Isolation and Chromosomal Localization of Highly Repeated DNA Sequences in *Drosophila melanogaster*

M. BOTCHAN†*, R. KRAM*, C. W. SCHMID*, AND J. E. HEARST*†

*Department of Chemistry and †Group in Biophysics and Medical Physics, University of California, Berkeley, Calif. 94720

Communicated by Jerome Vinograd, February 24, 1971

ABSTRACT The nuclear DNA of *D. melanogaster* contains DNA sequences that are repeated between ten and a hundred times more often than the next class of redundant DNA. This DNA, as a renatured duplex isolated on the basis of its renaturation kinetics, has a buoyant density of 1.691 (g/ml). In its native state it bands within the unique nuclear DNA peak ($\rho = 1.701$). These sequences have been localized by "in situ" hybridization in the chromocenter of the chromosomes of the salivary gland. The properties of centromeres are discussed in terms of the occurrence of repeated sequences at this locus.

At present, our understanding of the organization of DNA sequences in the higher organism is very poor. The presence of multiple copies of certain sequences is one feature that has been established through the study of renaturation kinetics of denatured DNA (1, 2). Gene classes have been defined (1), and are taken as those sequences with similar concentrations per genome. All those sequences within a class that can renature with each other are called a family. Relationships within and between families remain obscure.

The satellite DNAs of rodents represent a discrete set of sequences, with the highest extent of repetition (3-5) that is found in their genomes. The other repeated sequences belong to an intermediate class, with a continuous distribution of moderately repeated families, until a second discontinuity in the extent of repetitions presumably reveals sequences that represent single-copy DNA. Though representing 10% of the total DNA, the mouse satellite has a complexity of 300 base pairs. The basic unit probably contains a sequence of from 8 to 13 base pairs, distributed in some manner between sequences that have been derived by mutation from this basic sequence (6). There is evidence that this class of highly-redundant DNA is not transcribed (4). Recently, the mouse satellite DNA has been located in the centromeric heterochromatin of the metaphase chromosomes (7, 8). This observation is consistent with the earlier demonstration that the mouse satellite DNA is associated with the less soluble fraction of metaphase chromosomes (9) and with interphase heterochromatin (10). These facts suggest a structural function for this DNA. They also suggest a possible involvement of this DNA in the roles classically defined for the centromere, namely, chromosome pairing and chromosome attachment to the spindle of the mitotic apparatus. If this DNA does serve these critical functions, then clearly there must be analogous material in other organisms. This investigation sought to localize the position on chromosomes of the fastest renaturing DNA sequences in a simpler organism.

Laird and McCarthy (11) have characterized in *D. melanogaster* a collection of repeated DNA sequences whose extent of repetition suggested to us that they were in the intermediate class of moderately repeated sequences. The occurrence of

such sequences has been found in all eukaryotes studied. In this paper, we present evidence that the genome of *D. melanogaster* can also be divided into three classes of DNA, and that the sequences of highest repetition are concentrated within the centromeric heterochromatin.

METHODS

DNA extraction

D. melanogaster DNA was prepared from postgastrula to prelarva embryos (12), with the addition of a 4-hr digestion with 0.5 mg pronase per ml (self-digested for 2 hr) after the RNase and α -amylase treatments. As discussed in the results, cytoplasmic DNA contaminants were found in crude nuclear preparations ($600 \times g$ pellet). Nuclear DNA with less than 0.3% cytoplasmic DNA contamination was obtained from nuclei purified by sedimentation through 2.2 M sucrose for 1 hr at $45,000 \times g$ at 4°C. Contamination with either bacterial or yeast nucleic acid was ruled out because ^3H -labeled whole *Escherichia coli* and yeast cells added to the starting material did not add any labeled DNA to the final extracts.

"In situ" hybridization

Complementary RNA was synthesized by *in vitro* transcription with RNA polymerase (13); [^3H]ATP and [^3H]UTP (20.5 Ci/mmol) were used as labels. After DNase treatment and phenol extraction, the RNA was further purified on a Sephadex G-50 column and by passage through a nitrocellu-

FIG. 1. *D. melanogaster* nuclear DNA that contains cytoplasmic contaminants. A buoyancy gradient based on $1/\beta_B = 9.35 \times 10^{-10} \text{ g sec}^2/\text{cm}^5$ was used to calculate (15) density differences; 99% *E. coli* [^{15}N]DNA was used as a marker of buoyant density 1.727 ± 0.0004 , relative to a primary standard, unlabeled *E. coli* DNA, assigned a buoyant density of 1.710. The cytoplasmic peaks contaminate this nuclear DNA sample by 3%.

FIG. 2. Fractionation of *D. melanogaster* nuclear DNA on hydroxylapatite after denaturation and renaturation for a *Cot* of 0.5. Samples (1 mg/ml) were saturated with N_2 , placed in an ice bath, and sonicated at five 1-min intervals. The DNA double-strand molecular weight, determined by velocity sedimentation, was 250,000. The DNA was heat-denatured for 5 min at 95°C and renatured at 60°C in 0.12 M phosphate buffer, then poured into a slurry of hydroxylapatite at 70°C. Elution was performed stepwise at 70°C with 0.03 M, 0.12 M, 0.16 M, and 0.3 M phosphate buffers (Na_2HPO_4 - NaH_2PO_4), pH 6.8.

lose filter. The calculated specific activity was 9.8×10^7 dpm/ μ g.

The giant polytene chromosomes of the salivary glands were obtained from third-instar larvae and processed for hybridization (14). Hybridization was performed at 60°C for 10 hr with 50 μ l of complementary RNA at a concentration of 0.018 μ g/ml. Nonradioactive *E. coli* RNA was added as a carrier, to a concentration of 50 μ g/ml.

RESULTS

CsCl banding of *D. melanogaster* DNA

The CsCl banding pattern of DNA extracted from a crude nuclear pellet (Fig. 1) shows, in addition to a predominant main band ($\rho = 1.701$), two distinct satellites at $\rho = 1.687$ and $\rho = 1.679$. Both satellites appear to be extranuclear components, as they disappeared when the nuclei were pelleted through 2.2 M sucrose. The main satellite ($\rho = 1.687$) is identical to the one described for *D. melanogaster* by Travaglini *et al.* (16), and identified by them as mitochondrial. The very light satellite is intriguing, since Fansler *et al.* (17) have also characterized a cytoplasmic "dA-dT" with a density of 1.669. Naturally occurring "poly(dA-dT)" has been reported only in some species of crab and in certain petite mutants of yeast. The molecular weight of the DNA used in these experiments was 25×10^6 . It is conceivable that for lower molecular weights other satellites might appear.

In order to locate the position of the different classes of repeated sequences in the density gradient, and to examine the complexity of the cytoplasmic satellite, the DNA was banded in CsCl; material with $\rho \leq 1.695$ was pooled and rebanded.

FIG. 3. Melting curves of various samples of *D. melanogaster* nuclear DNA. ● 0.12 M elution peak from hydroxylapatite fractionation; Δ 0.3 M elution peak; ○ sheared, unfractionated DNA; □ unsheared, unfractionated DNA. All samples were dialyzed at 4°C against 0.12 M phosphate buffer (pH 6.8) prior to measurements.

The first fractionation enriched the satellite and light portion of the main band. From the second cycle of centrifugation, three fractions were investigated: fraction I, $\rho \leq 1.685$; fraction II, $1.690 \leq \rho \leq 1.694$; fraction III, $\rho > 1.701$. The DNA was sheared and the kinetics of renaturation of each fraction was measured. At a double-strand molecular weight of 200,000 in 0.18 M Na^+ , the major component of fraction I had a renaturation rate-constant between 44 and 30, which corresponds to a complexity of 2.1 – 2.9×10^4 base pairs. This complexity is consistent, within the precision of the measurements, with this DNA being mitochondrial (16). Sequences with a complexity of 200–6000 base pairs constitute 35% of fraction II. Less than 5% of fraction II had a complexity similar to the major component of fraction I. In fraction III, 5% of the material exhibited a complexity of about 6000 base pairs. The remaining material in all fractions exhibited higher complexities, as was described by Laird and McCarthy (11).

The fastest-renaturing components are, therefore, in the light part of the main band. This DNA was isolated on the basis of its kinetics of reassociation, with the use of a hydroxylapatite column.

Isolation of a very rapidly renaturing nuclear DNA on hydroxylapatite

We emphasize that subsequent experiments were performed on nuclear DNA, isolated from highly purified nuclei, on DNA that contained less than 0.3% of the ($\rho = 1.687$) cytoplasmic component.

To prevent the formation of "networks" during renaturation, and to free the reiterated sequences from attachments to "unique DNA," the total nuclear DNA was sheared before denaturation and renaturation. The renaturation time was adjusted to prevent reassociation of the families, having *Cot* values of 1–10, that were observed by Laird and McCarthy (11). The renatured DNA was separated from single strands on hydroxylapatite (4) (Fig. 2). We reproducibly isolated 7–8% of the total nuclear DNA as double strands that were eluted with 0.3 M phosphate. When rechromatographed at 70°C on hydroxylapatite, the rapidly renaturing material eluted as a single homogeneous peak at a salt concentration of 0.20 M phosphate in a linear salt gradient. Evidence for a high degree of base pairing in this fraction was obtained from its

FIG. 4. A solution of 45 $\mu\text{g}/\text{ml}$ of the very rapidly renaturing DNA from *D. melanogaster* was incubated for 12 hr at 59°C. 0.1 ml of this solution was placed in 1 ml of CsCl solution and banded with *E. coli* marker [^{15}N]DNA.

thermal denaturation profile (Fig. 3). The isolated material has a lower melting temperature ($T_m = 79^\circ\text{C}$) than does the unfractionated, sheared nuclear DNA ($T_m = 85^\circ\text{C}$). This is expected from the lower G + C content observed for the rapidly renaturing material in the CsCl fractionation experiments. A striking feature of the reassociation is that, although the material isolated from the column is of low molecular weight and gives an extremely broad band in CsCl, when it is renatured for 12 hr at 45 $\mu\text{g}/\text{ml}$ at 59°C, a sharp band appears, on top of a broader band, at $\rho = 1.691$ (Fig. 4). This indicates that the small double strands have polymerized to higher molecular weights of a fairly unique composition. This may be expected, since renaturation after random shearing will have produced "sticky ends," by allowing the association of single strands that were broken from the larger molecule in different places in the sequence. Preliminary evidence from electron microscopy supports this conclusion. Since this material has not been isolated as a native duplex, its "native" T_m is unknown, but from its density of 1.691 after renaturation, a G + C content of 32% and a T_m of 82°C can be predicted. Despite the possibility of marked sequence dependence, the observed T_m of 79°C indicates a high degree (no more than a few per cent mismatching) of base pairing in the renatured duplex.

Kinetics of reassociation of the very rapidly renaturing material

The rate of reassociation was plotted according to Wetmur and Davidson (2). From the data of Fig. 5, the isolated rapidly renaturing DNA is composed of at least two families of sequences, 55% having a k_2 of 3400 [(Cot) $_{1/2}$] = 6×10^{-4}] and 45% having a k_2 of 110 [(Cot) $_{1/2}$] = 2×10^{-2}]. The rate constants apply to the components, purified from each other.† The corresponding complexities are 260 and 8100 nucleotide base pairs. For comparison, a small virus (e.g., SV40) has a complexity of 4000. Our rate constants remain constant over a concentration range of a factor of 5; therefore, we conclude that the renaturation process obeys second-order kinetics.

The customary interpretation (3–5) of this type of kinetic data is (a) to assume a single repeating unit equal to the complexity and (b) to calculate the number of repetitions of this unit that fits the size of this class. The complexity is calculated

† The definition of k_2 according to Wetmur and Davidson (2) and used here is: $1/(\text{fraction reassociated}) = k_2(0.5 \times \text{moles phosphate})t + 1$, which gives a k_2 of twice the value as defined by Britten and Kohne: $1/(\text{fraction reassociated}) = k_2[\text{moles phosphate}] + 1$. We give values for (Cot) $_{1/2}$ as defined by Britten and Kohne, in order to facilitate direct comparison with the Cot literature.

FIG. 5. The smooth curve through the data points represents the theoretical curve for two rate constants of 3400 and 110 representing, respectively, fractions of the total material $\phi_1 = 0.55$ and $\phi_2 = 0.45$. The rate constants k_2 have the units liter $\times$ sec $^{-1} \times$ (mole phosphate) $^{-1}$. The limiting slope at long times gives an approximate rate constant for the slow component according to the Wetmur and Davidson (2) equations. It can be shown that the limiting slope at zero time is equal to $2.04 \times 10^{-4} (k_1\phi_1^2 + k_2\phi_2^2)$ for two components or, in general, $2.04 \times 10^{-4} (\sum k_i\phi_i^2)$.

The parameters used in the analysis of the data were: $A_0(100^\circ\text{C}) = 0.157$, $A_\infty(60^\circ\text{C}) = 0.78 \times 0.157 = 0.123$, where the 0.78 was taken from the denaturation curve of this sample of very rapidly renaturing DNA. The zero-time intercept is 36, which equals $[A_0(60^\circ\text{C}) - A_\infty(60^\circ\text{C})]^{-1}$, as it must if there is not another still faster renaturing component in the sample. $A_0(60^\circ\text{C}) = 0.151$, assuming a loss of 4% in absorbance of denatured DNA due to base-stacking changes from 100°C to 60°C.

The reassociation kinetics were studied in a Cary 15 spectrophotometer using a water-jacketed quartz cell connected to two water baths in parallel. One water bath was kept at 95°C and the other at 60°C. The sample was denatured at 95°C and, at time zero, the water bath at 60°C was switched into the cell. Temperatures in the cell were measured with a copper-constantan thermocouple and showed that thermal equilibrium was reached in 30 sec.

by assuming an inverse proportionality between rate constant and complexity (viral or bacterial DNA is taken as a standard). According to this interpretation, there are about 17,000 copies of a sequence 260 base-pairs long and 450 copies of a sequence 8100 base-pairs long in the *Drosophila* genome of 10^8 base pairs. More precisely, the complexity is the sum of all base pairs in discrete sequences of a given class. Therefore, we could with equal justification say that a complexity of 260 might mean in this case two sequences, each of 130 base-pairs long, that are repeated 17,000 times. Clearly, there are a tremendous number of ways to divide this number and the application of this approach is somewhat ambiguous. Its value is in measuring the total sequence diversity of a collection of DNA molecules, and a complexity value of 78,400 base pairs is clearly low, for the 8×10^6 base pairs contained in this rapidly renaturing material.

Chromosomal localization of the very rapidly renaturing material

Recently developed techniques of *in situ* hybridization provide a cytological tool for the localization of specific fractions of the genome. This technique was applied to labeled RNA produced *in vitro* from this material and annealed to giant chromosomes of the salivary gland. Autoradiograms (Fig. 6) de-

FIG. 6. Autoradiographic localization of the rapidly renaturing sequences in the chromocenter of salivary polytene chromosomes: four chromosome sets, with arrows pointing to the chromocenter, are shown.

monstrate a striking localization of grains within the chromocenter, a heterochromatic mass at which all chromosomes are joined at their centromeres. The chromocenter also includes heterochromatic segments occurring on either side of the centromeres and, in males, the Y chromosome. When the slides are developed after 5 days, the grains appear only in a small region of the chromocenter; however, upon longer exposures, the area covered by the grains increases, but remains within the chromocenter. It was interesting to observe that, in the nuclei of nonpolytene, neighboring fat-body cells, there seem to be an equal number of grains as in the chromocenter of the giant chromosomes. § This is consistent with the histophotometric determinations of DNA content in these polytene chromosomes, as Rudkin (18) has found that the centric heterochromatin does not undergo the polytenic replication. The specificity of the autoradiographic localization of these sequences is confirmed by the following control. With the DNA eluted with 0.12 M phosphate from the hydroxylapatite as a primer for the synthesis of radioactive RNA, and in a parallel experiment, scattered grains were seen on the chromosomal arms, but not localized at the chromocenter. This control does not rule out the occurrence of moderately repeated DNA sequences in the chromocenter because of their lower concentration in the 0.12 M fraction, which is primarily unique DNA. The possibility that other copies of rapidly renaturing material exist in lower concentration at other locations of the chromosome cannot be ruled out, since a critical amount (estimated at 10^8 daltons) of material must be concentrated at a particular locus to be detected in a 10-day exposure. This point is particularly important since occasional grains were noticed on the chromosomal arms, though no obvious pattern of grains was obtained.

DISCUSSION

The original inference that repeated sequences are localized in the chromocentric region of *D. melanogaster* was drawn from cytological experiments (7, 19). RNA complementary to the total nuclear DNA was used under conditions where, pre-

sumably, only reassociation between reiterated sequences could occur. Rae (20) has already confirmed this point by actually isolating the total repeated sequences on hydroxylapatite. We report here that, in *D. melanogaster* as in the mouse, the kinetic analysis of DNA renaturation reveals sequences whose highest level of repetition allows them to renature as a discrete class of DNA within the genome. These sequences have been overlooked in previous kinetic characterization of the *D. melanogaster* genome. This specific fraction of DNA sequences is concentrated in a discrete region within the chromocenter, which suggests that the sequences may be localized, as in mouse, at the centromeres. When this fraction is sheared to a double-strand molecular weight of 250,000, the renatured buoyant density ($\rho = 1.691$) is distinct from that of the main band ($\rho = 1.701$); however, the isolation of rapidly renaturing material from high molecular weight DNA provides a different density pattern, with information regarding the length and organization of its sequences in the genome (manuscript in preparation). It seems probable to us that other organisms with no distinct satellites have a discrete class of the most repetitious DNA buried within the main-band DNA. For this class, both its discreteness (as measured by renaturation kinetics) and its distinct chromosomal localization would suggest that it may have a functional role different from that of the "intermediate" class.

Though speculations as to the function of the rapidly renaturing fraction remain tentative, it seems reasonable to believe that some or all of the properties of centromeres can be understood in terms of the presence of highly repeated sequences. In general, since the morphology of the chromosomes is part of the genetic heritage of the cell, the DNA sequences must contain the information for directing the formation of this structure. Chromosomal morphology may be simply determined by specific DNA associations, either through protein cross-links or through direct DNA-DNA interactions. It is difficult to escape the hypothesis that the high degree of sequence homology at the centromeres results in extensive interactions between these sequences. Fig. 7 represents a speculative view [based on an earlier suggestion (21) of the possible involvement of mouse satellite sequences in chromo-

§ Similar observations have been made in *D. virilis* (Gall, J. G., and R. Eckhardt, private communication).

A. Interphase configuration of centromeric heterochromatin DNA composed of tandem repeats of repetitive DNA (a) separated by "major" spacers (b).

B. Mitotic configuration of centromeric heterochromatin.

FIG. 7. The model should be taken in its abstract form. The length of the major "spacers" can be widely variable, as well as the number of repetitions in a tandem array. Minor "spacers" may also occur within the repetitions. Different distributions and values of minor and major spacers can account for different sizes and shapes of the metaphase centric heterochromatin. The interaction between the sequences can be mediated either through proteins specific for these sequences or directly through DNA-DNA interactions.

somal folding] of how these sequences might affect chromosomal folding. The suggestion that this region of the chromosome is capable of extensive interactions is further supported by numerous cytological observations (22, 23) of somatic centromeric association of nonhomologous chromosomes, and of chromosome polymorphism, which is caused by centric inversion, "fusion", and "fission".

Biologically, perhaps the most important aspect of the aggregative behavior of centromeres is manifested in the pairing of homologous chromosomes at meiosis. This specific process demands a certain amount of heterogeneity in the total centromeric DNA. The specificity could be provided by unique "spacer" sequences on different chromosomes. However, as we have emphasized in our interpretation of the renaturation kinetic data, the possibility that there are several families within the very rapidly renaturing fraction, each associated with a particular chromosome, cannot be ruled out. This interpretation of the kinetic data is valid even for the kinetically homogeneous mouse satellite.

The work of Hennig and Walker (24) suggests that any role that the rapidly renaturing centromeric DNA plays in the physiology of the cell must allow for the fact that this material seems to be in rapid change in evolutionary history. Assigning a housekeeping function to this DNA would be consistent with these results (24), since rearrangement of this material could lead to chromosome polymorphism, through centric fusion, fission, or inversion. The generation of new linkage groups would obviously have an effect upon the direction of evolution.

We thank Dr. M. Chamberlin for the gift of *E. coli* RNA polymerase, Dr. J. G. Gall for a demonstration of *in situ* hybridization techniques, Mr. John Tang for his energetic technical help, and Mr. B. P. Dorman for many helpful conversations. Dr. R. Kram is on leave from the Free University of Brussels on a Charles Pecher Fellowship. This work is supported in part by

USPH Grant no. GM 15661. Michael Botchan was supported by an NIH Predoctoral Biophysics training grant. Carl W. Schmid was supported by a Predoctoral Fellowship, NIH no. 1-F01-GM 46,314-01.

1. Britten, R. J., and D. E. Kohne, *Science*, **161**, 529 (1968).
2. Wetmur, J. G., and N. Davidson, *J. Mol. Biol.*, **31**, 349 (1968).
3. Waring, M., and R. J. Britten, *Science*, **154**, 791 (1966).
4. Flamm, W. G., P. M. B. Walker, and M. McCallum, *J. Mol. Biol.*, **40**, 423 (1969).
5. Flamm, W. G., P. M. B. Walker, and M. McCallum, *J. Mol. Biol.*, **42**, 441 (1969).
6. As quoted in Southern, E. M., *Nature*, **227**, 794 (1970).
7. Pardue, M. L., and J. G. Gall, *Science*, **168**, 1356 (1970).
8. Jones, K. W., *Nature*, **225**, 912 (1970).
9. Maio, J. J., and C. L. Schildkraut, *J. Mol. Biol.*, **40**, 203 (1969).
10. Yasmineh, W. G., and J. J. Yunis, *Exp. Cell. Res.*, **59**, 69 (1970).
11. Laird, C. D., and B. J. McCarthy, *Genetics*, **63**, 865 (1969).
12. Ritossa, F. M., and S. Spiegelman, *Proc. Nat. Acad. Sci. USA*, **53**, 737 (1965).
13. Burgess, R. R., *J. Biol. Chem.*, **244**, 6160 (1969).
14. Gall, J. G., and M. L. Pardue, *Methods in Enzymology XIIC* (Academic Press, Inc., New York, in press).
15. Schmid, C. W., and J. E. Hearst, *Biopolymers* (in press).
16. Travaglini, E. C., J. Petrovic, and J. Schultz, *J. Cell. Biol.*, **39**, 136A (1968).
17. Fansler, B. S., E. C. Travaglini, L. A. Loeb, and J. Schultz, *Biochem. Biophys. Res. Commun.*, **40**, 1266 (1970).
18. Rudkin, G. T., *Genetics*, **52**, 470 (1965).
19. Jones, K. W., and F. W. Robertson, *Chromosoma*, **31**, 331 (1970).
20. Rae, P. M. M., *Proc. Nat. Acad. Sci. USA*, **67**, 1018 (1970).
21. Walker, P. M. B., W. G. Flamm, and A. McLaren, *Handbook of Molecular Cytology*, ed. Lima-de-Faria, (North Holland, Amsterdam, 1969), Vol. 52.
22. White, M. J. D., *Annu. Rev. Genetics*, **3**, 75 (1969).
23. Hsu, T. C., and R. A. Mead, in *Comparative Mammalian Cytogenetics*, ed. Benirschke (Springer, 1969), p. 8.
24. Hennig, W., and P. M. B. Walker, *Nature*, **225**, 915 (1970).