Cold Spring Harbor Symposia on Quantitative Biology

# The Organization of the Histone Genes in *Drosophila melanogaster:* Functional and Evolutionary Implications

R. P. Lifton, M. L. Goldberg, R. W. Karp, et al.

*Cold Spring Harb Symp Quant Biol* 1978 42: 1047-1051
Access the most recent version at doi:10.1101/SQB.1978.042.01.105

# The Organization of the Histone Genes in *Drosophila melanogaster:* Functional and Evolutionary Implications

R. P. LIFTON, M. L. GOLDBERG, R. W. KARP, AND D. S. HOGNESS

*Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305*

The expression of the genes for each of the five histone proteins is tightly regulated with respect to the cell cycle, the developmental stage, and the expression of the other histone genes. In several species studied, the synthesis of all five proteins occurs exclusively in concert with the DNA synthetic phase of the cell cycle (Robbins and Borun 1967; Kedes et al. 1969; Perry and Kelley 1973); the four histones of the nucleosome core are found in equimolar amounts, whereas H1 protein is found in less than molar quantity (for review, see Kornberg 1974). Variant histone proteins associated with specific developmental stages have been observed (Cohen et al. 1975; Arceci et al. 1976; Newrock and Cohen, this volume), indicating that there must be genes for these variant proteins which are differentially expressed according to precise developmental programs. *Drosophila melanogaster* is an appealing organism in which to study these genes and their expression for several reasons: (1) Its hereditary mechanics are understood in detail; (2) its polytene chromosomes allow high-resolution cytological mapping of the genome; and (3) it is easily accessible to experimental manipulation by virtue of the availability of tissue culture cells, a short life cycle that consists of well-defined developmental stages, and a small genome that facilitates the isolation of specific genes in cloned DNA segments (Dm segments). In this paper we review our knowledge of the organization of the *D. melanogaster* histone genes and discuss some of the functional and evolutionary implications of this organization.

## Characterization of the *D. melanogaster* Histone Gene Repeat Unit

The amino acid sequences of the histone proteins are highly conserved across species barriers; consequently, histone mRNAs, readily isolated from sea urchins, can hybridize to DNA from a wide variety of organisms (Kedes and Birnstiel 1971; Farquhar and McCarthy 1973). The histone genes of *D. melanogaster* were therefore first isolated by screening a collection of ColE1 hybrid plasmids containing randomly generated segments of *D. melanogaster* chromosomal DNA (cDm plasmids) for those that can hybridize with sea urchin histone mRNA (Karp and Hogness 1976, and in prep.). The plasmid cDm500 was obtained from this screen. Figure 1A shows a restriction map of the Dm500 segment carried by this hybrid and reveals that it consists of a 4.8-kb sequence repeated in tandem 1.8 times. The reassociation kinetics of this repeat unit in the presence of a vast excess of total *D. melanogaster* DNA indicates that its sequences are repeated approximately 100 times per haploid genome. Virtually all copies of this DNA sequence are located in region 39DE of *D. melanogaster* salivary gland polytene chromosomes, as shown by in situ hybridization using $^3$H-labeled cDm500 DNA (Karp and Hogness 1976, and in prep.).

The organization of genes within the repeating unit of Dm500 has been elucidated. Five poly(A)$^-$ RNAs present in *D. melanogaster* embryos and tissue-culture cells are complementary to this DNA, and each exhibits a length in the range expected for histone mRNAs. These RNAs were individually purified by annealing them to denatured cDm500 DNA covalently linked to cellulose powder (Noyes and Stark 1975) and subsequent gel electrophoresis. The purified RNAs were then hybridized to sets of restriction fragments from cDm500 to map the sequences in the DNA that are homologous to each of the five classes of RNA (R. Lifton et al., in prep.). Figure 1B shows that all five of the genes coding for these RNAs are confined to a single repeat unit. These genes were then identified by determining the nucleotide sequence of a portion of the DNA within each homology region and showing that such a sequence could be conceptually translated to yield the amino acid sequence of part of one of the histone proteins (M. Goldberg et al., in prep.).

The orientation of each gene, that is, the 5′-to-3′ direction of the coding sequence and hence the direction of transcription, was determined by annealing the individual RNA species to separated DNA strands of a hybrid λ phage containing the 4.8-kb unit (R. Lifton et al., in prep.). Figure 1B shows that the direction of transcription of each successive gene alternates as one proceeds to the right across the map from the H3 gene to the H1 gene. Three genes (H3, H2A, H1) are therefore transcribed from one DNA strand, and two (H4, H2B) from the other strand. These directions are also deducible from the nucleotide sequence of the DNA in each gene (M. Goldberg et al., in prep.), and they agree with this direct experimental determination.

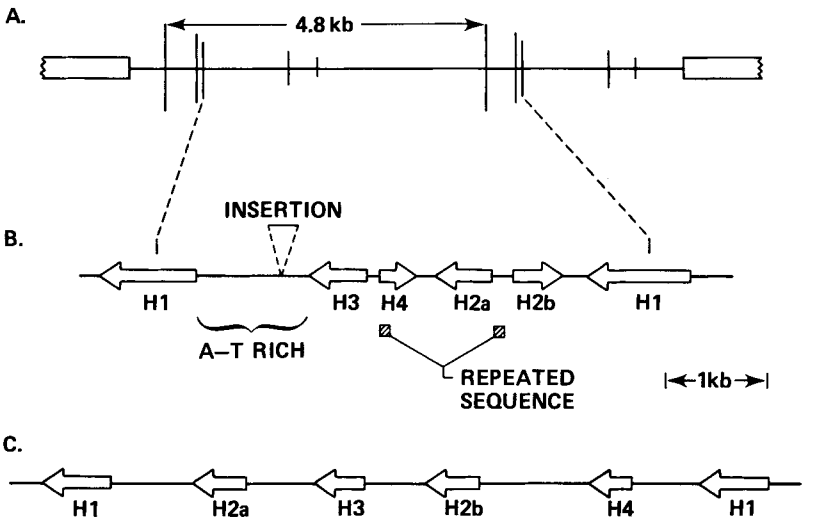


**Figure 1.** Maps of histone gene repeating units. *(A)* Restriction map of cDm500. The open rectangles and horizontal line represent, respectively, the ColE1 and Dm500 segments joined in cDm500 by poly(dA)-poly(dT) connectors according to the method of Wensink et al. (1974). The vertical lines represent restriction enzyme sites in the Dm500 segment. Starting with the leftmost site and progressing to the right, the sites are, respectively, *Bgl*II, *Bam*I, *Hind*III, *Hpa*I, and *Sst*I. Each of these sites is repeated in the Dm500 segment with a spacing of 4.8 kb. *(B)* Arrangement of the histone genes in *D. melanogaster* repeat units. The two classes of units are represented, one class differing from the other by an insertion in the spacer between H1 and H3. The scale in kilobases is indicated by the double arrow and applies to both *B* and *C*. *(C)* Arrangement of the histone genes in *S. purpuratus* (Cohn et al. 1976).

Is the histone gene organization in cDm500 representative of all 100 copies of this sequence in the *D. melanogaster* genome, or is there heterogeneity in the topographies of either the genes or their intervening spacers? Using labeled cDm500 DNA as a probe for homologous sequences in restriction enzyme digests of total *D. melanogaster* DNA blotted onto nitrocellulose by the procedure of Southern (1975), we have found two types of repeating unit. One corresponds to the 4.8-kb unit contained in cDm500, and the other has the same gene organization but contains an additional block of 270 base pairs in the spacer separating H1 from H3 (Fig. 1B). This latter type of repeat unit is three times as frequent as the former in the *D. melanogaster* genome (Karp and Hogness 1976, and in prep.). These two types of units constitute the vast majority of the *D. melanogaster* histone gene sequences.

### Models for Histone Gene Transcription

The map of the *D. melanogaster* histone genes enables us to exclude certain models for the mode of expression of these genes. Since the five genes are not all transcribed from the same DNA strand, we can rule out the possibility that coordinate expression of these genes is effected by a single polycistronic transcript that is cleaved to yield the five mature RNAs. Assuming that the RNA polymerase cannot switch from one strand to the other during transcription of these genes, the five RNAs must derive from at least two transcripts. Such transcripts could arise in a variety of ways. For instance, promoters lying outside the cluster of histone gene repeat units could initiate wholesale transcription of both DNA strands such that the individual transcripts cover many repeat units. At the other extreme, two polycistronic transcripts could extend across just enough of each strand to include the three or two RNAs derived from that strand. Since genes of opposite orientations are interdigitated throughout the repeat unit, any two-transcript model requires the transcription of both sense and antisense strands within the units.

A model in which each of the five genes is individually transcribed to yield RNAs that closely approximate the mature message seems particularly suited to the organization of the *D. melanogaster* histone genes. In such a scheme, transcription would originate in both directions from the spacers lying between gene pairs H3-H4 and H2A-H2B, whose members have their 5′ ends in close proximity. By contrast, the 5′ end and presumptive promoter of H1 mRNA lies in relative isolation from the other genes, separated by approximately 1200 base pairs from the 3′ end of H3. Since histones H2A, H2B, H3, and H4 are coordinately synthesized and utilized in equimolar amounts, one might expect the two proposed sites for the bidirectional promoters to share common DNA sequences. This may in fact be the case. When single strands of a linear DNA consisting of one repeat unit are observed in the electron microscope, hairpin snap-back structures characteristic of a nontandem inverted repeat sequence are seen. The regions of homology are roughly 100 base pairs in length and are at, or very close to, the two proposed locations of bidirectional promoters (Fig. 1B) (R. Lifton et al., in prep.). In this model, H1 would have its own distinct promoter to account for the nonstoichiometric amounts of both H1 protein (Kornberg 1974) and mRNA (R. Lifton et al., in prep.) found in the cell.

Preliminary evidence from our laboratory is consistent with this model. Hybridization of electrophoretically separated total RNA from cultured cells labeled by a 12-minute pulse of [$^3$H]uridine to cDm500 DNA revealed no histone sequences in transcripts perceptibly larger than the mature histone

mRNAs. In addition, we have looked for histone-specific transcripts in total cell RNA by the sensitive RNA-blotting technique developed by Alwine et al. (1977), using $^{32}$P-labeled cDm500 DNA as the probe. Again, the sequences homologous to cDm500 are confined to molecules the size of the histone mRNAs themselves (R. Lifton, unpubl.). Definitive proof that the primary transcripts are very similar to the mature histone mRNAs will require the demonstration of a nucleoside 5′ triphosphate on such species.

### Comparison of *D. melanogaster* and Sea Urchin Histone Genes

The organization of the histone genes in several species of sea urchin has been elucidated (Fig. 1C). In each, all five genes are contained in a unit that is tandemly repeated, and all the genes are transcribed in the same direction. The gene order is identical in all species examined (for review, see Kedes 1976). Bearing in mind that the branches of the phylogenetic tree which lead, respectively, to the sea urchins and the fruit flies diverged roughly 600 million years ago (Dickerson 1971), it is of interest to ask what structural and organizational features have been conserved since this evolutionary divergence, since we may suppose that conserved features will be those most crucial for the maintenance of proper gene expression.

From a cursory comparison of the sea urchin and *D. melanogaster* histone genes, one sees that the unit of organization is the same; each of the five genes is represented once per unit. The organization within this unit, however, is quite different. The *D. melanogaster* histone genes are transcribed from both DNA strands, and the gene order is completely rearranged with respect to that in sea urchins.

Although the histone protein sequences are rigidly conserved in evolutionary time, we find it curious that the arrangement of the histone genes within the repeat unit is capable of undergoing extensive changes. In considering these changes, it is of interest to determine the shortest paths of interconversion between the arrangements in sea urchins and *D. melanogaster*. At least five breaks are required to interconvert the two gene orders. Among the various five-break paths for gene-order conversion, only two will also interconvert the transcriptional orientations. These are indicated in Figure 2. Each involves a single intermediate (or common precursor) that can be converted to the sea urchin or *D. melanogaster* arrangement by a two- or three-break path.

The finding that both organisms carry repeat units containing all five genes arranged in such a way that they can be interconverted by relatively simple paths leads us to suggest that the five histone genes were linked in those species whose descendants subsequently diverged to give rise to the Protostomia and Deuterostomia. Assuming that evolution is economical, we might then expect the ancestral gene arrangement to be one of the four shown in Figure 2; i.e., we might expect to find one of these four arrangements in the primitive bilaterally symmetrical organisms whose prototypes evolved into the Protostomia and Deuterostomia. In any case, it would be important to a determination of the time and chromosome mechanics involved in the evolution of the two known arrangements of histone genes to investigate such arrangements in species that lie below *D. melanogaster* and sea urchins on these two branches of the phylogenetic tree.

The rearrangement of the histone genes has been accompanied by the divergence of the spacers which separate these coding sequences. In the sea urchin

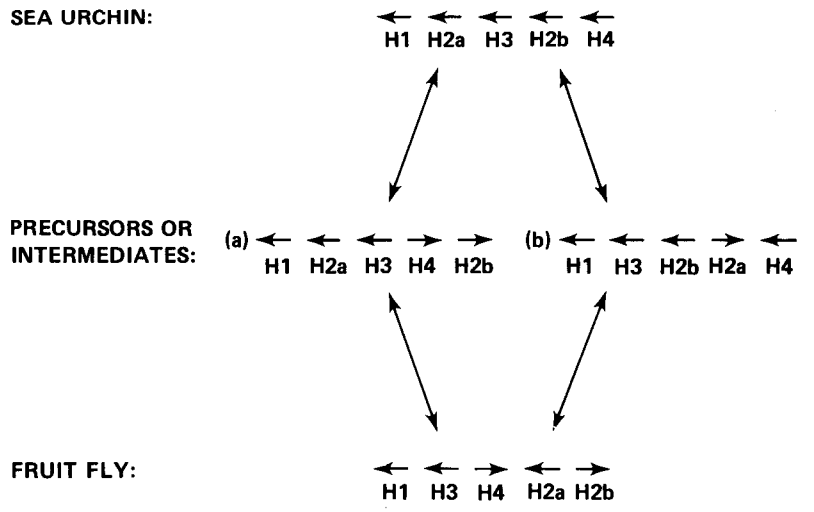


**Figure 2.** Two possible evolutionary paths for interconverting sea urchin and fruit fly histone gene organizations. The histone gene organization of sea urchin and fruit fly can be interconverted by rearrangements involving five chromosomal breakages involving either of two intermediate topographies. The sea urchin histone gene arrangement can be converted to the arrangement in structure *a* by inverting the H2B-H4 segment with respect to the other genes (two breaks). This structure can then be converted to the fruit fly organization by cutting out H2A and inserting it between H4 and H2B (three breaks). The same end result can be accomplished by cutting out H2A from the sea urchin genes, inverting it, and inserting it between H2B and H4 (three breaks). This structure *(b)* is then converted by an inversion of segment H2B, H2A, and H4 to the fruit fly topography. The path arrows are drawn in both directions since any of the four arrangements could represent the ancestral histone gene topography.

*Strongylocentrotus purpuratus,* the five spacers vary in length from 600 to 1200 base pairs (Fig. 1C). All of these spacers are A-T-rich in *Psammechinus miliaris,* as indicated by their preferential denaturation in formaldehyde (Portman et al. 1976). In *D. melanogaster,* four spacers are quite short (in the range of 0–250 bp) and the fifth is long (1200 bp). When the *D. melanogaster* repeat unit is spread for electron microscopy in 80% formamide, a single region of denaturation is observed; this region comprises the long spacer that lies between H3 and H1 (M. Goldberg et al., in prep.). These observations therefore yield a rule that applies to the repeat units in sea urchins and in *D. melanogaster:* Whenever adjacent genes are transcribed in the same direction, they are separated by a relatively long A-T-rich spacer. At present we can do little more than note this correlation, since any attempt to evaluate its significance will depend upon the location of the promoters and terminators for the transcription of the respective genes, and these loci have not been determined.

Since the five histone genes in *D. melanogaster* do not function as a single transcriptional unit, one might reasonably ask why these genes are closely linked. One might argue that this linkage simply reflects evolutionary history if, as has been suggested, these genes evolved from a common ancestral gene by repeated tandem duplication followed by divergence (Temussi 1975). However, a strong selection pressure must exist if this linkage has been maintained for roughly 1.2 billion years since these genes arose (Dickerson 1971) or at least, as we have postulated, the last 600 million years. Two functional advantages of this linkage come easily to mind. First, since these genes are reiterated, their linkage in a single unit provides a simple means of maintaining an equal dosage of each gene. Changes in the number of copies of repeat unit arising by unequal crossing-over will equally affect all five genes, thereby avoiding imbalances in dosage that could arise if, for example, each gene were individually reiterated in tandem. Linkage may also facilitate the coordinate expression of these genes. One can imagine that local events, such as the regional unfolding of the chromatin, could render all five previously inaccessible genes simultaneously available for transcription (see Finnegan et al., this volume, for a general consideration of this kind of control). The linkage of the histone genes could thus help to ensure both the simultaneous synthesis of histone proteins and their synthesis in proper molar ratios.

### The Organization of Histone Gene Repeat Units in 39DE

The 100 copies of the *D. melanogaster* histone genes appear to span most of 12 chromomeres in the 39DE region in the left arm of chromosome 2 (Pardue 1975; R. Karp and D. Hogness, in prep.). Partial digests of total *D. melanogaster* DNA indicate that these repeats generally occur in tandem; arrays comprising up to five tandem repeats have been observed (R. Karp and D. Hogness, in prep.). How are the two types of repeat units (i.e., the 4.8-kb and 5.0-kb units) arranged with respect to one another in this region? One can imagine several distinct possibilities for this arrangement. The two types could represent a genetic polymorphism having no functional significance. In this case one might expect recombination to lead to the random interspersion of the two types of repeat units. Alternatively, if each type has a different physiological role, one might expect to find these units arranged in some specific manner. To study this problem we have made use of the fact that the 4.8-kb unit contains no *Eco*RI cleavage site, whereas the 5.0-kb unit contains such a site within the insertion shown in Figure 1B. In probing the total *Eco*RI-cut *D. melanogaster* genome for histone gene sequences, we never find the 4.8-kb unit lying immediately between two 5.0-kb units (R. Karp and D. Hogness, in prep.). We conclude that the two types are not randomly interspersed, since roughly 15 of the 25 4.8-kb units would lie between 5.0-kb units if random interspersion occurred, and these would be easily detectable. These two types of units are not, however, completely segregated. We have cloned other histone gene units. Two plasmids each contain three full repeat units in tandem, and each contains both the 4.8-kb and the 5.0-kb units (R. Lifton, unpubl.). These combined results indicate that the two units are interspersed, but that this interspersion is decidedly nonrandom. These findings support the notion that each of the two types of units has physiological significance to the fly, though speculation regarding the different functions each type might serve is thus far unrestrained by data.

As previously mentioned, there are 12 bands and interbands in the 39DE region. Do the histone gene repeat units traverse all of these chromomeres in a single tandem array consisting of all 100 repeats, or are these arrays interrupted by sequences unrelated to the histone repeat unit or by inversions of some repeat units? As indicated above, tandem arrays of at least five repeat units exist. However, two lines of evidence suggest that interruptions do in fact occur. First, recall that the 4.8-kb units are not found between two 5.0-kb units. If uninterrupted tandem arrays are maintained throughout 39DE, then the 4.8-kb units should be found clustered in groups of two, three, or more units in tandem, bounded on either side by 5.0-kb units. We know that this is not the case, since the Southern blots of *Eco*RI-cut total *D. melanogaster* DNA do not reveal distinct fragments containing histone DNA sequences with lengths that are integral multiples of

the unit length. Instead, DNA fragments containing homology to 4.8-kb units are found over a wide length range (R. Karp and D. Hogness, in prep.). The simplest explanation of this observation is that the tandem arrays are interrupted so that the clusters of 4.8-kb units are bounded on at least one side by sequences of variable length or composition that are not homologous to those in the histone repeat unit.

In support of this notion, we have observed that when total *D. melanogaster* DNA is cleaved with *Hin*dIII, which cuts both types of units only once, and then blotted to nitrocellulose filters, we detect many fragments other than those of 4.8 kb and 5.0 kb that contain sequences homologous to the histone repeat unit. These fragments are all present at much lower frequencies per genome relative to the 4.8-kb and 5.0-kb fragments, and most cannot be explained by the gain or loss of a *Hin*dIII site in variant repeat units (R. Karp, unpubl.). Since there is no evidence that regions outside of 39DE have homology to histone DNA, we favor the interpretation that at least some of these extra fragments represent sites where the tandem array is interrupted. Indeed, we have recently cloned two Dm segments that contain histone repeat sequences adjacent to blocks of sequences that are not homologous to histone DNA (R. Karp, unpubl.).

What could be the function of these other sequences? If the formation of band-interband boundaries is determined by specific DNA sequences, such sequences must be present somewhere along the histone gene arrays in order to account for the multiple chromomeres in the 39DE region. It is tempting, therefore, to speculate that these intervening sequences could be involved in such a function. Alternatively, the information required for band-interband boundaries must lie in the histone repeat units themselves, for example, in the 270-bp insertion.

There are a variety of indications that individual chromomeres represent single functional units (Beermann 1972). Does this concept hold in the case of the chromomeres that contain the histone genes? Developmental variants of the histone proteins exist in sea urchins, and there is evidence for multiple forms of the *D. melanogaster* H1 protein (see Newrock and Cohen, this volume). Perhaps each chromomere in the 39DE region contains a developmentally specific set of histone genes that is coordinately activated by the local unfolding of the chromatin of that band. In such a case, we might expect to find all members of each type of variant repeat clustered in a single tandem array bounded on both sides by blocks of nonhistone repeat unit sequences. This macroorganization could then allow the independent regulation of expression of the repeat units containing different histone variants. By an approach combining cloning and cytogenetic techniques, we hope to be able to test these hypotheses.

**Acknowledgments**

This work was supported by research grants from the National Science Foundation and the National Institutes of Health. R. P. L. is a trainee in the Medical Scientists' Training Program of the U. S. Public Health Service. M. L. G. and R. W. K. were predoctoral fellows of the National Science Foundation and are currently supported by a training grant from the National Institutes of Health.

## REFERENCES

ALWINE, J. C., D. J. KEMP, and G. R. STARK. 1977. A method for the detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci.* **74**: 5350.

ARCECI, R. J., D. R. SENGER, and P. R. GROSS. 1976. The programmed switch in lysine-rich histone synthesis at gastrulation. *Cell* **9**: 171.

BEERMANN, W. 1972. Chromosomes and genes. *Results Probl. Cell Differ.* **4**: 1.

COHEN, L. H., K. N. NEWROCK, and A. ZWEIDLER. 1975. Stage-specific switches in histone synthesis during embryogenesis of the sea urchin. *Science* **190**: 994.

COHN, R., J. C. LOWRY, and L. H. KEDES. 1976. Histone genes of the sea urchin *(S. purpuratus)* cloned in *E. coli:* Order, polarity and strandedness of the five histone-coding and spacer regions. *Cell* **9**: 147.

DICKERSON, R. E. 1971. The structure of cytochrome C and the rates of molecular evolution. *J. Mol. Evol.* **1**: 26.

FARQUHAR, M. N. and B. J. MCCARTHY. 1973. Evolutionary stability of the histone genes of sea urchins. *Biochemistry* **12**: 4113.

KARP, R. and D. S. HOGNESS. 1976. Isolation and mapping of histone genes contained in cloned segments of *Drosophila* DNA. *Fed. Proc.* **35**: 1623.

KEDES, L. H. 1976. Histone messengers and histone genes. *Cell* **8**: 321.

KEDES, L. H. and M. L. BIRNSTIEL. 1971. Reiteration and clustering of DNA sequences complementary to histone messenger RNA. *Nat. New Biol.* **230**: 165.

KEDES, L. H., P. R. GROSS, G. COGNETTI, and A. L. HUNTER. 1969. Synthesis of nuclear and chromosomal proteins on light polyribosomes during cleavage in sea urchin embryo. *J. Mol. Biol.* **45**: 337.

KORNBERG, R. D. 1974. Chromatin structure: A repeating unit of histones and DNA. *Science* **184**: 868.

NOYES, B. E. and G. R. STARK. 1975. Nucleic acid hybridization using DNA covalently coupled to cellulose. *Cell* **5**: 301.

PARDUE, M. L. 1975. Repeated DNA sequences in the chromosomes of higher organisms. *Genetics* **79**: 159.

PERRY, R. P. and D. E. KELLEY. 1973. Messenger RNA turnover in mouse L cells. *J. Mol. Biol.* **79**: 681.

PORTMAN, R., W. SCHAFFNER, and M. BIRNSTIEL. 1976. Partial denaturation of cloned histone DNA from the sea urchin *Psammechinus miliaris. Nature* **264**: 31.

ROBBINS, E. and T. W. BORUN. 1967. The cytoplasmic synthesis of histones in HeLa cells and its temporal relationship to DNA replication. *Proc. Natl. Acad. Sci.* **57**: 409.

SOUTHERN, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**: 503.

TEMUSSI, P. A. 1975. Automatic comparison of the sequence of calf thymus histones. *J. Theor. Biol.* **50**: 25.

WENSINK, P. C., D. J. FINNEGAN, J. E. DONELSON, and D. S. HOGNESS. 1974. A system for mapping DNA sequences in the chromosomes of *Drosophila melanogaster. Cell* **3**: 315.