ELIMINATION OF THE SECONDARY STRUCTURE EFFECT IN GEL SEQUENCING OF NUCLEIC ACIDS

N. S. AMBARTSUMYAN and A. M. MAZO
Institute of Molecular Biology, USSR Academy of Sciences, Moscow 117984, USSR

Received 10 March 1980

1. Introduction

Two basic approaches are now used in sequencing nucleic acids: (1) the introduction of an end label into the molecule and subsequent specific, chemical or enzymatic digestion of polynucleotides [1–4]; (2) that based on enzymatic synthesis using DNA and RNA polymerases [5–9]. In either case, the products are separated by thin-layer electrophoresis in polyacrylamide gels with 7 M urea [1,10]. The disadvantage that the two methods have in common is abnormal mobility of certain oligonucleotides in electrophoresis (compression) which is usually attributed to their stable secondary structure. These structured regions may be melted due to the high temperature of sequencing gel [4].

Here we have modified cytosine residues in DNA by a mixture of methoxyamine and bisulfite. This treatment destroys the formation of GC base pairs and thus eliminates the effect of the secondary structure on the mobility of DNA fragments in gels. We have found that if cDNA synthesized on a template of phage MS2 RNA using reverse transcriptase in the presence of dideoxynucleoside triphosphates is subjected to such a treatment, further electrophoretic separation is considerably improved and certain ambiguities in the sequence are clarified.

This strategy in eliminating the 'compression' has been used earlier in RNA sequencing [11]. Therefore, chemical modification that blocks the formation of secondary structure is of universal significance in gel-sequencing determination of the primary structure of nucleic acids.

2. Materials and methods

2.1. *Materials*

Phage MS2 RNA was a generous gift of Dr V. Berzin' (Institute of Organic Synthesis, Latvian SSR Academy of Sciences). The oligonucleotide $^{5'}$CTCATGTT$^{3'}$ complementary to an intercistronic region between the genes for coat protein and for RNA replicase of phage MS2 was a kind gift from N. F. Sergeyeva and V. Veiko (Moscow State University). Protein S1 from the small subunit of *Escherichia coli* ribosomes was prepared in the laboratory of Professor S. E. Bresler (P. Konstantinov Institute of Nuclear Physics, USSR Academy of Sciences). RNA-dependent DNA polymerase was a generous gift of Professor J. Beard (Life Sciences, USA). Deoxynucleoside triphosphates (dNTP) and dideoxynucleotide triphosphates (ddNTP) were purchased from P-L Biochemicals, [α-^{32}P]dATP, dCTP and dGTP (400 Ci/mmol), from the Radiochemical Centre (Amersham). Reagents for electrophoresis were from Reanal (Hungary). All other reagents used in experiments were of the highest purity grade.

2.2. *DNA synthesis in the presence of chain terminators*

The reaction was conducted in two variants:

(1) The incubation mixture (10 μl) contained 50 mM Tris–HCl (pH 8.3), 6 mM $MgCl_2$, 4 mM dithiotreitol, 400 μg MS2 RNA/ml, a 40-fold molar excess of the primer, 5 μM dNTP, and 1 μCi of one of the labeled triphosphates. ddNTP were added to the incubation mixture at the following

molar ratios: ddTTP:dTTP, 2:1; ddATP:dATP, 4:1; ddGTP:dGTP, 1:1; ddCTP:dCTP, 2:1. The reaction was started by adding the enzyme (3 units of activity), and was performed for 30 min at 46°C, then the concentration of 3 dNTP was raised to 100 μM, the dNTP for which ddNTP was added to 25 μM and ddNTP to the above ratio, and 1 unit of activity of the enzyme was added. The incubation was carried on for a further 1.5 h at 46°C.

(2) The incubation mixture (10 μl) was the same with the exception of ddNTP. The incubation was for 12 min at 46°C with the following addition of a mixture containing 100 μM of the 3 dNTP 5 μM of the triphosphate for which the terminator was introduced, and ddNTP at the following molar ratios: ddATP:dATP, 10:1; ddTTP:dTTP, 5:1; ddCTP:dCTP, 2.5:1; ddGTP:dGTP, 5:1. After incubation for 60 min at 46°C, 1 unit of activity of the enzyme was added and the total mixture was incubated for a further 60 min. Once the synthesis was over, the samples were placed in ice, and the material was precipitated with ethanol adding yeast carrier tRNA

2.3. *Modification with a mixture of methoxyamine and bisulfite*

The dried samples were dissolved in 10 μl of a solution of 1.5 M methoxylamine, 1.5 M KCl (pH 5.0) and 1 M $Na_2S_2O_5$ and kept for 4 h at 37°C [11]. The samples were diluted with water to 1 ml, 20 μg carrier tRNA was added and the material was precipitated by adding 1:10 (v/v) cetyltrimethylammonium bromide (10 mg/ml) in 1 M Na-phosphate buffer (pH 5.0). After centrifugation at 6000 rev./min for 15 min the pellet was dissolved in 2 M NaCl and precipitated with ethanol; the precipitate was dissolved in 0.2 M sodium acetate (pH 5.0) and again precipitated with ethanol.

2.4. *Preparation of samples for electrophoresis*

The precipitates were dissolved in 5 μl H_2O, and 5 μl formamide with 0.01% xylencyanole was added. The mixture was heated at 100°C for 1 min, cooled in an ice bath and then applied to the gel. Electrophoresis in 10% polyacrylamide gel with 7 M urea was performed as in [1]. The dimensions of the gel were 60 × 20 × 0.06 cm. After electrophoresis, the gel was covered with Saran-Wrap and exposed with RT-1 X-ray film at −70°C.

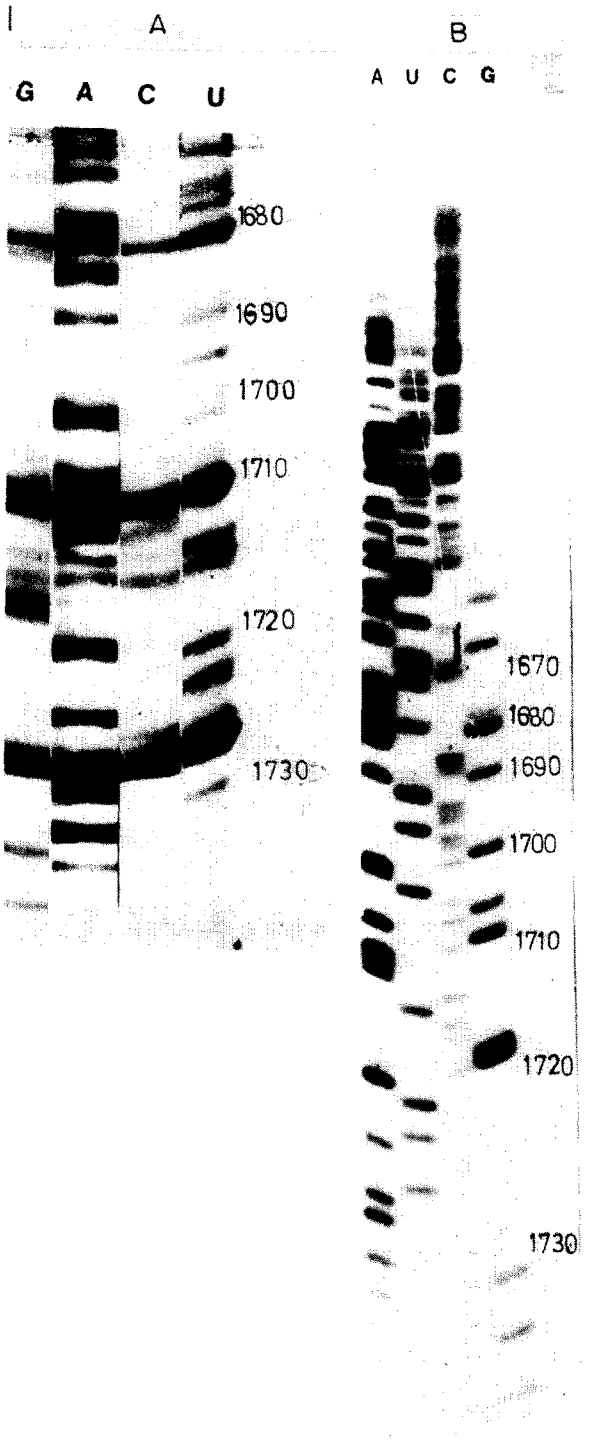

Fig.1. Autoradiograms of a part of the gel with the sequence of phage MS2 RNA. (A) Conditions of synthesis: section 2.2.(1). (B) Samples after synthesis were treated with a mixture of methoxyamine and bisulfite. Electrophoresis was run at 1500 V for 4 h. Nucleotides are numbered from the 5'-terminus of the molecule.

3. Results and discussion

In contrast to DNA polymerase, reverse transcriptase has higher affinity for dideoxynucleoside triphosphates (terminators) [12]. Therefore, for reverse transcriptase to synthesize effectively sufficiently long products (cDNA), the ddNTP: dNTP ratio varies from 0.25:1–10:1, whereas in the case of DNA polymerase, the ratio is 100:1 [9,13–16]. Moreover, this ratio must be determined for each terminator. If work is done with various batches of terminators and different templates, the conditions of the synthesis must also be changed [13–16].

Fig.1A shows part of the gel with a known sequence, A_{1675}–G_{1738}, of RNA phage MS2 [17]. Here, we preliminarily selected the ratio between the terminator and the dNTP for each of the 4 pairs. However, though we made an attempt to optimize the conditions, one can see that the intensity of the bands varies among different tracks, thus complicating the sequencing. Furthermore, bands common for all the terminators (1728, 1710, 1680) can be discerned in several positions in the gel; these bands do not make it possible to establish the nature of bases in these positions of the sequence.

We presumed that such common bands appeared either as the result of stable secondary structure present in these regions on the RNA template (thus complicating their transcription by reverse transcriptase), or originated from 'compression' in the gel caused by the RNA template incompletely separated from cDNA. However, attempts to get rid of these bands by either using alkaline hydrolysis of the RNA template (0.3 M KOH, 37°C, 16 h) or adding to the incubation mixture protein S1 from the small subunit of *E. coli* ribosomes that melts the secondary structure of RNA was unsuccessful (not shown).

These common bands almost entirely disappear from all columns if fresh RNA preparations are used (fig. 2A). We believe that specific breaks appear in the RNA molecule in the course of its storage; these breaks account for significant 'structural stops' in cDNA synthesis.

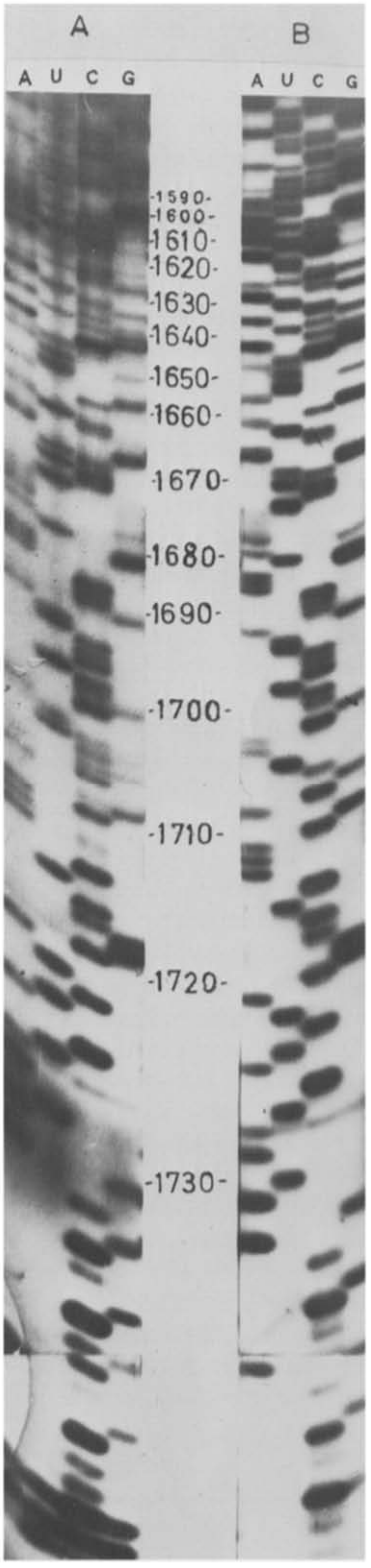

Fig.2. Autoradiograms of the gel with the sequence of phage MS2 RNA. (A) Conditions of synthesis: section 2.2.(2). (B) Samples after cDNA synthesis were treated with a mixture of methoxyamine and bisulfite. Conditions of the electrophoresis and numbering of nucleotides were as in fig.1.

As can be seen in fig.1B, if terminators are added at the beginning of the reaction (variant (1), section 2.2), differences in the length of cDNAs produced are observed, besides differences in the intensity of bands in different tracks. In order to eliminate these differences, we changed the conditions of cDNA synthesis (fig.2A). In this case, terminators were added 12 min after the beginning of the synthesis (variant 2, section 2.2). Under these conditions, the difference in the intensity of bands was considerably diminished, as well as that in the length of the products. The effectiveness of termination however did not change. The addition of terminators (with the increase in their relative content) sharply decelerated the synthesis; therefore the length of the products was predetermined mainly at the first stage of the reaction (without ddNTP) if terminators were added after the beginning of the reaction.

Fig.2A shows that though considerable improvement has been achieved as compared to fig.1, the sequence can be read for a relatively short region and, moreover, some regions remain ambiguous (1703–1708, 1718, and upstream from 1670).

In order to improve separation in the gel, we modified cDNA with a mixture of methoxyamine and bisulfite which resulted in the selective and quantitative conversion of cytidine into 5,6-dihydro-6-sulfo-N^4-methoxycytidine [18]. As can be seen in fig.1B, such a modification significantly improves the separation of polynucleotides. The sequence is read easily and unambiguously; moreover, the sequence being read is much longer in fig.2B than in fig.1B, 2A, and reaches 150 nucleotides.

As we have shown, the modification makes the secondary structure of tRNA and 5 S RNA unfold considerably, which may be attributed apparently to weakening of GC base pairs. As a result, 'compressions' caused by the secondary structure of RNA disappear from the sequencing gel.

Therefore, the strategy for elimination of defects in electrophoretic separation of polyribonucleotides in [19] is applicable, as has been shown here, for DNA and may be of universal significance in gel-sequencing of nucleic acids.

The sequence of the region 1710–1580 slightly differs from that determined in [17]. According to our data, positions A_{1699} and G_{1663} [17] are occupied by cytosines. Presumably, these differences may be attributed to strain differences in phage MS2.

When this manuscript was in preparation, we discovered a publication [19] in which 'compression' of the products of Qβ replicase synthesis was eliminated by adding ITP rather than GTP as a precursor to the incubation mixture.

Acknowledgements

The authors are greatly indebted to Professor L. L. Kisselev for his stimulating discussions and continued support in this work, to Professor J. Beard for a generous gift of reverse transcriptase, to V. Berzin', N. Sergeyeva and V. Veiko for kind gifts of the phage MS2 RNA preparations and the oligonucleotide $^{5'}$CTCATGTT$^{3'}$.

References

[1] Maxam, A. M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–564.
[2] Donis-Keller, H., Maxam, A. M. and Gilbert, W. (1977) Nucleic Acids Res. 4, 2527–2538.
[3] Simoncsits, A., Brownlee, G. G., Brown, R. S., Rubin, J. R. and Guilley, H. (1977) Nature 269, 833–836.
[4] Peattie, D. A. (1979) Proc. Natl. Acad. Sci. USA 76, 1760–1764.
[5] Sanger, F. and Coulson, A. R. (1975) J. Mol. Biol. 94, 4441–4448.
[6] Brownlee, G. G. and Cartwright, E. M. (1977) J. Mol. Biol. 114, 93–117.
[7] Kramer, F. R. and Mills, D. R. (1978) Proc. Natl. Acad. Sci. USA 75, 5334–5338.
[8] Axelrod, V. D., Vartikyan, R. M., Aivazashvili, V. A. and Beabealashvili, R. Sh. (1978) Nucleic Acids Res. 5, 3549–3563.
[9] Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 76, 731–735.
[10] Sanger, F. and Coulson, A. R. (1978) FEBS Lett. 87, 107–110.
[11] Mazo, A. M., Mashkova, T. D., Avdonina, T. A., Ambartsumyan, N. S. and Kisselev, L. L. (1979) Nucleic Acids Res. 7, 2469–2482.
[12] Smoler, D., Molineux, I. and Baltimore, D. (1971) J. Biol. Chem. 246, 7697–7700.
[13] McGeoch, D. J. and Turnbull, M. T. (1978) Nucleic Acids Res. 6, 4007–4024.
[14] McGeoch, D. J. and Dolan, A. (1979) Nucleic Acids Res. 6, 3199–3211.
[15] Zimmern, D. and Kaesberg, P. (1978) Proc. Natl. Acad. Sci. USA 75, 4257–4261.
[16] Bina-Stein, M., Thoren, M., Salzman, N. and Thompson, J. A. (1979) Proc. Natl. Acad. Sci. USA 76, 731–735.
[17] Fiers, W., Contreras, R., Duernick, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G. and Ysebaert, M. (1976) Nature 260, 500–507.
[18] Sverdlov, E. D., Monastyrskaya, G. S., Gushkova, L. L., Levitan, T. L., Scheichenko, V. I. and Budovsky, E. I. (1974) Biochim. Biophys. Acta 340, 153–165.
[19] Mills, O. and Kramer, F. R. (1979) Proc. Natl. Acad. Sci. USA 76, 2232–2235.