THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 255, No. 3, Issue of February 10, pp. 1138–1145, 1980
Printed in U.S.A.

Genome Organization and Characterization of the Repetitive and Inverted Repeat DNA Sequences in *Neurospora crassa**

(Received for publication, September 5, 1979)

Robert Krumlauf‡ and George A. Marzluf§

From the Developmental Biology Program and Department of Biochemistry, Ohio State University, Columbus, Ohio 43210

Regulatory circuits have been defined in *Neurospora crassa* within which multiple unlinked genes are controlled in a positive fashion by major regulatory genes. It has been suggested that repetitive DNA sequences situated adjacent to structural genes may serve as regulatory sites. Experiments are presented which examine the amount and organization of the repetitive sequences of *Neurospora*. The genome consists of 2.7 × 10^7 base pairs and is composed of 2% foldback, 8% repetitive, and 90% single copy sequences. Our results indicate that the repeated DNA sequences are organized in stretches of 10,000 base pairs or longer. Experiments using pure ribosomal DNA isolated from cloned recombinant DNA demonstrated that rDNA represents 7% of total nuclear DNA and 88% of the repetitive component. The rDNA sequence is repeated about 185 times per nucleus. Since the ribosomal DNA units are believed to be tandemly repeated many times, their clustering can largely account for the long repetitive stretches and their lack of interspersion with single copy DNA. These results argue that repetitive DNA sequences identified by renaturation experiments do not serve a regulatory function in *Neurospora*. About 350 foci of inverted repeat, or foldback, DNA sequences occur in the genome and are interspersed with single copy DNA. The relatively small number of foldback foci suggest that inverted repeats also cannot be the primary recognition elements for control of gene expression in *Neurospora*. We suggest that very short repeated sequences, undetected by renaturation experiments, may serve as recognition sites for eukaryotic gene regulation.

The fungus *Neurospora crassa* is a useful experimental organism for the study of eukaryotic regulatory systems, using the classic approaches of biochemical genetics. A number of regulatory circuits have been defined within which one or more control genes regulate the expression of a group of related but unlinked structural genes (1, 2). Numerous unlinked genes which encode enzymes of nitrogen metabolism are, as a group, subject to ammonium repression and positively controlled by *nit-2*, a major nitrogen regulatory gene (3, 4). Similar types of positive control systems have been demonstrated for phosphorus metabolism (2), sulfur metabolism (5), and synthesis of branched chain amino acids (6). The structural genes, which encode three enzymes for quinate metabolism, are clustered and regulated by a closely linked control gene (7, 8). Although these three *qa* genes are tightly linked, they are transcribed as separate units. It has been postulated that each structural gene within a given control circuit possesses a similar adjacent DNA sequence which serves as a recognition site for a common regulatory protein (9). Such considerations suggest that each control circuit should contain a family of repeated DNA sequences, the degree of repetition corresponding to the number of structural genes within the circuit. However, very little is known about the nature and functional significance of repetitive DNA in the *Neurospora* genome so that it is difficult to judge whether it might have such a regulatory role.

The genomes of eukaryotic organisms are composed of two general classes of DNA sequences, the repetitive and single copy components. The arrangement of the repetitive and unique sequences displays a highly ordered pattern of sequence organization in most of the organisms examined (10, 11). This pattern is typified by short repetitive sequences, 200 to 400 nucleotide base pairs (Nbp) in length, interspersed among single copy sequences at intervals of 1,000 to 2,000 Nbp[1] (11). The occurrence of this short period interspersion pattern across a broad phylogenetic spectrum suggests a common function for repetitive sequences, perhaps in controlling gene expression. The repetitive sequences adjacent to single copy DNA have been proposed by Britten and Davidson (12, 13) to be involved in the regulation of gene expression, by possibly serving as recognition sites for the control of transcription. Studies in sea urchins and rats lend support to this idea and indicate that nearly 80% of the sequences transcribed into mRNA are contiguous to repetitive sequences (14, 15). More recently, they have proposed an alternative model which suggests that repetitive sequences may also control the pattern of gene expression post-transcriptionally (16) by forming intermolecular RNA-RNA duplexes between repetitive sequence transcripts and complementary sequences also present in structural gene transcripts. These duplexes are proposed to regulate the turnover and processing of nuclear transcripts into functional mRNA's.

Several studies have questioned whether the interspersion pattern of genome organization occurs in certain fungi (17, 18). Recently, we have characterized the sequence complexity of the components in the *Neurospora* genome and obtained results which indicated that the repetitive sequences were not

* This work was supported by Grant GM-23367 from the National Institutes of Health and Grant PCM 78-16021 from the National Science Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Supported by a National Institutes of Health Predoctoral Traineeship in Developmental Biology. Present address, Beatson Institute for Cancer Research, Glasgow, Scotland.

§ Supported by National Institutes of Health Career Development Award GM-00052. To whom communications should be addressed.

[1] The abbreviations used are: Nbp, nucleotide base pairs; C_0t, initial DNA concentration (mol/liter) × time (s); PB, sodium phosphate buffer; Nt, nucleotides; K, reassociation second order rate constant (M^{-1} s^{-1}); Pipes, 1,4-piperazinediethanesulfonic acid; chloramphenicol, D-(−)-*threo*-2,2-dichloro-*N*-[β-hydroxy-α-(hydroxymethyl)-*P*-nitrophenethyl]acetamide.

organized in a short period interspersion pattern (19). In this paper, we present experiments that characterize the nature of the repetitive DNA sequences of *Neurospora* and their organization in the genome. We also examine the properties of foldback DNA sequences to determine whether either of these sequence components could play a role in the coordinate control of gene expression in *Neurospora*.

MATERIALS AND METHODS

Isolation of Nuclear DNA—N. crassa strain 74-OR23-1a was grown for 14 h in Vogel's minimal medium (20) and harvested by filtration. Nuclei were isolated from mycelia according to the procedure of Hautala *et al.* (21) except that the homogenization step was performed in a CO_2-cooled Braun homogenizer using glass beads (19). DNA was isolated from the nuclei and purified by preparative CsCl density centrifugation. The details of this procedure have been described (19). The properties of the purified nuclear DNA were (i) $A_{260/280} = 2.00$; (ii) $A_{260/230} = 2.24$; (iii) buoyant density in CsCl, 1.711 g/ml; (iv) single-stranded fragment length of 13,000 nucleotides (Nt); and (v) hyperchromicity as a percentage of final absorbance, 26%.

Total nuclear DNA was labeled *in vivo* by growing 500 ml cultures of mycelia in minimal medium minus phosphate supplemented with 3 mCi of carrier-free $Na^{32}PO_4$ (New England Nuclear) and 0.01% HK_2PO_4. The labeled DNA was purified from isolated nuclei as described above. The specific activity of the isolated DNA was approximately 30,000 cpm/μg.

Isolation of Recombinant Plasmid DNA and Cloned Ribosomal Repeat Fragment—Escherichia coli C600 cells carrying a recombinant plasmid of pBR322 containing the coding regions of the ribosomal DNA (rDNA) repeat unit of *N. crassa*, designated pMF2, was graciously provided by S. Free and R. Metzenberg (22). The following experiments were performed under P2, EK1 containment conditions according to the National Institutes of Health "Guidelines for Recombinant DNA Research." An overnight culture of cells containing the pMF2 plasmid was diluted into 1 liter of L broth medium (10 g of bacto-peptone, 5 g of yeast extract, 10 g of NaCl, 1 g of D-glucose, and 1 ml of 1 N NaOH/liter). The cells were grown at 37°C to an A_{595} of 0.8 and then supplemented with 100 μg/ml of chloramphenicol and incubated an additional 16 to 18 h to amplify the plasmid. The cells were harvested by centrifugation and a cleared lysate was prepared by established procedures (23). To the cleared lysate, 1 g of CsCl and 400 μg of ethidium bromide were added per ml of lysate. This preparation was centrifuged in a Beckman Ti 60 rotor at 20°C for 60 h at 42,000 rpm. The bands of plasmid and chromosomal DNA were visualized with ultraviolet light. The lower plasmid band was removed by piercing the tube and collecting fractions. The ethidium bromide was removed by extracting with CsCl-saturated isopropyl alcohol. The DNA was then dialyzed against 20 mM Tris-HCl, pH 7.6, 1 mM EDTA, and stored at 4°C. Typically, 1 mg of plasmid DNA was obtained from a 1-liter culture.

The ribosomal DNA fragment was originally cloned in the ampicillin resistance gene of pBR322. Therefore, by restricting 200 μg of the plasmid DNA with 100 units of *Pst*I endonuclease (Bethesda Research Laboratories) at 37°C for 5 h, the recombinant plasmid was separated into two fragments: a 2.6×10^6-dalton pBR322 fragment and a 3.9×10^6-dalton *Neurospora* rDNA fragment (22). The two fragments were separated by preparative horizontal electrophoresis in 50 mM Tris, pH 7.9, 5 mM sodium acetate, 1 mM EDTA, and 1% agarose at 40 V for 16 h. The gel was stained with ethidium bromide (0.5 μg/ml) for 20 min and the DNA was visualized under ultraviolet light. The region of the gel containing the rDNA fragment was removed and the DNA was isolated using a modification of the procedure of Jeffreys and Flavell (24). The gel was dissolved in 10 M $NaClO_4$, 10 mM Tris-HCl, pH 7.6. The solution was then passed over a hydroxyapatite column twice to bind the DNA. The column was rinsed with several volumes of 5.3 M $NaClO_4$, 10 mM Tris-HCl, pH 7.6, to remove all traces of agarose and contaminants. The column temperature was then raised to 60°C and the $NaClO_4$ was removed by washing the column several times with 0.03 and 0.12 M sodium phosphate buffer (PB), pH 6.8. The DNA was eluted from the column by washing with 4 ml of 0.5 M PB. The DNA was then dialyzed against 2 mM Tris-HCl, pH 7.6, 1 mM EDTA, and stored at 4°C. Approximately 75% of the DNA was recovered from the gel with this procedure.

DNA Labeling—Total nuclear DNA and the purified rDNA fragments were labeled *in vitro* by nick translation based on the procedure of Rigby *et al.* (25). A 200-μl reaction mixture containing 50 mM PB, pH 7.6, 5 mM $MgCl_2$, 15 μM 5′-dGTP, 5′-dCTP, and 5′-dATP, 250 μCi of [^{3}H]5′-dTTP (New England Nuclear, 60 Ci/mmol), 1 to 2 μg of DNA, 1.0 ng of DNAse I (Worthington), and 100 units of DNA polymerase I (Boehringer Mannheim) was incubated at 14°C for 2 to 3 h. The details of the removal of the unincorporated label and any foldback DNA introduced by the labeling procedure have already been described (19). The specific activity of the DNA labeled under these conditions was 1 to 2×10^7 cpm/μg of DNA.

Shearing and Sizing of DNA—DNA preparations 400 nucleotides long were obtained by sonicating DNA samples in 0.5 M NaCl, 0.015 M sodium citrate, pH 7.6, at 4°C using a Branson 350 Sonifier equipped with a microprobe. The solutions were given 30-s pulses to a total of 3 min at a power setting of 5.0. *N. crassa* (^{32}P)DNA was sheared under varying conditions in a Virtis 60 homogenizer (26) at 4°C. A series of labeled fragments of sharply defined fragment lengths were then isolated from preparative alkaline sucrose gradients of this sheared DNA. All fragment lengths were determined by sedimentation through linear gradients containing 5 to 11% (w/w) sucrose in 0.1 M NaOH, 0.9 M NaCl, and 3 mM EDTA, with the assistance of a computer program and molecular weight markers to calibrate the gradients (27).

Reassociation Analysis—DNA-DNA reassociation experiments were performed in 0.12 M PB at 60°C in siliconized sealed glass capillaries. Samples were denatured in a boiling water bath for 10 min, incubated at 60°C to the desired C_0t (initial DNA concentration (mol/liter) × time (s)) value, then the reaction was stopped by rapid freezing. Chromatography on hydroxyapatite was used to separate single-stranded DNA fragments from those containing duplex regions (26). Each sample was diluted to 0.03 M PB and bound to hydroxyapatite at 60°C. The single-stranded DNA fraction was eluted with 0.14 M PB, 0.2% sodium lauroyl sarcosine and the duplex fraction was subsequently eluted with 0.5 M PB, 0.2% sodium lauroyl sarcosine. The fraction of DNA reassociated was determined from the total counts recovered in the single-stranded and duplex fractions. The curve fits to the data and the rate parameter values were calculated using the nonlinear least squares computer program of Pearson *et al.* (28).

Renaturation and Digestion with S1 Nuclease—All DNA samples to be treated with S1 nuclease (Sigma) were reassociated in 0.18 M NaCl, 0.01 M 1,4-piperazinediethanesulfonic acid (Pipes), pH 6.8. After incubation to the desired C_0t those samples to be digested under mild conditions (29) were adjusted to 0.12 M NaCl, 0.025 M sodium acetate, 5 mM Pipes, 0.1 M $ZnSO_4$, 5 mM dithiothreitol, pH 4.4. S1 nuclease was added in the ratio of 10 units/μg of DNA and incubated 45 min at 37°C. The reaction was terminated by the addition of 1.0 M PB to a final concentration of 0.12 M PB and passed over hydroxyapatite at 60°C. The S1-resistant bound DNA was washed twice with 0.12 M PB and then eluted with 0.5 M PB. The size distribution of the resistant duplexes was measured on a column (90 × 1.5 cm) of agarose A-50 (Bio-Rad). The gel bed was poured around a support of siliconized 6-mm glass beads as described by Britten *et al.* (26). Samples were chromatographed in 0.12 M PB using $Na^{32}PO_4$ as an inclusion marker and blue dextran as an exclusion marker.

Preparations of foldback DNA were prepared by reassociating the DNA to a C_0t of 10^{-4} to 10^{-5} followed by digestion with S1 nuclease under more stringent conditions. These samples were reassociated in the same buffer, but after incubation they were adjusted to 0.12 M NaCl, 0.07 M sodium acetate, 5 mM Pipes, 2.5 mM $ZnSO_4$, 5 mM dithiothreitol, pH 4.4. S1 nuclease was added in the ratio of 50 units/μg of DNA and incubated 1 h at 45°C. The reaction was terminated by adding 1.0 M PB as before. The S1-resistant duplexes were either separated on hydroxyapatite or precipitated with 10% trichloroacetic acid plus carrier and collected on 0.45 μ nitrocellulose filters (Millipore).

Filter Hybridization—DNA samples were electrophoresed for 14 h at 30 V in 1% agarose, 50 mM Tris, pH 7.9, 5 mM sodium acetate, and 1 mM EDTA. The gel was stained with 0.5 μg/ml of ethidium bromide, visualized with ultraviolet light and photographed on Kodak Royal X Pan film for reference. The DNA bands were transferred to nitrocellulose filter paper (HAWP, Millipore) by the blotting method of Southern (30). The fragments were transferred in 24 h and the filter was washed in 0.3 M NaCl, 0.03 M sodium citrate, pH 7.6, blotted dry, and baked for 2 h in a vacuum oven at 80°C. The filter was prepared for hybridization according to the procedure of Jeffreys and Flavell (24). The filter was then transferred to a set of tightly interlocking siliconized glass tubes and immersed in 2.0 ml of 0.2% Ficoll 400 (Sigma), 0.2% polyvinylpyrollidone (Sigma), 0.2% bovine

serum albumin, 50 μg/ml of denatured 350 nucleotide salmon sperm DNA (Sigma), 0.1% sodium dodecyl sulfate, and 5×10^5 cpm of the nick-translated [^{3}H]rDNA fragment in 0.45 M NaCl, 0.045 M sodium citrate, pH 7.6, at 65°C. The tubes were sealed and allowed to incubate for 3 days. After hybridization, the filter was washed at 65°C for a total of 40 min with six changes of the complete hybridization solution without the probe. The filter was then washed in 0.015 M NaCl, 0.0015 M sodium citrate, pH 7.6, and 0.1% sodium dodecyl sulfate for 30 min to remove nonspecific hybrids and blotted dry. The filter was dipped in a solution of 2,5-diphenyloxazole in toluene (20%, w/v) and dried in air. Radioactivity was detected by fluorography with Kodak X-Omat R film at −70°C (30).

RESULTS

Sequence Components of the Neurospora Genome—We have characterized the sequence complexity of the components in the *Neurospora* genome (19). The reassociation kinetics of total nuclear DNA fragments (400 Nt) and the best least squares solution for the results is shown in Fig. 1 *A*. The analysis reveals that two second order kinetic components are present, repetitive and single copy DNA sequences. The single copy component represents 90% and the repetitive 7% of the total DNA binding to hydroxyapatite. The remaining 3% of the sequences are early binding or foldback DNA. Additional experiments were used to accurately determine the repetitive and the single copy rate constants as shown in Fig. 1 *B*. A ^{32}P-labeled highly enriched repetitive fraction and also a single copy fraction were isolated by reassociation of total (^{32}P)DNA to the appropriate C_0t values followed by fractionation on hydroxyapatite. These labeled isolated kinetic components were then reassociated as tracers, each with an excess of unlabeled total DNA. The repetitive tracer reassociated as a single second order component with a rate in whole DNA of 5.6 M^{-1} s^{-1}; a small amount of contaminating unique DNA sequences were also detected in this experiment. The single copy tracer reassociated with a rate of 0.04 M^{-1} s^{-1}. With these accurate values for the rate constants for reassociation of the repetitive and single copy sequences, the results of reassociation of total nuclear DNA were then recalculated to determine the exact fraction and complexity of each component in the genome.

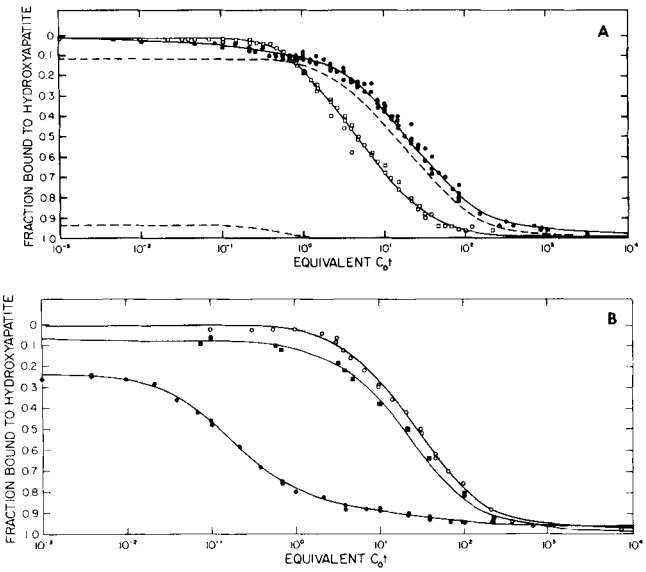

FIG. 1. **Reassociation kinetics of *Neurospora* nuclear DNA.** *A*, the best least squares solution of the reassociation of 400-Nt ^{32}P-labeled total DNA fragments (●) indicates two second order components shown by the *dashed curves*: 7% (repetitive) with a K of 6.43 M^{-1} s^{-1} and 90% (single copy) with a K of 0.043 M^{-1} s^{-1}. The root mean square of the solution is 2.00% and approximately 3% of the DNA (foldback) bound to hydroxyapatite at a C_0t of 10^{-3}. The best least squares solution of the reassociation of 400-Nt ^{3}H-labeled *E. coli* DNA fragments (□), used as a kinetic standard, indicates a single second order component: 98% with a K of 0.23 M^{-1} s^{-1} and root mean square of 2.19%. *B*, slave mini-C_0t reassociations of isolated repetitive and single copy components. The repetitive component was isolated as the fraction of 400-Nt ^{32}P-labeled DNA bound to hydroxyapatite (9.9%) at a C_0t of 0.75, a value at which 80% of the repetitive and 3% of the single copy components were expected to form duplexes. The isolated 400-Nt ^{32}P repetitive sequences were reassociated in the presence of a 350-fold excess of unlabeled 400-Nt total DNA fragments (●). Two second-order kinetic components were resolved from the best least squares solution of the data: 67% (repetitive) with a K of 5.58 M^{-1} s^{-1} and 8.5% (residual single copy) with a K of 0.01 M^{-1} s^{-1}. At a C_0t of 10^{-3}, 25% of the tracer (foldback) was already bound to hydroxyapatite. The root mean square of this solution is 1.09%. The single copy tracer was isolated as that fraction of 400-Nt ^{32}P-labeled total DNA not bound to hydroxyapatite (66.5%) at a C_0t of 15.0. The 400-Nt ^{32}P-labeled single copy tracer was reassociated with a 1000-fold excess of unlabeled 400-Nt total DNA fragments (○). The *curve* represents the best least squares solution of the data indicating only one second order component, 98% (single copy) with a K of 0.039 M^{-1} s^{-1} and a root mean square of 2.03%. The reassociation of the unlabeled 400-Nt total DNA driver in the above reassociations is shown for comparison (■).

The analysis of the combined data from the experiments shown in Fig. 1 is summarized in Table I. We calculate the genome size of *Neurospora* to be 2.7×10^7 Nbp, using the reassociation of *E. coli* DNA as a kinetic standard (26). The single copy component has sufficient complexity to contain 18,000 average-sized structural genes. However, the 15,300 Nbp complexity and 140 average reiteration frequency of the repetitive DNA is low and suggests that this component might largely consist of sequences coding for ribosomal and transfer RNA species.

Length and Organization of the Repetitive DNA—A sensitive experiment to determine the length of repetitive sequences and the extent of their interspersion with single copy DNA is presented in Fig. 2. Long single-stranded fragments (8050 Nt) were reassociated to a C_0t of 0.3 at which 90% of the repetitive DNA, but only 4% of the single copy DNA, form duplexes. This preparation was then treated with S1 nuclease which specifically removes all single-stranded DNA (98%) under conditions in which duplex regions with as much as 10 to 20% mismatched base pairs are not attacked (29). The S1-nuclease-resistant duplexes were isolated by binding to hydroxyapatite and their size profile was determined by chromatography on Bio-Gel agarose A-50. The difference in the

TABLE I

Composition of the Neurospora genome

Component	Fraction of genome[a]	K[b]	$C_0t½$[b]	K_p[c]	Repetition frequency[d]	Kinetic complexity[e] (Nbp)
		M^{-1} s^{-1}		M^{-1} s^{-1}		
Foldback	0.02					
Repetitive	0.08	5.6	0.178	70.0	140	1.5×10^4
Single copy	0.90	0.040	25.0	0.045	1	2.42×10^7

[a] Determined by the best least squares computer solution of the combined data from reassociation of the repetitive and single copy components and the renaturation of 400-Nt ^{32}P-labeled total DNA fragments.

[b] The second order rate constant for reassociation of 400-Nt DNA fragments in whole DNA. $C_0t½$ is the inverse of K.

[c] The second order rate constant for reassociation of 400-Nt fragments of each component if pure. Calculated by dividing the rate for the component in whole DNA by the fraction of the genome which that component represents.

[d] Calculated from the ratio of the rate constant for the repetitive and single copy components in whole DNA.

[e] The complexity of each component in the genome. Computed from the genome size of *E. coli*, 4.2×10^6 Nbp, and the ratio of the measured rate constant for reassociation of 400-Nt *E. coli* DNA fragments and the rates constant (if pure) for each component.

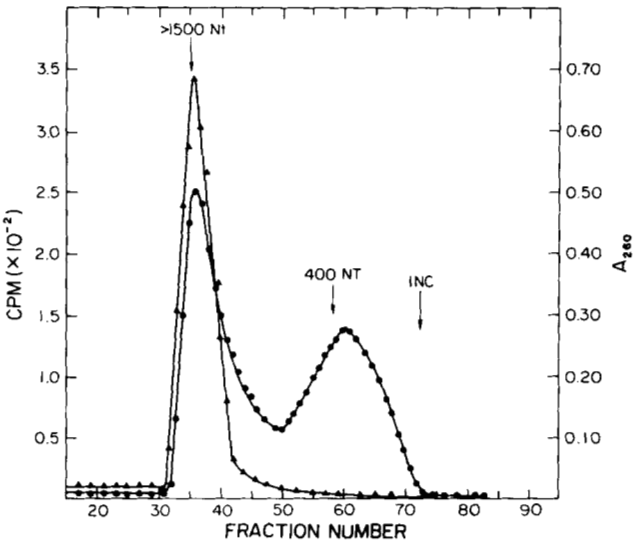

FIG. 2. **Profile of repetitive DNA duplexes resistant to S1 nuclease.** Fragments of total *Neurospora* DNA 8050 Nt long were reassociated to a repetitive C_0t of 0.3 at 60°C in 0.18 M NaCl, 0.01 M Pipes, pH 6.8. Under these conditions, 21% of the fragments contain duplex regions and bind to hydroxyapatite. After incubation, the samples were directly treated with S1 nuclease for 45 min at 37°C according to the mild conditions of Britten *et al.* (29) as described under "Materials and Methods." The S1 nuclease-resistant repetitive duplexes were isolated by fractionation on hydroxyapatite and represented 18.1% of the input DNA. The size profile of these duplexes was measured by chromatography on agarose A-50 (▲). Approximately 95% of all the resistant duplexes from *Neurospora* are excluded from the agarose. To test the assay conditions, mouse 3T6 fibroblast DNA was used as a control. Fragments of 3T6 DNA 10,100 Nt long were reassociated to a repetitive C_0t of 50 and treated in the same manner. The size distribution of the 3T6 S1 nuclease-resistant duplexes on agarose A-50 is shown (●). Only 34% of the S1-resistant duplexes are excluded from the agarose column. The *arrows* indicate the elution position of ^{32}P-labeled DNA markers and the $Na^{32}PO_4$ inclusion (*INC*) marker.

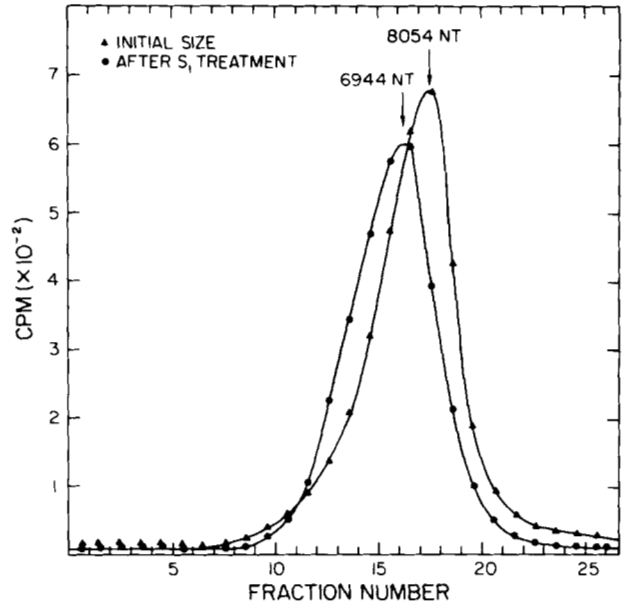

FIG. 3. **Size of repetitive DNA in duplexes resistant to S1 nuclease.** The single-stranded length of DNA in repetitive duplexes was determined by alkaline sucrose gradient centrifugation as described under "Materials and Methods." The size of the initial fragments (▲) was determined to be 8054 Nt and the size after treatment with S1 nuclease (●) was 7944 Nt.

amount of binding to hydroxyapatite of such duplexes treated (19.1%) and not treated (21.0%) with S1 nuclease is due to the presence of single-stranded tails in the latter. Therefore, 91% (19.1/21.0) of the length of fragments which reassociated at the repetitive C_0t were in duplex structures, suggesting that the repeated sequences are at least 7300 Nbp (8050 × 0.91) in length. Fig. 2 shows that the S1-resistant repetitive duplexes of *Neurospora* eluted as a single peak at the excluded volume of the agarose column, which corresponds to sizes of 1500 Nbp or greater. No duplex structures with a length of about 300 Nbp, which is characteristic of the short period interspersion pattern of repetitive DNA, could be detected. The single-stranded length of the DNA in the repetitive duplexes recovered from the agarose column was determined by alkaline sucrose gradients as shown in Fig. 3. Their length was found to be 7000 Nt or 90% of the original fragment length, in excellent agreement with the 7300 Nt length calculated from the hydroxyapatite binding data. The 10% of the DNA in *Neurospora* repetitive duplexes that is sensitive to S1 nuclease is probably due to fragment overlap which results from the reassociation of randomly sheared fragments, and not from linkage to single copy sequences.

Mouse DNA, which we expected should possess the typical short interspersed repetitive sequence pattern observed in rat (31) and most other eukaryotes (10), was treated in an identical manner as a control. The S1-resistant repetitive duplexes of mouse 3T6 fibroblast DNA were resolved by the agarose column into two size classes as shown in Fig. 2. A minor fraction (34%) of the duplexes was excluded from the column while the majority chromatographed as a class of fragments with a modal single-stranded length of 325 Nt, characteristic of the short period interspersion pattern. Thus, the control makes it clear that *Neurospora* lacks such short repeated elements.

The sensitivity of this experiment and the criterion of the reassociation would have enabled us to detect any significant amount of repetitive duplexes greater than 25 Nbp present in the genome. This experiment demonstrates clearly that short interspersed repetitive sequences are not a characteristic of the *Neurospora* genome, but rather that repeated DNA occurs in long stretches of at least 8000 Nbp and perhaps much longer. In other experiments, we have obtained results that suggest the repetitive sequences are at least 10,000 Nbp in length and that very little of the single copy sequences are contiguous to repeated DNA in fragments of this length (19).

Measurement of the rDNA Repeat Frequency—Since these experiments demonstrated that the repeated DNA of *Neurospora* is clustered in very long stretches and is not adjacent to single copy sequences, we were interested in determining its functional significance. The size of the *Neurospora* rDNA repeat unit which encodes the 17 S, 5.8 S, and 25 S rRNA species plus spacer regions has been established to be about 6 megadaltons (22, 32). A simple calculation based upon this unit size and the possibility that the rDNA sequence may be repeated between 100 and 200 times, as found in yeast, revealed that between 5 and 8% of the total genome could be devoted to ribosomal DNA. To directly test this possibility, we utilized pure rDNA, obtained as a cloned *Neurospora* sequence containing the major portion of the ribosomal repeat unit inserted into plasmid pBR322 (22). The pure rDNA sequences were used in reassociation experiments to determine the amount of ribosomal DNA and its repeat frequency in the genome. Fig. 4 *A* describes the separation and isolation of the *Neurospora* rDNA from the recombinant plasmid. We verified that the isolated rDNA has the correct structure and sequence by restriction analysis. The rDNA fragment was labeled *in vitro* by nick translation and its specificity was

demonstrated by filter hybridization (Fig. 4 *B*). Nearly all of the labeled probe (96%) hybridized to DNA bands containing the rDNA sequences and no significant hybridization to other sequences could be detected. The labeled rDNA sequence was reassociated with an excess of total nuclear DNA as shown in Fig. 5. The ratio of the rate of reassociation of this tracer to the rate of reassociation of single copy DNA sequences is a direct measure of the rDNA repeat frequency. The best least squares solution of this experiment revealed that the tracer reassociated as a single second order component with a K of 7.9 M^{-1} s^{-1}. The rate of reassociation for single copy DNA is 0.04 M^{-1} s^{-1} (Table I). Therefore, these results indicate that the rDNA sequences are repeated 197 times (7.9/0.04) in the genome. Since the entire repeat unit is 8600 to 9700 Nbp (22, 32), the rDNA sequences represent 1.7 to 1.9 × 10^6 Nbp which comprises 6.3 to 7.1% of the total DNA.

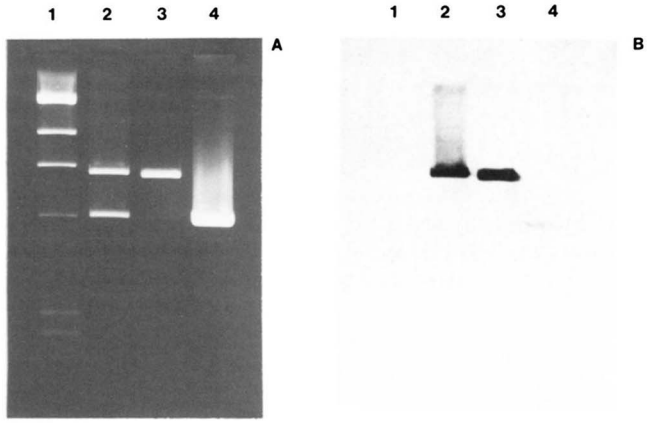

FIG. 4. **Isolation of labeled rDNA.** *A*, DNA samples were electrophoresed in 1% agarose gels, stained with ethidium bromide, visualized with UV light, and photographed. *Lane 1* represents molecular weight markers prepared by digesting λ DNA with *Hin*dIII. *Lane 2* is the recombinant plasmid DNA, pMF2, restricted with *Pst*I. The *upper band* is the 3.9 × 10^6-dalton rDNA repeat fragment and the *lower band* is the 2.6 × 10^6-dalton pBR322 fragment. *Lane 3* is the purified rDNA repeat fragment separated from the *Pst*I digest of the recombinant plasmid pMF2 by preparative electrophoresis and isolated with hydroxyapatite. *Lane 4* is authentic plasmid pBR322 digested with *Pst*I. *B*, hybridization of labeled rDNA. The DNA fragments in *A* were denatured and transferred to a nitrocellulose filter via the method of Southern (30). The purified rDNA was ^{3}H-labeled *in vitro* by nick translation and hybridized to the filter in 3 × SSC according to Jeffreys and Flavell (24). The washed filter was then fluorographed. The probe hybridized only to the upper rDNA band in *Lane 2* and the purified rDNA fragment in *Lane 3*. No hybridization can be detected to any of the markers in *Lane 1* or the pBR322 fragments in *Lanes 2* and *4*. The probe is therefore specific for the rDNA repeat sequences.

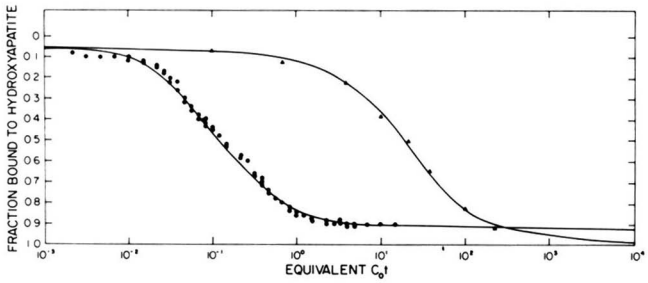

FIG. 5. **Reassociation of ^{3}H-labeled rDNA tracer with total nuclear DNA.** The purified rDNA repeat fragment was isolated and labeled *in vitro* as described (see "Materials and Methods"). Labeled fragments 400-Nt long were isolated by preparative alkaline sucrose centrifugation and reassociated with a 500-fold excess of unlabeled 400-Nt total nuclear DNA fragments (●). The best least squares solution of this data yields a single second-order component: 88.4% (rDNA genes) with a K of 7.9 M^{-1} s^{-1} and root mean square of 1.05%. The reassociation of the unlabeled 400-Nt total DNA driver is shown (▲).

Total Amount of rDNA in the Genome—A completely independent method was used to confirm the repeat frequency and amount of rDNA in the genome. In this experiment, total nuclear DNA from *Neurospora* was labeled by nick translation and the foldback sequences were removed. This ^{3}H total DNA was reassociated with a large excess of the recombinant plasmid DNA containing the rDNA sequences to determine what fraction of the total DNA represents these ribosomal DNA sequences. At saturation, 4.2% of the labeled total DNA was driven into duplexes and therefore contained rDNA sequences (Fig. 6). Since the cloned rDNA contains only 3.9 × 10^6 daltons of the entire 6.0 × 10^6-dalton rDNA repeat unit, 6.3% (4.2% × 6.0/3.9) of the genome is actually rDNA. These values agree very well with those obtained from Fig. 5. Therefore, 6.3 to 7.1% of the total nuclear DNA is rDNA and there are about 185 copies of the repeat unit per nucleus. The combined results of these experiments make it obvious that nearly all (90%) of the repeated DNA can be accounted for as rDNA and the small remaining amount may represent sequences coding for tRNA, 5 S RNA, and possibly histones.

Characteristics of Foldback DNA—A sizeable fraction of *Neurospora* DNA contains inverted repeats or foldback DNA that will bind to hydroxyapatite following denaturation and rapid reassociation (C_0t of 10^{-5}). The amount of foldback DNA was determined by rapidly reassociating (^{32}P)DNA fragments 2400, 5100, 8200, and 13,300 Nt long followed by treatment with S1 nuclease under stringent conditions. The S1-resistant fragments were collected on hydroxyapatite and sized. The data is summarized in Table II. The fraction of DNA bound to hydroxyapatite before treatment with S1 nuclease increases with the fragment length due to the presence of single-stranded tails linked to the foldback sequences, as is also shown in Fig. 7. However, the amount (1.5 to 1.8%) and length (505 to 710 Nt) of S1-resistant (foldback) DNA in the genome remains relatively constant for each fragment length. It should be noted that since the difference in the amount of foldback duplexes binding to hydroxyapatite before and after treatment with S1 nuclease is a measure of the actual portion of the fragments that are double stranded, we can predict the length of foldback sequences in a given fragment length from these values. As shown in Table II, the lengths predicted from the fraction of each fragment resistant

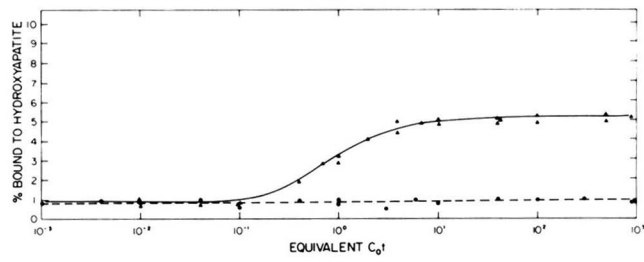

FIG. 6. **Saturation of labeled total nuclear DNA with purified rDNA.** Total *Neurospora* nuclear DNA was labeled by nick translation and the foldback sequences were removed. The 425-Nt ^{3}H-labeled total DNA was reassociated in the presence of a 100-fold excess of unlabeled 475-Nt sheared recombinant plasmid DNA containing the rDNA sequences (▲). The least squares solution of this curve yields a final saturation value of 5.1% of total DNA. A control demonstrated that the labeled repetitive sequences did not reassociate with themselves under these conditions. The 425-Nt ^{3}H-labeled tracer was reassociated with a 1000-fold excess of unlabeled 400 Nt-calf thymus DNA (●). No detectable increase in the amount of tracer bound to hydroxyapatite occurred during the incubation, and 0.9% of the label bound at the earliest C_0t values. Therefore, 4.2% (5.1 minus 0.9%) of the total nuclear DNA were driven to duplexes by pure rDNA.

TABLE II

Properties of Neurospora foldback DNA

Total *Neurospora* (^{32}P)DNA was denatured then reassociated to a C_0t of 10^{-5} to allow the foldback sequences to renature. The foldback sequences were isolated by chromatography on hydroxyapatite, both before and after being treated with S1 nuclease under stringent conditions as described under "Materials and Methods." The size of the S1-resistant fragments was determined by alkaline sucrose centrifugation.

	Initial fragment size in nucleotides			
	2,400	5,100	8,200	13,300
Fraction of DNA bound to hydroxyapatite	0.039	0.072	0.095	0.158
Fraction of DNA bound to hydroxyapatite after S1 digestion	0.015	0.016	0.018	0.018
Amount of foldback DNA[a] per genome in base pairs	4.05×10^5	4.32×10^5	4.86×10^5	4.86×10^5
Single-stranded size of S1-resistant DNA in base pairs				
Experimentally determined	505	620	700	710
Predicted from hydroxyapatite[b] binding data	460	565	780	760
Number of inverted repeats per genome	805	700	690	685
Average distance between foldback foci in base pairs	6.71×10^4	7.7×10^4	7.83×10^4	7.88×10^4

[a] Based on a genome size of 2.7×10^7 base pairs.

[b] Calculated by (fraction of DNA bound to hydroxyapatite after S1/fraction bound to hydroxyapatite) × (fragment length/2).

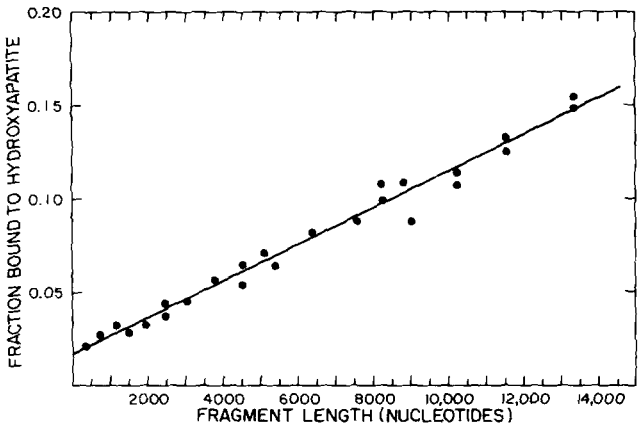

FIG. 7. **Fraction of *Neurospora* DNA containing foldback sequences as a function of length.** Total (^{32}P)DNA fragments of various lengths were denatured and rapidly reassociated to a C_0t of 10^{-5} and the fraction of DNA containing the duplex foldback sequences was isolated by fractionation on hydroxyapatite at 60°C. The *line* through the data represents the best linear least squares solution to the data. The fit has a slope of 0.0000095 and a y intercept of 0.017.

to S1 nuclease are in good agreement with those experimentally determined.

At increasingly larger fragment lengths, the length of the S1-resistant DNA and the fraction of foldback DNA plateau at about 700 Nt and 1.8%, respectively, since very few of the inverted repeats would be interrupted at these larger fragment lengths. Assuming that the inverted repeats have an average modal size of 700 Nbp and that they comprise 1.8% of the genome, we calculate that approximately 700 inverted repeats occur in the genome (Table II). Since an inverted repeat on one strand of the DNA must have an identical repeat on the complementary strand, there would actually only be 350 unique foldback foci within the genome. If these foldback foci were uniformly distributed throughout the genome, the average distance between foci would be about 78,000 Nbp.

In order to estimate the actual distribution of foldback foci, the fraction of total *Neurospora* DNA containing foldback sequences as a function of increasing fragment length was determined as shown in Fig. 7. The increased fraction of DNA bound to hydroxyapatite as the length increases clearly indicates that the foldback DNA is interspersed with single copy DNA. From these results, the distribution of the foldback foci was estimated by the use of equations derived by Wilson and Thomas (33) which describe the results predicted for both a uniform and random distribution pattern of foldback sequences in the genome. The linear relationship obtained in Fig. 7 implies that the inverted repeats are interspersed at regular intervals in a uniform pattern. Furthermore, the reciprocal of the slope of this line describes the average spacing interval in this pattern. We calculate this value to be 105,000 Nbp ($1/9.5 \times 10^{-6}$) which is slightly larger than that estimated (78,000 Nbp) from the size and fraction of S1-resistant DNA. The simplest interpretation of these results is that the inverted repeats are regularly spaced and show only slight clustering.

DISCUSSION

The results of our experiments indicate that *Neurospora* has a very simple genome but includes many of the features common in higher eukaryotes. The genome has a size of 2.7×10^7 Nbp and is composed of three sequence components: 2% foldback, 8% repetitive, and 90% single copy DNA. The repetitive sequences have a complexity of only 15,300 Nbp and are reiterated about 140 times, which suggest they might be primarily rDNA sequences. The experiments using cloned rDNA sequences directly demonstrated that this was indeed the case. The large repeating unit which contains the coding sequences for 17 S, 5.8 S, and 25 S rRNA and the spacer regions is present in about 185 copies per genome and comprises approximately 7% of total nuclear DNA and 90% of the repeated DNA. The repetitive sequences may also code for tRNA and 5 S RNA species. Considering the complexity of the rDNA repeat and these additional RNA species, we can clearly account for nearly all of the complexity and amount of repeated DNA as structural RNA coding sequences.

The *Neurospora* repetitive DNA is not organized in a short period interspersion pattern. The repetitive and single copy sequences are not linked to any appreciable extent in the genome. Additional results have confirmed that the repeated sequences are organized in lengths of at least 10,000 Nbp and perhaps even much longer. As an upper limit, only 0.6% of the single copy sequences are contiguous to repeated DNA in fragments up to 10,000 Nbp long (19). Fig. 2 definitely shows that no 200 to 400 Nbp class of repeated DNA exists in *Neurospora*. The long length of repetitive DNA stretches in *Neurospora* does not imply a long period interspersion pattern similar to that found in *Drosophila* (34); rather, *Neurospora* repetitive DNA does not show any significant interspersion,

presumably as a consequence of the organization of the rDNA sequences. We believe the rDNA units are tandemly repeated many times and perhaps all reside on a single chromosome, as is the case in yeast (35, 36). Their clustering largely accounts for the long stretches of repetitive DNA and their lack of interspersion with single copy DNA. In view of our results with *Neurospora* and those with other fungi (17–19, 35) we believe that this pattern of organization, which results from the rDNA content and its organization, may be a general feature of fungal genomes.

The foldback DNA sequences, in contrast to the repetitive sequences, are interspersed throughout the genome at regular intervals. The foldback sequences represent 1.8% of the total DNA and have a modal single-stranded size of about 700 Nbp, which corresponds to approximately 700 inverted repeats and 350 unique foldback foci per genome. The average spacing between these foci is 78,000 to 105,000 Nbp. If these are the primary sequences involved in regulating gene expression, it would place an approximate upper limit on the number of regulated genes at 350. The single copy sequences of *Neurospora* contain the potential for about 18,000 different genes. Recent results have shown that approximately 2000 mRNAs, in three different abundance classes, are expressed in rapidly growing vegetative cells.[2] In addition, there are a large number of other genes which are expressed throughout various developmental stages and in response to metabolic limitations. Although it is difficult to estimate the number of genes subject to regulation, we predict that this number will surely be greater than 350 in *Neurospora* and may in fact represent several thousand genes. Therefore, it seems to us to be unlikely that the foldback sequences are the primary elements responsible for the coordinate control of gene expression. Any role for foldback DNA elements in regulation would predict that the sequences adjacent to them are functionally expressed. It would also imply that the genes subject to regulation are spaced at very large regular intervals in the genome, which seems unlikely. We are presently examining whether those sequences adjacent to foldback DNA are transcriptionally active.

A striking difference seems to occur in the size of the primary nuclear transcripts between fungi and higher eukaryotes (37, 38), since the latter have primary transcripts which are much larger than the mRNA derived from them. Furthermore, with the exception of yeast tRNA genes (39), no evidence exists for the interruption of unique nuclear genes in lower eukaryotes. There are, however, many similarities in the chromatin structure, transcriptional machinery, and the processing of mRNA between fungi and higher eukaryotes. The degree to which the basic mechanisms that are responsible for differential gene expression in higher and lower eukaryotes are similar is uncertain. The lack of any short repetitive sequences detectable by reassociation analysis in a short or long period interspersion pattern strongly argues against a regulatory role for these repetitive sequences in *Neurospora*. However, it is well established that in *Neurospora* sets of unlinked genes are coordinately and positively controlled, apparently at the transcriptional level (1–8). Those repeated DNA sequences which we recognized in reassociation experiments are clearly rDNA and could not be responsible for gene regulation as suggested in the models of Britten and Davidson (12, 13, 16). Therefore, any model that bases eukaryotic gene regulation on the classes of repetitive sequences identified under renaturation conditions must provide alternative mechanisms for regulation in *Neurospora* and other fungi. It seems likely that very short repetitive regulatory elements do exist which are contiguous with structural genes. It is noteworthy that a biologically functional cloned lactose operator of *E. coli* is comprised of only 17 Nbp (40). If such repetitive sequences do occur (≤25 Nbp), they would not have been detected by the methods employed in these studies. A definitive answer to this question can probably be most directly obtained by cloning and partial sequencing of several *Neurospora* structural genes which are members of a single regulatory circuit.

[2] L. C. Wong and G. A. Marzluf, manuscript in preparation.

Acknowledgments—We thank Dr. Steve Free and Dr. Robert Metzenberg for providing the rDNA recombinant clone and Dr. Lee Johnson for the gift of mouse 3T6 fibroblast DNA. We also thank Dr. William Timberlake for helpful discussions and Dr. Brian Young for the computer program used to analyze sucrose gradients.

REFERENCES

1. Marzluf, G. A. (1977) in *Regulatory Biology*, (Marzluf, G. A., and Copeland, J. eds) pp. 196–242, Ohio State University Press, Columbus
2. Metzenberg, R. L., and Nelson, R. E. (1977) in *Molecular Approaches to Eucaryotic Genetic Systems*, ICN-UCLA Symposia on Molecular and Cellular Biology, Vol. VI, (Wilcox, G., Abelson, J., and Fox, C. F., eds) pp. 253–268, Academic Press, New York
3. Reinert, W. R., and Marzluf, G. A. (1975) *Mol. & Gen. Genet.* **139,** 39–55
4. Facklam, T. J., and Marzluf, G. A. (1977) *Biochem. Genet.* **16,** 343–354
5. Burton, E. G., and Metzenberg, R. L. (1972) *J. Bacteriol.* **109,** 140–151
6. Olshan, A. R., and Gross, S. R. (1974) *J. Bacteriol.* **118,** 374–384
7. Reinert, W. R., and Giles, N. H. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 4256–4260
8. Jacobson, J. W., Hautala, J. A., Lucas, M. C., Reinert, W. R., Stroman, P., Barea, J. L., Patel, V. B., Case, M. F., and Giles, N. H. (1977) in *Molecular Approaches to Eucaryotic Genetic Systems*, ICN-UCLA Symposia on Molecular and Cellular Biology, Vol. VI, (Wilcox, G., Abelson, J., and Fox, C. F., eds) pp. 262–284, Academic Press, New York
9. Hanson, M. A., and Marzluf, G. A. (1975) *Proc. Natl. Acad. Sci. U. S. A.* **72,** 1240–1244
10. Davidson, E. H., Galau, G. A., Angerer, R. C., and Britten, R. J. (1975) *Chromosoma (Berl.)* **51,** 253–259
11. Davidson, E. H., Hough, B. R., Amenson, C. S., and Britten, R. J. (1973) *J. Mol. Biol.* **77,** 1–23
12. Davidson, E. H., and Britten, R. J. (1973) *Q. Rev. Biol.* **48,** 565–613
13. Britten, R. J., and Davidson, E. H. (1969) *Science* **165,** 349–357
14. Davidson, E. H., Hough, B. R., Klein, W. R., and Britten, R. J. (1975) *Cell* **4,** 217–238
15. Holmes, D. S., and Bonner, J. (1974) *Proc. Natl. Acad. Sci. U. S. A.* **71,** 1108–1112
16. Davidson, E. H., and Britten, R. J. (1979) *Science* **204,** 1052–1059
17. Hudspeth, M. E. S., Timberlake, W. E., and Goldberg, R. E. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 4332–4336
18. Timberlake, W. E. (1978) *Science* **202,** 973–975
19. Krumlauf, R., and Marzluf, G. A. (1979) *Biochemistry* **18,** 3705–3713
20. Davis, R. H., and DeSerres, F. J. (1970) *Methods Enzymol.* **17A,** 79–108
21. Hautala, J. A., Conner, B. H., Jacobson, J. W., Patel, G. L., and Giles, N. H. (1977) *J. Bacteriol.* **130,** 704–713
22. Free, S. J., Rice, P. W., and Metzenberg, R. L. (1979) *J. Bacteriol.* **137,** 1219–1226
23. Guerry, P., LeBlanc, D. J., Falkow, S. (1973) *J. Bacteriol.* **116,** 1064–1066
24. Jeffreys, A. J., and Flavell, R. A. (1977) *Cell* **12,** 429–439
25. Rigby, P. W. J., Diekmann, M., Rhodes, C., and Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251
26. Britten, R. J., Graham, E. E., and Neufeld, B. R. (1974) *Methods Enzymol.* **29E,** 363–418
27. Hell, A., Birnie, G. D., Slimming, T. K., and Paul, J. (1972) *Anal. Biochem.* **48,** 369–377
28. Pearson, W. R., Davidson, E. H., and Britten, R. J. (1977) *Nucleic Acids Res.* **4,** 1727–1735
29. Britten, R. J., Graham, D. E., Eden, F. C., Painchaud, D. M., and Davidson, E. H. (1976) *J. Mol. Evol.* **9,** 1–23

30. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517
31. Pearson, W. R., Wu, J.-R., and Bonner, J. (1978) *Biochemistry* **17,** 51–59
32. Cox, R. A., and Peden, K. (1979) *Mol. & Gen. Genet.* **174,** 17–24
33. Wilson, D. A., and Thomas, C. A. (1974) *J. Mol. Biol.* **84,** 115–144
34. Manning, J. E., Schmid, C. W., and Davidson, N. (1975) *Cell* **4,** 141–155
35. Petes, T. D., and Botstein, D. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5091–5095
36. Petes, T. D. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76,** 410–414
37. Hereford, L. M., and Rosbash, M. (1977) *Cell* **10,** 453
38. Timberlake, W. E., Shumard, D. S., and Goldberg, R. B. (1977) *Cell* **10,** 623–632
39. Knapp, G., Beckmann, J. S., Johnson, P. F., Fuhrman, S. A., and Abelson, J. (1978) *Cell* **14,** 221–236
40. Bahl, C. P., and Wu, R. (1978) *Gene* **3,** 123–134