ELSEVIER

Biophysical Chemistry 100 (2003) 607–613

Biophysical Chemistry

www.elsevier.com/locate/bpc

Quantitative approaches to problems of eukaryotic gene expression

Gary Felsenfeld*

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0540, USA

Received 24 December 2001; accepted 28 February 2002

Abstract

During the past several years there has been intense interest in the mechanisms by which gene expression is regulated within the eukaryotic nucleus. We have made use of an avian β-globin locus to study various aspects of this problem, some of which are amenable to quantitative analysis. In the course of this work we have identified the transcription factor GATA-1, which is an essential regulatory factor for virtually all erythroid-specific genes, and studied the structure of its complex with its specific DNA binding site. The way in which GATA-1 forms tight interactions with DNA led to an understanding of how other zinc finger proteins of this class bind to DNA. We have extended such studies to examine interactions with DNA packaged as chromatin, and to studies of chromatin structure and function at higher levels of organization.

Keywords: Gene expression; Eukaryotic; Nucleus; Chromatin

1. Introduction

The study of regulation of transcription and replication in eukaryotic cells has increasingly revealed a very complex set of molecular structures and mechanisms. It is, therefore, difficult to identify problems that are amenable to analysis by the methods of physical chemistry. Systems sufficiently well defined and 'clean' enough to make such measurements worth while are often too simplified to reflect the realities of the intracellular environment. During the past years our laboratory has explored systems too complicated for the physical chemist, partly in hope that intimate knowledge would make it clear what simplifications were justified. This work has focused to a large extent on the regulation of gene expression at β-globin gene loci, but the intent has been to use that as a base from which to explore especially the relationship between chromatin structure and control of transcription.

We chose the chicken β-globin locus in part because at the time we began this work its chromatin structure had already received some attention. Furthermore, chicken erythroid cells are nucleated, and the nuclei of these cells have rather low levels of proteases and nucleases, so that relatively undergraded chromatin can be isolated. The β-globin locus contains four globin genes, two embryonic genes (ε and ρ) expressed in the primitive lineage, and two adult genes (β^H and β^A) in the definitive lineage (Fig. 1). Our earliest

*Corresponding author. Tel.: +1-301-496-4173.
E-mail address: gary.felsenfeld@nih.gov (G. Felsenfeld).

0301-4622/03/$ - see front matter
PII: S0301-4622(02)00309-5

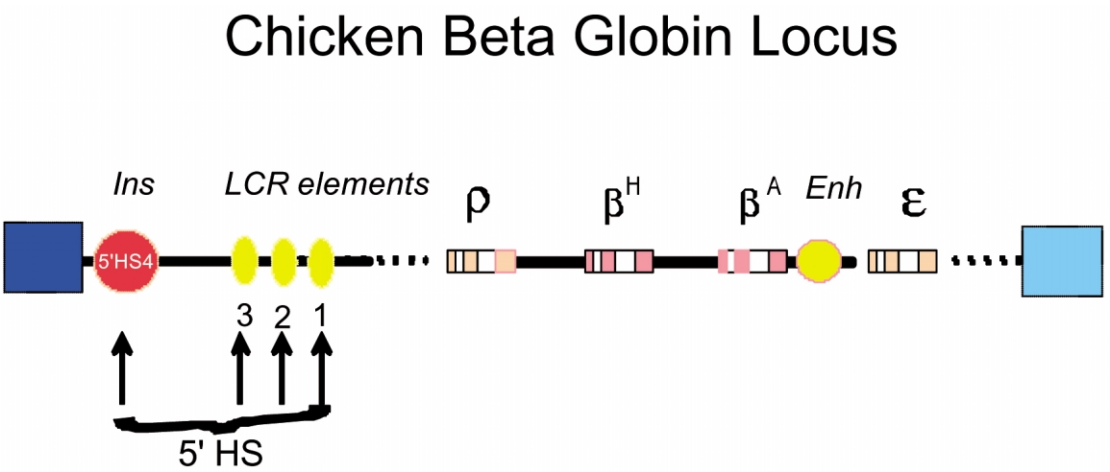

Fig. 1. Organization of the chicken β-globin locus, showing the four members of the β-globin family. Two of the genes (β^H and β^A) are expressed in the definitive lineage, the other two (ε and ρ) in the primitive lineage. A strong enhancer lies between β^A and ε. This enhancer also serves as one element in the locus control region (LCR) of the globin locus. The other three LCR components (1, 2 and 3) are located upstream of the genes. Still further upstream lies a nuclease hypersensitive site (5′HS4) marking the 5′ insulator or boundary element of the locus, and beyond that an extended region of condensed chromatin [18,19].

studies were designed to identify transcriptional regulatory elements and their associated DNA binding proteins [1–5]. Of particular interest was a DNA binding motif with the consensus sequence (A/T)GATA(A/G) that was found to be present in proximity to each of the globin genes [6,7]. We purified the single protein that bound to it, and we and others cloned the corresponding gene in chicken, mouse, and human, which was ultimately given the name GATA-1. GATA-1 turned out to be the first member of an important family of proteins that play central roles in a variety of developmental regulatory processes. GATA-1 is important in erythroid development, activating transcription from a wide variety of erythroid genes. It is a 304 amino acid protein with two zinc fingers of the cys_2–cys_2 class. At sites with a single (A/T)GATA(A/G) motif, GATA-1 typically interacts with DNA through the C-terminal zinc finger. This does not seem to be enough to account for the observed binding strength, and in fact there is an important additional interaction with DNA involving the quite basic C-terminal tail of the protein. NMR studies of a complex with a truncated GATA-1 carrying only the C-terminal finger and tail show that whereas the α-helical region near the finger makes typical contacts with bases in the DNA major groove, this tail makes additional contacts in the minor groove ([8]; Fig. 2). Deletion of the tail dramatically reduces affinity for DNA.

Variations in this binding scheme appear in other members of the GATA family. In GATA-1 only the C-terminal finger with its basic extension is necessary and sufficient for binding to DNA; the N-terminal finger alone is neither necessary nor sufficient. In contrast, the N-terminal fingers of GATA-2 and GATA-3 are able to bind independently [9]. This is because each possesses two basic regions, one on either side of the finger, both of them required for binding. We found a rather similar arrangement in an unrelated protein, the GAGA-binding regulatory factor from *Drosophila* [10]. This protein, the product of the trithorax-like gene, has a Cys_2–His_2 zinc finger. It had previously been thought that single fingers of this class were unable to bind DNA specifically, but GAGA protein, through this zinc finger, recognizes sites containing a central GAGAGAG motif. We showed that this specific binding involves both α-helical contacts in the DNA major groove and contacts from an adjacent N-terminal basic region, once again in the minor groove. GATA-1 itself

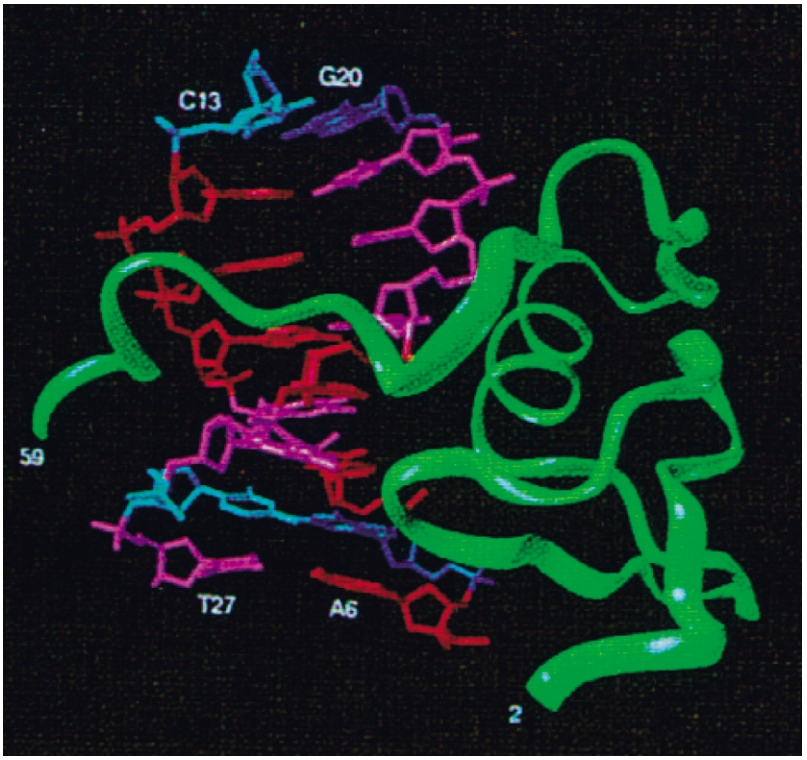

Fig. 2. Structure of the C-terminal region of GATA-1, as determined by NMR, complexed with its target site on DNA (from [8]). An α-helical domain makes contacts in the major groove of DNA. The C-terminus of the GATA-1 fragment extends toward the left in this figure, making additional contacts in the minor groove that are critical for strong binding.

provides an example of the extreme versatility of even the simplest regulatory proteins in their interactions with DNA. Although the C-terminal finger is sufficient as well as necessary for binding to a single GATA consensus site, the N-terminal finger can be engaged when a second binding site is present, leading to tighter binding [11]. Thus, increased selectivity can be achieved even with this relatively small molecule.

The problem of protein–DNA interactions becomes more complicated when the role of chromatin is considered. The fundamental chromatin subunit is the nucleosome. A 'core' particle consists of 147 bp of DNA wrapped in two superhelical turns around an octamer of the basic proteins called histones. Such a structure presents an impediment to the binding of transcription factors and a barrier to transcription. Numerous recent studies make it clear that there are multiple factors that can alter nucleosome structure or chemically modify the histones to change the way in which they interact with associated DNA [12]. However, certain interactions between unmodified nucleosomes and transcription factors also illuminate the nature of the initial complexes that are formed. To study GATA-1 interactions with chromatin we reconstituted nucleosome core particles with DNA containing six single GATA-1 sites [13]. Most experiments were carried out with the C-terminal part of the DNA binding domain, which as noted above can bind strongly to such sites. We found that this single finger was also able to form stable complexes with the nucleosome DNA, involving occupancy of multiple site (Fig. 3). Surprisingly, although the nucleosome structure was perturbed as judged by increased sensitivity to nucleases, the native structure could be restored by competing away the bound GATA finger with DNA that

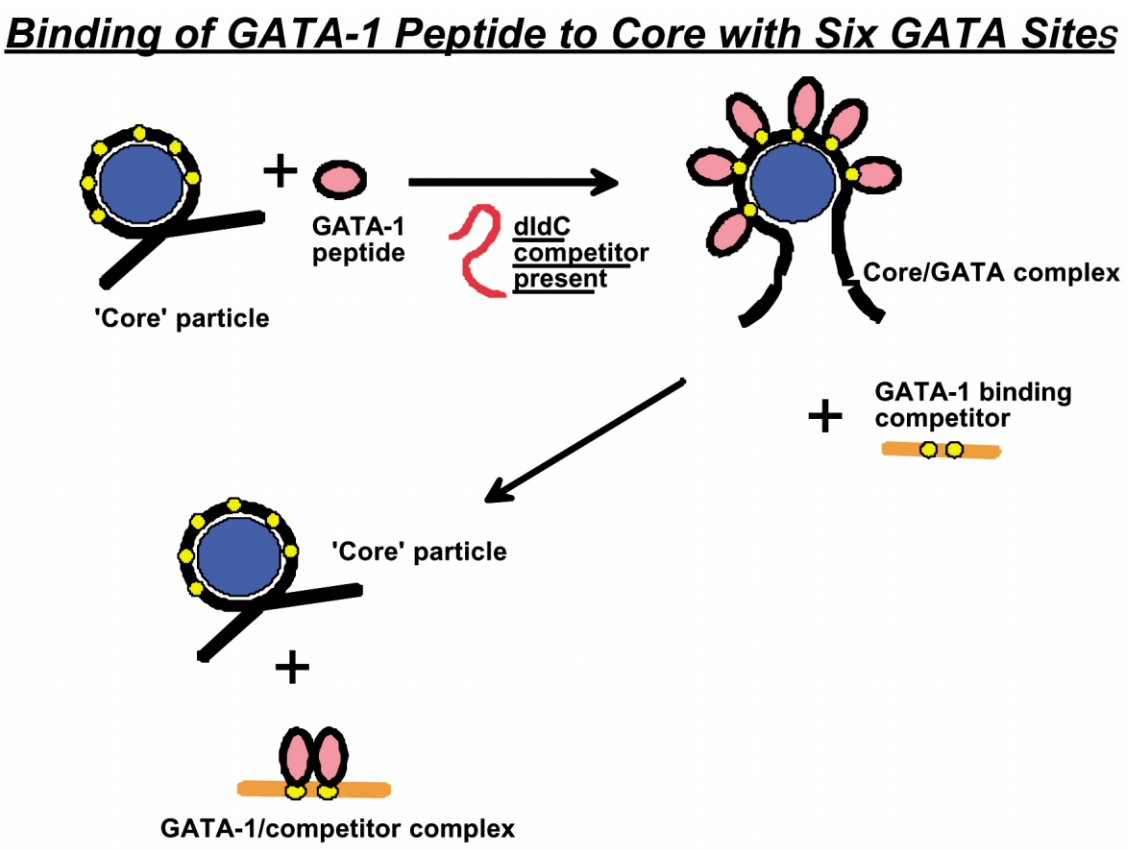

Fig. 3. The effect of adding GATA-1 peptide containing the C-terminal region to nucleosome core particles reconstituted with DNA carrying six GATA-1 binding sites. The resulting complex is more susceptible to attack by micrococcal nuclease than the core particle itself. If DNA competitor containing two GATA-1 sites is added, the original core particle is regenerated. Since loss of DNA from the core is irreversible under these conditions, the results show that a stable complex involving histones, DNA and GATA-1 peptide must have formed (from Boyes et al. [13]).

contained GATA sites. The interaction was, thus, completely reversible. These results are quite consistent with what is known about the ability of nucleosome-bound DNA to 'breath' free of the nucleosome surface at the DNA termini [14–16], while the central DNA region remains attached to the histone octamer. This transient release makes available binding sites that might otherwise be blocked; the probability of access decreases as one moves toward the center of the nucleosome. The nucleosome complex with GATA is probably held together by unperturbed DNA-histone contacts toward the center of the particle. Within the cell such reactions may well contribute to the interaction between transcription factors and regulatory sites on DNA; cells also contain complexes capable of mobilizing nucleosomes so that (perhaps randomly) sites become more accessible. The combination of the two processes may be involved in the initial stages of recognition.

Although it is possible to do informative physical chemistry of simple chromatin structures, it is more difficult to produce well-defined systems at higher levels of chromatin organization. Nonetheless, some reliable information can be obtained, although interpretations may be more ambiguous. One problem that has attracted our attention concerns the generation of nuclease hypersensitive

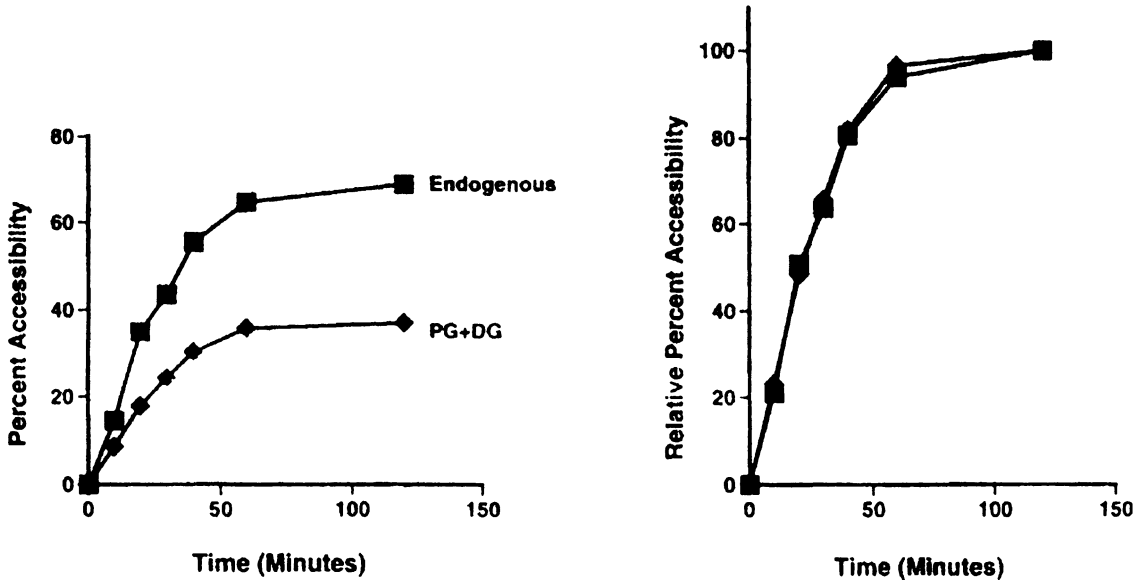

Fig. 4. Digestion kinetics for the action of the restriction enzyme Pvu II on nuclei from avian 6C2 cells. The cells are stably transformed with a reporter carrying the α-globin gene and its downstream enhancer (see Fig. 1), but with the pair of GATA-1 binding sites in the enhancer mutated (left, curve marked 'PG+DG'). The endogenous enhancer in 6C2 sites is hypersensitive to nucleases, and the sensitivity of the Pvu II site reflects this (left, curve marked 'Endogenous'). The mutation results in a reduction by almost half in the fraction of sites that are accessible. However, when the two curves are normalized to the same plateau value (right) it is apparent that the rate of cleavage of the accessible sites is identical for the mutated construct and the endogenous site (from Boyes and Felsenfeld [17]).

sites within chromatin. These sites typically mark points where nucleosomes have been disrupted or displaced, making DNA more accessible to regulatory factors and to nucleases. The binding of regulatory factors often will distort the DNA in such a way as to make it more easily cleaved by enzymes, adding to the hypersensitive response. Typically the hypersensitive response is stronger (sites are more readily cleaved) when more transcription factors are bound. The significance of this result is clouded by the fact that experiments are usually carried out with DNaseI, an enzyme that eventually cleaves all DNA and, therefore, makes it difficult to carry out a quantitative experiment. To avoid this problem, we turned to restriction enzymes for our investigation of the behavior of an enhancer that activates transcription of both the β- and ε-globin genes in chicken [17]. Within this enhancer are binding sites for two important erythroid factors, NF–E2 and GATA-1 (two sites for the latter). We constructed reporters that contained either the wild type enhancer or enhancers from which one or more of these binding sites had been deleted. These were stably transformed into 6C2 cells, a chicken erythroid cell line. Nuclei were isolated and digested with restriction enzymes expected to cut within this region if it is in a hypersensitive state, but not otherwise. Analysis of the kinetics of cleavage (Fig. 4) shows that as the strength of the enhancer is reduced by mutating factor binding sites, there is a decrease in the number of sites available for cutting. This is reflected in the decreased plateau values. However, the rate of cleavage of those sites that are available remains the same: If all data are scaled to the plateau values, the curves are quite similar. We conclude that the lowered hypersensitivity that arises when enhancer strength is reduced derives from a reduction in the number of accessible sites. Those that are available are cut at the full rate; the

rest are completely unavailable. This behavior could be connected to nucleosome positioning. It suggests that chromatin structure can be used to change continuously variable signals into sharp all-or-none transitions.

At higher levels of organization within the nucleus, new problems arise. The genes in eukaryotic and particularly vertebrate genomes are often interspersed with regions of condensed chromatin. In addition, genes or gene families are often sufficiently close together in the genome that powerful enhancers, capable of functioning at considerable distances from promoters, might be able to activate genes from adjacent families inappropriately. During the past several years we have been studying the properties of insulators, a varied group of sequence elements capable either of preventing the encroachment of condensed chromatin on active regions (and consequently work against certain kinds of position effect), or of blocking the action of an enhancer on a promoter when placed between them. The chicken β-globin locus once again proves a useful system in which to study these phenomena. We have identified at its 5′ end, just at the point where the 'open' chromatin domain ends and condensed chromatin begins [18,19], a compound sequence element that possesses both properties [20–22]. The ability to block enhancers is conferred by binding of CTCF, a protein with 11 zinc fingers, also found at critical enhancer-blocking sites elsewhere in the genome [23]. How CTCF carries out this function is as yet unknown, but it seems likely that when its mechanism of action is understood we will also understand a great deal more about how enhancers work. The ability to protect against position effect involves other proteins as yet uncharacterized, but surprisingly we have more clues about this function, which may be connected to an ability to maintain a high state of histone acetylation in adjacent nucleosomes [24].

Few of these higher order structures are amenable to analysis by the methods of physical chemistry. Recently, however, we have begun a study of the extended fragment of condensed chromatin found upstream of the β-globin locus. It is proving possible to excise this fragment from the nucleus and follow its hydrodynamic properties by using either Southern blotting or PCR techniques. It is our hope that with the availability of gene-specific probes of high sensitivity and with the aid of high-resolution preparative ultracentrifugation it may be possible to carry out precise physical measurements on specific genomic sequences as they are packaged within the nucleus.

2. Conclusion

I have tried to follow the thread of physicochemical reasoning and method that runs through some of the work we have carried out over recent years. Even in the case of very complicated systems some parts of the problem will yield to such an approach. John Edsall was one of the earliest to recognize the coming power of physical and molecular methods in solving biological problems, and one of the first to integrate contemporary advances in physics into his thinking. I was very fortunate that he was my tutor during three undergraduate years at Harvard. Tutorial hours with him were the most exciting and enduring experiences of my education. He taught me his way of thinking and strongly influenced me in choosing a scientific career. For both I am deeply grateful.

References

[1] J.D. McGhee, W.I. Wood, M. Dolan, J.D. Engel, G. Felsenfeld, Cell 27 (1981) 45–55.
[2] B.M. Emerson, J.M. Nickol, P.D. Jackson, G. Felsenfeld, PNAS 84 (1987) 4786–4790.
[3] B.M. Emerson, J.M. Nickol, T.C. Fong, Cell 57 (1989) 1189–1200.
[4] B.M. Emerson, C.D. Lewis, G. Felsenfeld, Cell 41 (1985) 21–30.
[5] B.M. Emerson, G. Felsenfeld, PNAS 81 (1984) 95–99.
[6] T. Evans, G. Felsenfeld, Cell 58 (1989) 877–885.
[7] T. Evans, M. Reitman, G. Felsenfeld, PNAS 85 (1988) 5976–5980.
[8] J.G. Omichinski, G.M. Clore, O. Schaad, et al., Science 261 (1993) 438–446.
[9] P.V. Pedone, J.G. Omichinski, P. Nony, C. Trainor, A.M. Gronenborn, G.M. Clore, G. Felsenfeld, EMBO J. 16 (1997) 2874–2882.
[10] P.V. Pedone, R. Ghirlando, G.M. Clore, A.M. Gronenborn, G. Felsenfeld, J.G. Omichinski, Proc. Natl. Acad. Sci. USA 93 (1996) 2822–2826.
[11] C.D. Trainor, J.G. Omichinski, T.L. Vandergon, A.M. Gronenborn, G.M. Clore, G. Felsenfeld, Mol. Cell. Biol. 16 (1996) 2238–2247.
[12] J.D. Aalfs, R.E. Kingston, Trends Biochem. Sci. 25 (2000) 548–555.

[13] J. Boyes, J. Omichinski, D. Clark, M. Pikaart, G. Felsenfeld, J. Mol. Biol. 279 (1998) 529–544.
[14] K.J. Polach, J. Widom, J. Mol. Biol. 254 (1995) 130–149.
[15] K.J. Polach, J. Widom, Methods Enzymol. 304 (1999) 278–298.
[16] R.U. Protacio, K.J. Polach, J. Widom, J. Mol. Biol. 274 (1997) 708–721.
[17] J. Boyes, G. Felsenfeld, EMBO J. 15 (1996) 2496–2507.
[18] T.R. Hebbes, A.L. Clayton, A.W. Thorne, C. Crane-Robinson, EMBO J. 13 (1994) 1823–1830.
[19] M.N. Prioleau, P. Nony, M. Simpson, G. Felsenfeld, EMBO J. 18 (1999) 4035–4048.
[20] J.H. Chung, A.C. Bell, G. Felsenfeld, Proc. Natl. Acad. Sci. USA 94 (1997) 575–580.
[21] A.C. Bell, A.G. West, G. Felsenfeld, Science 291 (2001) 447–450.
[22] A.C. Bell, A.G. West, G. Felsenfeld, Cell 98 (1999) 387–396.
[23] E.M. Klenova, R.H. Nicolas, H.F. Paterson, et al., Mol. Cell Biol. 13 (1993) 7612–7624.
[24] M.D. Litt, M. Simpson, F. Recillas-Targa, M.N. Prioleau, G. Felsenfeld, EMBO J. 20 (2001) 2224–2235.