ISOLATION AND CHARACTERIZATION OF THE *dnaA* GENE OF *RICKETTSIA PROWAZEKII*

R.T. WAITE, E.I. SHAW, H.H. WINKLER, D.O. WOOD*

Laboratory of Molecular Biology, Department of Microbiology and Immunology, University of South Alabama College of Medicine, Mobile, Alabama 36688, USA

Received April 21, 1998; accepted April 27, 1998

Summary. – The *dnaA* gene encoding the initiator protein of DNA replication was isolated from the obligate intracellular bacterium, *Rickettsia prowazekii.* Comparison of the deduced amino acid sequence of *R. prowazekii* DnaA with other bacterial DnaA proteins revealed extensive similarity. However, the rickettsial sequence is unique in the number of basic lysine residues found within a highly conserved portion of the putative DNA binding region, suggesting that the rickettsial protein may recognize a DNA sequence that differs from the consensus DnaA box sequence identified in other bacteria. Consensus DnaA box sequences, found upstream of many bacterial *dnaA* genes, were not identified upstream of rickettsial *dnaA* gene. In addition, gene organization within this region differed from that of other bacteria. The putative start of transcription of the rickettsial *dnaA* gene was localized to a site 522 nucleotides (nt) upstream of the DnaA start codon.

Key words: Rickettsia prowazekii; *dnaA* gene; initiator protein

Introduction

Rickettsia prowazekii, the etiologic agent of epidemic typhus, is an obligate intracellular parasitic bacterium that grows only within the cytoplasm of a eukaryotic host cell. As might be expected, this organism is well adapted to exploit this rich intracytoplasmic environment (Weiss 1982; Winkler, 1990). However, the rickettsiae exhibit a relatively slow generation time of 8 – 12 hrs. Such slow growth may have evolved to maximize the number of rickettsiae produced within a host cell (Winkler, 1995). Whatever the mechanisms may be that control rickettsial growth, replication of the genome is an essential process that must be completed prior to cell division. In *Escherichia coli*, the timing of DNA replication and cell division is controlled at the initiation step of DNA replication. In order to examine this critical process in the rickettsiae, we have begun the isolation and characterization of genes coding for components involved in the initiation of chromosomal replication in *R. prowazekii.*

In *E. coli,* the *dnaA* gene codes for the initiator protein DnaA, a protein indispensable for normal chromosomal replication (Skarstad and Boye, 1994). DnaA binds to 9 bp long recognition sites within *oriC* opening up the DNA for formation of the replication complex. In addition, it also serves as a versatile transcriptional regulator that can function as a repressor, activator or transcriptional terminator (Messer and Weigel, 1997). DnaA is highly conserved among bacterial species and a similar organization of specific genes contiguous to *dnaA* is observed in many species as well (Skarstad and Boye, 1994). Often the chromosomal origin of replication (*ori*) is found in this region (Ogasawara and Yoshikawa, 1992). However, variations of this common gene arrangement have been identified (Old *et al.,* 1993; Margolin *et al.,* 1995). This study addresses the unique characteristics of this gene in the obligate intracytoplasmic bacterium *R. prowazekii.*

*Corresponding author.

Abbreviations: aa = amino acid; EDTA = ethylenediamine tetraacetate; nt = nucleotide; *ori* = origin of replication; ORF = open reading frame; PCR = polymerase chain reaction; p.i. = post inoculation; SDS = sodium dodecyl sulfate

Materials and Methods

Bacterial strains and plasmids. E. coli strain XL1-Blue [*recA endA1 gyrA96 thi hsdR17 supE44 relA1 lac* (F' *proAB lacIqZ* M15 Tn10)] was used as the host strain in these studies. *E. coli* was maintained as frozen stock at -70°C in nutrient broth containing 12.5% glycerol and grown in Luria-Bertani medium (Davis *et al.,* 1980). When required, the antibiotics ampicillin and kanamycin were each added to a final concentration of 50 μg/ml. The Madrid E strain of *R. prowazekii* was propagated in six-day-old embryonated, antibiotic-free hen eggs by inoculation with a seed pool (passage 218). Eight days post inoculation (p.i.), infected yolk sacs were harvested and rickettsial suspensions were prepared as previously described (Winkler, 1976). The rickettsiae were further purified by a variation (Plano and Winkler, 1989) of Renografin density gradient centrifugation (Dasch and Weiss, 1977; Hanson *et al.,* 1981). Vectors used in this study included pBluescript SKII, Lambda Zap II (Stratagene) and PT7Blue (Novagen, Inc.).

DNA isolation and sequencing. Standard procedures for DNA isolation and sequencing were used (Sambrook *et al.,* 1989). Chromosomal DNA was isolated from *R. prowazekii* and *E. coli* using the procedure of Marmur (1961). Small scale plasmid preparations were purified from *E. coli* using the Wizard Miniprep DNA Purification Kit from Promega or the Circleprep Purification Kit from BIO101. The pZ523 spin columns purchased from 5 prime-3 prime, Inc. or the cesium chloride density gradient centrifugation method (Sambrook *et al.,* 1989) were used to purify larger amounts of plasmid DNA. Restriction enzymes were purchased from Life Technologies or New England Biolabs, Inc. DNA fragments were purified from agarose gels using the Geneclean II Kit from BIO101. The Lambda ZapII *R. prowazekii* clone bank generated in this study was screened with DNA probes labeled with [α-^{32}P]dATP (ICN) using the Random Primers DNA Labeling System (Life Technologies). Manual DNA sequencing was accomplished by the dideoxynucleotide chain-termination method (Sanger *et al.,* 1977) using the Sequenase Version 2.0 Kit supplied by Amersham. Automated sequencing was performed using the Applied Biosystem Taq Cycle Sequencing Kit (Applied Biosystem) as designed for use with the Perkin-Elmer 9600 Thermocycler. The sequencing reactions were analyzed on the University of South Alabama College of Medicine Biopolymer Center's Applied Biosystem Model 370A Automated DNA Sequenator. DNA sequencing was facilitated by the generation of nested deletions within the target sequence using the Erase-a-Base System (Promega). DNA and protein sequence analyses were performed using the GCG collection of programs (Devereux *et al.,* 1984) maintained on the University of South Alabama SPARC System 10 Model 30 SUN Computer.

Polymerase chain reaction (PCR) amplifications were conducted using the GeneAmp DNA Amplification Kit (Perkin-Elmer). Deoxyoligonucleotide primers were purchased from the University of South Alabama College of Medicine Biopolymer Center, Oligos Etc. or Integrated Biotechnologies. The PCR volume of 100 μl contained 100 ng of template DNA with primers added to a final concentration of 1 μmol/l. The amplifications were carried out on an Ericomp Thermocycler or a Perkin-Elmer 9600 Thermocycler.

RNA isolation and assays. Total RNA (rickettsial RNA and host cell RNA) was isolated from rickettsiae-infected L929 cells as previously described (Cai and Winkler, 1993) using the hot-phenol method (Emory and Belasco, 1990). For controls, RNA was also isolated from mock-infected L929 cells. For RNase protection assays, a specific antisense RNA probe, labeled with [α-^{32}P]UTP (ICN), was generated by *in vitro* run-off transcription with T7 RNA polymerase using the Riboprobe System (Promega). An RNase protection assay was performed according to standard methods (Sambrook *et al.,* 1989). Briefly, total RNA (100 μg) was hybridized with 10^6 cpm of labeled probe overnight at 30°C. Post hybridization, 1000 U of Rnase T1 (Life Technologies) was added in 300 μl of a digestion buffer containing 10 mmol/l Tris.HCl pH 7.4, 5 mmol/l ethylenediamine tetraacetate (EDTA) and 300 mmol/l NaCl, and incubated for 1 hr at 30°C. The reaction was terminated by the addition of sodium dodecyl sulfate (SDS) to a final concentration of 0.5% and proteinase K to a final concentration of 4 mg/ml and incubation for 30 mins at 37°C. The reaction mixture was extracted with phenol and precipitated with ethanol and the protected fragments were separated on a 6% polyacrylamide-urea gel and analyzed by autoradiography. Primer extension assay was conducted using the Primer Extension System from Promega and standard methods as described previously (Cai *et al.,* 1995). The primer used for extension in this assay (DW96, 5'-GAATGGTTATTAGAGTGC-3') was end-labeled using ^{32}P-ATP (>7,000 Ci/mmole; ICN) and T4 polynucleotide kinase.

Results

Isolation and characterization of the R. prowazekii dnaA gene

A PCR approach, successfully used to isolate the *R. prowazekii gyrA* gene (Wood and Waite, 1994), was used to identify the rickettsial *dnaA* gene. Two DnaA regions of conserved amino acids (aa), identified by comparing characterized DnaA protein sequences from other bacterial species, were selected and degenerate oligodeoxyribonucleotide primers specific for these regions were synthesized. Primer DW30 (5'-CARGARGARTTYTTYCAYAC-3') corresponded to aa 247-253 and primer DW37 (5'-GTNGTRTGRTC NCKNCCNCCRAA-3') corresponded to aa 429-436 of *E. coli* DnaA, respectively. When used in a PCR with *R. prowazekii* DNA as template, a 573 bp fragment was amplified that when sequenced exhibited a deduced aa sequence with 55% identity to a portion of *E. coli* DnaA (data not shown).

To accomplish the isolation of the entire rickettsial *dnaA* gene, the amplified fragment was used to screen size-specific (5 – 8 kb) *R. prowazekii Hin*dIII clone banks. A clone (MOB833) that hybridized strongly to the probe was identified and the recombinant plasmid isolated from this clone was designated pMW833. The sequence of this rickettsial insert starting at the *Hin*dIII site indicated in Fig. 1 and of an additional clone containing further upstream sequence (see below) is presented in Fig. 1. Anal-

ysis of this sequence revealed a large open reading frame (ORF) of 1,392 nt that begins with a GTG and terminates with a TAA and codes for a protein exhibiting 44% identity to *E. coli* DnaA. The GTG codon was designated as the start codon based on best fit alignment with other DnaA proteins, the conserved aa found between this codon and the next probable translational start at position 1481–1483 and the fact that the GTG codon is used as the initiating codon in other bacterial *dnaA* genes (Skovgaard and Hansen, 1987). Although not completely matching the

```
CTACTTGTATTTCGTGTAAAGCTATTTTCTACTAACAACATAGAAAGTTCATCCATAGTAAATTTTATTTTAGGATATAATTTCAAAAAATATTTAGAAA  100
AAAAATCTTTTGGTGGGAGTTTGTTTAATAAATTTGCTGCATGAGTAATAAGCAATTTCCCCTTACTATTTGTTATTAGTTTTGTTAATGTTTCCGCACG  200
TCTAGATACGATATTTGCGTTCGGTGACGTATGATCATAAGGAATTGTATCATAGCTTGGAAAGTAATAAATATTGTCATTTGAAGAAAAAAATAACGCT  300
TGCTTATATAATTGCAATGCCTCTTCTTCATTACTCACACTTAAAATGAAATCTTGATTAAGATTCTTAGTAAAATTATCAATTACATAGAAACATTTGG  400
CAGTTGCCGGAAATTTTTGTTGGATCATTCTACAGGAGCTTATATTTTTGCTATTTTGCGTAATTTGTCAATTATTTCGGGATCTAGATAAGAAGGTGCA  500
                    <---Mfd
GAAGGTTTATTATTAATCCAATTATAAAGATCATTATCATTTTGATCAAGTATTAAAGTATAACTTTTAAGTTTCTTTTCATCCATTGAAGATAAGTACT  600
TTTCGGCAAAGCGACCAAGTATATAATCCATTTCTCTGCAACCACGCTTTTTACTACGATAAAAAAGTTTCTTTTGTAAAGCGTTTTTATTTAATTTATT  700

CATCTATAACATCTTGCAAAATTTTTTTATAATAATATTATTAAAATAATAAGGTTCTATGAACATCATTTTCAATGATATAGAGCCTAGATACCGAAAT  800

CATTTTAAAATACTTGAATAATTAAGCACTCTAATAACCATTCTACATTGGGAAAGTTCTATTAACTTTTATTATCAATAATAGAAAAAAACGTAAATAT  900
                                                                                                      M
                    HindIII
GTTGGTTCAAAATATCTGTTCTACAAAAGCTTATAATATGCTAATTTTAAATAATAATGCTTTTCTTGTAGATGTTAGAACACAAGAAGAATGGAAACAA 1000
 L  V  Q  N  I  C  S  T  K  A  Y  N  M  L  I  L  N  N  N  A  F  L  V  D  V  R  T  Q  E  E  W  K  Q
GTTGGGATACCTCATTTAGATAATAAGAATAAAGTAATTTTTCTAAGTTTGCAATTGAATAAAAACTTTGAAGATAATTTTTTTATCTATTATTAACGAGA 1100
V  G  I  P  H  L  D  N  K  N  K  V  I  F  L  S  L  Q  L  N  K  N  F  E  D  N  F  L  S  I  I  N  E  K
AAATAGATACAGCCATATTTTTTCTGTGTCGTTCAGGATATAGGTCATTCATTGCAGCAAATTTTATAGCTAATATTGGTTATAAAAATTGTTATAATAT 1200
  I  D  T  A  I  F  F  L  C  R  S  G  Y  R  S  F  I  A  A  N  F  I  A  N  I  G  Y  K  N  C  Y  N  I
AAGTGACGGTTTTGAAGGTAATAATCAAGATAAAGGCTGGAAACAAAATAACTTACCGTGGCAGTTTTAAGTGAGCACTAATCAAATACCTTTAACAGAT 1300
 S  D  G  F  E  G  N  N  Q  D  K  G  W  K  Q  N  N  L  P  W  Q  F  *  M  S  T  N  Q  I  P  L  T  D
                                                                  DnaA ------>
CAGGGTGATAATTATGTAAATGTCTGGAGTTACGTTGCTCAAGATCTTTACAATCATTACGGTGAAACTCTATATAATAGTTGGTTTAGTAAGGTCAATT 1400
Q  G  D  N  Y  V  N  V  W  S  Y  V  A  Q  D  L  Y  N  H  Y  G  E  T  L  Y  N  S  W  F  S  K  V  N  F
TTATAGAGTCTTCATTAAATACCGTTATTCTGTGTGCACCTACTAATTTTATTAGAGATTGGATAAAATCCAAATATGCTATGGTCATATTGCAACTATT 1500
  I  E  S  S  L  N  T  V  I  L  C  A  P  T  N  F  I  R  D  W  I  K  S  K  Y  A  M  V  I  L  Q  L  F
CCAACATTATAATAATGCTATTAAGTCAATTGAAATAATTACTAAAGAGTTACCTGGAACAACACAAGCAGTGATAGAATCACCTACTAAGACTTTTGCC 1600
 Q  H  Y  N  N  A  I  K  S  I  E  I  I  T  K  E  L  P  G  T  T  Q  A  V  I  E  S  P  T  K  T  F  A
GATATCGGCAACAGTGCACTAAATTCAGAAAATATCGTTTCAACACTTGATGTACGTTTTACTTTCGATAATTTTGTGGTTGGAGTACCAAATGAACTAG 1700
D  I  G  N  S  A  L  N  S  E  N  I  V  S  T  L  D  V  R  F  T  F  D  N  F  V  V  G  V  P  N  E  L  A
CTTATGCAGCAGCAAGAGCTGTGGCAGAATCATCAGGGGCAGTGTTTGAATCTAATCCACTTTTTCTATATGGCGGAGTTGGACTCGGTAAAACTCATTT 1800
  Y  A  A  A  R  A  V  A  E  S  S  G  A  V  F  E  S  N  P  L  F  L  Y  G  G  V  G  L  G  K  T  H  L
AATGCATGCAATTGGTTGGTACATCAAACAAAATAACCCAAGGCGCAAAGTAATATACATGTCAGCAGAAAAATTTATGTATCAGTTTGTTAAAGCTCTG 1900
 M  H  A  I  G  W  Y  I  K  Q  N  N  P  R  R  K  V  I  Y  M  S  A  E  K  F  M  Y  Q  F  V  K  A  L
CGTAATAAAGAAGTAATTTTATTTAAAGAGAAATTCCGATCAGTTGACGTACTAATGATTGATGACATTCAATTTATCTGCGGTAAAGATAGTACACAGG 2000
R  N  K  E  V  I  L  F  K  E  K  F  R  S  V  D  V  L  M  I  D  D  I  Q  F  I  C  G  K  D  S  T  Q  E
AAGAATTTTTCCACACTTTTAATACATTGATTGATAATAACCGTCAAATGGTTATTTCTTGCGATAGGTCACCTTCAGATCTAGATAATATTGAAGATCG 2100
  E  F  F  H  T  F  N  T  L  I  D  N  N  R  Q  M  V  I  S  C  D  R  S  P  S  D  L  D  N  I  E  D  R
CATCAAATCTCGCCTTGGATGGGGCTTAGTTGCAGATGTGCATAGTACTACTTATGAATTACGCCTCGGCATTCTAGAATCTAAAATCGAGCAAATGAAT 2200
 I  K  S  R  L  G  W  G  L  V  A  D  V  H  S  T  T  Y  E  L  R  L  G  I  L  E  S  K  I  E  Q  M  N
GTTAAAATACCAAAAGATGTGATAAATTTTTTAGCATCTAAAATTGTTTCAAACGTTAGAGAACTTGAAGGAGCTTTAAATAAAGTGATTGCTCATTCTA 2300
V  K  I  P  K  D  V  I  N  F  L  A  S  K  I  V  S  N  V  R  E  L  E  G  A  L  N  K  V  I  A  H  S  N
ATTTTACTTTAAAAGCAATTACGCTTGAAAATACACAAAATATTTTGCGAGATTTATTGCGTTCTAATGAAAGGATAATTACTGTTGAAGATATTCAAAA 2400
  F  T  L  K  A  I  T  L  E  N  T  Q  N  I  L  R  D  L  L  R  S  N  E  R  I  I  T  V  E  D  I  Q  K
AAAAGTAGCTAGTCGTTATAATATTAAATTATCTGATATGTATTCTTCACGCAGGTTGCGAGAAGTCGCAAGACCACGTCAAATCGCTATGTATCTTAGT 2500
 K  V  A  S  R  Y  N  I  K  L  S  D  M  Y  S  S  R  R  L  R  E  V  A  R  P  R  Q  I  A  M  Y  L  S
AAAACATTAACGCCGAAAAGTCTTGCAGATATTGGTAAAAAATTCGGTAAAAAAGATCATACTACGGTTATGCATGCTATTAAGAAAGTAGAAGAATTAC 2600
K  T  L  T  P  K  S  L  A  D  I  G  K  K  F  G  K  K  D  H  T  T  V  M  H  A  I  K  K  V  E  E  L  L
TCGAAAATGATATAGAATTACGTGAGGAAATTAATTTGTTAATGAAAATATTGCAAAATTAAATACGCATAATATATTCATGTATAACACGATTTGATAG 2700
  E  N  D  I  E  L  R  E  E  I  N  L  L  M  K  I  L  Q  N  *
CAATTTCCAGTTTTAAAAAACTAAAGTTATTTAGTTTGAAATTGTTTTGCTTGATCAAGTATTCAAATAACATGATAATGTTGAACGTGCACTCTAGATC 2800
TATGAAACTAGAACCTCCACATACATGATATAGATAATATATACAAAACGACCTACAAAATGAAGTTTTTATTCTCTGCTCGAGCGCTAATGAATTAATA 2900
TATAAAAAGAAATTATAAATAATGTACATTTTATAGGATCAC  2942
```

Fig. 1

Nucleotide sequence of the *R. prowazekii dnaA* gene

The predicted amino acid sequence is given in single letter code. The start and orientation of DnaA and Mfd are indicated. The sequence complementary to the Mfd start codon is underlined. The putative promoter region sequences are indicated by double underlines while the transcriptional start site nucleotides are shadowed. The nucleotide sequence reported in this paper was submitted to the GenBank database and assigned Acc. No. U55213.

B. burgdorferi	inien*kkillrelkithkdieghskkpeitka*hiyay*lrnf*el*tve**kiig*kt*s**lysin*i.drdrnndkeinnlite*mnkikkn	486
T. maritimi	..idelieivakvtgvpreeilsns*nvkalta*rig*yvaknylks**rt*aek*.n*s*pv*vdsvk*vkdsllkgnkqlkalidevigeisrralsg	440
E. coli	vtidn*qktvaey*kikvadllskr*srsvarp**m**a*akel*nh**pe**da*g******lh*cr*i.eqlreeshdikedfsn*irtlss	467
S. typhimurium	vtidn*qktvaey*kikiadllskr*srsvarp**m**a*akel*nh**pe**da*g******lh*cr*i.eqlreeshdikedfsn*irtlss	466
S. marcescens	vtidn*qktvaey*kikvadllskr*srsvarp**m**a*akel*nh**pe**da*g******lh*cr*i.eqlreeshdikedfsn*irtlss	464
P. mirabilis	vtidn*qktvaey*kikvadllskr*srsvarp**m**a*akel*nh**pe**da*g******lh*cr*i.eqlreeshdikedfsn*irtlss	466
V. harveyi	vtidn*qktvaey*kikvadllskr*srsvarp**l**a*akel*nh**pe**da*g******lh*cr*i.eqlreeshdikedfsn*irtlss	468
B. aphidicola	vtian*qktvaey*kikvadllsrr*srsvarp**m**amakel*nh**pe**da*s******lh*cr*i.eqlrkenhdikedfsn*irtlsv	431
P. putida	vsvdn*qrtvaey*kikisdllskr*srsvarp**v**a*skel*nh**pe**dm*g******lh*cr*i.nelkesdadiredykn*lrtltt	506
H. influenzae	vtien*qkvvaey*rikvsdlksks*arsvtrp**i**a*akel*nr**pe**ra*.d******ln*crev.pkfreqdnsiqedwan*irtlsa	454
Synechocystis	aapet*itivaqh*qlkveellsns*rrevsla**vg*y*mrqh*dl**pr**ea*g*k*****myscd*i.tqlqqkdwetsqtlts*shriniagqapes	446
P. marinus	vtpkqvldkvaev*kvtpdemrsas*rrpvsqa**vg*y*mrqg*nl**pr**dt*g*k*****my*ieqv.ekklssdpqiasqvqkirdllqidsrrkr	461
M. leprae	isaat*mtataey*dttieelrgpgktralaqs**i**y*crel*dl**pk**qa*.******my*qr*i.lsemaerrevfdhvke*ttrirqrskr	521
M. tuberculosis	isaat*maataey*dttveelrgpgktralaqs**i**y*crel*dl**pk**qa*.******my*qr*i.lsemaerrevfdhvke*ttrirqrskr	507
M. smegmatis	istaa*mavtaey*ettveelrgpgktralaqs**i**ymcrel*dl**pk**qa*.*****rrdvrerr....saarwpsavrcsttsgthhahpparqtlsset	...495
S. aureus	itiqd*qkivgqy*nvriedfsakk*tksiayp**i**y*srel*df**pk**ee*g******ih*he*i.skdlkedpifkqeven*ekeirnv	453
B. subtillis	itike*qrvvgqq*nikledfkakk*tksvafp**i**y*srem*ds**pk**ee*g******ih*he*i.sklladdeqlqqhvkeikeqlk	446
M. luteus	itpel*lhatgey*nltleeltsks*trtlvta**i**y*lrel*em**pk**qvlg******ih*dr*i.relmaerrtiynqvte*tneikrkqrga	515
S. coelicolar	itsta*mgatady*gltvedlcgts*gralvta**i**y*crel*dl**pk**al*g******mh*dr*i.rnlmaerrsiynqvte*tnrikng	656
R. melitoti	vried*qrvvakh*nvsrqelvsnr*trvivkp**i**y*sktl*pr*fpe**rr*g******lh*vr*i.eelisadtklsheiel*krline	507
C. crescentus	itidd*qkataeh*gmkqadllser*nravarp**a**w*akql*tr**pd**rr*g******lh*vrri.ealraedsalshdlet*trklrg	490
M. pneumoniae	fdpyllienvcrr*nvpvdsvlsen*kaelvrv*dvcnyilrekykmqfqq**ki*.k*n*s**va*vkrv.akmiekddslqdiits*vi	439
M. genitalium	fdpyllienvcrr*nvpmdsvlsen*kaelvrv*dvcny*lrqkynmqfqq**ki*.k*s*ss*lm*vkrv.akmiendsslrdvits*vi	437
M. capricolum	lnvkk*kevvsek*gisvnaidgka*sksivta*hi**y*tkeilnht*aq**ee*g******in*er*i.emmlkkdkqlkktvdi*knkiltk	450
S. citri	itpkk*kqivads*nitikammsks*vsnvmqa**l**yfcrtlldepftr**te*g*k*****mnsvk*v.eahistnkefkhlvnairrkiegr	450
R. prowazekii	itved*qkkvasr*niklsdmyssr*lrevarp**i**y*sktl*pk**ad**kk*gkk*****mh*ik*v.eellendielreeinl*mkilqn	463
Consensus	**-----I-------Y-----------R-------RQ-AM-L--- T--SL--IG--F-GRDHTTV--A--K-----------------L----------**	

Fig. 2

Comparison of DnaA domain IV amino acid sequences (carboxy-terminal end) of 26 bacterial DnaA proteins

The GCG program PILEUP (Devereux *et al*., 1984) was used to generate the alignment. Stars indicate aa that match the consensus sequence shown at the bottom. Gaps introduced to maximize alignment are indicated by dots. Consensus bases are shown in uppercase letters and represent a base that is found in at least 20 of the 26 sequences. The DnaA domain region is as defined by Ogasawara *et al.* (1991). Rickettsial lysine residues are double underlined. Numbers refer to the last aa of each protein. The sequence of *M. smegmatis* is shown only to aa 469. The following sequences [Acc. No.] were obtained from the GenBank data base: *B. burgdorferi* [L14948], *T. maritima* [U24145], *E. coli* [J01602], *S. typhimurium* [M17352], *S. marcescens* [M17353], *P. mirabilis* [M58352], *V. harveyi* [L47617], *B. aphidicola* [M80817], *P. putida* [X14791], *H. influenzae* [L45631], *Synechocystis* [L36958], *P. marinus* [U44977], *M. leprae* [L39923], *M. tuberculosis* [X92504], *M. smegmatis* [U17833], *S. aureus* [D89066], *B. subtilis* [X02369], *M. luteus* [M34006], *S. coelicolar* [M82836], *R. melitoti* [L39265], *C. crescentus* [U01667], *M. pneumoniae* [U34816], *M. genitalium* [U39734], *M. capricolum* [D90426], *S. citri* [Z19108].

E. coli-like ribosomal binding site, AGGAGGT, (Shine and Dalgarno, 1974), there is a purine-rich sequence, GTGGCAG, appropriately placed upstream of this start codon. Translation of the ORF would produce a protein of 463 aa with M_r of 52,972, a size comparable to *E. coli* DnaA (Skarstad and Boye, 1994). Codon usage in the *R. prowazekii dnaA* gene is similar to that of other characterized *R. prowazekii* genes in its preference for (A+T)-rich codons (Winkler and Wood, 1988; Andersson and Sharp, 1996). Restriction mapping revealed that the rickettsial insert of pMW833 was only 2.0 kb rather than the expected 5 – 8 kb and that one of the predicted terminal *Hin*dIII sites of the insert had been deleted.

Homology of the R. prowazekii dnaA gene product to known DnaA proteins

Comparison of the rickettsial DnaA sequence to that of other characterized DnaA proteins revealed extensive homology. The highest level of overall identity was with *E. coli*, other enteric species and *Pseudomonas putida* DnaA proteins which exhibited approximately 44.5% identity to *R. prowazekii* DnaA. The lowest level of identity, approximately 28.5%, was seen with *Mycoplasma* species. *R. prowazekii* DnaA adheres to the domain organization observed by Ogasawara *et al.* (1991) with domains 1 and 2 exhibiting the least similarity to other DnaA proteins while domains 3 and 4, which contain the ATP and DNA binding motifs, respectively, are highly conserved. However, it is unique among DnaA proteins in the number of basic lysine residues found surrounding a highly conserved stretch of amino acids, GRDHTTV, located in DNA-binding domain IV (Fig. 2). One amino acid within this region of *E. coli* DnaA, Thr 435, has been shown to be involved in sequence-specific DNA binding activity (Sutton and Kagun, 1997).

Gene organization surrounding R. prowazekii dnaA

In order to determine if the gene organization surrounding the *R. prowazekii dnaA* gene was similar to that observed in other characterized bacteria, additional clones hybridizing to the rickettsial *dnaA* PCR amplified probe were identified in a Lambda ZapII clone bank. DNA sequence analysis of these clones resulted in additional upstream sequence (Fig. 1). Interestingly, no clones containing downstream sequence were obtained. This was also true for the original *dnaA* clone, pMW833. Restriction mapping revealed that the rickettsial insert of pMW833 was only 2.0 kb rather than the expected 5 – 8 kb and that one of the predicted terminal *Hin*dIII sites of the insert had been deleted. This suggests that there is a sequence downstream of rickettsial *dnaA* gene that cannot be cloned in this *E. coli* system. Using the sequence data presented in this paper and additional upstream sequence obtained form their *R. prowazekii* genome sequencing project, Andersson and Andersson (1997) recently identified the *mfd* gene analog (Fig. 1). Our data identify an additional ORF of 123 aa just upstream of *dnaA* gene. However, this ORF shows no significant homology to any known gene product. This information demonstrates that the gene organization of the *R. prowazekii dnaA* region differs from that of other characterized *dnaA* genes.

Localization of the rickettsial dnaA transcriptional start site

The RNase protection assay was used to localize the start of the rickettsial *dnaA* transcript (Fig. 3). The antisense probe used in this assay spanned nt 567-1010 (444 rickettsia-specific nt) in Fig. 1. An approximately 260 nt long fragment of the probe was protected localizing the start of the transcript to a region 520 nt upstream of the DnaA start codon. The primer extension analysis (Fig. 4) identified A at positions 748 and 749 as the first bases of the detected *dnaA* RNA. Appropriately positioned upstream of these A are sequences (TATTAT and TTGCAA) homologous to *E. coli* consensus promoter -10 (TATAAT) and -35 (TTGACA) regions.

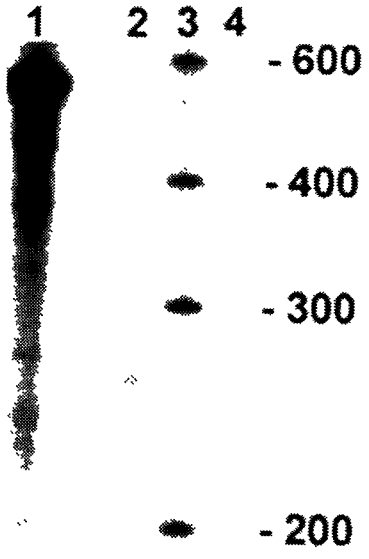

Fig. 3
RNase protection assay identifying the rickettsial *dnaA* transcript
Lane 1: undigested full-length probe; lane 2: total RNA from rickettsia-infected L929 cells; lane 3: single-stranded RNA ladder with sizes indicated; lane 4: RNA from mock-infected L929 cells.

Fig. 4
Primer extension assay of the *dnaA* transcript
Oligonucleotide DW96 was used as the extending primer. DNA sequencing reactions, for molecular size markers, are shown on the left. This sequence was generated using single-stranded M13mp18 as template and the -40 universal primer. The two T migrating at the same positions as the primer extension products correspond to sizes of 95 and 96 nt. This would place the start of transcription for the rickettsial RNA at the two nt indicated in Fig. 1.

Discussion

While certainly not the slowest growing bacterium, the *R. prowazekii* generation time of 8 – 12 hrs is surprising considering the cytoplasmic environment in which this organism grows. Presumably this slow growth has evolved to maximize the number of rickettsiae produced from each host cell (Winkler, 1995). Whatever may be the reasons for this extended generation time, the timing of DNA replication remains a critical process in rickettsial growth. The point of control for this process in other prokaryotes is at the initiation of chromosomal replication which involves the initiator protein DnaA and its interaction with the *ori*. Thus, we focused our initial studies on the isolation and characterization of the rickettsial *dnaA* gene analog.

Isolation of the *R. prowazekii dnaA* gene revealed a gene organization unlike any previously described. Gene organization surrounding the *dnaA* gene in other bacteria is highly conserved and usually includes *ori* C (Ogasawara and Yoshikawa, 1992). However, ORFs linked to the rickettsial *dnaA* gene do not resemble any gene products associated with the *dnaA* gene in other bacteria. Unfortunately, comparisons could be made primarily only for sequences upstream of rickettsial *dnaA* gene. We were unable to isolate clones containing the region immediately downstream of the gene suggesting that this region is lethal when introduced into *E. coli*. Additional efforts are underway to isolate downstream fragments and identify the genes present in this region.

Isolation of the rickettsial *dnaA* gene also permitted the comparison of the deduced product to that of characterized initiator proteins. DnaA is a highly conserved protein especially in the carboxy-terminal half of the molecule (Skarstad and Boye, 1994; Messer and Weigel, 1996). Sequences associated with ATP and DNA binding are easily identified. However, the rickettsial DNA binding sequence exhibited some unique characteristic that may affect the binding of this protein to DNA. The large number of positively charged lysine residues within one of the highly conserved areas of this region is especially striking. This raises the possibility that rickettsial DnaA may recognize a DNA sequence different from the consensus DnaA-box sequence identified for other characterized DnaA proteins. Thus, examining the rickettsial genome for consensus DnaA binding sites as a method for identifying DnaA-regulated genes or the *ori* may not be productive. Certainly, no consensus DnaA boxes, which are commonly found upstream of many *dnaA* genes, could be identified upstream of the rickettsial gene. In addition, we have been unsuccessful in isolating an autonomous replicating sequence of the rickettsial genome that can replicate in *E. coli* suggesting that the rickettsial *ori* C differs, possibly in its binding of *E. coli* DnaA, from the *E. coli ori* C. The isolation of the rickettsial *dnaA* gene now provides a means for expressing and purifying sufficient quantities of this protein or appropriate peptides for use in binding assays that would identify rickettsial target sequences. Such studies are essential for the isolation and characterization of the *R. prowazekii* origin of replication and subsequent understanding of the critical replication events occuring during intracytoplasmic growth of this unique bacterium.

Acknowledgements. The excellent technical assistance of Ms. A. Hines is gratefully acknowledged. This work was supported by Public Health Service grant No. AI20384 from the National Institute of Allergy and Infectious Disease.

References

Andersson JO, Andersson SGE (1997): Genomic rearrangements during evolution of the obligate intracellular parasite *Rickettsia prowazekii* as inferred from an analysis of 52015 bp nucleotide sequence. *Microbiology* **143**, 2783–2795.

Andersson SGE, Sharp PM (1996): Codon usage and base composition in *Rickettsia prowazekii* *J. Mol. Evol.* **42**, 525–536.

Cai J, Winkler HH (1993): Identification of *tlc* and *gltA* mRNAs and determination of in situ RNA half-life in *Rickettsia prowazekii*. *J. Bacteriol.* **175**, 5725–5727.

Cai J, Pang H, Wood DO, Winkler HH (1995): The citrate synthase-encoding gene of *R. prowazekii* is controlled by two promoters. *Gene* **163**, 115–119.

Dasch GA, Weiss E (1977): Characterization of the Madrid E strain of *Rickettsia prowazekii* purified by renografin density gradient centrifugation. *Infect. Immun* **15**, 280–286.

Davis RH, Bostein D, Roth JR (1980): *Advanced Bacterial Genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Devereux J, Haeberli P, Smithies O (1984): A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387–395.

Emory SA, Belasco JG (1990): The *ompA* 5' untranslated RNA segment functions in *Escherichia coli* as a growth-rate-regulated mRNA stabilizer whose activity is unrelated to translational efficiency. *J. Bacteriol.* **172**, 4472–4481.

Hanson BA, Wisseman CL,Jr, Waddel A, Silverman DJ (1981): Some characteristics of heavy and light bands of *Rickettsia prowazekii* on renografin gradients. *Infect. Immun.* **34**, 596–604.

Margolin W, Bramhill D, Long SR (1995): The *dnaA* gene of *Rhizobium meliloti* lies wihin an unusual gene arrangement. *J. Bacteriol.* **177**, 2892–2900.

Marmur J (1961): A procedure for the isolation of deoxyribonucleic acid from microorganisms. *J Mol Biol.* **3**, 208–218.

Messer W, Weigel C (1996): Initiation of chromosome replication. In Neidhardt FC, Curtiss RIII, Ingraham J et al. (Eds): *Escherichia coli and Salmonella. Cellular and Molecular Biology*. ASM Press, Washington, DC.

Messer E, Weigel C (1997): DnaA initiator-also a transcription factor. *Mol Microbiol.* **24**, 1–6.

Ogasawara N, Yoshikawa H (1992): Genes and their organization in the replication origin region of the bacterial chromosome. *Mol. Microbiol.* **6**, 629–634.

Ogasawara N, Moriya S, Yoshikawa H (1991): Initiation of chromosome replication: structure and function of *oriC* and DnaA protein in eubacteria. *Res. Microbiol.* **142**, 851–859.

Old IG, Margarita D, Saint Girons I (1993): Unique genetic arrangement in the *dnaA* region of the *Borrelia burgdorferi* linear chromosome: Nucleotide sequence of the *dnaA* gene. *FEMS Microbiol. Lett.* **111**, 109–114.

Plano GV, Winkler HH (1989): Solubilization and reconstitution of the *Rickettsia prowazekii* ATP/ADP translocase. *J. Membrane Biol.* **110**, 227–233.

Sambrook J, Fritsch EF, Maniatis T (1989): *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanger F, Nicklen S, Coulson AR (1977): DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.

Shine J, Dalgarno L (1974): The 3' terminal sequence of *E. coli* 16S ribosomal RNA: complementary to nonsense triplets and ribosomal binding sites. *Proc. Natl. Acad. Sci. USA* **71**, 1342–1346.

Skarstad K, Boye E (1994): The initiator protein *DnaA*: evolution, properties and function. *Biochim. Biophys. Acta* **127**, 111–130.

Skovgaard O, Hansen FG (1987): Comparison of *dnaA* nucleotide sequences of *Escherichia coli*, *Salmonella typhimurium*, and *Serratia mercescens* *J. Bacteriol.* **169**, 3976–3981.

Sutton MD, Kagun JM (1997): Threonine 435 of *Escherichia coli* DnaA protein confers sequence-specific DNA binding activity. *J. Biol. Chem.* **272**, 23017–23024.

Weiss E (1982): The biology of rickettsiae. *Ann. Rev. Microbiol* **36**, 345–370.

Winkler HH (1976): Rickettsial permeability: an ADP-ATP transport system. *J. Biol Chem.* **251**, 389–396.

Winkler HH (1990): *Rickettsia* species (as organisms). *Ann. Rev. Microbiol.* 44, 131–153.

Winkler HH (1995): *Rickettsia prowazekii*, ribosomes and slow growth. *Trend Microbiol.* **3**, 196–198.

Winkler HH, Wood DO (1988): Codon usage in selected AT-rich bacteria. *Biochimie* **70**, 977–986.

Wood DO, Waite RT (1994): Sequence analysis of the *Rickettsia prowazekii gyrA* gene. *Gene* **151**, 191–196.