HIGH SEQUENCE DIVERSITY IN THE RNA SYNTHESIZED AT THE LAMPBRUSH STAGE OF OÖGENESIS[*]

By Eric H. Davidson and Barbara R. Hough

THE ROCKEFELLER UNIVERSITY

Communicated by Alfred E. Mirsky, March 7, 1969

Abstract.—Many diverse RNA's are synthesized in the lampbrush stage oöcyte of *Xenopus*, as shown by the presence of different nucleotide sequences in the RNA population. This fact has been established by hybridizing lampbrush stage oöcyte RNA with an isolated *nonrepetitive* fraction of *Xenopus* DNA.

Much of our knowledge of oöcyte informational RNA derives from RNA-DNA hybridization experiments.[1-4] These hybridization experiments, however, like those so far reported for virtually every other animal cell system, have been confined to RNA species deriving from the internally redundant portion of the genome. It is impossible, in such experiments, to estimate the sequence diversity (i.e., the amount of genetic information) present in the hybridized RNA since the size of the sequence families active in transcription is always unknown. The recent studies of Britten and Kohne[5,6] have shown clearly that for the large genomes of higher organisms, much higher nucleic acid concentrations and much longer annealing periods than those generally in use would be required for the population of DNA sequences present only once in the genome to associate in a sequence-specific manner by base pairing; this process (in solution) follows the kinetics of a second-order collision reaction.[5-8] In the present paper we report experiments designed particularly to estimate the sequence diversity in the RNA transcribed during the lampbrush chromosome stage in *Xenopus*, by hybridizing this RNA with the nonrepetitive DNA fraction.

Methods.—Preparation of RNA from manually defollicled lampbrush-stage *Xenopus* oöcytes was carried out exactly as described by us previously.[1,2,9] The oöcytes were labeled *in vivo* by injection of 3 mC carrier-free phosphate into the dorsal lymph sac 6 days prior to their extraction, beginning 1 day after induction of ovulation with "Antuitrin S" (Parke-Davis). To label the RNA *in vitro*, we placed it with three to ten times its mass of H^3-dimethyl sulfate (110 mC/mM) under the conditions specified by Smith *et al.*,[10] and as in our previous report.[3] The RNA was then purified of adherent counts by passage through a Sephadex G-50 column and precipitated several times in ethanol.

DNA was extracted from *Xenopus* red blood cells after the hemoglobin had been removed from them by treatment with 0.6% w/v saponin in 0.15*M* NaCl–0.015 *M* Na citrate. DNA was extracted and purified with pronase and ribonuclease treatment according to the usual procedures.[11] It was then sheared by being forced through a needle valve under a pressure of 50,000 lb/in^2. (50 K sheared DNA). This was accomplished in the apparatus built for that purpose by Dr. Roy Britten, who very generously made this equipment available to us. The resulting fragments are about 400–500 nucleotides long.[5] When necessary, DNA fractions were concentrated by lyophilization, after dialysis against running distilled water.

Hydroxyapatite chromatography was carried out in a water-jacketed column at 60°C, as described by Britten and Kohne,[5] except that 0.5 *M* phosphate buffer was used for elution of double-stranded material. The buffer (PB) consists of equimolar quantities of

NaH_2PO_4 and Na_2HPO_4.[6] Hydroxyapatite (HAP) was pretreated by boiling in 0.12 *M* PB for 30 min. All the data reported here were obtained with one batch of HAP, Biorad lot no. 587500, except for Figure 1, in which two other batches were used as well. HAP columns were adjusted to 1 cc packed wet HAP/2 ODU DNA.

Annealing procedures at high Cots (see following paragraph for definition of this term) were carried out in sealed glass capillary tubes, immersed, after an initial treatment at 100° for 10 min, in a 60°C water bath. All salt concentrations were checked with a refractometer. Concentrated RNA-DNA mixtures were prepared by dissolving the desired amount of ethanol-precipitated RNA in a small volume (30–70 λ) of concentrated DNA dissolved in 0.24 *M* PB.

Character of the Xenopus Genome.—A hydroxyapatite reassociation curve for 50 K sheared *Xenopus* DNA is shown in Figure 1. To obtain this curve, various DNA solutions were denatured at 100°C and allowed to reassociate at 60°C. The ordinate represents the fraction of the DNA remaining bound to the HAP column, i.e., behaving in the column as a double-stranded structure under the conditions specified in *Methods.* The abscissa represents the controlling parameter of the reassociation reaction, viz., the concentration of 50 K sheared DNA (in moles of nucleotide/liter) times the time of annealing at 60°C in 0.12 *M* PB (in seconds). This parameter is represented by the term Cot, coined by Britten and Kohne, who originated this method of studying DNA reassociation.[5,6] Figure 1 indicates that a fraction renaturing more rapidly than we could measure (i.e., $<$ Cot 10^{-3}) exists in the *Xenopus* genome, amounting to at least 10 per cent of the DNA. Reassociation then occurs progressively over a broad range of Cots (five decades). At a Cot of 2–3 $\times 10^2$ reassociation of the slowest reacting fraction, amounting to about 50 per cent of the starting material, begins to dominate the over-all reaction. The

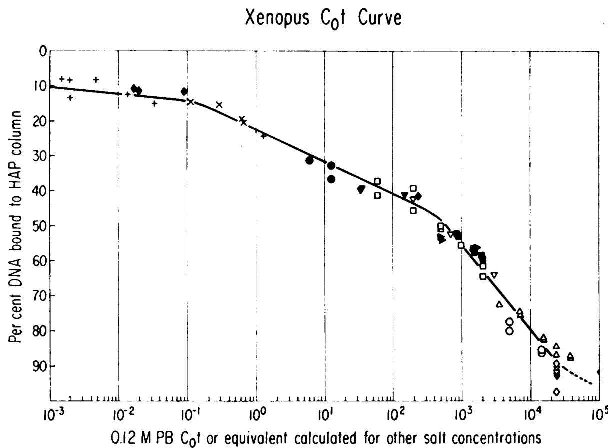

FIG. 1.—*Xenopus* Cot curve. DNA concentrations incubated, with PB molarity: (x) 9.3 μg/ml, 0.12 *M* (◆) 16.6 mg/ml, 0.12 *M* (+) 20 μg/ml, 0.12 *M* (●) 830 μg/ml, 0.14 *M* (▶) 1842 μg/ml, 0.24 *M* (▽) 2159 μg/ml, 0.29 *M* (■) 2611 μg/ml, 0.29 *M* (◇) 3029 μg/ml, 0.17 *M* (▼) 3133 μg/ml, 0.12 *M* (◇) 5626 μg/ml, 0.24 *M* (O) 5664 μg/ml, 0.24 *M* (△) 8133 μg/ml, 0.32 *M* (□) 8860 μg/ml, 0.12 *M*. An empirical relationship[14] was used to calculate "equivalent Cots," to correct for the effects of salt concentrations above 0.12 *M* PB on the reassociation rate.

DNA fraction remaining single-stranded at Cot 300, isolated on HAP and reannealed to Cot 300, contains very little rapidly reassociating material; thus, over 82 per cent of this fraction elutes as single-stranded material after the second annealing cycle.

The midpoint of the more rapidly descending segment of the curve beyond Cot 300 is in the range Cot 2–3 $\times 10^3$ (ref. 12). This value is similar to that established for the reassociation of the unique sequences of bovine DNA under similar conditions.[5, 6] Since the bovine genome is about the same size as that of *Xenopus*, 3pg (1 C),[11, 13] and reassociation rate for any nonrepeated DNA is inversely proportional to genome size,[5, 6] we conclude that the slowly reassociating fraction of the *Xenopus* genome consists mainly of nonrepeated sequences. Furthermore, in a "calibration" experiment[5] in which labeled *E. coli* DNA was added to the *Xenopus* annealing mixture, and thus exposed to the same viscosity, salt, and other environmental conditions, half reassociation for the slow *Xenopus* fraction was estimated[12] to occur at a Cot 740 times greater than for the *E. coli* DNA: the ratio of genome sizes is about 700.[5, 11] About 50 per cent of the *Xenopus* genome thus appears to consist of unique sequences, i.e., sequences different enough to prevent the formation of duplexes stable under the conditions used to obtain Figure 1. The 40 per cent of the DNA reassociating at Cot $<10^2$ is evidently composed of internally redundant sequence families, apparently ranging in size from a few copies to perhaps 10^3–10^4 copies. We were not able to study the character of the most rapidly renaturing 10 per cent fraction.

Melting of nonrepetitive DNA duplexes: Nonrepetitive DNA was separated on HAP as the fraction remaining single-stranded at Cot 2500, and then concentrated. Figure 2*a* shows the optical melting profile of the duplex structures isolated by salt elution from a HAP column after incubation of this DNA to an equivalent Cot of 9000. A difference of about 4°C separates the steeply rising portion of this melting curve from that of nonsheared native DNA melted simultaneously at the same salt concentration. The curve for the reassociated nonrepetitive DNA also contains a "foot" that is absent from the native DNA melt. Both of these features are expected as a result of the 50 K shearing to which the fractionated DNA has been subjected, according to data published by Britten and Kohne.[5] These authors show that the 50 K sheared DNA of *E. coli*, which lacks repeated sequences, melts 3–4°C below nonsheared *E. coli* DNA and also displays the slowly rising absorption component, or "foot," seen in Figure 2*a*. This figure also shows that the total hyperchromicity of the two DNA's is the same—about 138 per cent of the starting OD_{260} values. Figure 2*b* compares the curve of Figure 2*a* to that of the repetitive fraction of *Xenopus* DNA, obtained by annealing to Cot 88 and harvesting the reassociated structures. The internally redundant fraction forms double-stranded structures which contain more base-pair mismatching and which include more regions not engaged in base pairing, since the total hyperchromicity is lower. These effects are to be attributed to imperfect repetition within the repetitive sequence families, as discussed extensively by Britten and Kohne.[5, 6] A HAP melting curve for nonrepetitive *Xenopus* DNA is presented in Figure 2*c* (see figure legend for details). A T_m of about 82.5°C can be estimated from this curve (see also Fig. 4*b*).

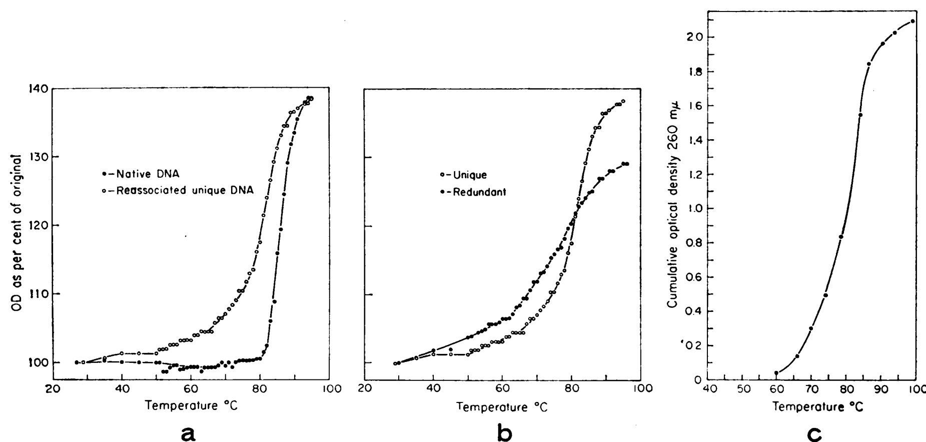

FIG. 2.—Melting profiles of 50 K sheared nonrepetitive DNA duplexes allowed to reassociate to an equivalent Cot of 9000. Optical melts (*a* and *b*) carried out in 0.12 *M* PB in the heated cell block of a Zeiss spectrophotometer. (*a*) Comparison of the melting curve of native unsheared *Xenopus* DNA (T_m = 86°C), and (*b*) that of a reassociated, internally redundant *Xenopus* DNA fraction (Cot 88.5). (*c*) HAP column melt of nonrepetitive DNA fraction in 0.12 *M* PB: temperature was raised and the column equilibrated in 4° steps, monitored by an inserted thermocouple. Six-ml fractions (six to seven times the column volume) were collected and read in 4 cm-cuvettes.

We conclude from Figure 2 that the DNA fraction remaining single-stranded at Cot 2500 reassociates to form well-matched duplexes indistinguishable, from those expected if the sequences present in the annealing mixture are unique.

Hybridization of newly synthesized P^{32}-lampbrush stage oöcyte RNA with nonrepetitive Xenopus DNA: In the following experiments, the RNA inputs are one of five times the amount of DNA used, and saturation of the nonrepetitive DNA fraction is not attained. Concentrations were adjusted to provide DNA Cots of over 21,000, and the mixtures (30–70 λ), sealed in capillary tubes, were incubated for six days at 60°C. The RNA itself withstands such incubations, as shown by the essentially unchanged Sephadex G-200 profile of a six-day incubated preparation (Fig. 3*a*). The incubation mixtures were diluted and treated with RNase A, with the results illustrated in Figure 3*b*. The effect of this treatment is to remove most of the highest-molecular-weight RNA components from the exclusion peak, which contains the DNA. Figure 4*a* shows the separation of a G-200 exclusion peak similar to that in Figure 3*b* into double- and single-stranded fractions. Most of the DNA can be observed to elute in the second peak (as expected after reassociation to Cot 21,000), and associated with this DNA are P^{32} counts representing hybridized RNA. A certain fraction of the RNA is also eluted with the single-stranded DNA. It is unknown whether this represents free RNA fragments which have resisted RNase treatment[14] or RNA fragments associated weakly with the DNA and carried with it in the Sephadex column. In any case it is clear that the HAP column provides the more stringent criterion in that less RNA behaves as part of an RNA-nonrepetitive DNA duplex in the HAP procedure than would be estimated according to the criterion of RNase resistance alone.

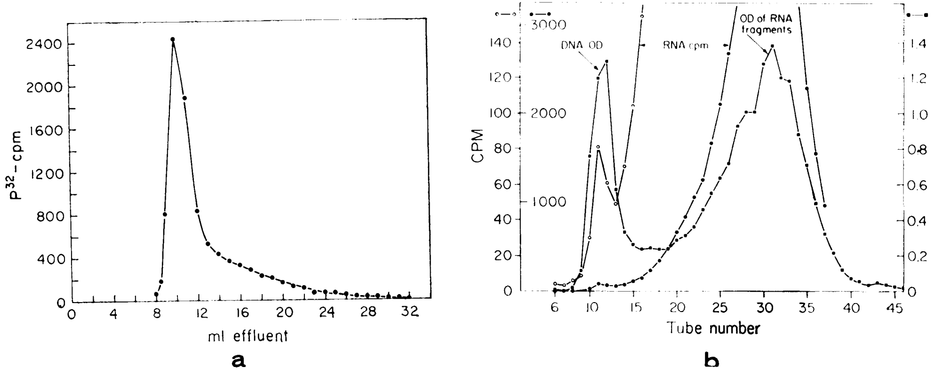

FIG. 3—(*a*) P^{32}-oöcyte RNA eluted with 0.12 *M* PB from a 35 × 1-cm Sephadex G-200 column, after 6-day incubation in 0.24 *M* PB at 60°C. The peak occurs at the exclusion volume of the column. (*b*) Elution from same G-200 column of mixture incubated at Cot 21,000 and containing 636 μg P^{32}-RNA, sp. Act. 752 CPM/μg, and 338 μg DNA, after ribonuclease treatment (20 μg/ml, 2 hr at 25°C). The mixture had been incubated to Cot 21,000, and then diluted in 0.24 *M* PB to 0.70 ml. 0.2 ml of 0.8 ml fractions were counted. Two CPM scales are shown in order to display the small fraction of the RNA which is associated with the DNA in the exclusion peak.

Figure 4*b* describes the HAP melting profile of the hybrids retained by the column. There exists a newly synthesized oöcyte RNA fraction bound to the unique DNA sequences in such a way that the RNA separates from the DNA at temperatures only 2°–4°C lower than those at which the DNA duplexes unwind. Experiments of Kohne with *E. coli* ribosomal RNA-DNA hybrids[15] and other empirical hybridization data obtained with nonrepetitive DNA's[16, 17] indicate that at least a 3°–4°C difference in T_m is to be expected in comparing the thermal stability of RNA-DNA hybrids with that of DNA-DNA duplexes. It can be concluded that Figure 4 demonstrates the formation of bona fide hybrids between newly synthesized lampbrush-stage oöcyte RNA and nonrepetitive DNA.

Estimation of the amount of RNA in the hybrids: In order to interpret hybridized RNA counts in terms of quantity, the RNA was labeled *in vitro* with H^3-dimethyl sulfate.[3, 10] All species of RNA appear to be labeled impartially by this procedure, except for minor variations due to differences in guanine content. By labeling the same RNA preparation *both* metabolically (with P^{32}, by *in vivo* injection) and chemically (by methylation) additional information is available: the ratio of H^3/P^{32} in the hybridized RNA relative to that in the totally conserved ribosomal RNA of the starting preparation[9, 18] can be used to provide an estimate of the extent to which the RNA species engaged in hybrid formation are undergoing rapid turnover in the lamp brush stage oöcyte.

Figure 5*a* shows a separation of nonrepetitive DNA-lampbrush stage oöcyte RNA hybrids similar to that in Figure 4*a*, except that the RNA had been doubly labeled. The double-stranded fraction of a duplicate preparation was melted from the HAP column, yielding a double melting curve identical in appearance to that in Figure 4*b*. The P^{32} and H^3 melting curves (not shown) are perfectly superimposable. Thus, transfer of H^3 counts to DNA does not occur; if methylation at the level attained here has damaged the ability of the RNA to

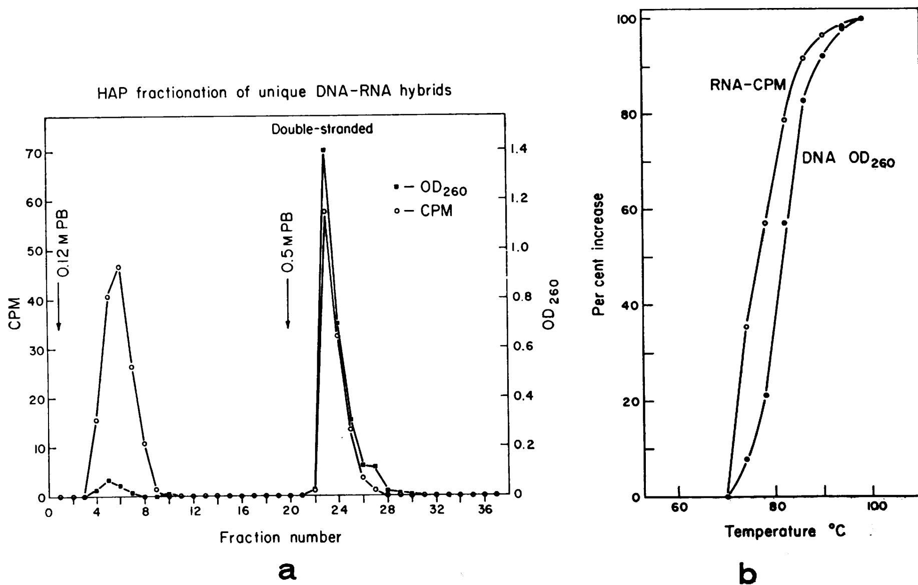

FIG. 4.—(*a*) HAP fractionation of nonrepetitive DNA-P^{32}-RNA hybrids (tubes 10–12, Fig. 3*b*). (*b*) Integral plot of HAP melt of P^{32}-oöcyte RNA-nonrepetitive DNA hybrids. The procedure was that described in the legend to Fig. 2*c*, except that, after the optical density was assayed, carrier yeast RNA and bovine γ-globulin were added and the radioactive RNA was coprecipitated with cold TCA. The precipitate was collected on HAWP Millipore filters under vacuum for assay of radioactivity[2] in the scintillation counter.

engage in hybridization, the effect is so small as to be imperceptible. The H^3 melting curve is plotted incrementally rather than integrally in Figure 5*b*. Only above 78°C is a constant ratio of RNA to DNA rendered single-stranded with each temperature increment. This, then, provides our most stringent criterion for extent of base pairing in the hybrids. Some of the RNA released at lower temperatures may have contained loops, nicks, free ends, or paired stretches of less than a certain length (the DNA can be observed to have melted more or less symmetrically). As the insert in Figure 5*b* shows, the H^3/P^{32} ratio remains approximately constant across the whole temperature range. This observation has been repeated several times and the ratio H^3/P^{32} is always close to, or very slightly higher than, that of the ribosomal RNA in the starting preparation. At least a considerable majority of the diverse RNA species in synthesis in the lampbrush-stage oöcyte thus appear to be retained, rather than turning over rapidly, for the duration of our seven-day labeling period. We have previously shown that in *Xenopus* most of the informational RNA transcribed from the repetitive fraction of the genome at the lampbrush stage is retained throughout oögenesis[1, 2] and in fact persists well into embryogenesis.[2]

In the experiment of Figure 5*b*, a conservative estimate of the amount of RNA hybridized with the 80 μg of reassociated nonrepetitive DNA present can be made by taking into account the total RNA released from the double-stranded state above 70°, minus the "excess" counts between 70° and 78°. The latter were

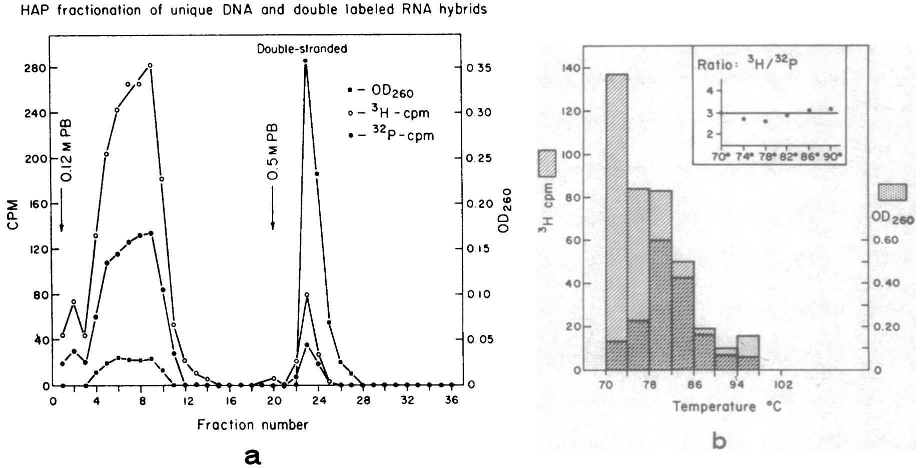

FIG. 5.—(*a*) HAP fractionation of nonrepetitive DNA and doubly labeled RNA hybrids. The RNA/DNA ratio was 2.4. No RNase treatment was applied in this experiment since heavy metal ions in the DNA had resulted in partial though not complete lysis of most of the free RNA, resulting in an elution pattern like that of Fig. 3*a*. (As pointed out in the text, RNase resistance is not a stringent criterion for hybridization with nonrepetitive DNA, compared to HAP column fractionation.) The partially degraded nonhybridized RNA is nearly totally eluted at 60°C in 0.12 *M* PB. (*b*) Incremental plot of melting curve obtained with a duplicate preparation, using the procedure described in the legend to Fig. 4*b*.

estimated on the basis of the ratio of RNA to DNA in the 78°–82° fraction. Calculated directly from the H^3-specific activity, this quantity is 0.11 μg RNA.

Discussion.—The analysis of nucleic acid reassociation kinetics provided by Britten and Kohne[6] has greatly facilitated quantitative handling of RNA-DNA hybridization procedures. In the experiments reported here, we have utilized methods and criteria mainly from these authors to isolate the nonrepetitive fraction of the *Xenopus* genome and to study the sequence diversity of the RNA synthesized in lampbrush-stage oöcytes. Despite the unusual technical problems encountered in working at very high Cots, the criteria for RNA hybridization with nonrepetitive DNA are more clear-cut than for RNA hybridization with internally redundant DNA fractions. In particular we have reference to the relatively stringent criterion available in the application of HAP melting curves to define the hybrids: It has been estimated that a mismatch frequency of less than 1 per cent would lower the T_m of a DNA duplex by 1°C.[19] The extent of mismatching in RNA hybridized with internally redundant DNA is likely to be at least several times higher than 1 per cent with procedures now in use. About half the *Xenopus* genome is internally redundant (Fig. 1), and the *maximum* amount of the repetitive DNA which could be functional in transcription in lampbrush-stage oöcytes is about 6 per cent according to saturation experiments[1, 2] performed in liquid systems at low Cots.[20]

Saturation of the nonrepetitive sequences with oöcyte RNA has not been attained in the present study. It is already clear, however, that the diversity of information transcribed during the lampbrush stage is very high. Consider only the extremely conservative estimate that about 0.14 per cent by weight of the

nonrepetitive DNA is hybridized in the experiment of Figure 5b. Since the DNA sequences involved are unique, we may translate this mass of DNA directly into numbers of different units of genetic information (genes), assuming some number of nucleotide pairs as equal to one such informational unit. A convenient, though not necessarily typical, example is the gene for the 146 amino acid hemoglobin β-chain: 0.28 per cent of the nonrepetitive DNA (double-stranded value) active is equivalent to about 10^4 such genes. This is clearly an underestimate, since it is based on a subsaturation value. Nonetheless, the presence of enormous *informational diversity* in the RNA of the lampbrush-stage oöcyte is already apparent.

We are much indebted to Professor Alfred E. Mirsky of this University and to Dr. Roy J. Britten of the Carnegie Institution of Washington for critical discussion and advice throughout this project. We are grateful to Dr. Britten for permitting us the use of his high-pressure pump for shearing DNA as well. In addition, it gives us great pleasure to acknowledge the intelligent and conscientious technical assistance of Miss Maggie Chamberlin.

* This investigation was supported by American Cancer Society grant E334C and by National Institute of Child Health and Development grant HD-02412-03.

[1] Davidson, E. H., M. Crippa, F. R. Kramer, and A. E. Mirsky, these PROCEEDINGS, **56**, 856 (1966).

[2] Crippa, M., E. H. Davidson, and A. E. Mirsky, these PROCEEDINGS, **57**, 885 (1967).

[3] Davidson, E. H., M. Crippa, and A. E. Mirsky, these PROCEEDINGS, **60**, 152 (1968).

[4] Gross, P. R., L. I. Malkin, and M. Hubbard, *J. Mol. Biol.*, **13**, 463 (1965).

[5] Britten, R. J., and D. E. Kohne, in *Carnegie Institution of Washington Yearbook*, **65**, 78 (1966).

[6] Britten, R. J., and D. E. Kohne, *Science*, **lol**, 529 (1968).

[7] Wetmur, J. G., and N. Davidson, *J. Mol. Biol.*, **31**, 349 (1968).

[8] Thrower, K. J., and A. R. Peacocke, *Biochem. J.*, **109**, 543 (1968).

[9] Davidson, E. H., V. G. Allfrey, and A. E. Mirsky, these PROCEEDINGS, **52**, 501 (1964).

[10] Smith, K. D., J. L. Armstrong, and B. J. McCarthy, *Biochim. Biophys. Acta*, **142**, 323 (1967).

[11] Dawid, I. B., *J. Mol. Biol.*, **12**, 581 (1965).

[12] The position of this phase of the curve is likely to have been affected progressively by viscosity,[8] but on the log Cot abscissa such corrections are not so extensive as to alter the interpretation of the curve. A slight uncertainty also exists with respect to estimation of the midpoint because reassociation never goes to absolute (100%) completion, and a few per cent of the starting OD_{260} may represent damaged DNA or UV absorbing impurities.

[13] Mirsky, A. E., and H. Ris, *Nature*, **163**, 666 (1949).

[14] Britten, R. J., in *Carnegie Institution of Washington Yearbook*, **67**, 332 (1969).

[15] Kohne, D. E., *Biophys. J.*, **8**, 1104 (1968).

[16] Moore, R. L., and B. J. McCarthy, *J. Bacteriol.*, **94**, 1066 (1967).

[17] Bolton, E. T., and B. J. McCarthy, *J. Mol. Biol.*, **8**, 201 (1964).

[18] Brown, D. D., and E. Littna, *J. Mol. Biol.*, **8**, 669 (1964).

[19] Bolton, E. T., R. J. Britten, D. B. Cowie, R. B. Roberts, P. Szafranski, M. J. Waring, in *Carnegie Institution of Washington Yearbook*, **64**, 322 (1965).

[20] Since an RNA gene product of a sequence belonging to an internally redundant sequence family can bind to sequences within the family other than that specifically responsible for its synthesis, empirical RNA saturation of a redundant DNA fraction can set a maximum limit on the amount of DNA which could have been active in transcription, but not an estimate of the amount of DNA actually responsible for the synthesis of the RNA studied. Saturation levels in our experiments (equivalent Cots 20–60) were the same in 0.9 *M* Na^{+1} as in 1.8 *M* Na^{+2}. This finding is not in accord with one of Church and McCarthy,[21] whose studies were performed in filter systems in which hybridization of only the most highly redundant sequences appear to have been monitored.

[21] Church, R. B., and B. J. McCarthy, *Biochem. Genet.*, **2**, 55 (1968).