Diversity of Immunodominant 56-kDa Type-specific Antigen (TSA) of Rickettsia tsutsugamushi

SEQUENCE AND COMPARATIVE ANALYSES OF THE GENES ENCODING TSA HOMOLOGUES FROM FOUR ANTIGENIC VARIANTS*

(Received for publication, September 13, 1991)

Norio Ohashi‡, Hiroko Nashimoto‡, Hideo Ikeda‡, and Akira Tamura‡

From the ‡Department of Microbiology, Niigata College of Pharmacy, Niigata 950-21, the ‡Department of Biosciences, Teikyo University, Utsunomiya 329, and the §Department of Molecular Biology, Institute of Medical Science, University of Tokyo, Tokyo 108, Japan

There are several antigenic variants in Rickettsia tsutsugamushi, and a type-specific antigen (TSA) of 56-kilodaltons located on the rickettsial surface is responsible for the variation. The primary structures of the protein in two variants, Gilliam and Karp, have been reported independently by us and Stover *et al.* by cloning and sequencing the corresponding genes (Ohashi, N., Nashimoto, H., Ikeda, H., and Tamura, A. (1990) Gene (Amst.) 91, 119–122; Stover, C. K., Marana, D. P., Carter, J. M., Roe, B. A., Mardis, E., and Oaks, E. V. (1990) Infect. Immun. 58, 2076–2084). In the present study, genes encoding the TSA homologues of the other four variants, Kato, Kawasaki, Kuroki, and Shimokoshi, which are all distinguishable serologically, were cloned and sequenced, and consequently, it became possible to compare the primary structures of the six antigenic variants. The sequence analyses revealed a complete open reading frame encoding 55,308–56,745-dalton proteins with 521–532 amino acids, in which a putative signal peptide consisting of 22 amino acids was recognized at the NH₂-terminal end. Transcription of the gene is regulated by several tandem promoters. All TSA molecules have the characteristics of transmembrane proteins with alternating hydrophobic and hydrophilic regions, and contain four variable domains with spans of 16–40 amino acids which are located in the hydrophilic regions in the molecule and show different amino acid sequences among the strains. Phylogenetic classification among the R. tsutsugamushi strains based on TSA homologues supports the antigenic relationships known in the closely and distantly related strains.

Rickettsia tsutsugamushi, a causative agent of scrub typhus, is an obligate intracellular bacterium. In this species of rickettsiae, several antigenic variants are recognized. The antigenic types of rickettsiae corresponding to the Gilliam, Karp, and Kato strains (1) are generally recognized, and these strains are frequently used as prototype strains. However, we recently isolated the additional antigenic types of Shimokoshi (2), Kawasaki (3), and Kuroki (4, 5) in Japan. These variants are distinguishable by serological cross-tests with strain-specific polyclonal or monoclonal antibodies. On the other hand, we demonstrated that these phenotypic variations of the rickettsiae depend on the antigenicity of an immunodominant 56-kDa major protein located on the rickettsial outer membrane (3, 5–7), and we defined this protein as a type-specific antigen (TSA) of this organism. Additionally, we found that the rickettsial strains which have TSA molecules antigenically distinguishable from the Gilliam, Karp, and Kato strains are all avirulent in mice (8). This suggests the possibility that the diversity of TSA molecules is associated with rickettsial pathogenesis. Furthermore, TSA is easily eliminated from the rickettsial surface by trypsin treatment of intact rickettsia (6), and the enzyme-treated rickettsiae lose the ability to adsorb to the host cell surface (9), suggesting that this TSA molecule plays some biological role in the initial steps of rickettsial infection. From this evidence, we recognize the importance of analyzing the molecular structure of TSA and to compare the structures among strains.

Our recent comparative analysis of the NH₂-terminal amino acid sequences of three TSA homologues purified from the prototype Gilliam, Karp, and Kato strains, revealed seven amino acid substitutions within 35 amino acid residues which are elicted by codon wobble, and in the Karp strain by one amino acid deletion (10). In addition, we previously succeeded in cloning and sequencing the gene (tsg56) encoding the TSA of the Gilliam strain, and the complete primary structure of the TSA was deduced from the nucleotide sequence (11). Independently, Stover *et al.* (12) reported the gene structure and the deduced amino acid sequence of the TSA homologues in the Karp strain. The two TSA molecules show 81.3% sequence similarity. It is very difficult in rickettsia to employ manipulation techniques generally used in common bacteria for analyses of the molecular basis of their pathogenesis. Hence, our initial approach to rickettsial pathogenesis was concentrated in a comparison of the complete primary structures of the TSA homologues of representative virulent and avirulent strains.

In the present study, we have cloned and sequenced four genes encoding the TSA homologues from the virulent strain Kato and from the three avirulent strains Kawasaki, Kuroki,

*This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan (to H. I., A. T., and N. O.), from the Nissan Science Foundation (to H. I.), and from the Naito Foundation (to A. T.). The costs of publication in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) M63380, M63381, M63382, and M63383.

‡To whom correspondence should be addressed. Tel.: 81-33443-8111; Fax: 81-33443-8704 and 81-33443-3899.

†The abbreviations used are: TSA, type-specific antigen; SDS, sodium dodecyl sulfate; kb, kilobase pair(s); bp, base pair(s); ORF, open reading frame; SDS, sodium dodecyl sulfate.
and Shimokoshi. A comparison of the deduced amino acid sequences among the six TSA homologues, including those of virulent strains Gilliam (11) and Karp (12), will be presented.

EXPERIMENTAL PROCEDURES

**Bacterial Strain**—The *R. tsutsugamushi*, Kato, Kawaski, Kuroki, and Shimokoshi strains used throughout this study were grown in suspension cultures of the L929 mouse fibroblast cell line and were purified by Percoll (Pharmacia P-L Biotechnology, Sweden) density gradient centrifugation. The procedures of cultivation and purification of the rickettsiae were essentially the same as described previously (15). The purified rickettsiae were used to prepare genomic DNA.

**Escherichia coli** strain JM103 (14) derived from JM183 and E. coli strains MV1184 and JM109 were used as the host strains for pUC plasmids and M13 phages.

**Enzymes**—Restriction endonucleases and modification enzymes were obtained from Takara Shuzo Co. Ltd., Kyoto, Japan, or Boehringer Mannheim, Germany.

**Southern Hybridization Analysis**—Genomic DNA of *R. tsutsugamushi* was isolated from purified rickettsiae by SDS lysis, Pronase digestion, CsCl density gradient centrifugation or RNase treatment to remove RNA, phenol-chloroform extraction, and ethanol precipitation, according to the modified procedure of Myers and Wiseman (15). The DNA was digested with HindIII or PstI endonucleases, electrophoresed, and transferred to nitrocellulose by the standard methods (16). The 2.3-kb HindIII fragment from a clone, pTG56-1, of the Gilliam strain (11) was labeled with [α-32P]dATP by the random primer method using commercially prepared reagents (Amer sham International plc., United Kingdom (U.K.)), and used as a DNA probe. Hybridization was performed under appropriately stringent conditions in solutions with 0.1% SDS, 50 mM Tris-hydrochloride, pH 7.5, 1 mM EDTA, 20 to 50% formamide, 20 μg/ml denatured salmon sperm DNA, 1× Denhardt's (0.02% each of Ficoll, polyvinylpyrrolidone, and bovine serum albumin), and 3× SSC (1× SSC containing 0.15 M sodium chloride and 0.015 M sodium citrate) at 42°C for 20 h. The nitrocellulose sheet was washed in 0.1× SSC with 0.1% SDS and then exposed to Fuji x-ray film at -70°C.

**Isolation of Genes Encoding TSA**—The HindIII fragments of DNA detected by genomic Southern hybridization analysis as described above were ligated into pUC19 or pUC118 plasmid vectors, and the recombinant plasmids were introduced into competent cells of *E. coli* JM109 or MV1184. By colony hybridization tests according to standard procedures (16) with the radiolabeled 2.3-kb fragment from pTG56-1 (11) as a probe, the radio-positive clones were isolated from the transformants. The isolated clones which have an insert length of 2.5 kb from the Kato and Shimokoshi strains, 1.8 kb from the Kawaski strain, 3.6 kb from the Kuroki strain, and 3.8 kb from the Shimokoshi strain, are designated as pT56H, pT56H, pTW56H, or pTW56H, respectively, as shown in Fig. 1. In clone pTW56H, the 1.8-kb DNA insert lacked the sequence encoding the COOH-terminal region of the TSA of the Kawasaki strain, and 3.6 kb from the Kuroki strain, are designated as pTT56H, pTT56H, pTW56H, or pTR56H, respectively, as shown in Fig. 1. In clone pTW56H, the 1.8-kb DNA insert lacked the sequence encoding the COOH-terminal region of the TSA of the Kuroki strain, 3.1-kb PstI fragment of the Kawasaki strain, which contained the sequence that partially overlapped with that of the 1.8-kb HindIII fragment, was isolated and cloned by the same procedures as described above with the WPH1.0 fragment as a probe. This clone was designated as pTW56P.

**Sequence Determination and Analysis**—The cloned DNAs from the rickettsiae were subcloned into M13 mp18/mp19 phages or pUC118 plasmid vectors. A recombinant pTW56P which has overlapping sequence with that of pTW56H, was cloned by using a WPH1.0-fragment probe (II). Open and closed boxes show open reading frames of the TSA protein. Arrows indicate the direction and extent of the sequencing in the DNA fragments subcloned in M13 mp18/mp19 or pUC118, and the arrows with blackened squares indicate sequencing by the use of synthetic oligonucleotide primers. Abbreviations D, H, P, and R are the restriction sites of DraI, HindIII, PstI, and Rsal, respectively.

**RESULTS**

**Gene Structures of Rickettsial TSA Homologues**—We sequenced 2,272 bp (Kato), 2,260 bp (Kawaski), 2,284 bp (Kuroki), and 2,303 bp (Shimokoshi) in the cloned rickettsial

FIG. 1. Restriction maps and sequencing strategy of the *R. tsutsugamushi* DNA inserts including the tss56, tus56, tar56, or tss6 genes in the recombinant clones: pTT56H (Kato), pTW56H and pTW56P (Kawaski), pTR56H (Kuroki), pTS56H (Shimokoshi). These rickettsial DNA fragments were cloned in pUC19 or pUC118 vectors. A recombinant pTW56P which has overlapping sequence with that of pTW56H, was cloned by using a WPH1.0-fragment probe (III). Open and closed boxes show open reading frames of the TSA protein. Arrows indicate the direction and extent of the sequencing in the DNA fragments subcloned in M13 mp18/mp19 or pUC118, and the arrows with blackened squares indicate sequencing by the use of synthetic oligonucleotide primers. Abbreviations D, H, P, and R are the restriction sites of DraI, HindIII, PstI, and Rsal, respectively.
DNAs. The nucleotide sequences consist of 556–592-bp of 5′-noncoding region, 1,563–1,596 bp of ORF, and 129–148 bp of 3′-noncoding region. The base composition of the sequenced DNAs showed high A + T contents (65%), especially prominent in the 5′- and 3′-noncoding regions (71–80%). This reflects a high A + T-rich (70%) genomic DNA in R. tsutsugamushi (19).

In the primer extension analysis of total RNA from infected L cells with two kinds of primers to determine the initial site of transcription (Fig. 2), three major (P1–P3) and some minor (P4 and asterisks) transcripts were recognized at 45, 52, and 64 h of postinfection during the developmental growth of rickettsiae (Kato strain). No extension products were observed in the RNA of uninfected cells at any different times, indicating that the transcripts are rickettsia-specific (data not shown). The putative sequences of the −35 and −10 promoter regions seen at 7–37 bp upstream from the 5′-end of four transcripts were TATACT and TATAAA (P1), TTTACA and TATACT (P2), TTGCGT and TATCTA (P3), and TTGATT and TATAAT (P4), respectively (Fig. 3), of which the sequences were similar to those of the other prokaryotic consensus promoter sequences (20). The identical sequences were seen in the corresponding areas of the 5′-flanking regions from the other rickettsial strains, including Gilliam (11) and Karp (12), except for the tss56 gene (Shimokoshi), such as TGATT at −35 of P1 (−10 region of P2), TTTATG at −35 of P2, and TTGAT at −35 of P3, and in the tsr56 gene (Kuroki), such as TAATTA at −10 of P4. Thus, transcription of the TSA genes in the six strains seems to be regulated by similar multiple tandem promoters. The spaces between the −35 and −10 regions were 17 bp in P1 and P3, 18 bp in the P2 of six strains, and 17 bp in the P4 of five strains, except 16 bp in that of Karp (12). These are consistent with the optimal spacing (17 ± 1) for prokaryotic promoters (21). The spaces between the −10 regions and 5′-end nucleotides of transcripts were 6–8 bp in all cases. We (11) and Stover et al. (12) presumed previously that the promoter regions in the tsg56 (Gilliam) and sta56 (Karp) genes correspond to P2 and P4 regions, respectively. Thus, we find that the mRNAs of different sizes are generated by transcription from tandem promoters, and similar multiple tandem promoters in bacterial outer membrane genes were reported in the case of the MOMP gene in Chlamydia (22, 23).

The sequence AGGAG at 7 bp upstream from the start codon, which was identified as a potential ribosome-binding site in the previous study (11), is recognized in the genes from all strains (Fig. 3 and 4). In the 3′-flanking region, Stover et al. (12) reported a potential hairpin-like loop structure with a free energy of −20 Kcal in the sta56 gene (Karp), which has the characteristics of a rho-independent transcriptional terminator. The identical sequences are also seen in the tss56 (Kato) (Fig. 4A) and tsg56 (Gilliam) genes (11). The corresponding structures in the tsr56 (Kuroki) and sta56 (Shimokoshi) genes consist of sequences with an incomplete inverted repeat including 1 or 2 unpaired base(s), and the free energy of formation for a potential base-paired region was calculated as −14 to −16 Kcal (Fig. 4, C and D). In the tsu56 gene (Kawasaki), 5 mismatched bases are included in the area corresponding to the inverted repeat in the other strains, indicating that this may not form a hairpin-like structure. In the coding regions, many of the codons used prefer A or U at the third position as in the case of other rickettsial genes (24).

Structure of TSAs—All four genes encode proteins with M, 56,112 (Kato), 55,969 (Kawasaki), 56,754 (Kuroki), and 55,308 (Shimokoshi) composed of 521 to 532 amino acids (Fig. 4). The presence of a putative signal peptide in the genes of tsg56 (Gilliam) and sta56 (Karp) was demonstrated previously (11, 12). Because the sequence of 35 amino acids at the NH2-terminus of mature TSA in the Kato strain, which was determined chemically in our previous study (10), is seen in the coding region of the tss56 gene (Kato) (indicated by a single underline in Fig. 4A), the sequence composed of 22 amino acids preceding the NH2-terminal sequence of mature TSA is predicted to be a signal peptide. From comparison of the amino acid sequences in the NH2-terminal region of TSAs, the sequences of 22 amino acid residues on the NH2-terminal side which are encoded by the tsu56 (Kawasaki), tss56 (Kuroki), and tsg56 (Shimokoshi) genes are also predicted as signal peptides (Fig. 4, B–D). The amino acid sequences of the signal peptides of five strains (except that of the Shimokoshi strain) are identical, while purine-to-purine replacements of nucleotides are recognized at four positions in corresponding regions of the genes. In the signal peptide of the Shimokoshi strain, two amino acids differ from those of the other strains, such as the substitution of Lys to Arg and Ser to Asn at the second and 21st positions, respectively. These signal peptides from all strains consist of a basic region and a hydrophobic core, of which the characteristics are seen in the signal peptides of other prokaryotic cells (25). The processing site of the signal peptide is at the bond between the 22nd and 23rd amino acids, with Ile at the NH2-terminal amino acid of mature TSA in all cases. Three COOH-terminal amino acids of the signal peptide are Ala-Ser-Ala in five strains and Ala-Asn-Ala in the Shimokoshi strain, and this Ala-X-Ala sequence is included in the preferred amino acid

![Fig. 2. Primer extension analysis to determine the 5′-end of the transcript of the tss56 gene (Kato strain). Two kinds of primers were used here. The primer positions corresponding to the tss56 gene are shown in Fig. 3. A, the pattern of transcripts obtained with primer 1. B, the pattern of transcripts obtained with primer 2. Lanes 1 and 8 (G), 2 and 9 (A), 3 and 10 (T), and 4 and 11 (C) show the sequence analysis of a recombinant M13 clone with the same primers. Lanes 5 and 12, RNA extracted from rickettsia-infected cells at 45 h of postinfection; lanes 6 and 13, at 52 h; and lanes 7 and 14, at 65 h. Arrowheads and symbols in parentheses indicate the initiation sites and identification of promoters for the four transcripts, respectively. The other minor transcripts are shown by asterisks. The plus strand sequences at the area shown on the left and these correspond to the sequences in Fig. 3.](image-url)
sequence of signal peptides at the processing sites proposed by Oliver (25).

The calculated Mr of the mature TSAs from the deduced sequences are 53,861 (Kato), 53,718 (Kawasaki), 54,504 (Kuroki), and 53,002 (Shimokoshi), and these values are almost equivalent to the size, 54-58 kDa, estimated by SDS-polyacrylamide gel electrophoresis (3, 5, 6). The isoelectric points of the mature TSAs predicted from the amino acid sequences (pl 5.0-5.4) are also equivalent to the values estimated, pl 5.3-5.7, by isoelectrofocusing-acrylamide gel electrophoresis of a lysate of purified rickettsiae.2

Amino acid sequence alignments of the TSAs of the six antigenic variants are shown in Fig. 5. In this comparative analysis, substitutions or deletions of one or several contiguous amino acid residues are recognized all over the molecules, but the significant differences in the sequence among strains are seen in the regions indicated by arrowheads and asterisks as shown in Fig. 5, and the sequence is numbered beginning at the first initiation point (+1) upstream from the start codon (barred ATG). Arrows indicate two oligonucleotide primers (minus strand) used for determination of transcriptional initiation sites in the tat56 gene. Nucleotide sequences of the Gilliam and Karp strains from the report of Ohashi et al. (11) and Stover et al. (12). Gaps are indicated by dashes.

FIG. 3. Alignment of nucleotide sequences of promoter region in the TSA gene of six *R. tsutsugamushi* strains. The putative -10 and -35 regions of four promoters and ribosome-binding site (RBS) are underlined. The initiation site of each tat56 transcript is indicated by arrowheads and asterisks as shown in Fig. 2, and the sequence is numbered beginning at the first initiation point (+1) upstream from the start codon (barred ATG). Arrows indicate two oligonucleotide primers (minus strand) used for determination of transcriptional initiation sites in the tat56 gene. Nucleotide sequences of the Gilliam and Karp strains from the report of Ohashi et al. (11) and Stover et al. (12). Gaps are indicated by dashes.

sequence of signal peptides at the processing sites proposed by Oliver (25).

The calculated Mr of the mature TSAs from the deduced sequences are 53,861 (Kato), 53,718 (Kawasaki), 54,504 (Kuroki), and 53,002 (Shimokoshi), and these values are almost equivalent to the size, 54-58 kDa, estimated by SDS-polyacrylamide gel electrophoresis (3, 5, 6). The isoelectric points of the mature TSAs predicted from the amino acid sequences (pl 5.0-5.4) are also equivalent to the values estimated, pl 5.3-5.7, by isoelectrofocusing-acrylamide gel electrophoresis of a lysate of purified rickettsiae.2

Amino acid sequence alignments of the TSAs of the six antigenic variants are shown in Fig. 5. In this comparative analysis, substitutions or deletions of one or several contiguous amino acid residues are recognized all over the molecules, but the significant differences in the sequence among strains are seen in the regions indicated by arrowheads and asterisks as shown in Fig. 5, and the sequence is numbered beginning at the first initiation point (+1) upstream from the start codon (barred ATG). Arrows indicate two oligonucleotide primers (minus strand) used for determination of transcriptional initiation sites in the tat56 gene. Nucleotide sequences of the Gilliam and Karp strains from the report of Ohashi et al. (11) and Stover et al. (12). Gaps are indicated by dashes.

amino acids are substituted from the sequence of the Karp strain.

VDII consists of 16-26 amino acids. In this domain, the amino acid sequences of the Gilliam and Kawasaki strains show high homologies of 86.4% (89.4% in the nucleotide sequence), and there are only three amino acid substitutions. However, the similarities of the other strains show greater variation. In particular, the sequence of the Shimokoshi strain in this area shows the deletion of several amino acids in comparison with those of the other strains.

VDIII consists of 24-35 amino acids. High conservation is seen in the amino acid sequences between the Gilliam and Kawasaki strains, and between the Karp and Kuroki strains. The similarities are 82.8 and 77.1% in amino acid sequences (88.5 and 85.7% in nucleotide sequences), respectively. The homologies between the Kato and Gilliam strains, and the Shimokoshi and Karp strains, are 51.5 and 60.0% in amino acid sequences, respectively.

VDIV is the largest variable domain, consisting of 34-40 amino acids. In this domain, the sequences of the Gilliam, Karp, and Kuroki strains are similar to each other with two to four amino acid substitutions; the homologies are 88.2-94.1% in amino acid sequences (96.1-97.1% in nucleotide sequences). The amino acid sequences of the Kato, Kawasaki, and Shimokoshi strains in this region are unique.

Hydropathy analysis shows that the TSAs of all the antigenic variants have alternative hydrophilic and hydrophobic motifs which are characteristic of transmembrane proteins (Fig. 7). All variable domains mentioned above locate mostly in the hydrophilic regions. Comparison of hydropathic profiles in each domain among the strains shows: (i) the profiles in the VDI region can be divided into two types, such as the Gilliam, Kuroki, Karp, and Kawasaki types, and the Kato and Shimokoshi types; (ii) the patterns in the VDII region are almost unique in each TSA protein; (iii) the profiles in the VDIII region are similar to each other except that of the Shimokoshi strain; (iv) the VDIV region locates in the largest hydrophilic area in the molecule and the patterns are more
Fig. 4. Nucleotide sequences of the complete coding regions of four TSA homologues in respective R. tsutsugamushi antigenic variants and the deduced amino acid sequences. A, tst56 gene of Kato strain. B, tsw56 gene of Kawasaki strain. C, tsr56 gene of Kuroki strain. D, tss56 gene of Shimokoshi strain. The numbering of bases in the tsw56, tsr56, or tss56 genes is begun at the adenine corresponding to the initiating base (+1) of the tst56 gene. The long underline in the tst56 gene indicates the NH$_2$-terminal amino acid sequence of mature TSA determined chemically by Ohashi et al. (10). Vertical arrows in the tsw56, tsr56, and tss56 genes show the putative cleavage sites between the signal peptide and NH$_2$ termini of mature TSA. Inverted repeats in A, C, and D are indicated by one set of opposing arrows.

similar to each other. In the amino acid sequences of the TSA molecules, a unique Gln-rich sequence is seen at the 342nd to 362nd amino acid positions in Fig. 5, and this region is hydrophilic; but in the Shimokoshi strain, the Gln sequence is short and is recognized as a small hydrophilic peak.

Homology of TSAs from Different Antigenic Variants—

Homologies of nucleotide sequences in the ORF regions, and of the amino acid sequences in all the TSA homologues (include signal peptides) (Table I) are summarized as follows: (i) the sequences of the Gilliam, Karp, and Kuroki strains have relatively high homologies (86–89% in nucleotide sequences and 81–82% in amino acid sequences); (ii) the se-
**Fig. 5. Alignment of deduced amino acid sequences of TSA homologues from six antigenic variants of **R. tsutsugamushi**.** Aligned positions of identical amino acids with the TSA of the Gilliam strain are shown with asterisks. The sequence of TSA from the Karp strain is from the report of Stover et al. (12). Gaps indicated by dashes were introduced for optimal alignment of all six TSA homologues. *Brackets* indicate four variable domains (VDI, II, III, IV).

**Fig. 6. Schematic diagram of diversity of TSA homologues.** The numbers at the top show each variable domain. The patterns with *fine* and *coarse* lateral stripes indicate high and some homology in the amino acid sequences, respectively, with that of the Gilliam strain in each domain, and *black patterns* show the highly homologous sequences with that of the Karp strain. The *black and white checkered patterns* represent sequence profile in each strain.

Sequence of the Kawasaki strain shows higher homology to that of the Gilliam strain than to those of the other strains (83% in nucleotide and 76.2% in amino acid sequences); (iii) the sequence of the Shimokoshi strain has the lowest homology relative to those of the other strains (71-73% in nucleotide and 69-62% in amino acid sequences); (iv) the sequence of the Kato strain shows an intermediate degree of homology with those of the others, except for Shimokoshi strain (75-79% in nucleotide and 66-72% in amino acid sequences). These relations are represented with a phylogenetic tree.
protein probe would be expected to detect several genomic restriction
endonuclease fragments which have sequences homologous to
the branching points, respectively.

Emini et al. (27) and Jameson and Wolf (28), respectively,
suggests the possibility that all VD regions locate on the
rickettsial surface and serve as antigenic determinant(s) (data
not shown).

If the molecular basis of genetic variation in R. tsutsugamushi
is based upon a repertoire of unique genes, as in
the case of gene conversion of pili in Neisseria (29), a TSA gene
probe would be expected to detect several genomic restriction
donorabstractcise fragments which have sequences homologous to
the probe, each corresponding to a gene in a different genomic
context. Southern hybridization of restriction endonuclease
digests of the genomic DNA from several antigenic types of
rickettsiae with the cloned probe showed only one hybridizable
restriction fragment (data not shown). The result suggests
that the TSA genes reside in a similar genomic context
in each antigenic variant and that only one TSA allele is
represented in each R. tsutsugamushi. The antigenic variation
in R. tsutsugamushi may be an allelic variation of the TSA
gene, as in the case of the MOMP genes in Chlamydia trachomatis (30).

The question arises as to how the polymorphisms in the
variable domains of the TSA molecules arise. As shown in
Fig. 6, the differences or similarities among strains is seen in
the sequences of each of the domains. If diversity is solely due
to the accumulation of point mutations during long evolution-
ary periods, consistent allele genealogy should be seen among
the gene structures of different strains in each variable do-
main. However, a reliable allelic genealogy could not be found
in comparisons of the nucleotide sequences of the variable
domains among strains, indicating that the diversity in each
domain is not caused solely by the accumulation of point
mutations. Possibly, this diversity may be due to a combina-
tion of point mutations and recombinational processes. For
example, the Kuroki strain seems to be a chimera of the
Gilliam and Karp strains, because the sequence of the Kuroki
strain is similar to that of the Gilliam strain in VDI, with the
Karp strain in VDIII, and with both the Gilliam and Karp
strains in VDIV. On the other hand, the sequences of the
Gilliam strain in VDII and VDIII are well conserved in the
Kawasaki strain, but the sequences in VDI and VDIV of both
strains do not resemble each other. Therefore, we may con-
clude that the homoplasy results from the shuffling of poly-
morphic segments between alleles by recombinational pro-
cesses. A similar diversity resulting from shuffling of polymor-
phic segments was reported in the case of the class II genes
of the murine major histocompatibility complex (31). In R.
tsutsugamushi, Gotoh (32) reported antigenic conversion of
progenies by mixed infection of two antigenic types of rick-
ettsiae in mice. This observation can be understood if recom-
bination occurs between two antigenically different strains.

Another speculation regarding the genetic variation of the
TSA genes in this rickettsia is the possibility that these
diversities are generated by a special bias during growth in
their host cells. Our previous observations showed that the
antigenic type of the rickettsiae seems to be related to the
species of vector mites of the rickettsia. We found that all
rickettsiae isolated from patients in the southwestern endemic
area in Japan, where Leptotrombidium scutellare is present as a
predominant vector, are immunologically classified into either of the Kawasaki or Kuroki types (4). In contrast, almost
all isolates in northeastern Japan, where L. pallidum is a main
vector, are classified into either of the Gilliam or Karp types
(7). This suggests that Kawasaki and Kuroki type rickettsiae
are transmitted by L. scutellare, while Gilliam and Karp types
by L. pallidum. These data lead us to the possible speculation
that the diversity of the TSA gene is derived by the accumu-

| Strains   | % of nucleotide/variable domain | % of amino acid/variable domain |
|-----------|---------------------------------|---------------------------------|
| Karp      | 81.6                            | 86.3                            |
| Kuroki    | 81.3                            | 87.8                            |
| Gilliam   | 68.8                            | 76.2                            |
| Kawasaki  | 69.9                            | 71.5                            |
| Kato      | 60.9                            | 61.2                            |
| Shimokoshi| 60.2                            | 59.2                            |

a Numbers above the diagonal indicate the homologies of nucleotide sequences.
b Numbers below the diagonal indicate the homologies of amino acid sequences.
lation of special selective mutations which are biased by host cell control during the long term symbiosis of vertically transmitted rickettsiae in the vector mites.

We have previously described that the Kawasaki and Kuroki strains show some cross-reaction with the Gilliam and Karp strains, respectively, in serological tests, but the Shimokoshi strain is considerably unique among the variants (2, 4). These relations were genetically ascertained by comparison of homologies in the nucleotide or amino acid sequences in the entire ORF of TSA from the variants (Table I) and in the phylogenetic relationship (Fig. 8). In particular, the homology in each variable domain from six variants may be the reasons for the immunological cross-reactivities.

In the TSA gene, the homology of the 5'- and 3'-flanking regions is higher than that of the region coding the six variants. The high homology of flanking regions suggests that transcriptional regulation of TSA gene is similar among the variants. Indeed, similar sequences in the −35 and −10 regions of four tandem promoters and in hairpin-loop structure were seen in each variants (Fig. 3 and 4). The tandem promoters may be required to supply massive amounts of protein products, because TSA is the most abundant surface protein, probably a structural protein, of *R. tsutsugamushi* (~20% of rickettsial total protein). Thus, the conservation of flanking region seems to be essential for *R. tsutsugamushi*.

As for the function of TSA, we previously observed the suggestive evidence that TSA is a protein which participates in the attachment of rickettsiae to eukaryotic host cells in the initial step of rickettsial infection (6, 9). Stover et al. (12) reported that the 165 amino acids at the NH2 terminus of TSA in the Karp strain including a signal peptide encoded by the sta56 gene has a moderate degree of homology with human Sendai virus and influenza virus M2 proteins. These cysteines may participate in the NH2-terminal side of the VDI region, the amino acid sequences of the regions (except VDI) are conserved in the entire ORF of TSA from the variants (Table I) and in the phylogenetic relationship (Fig. 8). In particular, the homology in each variable domain from six variants may be the reasons for the immunological cross-reactivities.

In a comparative analysis of TSA primary structures between the virulent strains (Gilliam, Karp, and Kato) and the avirulent strains (Kawasaki, Kuroki, and Shimokoshi), critical sequences which are specific to either of the virulent or avirulent strains are not observed. However, the cysteine content of the mature TSA molecules of each strain are different, i.e. Gilliam (2 residues), Karp (3 residues), Kato (4 residues), Kawasaki (4 residues), Kuroki (5 residues), and Shimokoshi (2 residues). These cysteines may participate in forming the tertiary structures of the molecules by S–S bonding. Therefore, there is the likelihood that there are differences in the tertiary conformation of the individual molecules, and this may be related to the virulence of this rickettsia.

Acknowledgments—We thank Drs. A. Kawamura, H. Tanaka, and S. Nogami for their support in this project and Dr. T. Takeda for advice in preparation of synthetic oligonucleotides.

REFERENCES

1. Weiss, E., and Moulder, J. W. (1984) in *Berger's Manual of Systematic Bacteriology* (Krieg, N. R., and Holt, J. G., eds) Vol. 1, pp. 688–698, Williams & Wilkins, Baltimore
2. Tamura, A., Takahashi, K., Tsuruhara, T., Urakami, H., Miyamura, S., Sekikawa, H., Kennemou, M., Shibata, M., Abe, S., Hirasawa, H., and Nishimura, N., 671–675
3. Yamamoto, S., Kawahata, N., Tamura, A., Urakami, H., Ohashi, N., Murata, M., Yoshida, Y., and Kawamura, A., Jr. (1986) *Microbiol. Immunol.* 30, 611–620
4. Yamamoto, S., Kawahata, N., Oura, K., Murata, M., and Minamishima, Y. (1989) *Jpn. Assoc. Infect. Dis.* 63, 109–117
5. Ohashi, N., Tamura, A., Sakurai, H., and Yamamoto, S. (1990) *J. Clin. Microbiol.* 28, 2111–2113
6. Tamura, A., Ohashi, N., Urakami, H., Takahashi, K., and Oyanagi, M. (1985) *Infect. Immun.* 48, 671–675
7. Murata, M., Yoshida, Y., Osono, M., Ohashi, N., Oyanagi, M., Urakami, H., Tamura, A., Nogami, S., Tanaka, H., and Kawamura, A., Jr. (1986) *Microbiol. Immunol.* 30, 599–610
8. Tamura, A. (1988) *Jpn. J. Bacteriol.* 43, 629–639
9. Tamura, A. (1988) *Microbiol. Sci.* 5, 228–232
10. Ohashi, N., Tamura, A., Ohta, M., and Hayashi, K. (1989) *Infect. Immun.* 57, 1427–1431
11. Ohashi, N., Nishimoto, H., Ikeda, H., and Tamura, A. (1990) *Gene* (Amst.) 91, 119–122
12. Stover, C. K., Marana, D. P., Carter, J. M., Roe, B. A., Mardis, E., and Oaks, E. V. (1990) *Infect. Immun.* 58, 2076–2084
13. Tamura, A., Urakami, H., and Tsuruhara, T. (1982) *Microbiol. Immunol.* 26, 321–329
14. Chiba, M., Shimizu, H., Fujimoto, A., Nishimoto, H., and Ikeda, H. (1989) *J. Biol. Chem.* 264, 12755–12790
15. Myers, W. F., and Wiseman, C. L., Jr. (1980) *Int. J. Syst. Bacteriol.* 30, 143–150
16. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: a Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
17. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* 74, 5463–5467
18. McNight, S. L., and Kingsbury, R. (1982) *Science* 217, 316–324
19. Kumura, K., Minamishima, Y., Yamamoto, S., Ohashi, N., and Tamura, A. (1991) *Int. J. Syst. Bacteriol.* 41, 247–248
20. Hawley, D., and McClure, W. R. (1983) *Nucleic Acids Res.* 11, 2237–2255
21. von Hippel, P. H., Bear, D. G., Morgan, W. D., and McSwiggen, J. A. (1984) *Annu. Rev. Biochem.* 53, 389–446
22. Stephens, R. S., Wagner, E. A., and Edman U. (1988) *J. Bacteriol.* 170, 744–750
23. Yuan, Y., Zhang, X.-Y., Manning, D. S., Caldwell, H. D. (1990) *Infect. Immun.* 58, 2850–2855
24. Winkler, H. H., and Wood, D. O. (1988) *Biochimica* 70, 977–986
25. Oliver, D. (1985) *Annu. Rev. Microbiol.* 39, 615–648
26. Shishido, A. (1962) *Jpn. J. Microbiol.* 5, 228–232
27. She, J. X., Boehme, S. A., Wang, T. W., Bonhomme, F., and Wakeland, E. K. (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 453–457
28. Gotz, J. (1989) *Jpn. J. Bacteriol.* 23, 437–445
29. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* 157, 105–132
30. Nei, M. (1987) *Molecular Evolutionary Genetics*, pp. 293–298, Columbia University Press, New York, NY