



**HAL**  
open science

## From crystal and NMR structures, footprints and cryo-electron-micrographs to large and soft structures: nanoscale modeling of the nucleosomal stem.

Sam Meyer, Nils B Becker, Sajad Hussain Syed, Damien Goutte-Gattat, Manu Shubhdarshan Shukla, Jeffrey J Hayes, Dimitar Angelov, Jan Bednar, Stefan Dimitrov, Ralf Everaers

### ► To cite this version:

Sam Meyer, Nils B Becker, Sajad Hussain Syed, Damien Goutte-Gattat, Manu Shubhdarshan Shukla, et al.. From crystal and NMR structures, footprints and cryo-electron-micrographs to large and soft structures: nanoscale modeling of the nucleosomal stem.. *Nucleic Acids Research*, 2011, 39 (21), pp.9139-54. 10.1093/nar/gkr573 . hal-00726331

**HAL Id: hal-00726331**

**<https://hal.science/hal-00726331>**

Submitted on 29 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# From crystal and NMR structures, footprints and cryo-electron-micrographs to large and soft structures: nanoscale modeling of the nucleosomal stem

Sam Meyer<sup>1,\*</sup>, Nils B. Becker<sup>1,2</sup>, Sajad Hussain Syed<sup>3,4</sup>, Damien Goutte-Gattat<sup>3</sup>, Manu Shubhdarshan Shukla<sup>3,4</sup>, Jeffrey J. Hayes<sup>5</sup>, Dimitar Angelov<sup>4</sup>, Jan Bednar<sup>6,7,8</sup>, Stefan Dimitrov<sup>3</sup> and Ralf Everaers<sup>1</sup>

<sup>1</sup>Laboratoire de Physique and Centre Blaise Pascal, École Normale Supérieure de Lyon, Université de Lyon, CNRS UMR 5672, 46 allée d'Italie 69364 Lyon cedex 07, France, <sup>2</sup>FOM Institute AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands, <sup>3</sup>INSERM; Université Joseph Fourier - Grenoble 1, Institut Albert Bonniot, U823, France, <sup>4</sup>Université de Lyon, Laboratoire de Biologie Moléculaire de la Cellule, CNRS-UMR 5239/INRA 1237/IFR128 Biosciences, France, <sup>5</sup>Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester NY 14642, USA, <sup>6</sup>Univ. Grenoble 1/CNRS, LIPhy UMR 5588, Grenoble, F-38041, France, <sup>7</sup>Charles University in Prague, First Faculty of Medicine, Institute of Cellular Biology Albertov 4, 128 01 Prague 2, Czech Republic and <sup>8</sup>Department of Cell Biology, Institute of Physiology, Academy of Sciences of the Czech Republic, 128 01 Prague 2, Czech Republic

Received November 30, 2010; Revised June 17, 2011; Accepted June 25, 2011

## ABSTRACT

The interaction of histone H1 with linker DNA results in the formation of the nucleosomal stem structure, with considerable influence on chromatin organization. In a recent paper [Syed, S.H., Goutte-Gattat, D., Becker, N., Meyer, S., Shukla, M.S., Hayes, J.J., Everaers, R., Angelov, D., Bednar, J. and Dimitrov, S. (2010) Single-base resolution mapping of H1-nucleosome interactions and 3D organization of the nucleosome. *Proc. Natl Acad. Sci. USA*, 107, 9620–9625], we published results of biochemical footprinting and cryo-electron-micrographs of reconstituted mono-, di- and tri-nucleosomes, for H1 variants with different lengths of the cationic C-terminus. Here, we present a detailed account of the analysis of the experimental data and we include thermal fluctuations into our nano-scale model of the stem structure. By combining (i) crystal and NMR structures of the nucleosome core particle and H1, (ii) the known nano-scale structure and elasticity of DNA, (iii) footprinting information on the location of protected sites on the DNA backbone and (iv) cryo-electron micrographs of

reconstituted tri-nucleosomes, we arrive at a description of a polymorphic, hierarchically organized stem with a typical length of  $20 \pm 2$  base pairs. A comparison to linker conformations inferred for poly-601 fibers with different linker lengths suggests, that intra-stem interactions stabilize and facilitate the formation of dense chromatin fibers.

## INTRODUCTION

Fifty years after the discovery of the molecular structure of the DNA double helix, it is still unclear how DNA is arranged on the mesoscale between 10 and several hundred nanometers where essential processes such as replication, transcription and repair occur (1). In eukaryotic cells DNA is coated with at least an equal mass of proteins, forming a complex called chromatin (2). Its fundamental repeating unit, the nucleosome (3), is made of  $\sim 200 \pm 40$  base pairs (bp) of DNA and an octamer of core histones containing two copies of each H2A, H2B, H3 and H4 (4).

The structure of the relatively rigid nucleosome core particle (NCP) has been resolved by crystallography at almost atomic resolution (5,6). It comprises 145–147 bp of DNA, wrapped in  $\sim 1.65$  left-handed superhelical

\*To whom correspondence should be addressed. Tel: +33 4 72 72 86 34; Fax: +33 4 72 72 86 36; Email: sam.meyer@ens-lyon.fr

turns around the core histones, with 14 well-defined contact points (7,8).

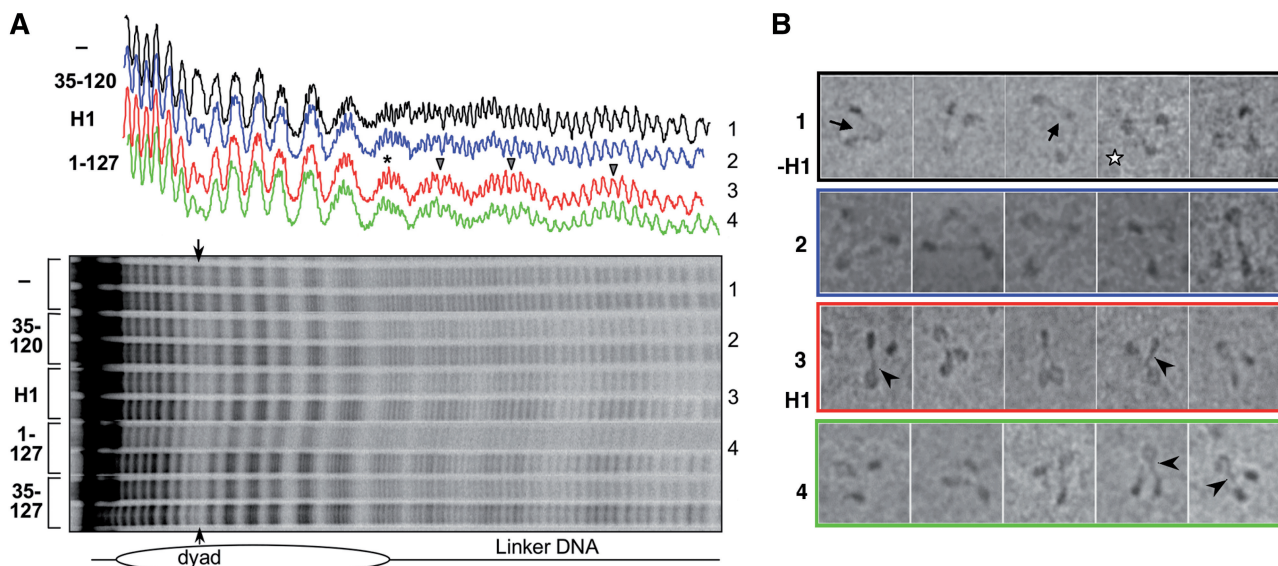
Comparably detailed structural information does not exist for the *soft* parts of the nucleosome: thermal fluctuations displace their elements on lengthscales larger than atomic distances, preventing the use of standard high-resolution techniques. Filling this gap remains an important challenge, since the flexible elements are essential in the formation and the maintenance of chromatin fibers and mitotic chromosomes (9–11): the linker DNA between NCPs, comprising 10–90 bp of DNA, via their mechanical properties (12–15); the flexible NH<sub>2</sub>-tails of the core histones, introducing subtle interactions between NCPs (16–22); and the linker histone (H1/H5) (23–26). Here we focus on the linker histone, which plays a key role in the organization of linker DNA in chromatin fibers. It consists of a central globular domain (gH1) (27), flanked by basic amino- and C-terminal tails, and it binds DNA close to the entry of the NCP, resulting in a characteristic ‘stem’ structure (28).

In a recent study (23), we have mapped the histone H1–DNA interactions within the nucleosome at a single base resolution. The physiologically relevant linker histone chaperone (NAP-1) was used to reconstitute histone H1 and truncated mutants on precisely positioned nucleosomal templates containing the 601 sequence (29). The resulting complexes were validated and investigated by a combination of cryo-electron-microscopy (CEM) and  $\bullet$ OH footprinting techniques. The footprints (Figure 1A) showed that binding of the globular domain protects the first 10 bp of the linkers as well as the DNA at the NCP dyad against  $\bullet$ OH-induced cleavage. Binding either full-length H1 or the 1-127 COOH terminus truncation mutant causes the appearance of the characteristic stem

structure in CEM images of tri-nucleosomes and an additional 10-bp repeat in the  $\bullet$ OH cleavage pattern in the stem region of the linker DNA. The raw experimental data clearly identify sections of the nucleosomal DNA affected by the stem formation. As in the case of scattering or NMR experiments, further interpretation of the biochemical data requires the use of macromolecular models.

In the present article, we develop a three-dimensional, dynamical coarse-grain model of the nucleosomal stem formed by the histone H1/H5 and the in- and outgoing linker DNA chains. The model integrates our CEM and footprinting results (23) together with crystal and NMR structures for the NCP (7) and the linker histone (27) and knowledge on the (sequence-dependent) B-DNA structure and elasticity (30–34), and can be related to experiments on model chromatin fibers (35).

The article is organized as follows: in the ‘Materials and Methods’ section and the ‘Results’ section we cover (i) the numerical analysis of the footprinting gels, (ii) the test of previously proposed atomistic models for the placement of the globular gH1 domain (24–26), (iii) the determination of the most likely coarse-grain conformation of the H1-bound linker DNA stem in the state of maximal protection, and (iv) the comparison of a thermal ensemble of fluctuating stem structures to footprinting and CEM data. In the ‘Discussion’ section, we present the emerging picture of the nucleosomal stem as a dynamic, polymorphic, hierarchically organized structure, composed of a ‘root’ where gH1 binds to the first  $\sim$ 10 bp of the DNA linkers, a ‘trunk’ formed by the association of the subsequent  $10 \pm 2$  bp with the cationic C-terminus of H1, and a flexible ‘crown’ or outer stem where the branching linkers exhibit substantial fluctuations, while preserving well-defined preferential contacts. In particular, we show



**Figure 1.** Illustration of the available data from (23). (A)  $\bullet$ OH-footprinting gels of mononucleosomes in the linker region, and corresponding intensity profiles: (1) without H1, (2) truncated mutant 35-120 of H1 (gH1), (3) full H1, (4) truncated mutant 1-127. The dyad region is protected by all H1 mutants, as well as the first 10 bps of the linker. Full H1 and mutant 1-127 exhibit further periodic protections on the linker. (B) CEMs of trinucleosomes: (1) without H1, (2) gH1, (3) H1, (4) 1-127 mutant of H1. Arrowheads indicate visible stems, the star indicates a shape incompatible with the presence of a stem.

that intra-stem interactions stabilize linker conformations closely related to those inferred from experiments on reconstituted poly-601 fibers (35). In the final section, we briefly conclude.

## MATERIALS AND METHODS

Experimental methods are described in (23). In the following, we present a detailed report of:

- the numerical analysis of the footprinting gels, which allows us to extract the protection patterns with single base resolution;
- the atomistic and coarse-grain descriptions of DNA and histone proteins;
- the investigated atomistic models of gH1 placement;
- the estimation of ●OH footprinting patterns from atomistic and coarse-grain structures;
- the minimization of the stem elastic energy under the constraint of the experimentally observed protection patterns and
- the construction of the ensembles of fluctuating nucleosomes.

Except for the last point, these methods (and some of the corresponding results) were already briefly presented in (23).

### Footprint analysis

The raw intensity signal (Figure 1A) shows four main features:

- The smallest oscillations are single-nucleotide bands. They can be separated reliably only in a region with sufficient contrast.
- Oscillations with a period of around 10 bands reflect protection from ●OH attack.
- Trends resulting from logarithmic migration in the gel.
- The trace contains a background exposure level.

These raw traces are processed as described below. The steps of the process are illustrated on Figure 2. More details are available in the Supplementary Data.

*Trace alignment.* The varying width of individual bands (7–14 pixels in this example) results from a combination of logarithmic migration and irregularities in the gel material. In a first step we determine the non-linear relation between migrated distance (in pixels) and base number.

To determine the positions in pixels  $x(n)$  of individual bands (numbered by  $n$ ), we first de-trended the intensity traces by subtracting a suitable moving average. We then iteratively maximized the correlation between this signal and a modulated cosine function  $A(x)\cos[2\pi n(x)]$ , where  $A$  is slowly varying. In each iteration, the running phase of the cosine is adjusted,  $n(x) \rightarrow n(x)+\delta(x)$ , to improve the correlation between signal and modulated cosine, in a moving window of 7 bands length. The width of the moving window allows to assign bands even in short intermediate regions without sufficient contrast, by using the fact that band widths do not change abruptly. To ensure

that no base pair has been missed, we checked by eye the maximum positions and the final base pair-pixel correspondence  $x(n)$ . The unresolved regions, where the bands could not be reliably positioned (irregular band widths), were excluded of the analysis (the resolved region of the linker is shown in the inset of Figure 2, right upper panel).

To relate ●OH protected areas to absolute sites on the nucleosome (with the dyad base pair centered at 0), rather than band numbers only, we identified absolute lengths of DNA fragments on the gels by using a combination of molecular weight markers present in the mononucleosome gels, laser UV irradiation and Fpg glycosylase treatment (36), and comparison with absolute positions determined in dinucleosome gels. The unique positioning of the 601 sequence on the nucleosome then allowed to assign DNA lengths to nucleosomal sites.

*Intensity per base-pair.* After removing a constant level of background noise, the raw intensity signal measures the amount of DNA of a given molecular weight. By integrating over the width of each band, we obtain the irradiation intensity per band as a function of band number (since we consider only regions with well-separated bands, integration instead of fitting of multiple peaks introduces negligible errors).

*Relative accessibility.* To eliminate the global trends in the trace amplitudes, we then generated a signal which represents the local accessibility of a nucleotide compared with its neighbors. In this final processing step, the intensity of each base pair is divided by the mean of the 3 maximum intensities in a sliding window. The window width was set to a value between 7 and 20 in the presented data. In effect, the ~10 bp oscillations are rescaled to values roughly between 0 and 1: see Figure 2 and the final complete signal for the mononucleosome on Figure 4A.

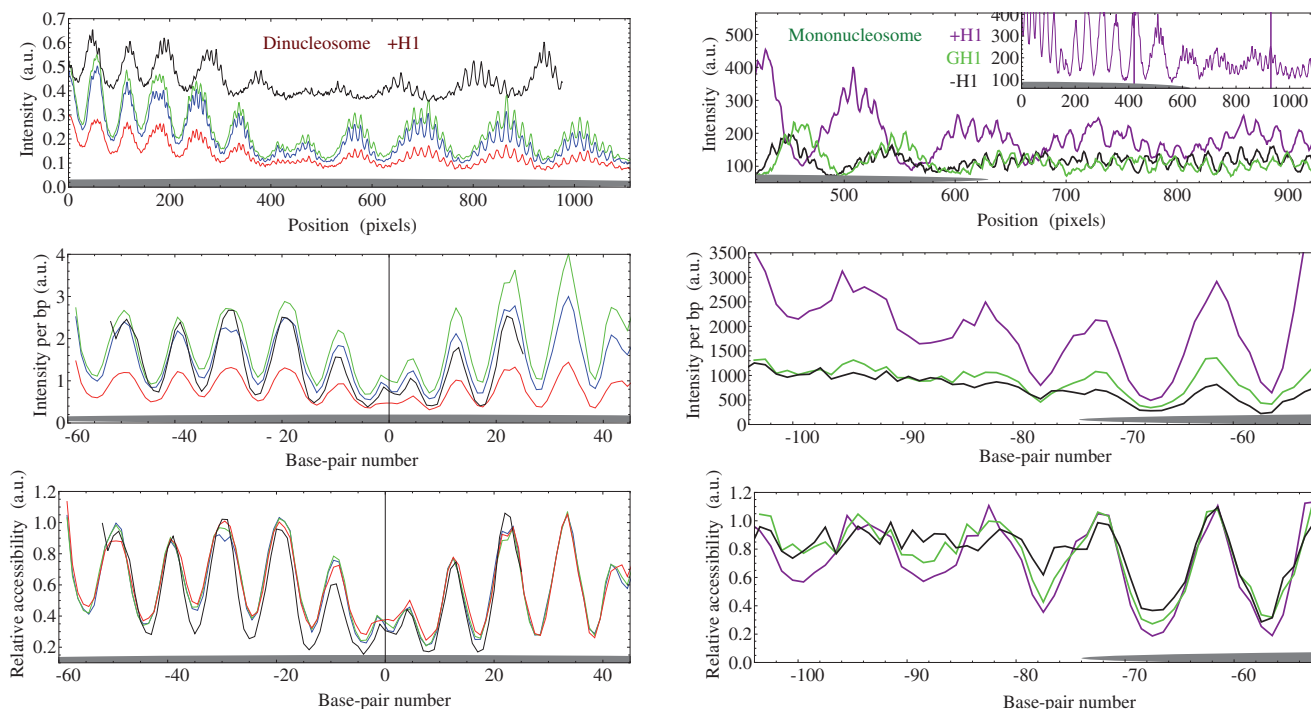
Applied to signals exhibiting no oscillations (typically, the -H1 trace on the linker), this step effectively amplifies the noise (black trace on Figure 2, compare the middle and lower right panels). We therefore exclude regions where this effect prevents a reliable interpretation of the generated signal (left end of -H1 trace on Figure 4A).

All processing steps were implemented in Mathematica Edition: Version 7.0, Wolfram Research Inc (2008).

*Exploiting the 2-fold symmetry.* The mononucleosome accessibility profiles were measured for only one of the two non-equivalent strands. However, the nucleosome structure has an approximate 2-fold symmetry axis, which allows to deduce an accessibility profile for the complementary strand, as follows: The 147-bp nucleosome core particle structure NCP147 (7) shows that the 2-fold (dyad) axis traverses the central base pair. Thus by rotating the DNA loop by a half turn around the dyad axis while keeping the histone core in place, one generates a second approximately symmetric conformation in which the two strands change roles.

We make the hypothesis that these two conformations are equally represented in our experiments and take the accessibility profile of the complementary strand equal to





**Figure 2. Post-processing of the footprinting gels.** Upper panels: available raw data for di- and mononucleosomes. Intensity variations in parts of the traces (see window in the inset, right panel) allow the resolution of bands corresponding to strands with a specific length. Outside these windows, one can only discern an oscillation with the 10 bp helical period of DNA. The key step in the quantitative analysis of the gels is the identification of individual bands, i.e. the mapping from pixels to base pairs (see details in the Supplementary Data). **Middle panels:** intensity per base pair, obtained by the integration of the raw signal over the bp width. **Bottom panels:** relative accessibilities: the intensity per base pair is rescaled by the maxima in a moving window (of 7–20 bp width). The resulting signal represents the relative accessibility of a site compared to its neighbors (in the same trace). **Left hand-side:** H1-bound dinucleosome traces in the NCP dyad region—red, green, blue—from the same gel, with different  $\bullet$ OH concentrations; black: complementary strand, from another gel. The consistency of the resulting signals (from independent traces with different relative noise) shows the robustness of the procedure (maximum difference  $\sim 0.1$ ). However, this signal does not represent the *absolute* accessibility, so that the amplitudes of different traces (or regions) cannot be directly compared (see text). **Right hand-side:** available mononucleosome traces in the linker region: -H1 (black), gH1 (green), +H1 (purple); the signals are shown in the resolved region where individual bands could be identified for the three signals (see window in the inset): only one available trace for each. For a non-oscillating signal (in particular, the left end of the -H1 black trace), the last step effectively amplifies pure noise. We therefore excluded from the subsequent analysis the data for which this effect prevents any reliable interpretation.

that of the measured strand (both read from 5' to 3'). This hypothesis is supported by the co-localization of protected sites from both strands (Figure 4B), and by the close agreement of the two dinucleosome signals obtained independently from the complementary strands (Figure 2, left hand-side).

### DNA and histone modeling

We have modeled DNA and histones at two different resolutions. (Figure 3).

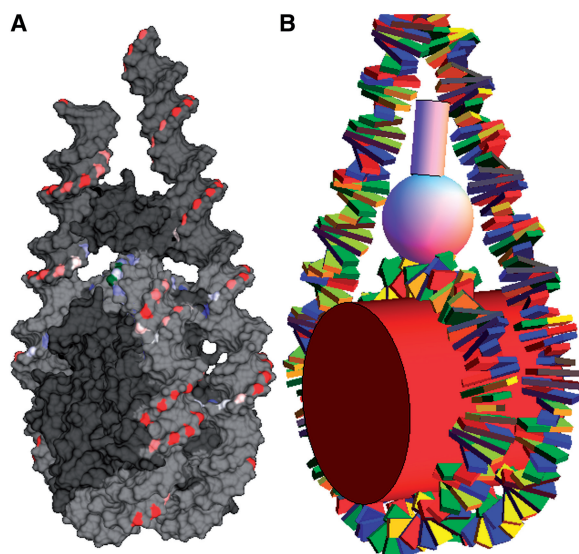
We built atomistic models where corresponding experimental information was available: on the nucleosome core and for gH1 (24–26,37). For the core DNA, we used the NCP147 nucleosome structure (7). Linker DNA was added in straight regular B-DNA conformation or in bent conformation, depending on the particular nucleosome model. These were constructed using a purpose-built Mathematica library for rigid base pair DNA manipulations (38) [with the 'MP' hybrid parameter set as described in (32,39) from (30,31)], with the 601 sequence used in the experiments, and then translated into pseudo-atomistic structures using the 3DNA program (38). We can localize the attacked sites with Ångström resolution in

the molecular models. For visualization, we used software Chimera (41).

For the modeling of the stem, no structural models are available. We used a coarse-grained model of DNA (the rigid base pair model, see references above). Histones are modeled as spheres and cylinders (see 'Accessibility profiles' section).

### Atomistic models of gH1 placement

**Three-contact model.** The three-contact structure was proposed by Fan *et al.* (24) as a result of exhaustive rigid molecular docking for given DNA linker configuration. We rebuilt their structure manually by matching the orientations of the protein alpha-helices and the protein-DNA contacts for each of the three contact sites (see Figure 5A). Deviating from the choice (24) of using a gH5 X-ray structure (27), we considered the solution NMR structure ensemble of gH1 (37) since it corresponds to our experimental system, has sufficient resolution for the present purpose, and allows to assess the structural variability of the protein. Specifically, while the protein fold is stable, the protein loop regions and lysine side chain orientations are highly variable; we chose conformer



**Figure 3.** Structural models used in the modeling of DNA and histones, and for the computation of structure-derived accessibility patterns. **(A)** Atomistic model (for the placement of gH1). NCP components from the crystal structure (7), NMR linker histone structure (37), and straight or bent linker B-DNA pseudo-atomic coordinates (40). **(B)** Coarse-grained model (for the modeling of the stem). Rigid base pair model of DNA (38), with NCP structure obtained from **(A)**. The histones are modeled as cylinders (core octamer and H1 tail) and spheres (gH1).

8 in the ensemble (PDB code 1ghc) since it accommodates the predicted contacts well (24). At the same time, its relatively extended loop conformation does not necessitate inward bending of the DNA linkers to establish three contacts, in contrast to the somewhat more compact gH5 conformation (24). The contacting residues are Lys47, Lys51 and Ser52 (site I, orange); Lys63 (site II, red); and Lys18, Arg20, Arg72 and the C-terminal Arg75 (site III, purple). Note that these residue numbers are those of chicken H1 used in NMR studies; they are offset by 22 AA from those in H5 used in the crystallography studies, and by 37 AA from the human H1 used in the experiments.

*Two-contact models.* Zhou *et al.* (25) proposed an arrangement of linker histone onto the nucleosome based on cross-linking experiments with mutated gH5. In this model, the linker histone globular domain contacts core DNA from the major groove, at around 2 bp distance from the dyad. It also contacts one of the DNA linkers. The spatial arrangement was rebuilt by deforming one of the linkers in the DNA model, and matching the location and helix orientations of the docking solution shown in (42) manually. We used the same molecular model (1ghc, conformer 8) for the H1 globular domain as for the three-contact model, whose shape gives close-fitting molecular contacts also in this arrangement, see Figure 5B. The gH1 -helices I (cyan), II (purple) and III (magenta) are colored as in (42); the C-terminal Lys75 is shown in purple, Lys63 is shown in red, contacting linker DNA. The residues Ser7,19,49 mutated in (42) are shown in orange.

Brown *et al.* (26) proposed another molecular model for linker histone placement refined by rigid docking. Here, the linker histone globular domain contacts core DNA from the major groove, at around 5 bp distance from the dyad, and one DNA linker. Note that globular domain positions in the two models A, B are on opposite sides of the dyad. Again the docking solution was reproduced manually by matching the reported helix orientations (different from model A) and contact residues; it is shown in Figure 5C. Residues contacting core DNA about 5 bp away from the dyad are Lys47, Lys51 and Ser52 are colored light green; residues contacting one DNA linker are Arg20, Arg72 and Lys75 (leftmost) are colored purple. The viewing direction is the superhelical axis.

Both two-contact models should be interpreted as showing one of two symmetric coexisting configurations, forming a contact with either of the linkers.

### Accessibility profiles

*Semi-quantitative •OH footprinting predictions from atomistic models.* The reactivity of attack sites is determined by their respective solvent accessible surface areas (43). The solvent accessible surface is computed by rolling a 1.4 Å sphere representing water, or the similar-sized •OH radical, over van der Waals spheres of the atoms in the molecular model. The variations of surface accessible areas due to DNA conformation and to contacts formed with protein side-chains have been used successfully to predict the position-dependent relative accessibilities observed in •OH-footprints (44,45).

The corresponding MSMS program (46) is implemented in the molecular visualization system Chimera (41) and can be used to compute the solvent-accessible area of each atom in a structural model (this area vanishes for interior atoms).

Lacking a thermal ensemble of structures and the resolution of single protons in our models, we somewhat simplified the procedure, considering solvent accessible surface areas of C5' atoms directly, and using (44,45) 'unified van der Waals radii' (47) to account implicitly for the hydrogens. To mimic the smoothing effect of thermal fluctuations, we increased the probe radius to 3 Å and calculated a moving average over the resulting trace with a 3 bp window. Finally, the predicted accessibility patterns for the two strands were averaged to account for the strand-exchange symmetry observed in experimental footprints.

*Semi-quantitative •OH footprinting predictions from coarse-grain models.* Protection patterns for rigid base pair models can be determined by reinserting atomic details using the 3DNA program (40) and then following the procedure described in the preceding section. However, this approach is too time-consuming for the analysis of large ensembles of structures and we have therefore developed a procedure to estimate protection patterns on the coarse-grain level of our elastic model.

Inspired by the MSMS algorithm, we consider a 'solvent-accessible sphere' around the C5' atom of each bp, of radius 9 Å. The protection signal associated to the

bp is the fraction of this sphere covered by a protecting macromolecule, either a protein or DNA. Proteins are modeled as spheres or cylinders, and DNA as a cylinder of radius 1.1 nm, centered and oriented by the closest bp and 2 nm long. To simplify the calculation of the overlap surface, cylinders are approximated by the tangent sphere at the intersection center point and of same radius (see Supplementary Data). Specific proteins were modeled with following parameters (Figure 3B):

- Core histones: cylinder aligned on the superhelical axis of wrapped DNA, of radius 3.25 nm and length 6 nm.
- gH1: Sphere centered on the dyad axis, 5.8 nm from the NCP center, and of radius 1.5 nm.
- H1 tail: Cylinder of radius 0.55 nm, and length 3 nm, tangent to gH1 sphere.

*Comparison of experiment and model-derived accessibility profiles.* The simplifications in the calculation of the solvent-accessible area appear reasonable since we focus on the positions of protected sites, and aim for a semi-quantitative measure for relative (not absolute) protection. This level of precision seems adequate for a comparison to an experimental signal, which also only represents an approximation of the DNA accessibility, relative to a free double-helix, determined via a comparison within a *local* window. It is justified *a posteriori* by the good correlation between predicted and measured accessibility profiles (Supplementary Data). In particular, the phase of the protected sites is accurately reproduced by both methods.

### Stem nanomechanics

For energy minimization, we employed the rigid base pair model of DNA elasticity (see above). Relaxation was conducted with sequence-independent elastic parameters (a further sequence-dependent relaxation showed that the global effects of sequence-dependence elasticity were negligible in this case). DNA volume exclusion was incorporated by placing purely repulsive, truncated Lennard-Jones spheres with 2.05 nm diameter around each base pair. To enforce contacts between the two DNA linkers at corresponding maximally protected sites, linear springs were introduced between the C5' atom positions at the minima of the accessibility profile. The positions were at distances 88.5 bp, 99.5 bp and 109.5 bp on each side of the dyad (see the accessibility profile Figure 4A and the corresponding springs, Figure 6B).

The springs had 0.7 nm rest length [chosen to leave space for the rolling probe of the MSMS algorithm (46)], except for the one connecting positions  $\pm 88.5$  bp, where the rest length was set to 1.3 nm to allow for insertion of lysines from the H1 C-terminal region. An alternative model where rest lengths were set to 0 showed only slightly different global conformations. An additional spring enforced unchanged linker separation at the height of the globular domain of H1 ( $\pm 80.5$  bp from the dyad), and was given a 3 nm rest length corresponding to the three-contact model (see Figure 5A). The initial

configuration was chosen with straight linkers (as in Figure 6B), and a conjugate gradient descent was carried out until convergence, keeping only the linker-core junction base pairs fixed.

The resulting 'stem' structure is shown on Figure 6C.

### Fluctuating nucleosomes

*Construction of the ensembles.* **-H1:** Ensemble of nucleosome conformations without linker histone. The rigid part is the central part of the core DNA in the crystal structure (PDB ID 1kx5). To account for observed dynamical properties of the NCP, we allow a partial unwrapping of the nucleosome: the three last anchor points of each end can be detached, with a dissociation energy of  $1k_B T$  per site, chosen from experimental values (48,49). The linkers fluctuate freely from the last attached anchor point. Here, we also excluded configurations where either linker [modeled as a sequence of 2.2 nm-diameter cylinders (Supplementary Data)], overlaps with the core histones (modeled as a cylinder of length 6 nm and diameter 6 nm, smaller than the most commonly accepted value of 6.5 nm). Note that the last protection of the NCP seems weaker in the -H1 case (Figure 1), which we interpret as a signature of partially unwrapped nucleosomes.

**gH1:** Ensemble of nucleosome conformations with the globular part of the linker histone (no tails). Following the assumption of a symmetrical binding of gH1, NCP and the first 9 bp of each linker are kept rigid (molecular contact at bp 8 from the NCP) in the conformation obtained from the relaxation, in which they contact gH1 (see Figure 7B).

**H1:** Ensemble of nucleosome conformations with 1-127 mutant linker histone. The electrostatic forces responsible for the stem are localized in the H1- globular and tail region. Yet the exact extension of that constraint region is *a priori* not obvious (not well-defined). We therefore generated three different ensembles, where the rigid region extends over different lengths in that range (16, 20 and 24 linker bp, see Figure 7).

*Trinucleosomes* were reconstituted by aligning 3 fluctuating mononucleosomes (of same sequence, taken in the same direction).

A quantitative comparison between CEMs and model trinucleosomes is delicate, because 3D-projection as well as fluctuations contribute to the apparent variety of conformations, so that it is difficult to avoid the introduction of a statistical bias in the selection of images.

In the CEM experiment, the 15 images of each experimental case had been selected manually from a larger ensemble of pictures, as the ones where all three nucleosomes were well-separated on the images, and the linkers were apparent. In contrast, in most images, two nucleosomes partly hide each other, or the linker DNA cannot be seen because it lies along the microscope optical axis.

To reproduce at best the experimental conditions, we first generated a large set of images of fluctuating trinucleosomes, seen from random directions. We then selected the first 15 of these images satisfying the same kind of criteria as in the experimental case.



Both experimental and model-derived images were then separated into groups according to the following criteria, ranging from ‘open’ structures towards more ‘closed’ ones. Red: the central nucleosome is widely unwrapped; Yellow: the central nucleosome is facing, and its in- and outgoing linkers start from separate points; Green: the central nucleosome is facing, the in- and outgoing linkers cross each other, and the external nucleosomes are in profile; Light blue: no well-defined orientation of the nucleosomes, with seemingly short linkers; Dark blue: the central nucleosome is in profile, the external nucleosomes are facing.

## RESULTS

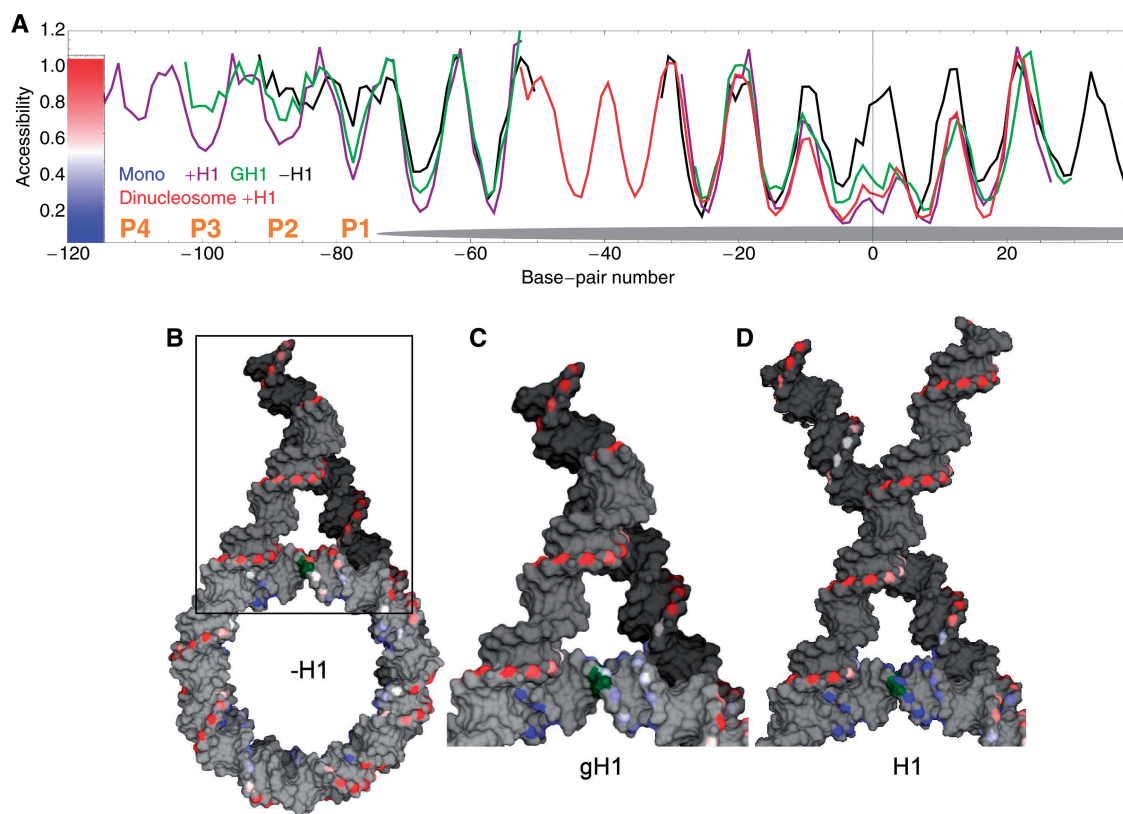
### (i) Relative accessibility and 3D structures

*Gel analysis.* The first part of our modeling is based on hydroxyl-radical footprints of mononucleosomes containing the 601 sequence ensuring well-defined positioning, and different mutants of the H1 linker histone. The proper association of the linker histone H1

(or truncated mutants) was ensured by the presence of the linker histone chaperone NAP-1, and the resulting complexes were validated by a combination of CEM and footprints. For all experimental details, please refer to (23).

The raw intensity traces of the gels are shown in Figure 1A. In a first step, we extracted the relevant information (the ~10-bp periodic protection pattern) from other features introduced by the experimental method. The traces were processed by automated band counting, band-wise integration and finally rescaling within a moving window (see ‘Materials and Methods’ section for details). The resulting signal represents the relative •OH accessibility per nucleotide, corrected for global trends and for irregularities in the gel (Figure 4A). We are able to identify the location and phasing of protected sites with single bp resolution, and to provide qualitative information on the degree of protection.

In absence of H1 (-H1), the linker DNA is fully unprotected from •OH attack. The core DNA exhibits a 10-bp periodic protection, which we attribute to the core histones. The presence of the globular part of H1 (gH1,



**Figure 4.** (A) Relative accessibility as obtained from the processing of the gels: mononucleosome without H1 (black), with gH1 (green), with H1 (purple) and dinucleosome with H1 (red). The position of the NCP is indicated by a gray ellipse, and the protected sites P1–P4 are indicated in orange. The correspondance with the color-coding scheme used on (B)–(D) is shown on the left. On the linker DNA, the protection is weaker than on the NCP, so that protected sites appear white rather than blue. (B) Model of the nucleosomal DNA without H1 (-H1) with color-coding of the C5' atoms from blue (protected) to red (accessible). C5' atoms without footprinting data and all other DNA are shown in gray, and the dyad is indicated in green. View from the NCP superhelical axis. The protected sites are facing the histone octamer, validating the quantitative positioning of the protection trace. The co-localization of the protected sites from both strands supports the symmetry hypothesis. (C) Same for gH1-bound nucleosomal DNA, shown with straight linker DNA. New protections can be identified in the dyad region (C5' atoms facing outside the NCP), and on the first turn of the linker DNA. (D) Same for H1-bound nucleosomal DNA, with straight linker DNA arms of length 38 bp, exhibiting additional protection on the linker (white spots).



mutant 35-112) protects the dyad region of the core DNA and a site in the first helical turn of the linker DNA (quoted site P1). Mutants with C-terminus truncated before AA 127 exhibit the same pattern. We interpret these protected sites as positions of DNA-gH1 contact or vicinity. Mutants truncated at AA 127 further exhibit the same pattern as that of complete H1: in addition to gH1 protections, the linker DNA exhibits a 10-bp periodic protection (P2, P3, P4). We interpret this pattern as a signature of the stem structure, in which case the linker DNA is protected either by H1 tail or by direct vicinity of the other linker branch.

**3D-rendering of the accessibility.** The  $\bullet\text{OH}$  radicals used in the footprinting are known to primarily attack the C5' carbon atoms of the backbone sugars (43), allowing us to pinpoint protected sites in 3D molecular models of the nucleosome with Ångström resolution.

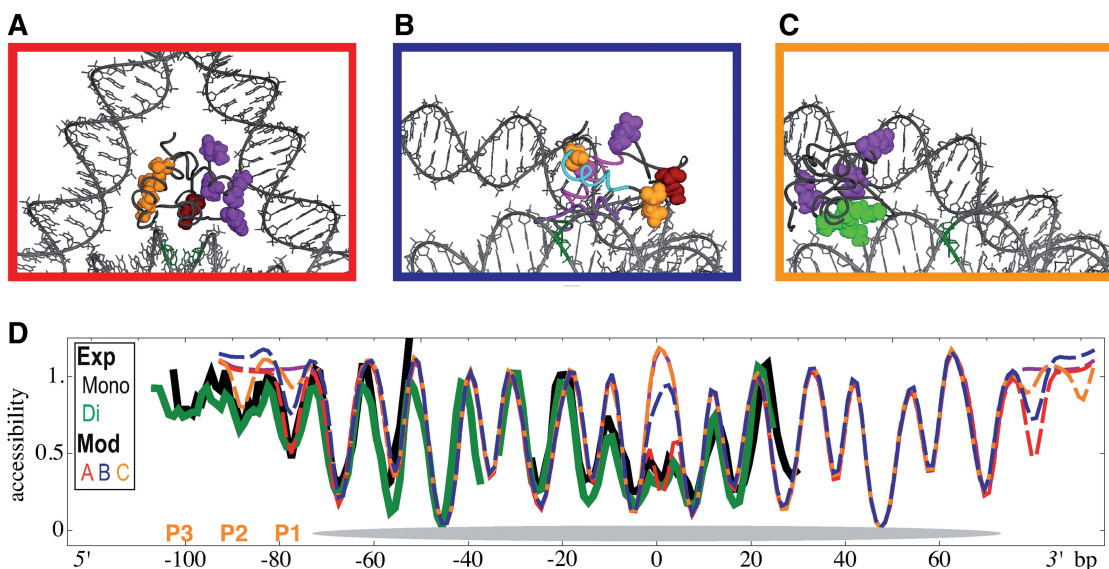
The molecular visualization package Chimera (41) allows the rendering of molecular structures using a color code for user-defined atom attributes. This feature was used to present the relative accessibility signals from the footprinting experiments by color coding the deoxyribose C5' atoms from blue (least accessible) over white to red (most accessible). Footprints were measured for one of the strands. Color-coding on both strands of DNA was displayed, by exploiting the approximate 2-fold symmetry of the nucleosome. Bases for which no single nucleotide resolution footprinting was available were not colored. The coloring scheme is first used to show the protection pattern of H1-less nucleosomes in Figure 4B. As expected,

this '3D-gel' shows directly that the 10.5-base periodicity of the experimental accessibility signal places all protected sites on one side of the double helix, facing inwards the NCP, whereas the external C5' atoms remain unprotected, as well as the linker DNA.

## (ii) Molecular modeling of gH1 placement

Addition of the globular domain gH1 induces additional protected sites on the core DNA and at the entering and exiting linkers [see Figure 1 and Figure 4A and C. Also visible in the movies of (23), Supplementary Data]. We addressed the question whether existing models of linker histone placement are compatible with the observed protection patterns. Three specific models were considered (23): a three-contact model (24) where the linker histone is placed between and contacting both the entry and exit linker DNA and the dyad. Alternatively, in the two-contact models by Zhou *et al.* (25) and Brown *et al.* (26), the linker histone is placed between one linker and a site on core DNA, contacting only a single linker (Figure 5A–C).

To establish a semi-quantitative relation between these structural models and  $\bullet\text{OH}$  footprints, we calculated footprint predictions for different structural models. C5' atoms reactivity is determined by their solvent-accessible surface area (43). Using a simplified method inspired by existing algorithms, we were able to compare these structure-derived predictions with the measured profile, as shown in Figure 5D.



**Figure 5.** Comparison of the experimentally observed protection in the presence of gH1 with relative accessibilities calculated from three different structural models for the location of the globular domain: (A) Three-contact nucleosome configuration by Fan *et al.* (24). The viewing direction is the superhelical axis. See 'Materials and Methods' section for details. (B) Two-contact nucleosome configuration proposed by Zhou *et al.* (25). Contact is established with core DNA at 1–4 bp from the dyad, and with one DNA linker (the other linker is not shown). The viewing direction is the superhelical axis. (C) Two-contact nucleosome configuration by Brown *et al.* (26). Contact is established about 5 bp away from the dyad, and with one linker DNA. (D) Experimentally observed protection: thick solid lines: black (mononucleosome), green (dinucleosome). Structure-derived accessibilities: dashed lines: red (three-contact, A), blue [two-contact by Zhou *et al.* (25), B] and orange [two-contact by Brown *et al.* (26), C], and based on a pure mononucleosome without H1 (magenta) [data already published in (23)]. The predictions differ in the protection at the dyad where the two-contact models show no or very weak protection, and at the entry/exit linkers. The measured relative accessibility for a gH1-bound mononucleosome is shown in black, and that of a gH1-bound dinucleosome is shown in green.

The two-contact model by Brown *et al.* (26) fails to reproduce protection at the dyad. It also incorrectly predicts stronger protection at  $-90$  bp than at  $-80$  bp. Since the contacts between gH1 and the core DNA at about 10 bp distance from the dyad are in the major groove, they do not protect the DNA backbone C5' atoms from  $\bullet$ OH attack. As a result, there is no footprint of this model on core DNA at all. The two-contact model by Zhou *et al.* (25) gives better predictions for linker DNA, generating a protected site at  $-80$  bp (P1). However it fails to reproduce the strong protection pattern at the dyad, despite the proximity; here again, protein contacts in the major groove cannot generate sufficient  $\bullet$ OH protection. In contrast, the three-contact model is compatible with the experimentally observed protection pattern, reproducing both the characteristic double-peak dyad protection at bp 2 and the protected site at  $-80$  bp.

### (iii) Fully-protected stem structure based on DNA nano-mechanics

For the linker DNA stem, models based on high-resolution experimental studies are not available.

We argue that under the assumption that the protected sites in the stem arise from DNA–DNA contacts (50), knowledge of (i) the detailed register of the protected sites along the stem and (ii) DNA nanoscale structure and elasticity are necessary and sufficient to extract valuable structural information.

To see that both types of information are required, consider first stem structures with juxtaposed, weakly deformed DNA entry and exit linkers. In general, these structures will fail to create contacts at the observed sites of maximal protection, see Figure 6A. Second, we note that any model for the center-line trajectory of the DNA linkers can be modified such that the protected sites on the two linkers face each other; this can be achieved simply by twisting DNA appropriately along its contour. It is thus impossible to conclude on a particular shape of the linker DNA center-lines purely on the basis of the geometric

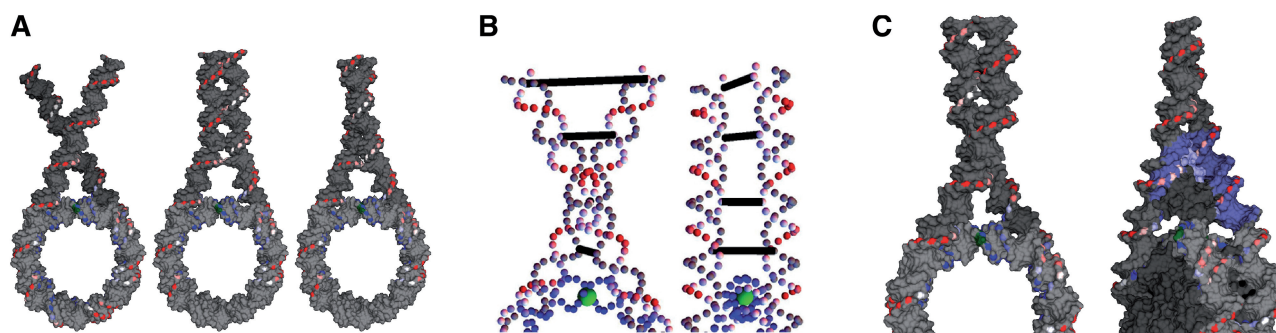
arrangement of protection patterns, neglecting the physical properties of DNA.

To construct a model for the linker stem structure we therefore minimized the DNA nanoscale elastic energy under the constraint that the alignment of the linkers in space reproduces the observed protection pattern (see ‘Materials and Methods’ section for details). The resulting stem structure is shown in Figure 6C. The linkers come together  $\sim 20$  bases outside the core particle, slightly curving into a two-start superhelical stem with a large pitch of around 100–120 bp, and extending at least to 40 bp from the NCP. This structure has, as the core particle itself, a 2-fold symmetry.

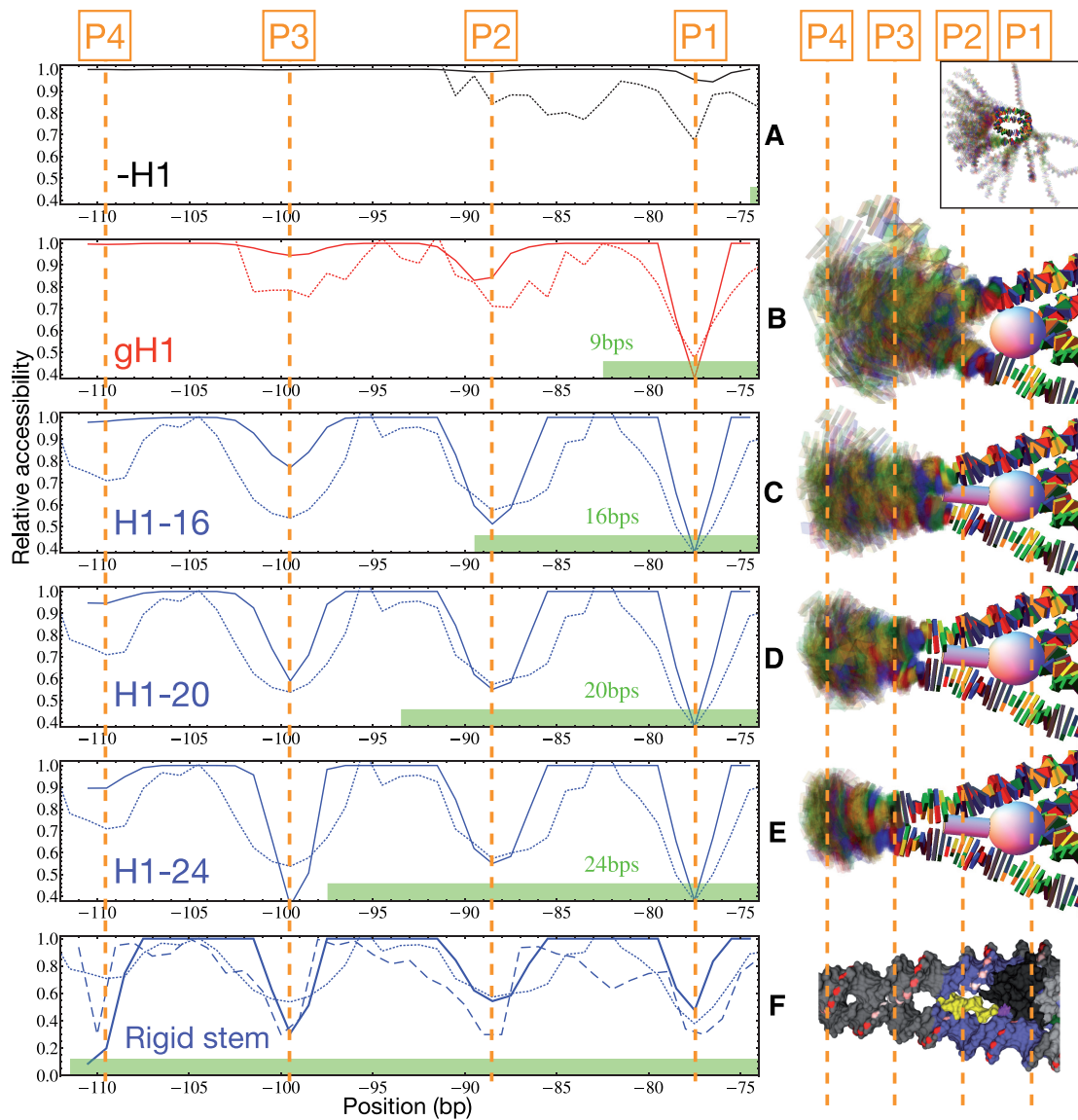
### (iv) Soft structure based on nanoscale modeling of fluctuations

By construction, there is perfect agreement for the location and phasing of the protected sites in the experimental traces and in the structure-derived accessibility profile of the proposed stem structure (Figure 7F). There is also qualitative agreement between the calculated and the observed degree of protection, even though one might argue that the model fails to reproduce the apparently weaker protection at the outermost P4 site. Since neither model nor experiment yield *quantitative* predictions or measurements of the DNA accessibility in the complex relative to naked DNA, it is difficult to refine the modeling or to draw more definite conclusions on the basis of the footprints alone.

However, a serious objection to the model presented thus far and in (23) can be raised from a direct visual inspection of the CEM images of trinucleosomes: stems of corresponding size can only be distinguished in some of the CEM images (Figure 1B, row 3, image 2). Most images show considerable angles between the in- and outgoing linkers of the central nucleosome, i.e. conformations for which we would not expect any mutual protection (Figure 1, row 3, images 1, 3, 4, 5: note that the linker DNA between two successive NCPs is here 53 bp long,



**Figure 6.** Nanoscale modeling of the stem (A) H1-protection color-coded nucleosomes with straight linkers (left) and two geometrical stem models. For both of them, protected sites (white spots on the linker) do not face each other. These models do not account for the observed protection pattern. (B) To determine the most likely stem structure compatible with the observed protection pattern, we minimized the DNA nanoscale elastic energy under the constraint that the most protected sites face each other. The initial configuration of the relaxation is illustrated here, with straight linkers. 4 springs (black cylinders) enforce contact between the protected sites. Only C5' atoms were depicted, color-coded according to the experimental protection pattern (gray when no signal). The green sphere represents the dyad. Views from the NCP superhelical axis and perpendicular. (C) Stem structure obtained as a result of the relaxation [already shown in (23)]. Views in direction of the superhelical axis and 30° apart (dark gray histones, light blue histone tail with arbitrary conformation). The DNA is colored in blue within possible extension of the truncated H1 tail.



**Figure 7.** Left hand-side: comparison of the (coarse-grain) model-derived ensemble-averaged relative accessibility (solid line) with the corresponding experimental relative accessibility (dotted line). The green bars indicate the rigid part. Vertical orange dotted lines are the maximally protected sites in the experimental data. Right hand-side: superposition of 40 snapshots of the fluctuating linkers (vertical orange dotted lines: same as on the left hand-side). (A) -H1 ensemble; (B) gH1 ensemble; (C)–(E) H1 ensembles with 16, 20, 24 bp kept rigid, respectively. As expected by construction, the most rigid stem model (24 rigid bp) reproduces the observed pattern. The apparent effect of the fluctuations is to weaken the mean protection, so that the protection of the 2 external sites (P3, P4) fades in the softest ensemble (16 rigid bp). (F) Rigid fully-protected stem structure. Left: additional dashed line: atomistic structure-derived accessibility. Right: Molecular model of the rigid stem, with DNA shown gray, color-coded protected sites (see Figure 4), gH1 shown black, H1 tail shown yellow (arbitrary conformation) and DNA within possible extension of the truncated H1 tail shown blue.

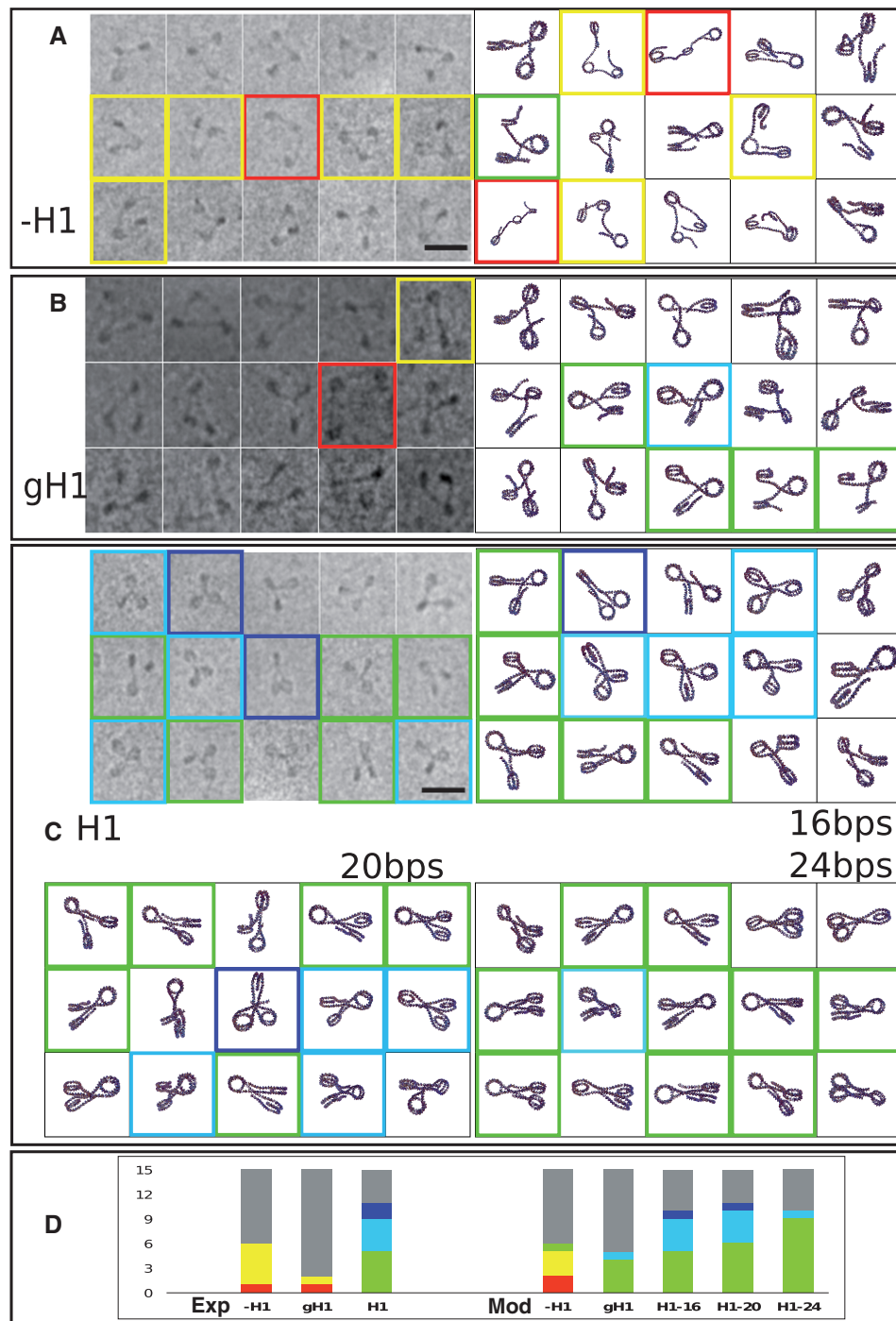
so that a 40 bp-long stem would fill nearly all of it. More images in Figure 8).

To resolve this apparent contradiction, we note that the stem structure provided by the relaxation is the *most likely* structure accounting for protection at the observed sites. However, there is no reason to assume that the stem is rigid and always found in this conformation. Instead of

considering our fully protected structure as ‘the’ stem structure, we now develop a description of the stem in terms of an *ensemble of thermally fluctuating, partially protected structures*.

How and where do we expect fluctuations to occur? Can we obtain meaningful results without a detailed modeling of the physical interactions responsible for the stem





**Figure 8.** Comparison of CEM trinucleosome pictures and model-derived trinucleosome snapshots, chosen randomly and appropriately projected. Images with five typical shapes of trinucleosomes were colored: from red (open structure) to blue (joined linkers). (A) Experimental and model trinucleosomes chosen in the -H1 ensemble. (B) Same in the gH1 ensemble. (C) Experimental and model trinucleosomes chosen in the H1 ensembles, with rigid parts of indicated lengths (softer 16bp, 20bp, most rigid 24bp). (D) Repartition of trinucleosome shapes (indicated by colors) in the experimental C-EMs and the pictures from the model trinucleosome ensembles. Pictures which could not be assigned a color appear gray. Only the two softer models can account for the most open conformations observed in the pictures. These observations together with Figure 7 suggest a typical rigidity extension of ~20 linker bp, accounting for both experimental data. Experimental data already published in (23).

formation (gH1 docking, tail-linker interactions including a full exploration of the conformational freedom of the histone tails, ion mediated DNA–DNA interactions, zipper-motives etc)?

We note that the DNA deformation free energy in the fully protected stem structure is small and not uniformly

distributed: 90% of a total of  $2k_B T$  are localized in the first ~20 bp of the stem, close to gH1. While the binding of gH1 alone seems ineffective, H1 variants with a small fraction of the tail induce a noticeable level of stem-like associations of the incoming and outgoing linker. In modeling fluctuations in this structure, it seems reasonable



to assume that the effect of H1 decreases with physical distance from the molecule.

We therefore expect the stem to open and close in a zipper-like fashion from a nucleotide around the region where H1 is localized. The simplest way to build a range of corresponding models is to divide our original, fully protected structure into two zones: completely rigid up to a variable position beyond the identified docking points with the H1 globular domain and completely free beyond. By construction, our minimal-energy structure of the fully protected stem is part of all ensembles.

We have considered three variants of the stem with 16, 20, 24 rigid bp respectively. For comparison, we also built ensembles for gH1 (9 rigid bp) and -H1 nucleosomes (allowing partial unwrapping of the core DNA, see 'Materials and Methods' section). For each model we have generated representative mono- and trinucleosome conformations. The fluctuating part of each ensemble is the result of thermal fluctuations of DNA modeled as a rigid base pair chain (30–33), with no adjustable parameter. The superpositions of aligned mononucleosome conformations shown in Figure 7 illustrate the range of fluctuations in the different ensembles with a crossover from extremely floppy to very rigid linkers. For each ensemble we have analyzed protection patterns (Figure 7) and trinucleosome snapshots (Figure 8), to see if we can reconcile the experimental footprinting and CEM data. The mononucleosome protection patterns shown in Figure 7 were computed directly on the nano-scale structures using a coarse-grained variant of the solvent-accessible area method. As to the trinucleosome conformations, we obviously do not expect to match the cryo-EM pictures one-by-one in a comparison to a correspondingly small ensemble of appropriately projected model conformations. Rather, we have classified snapshots into five categories from open structures (red) to more closed ones (blue), allowing us to compare ensembles according to a coarse statistics (we give the number of configurations of each color group, red-yellow-green-light blue-dark blue). For the definitions of the groups, see 'Materials and Methods' section, last section. In the following we discuss the results for the various ensembles, where an increasing fraction of the linker DNA is held fixed in a rigid 'root' of the nucleosomal stem.

**-H1:** the gels show a protected site just at the level of statistical noise, at the position of the gH1 domain (P1, -79 bp), and no other protected sites (Figure 7A). The -H1 ensemble exhibits a weak protection near the same position, generated by those nucleosomes where one linker arm transiently wraps around the core histones.

Furthermore, the ensemble reproduces the characteristic open trinucleosome conformations observed in the cryo-EM experiments. The repartition of the snapshots among the 'color groups', ranging from red (open structures) to blue, are: Experimental: 1-5-0-0-0; Model: 2-3-1-0-0 (Figure 8).

**gH1:** experimental traces show weakly protected sites at the same positions as the ones observed for H1 (P2, P3 in the resolved region of that gel lane). The predicted positions for transiently protected sites are those observed,

generated by particular conformations in which linkers spontaneously protect each other.

The model-derived trinucleosome pictures reproduce the tendency toward more closed conformations observed in the CEM pictures. Because of the apparent variety of shapes and the poorer quality of the images, it is difficult to draw any conclusion (Exp: 1-1-0-0-0, Mod: 0-0-4-1-0).

**H1:** Figure 7C–E show the protection patterns derived from the three ensembles with 16, 20, and 24 bp constrained to their positions in our minimal energy conformation for the fully protected stem. The extension of the rigid parts are indicated by green bars at the bottom of the graphs. Within these zones, the ensembles reproduce by construction the experimental protection patterns as well as the minimal elastic energy conformation of the fully protected stem (Figure 7F). As expected, the fluctuations reduce the protection in the outer zones. Interestingly, the transient contacts in the fluctuating ensembles are preferentially located at the experimentally observed positions. This is not an accident, but a direct consequence of the small elastic deformation energies in the outer part of our fully protected stem structure: the preferred positions of transient contacts in our model are controlled by the elastic properties of the linker DNA and the boundary conditions imposed by the rigidly held part. While the protection at P3 remains discernible in all three ensembles, the most flexible variant with only 16 fixed bp generates too little protection at P4 to be compatible with the experimental data. In contrast, the comparison of the model-derived trinucleosome images with the cryo EM pictures rather favors the softer models of the stem. The 16 and the 20 rigid bp models generate comparable distributions and varieties of structures, while the conformations from the 24-rigid bp model appear too uniform. (Exp: 0-0-5-4-2; Mod: H1-16 0-0-5-4-1, H1- 20 0-0-6-4-1, H1-24 0-0-9-1-0).

The comparisons show that we arrive at a coherent description, if we assume that the nucleosomal stem includes  $20 \pm 2$  bp of linker DNA: (i) closed conformations occur with sufficient probability to explain the experimentally observed protection; (ii) the corresponding low resolution structures of tri-nucleosomes reproduce the conformational statistics of the corresponding cryo-pictures.

## DISCUSSION

Based on our results, we view the nucleosomal stem as a dynamic, polymorphic, hierarchically organized structure composed of several parts:

*A root* comprising the globular part of H1 and the first 10 bp of the linkers, where gH1 preferentially establishes three contacts: one at the dyad of the histone octamer and two with the linkers, which remain on average straight and symmetric up to bp 8. The formation of the root considerably reduces the range of fluctuations of the linkers and suppresses unwrapping from the histone core.

*A trunk* or relatively rigid inner part of the stem, comprising the linkers up to  $bp 20 \pm 2$ , in direct contact with the cationic amino acids of the C-terminus of H1. It is

in this region, that the DNA is deformed to partially align the two linkers. In our nanomechanical model, 90% of the elastic energy of the fully-protected stem structure of  $\sim 2k_B T$  are located in the trunk. In our experiments, the formation of the trunk required a tail length of at least 15 AA (H1-127), indicating a biochemical control mechanism for this step.

A *flexible crown* or outer stem where the branching linkers exhibit substantial fluctuations, while preserving well-defined preferential contacts. We note that by imposing a boundary condition on the linker conformation, the influence of the trunk structure may extend beyond the region of direct interactions between H1 and the linker DNA. In particular, the linkers might appear connected without there being strong direct interactions. In our experiments, this influence extends to at least 40 bp away from the NCP, reaching the typical linker lengths of native fibers (40 bp for a *half-linker*). Our experiments (23) also suggest that the full C-terminus, while possibly stabilizing further the stem, does not qualitatively modify its structure. As a natural explanation for this effect, we suggest that the terminus may remain confined in the trunk region.

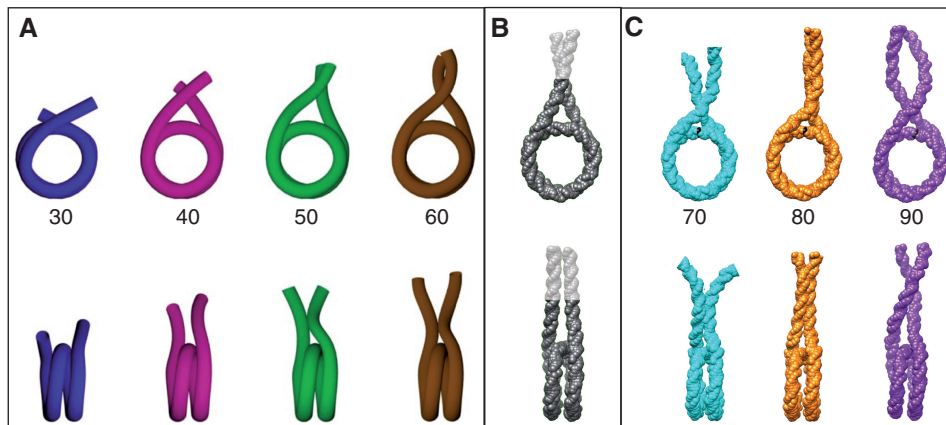
The hierarchical organization implies that the stem opens and closes in a zipper-like fashion. Under our experimental conditions, thermal fluctuations mainly affect the branching linkers in the crown. The response to external forces depends on their magnitude. While weak forces should essentially deform the crown, larger forces should disrupt the trunk and eventually the root before unwrapping the core particle (51). Such forces may be exerted in a controlled manner in single-molecule experiments (52). They also arise during the condensation of the chromatin fiber, where H1 might accommodate a fairly large range of linker conformations, if other interactions (18,20,21) compensate the free energy cost of a stem deformation or a partial stem disruption. We note, that this view of the stem is a natural extension of the current, dynamic picture of the nucleosome core particle: instead of a passive and rigid wrapping of 147 bp of DNA in a conformation resembling crystal structure(s) (7,53), DNA and histone octamers form a highly dynamic complex, where DNA spontaneously unwraps and rewraps from the ends (48,49,54) with actively created (55), diffusing (56,57) defects ensuring mobility (58).

How do our results match those of previous studies? There is a remarkable diversity of experimental (25,26) and modeling (24,59–62) results for the mode in which gH1 binds in what we now refer to as the ‘root’ of the stem. The structures presented here are derived from experiments where mono-, di- and tri-nucleosomes were reconstituted following a carefully elaborated protocol recreating *in vivo* conditions such as the presence of the chaperone NAP-1. If we believe them to represent a free energy minimum for systems dominated by intra-stem interactions, it is an interesting question, whether other binding modes might serve to stabilize alternate structures in nucleosomes under external constraints. If one interprets the multitude of predicted gH1 binding modes in the root as a *feature* of the molecule and not as a failure of the employed modeling schemes, then the ability of the

stem to adapt to external constraints might be even larger than apparent from our experiments.

Little was known about the trunk region. Neglecting fluctuations, Bharath *et al.* (60) used structural analogies and bioinformatics methods to predict a placement of gH1 close to the two-contacts model of Zhoul *et al.* (25). For the trunk, they proposed a particular conformation of the C-terminus making contact with the linker DNA up to 24 bp away from the NCP, beyond which the sharply bent linkers were supposed to diverge in the crown. While the predicted extension of the H1-DNA contacts is rather close to that of our trunk, the details of the proposed stem structure are incompatible with our experimental results. This holds for the resulting protection pattern close to the DNA dyad as well as for the predicted divergence of the two linkers beyond the contact zone with the H1 tail, which is difficult to reconcile with the experimentally observed protections P3 and P4 (see Figure 7).

Interestingly, our results shed some new light on the question, which factors determine the structure of dense chromatin fibers. Is it the structure and elasticity of the stem, which discriminates between the possible helical arrangements of nucleosomes? Or do the linkers simply adapt to whatever constraints follow from the packing and the interactions of the core particles (63), thereby possibly frustrating the formation of a proper stem? Robinson *et al.* (35) reported the properties of well-defined chromatin fibers reconstituted *in vitro* from poly-601 templates in presence of the linker histone H5. The templates were prepared for a range of well-controlled linker lengths (multiples of 10, from 30 to 90 bp). Based on these experiments, Wong *et al.* (59) used DNA elastic models to infer the linker DNA paths in the Robinson *et al.* (35) reconstituted fibers. The resulting Wong *et al.* (59) ‘most favorable’ structures shown in Figure 9 (A and C) belong to a remarkable variety of helix families. Nevertheless, the authors argued that (in our language) the DNA conformation in the root is approximately conserved among the structures: *a posteriori*, this feature allows for a common asymmetric three-contact binding mode of gH1/5, rather similar to the one inferred from our experiments. The only exception is the most favorable structure for the shortest linkers (30 bp or 15 bp for a *half-linker*), where the steric constraints seem indeed too strong to allow for the formation of the structure preferred by gH1 and where Wong *et al.* (59) propose an alternative binding mode. Otherwise, it is striking to see on Figure 9A, how, with increasing linker length, the Wong *et al.* (59) conformations reproduce larger and larger parts of the mono-nucleosomal stem shown in Figure 9B. Linker lengths 40 bp and larger thus possibly allow for the formation of a (deformed) trunk and further stabilization by the H1 tail, whereas they exhibit substantial variations in the external part corresponding to our crown (Figure 9C). At least qualitatively, the comparison of the two completely independent studies suggests a remarkable agreement between linker conformations required for achieving nucleosome packing in dense chromatin fibers and those stabilized by intra-stem interactions. Furthermore, the comparison illustrates the role of the polymorphic structure and hierarchic organization of the stem in stabilizing



**Figure 9.** Comparison of our mono-nucleosome stem structures to the inferred (59) linker conformations in model chromatin fibers reconstituted from poly-601 templates (35). View along the superhelical axis (top) and perpendicular (bottom). For the fiber conformations, only the first half of both linkers is shown. (A) Wong *et al.* (59) most favorable structures for linker length 30, 40, 50 and 60 bp (from left to right), corresponding to a fiber diameter of 35 nm. Picture courtesy of Julien Mozziconacci. (B) Ground state of the mono-nucleosome stem: the root and trunk are those of the fully-protected structure, and the flexible crown or outer stem (shown in brighter colors) is straight. (C) Wong *et al.* (59) most favorable structures for linker length 70, 80, 90 (from left to right), corresponding to a fiber diameter of 45 nm. The black dot indicates the dyad. Original pictures were not available: the linker conformations were rebuilt approximately, by hand, from the available data provided by Wong *et al.* (59). For all but the shortest linkers (30 bp), the root part is approximately conserved among the structures, allowing for a common three-contact asymmetrical binding mechanism of gH1. Longer linkers (50 bp and beyond) adopt a conformation compatible with the formation of a trunk and additional stabilization by the H1 tail.

a wide variety of fibers. From a kinetic point of view, the emerging picture is that of a folding funnel (64), where the local stem formation helps to achieve the cooperative fiber folding. Some of these features may already be included in a coarse-grain model of the chromatin fiber (14,65), which is compatible with our description of the root of the stem. Quantitatively, we crucially miss experimental or computational estimates of the orientation-dependent interactions between the core particles and of the (deformation-dependent) stem formation and interaction free energies to be able to embark on the systematic modeling of fiber conformations.

## SUMMARY AND CONCLUSION

While the structure of the nucleosome core particle is known with Ångström resolution, much less structural information is available for the soft parts of the nucleosome (linker DNA, linker histone H1/H5, histone tails), which strongly influence the structure of the chromatin fiber. The relevant structures are huge on the molecular scale and their intrinsic softness is a major obstacle to diffraction or NMR studies. On the other hand, they are too small to be imaged with the required resolution. Here, we have presented a detailed report of our attempt to develop a three-dimensional, dynamical coarse-grain model of the nucleosomal stem formed by the histone H1/H5 and the in- and outgoing linker DNA. We have combined available crystal and NMR structures (NCP and gH1, respectively) and the knowledge of the (sequence-dependent) B-DNA structure and elasticity with the results of our CEM and ●OH footprinting experiments for carefully reconstituted model nucleosomes (23).

This combination lead to a surprisingly clear picture of the nucleosome and of how H1 binds to and organizes

nucleosomal DNA. In a first step of our analysis, we have shown that the calculated footprinting pattern of a three-contact gH1-nucleosome structure based on a model (24) for the placement of H5 (gH5) (Figure 5A) matches very well the experimental pattern, whereas suggested two-contact models (25,26) were incompatible with the strong protection observed at the dyad.

In a second step, we reasoned that the observed periodic protections on the linker DNA stem are a signature of DNA–DNA contacts, and that the precise positioning of the protected sites provides a valuable information. To model the stem structure, we have used a nanomechanical model of DNA, compatible with the qualitative character of the measured protection amplitude. We have aligned the linkers in space in a way that their mutual protection reproduces the measured accessibility profile and have assumed that the most likely structure has minimal DNA elastic energy. The resulting stem structure is shown in Figure 6C. The linkers come together ~20 bases outside the core particle, slightly curving into a two-start superhelical stem with a large pitch of around 100–120 bp, and extending at least to 40 bp from the NCP. This structure has, as the core particle itself, a 2-fold symmetry.

In the third step, we have developed a description of the stem in terms of an ensemble of fluctuating, partially-protected structures. We have shown that (i) transient contacts are sufficiently well-defined to cause an experimentally observable partial protection if we assume a stem length of 20 or 24 bp, and (ii) the CEM pictures of reconstituted tri-nucleosomes are best reproduced by ensembles of fluctuating stems with a rigid part of 16 or 20 bps. Combining these results, we therefore estimate that the rigid part of the stem incorporates  $20 \pm 2$  bps of linker DNA. Interestingly, this



corresponds to the extent of the linker DNA which is close enough to the globular part of H1 to directly interact with the (truncated) COOH-terminus of H1, which plays a crucial role in the stem formation.

In the final step, we have developed a picture of the nucleosomal stem as a dynamic, polymorphic, hierarchically organized structure, composed of a 'root' where gH1 binds to the first ~10 bp of the DNA linkers, a 'trunk' formed by the association of the subsequent  $10 \pm 2$  bp with the cationic C-terminus of H1, and a flexible 'crown' or outer stem where the branching linkers exhibit substantial fluctuations, while preserving well-defined preferential contacts. To understand the role of H1 in the folding of compact chromatin fibers, we have compared our (thermal) ensemble of *mono*-nucleosome stem conformations to an ensemble of 'most favorable' linker conformations (59) inferred from experiments on reconstituted poly-601 *fibers* (35). Remarkably, for a wide range of linker lengths, the inferred linker conformations in the poly-601 fibers appear compatible with slightly deformed variants of the *mono*-nucleosomal stem. For shorter linkers, where the packing constraints become more severe, partial binding of H1 can still contribute to the stabilization of dense fibers. Finally, we hypothesize that the cooperative folding of dense chromatin fibers is facilitated by the hierarchical of the stem and the tendency of local intra-stem interactions to stabilize linker conformations, which prevent non organization-local steric clashes between nucleosomes.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

S.D. acknowledges La Ligue Nationale contre le Cancer (Equipe labellisée La Ligue) and the Indian-French program ARCUS. We acknowledge Julien Mozziconacci and Jean-Marc Victor (51) for providing us their data and for helpful discussions.

## FUNDING

Grants from Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique; Agence Nationale de la Recherche 'EPIVAR' N 08-BLAN-0320-02 (to S.D.); ANR-09-BLAN-NT09-485720 'CHROREMBER' (to D.A., S.D. and J.B.); Chair of excellence program from the ANR (to R.E.); Grant Agency of the Czech Republic (LC535,30 MSM0021620806 and AV0Z50110509). Funding for open access charge: INSERM.

*Conflict of interest statement.* None declared.

## REFERENCES

- Ball, P. (2003) Portrait of a molecule. *Nature*, **421**, 421–422.
- Kornberg, R.D. (1997) Structure of chromatin. *Annu. Rev. Biochem.*, **46**, 931–954, 1977.
- Kornberg, R.D. (1974) Chromatin structure: a repeating unit of histones and DNA. *Science*, **184**, 868–871.
- Van Holde, K.E. (1989) *Chromatin*. Springer, New York.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251–260.
- Muthurajan, U.M., Bao, Y., Forsberg, L.J., Edayathumangalam, R.S., Dyer, P.M., White, C.L. and Luger, K. (2004) Crystal structures of histone sin mutant nucleosomes reveal altered protein-DNA interactions. *EMBO J.*, **23**, 260–271.
- Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W. and Richmond, T.J. (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.*, **319**, 1097–1113.
- Richmond, T.J. and Davey, C.A. (2003) The structure of DNA in the nucleosome core. *Nature*, **423**, 145–150.
- Makarov, V.L., Dimitrov, S.I., Tsaneva, I.R. and Pashev, I.G. (1984) The role of histone H1 and non-structured domains of core histones in maintaining the orientation of nucleosomes within the chromatin fiber. *Biochem. Biophys. Res. Commun.*, **122**, 1021–1027.
- de la Barre, A.E., Angelov, D., Molla, A. and Dimitrov, S. (2001) The N-terminus of histone H2B, but not that of histone H3 or its phosphorylation, is essential for chromosome condensation. *EMBO J.*, **20**, 6383–6393.
- Claudet, C., Angelov, D., Bouvet, P., Dimitrov, S. and Bednar, J. (2005) Histone octamer instability under single molecule experiment conditions. *J. Biol. Chem.*, **280**, 19958–19965.
- Woodcock, C.L., Grigoryev, S.A., Horowitz, R.A. and Whitaker, N. (1993) A chromatin folding model that incorporates linker variability generates fibers resembling the native structures. *Proc. Natl Acad. Sci. USA*, **90**, 9021–9025.
- Zlatanova, J., Leuba, S.H. and van Holde, K. (1998) Chromatin fiber structure: morphology, molecular determinants, structural transitions. *Biophys. J.*, **74**, 2554–2566.
- Wedemann, G. and Langowski, J. (2002) Computer simulation of the 30-nanometer chromatin fiber. *Biophys. J.*, **82**, 2847–2859.
- Schiessel, H. (2003) The physics of chromatin. *J. Phys. Cond. Mat.*, **15**, 699–774.
- Arents, G. and Moudrianakis, E.N. (1995) The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc. Natl Acad. Sci. USA*, **92**, 11170–11174.
- Mangenot, S., Leforestier, A., Vachette, P., Durand, D. and Livolant, F. (2002) Salt-induced conformation and interaction changes of nucleosome core particles. *Biophys. J.*, **82**(1 Pt 1), 345–356.
- Dorigo, B., Schalch, T., Bystricky, K. and Richmond, T.J. (2003) Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *J. Mol. Biol.*, **327**, 85–96.
- Muhlbacher, F., Schiessel, H. and Holm, C. (2006) Tail-induced attraction between nucleosome core particles. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **74**(3 Pt 1), 031919.
- Arya, G. and Schlick, T. (2006) Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proc. Natl Acad. Sci. USA*, **103**, 16236–16241.
- Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M.J., Davie, J.R. and Peterson, C.L. (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*, **311**, 844–847.
- Robinson, P.J.J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R.G. and Rhodes, D. (2008) 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. *J. Mol. Biol.*, **381**, 816–825.
- Syed, S.H., Goutte-Gattat, D., Becker, N., Meyer, S., Shukla, M.S., Hayes, J.J., Everaers, R., Angelov, D., Bednar, J. and Dimitrov, S. (2010) Single-base resolution mapping of H1-nucleosome interactions and 3D organization of the nucleosome. *Proc. Natl Acad. Sci. USA*, **107**, 9620–9625.
- Fan, L. and Roberts, V.A. (2006) Complex of linker histone H5 with the nucleosome and its implications for chromatin packing. *Proc. Natl Acad. Sci. USA*, **103**, 8384–8389.



25. Zhou, Y.B., Gerchman, S.E., Ramakrishnan, V., Travers, A. and Muyltermans, S. (1998) Position and orientation of the globular domain of linker histone H5 on the nucleosome. *Nature*, **395**, 402–405.
26. Brown, D.T., Izard, T. and Misteli, T. (2006) Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo. *Nat. Struct. Mol. Biol.*, **13**, 250–255.
27. Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L. and Sweet, R.M. (1993) Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature*, **362**, 219–223.
28. Bednar, J., Horowitz, R.A., Grigoryev, S.A., Carruthers, L.M., Hansen, J.C., Koster, A.J. and Woodcock, C.L. (1998) Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc. Natl Acad. Sci. USA*, **95**, 14173–14178.
29. Lowary, P.T. and Widom, J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.*, **276**, 19–42.
30. Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M. and Zhurkin, V.B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl Acad. Sci. USA*, **95**, 11163–11168.
31. Lankas, F., Sponer, J., Langowski, J. and Cheatham, T.E. 3rd (2003) DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.*, **85**, 2872–2883.
32. Becker, N.B., Wolff, L. and Everaers, R. (2006) Indirect readout: detection of optimized subsequences and calculation of relative binding affinities using different DNA elastic potentials. *Nucleic Acids Res.*, **34**, 5638–5649.
33. Becker, N.B. and Everaers, R. (2009) DNA nanomechanics in the nucleosome. *Structure*, **17**, 579–589.
34. Lavery, R., Zakrzewska, K., Beveridge, D., Bishop, T.C., Case, D.A., Cheatham, T. 3rd, Dixit, S., Jayaram, B., Lankas, F., Laughton, C. et al. (2010) A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res.*, **38**, 299–313.
35. Robinson, P.J.J., Fairall, L., Huynh, V.A.T. and Rhodes, D. (2006) Em measurements define the dimensions of the “30-nm” chromatin fiber: evidence for a compact, interdigitated structure. *Proc. Natl Acad. Sci. USA*, **103**, 6506–6511.
36. Angelov, D., Lenouvel, F., Hans, F., Muller, C.W., Bouvet, P., Bednar, J., Moudrianakis, E.N., Cadet, J. and Dimitrov, S. (2004) The histone octamer is invisible when NF-kappa binds to the nucleosome. *J. Biol. Chem.*, **279**, 42374–42382.
37. Cerf, C., Lippens, G., Ramakrishnan, V., Muyltermans, S., Segers, A., Wyns, L., Wodak, S.J. and Hallenga, K. (1994) Homo- and heteronuclear two-dimensional NMR studies of the globular domain of histone H1: full assignment, tertiary structure, and comparison with the globular domain of histone H5. *Biochemistry*, **33**, 11079–11086.
38. Calladine, C.R. and Drew, H.R. (1997) *Understanding DNA: the Molecule and How it Works*, 2nd edn. Academic Press, San Diego.
39. Becker, N.B. and Everaers, R. (2007) From rigid base pairs to semiflexible polymers: coarse-graining DNA. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **76**(2 Pt 1), 021923.
40. Lu, X.-J. and Olson, W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
41. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
42. Haijun, Z. and Zhong-can, O.-Y. (1998) Bending and twisting elasticity: a revised marko-siggia model on DNA chirality. *Phys. Rev. E*, **58**, 4816–4819.
43. Balasubramanian, B., Pogozelski, W.K. and Tullius, T.D. (1998) DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone. *Proc. Natl Acad. Sci. USA*, **95**, 9738–9743.
44. Pastor, N., Weinstein, H., Jamison, E. and Brenowitz, M. (2000) A detailed interpretation of OH radical footprints in a TBP-DNA complex reveals the role of dynamics in the mechanism of sequence-specific binding. *J. Mol. Biol.*, **304**, 55–68.
45. Strahs, D., Barash, D., Qian, X. and Schlick, T. (2003) Sequence-dependent solution structure and motions of 13 tata/TBP (tata-box binding protein) complexes. *Biopolymers*, **69**, 216–243.
46. Sanner, M.F., Olson, A.J. and Spehner, J.C. (1996) Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*, **38**, 305–320.
47. Tsai, J., Taylor, R., Chothia, C. and Gerstein, M. (1999) The packing density in proteins: standard radii and volumes. *J. Mol. Biol.*, **290**, 253–266.
48. Li, G., Levitus, M., Bustamante, C. and Widom, J. (2005) Rapid spontaneous accessibility of nucleosomal DNA. *Nat. Struct. Mol. Biol.*, **12**, 46–53.
49. Montel, F., Fontaine, E., St-Jean, P., Castelnovo, M. and Faivre-Moskalenko, C. (2007) Atomic force microscopy imaging of SWI/SNF action: mapping the nucleosome remodeling and sliding. *Biophys. J.*, **93**, 566–578.
50. Kimball, A., Guo, Q., Lu, M., Cunningham, R.P., Kallenbach, N.R., Seeman, N.C. and Tullius, T.D. (1990) Construction and analysis of parallel and antiparallel holliday junctions. *J. Biol. Chem.*, **265**, 15348.
51. Kulic, I.M. and Schiessel, H. (2004) DNA spools under tension. *Phys. Rev. Lett.*, **92**, 228101.
52. Kruthof, M., Chien, F.-T., Routh, A., Logie, C., Rhodes, D. and van Noort, J. (2009) Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. *Nat. Struct. Mol. Biol.*, **16**, 534–540.
53. Makde, R.D., England, J.R., Yennawar, H.P. and Tan, S. (2010) Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature*, **467**, 562–566.
54. Koopmans, W., Brehm, A., Logie, C., Schmidt, T. and van Noort, J. (2007) Single-pair FRET microscopy reveals mononucleosome dynamics. *J. Fluoresc.*, **17**, 785–795.
55. Shukla, M.S., Syed, S.H., Montel, F., Faivre-Moskalenko, C., Bednar, J., Travers, A., Angelov, D. and Dimitrov, S. (2010) Remosomes: RSC generated non-mobilized particles with approximately 180 bp DNA loosely associated with the histone octamer. *Proc. Natl Acad. Sci. USA*, **107**, 1936–41.
56. Kulic, I.M. and Schiessel, H. (2003) Chromatin dynamics: nucleosomes go mobile through twist defects. *Phys. Rev. Lett.*, **91**, 148103.
57. Kulic, I.M. and Schiessel, H. (2003) Nucleosome repositioning via loop formation. *Biophys. J.*, **84**, 3197–3211.
58. Schiessel, H., Widom, J., Bruinsma, R.F. and Gelbart, W.M. (2001) Polymer reptation and nucleosome repositioning. *Phys. Rev. Lett.*, **86**, 4414–4417.
59. Wong, H., Victor, J.-M. and Mozziconacci, J. (2007) An all-atom model of the chromatin fiber containing linker histones reveals a versatile structure tuned by the nucleosomal repeat length. *PLoS ONE*, **2**, e877.
60. Srinivas Bharath, M.M., Chandra, N.R. and Rao, M.R.S. (2003) Molecular modeling of the chromatosome particle. *Nucleic Acids Res.*, **31**, 4264–4274.
61. Cui, F. and Zhurkin, V.B. (2009) Distinctive sequence patterns in metazoan and yeast nucleosomes: implications for linker histone binding to AT-rich and methylated DNA. *Nucleic Acids Res.*, **37**, 2818–2829.
62. Pachov, G.V., Gabdoulline, R.R. and Wade, R.C. (2011) On the structure and dynamics of the complex of the nucleosome and the linker histone. *Nucleic Acids Res.*, **39**, 5255–5263.
63. Depken, M. and Schiessel, H. (2009) Nucleosome shape dictates chromatin fiber structure. *Biophys. J.*, **96**, 777–84.
64. Dill, K.A. and Chan, H.S. (1997) From Levinthal to pathways to funnels. *Nature Str Mol. Biol.*, **4**, 10–19.
65. Stehr, R., Kepper, N., Rippe, K. and Wedemann, G. (2008) The effect of internucleosomal interaction on folding of the chromatin fiber. *Biophys. J.*, **95**, 3677–3691.