Volume 15 Number 20 1987 Nucleic Acids Research

Sequence features of the replication terminus of the *Bacillus subtilis* chromosome

C.M.Carrigan, J.A.Haarsma, M.T.Smith and R.G.Wake

Department of Biochemistry, University of Sydney, NSW, Australia

Received August 12, 1987; Revised and Accepted September 25, 1987

ABSTRACT

The sequence of 1267 nucleotides spanning the replication terminus, terC, of the Bacillus subtilis 168 chromosome has been determined. The site of arrest of the clockwise fork, which defines terC, has been localized to a 30-nucleotide portion (approximately) within this sequence. The arrest site occurs in an A+T-rich region between two open reading frames and very close to one of two imperfect inverted repeats (47-48 nucleotides each) which are separated by 59 nucleotides. The closeness of approach of the arrested clockwise fork to the first imperfect inverted repeat encountered in this region raises the possibility of a role for the inverted repeats in the mechanism of fork arrest.

INTRODUCTION

Termination of chromosome replication in Bacillus subtilis 168 occurs at a site, terC, which is approximately opposite the origin of replication, oriC (1,2). The clockwise replication fork reaches this site first and is blocked, or severely impeded, in its movement (3). The anticlockwise fork arrives a few minutes later to presumably fuse with the arrested fork and so complete termination. terC has been localized to a particular 1.95 kb PstI - EcoRI segment of DNA, which has also been cloned; and the site of fork arrest has been estimated to lie approximately 1.2 kb from the EcoRI end of this segment (4).

In this paper we present the nucleotide sequence of approximately 1.3 kb of DNA spanning terC and examine it for structural features. Analysis of single strands derived from the arrested clockwise fork has fixed more precisely the location of the fork junction within the sequence. The most striking feature of the sequence in the vicinity of the site of fork arrest is an A+T-rich region containing two imperfect inverted repeats. This region is located between two open reading frames (ORFs) and the possibility that it has a special role in the termination process is discussed.

MATERIALS AND METHODS

Bacterial strains and plasmids

B.subtilis strain GSY1127, a class II stable merodiploid (*hisH2 ilvC/ ilvC*$^+$) was obtained from C. Anagnostopoulos. Plasmid pWS10, which contains the *B.subtilis terC* region, was constructed in this laboratory (4).

DNA preparations

For analysis of forked DNA, or its single strand products, BamHI digested DNA was obtained from exponentially growing GSY1127 using the 'osmolysate' procedure described previously (3) but at a three fold cell concentration. To obtain BamHI + EcoRI digests for single strand analysis in alkaline gels the BamHI osmolysate was firstly fractionated in a preparative CsCl density gradient and the DNA band concentrated ten fold by pressure dialysis against 10 mM Tris HCl (pH 8) + 1 mM EDTA. It was then digested with EcoRI under standard conditions. DNA from spores of GSY1127 was obtained as described previously (3).

DNA sequencing and data analysis

DNA sequencing was performed by the dideoxynucleotide chain termination method (5). To obtain most (~ 90%) of the sequence suitable DNA fragments were sub-cloned into M13mp10 or M13mp11 vectors and sequenced according to the procedure outlined by Amersham International, U.K. Part of the sequence was obtained by dideoxy sequencing of double strand DNA cloned in the Bluescript M13 vectors according to the instructions provided by Stratagene (San Diego, U.S.A.). The complete sequence reported here was done in both directions.

DNA sequences were analysed for various features with the SEQNCE general purpose analysis program (Delaney Software Ltd., Vancouver, Canada). Secondary structure was analysed with the ZUKER program (PCFOLD: version 3.0, adapted for IBM PC by D. Brunelle) provided by the Molecular Biology Computer Research Resource (Dana Farber Cancer Institute, Harvard Medical School). Protein sequences were compared with the Genbank and EMBL nucleic acid databases translated in all 6 reading frames using the program of M. Kanehisa provided by the CSIRO Division of Molecular Biology.

Agarose gel electrophoresis, transfer of DNA to membranes and hybridization

Electrophoresis of BamHI + EcoRI digests of DNA were carried out in 2% alkaline agarose gels (6). Procedures for transfer of DNA to nylon membranes, hybridization with ^{32}P-labelled cloned DNA and autoradiography were as described previously (7).

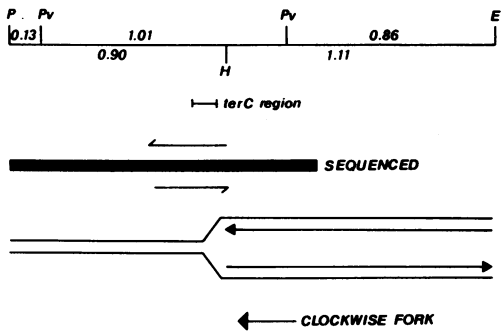

Fig. 1. Features of the B.subtilis terC - containing PstI - EcoRI insert of pWS10. The top section shows a restriction map of the insert; the terC region is located approximately 1.2 kb from the EcoRI end. The clockwise-moving replication fork generated at the origin of the chromosome enters this segment from the EcoRI end and stops at terC (bottom section). The heavy line (middle section), covering 1.3 kb and spanning terC, was sequenced. The arrowed lines above and below the sequenced portion show the directions of dideoxy sequencing through the terC region. P, PstI; Pv, PvuII; H, HindIII; E, EcoRI. All sizes are in kb. The overall size of this segment of DNA has been estimated at 1.95 kb, while the sum of the smaller portions of it give a value closer to 2.0 kb.

RESULTS

Nucleotide sequence

Figure 1 (top section) shows a restriction map of the terC - containing B.subtilis insert of the plasmid pWS10. The clockwise replication fork enters this segment as it resides in the bacterial chromosome from the right (or EcoRI end) and is arrested (bottom section) at around the position indicated, terC, approximately 1.2 kb from the EcoRI end. The portion shown as a heavy line just below the map was sequenced. The sequence of 1267 nucleotides is shown in Figure 2 (the clockwise fork enters this sequence from the bottom). It contains two ORFs reading in the PstI → EcoRI direction (opposite to movement of the clockwise fork). They are shown diagramatically in Figure 3. ORF1 (177 amino acids) is incomplete and is followed by a possible transcription termination site. ORF2 (122 amino acids) is complete. It is preceded closely by the consensus Shine-Dalgarno (SD) sequence AAGGAG which is complementary to the 3'end of B.subtilis 16S RNA (see ref. 8), and is followed by a possible transcription termination site. No attempt has been made to identify possible transcription start sites associated with ORF2. There is an additional complete ORF (73 amino acids) reading opposite to ORF1 and ORF2. It

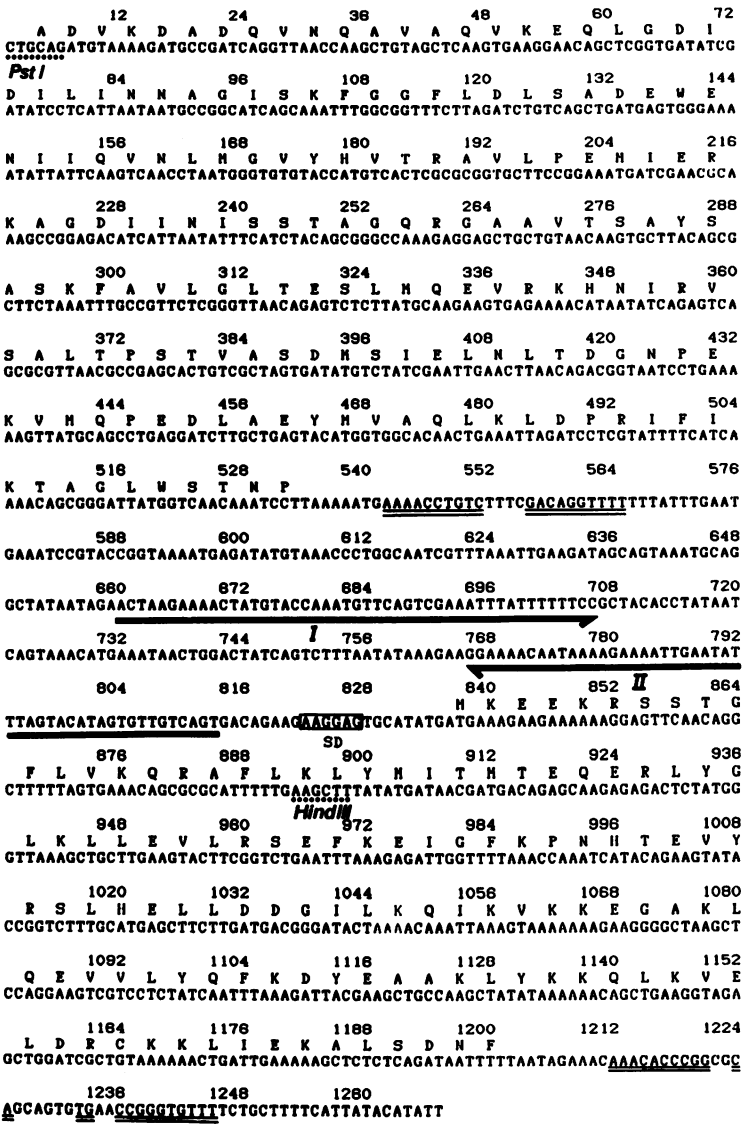

Fig. 2. Nucleotide sequence of approximately 1.3 kb of DNA spanning terC. The sequence extends from the PstI end of the insert of pWS10 and covers 1267 nucleotides (heavy line in Fig. 1). The boxed sequence is a potential ribosome-binding site (SD sequence); and the double-underlining indicates potential transcription termination sites. Two ORFs are defined by the amino acid sequences shown. The heavy arrows, labelled I and II, define two imperfect inverted repeats located between the ORFs. The PstI and HindIII sites (see Fig. 1) are identified.

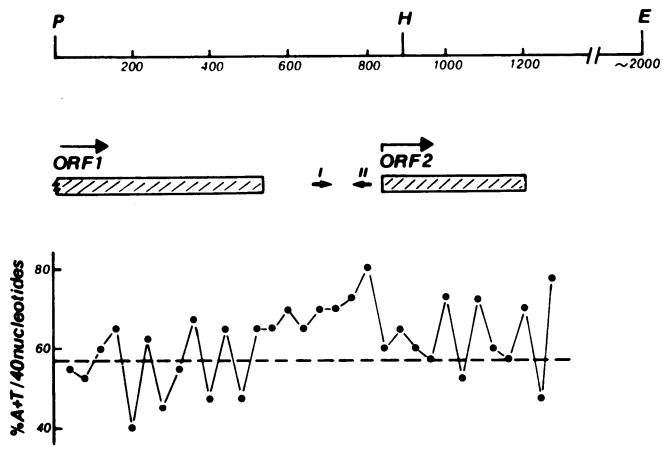

Fig. 3. Sequence features of the 1267 nucleotide segment of DNA spanning terC. The top section shows a scale (nucleotide number) of the sequenced region which extends from the PstI site (position 1) and through the HindIII site (895); terC is located around position 800 in the sequence. The middle section shows the regions (shaded) covered by the two ORFs; ORF1 is incomplete. The arrowed lines (I and II) between ORF1 and ORF2 define the inverted repeats. The bottom section shows the A+T distribution through the sequence; the dotted line is at the level of the A+T content of total B.subtilis DNA, P, PstI; H, HindIII; E, EcoRI.

occurs within ORF1 but there is no associated consensus SD sequence within a distance of 20 nucleotides from the putative initiation codon.

The leftwards extremity of ORF2 extends 60 nucleotides to the left of the HindIII site shown in Figures 1 and 3. This places it close to the arrest site, terC. The non translatable region of 301 nucleotides between ORF1 and ORF2 could define terC and it exhibits some special features. It is relatively rich in A+T (Fig. 3, bottom section). More striking is the presence of two imperfect inverted repeats within the 660-813 portion of the sequence. Each of the inverted repeats (labelled I and II) extends for 47-48 nucleotides. They show 77% homology to one another (Fig. 4) and are separated by 59 nucleotides.

Refined sequence location of the arrested fork

The earlier localization of terC to a region approximately 1.2 kb to the left of the EcoRI site of the pWS10 insert (see Fig. 1) was based upon the size of the double strand 'arm' released from the EcoRI + PstI-derived arrested fork either during the isolation of the DNA or after S1 nuclease treatment (4). A more reliable and refined estimate for location of the fork

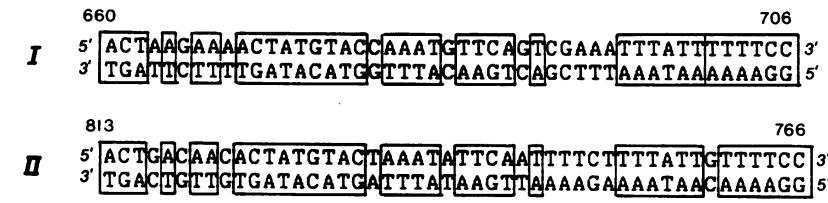

Fig. 4. Homology within the inverted repeats identified as I and II in Figs. 2 and 3. The boxed segments are regions of perfect homology.

junction has been obtained from the size of the single strands derived from the forked molecule present in an EcoRI + BamHI digest. The first BamHI site to the left of the EcoRI site shown in Figure 1 is 10.9 kb distant from it. Thus, the arrested fork in such a digest would yield single strands of approximately 1.2 and 10.9 kb. Figure 5 compares the single strand forms of DNA in EcoRI + BamHI digests of DNA from exponentially growing cells of the merodiploid strain GSY1127 (see ref. 7) and from spores of the same strain. Spore DNA contains completed non-replicating chromosomes and is therefore devoid of forked molecules. The DNA from exponential cells where the arrested fork is present in a portion of the chromosomes (lane 2) shows single strand material in the 1.2 kb size range (labelled X) which is absent from the spore DNA as expected. It covers the range 1.12-1.21 kb and appears to comprise two poorly resolved species. Lane 1 in Figure 5 contains the 1.11 kb EcoRI - HindIII segment of the pWS10 insert (see Fig. 1). The smaller of the two poorly resolved species is marginally larger than this segment. The use of appropriate digests of pWS10 as standards, sometimes in the same lane as the DNA to be analysed, established the smaller of the two single strand species to be 10-40 nucleotides longer than the EcoRI - HindIII portion of the pWS10 insert (Fig. 1), and the larger to be 70-100 nucleotides longer than it. This means that the junction of the arrested fork extends at the most 70-100 nucleotides upstream of the HindIII site i.e. to the 795-825 region of the sequence, and in the vicinity of inverted repeat II (see Fig. 2).

In Figure 5 the larger of the unique species, X, in lane 2 is the more prominent. In other experiments, using different DNA preparations, the relative amounts of the two species were not as markedly different. It should also be pointed out that the material spreading downwards from the 10.9 kb species (very obvious here because of the heavy exposure for autoradiography) frequently showed banding within it, even in spore DNA. This banding was mostly at positions above 2.0 kb. It probably reflects over-digestion of the

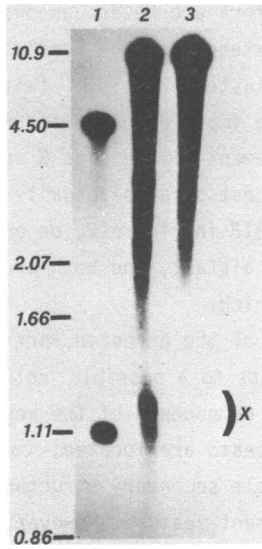

Fig. 5. Single strand composition of the arrested fork localized within the EcoRI - BamHI segment of the B.subtilis chromosome. EcoRI + BamHI digests of highly purified DNA from exponentially growing cells of GSY1127 containing the arrested fork (lane 2) and spore DNA (lane 3) were fractionated in a 2% agarose gel in alkali. The DNA was transferred to a nylon membrane, hybridized with ^{32}P-labelled pWS10 DNA and autoradiographed. Lane 1 shows one of the lanes containing size standards, in this case an EcoRI + HindIII digest of pWS10. Material in the 1.2 kb region (labelled X) is present in lane 2 and absent from lane 3. Sizes are in kb.

DNA and is certainly unrelated to fork arrest at terC. It has been shown that deletion of a substantial chromosomal segment to the left of and up to the PstI site in Figure 1 has no noticeable effect on fork arrest (unpublished data). Thus, the single strands from the 'arm' region of an EcoRI-derived arrested fork could not be longer than 2.0 kb.

DISCUSSION

Termination of chromosome replication in B.subtilis involves the arrest of the clockwise moving fork at a unique site, terC. Knowledge of the nucleotide sequence spanning terC is obviously important in understanding the molecular mechanism of fork arrest. While the minimum sequence needed for this aspect of termination has not yet been established it is likely that the sequence close to and just ahead of the arrested fork plays a crucial role. The 1267 nucleotide sequence described here spans terC. While the newly synthesized

strands within the arrested fork are heterogeneous in size, the longest ones (possibly leading strands) extend from the EcoRI site into a region between two ORFs (Fig. 3) and terminate very close (within 20 nucleotides) to the first of two 47-48 nucleotide imperfect inverted repeats that is encountered in the direction of fork movement (II in Figs 2 and 3). The present data do not indicate whether the longest strands actually stop short of this inverted repeat i.e. before position 813 in Figure 2, or enter it. The other inverted repeat, I, is 59 nucleotides distant, and both are located within a DNA segment that is relatively A+T rich.

The closeness of approach of the arrested fork to the first inverted repeat that it encounters points to a possible role for it in arresting fork movement. The single strand components of the region between the two ORFs, in which both inverted repeats are located, can be folded with the ZUKER computer program into a stable secondary structure within which the paired inverted repeats are a prominent feature. However, while *in vitro* synthesis by mammalian DNA polymerase α is arrested at stable secondary structures (hairpins) formed by palindromic sequences within single strand templates, it appears that such structures do not form *in vivo* (9). The sequencing experiments here showed that the Klenow fragment of DNA polymerase I had no difficulty in replicating the *terC* - containing single strand templates in both orientations (see Fig. 1). The question arises, whether there is some feature of chromosomal organisation in the vicinity of *terC* and/or the replication machinery itself that causes a stable secondary structure, incorporating the imperfect inverted repeats, to form. Of course, it is possible that the inverted repeats function to block fork movement through a mechanism not involving secondary structure formation within the separated template strands. They might even function independently of one another, each arresting a fork moving in one particular direction.

It is appropriate to point out that the nucleotide sequence surrounding the replication terminus of the *Escherichia coli* plasmid R6K was described some time ago and found not to contain any 2-fold rotational symmetry (10). But use of the ZUKER program has enabled the detection of two 20-nucleotide imperfect inverted repeats (90% homology) separated by 73 nucleotides. The explanation favoured for arrest of the fork at the R6K terminus was interaction of a protein with the terminus sequence (10) and this must be considered a possibility for *B.subtilis*. At present there is no information on the possible expression of the complete open reading frame, ORF2. Also, the possibility that its putative protein product has a role in termination cannot be

ruled out. A search of the Genbank and EMBL nucleic acid data bases for a translatable protein (all reading frames) related to the product of ORF2 was negative.

Current work is aimed at establishing the minimal sequence required for fork arrest and the precise sequence extremities of the leading and lagging strands within the arrested fork. Such information will help to clarify the possible role of the inverted repeat region in the termination of replication process.

ACKNOWLEDGEMENTS

We thank Dr A. Reisner for advice on the computer analysis of DNA sequences. This work was supported by the Australian Research Grants Scheme and the University of Sydney Cancer Research Fund.

REFERENCES

1. Weiss, A.S. and Wake, R.G. (1983) J. Mol. Biol. 171, 119-137.
2. Iismaa, T.P., Smith, M.T. and Wake, R.G. (1984) Gene 32, 171-180.
3. Hanley, P.J.B., Carrigan, C.M., Rowe, D.B. and Wake, R.G. (1987) J. Mol. Biol. (in press).
4. Smith, M.T., Aynsley, C. and Wake R.G. (1985) Gene 38, 9-17.
5. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
6. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
7. Weiss, A.S. and Wake, R.G. (1984) Cell 39, 683-689.
8. Hager, P.W. and Rabinowitz, J.C. (1985) In The Molecular Biology of the Bacilli Vol. II (Dubnau, D.A., ed). pp 1-32, Academic Press, New York.
9. Weaver, D.T. and De Pamphilis, M.L. (1984) J. Mol. Biol. 180, 961-986.
10. Bastia, D. Germino, J., Crosa, J.H. and Ram, J. (1981) Proc. Natl. Acad. Sci. USA 78, 2095-2099.