Positions of Sea Urchin (*Strongylocentrotus purpuratus*) Histone Genes Relative to Restriction Endonuclease Sites on the Chimeric Plasmids pSp2 and pSp17†

David S. Holmes,* Ronald H. Cohn, Laurence H. Kedes,‡ and Norman Davidson

ABSTRACT: The positions of the several sea urchin histone genes on the eukaryotic fragments of the chimeric plasmids pSp2 and pSp17 have been mapped relative to the *Eco* RI and *Hind* III restriction endonuclease sites on the plasmids. Two principal mapping methods using the electron microscope have been used: (a) the R-loop procedure and a new modification thereof to map the genes on duplex DNA; (b) the gene 32-ethidium bromide technique to visualize RNA-DNA hybrids on single strands of DNA. It is known that there are two histone genes, H3 and H2A, on pSp17. There are two *Eco* RI sites at the two junctions of the procaryotic segment with the eucaryotic segment on the plasmid. We show, by an electron microscope method, that for H2A, with a length of 0.52 kilobases (kb), one end of the gene is situated 0.02 to 0.03 kb from one RI site, and that there is a *Hind* III site within this gene at about 0.13 kb from the end proximal to the RI site. The distance from the other RI site to the proximal end of the other gene, H3, is about 0.19 kb. There are three histone genes, H2B, H1, and H4, on pSp2. The H2B gene is situated about 0.42 kb from the RI site closest to the procaryotic *Hind* III site, whereas the H1 gene maps at 0.76 kb from the other RI site of this plasmid. The H4 gene lies between H2B and H1. The measured lengths of genes and spacers agree with the values previously observed in circular molecules. When double-stranded plasmid DNA is incubated with histone mRNA in a high-formamide solvent, both the rate and the extent of R-loop formation increase as the incubation temperature is raised up to a temperature just below that at which strand dissociation of the duplex DNA occurs.

The chimeric plasmids pSp2 and pSp17 contain histone genes of the sea urchin *Strongylocentrotus purpuratus* inserted into the *Eco* RI restriction site of the prokaryotic vector pSC 101 (Kedes et al., 1975a). The eukaryotic components of these plasmids contain three and two sites, respectively, that hybridize with total histone mRNA. These two components represent adjacent segments and together are one histone gene repeat unit of *S. purpuratus* (Kedes et al., 1975b). In a preceding paper (Wu et al., 1976), the relative positions of genes and spacers on the plasmids were determined by observing, in the electron microscope, the relative positions of RNA-DNA hybrid regions on single strands of intact plasmid DNA using the gene 32-ethidium bromide method for distinguishing hybrid regions (Wu and Davidson, 1975). In addition, the relative order of the genes on the entire repeat unit of sea urchin DNA and on the plasmids, and the locations of the several spacer regions have been determined by restriction endonuclease hybridization studies (Kedes et al., 1975b; Cohn et al., 1976).

In the present paper, we have extended these observations by making quantitative assignments of the positions of the coding and spacer regions relative to the *Eco* RI and *Hind* III restriction endonuclease sites on the plasmids. For this purpose, we have used both the gene 32-ethidium bromide procedure and the newly developed R-loop method of mapping (Thomas et al., 1976). The latter method makes it possible to map RNA coding regions on a duplex DNA strand. R loops are formed by incubation of RNA strands with the duplex DNA containing a sequence complementary to the RNA in a high formamide solvent under conditions very close to the strand-dissociation temperature of the DNA. This process is illustrated in Figure 1a.

In addition, R loops have been formed by the alternate procedure depicted in Figure 1b, in which mRNA is hybridized to denatured single-strand DNA under conditions of high temperature and high formamide that permit formation of RNA-DNA hybrids but suppress DNA-DNA reassociation. Subsequently, the incubation temperature is lowered allowing DNA-DNA reassociation to occur (J. Casey and N. Davidson, personal communication). In general, DNA-DNA reassociation does not displace the RNA-DNA hybrids. Thus, the reassociated DNA contains R loops (N. D. Hershey, personal communication).

The scope, limitations, and usefulness of the R-loop method for gene mapping and gene enrichment experiments are not yet fully explored. We have, accordingly, made a study of the factors affecting the rate and extent of R-loop formation in our system.

Materials and Methods

Closed circular plasmid DNA and sea urchin histone mRNA were prepared as described previously (Kedes et al., 1975b). The mRNA was stored at −70 °C in 0.01 M sodium acetate, pH 5.5, and the DNA at 4 °C in 0.01 M Tris–0.01 M EDTA, pH 9.2.[1]

Digestion to completion of closed circular plasmid DNA with *Hind* III was carried out as follows. Twenty-five microliters of plasmid DNA at about 8.0 μg/mL was incubated at

† From the Department of Chemistry, California Institute of Technology, Pasadena, California 91125 (D.S.H. and N.D.) and the Stanford University School of Medicine and Veterans Administration Hospital, Palo Alto, California 94304 (R.H.C. and L.H.K.). *Received August 10, 1976.* Contribution No. 5344 from the Department of Chemistry, California Institute of Technology. This research has been supported by grants from the National Institutes of Health, the Damon Runyon-Walter Winchell Cancer Fund (D.S.H.), and the Veterans Administration.

* Present address: Department of Biological Sciences, State University of New York, Albany, N.Y. 12222.

‡ Investigator of the Howard Hughes Medical Institute.

37 °C for 2 h in 6 mM $MgCl_2$–8 mM Tris, pH 7.4–60 mM NaCl and 8 units of *Hind* III (New England Biolabs). The reaction was quenched by addition of 5.0 μL of 0.1 M EDTA–0.2 M Tris, pH 9.2, and stored at −70 °C. This resulted in a uniform population of 98% linear molecules, as judged by electron microscopy. Partial digestion of closed circular plasmid DNA was carried out in 50 μL of the same buffer at 37 °C, with 13 μg/mL DNA and 0.5 unit of *Hind* III for 10 min. The reaction was quenched as above and stored at −70 °C, resulting in 74% linear molecules.

Eco RI digestion of pSp2 or pSp17 was carried out in 50 mM NaCl–100 mM Tris, pH 7.8–5 mM $MgCl_2$ with 1 unit of *Eco* RI (gift of H. Boyer) per μg of DNA at 37 °C for 30 min.

R-Loop Formation. R-loop formation using double-stranded DNA was carried out in TFN (70% formamide (Matheson, Coleman, and Bell Co.)–0.5 M NaCl–0.1 M Tris, pH 8.3–25 mM EDTA). Typically, 25 μL of plasmid DNA at 2 μg/mL and mRNA at 2 μg/mL was dialyzed against the above buffer. Two microliter aliquots were sealed in 5-μL micropipettes and incubated at 58 °C for 3 h (except where stated). The incubation was terminated by pipetting into 50 μL of 50% formamide–0.08 M Tris, pH 8.6–8 mM EDTA at room temperature and prepared for electron microscopy by the formamide-spreading procedure of Davis et al. (1971).

Starting with single-stranded DNA, R loops were formed as follows: mRNA (28 μg/mL) was incubated with single-stranded DNA (1 μg/mL) in 2× SSC–80% formamide at 50 °C for 1 h, followed by incubation for 1 h at 25 °C. The material was prepared for electron microscopy as described above.

Gene 32–EtdBr Spreading Procedure. Hybrids were prepared for electron microscopy by a modified procedure of Wu and Davidson (1975). One microliter of *Hind* III treated pSp2 or pSp17 DNA at about 13 μg/mL was denatured by the addition of 1 μL of 0.4 N NaOH and incubation at 37 °C for 10 min, followed by neutralization at 4 °C with 1 μL of 1.0 N Tris-HCl. After the addition of 7 μL of mRNA at 56 μg/mL, the solution was dialyzed into 2× SSC–80% formamide at 4 °C. After allowing for a small volume change, the final DNA and RNA concentrations were about 1 μg/mL and 28 μg/mL, respectively. Hybridization was carried out at 50 °C for 1 h and terminated by plunging into dry ice–ethanol, followed by dialysis at 4 °C against 0.01 M phosphate buffer, pH 6.8. The volume change was minimized by using a modified microdialysis apparatus (D. S. Holmes, manuscript in preparation). Five microliters of the dialysate was added to 13 μL of 0.01 M phosphate buffer, pH 6.8, and 2 μL of gene-32 solution at 500 μg/mL and incubated at 37 °C for 5 min. After the addition of 2 μL of 2% gluteraldehyde (stored under argon), the incubation was continued at 37 °C for a further 10 min. The solution was then processed for electron microscopy in one of two ways: (1) DNA–RNA hybrids made from *Hind* III treated pSp2 and pSp17 were passed through a column of Sepharose 2B (ca. 300 μL) in 0.01 M phosphate buffer, pH 6.8, at room temperature and 1-drop (ca. 20 μL) fractions were collected. Ten microliters was taken from the pooled excluded fractions (fractions 7–8) and added to 4 μL of 1 M Tricine, pH 7.5, 3 μL of EtdBr at 1.13 mg/mL, and 23 μL of 0.01 M phosphate buffer, pH 6.8. The DNA was picked up by the drop procedure and grids were prepared for electron microscopy as described by Wu and Davidson)1975), except that the times of the washes in water, 70% ethanol, and 90% ethanol were reduced to 15 s each, and the grids were shadowed with platinum-palladium. We found that the introduction of the step involving chromatography on Sepharose 2B reduced the background on the grids, presumably by removal of the majority of unbound gene 32 and excess mRNA. (2) DNA–RNA hybrids made from *Eco* RI treated pSp2 or pSp17 DNA were retained in the included volume of the Sepharose 2B column, and were poorly resolved from histone mRNA. Therefore, after hybridization as described above, excess mRNA was removed by chromatography on hydroxylapatite. This step is not essential but does improve the visualization of the hybrids. Ten microliters of the 80% formamide–2× SSC solution containing the hybrids was diluted with 90 μL of 0.01 M sodium phosphate buffer, pH 6.8, and passed at room temperature over a 20-μL column of hydroxylapatite previously equilibrated in the phosphate buffer. Single-stranded nucleic acid was removed by 6 column volumes of 0.2 M sodium phosphate buffer, pH 6.8, and double-stranded nucleic acid by 4 column volumes of 0.5 M sodium phosphate buffer, pH 6.8. After dialysis against 0.01 M sodium phosphate buffer, pH 6.8, the hybrids were prepared for electron microscopy as described above but without the Sepharose 2B chromatography step. About 50–70% of the hybrids were recovered at this step, as judged by electron microscopy using DNA of known concentration as a standard.

Electron micrographs were contrast enhanced by printing on Agfa film BEH 1 No. 6.

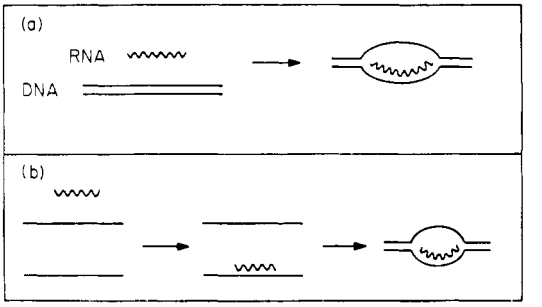

FIGURE 1: Schematic representation of sequence of events in formation of R loops on DNA that is initially either (a) double stranded or (b) single stranded. (a) A sample of double-stranded DNA plus RNA complementary to a segment of one of the DNA strands is incubated in high formamide at high temperature. Local denaturation of the DNA permits RNA to hybridize with one of the denatured DNA strands forming a stable loop with one side composed of a DNA–RNA hybrid and the other composed of the displaced DNA strand. (b) The DNA is denatured followed by the addition of RNA. Incubation in high formamide at high temperature suppresses the formation of DNA–DNA reassociation but permits DNA–RNA hybridization. Subsequent lowering of the temperature in the high formamide allows DNA–DNA reassociation to occur. The RNA–DNA hybrid is generally not displaced, thus forming an R loop.

Results and Discussion

A Summary Map. In order to present the rather complex data in a readily understandable way, we first present the final interpretation of all the measurements in the map of Figure 2. As shown in the circular maps in the upper part of the figure, both pSp17 and pSp2 consist of segments of sea urchin DNA bounded on each side by *Eco* RI sites and joined at these sites to the prokaryotic vector pSC 101. The sea urchin part of pSp17 contains the histone genes H2A and H3; that of pSp2 contains H2B, H4, and H1. (It has recently been shown that

[1] Abbreviations used are: TFN, 70% formamide, 0.1 M Tris, 0.5 M NaCl, 25 mM EDTA, pH 8.3; kb, kilobases, 1000 nucleotides or nucleotide pairs; EtdBr, ethidium bromide; r_0t, RNA concentration × time (moles of nucleotides × second/liter); EM, electron microscopy; n, number of molecules measured; Tris, 2-amino-2-hydroxymethyl-1,3-propanediol; EDTA, (ethylenedinitrilo)tetraacetic acid; SSC, standard saline citrate.

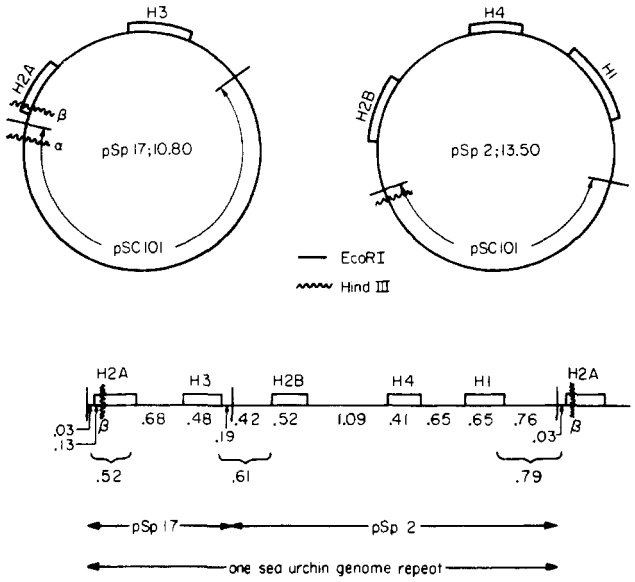

FIGURE 2: Positions and lengths of the sea urchin histone genes on the plasmids pSp17 and pSp2 and on one repeat unit in the sea urchin genome. The relative position of the pSp2 and pSp17 fragments on the sea urchin genome was determined by restriction endonuclease mapping (Cohn et al., 1976); the quantitative lengths of genes and spacers are based on the present and previous (Wu et al., 1976) EM studies. All lengths are in kilobases (kb). The standard deviations of the various length features in the lower figure are 0.03 ± 0.02 kb, 0.13 ± 0.03, 0.52 ± 0.04, 0.68 ± 0.10, 0.48 ± 0.09, 0.19 ± 0.05, 0.42 ± 0.15, 0.52 ± 0.07, 1.09 ± 0.08, 0.41 ± 0.08, 0.65 ± 0.07, 0.65 ± 0.11, 0.76 ± 0.09.

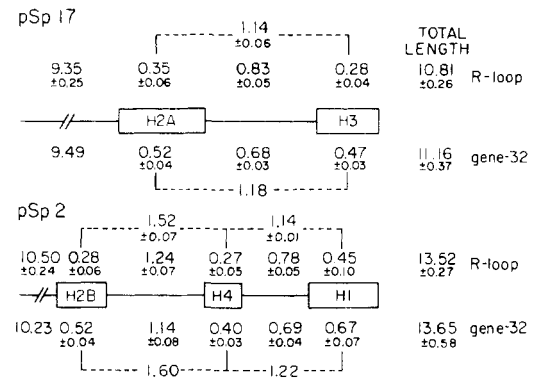

FIGURE 3: Summary of length measurements in kb, of the R loops, the spacings between the R loops, and the center to center distances between the R loops on open circular pSp17 and pSp2. The data is derived from circular molecules generally similar in appearance to the linears shown in Figure 4. Measurements derived from Figure 3 of Wu et al. (1976) using the gene 32-EtdBr method are included for comparison. The length of an R loop is taken as the average of the two strands. Assignment of coding specificity was inferred from R-loop lengths as described in the text.

the genes previously denoted on the basis of electrophoretic mobility, as B1, B2-3, and B4, in the previous EM mapping study (Wu et al., 1976) actually code for histones H3, H2A, and H2B (Sures et al., 1977). These new assignments are shown in the figures.)

There is a *Hind* III site in pSC 101 close to the Eco RI site; the distance between them has been estimated as 0.03 kb (H. Boyer, personal communication). The restriction endonuclease study (Cohn et al., 1976) and the present study both show that there is a *Hind* III site on pSp17 within the histone gene, which is closer to the prokaryotic *Hind* III site. H3 and H2A are almost identical in length. The restriction endonuclease map study shows that the gene that contains the *Hind* III site is H2A. The quantitative assignment of position of the site can then be made by electron microscopy. The two *Hind* III sites are denoted α and β, respectively, in the figure.

The relative positions of the genes and of the *Eco* RI sites on a single repeat in the sea urchin genome were determined by restriction endonuclease mapping (Cohn et al., 1976). The present determination of distances to the RI sites permits construction of the quantitative map in the lower part of the figure. This map shows the lengths of the spacers and of the genes, as well as their relative positions, on one repeat unit of the genome.

Length Measurements. Using double-stranded circular ϕX174 as an internal standard (length taken as 5.20 kb), the duplex lengths of open circular pSC101, pSp2, and pSp17 in standard formamide cytochrome *c* spreads were measured as 8.96 ± 0.30 kb (n = 19), 13.50 ± 0.27 kb (n = 55), and 10.81 ± 0.25 kb (n = 45), respectively. Therefore, the lengths of the inserted eukaryotic fragments in pSp2 and pSp17 are calculated to be 4.54 and 1.85 kb, respectively. The chimeric plasmids were digested to completion with *Eco* RI restriction endonuclease. The lengths of the two eukaryotic fragments and of the pSC101 vector were directly measured as 4.47 ± 0.19 (n = 50), 1.90 ± 0.09 (n = 50), and 9.00 ± 0.16 kb (n = 50) in agreement with the measurements above. These results agree, within experimental error, with the lengths of the single-strand circles measured by the gene 32-EtdBr method (Wu et al., 1976).

R-Loops and Gene Mapping. In a later section, we present in detail our results on the rate and extent of R-loop formation. The essential point here is that there are conditions for incubating linear or open circular pSp2 and pSp17 DNA with histone mRNA, which give a high frequency of formation of R loops. In order to study the usefulness of this technique of gene mapping, we first carried out studies on the distribution of the genes on circular plasmids to compare with the gene 32 mapping on single-stranded circles (Wu et al., 1976).

R-loop formation was carried out by incubation of total *S. purpuratus* histone mRNA with a mixture of open circular pSp2 and pSp17 DNA in TFN (70% formamide-0.5 M NaCl-0.1 M Tris, pH 8.3-25 mM EDTA) at 56 °C for 130 min. These conditions are close to optimal for R-loop formation in this system. The molecules were then diluted into a standard 50% formamide spreading solution for mounting for electron microscopy (Davis et al. 1971). Many molecules were seen with two and three R loops on pSp17 and pSp2, respectively. Some molecules were observed with fewer R loops than this, but only $^{3}/_{459}$ had R loops that did not conform to the patterns described below and only $^{3}/_{700}$ had detectable loops when no mRNA was included in the reaction. Further details of the statistics of R-loop frequency are given in a later section.

In some cases, it is possible to discriminate between the duplex and the single strand of an R loop. Very occasionally a small segment, presumably of single-stranded RNA, could be detected extending from one or other of the forks. Figure 3 summarizes the data on the lengths and spacings between the several R loops on circular molecules. The lengths and center-to-center spacings of the RNA-DNA hybrids observed by the gene 32-EtdBr procedure on circular DNA (Wu et al., 1976) have been included in this figure. A comparison shows that, in this system, any one R loop has a fairly homogeneous size distribution, but is characteristically only 50-70% of the full length of the gene, as judged by the gene 32-EtdBr method. However, the relative positions of the R loops are reproducible. The center-to-center distances agree well with those observed by the gene 32-EtdBr procedure. These em-

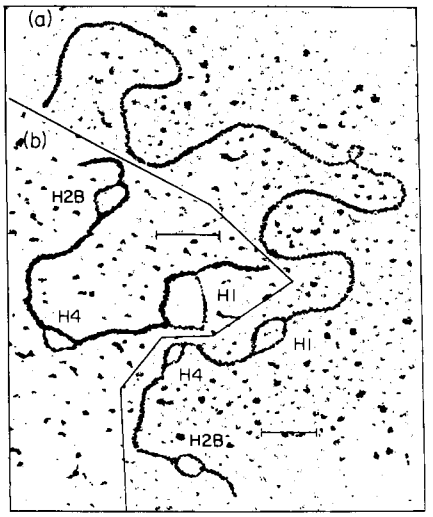

FIGURE 4: Electron micrographs of R loops formed on (a) pSp2 digested to completion with *Hind* III and (b) the eukaryotic fragment of pSp2 produced by digestion of pSp2 to completion with *Eco* RI. The bar represents 0.5 kb. R loops were formed in a by the method represented in Figure 1a. RNA (2 μg/mL) was incubated with DNA (2 μg/mL) in TFN at 58 °C for 6 h. R loops in b were formed by the method represented in Figure 1b. RNA (28 μg/mL) was incubated with DNA (1 μg/mL) in 2× SSC–80% formamide at 50 °C for 1 h, followed by incubation for 1 h at 25 °C. The DNA in both cases was then prepared for electron microscopy by the formamide cytochrome *c* spreading technique of Davis et al. (1971).

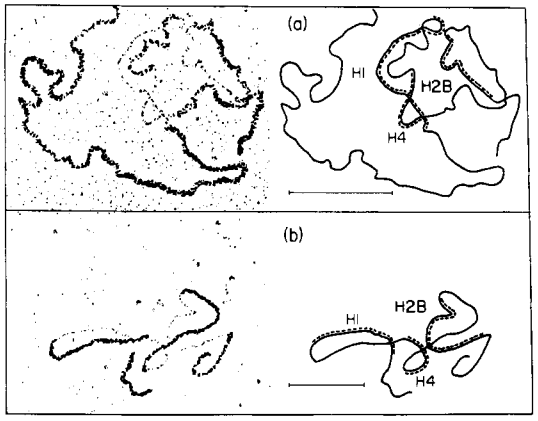

FIGURE 5: Electron micrographs of histone–mRNA hybrids on pSp2 visualized by the gene 32–EtdBr procedure. pSp2 digested to completion with (a) *Hind* III and (b) *Eco* RI. The bar represents 0.5 kb. RNA (28 μg/mL) was incubated with DNA (1 μg/mL) in 2× SSC–80% formamide at 50 °C for 1 h and prepared for electron microscopy, as described under Materials and Methods.

pirical observations are used in interpreting the data reported below on the positions of the R loops relative to restriction endonuclease sites.

We decided that it was useful in mapping the genes relative to restriction endonuclease sites to apply both the R-loop method and the gene 32–EtdBr method. The latter method has the advantage that the genes are of full length. However, it is more difficult because there is a certain amount of intrastrand tangling and crossing over which sometimes makes it difficult to observe relative positions on long strands. These problems are less serious for the R-loop mapping method with duplex DNA. On the other hand, as reported below, it is not always possible to form R loops by the direct procedure of Figure 1a; some duplex segments denature under the conditions needed for R-loop formation. The identification of a restriction endonuclease site and the statement that there are no other such sites on a given DNA molecule can be made with much greater confidence when the results of the two methods are compared.

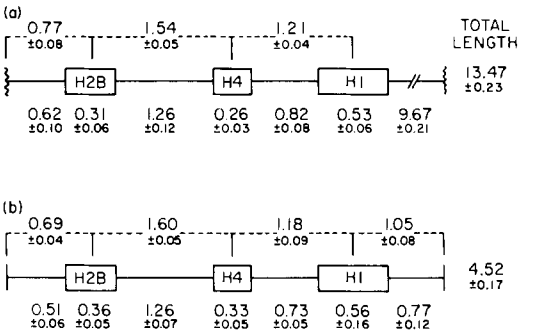

FIGURE 6: Length measurements (in kb) of R loops and position of (a) the *Hind* III site and (b) the *Eco* RI sites relative to the R loops in pSp2. The data in (a) was derived from electron micrographs of molecules, such as the one shown in Figure 4a. The data in b was derived from micrographs, such as the one shown in Figure 4b. (⟆) represents a *Hind* III site and (|) represents an Eco RI site. Assignment of coding specificity was inferred from R-loop lengths as described in the text.

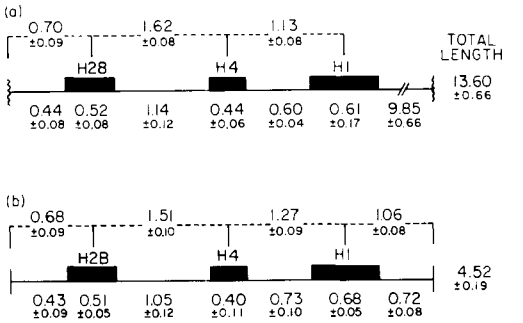

FIGURE 7: Summary of data on the lengths of the DNA–RNA hybrids, their spacings, and positions relative to (a) the *Hind* III ($n = 44$) and (b) the *Eco* RI sites ($n = 50$) on pSp2. Data based on molecules such as those shown in Figure 5a and b. (⟅) represents a *Hind* III site and (|) represents an *Eco* RI site.

The Eco RI and Hind III Sites on pSp2. pSp2 was digested to completion with *Hind* III. Duplex DNA was hybridized to histone mRNA for R-loop formation and single-stranded DNA was hybridized to mRNA and mounted by the gene 32 method. Electron micrographs are shown in Figures 4a and 5a, respectively. These digestions give full length linear molecules, consistent with there being a unique *Hind* III site within pSC 101 (Cohn et al., 1976).

Statistical summaries of the data are presented in Figs. 6a and 7a. In the R-loop method, the genes are again 50 to 60% of full length, but the pattern of spacers and gene lengths is such that an unambiguous assignment is possible. Both methods show that the H2B gene is the one that is proximal to the *Hind* III site.

Eco RI digestion of pSp2 liberates the sea urchin DNA fragment of length 4.52 ± 0.19 kb, consistent with the experiment reported above. The histone genes were mapped on this fragment using the gene 32 method. A micrograph is shown in Figure 5b, a summary of the data in Figure 7b.

When the duplex RI digest was incubated with mRNA under conditions optimal for R-loop formation, the eukaryotic fragment denatured but the pSC 101 component did not. The significance of this observation is discussed later. R loops were successfully formed on the RI fragment by the procedure illustrated in Figure 1b. The DNA was denatured and incubated

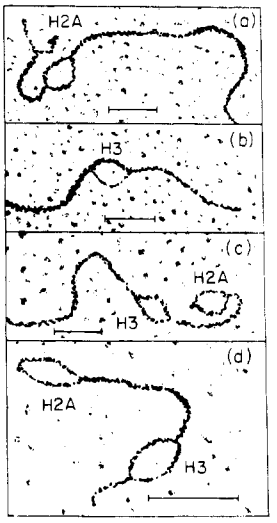

FIGURE 8: Electron micrographs of R-loops on pSp17. The bar represents 0.5 kb. (a) *Hind* III digestion to completion with the *Hind* III cut at the β position within the H2A gene. The H2A site is forked, while the H3 site has an R loop. (b) *Hind* III digestion to completion with the *Hind* III cut at the α position. The H2A site is not occupied, while the H3 site has an R loop. (c) Partial digestion with *Hind* III with the *Hind* III cut at the α position. Both the H3 and H2A sites have R loops. (d) *Eco* RI digestion to completion. Both the H3 and H2A sites have R loops. (a–c) R loops formed as shown in Figure 1a; (d) R loops formed as shown in Figure 1b. Note that the molecules in a, b, and c have been artificially truncated at one end each by the edges of the photographs.

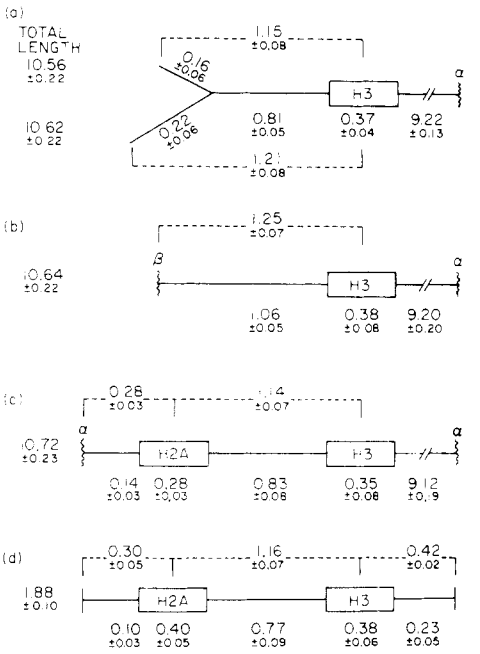

FIGURE 9: Summary of the statistics describing the size of the R loops and their positions relative to the *Eco* RI and *Hind* III sites of pSp17; lengths in kb. The data a to d are based on molecules such as those shown in Figure 8a–d, respectively. ({) represents a *Hind* III site and (|) represents an *Eco* RI site. Number of molecules measured: (a) $n = 53$, (b) $n = 46$, (c) $n = 32$, (d) $n = 46$.

with mRNA in a high-formamide solvent at a high temperature, as described under Materials and Methods. These conditions permit RNA–DNA hybridization but prevent almost completely DNA–DNA reassociation. After a suitable r_0t is reached, the temperature of incubation was lowered, allowing DNA–DNA reassociation to occur. Under these conditions, over 90% of the RNA–DNA hybrids formed stable R loops. An example of an electron micrograph of such R loops is shown in Figure 4b, and the measurements are summarized in Figure 6b. It may be seen that the R loops formed by this method are longer than those formed by direct incubation with duplex DNA, although still only 70–80% of full length.

Both the gene 32 and the R-loop data show that the H2B gene is proximal to the RI site that is close to the pSC 101 *Hind* III site. We calculate that the spacing between the sites is either 0.08 ± 0.06 (Figure 6) or 0.02 ± 0.09 kb (Figure 7). These numbers are based on the difference between large numbers and are not accurate, but the result that the two sites are close to each other is consistent with the direct measurement of fragment length by gel electrophoresis of 0.03 kb (H. Boyer, personal communication).

The Eco RI and Hind III Sites on pSp17. Both restriction endonuclease mapping (Cohn et al., 1976) and the present EM study show that there are two *Hind* III sites on pSp17, one within pSC101 and one within a gene on the eukaryotic fragment. The two sites are close to each other, and the restriction mapping shows that it is the H2A gene that contains the site. This situation is depicted in Figure 2.

We define the two sites as α and β and denote fragments by reading clockwise around the circular map (Figure 2). A complete digest will give the short fragment $\alpha\beta$ and the long fragment $\beta\alpha$. A partial digest will also give $\alpha\alpha$ and $\beta\beta$. R loops were formed on the DNA produced by a partial digestion with *Hind* III. Molecules, such as that illustrated in Figure 8c containing two R-loop regions, were observed and are presumed to be of the type $\alpha\alpha$. Statistical data for these molecules (Figure 9c) gives a distance of 0.28 ± 0.03 kb from the center of the R loop, corresponding to the H2A gene, to the α *Hind* III site.

For a molecule of the type $\beta\alpha$ (or $\beta\beta$), with a restriction cut within a gene, the corresponding R loop would presumably appear as a fork, possibly with a protruding single-stranded filament of RNA. We observe such molecules in complete digests, as illustrated in Figure 8a. A summary of the experimental data for molecules of this type is given in Figure 9a. Usually one side of the fork, presumably corresponding to the DNA–RNA hybrid, was longer than the other side, corresponding to the displaced DNA strand. However unambiguous discrimination between DNA–RNA hybrids and single strands of DNA could not be made with sufficient frequency to provide reliable statistics of the position of the β *Hind* III site by this method. Alternatively measurements were made on molecules, $\beta\alpha$, in which gene H3 is occupied by an R loop but gene H2A is not, and is, therefore, not forked. An example of such a molecule is shown in Figure 8b. The appropriate statistical data are presented in Figure 9b. The distance from the center of the H3 R loop to the β *Hind* III site is measured as 1.25 ± 0.07 kb. Circular pSp17 strands, subjected to R-loop formation, were included in this experiment as an internal standard for the center-to-center distance between the two genes H3 and H2A; this distance was measured as 1.14 ± 0.07 kb in agreement with the value reported in Figure 3.

Since the gene H2A has a length of 0.52 kb, as measured by the gene 32 procedure, and since the center of an R loop is at or close to the center of a gene (see below in Rate of R-Loop Formation), it is calculated that the β site is 0.15 kb from the end of the gene closest to the α site.

Molecules of the type $\beta\beta$, formed by partial digestion, might contain a fork at both ends, corresponding to the cleaved gene H2A. Instances of this were too rare ($^2/_{470}$) to provide reliable measurements. It is possible that the fork corresponding to the

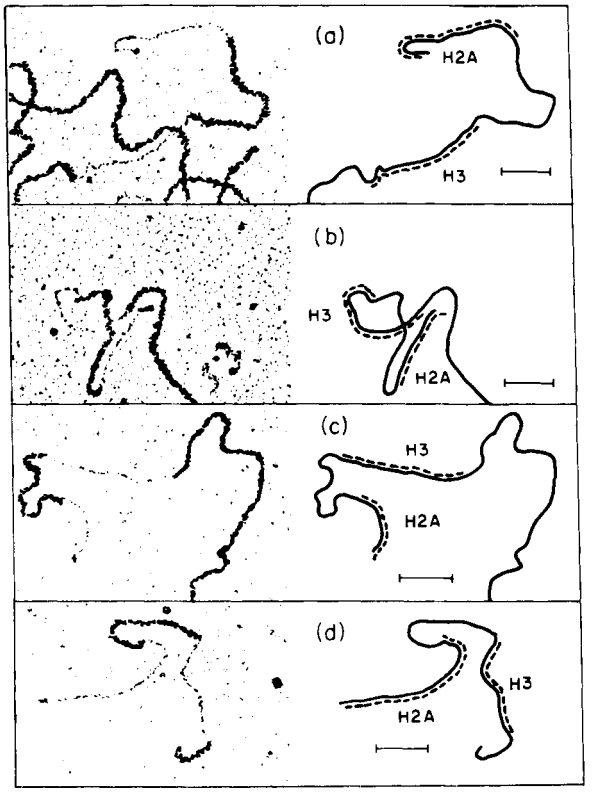

FIGURE 10: Electron micrographs of histone mRNA-DNA hybrids on single strands of pSp17 visualized by the gene 32-EtdBr procedure. The bar represents 0.5 kb. (a and b) represent structures formed by partial *Hind* III digestion with cleavage at the α and β *Hind* III sites, respectively, (c) complete digestion with *Hind* III with cleavage at the β *Hind* site, and (d) complete digestion with *Eco* RI. The conditions of hybridization are given in the legend to Figure 5.

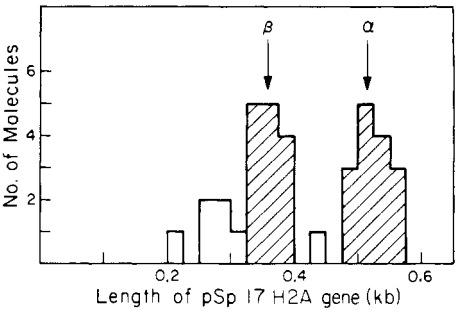

FIGURE 11: Histogram of the length of hybrids formed between histone mRNA and gene H2A on molecules of single-stranded pSp17 that have been partially digested with *Hind* III. Data are from molecules such as those shown in Figure 10a,b. The hatched areas represent data selected for computing the mean lengths reported in the text.

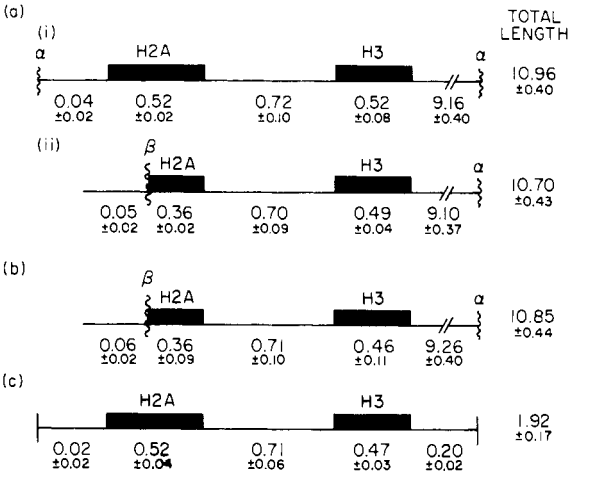

FIGURE 12: Summary of the statistics describing the lengths of histone genes and their positions relative to the *Hind* III and *Eco* RI sites of pSp17, as determined by the gene 32-EtdBr procedure. ( ) represents a *Hind* III site and (|) represents an *Eco* RI site. Length measurements of pSp17 (a) partially digested with *Hind* III with cleavage at (i) the α *Hind* III site or (ii) the β *Hind* III site. Data from molecules such as those shown in Figure 10a,b respectively; (b) digested to completion with *Hind* III with cleavage at the β *Hind* III site. Data from molecules such as that shown in Figure 10c; (c) digested to completion with *Eco* RI. Data from molecules such as that shown in Figure 10d. Conditions of hybridizations are described in the legend to Figure 5, and preparation for electron microscopy described under Materials and Methods.

shorter of the ends of the cleaved gene is too short to detect in the electron microscope, or possibly the RNA is displaced by branch migration of the DNA.

The observation of a fork, as in Figure 8a, is strong qualitative evidence for a restriction site within the gene. Further qualitative and quantitative evidence was obtained by gene 32-EtdBr studies. Linear molecules of pSp17 were generated by partial or complete digestion with *Hind* III or by complete digestion with *Eco* RI. Figure 10 shows electron micrographs of the structures formed on these molecules after hybridization with total histone mRNA followed by visualization in the electron microscope by the gene 32-EtdBr procedure.

Figure 11 shows a histogram of the lengths of gene H2A on a *Hind* III partial digest of pSp17. There are two predominant frequency classes, with means of 0.36 ± 0.02 and 0.52 ± 0.02 kb, respectively. We propose that the shorter structures represent hybrids of gene H2A that have been cleaved internally with *Hind* III at the β site and that the longer structures represent the full length H2A with cleavage at the external α *Hind* III site. Digestion to completion with *Hind* III produces linear molecules that have only the shorter hybrid at the gene H2A site, consistent with this interpretation. These measurements, together with the other parameters summarizing these various structures, are given in Figure 12.

The lengths of the two genes and the spacers between them are in good agreement with the equivalent structures on circular pSp17 (Wu et al., 1976). Wu et al. concluded that the two genes on pSp17 are slightly different in length and assigned values of 0.52 ± 0.04 and 0.47 ± 0.04 kb. While the experimental standard deviations are in some cases rather large, the data in Figure 12 support the conclusion that the two genes differ in length, and suggest that the H2A gene is larger than the H3 gene with lengths of 0.52 ± 0.03 and 0.47 ± 0.05 kb, respectively. If gene H2A has a length of 0.52 kb, and the duplex region observed when histone mRNA is hybridized to the gene cut at the β site has a length of 0.36 kb, we would expect a protruding RNA filament of length 0.16 kb, provided the RNA is full length. A small "knob" of single-strand material was observed to terminate the foreshortened gene. This knob measured only 50 ± 20 nucleotides in length. However, extensive measurements of defined short RNAs have not yet been made using the gene 32-EtdBr spreading procedure. Consequently, we do not consider that this measurement is necessarily reliable. The knob of single-stranded material terminating the full length gene H2A is presumably the flanking DNA between the end of the gene and *Hind* III site. Using the data of Figure 12, we calculate that there are about 200 nucleotides between the α and β *Hind* III sites.

By assuming that the center of an R loop is coincident with the center of a gene, the R-loop data give a value of 0.17 ± 0.12

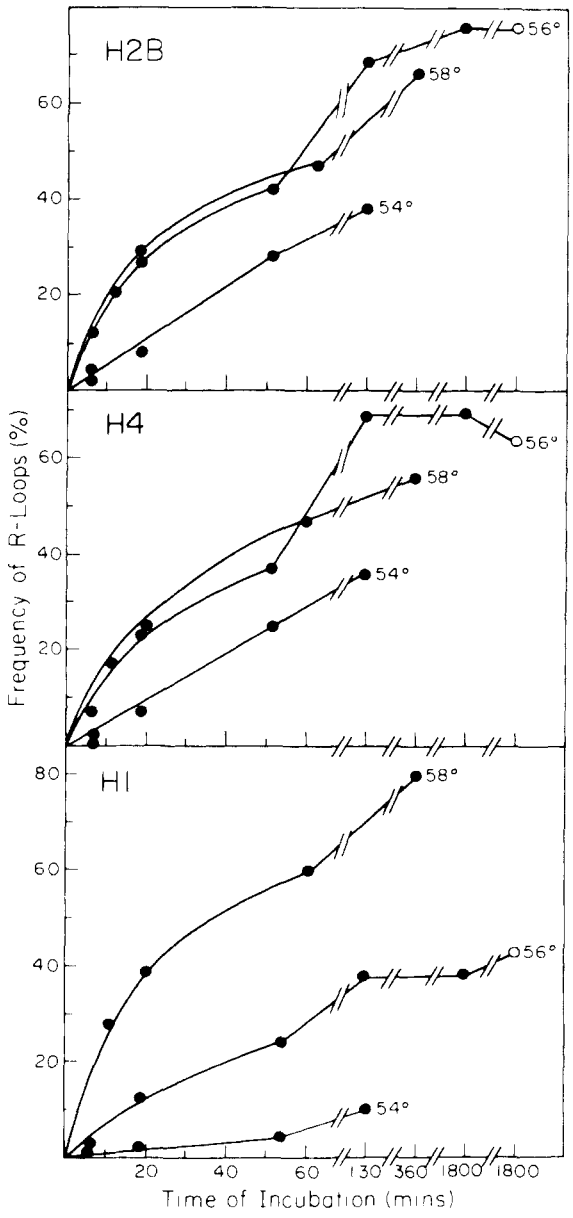

FIGURE 13: Frequency of R loops on pSp2 as a function of time and temperature of incubation. R loops were formed by the procedure depicted in Figure 1a. (●) 2 μg/mL or (○) 15 μg/mL of RNA were incubated with 2 μg/mL pSp2 DNA (digested to completion with *Hind* III) in TFN. DNA was prepared for electron microscopy by the formamide–cytochrome *c* technique of Davis et al. (1971).

kb for the distance between sites α and β. From the gene 32 data in Figure 12, we calculate a length of 0.20 ± 0.03 kb, thus supporting the assumption. Gel electrophoresis data give 0.23 kb (Cohn et al., 1976).

The gene 32 data in Figure 12 indicate that the proximal RI site is 0.02 kb from the α *Hind* III site. An approximate determination of the position of the RI site was also made by two R-loop procedures. R loops were produced on pSp17 either by partial restriction with *Eco* RI to generate molecules with only one cleaved *Eco* RI site or by complete digestion with *Eco* RI to liberate the 1.90 kb sea urchin fragment, denaturation, incubation with mRNA to form RNA–DNA hybrids, followed by DNA–DNA renaturation as depicted in Figure 1b. An example of the structure formed by the latter method is shown in Figure 8d and the statistical data from such electron micrographs is summarized in Figure 9d. Data provided by the other method is in good agreement with the assignments given in Figure 9d and is not reported here. This data indicates that the RI site is 0.30 kb from the center of the H2A gene or 0.04 kb from the end.

As already stated, a summary of all the plasmid mapping data is given in Figure 2. In this figure, we have used data derived from the various gene 32–EtdBr measurements described above to provide lengths for the histone genes and have used a number-weighted average derived from both the R-loop and gene 32–EtdBr measurements to locate the *Eco* RI and *Hind* III sites relative to the histone genes.

Rate of R-loop Formation. We have measured the rate of R-loop formation on relaxed (linear and open circular) and on supercoiled pSp2 and pSp17 DNA under different conditions in TFN (0.1 M Tris, pH 8.3–70% formamide–0.5 M NaCl–25 mM EDTA).

Measurements made by counting perceptible R loops in the electron microscope of complete *Hind* III digests are presented in Figure 13. Additional experiments were performed on open circular plasmids. The total number of R loops formed was scored, but, of course, the specific genes involved were not identified. These data, which are not presented in detail, are in agreement with the results presented in Figure 13. Overall, the data lead to the following conclusions.

(a) The rate and extent of R-loop formation, in general, increase as the temperature is raised from 54 to 58 °C. On prolonged incubation, 60 to 80% of all the genes formed R loops at 58 °C. At 59 °C, both pSp2 and pSp17 begin to denature. About half the molecules are denatured by incubation at 61 °C.

(b) The rates are different for different genes. In particular, the rate of formation of the R loop corresponding to the histone H1 gene is less than that of the other R loops on pSp2 at 54 and 56 °C but is greater at 58 °C.

(c) At lower temperatures (56 °C in Figure 13), the reactions do not go to completion on prolonged incubation, even at a very high RNA concentration.

(d) R loops were infrequently observed on supercoiled molecules, but, just as for relaxed DNA, the rate was greater at 58 °C than at lower temperatures. There is some difficulty in identifying supercoiled molecules in the electron microscope; thus, we did not collect extensive data.

Due to the complexity of the system studied by us, it is not ideal for a quantitative investigation of the kinetics and mechanism of R-loop formation. Studies on a yeast rRNA gene in a λ DNA vector have been reported by Thomas et al. (1976). It is probable, as argued below, that the rate and extent of R-loop formation will be dependent on the base composition and the sequences of the genes involved and of the flanking sequences on the DNA duplexes, and will therefore be different for different cases. R-loop formation has potentially great practical usefulness for gene mapping and gene enrichment experiments. We therefore think that it is useful to report our empirical observations in some detail.

Since R loops can form, RNA–DNA hybrids must be more stable than DNA–DNA duplexes in high formamide solvents. We, as well as Thomas et al. (1976), observe that the rate of R-loop formation increases as the temperature is raised up to a temperature just below that where the DNA undergoes denaturation and strand dissociation. We propose, as have Thomas et al. (1976), that the first step in R-loop formation is local denaturation or breathing of all or part of the DNA sequence complementary to the RNA. The temperature at which the DNA undergoes complete strand dissociation will be the melting temperature of the one or several most G + C

rich segments of the total molecule. The probability of the gene region, or a portion thereof, opening up at a given temperature depends on its G + C content. Therefore, the optimum conditions for R-loop formation will depend on the base composition of the gene and on the base composition of the flanking sequences. R-loop formation may not be possible for a gene which has a higher G + C content than any other segment of the DNA molecule in which it is contained.

It has been estimated that the G + C content of the histone genes lies between 50 and 57% and that the spacers have low (30–40%) G + C content (Kedes and Birnstiel, 1971; Grunstein et al., 1976). The average base composition of the flanking pSC101 sequences is 50% (Cohen and Chang, 1973). We propose that there are some segment(s) within pSC101 which have a higher G + C content that stabilize the pSp2 and pSp17 plasmids for R-loop formation. This hypothesis is suggested by the observation that at 58 °C in TFN buffer, which is optimal for R-loop formation of the plasmids, the eukaryotic fragments (prepared by Eco RI digestion) are almost fully denatured.

The observation that the rate of R-loop formation for the H1 gene is relatively slow at low temperature suggests that this gene has a particularly high G + C content. On the other hand, the estimates from the amino acid composition of calf thymus H1 and from the measured base composition of H1 mRNA from the sea urchin *Lytechinus pictus* (Grunstein et al., 1976) are both in the range of 51 to 53%, which is no higher than the estimates for the other histone genes. Alternatively, H1 mRNA may have some particular secondary structure that must be denatured by the higher temperatures before it can form R loops efficiently. A third possibility is that the threading of H1 mRNA between the DNA strands at lower temperatures is restricted by virtue of its length, leading to structures that would not be scored as loops in the electron microscope.

In our experiments, the length of the R loops was about 50 to 70% of the full length of the genes. White and Hogness (1977) also observed R loops on a cloned segment of *Drosophila melanogaster* rDNA that were 80–90% of the full length of the ribosomal gene. In this case, tails of RNA were frequently observed protruding from the ends of the R loops. Protruding histone mRNA of a similar proportional length would be only about 100–150 nucleotides in length and would probably not be observed in the electron microscope, especially if it was distributed at both ends of the R loops. Since the measurement of the center of R loops to *Hind* III and RI restriction sites agrees within experimental error with the same measurement from the center of genes as measured by the gene 32 procedure, it is apparent that the center of the R loop is coincident with the center of the gene. Thus, it is probably that protruding RNA filaments, if present, are distributed at both ends of the R loops. This is not necessarily a general conclusion for all R loops. Displacement of RNA could reflect its base composition, which need not be the same at both the 5′ and 3′ ends. The position of a short R loop relative to the gene may vary from system to system.

Another explanation of short R loops is degradation of the RNA either prior to or during incubation. This hypothesis could not be tested because of the limited amount of RNA available. We think it is more probable however that there are effects not yet understood, related to the distribution of sequences within the genes, that prevent the growth of an R loop to full length.

The data in Figure 13 indicate that prolonged incubation of RNA with plasmid DNA at 54 and 56 °C does not lead to 100% R-loop formation. At 56 °C, raising the RNA concentration by a factor of 10 does not lead to greater than 43 and 60% saturation of H1 and H4 genes. The R loops that did form at 54 and 56 °C were not significantly shorter than those formed at 58 °C. These observations suggest that there is an alternative process, possible formation of very short RNA–DNA complexes that are inhibited from further winding, that make some of the genes unavailable for full R-loop formation at temperatures below the optimal temperature.

A supercoiled DNA of length 13.50 kb (pSp2) with a typical superhelix density of −0.07 can relax by unwinding of about 1000 nucleotides. Unwinding provides an additional thermodynamic driving force for limited R-loop formation for supercoiled DNA vis-a-vis relaxed DNA. We interpret the fact that R-loop formation for the histone genes on supercoiled pSp2 and pSp17 is, in fact, slower than for the relaxed plasmids as follows. In partially denaturing solvents such as that used for R-loop formation, a supercoil will unwind readily up to the point of relieving the superhelix density, in the example above for 7% of the base pairs, but the thermodynamic probability of further unwinding is less for a supercoil than for a relaxed DNA. Therefore, unless the genes that are going to participate in R-loop formation are included in the 7% of the sequences that are the mostly A + T rich and most readily denatured, they are less likely to breath and be available for hybridization. Therefore, R- or D-loop formation will occur readily for a supercoiled DNA incubated with a sample of single-stranded polynucleotides that represent the entire supercoiled genome, but, in general, will not occur readily for a specific sequence (Holloman et al., 1975; Liu and Wang, 1975).

Measurements by J. Casey (personal communication) on the rate of association of RNA with single strands of DNA in high formamide solvents suggest that the $r_0t_{1/2}$ for an RNA, with the complexity of an average histone gene (length and complexity ≈ 500 nucleotides) in an RNA excess reaction with single-stranded DNA in a buffer similar to TFN, is approximately 0.001 mol L/s. At 58 °C, the time for 50% R-loop formation on any one histone gene is about 30 min at an RNA concentration of 2 μg/mL (Figure 13). If the concentration of any one histone mRNA species is ⅕ of the total, the $r_0t_{1/2}$ per gene is calculated as 0.002. Since the histone RNA is not pure, the actual $r_0t_{1/2}$ may be less than this. Therefore, under optimal conditions the rate of R-loop formation may be approximately equal to the rate of association of RNA with single-stranded DNA.

Acknowledgments

We are especially indebted to Dr. James Casey for communicating his results on relative rates of RNA–DNA and DNA–DNA reactions in high-formamide solvents prior to publication, to Jean Lowry for technical assistance, and to Drs. N. D. Hershey and M. Wu for advice and counsel.

References

Cohen, S. Y., and Chang, A. C. Y. (1973), *Proc. Natl. Acad. Sci. U.S.A. 70,* 1293.

Cohn, R., Lowry, J. C., and Kedes, L. (1976), *Cell 9,* 147.

Davis, R. W., Simon, M. N., and Davidson, N. (1971), *Methods Enzymol. 21,* 413.

Grunstein, M., Schedl, P., and Kedes, L. H. (1976), *J. Mol. Biol. 104,* 351.

Holloman, W. K., Wiegard, R., Hoessli, C., and Radding, C. M. (1975), *Proc. Natl. Acad. Sci. U.S.A. 72,* 2394.

Kedes, L. H., and Birnstiel, M. L. (1971), *Nature (London) New Biol. 230,* 165.

Kedes, L. H., Chang, A. C. Y., Houseman, D., and Cohen, S.

N. (1975a), *Nature (London) 255,* 533.
Kedes, L. H., Cohn, R. H., Lowry, J. C., Chang, A. C. Y., and Cohen, S. N. (1975b), *Cell 6,* 359.
Liu, L. F., and Wang, J. C. (1975), *Biochim. Biophys. Acta 395,* 405.
Syres, I., Maxam, A., Cohn, R. H., and Kedes, L. H. (1977), *Cell 9,* 495.
Thomas, M., White, R., and Davis, R. (1976), *Proc. Natl. Acad. Sci. U.S.A. 73,* 2294.
White, R., and Hogness, D. (1977), *Cell* (in press).
Wu, M., and Davidson, N. (1975), *Proc. Natl. Acad. Sci. U.S.A. 72,* 4506.
Wu, M., Holmes, D. S., Davidson, N., Cohn, R., and Kedes, L. H. (1976), *Cell 9,* 163.

Novikoff Hepatoma Deoxyribonucleic Acid Polymerase. Identification of a Stimulatory Protein Bound to the β-Polymerase†

Dale W. Mosbaugh, David M. Stalker, Gregory S. Probst,‡ and Ralph R. Meyer*

ABSTRACT: The Novikoff hepatoma DNA polymerase-β sediments as a 7.3S form in crude extracts but during purification sediments as a 4.1S form (after diethylaminoethyl-Sephadex chromatography) or as a 3.3S form (after DNA-cellulose chromatography). If 0.25 M ammonium sulfate or 0.5 M NaCl is included in the sucrose gradients, the 7.3S form sediments at 3.3 S; after removal of the salt, it sediments again at 7.3 S, indicating the reversibility of the aggregation phenomenon. By careful adjustment of ionic strength in the gradient, four distinct and reproducible forms of the enzyme sedimenting at 7.3, 5.8, 4.1, and 3.3 S can be generated. The isoelectric point of the DNA polymerase also changes during purification; the 7.3S form has a p*I* of 7.5, while the 4.1S form isoelectrically focuses at a pH of 8.5. During DNA-cellulose chromatography, the Novikoff β-polymerase is separated from a stimulatory factor designated as Novikoff factor IV. Factor IV is a protein as shown by its sensitivity to protease and resistance to nucleases. It is responsible for converting the 3.3S enzyme to the 4.1S form since the 3.3S homogeneous DNA polymerase-β sediments at 4.1 S in the presence of factor IV. Factor IV confers stability to the polymerase in low ionic strength buffers as well as stability to heat denaturation. Factor IV has the ability to increase the activity of the 3.3S homogeneous polymerase by about fourfold.

The β class of DNA polymerases (EC 2.7.7.7) from higher eukaryotes is characterized, among other properties, by a relatively small size (see recent reviews by Loeb, 1974; Bollum, 1975; Weissbach, 1975). In mammalian cells, homogeneous β-polymerases obtained from calf thymus (Chang, 1973), human KB cells (Wang et al., 1974), Novikoff hepatoma (Stalker et al., 1976), and guinea pig liver (Kunkel, Tcheng, and Meyer, in preparation) have molecular weights of 32 000 to 45 000. In crude cellular extracts or partially purified preparations, β-polymerases have been shown, under low salt conditions, to sediment in sucrose gradients as a 6–8S complex of proteins or to filter through Sephadex gels as an aggregate form (Hecht, 1973a,b; Hecht and Davidson, 1973; Lazarus and Kitron, 1973; Probst, 1974; Wang et al., 1975). They can be converted to a 3–4 S or low-molecular-weight form if appropriate salt concentrations are included in the sucrose gradients or Sephadex buffers. This aggregate form may represent a native conformation of proteins involved in DNA metabolism in vivo, particularly in light of recent studies with prokaryotic systems (Geider and Kornberg, 1974; Schekman et al., 1975; Hendler et al., 1975) which indicate that DNA replication and repair involve multienzyme complexes. We have previously purified to homogeneity and characterized the DNA polymerase-β from the Novikoff hepatoma (Stalker et al., 1976; Mosbaugh et al., 1976) and have identified three proteins isolated from this tumor which markedly and specifically stimulate this enzyme (Probst et al., 1975). In order to determine whether these stimulatory proteins are part of a polymerase complex and to identify other proteins involved in DNA synthesis, we are attempting to resolve, purify, and characterize the individual proteins found in the 6–8S β-polymerase complex. In the present report we (a) describe the changes in sedimentation behavior and isoelectric pH of the Novikoff hepatoma DNA polymerase-β during purification; (b) identify a new stimulatory protein, factor IV; and (c) provide evidence that factor IV binds to the homogeneous β-polymerase and is probably one of the components of the 6–8S complex. An abstract of this work has been published (Stalker et al., 1975).

Experimental Procedure

Materials

All chemicals were of analytical or reagent grade. Unlabeled deoxyribonucleoside triphosphates, phenylmethyanesulfonyl fluoride, and calf thymus DNA were purchased from Sigma. Enzyme grade ammonium sulfate, BSA,[1] and [^{3}H]dTTP were obtained from Schwarz/Mann. New England Nuclear Corp.

† From the Department of Biological Sciences, University of Cincinnati, Cincinnati, Ohio 45221. *Received October 14, 1976.* Supported by a grant from the National Institutes of Health (CA 17723). This is paper 3 in this series; papers 1 and 2 are Stalker et al. (1976) and Mosbaugh et al. (1976), respectively.

‡ Present address: Department of Toxicology, Eli Lilly & Co., Inc., Indianapolis, Indiana.

[1] Abbreviations used are: BSA, bovine serum albumin; dNTP, unlabeled deoxyribonucleoside triphosphates; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetate; PMEG buffer, 0.02 M potassium phosphate at pH 7.5, 0.005 M β-mercaptoethanol, 0.001 M EDTA, and 10% (w/v) glycerol; PMG buffer, PMEG with EDTA omitted; Cl_3AcOH, trichloroacetic acid; TMEG, 0.02 M Tris-HCl at pH 8.0, 0.005 M β-mercaptoethanol, 0.001 M EDTA, 10% (w/v) glycerol; TME buffer, TMEG with glycerol omitted; BBOT, 2,5-bis[2-(5-*tert*-butylbenzoxazolyl)]thiophene.