Protein sequencing established in 1950s: degradation based
Protein Sequencing

The Nobel Prize in Chemistry 1958

"for his work on the structure of proteins, especially that of insulin"

Frederick Sanger

United Kingdom

University of Cambridge
Cambridge, United Kingdom
Sequencing ~1960

• Protein: Yes: sequential end chemical degradation based

• DNA: NO

• OLIGO SYNTHESIS: 2 bases
Sequencing Technology
1960-1970

1965 (first RNA sequence determined)-
Generally degradation based
RNA easier than DNA
Longest DNA 12nt (lambda cohesive ends, 1971)
Limited ability to get enough overlap to deduce long sequences
Time consuming steps
Complex interpretation of gel data
Limited priming capabilities limit what can be sequenced
Limited purified small DNA molecules...why were promoter elements so popular?
Sequencing Approaches
1960-1970

• Convert DNA to RNA and use degradation techniques (RNAses which cleave frequently) to compare to known fingerprints of small fragments (1-5 nt, major)
• Replace dNTP with NTP to make susc. To specific cleavage (NTP backbone more frail)
• Partial exonuclease digestion (limited use)
• EM, look directly at the bases in sequence as DNA and ribosomes were visualized: never generated any sequence
Info available, circa 1970

• Method of DNA polymerase catalyze polymerization known (Kornberg, 1959 nobel prize), primed DNA template dependent...

• Polynucleotide kinase known

• Synthesis of deoxyribonucleotides, dideoxyribonucleotides, and analogues known
Info available, circa 1970

- Restriction enzymes available (1972)
- Oligosynthesis available (1969)
- Reverse transcriptase (1972 used widely) available (hard to label mRNA directly)
The Nucleotide Sequence of the lac Operator

(regulation/protein-nucleic acid interaction/DNA-RNA sequencing/oligonucleotide priming)

WALTER GILBERT AND ALLAN MAXAM

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138

Communicated by J. D. Watson, August 9, 1973

ABSTRACT The lac repressor protects the lac operator against digestion with deoxyribonuclease. The protected fragment is double-stranded and about 27 base-pairs long. We determined the sequence of RNA transcription copies of this fragment and present a sequence for 24 base pairs. It is:

5'--TGGAATTGTGAGCCGATAAACAATT3'
3'--ACCCTTAAACCTGGCTATTTGTTAAA5'

The sequence has 2-fold symmetry regions; the two longest are separated by one turn of the DNA double helix.

The lactose repressor selects one out of six million nucleotide bind again to the repressor, and is about 27 base-pairs long. Here we shall describe its sequence.

METHODS

Sonicated DNA Fragments. Sonicated [\textsuperscript{32}P]DNA fragments were made by growing a temperature-inducible lysogen of \textit{\textlambda}c1857\textit{plac537} at 34° in a glucose–50 mM Tris·HCl or TES (pH 7.4) medium in 3 mM phosphate, heating at 42° for 15 min at a cell density of \(4 \times 10^8\)/ml, then washing and resuspending the cells at a density of \(8 \times 10^8\)/ml in the same medium. The labeled 1000–2000-MDa DNA...
Sequencing of the lac Operator…
PNAS 1973

- Lac repressor available to bind operator
- Sonicate DNA labeled in vitro with $\text{H}^{32}\text{PO}_4$
- Bind lac repressor and purify bound products
- Determine pyrimidine tracts (C,T)
- Pyrimidine tracts flanked by either G or A, can roughly quantitate relative abundance by two D gels
Fig. 3. (a) Two-dimensional electrophoresis of fragments resulting from alkaline digestion of rAp-substituted DNA synthesized from an M13 template and labeled with [α-32P]dTTP. The ribosubstituted DNA was synthesized as in Fig. 1a, except rATP is substituted for rGTP, with the appropriate changes made in the deoxyribonucleotide triphosphates. Maximum incorporation (16%) occurred after 30 min of incubation. The sample was prepared as described in Fig. 1a, except piperidine digestion was for 8 hr.

(b) Line drawing of the autoradiograph shown in Fig. 3a. Some very preliminary sequence assignments are indicated.

(c) Fragments resulting from alkaline and alkaline phosphatase digestion of [α-32P]dTTP labeled, rAp-substituted DNA synthesized from M13 template. The ribosubstituted DNA is the same product described in Fig. 3a. After alkaline digestion with 10% piperidine for 16 hr, the sample was treated with alkaline phosphatase as in Fig. 2b.

(d) Line drawing of the autoradiograph shown in Fig. 3c. Some very preliminary sequence assignments are indicated.
<table>
<thead>
<tr>
<th>Tract</th>
<th>Moles</th>
<th>Yields</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCp</td>
<td>4–5</td>
<td>(4.6)</td>
</tr>
<tr>
<td>pTp</td>
<td>7</td>
<td>(7.0)</td>
</tr>
<tr>
<td>pTptp</td>
<td>4</td>
<td>(3.8)</td>
</tr>
<tr>
<td>pTpCpCp</td>
<td>1</td>
<td>(2.0)</td>
</tr>
<tr>
<td>pCpTpCp</td>
<td>1</td>
<td>(0.7)</td>
</tr>
<tr>
<td>pTptpCpCp</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Pyrimidine tracts were isolated and fingerprinted. The sequences were determined by partial digestion of phosphatase-treated material by spleen and by venom phosphodiesterase. The relative molar yields are the averages of three experiments, taking the TCC and CTC isostichs together as 2 mol/mol of operator.
Sequencing of the lac Operator…

PNAS 1973

• RNA synthesis… Short DNA fragment converted to RNA with RNA pol holoenzyme (primed or unprimed)

• Primers: short synthetic di-nucleotides, GGAAU from RNAse A digestion of tRNA (gly) from Staph Epi
Sequencing of the lac Operator…
PNAS 1973

• Labeled RNA separated on polyacrylamide
• Digested with RNAse A (U and C)
• Or RNAse T1 (only G cleaved)
• Run 2D gels to determine which short sequences present after digestion
• Specific primed sequence products(low NTP) also generated and analyzed
Fig. 1. (a) Two-dimensional electrophoresis of fragments from alkaline digestion of rGp-substituted DNA synthesized from an M13 template and labeled with $[\alpha^{32}P]dTTP$. The ribosubstituted DNA was synthesized in a 0.6-ml mixture containing 50 $\mu$g/ml of M13 DNA; 50 $\mu$g/ml of bovine serum albumin; 83 $\mu$g/ml of DNA polymerase I; 67 mM Tris-HCl (pH 7.4); 1 mM 2-mercaptoethanol; 0.67 mM MnCl$_2$; 330 $\mu$M rGTP; 25 $\mu$M $[\alpha^{32}P]dTTP$, 6.8 Ci/mmol; 33 $\mu$M (each) dATP and dCTP. 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 $\mu$l 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 $\mu$l of a 3-fold concentrated tracking dye mixture was spotted on either side of the origin. The strips were then run in a two-dimensional apparatus, first electrophoresing in the pH 3.5 direction and then in the pH 3.6 direction.
Table 2. Sequence analysis of some larger fragments

<table>
<thead>
<tr>
<th>Fragment</th>
<th>Source</th>
<th>Digestion products</th>
<th>RNase A</th>
<th>RNase U2</th>
<th>RNase T1</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUAAACAUUOHN</td>
<td>RNase T1</td>
<td>AU(A) AAC(A) AAU(U)</td>
<td>A CAA UAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CUCACAUUUCACOH</td>
<td>RNase T1</td>
<td>2C(A) 2U(C) C(U)</td>
<td>CAA (C,C,U)A</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AC(A) AAU(U) C(C)</td>
<td>(C,C,U,U)OH</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>U(U) U(A) C(C)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C(G) G(C) AU(C)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AAU(U) U(G) G(U)</td>
<td>AA UUG</td>
<td></td>
<td>G(G) G(A) AAU(U)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>UG AG CG(G)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>UG AAGUG</td>
</tr>
<tr>
<td>AAUUG(U)</td>
<td>RNase T1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGAUO(U)</td>
<td>RNase A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UGACCG(G)</td>
<td>partial RNase T1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAUUGUG</td>
<td>partial RNase T1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGAUAAC</td>
<td>CMCT-blocked RNase A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GUGAGC</td>
<td>CMCT-blocked RNase A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The long fragments with 3’-hydroxyl termini were isolated from RNase T1 digests of gel bands from primed syntheses. The other oligonucleotides were isolated from unprimed total syntheses. The fingerprinting and sequencing techniques are as described by Barrell (11) and in the Methods, except that the CMCT blocking was 25 mg/ml of CMCT in 10 mM borate (pH 8.5)–3.5 M urea overnight at 32°. Nucleotides determined through nearest-neighbor analysis are shown in parentheses.
Sequencing of the lac Operator…
PNAS 1973

- Multiple very short sequences were assembled knowing length and needing to have all pieces of puzzle fit… pyrimidine tracts and purine positions, products from whole RNA, products from primed RNA
- 24 nt of sequence determined
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 ... pGpTpPypPupApCp ... (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

mination 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 deamination 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 Chamb"
Birth of Sequencing Technology for any region

1973-1977

Introduction of maxam/gilbert and sanger methods
A new method for sequencing DNA

(DNA chemistry/dimethyl sulfate cleavage/hydrazine/piperidine)

ALLAN M. MAXAM AND WALTER GILBERT

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138

Contributed by Walter Gilbert, December 9, 1976

ABSTRACT DNA can be sequenced by a chemical procedure that breaks a terminally labeled DNA molecule partially at each repetition of a base. The lengths of the labeled fragments then identify the positions of that base. We describe reactions that cleave DNA preferentially at guanines, at adenines, at cytosines and thymines equally, and at cytosines alone. When the products of these four reactions are resolved by size, by electrophoresis on a polyacrylamide gel, the DNA sequence can be read from the pattern of radioactive bands. The technique will permit sequencing of at least 100 bases from the point of labeling.

We have developed a new technique for sequencing DNA molecules. The procedure determines the nucleotide sequence of a terminally labeled DNA molecule by breaking it at adenine, guanine, cytosine, or thymine with chemical agents. Partial cleavage at each base produces a nested set of radioactive fragments extending from the labeled end to each of the positions of that base. Polyacrylamide gel electrophoresis reveals...

THE SPECIFIC CHEMISTRY

A Guanine/Adenine Cleavage (2). Dimethyl sulfate methylates the guanines in DNA at the N7 position and the adenines at the N3 (3). The glycosidic bond of a methylated purine is unstable (3, 4) and breaks easily on heating at neutral pH, leaving the sugar free. Treatment with 0.1 M alkali at 90° then will cleave the sugar from the neighboring phosphate groups. When the resulting end-labeled fragments are resolved on a polyacrylamide gel, the autoradiograph contains a pattern of dark and light bands. The dark bands arise from breakage at guanines, which methylate 5-fold faster than adenines (3).

This strong guanine/weak adenine pattern contains almost half the information necessary for sequencing; however, ambiguities can arise in the interpretation of this pattern because the intensity of isolated bands is not easy to assess. To determine the bases we compare the information contained in this column...
Fig. 1. Strand separation of a restriction fragment: 1.5 μg of a 64-base-pair DNA fragment (75 pmol of 5' ends) was phosphorylated with [γ-32P]ATP (800 Ci/mmole) and polynucleotide kinase, denatured in alkali, layered onto a 0.3 cm × 3 cm surface of an 8% polyacrylamide slab gel (see under Techniques), and electrophoresed at 200 V (regulated) and 20 mA (average), until the xylene cyanol (XC) dye moved 9 cm. The gel on one glass plate was then tightly covered with Saran Wrap and exposed to Kodak XR-5 x-ray film for 10 min.
FIG. 2. Autoradiograph of a sequencing gel of the complementary strands of a 64-base-pair DNA fragment. Two panels, each with four regions, show the sequence of the DNA.
FIG. 2. Autoradiograph of a sequencing gel of the complementary strands of a 64-base-pair DNA fragment. Two panels, each with a different set of nucleotides, are shown.
DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage φX174)

F. SANGER, S. NICKLEN, AND A. R. COULSON

Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England

Contributed by F. Sanger, October 3, 1977

ABSTRACT A new method for determining nucleotide sequences in DNA is described. It is similar to the “plus and minus” method [Sanger, F. & Coulson, A. R. (1975) J. Mol. Biol. 94, 441–448] but makes use of the 2′,3′-dideoxy and arabinonucleoside analogues of the normal deoxynucleoside triphosphates, which act as specific chain-terminating inhibitors of DNA polymerase. The technique has been applied to the DNA of bacteriophage φX174 and is more rapid and more accurate than either the plus or the minus method.

The “plus and minus” method (1) is a relatively rapid and simple technique that has made possible the determination of the sequence of the genome of bacteriophage φX174 (2). It depends on the use of DNA polymerase to transcribe specific regions of the DNA under controlled conditions. Although the method is considerably more rapid and simple than other available techniques, neither the “plus” nor the “minus” method is completely accurate, and in order to establish a sequence both must be used together, and sometimes confirmatory data are necessary. W. M. Barnes (J. Mol. Biol., in press) has recently developed a third method, involving ribo-substitution, which has certain advantages over the plus and minus method, but this has not yet been extensively exploited.

Another rapid and simple method that depends on specific chemical degradation of the DNA has recently been described by Maxam and Gilbert (3), and this has also been used extensively for DNA sequencing. It has the advantage over the plus and minus method that it can be applied to double-stranded DNA, but it requires a strand separation or equivalent fractionation of each restriction enzyme fragment studied, which makes it somewhat more laborious.

This paper describes a further method using DNA polymerase, which makes use of inhibitors that terminate the newly synthesized chains at specific residues.

Principle of the Method. Atkinson et al. (4) showed that the inhibitory activity of 2′,3′-dideoxythymidine triphosphate (ddTTP) on DNA polymerase I depends on its being incorporated into the growing oligonucleotide chain in the place of the 3′-hydroxyl. However, the ddT contains no 3′-hydroxyl group, and this molecule is not incorporated into DNA when used as a substrate for DNA synthesis. The inhibition arises from the presence of a stereosomer of ribose in which the 3′- hydroxyl group is oriented in trans position with respect to the 2′-hydroxyl group. The arabinosyl (ara) nucleotides act as chain terminating inhibitors of Escherichia coli DNA polymerase I in a manner comparable to ddT (4), although synthesized chains ending in 3′ araC can be further extended by some mammalian DNA polymerases (5). In order to obtain a suitable pattern of bands from which an extensive sequence can be read it is necessary to have a ratio of terminating triphosphate to normal triphosphate such that only partial incorporation of the terminator occurs. For the dideoxy derivatives this ratio is about 100, and for the arabinosyl derivatives about 5000.

METHODS

Preparation of the Triphosphate Analogues. The preparation of ddTTP has been described (6, 7), and the material is now commercially available. ddA has been prepared by McCarthy et al. (8). We essentially followed their procedure and used the methods of Tener (9) and of Hoard and Ott (10) to convert it to the triphosphate, which was then purified on DEAE-Sephadex, using a 0.1–1.0 M gradient of triethylamine carbonate at pH 8.4. The preparation of ddGTP and ddCTP has not been described previously; however, we applied the same method as that used for ddATP and obtained solutions having the requisite terminating activities. The yields were very low, and this can hardly be regarded as adequate chemical characterization. However, there can be little doubt that the activity was due to the dideoxy derivatives.

The starting material for the ddGTP was N-isobutyl-5′-O-monomethoxytrityldeoxyguanosine prepared by F. E. Baralle (11). After tosylation of the 3′-OH group (12) the compound was converted to the 2′,3′-dideoxy derivative with sodium methoxide (8). The isobutyl group was partly removed during this treatment and removal was completed by incubation in NH₃ (specific gravity 0.88) overnight at 45°C. The dideoxy derivative was reduced to the dideoxy derivative (8) Be prepared by the method of Atkinson et al. (4)
FIG. 1. Autoradiograph of the acrylamide gel from the sequence determination using restriction fragments A12d and A14 as primers on the complementary strand of φX174 DNA. The inhibitors used were (left to right) ddGTP, ddATP, ddTTP, and araCTP. Electrophoresis was on a 12% acrylamide gel at 40 mA for 14 hr. The top 10 cm of the gel is not shown. The DNA sequence is written from left to right and upwards beside the corresponding bands on the radioautograph. The numbering is as given in ref. 2.
The Nobel Prize in Chemistry 1980

"for his fundamental studies of the biochemistry of nucleic acids, with particular regard to recombinant-DNA"

"for their contributions concerning the determination of base sequences in nucleic acids"

Paul Berg 1/2 of the prize
Walter Gilbert 1/4 of the prize
Frederick Sanger 1/4 of the prize
Invention of PCR

- Science 1985 Dec 20;230(4732):1350-4

Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N.

Two new methods were used to establish a rapid and highly sensitive prenatal diagnostic test for sickle cell anemia. The first involves the primer-mediated enzymatic amplification of specific beta-globin target sequences in genomic DNA, resulting in the exponential increase (220,000 times) of target DNA copies. In the second technique, the presence of the beta A and beta S alleles is determined by restriction endonuclease digestion of an end-labeled oligonucleotide probe hybridized in solution to the amplified beta-globin sequences. The beta-globin genotype can be determined in less than 1 day on samples containing significantly less than 1 microgram of genomic DNA.
Thermostable pol

• Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Related Articles, Links
Oligonucleotide Synthesis

• 1969 chemical synthesis of deoxynucleotide polymers...

1972-1976 Improvements in technique allow 15-20 mers

By 1979, oligo houses formed to commercially produce oligos.