The Organization of the Main Component DNA of a Crustacean Genome with a Paucity of Middle Repetitive Sequences

Christie A. Holland and Dorothy M. Skinner

The University of Tennessee-Oak Ridge Graduate School of Biomedical Sciences
and Biology Division[1], Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, U.S.A.;

[1] Operated by Union Carbide Corporation for the Energy Research and Development Administration

Abstract. The frequency classes and organization of the main component (mc) DNA of a crustacean, the land crab, *Gecarcinus lateralis*, have been characterized. The reassociation kinetics of 380 nucleotide long mcDNA fragments show that approximately 50% contain sequences repeated more than 800 times. Present in few, if any, copies are sequences repeated from 2 to 800 times. The remainder of the DNA reassociates as single copy sequences with a rate constant consistent with the organism's genome size. The reassociation kinetics of highly sheared DNA fragments of every true crab studied (Vaughn, 1975; Christie et al., 1976) are similar to each other and different from those of other invertebrate DNAs (Goldberg et al., 1975). Each of these genomes has a paucity of sequences repeated from 10 to 800 times and an abundance of highly repeated sequences. To determine if sequences repeated more than 800 times are interspersed with single copy sequences, we examined the arrangement of repetitive and non-repetitive sequences in mcDNA. The reassociation and melting properties of partially duplex mcDNA fragments of increasing lengths show that at least 75% of the DNA is organized in an interspersed pattern. In this pattern, single copy sequences with an average length of 800–900 nucleotides are interspersed with repetitive sequences. S1 nuclease digestion of reassociated 3100 nucleotide fragments indicates that ~44% of the mcDNA is repetitive and that one-third of the repetitive sequences (average length = 285 nucleotides) are interspersed with single copy sequences. We conclude that repetitive sequencies are interspersed with most of the single copy sequences in an interspersion pattern similar to that of *Xenopus* rahter than to that of another arthropod, *Drosophila*.

Introduction

Repetitive and single copy sequences are interspersed in a characteristic and highly ordered arrangement in the genomes of vertebrates (*Xenopus laevis*, Da-

vidson et al., 1973; *Homo sapiens,* Schmid and Deininger, 1975), invertebrates (a sea urchin, *Strongylocentrotus purpuratus,* Graham et al., 1974; a gastropod mollusc, *Aplysia californica,* Angerer et al., 1975), and plants (cotton, *Gossypium hirsutum,* Walbot and Dure, 1976; tobacco, *Nicotiana tabacum,* Zimmerman and Goldberg, 1977). In such genomes, more tthan half of the DNA is organized in a pattern of interspersed repetitive sequences of 300–500 base pairs and single copy sequences of 1000–2500 base pairs. Another smaller fraction of the DNA is organized in a long period interspersion pattern in which repetitive sequences are interspersed with single copy sequences greater than 3500 base pairs. The remainder of the sequences are organized as non-interspersed single copy sequences, in tandem repeats (satellites), or in sequences that form intramolecular duplexes.

The organization of sequences in the genomes of *Drosophila melanogaster* (Manning et al., 1975; Crain et al., 1976a), and the honeybee, *Apis mellifera* (Davidson et al., 1976) is significantly different. In these genomes, long repetitive sequences appear to be interspersed with single copy sequences >2200 nucleotides long. Although this pattern of genome organization has not been linked to functional or molecular differences in the organisms, there is speculation that the very small genome sizes of *Drosophila* (0.12 pg) and the honeybee (0.3–0.4 pg) dictate this unusual pattern (Davidson et al., 1976). This pattern does not occur throughout the Class *Insecta* since the larger genomes of the wild silk moth, *Antheraea pernyi* (0.8 pg; Davidson et al., 1976) and the housefly, *Musca domestica* (0.89 pg; Crain et al., 1976b) have a short period interspersion pattern similar to that of *Xenopus laevis.*

We have examined the organization of sequences of the DNA of another arthropod, the crustacean, *Gecarcinus lateralis.* The optimal conditions for reassociation studies on main component (mc) DNA, are very close to the mid-point of the melting temperature for the "poly(dA-dT)" satellite (20% of the total DNA; Skinner, 1967) which is consequently partially, but only partially, denatured. This introduces a variability that is difficult to control if *total* DNA is used. Consequently, for our studies on the sequence organization of *G. lateralis* DNA, both poly(dA-dT) and a G+C-rich satellite (3% of the total DNA; $\rho = 1.721$ g/cm^3) have been removed.

Materials and Methods

1. Purification, Centrifugation and Labeling of mcDNA. DNA was isolated from testes and regenerating limbs of the land crab, *Gecarcinus lateralis* (Holland and Skinner, 1976), or by a batchwise procedure. The latter involved adsorbing the DNA to hydroxyapatite in 0.20 M PB, (PB is phosphate buffer prepared with equimolar concentrations of monobasic and dibasic sodium phosphate), 8 M urea, 0.01 M EDTA, 0.25% SDS, 1 M NaClO$_4$ (Britten et al., 1974) and washing in 8 M urea and 0.24 M PB, followed by centrifuging ~5 min at low speeds (400–600 × g). The subsequent washes were in 0.01 M PB. The DNA was eluted in 0.5 M PB. One to 1.2 mg of total DNA was fractionated in 18 ml relaxed CsCl gradients (Beattie and Skinner, 1972) in a Ti60 rotor centrifuged for 60 h at 32,000 rpm, 25° C. Purified mcDNA had a unimodal transition on melting (see Fig. 3) with a T_m of 86.5° C in 0.12 M PB. The single stranded length of mcDNA prepared by this method was 6000 nucleotides when sized by boundary sedimentation in a Model E ultracentrifuge (Markham, 1967). Analytical Hg^{2+} or Ag$^+$-Cs$_2$SO$_4$ samples were prepared and centrifuged as described by Skinner and Beattie (1973).

Labeled DNA was isolated from regenerating limbs incubated in the presence of ^{3}H-thymidine for 16 h (Holland and Skinner, 1976). Since only small quantities of in vivo labeled DNA were obtained, in vitro labeled DNA was synthesized from a mcDNA primer-template by DNA polymerase I in polymerase diluent (Englund, 1971). The reaction mixture consisted of unsheared mcDNA (30 nmoles), four cold dNTPs (30 μM each, PL Biochemicals), ^{3}H-dTTP (7×10^{-7} μmoles), potassium phosphate, pH 7.3 (20 μmoles), MgCl$_2$ (2 μmoles), 1–2 units of DNA pol I (Boehringer, or the gift of S. Mitra or C. Richardson), or T5 DNA polymerase (the gift of S. Das). The reaction mixture, containing equal quantities of the four dNTPs was incubated at 37° C (Englund, 1971) or at 11° C with the conditions of Angerer et al. (1976). The latter conditions produced a labeled product ~200 nucleotides long while the former conditions produced a product ~600 nucleotides long and were routinely used. The reaction mixture was extracted twice with redistilled phenol equilibrated with 1×10^{-3} M borate and the labeled product was desalted on Sephadex G-100. With the conditions used 0.16 nmole of DNA was synthesized per h. The incorporation of ^{3}H-thymidine was linear for at least 10 h by which time 16% of the total radioactivity had been incorporated into an acid insoluble product.

2. Characteristics of E. coli DNA Polymerase I labeled DNA. When main component DNA, 6000 nucleotides long, was labeled in vitro, most of the labeled product was 300–1500 nucleotides long (Fig. 1a) The difference in the size of the primer-template and the labeled product (see also Kunkel et al., 1976; Angerer et al., 1976) could be due to degradation of the primer-template by a nuclease. This possibility was excluded by re-sizing several preparations of labeled mcDNA by boundary sedimentation rather than by monitoring radioactivity in an alkaline sucrose gradient. It was found that the bulk of the primer-template (A_{260nm}) remained at 6000 nucleotides. The size of the newly synthesized labeled DNA molecules could not be detected in the boundary sedimentation analyses since they did not comprise a significant fraction of the total absorbancy (0.02 A_{260nm}/4.0 A_{260nm}).

To rule out the possibility that the labeled molecules had been partially degraded by the endogenous $5' \rightarrow 3'$ exonuclease activity of Pol I (Klett et al., 1968), mcDNA was labeled under similar reaction conditions using T5 DNA polymerase which lacks $5' \rightarrow 3'$ exonuclease activity (Das and Fujimura, 1976). Again, the radioactivity sedimented as it had when the DNA was labeled with Pol I.

Although the reaction conditions did not favor their insertion, ribonucleotides can serve as substrates for a Pol I reaction (Berg et al., 1963). The insertion of a few rNTPs could lead to the formation of alkali-labile bonds. To examine the possibility that the dNTPs were contaminated with rNTPs (Bollum, 1962), the labeled product was displayed on a non-alkaline denaturing gradient (4M urea, 85% formamide, 0.01 M Tris; L. Angerer and N. Davidson, personal communication). The distribution of the ^{3}H-DNA was identical to that in Figure 1. Taken together, these data indicate that most of 3'-OH primers available for the initiation of the Pol I reaction are attached to fragments of 600 nucleotides or less.

3. Shearing and Sizing of DNA Fragments. Unlabeled DNA was sheared in a Virtis 60 homogenizer (Davidson et al., 1973) with one blade using the large cup. The shearing conditions used and the fragment lengths obtained are listed in Table 1. Since Davidson et al. (1973) used different conditions to obtain the fragment lengths listed in Table 1, we conclude that the length of the sheared products depends on the length of the starting DNA. After shearing, all solutions were brought to 0.12 M PB. The DNA fragments were collected on hydroxyapatite and eluted with 0.5 M PB. The DNA was dialyzed into 10^{-4}M NaCl, 10^{-4}M EDTA and stored over chloroform at 10° C in siliconized sterile vials.

The lengths of the DNA fragments were determined in one of two ways. Unlabeled fragments were sedimented in a Model E analytical ultracentrifuge in 0.9 M NaCl-0.1 M NaOH or in 1.0 M NaCl (Markham, 1967). Labeled fragments were sedimented in alkaline sucrose (5–25%, 0.1 M or 1.0 M NaOH) isokinetic gradients (Noll, 1967). As a marker, ^{14}C-*E. coli* labeled in vivo, sheared, and previously sized by boundary sedimentation in a Model E analytical ultracentrifuge was either present in each gradient or centrifuged in gradients made in parallel. Sucrose gradients were centrifuged at 45 K rpm for 8–16 h at 20° C in a SW 50.1 rotor. The weight-average molecular weights were determined using the equations of Studier (1965). Fragment sizes reported are an average of at least two determinations.

Table 1. Conditions used for shearing mcDNA

DNA fragment length[a] (N.T.)	Shearing solution	External milieu	Speed[b] (K RPMs × 25 min)
380	54% glycerol (0.01 M PB)	Dry ice-Ethanol	22–25
780	66% glycerol (0.01 M PB)	Ice water	17–18
1680	(0.01 M PB)	Ice water	35
2200	(0.01 M PB)	Ice water	17

[a] After shearing the DNA fragments were sized by boundary sedimentation in a Model E ultra-centrifuge as outlined in Materials and Methods
[b] The speed of the Virtis was determined by monitoring the rotation of the shaft with a Strobotac

4. Reassociation of DNA Fragments. For reassociation experiments, labeled and unlabeled 380 nucleotide long fragments were brought to 0.12 M PB or 0.4 M PB. The unlabeled molecules were present in at least a 100-fold excess. The mixture was sealed in capillaries or glass ampules, heat-denatured at 110° C for 3 min and incubated at criterion (Britten et al., 1974) for the length of time necessary to produce the desired equivalent Cot (see Britten et al., 1974). The samples were quick frozen in dry ice-ethanol and stored at −20° C until analysis. They were rapidly thawed in a 60° C water bath, diluted into 2 mls of 0.12 M PB containing sufficient quantities of *E. coli* 300 nucleotide fragments to maintain a constant DNA to hydroxyapatite ratio of 50 µg/0.4 gm and loaded on a hydroxyapatite column (Bio Rad, DNA grade, lot number 14115). The column was washed successively with 0.12 M PB and 0.4 M PB. The elutes were counted in mixtures of 0.06 M PB (7 mls) and "Instagel" (Packard; 10 mls). Only those data points in which 95%–105% of the radioactivity was recovered are included.

5. Thermal Dissociation. DNAs were melted in 0.12 M PB in a Gilford spectrophotometer. Hyper-chromicity was calculated according to $H = (A_{260\,nm\,at\,100°\,C} = A_{260\,nm\,at\,55°\,C})/A_{260\,nm\,at\,100°\,C}$. Transitions were calculated by determining the temperature at which 17% of the final hyperchromicity was achieved and temperature from the temperature at which 83% of the final hyperchromicity was achieved.

6. S1 Nuclease Digestion. A commercial preparation of single-strand-specific nuclease (S1 nuclease, Sigma Chemicals; Vogt, 1973) was further purified by passage over Sephadex G-150. It was then characterized (Britten et al., 1976) to determine the quantity of enzyme necessary to digest all single stranded DNA regions. Control experiments showed that with the conditions used the enzyme did not digest double stranded DNA (i.e., the single stranded size, measured in an alkaline sucrose gradient, of a double stranded DNA was not reduced after digestion).

DNA fragments to be digested with S1 nuclease were denatured and reassociated in 0.2 M NaCl, 0.01 M Pipes buffer (pH 6.7) at 64° C. They were then adjusted to 0.15 M NaCl, 0.1 mM ZnSO₄, 0.025 M sodium acetate, 0.005 M Pipes (pH 4.4), and 0.025 M β-mercaptoethanol. S1 nuclease was added and the reaction mixture was incubated for 45 min at 37° C to achieve complete digestion of all single stranded ragions (0.33 mls of enzyme/mg of DNA). The reaction was terminated by chilling the sample and bringing it to 0.12 M PB. The duplex fragments were collected on hydroxyapatite and the A_{260nm} of the fraction eluting in 0.12 M PB was corrected for the increase in absorbancy of mononucleotides by dividing it by 1.33 (Britten et al., 1976).

7. Agarose-50 Chromatography. The sizes of the undigested duplex fragments were determined on a 2.2 × 36 cm agarose-50 column (Bio-Gel, 100–200 mesh) poured onto a support of 5 and 6 mm glass beads (Sachs and Painter, 1972). Long ¹⁴C-labeled T7 DNA, unlabeled duplex 285 nucleotide-long fragments and ³H-uridine were used to calibrate the column as excluded, internal, and included markers respectively.
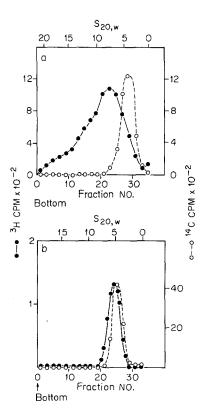
Results

1. The Reassociation Kinetics of 380 Nucleotide Long mcDNA Fragments

The reassociation kinetics of 380 nucleotide long fragments have been examined to determine the range of repetition frequencies present. Pol I-labeled 380 nucleotide long fragments were selected from alkaline preparative gradients (Fig. 1a). Their homogeneous size was demonstrated by sedimenting them in a second gradient (Fig. 1b). When reassociated to C_0t 10^{-5}, 40 to 50% of the Pol 1 labeled 380 nucleotide long fragments bound to hydroxyapatite in 0.12 M phosphate buffer (PB).

A large fraction of the DNA labeled after extensive synthesis by Pol I reassociates with zero-order kinetics (Masamune and Richardson, 1971). These sequences are thought to be synthesized on a displaced strand of the primer template (branch migration) and to have intramolecular complementarity. Since branch migration occurs at random, the DNA sequences that do not rapidly reassociate are thought to be synthesized directly from the primer-template and to represent all the sequences of the primer-template.

To select the sequences synthesized directly from the primer-template, (as opposed to those synthesized during branch migration) the Pol I labeled molecules were reassociated to C_0t 10^{-5} in 0.12 M PB and only those fragments

Fig. 1a and b. Sedimentation of Pol I labeled mcDNA. The single stranded length of unsheared Pol I labeled mcDNA (●) (see Materials and Methods) was determined by sedimentation in isokinetic alkaline sucrose gradients (a). The homogeneous size of Pol I labeled 380 nucleotide long fragments (●) selected from an alkaline preparative gradient (as in a) was demonstrated by sedimenting them in a second gradient (b). In vivo labeled [14]C-*E. coli* DNA (--o--) previously sized by boundary sedimentation in a Model E (3.8S) was included in both gradients and its position used to determine the molecular weight of the ^{3}H-mcDNA (Studier, 1965)
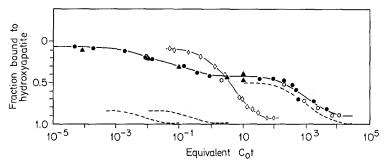
Fig. 2. Reassociation kinetics of 380 nucleotide fragments of mcDNA. *G. lateralis* mcDNA of 380 N.T. fragment length was reassociated as described in Materials and Methods. The fraction of the fragments containing duplex regions was measured at the indicated equivalent C_0ts by hydroxyapatite chromatography. The points are values determined by optical density (○) or by radioactivity in Pol I labeled (●) or in vivo labeled (▲) 380 N.T. fragments. The lines drawn are a computer fit by least squares analysis yielding the second order coponents listed in Table 2. As a standard, in vivo labeled ^{14}C-*E. coli* DNA was sheared to 300 N.T. and reassociated under identical conditions (◇)

that did not bind to hydroxyapatite were used in subsequent reassociation experiments. The reassociation kinetics of Pol I labeled fragments prepared in this way are similar to those of unlabeled and in vivo labeled mcDNA fragments all of which fall on the same curve (see Fig. 2).

Figure 2 shows the percentage of labeled or unlabeled 380 nucleotide long fragments that bind to hydroxyapatite after reassociation to various C_0ts. As a control, 300 nucleotide long fragments of in vivo labeled *E. coli* DNA were reassociated under identical conditions; they had a $C_0t_{1/2} = 4.2$ (Fig. 2, Table 2). The curves shown for the crab mcDNA fragments are a least squares analysis, fit by computer, for three second order components (computer program modified from Britten et al., 1974).

The slowest component reassociates at approximately the rate predicted for sequences that are present only once in the *G. lateralis* haploid genome (1 pg; Musich and Skinner, 1972), namely $0.00118 \ M^{-1} \ sec^{-1}$. At least half of the 380 nucleotide long fragments reassociated with this rate as a homogeneous frequency class.

The remaining DNA reassociates between $C_0t \ 10^{-5}$ and $C_0t \ 1$. The best fit by a computer program of Britten et al. (1974) divides these data into two classes of sequences: one highly repetitive class comprises 18% of the genome and reassociates with a rate of $179 \ M^{-1} \ sec^{-1}$; the other less highly repetitive class comprises 17% of the genome and reassociates with a rate constant of $5.59 \ M^{-1} \ sec^{-1}$. At the opposite extreme, there may be a range of repetitive sequences that reassociate with rates from 4700 to 152,000 times the rate of the non-repetitive sequences.

These data, summarized in Table 2, indicate that mcDNA is composed of at least three classes of sequences. The estimate that 50% of the 380 nucleotide long fragments contain only single copy sequences is slightly lower than the actual percentage of single copy sequences in the genome since these sequences are interspersed with repeated sequences. The majority of the remaining se-

Table 2. Repetition frequencies of *G. lateralis* main component DNA

	Fast[d]	Intermediate[d]	Single copy
Fraction of mcDNA[a]	0.18	0.17	0.50
Average rate constant[a] ($M^{-1} sec^{-1}$)	179.0	5.50	0.00118
Observed $C_0t_{1/2}$	5.59×10^{-3}	1.78×10^{-1}	8.47×10^{2}
$C_0t_{1/2}$ pure[b]	1.00×10^{-3}	3.00×10^{-2}	4.23×10^{2}
Kinetic complexity[c] (daltons)	6.66×10^{5}	2.00×10^{7}	2.82×10^{11}
Average repetition frequency	152,000	4,700	1

[a] The size of each component and the average rate constant are based on the least squares fit to the reassociation of 380 nucleotide long (N.T.) fragments shown in Figure 2
[b] $C_0t_{1/2}$ pure $= C_0t_{1/2}$ observed X fraction of the mcDNA
[c] Calculated by comparison of the $C_0t_{1/2}$ pure of each component of mcDNA to that measured for *E. coli* DNA ($C_0t_{1/2} = 4.2$) using 2.8×10^{9} daltons (Cairns, 1963) as the genome size of *E. coli*
[d] Since single copy sequences are interspersed with repetitive sequences the quantities of these components are a slight overestimate

quences in the mcDNA are repeated more than 800 times. No measurable reassociation is detected between C_0t 1 and C_0t 100 which results in an apparent plateau in the C_0t curve. The scatter inherent in the measurement of DNA reassociation kinetics would permit the reassociation of as much as 5 to 10% of the DNA to escape detection. Nonetheless, it is clear that sequences repeated from 10 to 800 times are rare in the mcDNA of *G. lateralis*. In many organisms, these are the repetitive frequencies of the rRNA genes (*S. cerevisiae, D. melanogaster,* chicken, hamster, rat, *X. laevis, Homo sapiens;* Birnstiel et al., 1971; Tartof, 1975), the 5 S RNA genes and tRNA genes (*S. cerevisiae* and *D. melanogaster;* Tartof, 1975) and the histone genes (sea urchin, Kedes, 1976; mouse, Jacob, 1976). Of the genes in this group, we have data in *G. lateralis* only on those coding for rRNA which we find to be present in a considerably higher frequency class. [The rRNA genes ($\rho = 1.704$ g/cm³) of *G. lateralis* comprise approximately 2.5% of the genome (Skinner, 1969). Assuming the combined molecular weight of 28S and 18S rRNA to be 3.34×10^{-18} g, there would be 3700 copies of the rRNA genes per haploid genome (1.25×10^{-14} g of rRNA/ 3.34×10^{-18} g per cistron)].

The reassociation kinetics of highly sheared DNA fragments of *Cancer borealis, Libinia emarginata,* (Vaughn, 1975), *Callinectes sapidus* (Vaughn, personal communication) and *Geryon quinquedens* Christie et al., 1976) are very similar to those of *G. lateralis*. Clearly, most of the repetitive sequences in the DNAs of the true crabs are of high reiteration frequency while sequences repeated only a few times are relatively rare. Since crabs have few sequences of the repetition frequency that are interspersed with single copy sequences in many other organisms (Davidson et al., 1976), we have studied the organization of the more highly repetitive sequences in the mcDNA.

2. Thermal Dissociation Properties of Reassociated Fragments

Evidence for the interspersion of repetitive and single copy sequences has been obtained from observation from observations on the optical melting properties

of self-reassociated fragments of increasing length. If single copy sequences are interspersed with repetitive sequences, then after reassociation of the repetitive sequences, many long fragments that contain duplex regions should also have relatively extensive single stranded regions. If, as in other organisms (Davidson et al., 1976), a major fraction of the repetitive sequences are interspersed with single copy sequences, then as the length of the fragments increases, the fraction of DNA bound to hydroxyapatite should increase. In this case, the hyperchromicity, a function of the fraction of nucleotides in duplex (Graham et al., 1974), should decrease as the fragment length increases.

Unlabeled mcDNAs of three different fragment lengths were reassociated in 0.12 M PB at 60° C to C_0t 10. C_0t 10 was chosen to insure that all repetitive sequences were in duplex form and that a significant quallity of DNA could be used for the hyperchromicity measurement. The fragments containing duplex regions were collected on hydroxyapatite (60° C, 0.12 M PB) and eluted with 0.5 PB, dialyzed to 0.12 M PB and melted. The fraction of mcDNA bound to hydroxyapatite increases markedly with increasing fragment length; 84% of the 2200 nucleotide long fragments contain a repetitive sequence as compared to 51% of the 420 nucleotide long fragments (Table 3). Since binding depends on the presence of duplex structures in each fragment and in mcDNA only repetitive sequences form duplexes by C_0t 10, most of the single copy sequences in mcDNA must be located adjacent to repetitive sequences in the 2200 nucleotide fragments.

Table 3. Characteristics of reassociated fragments of *G. lateralis* main component DNA

Fragment length (N.T.)	420	1680	2200	"Unsheared"
Fraction bound to HAP	0.512	0.620	0.835	—
Hyperchromicity (%)	17.50	16.13	13.80	28.0
Melting temperature (° C)	79.0	77.0	79.0	85.5
Transition (° C)	6.5	7	6.7	—
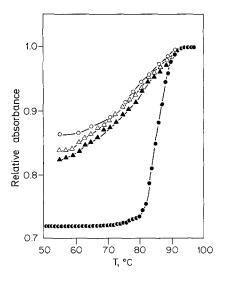
Fig. 3. Thermal dissociation of reassociated mcDNA. DNA fragments of 420 (▲), 1680 (△), and 2200 (○) N.T. were reassociated to C_0t 10. The partially duplex molecules were isolated by hydroxyapatite chromatography, dialyzed into 0.12 M PB and melted. As a control, unsheared mcDNA (●) in 0,12 M PB was melted in parallel

Figure 3 shows the melting curves of reassociated 420, 1680, and 2200 nucleotide long fragments and of unsheared mcDNA. As expected, the hyperchromicities of the reassociated duplex decrease slightly with increasing fragment length providing further evidence that as the fragment length increases the fraction of each that is single stranded also increases. Since the T_ms and transitions of each size class of duplexes are very similar, similar sequences have reassociated. Taking a $1°C$ depression in T_m as equivalent to 1.5% base mismatch (Bautz and Bautz, 1964) these sequences contain on the average 10.5% sequence divergence.

If the genome organization of *G. lateralis* were similar to that of *D. melanogaster*, the fraction of DNA bound to hydroxyapatite after reassociation would have increased very little as the fragment length increased as much as 5-fold. Reassociated partially duplex molecules of *D. melanogaster* DNA of the sizes tested here have practically identical hyperchromicities since the average repetitive sequence length for *D. melanogaster* is > 5600 nucleotides (Manning et al., 1975; Crain et al., 1976a). Our data are not consistent with a *Drosophila*-like pattern of genome organization but are very similar to those for *Xenopus*. The following experiments were performed to determine the specific lengths of the interspersed single copy and repetitive sequences.

3. Interspersion of Foldback Sequences

DNA sequences that form intramolecular duplexes at Cots $< 10^{-4}$ can be clustered highly repetitive sequences such as cryptic satellites or short palindromes (Davidson et al., 1973; Wilson and Thomas, 1974; Perlman et al., 1976). To determine if mcDNA contains a sufficient quantity of cryptic satellites to account for foldback sequences, mcDNA was centrifuged in Hg^{2+} or Ag^+-Cs_2SO_4 gradients with a range of ratios of heavy metals to total DNA phosphorus from 0.1 to 0.7 (Skinner and Beattie, 1973). Although the long mcDNA used for all these studies had a unimodal transition when melted (Fig. 3) and formed a symmetrical peak in neutral CsCl (Fig. 4b), a cryptic satellite was isolated in Ag^+-Cs_2SO_4 gradients at a ratio of 0.51 (Fig. 4c). Since it comprised less than 1% of the mcDNA it could not account for a significant fraction of the reassociated molecules.

Figure 5 shows the results of an experiment in which in vivo labeled mcDNA fragments of discrete increasing lengths were denatured, reassociated to C_0t $< 10^{-4}$, and passed over hydroxyapatite. DNA labeled in vivo as described in Materials and Methods was used for this experiment since Pol I labeled mcDNA fragments contain intramolecular duplexes that are produced during labeling. (See Characteristics of *E. coli* DNA Polymerase I Labeled DNA.)

The increase in the percentage of foldback DNA bound to hydroxyapatite with increasing fragment length (Fig. 5) would be expected if stretches of unreassociated single stranded DNA are adjacent to foldback sequences. From the linear relationship between the percentage of the fragments bound to hydroxyapatite and the fragment lengths, we conclude that in mcDNA the class of sequences that undergoes intramolecular reassociation is interspersed with other
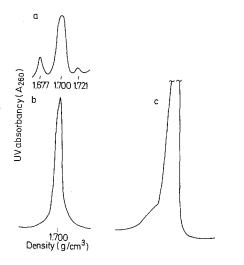
Fig. 4a–c. Tracings of Model E analytical ultracentrifuge runs of *G. lateralis* DNA. Total (a) was centrifuged in a preparative neutral CsCl gradient; unlabeled, 6000 N.T. mcDNA isolated from it was displayed in either neutral CsCl (b) or in Ag^+-Cs_2SO_4 (r=0.51) (c)
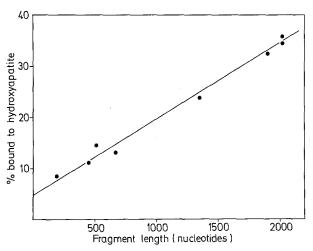

Fig. 5. Interspersion of foldback sequences. Purified in vivo labeled mcDNA of specific fragment lengths was denatured and reassociated to C_0t 10^{-4}, passed over hydroxyapatite and the duplex containing fragments eluted with 0.5 M PB. All fractions were counted as described in Materials and Methods

sequences ≥2000 nucleotides long. The intercept of the line with the Y-axis shows that approximately 5% of the mcDNA is foldback sequences. The organization of zero-time binding sequences in the DNA of tobacco (Zimmerman and Goldberg, 1977), cotton (Walbot and Dure, 1976), and *Aplysia* (Angerer et al., 1975) is similar to that of *G. lateralis*; by contrast, the foldback sequences in the genome of *Xenopus* (Davidson et al., 1973) and a sea urchin (Graham et al., 1974) are interspersed with sequences of a defined length.

4. The Spacing of the Interspersed Repetitive Sequences

The average length of the single copy sequences that are interspersed with repetitive sequences was determined by measuring the percentage of fragments of increasing length that bind to hydroxyapatite after reassociation with an excess of 380 nucleotide long fragments to a C_0t where all the repetitive but none of the single copy sequences have reassociated (Graham et al., 1974). If the repetitive sequences are interspersed with single copy sequences, many of the short fragments will not contain repetitive sequences and will not bind to hydroxyapatite. As the fragment length increases, the fraction of DNA containing a repetitive sequence and binding to hydroxyapatite increases as shown previously (Graham et al., 1974). When the fragments are longer than the interspersed single copy sequences, the percent bound to hydroxyapatite will not increase further.

Figure 6 shows the results of an experiment in which in vivo labeled mcDNA of specific fragment lengths was denatured and reassociated to Cot 10 in the presence of a 100-fold excess of unlabeled 380 nucleotide long fragments to prevent self-reassociation of the labeled tracer and to drive to completion the reassociation of all repetitive sequences in the labeled fragments. None of the single copy sequences but all of the repetitive sequences should be in duplex form at C_0t 10 (Fig. 2). Zero-time binding (foldback) sequences were not removed from the DNA since they are interspersed with other sequences (Fig. 5); rather, a correction was made for the contribution of the zero-time binding DNA at each fragment length.
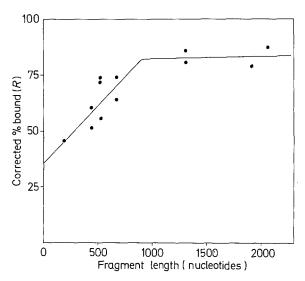

Fig. 6. Interspersion of repetitive sequences. In vivo labeled mcDNA of specific fragment lengths was mixed with unlabeled mcDNA (380 N.T. fragments) in excess and reassociated to C_0t 10. The fraction containing duplex regions (F) was measured by adsorption to hydroxyapatite and corrected for the fraction of foldback sequences (Z), to obtain the corrected binding (R) ($R = F\text{-}Z/1\text{-}Z$)

The line in Figure 6 is the best fit of the least squares solution for these data. Although the scatter of the data points makes several other solutions possible, the calculated length of the interspersed single copy sequences does not change markedly, since most solutions show a change in slope of the line at an average value of 800–900 nucleotides. These data show that the interspersed single copy sequences are, on the average, 800–900 nucleotides long and that more than 50% of the mcDNA is organized in a pattern of interspersed repetitive and single copy sequences.

5. Length of the Interspersed Repetitive Sequences.

To determine the quantity and the length of the interspersed repetitive sequences, 3100 nucleotide long fragments were reassociated to C_0t 10 in Pipes buffer (see Materials and Methods) and digested with S1 nuclease. The S1 resistant duplexes were collected on hydroxyapatite, washed with 0.12 M PB and eluted with 0.5 M PB. Forty-six percent of the DNA was eluted in the high salt fraction (Table 4, Experiments 1 and 2). Since approximately 2–3% of the S1 resistant DNA might be single stranded, the best estimate obtained from this experiment is that $\sim 44\%$ of the mcDNA sequences are repetitive and reassociate before C_0t 10. This agrees quite well with the data obtained independently from both the C_0t curve (Fig. 2; 43%) and from the interspersion experiment shown in Figure 6 (36.3%) when it is considered that extrapolation of the interspersion curve to short fragment lengths often provides an underestimate (Efstratiadis et al., 1976). From these data we calculate that approximately 58% of the G. lateralis genome is multiply reiterated (75% of the genome is mcDNA, 44% of which is repetitive; the remaining 25% of the genome is isolable as highly repetitive satellites).

After dialysis into 0.12 M PB, the S1 resistant duplexes (0.5 M fraction) were chromatographed on an agarose A-50 column. Two-thirds of the duplexes were excluded from the column while approximately one-third co-chromatographed with a 285 nucleotide long marker (Fig. 7a). We calculate that approximately 15% (0.33 × 0.46) of the mcDNA forms short S1 resistant duplexes (Table 4, Experiments 1 and 2).

Table 4. Duplex content of mcDNA fragments

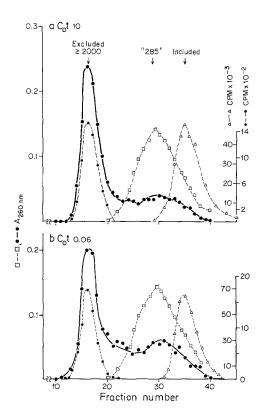|	C_0t 10		C_0t 0.06	
Experiment	1	2	3	4
A. Total[a]	0.460	0.453	0.337	0.299
B. 285 N.T.[b]	0.333	0.326	0.331	0.356
A × B	0.153	0.152	0.112	0.106

[a] Fraction of DNA bound to hydroxyapatite after S1 digestion

[b] Fraction of S1 resistant DNA co-chromatographing with a 285 N.T. marker

To exclude the possibility that the reassociated duplexes were underdigested by S1 nuclease, several control experiments were performed using incubations and digestion conditions identical to those used for the experiment. In one experiment performed in triplicate, in vivo labeled, 2000 nucleotide long ^{14}C-Tt phage DNA fragments were denatured and digested with 1 nuclease. After digestion, only 2–3% of the DNA was acid precipitable indicating that $\sim 97\%$ of the single stranded DNA had been digested. In another experiment, unlabeled 3000 nucleotide long fragments of human DNA were denatured, reassociated to C_0t 0.2 and digested. Approximately 30% of the DNA was S1 resistant and 33% of these duplexes were included in an A-50 column. These values agree well with those predicted from published work on the organization of the human genome (Schmid and Deininger, 1975). In a third experiment, very small quantities of reassociated duplexes of crab DNA were treated with S1 in a large volume. Even though the same enzyme to DNA ratio was used as in previous experiments, the effective enzyme concentration was reduced 100-fold. In this instance, the DNA appeared to be underdigested in that the excluded peak was not discrete and had a broad trailing edge. These data lead us to conclude that under the conditions described, S1 digests 97–98% of the single stranded ends while leaving duplex regions undigested.

To determine the length of the repetitive sequences reassociating before C_0t 0.1, 3100 nucleotide long fragments were reassociated to C_0t 0.06 and

Fig. 7a and b. Chromatography on Agarose A-50 of repetitive mcDNA sequences. The mcDNA of 3100 N.T. was reassociated to C_0t 10 (**a**) or C_0t 0.06 (**b**), digested with S1 nuclease as described in Materials and Methods and the duplex fragments collected on hydroxyapatite and eluted with 0.5 M PB. After dialysis into 0.12 M PB, a 3 ml sample was chromatographed on a column of agarose A-50, calibrated immediately before use with long native ^{14}C-T7 DNA (●----●) as an exclusion marker, ^{3}H-uridine (△----△) as an inclusion marker and 285 N.T. mcDNA fragments (□----□)

digested with S1 nuclease as in the previous experiment. Since the rate of reassociation increases as the square root of the fragment length (Britten et al., 1974), reassociating 3100 nucleotide long fragments to C_0t 0.06 is equivalent to reassociating 380 nucleotide long fragments to C_0t 0.17 ($\sqrt{3100}/\sqrt{380}=2.86$). From the C_0t curve, we predicted that all of the fast repetitive sequences and a fraction of the intermediate repetitive sequences ($\sim 50\%$) should reassociate. The results of the experiment (Fig. 7b; Table 4, Experiment 3) showed that approximately 34% of the mcDNA was S1 resistant, and that one-third of the S1 resistant duplexes ($\sim 11\%$ of the mcDNA) co-chromatographed with a 285 nucleotide long marker. Approximately 4% of the mcDNA reassociated between C_0t 0.06 and C_0t 10 as short interspersed sequences.

The experiment outlined in Figure 8 was performed to confirm that most of the short interspersed sequences reassociate before C_0t 0.06. When 3100 nucleotide long mcDNA fragments were reassociated to C_0t 0.06 in 0.12 M PB, 53.9% of the fragments contained duplex regions and bound to hydroxyapatite. This is approximately the percentage expected if the long repetitive sequences reassociate as the fast component and the short interspersed repetitive sequences reassociate with the intermediate class of sequences. Approximately 20% of the mcDNA should reassociate at a low C_0t as tandem or clustered repetitive sequences (Fig. 6); half of the short interspersed repetitive sequences (7.5%
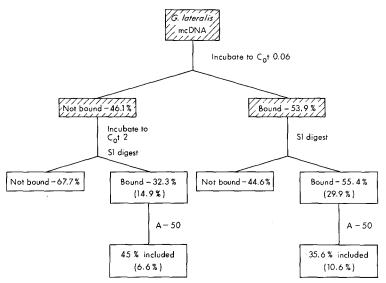

Fig. 8. Preparation of repetitive sequences that reassociate between C_0t 0.06 and C_0t 10. *G. lateralis* 3100 N.T. long mcDNA fragments were denatured, reassociated to Cot 0.06, and duplex containing fragments captured on hydroxyapatite. The partially duplex fragments (*Bound*-right column) were than eluted with 0.5 M PB, dialyzed into the buffers outlined in Materials and Methods and digested with S1 nuclease. The S1 resistant molecules were then chromatographed on A-50. The results of the experiments with the bound fraction are listed in Table 4, Experiment 4. After reassociation to C_0t 2, the molecules not bound to hydroxyapatite were treated similarly (*Not Bound*—left column). All steps enclosed in the shaded boxes were performed with 3100 N.T. long fragments. The numbers in parentheses are the percentages of the starting mcDNA

of the mcDNA) should reassociate and be attached to half of the single copy sequences (25–30% of the mcDNA).

Aliquots of both the bound and unbound fractions were resized and shown to be 3100 nucleotides. The bound fraction was dialyzed into the buffers described in Materials and Methods for digestion with S1 nuclease. Approximately 55% of the bound fraction (30% of the starting mcDNA) was S1 resistant. The single-stranded, S1-sensitive sequences are presumably single copy sequences that are linked to one or more repetitive sequences in 3100 nucleotide fragments. The results of this experiment (Table 4, Experiment 4) agree with the previous predictions and are practically identical to those obtained when the reassociated fragments were digested with S1 nuclease before hydroxyapatite chromatography (Table 4, Experiment 3). They demonstrate that $\sim$11% of the mcDNA has reassociated as short interspersed sequences by C_0t 0.06.

The fraction of the reassociated DNA that did not bind to hydroxyapatite was dialyzed into the appropriate buffers, denatured, reassociated to C_0t 2 and digested with S1 nuclease. Approximately 32% of the DNA was in S1 resistant duplexes indicating that 15% of the mcDNA sequences reassociate between C_0t 0.06 and C_0t 2 (Table 4, Experiment 2). When chromatographed on agarose A-50, 45% of the duplexes were 285 nucleotides long. Thus 6.6% (0.45×0.15) of the mcDNA sequences reassociate as short interspersed repetitive sequences between C_0t 0.06 and C_0t 2. The results are in agreement with the previous experiment and support the conclusion that $\sim$11% of the short interspersed sequences reassociate before C_0t 0.06 and 4–7% reassociate at a higher C_0t.

Discussion

At least 55% of the DNA of the land crab is composed of highly repetitive sequences. Of these, approximately 25% are isolable as distinct satellites in neutral CsCl gradients (Skinner, 1967). One of these, "crab poly(dA-dT)" would presumably undergo extensive intramolecular reassociation at criterion (T_m-25°) at very low C_0ts. Preliminary observations show that the second satellite, which is G+C-rich and comprises 3 to 5% of the DNA, is also highly repetitive in that it reassociates before C_0t 10^{-2} (Holland and Skinner, unpublished observations). In addition to these two satellite DNAs, at least 40% of the sequences of mcDNA are repeated more than 800 times; approximately 33% of the repetitive sequences (19% of the mcDNA or 14% of the total DNA) is interspersed with single copy sequences in a short period interspersion pattern.

In addition to those in *G. lateralis*, most repetitive sequences in the total DNAs of four other crabs examined by us (Christie and Skinner, in preparation) and by others (Vaughn, 1975) are present more than 800 times. These crab genomes also contain few, if any, sequences reassociating between C_0t 1 and C_0t 100. The dearth of such sequences in all five crabs divides their C_0t curves into clearly distinguishable highly repetitive and single copy components. The existance of specific frequency classes is unusual in invertebrate genomes which

are more often composed of detectable quantities of every repetition frequency (Sachs and Clever, 1972; Lagowski et al., 1974; Goldberg et al., 1976).

At least two types of data suggest that repetitive sequences are in some way maintained in crustacean genomes. First, several satellite DNAs are composed almost entirely of uniform, simple repeating sequences. For example, essentially all of Satellite I from the hermit crab, *Pagurus pollicaris*, has the sequence $\frac{(-T-A-G-G-)n}{(-A-T-C-C-)n}$ (Skinner et al., 1974) and more than 90% of Satellite II from the same organism appears to be a simple repeat of a 31-mer (Chambers et al., in preparation). In addition, approximately 20% of the genome of *G. lateralis* (and several other crustaceans; Beattie and Skinner, 1972) is composed of a light satellite in which more than 90% of the sequence is alternating adenylate and thymidylate residues. These are in marked contrast to several mammalian satellite DNAs which contain a simple repeating sequence and variations thereon, leading to a repeat unit of 200 to 300 nucleotides (guinea pig, Southern, 1970; mouse, Biro et al., 1975). Secondly, a significant amount of homology exists between highly repetitive sequences of crustaceans. In *G. lateralis* DNA, approximately half of the repetitive sequences in the main component and 5–10% of those in the $G+C$-rich satellite are represented in the DNA of other true crabs (Graham and Skinner, 1973). In that study, the hybrids formed between repetitive mcDNA or the $G+C$-rich satellite and heterologous crustacean DNAs showed relatively narrow thermal transitions and high thermal stabilities providing evidence that the conserved sequences have undergone relatively little divergence. Similar homologies between DNA sequences of *L. emarginata* and those of other crustacean DNAs have been observed (Vaughn and Traeger, 1976). The sequence homogeneity of a highly repetitive sequence within an organism on the one hand and conservation of repetitive sequences between divergent organisms on the other, suggest that crustaceans maintain repetitive sequences. In addition, the data reported here suggest that 800 or more copies of most repetitive sequences are maintained.

Our data also demonstrate that repetitive sequences are interspersed with single copy sequences in a short period interspersion pattern which is very similar to that of *Xenopus* (Davidson et al., 1973), the sea urchin (Graham et al., 1974) and other invertebrates (Angerer et al., 1975; Goldberg et al., 1975). At least 50% of the mcDNA is arranged in a pattern of interspersed repetitive sequences.

References

Angerer, R.C., Davidson, E.H., Britten, R.J.: DNA sequence organization in the mollusc Aplysia californica. Cell **6**, 29–39 (1975)

Angerer, R.C., Davidson, E.H., Britten, R.J.: Single copy DNA and structural gene sequence relationships among four sea urchin species. Chromosoma (Berl.) **56**, 213–226 (1976)

Bautz, E.K.F., Bautz, F.A.: The influence of noncomplementary bases on the stability of ordered polynucleotides. Proc. nat. Acad. Sci. (Wash.) **52**, 1476–1481 (1964)

Beattie, W.G., Skinner, D.M.: The diversity of crustacean satellite DNAs. Biochim. biophys. Acta (Amst.) **281**, 169–178 (1972)

Berg, P., Fancher, H., Chamberlin, M.: The synthesis of mixed polynucleotides containing ribo- and deoxyribonucleotides by purified preparations of DNA polymerase from E. coli. In: Informational macromolecules (H.J. Vogel, V. Bryson and J.O. Lampen, eds.), pp. 467–484. New York: Academic Press 1963

Biro, P.A., Carr-Brown, A., Southern, E.M., Walker, P.M.B.: Partial sequence analysis of mouse satellite DNA: evidence for short range periodicities. J. molec. Biol. **94**, 71–86 (1975)

Birnstiel, M.L., Chipchase, M., Speirs, J.: The ribosomal RNA cistrons. Prog. Nucleic Acids Res. molec. Biol. **11**, 351–389 (1971)

Bollum, F.J.: Oligodeoxyribonucleotide-primed reactions catalyzed by calf thymus polymerase. J. biol. Chem. **237**, 1945–1949 (1962)

Britten, R.J., Graham, D.E., Eden, F.C., Painchaud, D.M., Davidson, E.H.: Evolutionary divergence and length of repetitive sequences in sea urchin DNA. J. molec. Evol. **9**, 1–23 (1976)

Britten, R.J., Graham, D.E., Neufeld, B.R.: An analysis of repeating DNA sequences by reassociation. In: Methods in enzymology, **29E**, (L. Grossman and K. Moldave, eds.), pp. 363–418. New York: Academic Press 1974

Cairns, J.: The chromosome of E. coli. Cold Spr. Harb. Symp. quant. Biol. **28**, 43–46 (1963)

Christie, N.T., Holland, C.A., Skinner, D.M.: Repetitive sequences in the DNA of a highly evolved crab and a primitive crab. J. Cell Biol. **70**, 246a (1976)

Crain, W.R., Davidson, E.H., Britten, R.J.: Contrasting patterns of DNA sequence arrangement in Apis mellifera (Honeybee) and Musca domestica (Housefly). Chromosoma (Berl.) **59**, 1–12 (1976b)

Crain, W.R., Eden, F.C., Pearson, W.R., Davidson, E.H., Britten, R.J.: Absence of short period interspersion of repetitive and non-repetitive sequences in the DNA of Drosophila melanogaster. Chromosoma (Berl.) **56**, 309–326 (1976a)

Das, S.K., Fujimura, R.K.: Exonuclease associated with bacteriophage T-5-induced DNA polymerase. J. Virol. **20**, 70–77 (1976)

Davidson, E.H., Galau, G.A., Angerer, R.C., Britten, R.J.: Comparative aspects of DNA organization of metazoa. Carnegie Inst. Wash. Year Book **74**, 673–678 (1976)

Davidson, E.H., Hough, B.R., Amenson, C.S., Britten, R.J.: General interspersion of repetitive with non-repetitive sequence elements in the DNA of Xenopus. J. molec. Biol. **77**, 1–23 (1973)

Efstratiadis, A., Crain, W.R., Britten, R.J., Davidson, E.H., Kafatos, F.C.: DNA sequence organization in the lepidopteran Antheraea pernyi. Proc. nat. Acad. Sci. (Wash.) **73**, 2289–2293 (1976)

Englund, P.T.: DNA polymerase from E. coli. In: Procedures in nucleic acid research (G. Cantoni and D. Davies, eds.), pp. 864–874. New York: Harper and Row 1971

Goldberg, R.B., Crain, W.R., Ruderman, J.V., Moore, G.P., Barnett, T.R., Higgins, R.C., Gelfand, R.A., Galau, G.A., Britten, R.J., Davidson, E.H.: DNA sequence organization in the genomes of five marine invertebrates. Chromosoma (Berl.) **51**, 225–251 (1975)

Graham, D.E., Neufeld, B.R., Davidson, E.H., Britten, R.J.: Interspersion of repetitive and non-repetitive DNA sequences in the sea urchin genome. Cell **1**, 127–137 (1974)

Graham, D.E., Skinner, D.M.: Homologies of repetitive DNA sequences among Crustacea. Chromosoma (Berl.) **40**, 135–152 (1973)

Holland, C.A., Skinner, D.M.: Interactions between molting and regeneration in a land crab. Biol. Bull. **150**, 222–240 (1976)

Jacob, E.: Histone-gene reiteration in the genome of mouse. Europ. J. Biochem. **65**, 275–284 (1976)

Kedes, L.H.: Histone messengers and histone genes. Cell **8**, 321–331 (1976)

Klett, R.P., Cerami, A., Reich, E.: Exonuclease VI, a new nuclease activity associated with E. coli DNA polymerase. Proc. nat. Acad. Sci. (Wash.) **60**, 943–950 (1968)

Kunkel, L.M., Smith, K.D., Boyer, S.H.: Human Y-chromosome-specific reiterated DNA. Science **191**, 1189–1190 (1976)

Lagowski, J.M., Wong Yu, M., Forrest, H.S., Laird, C.D.: Dispersity of repeat DNA sequences in Oncopeltus fasciatus, an organism with diffuse centromeres. Chromosoma (Berl.) **43**, 349–373 (1973)

Manning, J.E., Schmid, C.W., Davidson, N.: Interspersion of repetitive and nonrepetitive DNA sequences in the Drosophila melanogaster genome. Cell **4**, 141–155 (1975)

Markham, R.: The ultracentrifuge. In: Methods in virology (K. Maramorosch and H. Koprowski, eds.), pp. 1–39. New York: Academic Press 1967

Masamune, Y., Richardson, C.C.: Strand displacement during deoxyribonucleic acid synthesis at single strand breaks. J. biol. Chem. **246**, 2692–2701 (1971)

Musich, P.R., Skinner, D.M.: A cytological study of the DNA of the Bermuda land crab. J. Cell Biol. **55**, 184a (1972)

Noll, H.: Characterization of macromolecules by constant velocity sedimentation. Nature (Lond.) **215**, 360–363 (1967)

Perlman, S., Phillips, D., Bishop, J.: A study of foldback DNA. Cell **8**, 33–42 (1976)

Sachs, D.H., Painter, E.: Improved flow rates with porous sephadex gels. Science **175**, 781–782 (1972)

Sachs, R.I., Clever, U.: The unique and repetitive DNA sequences in the genome of Chironomus tentans. Exp. Cell Res. **74**, 587–591 (1972)

Schmid, C.W., Deininger, O.L.: Sequence organization of the human genome. Cell **6**, 345–358 (1975)

Skinner, D.M.: Satellite DNAs in the crabs Gecarcinus lateralis and Cancer pagurus. Proc. nat. Acad. Sci. (Wash.) **58**, 103–110 (1967)

Skinner, D.M.: Deoxyribonucleic acid sequences complementary to ribosomal ribonucleic acid in a crustacean. Biochemisty **8**, 1467–1473 (1969)

Skinner, D.M., Beattie, W.G.: Cs_2SO_4 gradients containing both Hg^{2+} and Ag^+ effect the complete separation of satellite DNAs having identical densities in neutral CsCl gradients. Proc. nat. Acad. Sci. (Wash.) **70**, 3108–3110 (1973)

Skinner, D.M., Beattie, W.G., Blattner, F.R., Stark, B.P., Dahlberg, J.E.: The repeat sequence of a hermit crab satellite deoxyribonucleic acid is $(-T-A-G-G-)_n \cdot (-A-T-C-C-)_n$. Biochemistry **13**, 3930–3937 (1974)

Southern, E.M.: Base sequence and evolution of guinea pig α-satellite DNA. Nature (Lond.) **227**, 794–798 (1970)

Studier, F.W.: Sedimentation studies of the size and shape of DNA. J. molec. Biol. **11**, 373–390 (1965)

Tartoff, K.D.: Redundant genes. Ann. Revs. Genet. 355–385 (1975)

Vaughn, J.C.: DNA reassociation kinetics and chromosome structure in the crabs Cancer borealis and Libinia emarginata. Chromosoma (Berl.) **50**, 243–257 (1975)

Vaughn, J.C., Traeger, F.J.: Conservation of repeated DNA base sequences in the Crustacea: A molecular approach to decapod phylogeny. J. molec. Evol. **7**, 111–131 (1976)

Vogt, V.M.: Purification and further properties of single-strand-specific nuclease from Aspergillus oryzae. Europ. J. Biochem. **33**, 192–200 (1973)

Walbot, V., Dure, L.S.: Developmental biochemistry of cotton seed embryogenesis and germination. VII. Characterization of the cotton genome. J. molec. Biol. **101**, 503–536 (1976)

Wilson, D.A., Thomas, C.A.: Palindromes in chromosomes. J. molec. Biol. **84**, 115–144 (1974)

Zimmerman, J.L., Goldberg, R.B.: DNA sequence organization in the genome of Nicotiana tabacum. Chromosoma (Berl.) **59**, 227–252 (1977)