ABSTRACT

Although the overall structure of the fifth histone (linker histone, H1) is understood, its location on the nucleosome is only partially defined. Whilst it is clear that H1 helps condense the chromatin fibre, precisely how this is achieved remains to be determined. H1 is not a general gene repressor in that although it must be displaced from transcription start sites for activity to occur, there is only partial loss along the body of genes. How the deposition and removal of H1 occurs is in particular need of further study. Linker histones are highly abundant nuclear proteins about which we know too little.

The functional role of the fifth histone (H1, for short in what follows) has proven the most difficult to understand. While the structure of the octameric 147 bp core particle [1] rapidly led to an enhanced appreciation of core histone function, in particular regarding the roles of their different structural domains, no equivalent structure has been obtained for an H1-containing nucleosome. The main reason is that reconstitution of the very basic H1 molecule with an extended core particle of high net negative charge is subject to artifactual non-specific interactions and, furthermore, there is no obvious functional assay for correct binding. Additionally, chromatin-bound H1 is known to rapidly exchange [2,3] (in vivo, at least) so the conditions for generating a stable in vitro H1-containing chromatosome are not yet defined. Even more uncertain are methods for generating H1-containing higher order chromatin structures, such as the so-called 30 nm supercoil (30 nm fibre). So the primary issues in need of resolution are to understand precisely how H1 is bound to the nucleosome and how this helps the formation and stabilisation of the 30 nm fibre. Only then will we be able to fully comprehend how chromatin structures are controlled to facilitate the various transactions to which the DNA is subject.

Structure of linker histones. It was early appreciated that H1 plays a role in condensing chromatin and studies of its free solution structure gave hints as to how
it achieves this. Limited proteolysis showed that most H1 species consist of 3 domains [4]. The central globular domain (GH1, ~80 residues, d~3 nm) is the only folded element in the ~200 amino acid protein in free solution. It adopts the winged helix fold [5]. The short N-terminal domain (NTD, ~35 residues) is very basic but only in its second half: in contrast, the N-terminal part is somewhat apolar, even acidic. The NTD is disordered in free solution and does not play an obvious role in chromatin condensation [6] but is subject to post-translational modifications (PTMs), so may play a regulatory/signalling role in H1 function. Linker histones specific for condensed chromatin typically have a shorter NTD but in oocyte specific H1s it is often longer. The C-terminal domain (CTD) is also disordered in free solution but is much longer: ~100 residues in the canonical mammalian species but >200 residues in some oocyte-specific linker histones. Since the CTD is very lysine rich it is assumed to play the main role in condensing the nucleosomal fibre, a function modulated by phosphorylation at multiple sites, i.e. the addition of negative charges [6-9]. An important but difficult question to which a definitive answer is needed is the extent to which the NTD and CTD adopt defined secondary structures when bound into the chromatin fibre [10,11].

**H1s and their variants** The canonical mammalian replication-dependent (RD) H1 subtype genes (H1.1 to H1.5 and the testis-specific H1t in humans) are encoded within multiple histone clusters together with those of the core histones, synthesised only during S-phase [12]. So despite the substantial structural and functional distinctions between the core and linker histones, the H1s are an intrinsic component of the chromatin of higher eukaryotes – quite different from the situation in yeast (*Saccharomyces cerevisiae*) where there are only two copies of each of the core histone genes [13,14] and the presumed linker histones look very different from mammalian H1s.

As with the core histones H2A and H3, in mammalian cells there are also a number of replacement (variant) H1s, synthesised throughout the cell cycle from genes that sometimes contain introns and generate polyA+ mRNAs. Variant H1s play
very specific roles: for example, H5 from the nucleated erythrocytes of birds, and its mammalian homologue H1.0, are both a feature of highly compacted chromatin, probably playing a condensing role by including multiple arginines (largely absent from the canonical mammalian H1 subtypes). Oocyte-specific variants, for example xenopus B4 and mammalian H1oo, have CTDs with a substantial admixture of acidic residues (E and D) – amino acids totally absent from the CTDs of canonical mammalian H1s.

**H1 PTMs** Extensive phosphorylation of serine and threonine in the NTD and especially the CTD of canonical H1s was documented early and linked to chromatin condensation at mitosis [15]. ADP-ribosylation of the NTD was also early recognised [16] and might be thought of a component of the histone code but H1s were initially not thought to be acetylated or methylated at lysines. Mass spectrometry has shown not to be the case [17]: for example, K26 of human H1.4 in the NTD can be methylated and the adjacent residue S27 phosphorylated. The same H1 subtype can be acetylated at K34, a residue on the border of the NTD and globular domains. Other modifications have been noted within GH1, notable citrullination of R54 in ESCs [18], a conserved residue within DNA binding Site II (see below).

**Histone H1 in the 30 nm fibre.** The first and much vexed question was the location of GH1 on the nucleosome. On the assumption that each nucleosome carries a single H1 molecule [19] and the observation that GH1 alone is able to protect an extra 10 bp at each end of the core particle, giving the 167 bp chromatosome, it was proposed that GH1 binds symmetrically to the DNA on the dyad axis, making contact with both exiting duplexes and the central gyre: a 3-contact model [4]. Although a symmetrical location was supported by DNasel footprinting experiments on native chromatin [20], it provoked the question as to what determines the orientation of the linker histone on the nucleosome, so the model was widely challenged both experimentally and by molecular dynamics calculations of preferred orientations [21-23]. General agreement emerged that GH1/5 carries just two DNA binding sites [24,25]: Site I on
Helix 3 and Site II on the Loop between Helix 1 and Helix 2. This led to proposed interactions being with the central DNA gyre and just one of the exiting duplexes, i.e. a slightly off-axis binding site – as found in a DNA-GH5 crosslinking experiment [26]. Such locations did not however offer any explanation for symmetrical MNase protection of 2 x 10 bp beyond the core particle.

To extend knowledge of how H1 (in particular GH1/5) is bound to the nucleosome, i.e. define where it sits in the compacted state, neutron scattering contrast measurements were made of the 30 nm fibre containing native and deuterated H1 [27]. Plots of the square of the radius of gyration against the inverse contrast made it clear that the increased contrast in the deuterated sample must come from internally located H1. Whatever the precise model of H1 binding, this indicated that the DNA exit points of the nucleosomes must be on the inside of the fibre. Although this result did not lead to a description of nucleosome connectivity within the fibre (continuous solenoid vis-à-vis cross-linker models), it excluded all models with outward facing nucleosomes.

Appreciation of a possible third DNA binding site on GH1/5 [28,29,22,30] led to continuing interest in a symmetrical location and this was supported by hydroxyl radical footprinting studies (giving higher resolution than DNasel [20]) that showed a clear shadow from GH1 protection right on the dyad [31]. As the latter studies were conducted on reconstituted short oligo-nucleosomes, it was also possible to investigate the linker DNA. Strikingly, this showed protection by H1 of ~50 bp (170 Å) of linker with an alternating period of 10 bp, implying extended shielding from one side – presumed to result from the CTD as deletion of the 35-residue NTD did not affect the protection. The success of these footprinting experiments [31] may well have been due to the use of a chaperone (NAP1) for loading the H1s onto the chromatin template, an approach full of perspective.
Further evidence for the location of GH1 came from a cryo-EM study [32], at ~11 Å resolution, of crosslinked 12-mer arrays of H1.4–containing nucleosomes, which appear as three tetranucleosomal units stacked to form a zig-zag 2-start helix, each unit having fully extended DNA linkers. Fitting the crystal structure of the 167 bp repeat tetranucleosome (that lacks H1) [33] to the averaged cryo-images allowed an assessment of the location of H1. Well-defined density was observed between the exiting DNA duplexes: fitting this to the crystal structure of free GH5 [5] led to a 3-contact model for the globular domain in which Site III, (comprising the two β–strands of Loop 3), is in contact with the central DNA gyre of the nucleosome and Sites I and Site II make contact with the two exiting DNA duplexes. This location put GH1 somewhat asymmetrically placed, the Site III contact being about one quarter of a turn (~8 Å) from the dyad axis.

Even more recently, however, a crystal structure of a reconstituted GH5-containing 167 bp chromatosome at 3.5 Å resolution was published [34] showing the H5 globular domain located essentially symmetrically on the dyad axis, with 3 contact regions: Loop 3 (i.e. Site III, e.g. S90 and V87) is in contact with the central DNA gyre (major groove); Loop 1 (e.g. R42) is in contact with the minor groove of one exiting DNA duplex and Helix 3 (e.g. Q67) is edge on into the minor groove of the other exiting duplex, in addition to several K/R-phosphate contacts. This is essentially the symmetrical model originally proposed in our 1980 article [4] but this does not imply that all H1 globular domains are necessarily placed so symmetrically. The authors of the GH5 chromatosome structure themselves previously used several other physical techniques, i.e. not crystallography, to study the nucleosomal binding of the globular domain of Drosophila H1 [35] and concluded that binding is asymmetric: contact is made with the central DNA gyre and with one exiting duplex. This conclusion was reinforced in their latest publication [34] in which they compared the sedimentation of a 12-nucleosome array carrying GH5 or Drosophila GH1: the GH5 array was seen to be substantially more compact, implying that the binding mode of linker histones is variable and can influence the higher order folding of chromatin. This is unquestionably an attractive hypothesis: the existence of
multiple subtypes and variants of the linker histones are clearly related to disparate functions that must surely imply variations of structure in the resulting chromatin.

A primary aim of structural studies of H1 in the context of the fibre has always been to understand how it compacts the nucleosomal beads-on-a-string structure. Since H1-containing nucleosomal arrays adopt a rigid zig-zag structure that compacts to form the 30 nm fibre [36-39], the first assumption was that GH1 directs the exit angles of the DNA from the nucleosome and the basic CTD covers the linker DNA, the resulting charge reduction permitting folding and compaction. Such a general principle of H1 involvement in folding could hold for all models of nucleosome connectivity within the fibre. However, such a mechanism suggests that without H1 the 30 nm fibre might not form at all – but this is not the case: the presence of divalent cations is sufficient to compact an H1-depleted regular nucleosomal array [40]. A possible explanation for this is seen in AFM observations that the DNA duplexes exiting from nucleosomes lacking H1 cross at right angles, provided 4-10 mM Mg++ is present [41]. Thus Mg++ could help order the DNA architecture within the fibre but how such divalent cations might be sufficient to counteract repulsions between linker DNAs on the inside of the fibre remains an open question.

The precise geometry of the nucleosomes and the trajectory of the linker DNA in the folded 30 nm chromatin fibre has been the subject of much study and dispute. The 167 bp repeat tetranucleosome, lacking H1, [33], as well as the recent cryo-EM structure [32] showed a stacked arrangement with fully extended DNA linkers. This has lent support to models in which the linker DNA between adjacent nucleosomes criss-crosses the fibre and several arrangements are physically possible [42]. *In vitro* studies of longer fibres typically use arrays of a core octamer-positioning sequence, such as the Widom 601 sequence [43] – which allows for variations in the spacing, i.e. DNA linker length – reconstituting with core histones and H1. The resulting ‘fibre’ has been observed by both negative staining and cryo-EM and a helix of nucleosomes proposed, the dimensions of which depend on the NRL used in the constructs [44-
The EM images imply that the structure is not particularly rigid and thus not perfectly regular, even over short distances; additionally, the compaction of the nucleosomal fibre obscures the path taken by the centrally located linker DNA. Nevertheless, the Rhodes laboratory propose that the inter-digitated helix that best fits their data is a 1-start solenoid with continuously wound linker DNA [45]. The recent cryo-EM study [32] presents a clear cross-linker structure, moreover not just for a 167 bp NRL but also for the more typical 187 bp.

The conflict between bent linker and straight, cross-linker, models is however by no means fully resolved, not least because most observations to date have been made on chromatin fragments on surfaces or in the solid state. High resolution biochemical experiments in solution would be helpful to define the path of the inter-nucleosomal linker DNA, i.e. define the internal architecture of the fibre. It is worth noting here that as well as on EM grids there is good solution evidence for in vitro 30 nm fibres [48], to complement observations of the fibre in the highly condensed chromatin of sperm and nucleated erythrocytes [47]. However, no dominant scattering peaks corresponding to this distance were obtained from active somatic cells, either in interphase or even at mitosis [49-51].

**Rapid dynamics of H1** FRAP/FLIP experiments have shown that a large proportion of the H1 in mammalian nuclei exchanges rapidly between sites on euchromatin with an average residence time of \(~4\) min, a process accelerated after treatment of cells with TSA to boost histone acetylation levels [2,3]. This in in sharp contrast to the core histones for which essentially no exchange is observed over the same time period. The residues on the surface of GH1 responsible for H1.0 binding to chromatin have been mapped *in vivo* by FRAP, measuring the exchange rates of a large set of GH1.0 point mutants [23]. Two clusters of basic residues were highlighted for which mutation to alanine led to much increased dissociation rates, indicating that these represent nucleosomal binding sites – in agreement with the two binding sites previously defined, Site I and Site II. This led the authors to favour an off-axis model in which GH1.0 is in contact with the central DNA gyre and just one exiting duplex.
Additionally, mutations of A and S to D in the β-loop wing element of the fold also led to much increased dissociation rates [23], perhaps also indicative of a third binding site, that would be consistent with symmetric on-axis 3-site binding models. The importance of Site II was seen from a R42A mutation that led to a 3-fold increase in the dissociation of GH1.0 [23] and this was strikingly confirmed by the recent demonstration that this residue is subject to citrullination by PADI-4 (conversion of arginine to citrulline) in pluripotent mouse ESCs, leading to displacement of the H1 from chromatin [18]. This modification must be a factor contributing to the rapid mobility of H1s and the consequent ‘open chromatin’ characteristic of mESCs [52].

Such dynamic binding of H1 [2,3,23] seems inconsistent with the ubiquitous presence in cells of a tightly wrapped 30 nm fibre having internal H1, so the short residence times observed can be regarded as indirect evidence that the 30 nm fibre is itself dynamic in nature, at least in vivo, which might also explain the failure to detect it in the nuclei of most cell types [49-51].

**The functionality of H1 in chromatin.** Although the stoichiometry of H1 in vivo is about one linker histone molecule per nucleosome [19], it is variable and appears related to the nucleosome repeat length (NRL). For example, neuronal chromatin has a NRL of 162 bp and 0.45 molecules of H1 per nucleosome, whereas in glial cells the NRL is 201 bp and the stoichiometry is 1.04 [53] and in natural retinal maturation the NRL gradually increases from 190 to 206 bp as the amount of H1 increases by 23% [54]. Knock-outs of individual H1 subtypes conducted in mouse ES cells led finally to only ~50% of wt levels of H1 and it was observed that the NRL fell in proportion to the amount of H1 lost [55,56]. In fact, a plot of all available in vivo data showed a good linear relationship from yeast (NRL=165 bp, no conventional H1) to chicken erythrocyte nuclei (NRL=211 bp, 1.3 H1s/nucleosome) [57].

The role of the H1 molecule in this relationship seems to be the maintenance of electrostatic charge homeostasis, outcome being that its binding keeps adjacent nucleosomes apart and achieves their regular spacing. In the knock-out experiments
the reduced amounts of the prime agent of chromosome condensation that led to artificially lowered NRL values might be expected to be accompanied by substantial increases in transcription – but that was not found to be the case [56]. This could be the consequence of H1 depletion being concentrated in inactive regions of the genome but, strikingly, in ESCs (having a quite low average NRL of 185 bp and only about 0.5 molecules of H1 per nucleosome) there was no clear distinction between observed NRLs for active genes and inactive heterochromatin [56]. There is therefore no general link between the capacity for transcription and the amount of H1 present.

There is however a link between local levels of H1 and actual/potential transcriptional activity [58]. Experiments to map the distribution of linker histones across the genome show reductions in the concentration of H1 at DNaseI hypersensitive sites (DHS), e.g. at gene promoters and enhancers. Clear-cut results were obtained in Drosophila and human cells using the DAM-methylase mapping protocol. Although not yielding particularly high resolution, very clear ‘dips’ of low H1 occupancy were noted at the TSS of active genes and at multiple cis-regulatory sites [59, 60]. Such reductions are not simply a consequence of nucleosome depletion, as no concomitant loss of core histone was seen. Similar ‘dips’ were noted when antibodies against H1 subtypes and variants were used with several chicken cell types [61]. Such observations indicate that H1 needs to be displaced from cis-regulatory elements to facilitate their function. Along the body of active genes, however, substantial amounts of H1 remain, though at depleted levels [62].

H1 and chromatin remodelling

The important question is then: how is access gained to H1-containing compacted (‘closed’) chromatin in order to initiate transcription? The answer lies in a subset of transcription factors – pioneer transcription factors (PTFs) – that can access their target sites on wrapped, nucleosomal DNA in compacted chromatin, rather than requiring exposed targets on inter-nucleosomal DNA, i.e. the binding of PTFs is not inhibited by the presence of H1. Striking early examples of PTFs were the factors
HNF-3 (FoxA) and GATA-4 that were shown to target albumin gene enhancer elements in reconstituted H1-compacted fibres, leading to opening of the chromatin in the absence of remodelers [63]. An opening mechanism was proposed in which the C-terminal domain of HNF-3 binds to core histone tails, preventing them from stabilising the chromatin higher order structure. Hormone receptors were early recognised as binding to their targets on nucleosomal DNA [64] and the PTF list is now much extended, in particular to include the pluripotency-inducing factors Oct4, Sox2 and Klf4 [65]. These represent a broad range of DBD folds, so there is no universal mechanism by which PTFs open up H1-compacted chromatin but it is clear that binding of PTFs is followed by recruitment of acetyltransferases (HATs), remodeling complexes and other key factors that open up and re-shuffle the chromatin, leading to local loss of H1 [66].

Some chromatin remodelling steps have been individually investigated: for example, it has been shown that H1-containing nucleosomes and more importantly, H1-containing nucleosomal arrays can be remodelled by ACF (but not by CHD1) [67]. The same is true for the SWI/SNF complex but only if the H1 is phosphorylated [68]. The ability of ACF to remodel H1-containing chromatin elements certainly depends on the type of linker histone: the xenopus oocyte-specific B4 histone fails to inhibit accessibility of the linker DNA and thus permits remodelling. This is in contrast to the somatic histone H1A which is able to block remodelling [69] – at least for a di-nucleosomal substrate – a difference that could be a direct consequence of their very different CTDs.

Displacement of H1 can be regarded as a prerequisite for transcription [63,66] and reciprocal binding of PARP1 following loss of H1 at promoters has been described [70], as has the recruitment of the variant histone H2A.Z [71-73] following binding of a PTF. An active promoter is therefore envisaged as lacking H1, with several other factors having gained access.

A more recent in vitro-assembled H1-containing transcriptional system [74] involved chromatin assembled on a 5 kb plasmid containing 540 bp from the rat PEPCK promoter, which can be induced by addition of retinoic acid (RA) in
consequence of incorporating two RAREs. Reconstitution used unacetylated core histones and H1 in the presence of chaperones and the state of compaction of the template was monitored by EM, showing condensed fibres of 26-28 nm diameter. In this state the template was refractory to transcription, however following treatment with the PTF RAR/RXR, addition of the remodeler SWI/SNF and the HAT p300 was then sufficient to displace nucleosomes, release the H1 and permit binding of NF1 – with subsequent transcription. The sequence of these in vitro activating events followed the main features of the regulated expression of the PEPCK gene in vivo.

**Multiple subtypes of H1** Whilst the variant linker histones such as H5, H1.0 and B4 have been the subject of multiple studies, the function of the replication-dependent subtypes that make up the bulk of the H1 presents a more difficult problem, not least because their primary structures vary little [75]. Despite the fact that linker histone subtypes appear to act as redundant proteins in knock-out mice [55], suggesting no functional distinction between such subtypes, different tissues are nevertheless characterized by the action of specific linker histone subtypes and changes in subtype expression are observed during development [76-79].

One approach to this question has been the use of subtype specific antibodies in ChIP studies - with rather mixed results. In human cells H1.5 (not one of the most abundant RD subtypes) binds in blocks of enrichment to certain genic and intergenic regions, leading to the binding of Sirt1 and the dimethylation of H3K9, suppressing their activity [77]. Using knock-in FLAG-tagged H1 species in ChIPs with mouse ESC, blocks of H1c and H1d enrichment (abundant RD subtypes) were seen at satellite DNA sequences and their depletion noted at active promoters [80].

Some functional distinctions between the RD subtypes have been documented in gene expression studies [81] that frequently show recruitment of specific H1 subtypes to genes subject to repression in consequence of the interaction of the subtype with an established transcription factor (TF). In one case a pioneer transcription factor, FoxP3, interacts with human H1.5 to repress IL-2 expression in
T-cells [82]. Other examples are mouse H1b interacting with Msx1 to negatively regulate MyoD expression in myoblasts [83] and human H1.2 interacting with p53 to direct H1.5 to genes targeted by p53 in transfected HeLa cells [84].

As functional distinctions between different H1 subtypes have been clearly made, the challenge now is to refine these and determine the molecular mechanisms whereby these distinctions are brought about. One possibility is that variations in the primary structure of the common RD subtypes give rise to differences in their epigenetic modifications. A striking example of this is the methylation of K26 in human H1.4 that leads to recruitment of the repressor HP1, provided the adjacent S27 is not phosphorylated. The K26/S27 pair (representing a phospho-switch) is unique to the H1.4 subtype in this region of the NTD [85]. The sequence variability and frequent modification of this domain make it the primary candidate for the regulatory element of the linker histones.

REFERENCES

10) H. Fang, D.J. Clark, J.J. Hayes, DNA and nucleosomes direct distinct folding of a linker histone H1


26) Y.-B. Zhou, S.E. Gerchman, V. Ramakrishnan, A. Travers, S. Muyldermans, Position and orientation


41) I. Nazarov, Personal Communication.


43) P.T. Lowary, J. Widom, New DNA sequence rules for high affinity binding to histone octamer and


59) U. Braunschweig, G.J. Hogan, L. Pagie, B. van Steensel, Histone H1 binding is inhibited by histone


