Proc. Nat. Acad. Sci. USA
Vol. 70, No. 12, Part I, pp. 3581–3584, December 1973

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′--T G G A A T T G T G A G C G G A T A A C A A T T 3′
3′--A C C T T A A C A C T C G C C T A T T G T T A A 5′

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 sequences in the *Escherichia coli* genome and binds to it to prevent the expression of the genes for lactose metabolism. How does this protein, a 150,000-dalton tetramer of identical subunits, recognize its target? To answer this question we have determined the sequence of the repressor-binding site: the operator.

Genetically the operator is the locus of operator constitutive (o^c) mutations, *cis*-dominant changes that render the adjacent structural genes less sensitive to repressor control (1). The isolation of repressors (2, 3) and the demonstration that o^c mutations interfere with the binding of these proteins to DNA (4, 5) showed that the operator is the DNA binding site for the repressor protein.

We have isolated the lactose operator as a double-stranded fragment of DNA that has been protected by the lactose repressor against digestion by pancreatic DNase (6). The first step in this purification isolated DNA fragments, about 1000 base-pairs long, that carry the *lac* promoter–operator region. The *lac* repressor, when mixed with sonicated fragments of DNA from a phage that carries the *lac* genes, binds to a cellulose nitrate filter only those fragments containing the operator. IPTG (isopropyl-β,D-thiogalactoside, a synthetic inducer of the *lac* operon) specifically elutes from the filter those fragments that are found to the repressor (7). The *lac* repressor binds to these DNA fragments, as it does to whole phage DNA, with a dissociation constant of 10^{-13} M in 0.01 M salt and 0.01 M Mg^{++} (8). [The repressor binds IPTG with a constant of 1.3×10^{-6} M (2).] Since the repressor dissociates from the operator with a half-life (under our conditions) of 15 min, we treat a mixture of repressor and sonic fragments with DNase for a very short time, 1 min, and isolate any protected region before the repressor can release the DNA. The protected fragment we obtain is double-stranded, can bind again to the repressor, and is about 27 base-pairs long. Here we shall describe its sequence.

Abbreviations: TES, *N*-tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid; CMCT, 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide metho-*p*-toluenesulfonate; IPTG, isopropyl-β, D-thiogalactoside.

METHODS

Sonicated DNA Fragments. Sonicated [^{32}P]DNA fragments were made by growing a temperature-inducible lysogen of λ*c*1857p*lac*5*S*7 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 4×10^8/ml, then washing and resuspending the cells at a density of 8×10^8/ml in the same medium with 0.1 mM phosphate. 100 mCi of neutralized $H_3{}^{32}PO_4$ was added to 10 ml of cells, and the incorporation was continued for 2 hr at 34°. The cells were washed, suspended in 2 ml of TE buffer [10 mM Tris·HCl (pH 7.5)–1 mM EDTA], sonicated with six 15-sec bursts, and extracted with phenol. The aqueous phase was extracted with ether, and the residual ether was removed with a stream of N_2. The mixture of radioactive DNA and RNA was then made up to about 6 ml with binding buffer [containing 10 mM $MgCl_2$, 10 mM KCl, 10 mM Tris·HCl (pH 7.5), 0.1 mM EDTA, 0.1 mM dithiothreitol, 5% dimethylsulfoxide, and 50 μg/ml of bovine-serum albumin]. 20 μg of *lac* repressor was added and 1-ml aliquots of this mixture were filtered on six separate 25-mm Schleicher and Schuell B-6 filters. The filters were presoaked in binding buffer without bovine-serum albumin; the filtration was fast. Each filter was rinsed twice with 0.3 ml of binding buffer, then transferred to a second filtration apparatus and eluted twice with 0.25 ml of 0.1 mM IPTG in binding buffer. The eluate was extracted with phenol and dialyzed against TE buffer. This one-step process yields sonic fragments containing the *lac* operator, which repressor will bind again to filters with about 50–70% efficiency. The overall yield is about 0.2% of the input label.

Unlabeled sonicated fragments were made by purifying 100–300-mg amounts of phage. A 150-liter fermentor culture of a λ*c*I857S7p*lac*5 lysogen was heat-induced at $A_{550} = 0.6$ in M63 minimal medium supplemented with glucose and casamino acids. After 3 hr the bacteria were harvested and frozen. 300–700 g of bacteria were thawed, lysed by suspension in phage buffer [0.1 M NaCl–10 mM Tris·HCl (pH 7.4)–10 mM $MgSO_4$] containing 0.1 μg/ml of DNase, and centrifuged at 17,000 × *g* for 30 min. The phages were precipitated with 2.1% NaCl and 6.5% polyethylene glycol. The precipitated phages were dissolved in phage buffer, run on a block CsCl gradient, and then on an equilibrium CsCl gradient. After brief dialysis to remove CsCl, purified phages were made up to 20 mM EDTA and sonicated with eight 15-sec bursts at full power, and extracted with phenol. The DNA was dialyzed against TE buffer.

The DNA solution was made up to binding buffer, and 5-ml

TABLE 1. *Pyrimidine tracts from the lac operator*

Tract	Moles	Yields
pCp	4–5	(4.6)
pTp	7	(7.0)
pTpTp	4	(3.8)
pTpCpCp	1	(2.0)
pCpTpCp	1	
pTpTpCpCp	1	(0.7)

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.

aliquots of DNA with repressor were filtered on a 47-mm B-6 filter. The filter capacity is 25 mg of DNA with 0.25 mg of active repressor (a 2-fold excess). The filter was washed with 3 ml of binding buffer and then eluted with 1 mM IPTG.

Operator Fragments. ^{32}P-Labeled operator was made by mixing sonic ^{32}P-labeled fragments with 2 μg of pure unlabeled *lac* fragments and 1 μg of repressor in 0.5 ml of binding buffer, digesting with 0.05–0.1 mg of DNase for 1 min, filtering rapidly on a B-6 filter, washing with 0.5 ml of 0° binding buffer without Mg^{++} and containing 20 mM EDTA, and shaking the filter on a Vortex mixer in 0.2 ml of 50 mM NH_4HCO_3 with 0.1% Na dodecyl sulfate.

Unlabeled operator was prepared by mixing 300 μg of sonic *lac* fragments, 10^6 cpm of ^{32}P-labeled sonic *lac* fragments (as tracer), and 150 μg of active repressor (300–600 μg of repressor) in 3 ml of binding buffer, and treating with 1 mg of DNase for 1 min at room temperature. The mixture was filtered on a 47-mm Schleicher and Schuell B-6 filter, washed with 3 ml of 20 mM EDTA, and shaken on a Vortex mixer in 1 ml of 50 mM NH_4HCO_3 with 1% Na dodecyl sulfate.

In both cases, the Na dodecyl sulfate rinses were run on a Sephadex G-100 column in 50 mM NH_4HCO_3 and 0.1% Na dodecyl sulfate. The operator eluted as a peak at 1.4 times the void volume. The peak fractions were pooled, the Na dodecyl sulfate was extracted with phenol, the phenol was extracted with ether, and the material was concentrated by lyophilization, which removed the NH_4HCO_3. There was about a 0.2–0.5% yield of material from the sonic *lac* fragment stage. We finally recover 100,000–200,000 Cerenkov cpm from 100-mCi incorporations.

Pyrimidine Tracts. DNA was depurinated according to Burton (9). Operator [^{32}P]DNA (10,000–100,000 cpm) and 50 μg of calf-thymus DNA were incubated with 66% formic acid and 2% diphenylamine in 200 μl for 18 hr at 37° in the dark. 200 μl of H_2O was added, and the mixture was extracted four times with 2 ml of ice-cold ether to remove diphenylamine, lyophilized repeatedly to remove formic acid, and dried to a compact residue on a polyethylene sheet.

RNA Synthesis. Reaction mixtures contained, in a volume of 20–50 μl: 10 mM $MgCl_2$, 20 mM Tris·HCl (pH 7.8), 70 mM KCl, 0.1 mM K_2HPO_4, 0.1 mM dithiothreitol, 250 μg/ml of bovine-serum albumin, generally 3–5 pmol of denatured operator strands, and 10–30 pmol of RNA polymerase (holoenzyme or core). For labeling with a single NTP, the mixtures also contained one [α-^{32}P]NTP (5–50 μM) at specific activities of 20–150 mCi/μmol (New England Nuclear Corp.), and three other cold NTPs at 300 μM. Reactions using these NTP concentrations were quite efficient, incorporating 20–70% of the label, about a 30-fold net synthesis. For primed synthesis, the mixtures contained either all four labeled NTPs at 5 μM or only GTP at 5 μM, the others at 20 μM, and dinucleotide primers at 100–500 μM or oligonucleotide primers at 50–100 μM. Reactions using these lower NTP concentrations with primers were less efficient, incorporating about 10% of the label, but synthesized better-defined species of RNA.

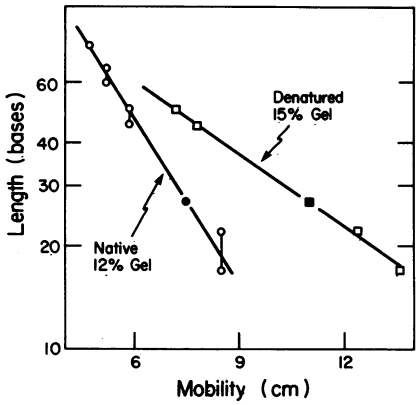

FIG. 1. Gel electrophoresis of operator fragments. ^{32}P-Labeled operator fragments were subjected to electrophoresis: native fragments (●) on a 12% polyacrylamide gel in 90 mM Tris·borate (pH 8.3)–5 mM $MgCl_2$ with double-stranded synthetic DNA markers of length (○) 77/77, 66/60, 50/45, and 22/17, or after denaturation (■) on a 15% polyacrylamide gel in 90 M Tris-borate (pH 8.3)–2.5 mM EDTA–7 M urea with DNA strands of length (□) 50, 45, 22, and 17. This experiment was done by Dr. T. Maniatis (14); the synthetic DNAs were the generous gift of Drs. H. Van de Sande and H. G. Khorana.

The reaction mixtures were incubated at 35° for 8–15 hr. 5 μl of 0.1 M EDTA and 20 μg of tRNA (Schwarz) were added and the reaction mixture was extracted with phenol. The RNA was isolated either by passage through a Sephadex G-25 or G-50 column in 10 mM NH_4HCO_3, followed by lyophilization, or by addition of 0.1 volume of 20% NaOAc and precipitation with 2 volumes of ethanol in a silicated tube, chilling in dry ice–acetone. The precipitate was washed with ice-cold 95% ethanol and dried.

Primers. Dinulceoside monophosphates were obtained from Sigma. The pentanucleotide ApUpCpCpG was synthesized enzymatically with primer-dependent polynucleotide phosphorylase by the method of Thach (10). GpGpApApU was recovered from a two-dimensional RNase A fingerprint of $tRNA_{I_A}^{Gly}$ isolated from *Staphylococcus epidermis* by Dr. Richard Roberts.

Polyacrylamide Gels. RNA molecules were resolved on 12% or 18% (w/v) polyacrylamide gels made in TBE buffer [90 mM Tris-borate (pH 8.3)–2.5 mM EDTA] and 7 M urea. Samples were heated at 100° for 3 min in $^1/_{10}$ TBE buffer with 7 M urea. Slabs, 20 × 20 × 0.2 cm, were run at 30 mA (300 V) for 3–4 hr at room temperature. After autoradiography, the films were aligned with the gels; sections corresponding to bands were cut out, crushed, suspended in 1 ml of

TABLE 2. *Sequence analysis of some larger fragments*

Fragment	Source	Digestion products		
		RNase A	RNase U2	RNase T1
AUAACAAUU$_{OH}$	RNase T1	AU(A) AAC(A) AAU(U)	A CAA UAA	
CUCACAAUUCCA$_{OH}$	RNase T1	2C(A) 2U(C) C(U) AC(A) AAU(U) C(C)	CAA (C,C,U)A (C,C,U,U)N$_{OH}$	
UUAUCCG(C)	RNase T1	U(U) U(A) C(C) C(G) G(C) AU(C)		
AAUUG(U)	RNase T1	AAU(U) U(G) G(U)	AA UUG	
GGAAU(U)	RNase A			G(G) G(A) AAU(U)
UGAGCG(G)	partial RNase T1	U(G) GAGC(G) G(G)		UG AG CG(G)
AAUUGUG	partial RNase T1	U G GU AAU		UG AAUUG
GGAUAAC	CMCT-blocked RNase A	GGAU AAC		
GUGAGC	CMCT-blocked RNase A	GU GAGC		

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*.

0.5 M NH_4OAc, 10 mM $Mg(OAc)_2$ 0.1% Na dodecyl sulfate, and 20 μg/ml of tRNA, and left overnight at 25°. The polyacrylamide fragments were filtered off by passing the solution, by gentle centrifugation, twice through a plug of glass-wool packed into the bottom of a small, punctured, plastic tube fitted into the top of a silicated centrifuge tube. The RNA was then precipitated with 2 volumes of ethanol.

Fingerprinting. The fingerprinting and sequencing techniques are those described in Barrall (11), except that the first dimension on cellulose acetate was in 5% acetic acid–0.1% pyridine–7 M urea–1–5 mM EDTA; the EDTA eliminated streaking, especially of pyrimidine tracts. Transfer to DEAE–paper for the second dimension was facilitated by placing a pad of dry Whatman 3MM paper strips beneath the usual layers of DEAE–paper, cellulose acetate, and wet strips, and washing through extensively with H_2O. All the radioactive label is found on the DEAE–paper, and all the urea is absorbed by the bottom pad and need not be washed away afterwards. For the first dimension, the yellow dye is run to 38 cm for a pancreatic fingerprint or to 42 cm for a T1 fingerprint. The second dimension is run in 7% formic acid until the blue dye has migrated 21 cm (pancreatic) or 27 cm (T1).

RESULTS

The operator fragment, protected by the repressor from digestion with pancreatic DNase, is a double-stranded DNA molecule that has a T_m of 67° in 0.15 M NaCl–0.015 M Na-citrate. (This is easily measured because *lac* repressor does not bind the denatured operator to a cellulose nitrate filter, and the denatured fragments themselves do not stick to the filter.) It is 60% A + T. Fig. 1 shows that the operator fragments move in gel electrophoresis as though they are 27 base-pairs long, in both the native and denatured states.

We began to determine the sequence of the operator by isolating pyrimidine tracts from ^{32}P-labeled material. Table 1 lists these pyrimidine tracts and their relative amounts. Unfortunately the longest pyrimidine run is pTpTpCpCp—not long enough to establish much sequence. The number of tracts is roughly consistent with our estimated size of th operator.

The final yields of ^{32}P-labeled operator were too limited for us to make progress with a direct attack on the sequence of the DNA; instead, we transcribed the operator fragment with RNA polymerase and determined the sequence of the RNA product. To do this, we scaled up the purification, beginning with 300 mg of purified phage to obtain, ultimately, micrograms of operator. A typical RNA reaction mixture contained 3 pmol of denatured operator DNA and 20 pmol of RNA polymerase, either holoenzyme or core. With the nearest-neighbor information from α-^{32}P transfer experiments, we were able to work out the structure of the oligonucleotides from complete digests, as well as to obtain some overlapping sequences from partial RNase T1 and 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide metho-*p*-toluenesulfonate (CMCT)-blocked RNase A digests. Table 2 lists some of the longer fragments. However, because the synthesis produces a mixture of products initiating and terminating at many different points on both strands, we were not able to establish a sequence.

Partial Synthesis. We resolved the RNA product into single molecular species by using oligonucleotides to direct the synthesis of specific regions. Downey and So (12) have shown that when the NTP concentration is lowered to 5 μM, the RNA polymerase can no longer initiate normally with a triphosphate (because the K_m for initiation is much greater than the K_m for elongation). Under these conditions, on denatured DNA, the polymerase will initiate if it is presented with a dinucleotide or oligonucleotide complementary to the template which it can incorporate into the beginning of an RNA chain. (The oligonucleotide does not have to form a stable association with its complementary DNA sequence.) We exploited this fact to force the polymerase to synthesize from a single strand of our denatured operator DNA and to begin synthesis at a specific point. Although the oligonucleotide priming selects one strand rather than the other, the product still has ragged 3′ ends. These products resolve into a

TABLE 3. *Products of oligonucleotide-primed synthesis*

One strand	
*GGAAU*UGUGAGCGGAUAACAAUU$_{OH}$	
The other strand	
*UA*UCCGCUCACAAUUCCA$_{OH}$	
*AUCCG*CUCACAAUUCCA$_{OH}$	
*GUUA*UCCGCUCACAAUUCCA$_{OH}$	

Single bands on polyacrylamide gels were isolated from primed syntheses. The primers are *italicized*. The GpU-primed molecule was isolated from a two-dimensional gel as described by de Wachter and Fiers (20), except that the first dimension was a 4% gel in 10 mM citric acid and the second dimension a 12% TBE–7 M urea gel.

series of bands during electrophoresis on 12% or 18% polyacrylamide gels in 7 M urea. The oligonucleotide GpGpApApU isolated from *Staphylococcus epidermis* $tRNA_{I_A}^{Gly}$ (a gift and a suggestion from Dr. Richard Roberts), primed the synthesis of the RNA molecule from one strand. The dinucleotide UpA and the pentanucleotide ApUpCpCpG primed synthesis from the other strand. GpU primed on both strands. Table 3 shows these results. The two sequences overlap and are complementary. Matching them up, written as DNA sequences, and filling out the gaps by complementary base pairing, produces a 24-base long sequence:

5′ TGGAATTGTGAGCGGATAACAATT 3′
3′ ACCTTAACACTCGCCTATTGTTAA 5′

The squared ends of the sequence are hypothetical; we do not know exactly where the DNase cuts. There are a few more bases on the DNA fragment to the left of this sequence. The pyrimidine tracts can be accommodated in the sequence (with the exception that the yields of pCp and pTp are too high); although they fail to provide any overlaps, they serve to confirm the correctness of some of the oligonucleotide sequences. We believe the orientation of the fragment to be correct as written, with the *i* gene to the left and the *z* gene to the right, in order to be consistent with the *lac* messenger sequence described by Maizels (13).

DISCUSSION

The sequence for the operator fragment shows symmetries. A total of 16 bases lie in a 2-fold symmetrical pattern, which is not centered on the fragment. Six base pairs to the left are symmetrically oriented with respect to six base pairs to the extreme right of the fragment. There are additional symmetrically arranged base pairs in a nine-base stretch between the two major regions. These symmetry regions are boxed below:

TGG[AATTGT]G[A]G[C]G[G]A[T]A[AGAATT]
ACC[TTAACA]C[T]C[G]C[C]T[A]T[TGTTAA]

(However, two to three of these matched base pairs might be symmetrical by chance.) This symmetry could permit the repressor protein to interact with DNA on a 2-fold symmetry axis: either two subunits could interact, each with the appropriate half of the symmetry region; or (as suggested by Dr. Thomas Steitz) all four subunits could interact, two with the left-hand region, the other two, related by a 2-fold axis, with the right-hand region. A 2-fold symmetry was suggested on genetic grounds by Sadler and Smith (14). The two longer regions are about a turn apart and can be approached from one side of the DNA helix. If these regions do reflect the interaction of the protein, then the DNase has shown some sequence specificity in cutting asymmetrically.

The structure of the *lac* operator is different from the λ operators isolated by Pirrotta (15) and analyzed by Maniatis and Ptashne (16). Maniatis, and Ptashne find a series of concatenated binding sites, 35–100 base pairs in length, while we isolate, under all conditions, only a single small fragment protected by the *lac* repressor.

The operator sequence does not immediately show how the repressor interacts with DNA. Mechanisms by which the repressor might detect the operator include: hydrogen bonding to the outside of the bases in the large or small groove of the DNA double helix, feeling variations in the positions of the phosphates that might be dictated by the A + T content affecting the structure (17), or opening up the DNA to see the bases directly. We discuss these possibilities more fully elsewhere (18).

The knowledge of both the nucleotide sequence of the operator and the complete amino-acid sequence of the repressor (19) should lead toward an understanding of this DNA–protein interaction. We anticipate that the sequences of o^c mutant operators will ultimately reveal which bases determine the structure that the repressor recognizes and will cast light on the actual interaction.

We thank Mary Archer and Joanna Knobler for technical assistance, Terry Platt and Ron Ogata for *lac* repressor, Charlotte Hering for RNA polymerase, and Richard Roberts for conversations about sequencing. This work was supported by USPHS Grant GM 09541 from the National Institute of General Medical Sciences. W.G. is an American Cancer Society Professor of Molecular Biology.

1. Jacob, F. & Monod, J. (1961) *J. Mol. Biol.* **3**, 318–356.
2. Gilbert, W. & Müller-Hill, B. (1966) *Proc. Nat. Acad. Sci. USA* **56**, 1891–1898.
3. Ptashne, M. (1967) *Proc. Nat. Acad. Sci. USA* **57**, 306–313.
4. Ptashne, M. (1967) *Nature* **214**, 232–234.
5. Gilbert, W. & Müller-Hill, B. (1967) *Proc. Nat. Acad. Sci. USA* **58**, 2415–2421.
6. Gilbert, W. (1972) in *Polymerization in Biological Systems*, Ciba Foundation Symposium (ASP, Amsterdam), Vol. 7 (new series), pp. 245–256.
7. Bourgeois, S. & Riggs, A. D. (1970) *Biochem. Biophys. Res. Commun.* **38**, 348–354.
8. Riggs, A. D., Suzuki, H. & Bourgeois, S. (1970) *J. Mol. Biol.* **48**, 67–83.
9. Burton, K. (1967) in *Methods in Enzymology*, eds. Grossman, L. & Moldave, K. (Academic Press, New York), Vol. XIIA, pp. 222–224.
10. Thach, R. (1966) in *Procedures in Nucleic Acid Research*, eds. Cantoni, G. L. & Davies, D. R. (Harper and Row, New York), Vol. 1, pp. 520–534.
11. Barrell, B. G. (1971) in *Procedures in Nucleic Acid Research*, eds. Cantoni, G. L. & Davies, D. R. (Harper and Row, New York), Vol. 2, pp. 751–779.
12. Downey, K. M. & So, A. G. (1970) *Biochemistry* **9**, 2520–2525.
13. Maizels, N. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 3585–3589.
14. Sadler, J. R. & Smith, T. F. (1971) *J. Mol. Biol.* **62**, 139–169.
15. Pirrotta, V. (1973) *Nature New Biol.* **244**, 13–16.
16. Maniatis, T. & Ptashne, M. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 1531–1535.
17. Bram, S. (1971) *Nature New Biol.* **232**, 174–176.
18. Gilbert, W., Maizels, N. & Maxam, A. (1973) *Cold Spring Harbor Symp. Quant. Biol.* **38**, in press.
19. Beyreuther, K., Adler, K., Geisler, N. & Klemm, A. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 3576–3580.
20. de Wachter, R. & Fiers, W. (1972) *Anal. Biochem.* **49**, 184–197.