The nucleotide sequence of the replication origin of plasmid NTPl

J.N.Grindley and D.Nakada

Department of Biochemistry, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15261, USA

Received 27 April 1981

Abstract

The sequence of the DNA of the origin region of NTP1 has been obtained. Analysis of the sequence indicates that: (1) there is great sequence homology in the DNA upstream from the origin in NTP1, Co1E1, CLODF13, PBR345 AND PBR322; (2) only seven base pairs of NTP1 are identical with the sequence downstream from the origin in ColEl, although some homology exists for 140 bases downstream; (3) two ten base pair direct repeats are present in NTP1 which are also conserved in all four plasmids named above; (4) probably no polypeptide greater than fifteen amino acids in length is encoded by the NTP1 origin region, since no single open reading frame is conserved in all five plasmids.

Introduction

The plasmid NTP1 is an 8335 base pair (B.P.) multicopy plasmid which codes for a -lactamase (1) as part of a tranposon Tnl701 (2). A map of the HincIl and HaeII restriction enzyme cleavage sites and major promotors has been published. NTP1 was isolated in England from Salmonella typhimurium phage type 29 in 1965 (3) but appears to be homologous (Grindley and Nakada, unpublished observations) to RSF1030, despite the apparently independent sources of the two plasmids (4).

In this paper deletion derivatives of NTP1 are described. The composition of these smaller replicons strongly suggests that the HaeIIE fragment of the original plasmid NTP1 is essential for survival of the smaller replicons. In vitro, RSF1030 has been shown to replicate undirectionally from an origin located in or near the region equivalent to HaeIIE (5). The sequences of the DNA of NTP1 of this fragment and an adjacent fragment have been obtained. They are compared here with the sequence of the origin regions of other small, multicopy plasmids which have the same

host cell requirements, for replication of their DNA.

Materials and Methods

Materials

Restriction enzymes were purchased from New England Biolabs or B.R.L. T_A polynucleotide kinase was from P.L. Biochemicals; calf alkaline phosphatase was obtained from Boehringer Mannheim, as was T_A DNA ligase.

Isolation of plasmid DNA

Cells harboring plasmids and grown in minimal medium with casamino acids and with glycerol as carbon source were lysed with Triton X after lysozyme-EDTA treatment. Cleared lysates (6) were subjected to two sequential ethidium-bromide caesium-chloride equilibrium gradients, or to sucrose density gradients after deproteinisation with phenol.

Mapping Procedures

Restriction enzyme cleavage sites on the deletion derivatives of NTP1 were mapped using conventional procedures of partial and double digestions or by the Smith and Bernsteil method (7).

Sequencing Procedures

These were essentially as described by Maxam and Gilbert (8).

Results

Isolation of NTP1 deletion derivatives

Figure ¹ shows the steps by which the deletion plasmid pJN86 was derived from NTPI. In the first step, NTP1 was digested to apparent completion by Hae II, as judged by the absence of partial digestion products in the digest displayed on an agarose gel. The band comprising the Hae II A fragment was isolated from the gel, the DNA extracted and treated with T_4 ligase, then used to transform E.coli. The plasmid DNA from several resulting penicillin-resistant transformants was isolated and the restriction cleavage patterns examined. In all cases the plasmid DNA was as exemplified by pJN70; that is they were formed of both Hae II A and E fragments.

In the second step, Hinc II partial fragments about 3500 B.P. in size were isolated from pJN70 DNA. These were treated

Figure 1. Restriction Enzyme cleavage map of NTP1 and its deletion derivatives pJN70 and pJN86. Filled arrows indicate Hincll cleavage sites, open arrows show Haell sites. The heavy horizontal line indicates material derived from Tnl701, the cross-hatched block shows the position of the β -lactamase gene. The zigzag arrow shows the origin and direction of replication of the plasmid.

with ligase and used to transform E.coli. The penicillinresistant transformants which resulted were screened for plasmids smaller than pJN70. Most of the smaller plasmids examined were shown to be composed of three Hinc II fragments (see figure 1). Of these plasmids, pJN86 was chosen for further mapping. Part of this map is shown in Figure 2.

Sequencing the DNA of pJN86 in the Hae II E region

The strategy employed for sequencing the pJN86 Hae II E region using the Maxam-Gilbert chemical degradation method (8) was as diagrammed in Fig. 2. It can be seen that the DNA sequence for this region (given in Fig. 3) was obtained for both strands in most areas, or for the same DNA strand from two overlapping restriction fragments. The exceptions are in the areas of nucleotides 93-100; 211-249 and 558-576. The sequence

Figure 2. Restriction Enzyme Cleavage sites and extent of DNA sequence obtained from specific fragments in the Origin region of pJN86 (NTP1).

of both strands for the region 93-100 was obtained from pJN89, a plasmid analogous to pJN86 in cleavage map pattern and derivation save that it was obtained from pNG77, a mutant of NTP1 temperature-sensitive for DNA replication (9). Similarly, the sequences of the DNA in the other two regions have been confirmed using pJN89 DNA.

Discussion

Locating the Origin of Replication of NTP1

Replication of RSF 1030 (and thus probably NTP1) like that of many other small multicopy plasmids, is dependant on the hostencoded products of the dnaB, dnaC, dnaG, dnaI, and dnaP genes (10). In addition, the host RNA polymerase and DNA polymerases ^I and III are required (10, 11). Although the sequence of the DNA constituting the origin region of some of these small plasmids is known, the functions encoded by this region are understood only in gross terms. As a prelude to further studies on the specific sequences required for DNA replication to initiate, we decided to sequence the origin DNA of NTP1 and compare it with the known sequences of related plasmids.

Evidence that the pJN86 regions around the Hae II E fragment whose sequence is shown in Fig. 3, covers the point at which NTP1 DNA replication initiates, is derived from four sources.

Figure 3. Sequence of the DNA of the origin region of pJN86 (NTP1). Zigzag arrow indicates the origin of replication and the direction of replication. The underlined region is the position of a semi-conserved direct repeat of a symmetrical sequence. *= the start of RNA_{100nt} (Frederick, Yamada and Nakada, in press). The Hael1 site at nucleotide 91 is that between Haell E and Haell A, the DNA to the left being TnI701 material.

Nucleic Acids Research

Firstly, we have shown in this paper that the Hae II A fragment alone cannot yield a viable plasmid, but that partial digestion products containing also the E fragment, but not the smaller adjacent F fragment, are selected. Secondly, Conrad, Wold, and Campbell (5) have located the origin of replication of the homologous plasmid RSF1030 within this area of the genome. Thirdly, the sequence of pJN86 DNA is almost completely identical to the sequence of ColEl DNA upstream from the point where DNA replication initiates, for some 300 bases. This region is also highly conserved in PBR345 (12) PBR322 (13) and CLODF13 (14) except that the latter plasmid has a 45 B.P. insert some 40 B.P. from the point of initiation. In contrast, the sequence of pJN86 remains identical with that of ColEl for only about ⁷ B.P. downstream from the initiation point (i.e. in the direction of replication) although substantial similarity remains for at least 140 B.P. downstream. Fourthly, Naito and Uchida recently reported mutants of ColEl/ hybrids which are defective for ColEl-type replication (15). One mutation maps within the Hae II E fragment at the base of a potential hairpin structure (in a position equivalent to base 426 in pJN86).

There is evidence that suggests that this region, which is so highly conserved in all four plasmids, does not suffice for initiation of plasmid DNA replication. Inselberg showed that the ColEl, Hae II E fragment in a pSC101/Co1E1 hybrid must be maintained in its correct position and orientation for ColEl-type replication to be recovered, suggesting that continuity of sequence over the Hae II cleavage site into the adjacent fragment is required (16). Itoh and Tomizawa (17) showed that the promotor for the RNA primer of ColEl replication lies 555 B.P. upstream from the point at which replication initiates, outside the Hae II E fragment and the region conserved in pJN86.

The pJN86 sequence given in Fig. ³ also shows the region coding for a small, prevalent RNA species whose properties and possible function are discussed by Frederick, Yamada and Nakada (18). A similar small RNA species was first described for ColEl by Levine and Rupp (19) and for RSF1030 by Conrad and Campbell (20). Tomizawa et al (21) have shown that RNA primer formation for ColEl DNA replication is inhibited by the small ColElspecified RNA but not by that specified by RSF1030.

Secondary and Repeated Structures

We have used the Staden Computer programme (22) to scan the pJN86 DNA sequence for stem and loop structures, Shine-Dalgarno sequences (23) and open reading frames.

Backman et al. (24) reported two potential stem and loop structures near the origin of replication of PBR345. They are conserved in ColEl, PBR322 and pJN86 but the longer of the two is not conserved in CLODF13 because of the 45 B.P. insertion.

Ohmori and Tomizawa (25) showed that two inverted repeated sequences existed about 500 and 600 B.P. upstream from the ColEl origin. Neither of these are conserved intact in pJN86.

Nomura and Ray (26) showed that the ColEl Hae II E fragment cloned in phage M13 would promote lagging (H) strand initiation of the chimeric single-stranded DNA template, in the presence of the dnaB and dnaG gene products. This is in agreement with the reuirements noted by Staudenbauer et al. (27) for ColEl H-strand synthesis. Thus, ColEl H-strand initiation might require a single-stranded template and involve a mechanism similar to priming of 0X174 negative (complementary) strand initiation. The exact location of the ColEl site was not determined. However, complementing studies by Zipursky and Marians (28) established that a 156 B.P. L-strand region of the PBR322 Hae II E fragment, supported factor Y ATPase activity. Factor Y is also essential for priming OX174 negative strand synthesis. Thus, this factor Y effector site lies within the region whose DNA sequence is substantially conserved between all these small plasmids.

Zipursky and Marians also reported a second site on PBR322 supporting factor Y ATPase activity. This site, downstream from the origin of the H-strand of the Hae II B fragment, was postulated to be the site where discontinuous synthesis of the Lstrand started: at the termination site of the 6SL fragment which is primed by RNA polymerase from the origin (17, 27). Again Zipursky and Marians could find no sequence homology with either the OX174 site or the other PBR322 site. However, a hairpin in the ColEl DNA sequence was noted by Bastia (29) to bear a striking resemblance to the hairpin in the intergenic region between genes F and G of OX174 and G4 which forms the recognition site (30) for protein n' (alias factor Y). The

significance of this homology and its correlation with the recognition site for factor Y remains unclear because replication of OX174 requires Y factor activity while that of G4 does not. This hairpin starts only ⁹ B.P. from, and covers the Hae II site which divided Hae IIE from the ColEl equivalent of PBR322 Hae II B. It ends in a Pvu II recognition site. This ColEl structure is not totally conserved in PBR322 although both restriction sites are present. Moreover, Zipursky and Marians note that Pvu II cleavage of PBR322 inactivates the Hae II B effector region, suggesting that this region of the fragment is important. Restriction mapping and preliminary sequence data of the possibly equivalent region of pJN86 (i.e., the region of Hinc II D which is not part of Tn1701) suggests that there is no extensive or precise sequence homology between the two plasmids in this region. However, Sims and Benz (31) have presented evidence that two regions of 115 bases of OK were protected against nuclease digestion by dnaG binding to initiate replication on OK, and that this binding may involve tertiary structute of the DNA. Even if the same mechanism of initiation is involved and the same proteins bind to the DNAs, sequence homology between initiation sites may not exist.

Backman et al. (24) noted the presence of two 10 B.P. direct repeats about 400 B.P. upstream from the origin of PBR342, which are more or less conserved in ColEl, PBR322, CLODF13 and pJN86. Such repeated synmetrical sequences resemble known DNA-protein sites of interaction such as that of the repressor (32) and that postulated for the protein of the plasmid R6K (33). Potential Polypeptides encoded by the pJN86 origin region

The replication of these small multicopy plasmids appears not to require the presence of a plasmid-encoded protein (6, 34) but two results implicate proteins in the control of plasmid replication. Firstly, mutant plasmids temperature-sensitive for DNA replication have been isolated (9, 14). Secondly, extracts from chloramphenical-treated ColEl⁺ cells replicate ColEl DNA better than do untreated extracts from similar cells (35). It is reasonable to expect that a control protein might be encoded by the region around the origin of replication. We therefore searched for open reading frames for polypeptides longer than 15

amino acids, in this region of the pJN86 sequence. Such polypeptides are listed in Table. 1.

Comparing these with the potential polypeptides listed by Ohmori and Tomizawa (25) for ColEl we find the pJN86-5 is

Table 1. Polypeptides potentially encloded by the DNA of the pJN86 origin region

identical with the ColEl L-A profein. In PBR322, the coding region for this potential polypeptide would yield one ⁴ amino acids longer, but it would not be encoded by CLODF13 because the GTG start codon is CGG. Similarly, while pJN86-6 is the same as ColEl L-B, appears in PBR345, and suffers only one amino acid change in PBR322, in CLODF13 the GTG start has been abolished. The potential polypeptide starting at 518 in pJN86 (-8) is homologous to L-D of ColEl but has no equivalent in CLODF13: that equivalent to L-E (52 amino acids) of ColEl is only 36 amino acids long in pJN86, 84 amino acids long in PBR322. The remaining L-strand specified polypeptides potentially specified by ColEl have no equivalent in pJN86, nor does ColEl encode the others that pJN86 may specify from this strand.

On the opposite strand, pJN86 could code for four polypeptides initiating at ATG, and 6 starting with GTG (see Table 1). Of the ATG starts, two potential polypeptides of pJN86 may be encoded by ColEl. The coding sequence for pJN86-9 is identical with ColEl H-B except that the former starts with ATG. There are three other codon changes relative to ColEl H-B, only two of which result in amino acid substitutions. This frame also appears open in CLODF13. The potential polypeptide pJN86-19 is equivalent to ColEl H-F, and may also be encoded by CLODF13. However, the ATG start codon is absent in PBR322.

Of the pJN86 open reading frames beginning with GTG on this strand, two (14 and 15) have ColEl equivalents. While pJN86-15 would yield a polypeptide similar to that of ColEl, it would contain three different amino acids at the N-terminal end.

Comparing all these multicopy plasmids, we infer that it is unlikely that even a small polypeptide is specified by their origin regions. For pJN86, additional weight is given to this argument by the absence of Shine-Dalgarno sequences (23) prior to a start codon. Proteins specified by pJN86 have been examined in minicells and maxicells on 2-dimensional gels, but polypeptides less than 100 amino acids would not have been resolved (unpublished data of this laboratory). We have not specifically attempted to detect the most likely polypeptide, pJN86-9. Further conclusions regarding the involvement of NTP1-encoded polypeptides in plasmid replication will only be possible when

the DNA sequence of plasmids defective in replication is known. Such experiments are now being conducted in this laboratory.

Acknowledgements

We wish to thank the students attending a course in DNA sequencing in the Biological Sciences Department of Pittsburgh University and Drs. Nigel Grindley and Cathy Clarke who helped to teach the course which generated the sequence given here. This work was supported by U.S. Public Health Service Grant #GM 25227 to Dr. Nigel Grindley, by National Science Foundation Grant #PCM79-04591 to Dr. D. Nakada and by U.S. Public Health Service Grant #GM 28137 to Drs. D. Nakada and J.N. Grindley.

Dr. D. Nakada passed away while this manuscript was in preparation. He was greatly loved both for his enthusiasm for science and for his strong affection for his fellow scientists. Dai will be remembered dearly by all of us who were privileged to know him.

References

- 1. Grindley, J.N., Blumberg, R. and Nakada, D. (1977) J. Bacteriol. 130, 852-859.
- 2. Yamada, Y., Calame, K.L., Grindley, J.N. and Nakada D. (1979) J. Bacteriol. 137, 990-999.
- 3. Anderson, E.S., and Lewis, M.J. (1965) Nature Lond. 208, 843-849.
- 4. Heffron, F., Sublett, R., Hedges, R.W., Jacob, A. and Falkow, S. (1975), J. Becteriol. 122, 250-256.
- 5. Conrad, S.E., Wold, M. and Campbell, J.L. (1979) Proc. Natl Acad. Sci. U.S.A. 76, 736-740.
- 6. Clewell, D.B. and Helinski, D.R. (1969) Proc. Natl. Acad. Sci. U.S.A. 62, 1159-1166.
- 7. Smith, H.O. and Bernsteil, M.L. (1976) Nucleic Acids Res. 3, 2387-2390.
- 8. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
- 9. Grindley, N.D.F., Grindley, J.N. and Kelley, W.F. (1978) in "Microbiology 1978," ed. D. Schlesinger, Publ. Am. Soc. Microbiol. pp. 71-73.
- 10. Conrad, S.E., and Campbell, J.L. (1979) Nucleic Acids Res. 6, 3298-3302.
- 11. Grindley, N.D.F. and Kelley, W.S. (1976) Molec. Gen.
- Genet. 143, 311-318.
12. Bolivar, F., Betlach, M.C., Heyneker, H.L., Shine, J., Rodriguez, R. and Boyer, H.W. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5265-5296.
- 13. Sutcliffe, J.G. (1978) Cold Spring Harbor Symp. Quant. Biol. 43, 69-76.

14. Stuitje, A.R., Veltkamp, W., Maas, J. and Heyneker, H.L. (1980) Nucleic Acids Res. 8, 1459-1473. 15. Naito, S. and Uchida, H. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 6744-6748. 16. Inselberg, J. (1980) J. Bacteriol. 143, 1073-1076.
17. Itoh. T. and Tomizawa. J. (1980) Proc. Natl. Acad. Itoh, T. and Tomizawa, J. (1980) Proc. Natl. Acad. Sci.
U.S.A. 77, 2450-2454. $U.S.A.$ $17, 2450-2454.$ 18. Frederick, C., Yamada, Y. and Nakada, D. (1980) Nucleic Acids Res. (in press). 19. Levine, A. and Rupp, W.D. (1978) Microbiology. 163-166. 20. Conrad, S.E. and Campbell, J.L. (1979) Cell. 1<u>8</u>, 61–71. 21. Tomizawa, J., Itoh, T., Selzer, G. and Som, T. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 1421-1425. 22. Staden, R. (1977) Nucleic \overline{A} cids Res. $\overline{4}$, 4037-4051.
23. Shine. J. and Dalgarno. L. (1974) Proc. Natl. Acad. Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U.S.A. <u>71</u>, 1342-1346. 24. Backman, K., Betlach, M., Boyer, H.W., Yanofsky, S. (1978) Cold Spring Harbor Sym. Quant. Biol. 43, 69-76. 25. Ohmori, H. and Tomizawa, J. (1979) Molec. Gen. Gent. 176, 161-170. 26. Nomura, N. and Ray, D.S. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 6566-6570. 27. Staudenbauer, W.L., Scherzinger, E. and Lanka, E. (1979) Molec. Gen. Genet. 177, 113-120. 28. Zipursky, S.L. and Marians, K.J. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 6521-6525. 29. Bastia, D. (1 $\overline{977}$) Nucleic Acids Res. $\overline{4}$, 3123-3142.
30. Shlomai, J. and Kornberg. A. (1980) Proc. Natl. Aca 30. Shlomai, J. and Kornberg, A. (1980) Proc. Natl. Acad. Sci. U.S.A. <u>77</u>, 799–803. 31. Sims, J. and Benz, E.W. (1980) Proc. Natl. Acad. Sci. U.S.A. <u>77</u>, 799–803. 32. Ptashne, M., Backman, K., Humayun, M.A., Jeffrey, A., Manier, R., Meyer, B. and Sauer, R.T. (1976) Science. 194, 156-161. 33. Stalker, D.M., Kolter, R. and Helinski, D.R. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 1150-1154. 34. Crossa, J.H., Luttrop, L.K. and Falkow, S. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 654-658. 35. Tomizawa, J., Sakahibara, Y. and Kakefuda, T. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 1050-1054.