Structure-independent nucleotide sequence analysis

(inosine substitution/nucleoside triphosphate terminators/slab gel electrophoresis/Qβ replicase/MDV-1 RNA)

DONALD R. MILLS AND FRED RUSSELL KRAMER

Institute of Cancer Research and Department of Human Genetics and Development, College of Physicians and Surgeons, Columbia University, New York, New York 10032

Communicated by Sol Spiegelman, February 26, 1979

ABSTRACT Substitution of inosine for guanosine in the nucleic acid fragments synthesized for the sequencing of RNA effectively prevents the formation of secondary structures during electrophoretic analysis. Consequently, the mobility of each fragment in the sequencing gel is a strict function of its molecular weight. Inosine substitution should markedly improve the resolution that can be obtained in the sequencing of DNA as well as RNA.

Rapid nucleotide sequencing procedures for DNA (1–5) and RNA (6–9) depend on acrylamide gel electrophoresis in the presence of urea (10, 11) to separate nucleic acid fragments by molecular weight. In general, fragments differing in length by a single nucleotide can be resolved. Occasionally, a series of different-length fragments migrate at about the same rate and are seen in the gel as a region of “band compression” (1, 3, 9). This occurs when base-paired “hairpin” structures are present at the 3′ ends of the fragments (9). Although the gels contain urea and the separations are carried out at high temperatures to discourage the formation of secondary structures (12), the stronger hairpins persist, and the nucleotide sequence cannot be completely determined.

In this paper we describe a modified sequencing procedure that depends on the synthesis of nucleic acid fragments that do not form secondary structures during electrophoresis. The electrophoretic mobility of these fragments is a strict function of molecular weight. The method depends on the substitution of 3′-dITP for GTP (or dITP for dGTP) during synthesis of the fragments. Inosine specifically replaces guanosine in the products of all nucleic acid polymerases that have been tested (13–17). Because I-C base pairs are linked by only two hydrogen bonds (18), rather than the three that occur in G-C base pairs, the secondary structures that form in inosine-substituted nucleic acids are relatively weak. The magnitude of this destabilization is illustrated by the fact that the melting temperature of poly(dI-dC) is 55°C lower than that of poly(dG-dC) (19). Thus, inosine-substituted fragments should be much less likely to retain secondary structures during electrophoresis in the presence of denaturing agents such as 7 M urea.

To test this hypothesis, we prepared both inosine-substituted and control (guanosine-containing) fragment sets of MDV-1 (→) RNA for sequence analysis. MDV-1 (→) RNA is well suited for this study for three reasons: it is synthesized in vitro by Qβ replicase (20); its complete nucleotide sequence has been determined (21) by the classical procedures of Sanger and his colleagues (22) and has been confirmed by the newer chain-termination procedure (9); and the location of its secondary structures has been determined with the aid of reagents, such as sodium bisulfite, that react specifically with nucleotides that are in a single-stranded conformation (unpublished data).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

electrophoretic separation of inosine-substituted fragments was compared with that of control fragments. The results show that inosine substitution effectively eliminates secondary structures during electrophoresis.

MATERIALS AND METHODS

Nucleotides. Unlabeled ribonucleoside triphosphates were obtained from P-L Biochemicals, [γ-32P]GTP from ICN, and cordycepin triphosphate (3′-dATP) from Miles. We synthesized 3′-deoxyuridine from 2,2′-anhydro-1-(β-D-arabinofuranosyl)uracil (Sigma) by the procedure of Brown et al. (23), converted it to the 5′-monophosphate by the method of Yoshikawa et al. (24), and then converted this compound to 3′-dUTP by the procedure of Hoard and Ott (25).

Qβ Replicase. This RNA-directed RNA polymerase (26) was isolated from Qβ bacteriophage-infected Escherichia coli Q13 by the procedure of Eoyang and August (27), with the hydroxylapatite step omitted.

MDV-1 RNA. "Divalent" RNA is a low-level contaminant of Qβ replicase preparations (28). It was originally isolated from a Qβ replicase reaction that was not provided with exogenous template RNA (20). It consists of two antiparallel, single-stranded complements, each 221 nucleotides long. Its biological origin and cellular function are not yet known. However, it is an excellent template for Qβ replicase. The synthesis of MDV-1 RNA and the separation of the complementary (+) and (−) strands by electrophoresis in the presence of Mg2+ have been described (29). "Wild-type" MDV-1 RNA has some sequence heterogeneity at a number of its nucleotide positions. In order to use a homogeneous template in these experiments, we used a mutant of MDV-1 RNA that had been "cloned" from one strand of RNA (28). The sequence of this mutant differs from the published sequence of MDV-1 (→) RNA (21) at positions 87 and 88 in the (−) strand (see Fig. 2). For simplicity, this mutant RNA is referred to as "MDV-1 RNA" throughout this paper.

Synthesis of [5′-32P]MDV-1 (→) RNA Fragments. A two-stage reaction was used (30). In the first stage, a 250-μl reaction mixture containing 6 μg of MDV-1 (→) RNA, 25 μg of Qβ replicase, 100 μM [γ-32P]GTP, 200 μM ATP, 40 μM CTP, 12 mM MgCl₂, and 84 mM Tris-HCl (pH 7.5) was incubated at 37°C for 3 min. The absence of UTP results in the cessation of chain elongation when the first uridine is required. In the second stage, the reaction was divided into four aliquots and each aliquot was brought to a volume of 500 μl. All four reactions contained 200 μM CTP, 12 mM MgCl₂, and 84 mM Tris-HCl (pH 7.5). In addition, the first reaction contained 50 μM ATP, 300 μM 3′dATP, 200 μM UTP, and 12.5 mM ITP; the second reaction contained 100 μM UTP, 600 μM 3′dUTP, 200 μM ATP, and 12.5 mM GTP; the third reaction contained 50 μM ATP, 300 μM 3′dATP, 200 μM UTP, and 12.5 mM ITP; and the fourth reaction contained 100 μM UTP, 600 μM 3′dUTP, 200 μM ATP, and 12.5 mM ITP. All four reactions were incu-
bated at 37°C for 10 min and then terminated by bringing each mixture to 1 mg of sodium dodecyl sulfate per ml, 10 mM EDTA, and 400 mM NaCl. Each mixture was extracted with an equal volume of phenol/cresol solution (31). Deproteinization results in the annealing of partially synthesized product strands to their template strands (32). The RNA in each reaction was isolated by gel filtration and precipitation with ethanol, as described (28), and then dissolved in 16 μl of 7 M urea/5 mM Tris-borate, pH 8.3/200 μM Na₂EDTA containing bromophenol blue and xylenol cyanol tracking dyes (Eastman) each at 500 μg/ml. These solutions were then heated at 100°C for 15 sec to melt the partially synthesized product strands free of their templates. Samples (5 μl) of each RNA preparation were then analyzed in parallel by electrophoresis on 8% polyacrylamide slab gels containing 7 M urea.

Electrophoresis. Polyacrylamide slab gels, 405 mm high × 306 mm wide × 0.4 mm thick with slots of a 10 × 0.4 mm cross section, were cast between glass plates in an apparatus obtained from Dan-Kar Plastic Products (Reading, MA). The solution poured into the mold contained 77.3 mg of acrylamide and 2.67 mg of N,N'-methylenebisacrylamide per ml (both obtained from Bio-Rad Laboratories), 600 μg of ammonium persulfate per ml, 50 mM Tris-borate (pH 8.3), 2 mM Na₂EDTA, 7 M urea, and 0.06 μl of N,N,N',N'-tetramethylthelyenediamine (Eastman) per ml. Gels were run in 50 mM Tris-borate, pH 8.3/2 mM Na₂EDTA. Each chamber of the electrophoresis apparatus contained 400 ml of the buffer. The buffer was not recirculated and the gel was not cooled. Gels were prerun at 1000 V for 1 hr; the samples were then loaded and run at 1000 V. After electrophoresis, one glass plate was removed and the gel was covered with SaranWrap and autoradiographed at −80°C with Kodak X-Omat R film in the presence of a Cronex Lightning-Plus intensifying screen (Du Pont).

EXPERIMENTAL DESIGN

Two-Stage Reaction Scheme. Synthesis of MDV-1 RNA by Qβ replicase does not require a primer oligonucleotide, and it always begins with the same 5'-guanosine residue (20), ensuring that all the product strands possess the same 5' end. Although ITP can be substituted for GTP during chain elongation (14), ITP cannot be substituted for GTP in the chain initiation step (33). We therefore used a two-stage reaction (30) to synthesize MDV-1 (−) RNA. In the first stage, pure MDV-1 (+) RNA was incubated with Qβ replicase in the presence of [γ-32P]GTP, ATP, and CTP. Because UTP was omitted from this reaction, chain elongation ceased at the 12th nucleotide, where the first uridine was required. In the second stage, UTP was added, allowing synthesis to continue, and the GTP concentration was diluted 1:1000 by the addition of ITP. Therefore, chain initiation occurred in the presence of GTP, and chain elongation was continued in the presence of ITP. Because all product strands begin with a GTP residue (20), the [γ-32P]GTP present during the first stage labeled the 5' ends.

Synthesis of Specifically Terminated Fragments. We ran four reactions to synthesize fragments for sequence analysis. Each reaction contained a specific 3'-deoxyribonucleoside triphosphate terminator (9). Two types of terminator reactions were run: one contained 3'-dUTP, and the other contained 3'-dATP. The inclusion of a terminator results in the synthesis of different-length MDV-1 (−) RNA fragments that have the same nucleotide at their 3' ends. One reaction of each type contained ITP, and the other contained GTP and served as a control. All four reactions synthesized approximately the same amount of product strands. The terminated fragment sets from each of the four reactions were isolated and analyzed in parallel by electrophoresis on 8% polyacrylamide slab gels containing 7 M urea.

RESULTS AND DISCUSSION

Interpretation of the Sequencing Gels. Fig. 1 shows an autoradiograph of the gel obtained after 5 hr of electrophoresis. Ideally, there should be one distinct band for each adenosine or uridine that occurs in MDV-1 (−) RNA, and the correct order of the adenosine and uridine residues should be indicated by the order in which the bands occur in the two tracks in each panel. A comparison of the results shown in the control tracks (Fig. 1 left) with the actual order of adenosine and uridine residues in the sequence of MDV-1 (−) RNA, shown in Fig. 2, indicates that there are regions of the sequence that are poorly represented by the pattern of bands in the autoradiogram. For instance, the five adenosines that occur in the region between nucleotides 92 and 101 are compressed into a single thick band. Similarly, two adenosines and two uridines in the region between nucleotides 119 and 129 are compressed; and another region of band compression can be seen between nucleotides 141 and 149. Finally, the two fragments terminating in adenosines at positions 61 and 62 migrated farther down the gel than did the shorter fragment terminating in a uridine at position 57. All of these hard-to-interpret regions are located at places in the sequence where the fragments can form a 3'-ter-
minal hairpin. *Thus, the persistance of secondary structures during electrophoresis not only causes band compression but can also result in a longer fragment migrating farther than a shorter one.

In contrast, the bands obtained with the inosine-substituted fragments (Fig. 1 right) are all in the expected order, and there are no regions of band compression. The adenosines between nucleotides 91 and 101 and the adenosines and uridines between nucleotides 119 and 129 are well resolved. In general, the spacing between all the bands correctly reflects the location of each nucleotide in the sequence of MDV-1 (−) RNA. To observe the resolution of the larger fragments, the inosine-substituted RNAs were separated for a longer period of time. Fig. 3 shows an autoradiograph obtained after 10 hr of electrophoresis. The sequence of adenosines and uridines in MDV-1 (−) RNA can be read unambiguously to the end of the molecule. We conclude that inosine substitution markedly improves the accuracy of sequencing procedures that depend on the electrophoretic separation of nucleic acid fragments.

**Relationship between Mobility and Molecular Weight.** Lehrach and his colleagues (34) have demonstrated that a homologous series of RNA fragments that are in the form of random coils (i.e., that do not have secondary structures) should obey the following relationship: the logarithm of the electrophoretic mobility of each fragment should be inversely proportional to the square root of its molecular weight. We measured the distance migrated by each band shown in Fig. 3 and calculated the molecular weight of each of these fragments from its known nucleotide sequence. The square root of each fragment's molecular weight was then plotted against the logarithm of the distance it had migrated (Fig. 4). All the points fell on a straight line. This result indicates that these fragments are structureless, random coils. This plot also provides another demonstration of the resolution that was obtained. Not only can the correct sequence of adenosines and uridines be read from the order of the open and filled symbols along the line, but also the size of the gaps between these points correctly reflects the number of guanosines and cytidines that occur in each region. We therefore conclude that the substitution of inosine for guanosine eliminates secondary structures that interfere with

*Although nucleotides 61 and 62 occur at the top of a hairpin loop in full-length MDV-1 (−) RNA, the two partially synthesized fragments terminating at these nucleotides can form a strong 5' terminall hairpin by the pairing of nucleotides 29 through 39 with nucleotides 49 through 60."
are used to initiate in vitro DNA synthesis (36). Inosine substitution should eliminate the need to run gels at high temperatures, and it may obviate the need to sequence both complementary strands of a DNA molecule. In summary, inosine substitution should improve the resolution of all sequencing techniques that involve the synthesis and electrophoretic separation of DNA or RNA fragments.

We thank S. Spiegelman and Carl Dobkin for helpful advice. This work was supported by American Cancer Society Grant NP-229, National Science Foundation Grant PCM-76-22220, and National Institutes of Health Grant CA-23767.
