Mol Gen Genet (1981) 181:8–12

MGG

Nucleotide Sequence of the Region of the Origin of Replication of the Broad Host Range Plasmid RK2

David M. Stalker, Christopher M. Thomas*, and Donald R. Helinski
Department of Biology, B-022, University of California, San Diego, La Jolla, California 92093, USA

Summary. A DNA sequence consisting of 617 base pairs (bp) from the region of the origin of replication of the broad-host range plasmid RK2 has been determined. Included within this sequence is a 393 bp *Hpa*II restriction fragment that provides a functional origin or replication when other essential RK2 specified functions are provided in *trans*. Also contained in this sequence is a region, distinguished functionally from the replication origin, which is involved in the expression of inc_2 incompatibility, i.e., the ability of derivatives of RK2 to eliminate a resident RK2 plasmid. The 617 bp sequence includes eight 17 base pair direct repeats with 5 located within the region required for a functional replication origin and 3 within the region involved in inc_2 incompatibility. In addition, a 40 bp region rich in A–T followed by a 60 bp stretch having a high G+C content is present. Deletion evidence indicates that the A–T rich and possibly the G+C regions are required for a functional replication origin. Based on the evidence contained in this and the preceding paper (Thomas et al. 1980b) a model will be presented for the involvement of these specific sequences in the initiation of RK2 DNA replication, plasmid maintenance and plasmid incompatibility.

Introduction

Plasmid replicons are stably inherited extrachromosomal elements that exhibit a variety of replication modes. Because of these differences in replication properties and their relatively small size, plasmids provide convenient model systems for studying the regulation of DNA replication in bacteria. Of primary importance in studying these plasmid systems is the determination of the nucleotide sequence of regions that function in the control of initiation of replication of the specific plasmid replicon.

The DNA sequences of functional origins of DNA replication have been determined for a variety of prokaryotic and eukaryotic replicons. Included in the prokaryotic origins sequences are ColE1 (Tomizawa et al. 1977; Oka et al. 1979), RI and Rdrd19 (Oertel et al. 1979; Rosen et al. 1979), R6K (Stalker et al. 1979), the single-strand DNA phages G4 (Sims and Dressler 1978) and ØX174 (Sanger et al. 1978), the lambdoid bacteriophages (Grosshield and Hobom 1979), and the chromosomal origins of *E. coli* (Meijer et al. 1979; Sugimoto et al. 1979) and *S. typhimurium* (Zyskind et al. 1980). Eukaryotic origins sequenced include the papovaviruses polyoma (Soeda et al. 1978; Friedman et al. 1978), SV40 (Fiers et al. 1978; Reddy et al. 1978) and BK (Dhar et al. 1978), and the origin of DNA replication of human mitochondrial DNA (Crews et al. 1979).

We have chosen to determine the nucleotide sequence of the functional origin of replication for the self-transmissible 56 kilobase (kb) plasmid RK2, which confers antibiotic resistance to ampicillin, kanamycin and tetracycline (Meyer et al. 1977a). Plasmid RK2 replicates in *Escherichia coli* unidirectionally from a unique origin (Meyer et al. 1977b). Recent evidence has shown that a 750 base pair (bp) *Hae*II DNA fragment linked to a DNA fragment specifying kanamycin resistance can act as a functional replication origin when other segments of the RK2 genome are provided in *trans* in the same cell (Figurski and Helinski 1979; Thomas et al. 1979). The preceding paper describes a further reduction in size of this origin to a 393 bp *Hpa*II restriction fragment (Thomas et al. 1980b). We have determined the nucleotide sequence of this *Hpa*II origin fragment and also a 244 base pair adjacent region that is involved in specifying certain incompatibility properties of RK2.

The sequence of the RK2 origin of replication is of particular interest due to the fact that this plasmid has the ability to transfer between and be maintained in a wide range of gram negative bacteria (Beringer 1974; Datta and Hedges 1972). If the single origin identified in molecules replicating in *E. coli* is the only functional origin on the plasmid, then it must be utilized by replication enzyme systems from a variety of hosts. RK2 also is a tightly controlled replicon by virtue of its low copy number (Meyer et al. 1977a) but differs from other stringently regulated plasmids in that physically separable *trans*-acting functions have been shown to be essential for the initiation of RK2 replication.

Material and Methods

Strains and Plasmids. Plasmid pCT7 was obtained from the *E. coli* strain GMII9 *dam6, dam3 met* B1 *lac* Y1 *sup* E44 tsx 78. The plasmid was extracted and purified by two cycles of cesium chloride/ethidium bromide gradients as described by Meyer et al. (1977a).

Enzymes and Radiolabels. All restriction enzymes used were prepared by the unpublished method of Pat Greene or purchased from New England Biolabs or Bethesda Research Laboratories. The Klenow fragment of *E. coli* DNA polymerase I and T4 polynucleotide kinase were obtained from Boeringer-Mannheim and P-L Laboratories, respectively. Bacterial alkaline phosphatase was from Worthington. Radiochemical Centre (Amersham) was the source of [α-^{32}P] dNTPs (350 Ci/mmol) and [γ-^{32}P] ATP (~7,000 Ci/mmol) was kindley supplied by Mel Simon.

* *Present address:* Department of Genetics, University of Birmingham, P.O. Box 363, Birmingham B15 2TT, England

Offprint requests to: D.M. Stalker

0026-8925/81/0181/0008/$01.00

Isolation and Labeling of DNA Fragments. DNA restriction fragments were purified on $15 \times 7 \times 0.3$ cm 6% polyacrylamide gels run at 80 volts after digestion of pCT7 DNA with the appropriate restriction enzyme. The DNA fragments were removed from the gel and eluted in buffer containing 0.5 M ammonium acetate, 0.01 M magnesium acetate, 0.1 mM EDTA and 1% sodium laurylsulfate at 65° C for approximately 15 h (Gilbert and Maxam 1977). The fragments were precipitated and subjected to end-labeling reactions with Klenow DNA polymerase and [α-^{32}P]dNTPs for the 3′-ends (Klenow et al. 1971), or T4 polynucleotide kinase and [γ-^{32}P]ATP for 5′-ends (Murray 1973) after dephosphorylation of the 5′-end with bacterial alkaline phosphate. The resulting labeled fragments were then re-digested with a second restriction enzyme and re-purified on 6% polyacrylamide gels or subjected to strand separation on $36 \times 26 \times 0.1$ cm 7% polyacrylamide gels (60:1 ratio of acrylamide to bis). Prior to strand separation, the DNA samples were denatured in 60% dimethylsulfoxide at 90° C for 2 min, chilled in ice and immediately loaded. Gels were run at 160 volts with buffer containing 0.045 M Tris, 0.045 M boric acid and 0.0012 M EDTA. Autoradiographs were then obtained and the single-strand fragments eluted as described above.

DNA Sequencing. Chemical modification and degradation of the labeled DNA fragments was carried out essentially as described by Gilbert and Maxam (1977) with minor modifications (Stalker et al. 1979). Sequencing gels ($36 \times 15 \times 0.04$ cm) were 25%, 12% and 8% polyacrylamide and allowed determination of a given DNA sequence of up to 200 bases from a labeled end. Autoradiographs were obtained at −70° C with Ilford fast tungstate intensifying screens.

Results

In order to obtain the nucleotide sequence of the functional origin of RK2 replication, the 750 bp *Hae*II origin fragment of plasmid pCT7 (Fig. 1) (Thomas et al. 1979, 1980a) was isolated as described in Materials and Methods. This *Hae*II fragment was subjected to the sequencing strategy shown in Fig. 1. DNA was digested with the restriction enzyme *Mbo*I, labeled at the 3′ and 5′ ends, then re-digested with *Hpa*II and the fragments re-purified by polyacrylamide gel electrophoresis. Additionally, the origin fragment was cleaved with *Hpa*II, end-labeled, re-cleaved with *Mbo*I and re-purified. These fragments were then subjected to nucleotide sequencing. The small (95 and 90 bp) *Hpa*II fragments located within the *Hae*II fragment were strand-separated before the sequence determination. The 5′-ends of the single *Alu*I restriction site were also labeled and the two resulting *Alu*I/*Hae*II DNA fragments were strand-separated and sequenced. Finally, the 118 base pair *Taq*I fragment was isolated from a total *Taq*I digest of pCT7 and labeled at the 3′-end before strand separation for sequencing. This strategy allowed us to obtain significant over-lapping between sequences of the individual fragments and facilitated determination of each base in both complementary DNA strands.

The entire 617 base pair sequence containing the RK2 functional replication origin is displayed in Fig. 2. This sequence is 61% rich in G–C base pairs which is very high when compared to the A–T rich origins of ColE1 (Oka et al. 1979), R6K (Stalker et al. 1979), and the *E. coli* chromosome (Meijer et al. 1979). Important features include a possible RNA polymerase recognition sequence at positions 290 through 296. The sequence GATGATG differs by a single base from the one originally described by Pribnow (1975) for the binding of RNA polymerase complexes to initiate transcription. From positions 501–550, there exists a region which is 74% rich in A–T base pairs followed by a 79% G+C containing region starting at base pair 551 and ending at 617. Regions rich in adenine and thymine, low in their thermal stability, are common to a variety of replication origins (Grosschedl and Hobom 1979; Stalker et al. 1979). Sequences of high G+C content exhibit high thermal stability and have been postulated to propagate their energy of stability to adjacent A–T rich regions to facilitate their "flexibility" (Sobell 1978). As shown in Fig. 3, this 67 bp high G–C rich region can also undergo extensive intrastrand base-pairing. Hairpin loops can be formed in four different ways dependent upon the degree of folding. The guanine and cytosine stretches within this region make these inverted structures extremely stable with respect to their free energy of formation.

The most striking feature of this entire sequence is the presence of eight 17 bp direct repeats. These repeats are not in tandem, but are interspersed throughout the sequence, separated by regions containing 5–6 bases in all except one junction. These direct repeats can be placed into two main groups. Five of the repeats are clustered within the 393 bp *Hpa*II fragment which provides a functional origin of replication while the other three repeats reside outside this fragment (positions 144 through 207) in a region that is required for inc_2 incompatibility (Thomas et al. 1980b). Extensive direct repeat regions have been found in the replication origins of plasmid R6K (Stalker et al. 1979), the lambdoid phages (Grosshield and Hobom 1979) and the chromosome of *E. coli* (Sugimoto et al. 1979). The presence of direct repeats in the origin region of several well defined replicons suggests a functional role for these repeats in the initia-

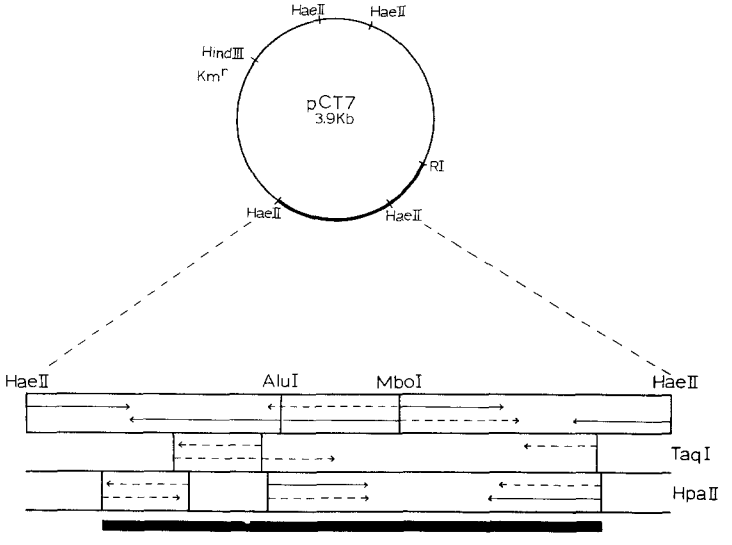

Fig. 1. Partial restriction map of the 750 base pair *Hae*II origin fragment from plasmid pCT7 indicating the strategy of DNA sequence determinations. Relevant restriction endonuclease sites are indicated. Solid arrows refer to sequences obtained from 5′-end labels while dotted arrows show sequences obtained by 3′-end labeling. The solid bar indicates the segment of DNA where both DNA strands have been sequenced

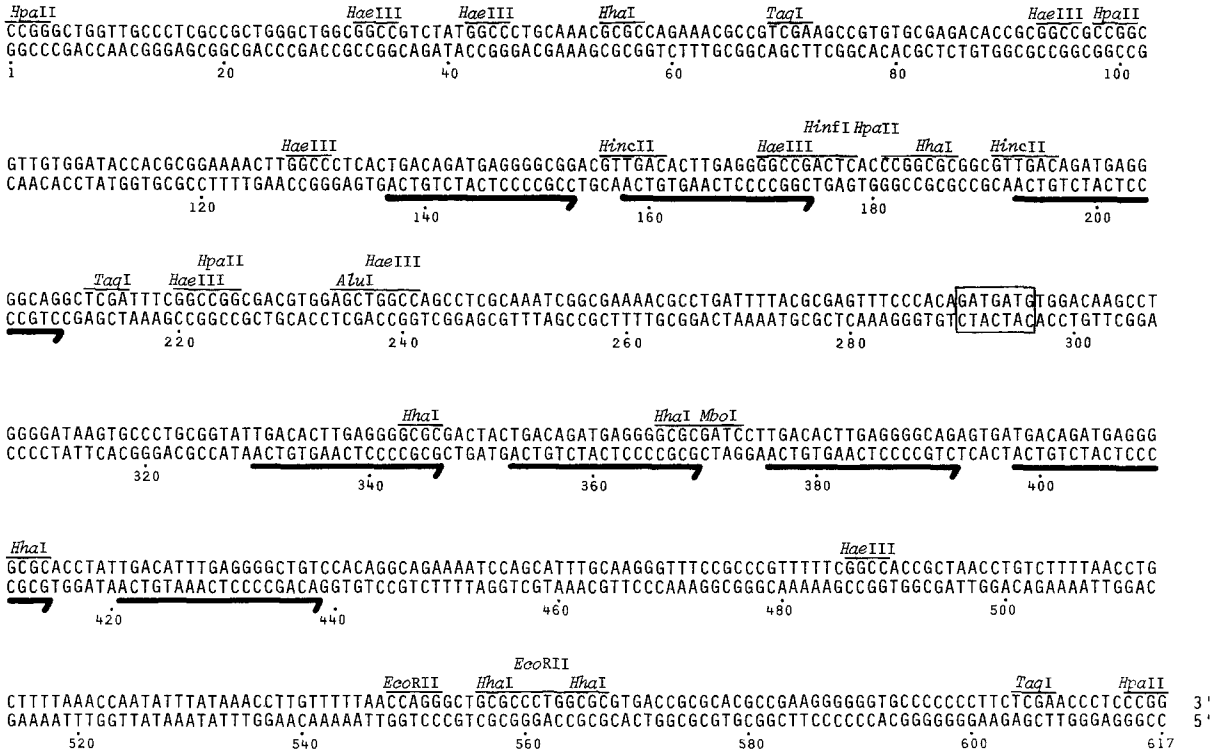

Fig. 2. The 617 base pair nucleotide sequence from a region of the *Hae*II origin fragment. The base pairs are numbered from 1–617 and appropriate restriction endonuclease sites are indicated. Arrows underneath sequences refer to the position of the 17 base pair direct repeats. The nucleotide sequence within the box corresponds to an RNA polymerase recognition sequence (Pribnow 1975)

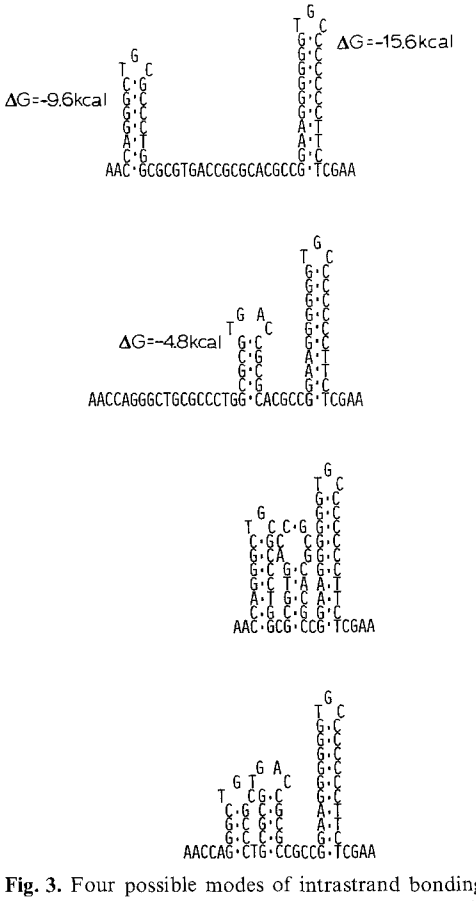

Fig. 3. Four possible modes of intrastrand bonding of the 79% G–C rich region (nucleotide pairs 546–617 in Fig. 2) are shown. Free energies were calculated according to Tinoco et al. 1971

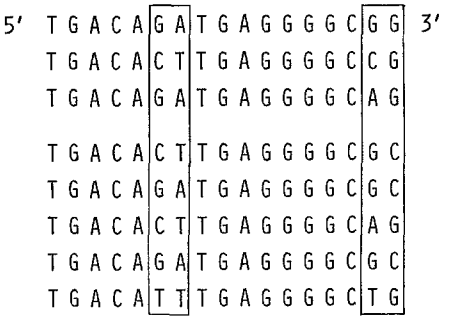

Fig. 4. Nucleotide sequence of the eight 17-bp direct repeats which are read from left to right in Fig. 2. The top 3 repeats are located outside the 393-bp *Hpa*II which contains the 5 other repeat units. Boxes indicate base positions where different homologies occur

tion of replication. The repeat units of RK2 which show a very high degree of homology between each other are displayed in Fig. 4. The differences between the sequences are found primarily (excepting the terminal bases) at the sixth and seventh positions of each unit, with GA and CT dinucleotide pairs alternating between seven of the eight repeat units. The last repeat contains a TT dinucleotide at these two positions.

The largest polypeptide that can be coded for by the RK2 origin region is 62 amino acids in length beginning with an AUG start codon at position 460 and ending with a TAA termination codon at 275; this would constitute a protein of approximately 7,400 daltons. There is also a possible ribosome binding site (GGA) located eleven bases upstream from the AUG start (Shine and Delgarno 1976). Using mini-cells obtained from a mini-cell producing strain carrying plasmid pCT7, we were unable to identify a protein in the size range of 8,000 daltons.

All other reading frames in this sequence are closed by termination codons.

Discussion

The nucleotide sequence of a 617 bp segment of a *Hae*II fragment obtained from the broad-host range plasmid RK2 has been determined. This region contains a 393 bp *Hpa*II fragment that functions as the RK2 replication origin (Thomas et al. 1980b). Although this *Hpa*II fragment, as in the case of the larger *Hae*II fragment, is not a complete replicon, it provides a functional replication origin when two other regions of the RK2 genome (trfA and trfB) are provided in *trans* in the same cell (Thomas et al. 1980a). One or both of these regions provide trans-acting factors that are required for the initiation of RK2 DNA replication and may be involved in the regulation of this event. Potential sites for the action of the trans-acting factors within the RK2 replication origin are illustrated in Fig. 5. The Pribnow box at base pair positions 290–296 may provide the signal for the start of an RNA transcript that conceivably would terminate at the sequence GC-CCGTTTTTCGGCC (base pairs 475–489) which is similar to several rho independent transcription terminators that have been described previously (Gilbert 1976). This RNA transcript would be approximately 175 bp in length and includes the cluster of five direct repeats. It is possible that this transcript is required for the initiation of replication by serving as an RNA primer (possibly after being processed) or providing a necessary transcriptional activation step similar to that required for the initiation of λ DNA replication (Dove et al. 1971). One or both of the trans-acting factors (*trf*A and *trf*B), required for RK2 replication, conceivably could function at the region of the Pribnow box and/or the five direct repeat sequences to initiate or modulate the initiation of RK2 replication. The results of analyses of RK2 deletion derivatives have suggested that the high G-C containing region (and possibly the A–T rich region as well) is required for a functional replication origin (Thomas et al. 1980b). The increased stability of these G-C structures might function to destabilize the A–T rich region and facilitate binding of the DNA to a protein(s) essential for the formation of complex of initiation and/or an elongation replication complex. Replication would then proceed unidirectionally from the RK2 origin region.

The 750 bp *Hae* II origin fragment, when cloned into a functional replicon such as ColE1 (pCT7), is eliminated, but not readily, by an incoming RK2 plasmid (inc_1) and will eliminate other RK2 replicons (inc_2) (Thomas et al. 1980b). However, plasmid constructs consisting of the 393 bp *Hpa* II origin joined to functional replicons ColE1 (pCT42) and R6K (pCT43) do not eliminate an RK2 replicon nor are these plasmids subject to elimination (Thomas et al. 1980b). This suggests that the DNA segment containing the group of 3 direct repeats (base pair positions 144 through 208) is obligatory for the expression of the inc_2 function when these origin fragments are inserted into functional replicons. This is further supported by the observed reduction in the expression of inc_2 incompatibility upon removal of one or two of the repeat units located in the region of the three direct repeats (Thomas et al. 1980b). The cluster of five direct repeats located within the *Hpa*II origin fragment also is involved in the expression of incompatibility by RK2 since plasmid derivatives (e.g., pCT47.1) containing a deletion extending from the single *Mbo*I site, which removes 3 repeats at positions 376–414, are less efficient in their expression of the inc_2 function (Thomas et al. 1980b). These observations suggest that all 8 direct repeat units are required for "full" expression of incompatibility by RK2.

The exact role of the 8 direct repeats in the maintenance of the plasmid and the expression of incompatibility is unknown at this time. The repeated sequences may function as recognition sites for plasmid and/or host encoded proteins essential for the initiation of RK2 replication. They also may play a role, possibly together with a plasmid and/or host encoded protein, in the association of the plasmid with a hypothetical cellular membrane site required for plasmid replication or segregation. The finding of a substantial decrease in stability of the RK2 plasmid derivatives, despite an increased copy number, upon deletion of the segment of DNA containing the cluster of three direct repeats (Thomas et al. 1980b) can be accounted for by a *cis*-acting role of this cluster of repeats in plasmid RK2 segregation. This proposed role of the repeats may be related to their observed involvement in the expression of RK2 incompatibility. A further dissection of this replication origin region and the testing for replication, maintenance and incompatibility properties of various combinations of components of this region should provide important new insights to the mechanisms responsible for the maintenance of this broad host range plasmid.

Acknowledgements. This work was supported by Public Health Service Grant AI-07194 from the National Institute of Allergy and Infectious Disease and Grant PCM 79-04635 from the National Science Foundation. David M. Stalker is a United States Public Health Service postdoctoral fellow (1-F32 GMO6542). Christopher M. Thomas was a recipient of a Smith Kline Corporation postdoctoral fellowship.

References

Beringer JG (1974) R factor transfer in *Rhizobium leguminosarum*. J Gen Microbiol 84:188–198

Datta N, Hedges RW (1972) Host range of R factors. J Gen Microbiol 70:453–460

Dhar R, Lai C-J, Khoury G (1978) Nucleotide sequence of the DNA replication origin for human papovavirus BKV: Sequence and structural homology with SV40. Cell 13:345–358

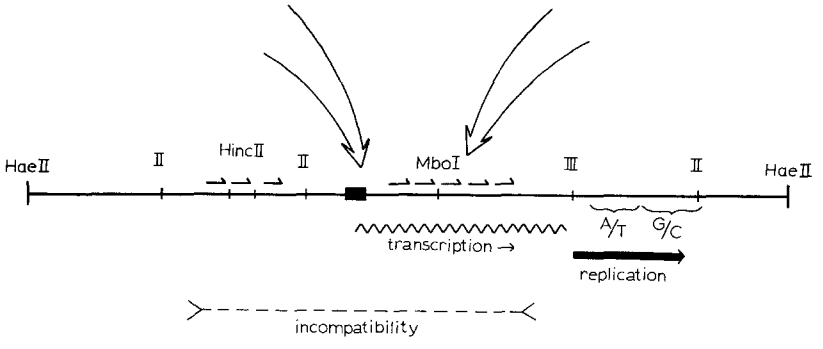

Fig. 5. A possible role for various regions of RK2 in the replication and incompatibility properties of the RK2 replication origin. Relevant restriction endonuclease sites are listed. Small arrows refer to the positions of the direct repeats. The location of a Pribnow box is indicated (■). A+T and G+C refer to adenine/thymine and guanine/cytosine rich regions of the sequence. The large arrows refer to *trans*-acting replication factors and their possible interaction at sites within the origin region

Dove WF, Inokuchi H, Stevens WF (1971) Replication control in phage lambda. In: Hershey AD (ed) The bacteriophage lambda. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, p 747

Fiers W, Contreras R, Haegeman G, Rogiers R, Vande Voorde A, Van Heuversweyn H, Van Herreweghe J, Volckaert G, Ysebaert M (1978) Complete nucleotide sequence of SV40 DNA. Nature 273:113–120

Figurski D, Helinski DR (1979) Replication of an origin containing derivative of plasmid RK2 dependent on a plasmid function provided in *trans*. Proc Natl Acad Sci, USA 76:1648–1652

Friedman T, LaPorte P, Esty A (1978) Nucleotide sequence studies of polyoma DNA. J Biol Chem 253:6561–6567

Gilbert W (1976) Starting and stopping sequences for the RNA polymerase. In: Losick R and Chamberlin M (eds) RNA polymerase. Cold Spring Harbor Laboratory, Cold Spring Harbor NY, p 193

Grosschedl R, Hobom G (1979) DNA sequences and structural homologies of the replication origins of lambdoid bacteriophages. Nature 277:621–627

Klenow H, Overgaard-Hansen K, Patkar SA (1971) Proteolytic cleavage of native DNA polymerase into two different catalytic fragments. Eur J Biochem 22:371–381

Maxam A, Gilbert W (1977) A new method for sequencing DNA. Proc Natl Acad Sci USA 74:560–564

Meijer M, Beck E, Hansen FG, Bergmans H, Messer W, von Meyenburg, K, Schaller H (1979) Nucleotide sequence of the origin of replication of the *Escherichia coli* K-12 chromosome. Proc Natl Acad Sci USA 76:580–584

Meyer R, Figurski D, Helinski DR (1977a) Physical and genetic studies with restriction endonucleases on the broad host range plasmid RK2. Mol Gen Genet 152:129–135

Meyer R, Helinski DR (1977b) Unidirectional replication of the P-group plasmid RK2. Biochim Biophys Acta 478:109–113

Murray K (1973) Nucleotide sequence analysis with polynucleotide kinase and nucleotide mapping methods. Biochem J 131:569–583

Oertel W, Kolleck R, Beck E, Goebel W (1979) The nucleotide sequence of a DNA fragment from the replication origin of the antibiotic resistance factor RIdrd19. Mol Gen Genet 171:277–285

Oka A, Nomura N, Morita M, Sugisaki H, Sugimoto K, Takanami M (1979) Nucleotide sequence of small ColE1 derivatives: Structure of the regions essential for autonomous replication and colicin E1 immunity. Mol Gen Genet 172:151–159

Pribnow D (1975) Nucleotide sequence of an RNA polymerase binding site at an early T7 promotor. Proc. Natl. Acad. Sci USA 72:784–788

Pritchard RH, Barth PT, Collins J (1969) Control of DNA synthesis in bacteria. Symposia of the Society for Gen Microbiol 19:263–297

Reddy VB, Thimmappaya B, Dhar R, Subramanian KN, Zain BS, Pan J, Ghosh PK, Celma ML, Weissman SM (1978) The genome of simian virus 40. Science 200:494–502

Rosen J, Ohtsubo H, Ohtsubo E (1979) The nucleotide sequence of the region surrounding the replication origin of an R100 resistance factor derivative. Mol Gen Genet 171:287–293

Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes JC, Hutchinson CA, Slocombe PM, Smith M (1977) Nucleotide sequence of bacteriophage ϕX174 DNA. Nature 265:687–705

Shine J, Dalgarno L (1975) Determinant of cistron specificity in bacterial ribosomes. Nature 254:34–48

Sims J, Dressler D (1978) Site-specific initiation of a DNA fragment: Nucleotide sequence of the bacteriophage G4 negative-strand initiation site. Proc. Natl Acad Sci USA 75:3094–3098

Smith DW, Deen LT, Vuong TM, Zyskind JW (1980) Recombinant DNA methodology and the study of DNA-cell surface interactions in bacteria. Microbiol in press

Sobell HM, Lozansky ED, Lessen M (1978) Structural and energetic considerations of wave propagation in DNA. Cold Spring Harbor Symp Quant Biol 43:11–20

Soeda E, Kimura G, Miura K (1978) Similarity of nucleotide sequences around the origin of DNA replication in mouse polyoma virus and simian virus 40. Proc Natl Acad Sci USA 75:162–166

Stalker DM, Kolter R, Helinski DR (1979) Nucleotide sequence of the region of replication of the antibiotic resistance plasmid R6K. Proc Natl Acad Sci USA 76:1150–1154

Sugimoto K, Oka A, Sugisaki H, Takanami M, Nishimura A, Yasuda Y, Hirota Y (1979) Nucleotide sequence of *Escherichia coli* K-12 replication origin. Proc Natl Acad Sci USA 76:575–579

Thomas CM, Meyer R, Helinski DR (1980a) Regions of the broad host range plasmid RK2 essential for replication and maintenance. J Bacteriol 141:213–222

Thomas CM, Stalker DM, Guiney D, Helinski DR (1979) Essential regions for the replication and conjugal transfer of the broad host range plasmid RK2. In: Timmis KN and Puhler A (eds), Symp. on Plasmids of Medical, Environmental and Commercial Importance. Elsevier/North Holland, Amsterdam, p 375

Thomas CM, Stalker DM, Helinski DR (1981) Replication and incompatibility properties of segments of the origin region of replication of the broad-host range plasmid RK2. Mol Gen Genet 181:1–7

Tinoco I Jr, Uhlenbeck OC, Levine MD (1971) Estimation of secondary structure in ribonucleic acids. Nature 230:362–267

Tomizawa J, Ohmori H, Bird RE (1977) Origin of replication of colicin E1 plasmid DNA. Proc Natl Acad Sci USA 74:1865–1869

Communicated by D. Sherrat

Received October 12, 1980