rr 201 GGT.AAGGGCA AAGGACAGCT TGTTGGAGTA GGTGTAGGTG CATTACTTGG  250
Rc 201    -                                                    250
Rt 197    -C C  T         A      C             C              246
Rp 198    C    -T         A      C             C              247

Rr 251 AGCAGTTCTT GGTGGACAAA TCGGTGCAGG TATGGATGAA CAGGATAGAA   300
Rc 251                                                         300
Rt 247 G                              A   C       G            296
Rp 248 G                           C  A           G            297

Rr 301 GACTTGC.AGA GCTTACCTCA CAGAGAGCTT TAGAAACAGC TCCTAGTGGT  350
Rc 301       -                                                 350
Rt 297 A   - T   A  A  A      A            T T         C       346
Rp 298     - T   A  A  A      A            T T   A     C       347

Rr 351 AGTAACGTAG AATGGCGTAA TCCGGATAAC GGCAATTACG GTTACGTAAC   400
Rc 351                                                         400
Rt 347       A          C      A     T       C T              396
Rp 348       A          G      A             C T          C   397

Rr 401 ACCTAATAAA ACTTATAGAA ATAGCACTGG TCAATATTGC CGTGAGTACA   450
Rc 401                                                         450
Rt 397                   G   C  T  A                    A      446
Rp 398                   G   C  TG A                    A      447

Rr 451 CTCAAACAGT TGTAATAGGC GGAAAACAAC AAAAAGCATA CGGTAATGCA  500
Rc 451                                                         500
Rt 447                      G              C A     T  A        496
Rp 448             A                  -     - -     T           494

Rr 501 TGCCGCCAAC CTGACGGACA ATGGCAAGTT GTGAATTGA 539
Rc 501                                           539   99.8%
Rt 497                                           535   88.7%
Rp 495      T         T  G                        533   88.1%
```

FIG. 2. Nucleotide sequence alignment for the rickettsial 17K genus-common antigen genes. Sequences were aligned to the *R. rickettsii* gene for maximal homology, and only nucleotides differing from those in the *R. rickettsii* sequence are shown. Deletions (−) and inserted nucleotides (.) are indicated. Rickettsial sequences are indicated for *R. rickettsii* (Rr), *R. conorii* (Rc), *R. typhi* (Rt), and *R. prowazekii* (Rp). The *R. rickettsii* −35 and −10 promoter regions, transcription start site (+1), and initiator methionine (fMET) are overlined and indicated for reference. The nucleotide homology to the *R. rickettsii* sequence is indicated at the end of each sequence.

that the obtained sequences were responsible for expressing the 17K genus-common antigens and that no polymerase incorporation errors were present, the entire sequence (encoding the full-length gene) from each of the four rickettsial species was reamplified, cloned, and sequenced. The nucleotide sequences from a second independent determination, for each of the four species of rickettsiae, agreed perfectly with those shown in Fig. 2. The cloned genes from each of the four species of rickettsiae directed the synthesis of the 17K genus-common antigen in *E. coli* DH5α as determined by immunoblotting techniques with the antipeptide serum already described (Fig. 3).

The 5′ flanking control regions as well as the coding regions for the 17K genus-common antigen genes from each of the various species of rickettsiae were well conserved. The observed homology to the *R. rickettsii* 17K-antigen gene was 99.8% for *R. conorii* (Fig. 2, line 2), 88.7% for *R. typhi* (Fig. 2, line 3), and 88.1% for *R. prowazekii* (Fig. 2, line 4). The −35 (nucleotides 1 to 6) and −10 (nucleotides 22 to 27) promoter regions previously identified for *R. rickettsii* were entirely conserved among all four species of rickettsiae (Fig. 2). The adenine residue (indicated as the +1 nucleotide in Fig. 2), where transcription initiates in *R. rickettsii*, was also

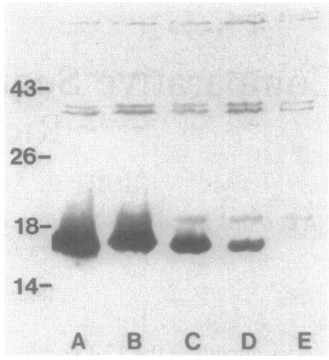

FIG. 3. Immunoblot showing cloned rickettsial 17K genus-common antigens expressed in *E. coli*. Samples were electrophoresed and reacted with antiserum as described in the legend to Fig. 1. The rickettsial sources of the gene expressed in *E. coli* are as follows: lane A, *R. rickettsii*; lane B, *R. conorii*; lane C, *R. typhi*; lane D, *R. prowazekii*; lane E, *E. coli* DH5α(pUC19). Molecular mass standards (in kilodaltons) are shown at left.

present in all the rickettsiae sequenced. Likewise, the presumed ribosome-binding site was well conserved, with only slight variations. The actual coding regions of the genes from *R. rickettsii* (Fig. 2, nucleotides 60 to 536), *R. conorii* (Fig. 2, nucleotides 60 to 536), *R. typhi* (Fig. 2, nucleotides 56 to 532), and *R. prowazekii* (Fig. 2, nucleotides 57 to 530) were also well conserved. At the amino acid level, the sequence of the 17K genus-common antigen of *R. conorii* was 100% homologous to the *R. rickettsii* sequence; *R. typhi* and *R. prowazekii* were 91.8% homologous (Fig. 4). The signal peptide sequences identified from the *R. rickettsii*-deduced amino acid sequence (residues 1 to 20) contain three amino acid substitutions for *R. typhi* (Fig. 4, line 3) and one for *R. prowazekii* (Fig. 4, line 4). Each of these changes results in substitution with similar hydrophobic amino acids. The tetrapeptide sequence Lys-Gln-Ala-Cys that targets the recombinant expressed antigen for lipid modification was found in the deduced amino acid sequence of all four rickettsiae (Fig. 4, residues 17 to 20).

The high degree of homology found in the coding regions for this genus-common antigen indicates a strong selective pressure for continued expression of this gene among a diverse group of rickettsiae. Furthermore, near-perfect conservation of regions controlling gene expression suggests that a specific level of expression from this genus-common

```
Rr    1 MKLLSKIMII ALATSMLQAC NGPGGMNKQG TGTLLGGAGG ALLGSQFGKG  50
Rc    1                                                        50
Rt    1       V  L    A                                   H   50
Rp    1                A          QS                      Q   50

Rr   51 KGQLVGVGVG ALLGAVLGGQ IGAGMDEQDR RLAELTSQRA LETAPSGSNV 100
Rc   51                                                        100
Rt   51                           SL      K L          S     I 100
Rp   51                           S         L          S     I 100

Rr  101 EWRNPDNGNY GYVTPNKTYR NSTGQYCREY TQTVVIGGKQ QKAYGNACRQ 150
Rc  101                                                        150
Rt  101          H                                      TT     150
Rp  101          H                    A           I     - T    149

Rr  151 PDGQWQVVN  159
Rc  151            159   100%
Rt  151            159    91.8%
Rt  150            158    91.8%
```

FIG. 4. Deduced amino acid sequences for the rickettsial 17K genus-common antigens. Sequences were aligned for maximal homology, with deletions indicated (−). The homology to the *R. rickettsii* amino acid sequence is indicated at the end of each sequence. Rickettsial sequences are those of *R. rickettsii* (Rr), *R. conorii* (Rc), *R. typhi* (Rt), and *R. prowazekii* (Rp).

antigen locus is beneficial for the rickettsiae. However, the role of this gene product in the life cycle of the rickettsiae has not been established.

The nucleotide sequence for the *R. typhi* 17K genus-common antigen gene contains 50 separate deletions or base changes compared with the *R. rickettsii* sequence. However, only 13 amino acid residues are affected, indicating that most of the mutations are silent. Similarly, only 13 amino acids from the *R. prowazekii*-deduced amino acid sequence differ from those of the 17K antigen of *R. rickettsii* despite 55 nucleotide changes or deletions. The vast majority of amino acid substitutions between the spotted fever and the typhus group 17K genus-common antigens occur in either neutral or hydrophilic regions. These may represent determinants found on the exterior portion of the antigen and may be surface-exposed epitopes that have undergone antigenic variation. The deduced amino acid sequence for *R. prowazekii* contains a deletion at residue 140; the glutamine residue found in *R. rickettsii*, *R. conorii*, and *R. typhi* is absent at this position in the sequence of *R. prowazekii* (Fig. 4).

Comparing nucleotide sequences for a common locus among a number of species of bacteria has assisted attempts at understanding the phylogeny of these organisms (4, 13). In the case of the 17K genus-common antigen, the gene is well conserved among members of both the spotted fever and the typhus group rickettsiae. For this reason, an ancestral form of the gene should have been the source from which this genus-common antigen evolved. The sequence comparisons provide data consistent with the fact that *R. prowazekii* and *R. typhi* evolved from a common source and are much more closely related to each other than to *R. rickettsii* or *R. conorii*. When the *R. rickettsii* and *R. conorii* 17K genus-common antigen genes and flanking regions were compared, only one nucleotide differed, despite the fact that these two rickettsiae were isolated on different continents. These data, coupled with the fact that DNA-DNA hybridizations between the entire genomes of the two species show greater than 90% homology (12), suggest that a separate species designation for *R. conorii* may not have been warranted. The high degree of similarity between *R. rickettsii* and *R. conorii* is also evident in serological cross-reactivity (1, 9) and cross-reactive T-cell responses (6, 7). Regardless, from the data presented in this report concerning the 17K genus-common antigen gene, it is obvious that *R. rickettsii* and *R. conorii* are highly related. In contrast, areas of divergence within the nucleotide sequence between the members of the spotted fever group (*R. rickettsii* and *R. conorii*) and the typhus group (*R. typhi* and *R. prowazekii*) can be seen in the 17K genus-common antigen gene (Fig. 2). These divergent areas are currently being used as group-specific priming sites for polymerase chain reaction amplification of subpicogram quantities of rickettsial DNA (T. Tzianabos, B. E. Anderson, and J. E. McDade, submitted for publication). This technique should allow detection of rickettsial DNA in blood from infected individuals and provide a much-needed specific and sensitive test for early diagnosis of rickettsial diseases.

We are grateful to Brian Holloway for synthesizing the oligonucleotide primers and the synthetic peptide used to generate specific antiserum to the genus-common antigen. We also thank William Bellini for helpful advice and discussion and Russell Regnery for providing *R. conorii* genomic DNA.

LITERATURE CITED

1. **Anacker, R. L., R. N. Philip, E. Casper, W. J. Todd, R. E. Mann, M. R. Johnston, and C. J. Nauck.** 1983. Biological properties of rabbit antibodies to a surface antigen of *Rickettsia rickettsii*. Infect. Immun. **40**:292–298.
2. **Anderson, B. E., B. R. Baumstark, and W. J. Bellini.** 1988. Expression of the 17-kilodalton antigen gene from *R. rickettsii*: transcription and posttranslational modification. J. Bacteriol. **170**:4493–4500.
3. **Anderson, B. E., R. L. Regnery, G. M. Carlone, T. Tzianabos, J. E. McDade, Z. Y. Fu, and W. J. Bellini.** 1987. Sequence analysis of the 17-kilodalton-antigen gene from *Rickettsia rickettsii*. J. Bacteriol. **169**:2385–2390.
4. **Fox, G. E., E. Stackebrandt, R. B. Hespell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. S. Tanner, L. J. Magrum, L. B. Zablen, R. Balakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Luehrsen, K. N. Chen, and C. R. Woese.** 1980. The phylogeny of prokaryotes. Science **209**:457–463.
5. **Hanahan, D.** 1985. Techniques for transformation of *E. coli*, p. 109–136. *In* D. M. Glover (ed.), DNA cloning: a practical approach, vol. 1. IRL Press, Oxford.
6. **Jarobe, D. L., C. S. Eisemann, and T. R. Jerrells.** 1986. Production and characterization of cloned T-cell hybridomas that are responsive to *Rickettsia conorii* antigens. Infect. Immun. **52**:326–330.
7. **Jerrells, T. R., D. L. Jarobe, and C. S. Eisemann.** 1986. Cross-reactive lymphocyte responses and protective immunity against other spotted fever group rickettsiae in mice immunized with *Rickettsia conorii*. Infect. Immun. **51**:832–837.
8. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
9. **Plotz, H., R. L. Reagan, and K. Wertman.** 1944. Differentiation between fievre boutonneuse and Rocky Mountain spotted fever by means of complement fixation. Proc. Soc. Exp. Biol. Med. **55**:173–176.
10. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
11. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76**:4350–4354.
12. **Walker, D. H., J. V. Lange, P. A. Hanff, R. C. Tidwell, and B. C. Hegarty.** 1984. Antigens of *Rickettsia rickettsii* and *Rickettsia conorii*, p. 244–247. *In* L. Leive and D. Schlessinger (ed.), Microbiology—1984. American Society for Microbiology, Washington, D.C.
13. **Weisburg, W. G., C. R. Woese, M. E. Dobson, and E. Weiss.** 1985. A common origin of rickettsiae and certain plant pathogens. Science **230**:556–558.
14. **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.