7124–7133 Nucleic Acids Research, 2011, Vol. 39, No. 16 Published online 19 May 2011
doi:10.1093/nar/gkr396

# Cross-linking of DNA through HMGA1 suggests a DNA scaffold

**Benjamin Vogel[1], Anna Löschberger[2], Markus Sauer[2] and Robert Hock[1,*]**

[1]Department of Cell and Developmental Biology and [2]Department of Biotechnology and Biophysics, University of Wuerzburg, Biocenter, Am Hubland, D-97074 Wuerzburg, Germany

Received March 3, 2011; Revised April 27, 2011; Accepted May 3, 2011

## ABSTRACT

**Binding of proteins to DNA is usually considered 1D with one protein bound to one DNA molecule. In principle, proteins with multiple DNA binding domains could also bind to and thereby cross-link different DNA molecules. We have investigated this possibility using high-mobility group A1 (HMGA1) proteins, which are architectural elements of chromatin and are involved in the regulation of multiple DNA-dependent processes. Using direct stochastic optical reconstruction microscopy (*d*STORM), we could show that overexpression of HMGA1a-eGFP in Cos-7 cells leads to chromatin aggregation. To investigate if HMGA1a is directly responsible for this chromatin compaction we developed a DNA cross-linking assay. We were able to show for the first time that HMGA1a can cross-link DNA directly. Detailed analysis using point mutated proteins revealed a novel DNA cross-linking domain. Electron microscopy indicates that HMGA1 proteins are able to create DNA loops and supercoils in linearized DNA confirming the cross-linking ability of HMGA1a. This capacity has profound implications for the spatial organization of DNA in the cell nucleus and suggests cross-linking activities for additional nuclear proteins.**

## INTRODUCTION

High-mobility group A (HMGA) proteins, which are considered as architectural chromatin proteins, belong to a network of dynamic chromatin proteins that transiently reside on DNA/chromatin, fall off and rebind. Alike all HMG proteins they permanently scan the chromatin for available binding sites in a stop and go fashion thus allowing a flexible modulation of chromatin structure and function (1,2). Their binding to chromatin either promotes or suppresses DNA-dependent processes such as transcription, replication and DNA repair. Deregulated expression of HMGA proteins impairs gene expression, affects differentiation processes and causes several diseases (3–7).

HMGA1 proteins play a key role in coordinating the assembly/disassembly of enhanceosomes, which need to be constructed on gene promoters to allow efficient transcription of many genes (4,5,8). Besides the modulation of local chromatin near regulatory DNA elements, HMGA proteins affect chromatin on a more global scale including chromatin loop formation, heterochromatin structure and chromosome condensation (2, 4).

The HMGA1a/b-splice variants and HMGA2-proteins, encoded by a separate gene, are characterized by three AT-hook DNA binding motifs, which preferentially bind to the minor groove of A/T-rich B-form DNA to both naked DNA and DNA on the nucleosomal surface (4,9,10). While a single AT-hook motif would be enough for binding, two AT-hook motifs considerably increase the binding of HMGA proteins to substrates and confer DNA binding with nanomolar affinity (9–14). The three individual AT-hooks do not contribute equally to the strength of binding due to variations in sequence and structure (9–11). An AT-hook consensus peptide comprised of 11 amino acids, and which is virtually identical to the second AT-hook found in HMGA1 proteins is considered as high affinity type 1 AT-hook whereas the type 2 AT-hooks I and III bind with lower affinity (9,10). The affinity of AT-hook II is about one to two orders of magnitude greater than that of AT hook III (10). Amino acids flanking the AT-hooks as well as multiple modifications fine tunes the binding of HMGA proteins (2,9,10,15). Besides their preference for A/T-rich B-DNA, HMGA1 binding is independent of the DNA sequence and also recognizes peculiar structural features such as synthetic four-way junctions, bent DNA and supercoiled DNA (8,12,14,16).

To gain more details about chromatin organization by HMGA1 proteins in the nucleus, we applied super-resolution fluorescence imaging using direct stochastic

*To whom correspondence should be addressed. Tel: +49 931 31 84264; Fax: +49 931 31 84252; Email: rhock@biozentrum.uni-wuerzburg.de

optical reconstruction microscopy (*d*STORM) in fixed cells (17–19). The results suggest that HMGA1 proteins localize in multiple small domains evenly distributed throughout the cell nucleus. Interestingly, in cells overexpressing HMGA1 proteins these domains are not enlarged by accumulation of additional HMGA1 molecules, but concentrated in multiple domains accumulated in densely packed DNA. This observation suggested that HMGA1 protein accumulation and DNA compaction correlate. The data assume that DNA clusters could be formed via local cross-linking of DNA through HMGA1 proteins.

So far, binding of HMGA1 proteins to their substrates has been considered only in a 1D way, meaning that one HMGA1 molecule binds to one DNA molecule containing multiple appropriately spaced A/T-tracts in close proximity using its three DNA binding AT-hooks. To investigate DNA cross-linking activities, we developed a novel DNA-capture assay to investigate this putative function. The results demonstrate that HMGA1 proteins indeed are able to cross-link individual DNA fibers. Experiments with mutated HMGA1 proteins show that AT-hooks together with conserved basic amino acids located between AT-hooks II and III are essential to maintain this activity. Electron microscopy studies revealed that HMGA1 proteins are able to create DNA networks, loops and coils in linearized DNA. Our findings imply that DNA binding proteins are able to cross-link DNA fibers and to create a DNA scaffold.

## MATERIALS AND METHODS

### *d*STORM microscopy

Cos-7 cells were grown in Lab-Tek II chambered coverglass in DMEM with 1% PenStrep, 1 mM L-Glutamine and 10% fetal calf serum (Invitrogen) at 37°C and 5% $CO_2$. Transient transfection of HMGA1a-eGFP was performed using Effectene (Qiagen). Cells were washed in PBS, fixed in 2% formaldehyde/PBS for 10 min, permeabilized for 10 min in 0.1% Triton X-100/PBS, immersed in ice-cold methanol for 3 min. Immunostaining was performed for 45 min using affinity purified rabbit HMGA1 antibodies (IG-1005, 1:150, Immunoglobe, Himmelstadt), mouse DNA antibodies (AC-30-10, 1:50, Progen, Heidelberg) and appropriate Alexa Fluor 647 secondary antibodies (Invitrogen). Cells were washed intensively overnight with excess of PBS, post-fixed with 2% formaldehyde for 5–10 min and washed again with PBS.

For *d*STORM imaging, Lab-Tek wells were filled with switching buffer, which is PBS (pH 7.4), containing oxygen scavenger [0.5 mg/ml glucose oxidase (Sigma), 40 µg/ml catalase (Roche Applied Science), 10% w/v glucose] and 50 mM mercaptoethylamine (Sigma). The *d*STORM setup, which contained an inverted microscope (Olympus IX-71) equipped with an oil-immersion objective (x60, NA 1.45; Olympus) was described previously (17–19). A 640 nm laser (Cube 640–100 C, Coherent) was used for excitation of Alexa 647 and a 488 nm laser (Sapphire 488 LP; Coherent) for eGFP. Photoswitching of Alexa Fluor 647 was performed with 0.05–0.5 $kW/cm^2$. A total of 10 000–12 000 frames were measured with a frame rate of 60–100 Hz. The fluorescence light was collected by the objective, filtered by appropriate bandpass filters (Chroma ET700/75, HQ 535/50 AHF analysentechnik AG), and imaged on an electron-multiplying charge-coupled device (EMCCD) camera (Ixon DV897; Andor). The images were analyzed and reconstructed with Rapid2STORM software (open source) as described by Wolter *et al.* (28). The fluorescence signals were analyzed applying asymmetry and photon thresholds, i.e. spots appearing too wide or too elliptical, as well as spots with less than 2000 or more than 6000 photons (HMGA1), or 300–2000 photons (DNA) per spot and frame were discarded from further analysis. It should be pointed out that the thresholds applied have no influence on the size of domains determined from reconstructed super-resolution images because each fluorophore is localized several times.

### Preparation of labeled DNA and cloning of constructs

A major satellite repeat (satDNA) was amplified from murine genomic DNA by PCR using phusion polymerase (Finnzymes), cloned in pDrive (Qiagen) and sequenced.

Terminally labeled satDNA with one molecule Cy5 or biotin (Cy5-satDNA or bio-satDNA, 563 bp) was produced by PCR using Taq polymerase (NEB) and 5′-Cy5 or 5′-biotinylated primers. Dual labeled Cy5/bio-satDNA was used to determine the binding efficiency to streptavidin. PCR products were quantified using a NanoQuant plate and a Tecan infinite M200 reader.

Human HMGA1a, HMGA1b, single- (R28G, R60G, R86G), double- (R28G R60G, R28G R86G, R60G R86G) or triple-mutated (R3xG) constructs were originally produced by Harrer *et al.* (1).

Truncated HMGA1a lacking the C-terminal 17 (ΔL90), 25 (ΔK82), 34 (ΔK71) or 43 (ΔG63) amino acids were produced by using appropriate primers. Point mutations K67G or RK73/74GG or K67G + RK73/74GG were created by site-directed mutagenesis. HMGN1 or HMGN2 were cloned from murine cDNAs, respectively.

All products were subcloned in pJet 2.1/blunt (Fermentas) cut with EcoRI and finally cloned in pET28a(+) (Novagen) and sequenced. Primers and sequences can be found in the Supplementary Data.

### Protein preparation

HMGA1 or HMGN proteins were expressed using Rosetta strain (Novagen) and induced with 0.5 mM IPTG at OD600 = 0.5. Extraction, purification and quantification of HMGA1 proteins were performed as described by Reeves and Nissen (29). HMG proteins were quantified using a Jasco V-650 spectrophotometer. For relative quantification of HMG proteins, a Tecan infinite M200 reader and extinction at 230 nm was used. Protein purity was controlled in SDS–PAGE and Coomassie staining.

### DNA cross-linking assay

Assays were performed in black 96-well plates pre-coated with streptavidin (Greiner bio-one). All incubations were at room temperature. Wells were pre-incubated with 150 µl PBS for 10 min. Supernatants were carefully

removed by aspiration with a suction pump and a gel-loader tip. 500 ng of bio-satDNA were resuspended in 150 µl high-salt buffer (2 M NaCl, 10 mM Tris–HCl, pH 7.5, 1 mM EDTA) and incubated for 1 h. Wells were washed with 150 µl high-salt buffer and three times with freshly prepared $PBST^+$ (PBS, 0.1% Triton X-100, 0.3 M sucrose). Ten microgram of HMG protein was incubated in 150 µl $PBST^+$ on a roller shaker for 30 min and 500 ng of Cy5-satDNA was added. The Cy5-satDNA/HMG solution was added to the washed wells and incubated overnight. After measuring the input Cy5 fluorescence, wells were washed seven times with 150 µl $PBST^+$ and remaining Cy5 fluorescence was measured again. Measurements were performed using a Tecan infinite M200 reader. Excitation for Cy5-satDNA was set to 641 nm and emission to 670 nm using a fixed gain. Each measurement was performed four times using five flashes. Relative fluorescence was calculated from mean absolute values (input Cy5 fluorescence/remaining fluorescence ×100%) and was defined as relative cross-linking efficiency (RCE). For statistical purposes, each experiment was at least reproduced four times. Statistical relevance was determined using a student's *t*-test. Control A (Ctrl A) was performed incubating only Cy5-satDNA, control B (Ctrl B) was performed using Cy5-satDNA in wells containing precoupled bio-satDNA without any protein and control C (Ctrl C) was performed incubating Cy5-satDNA/protein in wells without precoupled bio-satDNA. Proteinase K treatment was performed after HMGA1a cross-linking in $PBST^+$ by adding 50 ng/µl proteinase K for 6 h.

### Electron microscopy

For DNA spreading experiments a pDrive vector containing the major satellite DNA was linearized with HindIII and gel extracted (Peqlab). Incubations were at room temperature. One hundred and fifty nanograms DNA and 10 µg HMGA1a or R3xG were resuspended in 100 µl incubation buffer (10 mM Tris–HCl, pH 8.0, 1 mM EDTA, 25 mM NaCl) and incubated for 20 min. DNA staining for electron microscopy was principally made as described by Bock and Zentgraf (30). Two hundred microliters 0.2 M freshly prepared $NH_4OAc$ were mixed with 0.8 µl 1% cytochrome C in $H_2O$ and added to the DNA/protein solution. Resulting solution (100 µl) was incubated on parafilm for further 20 min. The drop was slightly touched with a parlodione-coated 300 mesh copper grid and then directly contrasted by immersion in 90% EtOH with 50 µM uranylacetate for 30 s and 90% EtOH for 30 s. The grid was air-dried on filter paper. Grids were rotary-shadowed using a Pt-Pd wire with an angle of 5–7° for contrast enhancement. HMGA1a/DNA samples were treated in the same way, but for higher complex count on the grid the DNA/protein solution was directly added to the grid for 30 s to 20 min, then prepared as described above. Grids were analyzed with a magnification of 8000–10 000 using a Zeiss EM10 equipped with a sheet film camera and Kodak SO163 electron image films.

## RESULTS

### HMGA1 proteins localize in defined domains, which correlate with DNA compaction

HMGA1 proteins preferentially localize to the more densely packed heterochromatin. We observed that transiently transfected Cos-7 cells display compaction of DNA as a consequence of HMGA1a-eGFP overexpression. To analyze the nature of chromatin clustering with higher resolution, we used *d*STORM. *d*STORM relies on the use of organic fluorophores and does not require the close proximity of a second activator fluorophore (20) thus enabling the use of commercially available standard fluorescent probes in fixed (17) and living cells (18,21). For *d*STORM microscopy, Cos-7 cells overexpressing HMGA1a-eGFP were immunostained with Alexa Fluor 647 labeled antibodies. Cells for *d*STORM were identified by morphological criteria using wide field fluorescence microscopy of eGFP with low HMGA1a-eGFP expression showing a diffuse nuclear fluorescence (Figure 1a and c) and cells with high HMGA1a-eGFP expression displaying distinct chromatin clusters (Figure 1b and d). In addition, wide field fluorescence images of the corresponding immunolabeled cells were taken prior to *d*STORM experiments (Figure 1a′–d′). Typically, we detected more than 1000 photons per switching cycle from Alexa Fluor 647 fluorophores allowing a localization precision of ∼20 nm in the imaging plane, experimentally determined from repetitive localizations of individual Alexa Fluor 647 labeled antibodies (17). The *d*STORM images (Figure 1a″–d″ and a‴–d‴) show a substantially improved resolution over the widefield fluorescence images and enable the identification of distinct HMGA1 domains with diameters of 50–100 nm (Supplementary Figure S1A). Furthermore, the *d*STORM images show that distances between adjacent HMGA1 domains of ∼100 nm can be easily resolved (Supplementary Figure S1A). When HMGA1a-eGFP is expressed at low levels, these HMGA1 domains are evenly distributed in the nucleus (Figure 1a″–a‴). This pattern can also be observed comparable to that of the endogenous HMGA1 in non-transfected cells (Supplementary Figure S1B). In widefield fluorescence pictures the distribution of the overexpressed HMGA1a-eGFP is more diffuse (Figure 1a) as compared to endogenous HMGA1 [Supplementary Figure 1SB(a′)].

In contrast, HMGA1a-eGFP overexpression leads to accumulation of HMGA1 domains in chromatin clusters (Figure 1b″–b‴). The increase in fluorescence intensity supports the idea of chromatin cluster formation in HMGA1a-eGFP overexpressing cells (Supplementary Figure S1C). *d*STORM experiments of immunostained DNA in cells with either low HMGA1a-eGFP expression (Figure 1c″–c‴) or high HMGA1a-eGFP expression (Figure 1d″–d‴) corroborate that HMGA1-containing clusters consist of compacted DNA. Chromatin clustering through HMGA1a-eGFP overexpression is not a specificity of the Cos-7 cell line and was also observed in the human cell line HepG2 (Supplementary Figure 1D). This finding raises the question how this DNA compaction can be achieved by HMGA1a mechanistically.

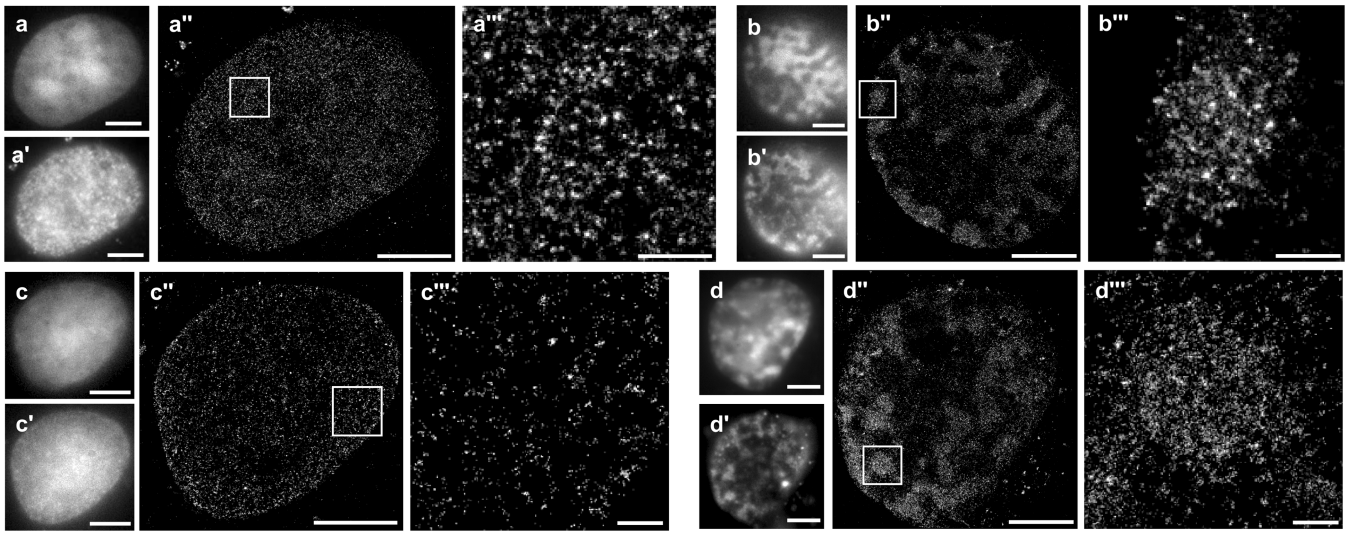


**Figure 1.** *d*STORM pictures acquired from Cos-7 cells transfected with HMGA1a-eGFP. Widefield microscopy was used to select cells with either low HMGA1a-eGFP expression (no chromatin clustering; a and c) or high HMGA1a-eGFP expression (chromatin clustering; b and d). Widefield fluorescence images of immunostainings using HMGA1 and DNA antibodies are shown in a′ and b′ (HMGA1) and c′ and d′ (DNA). Note that in a′ both endogenous HMGA1 and the fusion protein are detected by the immunostaining. Corresponding *d*STORM images are shown in a″–d″. Highlighted boxes are shown in a‴–d‴. Scale bars are 5 μm in a–d″ or 500 nm in a‴–d‴. Note the localization of HMGA1 in defined domains.

### DNA–DNA cross-linking activity of HMGA1 proteins

The presence of three AT-hook binding domains and the capacity of a multivalent recognition of binding sites suggest that HMGA1 proteins can bind to A/T-tracts not only on the same, but also on different DNA fibers. Therefore, we developed a novel assay to investigate this assumption. The experimental strategy is outlined in Figure 2. For this assay, a PCR product of 563 bp containing A/T-rich mouse major satellite sequences (satDNA) was used. In the experiments, we immobilized a terminally biotinylated satDNA (bio-satDNA) on the surface of streptavidin-coated 96-well plates. Of 500 ng bio-satDNA, 50–55% were routinely bound to the bottom of the wells. To such pre-treated wells a Cy5-end-labeled satDNA (Cy5-satDNA) and 10 μg purified recombinant HMG proteins were added. Coomassie staining of proteins after SDS–PAGE is shown in Supplementary Figure S2. After incubation the input Cy5 fluorescence was measured followed by measuring the remaining Cy5 fluorescence after several washing steps. The ratio of remaining Cy5 fluorescence over input was defined as RCE. In the presence of HMGA1 proteins and precoupled bio-satDNA, the RCE was either 20.4 ± 3.2% for HMGA1a or 28.6 ± 4.2% for HMGA1b (Figure 3A and E). To rule out non-specific binding of Cy5-satDNA control experiments were performed. To this end, Cy5-satDNA was either incubated in wells without bio-satDNA (Ctrl A) or with bio-satDNA (Ctrl B), but containing no proteins. The RCEs for the controls were 0.6 ± 0.3% for Ctrl A and 0.7 ± 0.3% for Ctrl B (Figure 3A). In further controls incubating Cy5-satDNA with either HMGA1a or HMGA1b in wells without immobilized bio-satDNA we obtained RCEs of 2.7 ± 1.7% (A1a—Ctrl C) and 2.8 ± 1.8% (A1b—Ctrl C), respectively. These controls (protein—Ctrl C) were routinely run in parallel for each further experiment. Together these results indicate that HMGA1 proteins capture Cy5-satDNA through cross-linking to immoblized bio-satDNA.

Significantly, Cy5 fluorescence of captured DNA was lost up to 98% after proteinase K treatment, supporting that cross-linking requires streptavidin-coupled bio-satDNA and HMGA1a proteins. Furthermore, we also used HMGN1 or HMGN2 proteins, which preferentially bind to nucleosomes. In these experiments, the RCEs were also at background levels either at 2.0 ± 0.1% for HMGN1 (N1—Ctrl C, 0.9 ± 0.3%) or 1.5 ± 0.2% for HMGN2 (N2—Ctrl C, 0.7 ± 0.1%) (Figure 3A). These results corroborate that the cross-linking capacity is HMGA1 specific. Of note, the RCE of HMGA1b was significantly higher as compared to HMGA1a ($P < 0.0001$). The HMGA1a- and HMGA1b-splice variants differ in 11 hydrophobic amino acids between AT-hooks I and II. This implies that either the hydrophobicity of these 11 amino acids or the increased distance between AT-hooks I and II reduces the RCE of HMGA1a.

### The role of the AT-hook motifs in DNA cross-linking

Owing to the fact that HMGA1 proteins possess 3 AT-hook binding domains, we investigated how the individual AT-hook motifs I–III contribute to cross-linking. We, therefore, used the cross-linking assay applying point mutated proteins, where the second arginines in the minor groove binding core consensus R-G-R of the AT-hooks were replaced by glycines (1). Recombinant proteins used in the assay were point mutated in either AT-hook I (R28G), AT-hook II (R60G) or AT-hook III (R86G), double-point mutated in AT-hooks I + II (R28G R60G), I + III (R28G R86G), II + III (R60G R86G) or

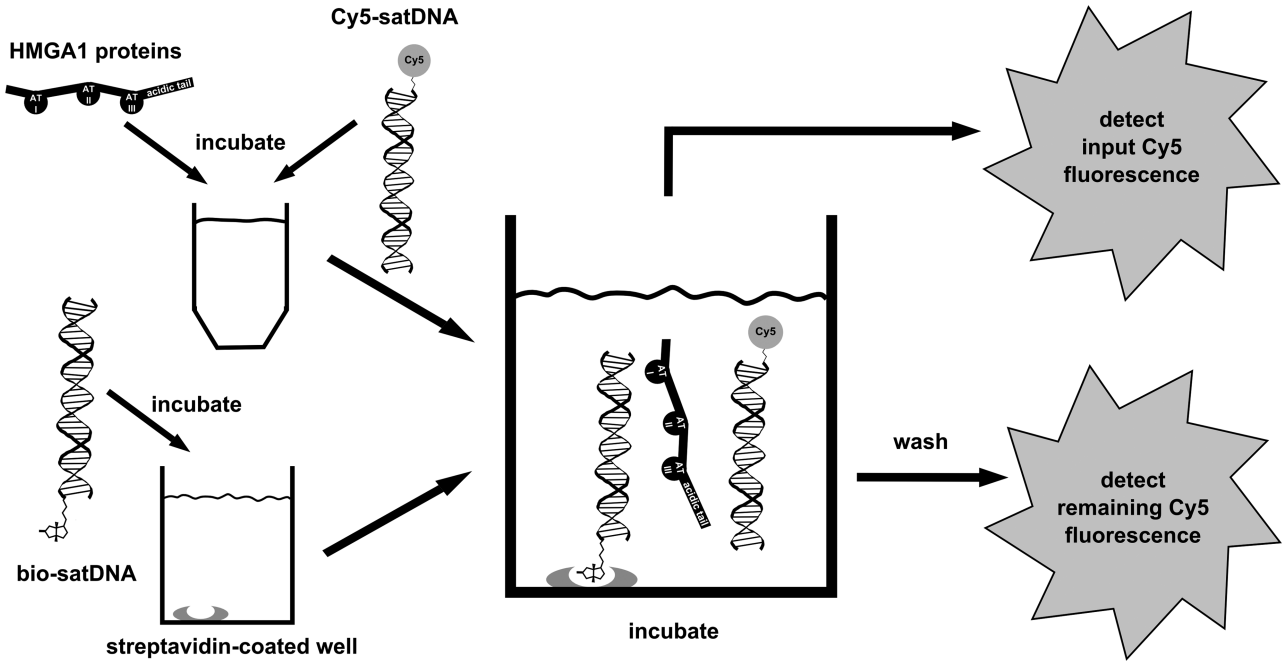


**Figure 2.** Principle of the DNA cross-linking assay. Streptavidin pre-coated wells are loaded with bio-satDNA. A protein/Cy5-satDNA solution is incubated. Input Cy5 fluorescence and, after intensive washing steps, remaining Cy5 fluorescence is detected. The ratio of nput / remaining Cy5 fluorescence $\times 100\%$ was defined as RCE.

triple-point mutated in AT-hook I + II + III (R3xG) (Figures 3B and 4A). All RCEs obtained for these proteins are summarized in Figure 3E. Compared to wild-type HMGA1a, point mutations of one AT-hook revealed significantly reduced cross-linking when either AT-hook I (R28G: 15.7 ± 1.6%, $P = 0.0002$), AT-hook II (R60G: 14.4 ± 4.3%, $P < 0.0001$) or AT-hook III (R86G: 15.0 ± 4.6%, $P = 0.0004$) was mutated (Figure 3B). This clearly shows that binding of all three AT-hooks is required for efficient cross-linking.

Double mutations in two out of three AT-hooks or mutation in all three AT-hooks further significantly diminished the RCE as compared to wild-type HMGA1a ($P < 0.0001$ for all mutants). As compared to the single AT-hook mutants the cross-linking efficiencies were further diminished in the double mutants ($P < 0.0001$ for I + II, $P = 0.0056$ for I + III and $P = 0.0012$ for II + III) or the triple mutant ($P < 0.0001$) (Figure 3B and E). Thus, the presence of more functional AT-hooks translates into an increased cross-linking efficiency.

Cross-linking was affected mostly, when AT-hooks I and II were double mutated and the RCEs decreased to those of the triple mutant (Figure 3B and E). The effect was diminished when an AT-hook III mutation was part of the double mutant protein. The results support that AT-hook I and/or AT-hook II play a more important role than AT-hook III. This is consistent with data, where AT-hooks I and II were found to be the major players in DNA binding *in vitro* and *in vivo* (1,11,22). However, these data also indicate that at least one AT-hook is required to capture DNA independently of the considered AT-hook motif. Moreover, the data also indicate that a minor, but residual cross-linking capacity exists, even when all three AT-hooks are point mutated. This is in accordance with previous mobility studies in living cells, where the triple-point mutated protein also possesses a weak, but evident residence time on chromatin (1). Cross-linking assumes multivalent binding of at least two motifs on either DNA molecule. The results of the R28G, R60G and R86G mutants indicate that two functional AT-hooks are capable of cross-linking DNA. However, the relative high RCE values of the proteins containing two mutated AT-hooks suggest that one AT-hook plus additional sequences outside the AT-hooks could be sufficient to convey this function.

### Conserved basic amino acids located between AT-hooks II and III are required for efficient DNA cross-linking

To seek for additional sequences we applied C-terminal truncated versions of HMGA1a in the cross-linking assay (Figures 3C and 4A). HMGA1a- or HMGA1b-proteins lacking the negatively charged C-terminal domain were found to bind A/T-rich DNA ~1.4-fold better than the native protein and such truncations were estimated to be 8–10-fold more effective at introducing negative supercoils in plasmid DNA (12). Consistent with these findings, removal of the acidic 17 amino acids from the C-terminus of HMGA1a (ΔL90) significantly increased the RCE ~1.6-fold (37.9 ± 3.0%, $P < 0.0001$) (Figure 3C and E). Removal of 24 amino acids including AT-hook III (ΔK82) resulted in a cross-linking efficiency of 23.5 ± 7.7% comparable to the full-length HMGA1a. Significantly, a further HMGA1a truncation (ΔK71) almost lost the cross-linking capacity and the RCE

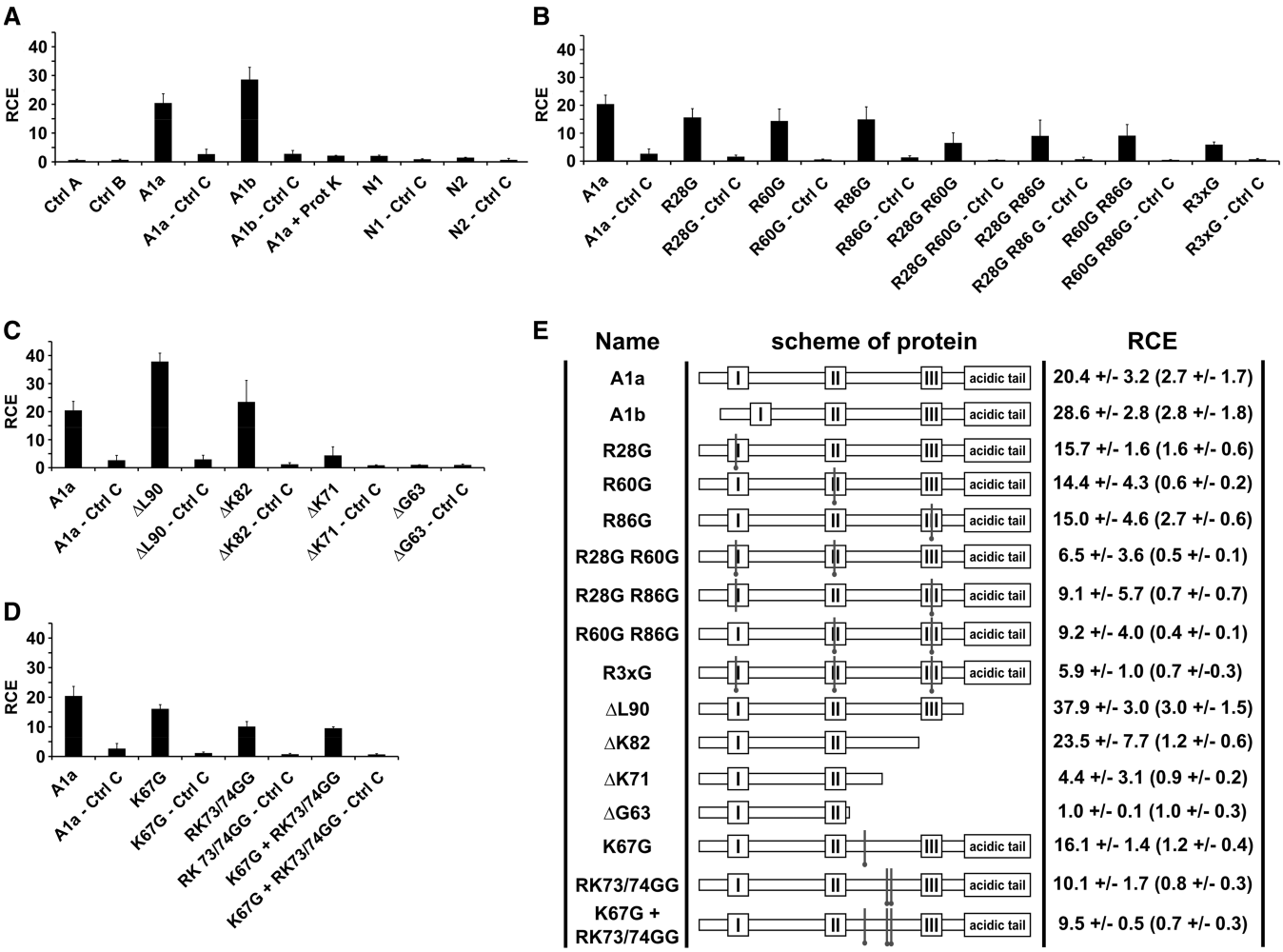


**Figure 3.** Results of DNA cross-linking assays. Bar charts of cross-linking assays showing RCE in (**A–D**). Standard deviations are included. (A) Specific cross-linking of DNA through HMGA1 proteins (A1a, A1b). Controls to exclude non-specific binding are shown with Cy5-satDNA in wells without precoupled bio-satDNA (Ctrl A), in wells containing precoupled bio-satDNA (Ctrl B) or without precoupled bio-satDNA incubated with protein and Cy5-satDNA (protein—Ctrl C). RCEs of HMG proteins N1 and N2 reveal no DNA cross-linking (N1—Ctrl C, N2 Ctrl C). Proteinase K treatment (A1a + Prot K) disrupts DNA cross-linking to 98%. (B) Cross-linking efficiencies of HMGA1a-AT-hook mutants. Compared to HMGA1a (A1a) single point mutations of one AT-hook (R28G; R60G; R86G) show significantly reduced RCEs. Double-mutants of two AT-hooks (R28G R60G, R28 R86G, R60G R86G) cross-link DNA with further diminished RCEs. Note that R28G R60G mutant has the lowest RCE. Triple-point mutation of all three AT-hooks (R3xG) again shows low RCEs. Corresponding controls without precoupled bio-satDNA incubated with protein and Cy5-satDNA (protein—Ctrl C) are given. (C) Cross-linking efficiencies of C-terminal truncated HMGA1a proteins suggest the requirement of a cross-linking domain between AT-hooks II and III. Deletion of the acidic tail (ΔL90) of HMGA1a leads to an increased RCE compared to the wild-type protein (A1a). Further deletion of A1a including the third AT-hook (ΔK82) gives RCEs again to wild-type levels (A1a). Further truncation decreases the RCE dramatically (ΔK71). Further truncation up to the second AT-hook (ΔG63) diminishes the RCE to background level. Corresponding controls without precoupled bio-satDNA incubated with protein and Cy5-satDNA (protein—Ctrl C) are given. (D) Point mutations in conserved amino acids between the second and the third AT-hook reduce DNA cross-linking. The point mutation K67G leads to a slightly decreased RCE compared to wild-type protein. Compared to wild-type protein the RCE of the RK73/74GG double mutant drops 50%. A triple-point mutated protein K67G + RK73/74GG shows a similar RCE as the double-point mutant RK73/74GG alone. Corresponding controls without precoupled bio-satDNA incubated with protein and Cy5-satDNA (protein—Ctrl C) were at background levels. This shows that the amino-acids R73 and K74 are essential for efficient DNA cross-linking. (**E**) Schemes of HMGA1 proteins and summary of RCE. Sites of point mutations are indicated. RCE values and SDs are listed accordingly. RCEs of corresponding controls without bio-satDNA, but incubated with protein and Cy5-satDNA (protein—Ctrl C) are given in brackets.

measured was 4.4 ± 3.1%. When applying a further truncated HMGA1a protein (ΔG63) comprised only of the first and second AT-hook to the assay, the cross-linking activity was completely lost and dropped to background levels (1.0 ± 0.08%). Thus, albeit functional AT-hooks are required to convey full cross-linking efficiency, even two AT-hooks are not sufficient when the sequence element between AT-hooks II and AT-hook III is missing. This indicates that the sequence element located between AT-hooks II and III is essential for DNA cross-linking.

Within this region spanning the amino acids 64–82 the basic amino acids K64, K67, K70, R73 and K74 are highly conserved (Figure 4B). We, therefore, created the point

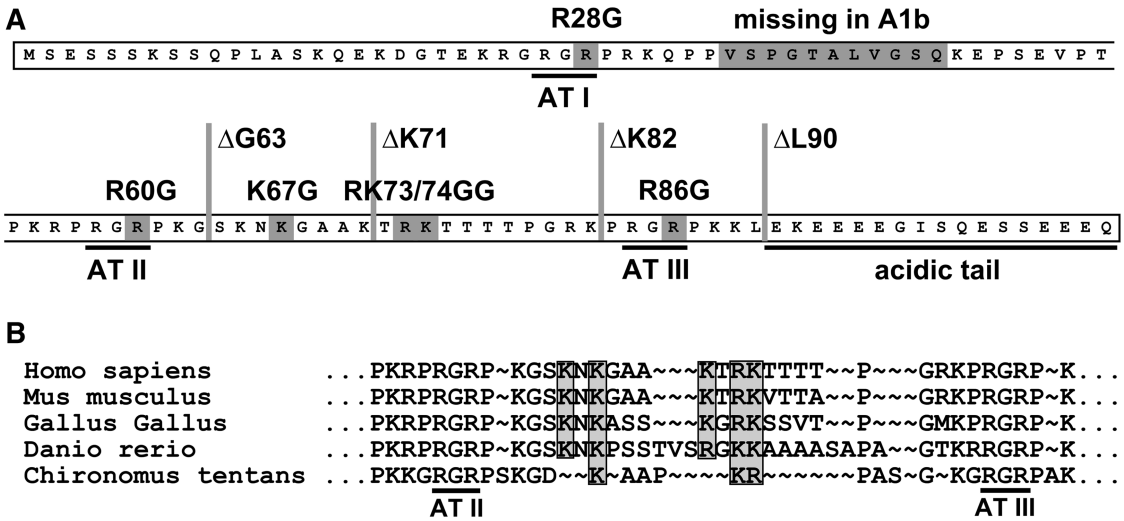


**Figure 4.** (**A**) HMGA1 mutations used indicated on the level of the amino acid sequence. (**B**) Species alignment of HMGA1 amino acid sequences of the region between AT-hooks II and III. Conserved amino acids are indicated (grey boxes). Compared are sequences from human (homo sapiens, CAI14992), mouse (mus musculus, NP_057869), chicken (gallus gallus, NP_989700), zebra fish (danio rerio, NP_998333) and midge (chironomus tentans, CAA85365).

mutated proteins HMGA1a K67G and HMGA1a RK73/74GG where we replaced the indicated basic amino acids by glycines (Figure 4A). Of note, both mutated proteins lost the full capacity to cross-link DNA. The RCEs significantly dropped either to 16.7 ± 1.4% (K67G) or to 10.1 ± 1.7% in the RK73/74GG mutant (Figure 3D and E). This suggests that K67 plays a minor role whereas R73 and K74 are major players in DNA cross-linking. This is consistent with the cross-linking capacity of a K67G + RK73/74GG triple-mutated protein of 9.5 ± 0.5% and compares to that of the RK73/74GG double-mutated protein. Together the results propose that the sequence element between amino acids 65–74, with R73 and K74 as an essential core is required for efficient DNA cross-linking by HMGA1a.

### HMGA1a creates DNA networks

The capacity to cross-link DNA was also visualized at ultra-structural level using electron microscopy. To this end, 150 ng of a linearized plasmid containing satDNA was contrasted with cytochrome C to visualize the DNA through the electron dense iron ion, which is part of the cytochrome C molecule. Spread preparations on parlodion coated copper grids revealed evenly distributed distinct single DNA molecules of 1145.7 ± 117.5 nm length in every sector of the grids [Figure 5A(a)]. Pre-incubation of 150 ng DNA with 10 μg HMGA1a prior to cytochrome C treatment and spreading resulted in the creation of few DNA network patches containing loops and coils [Figure 5A(b–f)]. In the absence of proteins, DNA crossings were found randomly in ~9% and loops randomly in ~5% of counted DNA molecules ($n = 361$). Significantly, in the presence of HMGA1a proteins the number of DNA crossings was increased to ~85% and loop formation was observed in 35% of counted molecules ($n = 92$).

This supports that HMGA1a proteins are able to cross-link DNA intra- and intermolecularly and are able to induce loop and coil formation [Figure 5A(c–f)]. Notably, an evenly distributed DNA network was also created after pre-incubation with the mutant HMGA1a bearing the point mutation in each AT-hook (R3xG) [Figure 5A(g)]. Thus, weak binding of AT-hooks still enables DNA cross-linking. This is in accordance with the findings obtained with the cross-linking assay presented above. However, loop formation was not observed indicating that creation of loops requires both, AT-hooks and the cross-linking domain of HMGA1a.

## DISCUSSION

*d*STORM suggested that HMGA1 proteins localize concentrated in defined domains with a maximal diameter of 50–100 nm. The localization in defined domains is consistent with previous findings that HMGA proteins localize to A/T-rich G/Q- and C-bands of human and mouse metaphase chromosomes (23). Together with our findings, this implies that the localization of HMGA proteins in defined chromosomal regions persists throughout the cell cycle. Nevertheless, in cells highly overexpressing HMGA1a, the DNA is compacted in clusters, which contain a crowd of spatially compacted HMGA1-containing dots. Likewise, calculation of the labeling intensities revealed that HMGA1 molecules accumulate in defined domains within clusters of tightly concentrated DNA. We, therefore, asked whether HMGA1 proteins could be responsible for DNA compaction and assumed that HMGA1 proteins could mediate or stabilize a compacted DNA scaffold through cross-linking of DNA.

Using a specifically developed cross-linking assay we confirmed this assumption. This assay revealed that efficient cross-linking requires at least some low affinity DNA

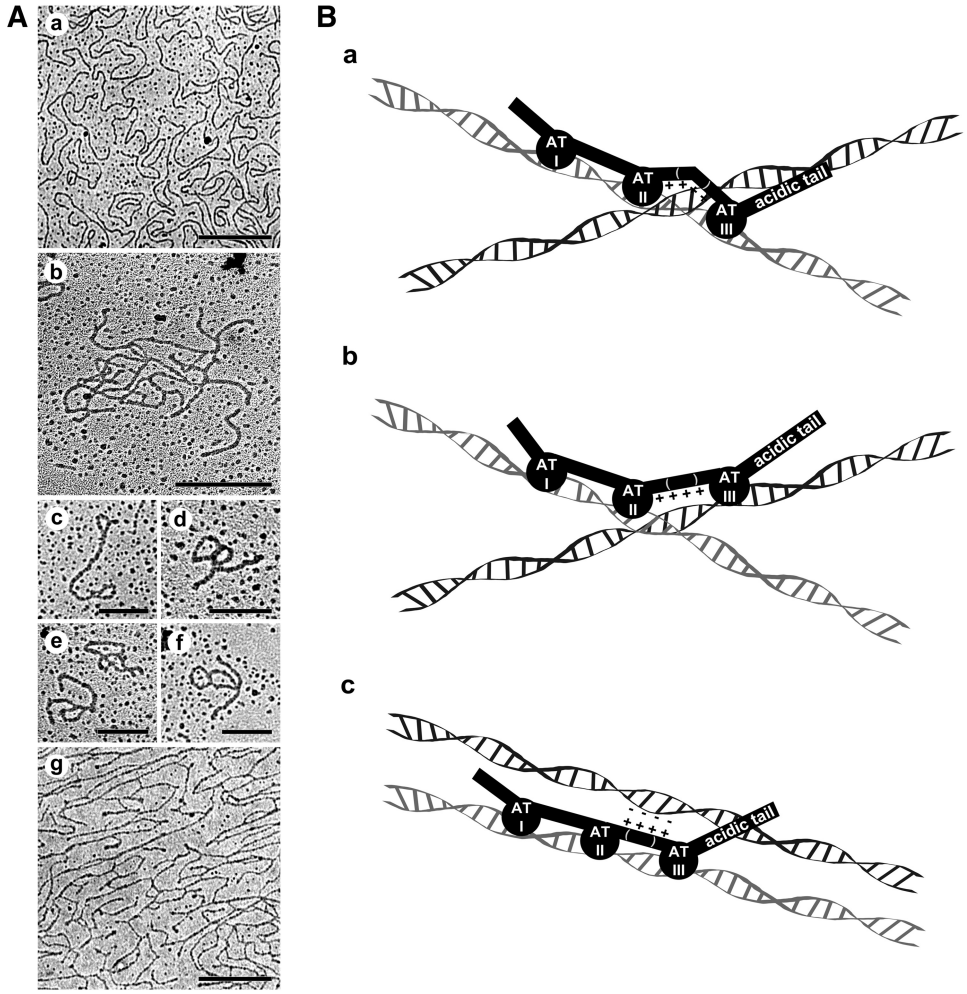


**Figure 5.** (**A**) Electron microscopic analysis of HMGA1–DNA interaction using DNA spread preparations. Naked DNA is shown in a. DNA incubated with HMGA1a is presented in pictures b–f. Note the DNA cross-linking in b and loops and coils in c–f. DNA incubated with the R3xG AT-hook triple mutant is shown in g. Low DNA binding and an intact region between AT-hooks II and III are sufficient to create a DNA network. Bars in a, b and g are 500 nm. Bars in c–f are 250 nm. (**B**) Models of DNA cross-linking through HMGA1. Either a clamp- or a push-button model is conceivable. In the clamp-model HMGA1 molecules bind to DNA with their AT-hooks, whereas the cross-linking domain-containing R73 and K74 overstretches an adjacent DNA fiber (a and b). The third AT-hook either contacts the same DNA molecule than AT hooks I and II (a) or contacts the second DNA molecule (b). In the push-button model, the cross-linking domain contacts a neighboring DNA fiber (c).

binding either of one functional AT-hook or three weakly binding mutated AT-hooks. In addition, an essential sequence element between AT-hooks II and AT-hook III containing the conserved basic amino acids R73 and K74 is crucial. We, therefore, suggest that this region plays an important role as cross-linking domain of HMGA1 proteins. Thus, in the wild-type protein both the AT-hooks and the cross-linking domain act together to convey full DNA cross-linking activity. The cross-linking activity is intrinsic and does not require energy. Our finding of this novel function of HMGA1 proteins is in full agreement with previous findings that HMGA proteins recognize and bind to non-B-form DNA including four-way junction DNA (14), bent and supercoiled DNAs (12) and distorted DNA on isolated nucleosomal core particles (16).

Soluble HMGA molecules are unstructured disordered random coils and only adopt a defined conformation when bound to DNA or chromatin assuming a kind of flexibility in selecting their binding substrates (10). Such flexibility has to be assumed for any cross-linking activity. Potential models of how HMGA1 proteins could cross-link adjacent DNA fibers are summarized in Figure 5B. Combining our results and previous findings either a clamp- or a push-button model are likely. In the clamp-model HMGA molecules bind to DNA with their AT-hooks, whereas the cross-linking domain containing R73 and K74 overstretches an adjacent DNA fiber [Figure 5B(a–b)]. In the push-button model, the cross-linking domain could loop out and contact a neighboring DNA fiber [Figure 5B(c)]. When unraveling the solution structure of AT-hooks bound to the minor groove of DNA, it previously was found that only the RGR core peptide fits deep into the minor groove (10). Six amino acids C-terminal of the core, including K65 and K67, only present C-terminal to AT-hook II, were found

to interact with the sugar–phosphate backbone on either edge of the minor groove (10). It is conceivable that in addition the basic amino acids K71, R73 and K74 similarly contact the sugar–phosphate backbone to capture adjacent DNA molecules. This would be consistent with both, a clamp- or a push-button model.

However, several previous findings rather support the clamp model. Binding of HMGA1 to four-way junction (4H) DNA occurs via contact of two out of three AT-hooks on opposite arms of the branch point and contacts across the 4H DNA branch point (14). A subfragment containing AT-hooks II and III plus the intervening peptide binds to 4H DNA with about the same affinity as the full-length protein (14). Thus the contacts over the branch point occur within the sequence between AT-hooks II and AT-hook III, which we also found to be relevant for the cross-linking activity. In addition, findings that HMGA1 proteins are able to introduce supercoils (DNA crossings) in plasmid DNA *in vitro* without requiring energy (4,12) are consistent with a clamp like model where DNA crossings, as occurring in supercoils, are stabilized. In support, supercoil-like DNA crossings were also observed in our electron microscopy studies using linearized DNA.

The capacity of DNA cross-linking has implications for several architectural functions that have been described for HMGA proteins. For example, together with histone H1 and topoisomerases II HMGA1 co-localizes at scaffold attachment regions (SARs). SARs are thought to be A/T-rich DNA elements at the base of large gene-containing DNA loops and considered as structural backbone of metaphase chromosomes (24). Interestingly, we found loop formation in linearized plasmids in our electron microscopy studies, which is in full agreement with these previous studies and with a localization of HMGA1 at the base of DNA loops.

Biochemical results demonstrated a multivalent DNA binding to the IFN-γ promoter (13) and HMGA1 is a key player in coordinating enhanceosome organization during the on/off regulation of interferon transcription [for a summary see (4)]. Interestingly, acetylation of K65 destabilizes the enhanceosome, whereas acetylation at K71 enhances IFN-γ transcription (25). Moreover, it was recently described that the histone acetyl transferases PCAF and p300 also acetylate K67 and K74 of HMGA1a (26). Since, especially these lysines are involved in DNA cross-linking it is conceivable that DNA cross-linking and its regulation through acetylation could be involved in stabilization/destabilization of the enhanceosome organization. Within this context, it is noteworthy that these lysines are hyperacetylated in PC-3 human prostate cancer cells and increased HMGA1a acetylation was observed in metastatic breast cancer cells (27). A hallmark of malignant tumors is HMGA1 overexpression. Therefore, it is intriguing to speculate that HMGA1 induced DNA cross-linking and a modulation through acetylation may influence DNA organization in tumor cells.

HMGA1 proteins are involved in the organization of heterochromatin, influence global chromatin structure during apoptosis or cellular senescence and are involved in chromosome architecture and compaction. It is conceivable that the spatial organization of a DNA scaffold through the cross-linking capacity of HMGA1 proteins plays a significant role in these processes. Moreover, besides HMGA1 proteins additional chromatin proteins could possess a similar intrinsic DNA cross-linking activity using identical or different domains.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Sebastian van de Linde and Steve Wolter for their assistance in *d*STORM experiments and data analysis. We thank the department of electron microscopy for technical assistance.

## FUNDING

This work was supported by a grant of the DFG (Deutsche Forschungsgemeinschaft) to RH. MS acknowledges funding by the Biophotonic Initiative (grant Ho1804/5-1) of the BMBF (Bundesministerium für Bildung und Forschung, grant 13N11019). Funding for open access charge: DFG (Deutsche Forschungsgemeinschaft) and the University of Wuerzburg.

*Conflict of interest statement*. None declared.

## REFERENCES

1. Harrer,M., Lührs,H., Bustin,M., Scheer,U. and Hock,R. (2004) Dynamic interaction of HMGA1a proteins with chromatin. *J. Cell. Sci.*, **117**, 3459–3471.
2. Catez,F. and Hock,R. (2010) Binding and interplay of HMG proteins on chromatin: Lessons from live cell imaging. *Biochim. Biophys. Acta*, **1799**, 15–27.
3. Hock,R., Furusawa,T., Ueda,T. and Bustin,M. (2007) HMG chromosomal proteins in development and disease. *Trends Cell. Biol.*, **17**, 72–79.
4. Reeves,R. (2010) Nuclear functions of the HMG proteins. *Biochim. Biophys. Acta*, **1799**, 3–14.
5. Sgarra,R., Zammitti,S., Lo Sardo,A., Maurizio,E., Arnoldo,L., Pegoraro,S., Giancotti,V. and Manfioletti,G. (2010) HMGA molecular network: From transcriptional regulation to chromatin remodeling. *Biochim. Biophys. Acta*, **1799**, 37–47.
6. Fedele,M. and Fusco,A. (2010) HMGA and cancer. *Biochim. Biophys. Acta*, **1799**, 48–54.
7. Brocher,J., Vogel,B. and Hock,R. (2010) HMGA1 down-regulation is crucial for chromatin composition and a gene expression profile permitting myogenic differentiation. *BMC Cell. Biol.*, **11**, 64.
8. Reeves,R. and Beckerbauer,L. (2001) HMGI/Y proteins: flexible regulators of transcription and chromatin structure. *Biochim. Biophys. Acta*, **1519**, 13–29.
9. Reeves,R. and Nissen,M.S. (1990) The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. *J. Biol. Chem.*, **265**, 8573–8582.
10. Huth,J.R., Bewley,C.A., Nissen,M.S., Evans,J.N., Reeves,R., Gronenborn,A.M. and Clore,G.M. (1997) The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat. Struct. Biol.*, **4**, 657–665.

11. Claus,P., Schulze,E. and Wiśniewski,J.R. (1994) Insect proteins homologous to mammalian high mobility group proteins I/Y (HMG I/Y). Characterization and binding to linear and four-way junction DNA. *J. Biol. Chem.*, **269**, 33042–33048.
12. Nissen,M.S. and Reeves,R. (1995) Changes in superhelicity are introduced into closed circular DNA by binding of high mobility group protein I/Y. *J. Biol. Chem.*, **270**, 4355–4360.
13. Maher,J.F. and Nathans,D. (1996) Multivalent DNA-binding properties of the HMG-1 proteins. *Proc. Natl Acad. Sci. USA*, **93**, 6716–6720.
14. Hill,D.A., Pedulla,M.L. and Reeves,R. (1999) Directional binding of HMG-I(Y) on four-way junction DNA and the molecular basis for competitive binding with HMG-1 and histone H1. *Nucleic Acids Res.*, **27**, 2135–2144.
15. Zhang,Q. and Wang,Y. (2010) HMG modifications and nuclear function. *Biochim. Biophys. Acta*, **1799**, 28–36.
16. Reeves,R. (2000) Structure and function of the HMGI(Y) family of architectural transcription factors. *Environ. Health Perspect.*, **108(Suppl. 5)**, 803–809.
17. Heilemann,M., van de Linde,S., Mukherjee,A. and Sauer,M. (2009) Super-resolution imaging with small organic fluorophores. *Angew. Chem. Int. Ed. Engl.*, **48**, 6903–6908.
18. Wombacher,R., Heidbreder,M., van de Linde,S., Sheetz,M.P., Heilemann,M., Cornish,V.W. and Sauer,M. (2010) Live-cell super-resolution imaging with trimethoprim conjugates. *Nat. Methods*, **7**, 717–719.
19. van de Linde,S., Löschberger,A., Klein,T., Heidbreder,M., Wolter,S., Heilemann,M. and Sauer,M. (2011) Direct stochastic optical reconstruction microscopy with standard fluorescent probes. *Nature Protocols*, in press doi:10.1038/nprot.2011.336.
20. Rust,M.J., Bates,M. and Zhuang,X. (2006) Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods*, **3**, 793–795.
21. Klein,T., Löschberger,A., Proppert,S., Wolter,S., van de Linde,S. and Sauer,M. (2011) Live-cell dSTORM with SNAP-tag fusion proteins. *Nat. Methods*, **8**, 7–9.
22. Frank,O., Schwanbeck,R. and Wiśniewski,J.R. (1998) Protein footprinting reveals specific binding modes of a high mobility group protein I to DNAs of different conformation. *J. Biol. Chem.*, **273**, 20015–20020.
23. Disney,J.E., Johnson,K.R., Magnuson,N.S., Sylvester,S.R. and Reeves,R. (1989) High-mobility group protein HMG-I localizes to G/Q- and C-bands of human and mouse chromosomes. *J. Cell. Biol.*, **109**, 1975–1982.
24. Saitoh,Y. and Laemmli,U.K. (1994) Metaphase chromosome structure: bands arise from a differential folding path of the highly AT-rich scaffold. *Cell*, **76**, 609–622.
25. Munshi,N., Agalioti,T., Lomvardas,S., Merika,M., Chen,G. and Thanos,D. (2001) Coordination of a transcriptional switch by HMGI(Y) acetylation. *Science*, **293**, 1133–1136.
26. Zhang,Q., Zhang,K., Zou,Y., Perna,A. and Wang,Y. (2007) A quantitative study on the *in vitro* and *in vivo* acetylation of high mobility group A1 proteins. *J. Am. Soc. Mass Spectrom.*, **18**, 1569–1578.
27. Edberg,D.D., Bruce,J.E., Siems,W.F. and Reeves,R. (2004) *In vivo* posttranslational modifications of the high mobility group A1a proteins in breast cancer cells of differing metastatic potential. *Biochemistry*, **43**, 11500–11515.
28. Wolter,S., Schüttpelz,M., Tscherepanow,M., VAN DE Linde,S., Heilemann,M. and Sauer,M. (2010) Real-time computation of subdiffraction-resolution fluorescence images. *J. Microsc.*, **237**, 12–22.
29. Reeves,R. and Nissen,M.S. (1999) Purification and assays for high mobility group HMG-I(Y) protein function. *Methods Enzymol.*, **304**, 155–188.
30. Bock,C.T. and Zentgraf,H. (1993) Detection of minimal amounts of DNA by electron microscopy using simplified spreading procedures. *Chromosoma*, **102**, 249–252.