Nucleotide Sequencing of DNA: Preliminary Characterization of the Products of Specific Cleavages at Guanine, Cytosine, or Adenine Residues

(bacteriophage M13/ribosubstitution/DNA polymerase I/electrophoresis/two-dimensional fingerprinting)

WINSTON SALSER, KIRK FRY, CLIFFORD BRUNK, AND RAYMOND POON

Department of Zoology and Molecular Biology Institute, University of California at Los Angeles, Los Angeles, Calif. 90024

Communicated by H. W. Magoun, November 8, 1971

ABSTRACT DNA synthesized in vitro from a phage M13 template has been cleaved at either guanine, adenine, or cytosine residues by ribosubstitution techniques. Fingerprints of the fragments obtained suggest that DNA sequencing will be possible with this technique.

Powerful techniques are now available for the determination of nucleotide sequences in RNA. The two-dimensional fingerprinting methods introduced by Sanger and his collaborators (1) have been especially important in increasing the rapidity with which moderately large nucleotide sequences can be determined; nevertheless, our ability to sequence very large nucleic acid molecules by the present techniques is extremely limited. DNA sequencing offers some important alternatives and advantages. For instance, there are available sequence-specific DNases, such as the Hemophilus restriction enzyme, endonuclease R, which cleaves at the sequence . . . pGpTp-PypPupApCp . . . (2) to yield specific fragments averaging about 1000 nucleotide-pairs in length. Various other specific nucleases are available to make either larger or smaller unique fragments, providing a much more powerful approach than the partial digestions with RNase T₁ that are used to obtain large specific fragments for RNA sequencing. Portions of the molecule especially sensitive to nuclease attack may be lost in partial digestions with T₁; RNase. Also, such digestions produce very complex mixtures that are difficult to work with because each of the nuclease-resistant sequences is usually distributed among fragments of several different sizes.

Another advantage of DNA sequencing comes from the fact that DNA polymerase I can initiate synthesis only by addition to the 3'-OH end of a primer molecule complementary to the DNA template (3). Specific primer oligonucleotides can be obtained in various ways, and by the use of these oligonucleotides, it should be possible to make DNA polymerase synchronously replicate the template, starting at a chosen initiation point, so that a specific region can be labeled for sequence studies. A similar approach has been used by Billeter et al., (4) in sequencing Q₈ RNA, but their approach is not a general one since it depends on the Q₈ RNA polymerase, which replicates only Q₈ RNA. DNA polymerase I, however, should provide a truly general approach that will allow synchronous in vitro replication of many different DNA molecules of biological interest.

Since it is not yet possible to take full advantage of such approaches, DNA sequencing is still in a much more primitive state than RNA sequencing. The longest DNA sequence published is 12 nucleotides, from the "sticky ends" of phage λ (5). The main difficulty that has prevented the determination of longer DNA sequences is the lack of any method for base-specific cleavage of DNA comparable to that provided by T₁ RNase, which hydrolyzes RNA at Gp residues. Since no base-specific DNases have been found, it has been necessary to use less specific chemical cleavages (such as depurination to cleave at A and G, or hydrazinolysis of pyrimidines to cleave at C and T) or to use partial digestion with such enzymes as micrococcal DNase II or pancreatic DNase I. Murray (6) studied cleavage patterns with pancreatic DNase in considerable detail, but "no conclusion could be drawn about base specificity of pancreatic deoxyribonuclease in either early or limit reactions"; thus, this enzyme would appear to have very limited value.

Cleavage at Gp residues

To obtain base-specific cleavages of DNA molecules, we have turned to the observation of Berg, Fancher, and Chamberlin (7) that DNA polymerase I will insert ribonucleotides into a DNA molecule if Mn⁺⁺ is substituted for Mg⁺⁺ in the reaction mixture. When rGTP is added in place of dGTP, the polynucleotide synthesized in the presence of Mn⁺⁺ has a ribose linkage at every G residue, and it can therefore be cleaved specifically at G residues by alkaline hydrolysis. We have used the two-dimensional electrophoretic separation method of Sanger, Brownlee, and Barrell (1) to analyze the products obtained when ribosubstituted DNA is cleaved at G residues. This analysis has enabled us to obtain much more information about the cleavage products than was possible in the earlier work of Berg and his collaborators (7, 8). As we had hoped, the fingerprint pattern obtained by cleavage of DNA at G residues by this technique (Fig. 1) is very similar to that obtained when RNA is cleaved at G residues with RNase T₁.

T and U residues have similar effects upon electrophoretic mobilities [as was already known from work with tRNA fragments containing a T residue, and from work of Murray (6) with the nonspecific fragments produced by DNases]. The major novelty of the fingerprints in Fig. 1 is that spots are "twinned", due to the presence of both 2' and 3' phosphate isomers of each fragment in the alkaline hydrolysates (T₁ RNase digestion of RNA yields only 3' phosphate groups). The most rapidly migrating member of the pairs appears to bear the 2' phosphate group. This was determined for the 2'- and 3'-GMP spots by their position relative to 2',3'-cyclic GMP, which was produced in some of our early experiments. It has been confirmed for 2' and 3' AMP (from Fig. 3) by thin
layer chromatography in two solvents with appropriate markers.

It should be noted that in Fig. 1a the radioactivity was introduced as $[^{32}P]dATP$, while in Fig. 1b $[^{32}P]dATP$ was used. Since the two preparations were otherwise identical, the dramatic differences in the intensity patterns of the two fingerprints reflect the relative amounts of DT or dA in each fragment. All of the spots in the lower graticule of Fig. 1a are labeled, even though they contain no T residues. Radioactivity is introduced into these fragments only by the transfer of label from T to G during the alkaline hydrolysis. For example, in the sequence ... prGpApCprGp*T ... (G indicates a ribo-G residue in a sequence of deoxynucleotides), there is only one radioactive phosphorous atom (indicated by *) that entered the polymer 5' to T. After alkaline hydrolysis, this $^{32}P$ will be found attached to G in the fragment ApCprGp*.

In Fig. 2 we show the results of an experiment similar to that in Fig. 1, except that the alkaline-hydrolyzed DNA fragments were treated with alkaline phosphatase before being electrophoresed. As expected, removal of the 2' and 3' phosphates eliminates the twinning of spots. Also, the fragments that are labeled only by transfer of $^{32}P$ from T to G (Fig. 1a) or A to G (Fig. 1b) are now unlabeled, since this phosphorous has been removed. For this reason Fig. 2a (radioactivity introduced as $[^{32}P]dATP$) and Fig. 2b (radioactivity introduced as $[^{32}P]dATP$) look very different, even though closer inspection reveals that the differences are the expected ones. Another effect of the phosphatase treatment is to substantially increase the mobility of the nucleotides in the second

**Fig. 1.** (a) Two-dimensional electrophoresis of fragments from alkaline digestion of rGp-substituted DNA synthesized from an M13 template and labeled with $[^{32}P]dATP$. The ribosubstituted DNA was synthesized in a 0.6-ml mixture containing 56 μg/ml of M13 DNA; 50 μg/ml of bovine serum albumin; 83 μg/ml of DNA polymerase I; 67 mM Tris-HCl (pH 7.4); 1 mM 2-mercaptoethanol; 0.67 mM MnCl$_2$; 330 μM rGTP; 25 μM $[^{32}P]dATP$; 6.8 Ci/mmole; 33 μM (each) dCTP and dTTP. The DNA polymerase I was purified to step 6 (refs. 9, 10). The M13 DNA template had been broken by repeated freezing and thawing and needed no added primer.

Incorporation of label was followed by determination of acid-precipitable radioactivity; the reaction was terminated when acid-precipitable counts reached a maximum (39%) after 90 min. Protein was extracted by the Sevag procedure (chloroform–octanol 9:1) and the aqueous phase was chromatographed on a Bio-gel P-2 (200-400 mesh) column to remove salt and unincorporated nucleotides. (Subsequently, we have used Bio-gel P-60 columns, which are more effective in removing mononucleotides.) The desalted sample from the Bio-gel column was dried and resuspended in 0.5 ml of 10% piperidine for alkaline hydrolysis (16 hr, 50°C). The sample to be electrophoresed was dried and redissolved in 3 μl of a standard tracking dye mixture (0.11% xylene cyanol FF, 0.11% acid fuchsin, and 0.22% orange G in water). The sample was then applied (1) to the origin of an oxoid strip, and 2 μl of a 3-fold concentrated tracking dye mixture was spotted on either side of the sample. The small amount of tracking dye in the sample makes it possible to detect the presence of any salt in the sample by an easily detectable "fanning" pattern of the dye immediately after the start of electrophoresis. Electrophoresis in the second dimension was on DEAE-cellulose in 7% formic acid, and the dried electropherogram was exposed to x-ray film (1). P and B denote the positions of the pink and blue dye markers in the second dimension.

(b) Similar to Fig. 1a, except that $[^{32}P]dATP$ was used in place of $[^{32}P]dATP$ to label the DNA. Maximum incorporation (40%) occurred at 90 min. The sample was digested with concentrated NH$_4$OH (rather than piperidine) for 16 hr.

**Fig. 2.** (a) Two-dimensional electrophoresis of fragments resulting from alkaline and alkaline phosphatase digests of $[^{32}P]dATP$-labeled rGp-substituted DNA synthesized from an M13 template. The ribosubstituted DNA was synthesized as described in Fig. 1a, except the volume of the reaction mixture was 0.90 ml and 33 μM $[^{32}P]dATP$ (1.85 Ci/mmole) was used.

Maximum incorporation (18.5%) occurred at 90 min. Digestion of DNA was with concentrated NH$_4$OH for 24 hr. After the sample was dried, alkaline phosphatase digestion was done with 4 μg of Worthington BAPC alkaline phosphatase in 10 μl of 0.02 M Tris-HCl (pH 8.9)–5 mM MgCl$_2$ for 30 min at 37°C in a small capillary tube.

(b) Two-dimensional electrophoresis of fragments resulting from alkaline and alkaline phosphatase digest of $[^{32}P]dATP$-labeled rGp-substituted DNA synthesized from an M13 template. The ribosubstituted DNA used for digestion was the same as that in Fig. 1b. Digestions were as described in Fig. 2b, but Worthington BAPF alkaline phosphatase was used.
dimension, as anticipated from the results of Sanger, Brownlee, and Barrell (1). This increased mobility is useful in permitting the separation of larger fragments. When the radioactive patterns expected when $^{32}$P is introduced only on T or only on A are taken into account, the fingerprint patterns obtained in Fig. 2 are quite similar to those obtained for RNA digested with a combination of RNase T1 and alkaline phosphatase [see Fig. 8 of Sanger and Brownlee (11)]. In Fig. 2a, this is most easily seen in the graticule of fragments containing two T residues, which form the characteristic diamond pattern expected.

**Cleavage at Ap residues**

In RNA sequencing, cleavages by RNase T1 at G have been the most important tool available, since other enzymes are less specific. RNase A, which cleaves at both U and C residues, gives an average fragment size of two nucleotides that is of limited value in providing the overlap data essential in determining a large sequence. Thus, one of the main advantages of DNA sequencing should be that ribosubstitution techniques can provide specific cleavages at nucleotides other than G. Fig. 3 shows the cleavage pattern obtained after base hydrolysis of rA-substituted DNA. The simple pattern obtained suggests that the cleavage is quite specific and should be a powerful tool for sequencing studies.

Since we have so far only labeled material substituted with ribo A with [α-$^{32}$P]dTTP, it has not been possible to determine the base composition of material eluted from each spot. Nevertheless, the rough overall placement of the fragments resulting from a cleavage at A may be deduced from indirect evidence, and is indicated by tentative assignments in Fig. 3b and d. In making these assignments, we assumed that the positions (relative to the dye markers) of fragments with base composition (G, Tn, Cm)A in Fig. 3 would be similar to the positions of the corresponding isomers (A, Tn, Cm)G in Figs. 1 and 2. It was further assumed that the effects of additional G residues on observed mobilities are similar to the effects of additional T residues, so that fragments (T, Cm)A will run in the vicinity of fragments (G, Cm)A, and so on. [This assumption is suggested by the known effects of additional G residues in RNA fragments from RNase A digestion (1), and by the recent work of Murray with fragments produced by DNases (6)].

These tentative assignments are consistent with data from spleen phosphodiesterase digests of 25 spots eluted from the fingerprint shown in Fig. 3b and of 29 spots from a fingerprint similar to that shown in Fig. 3a. Detailed interpretation of this data is reserved for a future paper, but it is worth mentioning that fragments with nucleotide composition (Tn, Cm)A appear to always have a slightly higher mobility in the first dimension and a slightly lower mobility in the second dimension than do the corresponding fragments (Gm, Cn)A. For instance, when spot 1 in Fig. 3d was digested with spleen phosphodiesterase, the radioactivity was all recovered in 3' TMP, as expected if this spot contains the fragment Tp*Tp*TpA. Spot 2 gave equal amounts of radioactive 3' TMP and 3' GMP, as expected if it contained equal quantities of the three isomers Gp*Tp*TpA, TpGp*TpA, and Tp*TpGpA. By an analogous argument, spot 3 appears to contain the isomers of (Gn, Tn)A. (Fragments with the sequence GpGpGpA would be unlabeled after the phosphatase treatments, since the $^{32}$P was introduced on dTTP in this experiment.) Addition of dC residues appears to slow the fragments slightly in both the first and second dimensions, as in the case of cleavage at G residues.

We have also obtained encouraging preliminary results with cleavage of rC-substituted DNA, but since we have not as yet characterized individual fragments, commentary on
the fingerprint pattern obtained is reserved for a subsequent communication.

Isolation of unique fragments

The largest fragments that are resolved in the fingerprint patterns presented here are about 7–10 residues in length (e.g., those in the region where \([T_2, A_4, C_2]G\) might be expected in Fig. 2a and 2b. It should be emphasized that even these spots on the electropherogram must each contain several different isomers when a molecule as large as M13 DNA is used as template. A nucleotide sequence the size of M13 DNA [M13 viral DNA is single-stranded, and has a molecular weight of about \(2.1 \times 10^6\)] would be expected to contain about 17 of the 210 possible isomers of \([T_2, A_4, C_2]G\). Of course, larger fragments should be considerably less frequent. By chance, one would expect to find only about 13 fragments greater than 20 nucleotides in length resulting from cleavage at G. About four of these should be greater than 25 nucleotides in length. We have used electrophoresis on 12.5\% and 20\% polyacrylamide gels to search for specific fragments in this size range after cleavage of rG- or rC-substituted DNA. Autoradiographs of the gels (Fig. 4) show that the specific fragments are produced. As expected, the banding pattern obtained after cleavage at G is quite different from that obtained after cleavage at C, but in both cases the largest fragments are in the range of about 30–40 nucleotides. We believe that the bands indicated by arrows represent pure, unique fragments because (a) they appear to be well separated from all smaller fragments, (b) the electrophoretic mobility of the largest fragments is that expected for fragments that one might obtain by chance from a DNA molecule of this size, and (c) the amount of radioactivity in each band is consistent with a unique fragment of this size. As judged from the amount of radioactivity they contain, the faster-running bands on the same gels appear to represent mixtures of several smaller fragments.

Fidelity of synthesis

If the ribosubstitution method is to be generally useful in sequence studies, the fidelity of the \textit{in vitro} synthesis must be high. It is known that \textit{in vitro} synthesis with DNA polymerase I can be extremely accurate because Goulian, Kornberg, and Sinsheimer (14) were able to obtain \textit{in vitro} synthesis with this enzyme of biologically active \(\phi X174\) DNA molecules. Whether base-pairs are made with a similar high degree of accuracy during synthesis under ribosubstitution conditions is not as well documented, but the available data suggest that the fidelity is good. Berg \textit{et al.} (7) examined this question in their original work on ribosubstitution by studying the base composition and nearest-neighbor frequencies of ribosubstituted products made from various templates of widely different base composition. Both the overall base composition and the nearest-neighbor frequencies of the product corresponded to that of the template in each case.

Our own data also suggest that the fidelity of the ribosubstitution system is good. From published data on T1 digests of RNA (1), we can predict what fingerprint pattern would be obtained after cleavage of DNA at rG residues (Fig. 1), and in fact we find that all of the expected spots are present with about the expected intensities. Equally important, the fingerprint is "clean" (i.e., very little material is present in unexplained spots; such unexplained spots would suggest incorrect cleavages).

A much more sensitive indicator of fidelity is provided by electrophoresis on polyacrylamide gels that allows us to examine the largest unique fragments from each cleavage. The larger fragments would have a proportionately greater chance of containing any errors occurring in the system; therefore, the fact that even the largest unique fragments can be obtained as distinct bands, in what appears to be good yield, suggests that the fidelity of the system is adequate for sequencing studies. It is worthwhile pointing out that a considerable number of random base-substitution errors can be tolerated in sequencing studies if necessary. If 30\% of the fragments from a particular sequence contain errors, then 30\% of the intensity will be removed from the corre-

**Fig. 4.** Autoradiographs of fragments resulting from alkaline digestion of ribosubstituted DNA separated on 12.5 and 20\% polyacrylamide gels. Gels were made as described by Jeppesen \textit{et al.} (13), with the concentrations of acrylamide and bisacrylamide proportionately increased for the 20\% gel. Track 1 shows the fragments resulting from alkaline digestion of rGp-substituted DNA on a 20\% polyacrylamide gel. For comparison, the same fragments are shown (Track 4) on a 12.5\% gel. Track 2 shows the riboguanosine-substituted DNA before digestion. Tracks 1 and 3 are the same as 4 and 5, except ribocytosine-substituted DNA has been used. Arrows denote purified fragments. Positions of the marker dyes are indicated by A, xylene cyanol FF; B, bromphenol blue; and Y, orange G.

responding spot on the fingerprint and will appear instead distributed among a large number of other "illegitimate" spots, easily identifiable as such because of their low intensity. The question of fidelity remains extremely important and we are currently developing other approaches to more carefully measure the accuracy of the \textit{in vitro} system.

Conclusions

The sequencing approach that we describe here has obvious uncertainties, chiefly because of our very limited experience with it. Still, it may be worthwhile to briefly discuss some of the possible advantages of this approach. There are obvious advantages in being able to cleave specifically at either G or at
A (also probably at C, and possibly T if the Kornberg enzyme will accept rTTP in place of dTTP—this point is currently being tested). In RNA sequencing, a great deal of time must be devoted to deciding how the T; RNase fragments were arranged in the original molecule, but in the DNA method the complete sequence should be more easily deduced because of the availability of overlap data provided by the additional sets of fragments cleaved at A or C.

Since the approach we have presented involves in vitro synthesis, it will be possible to label all four nucleotides concurrently or one at a time at the discretion of the investigator. The ability to perform a series of parallel experiments in which only one nucleotide is labeled at a time is extremely useful since, by analysis of the presence or absence of 32P in the 2'- or 3'-terminal phosphate of each fragment, one can ascertain the identity of the nucleotide at the 5' end of the neighboring fragment. (For a detailed demonstration of the advantages of this approach, see ref. 4.)

The main advantages of DNA sequencing, however, should come from: (a) the availability of a wide range of sequence-specific nucleases, such as the Hemophilus endonuclease R, and (b) the primer dependence of DNA polymerase I that should allow general application of synchronous replication techniques. Obviously, these techniques can be combined in several ways. For example, Danna and Nathans (15) have shown that endonuclease R degrades the genome of Simian Virus 40, a small oncogenic DNA virus, to eleven unique fragments that can be separated by polyacrylamide gel electrophoresis. Purified fragments obtained from such digestions of either polyoma or SV40 DNA could be further analyzed by partial digestions with exonuclease III, followed by resynthesis of 32P-labeled ribosubstituted DNA with DNA polymerase I. By independent control of the amount of exonuclease III digestion and the amount of resynthesis by DNA polymerase I (using conditions appropriate for synchronous replication from the primer produced by exonuclease III digestion), it should be possible to selectively label different parts of the fragments for sequencing.

It is clear that a great variety of techniques is available for DNA sequencing. Only further work will reveal which are the most useful in practice, but it does seem that there is some basis for believing that it may be possible to make worthwhile progress with molecules as large as polyoma or SV40 DNA. We believe that such detailed investigation of eukaryotic material is extremely important to expand our very limited knowledge of the control mechanisms and genetic "punctuation" signals in higher organisms.

We are especially indebted to Susie Kinoshita, who performed all of the electrophoretic separations, and to Dickson Hong and Avram Pauker for their excellent technical assistance. We thank Drs. Roberts Smith, John Abelson, Paul Berg, and James Dahlberg for helpful advice. This work was supported by grants NIH GM-16578 and NSF GB-27512 to W. S. and by a NSF postdoctoral traineeship and a USPHS traineeship (GM 1531) to K. F.