Purification and mapping of specific mRNAs by hybridization-selection and cell-free translation

(DNA immobilized on nitrocellulose/DNA-mRNA hybridization/adenovirus 2 DNA/hybrid arrest/cell-free protein synthesis)

ROBERT P. RICCIARDI, JACQUELINE S. MILLER, AND BRYAN E. ROBERTS

Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts 02115

Communicated by Alexander Rich, July 17, 1979

ABSTRACT We describe a simple procedure for isolating specific mRNAs and for mapping them to the regions of the DNA from which they originate. The method involves hybridization of total cytoplasmic RNA to restriction fragments of DNA that have been fractionated in agarose gels and immobilized on nitrocellulose filters. The hybridization-selected RNAs are eluted and translated in a cell-free system in order to identify their encoded polypeptides. Optimal hybridization and filter washing conditions are given for selection of mRNAs from adenovirus 2-infected cells and transformed cells.

Synthesis of individual polypeptides directed by purified mRNAs in a cell-free translation system provides compelling evidence for the expression of specific genes. The ability to purify functional mRNAs is a prerequisite for understanding some of the detailed mechanisms involved in gene regulation. Characterization of purified mRNAs by cell-free translation, in combination with other techniques, can yield information about the spatial and topological arrangement of transcripts along the genome, the location of protein coding regions within the mRNA, mRNA structure, and mRNA abundance.

Several methods for isolating specific mRNAs have been developed. DNA, either in liquid or bound to immobilized supports, has been used to hybridize and thereby sequester specific mRNAs from the plethora of cellular RNA species. These methods have included isolation of RNA-DNA hybrids by: selective binding to hydroxylapatite (1); selective exclusion through agarose (2) or Sepharose 4B (3); the use of DNA covalently bound to cellulose (4, 5) or Sepharose (6); DNA bound directly to nitrocellulose (7-9); DNA enzymatically synthesized and covalently bound to oligo(dt)-cellulose (10).

Here we describe an efficient method by which specific mRNAs can be purified and used to determine the location of these mRNAs with respect to their DNA coding regions. This method relies on hybridization of total cytoplasmic RNA to restriction fragments of DNA which have been immobilized on nitrocellulose filters. The hybridization-selected mRNAs are eluted from the DNA and identified by the polypeptide products that are synthesized in a reticulocyte cell-free system. The map positions of the mRNA transcripts on the DNA can be determined directly because the genomic coordinates of the DNA restriction fragments are known.

This procedure has several advantages. Most importantly, purification of the DNA restriction fragments is not required. Rather, DNA restriction fragments are fractionated by electrophoresis in agarose gels and then directly transferred to the nitrocellulose filter membrane by the method of Southern (11). In this way, many different DNA fragments may be easily handled in a single experiment. In addition, the nitrocellulose filters that contain the bound DNA fragments can be used several times. By this procedure, it is possible to identify the translation product of a specific mRNA which is otherwise obscured by the large number of polypeptides that are synthesized when total cellular RNA is used in the cell-free system. Applications of this technique are demonstrated by analysis of mRNAs from adenovirus 2-infected cells.

EXPERIMENTAL PROCEDURES

DNA and RNA Extraction. DNA was extracted from purified adenovirus 2 virions (12). RNA from adenovirus 2-infected cells and from adenovirus 2-transformed cells was prepared by washing cells twice with ice-cold phosphate-buffered saline. The cells were lysed with 0.65% Nonidet P-40 in 0.15 M NaCl/1.5 mM MgCl2/0.01 M Tris-HCl, pH 7.9. The nuclei were sedimented and the supernatant was centrifuged at 10,000 rpm for 10 min and adjusted to a final concentration of 1 mM EDTA (pH 7.9) and 0.5% NaDodSO4. The solution was extracted twice with chloroform/phenol, 1:1 (vol/vol), equilibrated with 0.01 M Na acetate, pH 6.0/100 mM NaCl/1 mM EDTA and twice with chloroform/isoamyl alcohol, 24:1 (vol/vol). The RNA was precipitated with 2.5 vol of ethanol/0.2 M Na acetate, pH 5.5. After the pellet was washed with 70% ethanol, the RNA was dissolved in water and frozen.

Preparation and Fractionation of DNA Restriction Fragments. Complete digestion of adenovirus 2 DNA was carried out with restriction enzymes purchased from and used according to procedures outlined by New England Biolabs. Either preparative amounts (50-200 μg) of total adenovirus 2 DNA or purified restriction fragments prepared by the method of Vogelstein and Gillespie (13) were digested to completion and fractionated by electrophoresis in a 1% agarose gel in 0.1 M Tris-HCl, pH 7.7/36 mM Na2HPO4/1 mM EDTA.

Selection and Isolation of RNA by Hybridization to DNA Immobilized on Nitrocellulose. DNA restriction fragments generated by endonuclease digestion of 1–3 μg of either total adenovirus 2 DNA or a purified fragment were fractionated by electrophoresis in a 20 × 23 × 0.6 cm, 1% agarose gel in the Tris/phosphate/EDTA buffer described above. After the gel was stained with ethidium bromide (0.5 μg/ml), fragments of DNA were visualized by fluorescence under UV light and their location was marked by making notches in the gel on each side of the DNA fragment. The DNA in the gel was then denatured, neutralized, and transferred to nitrocellulose filter paper (Schleicher & Schuell BA-85-R597) according to the method of Southern (11). During this transfer procedure, the salt concentration was maintained at 3 M NaCl and 0.3 M Na citrate. After transfer, the location of each DNA fragment on the nitrocellulose was marked with a blue ink ball-point pen by using the notches in the gel as references. The nitrocellulose was washed in 1.5 M NaCl/0.5 M Na citrate by gently rubbing across the area of the DNA fragment with a gloved finger to remove any adherent agarose. The filter was then soaked for

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.

Abbreviations: NaDodSO4, sodium dodecyl sulfate; Pipes, 1,4-piperazinediethanesulfonic acid.
60 min, with intermittent vortexing. Excess wash buffer was removed from the nitrocellulose by blotting with Whatman 3 MM filter paper. The nitrocellulose was air-dried overnight and baked at 75°C for 2 hr under reduced pressure.

The portions of nitrocellulose that contained isolated DNA fragments were excised, cut into small pieces, and put into 1.5-ml polyethylene microcentrifuge tubes. A 100-μl hybridization mixture consisted of 65% deionized formamide (Matheson, Coleman and Bell), RNA of varying concentrations (see figure legends), 10 mM 1,4-piperazinediethanesulfonic acid (Pipes) at pH 6.4, and 0.4 M NaCl. The reaction mixture was spun for 10 sec in a Microfuge, vortexed, spun again, and gently vortexed by finger to resuspend the filters. After incubation at 50°C for 1–2 hr with gentle shaking, the reaction mixture was removed by aspiration and the filters were washed in 1 ml of 0.15 M NaCl/0.015 M Na citrate/0.5% NaDODSO4 10 times. This wash buffer was carefully maintained at 60°C throughout the wash procedure. The tubes were vortexed for several seconds after each addition of wash buffer. The final wash was done twice in 1 ml of 2.0 mM EDTA at pH 7.9. Finally, 1 ml of 2.0 mM EDTA at pH 7.9 was added to the tubes, incubated at 60°C for 5 min, and then removed by aspiration.

The RNA was eluted from the hybrid by boiling the filters for 60 sec in 500 μl of double-distilled water and then quick-frozen in a methanol/dry ice bath. The liquid was removed and brought to a final concentration of 0.2 M Na acetate, and 20 μg of calf thymus tRNA (Boehringer Mannheim) was added. The RNA was precipitated with 2.5 vol of ethanol at -20°C. Prior to translation, the RNA was pelleted at 12,000 × g for 10 min at 4°C, washed twice with 70% ethanol, and dried under reduced pressure.

Alternatively, total DNA or purified fragments can be bound directly to the nitrocellulose. The DNA is denatured by boiling for 30 sec, quick-frozen, and thawed. The DNA is pipetted onto a small square of nitrocellulose which is air-dried and baked and the hybridization and wash procedure are as described.

Translation and Polypeptide Analysis. RNA was translated in the reticulocyte system (New England Nuclear) as described (14). Polypeptides were fractionated in NaDODSO4/10% polyacrylamide gels (15) and fluorography was carried out by the method of Bonner and Laskey (16).

Hybrid-arrested translation was as described (17).

Strand Separation by Agarose Gel Electrophoresis. DNA fragments were denatured as described by Sharp et al. (18). Approximately 1–2 μg of a purified fragment was dissolved in 34 μl of water and 12 μl of 1 M NaOH and incubated for 5 min at room temperature. Twelve microliters of Tris/phosphate gel buffer (described above) containing 0.5% bromophenol blue and 60% sucrose was added to the sample. The sample was immediately loaded onto a 20 × 23 × 0.6 cm 1% agarose gel maintained at 20 V. The voltage was increased to 150 V for 6 hr at 4°C. After the gel was stained with ethidium bromide (0.5–1.0 μg/ml) to locate the DNA strands, it was soaked in 3 M NaCl/0.5 M Tris-HCl, pH 7.0, for 20 min and the DNA was transferred to nitrocellulose as described above.

RESULTS

An example of this hybridization-selection is the assignment of late adenovirus 2 RNAs to specific locations along the genome. Adenovirus 2 DNA was digested with Bam HI; the four fragments were fractionated on an agarose gel and transferred to nitrocellulose filters by the blotting method of Southern (11). The portions of the filter containing individual immobilized DNA fragments were then utilized to select late cytoplasmic RNAs. The RNAs were eluted from the DNA hybrid and translated in the reticulocyte cell-free system. Fig. 1 shows the results of this analysis. The Bam B fragment (map coordinates, 0–29) selects for RNAs encoding the entire complement of late proteins. Fragment Bam D (map coordinates, 29–42) selects transcripts coding for polypeptides III, IIIa, and 52, 55K. The Bam C fragment (map coordinates, 42–59.5) selects mRNAs that code for polypeptides II, III, V, pVI, pVII, and HA as well as unidentified minor proteins. Finally, the Bam A fragment (map coordinates, 59.5–100) selects for transcripts encoding II, 100K, 72K, IV, 33K, pVI, pVIII, and HA in addition to other minor proteins.

In some cases, adjacent DNA fragments select mRNAs that direct the synthesis of the same polypeptide. For example, Bam D and Bam C each select RNAs encoding polypeptide III (Fig. 1, lanes 2 and 3) and fragments Bam C and Bam A each select for transcripts encoding for polypeptide pVI (Fig. 1, lanes 3 and 4). In other cases, nonadjacent fragments select functionally identical transcripts. For example, transcripts selected by Bam B are also selected by the nonadjacent fragments Bam C and

![Fig. 1.](image-url)
Bam A (Fig. 1, lanes 1, 3, and 4). Selection by adjacent fragments could mean either that the transcript spans the cleavage site of the restriction endonuclease or that the mRNA is derived from noncontiguous regions on the DNA. These two alternatives can be distinguished by hybridization-selection to smaller DNA fragments generated by subdigestion of larger fragments. Thus far, the smallest fragment successfully used in the hybridization selection has been a fragment of \( \approx 350 \) base pairs (map coordinates, 41–42) which selects only the mRNA encoding polypeptide III (Fig. 1, lane 6).

Selection by nonadjacent fragments can be explained if the sequences within the transcripts are derived from noncontiguous regions of the genome. Most adenovirus mRNAs, isolated from cells during late infection, contain a tripartite leader sequence of 150 nucleotides at their 5' termini that originate from three distinct regions in Bam B at map coordinates 16.6, 19.6, and 26.6 (19, 20). The three sequences that constitute the tripartite leader were separated by subdigestion of the DNA followed by electrophoresis in agarose. The fragments were then transferred to nitrocellulose. The individual fragments obtained by this procedure were fragments with map coordinates 8–17, 17–23.5, and 23.5–31.5. As shown in Fig. 1, each leader sequence selects for transcripts coding for most of the late polypeptides. In addition, fragment 8–17 exclusively selects RNAs coding for polypeptides IX and IVa2 (Fig. 1, lane 8) which are known to be coded by DNA in this region (20, 21). Selection of late mRNAs by the individual leader sequences demonstrates that a sequence homology of 50 nucleotides forms a sufficiently stable hybrid to allow for selection.

In addition to using DNA fragments to locate mRNAs to specific regions of the genome, hybridization-selection using separated strands of these fragments permits the assignment of specific transcripts to the strand of origin. DNA fragments are denatured in 300 mM NaOH and fractionated in a neutral 1% agarose gel. The transfer of DNA, hybridization-selection, and translation of mRNAs are carried out as described. Fragment EcoRI E (map coordinates, 84–89.7) exclusively selects a transcript coding for polypeptide IV (Fig. 2, lanes 5 and 6). The selection on separated strands of EcoRI E demonstrates that the faster migrating strand (F strand) hybridizes the transcript that codes for polypeptide IV whereas the slower migrating strand (S strand) does not. The EcoRI F fragment (map coordinates, 70.7–75.9) selects transcripts that code for polypeptides 100K, 72K, 33K, and pVIII (Fig. 2, lanes 3 and 4). The S strand of EcoRI F only selects the transcript that codes for the 72K polypeptide. The F strand of EcoRI F selects those transcripts that code for polypeptides 100K, 33K, pVIII, and 72K.

Although hybridization-selection defines the location of the constituent sequences of the mature mRNAs along the genome, it does not directly position the protein coding sequences within the transcript. Hybrid-arrested translation allows for the determination of protein coding sequences within the transcript by virtue of the fact that DNA-RNA hybrids formed within these coding sequences block the translation of that specific RNA (17). The use of hybridization-selection and hybrid-arrested translation in concert permits the location of specific mRNAs along the genome and the positioning, within this transcript, of the protein coding sequences. Fragment Bam C (42.0–59.5) selects mRNAs encoding polypeptides III, pVII, V, pVI, and II (Fig. 1, lane 3). The synthesis of all five polypeptides is arrested by fragment Bam C (Fig. 3, lanes 1 and 2). The RNA encoding polypeptide III also was selected by Bam D (29.0–42.0) (Fig. 1, lane 2). When RNA selected by Bam C was hybridized with DNA fragment Bam D, only the synthesis of polypeptide III was arrested (Fig. 3, lanes 3 and 4). This suggests that the coding sequence of the mRNA for polypeptide III lies near the 5’ terminus and to the left of the sequences coding for polypeptides pVII, V, and pVI and II.

**DISCUSSION**

DNA fragments fractionated by agarose gel electrophoresis, transferred, and immobilized on nitrocellulose filters are used to select mRNAs by hybridization. After a stringent washing procedure to remove nonspecifically bound RNAs, the hybridized RNAs are eluted and the polypeptides that they encode are defined by translation in a cell-free system. It should be noted that by this procedure mRNAs are functionally defined by their capacity to direct the synthesis of specific polypeptides in a cell-free system. We have demonstrated the efficiency of this method for locating the genomic origins of constituent sequences within mRNAs.

This is best illustrated by the mapping data pertinent to

---

**FIG. 2.** Fluorograph of \(^{35}S\)methionine-labeled late adenovirus polypeptides synthesized in a reticulocyte cell-free system and fractionated in a NaDODSO4/10% polyacrylamide gel. Details as in Fig. 1, except that adenovirus 2 DNA (1–3 \( \mu \)g) also was incubated in the presence of 0.26 M NaOH for 5 min at room temperature. Samples contained RNA selected by the separated strands of the following restriction fragments: lane 3, slow-migrating strand of EcoRI F (map coordinates, 70.7–75.9); 4, fast-migrating strand of EcoRI F; 5, slow-migrating strand of EcoRI E (84.0–89.7); 6, fast-migrating strand of EcoRI E; 1, cytoplasmic extracts of cells labeled with \(^{35}S\)methionine 35 hr after infection; 2, no added RNA.

Often the protein product of a specific mRNA cannot be convincingly visualized in a polyacrylamide gel among the many other polypeptides. This is exemplified early in adenovirus 2 infection when host cell protein synthesis is not yet inhibited and also in adenovirus 2-transformed cells in which specific viral proteins are expressed in addition to the plethora of host cell proteins. The translation of total cytoplasmic RNA isolated early during infection or from transformed cells results in the complex pattern of protein products depicted in Fig. 4, lanes 5 and 6, respectively. Hybridization-selection of cytoplasmic RNAs from these cells to total adenovirus 2 DNA and subsequent translation yielded a representative profile of early polypeptides (Fig. 4, lane 3) or a single predominant viral polypeptide from transformed cells (Fig. 4, lane 4).
polypeptide III. The mRNA encoding polypeptide III was selected by restriction fragments Bam B (0–29), Bam D (29–42), and Bam C (42–59.5) but not by Bam A (59.5–100) as shown in Fig. 1. This indicates that this mRNA originates within the left-most 59.5% of the genome. This region of the DNA was digested by other restriction enzymes, and the fragments generated were used to select mRNA encoding polypeptide III. The map coordinates of the fragments selecting this mRNA were 8–17, 17–23.5, and 23.5–31.5 (Fig. 1) and also 27.3–41, 41–42, 42.7–47.4, and 47.4–50.1 (data not shown). The fact that noncontiguous fragments select the mRNA encoding polypeptide III implies that its constituent sequences originate from disparate regions along the genome. Those DNA fragments (map coordinates 8–17, 17–23.5, and 23.5–31.5) contain the noncoding leader sequences located at map positions 16.6, 19.6, and 26.6 (19, 20); these three leader sequences, each approximately 50 nucleotides, occur at the 5’ proximal ends of most late mRNAs. The presence of these leader sequences accounts for the selection of these mRNAs, including that coding for polypeptide III, at these locations within the genome (Fig. 1).

Furthermore, the ability of DNA fragments containing these individual leader sequences to select these mRNAs demonstrates that a DNA-RNA hybrid of approximately 50 nucleotides is selected stably by this procedure. Location of the remainder of the transcript that encodes polypeptide III is indicated by the failure of the DNA fragments between map coordinates 28.5 and 37.3 and those to the right of 50.1 to select this transcript. However, the transcript is selected by the four fragments within map coordinates 37.3 and 50.1 (unpublished results). The protein coding regions within this transcript have been located between map coordinates 37.3 and 47.4 by hybrid-arrested translation (unpublished data). Furthermore, the DNA fragment containing the leader sequences (map coordinates 0–29) does not arrest the translation of the RNA encoding polypeptide III (17). Therefore, the mRNA that encodes polypeptide III contains sequences at its 5’ end that are derived from noncontiguous regions of the DNA. These sequences are directly fol-
lowed by sequences that are essential for its translation and then by a substantial 3′ untranslated portion.

This example clearly illustrates that hybridization-selection locates sequences within the genome that constitute the mRNA, whereas hybrid-arrested translation defines which sequences are essential for translation. Therefore, for the purpose of determining the genomic organization of mRNAs and the arrangement of their coding and noncoding regions, it is essential to correlate information derived from both hybridization-selection and hybrid-arrested translation.

We have shown hybridization-selection is a simple and reproducible method for isolation and mapping of functional mRNAs. After size fractionation on methylmercury hydroxide gels, the eluted RNAs which are defined by translation are indistinguishable in size from unselected RNAs (data not shown). However, successful application of the technique relies on a number of crucial methodological considerations. In these experiments, the smallest length of DNA sequence stably immobilized to nitrocellulose and used to select mRNA was approximately 350 base pairs (Fig. 1, lane 6). The minimal size of the DNA-mRNA hybrid required for reproducible selection was approximately 50 nucleotides (Fig. 1, lanes 8, 9, and 10). Two steps in this procedure are essential in the elimination of nonspecifically bound RNAs: (i) after the transfer of the DNA to the filter but prior to baking, the membrane is purged of any adhering agarose by gently massaging the surface of the filter; (ii) after selection, the filters are exhaustively washed with buffers carefully maintained at a predetermined temperature (in these experiments, the critical temperature was 60°C). It should be noted that the thermal stability of a DNA-RNA hybrid is a function of the G+C content of the DNA and therefore care should be taken to ensure that certain hybridized mRNAs are not selectively eluted during washing (9). The hybridization conditions outlined here are optimal for the selection of adenovirus 2 RNA. However, in other systems the concentration of formamide used and the time and temperature of incubation may be different. In order to maximize the selection of specific RNAs and to minimize nonspecifically bound RNAs, the appropriate concentration of RNA also should be determined experimentally.

This hybridization-selection procedure has a number of distinct advantages. Elimination of the requirement for the purification of DNA fragments is one obvious advantage. Also, many hybridization-selection reactions can be carried out simultaneously. Furthermore, nitrocellulose filters have been reused in selection reactions at least four times with no obvious diminution in hybridization-selection efficiency. Refined mapping of the RNA transcripts can be achieved by the use of restriction fragments of DNA as small as 350 base pairs, which can be immobilized on the nitrocellulose filter. Hybrids of DNA and mRNA as small as 50 nucleotides long can be reproducibly selected. The application of hybridization-selection is demonstrated in locating the constituent sequences of an mRNA on the genome.

There exist two possible but somewhat improbable complications that could confound a simple interpretation of the data obtained from hybridization-selection. A constituent sequence of an inverted repeat in the DNA which is transcribed into mRNA would dictate the possible existence of an homologous sequence outside the regions of transcription that would select that mRNA. Alternatively, the occurrence of overlapping transcripts that originate from opposite DNA strands may result in hybridization between these RNAs. If a portion of one of these transcripts is hybridized to a strand of DNA and if the unhybridized segment contains sequences homologous to the other RNA, then an RNA that is not complementary to the DNA fragment used could be selected.

Applications of hybridization-selection include: (i) selection of mRNAs by small DNA fragments, which permits refined genomic mapping of these RNAs; (ii) elucidation of specific mRNAs whose cell-free products are not obvious in the complex pattern of products directed by total RNA; (iii) discrimination of comigrating cell-free products originating from the translation of different mRNAs; (iv) preparative purification of specific mRNAs; (v) correlation of precise size and genomic organization of these transcripts with their coding and noncoding sequences by S1 nuclease analysis and hybrid-arrested translation on selected mRNAs (unpublished data); and (vi) identification of recombinant DNAs as to those containing constituent sequences of an mRNA encoding a specific polypeptide.