The formation and function of DNase I hypersensitive sites in the process of gene activation

Sarah C.R. Elgin  
*Washington University in St. Louis, selgin@wustl.edu*

Follow this and additional works at: [https://openscholarship.wustl.edu/bio_facpubs](https://openscholarship.wustl.edu/bio_facpubs)

**Recommended Citation**

[https://openscholarship.wustl.edu/bio_facpubs/227](https://openscholarship.wustl.edu/bio_facpubs/227)

This Article is brought to you for free and open access by the Biology at Washington University Open Scholarship. It has been accepted for inclusion in Biology Faculty Publications & Presentations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
Minireview

The Formation and Function of DNase I Hypersensitive Sites in the Process of Gene Activation*

Sarah C. R. Elgin
From the Department of Biology, Washington University, St. Louis, Missouri 63130

It is now 10 years since DNase I hypersensitive sites (DH sites) in genomic chromatin were first reported in a study of the structure of the Drosophila heat shock genes (1). Recent progress makes this a fitting time to review what we now know concerning the occurrence, formation, and function of these complex chromatin structures. Several comprehensive reviews have been published recently (2, 3); here we will use only selected examples to illustrate the major conclusions and outstanding questions. While it is now clear that all DH sites share the characteristic of being nucleosome-free regions of the chromatin fiber, it is also clear that there are many classes of DH sites, differing in form and function. We will focus on the 5'-promoter regions of three different inducible genes; the differences observed in chromatin structure imply different mechanisms for gene activation. Using directed mutagenesis and structural mapping, we should in the next few years be able to determine the mechanisms that generate such a promoter/enhancer structure, one accessible to the transcriptional apparatus in vivo.

General Occurrence of DH Sites

There is now no doubt that DH sites, in their most general form, simply represent discontinuities, or gaps, in the nucleosome array of the 100-Å chromatin fiber. Such sites were first detected in SV40 viral chromatin and in Drosophila genomic chromatin by their hypersensitivity to cleavage by DNase I (4-6). Subsequent investigation has shown that the DNA in such sites is generally accessible to all enzymes or reagents which will cut the double-stranded DNA, including restriction enzymes, micrococcal nuclease, endonuclease III, methylidium propyl-EDTA, Fe++, etc. (7, 8). Various studies have shown the region of a DH site (generally from 50 to 400 bp) to be nucleosome-free (7, 9). Analysis at higher resolution indicates, however, that while such sites always include segments of protein-free DNA, they can also contain internal regions associated with nonhistone chromosomal proteins (NHC proteins).

DH sites appear to be an essential feature of chromatin structure in eukaryotes, having been consistently observed in the chromatin of fungi, plants, and animals. DH sites have now been mapped at a number of specific positions of known function, including promoters, upstream activation sequences (UAS), enhancers of active or inducible genes, silencers of transcription, origins of replication, recombination elements, and structural sites within or around telomeres and centromeres (tabulated in Ref. 2). Additional sites have been mapped for which no function is readily apparent; however, a genetic assessment of such sites has frequently yielded interesting results. A typical pattern of DH sites for a gene showing tissue-specific expression, the chick lysozyme gene, is illustrated in Fig. 1 (10, 11). One commonly observes a cluster of DH sites close to the 5' end of the gene, present specifically in cells in which the gene is active or inducible; some of these are present in all cell types in which the gene is active, while others are present only in a subset of these cells. In this case sites 7 and 2 (at positions -0.1 and -6.1 kb relative to the

*Work in this laboratory discussed in this paper has been supported by the National Institutes of Health and the National Science Foundation.

1 The abbreviations used are: bp, base pair(s); NHC proteins, nonhistone chromosomal proteins; UAS, upstream activation sequences; kb, kilobases; HSCS, heat shock consensus sequence; HSF, heat shock transcription factor; MMTV, mouse mammary tumor virus; LTR, long terminal repeat.

![FIG. 1. Summary of patterns of DH sites](attachment:fig1.png)

**FIG. 1. Summary of patterns of DH sites**

The start of transcription) are strongly correlated with gene expression per se and coincide with the positions of known promoter and enhancer functions, respectively. Certain DH sites, however, are restricted to cells in the oviduct that express the gene in response to hormone stimulation, while a different set of additional sites is observed in macrophages, which express lysozyme constitutively. One oviduct-specific site (site 5 at -1.9 kb) is observed only when hormone is present; it appears to be directly involved in steroid hormone-stimulated induction of the gene (12). In contrast, transient expression assays indicate that site 3 (at -2.7 kb) is a macrophage-specific enhancer (13). The element at -2.4 kb (site 4) acts as a "silencer" in certain tissues (13). Flanking DH sites common to both of active and inactive tissues are also seen.

In cases where gene switching occurs as an intrinsic part of a developmental process, as in the different forms of globin synthesized in fetal, embryonic, and adult red blood cells, a switching in the pattern of DH sites associated with the different promoters is also observed, correlating the presence of particular 5'-DH sites with gene activity (7, 14). Appearance of the promoter-specific DH site at the chicken adult 8-globin gene is a relatively late step in the progression of chromatin structure changes associated with gene activation. However, the formation of this DH site occurs prior to the actual initiation of transcription of the gene, as shown by experiments with chick cells transformed with a temperature-sensitive avian erythroblastosis virus. When such transformed cells are maintained at the permissive temperature, their further development is blocked; however, on a shift to a nonpermissive temperature, they will continue to differentiate and ultimately will express hemoglobin. One arrested cell line was recovered that exhibited a chromatin structure including DH sites at the 8-globin promoter but no transcription; transcription occurred only when the developmental block induced by the virus was released by a temperature shift, allowing the remaining critical step(s) in the activation process to proceed (15). Many other studies support the conclusion that formation of tissue-specific DH sites precedes or accompanies tissue-specific gene expression (see Refs. 2 and 3).

The Structure of a DH Site

Recent technical advances have allowed the "dissection" of DH sites, including mapping of the protein-DNA interactions at the
nucleotide sequence level. We have recently completed an analysis of hp26, one of the heat shock genes of Drosophila. This gene is inactive in most tissues but can be activated within minutes by heat shock or other stress. There are two DH sites at, and just upstream of, the promoter (indicated as the proximal and distal sites in Fig. 2) (8); these sites include the sequences matching the regulatory heat shock consensus sequence (HSCS) (16). Following heat shock, a protected region can be discerned within each DH site, suggesting binding of the heat shock transcription factor (HSF) to the HSCS. Such an interpretation is consistent with the earlier results of Wu (17), who suggested (from ex0111 mapping studies of hp27 chromatin) that protein is bound in the TATA box region both prior to and after heat shock, while an additional protein is observed just upstream, protecting the HSCSs, only after heat shock. The TATA box, a general regulatory element critical for precise initiation of transcription, has been observed in vitro to interact with the RNA polymerase II transcription factor TFIIID (18).

Footprinting of hp26 chromatin at single base pair resolution, carried out using an indirect end-labelling strategy, has provided a more detailed picture. Prior to heat shock, the footprint of the TATA box binding protein can clearly be seen within the proximal DH site. Between the proximal and distal DH sites, a protected region of 150-160 bp is seen. Cleavage in this region by high levels of DNase I generates fragments with a 10-11-bp periodicity. Together, these results suggest that a precisely positioned nucleosome is present in this region. The proximal boundary for the nucleosome is a homopurine-homopyrimidine segment, capable of forming an S1 nuclease-sensitive DNA structure (most likely a triple helix (19)) in vitro (20); whether or not there is any protein binding at this site in vivo is difficult to discern. After heat shock, the footprints of the HSFs within the DH sites are clearly seen. Interaction of the TATA box protein with the DNA is now altered; several new, strong cleavage sites are seen at the downstream boundary of that footprint (21). Thus it appears that the proximal DH site is defined by the binding of (at least) the TATA box protein downstream and a precisely positioned nucleosome upstream. The same nucleosome defines the downstream boundary of the distal DH site; what defines its upstream boundary is unknown. Lis and his colleagues (22, 23) have shown that not only the TATA box binding protein, but also a molecule of RNA polymerase II, is already associated with the promoters of inactive hp70 and hp26 genes, resulting in a "poised" transcription complex. This complex is presumably activated by the heat shock transcription factor; HSF (22, 25). The above results suggest the model illustrated in Fig. 3.

The analysis of the hp26 DH sites illustrates several important points. First, the nature of the proximal site appears to be a consequence of both a specific NHC protein-DNA interaction (at the TATA box) and a precise positioning of a nucleosome, leaving a sequence between them which is too small to accommodate an additional nucleosome and hence remains open. Second, the key regulatory DNA sequences for this locus, the HSCSs, are in the open DNA regions, available to interact with the newly arriving (or newly activated) HSF. Thus one can argue that in this case the DH site is more than a marker of NHC protein-DNA interactions; it plays an important role. If, as implied, sequences packaged in nucleosomes cannot be readily "seen" in this system, the generation of a defined chromatin structure will greatly facilitate specific DNA-protein interactions by requiring the protein to search only a very limited portion of the genome, that portion lying within DH sites. A direct experimental test of this idea should be possible. In addition, the map of protein-DNA interactions suggests that folding of the DNA around a histone core (at position about -140 to -300) brings the two HSCSs close together via a small "loop" that is stabilized by a nucleosome. This provides an example of how features of chromatin structure may promote the cooperative interaction of multiple regulatory elements.

While in the above instance the site kept as nucleosome-free DNA in the inactive gene is clearly of functional significance, this certainly need not always be the case. In some instances, DH sites appear to be only the consequence of NHC protein binding, with no functional significance associated with the increased accessibility of the DNA. This appears to be the case at the yeast centromere, where DH sites are seen in the DNA flanking the functional protein-DNA complex; the DNA sequences in these flanking regions are not an essential component of the centromere (24, 25). There will no doubt be other such cases. Nonetheless, the accessibility of the DNA that is implied by nuclease sensitivity appears to be of functional significance in the case of hp26 detailed above and in most instances where NHC protein-DNA interaction is required for genetic activity. In fact, the accessibility of DH sites appears to have made them "targets of opportunity" for a variety of invasive events, including DNA damage and repair (e.g. Ref. 26) and insertion of transposable elements (e.g. Ref. 27).

Formation of DH Sites Accompanying Gene Activation

It should be noted that while in the case of hp26 the TATA box binding protein forms the downstream boundary of the highly accessible region that contains the proximal HSCSs, analysis at lower resolution shows that this entire region is relatively DNase I-sensitive, as are promoters of all active genes. The results suggest a need to exclude or remove nucleosomes from a transcription start site. The concept that nucleosomes could block the proper assembly of an RNA polymerase transcription complex is both reasonable and supported by experimental evidence. A careful mapping study of the mouse β-major globin gene in erythroid and nonerythroid cells has shown that whereas the promoter region is nucleosome-free in chromatin of active and inducible cells, that same region is covered by nucleosomes in nonerythroid cells (28). In vitro, neither SPS nor maximall RNA polymerase can utilize a nucleosome-bound promoter, although once properly assembled on a promoter, both enzymes can read through short linear fragments of nucleosome-associated DNA (29, 30). That NHC proteins can play a role in establishing nucleosome-free regions is indicated by the observation that the DH sites just 5′ of the chick β-globin gene can be reconstituted in vitro if and only if the DNA is incubated with a fraction of erythrocye nuclear proteins prior to nucleosome assembly by Xenopus oocyte extracts (31). Similarly, the assembly of a nucleosome array on a plasmid DNA template by Xenopus oocyte extracts will block subsequent transcription by HeLa cell extracts (32), but this effect can be reversed by prior incubation of the template DNA with the HeLa transcription extracts (33). Binding of transcription factor TFIIID may be necessary and sufficient to maintain a potential DH site during nucleosome assembly in some cases of the chicken DT site appears to be a "key player" in generating an active transcription complex.

Studies on two very different gene systems, the PHOS gene of yeast and the MMTV promoter in mouse cells, have revealed new insights concerning DH site formation and mechanisms of gene
Minireview: Formation and Function of DNase I Hypersensitive Sites

The DNase I hypersensitive sites (DH1 and DH2), positioned nucleosomes (large open circles), and the four 19-bp UAS elements (small solid circles) are shown. The scale is based with +1 referring to the initiation codon of the PHOS coding sequence. The region of transcription is indicated by a solid black bar; transcription is to the right. The upstream nucleosomes are numbered starting with −1 for the nucleosome immediately upstream of the start site; the TATA box lies in the center of nucleosome −1. This material was adapted from Ref. 35.

Changes in histone gene dosage which alter the ratio of H2A-H2B dimers to H3-H4 dimers have been shown to alter transcription patterns in yeast (38). Recently Grunstein and his colleagues (39) have created a yeast strain (UKY403) in which the single histone H4 gene is under control of the GAL1 promoter; growth under glucose conditions results in a depletion in H4 and the loss of about half of the chromosomal nucleosomes. This loss of nucleosomes results in induction of a subset of loci, including PHOS. Direct examination of the PHOS chromatin under these conditions shows the upstream nucleosome array to be destabilized (40). The results suggest that nucleosomes indeed serve as negative regulators in this case; access to the TATA box appears to be the critical parameter. Studies of PHOS in wild-type yeast on a finer scale indicate that the stability of the −2 nucleosome is important; when this segment of DNA is replaced by a fragment of African green monkey a-satellite DNA that generates a very stable nucleosome structure, induction of the gene under low phosphate conditions is blocked. If, however, the same region is replaced with a segment of pBR DNA that has a weak association with the histone octamer, the promoter is both leaky and inducible. Chromatin structure analyses show that the pBR segment undergoes a transition from a largely nucleosomal to a non-nucleosome state upon induction. The results indicate that both the precise positions and the stability of the nucleosomes in this 5′ region are critical in regulating expression of the gene.

A somewhat similar step in the process of activation, a displacement of a nucleosome from the regulatory region, has been suggested by Hager and his colleagues (41, 42) for the mouse mammary tumor virus LTR promoter. The promoter, normally inactive but inducible by steroid hormones, was studied on bovine papilloma virus-based episomal constructs. Two factors, probably nuclear factor 1 and the TATA factor, interact with this promoter in the region −82 to −4 bp in an in vitro analysis. Both factors are present in nuclear extracts of both control and induced cells (41). In the absence of hormone, a regular array of nucleosomes was mapped across the promoter region. In the presence of hormone, the pattern is unchanged except for the region between −60 and −250, which contains the hormone receptor binding site. This region becomes hypersensitive during induction, suggesting displacement of the nucleosome as a part of the activation process (42). A very interesting contrast to the results for yeast PHOS exists, however; in this case it appears that the hormone-receptor complex can “see” its binding site even when it is on a nucleosome. In reconstitution studies using a restriction fragment from the MMTV LTR, a nucleosome is observed at the same position (−76 to −219) as mapped in vivo (43). The glucocorticoid receptor can bind directly to this complex; no unfolding or dissociation of the nucleosome is observed. The footprinting pattern reveals the alterations in the DNA digestion pattern by the two interactions (43). Further studies in vivo and in vitro may reveal subsequent steps by which the nucleosome is displaced. It should be noted, of course, that in all of the cases discussed here the nucleosomes in vivo have been identified only on the basis of digestion patterns with micrococcal nuclease, DNase I, and restriction enzymes; no direct protein characterization has been carried out. Consequently, we do not know whether or not there are any special features (e.g., presence of minor histone subtypes, core histone modifications, binding of specific NHC proteins) of those critical nucleosomes which facilitate the required reorganization of promoter regions.

These studies, taken together, provide a striking contrast in implied mechanisms for transcriptional activation. In comparing the 5′ regulatory regions of hap26, PHOS, and the MMTV LTR, we find that it is possible to have 1) a chromatin structure in which all of the essential regulatory regions are positioned in nucleosome-free DNA, and no 5′-nucleosome displacement is necessary for activation (hap26); 2) a chromatin structure in which an initial responsive element is in nucleosome-free DNA, but nucleosome displacement is necessary to provide access to the TATA box for the transcription complex (PHOS); and 3) a chromatin structure in which a critical regulatory sequence is in a nucleosomal structure but is nonetheless accessible and capable of binding the signal protein (hormone-receptor complex) to start the process (including nucleosome displacement in the region which will eventually result in assembly of the transcription initiation complex (MMTV LTR). We do not know which of these modes of operation may be the more common or on what basis a given mechanism is established for a given regulatory problem. No doubt the list of possible activation mechanisms is still far from complete. The prospects for further experimental work are exciting.

Nucleosome Positioning

A very important conclusion from the above work is that nucleosomes are specifically positioned in regulatory regions of the genome and that such precise positioning is likely to be of critical importance. Certain induced promoters must be displaced by nucleosomes during the activation process; moreover, it appears that the pattern of nucleosomes places key DNA sequences in relatively inaccessible positions in some instances and in relatively accessible positions in others. We know that nucleosomes can assemble on a specific fragment of DNA in a sequence-dependent manner (e.g., Ref. 44), that nucleosomes do occupy defined positions within chromatin (as described above), and that certain homopolymers and specific DNA structures exclude nucleosome formation and thus can serve as boundaries for nucleosome arrays (e.g., Ref. 45). However, in no case do we have a complete set of data demonstrating how these features of DNA sequence and chromatin structure are utilized in generating a particular DH site. Space limitations preclude further discussion of the role of DNA structure in dictating chromatin structure, a topic of much current interest and research (see Refs. 2, 3, 44, and 46).

While placement of nucleosomes can help to define a DH site, the reciprocal is also true: creation of a DH site can help dictate the positions of an array of nucleosomes, as shown by Thoma (47, 48) using yeast episomes. Here a defined minichromosome (with two nucleosome arrays and two DH sites) has been altered either by the addition of short DNA segments within a nucleosome array or by the insertion of a new DH site. The former has no significant effect; if the length of DNA added to a DH site is too short, no DH site can be added, but the overall pattern does not change. In contrast, addition of a new DH site in the minichromosome appears to create a new boundary, and nucleosome positions are shifted to accommodate this. Given the local autonomy of DH site formation (e.g., Ref. 49) and the recent success in using transformation for chromatin structure analysis in higher eukaryotes (e.g., Ref. 50), answers to the question of the problem of DH site formation will probably be addressed successfully in the near future using this approach.

FIG. 4. Chromatin fine structure at the PHOS promoter. The DNase I hypersensitive sites (DH1 and DH2), positioned nucleosomes (large open circles), and the four 19-bp UAS elements (small solid circles) are shown. The scale is based with +1 referring to the initiation codon of the PHOS coding sequence. The region of transcription is indicated by a solid black bar; transcription is to the right. The upstream nucleosomes are numbered starting with −1 for the nucleosome immediately upstream of the start site; the TATA box lies in the center of nucleosome −1. This material was adapted from Ref. 35.

4 M. Han and M. Grunstein, personal communication.
5 C. Straka and W. Hö rz, personal communication.
Minireview: Formation and Function of DNase I Hypersensitive Sites

Conclusions

Where are we now? The work discussed above indicates that an appropriate chromatin structure is an intrinsic part of effective and efficient gene regulation. Specifically, we find that in eukaryotes (organisms that package their DNA in nucleosome arrays) key regulatory regions of active and inducible genes are maintained in an accessible configuration and that generation and/or maintenance of nucleosome-free DH sites can be an important step in allowing assembly of a transcription complex. It is most likely that patterns of DNA structure and of histone octamer-DNA interaction, as well as chromatin structure, are dependent on the interaction of factors of the transcription complex. It is most likely that patterns of DNA structure and of histone octamer-DNA interaction, as well as chromatin structure, are dependent on the interaction of factors of the transcription complex. It is most likely that patterns of DNA structure and of histone octamer-DNA interaction, as well as chromatin structure, are dependent on the interaction of factors of the transcription complex.

How should we proceed? The most interesting questions now revolve around the formation of DH sites. In the case of many constitutive sites, one should be able to analyze formation by combining in vitro studies of assembly with in vivo mapping (at the nucleotide sequence level) of the chromatin structures of carefully designed mutants. Such studies should allow one to elucidate cause and effect, as in the studies of PHOS regulation. More challenging problems arise in elucidating the sequence of events that result in developmentally regulated changes in chromatin structure; here defined cell lineages which are becoming available (e.g. cell lines with erythropoietic potential) will be an enormous help. Without doubt, however, some critical events determining chromatin structure patterns occur quite early in embryogenesis; here the invertebrates and lower vertebrates (i.e. Drosophila, sea urchin, Xenopus) with more accessible embryos and larger numbers of cells/embryo at the critical developmental decision points will be invaluable analytical models. Such analyses should help us to bridge the gap between the biochemical analysis of transcription and the observations of developmental biology, leading to an integrated picture of gene regulation.

Acknowledgments—I thank many colleagues for helpful discussion and criticism of the manuscript; I thank Tom Dietz for help with the figures.

REFERENCES
43. Perlman, T., and Wrangle, O. (1985) EMBO J. 4, 3703-3709
47. Thoma, F. (1986) J. Mol. Biol. 190, 177-190