Transcription factor access to chromatin

Miguel Beato* and Karin Eisfeld

Institut für Molekularbiologie und Tumorfororschung, Philippus Universität, E.-Mannkopff-Strasse 2, 35037 Marburg, Germany

Received January 24, 1997; Revised and Accepted July 18, 1997

ABSTRACT

The question of how sequence-specific transcription factors access their cognate sites in nucleosomally organized DNA is discussed on the basis of genomic footprinting data and chromatin reconstitution experiments. A classification of factors into two categories is proposed: (i) initiator factors which are able to bind their target sequences within regular nucleosomes and initiate events leading to chromatin remodelling and transactivation; (ii) effector factors which are unable to bind regular nucleosomes and depend on initiator factors or on a pre-set nucleosomal structure for accessing their target sequences in chromatin. Studies with the MMTV promoter suggest that the extent and number of protein–DNA contacts determine whether a factor belongs to one or the other category. Initiator factors have only a few DNA contacts clustered on one side of the double helix, whereas effector factors have extensive contacts distributed throughout the whole circumference of the DNA helix. Thus, the nature of DNA recognition confers to sequence-specific factors their specific place in the sequential hierarchy of gene regulatory events.

INTRODUCTION

DNA in the cell nucleus is organized and compacted in chromatin. The question of how eukaryotic transcription factors gain access to their target sequences in the chromatin context has attracted considerable attention. There are several levels of structural organization in chromatin, which could contribute to modulate accessibility of binding sites for transcription factors. In this review we will focus on the role of the first level of compaction of DNA in chromatin, its packaging into nucleosomes. The basic unit of chromatin is the nucleosome core particle, which consists of an octamer of core histones (a tetramer of histones H3 and H4 and two dimers of histones H2A and H2B) around which are wrapped 146 bp of DNA in 1.75 left-handed superhelical turns (1). In chromatin of differentiated cells a linker histone, H1, H5 or variants thereof, joins the core particle asymmetrically to yield a nucleosome in which an additional 20 bp are protected against nuclease digestion (2). In many promoter and enhancer regions nucleosomes have been found to be non-randomly positioned in that they are located over defined regions of DNA (for reviews see 3–5). A database for positioned nucleosomes is available from the EMBL Sequence Database, which in 1993 already contained 143 entries (6).

Two parameters are used to define the position of nucleosomes along DNA sequences, translational and rotational positioning. Translational positioning defines the region of DNA within a given sequence which is wrapped around the histone octamer, whereas rotational positioning refers to the angular orientation of individual base pairs on the nucleosome surface. The parameters determining nucleosome positioning are not well understood, but the bendability and anisotropy of the nucleotide sequence play a role at least for rotational positioning, as algorithms have been developed which predict with some confidence the orientation of DNA sequences in nucleosomes (see 7,8 and references therein). The parameters responsible for translational positioning are less clear, but the nucleotide sequence is also involved. In the classical example of positioned nucleosomes, namely the 5S rRNA gene from Lytechinus variegatus, the central 50–60 bp are responsible for translational positioning (9). The central DNA over the nucleosome dyad has to be deformed and underwound to provide the best fit (10–12) and this structural requirement may influence the translational position of nucleosomes by giving some DNA sequence a higher preference for the centre of the nucleosome. In addition, histone H1 can also influence nucleosome positioning (13,14).

Both translational and rotational positioning could influence the accessibility of factors to individual binding sites in chromatin. Translational positioning determines whether a factor binding site is included within the confines of a nucleosome particle or is rather located in the linker DNA joining two nucleosomes. When a binding site is located within a nucleosome its accessibility would be influenced by rotational orientation of the elements within the binding site which are contacted by the factor. Factor binding would be facilitated if these elements are exposed on the surface of the particle and may be more difficult when the elements are oriented towards the nucleosome interior, facing the histone octamer. On the other hand, when a binding site is located in the linker DNA its accessibility may be influenced by linker histones, HMG proteins or by other parameters which control the degree of compaction of the nucleosomal fibre.

IN VIVO INTERACTIONS

High resolution chromatin analysis and genomic footprinting experiments have led to the distinction between a minority of

*To whom correspondence should be addressed. Tel: +49 6421 28 62 86; Fax: +49 6421 28 62 86; Email: beato@imt.uni-marburg.de
Factors able to bind to their target sites when these are organized in regular chromatin and a majority of factors which lack this property and need remodelled or preset chromatin structures for binding (15). In discussing the literature we will focus on sequence-specific transcription factors and will not refer to results with general transcription factors. In particular, we will not discuss experiments performed with the TATA box binding protein TBP (16,17), as its strong DNA bending effects complicate interpretation of the results.

De novo initiators of chromatin remodelling

The steroid hormone receptors belong to the class of remodelling initiators which can bind to regularly organized chromatin (18). In many hormonally regulated genes hormone induction leads to a change in chromatin structure reflected in the appearance of DNase I hypersensitive regions over the promoter or enhancer regions. This has been interpreted as a consequence of either disruption, unfolding or rearrangement of nucleosomes (19–21). A classical example is induction of the MMTV promoter by glucocorticoids (19,22) or progesterone (23), but many other hormonally inducible systems have been described, including the chicken vitellogenin gene (24), the rat tyrosine aminotransferase gene (25) and the rabbit urotoglobin gene (26).

The MMTV promoter is organized into a regular array of positioned nucleosomes and there is no indication of proteins other than histones bound to DNA prior to hormone induction (23). This suggests that the events leading to chromatin remodelling are initiated by hormone binding to the receptor followed by interaction of the receptor with the target DNA in regular chromatin (23). Moreover, the hormone-responsive elements (HREs) to which the receptors bind are included within a positioned nucleosome (19,23,27) and, therefore, the receptors must be able to bind to their recognition sequences on the surface of nucleosomes in order to initiate remodelling. Activation can also be accomplished by moderate hyperacetylation of core histones in response to inhibitors of histone deacetylases (28), underlining the role of chromatin in transcriptional control of this promoter.

There is no other well-documented example of a factor able to bind to a regular positioned nucleosome. The rat albumin enhancer is organized into a regular nucleosomal array during liver embryogenesis, simultaneously with liver-specific expression of HNF3 (29,30). Although later in development and in the adult liver HNF3 is bound to the surface of a positioned nucleosome, there is no information about the structure to which the factor originally binds during early development and what other factors may be involved in generating a precise nucleosomal array. Moreover, the process has not been shown to be reversible. As HNF3 shares a structural motif, the winged motif, with linker histones, the possibility remains that it fulfils a histone-like function on the albumin enhancer.

Factors unable to bind to regular chromatin

Many transcription factors mediating gene induction interact with remodelled or preset chromatin structures generated prior to the inducing stimulus. This large group of factors could be divided in two subgroups: those which are able to bind to altered nucleosomes and those which bind to ‘nucleosome-free’ regions of chromatin.

Binding to altered nucleosomes. Induction of MMTV by glucocorticoids or progestins not only requires binding of the receptors to the HREs but also a synergistic interaction with other ubiquitous factors, in particular NF1, binding to adjacent sites (31). NF1 exhibits a higher affinity than the hormone receptors for naked promoter DNA but cannot interact with its target site in nucleosomes (32). This is likely due to the large number of DNA contacts found in the NF1 binding site (see Eisfeld et al., this issue). However, upon binding of the hormone receptors a change in the structure of the promoter nucleosome takes place which enables NF1 to access its binding site, while the nucleosome remains in place (23).

We suspect that many high affinity DNA binding proteins behave like NF1; they cannot bind to their sites in regular chromatin without the previous action of initiators which promote chromatin remodelling (33). SP1, for instance, a ubiquitous transcription factor which binds to GC boxes in many genes, is not found constitutively bound to inducible promoters prior to induction but occupies its site upon induction (A.Scholz, M.Truss and M.Beato, unpublished observation). However, it has not been shown that Sp1, or any other of these factors, can bind to remodelled nucleosomes without disrupting their structure.

An interesting example of this group is the binding of Amt1 from Candida glabrata to the metal regulatory element (MRE) in its own promoter (34). Amt1 is activated by copper binding and stimulates its own production by binding as a monomer to this MRE. The MRE is contained within a positioned nucleosome and rapid access of Amt1 to the binding site depends on a neighbouring stretch of 16 A-T base pairs, which makes the adjacent nucleosomal DNA more accessible to nucleases or other factors (35). No factor binding to the A-T stretch is detected and the sequence can be inverted without losing its permissive function, which may be more general (36). However, replacement of the A-T stretch by B-DNA sequences generates a regular nucleosome and precludes rapid binding of Amt1 to the MRE. It seems that the rigid A-T stretch generates a deformed nucleosome to which Amt1 can easily bind. Since the nucleosome remains in place upon Amt1 binding and transcriptional induction, this system resembles binding of NF1 to the nucleosome on the MMTV promoter following hormone-induced remodelling.

Binding to ‘nucleosome-free’ DNA in chromatin. According to the present view, the large majority of transcription factors belongs to this group, but one should keep in mind that ‘nucleosome-free’ does not necessarily mean histone-free, as demonstrated in the active ribosomal genes by psoralen and UV-laser crosslinking (37,38). Within this group one can distinguish factors exhibiting inducible chromatin binding from those constitutively bound to chromatin.

To the inducible class belong, for instance, Pho4 and the heat shock factor. The yeast factor Pho4 mediates induction of the PHO5 promoter as a consequence of low phosphate in the medium. This factor is unable to bind to its target sequence when organized in nucleosomes (39) and its action depends on the existence of a binding site in a 70 bp nucleosome-free region between positioned nucleosomes which is constitutively DNase I hypersensitive (40). How this structure is generated is not clear, but following binding of Pho4 to this element remodelling is initiated that allows binding of Pho4 to an additional nucleosomal site and also exposes binding sites for Pho2. These events eventually lead to disruption of four consecutive nucleosomes, two on each side of the internucleosomal Pho4 site (41), probably mediated by ATP-dependent chromatin remodelling activities.
which are recruited by the transactivation domain of Pho4 (42). If the Pho4 site in the ‘nucleosome-free’ region is deleted induction of the PHO5 promoter is strongly reduced (39). Thus although Pho4 can initiate chromatin remodelling, it depends for doing so on a preset chromatin structure, because it cannot bind to regular nucleosomes.

A similar situation is found in the promoters of the heat shock genes in *Drosophila* (21). The promoter of the hsp26 gene exhibits a positioned nucleosome flanked by two constitutive DNase I hypersensitive sites, each including binding sites for the heat shock factor (HSF) and for the GAGA factor. How this structure is generated is not known, but the promoter contains bound TFIIID over the TATA box prior to heat shock, demonstrating that the chromatin structure has been developmentally preset, probably in a process involving the GAGA factor (43). The positioned nucleosome could serve to bring the two HSF binding sites and the core promoter elements into close apposition, to facilitate rapid and efficient induction upon heat shock and HSF binding (44). A similar situation may prevail in the *Drosophila* alcohol dehydrogenase gene (45) and in the *Xenopus* vitellogenin gene (46). In this latter case organization of the promoter in nucleosomes has been shown to enhance its transcription in vitro (46). In all these cases the transactivating factor binds to nucleosome-free regions previously established by other factors.

In cases where transcription factors are found constitutively bound to ‘nucleosome-free’ DNA they likely bind to their cognate sites during chromatin replication at critical mitosis and set the stage for a rapid response of the corresponding genes. This kind of situation applies, for instance, to the array of proteins bound to the serum response element of the *c-fos* gene promoter, which is preformed prior to serum induction and thus guarantees a rapid response (47). To this class also belongs the GAL1/GAL10 regulatory region, in which GAL4 is constitutively bound to a nucleosome-free region as an inactive complex with GAL80 (48,49).

In infected T cells and macrophages the HIV-1 5′-long terminal repeat (LTR) exhibits constitutive DNase I hypersensitive regions located between positioned nucleosomes, which have been interpreted as ‘nucleosome free’ (50). One of these regions, between the start site of transcription and position −250, is about the size of a nucleosome and encompasses binding sites for many factors, including sites for NF-κB and Sp1. Upon induction of HIV transcription by TNFα, which leads to NF-κB activation, there is remodelling of chromatin manifested as disruption of a downstream nucleosome over the adjacent transcribed region (50). In transcription experiments with chromatin templates NF-κB p65, but not p50, activates chromatin remodelling and transcriptional derepression synergistically with Sp1 (51). However, the NF-κB and Sp1 sites are occupied *in vivo* regardless of activation of the promoter (52) and there is no evidence yet that NF-κB binds to a nucleosomally organized site *in vivo*. It is possible that the region identified as DNase I hypersensitive encompasses a preset nucleosome that, as in the case of induced MMTV, is more accessible to DNase I. Intriguingly, a potential glucocorticoid-responsive element has been identified around position −250, immediately upstream of the hypothetical nucleosome, with the NF-κB binding site (53,54). Nothing is known about the effect of glucocorticoids on remodelling of HIV chromatin structure but, like the MMTV promoter, the HIV-1 promoter can be activated by inhibitors of histone deacetylases, suggesting regulation at the level of chromatin (55). Thus the classification of this system awaits further experimental evidence.

How the preset chromatin structure is generated during mitosis and maintained through cell division is unknown, but there are proteins which possess the properties required to accomplish this function. In yeast a special factor, Gr2, has been postulated to act by serving as a barrier to organization of arrays of nucleosomes, thus generating nucleosome-free regions (56,57). In higher organisms several HMG domain-containing proteins have been described capable of organizing a regulatory DNA region into a stable structure, to serve as a platform for transcription factor binding (58), and there are indications that HMG1 can direct nucleosome positioning (59).

### IN VITRO BINDING STUDIES

Is there a molecular basis for the distinction between the different classes of factors in terms of their ability to interact with their target sites in chromatin? Data obtained with nucleosomes assembled *in vitro* on defined DNA fragments could help to understand these differences. Several transcription factors have been shown to be able to interact with their cognate sites when they are assembled into nucleosome core particles. To this class of factors belong the glucocorticoid receptor (32,60), the progesterone receptor (61,62), the thyroid hormone receptor (63), USF (64,65), Sp1 (65,66) and GAL4 (67). For all these factors the affinity for nucleosomally organized target sites is lower than for the same sites on free DNA. Other factors, such as HSF, are unable to bind to nucleosomally organized sites (67). However, caution is required when considering these results. First, except for the experiments with nuclear receptors and some experiments with 5S rRNA genes, none of the other binding sites have been included in natural sequences and most experiments have been performed without controlling nucleosome positioning and with artificial clustering of factor binding sites. The physiological significance of these studies awaits further evaluation. Second, binding results are very much dependent on the experimental conditions and a model based on kinetic measurements suggests that nucleosomes are unstable *in vitro* and that any site can be accessed by restriction enzymes provided longer incubation times are used (68,69).

With this in mind, I will concentrate on those factors for which *in vivo* data are available and the DNA binding mechanism is well known. Initial nucleosome binding studies were performed with the *Xenopus* 5S RNA gene reconstituted in positioned mono-nucleosomes and the transcription factor TFIIIA (70). TFIIIA is a zinc finger protein which binds to positions 45–97 within the gene. As the positioned nucleosome extends to position 78, a 35 bp overlap between the TFIIIA binding site and nucleosome must be postulated. A ternary complex can be formed containing the nucleosomally bound factor, as shown by difference probability plots of DNase I footprints. According to these data, TFIIIA would be able to bind to its binding site on nucleosomes invading the last 35 bp normally covered by the core particle (70). This result has been questioned, in that binding to a regular nucleosomal 5S RNA gene was not observed, but TFIIA was shown to interact with the 5S gene in acetylated nucleosomes (71) and in tetramers of histone H3 and H4 (72). Moreover, TFIIA cannot bind to human 5S rRNA genes organized in normal nucleosomes.
Unfortunately, no in vivo data are available on TFIIIA binding due to the existence of multiple copies of the SS rRNA gene.

The glucocorticoid receptor (32,60) and the progesterone receptor (C.Spangenberg and M.Beato, unpublished) are able to bind to their MMTV target sites in the natural MMTV promoter context within positioned nucleosomes in vitro provided rotational orientation over the sites exposes the major groove of the contacted DNA on the exterior of the particle. A similar dependence on rotational phasing has been detected for binding of the glucocorticoid receptor to HREs within an artificial sequence context on positioned nucleosomes (74). The affinity of the receptors for nucleosomally organized HREs is only 5-fold lower than for the same sequences in free DNA (75). The ability of steroid hormone receptors to bind nucleosomally organized HREs probably resides in the distribution of their DNA contacts. The results of DNA binding experiments (76–78; Eisfeld et al., this issue) and structural analysis of DNA co-crystals (79) show that the two monomers of the receptor homodimer contact the same face of the DNA double helix, through a narrow sector of the major groove (see Fig. 7 of Eisfeld et al., this issue). Therefore, the receptors could bind an exposed site even when the other side of the double helix is in contact with the core histones.

In addition to MMTV, many other steroid-inducible genes exhibit remodelling of chromatin over regulatory regions in response to hormone induction. The thyroid hormone receptor TRβ, as a heterodimer with RXR, can bind to a responsive element at position +264 on the promoter of the TRβ gene in chromatin in vivo as well as in vitro on rotationally positioned nucleosomes (63). This suggests that other members of the nuclear receptor family share the ability to recognize their target sites in chromatin.

The situation is very different with NF1. Like other high affinity DNA binding proteins, NF1 contacts the DNA double helix at many sites and almost completely embraces the circumference of the helix (80; Eisfeld et al., this issue). This type of DNA recognition appears to be incompatible with binding to nucleosomally organized sites, since a fraction of the contacted DNA sites will be inaccessible due to interactions with the core histones. This prediction has been experimentally confirmed. Both in the natural context in the MMTV promoter (32) as well as in artificially positioned nucleosomes (81), NF1 cannot bind its target sequences as long as they are included within the confines of a normal nucleosome. However, when the MMTV sequences are assembled on a tetramer of histones H3 and H4 (C.Spangenberg, K.Eisfeld, K.Luger, T.J.Richmonds, M.Truss and M.Beato, manuscript submitted) or when the NF1 binding site is moved to the linker DNA (Eisfeld et al., this issue), NF1 binding is observed. When the NF1 site is placed in the linker DNA it is accessible to NF1 even in nucleosomes with bound histone H1 (Eisfeld et al., this issue). If histone H1 could bind to the exterior of the nucleosome in a symmetrical manner or to the proximal linker DNA, it would be expected to interfere with NF1 binding to a nucleosome containing a 30 bp insertion. Thus, our finding supports an asymmetrical model for histone H1 binding to the core particle in which the linker histone is located at the distal end of the nucleosome inside the DNA superhelical turns (2).

Very likely a similar explanation applies to the lack of binding of Amt1 to its target site in regular nucleosomes, as this protein also embraces the circumference of the DNA double helix (34). Only when the nucleosome is disrupted by an adjacent stretch of 16 A residues can Amt1 enter its site efficiently (35). The structure of this specialized nucleosome remains to be established.

We suspect that many other high affinity and ubiquitous transcription factors which contact DNA at many sites and encircle the circumference of the double helix share the inability to interact with their target sites within regular nucleosomes and may depend on prior remodelling of the chromatin structure for DNA binding. Remodelling could involve a transient modification of the nucleosomal structure, for instance by acetylation (for a review see 5) or dissociation of some core histones (82), or a more stable disruption leading to the formation of ‘nucleosome-free’ regions which could be epigenetically transmitted during cell division.

We conclude that the nature of the interaction of a given factor with the DNA double helix determines whether it will be able to bind to its cognate site within the context of regularly organized nucleosomes or whether it will require either altered nucleosomes or ‘nucleosome-free’ DNA regions.

ACKNOWLEDGEMENTS

We thank Keith Koch and Dennis Thiele (Michigan) for providing the coordinates of the Amt1 DNA contacts, Jörg Klug for carefully reading the manuscript and members of the Beato group for helpful discussions. The experimental work was supported by grants from the Deutsche Forschungsgemeinschaft, the European Union and the Fonds der Chemischen Industrie.

REFERENCES
