11906–11920 *Nucleic Acids Research, 2019, Vol. 47, No. 22* *Published online 24 July 2019*
doi: 10.1093/nar/gkz614

Phosphorylation orchestrates the structural ensemble of the intrinsically disordered protein HMGA1a and modulates its DNA binding to the NFκB promoter

Bastian Kohl[1], Xueyin Zhong[1], Christian Herrmann[2] and Raphael Stoll[1,*]

[1]Faculty of Chemistry and Biochemistry, Biomolecular NMR Spectroscopy, Ruhr University of Bochum, Universitätsstraße 150, 44780 Bochum, Germany and [2]Faculty of Chemistry and Biochemistry, Protein Interactions, Ruhr University of Bochum, Universitätsstraße 150, 44780 Bochum, Germany

Received February 06, 2019; Revised June 14, 2019; Editorial Decision June 24, 2019; Accepted July 05, 2019

ABSTRACT

High Mobility Group Protein A1a (HMGA1a) is a highly abundant nuclear protein, which plays a crucial role during embryogenesis, cell differentiation, and neoplasia. Here, we present the first ever NMR-based structural ensemble of full length HMGA1a. Our results show that the protein is not completely random coil but adopts a compact structure consisting of transient long-range contacts, which is regulated by post-translational phosphorylation. The CK2-, cdc2- and cdc2/CK2-phosphorylated forms of HMGA1a each exhibit a different binding affinity towards the PRD2 element of the NFκB promoter. Our study identifies connected regions between phosphorylation sites in the wildtype ensemble that change considerably upon phosphorylation, indicating that these posttranslational modifications sites are part of an electrostatic contact network that alters the structural ensemble by shifting the conformational equilibrium. Moreover, ITC data reveal that the CK2-phosphorylated HMGA1a exhibits a different DNA promoter binding affinity for the PRD2 element. Furthermore, we present the first structural model for AT-hook 1 of HMGA1a that can adopt a transient α-helical structure, which might serve as an additional regulatory mechanism in HMAG1a. Our findings will help to develop new therapeutic strategies against HMGA1a-associated cancers by taking posttranslational modifications into consideration.

INTRODUCTION

During the last decade, the function of intrinsically disordered proteins (IDPs) has become an additional focus of structural research endeavours. Many studies show that IDPs are involved in a broad range of important biochemical processes, such as transcriptional regulation, chromosomal arrangement, virus integration, and large protein complex assembly (1–5). Their participation in several maladies like cancer, HIV, Alzheimer and Parkinson diseases was also discovered (6–8). In order to gain a deeper understanding of how IDPs work, several of their structural ensembles have been calculated. These include structural representations of α-Synuclein, $p27^{Kip1}$, Sic1, tau protein, p53, and the measles virus nucleocapsid (9–14). Different methods have been proposed to calculate structural ensembles of IDPs. These are based on software packages such as ENSEMBLE, Flexible Meccano, molecular dynamic simulation-based algorithms, and Bayesian statistics-based approaches (15–18). IDPs are highly dynamic, which makes them flexible for interactions with different classes of biomolecules, such as other proteins, DNA, and RNA (19,20). An excellent example for this promiscuity is the non-histone chromosomal high mobility group A1 protein (HMGA1), which can interact with duplex B-DNA, holiday junctions, RNA, and other proteins, such as the transcription factor NFκB1 (6,21–23).

HMG-proteins are the second most abundant class of chromosomal proteins and they are involved in chromosomal rearrangement, DNA unbending, and transcriptional regulation in higher eukaryotes (24–27). The HMGA protein family consists of three members: HMGA1a, HMGA1b and HMGA2 (28–30). HMGA1a is involved in several biochemical processes like embryogenesis, enhancersome assembly, neoplasia, and carcinogenesis. Overexpression of HMGA1a has been detected for several types of cancer including breast, colon, thyroid, ovary, and gastric malignancies (31–34). High levels of HMGA1a indicate a negative prognostic outcome for long term survival of cancer patients (35). Therefore, a better understanding of the structure of HMGA1a could be essential in the development of therapeutic strategies against various cancers.

HMGA family members are highly unstructured and contain three DNA-binding domains—denoted as AT-hooks (AT1, AT2, AT3)—and a glutamate rich, highly neg-

*To whom correspondence should be addressed. Tel: +49 234 32 25466; Fax: +49 234 32 05466; Email: raphael.stoll@ruhr-uni-bochum.de

atively charged C-terminal region (Figure 1A) (36). X-ray and NMR structures of HMGA1 have only been solved for fragments of AT2 and AT3, each bound to AT-rich duplex DNA (21,37). A structural investigation of the overall structure of HMGA1a has not yet been performed. However, a structural analysis of only AT2 and AT3 showed that HMGA1a does not adopt classical secondary structure elements upon binding to DNA (21). The AT-hook DNA binding domains consist of two major components that bind to AT-rich DNA sequences. The first component harbours the central RGR-motif, which directly binds into the minor groove of the target DNA. The second component consists of neighbouring pairs of positively charged residues, such as arginine or lysine, which form hydrophobic and electrostatic interactions with the DNA phosphodiester backbone. Additionally, the second AT-hook contains a C-terminal extension, which exhibits further contacts with the backbone of the DNA and increases the binding affinity by one or two orders of magnitude (21).

Due to their important cellular function, HMGA1 proteins are highly regulated. In fact, they are one of the most posttranslational modified proteins in the nucleus, including acetylation, methylation, and phosphorylation by a broad range of different enzymes (38). Earlier studies had shown that HMGA1a is one of the most phosphorylated nuclear proteins. HMGA1a is phosphorylated by several kinases including Casein kinase 2 (CK2), cyclin-dependent kinase 2 (cdc2/CCNB1), protein kinase C (PKC), and the homodomain-interacting protein kinase-2 (HIPK2) (39–45). *In vivo*, CK2 constitutively phosphorylates residues S99, S102 and S103, located at the negatively charged C-terminus. In HMGA1b, this reduces the binding affinity by a factor of three for AT-rich DNA (45). In contrast, the cdc2 kinase phosphorylates residues S36, T53 and T78, which are each adjacent to the three AT-hooks. These posttranslational modifications (PTMs) also impair binding to AT-rich DNA as they reduce the affinity by a factor of 4.5 (40,41). However, little is known about the impact of PTMs on the overall structure of HMGA1 in its uncomplexed form and, in particular, their structural impact on the AT-hooks. It could be shown by ion-mobility separation-mass spectrometry (IMS-MS) that the C-terminal phosphorylation of HMGA2 leads to a higher compactness of HMGA2 (46). Other studies revealed that various regions of HMGA1 in the complexed form are protected against proteolysis to a different degree (45).

Here, we have characterised the overall structure of wild-type HMGA1a and the effects of phosphorylation by CK2 and cdc2 on an atomic level using multidimensional nuclear magnetic resonance (NMR) spectroscopy and isothermal titration calorimetry (ITC). In order to analyse the secondary and tertiary structure of HMGA1a, long range contacts and structural features were analysed by paramagnetic relaxation enhancement (PRE), ^{3}J couplings, and chemical shift analysis (CSA). The effects of phosphorylation of HMGA1a were analysed through chemical shift perturbation (CSP) experiments and their effect on NFκB PRD2 DNA promoter affinity was analysed using ITC as well as NMR spectroscopy. In order to characterise the structure in solution of HMGA1a, an IDP structural ensemble has been calculated using the software package ENSEMBLE (17,18,47). The acquired data presented here provide a deeper understanding of the structural properties of HMGA1a, the structural effects of phosphorylation of different amino acids, and the functional consequences thereof for promoter DNA binding of its AT-hooks.

MATERIALS AND METHODS

Cloning, expression, and purification of HMGA1a proteins

For the expression and purification, the codon-optimized hmga1a gene was always freshly transferred into chemical competent *Escherichia coli* BL21 (DE3) T1R cells.

The hmga1a gene was cloned into pET9a vector including resistance genes for kanamycin. After transformation, 200 ml LB media [30 ug/ml Kanamycin (Roth)] overnight, the culture was inoculated and incubated at 37°C at 200 rpm in a 1 l flask. Then, 8 l LB-media were inoculated to $OD_{600} = 0.1$ and grown at 37°C and 160 rpm in 4 × 5 l chicane flasks to an $OD_{600} \approx 0.6$. Cells were harvested by centrifugation at 4000 rpm (Sorval GS3 rotor) at room temperature for 20 min and re-suspended in 2 l ^{15}N [0.5 g/l $^{15}NH_4Cl$ (Euroisotop]; 10 g ^{12}C-glucose)- or $^{15}N/^{13}C$ (0.5 g/l $^{15}NH_4Cl$; 2 g ^{13}C-glucose (Euroisotop))-high density minimal medium. Cells were adapted to minimal medium for 1 h at 37°C and 160 rpm before induction by adding 0.5 mM IPTG (isopropyl β-D-1-thiogalactopyranoside) per liter medium. Cells were harvested after 3 h at 37°C by centrifugation at 4000 g (Sorval GS3 rotor) at 4°C. Afterwards, cells were washed in RSB buffer and frozen in liquid nitrogen and stored at −80°C until usage. Protein purification was performed as described by Reeves *et al.* applying a TCA extraction procedure (48). Crude protein extracts were exchanged twice into buffer before further use. Protein purity was monitored by polyacrylamide-gel electrophoreses. Typically, samples contained >95% of pure protein.

Cysteine mutants for PRE measurements were produced by standard protocol for QuikChange mutagenesis.

Preparation of PRD2 DNA oligonucleotide

PRD2 Promoter element was purchased by Eurofins genomics. Single stranded DNA oligos (forw: GGGAAA TTCCTC; rev: GAGGAATTTCCC) were dissolved in Phospho-NMR buffer and mixed in a 1:1 ratio. Annealing of DNA was performed by heating the sample for 10 min to 98°C and then cooling down over 2 h with temperature gradient to room temperature. Formation of double stranded DNA was validated by 1D ^{1}H NMR spectra.

NMR assignment of backbone and sidechain resonances of HMGA1a and its CK2-phosphorylated form

In order to obtain backbone chemical shifts for $HMGA1a_{S64C}$ (^{15}N, ^{13}CO, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, $^{1}H_{\alpha}$, $^{1}H_{N}$), three-dimensional triple resonance spectra were recorded on a 1.5 mM $HMGA1a_{S64C}$ (U-^{15}N/U-^{13}C) sample in NMR buffer (50 mM HEPES, 150 mM NaCl, 10 mM TECEP, pH 7.0) at 298 K. All NMR spectra were recorded on Bruker DRX, AVANCE II, and AVANCE III HD spectrometers operating at 600 and 700 MHz ^{1}H Larmor frequencies, respectively. For assignment of chemical

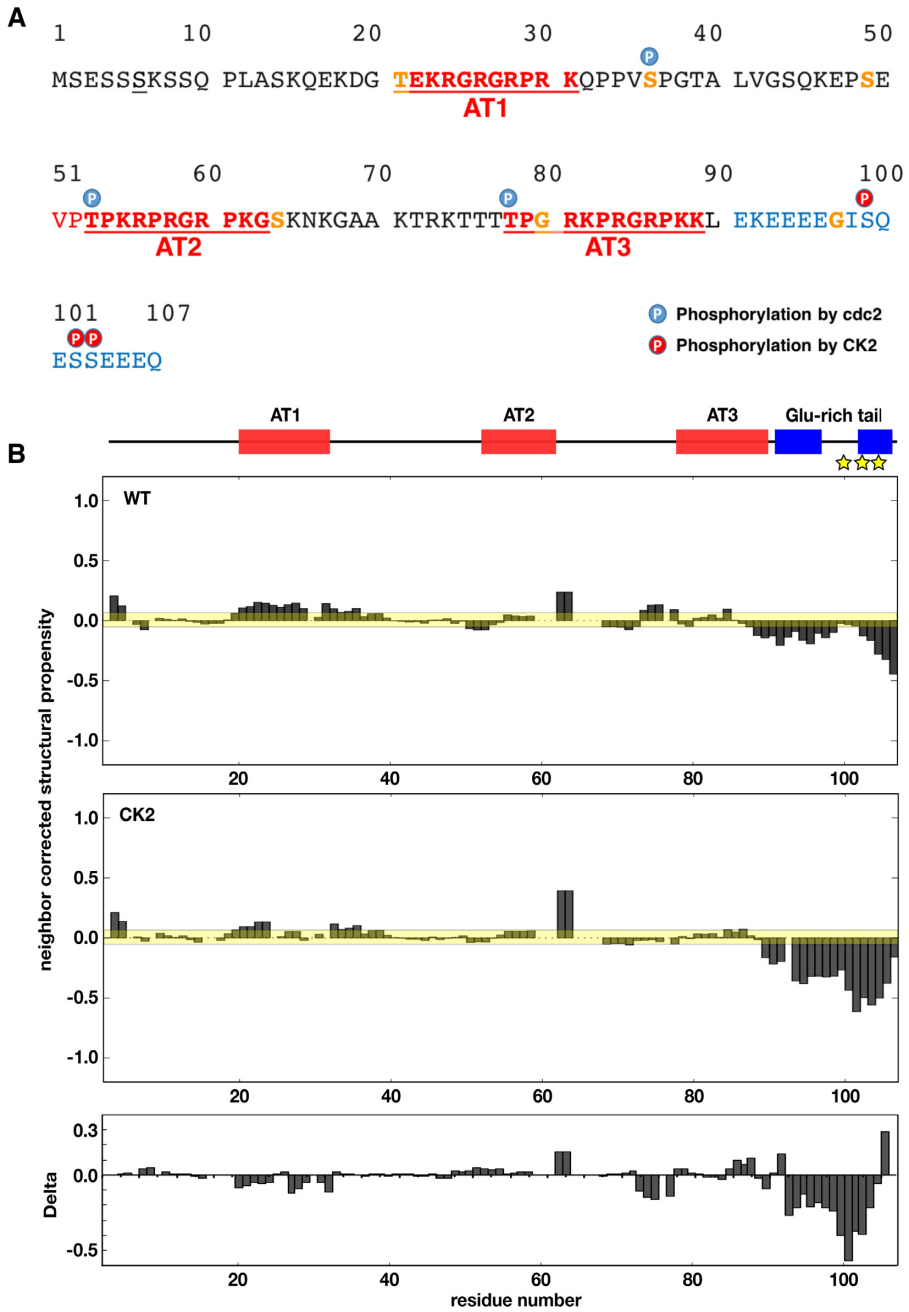

Figure 1. (**A**) Domain structure of HMGA1a. The phosphorylation sites of cyclin-dependent kinase 2 (cdc2) and Casein Kinase 2 (CK2) are shown in blue and red, respectively. PRE spin label positions are highlighted in orange. AT-hooks of HMGA1a are coloured in red and underlined. The charged C-terminus is coloured in blue. (**B**) Neighbour-corrected structural propensity analysis of HMGA1a: All values were plotted with a five-residue smoothing function. Values in range of [–0.1,0.1] are indicative of random coil structure. Helical propensities are in range of [0.1,1.0] and extended/β-sheet structures are in range of [–0.1,–1.0] WT: Propensity plot of HMGA1a$_{S64C}$ wt. CK2: Propensity plot of CK2-phosphorylated HMGA1a$_{S64C}$. Delta: Difference of ncSSP-values of wt HMGA1a and its CK2-phosphorylated form (ΔncSSP = ncSSP$_{WT}$ – ncSSP$_{CK2}$).

shifts of wt HMGA1a at pH 7, the following spectra were recorded: BT-HNCACB (^{1}H: 1024/^{15}N: 90/^{13}C: 140; ns: 16), BT-HNCOCACB (^{1}H: 1024/^{15}N: 90/^{13}C: 140; ns: 16), BT-HNCO (^{1}H: 1024/^{15}N: 90/^{13}C: 140; ns: 16), virtually decoupled IPAP-(H)CBCACON (^{13}C: 1024/^{15}N: 64/^{13}C: 180; ns: 16), virtually decoupled IPAP-(H)CANCO (^{13}C: 1024/^{15}N: 64/^{13}C: 128; ns: 32) , virtually decoupled IPAP 2D [^{13}C,^{15}N] NCO (^{13}C: 1024/^{15}N: 128; ns: 32) (49,50). The use of carbon-detected (H)CBCACON and (H)CANCO spectra simplified the assignment of sidechains due to the increased signal dispersion of carbon-detected 2D [^{13}C,^{15}N] NCO spectra compared to 2D [^{15}N,^{1}H] HSQC spectra (Supplementary Figure S2). Chemical shifts were referenced to DSS and indirect referencing for carbon and nitrogen was employed. Assignments have been cross-validated with the published assignments of HMGA1a at

pH 6 by Buchko *et al.* (51). Assignments of ^{1}H, ^{13}C, and ^{15}N NMR chemical shifts have been deposited in the Biological Resonance Bank (BMRB, http://www.bmrb.wisc.edu/) under accession numbers 27883 and 27884 for HMGA1a$_{S64C}$ and its CK2-phosphorylated form, respectively.

Spin-labeling of HMGA1a cysteine-mutants with the paramagnetic MTSL (*S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl-methanesulfonothioate) label

Cysteine-mutants of HMGA1a (U-^{15}N) were exchanged into DTT-buffer (50 mM HEPES, 20 mM DTT, pH 7.6) and incubated for 2 h at room temperature in order to reduce all thiol groups. During the next step, DTT was removed using 7 kDa ZEBA spin desalting column (GE), the protein was exchanged into reaction buffer (50 mM HEPES, 10 mM MTSL (Toronto Research Chemicals), pH 7.5), and labelled overnight in the dark at room temperature. Finally, the sample was exchanged into NMR buffer (50 mM HEPES, 150 mM NaCl, 10% D_2O, pH 7.0) using a 7 kDa ZEBA spin desalting column (GE). In order to obtain diamagnetic samples, the spin label was reduced by addition of 2 mM sodium ascorbate (Roth) and incubation for 12 h.

NMR acquisition of para- and diamagnetic spectra of MTSL-labeled HMGA1a and its CK2-phosphorylated CK2 form

For PRE analysis of HMGA1a (U-^{15}N), the cysteine-mutants T21C, S36C, S49C, S64C, G80C, and G97C were analysed. For its CK2-phosphorylated form, spectra of HMGA1a (U-^{15}N) cysteine mutants S64C and G97C were recorded. NMR samples contained 0.4 mM MTSL-labeled HMGA1a protein (U-^{15}N) in NMR buffer (50 mM HEPES, 150 mM NaCl, 10% D_2O, pH 7.0/CK2-form: additionally 2 mM NaF, 2 mM sodiumpyrophosphate, 2 mM β-glycerolphosphate) at pH 7. All 2D [^{15}H,^{1}H] HISQC spectra (ns 32, TD_{1H}: 2k, TD_{15N}: 256) under para- and diamagnetic conditions were recorded on a Bruker DRX 600 MHz spectrometer at 298 K (52). Data were processed with nmrPipe and analysed using the CCPNMR 2.4.2 software package (53–55). PRE derived constraints were calculated according to (56).

In vitro phosphorylation of HMGA1a by Casein kinase 2 and/or cdc2

For phosphorylation of HMGA1a (U-^{15}N and U-^{15}N/U-^{13}C), 600 μl of a 1 mM sample were exchanged into phosphorylation buffer (50 mM Tris–HCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 2 mM DTT, 0.01 % Brij 35, 20 mM ATP, pH 7.5) containing 1x phosphatase inhibitor cocktail (Thermo Scientific) using a 7 kDa ZEBA spin desalting column (GE). 25 000 Units of CK2 (New England Biolabs) and/or 50 μg of cdc2/CCNB1 (Abcam) were used in the phosphorylation reaction. The phosphorylation was performed for 12 h at 30°C. Afterwards, HMGA1a was exchanged into Phosphor-NMR buffer (50 mM HEPES, 150 mM NaCl, 20 mM DTT, 2 mM NaF, 2 mM beta-glycerolphosphat, 2 mM sodium pyrophosphate, 10% D_2O, pH 7.0) as described previously.

Chemical shift perturbation analysis of CK2-, cdc2-, and CK2/cdc2-phoshorylated HMGA1a

In order to investigate the sites, extent, and impact of phosphorylation of HMGA1a by CK2 and/or cdc2, a chemical shift perturbation analysis was carried out. For each HMGA1a form (U-^{15}N/U-^{13}C), a 2D [^{15}N,^{1}H] HSQC spectrum (ns: 8, TD_{1H}: 2k, TD_{15N}: 256) of a 1.5 mM sample in Phosphor-NMR buffer was measured at pH 7. Spectra were recorded on a Bruker AVANCE III HD 700 MHz spectrometer at 298 K. Data were processed by TopSpin 3.5 and referenced to DSS. Calculation of chemical shift distance was performed with CCPNMR 2.4.2 (53–55). For shift distance calculations, nitrogen shifts were rescaled by a factor of 0.15.

Measurement of 2D [^{15}N,^{1}H] sfHSQC spectra at different pH values

For each pH value (4.0 and 7.0), a 2D [^{15}N,^{1}H] sfHSQC spectrum (ns: 8, TD_{1H}: 2k, TD_{15N}: 128) of 0.4 mM sample (U-^{15}N) in NMR buffer was recorded. All spectra were acquired on a Bruker DRX 600 MHz spectrometer at 298 K. All data were processed by TopSpin 3.5 and spectra were analysed using CCPNMR 2.4.2 (53–55).

Measurement of 2D [^{13}C,^{15}N] NCO NMR spectra of HMGA1a in complex with the PRD2 DNA element

For both HMGA1a wt (U-^{15}N/U-^{13}C) and CK2 phosphorylated HMGA1a (U-^{15}N/U-^{13}C), virtually decoupled IPAP 2D [^{13}C,^{15}N] NCO (^{13}C: 1024/^{15}N: 128; ns: 64) spectra of the non-phosphorylated as well as the phosphorylated form of a 0.2 mM sample with 3.4× (for wt HMGA1a) and 3.5× (for CK2 phosphorylated HMGA1a) stoichiometric excess of DNA in Phosphor-NMR buffer were recorded (49,50). All spectra were acquired on a Bruker Avance III HD 700 MHz spectrometer at 298 K. Data were processed by TopSpin 3.5 and spectra were analysed using CCPNMR 2.4.2 (53–55).

Measurement of R1/R2 relaxation data and HN ^{3}J coupling constants of HMGA1a

T_1/T_2 relaxation data were measured on a 1 mM NMR sample of HMGA1a (U-^{15}N) in NMR buffer at pH 7.0. Standard Bruker pulse programs were used and data were recorded on a Bruker DRX 600 MHz spectrometer at 298K. For T_2 relaxation measurements, a CPMG-based pulse program was used. For T_2 time measurement relaxation delays of 15, 31, 79, 126, 174, 221, 269, 316, 364, 411, 443 and 491 ms were used. Delays of 20, 170, 330, 480, 630, 780, 930, 1080, 1230, 1380, 1600 and 2000 ms were used for T_1 relaxation measurements. Relaxation data were processed using TopSpin 3.5. The analysis of spectra and data fitting was carried out using CCPNMR 2.7 (53–55). In order to determine ^{3}J coupling constants, a standard Bruker 3D [^{15}N,^{1}H,$^{1}H_{\alpha}$] HNHA spectrum was recorded on a Bruker DRX 600 MHz spectrometer at 298 K. Spectra were processed using TopSpin 3.5 and the coupling constants were determined using the CCPNMR 2.7 software package (53–55).

Prediction of PRE-Profiles for HMGA1a using Flexible Meccano

For the prediction of random coil PRE-Profiles, the software package Flexible Meccano was used (15). For each PRE label position, a random coil ensemble of 10 000 structures and its PRE profile was generated. The intrinsic linewidth used during simulation was calculated based on the corresponding NMR-spectra.

Calculation of the structural ensemble of wt HMGA1a using ENSEMBLE

HMGA1a structural ensembles were calculated with the software ENSEMBLE (47). Briefly, four independent ensembles consisting of 100 structures each were selected out of structural pools of ∼20 000–80 000 conformers. Restraints were fitted sequentially into the ensemble in the following order: chemical shifts, PRE-data, R_2-relaxataion data, and finally ^3J couplings. The independent ensembles were grouped into eight structural clusters by an RMSD-based clustering method using the Clusco software package (12,57).

ITC measurement of HMGA1a in complex with the PRD2 element of the NFκB promoter

In order to measure binding affinities and enthalpies between both phosphorylated or non-phosphorylated HMGA1a (unlabelled) and the PRD2 DNA element, isothermal titration calorimetry was performed. All data were acquired on a microcal ITC 200. ITC data were acquired in Phosphor-NMR buffer at 298 K. In the reaction cell, a DNA concentration of 0.09 mM was used. In the syringe, an HMGA1a (wt or CK2-phosphorylated form) concentration of 0.45 mM was used. Data analysis was performed with the independent model fitting and minimization algorithm of Affinimeter software (www.affinimeter.com). An independent three binding site model was used to fit the data, which is in good agreement with our NMR data. This model is based on site-specific equilibrium constants and describes ligand binding to a multivalent receptor with a number of independent binding sites (58). For the fitting of the data, an *n*-value of 0.33 was assumed for each independent site and the concentration values for substances in cell and syringe were fitted within a range of ±10 %.

RESULTS

Secondary structure of wildtype HMGA1a

To determine as to whether HMGA1a contains residual secondary structure propensities, we acquired 3D proton- and carbon-detected triple resonance spectra at neutral pH to obtain chemical shifts and performed a neighbour-corrected secondary structure (ncSSP) analysis as shown in Figure 1B (59). HMGA1a only shows low propensity for secondary structures. Most of the ncSSP values are within the range from –0.1 to 0.1, which indicates random coil structure. Interestingly, the first AT-hook shows a small propensity for an α-helical structure of ∼0.15. The other AT-hooks do not possess a propensity for secondary structure over their entire length. In contrast to the remainder of the protein, the negatively charged C-terminal region shows, on average, a significant propensity for an extended/β-sheet structure of 20%.

Tertiary structure of HMGA1a

In order to investigate as to whether HMGA1a adopts an internal structure, which is based on long range contacts between different regions, we performed nitroxide spin labelling of HMGA1a. To this purpose, cysteine point mutations were introduced in six positions (T21, S36, S49, S64, G80, G97) of HMGA1a and labelled with (*S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl-methanesulfonothioate) (MTSL) (Figure 1A). The attachment of paramagnetic MTSL at specific sites in proteins can be used to measure transient interactions. Due to the paramagnetic nature of the MTLS spin labels, the HSQC peaks of ^{15}N amides that are in close proximity or in transient contact with the MTSL spin labels will become line-broadened because of enhanced transverse relaxation rates. This effect depends on the distance from the spin label, and can be measured by recording para- as well as diamagnetic spectra and plotting intensity ratios against the protein sequence (60). Based on these intensity ratios, distances between the label and the amide proton can be calculated for contacts up to 2 nm (56). In order to investigate as to whether the DNA binding AT-hooks exhibit long-range contacts to other regions of the protein, spin-labels were positioned close to them (T21, S36, S49, S64, G80, G97). These include residues T21 for AT1, S64 for AT2 and G80 for AT3. Additionally, residues at the C-terminus (G97) and between the AT-hooks (S36, S49) were MTSL-labelled to obtain more information on the function of the negatively charged carboxy-terminus. In Figure 2, the intensity ratios are plotted for all labelled positions against the amino acid sequence of HMGA1a. For amino acids T21 and S36, long range contacts could not be detected. This suggests that AT1 and the linker region between AT1 and AT2 lacks preferential contacts to other regions of HMGA1a. The obtained intensity profiles correlate well with the predicted non-contact profiles generated by Flexible Meccano (15). Serine 49, which is close to AT2, shows a weak contact to region ranging from 76 to 86 that includes AT3. Noteworthy, long range contacts to other regions in HMGA1a for AT2 and AT3 could indeed be detected. AT2 shows long range contacts to the C-terminal region of HMGA1a. Spin-labeling at position 64 leads to a strong decrease of signal intensities in the region ranging from E96 to E106 with a minimum of signal intensity observed for residues Q101 and S102. Based on Γ_2 rates of this region, this corresponds to an average distance of AT2 to the C-terminus of ∼1.8 nm in the HMGA1a ensemble. AT3 also shows contacts to the C-terminus of the protein. Spin labeling at position 80 leads to increased paramagnetic relaxation rates at the glutamate rich region of HMGA1a. These contacts of AT2 and AT3 can be confirmed by the spin label at position G97 that shows similar relaxation enhancement effects. In summary, paramagnetic relaxation analysis shows that HMGA1a has an internal domain organization character-

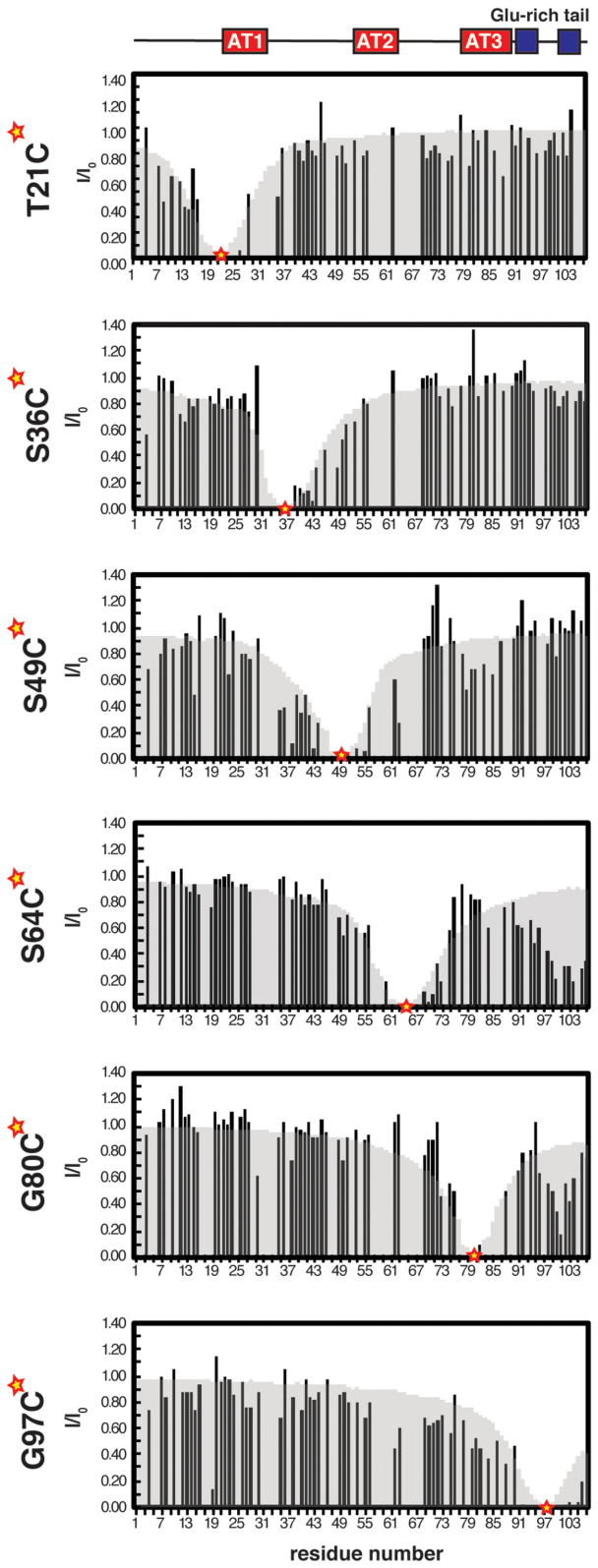

Figure 2. PRE-Profiles of wildtype HMGA1a. Gray background indicates random-coil PRE-Profiles for specific label position calculated with Flexible Meccano based on 10 000 structures and the intrinsic linewidth of each dataset. Yellow stars indicate the MTSL-label site.

Table 1. Overview of experimental input data for the ENSEMBLE calculation and RMSD values of the computed HMGA1a ensemble

Restraint Type	Number of restraints	RMSD
$^{13}C_\alpha$ chemical shift	87	0.51 ppm
$^{13}C_\beta$ chemical shift	66	0.51 ppm
$^{15}N_H$ chemical shift	71	1.08 ppm
$^{1}H_N$ chemical shift	68	0.14 ppm
$^{13}C'$ chemical shift	72	0.53 ppm
$^{1}H_\alpha$ chemical shift	38	0.09 ppm
PRE	55	NA
^{3}J	34	NA
^{15}N R_2	53	0.65
Total	545	NA

RMSD values are based on measured and back-calculated shift values for atoms based on the structure ensemble.

ized by long range contacts between AT2 and AT3 with the Glu-rich C-terminal region of HMGA1a. Interestingly, the N-terminal region of HMGA1a that includes the first AT-hook lacks any long-range contacts to other regions of the protein.

Structural model of HMGA1a calculated with ENSEMBLE

The NMR-based ensemble of HMGA1a is based on 545 NMR-derived restraints including chemical shifts, PRE-data, and ^{3}J coupling constants and was calculated with the software package ENSEMBLE. A compilation of the input data is shown in Table 1. The structural ensemble consists of four independent calculated ensembles of 100 structures each. Figure 3A shows the fractional contact map of the entire ensemble. Obviously, HMGA1a is not a completely random coil polypeptide and can adopt several conformations that are, for example, either very compact (Figure 3B, cluster 6) or fully extended (Figure 3B cluster 7). The crucial function of the negatively charged C-terminus for the structural regulation of the ensemble emerges through a detailed analysis of the contact map. Several regions, including AT2 and AT3, show contacts to the C-terminus. For a more detailed picture of the structural ensemble of HMGA1a, the structures have been analysed by C_α–C_α RMSD-based cluster analysis. Thereby, eight structural clusters could be assigned (Figure 3B). Clusters 1, 2 and 3 show open conformations and are more extended. They predominantly show contacts between neighbouring regions. Cluster 4 includes conformers that are more compact and contain contacts to the C-terminal region of HMGA1a as well as a region located between AT2 and AT3. Cluster 5 consists of conformers that exhibit contacts between AT3 and the C-terminus. Overall, the most compact structures are found in cluster 6. It shows a very compact conformation of HMGA1a with contacts between almost all regions. Clusters 7 and 8 are only populated to a low extent and are characterized by rather extended conformations.

Impact of phosphorylation by casein kinase 2 on the secondary and tertiary structure of HMGA1a

HMGA1a is one of the most posttranslational modified nuclear proteins (38). Phosphorylation of HMGA1a leads

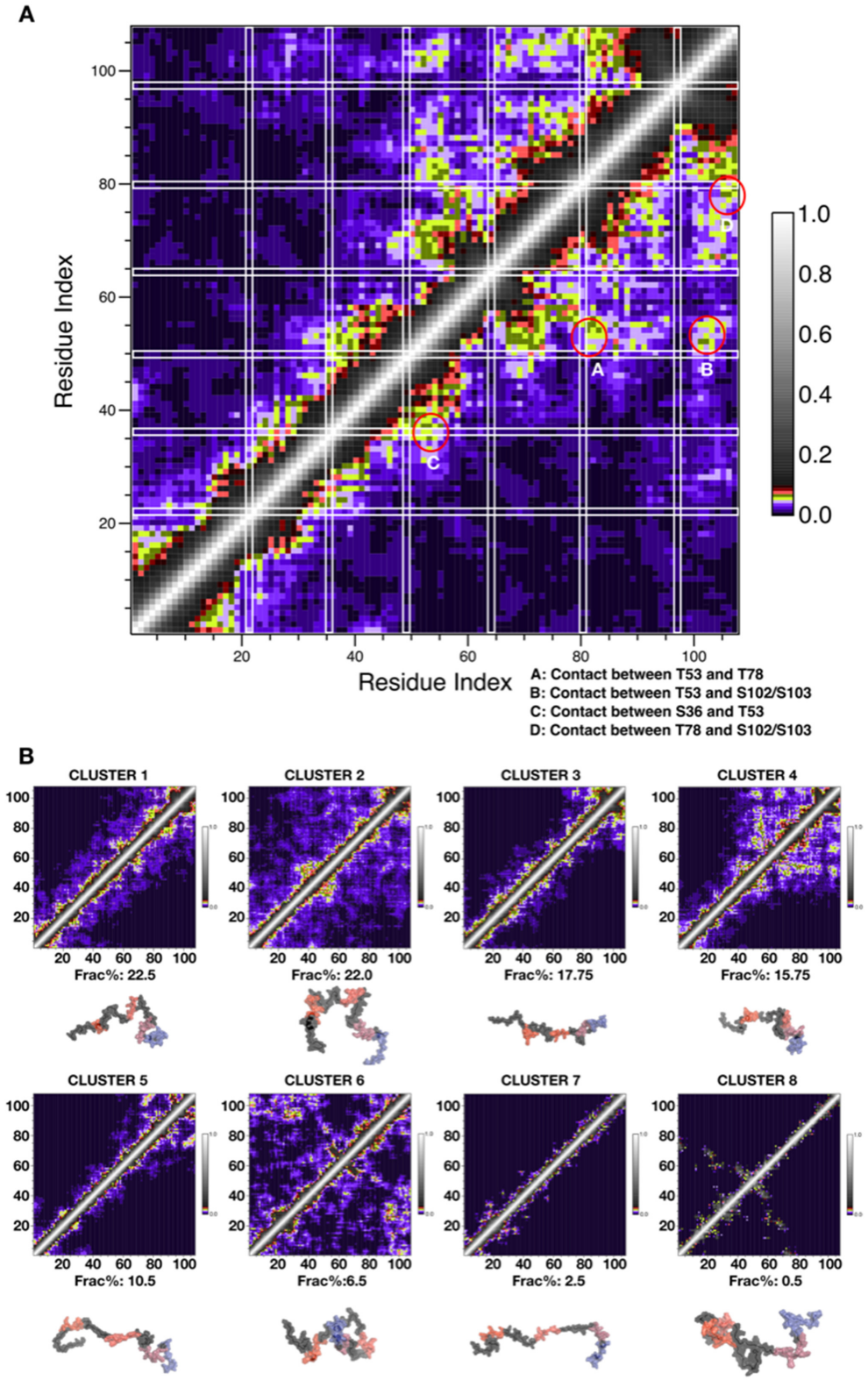

Figure 3. (**A**) Fractional contact plot of the HMGA1a structural ensemble calculated using the software package ENSEMBLE: Data of the ensemble are derived from four independent 100 structure ENSEMBLE calculations of 100 structures. White lines indicate PRE label positions. For the calculation of ensembles, PRE restraints of position S49, S64, G80 and G97 were used. Position T21 and S36 show no significant contacts. Red circles highlight contact regions of phosphorylation sites in HMGA1a. Position A indicates a contact between T53 and T78 that are both phosphorylated by cdc2. Position B indicates a contact between T78 and the negatively charged C-terminus (S102/S103). Position C depicts a contact between S36 and T53 that both are phosphorylated by cdc2. Position D indicates a contact between region of T78 and S102/S103 region. (**B**) Structural ensemble of HMGA1a calculated by ENSEMBLE: Structures of the four ensembles have been assigned to eight clusters. For each cluster, a fractional contact map is shown. Additionally, the median PDB-structure for each cluster is presented. The C-terminus is coloured in blue and the AT-hooks are shown in red.

to several effects in the cell, which include reduced binding affinity towards DNA and the alteration of protein complex stability, as observed for the enhancersome. In order to investigate the impact of PTMs on the overall structure, HMGA1a was phosphorylated by protein kinases CK2 and/or cdc2. Figure 4 shows the 2D [^{15}N,^{1}H] HSQC spectra of wt HMGA1a and the CK2-, cdc2-, and CK2/cdc2-phosphorylated forms. To validate the efficiency of CK2 phosphorylation we checked the resonance frequencies of S99, S102, and S103 observed in non-phosphorylated HMGA1a. After phosphorylation of HMGA1a, NMR signals for the non-phosphorylated form of these three amino acids could not be detected (Supplementary Figures S3 and S4), which indicates phosphorylation to nearly complete extent. Furthermore, the $^{13}C_{\beta}$ chemical shifts for residues S99, S102, and S103 change upon phosphorylation by approximately 2 ppm each, clearly suggesting three phosphorylation sites (61). Other possible phosphorylation sites like S14, T21, T78, and others did not show any chemical shift perturbations which would indicate phosphorylation. Based on the NMR data, CK2 thus yields a single new species of HMGA1a with three completely phosphorylated residues—S99, S102, and S103. The signal dispersion of the wt protein for protons is very small, consistent with its mainly intrinsically disordered character. Phosphorylation induces strong chemical shift perturbations towards lower field for several amino acids. In Figure 1B, the ncSSP profile of the CK2-phosphorylated form of HMGA1a is shown. Phosphorylation leads to a more extended/β-sheet like conformation of the C-terminus rich in negatively charged residues. A comparison of the secondary structure propensity plots for wt HMGA1a and its CK2-phosphorylated form reveals significant changes in the region ranging from 75 to 80 suggesting a propensity for more extended conformations within the random coil regime. AT1 loses its helical propensity and appears random coil in nature. To extract more information about the impact of the phosphorylation of HMGA1a by CK2, we performed a CSP analysis of the CK2-phosphorylated form shown in Figure 5B. Phosphorylation induces significant shift perturbations at the C-terminus and additionally in the region ranging from amino acid 70 to 86, which already showed secondary structure changes in the ncSSP analysis. This region harbours AT3 and one of two previously identified protein-protein interaction sites. Due to chemical exchange, the amide resonances of the AT-hooks are severely affected by line broadening. To overcome this problem and to investigate as to whether all AT-hooks are influenced by the C-terminal phosphorylation, we measured 2D [^{15}N,^{13}C] CON spectra to record proline resonances (Figure 5A, Supplementary Figure S1). The AT-hook regions are proline-rich and, therefore, these residues can be considered as important probes for the structural properties of the AT-hooks. Supplementary Figures S1 and 5A show the proline region of the 2D [^{15}N,^{13}C] CON spectrum and corresponding 1D slices thereof. After phosphorylation of HMGA1a by CK2, prolines 52, 54, 57, 79, 83, and 87 of the second as well as third AT-hook experience small chemical shift perturbations. Additionally, line broadening could be observed which are indicative for chemical exchange of these residues between open and more compact species of HMGA1a. These could suggest that the C-terminal modifications lead to structural alterations of these two AT-hooks due to transient long-range interactions, for which even clearer and stronger evidence is provided by the PRE-analysis (see below). Notably, no such effect could be detected for AT1.

PRE-analysis of internal structure of Casein Kinase 2 phosphorylated HMGA1a

To determine as to whether the phosphorylation of HMGA1a alters these tertiary structures, we measured PREs for amino acids S64 and G97. Based on the six PRE data sets on non-phosphorylated HMGA1a, results from the ENSEMBLE calculations on wildtype HMGA1a, and the CK2 phosphorylation sites, we carefully designed PRE spin label mutants S64C and G97C of HMGA1a as these would shed light on the kinase-mediated impact on the HMGA1a structural ensemble. In Figure 5C, the PRE-profiles of S64 and G97 are shown. The phosphorylation of HMGA1a by CK2 strengthens the long-range contact of the C-terminus to AT3 and its adjacent regions. The distance of the C-terminus to this region decreases and thus compresses the C-terminal region. This phosphorylation leads to a contact of G97 to the region ranging from residues 72 to 80 that is located right in front of AT3. In contrast, contacts of S64 to the identified wild-type contact region at the C-terminus are not altered. Only regions adjacent to S64 are affected, which indicatives a tighter contact of the C-terminus to the region ranging from residues 72 to 80. Thus, HMGA1a exhibits transient long-range contacts in despite of its IDP character.

Chemical shift perturbation analysis of cdc2-phosphorylated HMGA1a

In order to analyse the effects of HMGA1a phosphorylation by cyclin-dependent kinase 2 at position S36, T53, and T78, we compared the non-phosphorylated and cdc2-phosphorylated 2D [^{15}N,^{1}H] HSQC spectra and carried out a chemical shift perturbation analysis (Figure 5D). The phosphorylation by cdc2 induces chemical shift perturbations in several regions of HMGA1a. The largest changes could be detected for residues closer to the phosphorylation site. Additionally, the C-terminus rich in negatively charged residues is strongly affected and experiences substantial CSPs. This effect is in good agreement with the observed tertiary structure of HMGA1a by PRE-analysis. Phosphorylation of HMGA1a by cdc2 not only affects the phosphorylation sites themselves, but also alters the overall ensemble structure of HMGA1a.

Impact of dual phosphorylation of HMGA1a by Casein Kinase 2 and cyclin-dependent kinase 2

Several studies have shown that HMGA1a is phosphorylated by multiple kinases at the same time in the nucleus. Therefore, we investigated the impact of dual phosphorylation of HMGA1a by CK2 and cdc2, which play a crucial role during cell cycle control. It could been shown that this phosphorylation pattern could decrease the binding affinity of HMGA1a to AT-rich DNA by factor of 10

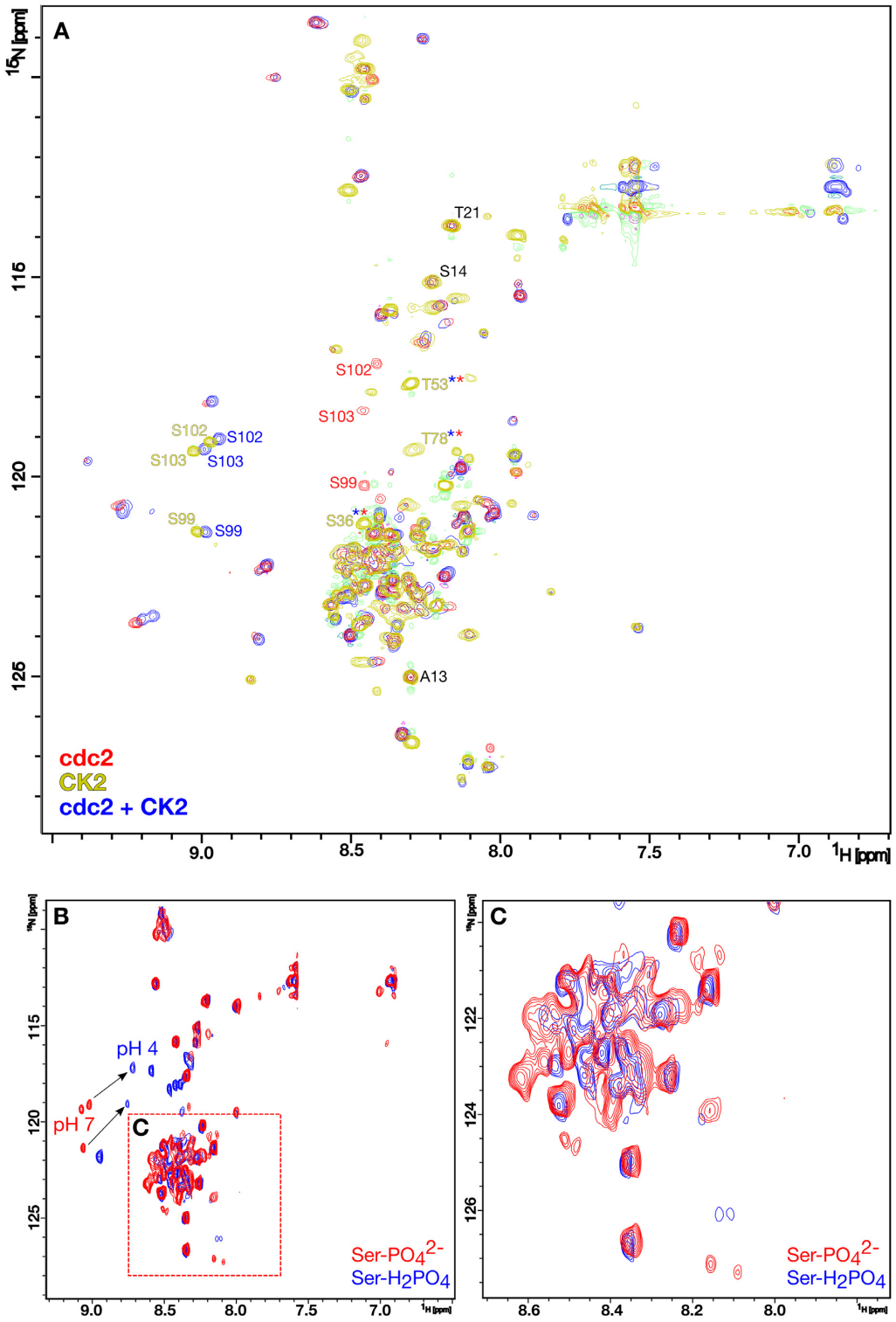

Figure 4. (**A**) Overlay of 2D [^{15}N,^{1}H] HSQC spectra of HMGA1a in different phosphorylation states: cdc2-form (red), CK2-form (green), cdc2/CK2-form (blue). Amino acid labels are coloured according to their phosphorylation state. Black amino acid labels represent unaffected amino acids in all states. (**B**) 2D [^{15}N,^{1}H] sfHMQC spectrum of the CK2-form at pH 4 (blue) and pH 7 (red). (**C**) Red labelled square: detailed view of the 2D [^{15}N,^{1}H] sfHMQC spectrum shown in B.

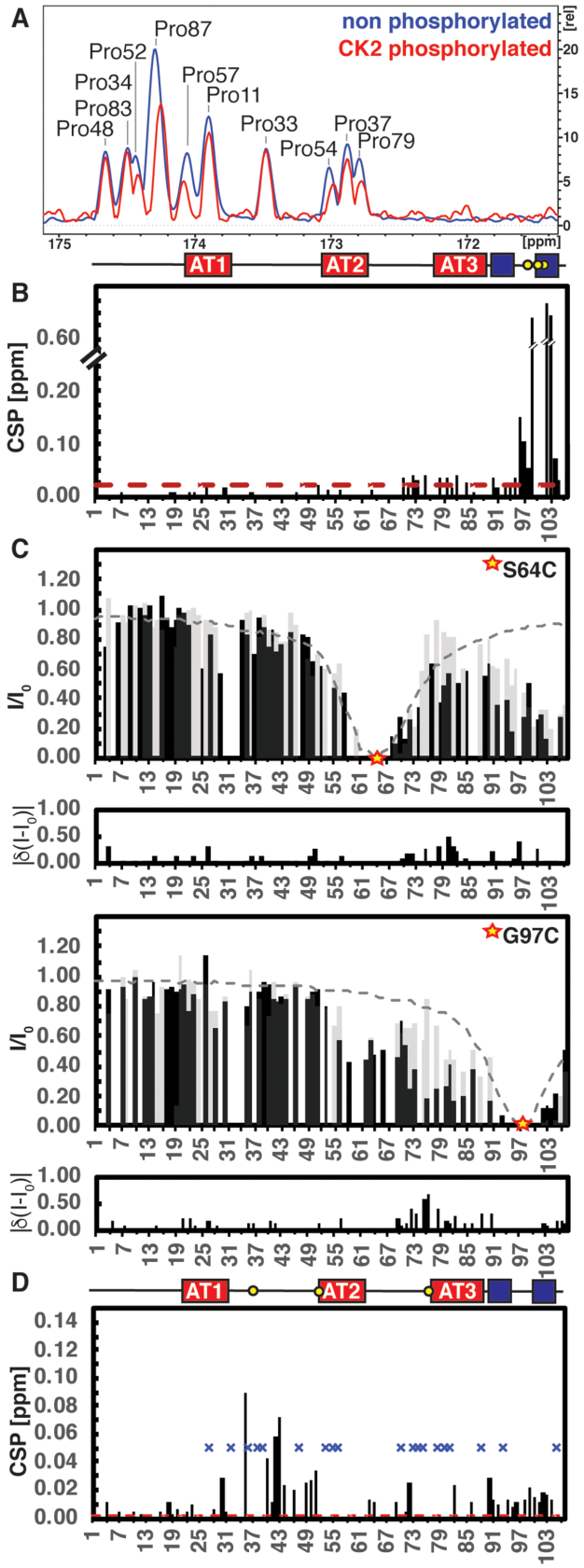

Figure 5. (**A**) 1D projection of the proline region of the 2D [^{15}N,^{13}C] CON spectra of HMGA1a$_{S64C}$ (red) and CK2-form (blue) of HMGA1a$_{S64C}$. (45). In Figure 4A, we present the superposition of the 2D [^{15}N,^{1}H] HSQC spectra of the CK2-, cdc2-, and CK2/cdc2-forms of HMGA1a. Our NMR spectra clearly show that resonances of phosphorylation sites in non-phosphorylated HMGA1a are not detectable in the 2D [^{15}N,^{1}H] HSQC spectrum of phosphorylated HMGA1a anymore and that phosphorylation is therefore almost complete (Supplementary Figures S3 and S4). The comparison of the three spectra clearly shows that the structural ensembles of these three forms are indeed different. The effects of phosphorylation by cdc2 and CK2 are not independent of each other. The amino acids S99, S102 and S103, which are phosphorylated by CK2, experience CSPs after dual phosphorylation in comparison to the CK2-form itself. This suggests that the C-terminus is affected by cdc2 phosphorylation and that the dual phosphorylated form adopts a different conformational ensemble than the single phosphorylated forms. On the contrary, other regions, such as the N-terminal region ranging from amino acids 13 to 21, do not exhibit CPSs after phosphorylation neither by CK2 or cdc2 nor by CK2 and cdc2 simultaneously. This clearly demonstrates that HMGA1a contains an electrostatic regulatory network, which is modulated by PTMs and regulates the structural IDP ensemble through controlled changes of the charge distribution within the protein.

Analysis of the general structural role of the C-terminus of HMGA1a

In order to investigate the impact of the electrostatic interaction of the C-terminus of HMGA1a on its overall structure, we reduced the pH-value of the NMR sample to 4. In Figure 4B, the superposition of the 2D [^{15}N,^{1}H] sfHMQC spectra of the CK2-form at pH 4 and pH 7 are shown. Obviously, the low pH leads to a strong decrease of the signal dispersion in the proton dimension, which is indicative for an even higher random-coil propensity of HMGA1a as compared to neutral pH. Several resonances in the crowded region between 8.2 and 8.6 ppm in proton dimension and 121.0 to 124.5 ppm in nitrogen dimension are even less well dispersed (Figure 4C). In conclusion, the charged C-terminus seems to be essential for the overall tertiary organisation of HMGA1a. Changes in charge distribution by PTMs, like phosphorylation, methylation, and acetylation, are not only the major regulatory

Please refer also to Supplementary Figure S1. (**B**) Chemical shift perturbation analysis of CK2-phosphorylated HMGA1a: The Red line indicates the 4σ-level. During analysis, all chemical shift perturbations above this level were considered to be significant. Yellow circles highlighted phosphorylation site S99, S102, and S103. (CSP values are reported in ppm). Yellow stars indicate the MTSL-label site. (**C**) PRE-profiles (I/I_0) of CK2-phosphorylated HMGA1a at position S64 and G97: The dashed lines indicate random coil PRE profile calculated using Flexible Meccano. Grey bars represent wt HMGA1 PRE profiles. Yellow circles indicate CK2 phosphorylation site at positions S99, S102, and S103. Yellow stars highlight MTSL-label sites. (**D**) Chemical shift perturbation analysis of cdc2-phosphorylated HMGA1a: The red line indicates 4σ-level. During analysis, all chemical shift perturbations above this level were taken as significant. A blue cross indicates an amino acid resonance that is significantly perturbed. Yellow circles label the phosphorylation site S36, T53, and T78 (CSPs are reported in ppm).

mechanisms that structurally regulate the entire ensemble of the IDP HMGA1a but—as shown for the CK2-phosphorylated form of HMGA1a—can also modulate its DNA binding affinity by up to three orders of magnitude as shown by our ITC data (Figure 6).

Analysis of the DNA-binding mode of HMGA1a after C-terminal phosphorylation by Casein Kinase 2

In order to investigate the impact of HMGA1a phosphorylation by CK2 in more detail, we acquired carbon detected NCO NMR spectra for a deeper insight into the structural changes of the AT hooks after DNA binding either in a phosphorylated or non-phosphorylated form. For this purpose, the prolines in the AT-hooks served as structural probes (Figure 6A). Our data show, that the chemical shifts of the prolines in the AT-hooks in wt and CK2-form of HMGA1a are same when bound to DNA, in contrast to the DNA-free wt and CK2-form of HMGA1a (Figure 5A). Apparently, the C-terminus of HMGA1a is replaced by the DNA. Hence, as prolines experience chemical shift perturbations upon CK2 phosphorylation, the C-terminus clearly modulates the DNA binding affinity of the AT-hooks, but does not changes the overall structure of the final DNA-protein complex.

In addition, we performed isothermal titration calorimetry (ITC) in order to obtain binding constants for the interaction between HMGA1a and DNA. Based on statistical analysis carried out using the AFFINImeter software package, the ITC data could be satisfactorily fitted assuming a model with site-specific equilibrium constants that describe ligand binding to multivalent HMGA1a with three independent DNA binding sites (58). Our ITC data reveal that the three AT hooks of HMGA1a exhibit different affinities towards the PRD2 DNA element. Based on the acquired data it is not possible to assign the measured K_D values to a specific AT-hook. But it can be claimed that two AT-hooks show K_D values of 10–40 nM and a third AT-hook in the submicromolar range (~0.1 μM). After phosphorylation of the C-terminus of HMGA1a, the affinity of two AT-hooks is reduced by one order of magnitude to the submicromolar range (0.2–0.6 μM) and the remaining one by even three orders of magnitude to the submillimolar range (approx. 0.5 mM) (Figure 6B).

DISCUSSION

NMR-based characterisation of the structural ensemble of the intrinsically disordered protein HMGA1a

Here, we present the first NMR-based model for full-length HMGA1a and shed light on the impact of its post-translational modifications, especially phosphorylation, on structural and functional changes. HMGA1a consists of four regions, including three AT-hooks and a negatively charged C-terminus. Our data reveal that HMGA1a is not totally random coil but rather adopts a complex structural ensemble. The analysis of the calculated structural ensemble based on our chemical shifts as well as R_2 relaxation data, PRE-profiles, and ^3J couplings shows that HMGA1a adopts different transient secondary and tertiary structures.

To obtain information about secondary structures we measured chemical shifts for $HMGA1a_{S64C}$, which revealed that most of its regions are random coiled. Interestingly, our study showed for the first time that one of the DNA-binding domains (DBDs), AT1, transiently adopts a structured α-helical conformation. Additionally, the C-terminus of HMGA1a shows an extended and/or β-sheet-like conformation.

To acquire tertiary structure information of HMGA1a, we performed PRE experiments with spin-labels attached to different sequence positions. Our analysis shows that the glutamate-rich negatively charged C-terminus is in transient contact with the second and the third AT-hook (Figure 3B: clusters 3, 4 and 5). Thus, HMGA1a loops back on itself, bringing its C-terminus and the second as well as the third AT-hook together. Our results provide a possible structural mechanism on how post-translational modification of the C-terminus influences the DNA-binding properties of HMG-proteins (45,46).

Phosphorylation of HMGA1a has been extensively studied during the last decade, but its structural effects on the IDP are still not fully understood (38). On the one hand, phosphorylation of HMGA1a by CK2 has been identified as a constitutively PTM pattern present in insect HMGA1a. On the other hand, different studies showed that the exposition of B-lymphocytes to the cytokine interleukin-4 leads to a phosphorylation of HMGA1a by CK2 after 15 min (62,63). This implies that this type of phosphorylation plays a crucial role during the regulation of HMGA1a. Functional studies showed that the C-terminal phosphorylation of HMGA1b leads to a three-fold reduction of the binding affinity towards DNA (45). Therefore, we took a closer look at the impact of phosphorylation by different kinases on the structural properties of HMGA1a.

Phosphorylation orchestrates the structural ensemble of the intrinsically disordered protein HMGA1a

Our structure ensemble reveals that contacts frequently occur between regions or adjacent regions that are phosphorylated by different kinases, e.g. cdc2 and CK2. Taking into account that contacts in IDPs are often mediated by polar and less by hydrophobic amino acids, this observation could explain how the structural ensemble of HMGA1a is regulated (64). The cluster analysis shows that more compact conformers of HMGA1a exhibit contacts in the vicinity of the CK2 and cdc2 phosphorylation sites located at AT2 and AT3 (see Cluster 4). These contacts could play a crucial role in the regulation of HMGA1a, because phosphorylation at these sites would alter the electrostatic interactions and possibly shift the structural ensemble towards different, either more open or compact conformers. Our data indicate that HMGA1a is regulated by an electrostatic contact network, which regulates the structural ensemble of the IDP (see Figure 7B).

To validate this model, we investigated the effect of the phosphorylation of HMGA1a by CK2 and/or cdc2. The overlay of the 2D [^{1}H, ^{15}N] HSQC NMR spectra clearly shows that phosphorylation at different sites induces strong chemical shift perturbations for both forms. Further, these effects do not occur only locally, but also affect residues

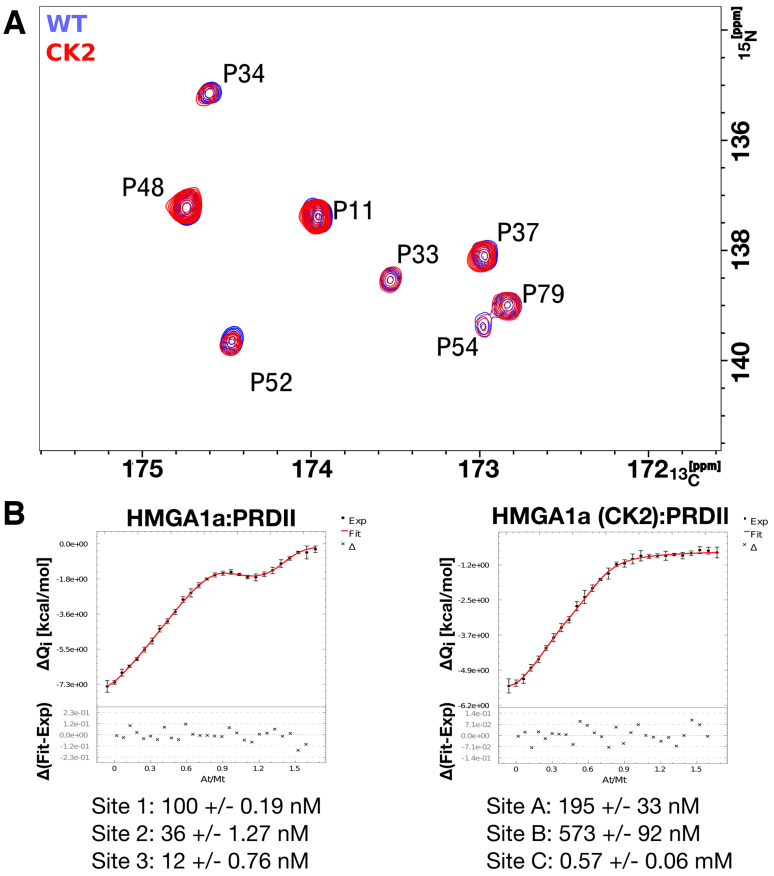

Figure 6. (**A**) 2D [^{15}N,^{13}C] CON spectra of HMGA1a wt (red) and CK2-form (blue) of HMGA1a in complex with the PRD2-element of the NFκB promoter DNA. (**B**) ITC-Data of HMGA1a and CK2-phosphorylated HMGA1a interacting with the PRD2 element of the NFκB promoter DNA.

that are located far apart in the amino acid sequence of HMGA1a (Figure 4A). Our data suggest that phosphorylation by CK2 (at S99, S102, and S103) strengthens the contact between the C-terminus and the region ranging from amino acids 71 to 80. In addition, this phosphorylation leads to a more extended/β-sheet-like structure of the C-terminus.

Interestingly, chemical shift perturbations of C-terminal residues reveal that the C-terminus of HMGA1a is also affected by phosphorylation at the cdc2 phosphorylation sites (i.e. S36, T53, T79). This observation correlates very well with our structural ensemble, which displays contacts between the phosphorylation sites T53 as well as T79 and the C-terminus. This clearly shows that PTMs can modulate the structural ensemble of HMGA1a, corroborating previously predicted potential structural effects of phosphorylation on SPXX and TPKK/R motifs (65). Our study clearly reveals that transient internal long-range contacts of intrinsically disordered HMGA1a play a crucial role in the modulation and regulation of the entire structural ensemble, with severe consequences for the DNA binding affinity of HMGA1a.

Phosphorylation of HMGA1a modulates its DNA binding to the NFκB PRD2 promoter element

Our ITC binding studies reveal that phosphorylation of HMGA1a by CK2 greatly reduces the binding affinity of all AT hooks for the PRD2 promoter DNA element. For other HMG-group proteins like HMG1-d, it was shown that the negatively charged C-terminus plays a crucial role for DNA binding (66). However, it was unclear as to why introducing negative charges at the C-terminus would affect the binding properties of HMGA1a. We are now able to explain the molecular basis of this long-range structural effect in the IDP protein HMGA1a upon phosphorylation of its C-terminus through CK2 and suggest a structural model that is in full concordance with our ITC and NMR data. In fact, we could show by PRE-experiments that phosphorylation of S99, S102, and S103 enhances contacts between the C-terminus and the region ranging from amino acids 71 to 80 of HMGA1a and that phosphorylation by CK2 even increases this effect. This region is lysine-rich and the positive amine group of its side chain can efficiently coordinate negative phosphate groups. Interestingly, similar contacts were identified for the phosphorylated C-terminus of HMGA2 by proteolysis and IMS-MS analysis (46).

This observation links the C-terminus to DNA binding regions and can explain the influence of phosphorylation of the C-terminus on the DNA binding affinity of AT2 and AT3 to the PRD2 element of the NFκB promoter. Our structural model of HMGA1a is based on the structural ensemble clusters which adopt more open or closed conformations. For example, in clusters 1 and 2, all AT-hooks

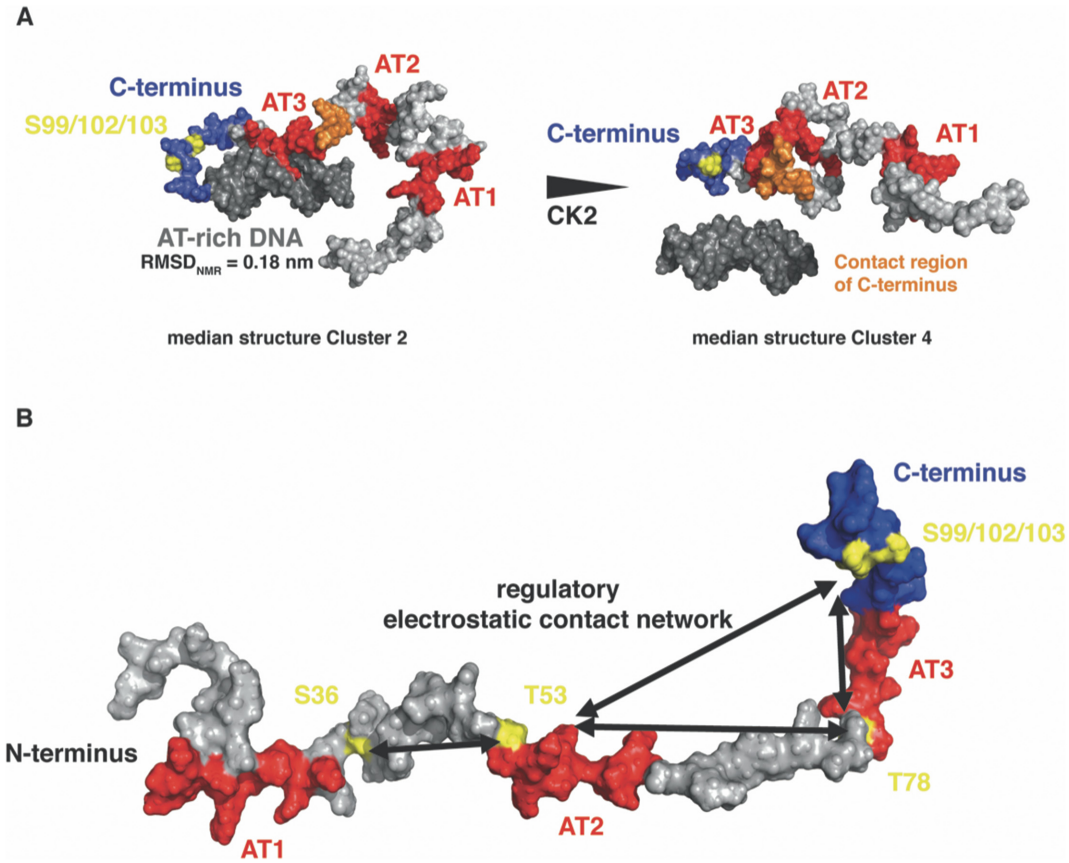

Figure 7. (**A**) NMR ensemble-based structural model for the impact of C-terminal phosphorylation of HMGA1a by CK2 on the binding affinity towards an AT-rich B-DNA duplex. (**B**) Regulatory electrostatic contact network of HMGA1a. Arrows indicate long range contacts identified by PRE analysis between internal phosphorylation sites (CK2, cdc2).

are exposed and thus more accessible to DNA. To illustrate this, we modelled the structure of cluster 2 docked to AT-rich DNA as shown in Figure 7A. Here, the AT-hook region was aligned with the published NMR structure of AT3 (21). The RMSD of the central binding motif of the third AT-hook is 1.8 Å, which indicates that the AT-hook in the median structure of cluster 2 indeed adopts a very similar conformation to the DNA-bound form. Presumably, this suggests a mechanism of conformational selection for the HMGA1a/DNA complex.

Furthermore, our cluster analysis of the non-phosphorylated HMGA1a ensemble reveals that clusters 3 and 4 contain more closed conformers whose C-terminus contact AT2 and/or AT3 as well as the linker region connecting AT2 and AT3 (Figure 3B).

A closer look at more closed HMGA1a conformers like clusters 3 and 4 sheds light on the mechanism of impaired DNA binding. The CK2-mediated stabilisation of contacts of the C-terminus to the region between AT2 and AT3 probably increases the number of closed conformers in the HMGA1a ensemble and impairs DNA binding by steric interference (see Figure 7A). This model is in good agreement with our observation that the C-terminal phosphorylation of HMGA1a does not influence the final binding mode of AT-hooks to the PRD2 element of the NFκB promoter (Figure 6A). We can conclude that reduced DNA-binding affinity is induced by perturbation of the structural ensemble towards a more compact, binding-impaired conformation as found in cluster 4 (Figure 3B). Evidently, the C-terminus of HMGA1a is crucial for the regulation of the IDP.

Interestingly, the identified region between AT2 and AT3 contains several lysine residues which can be acetylated by PCAF, which implies another important, yet complementary mechanism of regulation for the compaction of HMGA1a's C-terminus (67). Acetylation of this region would alter the electrostatic properties and could counteract the effect of phosphorylation. This proposed model of regulation shall be explored in future.

CONCLUSION

In this study, we present the first NMR-based structural characterization of free HMGA1a in solution. Our data show that HMGA1a is not a totally random coil polypeptide but adopts a tertiary structure that is based on transient long-range contacts and, as such, it bears all the traits of an IDP (Figure 7B). These contacts appear to play a crucial role during the regulation of HMGA1a by PTMs, such as phosphorylation. Based on our NMR data we present a mechanism for the impact of phosphorylation of HMGA1a on the structural ensemble. (Figure 7A,

B). We could show that CK2-phosphorylation of the C-terminus impairs DNA binding by steric interference. This model is fully corroborated by ITC-based promoter DNA binding studies of both wild-type and CK2-phosphorylated HMGA1a. Additionally, we present data on the structural properties of the first AT-hook. Presumably, this much improved structural understanding of full length wild-type and CK2-phosphorylated HMGA1a will ultimately even contribute to the development of new therapeutic strategies against HMGA1a-associated types of cancer.

DATA AVAILABILITY

Assignments for $HMGA1a_{S64C}$ and CK2-phosphorylated $HMGA1a_{S64C}$ have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under accession numbers 27883 and 27884, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

B.K. was supported by the RUB Research School^{Plus}. R.S. gratefully recognizes support by the DFG.

FUNDING

DFG [INST 213/757-1 FUGG and INST 213/843-1 FUGG]. Funding for open access charge: DFG Open Access Publication Funds of the Ruhr University of Bochum, Germany.
Conflict of interest statement. None declared.

REFERENCES

1. Arai,M., Sugase,K., Dyson,H.J. and Wright,P.E. (2015) Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 9614–9619.
2. Wright,P.E. and Dyson,H.J. (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.*, **16**, 18–29.
3. Theillet,F.-X., Kalmar,L., Tompa,P., Han,K.-H., Selenko,P., Dunker,A.K., Daughdrill,G.W. and Uversky,V.N. (2014) The alphabet of intrinsic disorder. *Intrinsically Disord. Proteins*, **1**, e24360.
4. Dunker,A.K., Brown,C.J., Lawson,J.D., Iakoucheva,L.M. and Obradovic,Z. (2002) Intrinsic disorder and protein function. *Biochemistry-US*, **41**, 6573–6582.
5. Dunker,A.K., Lawson,J.D., Brown,C.J., Williams,R.M., Romero,P., Oh,J.S., Oldfield,C.J., Campen,A.M., Ratliff,C.M., Hipps,K.W. *et al.* (2001) Intrinsically disordered protein. *J. Mol. Graph. Model.*, **19**, 26–59.
6. Eilebrecht,S., Wilhelm,E., Benecke,B.J., Bell,B. and Benecke,A.G. (2013) HMGA1 directly interacts with TAR to modulate basal and Tat-dependent HIV transcription. *RNA Biol.*, **10**, 436–444.
7. Iakoucheva,L.M., Brown,C.J., Lawson,J.D., Obradovic,Z. and Dunker,A.K. (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.*, **323**, 573–584.
8. Uversky,V.N., Oldfield,C.J. and Dunker,A.K. (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.*, **37**, 215–246.
9. Sivakolundu,S.G., Bashford,D. and Kriwacki,R.W. (2005) Disordered p27Kip1 exhibits intrinsic structure resembling the Cdk2/cyclin A-bound conformation. *J. Mol. Biol.*, **353**, 1118–1128.
10. Wells,M., Tidow,H., Rutherford,T.J., Markwick,P., Jensen,M.R., Mylonas,E., Svergun,D.I., Blackledge,M. and Fersht,A.R. (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 5762–5767.
11. Jensen,M.R., Communie,G., Ribeiro,E.A. Jr, Martinez,N., Desfosses,A., Salmon,L., Mollica,L., Gabel,F., Jamin,M., Longhi,S. *et al.* (2011) Intrinsic disorder in measles virus nucleocapsids. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 9839–9844.
12. Mittag,T., Marsh,J., Grishaev,A., Orlicky,S., Lin,H., Sicheri,F., Tyers,M. and Forman-Kay,J.D. (2010) Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. *Structure*, **18**, 494–506.
13. Schwalbe,M., Ozenne,V., Bibow,S., Jaremko,M., Jaremko,L., Gajda,M., Jensen,M.R., Biernat,J., Becker,S., Mandelkow,E. *et al.* (2014) Predictive atomic resolution descriptions of intrinsically disordered hTau40 and α-synuclein in solution from NMR and small angle scattering. *Structure*, **22**, 238–249.
14. Mukrasch,M.D., Bibow,S., Korukottu,J., Jeganathan,S., Biernat,J., Griesinger,C., Mandelkow,E. and Zweckstetter,M. (2009) Structural polymorphism of 441-residue tau at single residue resolution. *PLoS Biol.*, **7**, e34.
15. Ozenne,V., Bauer,F., Salmon,L., Huang,J.R., Jensen,M.R., Segard,S., Bernadó,P., Charavay,C. and Blackledge,M. (2012) Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics*, **28**, 1463–1470.
16. Marsh,J.A., Neale,C., Jack,F.E., Choy,W.Y., Lee,A.Y., Crowhurst,K.A. and Forman-Kay,J.D. (2007) Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure. *J. Mol. Biol.*, **367**, 1494–1510.
17. Marsh,J.A. and Forman-Kay,J.D. (2009) Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. *J. Mol. Biol.*, **391**, 359–374.
18. Marsh,J.A., Dancheck,B., Ragusa,M.J., Allaire,M., Forman-Kay,J.D. and Peti,W. (2010) Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators. *Structure*, **18**, 1094–1103.
19. Uversky,V.N. and Dunker,A.K. (2012) Multiparametric analysis of intrinsically disordered proteins: looking at intrinsic disorder through compound eyes. *Anal. Chem.*, **84**, 2096–2104.
20. Uversky,V.N., Oldfield,C.J. and Dunker,A.K. (2016) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recogn.*, **18**, 343–384.
21. 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.
22. Claus,P., Schulze,E. and Wisniewski,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.
23. Subramanian,D. and Griffith,J.D. (2002) Interactions between p53, hMSH2-hMSH6 and HMG I(Y) on Holliday junctions and bulged bases. *Nucleic Acids Res.*, **30**, 2427–2434.
24. Reeves,R. and Beckerbauer,L. (2001) HMGI/Y proteins: flexible regulators of transcription and chromatin structure. *Biochim. Biophys. Acta*, **1519**, 13–29.
25. Reeves,R. (2001) Molecular biology of HMGA proteins: hubs of nuclear function. *Gene*, **277**, 63–81.
26. Reeves,R. and Beckerbauer,L.M. (2003) HMGA proteins as therapeutic drug targets. *Prog Cell Cycle Res.*, **5**, 279–286.
27. Reeves,R. (2010) Nuclear functions of the HMG proteins. *Biochim. Biophys. Acta*, **1799**, 3–14.
28. Friedmann,M., Holth,L.T., Zoghbi,H.Y. and Reeves,R. (1993) Organization, inducible-expression and chromosome localization of the human HMG-I(Y) nonhistone protein gene. *Nucleic Acids Res.*, **21**, 4259–4267.
29. Nagpal,S., Ghosn,C., DiSepio,D., Molina,Y., Sutter,M., Klein,E.S. and Chandraratna,R.A. (1999) Retinoid-dependent recruitment of a

histone H1 displacement activity by retinoic acid receptor. *J. Biol. Chem.*, **274**, 22563–22568.
30. Johnson,K.R., Lehn,D.A. and Reeves,R. (1989) Alternative processing of mRNAs encoding mammalian chromosomal high-mobility-group proteins HMG-I and HMG-Y. *Mol. Cell Biol.*, **9**, 2114–2123.
31. Chiappetta,G., Botti,G., Monaco,M., Pasquinelli,R., Pentimalli,F., Di Bonito,M., D'Aiuto,G., Fedele,M., Iuliano,R., Palmieri,E.A. *et al.* (2004) HMGA1 protein overexpression in human breast carcinomas: correlation with ErbB2 expression. *Clin. Cancer Res.*, **10**, 7637–7644.
32. Tamimi,Y., van der Poel,H.G., Denyn,M.M., Umbas,R., Karthaus,H.F., Debruyne,F.M. and Schalken,J.A. (1993) Increased expression of high mobility group protein I(Y) in high grade prostatic cancer determined by in situ hybridization. *Cancer Res.*, **53**, 5512–5516.
33. Chiappetta,G., Manfioletti,G., Pentimalli,F., Abe,N., Di Bonito,M., Vento,M.T., Giuliano,A., Fedele,M., Viglietto,G., Santoro,M. *et al.* (2016) High mobility group HMGI(Y) protein expression in human colorectal hyperplastic and neoplastic diseases. *Int. J.Cancer*, **91**, 147–151.
34. Fedele,M., Pierantoni,G.M., Berlingieri,M.T., Battista,S., Baldassarre,G., Munshi,N., Dentice,M., Thanos,D., Santoro,M., Viglietto,G. *et al.* (2001) Overexpression of proteins HMGA1 induces cell cycle deregulation and apoptosis in normal rat thyroid cells. *Cancer Res.*, **61**, 4583–4590.
35. van der Zee,J.A., ten Hagen,T.L., Hop,W.C., van Dekken,H., Dicheva,B.M., Seynhaeve,A.L., Koning,G.A., Eggermont,A.M. and van Eijck,C.H. (2010) Differential expression and prognostic value of HMGA1 in pancreatic head and periampullary cancer. *Eur. J. Cancer*, **46**, 3393–3399.
36. 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.
37. Fonfria-Subiros,E., Acosta-Reyes,F., Saperas,N., Pous,J., Subirana,J.A. and Campos,J.L. (2012) Crystal structure of a complex of DNA with one AT-hook of HMGA1. *PLoS One*, **7**, e37120.
38. Zhang,Q. and Wang,Y. (2008) High mobility group proteins and their post-translational modifications. *Biochim. Biophys. Acta*, **1784**, 1159–1166.
39. Palvimo,J. and Linnala-Kankkunen,A. (1989) Identification of sites on chromosomal protein HMG-I phosphorylated by casein kinase II. *FEBS Lett.*, **257**, 101–104.
40. Nissen,M.S., Langan,T.A. and Reeves,R. (1991) Phosphorylation by cdc2 kinase modulates DNA binding activity of high mobility group I nonhistone chromatin protein. *J. Biol. Chem.*, **266**, 19945–19952.
41. Zhang,Q. and Wang,Y. (2007) Homeodomain-interacting protein kinase-2 (HIPK2) phosphorylates HMGA1a at Ser-35, Thr-52, and Thr-77 and modulates its DNA binding affinity. *J. Proteome Res.*, **6**, 4711–4719.
42. Reeves,R., Langan,T.A. and Nissen,M.S. (1991) Phosphorylation of the DNA-binding domain of nonhistone high-mobility group I protein by cdc2 kinase: reduction of binding affinity. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 1671–1675.
43. Schwanbeck,R. and Wisniewski,J.R. (1997) Cdc2 and mitogen-activated protein kinases modulate DNA binding properties of the putative transcriptional regulator Chironomus high mobility group protein I. *J. Biol. Chem.*, **272**, 27476–27483.
44. Xiao,D.M., Pak,J.H., Wang,X., Sato,T., Huang,F.L., Chen,H.C. and Huang,K.P. (2000) Phosphorylation of HMG-I by protein kinase C attenuates its binding affinity to the promoter regions of protein kinase C gamma and neurogranin/RC3 genes. *J. Neurochem.*, **74**, 392–399.
45. Piekielko,A., Drung,A., Rogalla,P., Schwanbeck,R., Heyduk,T., Gerharz,M., Bullerdiek,J. and Wiśniewski,J.R. (2001) Distinct organization of DNA complexes of various HMGI/Y family proteins and their modulation upon mitotic phosphorylation. *J. Biol. Chem.*, **276**, 1984–1992.
46. Maurizio,E., Cravello,L., Brady,L., Spolaore,B., Arnoldo,L., Giancotti,V., Manfioletti,G. and Sgarra,R. (2011) Conformational role for the C-terminal tail of the intrinsically disordered high mobility group A (HMGA) chromatin factors. *J. Proteome Res.*, **10**, 3283–3291.
47. Marsh,J.A. and Forman-Kay,J.D. (2016) Ensemble modeling of protein disordered states: Experimental restraint contributions and validation. *Proteins: Struct. Funct. Bioinformatics*, **80**, 556–572.
48. Reeves,R. (2004) HMGA proteins: isolation, biochemical modifications, and nucleosome interactions. *Methods Enzymol.*, **375**, 297–322.
49. Bermel,W., Bertini,I., Duma,L., Felli,I.C., Emsley,L., Pierattelli,R. and Vasos,P.R. (2016) Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. *Angew. Chem.*, **117**, 3149–3152.
50. Lescop,E., Schanda,P. and Brutscher,B. (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson.*, **187**, 163–169.
51. Buchko,G.W., Ni,S., Lourette,N.M., Reeves,R. and Kennedy,M.A. (2007) NMR resonance assignments of the human high mobility group protein HMGA1. *J. Biomol. NMR*, **38**, 185.
52. Yuwen,T. and Skrynnikov,N.R. (2014) CP-HISQC: a better version of HSQC experiment for intrinsically disordered proteins under physiological conditions. *J. Biomol. NMR*, **58**, 175–192.
53. Chignola,F., Mari,S., Stevens,T.J., Fogh,R.H., Mannella,V., Boucher,W. and Musco,G. (2011) The CCPN Metabolomics Project: a fast protocol for metabolite identification by 2D-NMR. *Bioinformatics*, **27**, 885–886.
54. Vranken,W.F., Boucher,W., Stevens,T.J., Fogh,R.H., Pajon,A., Llinas,M., Ulrich,E.L., Markley,J.L., Ionides,J. and Laue,E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
55. Skinner,S.P., Fogh,R.H., Boucher,W., Ragan,T.J., Mureddu,L.G. and Vuister,G.W. (2016) CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *J. Biomol. NMR*, **66**, 111–124.
56. Battiste,J.L. and Wagner,G. (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry-US*, **39**, 5355–5365.
57. Jamroz,M. and Kolinski,A. (2013) ClusCo: clustering and comparison of protein models. *BMC Bioinformatics*, **14**, 62.
58. Eva Muñoz,A.P. (2018) AFFINImeter Software: from its Beginnings to Future Trends-A Literature review. *J. Appl. Bioanal.*, **4**, 124–139.
59. Tamiola,K. and Mulder,F.A. (2012) Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. *Biochem. Soc . Trans.*, **40**, 1014–1020.
60. Sjodt,M. and Clubb,R.T. (2017) Nitroxide labeling of proteins and the determination of paramagnetic relaxation derived distance restraints for NMR studies. *Biol. Protoc.*, **7**, e2207.
61. Bienkiewicz,E.A. and Lumb,K.J. (1999) Random-coil chemical shifts of phosphorylated amino acids. *J. Biomol. NMR*, **15**, 203–206.
62. Wang,D.Z., Zamorano,J., Keegan,A.D. and Boothby,M. (1997) HMG-I(Y) phosphorylation status as a nuclear target regulated through insulin receptor substrate-1 and the I4R motif of the interleukin-4 receptor. *J. Biol. Chem.*, **272**, 25083–25090.
63. Wang,D.Z., Ray,P. and Boothby,M. (1995) Interleukin 4-Inducible phosphorylation of Hmg-I(Y) is inhibited by Rapamycin. *J. Biol. Chem.*, **270**, 22924–22932.
64. Marsh,J.A. and Forman-Kay,J.D. (2010) Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.*, **98**, 2383–2390.
65. Churchill,M.E. and Travers,A.A. (1991) Protein motifs that recognize structural features of DNA. *Trends Biochem. Sci.*, **16**, 92–97.
66. Dow,L.K., Jones,D.N., Wolfe,S.A., Verdine,G.L. and Churchill,M.E. (2000) Structural studies of the high mobility group globular domain and basic tail of HMG-D bound to disulfide Cross-Linked DNA. *Biochemistry-US*, **39**, 9725–9736.
67. 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.