Microbiology (1998), **144**, 433–439 Printed in Great Britain

Organization around the *dnaA* gene of *Streptococcus pneumoniae*

Anne-Marie Gasc,[1] Philippe Giammarinaro,[1] Stefan Richter[2]† and Michel Sicard[1]

Author for correspondence: Anne-Marie Gasc. Tel: +33 05 61 33 59 71. Fax: +33 05 61 33 58 86.
e-mail: gasc@ibcg.biotoul.fr

[1] Laboratoire de Microbiologie et de Génétique Moléculaire du CNRS, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France

[2] Max-Planck-Institut für Molekulare Genetik, Ihnestr. 73, D-14195 Berlin, Germany

The *dnaA* gene region of *Streptococcus pneumoniae* was cloned and sequenced. A tRNA gene, seven ORFs and three DnaA box clusters were identified. The order of the genes and intergene regions found was tRNA^{Arg}-*orf1*-DnaA box cluster 3-*htrA*-*spo0J*-DnaA box cluster 2-*dnaA*-DnaA box cluster 1-*dnaN*-*orfX*-*orfY*. Five ORFs are homologous to known bacterial genes. The tRNA^{Arg} gene and *orf1*, also called *orfL*, have already been described in pneumococci and have been reported to be preceded by the competence regulation locus *comCDE*. In *Escherichia coli*, *htrA* encodes a serine protease. In *Bacillus subtilis*, *spo0J* plays a role in sporulation and partition. *dnaA* encodes an initiator replication protein, very well conserved in several bacteria and *dnaN* encodes the β subunit of DNA polymerase III in *E. coli*. The function of *orfX* is unknown. The N-terminal part of another reading frame, *orfY*, revealed high homology with a GTP-binding protein. DnaA box clusters were found upstream and downstream from *dnaA*. The presence of two such clusters suggests that the chromosomal origin of *S. pneumoniae* is located within this region. The position of *dnaA*, and therefore the putative origin of replication, were localized on the physical map of *S. pneumoniae*.

Keywords: *Streptococcus pneumoniae*, *dnaA*, DnaA boxes, *dnaN*, *htrA*, replication origin

INTRODUCTION

The DnaA protein is the essential initiator of chromosome replication in eubacteria. It recognizes the replication origin by binding to 9 bp sequences known as DnaA boxes with the consensus sequence TTAT(C/A)CACA (Ogasawara *et al.*, 1990; Ogasawara & Yoshikawa, 1992; Fujita *et al.*, 1990, 1992; Schaper & Messer, 1995). The interaction of the DnaA protein with the DnaA box plays a central role in the initiation of chromosome replication. Different organization around *dnaA* genes has been reported in several species (Zweiger & Shapiro, 1994; Richter & Messer, 1995). However, at least six genes, *rnpA*, *rpmH*, *dnaA*, *dnaN*, *recF* and *gyrB*, are conserved in *Escherichia coli* and *Bacillus subtilis* (Ogasawara *et al.*, 1985). Even *Mycoplasma capricolum*, a species with a genome size of approximately 700 kb, has at least two conserved genes *dnaA*:*rpmH* and *dnaN* (Ogasawara & Yoshikawa, 1992). DnaA boxes are located near *dnaA* in several bacteria or elsewhere in bacteria such as *E. coli* or *Synechococcus* sp. (Liu & Tsinoremas, 1996).

Streptococcus pneumoniae is a pathogenic bacterium responsible for a high level of mortality in patients of all ages. Therefore, it is important to know how its replication is regulated. We have constructed a physical map of the *S. pneumoniae* chromosome showing that it is circular and 2·2 Mb in size (Gasc *et al.*, 1991). However, the origin of replication has not yet been determined. The aim of this work was to clone and sequence the *dnaA* gene and the neighbouring genes in this bacterium. The *dnaA* region was localized on the physical map of the chromosome. The sequence of approximately 6000 bp around the *dnaA* gene is presented showing seven ORFs, including *dnaA* and the tRNA^{Arg} gene. A comparison of the predicted amino acid sequences from these ORFs with known gene products from other species is given. Several DnaA

Abbreviation: SD, Shine–Dalgarno.

The GenBank accession number for the sequence reported in this paper is AF000658.

boxes were identified downstream and upstream from *dnaA*, defining a putative replication origin.

METHODS

Bacterial strains, growth conditions and DNA preparation. The pneumococcal strain used in this study is a derivative of R6, called 801 (Lefèvre *et al.*, 1979). The culture media have been described elsewhere (Claverys *et al.*, 1980). Cultures were incubated at 37 °C without aeration. Preparation of DNA and transformation procedures have been described previously (Martin *et al.*, 1985; Gasc *et al.*, 1995).

Determination of the nucleotide sequence. The fragments used are described in the text. The PCR fragments were obtained using the Tub polymerase from Amersham by 25

Table 1. Primers used to sequence the insert

The primers in the upper part are localized on one strand; the primers in the lower part are on the complementary strand. T45, B39 and B40 were synthesized using fragment A containing part of the *dnaA* gene. T31 and T70 were synthesized using the sequence of a λ Zap clone determined from T3 or T7 universal primers.

Primer	Sequence (5′–3′)	Localization
T31	GGG GAA TCC CTG CTG AAT CGT AAA	127
T32	CCT TAT CTG GTG TTT TTT CAT C	456
T33	CAC AAT TTC AAA AGA GTT ATC CA	670
T33a	CAG TAA TTC AAA CAT ATG GAG GC	750
T34	GGC TGT TAA CAA AGT AAA AGA TGC TG	965
T35	CTT TCT CTG ATA TTG CTG TCG CC	1222
T36	CCT GCA AAT GAT GCT ATC	1581
T37	GCT TTG GGA ATC CAG ATG G	1644
T38	CGG GAA AGA AGA AAC TAC CTC	1898
T39	CAA GAG ATG ATG GTC CCA GTC C	2265
T40	CAA CGG ATT ATG AAG AAG AT	2541
T41	CTA TCT AAA AAA GGC AG	2691
T42	CTT GCT ATC TAT GGT AAA ATA TC	2924
T43	ACA AGG TGA AGA AGC TAC AA	3250
T44	AAC GCT ATT GGA AAT GAA ATT C	3482
T45	AAC AGA TTG TCC TAA CGA GTG ATC G	3723
T46	GCC CGT CAA GTA GCC ATG TA	4163
T47	CCA CAG ATT TCA CAG ACT C	4439
T48	GGT TCG ATC CTT CTT GAA G	4723
T49	GCC GTT CTC TAC GCG AAT TTT CA	5120
T50	GGT TAC TGG TGA AGA TTT G	5463
T51	GGT AAA AAG GCT AAT CGT TGG	5771
T70	AAT AGT CGC AAA TGG GTA G	6096
T71	CAT ATC GCC CCA TCA TGA CAA C	5867
T72	GTA AAT GGA CGA ACA GCT GAG	5579
T73	GTT GGA ATC AAG CGA TCT G	5279
B36	GGG TAA TTT CTG ATT TGC CAC TGG	4849
B37	GTA ATA TTT AAT GCT TGT AG	4559
B38	CCC CCA AAT TCC TTC CC	4239
B39	AAT ACA TGG CTA CTT GAC GGG CC	4184
B40	CGT GAC AAG CCT CTC CTC GAG CCC	3790
B41	AGT CTT ACC AAG GCC TGG	3472
B42	GTG GGG TGT TAT TTC AGC G	3214
B43	CAC AGG TTG GGG ATA AAA GAA GA	2873
B44	GCT GTT TCT TTT GTT TTT TCT	2616
B45	GGG TGA ATC GGC CTC TAC GAG	2371
B46	CAC ATC TTT TCT CTT ATG TAA AGC	2002
B47	GAA TAT TGA GTC TTC TGA TGT CGC	1710
B48	CAT ATT CAG AAC CTA ACG GGC	1353
B49	CAG TAT CAT CAT TCC CAA AGA CGC	1053
B50	GCC TCC ATA TGT TTG AAT TAC TG	772
B51	GGT ATC GCA GAG TAT TCC AAA ACG	540

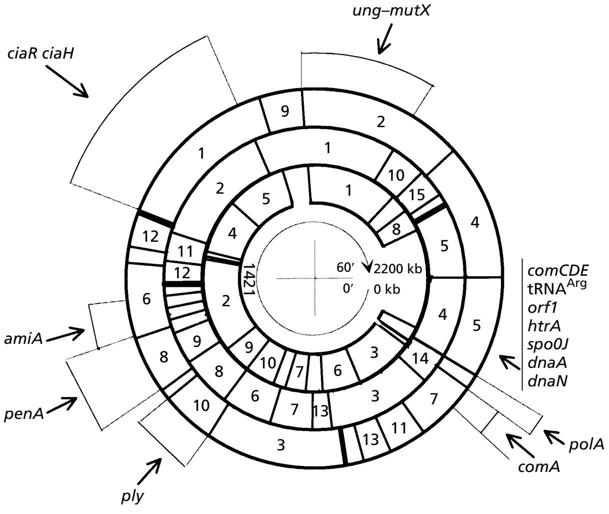

Fig. 1. Position of the *dnaA* region in the physical map of the *S. pneumoniae* R6 chromosome. The outer circle represents *Sma*I fragments, the middle circle *Apa*I fragments and the inner circle *Sac*II fragments. Spaces between fragments indicate contiguity has not been established.

cycles of 30 s at 94 °C, 2 min at 56 °C and 1 min at 72 °C following the manufacturer's recommendations. The fragments were purified using the Wizard Purification Kit (Promega). Sequencing was done using the Circumvent Thermal Cycle Sequencing Kit (Biolabs) or performed at Euro Sequence Genes Services, Evry, France, on an ABI377 Sequencer using the PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA polymerase (Applied Biosystems). Oligonucleotide primers and their positions are listed in Table 1.

Electrophoresis and hybridization. outhern blot hybridizations (Southern, 1975) were performed essentially as described by Sambrook *et al.* (1989). Field inversion gel electrophoresis has been described by Kauc *et al.* (1989). Fragments were labelled by the Megaprime Labelling Kit (Amersham).

RESULTS AND DISCUSSION

Cloning and sequencing of the *dnaA* region

Degenerate oligonucleotide primers were deduced from two highly conserved regions of the DnaA protein (foreward primer 5′ GARGARTTYTTYCAYACNTTYAAY; reverse primer 5′ ACWGTHGTRTGRTCYCTWCCWCC). PCR using the degenerate primers and chromosomal DNA of *S. pneumoniae* was carried out as described previously (Richter & Messer, 1995). A PCR fragment of 531 bp (fragment A) was obtained and sequenced. The deduced amino acid sequence showed a high degree of homology to the corresponding region of known DnaA proteins, suggesting that this fragment belongs to the pneumococcal *dnaA* gene. Fragment A, labelled with [α-^{32}P]dATP, was hybridized to *Sma*I, *Apa*I and *Sac*II fragments of the *S. pneumoniae* chromosome, separated by field inversion gel electrophoresis. Fig. 1 shows the position of the *dnaA* gene on the previously published physical map of *S. pneumoniae* (Gasc *et al.*, 1991).

To clone the region around the *dnaA* gene, we screened a *S. pneumoniae* DNA λ Zap (Short *et al.*, 1988) library constructed by B. Martin (this laboratory) for the presence of fragment A. We selected a clone (λ ZapA) that hybridized with fragment A. To determine the size of the cloned fragment, λ ZapA DNA was digested with *Bam*HI and *Sma*I which cut the polylinker on both sides of the fragment. A 6 kb fragment was obtained. To localize fragment A on the 6 kb fragment, we generated PCR products using the universal primers T3 or T7 and one of the three primers T45, B39 or B40 (Table 1) synthesized using the pneumococcal fragment A sequence. The localization of fragment A on the insert indicated that it carries sequences both downstream and upstream from the *dnaA* gene.

To determine the start point of the pneumococcal DNA sequence at both ends of the insert, we sequenced the PCR products prepared as described above. We synthesized new oligonucleotides (T31 and T70, Table 1)

Table 2. Genes of the insert

Gene	Localization of the gene	Size of the protein (aa)	Putative promoter		Putative SD sequence*
Consensus			TTGACA	TnTGnTATAAT	UAAGGAGGUG
tRNA^{Arg}	72–1				
orf1 or *orfL*	594–115				
htrA	765–1955	397	gTGtaA	TaaGtTATAAT	GGAGG (5)
spo0J	2016–2771	252	TTGACA	aagctTtacAT	GTG (7)
dnaA	2987–4345	453	TgGAaA	TaTGgTAaAAT	AAGGAGG (8)
dnaN	4507–5640	378	TTcACA	acTat ATctT	AAGGAG (5)
orfX	5707–5899	64	TTGAgc	TaTGaTATAAT	AAGGAG (6)
orfY	5986–6098 insert end	Not known	TTagCA	TgTGtTATAAT	GGAG (10)

* The number of nucleotides between the SD sequence and the initiation codon is indicated in parentheses.

```
                                                   Domain I >
E. coli          MS--LSLWQQCLARLQDELPATEFSMWIRPLQAE-LSDNTLALYAPNRFV    47
P. putida        MS--VELWQQCVELLRDELPAQQFNTWIRPLQVE-AEGDELRVYAPNRFV    47
B. subtilis      MENILDLWNQALAQIEKKLSKPSFETWMKSTKAHSLQGDTLTITAPNEFA    50
S. pneumoniae    LKEK-QFWNRILEFAQERLTRSMYDFYAIQAELIKVEENVATIFLP-RSE    48
                 ..    .*.. ..   .*.   .. .    .     ..   .    *
                      <Domain II
E. coli          LDWVRDKYLNNINGL--LTSFCGRIAPQLRFEVGTKPVTQ---TPQAAVT    92
P. putida        LDWVNEKYLGRLLEL--LGENGSGIAPALSLLIGSRRSSAPRAAPNAPVS    95
B. subtilis      RDWLESRYLHLIADT--IYELTGE-ELSIKFVI-----------PQNQDV    86
S. pneumoniae    MEMVWEKQLKDIIVVAGFEIYDAEITPHYIF-------------------    79
                  .  . .. *  .     .     .   .    .

E. coli          SNVAAPAQVAQTQPQRAAP-------------------STRSGWDNVPAP   123
P. putida        AAVAAS--LAQTQAHKTAPAAAVEPVAVAAAEPVLVETSSRDSFDAMAEP   143
B. subtilis      EDFMPKPQVKKA--------------------------------------    98
S. pneumoniae    -----------TKPQDTTSSQGEEATNLTL------------YDYSPKL   105
                    Domain II >  <Domain III
E. coli          A-----------EPTYRSNVNVKHT--------FDNFVEGKSNQLARAAA   154
P. putida        AAAPPSGGGRAEQRTVQVEGALKHTSYLNRTFTFDTFVEGKSNQLARAAA   193
B. subtilis      --------VKEDTSDFPQNM-------LNPKYTFDTFVIGSGNRFAHAAS   133
S. pneumoniae    VSIPYSDTG-----------------LKEKYTFDNFIQGDGNVWAVSAA   137
                                                **.*. *..*  * .*.
                                   ATP binding
E. coli          RQVADNPGGAYNPLFLYGGTGLGKTHLLHAVGNGIMARKPNAKVVYMHSE   204
P. putida        WQVADNPKHGYNPLFLYGGVGLGKTHLMHAVGNHLLKKNPNAKVVYLHSE   243
B. subtilis      LAVAEAPAKAYNPLFIYGGVGLGKTHLMHAIGHYVIDHNPSAKVVYLSSE   183
S. pneumoniae    LAVSEDLALTYNPLFIYGGPGLGKTHLLNAIGNEILKNIPNARVKYIPAE   187
                  .*...   .*****.*** *******..*.*.  ..  . *.*.* *.  .*

E. coli          RFVQDMVKALQNNAIEEFKRYYRSVDALLIDDIQFFANKE-RSQEEFFHT   253
P. putida        RFVADMVKALQLNAINEFKRFYRSVDALLIDDIQFFARKE-RSQEEFFHT   292
B. subtilis      KFTNEFINSIRDNKAVDFRNRYRNVDVLLIDDIQFLAGKE-QTQEEFFHT   232
S. pneumoniae    SFINDFLDHLRLGEMEKFKKTYRSLDLLLIDDIOSLSGKKVATOEEFFNT   237
                 .*...... .. .    .*.. **..* ******* .. *.   .*****.*

E. coli          FNALLEGNQQIILTSDRYPKEINGVEDRLKSRFGWGLTVAIEPPELETRV   303
P. putida        FNALLEGGQQVILTSDRYPKEIEGLEERLKSRFGWGLTVAVEPPELETRV   342
B. subtilis      FNTLHEESKQIVISSDRPPKEIPTLEDRLRSRFEWGLITDITPPDLETRI   282
S. pneumoniae    FNALHDKQKQIVLTSDRSPKHLEGLEERLVTRFSWGLTQTITPPDFETRI   287
                 **.* .  .*....*** **..  ..*.** .**.***.  ...**..***.

E. coli          AILMKKADENDIRLPGEVAFFIAKRLRSNVRELEGALNRVIANANFTG-R   352
P. putida        AILMKKADQAKVELPHDAAFFIAQRIRSNVRELEGALKRVIAHSHFMG-R   391
B. subtilis      AILRKKAKAEGLDIPNEVMLYIANQIDSNIRELEGALIRVVAYSSLIN-K   331
S. pneumoniae    AILQSKTEHLGYNFQSDTLEYLAGQFDSNVRDLEGAINDITLIARVKKIK   337
                 *** .*..      .. ..   ..* .. **.*.****.   ..  .    .
                 Domain III >  < Domain IV = DNA-binding domain
E. coli          AITIDFVREALRDLLAL-QEKL-VTIDNIQKTVAEYYKIKVADLLSKRRS   400
P. putida        DITIELIRESLKDLLAL-QDKL-VSVDNIQRTVAEYYKIKISDLLSKRRS   439
B. subtilis      DINADLAAEALKDIIPSSKPKV-ITIKEIQRVVGQQFNIKLEDFKAKKRT   380
S. pneumoniae    DITIDIAAEAIRARKQDVSQMLVIPIDKIQTEVGNFYGVSIKEMKGSRRL   387
                 .*. ..  *....  .     .. .....**  *.. . ... .. ...*

E. coli          RSVARPRQMAMALAKELTNHSLPEIGDAFGGRDHTTVLHACRKIEQLREE   450
P. putida        RSVARPRQVAMALSKELTNHSLPEIGDMFGGRDHTTVLHACRKINELKES   489
B. subtilis      KSVAFPRQIAMYLSREMTDSSLPKIGEEFGGRDHTTVIHAHEKISKLLAD   430
S. pneumoniae    QNIVLARQVAMYLSRELTDNSLPKIGKEFGGKDHTTVIHAHAKIKSLIDQ   437
                 ....  .**.** *..*.*. ***.**. ***.*****.**  **. *  .

E. coli          SHDIKEDFSNLIRTLSS   467
P. putida        DADIREDYKNLLRTLTT   506
B. subtilis      DEQLQQHVKEIKEQLK-   446
S. pneumoniae    DDNLRLEIESIKKKIK-   453
                 . ... . ...   ..
```

Fig. 2. Sequence alignment of four DnaA proteins from *E. coli, P. putida, B. subtilis* and *S. pneumoniae*. The alignment was done using PCGENE. Division in domains I–IV was defined according to Yoshikawa & Ogasawara (1991). The ATP-binding site and DNA-binding domain are shown. Asterisks indicate identical amino acids; dots indicate similar amino acids.

based on pneumococcal DNA sequences. The next step was PCR amplification of pneumococcal DNA from strain 801 using T31 and B39 primers to generate a 4 kb fragment. Another PCR amplification using T45 and T70 generated a 2·2 kb fragment. The PCR sequences were then determined.

Sequence analysis

Putative ORF, promoter and Shine–Dalgarno (SD) sequences. Amino acid sequences were deduced from the nucleotide sequences. We found seven putative ORFs on the λ ZapA insert (Table 2). A comparison with database sequences revealed some homologies. The *S. pneumoniae* genes were named according to their similarity to known genes. In the 6 kb fragment, six putative promoters were defined using the criteria proposed for *S. pneumoniae* by Sabelnikov *et al.* (1995) (Table 2). The prokaryotic mRNA ribosome-binding site (RBS) usually contains part or all of the SD sequence UAAGGAGGU (Shine & Dalgarno, 1974) and the spacing between the SD sequence and the initiation codon varies from 5 to 10 nt, as usually observed (Singer *et al.*, 1981; Chen *et al.*, 1994). According to these criteria, six putative SD sequences were proposed (Table 2).

***dnaA* gene.** The complete sequence of the DnaA protein is known for several bacteria. The size of the DnaA proteins varies from 437 to 656 aa. The length of the *S. pneumoniae* deduced DnaA protein is 453 aa (Table 2, Fig. 2).

The amino acid sequences of four DnaA proteins were compared (Fig. 2). The level of identity between *S. pneumoniae* DnaA and that of *B. subtilis*, another Gram-positive bacterium, is the highest (38·8 %). The level of identity with the DnaA proteins of *Pseudomonas putida* and *E. coli* is 33·3 and 32·5 %, respectively. DnaA proteins were divided into four domains with different degrees of conservation (Fujita *et al.*, 1992). These domains are conserved in *S. pneumoniae*. The N-terminal domain I, with moderate conservation, is followed by domain II, variable in sequence and length. The highly conserved domain III contains an ATP-binding site. The C-terminal domain IV, the DNA-binding domain (Yoshikawa & Ogasawara, 1991; Roth & Messer, 1995) is also highly conserved. A TTG codon was identified near the ribosome-binding site and it seems the best candidate to be the start codon. The secondary structure was also conserved compared to the DnaA protein predicted by Schaper & Messer (1997). DnaA from *S. pneumoniae* could therefore be included among the other proteins of the DnaA family.

The unusual initiation codon, TTG, has also been found in *E. coli* and in *B. subtilis* (Ogasawara *et al.*, 1994) where it is the third most frequent initiation codon after ATG and GTG. It is noteworthy that it has been recently reported that the start codon for *dnaA* genes in *Mycobacterium smegmatis*, *M. leprae* and *M. tuberculosis* is also TTG (Salazar *et al.*, 1996).

DnaA boxes. The DnaA protein binds specifically to the DnaA box TTAT(C/A)CACA and closely related sequences (Yoshikawa & Ogasawara, 1991; Fuller *et al.*, 1984; Schaefer & Messer, 1991; Fujita *et al.*, 1992). Interaction of DnaA with the DnaA boxes of the origin forms the initiation complex. DnaA boxes have been identified in the intragenic region flanking *dnaA* in most bacteria studied. In *S. pneumoniae* we found six DnaA boxes upstream (cluster 2) and three downstream (cluster 1) of *dnaA*. This seems to be a common feature of the chromosomal origins in Gram-positive bacteria (Ogasawara *et al.*, 1985; Fujita *et al.*, 1989, 1990). Four other DnaA boxes were localized upstream from *htrA* in a non-translatable region (cluster 3). All putative DnaA

Table 3. DnaA box localization in the insert

Cluster 3 is localized upstream from the *htrA* gene. Clusters 2 and 1 are upstream and downstream from *dnaA*, respectively. * refers to DnaA boxes on the complementary strand. Consensus sequences were TTATC/ACACA on one strand and TGTGG/TATAA on the complementary strand. Lower case letters indicate a deviation from these consensus sequences.

Position	DnaA box sequence	Cluster
631	TTATtCACA	3
686	TTATCCACA	3
698*	TGTGTAaAA	3
711*	TGTGTtTAA	3
2755	TTATCaACA	2
2806	TTATCCACt	2
2858	TTATCCcCA	2
2870*	TGTGGATAA	2
2892*	TGTGGATtA	2
2913*	TGTGGAaAA	2
4351*	TGTGGATAA	1
4380	TTATCCACA	1
4435	TTtTCCACA	1

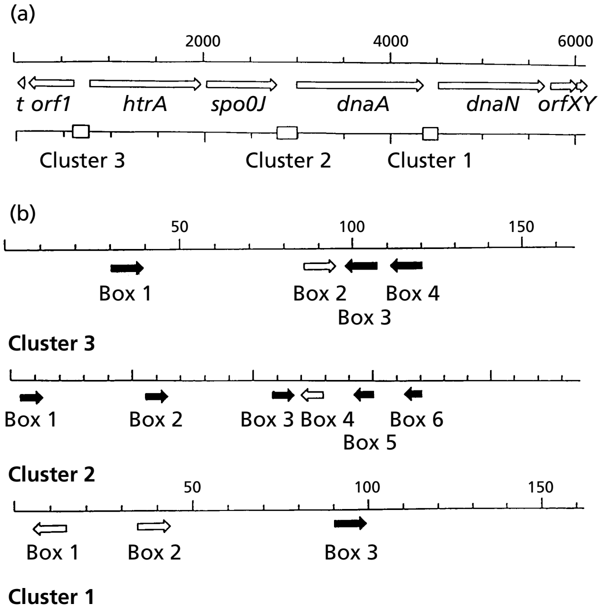

Fig. 3. Gene organization of the *dnaA* region of *S. pneumoniae* strain R801. (a) Schematic representation of the ORFs deduced from the nucleotide sequence. The arrows indicate the 5′–3′ direction. The open triangle (*t*) represents the $tRNA^{Arg}$ gene. Boxes (on the second line) between arrows indicate non-translatable regions containing clusters of multiple repeats of DnaA boxes. (b) Clusters of DnaA boxes. White arrows refer to consensus sequences TTAT(A/C)CACA or TGTG(G/T)ATAA; black arrows refer to DnaA box sequences that differ by one base from the consensus.

boxes are listed in Table 3. In *B. subtilis* a DnaA box region is also located about 3 kb upstream from *dnaA*. Another common feature of bacterial replication origins are AT-rich regions adjacent to DnaA boxes. These regions have been identified. This argues for the localization of the replication origin of *S. pneumoniae* near any one of the three clusters of DnaA boxes.

Gene organization around *dnaA*

Fig. 3 shows the genes and their organization in the 6 kb fragment containing the *dnaA* gene. The first nucleotides on the left of this insert (1–71) are identical to a pneumococcal sequence reported by Pestova *et al.* (1996) who designated it as a $tRNA^{Arg}$ gene. Only the last 2 nt of the gene are missing. This gene is followed by *orfL*, also named *orf1*. Interestingly *orfL* and $tRNA^{Arg}$ are preceded by the competence locus *comCDE* (Pestova *et al.*, 1996; Cheng *et al.*, 1997).

The gene immediately to the right of *orfL* is a gene homologous to *htrA* of *E. coli* encoding a serine protease (Lipinska *et al.*, 1989; Skorko-Glonek *et al.*, 1995). The active site of this protease contains a serine residue, a histidine residue and an aspartate residue. The homology is especially striking around the active site (Bass *et al.*, 1996).

Between *htrA* and *dnaA*, separated by 216 bp, where DnaA boxes are found, a gene homologous to *spo0J* (Mysliwiec *et al.*, 1991) has been identified. The identity between *spo0J* of *B. subtilis* and *S. pneumoniae* is 42·9 %. This gene is required for the initiation of sporulation and for normal chromosome partitioning during vegetative growth in *B. subtilis* (Mysliwiec *et al.*, 1991; Ireton *et al.*, 1994; Lin *et al.*, 1997; Glaser *et al.*, 1997). It is likely that the *spo0J* product of *S. pneumoniae*, being a non-sporulating bacterium, is required for accurate chromosome partitioning. In *B. subtilis*, *spo0J* is also upstream from *dnaA* but separated by 10 genes representing about 10 kb. Another divergence is the direction of translation. In *S. pneumoniae dnaA* and *spo0J* have the same orientation, whereas they are opposite in *B. subtilis*.

A gene homologous to *dnaN* has been found downstream from *dnaA* and DnaA box cluster 1. The identity between *dnaN* of *B. subtilis* and *S. pneumoniae* is 42·3 %. The gene encodes the β subunit of PolIII, a component of the core enzyme. This organization of *dnaA*, DnaA boxes and *dnaN* has been reported for several bacteria and the replication origin is often located upstream from this gene.

No significant homology was found by searching the databases with the amino acid sequence of the *orfX* product. For *orfY* only a part of the sequence is known. The N-terminal sequence of its product is highly homologous to the corresponding part of a putative GTP-binding protein of *B. subtilis* called YyaF. In this organism the corresponding gene is localized upstream from *spo0J* and separated by about 5·7 kb containing three unknown proteins.

Conclusions

In this study we have sequenced 6 kb of the *S. pneumoniae* genome that includes the *dnaA* gene. We have found that this gene is surrounded by several DNA boxes and is adjacent to the *dnaN* gene. As this organization is conserved in several bacteria (*B. subtilis*, *Streptomyces coelicolor*, *P. putida* and *M. capricolum*), we suggest that the origin of replication of the pneumococcal chromosome is also in this region. Another argument for this proposal is that AT-rich regions have been observed in this fragment. Computer analysis shows that there are 14 sequences of 20–33 nt containing more than 90 % AT. They are dispersed in the fragment with a tendency to occur near DNA boxes. This is a common feature of a replication origin and might participate in the destabilization of the DNA helix.

However, organization around *dnaA* is different in some bacteria such as *Synechococcus* sp., where *dnaA* and *dnaN* are not adjacent (Richter & Messer, 1995). The replication origin has been located near 11 DNA boxes upstream from *dnaN* (Liu & Tsinoremas, 1996).

The organization of the genes around *dnaA* in *Streptococcus pneumoniae* shares some similarities with that of other bacteria but is itself unique.

Several genes were identified in this cloned fragment by their homology with known genes. A serine protease encoded by *htrA* has not been described previously in *S. pneumoniae*. Several reports suggest that serine proteases are important components of the pathogenic mechanisms of other intracellular micro-organisms (Johnson *et al.*, 1991; Braun Breton & Pereia da Silva, 1993; Cameron *et al.*, 1994). It is quite possible that this serine protease is involved in *S. pneumoniae* virulence.

The occurrence of *spo0J* in a non-sporulating bacterium such as *S. pneumoniae* suggests that the function of this gene, reported to act either in sporulation or in septation in *Bacillus subtilis*, is primarily involved in the latter in *S. pneumoniae*. The proximity of the competence operon to the putative origin is noteworthy but its significance is unknown.

Localization of the putative origin of replication in the physical and genetic map of the chromosome will be useful for further studies on the molecular genetics of this pathogenic bacterium.

ACKNOWLEDGEMENTS

We thank Sandy Lacks for discussions and suggestions about the SD and promoter sequences of *Streptococcus pneumoniae*, and Bernard Martin for the generous gift of the λ Zap bank.

REFERENCES

Bass, S., Gu, Q. & Christen, A. (1996). Multicopy suppressors of Prc mutant *Escherichia coli* include two HtrA (DegP) protease homologs (HhoAB), DksA, and a truncated R1pa. *J Bacteriol* **178**, 1154–1161.

Braun Breton, C. & Pereia da Silva, L. H. (1993). Malarial proteases and red blood cell invasion. *Parasitol Today* **9**, 92–96.

Cameron, R. M., Stevenson, K., Inglis, H. F., Klausen, J. & Sharp, J. M. (1994). Identification and characterization of a putative serine protease expressed *in vivo* by *Mycobacterium avium* subsp. *paratuberculosis*. *Microbiology* **140**, 1977–1982.

Chen, H., Bjerknes, M., Kumar, R. & Jay, E. (1994). Determination of the optimal aligned spacing between the Shine–Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res* **22**, 4953–4957.

Cheng, Q., Campbell, E. A., Naughton, A. M., Johnson, S. & Masure, H. R. (1997). The *com* locus controls genetic transformation in *Streptococcus pneumoniae*. *Mol Microbiol* **23**, 683–692.

Claverys, J. P., Roger, M. & Sicard, A. M. (1980). Excision and repair of mismatched base pairs in transformation of *Streptococcus pneumoniae*. *Mol Gen Genet* **178**, 191–201.

Fujita, M. Q., Yoshikawa, H. & Ogasawara, N. (1989). Structure of the *dnaA* region of *Pseudomonas putida* : conservation among three bacteria, *Bacillus subtilis*, *Escherichia coli* and *Pseudomonas putida*. *Mol Gen Genet* **215**, 381–387.

Fujita, M. Q., Yoshikawa, H. & Ogasawara, N. (1990). Structure of the *dnaA* region of *Micrococcus luteus*: conservation and variations among eubacteria. *Gene* **93**, 73–78.

Fujita, M. Q., Yoshikawa, H. & Ogasawara, N. (1992). Structure of the *dnaA* and *dnaA*-box region in the *Mycoplasma capricolum* chromosome: conservation and variations in the course of evolution. *Gene* **110**, 17–23.

Fuller, R. S., Funnel, B. E. & Kornberg, A. (1984). The DnaA protein complex with the *E. coli* chromosomal replication origin (*oriC*) and other DNA sites. *Cell* **38**, 889–890.

Gasc, A.-M., Kauc, L., Baraillé, P., Sicard, M. & Goodgal, S. (1991). Gene localization, size and physical map of the chromosome of *Streptococcus pneumoniae*. *J Bacteriol* **173**, 307–312.

Gasc, A.-M., Geslin, P. & Sicard, A. M. (1995). Relatedness of penicillin-resistant *Streptococcus pneumoniae* serogroup 9 strains from France and Spain. *Microbiology* **141**, 623–627.

Glaser, P., Sharpe, M. E., Raether, B., Perego, M., Ohlsen, K. & Errington, J. (1997). Dynamic, mitotic-like behavior of a bacterial protein required for accurate chromosome partitioning. *Genes Dev* **11**, 1160–1168.

Johnson, K., Charles, I., Dougan, G., Pickard, D., O'Gaora, P., Costa, G., Ali, T., Miller, I. & Hormaeche, C. (1991). The role of a stress-response protein in *Salmonella typhimurium* virulence. *Mol Microbiol* **5**, 401–407.

Ireton, K., Gunther, N. W., 4th & Grossman, A. D. (1994). *spo0J* is required for normal chromosome segregation as well as the initiation of sporulation in *Bacillus subtilis*. *J Bacteriol* **176**, 5320–5329.

Kauc, L., Mitchell, M. & Goodgal, S. H. (1989). Size and physical map of the chromosome of *Haemophilus influenzae*. *J Bacteriol* **171**, 2472–2479.

Lefèvre, J. C., Claverys, J. P. & Sicard, A. M. (1979). Donor deoxyribonucleic acid length and marker effect in pneumococcal transformation. *J Bacteriol* **138**, 80–86.

Lin, D. C. H., Levin, P. A. & Grossman, A. D. (1997). Bipolar localization of a chromosome partition protein in *Bacillus subtilis*. *Proc Natl Acad Sci USA* **94**, 4721–4726.

Lipinska, B., Fayet, O., Baird, L. & Georgopoulos, C. (1989). Identification, characterization, and mapping of the *Escherichia coli htrA* gene, whose product is essential for bacterial growth only at elevated temperatures. *J Bacteriol* **171**, 1574–1584.

Liu, Y. & Tsinoremas, N. F. (1996). An unusual gene arrangement for the putative chromosome replication origin and circadian expression of *dnaN* in *Synechococcus* sp. strain PC7942. *Gene* **172**, 105–109.

Martin, B., Prats, H. & Claverys, J. P. (1985). Cloning of the *hexA* mismatch repair of *Streptococcus pneumoniae* and identification of the product. *Gene* **34**, 293–303.

Mysliwiec, T. H., Errington, J., Vaidya, A. B. & Bramucci, M. G. (1991). The *Bacillus subtilis spo0J* gene: evidence for involvement in catabolite repression of sporulation. *J Bacteriol* **173**, 1911–1919.

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., von Meyenburg, K., Hansen, F. G. & Yoshikawa, H. (1985). Conservation of genes and their organization in the chromosomal replication origin region of *Bacillus subtilis* and *Escherichia coli*. *EMBO J* **4**, 3345–3350.

Ogasawara, N., Fujita, M. Q., Moriya, S., Fukuoka, T., Hirano, M. & Yoshikawa, H. (1990). Comparative anatomy of *oriC* of eubacteria. In *The Bacterial Chromosome*, pp. 287–298. Edited by M. Riley & K. Drlica. Washington, DC: American Society for Microbiology.

Ogasawara, N., Nakai, S. & Yoshikawa, H. (1994). Systematic sequencing of the 180 kilobase region of the *Bacillus subtilis* chromosome containing the replication origin. *DNA Res* **1**, 1–14.

Pestova, E. V., Havarstein, L. S. & Morrison, D. A. (1996). Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. *Mol Microbiol* **21**, 853–862.

Richter, S. & Messer, W. (1995). Genetic structure of the *dnaA* region of the cyanobacterium *Synechocystis* sp. strain PCC6803. *J Bacteriol* **177**, 4245–4251.

Roth, A. & Messer, W. (1995). The DNA binding domain of initiator protein DnaA. *EMBO J* **14**, 2106–2111.

Sabelnikov, A. G., Greenberg, B. & Lacks, S. A. (1995). An extended -10 promoter alone directs transcription of the DpnII operon of *Streptococcus pneumoniae*. *J Mol Biol* **250**, 144–155.

Salazar, L., Fsihi, H., de Rossi, E., Riccardi, G., Rios, C., Cole, S. T. & Takiff, H. E. (1996). Organization of the origins of replication of the chromosomes of *Mycobacterium smegmatis*, *Mycobacterium leprae* and *Mycobacterium tuberculosis* and isolation of a functional origin from *M. smegmatis*. *Mol Microbiol* **20**, 283–293.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Schaefer, C. & Messer, W. (1991). DnaA protein/DNA interaction. Modulation of the recognition sequence. *Mol Gen Genet* **226**, 34–40.

Schaper, S. & Messer, W. (1995). Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. *J Biol Chem* **270**, 17622–17626.

Schaper, S. & Messer, W. (1997). Prediction of the structure of the replication initiator protein DnaA. *Proteins Struct Funct Genet* **28**, 1–9.

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

Short, J. M., Fernandez, J. M., Jorge, J. A. & Huse, W. D. (1988). λ Zap: a bacteriophage λ expression vector with *in vivo* excision properties. *Nucleic Acids Res* **16**, 7583–7600.

Singer, B. S., Gold, L., Shinedling, S. T., Colkitt, M., Hunter, L. R., Pribnow, D. & Nelson, M. A. (1981). Analysis *in vivo* of translational mutants of the rIIB cistron of bacteriophage T4. *J Mol Biol* **149**, 405–432.

Skorko-Glonek, J., Wawrzynow, A., Krzewski, K., Kurpierz, K. & Lipinska, B. (1995). Site-directed mutagenesis of the HtrA (DegP) serine protease, whose proteolytic activity is indispensable for *Escherichia coli* survival at high temperatures. *Gene* **163**, 47–52.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**, 503–517.

Yoshikawa, H. & Ogasawara, N. (1991). Structure and function of DnaA and the DnaA box in eubacteria: evolutionary relationships of bacterial replication origins. *Mol Microbiol* **5**, 2589–2597.

Zweiger, C. & Shapiro, L. (1994). Expression of *Caulobacter dnaA* as a function of the cell cycle. *J Bacteriol* **176**, 401–408.

Received 24 July 1997; revised 5 September 1997; accepted 6 October 1997.

†**Present address:** University of Chicago, Department of Molecular Genetics and Cell Biology, 920 East 58th Street, Chicago, IL 60637, USA.