6-1990

Chromatin structure and gene activity

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

Chromatin structure and gene activity

S.C.R. Elgin

Department of Biology, Washington University, St Louis, Missouri, USA

Current Opinion in Cell Biology 1990, **2**:437–445

Introduction

Twenty years ago, the histones were commonly described as 'structural proteins of chromatin and general repressors of template activity' without any clear idea of how they acted as such. We now know that both the distribution of nucleosomes along the chromatin fiber and the organization of higher order structure play a role in regulating gene expression. Progress made during the last year reflects our increasing ability to use genetic and gene-specific biochemical mapping techniques to establish functional relationships. We have identified the genes encoding many chromosomal proteins and can create or recover mutations in these genes, altering either the protein itself or the amount available; this has allowed us both to determine the importance of these proteins and to identify interactive domains. We can now manipulate the DNA sequences of a given gene and, following transformation of the gene back into the cell, establish the effects of these manipulations on chromatin structure and function. Transformation experiments have revealed the sensitivity of genes to their 'chromosomal environment', (i.e. have demonstrated euchromatic position effects) and the buffering of genes from such position effects has become a new and useful definition of 'domain' structure. The following review will focus on four areas that have been significantly advanced during the last 18 months by studies employing a combination of biochemical and genetic techniques.

Nucleosomes as repressors

It is now clearly established that nucleosomes can occupy specific positions along the DNA. Mapping experiments have shown the disposition of specific ordered nucleosome arrays over both inducible and stably inactive genes [chick globin (Benezra *et al.*, *Cell* 1986, 44:697–704) and *Drosophila* heat-shock genes (Cartwright and Elgin, *Mol Cell Biol* 1986, 6:779–791)], and also extending from telomeres (Gottschling and Cech, *Cell* 1984, 38:501–510) and around centromeres (Bloom and Carbon, *Cell* 1982, 29:305–317). Nucleosome positioning can be affected both by boundary elements (for example, other proteins that block nucleosome formation) and by the DNA sequence itself (Thoma and Simpson, *Nature* 1985, 351:250–252; Thoma, *J Mol Biol* 1986, 190:177–190). In the above examples, both effects can be inferred. Specifically bound non-histone chromosomal proteins (NHC proteins) are usually observed at or near the 5′ end of active/inducible genes; these sites [detected as DNAase I hypersensitive sites (DH sites)] serve as boundaries for nucleosome arrays. Nonetheless, the distribution of the nucleosomes across a gene is usually irregular, and shows local variation, presumably reflecting the DNA sequence. No data are available, however, on the very large stretches of DNA that can occur between transcripts in the eukaryotic genome; in large areas of unique sequence with relatively few boundaries, a random nucleosome distribution could occur.

Where there is specificity in organization, one can readily infer that nucleosome positioning might limit access to specific sequences. Both *in vivo* and *in vitro* studies have demonstrated that a functional competition exists between nucleosomes and the transcriptional apparatus for occupancy of the TATA box. In yeast (*Saccharomyces cerevisiae*), alteration of the stoichiometry of the dimer sets of core histones results in altered patterns of transcription (Clark-Adams *et al.*, *Genes Dev* 1988, 2:150–159). Moreover, depletion of nucleosomes leads to the activation of many genes in the absence of the appropriate inducing agents [1]. Recently, Grunstein and his colleagues (Kim *et al.*, *EMBO J* 1988, 7:2211–2219) have created a yeast strain (UKIY403) in which the single histone *H4* gene is present under control of the *GAL1* promoter; growth of this strain on glucose-containing media results in a depletion in H4 and the loss of about half of the chromosomal nucleosomes. This loss of nucleosomes results in activation of a set of loci, including *PHO5*, normally regulated through upstream activating sequence (UAS) elements [1]. Direct examination of the *PHO5* chromatin under these conditions shows the upstream nucleosome array to be destabilized (Han *et al.*, *EMBO J* 1988, 7:2221–2228). In wild-type chromatin, one observes a nucleosome associated with the TATA box when the gene is inactive; following induction, nu-

Abbreviations

ARS—autonomously replicating sequence; **DH site**—DNAase I hypersensitive site; **IE**—immediate-early; **LAR**—locus activation region; **MAR**—matrix attachment region; **mRNA**—messenger RNA; **NHC**—non-histone chromosomal [protein]; **RAP**—repressor–activator binding protein; **SAR**—scaffold attachment region; **snRNP**—small nuclear ribonucleoprotein particle; **UAS**—upstream activating sequence [element].

 ISSN 0955-0674

cleosomes are lost from this region (Almer *et al.*, *EMBO J* 1986, 5:2689–2691). The results suggest that nucleosomes indeed serve as negative regulators at such genes; access to the promoter appears to be a critical parameter.

A less drastic perturbation, a yeast mutant deficient in the H2A-H2B dimer, shows a localized disruption of nucleosome arrays; the chromatin structures of *CYH2*, *UB14*, and *CEN3* (the gene encoding the centromere of chromosome III) are dramatically disrupted, while *HIS4*, *GAL1*, and the telomeres appear normal [2]. This selectivity in nucleosome formation is intriguing. Whether it results from direct competition for histone binding, or reflects the timing of replication or other characteristics of the individual genes, is unknown. The ability to replace the wild-type histone genes of yeast with mutant constructs is beginning to allow the identification of functional domains within these proteins. No doubt this approach will be used to analyze such targeted chromatin assembly.

A direct demonstration that nucleosome positioning can affect the function of a *cis*-acting DNA element has been provided by the analysis of an autonomously replicating sequence (ARS1) by Simpson [3]. The 11 base pairs of the ARS1 element in the high-copy-number TRP1/ARS1 yeast plasmid are normally located in a DH site, near the edge of a positioned nucleosome. By deleting increasing amounts of DNA from the near edge of the nucleosome, a series of plasmids was created in which the ARS1 sequence was moved successively into the nucleosome core (see Fig. 1a). Deletion of 60–80 bp of DNA shifts the ARS1 into the central region of the nucleosome structure; these plasmids show a striking reduction (20-fold or more) in copy number after propagation, indicating a loss of ARS1 function. The nucleosome could be repositioned to uncover the ARS1 site by insertion of a synthetic α2 operator on the distal side; the α2 repressor is observed to organize the chromatin structure of plasmids containing this site (Roth *et al.*, *Mol Cell Biol* 1990, 10:2247–2260). As anticipated, this results in recovery of ARS1 function specifically in α-cells, where the α2 repressor is present (Fig. 1b). The results demonstrate that nucleosome association *in vivo* severely limits the functional accessibility of the ARS1 DNA. In a similar type of study, Wolffe and Drew [4], using curved DNA to manipulate nucleosome position, have shown that incorporation of a TATA box into a nucleosome will block transcription by T7 RNA polymerase *in vitro*. One would hope to see similar *in vivo* experiments carried out within the year to evaluate directly the effects of nucleosome binding on a TATA box and on upstream regulatory sites normally seen to function within DH sites.

In contrast, there are other instances in which the organization of DNA sequences on the surface of a nucleosome appears to be important for the interaction of the DNA with other chromosomal components. For example, the glucocorticoid hormone receptor-binding site in the mouse mammary tumor virus promoter associates with a histone octamer *in vitro* in such a way that it can directly bind the glucocorticoid receptor complex, producing a tripartite structure (Perlmann and Wrange, *EMBO J* 1988, 7:3073–3079; Pina *et al.*, *Cell* 1990, 60:719–731). This implies an exact positioning of the DNA on the histone octamer, leaving the critical binding sequences facing outward. A nucleosome in just this position is observed in the chromatin structure of the uninduced gene *in vivo*; following hormone treatment, this nucleosome is no longer observed (Richard-Foy and Hager, *EMBO J* 1987, 6:2321–2328). NF1, a transcription factor present in the nucleus, essential for efficient use of this promoter, is observed by *in vivo* exonuclease protection assays to bind to this region after hormone induction (Cordingly *et al.*, *Cell* 1987, 48:261–270); *in vitro*, NF1 binds only in the absence of nucleosome structure (Pina *et al.*, 1990). Again, this suggests that a nucleosome can serve as a repressor, blocking access until displaced as part of the gene activation process.

As mentioned above, depletion of functional nucleosomes results in a disruption of *CEN3* structure and loss of order in the surrounding region [5]. A 'surface' orientation of the critical sequence elements (CDE I and CDE III) of the centromere, which interact with other proteins, is again predicted. The precise chromatin structure observed in *S. cerevisiae* for these elements is not observed on this DNA transformed into *Schizosaccharomyces pombe*, suggesting that the NHC proteins (rather than the conserved core histones themselves) position the nucleosome *in vivo* [5].

Chromatin structures related to promoter activation

Many laboratories have demonstrated that nucleosome assembly *in vitro* can block the assembly of active transcription complexes on a promoter. However, if a stable preinitiation complex (based on the binding of factor TFIID to the TATA element) is formed on the promoter prior to nucleosome assembly, the promoter will remain active for subsequent transcription (e.g. Matsui, *Mol Cell Biol* 1987, 7:1401–1408; Workman and Roeder, *Cell* 1987, 51:613–622; Knezetic *et al.*, *Mol Cell Biol* 1987, 8:3114–3121; Lorch *et al.*, *Cell* 1987, 49:203–210; Losa and Brown, *Cell* 1987, 50:801–808). Such a strategy of preformation of a transcription complex, presumably during replication, appears to be used *in vivo* not only for constitutive genes, but also for some inducible genes. Using ultraviolet photocrosslinking followed by immunoprecipitation with antibodies against RNA polymerase II (Gilmour and Lis, *Mol Cell Biol* 1986, 6:3984–3989) and run-on transcription assays [6], Lis and his colleagues have established the preferential association of a molecule of RNA polymerase II at the 5′ end of several *Drosophila* genes, including both inducible genes (*hsp70* and *hsp26*) and constitutively expressed genes (polyubiquitin and β-tubulin). Negative results were obtained for the yolk protein 1 gene and gene 1 of locus 67B; neither of these genes are expressed in the SL2 cells used in this study (Rougvie and Lis, 1990, personal communication). There remains a question about whether the polymerase associated with the heat-shock genes is in a

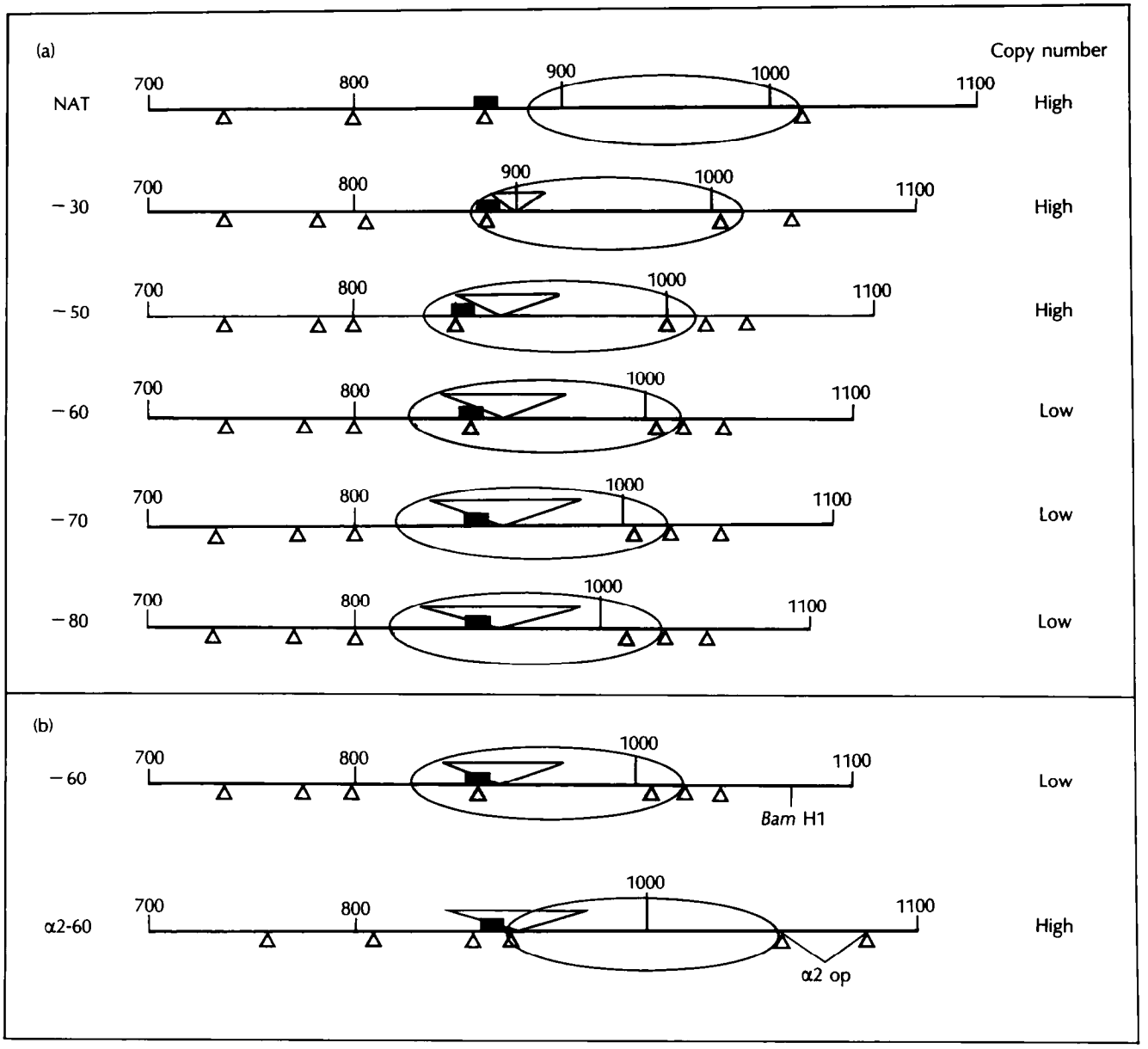

Fig. 1. Nucleosome positioning can affect the function of a *cis*-acting DNA element (ARS1) *in vivo*. (a) Small deletions (indicated by the open triangles) were used to move an ARS1 core sequence (indicated by a black box) progressively from the edge of a nucleosome to an interior location. Chromatin structure analysis was carried out using micrococcal nuclease. The locations of nuclease-cutting sites are indicated by the arrowheads. The inferred position of the nucleosome for each construct is shown as an open ellipse. The copy number of the mutant plasmids is indicated in the column at the right. There is at least a 20-fold difference between 'high' and 'low'. (b) Insertion of the α2 operator (shown by the lower triangle) results in the reorganization of chromatin structure in α cells (expressing the α2 repressor) as shown, restoring ARS1 function. Published by permission [3].

preinitiation complex or in an arrested elongation complex; the run-on transcription studies show a nascent RNA chain of approximately 25 nucleotides for *hsp70* and *hsp26*, but this synthesis might have occurred during the requisite nuclear preparation. Genomic footprinting has also provided evidence for the binding of the TATA factor to the inactive but inducible *hsp26*, *H3*, and *Sgs3* genes (Thomas and Elgin, *EMBO J* 1988, 7:2191–2202; [7]; Mathers and Meyerowitz, 1990, personal communication).

These considerations focus attention on the assembly processes that must occur during replication of the chromatin. Alberts and his colleagues have very recently reported that *in vitro* the T4 replication machinery can replicate through nucleosomes without histone displacement [8]. This should facilitate the replication of precise nucleosome arrays. (See Svaren and Chalkley, *Trends Genet* 1990, 6:1–4 for a more detailed discussion of persistence and change in chromatin structure during replication.) It should be noted that while nucleosomes must (often) be excluded from the TATA box, they nonetheless are an integral part of the promoter–enhancer complex; by their own position, nucleosomes will define nucleosome-free spaces, and by folding the DNA they

will bring distant regions into proximity. Thus, in looking for mechanisms to establish functional transcription complexes during replication, one might look either for DNA or protein elements that can position nucleosomes so as to leave the TATA box accessible, or for elements that promote efficient binding of TFIID. The studies cited above on the ARS1 illustrate the former strategy; recent studies on reconstitution of functional transcription complexes suggest that the latter strategy also plays an important role.

Several proteins that might facilitate TFIID binding to the TATA box are being characterized. Workman *et al.* [9] have shown that when the binding of TFIID is in direct competition with nucleosome assembly in an *in vitro* system, a transcriptional activator, the immediate-early (IE) protein of pseudorabies virus, stimulates TFIID binding to the adenovirus major late promoter. IE allows TFIID to bind more rapidly than the nucleosomes, resulting in subsequent formation of preinitiation complexes. Maximal transcription from this promoter requires a ubiquitous cellular upstream transcription factor, USF (MLTF) (which binds at about position −58) in addition to RNA polymerase II and general transcription factors. USF can also facilitate the formation of stable preinitiation complexes in a chromatin assembly system, greatly stimulating relative transcription. The effect appears to be the consequence of formation of preinitiation complexes, rather than nucleosomes, over an increased proportion of the promoters [10].

A similar role is being suggested for a *Drosophila* CT-binding protein, based on analyses at the *hsp26* promoter. Deletion or replacement of the $(CT)_n$ region leads to a four to fivefold reduction in inducible activity *in vivo*, indicating that this element functions as a positive transcriptional regulator (Glaser *et al.*, *J Mol Biol* 1990, 211:751–761). Presumably, this function involves the CT-binding protein [7] which appears to be the same as, or related to, the GAGA protein, a transcription factor active at *Ubx* and other promoters in *Drosophila* (Biggin and Tjian, *Cell* 1988, 53:699–711). *In vitro* binding studies have shown a cooperative interaction between the protein binding to the $(CT)_n$ element (at −85 to −116) and the protein(s) binding to the TATA box; kinetic studies suggest that the CT-binding protein facilitates assembly of the TATA-dependent complex (Dietz, Gilmour and Elgin, unpublished results). Thus, an abundant transcription factor may play a key role in 'setting up' the chromatin structure necessary for an active or inducible complex. In different cases this might be achieved through an interaction with TFIID, directly facilitating its binding, or through exclusion of nucleosomes, indirectly facilitating TFIID binding, or both.

A similar hypothesis can be derived from results of studies on the GRF2 protein (also known as REB1, Q, or Y) in yeast. Binding of this protein to DNA creates a nucleosome-free region of about 230 bp, with consequent ordering of nucleosome arrays at both sides (Fedor *et al.*, *J Mol Biol* 1988, 204:109–127). GRF2 binding sites are associated with many UASs and other functional elements in yeast. GRF2 by itself stimulates transcription only slightly (twofold) but it can act with a second weak activator to give 170-fold stimulation of transcription [11]. GRF2 may be acting synergistically by creating an effective (accessible, nucleosome-free) binding site for the second activator. The effect of GRF2 is distance-dependent (i.e. the second activator must be adjacent), as might be expected for a structural mechanism. Another synergistic transcription factor found in yeast, ABF1, binds not only to sequences in promoter regions, but also to sequences associated with some ARSs (Buchman and Kornberg, *Mol Cell Biol* 1990, 10:887–897). One is reminded of the observation that in mammals transcription factor OTF-1 is functionally identical to DNA replication factor NF-III (O'Neill *et al.*, *Science* 1988, 241:1210–1212). This suggests a possible relationship between control of transcription and initiation of replication, which might reflect the need to maintain an accessible chromatin structure at these sites.

Elongation and core nucleosome structure

Once proper initiation occurs, the transcription complex faces what would appear to be the equally formidable problem of 'reading through' the DNA template, which apparently has been, until that moment, organized in a nucleosome array and folded into a 300Å fiber. There is considerable evidence to indicate that both the nucleosomes themselves and the packaging of the larger domains differ between active genes and the rest of the chromatin, but we are only beginning to learn about the biochemical basis for these differences. As described last year in this journal (Patient and Allan, *Curr Opin Cell Biol* 1989, 1:454–457) rapid and reversible changes occur in nucleosomes on activation, sufficient to allow fractionation by affinity chromatography (Chen and Allfrey, *Proc Natl Acad Sci USA* 1987, 84:5252–5256). A striking biochemical feature of the 'active' nucleosomes obtained is a high level of acetylation of the N-terminal regions of the core histones, likely to alter contacts with DNA. In a different but complementary approach, immunoprecipitation of oligo- and mononucleosomes with an antibody against ε-N-acetyl lysine has shown a 15–30-fold enrichment of an active gene in the selected fashion (Hebbes *et al.*, *EMBO J* 1988, 7:1395–1402). Interestingly, histone acetylation has also been reported to alter the negative specific linking number change per nucleosome on closed circular DNA from -1.04 ± 0.08 to -0.82 ± 0.05 [12]. This release of constrained negative supercoiling might facilitate an opening of the double helix for transcription. The possibility of the nucleosomes functioning to some extent like an allosteric complex is intriguing, as even small shifts in core structure can alter the predicted linking number of the associated DNA (White *et al.*, *J Mol Biol* 1989, 207:193–199).

A further potential source of variation in nucleosome structure is the use of core histone variants. Variants of H2A, H2B and H3 have been recognized which are not under the S-phase regulation imposed on the bulk of histone synthesis. Several of the genes encoding these

'basal' proteins have now been cloned. One generally finds that these differ from the genes (of animals and protozoa) encoding replication-dependent histones by inclusion of introns and production of polyA$^+$ messenger RNA (mRNA). [In plants, it appears that most of the genes encoding histones produce a polyA$^+$ mRNA (Chaboute *et al.*, *Gene* 1988, 71:217–223).] That the variant genes are both ancient and essential has recently been established from studies of the H2A.Z family, which includes the hv1 protein of *Tetrahymena* and H2AvD of *Drosophila*. The *Tetrahymena* protein is associated with the transcriptionally active macronucleus, but not the transcriptionally inert micronucleus (Allis *et al.*, *Cell* 1980, 20:609–617). The *H2A.Z* gene clearly diverged from the replication-dependent *H2A.1* gene prior to the separation of yeast and *Tetrahymena* from the eukaryotic lineage [13]. H2A.Z is encoded by a single copy gene in *Drosophila*; loss of gene function is lethal (van Daal and Elgin, personal communication, 1990). This result supports earlier suppositions (based on the high level of conservation) that the histone variants are an essential component of chromatin, and must be taken into account in our assessment of variation in nucleosome structure and in the chromatin fiber.

The control of histone synthesis clearly is more complex (and interesting) than previously supposed. Regulation of replication-dependent histone synthesis uses control of initiation of transcription, 3′ processing of the transcript, and stability of the mRNA (Schumperli, *Trends Genet* 1988, 4:187–191). Variant histone synthesis is regulated in developmentally specific patterns. In at least some instances, this makes use of alternative modes of processing of the transcript, with both small nuclear ribonucleoprotein particle (snRNP)-mediated 3′ cleavage sites and polyA addition sites being encoded by the gene [14,15]. A similar result has been reported for a mouse H1 histone gene (Cheng *et al.*, *Proc Natl Acad Sci USA* 1989, 86:7002–7006). A compilation and alignment of the histone and histone gene sequences is available [16]. Clearly, the study of the histone variants is an important area touching on questions of specificity in nucleosome structure, nucleosome distribution, and nucleosome function.

Domains and compaction

There are many observations from cytological, genetic and molecular studies to suggest that the eukaryotic chromosome is structurally and functionally subdivided into domains of 10–100 kb of DNA. 'Domains' have been observed as bands in polytene chromosomes, loops of lampbrush chromosomes, loops of DNA emanating from histone-extracted nuclei or metaphase chromosomes, topological units, regions of general DNAase I sensitivity (for active genes), and genetic units. Whether or not the boundaries of these various units functionally coincide has not been determined; we have no coherent picture of the 'domain' as yet, but hope to gain one from the application of functional assays, made possible by transformation techniques. For such purposes, one can define a 'domain' as that piece of contiguous DNA which, when reinserted into the genome, allows accurate and fully regulated expression of the gene at normal levels, independent of the site of insertion. It has been suggested that domain boundary sites are of critical importance, and might act through attachment to a nuclear matrix or scaffold. [Such 'matrix attachment regions' (MARs) or 'scaffold attachment regions' (SARs) have been biochemically identified as those sequence elements remaining associated with a nuclear pellet after lithium diiodosalicylate extraction and nuclease digestion (Mirkovitch *et al.*, *Cell* 1984, 39:223–232), or as sequence elements that bind competitively to a nuclear pellet prepared by DNAase I digestion and 2M NaCl extraction (Cockerill and Garrard, *Cell* 1986, 44:273–282); the two approaches appear to identify the same set of AT-rich DNA fragments. It should be noted that both assays identify all potential binding sites; actual binding sites for a given cell type may well be only a subset of these.]

The most complete analysis currently available comes from the studies of Sippel and his colleagues on the chick lysozyme gene (Fig. 2) [17]. A broad domain of relatively DNAase I-sensitive chromatin has been mapped for this active/inducible gene. This region covers about 24 kb of DNA with the 4 kb transcript approximately in the middle. 'A elements', which behave as nuclear scaffold elements (SARs) in the lithium diiodosalicylate-extraction protocol, have been identified at the boundaries of this region (Phi-Van and Stratling, *EMBO J* 1988, 7:655–664). Because transformation is more difficult with large pieces of DNA, test constructs containing a reporter gene with an appropriate promoter and regulatory elements, bracketed by the putative boundary elements to be tested have frequently been made. Sippel's group, using a chloramphenicol acetyl transferase (CAT) reporter gene (C) coupled to the lysozyme enhancer (E) and promoter (P) as shown in Fig. 2, assessed the functional effects of the A elements by transient transfection or stable transformation in promacrophages that normally express this gene. In the former case, no effect was observed as a consequence of bracketing the gene with A elements. In the latter case, high-level CAT activity that was dependent on the copy number and significantly less affected by the site of insertion was observed [17]. This suggests that the transformed A elements, which again behave as SARs, create an independent domain for the newly inserted gene. Significant A-element-dependent stimulation of reporter-gene activity was also observed. Further experiments will be needed to delineate the role of the biochemically defined SARs within the large DNA fragments used, which might have some protective effect by virtue of their size. Similar strategies are being employed in *Drosophila* to explore the significance of the 'specialized chromatin structures', elements identified by chromatin structure analysis and cytological analysis which bracket the 87A7 heat-shock puff (Udvardy *et al.*, *J Mol Biol* 1985, 185:341–358).

Some caution is required in interpreting this type of experiment, however. In the above instance, the congruence of biochemical and genetic results strength-

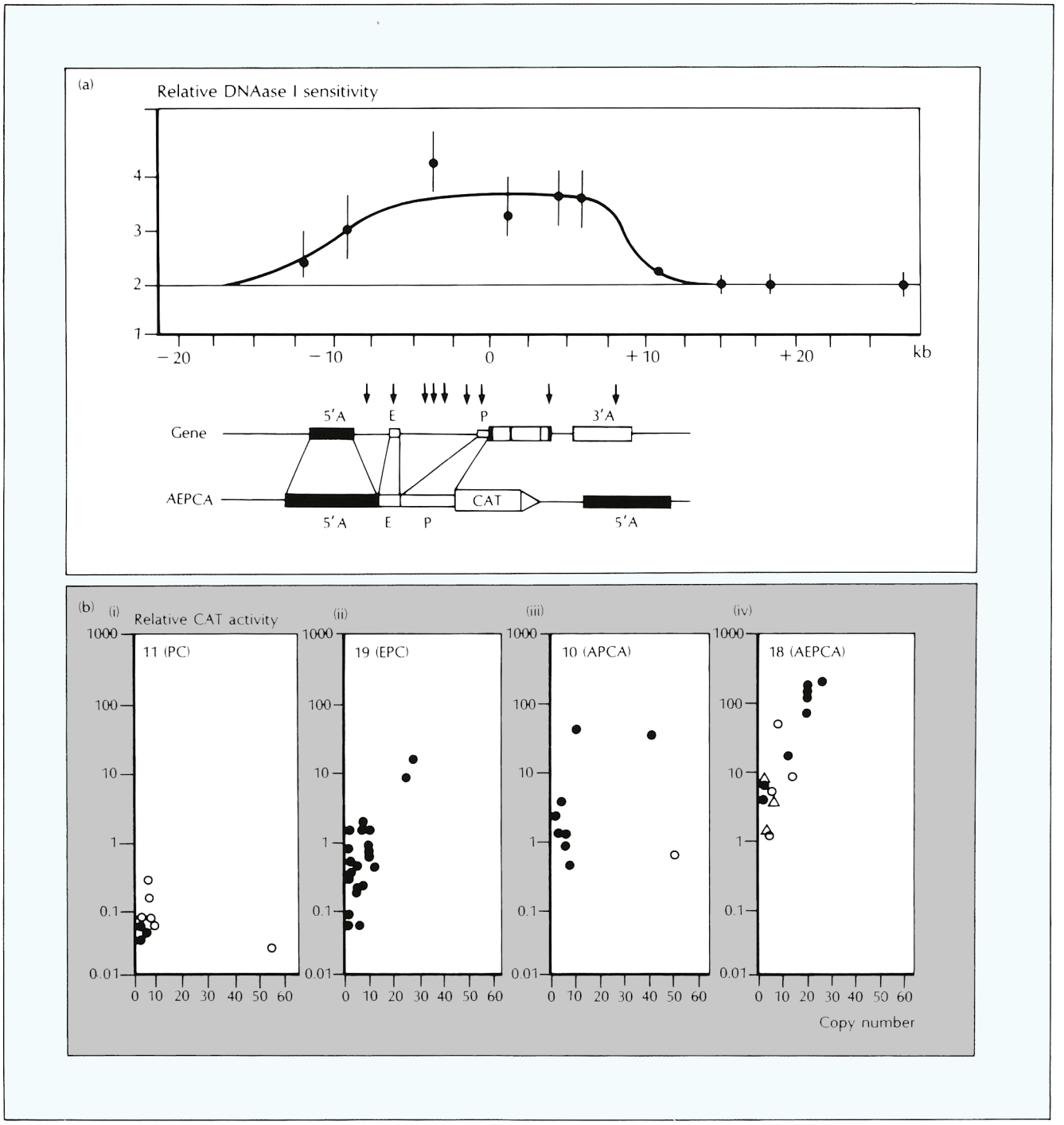

Fig. 2. Domain boundaries can protect transcription units from position effects. (a) Diagram of the genomic organization of the lysozyme gene and of the construct containing an AEPCA minidomain. The broad, DNAase 1-sensitive domain of the wild-type gene is shown in the top line [Fritton *et al*, *Architecture of Eukaryotic Genes* edited by Kahl G, VCH Publishers (UK) Ltd., 1988 pp 333–353]. In its genomic location, the lysozyme domain (−12 kb to +9 kb) is flanked by 5′ (solid black box) and 3′ (shaded box) DNA attachment elements (5′A and 3′A). The enhancer element E (nucleotides −6331 to −5722), promoter element P (nucleotides −579 to +15), and the coding region with exons and introns (box with filled and open bars) are detailed. Arrows mark the positions of DNAase 1-hypersensitive sites in the chromatin of various cell types. In the construct containing the AEPCA mini-domain (not to scale), the reporter gene CAT, which is linked to the lysozyme-gene promoter and enhancer, is flanked by two lysozyme gene 5′A elements (nucleotides −11.7 to −8.7 kb). (b) Effect of A elements on the activity of stably integrated reporter genes. Each dot represents one out of 58 stably transfected HD11/HBC1 cell clones of two independent series of experiments (open and closed symbols). Copy number of correctly inserted DNA is plotted versus relative CAT activity. DNA segments spliced together are the lysozyme enhancer (E) and promoter (P) joined to the reporter CAT gene (C), surrounded by copies of the 5′ attachment element (A) in the order indicated by the letters. (i) 11 clones with inserted PC-containing plasmid DNA; (ii) 19 clones of EPC-containing cells; (iii) 10 clones of APCA-containing cells; (iv) 18 clones of AEPCA-containing cells. Control experiments demonstrated that A elements in integrated constructs containing AEPCA retained the ability to attach to the nuclear matrix or scaffold obtained from lithium diiodosalicylate extracted nuclei. In AEPCA-containing cells, represented by circles, the upstream A element is in the sense orientation; those cells represented by triangles have the A element in the antisense orientation. Modified from [17].

ens the interpretation, but a second intensely studied case appears more complicated. It has been reported that a region containing several DH sites at the 5′ end of the human β-globin gene cluster can, in transgenic mice and cell lines, lead to high-level expression of linked genes in erythroid cells (Grosveld *et al.*, *Cell* 1987, 51:975–985; Forrester *et al.*, *Proc Natl Acad Sci USA* 1989, 86:5439–5443; Collis *et al.*, *EMBO J* 1990, 9:233–240). However, it is not clear how this locus activation region (LAR), also referred to as the dominant control region, functions *in vivo*. In contrast to the lysozyme A elements, some elements within the LAR have strong erythroid-specific enhancer activity in transient expression assays (Tuan *et al.*, *Proc Natl Acad Sci USA* 1989, 86:2554–2558). In addition, it is difficult to assess whether or not the LAR acts as a domain boundary, since the mapping of the human β-globin gene DNAase I-sensitive domain has not been completed. However, in naturally occurring mutations which delete the bulk of the LAR, the β-globin locus remains DNAase I resistant and late replicating in erythroid cells, suggesting that an intact LAR is required for the generation of an early replicating, DNAase I-sensitive domain (Forrester, Epner, Enver, Papayannopoulou, and Groudine, personal communication, 1990).

It has been extremely difficult to correlate results from genetic and/or biochemical analyses with cytological observations. But it has been possible to travel full circle in the case of the yeast silent mating type locus *HML*. In this instance, Gasser and colleagues [18] have been able to observe the formation of DNA loops by reaction of solubilized 'scaffold' proteins with DNA fragments encompassing *HMLα*. The sites of interaction are either silencer–silencer (E–I) or silencer–promoter (E–P and I–P) elements. Affinity purification of scaffold proteins has identified RAP-1 (repressor–activator binding protein; also known as TUF and GRF-1) as binding to these sites; competition experiments have shown that RAP-1 is essential for loop formation [18]. Repression of the mating type loci has been shown to require RAP-1, ABF1 (a factor which binds both to ARS sequences and to some promoters), the four SIR proteins, and, apparently, an intact nucleosome array. While the highly conserved N-terminal portion of histone H4 (amino acids 4–19) can be deleted without loss of viability in yeast, the deletion alters normal chromatin structure, lengthens the cell cycle, and derepresses the silent mating type loci (*HMLα* and *HMRa*) in a gene-specific fashion [19]. Genetic analysis indicates an interaction between a cluster of basic residues in H4 (16–19) and SIR 3 (Johnson, Kayne, Kahn and Grunstein, 1990, personal communication). Thus, the elements required to generate an 'inactive domain' for yeast include, at a minimum, potential DNA domain boundaries (E and I), proteins interacting with those sites, nucleosomes, and proteins interacting with the nucleosomes. The yeast silencing system should yield many further insights into the inactivation process.

Part of the rationale for the domain hypothesis is the notion that enhancer function must be limited in some way. It is hard to visualize how the creation of topological domains could interfere with the protein–protein contacts inferred in most looping models; indeed, topological continuity may not be required for enhancer function (Muller *et al.*, *Cell* 1989, 58:767–777; Dunaway and Dröge, *Nature* 1989, 341:657–679). However, the notion that domain boundaries should block enhancer function is certainly open to experimental test in any of the systems in which stable integration of manipulated genes can be obtained. Indeed, insertion of a mouse MAR element between a UAS and a *GAL1-lacZ* fusion gene results in a striking downregulation in *lacZ* expression (Fishel, Sperry and Garrard, personal communication, 1989). Similar experiments are reportedly in progress in other systems.

Genetic approaches are also leading to the identification of chromosomal proteins required for the formation of heterochromatin in higher eukaryotes. This form of packaging has long been established as a mechanism of gene inactivation, as seen in X-chromosome inactivation in mammals and position–effect variegation in *Drosophila* and other organisms (see review by Eissenberg, *Bio Essays* 1989, 11:14–17). It has been inferred that mutations that suppress or enhance position–effect variegation (the clonal inheritance of inactivation in some cells of a gene translocated next to heterochromatin) will identify loci that encode the chromosomal proteins involved or their modifiers. Indeed, deletions that reduce the number of histone gene copies have been shown to suppress position–effect variegation in some cases (Moore *et al.*, *Genetics* 1983, 104:327–344), but not in others (Rushlow *et al.*, *Genetics* 1984, 108:603–615). Several *Drosophila* loci have been identified that have haplo- and triplo-enhancer/suppressor effects, suggesting a stoichiometric contribution (Locke *et al.*, *Genetics* 1988, 120:181–198; Wustman *et al.*, *Mol Cen Genet* 1989, 217:520–527). One such region covers locus 29A which includes the gene for a protein (HP1) shown by immunofluorescent staining of chromosomes and nuclei to be associated with heterochromatin (James and Elgin, *Mol Cell Biol* 1986, 6:3862–3872; James *et al.*, *Eur J Cell Biol* 1989, 50:170–180). Molecular analysis has shown that a mutation in this gene, *Su(var)205*, which results in suppression of position–effect variegation, causes a defect in HP1 mRNA processing (Eissenberg, James, Foster-Hartnett, Hartnett, Ngan and Elgin, unpublished data). HP1 has a region of homology to *polycomb*, a known downregulator of many homeotic loci, which is found in association with these loci in the euchromatin [20]. Neither of these proteins has been characterized as a DNA-binding protein. In contrast, a second suppressor of position–effect variegation, *Suvar(3)7*, cloned from a chromosome walk, identifies a protein with multiple zinc fingers, implying direct DNA interactions [21]. It will be extremely interesting to learn with what chromatin components these proteins are interacting to generate a compact chromatin structure. New approaches to three-dimensional, time-lapse microscopy hold considerable promise for following the process of heterochromatin formation and chromosome condensation in living *Drosophila* embryos and other cells [22,23]. The next

few years should see a confluence of many of these lines of work, leading to the identification of elements that dictate chromatin structure and so set the pattern for formation of an open and active, or a condensed and inactive, domain.

Acknowledgements

Work from the author's laboratory cited above is supported by NIH grants GM 31532 and HD23844. I thank numerous colleagues, but especially members of the Elgin laboratory, Mark Groudine, and Robert Simpson for comments and criticisms on the draft text.

Annotated references and recommended reading

- • Of interest
- •• Of outstanding interest

1. •• HAN M, GRUNSTEIN M: **Nucleosome loss activates yeast downstream promoters *in vivo*.** *Cell* 1988, 55:1137–1145.
Nucleosome depletion in *S. cerevisiae* activates transcription from the *PHO5*, *CYC1*, and *GAL1* promoters lacking UAS elements. Nucleosome loss appears to increase transcription initiation.

2. • NORRIS D, DUNN B, OSLEY MA: **The effect of histone gene deletions on chromatin structure in *Saccharomyces cerevisiae*.** *Science* 1988, 242:759–761.
Loss of 1 of the 2 pairs of H2A–H2B genes in yeast results in a selective disruption of chromatin structure; nucleosome arrays over *HIS4*, *GAL1*, and telomeres appear normal, while the nucleosome array over *CYH2*, *UB14*, and *CEN3* is severely disrupted.

3. •• SIMPSON RT: **Nucleosome positioning can affect the function of a cis-acting DNA element *in vivo*.** *Nature* 1990, 343:387–389.
When an ARS1 element lies within the central portion of a nucleosome core particle, there is a more than 20-fold decrease in copy number, suggesting a loss of function caused by decreased accessibility.

4. • WOLFFE AP, DREW HR: **Initiation of transcription on nucleosomal templates.** *Proc Natl Acad Sci USA* 1989, 86:9817–9821.
Curved DNA segments are used to move the position of a nucleosome relative to a T7 promoter. Small changes in the position of the nucleosome lead to major changes in promoter efficiency.

5. • BLOOM K, KENNA M, SAUNDERS M: **Cis- and trans-acting factors affecting the structure of yeast centromeres.** *J Cell Sci* 1989, 12 (suppl):231–242.
Given the effects of nucleosome depletion, it is argued that *CEN3* structure is based on a nucleosome. The specific chromatin organization seen at *CEN3* in *S. cerevisiae* is not seen on this DNA in *S. pombe*.

6. • ROUGVIE AE, LIS JT: **The RNA polymerase II molecule at the 5′ end of the uninduced *hsp70* gene of *Drosophila melanogaster* is transcriptionally engaged.** *Cell* 1988, 54:795–804.
Not only is RNA polymerase II already present at the 5′ end of the inactive, but inducible, heat-shock gene *hsp70*, but it may already have initiated transcription, only to be arrested by some other regulatory feature of the locus.

7. • GILMOUR DS, THOMAS GH, ELGIN SCR: ***Drosophila* nuclear proteins bind to regions of alternating C and T residues in gene promoters.** *Science* 1990, 245:1487–1490.
While $(CT)_n$ elements can form a triple helix, they appear to act through the binding of a 66 kD protein which associates with linear DNA.

8. • BONNE-ANDREA C, WONG ML, ALBERTS BM: ***In vitro* replication through nucleosomes without histone displacement.** *Nature* 1990, 343:719–726.
The highly defined bacteriophage T4 replication system is used to demonstrate that, *in vitro*, a replication fork can pass nucleosomes without displacing them. It is suggested that nucleosomes are designed to open up transiently during replication.

9. • WORKMAN JL, ABMAYR SM, CROMLISH WA, ROEDER RG: **Transcriptional regulation by the immediate early protein of pseudorabies virus during *in vitro* nucleosome assembly.** *Cell* 1988, 55:211–219.
IE function requires the simultaneous function of TFIID and results in the formation of stable preinitiation complexes within nucleosome-assembled templates.

10. •• WORKMAN JL, ROEDER RG, KINGSTON RE: **An upstream transcription factor, USF (MLTF), facilitates the formation of preinitiation complexes during *in vitro* chromatin assembly.** *EMBO J* 1990, 9:1299–1308.
If present during assembly, USF increases the number of assembled minochromosomes which contain stable preinitiation complexes. USF appears to function by increasing the rate or stability of TFIID binding.

11. •• CHASMAN DI, LUE N, BUCHMAN AR, LAPOINTE JS, LORCH Y, KORNBERG R: **A yeast protein that influences chromatin structure of UAS_G and functions as a powerful auxilary gene activator.** *Genes Develop* 1990, 4:503–514.
GRF2 binds to sequences found in many UASs, in the 35S RNA enhancer, at centromeres and at telomeres; in binding, it creates a 230 bp nucleosome-free region. While only a weak stimulatory factor on its own, it combines with another weak activator to give 170-fold enhancement of transcription.

12. • NORTON VG, IMAI BS, YAU P, BRADBURY EM: **Histone acetylation reduces nucleosome core particle linking number change.** *Cell* 1989, 57:449–457.
With increasing levels of histone acetylation, the negative linking number change per nucleosome on closed circular DNA decreases from −1.04 to −0.82. This indicates that histone modification can release negative previously constrained supercoils; in effect, the nucleosome can function as a gyrase.

13. • VAN DAAL A, WHITE EM, ELGIN SCR, GOROVSKY MA: **Conservation of intron position indicates separation of major and variant H2As in an early event in the evolution of eukaryotes.** *J Mol Evol* 1990, 30:449–455.
The H2A.Z variant clearly diverged from the S-phase-regulated H2A.1 before *S. pombe* and *Tetrahymena* diverged from the other eukaryotes.

14. • MANNIRONI C, BONNER WM, HATCH CL: **H2A.X, a histone isoprotein with a conserved C-terminal sequence, is encoded by a novel mRNA with both DNA replication type and poly A3′ processing signals.** *Nucleic Acids Res* 1989, 17:9113–9126.
While H2A.X is observed to be a basal histone (expressed throughout the cell cycle), the mRNA shows both the polyA addition site typical of basal histones, and the conserved stem-loop and U7 binding sequences are used for termination of S-phase-regulated histones.

15. • CHALLONER PB, MOSS SB, GROUDINE M: **Expression of replication-dependent histone genes in avian spermatids involves an alternate pathway of mRNA 3′-end formation.** *Mol Cell Biol* 1989, 9:902–913.
A subset of the replication-dependent histone genes is expressed in chicken spermatids. These transcripts include the hairpin structures and purine-rich elements normally used in termination, but have a downstream polyA addition site which is used in these cells.

16. • WELLS D, MCBRIDE C: **A comprehensive compilation and alignment of histones and histone genes.** *Nucleic Acids Res* 1989, 17:r311–r346.
See title.

17. •• STIEF A, WINTER DM, STRATTING WEH, SIPPEL AE: **A nuclear DNA attachment element mediates elevated and position-independent gene activity.** *Nature* 1989, 341:343–345.

The A elements protect a test gene in a stable transforment, but have little effect bracketing a gene if assayed by transient transfection, indicating a structural role.

18. •• HOFMANN JR-X, LAROCH T, BRAND AH, GASSER SM: **RAP-1 factor is necessary for DNA loop formation *in vitro* at the silent mating type locus *HML*.** *Cell* 1989, 57:725–737.

The nuclear protein RAP-1, which fractionates with the nuclear scaffold, binds to E, I and promoter regions in *HML*. RAP-1 plays a role in creating specific DNA loops at *HML*, and is essential for repression of the mating type genes.

19. •• KAYNE PS, KIM U-J, HAN M, MULLEN JR, YOSHIZAKI F, GRUNSTEIN M: **Extremely conserved histone H4 N-terminus is dispensable for growth but essential for repressing the silent mating loci in yeast.** *Cell* 1988, 55:27–39.

Deletions in the hydrophobic core of H4 are lethal and block chromosomal segregation; deletions in the hydrophobic N-terminus are viable. The latter show a lengthened cell cycle and specific depression of the mating type loci.

20. • ZINK B, PARO R: ***In vivo* binding pattern of a trans-regulator of homeotic genes in *Drosophila melanogaster*.** *Nature* 1989, 337:468–471.

The protein product of *polycomb*, a downregulator of many homeotic genes, is shown to be associated with those loci in the polytene chromosomes of *Drosophila*.

21. • REUTER G, GIARRE M, FARAH J, GAUSZ J, SPIERER A, SPIERER P: **Dependence of position-effect variegation in *Drosophila* on dose of a gene encoding an unusual zinc-finger protein.** *Nature* 1990, 344:219–223.

The gene for *Suvar(3)7*, a suppressor of position effect variegation, has been recovered from a chromosome walk. It apparently encodes a protein with 5 widely-spaced zinc fingers.

22. • MINDEN JS, AGARD DA, SEDAT JW, ALBERTS BM: **Direct cell lineage analysis in *Drosophila melanogaster* by time-lapse, three-dimensional optical microscopy of living embryos.** *J Cell Biol* 1989, 109:505–516.

This work demonstrates that, by injection of fluorescently labeled histones, one can label and follow movements and divisions of each nucleus without perturbing development of the *Drosophila* embryo. Potentially, the technique could be used to follow nuclear localization of other chromosomal proteins.

23. • HIROSKA Y, MINDEN JS, SWEDLOW JR, SEDAT JW, AGARD DA: **Focal points for chromosome condensation and decondensation revealed by three-dimensional *in vivo* time-lapse microscopy.** *Nature* 1989, 342:293–296.

High-resolution light microscopy can be used to follow chromosome condensation and decondensation in nuclei stained with labelled histone [22]. Imaging at low light levels can be accomplished using a cooled charge-coupled device imager; three-dimensional data sets can be obtained using computational processing of optical sections.