
Assembly of an active chromatin structure during replication

Harold Weintraub

Hutchinson Cancer Center, 1124 Columbia Street, Seattle, WA 98104, USA

Received 18 June 1979

ABSTRACT

MSB cells were pulse labeled with ^3H -thymidine and the isolated nuclei digested with either staphylococcal nuclease (to about 40% acid solubility) or DNase I (to 15% acid solubility). The purified, nuclease resistant single-copy DNA was then hybridized to nuclear RNA (nRNA). The results of these experiments show that actively transcribed genes are assembled into nucleosome-like structures within 5-10 nucleosomes of the replication fork and that they also acquire a conformation characteristic of actively transcribed nucleosomes (ie, a DNase I sensitive structure) within 20 nucleosomes of the fork. Assuming DNA sequence specific interactions are required for establishing a DNase I sensitive conformation on active genes during each round of replication, our results indicate that a specific recognition event can occur very rapidly and very specifically in eukaryotic cells. The results are discussed in terms of the possible mechanisms responsible for propagating active, chromosomal conformations from mother cells to daughter cells.

INTRODUCTION

It is now very clear that almost all actively transcribed genes in higher eukaryotes are packaged into nucleosome-like structures as assayed by electron microscopy (1,2) as well as by their resistance to digestion by staphylococcal nuclease (3,4). However, nucleosomes associated with actively transcribed genes are conformationally distinct in that they are extremely sensitive to digestion by DNase I (5,6). Since the active nucleosome structure is cell type specific for those genes that are expressed in a tissue-specific way, it is assumed that the DNase I sensitive structure associated with these genes is dictated in some way by DNA sequence-specific interactions. The problem posed by such an assumption is that the cell replicates and hence, duplicates a DNase I sensitive structure at each division. If DNA sequence-specific interactions dictate the DNase I sensitive conformation, does replication then require a new recognition event for each transcription unit at each cell generation? Here I show that newly replicated DNA is rapidly assembled into an active DNase I sensitive conformation within 3 minutes of its synthesis. This

finding is discussed in terms of possible mechanisms responsible for transmitting a DNase I sensitive structure to daughter transcription units with each round of DNA replication.

RESULTS

Kinetics of assembly of an active chromatin conformation:

MSB cells (7) were pulse labeled for increasing periods of time with ³H thymidine; the nuclei were isolated, digested to 45% acid solubility with staph nuclease (Fig. 1A), or to 15% acid solubility with DNase I (Fig. 1B); and the purified, nuclease-resistant DNA hybridized to nuclear RNA (nRNA).

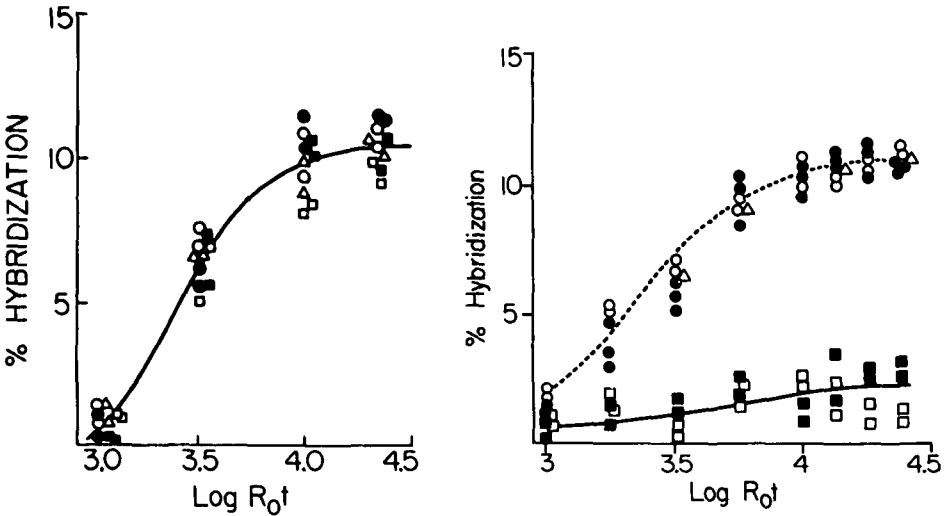


Figure 1: Kinetics of assembly of an active nucleosome conformation after chromosome replication:: MSB cells (7) were labeled with H3-thymidine (1Ci/ml) for 0.5 minute (closed circles), 1 minute (open triangles), 3 minutes (closed squares), or 15 minutes (open squares). Nuclei were digested to 45% acid solubility with staph nuclease (A) or to 15% acid solubility with DNase I (B) and the single-copy DNA isolated and hybridized (5) to an excess of nRNA (18 mgm/ml). Hybridization was measured as resistance to S₁ nuclease (5). Open circles show hybridization to total labeled DNA (10⁶ CPM/ugm). Specific activities of input DNA were estimated at about 2-5 x 10⁵ CPM/ugm. Reactions were in 10 ul with 4 x 10⁴ input CPM labeled DNA. Each curve represents the combined data from 3 different experiments. Single-copy DNA was purified by allowing the DNA to reassociate to a Cot of 100. The unhybridized DNA was isolated by passage over hydroxylapatite (Biorad) at 60°C in 0.12M phosphate buffer containing 0.1% SDS. This was repeated twice more and the DNA purified as previously described (5).

As controls, all preparations were shown to hybridize normally to DNA. Fig. 1 shows that by 1 minute, the newly-made DNA complementary to rRNA is in a "nucleosome-like" conformation as defined by its resistance to digestion by staph nuclease and its ability to hybridize normally to rRNA. At this point it is also resistant to DNase I suggesting that it is not yet assembled into a fully active nucleosome conformation. Nevertheless, by 3 minutes, the labeled DNA has acquired a DNase I sensitive conformation and no longer hybridizes to rRNA. Controls showed that this DNA was capable of hybridizing to a DNA driver. Given known rates of DNA chain elongation (7), it can be estimated that nascent DNA complementary to rRNA is assembled into DNase I sensitive chromatin very rapidly, some 10-20 nucleosomes behind the replication fork.

Assembly of an active chromatin conformation away from the replication fork

The previous experiment demonstrates that an active nucleosome conformation is usually assembled very close to the replication fork. It is experimentally possible to ask whether proximity to the fork is essential for the construction of an active nucleosome conformation. Previously, it was shown that when DNA replication occurs in the absence of protein synthesis, parental histone octamers segregate to one daughter DNA double helix while the other daughter DNA double helix is not packaged into nucleosomes (8-10), but appears as long stretches (ca. 70kb) of protein-free DNA (7) in the electron microscope. This type of histone segregation has been termed "conservative" segregation and has also been verified using density labeled histone octamers (and their multimers) (11). Recently, it was shown (12) that for the SV40 mini-chromosome the parental histones actually segregate with the "leading" strand for DNA replication (ie, the continuously replicated DNA strand). It was also shown that when cellular DNA is replicated in the presence of ^3H -thymidine and cycloheximide, the staph nuclease resistant labeled DNA (ie, the DNA that receives recycled parental histone octamers) hybridizes normally to DNA but fails to hybridize to rRNA. Thus, when cellular chromatin is synthesized in the absence of protein synthesis, parental nucleosomes segregate with the newly-replicated DNA strand that is anti-complementary (ie, has the same 5' 3' polarity) to nuclear RNA.

This experimentally induced situation allows us to ask whether an active nucleosome conformation can be assembled at a distance from the replication fork. Essentially, we ask whether the labelled, protein-free, cycloheximide DNA complementary to rRNA and sensitive to staph nuclease regains a staph nuclease resistant, nucleosome-like conformation after reversal of cyclohexi-

mid inhibition. If so, do these nucleosomes also acquire an "active" conformation as assayed by DNase I sensitivity? Previous work (7) with bulk chromatin has already shown that the free DNA generated by replication in the absence of protein synthesis is assembled into nucleosomes within 10 minutes of reversal. MSB cells were incubated with cycloheximide and ^3H -thymidine for 30 minutes followed by removal of ^3H -thymidine and reversal of cycloheximide for 10 minutes (the earliest time compatible with experimental manipulation). Nuclei were isolated; digested to a limit-digest (50% acid solubility) with staph nuclease; and the purified, resistant DNA hybridized to nRNA. Fig. 2A shows that this DNA hybridizes normally to nRNA. As a control, failure to

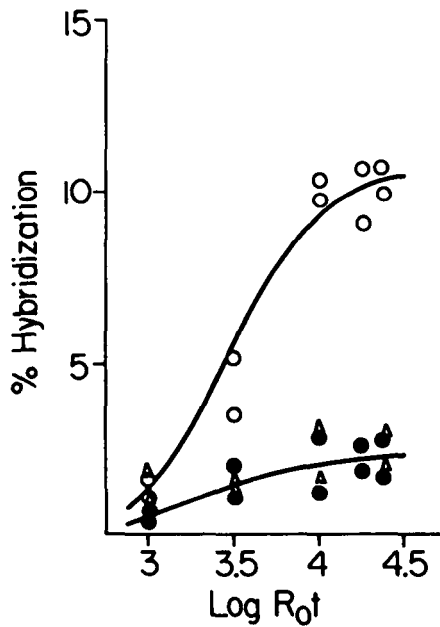


Figure 2: Assembly of an active nucleosome conformation at a distance from the replication fork: MSB cells were labeled with ^3H -thymidine (100u Ci/ml) for 30 minutes in the presence of cycloheximide (50 μm). The cells were then washed 3 times with medium containing 10^{-5}M cold thymidine and no cycloheximide. They were then incubated for 10 minutes. Nuclei were isolated and digested to 15% acid solubility with DNase I (closed circles) or to a limit digest (about 50% acid solubility) with staphylococcal nuclease (open circles). The purified, single-copy, nuclease-resistant DNA was then hybridized to an excess of nRNA (18 mgm/ml). Reactions were in a total volume of 10 μl with an input of 4×10^4 CPM of labeled DNA (about 2×10^5 CPM/ μgm). Triangles show hybridization to nRNA of staph nuclease resistant ^3H -DNA obtained from cells treated with ^3H -thymidine and cycloheximide for 30 minutes, but digested immediately without reversal of the cycloheximide block.

reverse the cycloheximide block resulted in the production of protected fragments that did not hybridize to nRNA (fig 2, triangles; see Seidman et al, (12)). Most important, the particles that become assembled into nucleosomes after reversal of cycloheximide also acquire a DNase I sensitive configuration within 10 minutes of reversal (Fig. 2B) since only 15% digestion of these nuclei with DNase I preferentially digests DNA sequences complementary to nRNA. Since the bulk of the free DNA made in the absence of protein synthesis is as long as 70kb (7), it follows from these experiments that an active nucleosome configuration can be assembled in vivo onto DNA as long as 70kb and need not necessarily be assembled at the replication fork or immediately behind the fork.

DISCUSSION

Establishment of a DNase-I sensitive structure precedes transcription

Previously it was pointed out (5) that each nucleosome in a transcription unit was preferentially sensitive to DNase I digestion. In order to explain how the information for DNase I sensitivity is distributed to each nucleosome, we postulated some type of DNA sequence-specific recognition event which was followed by a propagation event where the DNase I sensitive structure is propagated to each nucleosome in the transcription unit. Furthermore, it is likely that the starting and stopping signals for propagation are rather precise, since the domain for the DNase I sensitive structure has rather well-defined end-points (13).

It is possible that during gene activation, it is the first passage of an RNA polymerase molecule that leads to the propagation of the DNase I sensitive state down the length of the transcription unit; thus, the altered chromatin conformation would be viewed as a consequence and not necessarily a requirement for transcription. The kinetic experiments presented here (Fig. 1) showing that active genes acquire a DNase I sensitive conformation within 3 minutes of replication make this explanation unlikely. This follows from the fact that in using nRNA as a probe we are assaying the low abundant, high complexity class of transcribed DNA sequences. These sequences would be transcribed very rarely (several times per generation (14)) and on average, not within the 3 minute interval required to assemble a DNase I sensitive structure after replication. Thus, as an active gene is replicated, a DNase I sensitive conformation is assembled before an RNA polymerase ever transcribes that gene and it is therefore unlikely that RNA polymerase itself induces a stable DNase I sensitive conformation in newly-replicated active chromatin.

Since several nucleosomal alterations contribute to the DNase I sensitivity of active genes (15), these kinetic experiments measure only the completed structure and it is possible that some subset of structural elements are actually assembled much sooner than 3 minutes. Also these kinetic results monitor only the assembly of the labeled DNA strand that is complementary to rRNA (ie, the strand that does not receive parental nucleosomes (12)). Thus, it is very possible that the parental nucleosomal proteins associated with active genes segregate a DNase I sensitive structure to the opposite side of the fork with kinetics much faster than 3 minutes.

Propagation of an active nucleosome structure to daughter chromosomes

The sensitivity of actively transcribed genes to DNase I is one of the few biochemical tools presently available to probe and define the structure of active chromatin. In trying to understand how actively transcribed genes become assembled into a DNase I sensitive structure it is likely that at one point the gene must be recognized in some way. This process is then followed by a second step where the DNase I sensitivity is built into the structure and the gene is then competent for transcription. Thus, in a certain sense, the sensitivity to DNase I reflects the result of some type of gene recognition event. If this is true then with each round of DNA replication when two chromosomes must be assembled from one, it is possible that a specific type of gene recognition takes place. Since this is one of only a few known systems in eukaryotes where the problem of DNA sequence recognition may be experimentally approachable, I would like to present the problem here in some detail, especially since the subject has not been discussed extensively elsewhere.

While a number of mechanisms could explain how newly-replicated DNA acquires the information to become packaged into an active chromatin configuration, only two rather extreme cases will be considered here--mechanisms that do not require DNA sequence specific recognition and mechanisms that do require DNA sequence specific recognition.

DNA sequence-specific interactions at the replication fork: Repeated DNA Sequences

DNA sequence specific interactions, for example with specific proteins or DNA modification enzymes, may be responsible for assembling newly-replicated DNA into active chromatin. Because this type of model would probably be the more prevalent one, it is worth discussing in some detail and, in particular, emphasizing some of the very real conceptual problems the model must explain. If DNA sequence specific interactions are indeed responsible for

establishing DNase I sensitivity of actively transcribed genes then two important points must be considered. The first is whether DNA site-specific recognition proteins can rapidly find the right DNA site at each division for every active gene in the genome. (The data shown in Fig. 1 suggest that whatever the detailed mechanism, the assembly of active chromatin during replication occurs very quickly). The second point which is equally important is whether these proteins will also interact with the wrong DNA site. The first of these questions has been discussed at length by a number of investigators (16-18) who have pointed out that a DNA binding protein like the lac repressor has an affinity for both specific and non-specific DNA sites. Consequently, even though the affinity for specific sites might be extremely high, a vast excess of non-specific DNA binding sites can compete for the interaction with specific sites. Clearly though, at a given ratio of specific to non-specific DNA sites (this has been calculated to be 1 part in 10^9 for higher eukaryotes (16)), the specific sites will eventually be occupied as the protein concentration is increased. The problem at this point, however, is that higher protein concentrations will lead to an increased occupancy of non-specific DNA sites as well. Thus, assuming DNA sequence specific interactions are actually responsible for gene recognition, the cell must optimize in some way the ratio of binding to specific versus non-specific DNA sites.

In principle, binding to specific sites could be optimized by increasing the ratio of specific to non-specific sites (See the excellent discussion of Linn & Riggs (17) for a detailed analysis of the problem as it pertains to eukaryotic cells). There are several ways this could be done. One way is if most housekeeping genes had the same DNA sequence (ie, a "repeated" DNA sequence) which is recognized and used to establish a DNase I sensitive structure along the transcription unit with each round of replication and that the responsible protein is present in large numbers per cell (10^4 - 10^5 per cell). If there are 10^4 housekeeping genes per cell, specific recognition then becomes a matter of finding 1 sequence out of 10^5 (10^4 genes/ 10^9 bp per cell), an efficiency well within the capability of a protein like the lac repressor. As a result of increasing the ratio of specific to non-specific sites it becomes possible to increase the level of recognition protein without increasing the occupancy of non-specific sites. Because the protein level can then be set much higher, the probability of a correct interaction at any given locus increases correspondingly without significantly influencing the probability of an interaction with the wrong sites.

If there are 10^4 transcription units in a typical cell, it becomes impor-

tant to consider the probability that all 10^4 of these are recognized and assembled into an active chromatin structure after replication. Presumably, the failure to properly assemble just one gene out of 10^4 could lead to cell death. While the postulate of repeated DNA sequences certainly helps to reduce this probability, it can be estimated that given reasonable numbers for specific ($10^{-14}M$) and non-specific ($10^{-6}M$) binding constants and reasonable ratios of specific to non-specific DNA binding sites, there would still be a rather high probability that at least one transcription unit is not properly recognized at each replication. Thus, no matter what the actual mechanism or the actual binding constants chosen, it would seem rather difficult to envision a sequence-specific process that can properly assemble some 10^4 transcription units with no mistakes whatsoever. One additional factor that may be relevant however, is that most cells are diploid; consequently, even though there is a finite probability that at least one of 10^4 genes may not be properly recognized and assembled, there is an extremely low probability that this will occur twice, on both the maternal and paternal chromosomes. However, this solution is also not entirely satisfactory since haploid cell lines as well as haploid organisms are known to exist. Finally, while the assumption of a common DNA sequence can help to explain the faithful replication of an active chromatin structure for housekeeping genes, it does not readily explain the replication of an active chromatin structure for genes that are active in one cell type but not in others. For these genes, it would seem that if DNA sequence specific binding is responsible for replication of the active state, then recognition proteins must ultimately also have to recognize 1 part in 10^9 with each cell division, although it is possible or even likely that a completely different type of mechanism governs the replication of such genes.

Interaction with non-specific DNA binding sites

It is assumed that, at least for housekeeping genes, once a recognition event occurs, the assembly of an active chromatin structure follows and that this structure is stable, at least until the next round of DNA replication. In this view, continued "recognition" does not have to be occurring throughout the entire cell cycle. However, it is clear that more complicated explanations can also be put forth. In any case, the important problem that has not been considered in detail is what happens when an interaction takes place at the "wrong" DNA site? This is almost a certainty since it is difficult to imagine how a protein could have an absolute specificity for one and only one DNA sequence. Does the occurrence of non-specific interactions mean that all genes are randomly being turned on and off and that supposedly "inactive"

genes are promiscuously being activated in a constantly changing sub-population of cells? While a number of observations in the literature might be put forth to support this idea, there is one obvious way to prevent this non-specific activation of genes. This is to require that more than one independent interaction occur before a gene is assembled into an active nucleosome configuration. While such a situation would initially decrease the probability that a given gene is activated since more than one event must occur, the cell could easily compensate by increasing the level of recognition protein. As a result, the likelihood that a given, specific gene is properly recognized (and subsequently activated) would not be greatly diminished if multiple site binding were required, but the possibility that a different gene is promiscuously activated would become very small since two low probability events would have to occur at two (presumably adjacent) sites. In an extreme version of this logic it would be of ultimate benefit to have a large number of possible specific binding sites associated with each gene and to require that a recognition mechanism interact with a rather small subset of these sites before actual activation occurs (17). In this context, it is of some interest that *in vitro*, the T-Antigen from SV40 can bind to multiple, contiguous sites on the SV40 chromosome (19), although it is not known whether multiple binding is, in fact, required for its biological activity. In any case, it would seem that if specific protein - DNA interactions are in fact responsible for specifying an active chromatin structure for housekeeping genes after replication, then based on what we presently know about the nature of these interactions from the prokaryotic models, it would seem necessary that there be only a few of these proteins, that each is present in large numbers, and that each interacts with repeated DNA sequences associated with most housekeeping genes. Many of these predictions are now subject to experimental testing.

DNA sequence-specific interactions: Histone masking

In their original discussion, Linn and Riggs suggested a number of plausible solutions to the problem of how specific sites might be recognized given a large number of non-specific sites. One is that eukaryotic recognition proteins have evolved to give a much higher degree of specificity than the lac repressor. In addition, they suggested that the packaging of eukaryotic DNA into chromatin might mask non-specific sites. However, it is now likely that individual nucleosomes are not completely competent in masking DNA sites since the DNA is on the outside of the nucleosome and most of it is known to be readily accessible to nucleases. Moreover, recent experiments have shown that the lac repressor can still recognize the lac operator when this DNA is

folded into reconstituted nucleosomes (20). In contrast, it is possible that when nucleosomes are folded into a higher order structure, the so-called thick fiber, most of the DNA in such a structure may be relatively inaccessible to large proteins. If the native SV40 mini-chromosome is a valid model only a very restricted region of this compact chromosome can be efficiently attacked by nucleases (21, 22).

If most of the DNA in the native, intact nucleus is indeed opaque to proteins, how can any type of DNA recognition occur? One obvious possibility is that it occurs during DNA replication, where it has been estimated that packaging of newly-assembled nucleosomes into higher order structures occurs some 10-15 minutes after replication of the associated DNA (23). Thus, if it is assumed that only newly-replicated DNA is available to recognition proteins, it is possible to estimate that at the time when a given specific sequence is replicating these proteins would only have to recognize 1 part in about 10^7 , a capability comparable to that displayed by the lac repressor at a concentration of about 100 molecules of protein per cell. Consequently, by partitioning the genome and sampling only the newly-replicated DNA, the effective level of competing non-specific sites can be reduced. Of course, there may be a problem with kinetics if there is only a finite time period for interaction; however, since specificity is usually established by decreasing the "off-rate" the kinetic parameter should, if anything, add to the efficacy of the interaction.

Finally, it is worth mentioning that an interesting consequence of these considerations would be that major sequence-directed alterations in chromatin structure (for example, during the activation of a gene (24)) would usually have to occur very shortly after DNA replication.

Protein - templating; Semi-Conservatively segregated, symmetrical proteins

The previous discussion has emphasized some of the very real (but perhaps not insurmountable) problems associated with models that require DNA sequence specific interactions for assembly of DNase I sensitivity after replication. An example of a mechanism that does not require DNA sequence specific interactions is one where the active chromatin uses protein-protein interactions to essentially "template" its own structure during replication. The very compelling biological arguments for proposing such a process have been discussed at length by Tsanev and Sendov (25) and by Alberts et al. (26) who have also suggested that one mechanism for "templating" might be analogous to a crystallization process wherein the parental nucleosome structure responsible for DNase I sensitivity "seeds" the structure for both daughter

genes. A related type of protein templating model can also be envisaged. Here a multi-subunit non-histone protein binds symmetrically to the DNA at specific sites by virtue of symmetrically paired subunits that have a preference for binding to the individual DNA strands. At the time of replication, the protein splits, its two halves going to each of the two daughter strands, and a new subunit "fills in" the vacancy, thus reconstituting the original paired multi-subunit protein structure on both daughter chromosomes. Formalistically, this model is the same as the half-nucleosome model previously proposed (26, 27) to deal with the segregation problem and the propagation of chromosomal structural information to daughter cells. Once such a "semi-conservative" replication process occurs for the proposed multi-subunit non-histone protein, there is still the problem of how each nucleosome in the transcription unit acquires an altered structure. Here we assume that the symmetrical, semi-conservatively segregated protein is recognized by the nucleosome assembly system and as a consequence, the active structure is in some way propagated down the length of the transcription unit as proposed in the first part of the Discussion. Clearly, a direct and testable prediction of this hypothesis is that in contrast to the histones which assemble and segregate conservatively (28), some chromosomal proteins will segregate semi-conservatively.

In conclusion, it should be emphasized that in discussing how an active nucleosome conformation is propagated to successive generations of daughter chromatin molecules, only two extreme mechanisms have been considered--those that do and those that do not require DNA sequence specific interactions. Various combinations of these mechanisms as well as alternative models might also have been considered; nevertheless, it is hoped that the analysis presented here, while obviously incomplete, will help to crystalize the problem and perhaps more importantly, to define the problem in terms that will make future experiments more approachable. Clearly, a next step in this type of analysis is to develop systems and techniques to follow the replication of specific actively transcribed genes in well-defined in vitro systems.

ACKNOWLEDGEMENTS

I would like to thank the NIH (GM26176 and CA20317) for support and Pat Thomas for suggestions.

REFERENCES:

1. Foe, V.E., Wilkinson, L.E. & Laird, C.D. (1976) *Cell* 9, 131-146.
2. McKnight, S.L. & Miller, O.L. Jr., (1977) *Cell* 12, 795-804.
3. Lacy, E. & Axel, R. (1976) *Proc. Nat. Acad. Sci. USA* 72, 3978 - 3984.
4. Felsenfeld, G. (1978) *Nature* 271, 115-122.
5. Weintraub, H. & Groudine, M. (1976) *Science*, 193, 848-856.
6. Garel, A. & Axel, R. (1976) *Proc. Nat. Acad. Sci. USA* 73, 3966-3970.
7. Riley, D. & Weintraub, H. (1979) *Proc. Nat. Acad. Sci.* 76, 328-332.
8. Weintraub, H. (1973) *Cold Spr. Harb. Symp. Quant. Biol.* 38, 247-256.
9. Weintraub, H. (1976) *Cell* 9, 420-422.
10. Seale, R. (1976) *Cell* 9, 423-429.
11. Leffak, M., Grainger, R., & Weintraub, H. (1977) *Cell* 12, 837-845.
12. Seidman, M., Levine, A.J. & Weintraub, H. In press.
13. Flint, S.J. & Weintraub, H. (1977) *Cell* 12, 783-792.
14. Garel, A., Zolan, M., & Axel, R. (1977) *Proc. Nat. Acad. Sci. USA* 74, 4867 - 4871.
15. Groudine, M., Weisbrod, S. & Weintraub, H. (1979) Manuscript in preparation.
16. Von Hippel, P.H., Revzin, A., Gross, C.A. & Wang, A.C. (1974) *Proc. Nat. Acad. Sci. USA* 71, 4808-4812.
17. Lin, S. & Riggs, A.D. (1975) *Cell* 4, 107-111.
18. Yamamoto, K. & Alberts, B. (1975) *Cell* 4, 301-310.
19. Tjian, R. (1978) *Cell* 13, 165-179.
20. Chao, M.V., Gralla, J. & Martiason, H. (1979) *Biochem.* 18, 1068-1074.
21. Scott, W.A. & Wigmore, D.J. (1978) *Cell* 15, 1511-1518.
22. Varschavoky, A.J., Sundin, O. & Bohn, M. (1979) *Cell* 16, 453-466.
23. Worcel, A., Han, S. & Wong, M.L. (1978) *Cell* 15, 969-977.
24. Holtzer, H. (1975) *Q. Rev. Biophys.* 8, 523-545.
25. Tsanev, R. & Sendov, B. (1971) *J. Theor. Biol.* 30, 337-384.
26. Alberts, B., Worcel, A. & Weintraub, H. (1976) in "Proceedings of the International Symposium on the Eukaryotic Chromosome," M. Bradbury and K. Jahaverian, Eds.
27. Weintraub, H., Worcel, A. and Alberts, B. (1976) *Cell*.
28. Leffak, M. Grainger, R., and Weintraub, H. (1977) *Cell*.