**Correction.** In the article "Energy conformation study of Met-enkephalin and its D-Ala² analogue and their resemblance to rigid opiates" by Gilda H. Loew and S. K. Burt, which appeared in the January 1978 issue of *Proc. Natl. Acad. Sci. USA* (75, 7–11), the authors request the following changes. On page 10, lines 4–11 of the Discussion might be unintentionally misconstrued and we would like to clarify them. Thus, lines 10 and 11 should be deleted and lines 5–9 should be corrected to read: "from crystal and gas phase studies on small molecules and has been used almost exclusively for peptide conformations. Some differences found between the two methods were: (i) the PCILO method gave smaller energy differences among the various conformers, (ii) extended conformers were more favored with PCILO than with ECEPP, and (iii) variation of side-chain torsion angles with PCILO gave a more monotonically varying energy contour than ECEPP."

**Correction.** In the article "Isolation, characterization, and synthesis of a corticotropin-inhibiting peptide from human pituitary glands" by Choh Hao Li, David Chung, Donald Yamashiro, and C. Y. Lee, which appeared in the September 1978 issue of *Proc. Natl. Acad. Sci. USA* (75, 4306–4309), the authors request that the following correction be made. On p. 4309, paragraph 1, sentence 1, the phrase "... 10-fold molar concentration" should be changed to "... 10²-fold molar concentration." The corrected sentence should then read: "As shown in Table 3, α₂-ACTH-(7–38) is devoid of corticosteroidogenic activity but it inhibits corticosterone production as stimulated by α₂-ACTH in isolated rat adrenal cells by 95% at 10²-fold molar concentration."

**Correction.** In the article "Site-specific initiation of a DNA fragment: Nucleotide sequence of the bacteriophage G4 negative-strand initiation site" by John Sims and David Dressler, which appeared in the July 1978 issue of *Proc. Natl. Acad. Sci. USA* (75, 3094–3098), the reproduction of Fig. 3 was unsatisfactory. A better version is shown below.

![Fig. 3. Parts of two sequencing gels showing the positive strand (Left) and the negative strand (Right) at the initiation site. Synthesis of the negative strand begins at the position labeled +1. Sequencing was performed as described by Maxam and Gilbert (15), with the following modifications (A. Maxam, personal communication). The amount of calf thymus carrier DNA was decreased to \( \frac{1}{10} \) and the amount of tRNA in the stop solutions was decreased to \( \frac{1}{100} \); both reductions increase the sharpness of the bands. Also, magnesium acetate was omitted from the hydrazine stop solution to prevent the occasional formation of a precipitate which impairs the quality of the pyrimidine displays. Several of the bands of these two gels do not show the regular spacing characteristic of most sequencing displays. This is almost certainly due to the formation of intramolecular secondary structure. The ability to read the sequences, however, is not impaired.](image-url)
Site-specific initiation of a DNA fragment: Nucleotide sequence of the bacteriophage G4 negative-strand initiation site

(DNA replication/origin of replication/dna G protein/Okazaki fragment)

JOHN SIMS AND DAVID DRESSLER

The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138

Communicated by J. D. Watson, April 10, 1978

ABSTRACT The synthesis of the bacteriophage G4 negative strand is an example of the de novo initiation of a polynucleotide chain. This initiation is performed by the Escherichia coli replication protein dna G which selects a unique site on 5400-base positive-strand template. In this paper we present the nucleotide sequence of the G4 negative-strand initiation site. This is the template element recognized by the dna G priming protein.

In conjunction with the sequence of the nascent negative strand, obtained by Bouché, Rowen, and Kornberg [Bouché, J.-P., Rowen, L. & Kornberg, A. (1978) J. Biol. Chem. 253, 765-769], the present data provide a description of a dna G-dependent origin of replication in which one knows the place at which polymerization starts at the nucleotide level.

A general problem in DNA synthesis is the nature of the mechanism by which polynucleotide strands are initiated. The single-stranded DNA phages provide model systems for studying this problem. The first step in the life cycle of these viruses is the conversion of the infecting positive-strand circle to the duplex state, an event that involves the de novo initiation of the negative strand. In many respects this synthesis is similar to the formation of an Okazaki fragment during bacterial DNA replication.

Three different methods of initiating DNA strands de novo are used by the various single-stranded phages. In one class of phages (M13, fd, and f1), negative-strand synthesis is initiated by the bacterial RNA polymerase (1-3). A short oligoribonucleotide is synthesized in situ and subsequently elongated by DNA polymerase to give a full-length negative strand (4,5). Two other classes of phages, represented by ϕX174 and G4, do not use RNA polymerase to initiate their negative strands but instead use the same proteins that the host uses for replicating its own chromosome. Specifically, the enzyme that synthesizes the primer is dna G. In the case of bacteriophage ϕX, several proteins in addition to dna G are required for negative-strand initiation, and synthesis apparently is able to begin at many, not necessarily specific, locations (6-8). However, for G4 the situation is simpler. The dna G protein recognizes a specific DNA element on the positive strand and initiates negative-strand synthesis (9-11). The presence of a unique initiation site makes the class of single-stranded phages represented by G4 particularly useful as an experimental system in which to study DNA strand initiation.

Our general goal is to determine the nucleotide sequence of the negative-strand initiation site for several phages of the G4 class. This site serves in each phage as the nucleic acid counterpart of the dna G protein, providing a unique recognition element on the positive strand at which the replication enzymes begin negative-strand synthesis. A comparison of these sequences should reveal common features and hopefully will provide insight into the way in which this region is able to serve as an origin of DNA synthesis.

Bouché et al. (12) have already provided a picture of one part of the G4 negative-strand initiation site. They have determined the sequence of the short primer made in vitro by the dna G protein [which constitutes, depending on the specific conditions, the first 2 to 29 nucleotides of the nascent chain (13)]. Our experiments are designed to study the initiation site as a whole, including potential recognition elements adjacent to the primer sequence. Taken together, the data indicate both the overall structure of the initiation region and the exact nucleotide at which polymerization begins.

Location of the G4 negative-strand initiation site

The precise location of the negative-strand initiation site, and therefore the region of the G4 chromosome that we wished to sequence, was known to us from experiments carried out in our laboratory by D. Hourcade. His work followed the studies by Zechel et al. (9) who first showed in vitro that G4 possessed a unique origin for negative-strand synthesis. They mapped this origin to a region approximately 5% of the genome (or some 250 bases) from the single EcoRI site in G4.

Hourcade obtained a precise location for the origin through in vitro experiments (11). Briefly, cells were infected with G4 phage that had been irradiated with ultraviolet light. The thymine dimers introduced by the irradiation blocked the progress of the replication enzymes, allowing initiation but not completion of the negative strand. After recovery of the par-

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.
FIG. 2. (Upper) Electrophoresis. Lane A: Hin-Hpa 580, a 580-base-pair fragment containing the origin, was isolated from a polycracylamide gel after digestion of the G4 double-stranded DNA with HindII and Hpa II. The fragment was made radioactive by labeling its 5' ends with [γ-32P]ATP and polynucleotide kinase. Lane B: Uniformly labeled Hin-Hpa 580 was prepared from nick-translated G4 duplex rings. Upon digestion with Hae III, two fragments were released, 440 and 140 base pairs long. Lane C: Uniformly labeled Hin-Hpa 580 was cut with Hinf to give pieces 380, 110, and 100 base pairs long. Lane D: When end-labeled Hin-Hpa 580 was cut with Hinf, only the 380- and 110-base-pair-long pieces were radioactive, showing that they are the terminal fragments. (Lower) The composite restriction enzyme map in the region of the initiation site. The negative strand is identified as described in the text.

Sequencing the negative-strand initiation region
Knowing the location of the negative-strand initiation site to within 10 bases, we set out to sequence this regulatory DNA element. The first step was to construct a detailed restriction enzyme map of this area of the chromosome in order to identify fragments suitable for sequencing.

The Restriction Enzyme Map. As can be seen in Fig. 1, the region of the initiation site lies within the 2500-base-pair fragment A produced by HindII cleavage of G4. Furthermore, the initiation site will be contained in the end of this fragment nearer to the EcoRI cut. By digestion of duplex rings with both HindII and Hpa II, we were able to isolate the initiation region on a relatively small piece of DNA (called Hin-Hpa 580). This piece was then subdivided with other restriction enzymes (Fig. 2 upper).

Hae III cut once within Hin-Hpa 580 to produce pieces 440 and 140 base pairs long. Hinf made two cuts giving pieces of 380, 110, and 100 base pairs. Alu gave four pieces, 275, 245, 35, and 20 base pairs long. The use of polynucleotide kinase and [γ-32P]ATP to end-label the DNA allowed identification of the terminal pieces. Double digestion with pairs of enzymes revealed the order of the cuts. These data led to the restriction enzyme map for the origin region shown in Fig. 2 lower.

Finally, we needed to know which strand in Hin-Hpa 580 represented the template positive strand and which the negative strand. This information could be deduced from the in vivo data (11). Because these experiments had demonstrated that, after initiation, negative-strand synthesis proceeds through a Hae III cut and a HindII cut, on route toward the EcoRI site, the negative strand is identified as the chain with that order of cuts in the 5'-to-3' direction (Fig. 2 lower).

Sequencing Procedure. To sequence the initiation site, we used the technique developed by Maxam and Gilbert (15). This procedure requires polynucleotide strands labeled at single, defined ends. In brief, a restriction fragment is isolated, and then its terminal 5'-phosphoryl groups are removed with alkaline phosphatase and replaced with radioactive phosphates by incubation with [γ-32P]ATP and polynucleotide kinase. One thus obtains a fragment that is labeled specifically at two known positions, the 5' ends. A piece of DNA with a single labeled 5' end is then generated, either by separating the individual strands of the fragment on a polycracylamide gel or by cleaving the duplex fragments into two separable pieces with a second restriction enzyme.

In sequencing, the end-labeled DNA is subjected to one of four chemical treatments that promote random cleavage at a specific base. For example, limited reaction with hydrazine in the presence of 1 M NaCl followed by piperidine treatment leads to breakage at random cytosine residues. The denatured DNA then consists of a set of radioactive strands, each running from the 32P-labeled end to a random cytosine that was cleaved. This collection of chains is run on a slab gel, where separation occurs according to length. The set of bands that appear after autoradiography establish that, at defined distances from the 32P-labeled 5' end, there was a cytosine residue. Essentially analogous methods are used to obtain and display in parallel sets of chains ending in C + T, G, and A + G.

Nucleotide sequence of the initiation site
The restriction enzyme map in the region of the initiation site (Fig. 2 lower) shows several cuts that are useful for sequencing.
Two examples will be considered. An Alu site occurs about 75 base pairs upstream from the initiation site. This cut forms one end of a 275-base-pair piece that can be isolated as the sixth fragment present in a total Alu digest of G4 duplex rings. Alu 6 was thus purified, its two ends were labeled with 32P, and the fragment then was cut with Hae III. The larger of the two cleavage products contained a radioactive negative strand labeled about 75 bases from the initiation site. Analysis of this

![Sequence Diagram]

**FIG. 3.** Parts of two sequencing gels showing the positive strand (Left) and the negative strand (Right) at the initiation site. Synthesis of the negative strand begins at the position labeled +1. Sequencing was performed as described by Maxam and Gilbert (15), with the following modifications (A. Maxam, personal communication). The amount of calf thymus carrier DNA was decreased to \(1/10\) and the amount of tRNA in the stop solutions was decreased to \(1/5\); both reductions increase the sharpness of the bands. Also, magnesium acetate was omitted from the hydrazine stop solution to prevent the occasional formation of a precipitate which impairs the quality of the pyrimidine displays. Several of the bands of these two gels do not show the regular spacing characteristic of most sequencing displays. This is almost certainly due to the formation of intramolecular secondary structure. The ability to read the sequences, however, is not impaired.
chain (Fig. 3 right) gave 115 nucleotides of negative-strand sequence, running toward and then through the initiation site.

Fig. 2 lower also shows a cut that is particularly useful for sequencing the positive strand in the region of the initiation site—a Hinf cut about 45 base pairs downstream from the origin. When sequenced, the DNA labeled at this end gave 146 nucleotides from the positive strand, including the initiation site and the regions on either side (Fig. 3 left).

The data obtained by sequencing from the ends just described were mutually confirming. Together with additional sequencing from several other ends, they lead to the result shown in Fig. 4. This figure shows the base sequence for about 250 nucleotides upstream and 200 nucleotides downstream from the negative-strand initiation point. The sequence represents about 10% of the G4 genome.

In accord with the in vitro results (11) discussed above, the point at which negative-strand synthesis begins can be placed approximately 100 bases from the Hae III site (at the position marked by the horizontal bar in Fig. 4).

Fiddes et al. (16), in the course of a general sequencing study of G4, have also obtained a sequence for the region containing the negative-strand initiation site. Our two sequences for this region are in agreement.

The most important part of the initiation region sequence is the 28-base stretch:

\[5'-\text{ATGGACGCGAAGCCGCGGTCCTACT-3'}\]

This stretch is complementary to the sequence of the short oligonucleotide synthesized from G4 DNA in vitro by the \textit{dna} G protein (12). The presence of this sequence at the in vitro initiation site verifies one aspect of the primer hypothesis (1). The oligonucleotide synthesized by the \textit{dna} G protein and used to prime the formation of the negative strand is encoded intact at the physiological negative-strand initiation site. The 5' end of the primer synthesized in vitro is seen to occur 103 bases from a \textit{Hae} III cut; our in vitro experiments had located the start of negative-strand synthesis about 100 bases from this same \textit{Hae} III cut.

Thus, for G4, the in vitro, \textit{in vitro}, and DNA sequencing studies complement one another. Taken together, the data tell one exactly where polymerization starts and make the G4 negative-strand origin the first initiation site using the \textit{E. coli} replication protein \textit{dna} G for which the point of initiation of DNA synthesis is known at the nucleotide level.*

Properties of the initiation region

The specific features of the G4 negative-strand initiation site that allow it to serve as an origin of replication, although of considerable interest, are not yet clear. The primary sequence of bases could display the information necessary for recognition. Alternatively, the \textit{dna} G protein could interact with a structure of higher geometrical order. Several regions of symmetry, capable of forming hairpins, account for most of the initiation region. They are diagrammed in Fig. 5. The structure recognized by the \textit{dna} G protein might be the base-paired stem of one of these hairpins, the loop, or a more complicated tertiary structure involving one or more hairpins and perhaps using non-Watson-Crick base pairs such as those found in tRNA molecules.

A large part of the difficulty in deciding which aspects of the initiation site sequence are important for \textit{dna} G recognition results from the absence of comparable sequences. In Fig. 6, we note the intriguing and statistically significant similarity of the G4 sequence to the general region of the bacteriophage \lambda origin and to an area of the \phi X174 positive strand. Both templates are also used by the \textit{dna} G protein; however, no firm conclusions can be drawn because only in the case of G4 does one know where synthesis actually starts.

We anticipate that further progress will come through the study of other single-stranded phages that require only the \textit{dna} G protein for priming. For instance, studies similar to those discussed here using phages such as ST-1, \phi K, and \alpha-3 will provide sequences that can be compared directly to that of the G4 origin.

An intercistronic region

What is the relationship between the negative-strand initiation site and the genes of the G4 chromosome? Scanning the sequence for translational starting (AUG, GUG) and stopping (UAG, UAA, UGA) signals, one sees several examples in all three

* A recent report by Geider et al. (17) identifies the precise initiation site for the bacteriophage \textit{fd} negative strand. As discussed in the Introduction, this type of strand initiation is primed not by the \textit{dna} G protein but by RNA polymerase.
reading frames. When the predicted first 10 amino acids following AUG codons were written out and compared with the NH$_2$-terminal amino acid sequences of the proteins of the related phage $\phi$X174 (22, 23), it was apparent that those following the AUG at position −23 were similar to the amino acids at the NH$_2$ terminus of $\phi$X174 gene G. The sequence preceding the UAA stop codon at position +117 is nearly identical to that at the COOH terminus of $\phi$X174 gene F. Thus, the G4 negative-strand initiation site lies within a 136-base region between genes F and G, which may be compared to the 111-base intercistronic region between genes F and G that is found in $\phi$X174 (24). The two regions are similar in length but not particularly in sequence. Both show extensive secondary structure (see Fig. 5 and, for $\phi$X174, ref. 24). Intercistronic regions of this size in phages that maximize coding capacity by overlapping genes are surprising and apt to possess important functions. The data of this paper show that the region between genes G4 and F contains the recognition site at which the dna G protein initiates the synthesis of the negative strand.

We thank Eileen Herlihy for excellent technical assistance, P. Farabaugh, D. Hourcade, D. McConnell, U. Siebenlist, C. Sutcliffe, and R. Tizard for restriction enzymes, and Allan Maxam for helpful conversations. This study has been funded by Research Grants NF-57 from the American Cancer Society and GM-17088 from the National Institutes of Health and Research Career Development Award GM-70440 from the National Institutes of Health.