Dynamic Autoinhibition of the HMGB1 Protein via Electrostatic Fuzzy Interactions of Intrinsically Disordered Regions

Xi Wang [1], Harry M. Greenblatt [2], Lavi S. Bigman [2], Binhan Yu [1], Channing C. Pletka [1], Yaakov Levy [2*] and Junji Iwahara [1*]

*1 - **Department of Biochemistry and Molecular Biology,** Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, TX 77555-1068, USA*

*2 - **Department of Chemical and Structural Biology,** Weizmann Institute of Science, Rehovot 76100, Israel*

***Correspondence to Yaakov Levy and Junji Iwahara:** Koby.Levy@weizmann.ac.il (Y. Levy), j.iwahara@utmb.edu (J. Iwahara)*

https://doi.org/10.1016/j.jmb.2021.167122

Edited by Richard W. Kriwacki

Abstract

Highly negatively charged segments containing only aspartate or glutamate residues ("D/E repeats") are found in many eukaryotic proteins. For example, the C-terminal 30 residues of the HMGB1 protein are entirely D/E repeats. Using nuclear magnetic resonance (NMR), fluorescence, and computational approaches, we investigated how the D/E repeats causes the autoinhibition of HMGB1 against its specific binding to cisplatin-modified DNA. By varying ionic strength in a wide range (40–900 mM), we were able to shift the conformational equilibrium between the autoinhibited and uninhibited states toward either of them to the full extent. This allowed us to determine the macroscopic and microscopic equilibrium constants for the HMGB1 autoinhibition at various ionic strengths. At a macroscopic level, a model involving the autoinhibited and uninhibited states can explain the salt concentration-dependent binding affinity data. Our data at a microscopic level show that the D/E repeats and other parts of HMGB1 undergo electrostatic fuzzy interactions, each of which is weaker than expected from the macroscopic autoinhibitory effect. This discrepancy suggests that the multivalent nature of the fuzzy interactions enables strong autoinhibition at a macroscopic level despite the relatively weak intramolecular interaction at each site. Both experimental and computational data suggest that the D/E repeats interact preferentially with other intrinsically disordered regions (IDRs) of HMGB1. We also found that mutations mimicking post-translational modifications relevant to nuclear export of HMGB1 can moderately modulate DNA-binding affinity, possibly by impacting the autoinhibition. This study illuminates a functional role of the fuzzy interactions of D/E repeats.

Introduction

Eukaryotic gene regulation at a transcriptional level involves various DNA-binding proteins such as transcription factors, histones, and other architectural proteins in the nuclei. The majority of these proteins contain intrinsically disordered regions (IDRs) to a remarkable extent.[1] For example, the average total size of IDRs in human transcription factors is 50% of the total length of protein sequence.[2] Over the past two decades, it has been shown that IDRs can regulate functions of gene-regulatory proteins through various mechanisms, including post-translational modifications, protein-protein interactions, and intramolecular interactions.[1,3]

Some IDRs interact with DNA-binding domains within the same protein molecule.[4–19] Such

intramolecular interactions can cause autoinhibition and reduce the apparent binding affinity of gene-regulatory proteins for DNA. In fact, there are many reports on IDR-mediated autoinhibition for various gene-regulatory proteins, including C/EBPβ,[9] Ets-1,[4,11,12] FoxO3,[17] HMGB1,[7,14,18] p53,[5,8,15] RFX1,[6] Sox11,[19] UBF1,[16] and Ultrabithorax.[10] In some cases, autoinhibition is modulated by post-translational modifications (PTMs) such as phosphorylation.[4,9,11,12,15,20]

For some gene-regulatory proteins, autoinhibition is imposed by strongly negatively charged segments that contain only aspartate (D) or glutamate (E) residues.[6,13,14,16,18,19] These segments, which we refer to as "D/E repeats", are found among various protein families of transcription factors. Although some researchers suggested that D/E repeats may serve as DNA mimic,[21,22] formation of a DNA-like shape seems unlikely. Polyglutamate can form an α-helix at pH 4 or lower, but is predominantly a random coil at neutral pH.[23,24] The presence of aspartate residues in an negatively charged polypeptide seems to strengthen the propensity of a random coil.[25] Acting as strongly negatively charged IDRs, the D/E repeats may electrostatically inhibit the binding of the proteins to DNA in a dynamic fashion.

The human HMGB1 protein contains a 30-residue segment of D/E repeats at the C-terminus (the residues 186–215)[1]. In the cell nuclei, HMGB1 serves as a DNA chaperone and promotes activities of various DNA-binding proteins such as transcription factors, DNA-repair/recombination enzymes, and chromatin remodeling factors.[26–28] Although the biological role of the HMGB1 autoinhibition remains to be delineated, the competitive intramolecular interactions with the D/E repeats that destabilize interactions with DNA may be important for dynamic action of HMGB1 as a DNA chaperone.[26] In fact, the removal of the D/E repeats from the HMGB1 protein severely diminishes its activity to facilitate nucleosome remodeling.[29] It was also proposed that the D/E repeats play a role in displacement of histone H1 from linker DNA by HMGB1, a dynamic process that impacts the accessibility of chromatin.[30,31] Previous studies using nuclear magnetic resonance (NMR) and small-angle X-ray scattering (SAXS) showed that the D/E repeats of HMGB1 interact with its two DNA-binding domains (A-box, residues 9–79; B-box, residues 95–163) and create dynamic equilibrium between extended and collapsed structures.[14,18] However, thermodynamics of the conformational equilibrium and its relevance to specific interactions of HMGB1 with its target molecules remain to be delineated.

In this work, using experimental and computational approaches, we quantitatively investigated how the autoinhibition of HMGB1 occurs and impacts its specific interactions with a cisplatin-modified DNA duplex. By varying the salt concentration in a wide range, we were able to shift the equilibrium between the uninhibited and autoinhibited states toward either of them to the full extent. This allowed us to gain quantitative information about the conformational equilibrium of autoinhibition and its impact on the specific binding to distorted DNA. Our data show that the ensemble of relatively weak interactions between the D/E repeats and other parts of HMGB1 causes strong autoinhibition against specific binding to distorted DNA. Our experimental and computational data also show that in the autoinhibition process, the D/E repeats prefer positively charged IDRs over the two folded DNA-binding domains of HMGB1. Furthermore, we show that mutations which mimic post-translational modifications relevant to nuclear export of HMGB1 moderately modulate the DNA-binding affinity. Our current study on HMGB1 illuminates a functional role of the electrostatic fuzzy interactions between IDRs.

Results

Salt-concentration dependent autoinhibition against binding to DNA

HMGB1 is known to strongly bind to non-B-form DNA such as Holliday junction,[32,33] bulged DNA,[34,35] G-quadruplex,[36,37] and cisplatin-modified DNA.[38,39] Using a fluorescein amidite (FAM)-labeled 20-bp DNA duplex containing a cisplatin-modification at a GG dinucleotide (Figure 1(A)), we measured the specific DNA-binding affinities of the full-length HMGB1 protein and its Δ30 variant, which lacked the D/E repeats of the C-terminal 30 residues. The apparent dissociation constants ($K_{d,app}$) were determined from FAM fluorescence anisotropy data for the cisplatin-modified DNA duplex at various concentrations of the proteins. Figures 1(B) shows examples of the binding isotherm data from fluorescence anisotropy measurements at 300 and 800 mM KCl. Using this method, we determined the $K_{d,app}$ constants for the full-length HMGB1 and Δ30 variant proteins at various concentrations of KCl (Figure 1(C)). When the ionic strength was relatively low, the affinity of the Δ30 variant lacking the D/E repeats was significantly higher than that of the full-length HMGB1 protein. The lower affinity of the full-length HMGB1 protein represents autoinhibition via the D/E repeats that may compete with DNA. Interestingly, at KCl concentrations greater than 500 mM, the

1 In some literature on HMGB1, a residue-numbering scheme starting from the actual amino-terminal glycine (i.e., G1-K2-...) is used, whereas in others, a scheme starting from the initial methionine in the gene (i.e., M1-G2-K3-...) is used. In this paper, we use the latter.

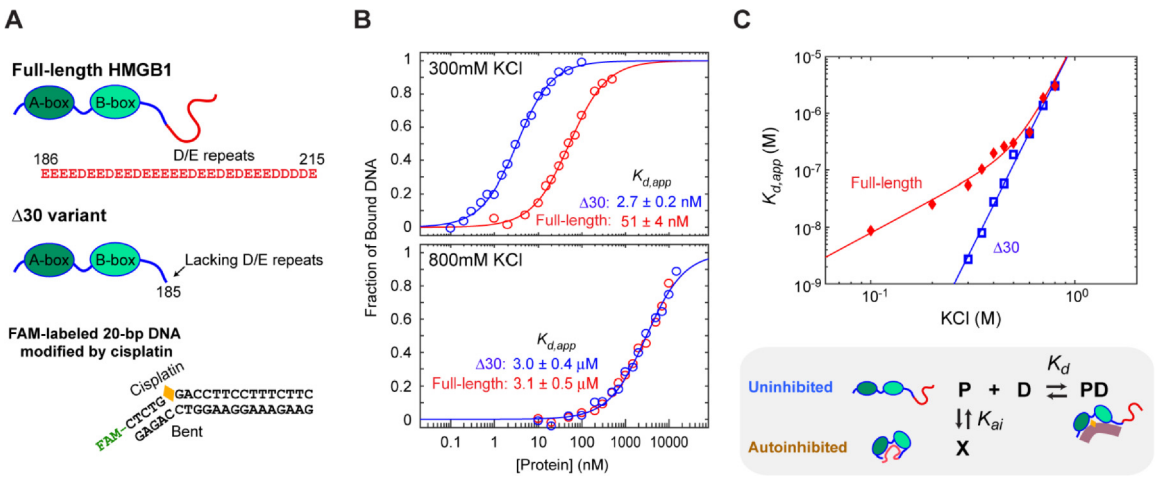

Figure 1. Salt concentration-dependence of HMGB1 autoinhibition examined through measurements of affinity for a cisplatin-modified 20-bp DNA duplex. (**A**) The full-length and Δ30 HMGB1 constructs and the FAM-labeled 20-bp DNA duplex modified by cisplatin at a GG dinucleotide step. (**B**) Binding isotherm data obtained through measurements of fluorescence anisotropy at various concentrations of proteins. Data at 300 mM and 800 mM KCl are shown. Data for the full-length HMGB1 protein and the Δ30 variant are shown in red and blue, respectively. (**C**) Logarithmic plots of the dissociation constant K_d for the complexes of the cisplatin-modified DNA and the proteins measured at various concentrations of KCl. Data for the full-length HMGB1 protein and the Δ30 variant are shown in red and blue, respectively. For the Δ30 variant data, the straight solid line was obtained through fitting with Eq. (1). For the full-length HMGB1, the solid curve was obtained through fitting with Eqs. (3) and (4) along with Eq. (1) parameterized with the Δ30 variant data. Through this fitting calculation, the equilibrium constant K_{ai} at 150 mM KCl was determined to be $(8.3 \pm 1.4) \times 10^2$.

autoinhibition disappeared and the affinities of the Δ30 variant and the full-length HMGB1 protein became virtually identical.

To understand why the autoinhibition by the D/E repeats disappear at high concentrations of KCl, we quantitatively analyzed the salt concentration dependence of the dissociation constants. For many protein-DNA complexes, the following empirical relation has been confirmed:[40]

$$\log K_d = s \log [M^+] + \log K_{d,1M} \quad (1)$$

where $[M^+]$ is the cation concentration and $K_{d,1M}$ is the dissociation constant at $[M^+]$ = 1 mol/L. This relation also means that a linear relationship exists between the binding free energy ΔG and log[KCl], which is supported by the counterion condensation theory.[41] Our experimental data for the Δ30 variant clearly show the linear relationship between logK_d and log[KCl] (Figure 1(C)).

In contrast, the full-length HMGB1 protein exhibited a significant deviation from the linear relationship between logK_d and log[KCl], as seen in Figure 1(C). Electrostatically driven autoinhibition can account for this deviation. The full-length protein in the free state undergoes dynamic equilibrium between the autoinhibited state (X) and the uninhibited state (P). In the autoinhibited state, the D/E repeats electrostatically interact with the positively charged regions of HMGB1 and inhibit DNA-binding, whereas in the uninhibited state, the D/E repeats are dissociated and allow the protein to bind to DNA. Here we define the equilibrium constant K_{ai} for the conformational equilibrium of autoinhibition as:

$$K_{ai} = [X]/[P]. \quad (2)$$

With this equilibrium constant, the apparent dissociation constant $K_{d,app}$ is given by:

$$K_{d,app} = (K_{ai} + 1)K_d \quad (3)$$

Since the autoinhibition involving the D/E repeats should occur primarily via electrostatic interactions, it is reasonable to consider that the free energy difference between the X and P states is also a linear function of log [KCl], as predicted by the counterion condensation theory on linear polyelectrolytes. This assumption leads to:

$$\log K_{ai} = a \log [KCl] + b \quad (4)$$

The population of the autoinhibited state X is lower at a higher concentration of KCl, and therefore the slope a should be a negative number. When $K_{ai} \ll 1$ and the autoinhibition is insignificant (i.e., $[X] \ll [P]$), $K_{d,app}$ becomes virtually identical to K_d. This occurs when the salt concentration is high. When $K_{ai} \gg 1$ and the autoinhibition is considerably strong, $\log K_{d,app} \approx \log K_{ai} + \log K_d$ and therefore the slope $\partial \log K_{d,app} / \partial \log [M^+]$ can be approximated by $s + a$, which is smaller than s due to the negative value of a. In fact, for the full-length HMGB1 protein, the slope is gentle at lower concentrations of KCl and

becomes as steep as that for the Δ 30 variant at higher concentrations of KCl (see Figure 1(C)).

For the full-length HMGB1 data, we performed a nonlinear least-squares fitting using Eqs. (1), (3), and (4), in which the parameters *a* and *b* (Eq. (4)) were optimized while the parameters for Eq. (1) were kept fixed at the values obtained for the Δ30 variant. The best-fit curves for the full-length HMGB1 and the Δ30 variant are shown in solid lines in Figure 1(C). The slope parameters *s* and *a* were determined to be 7.2 and −5.4, respectively. According to the counterion condensation theory, the absolute values of these slopes correspond to the number of ions that are thermodynamically released upon macromolecular association.[40,41] The value of the parameter *s* seems qualitatively consistent with the total number of interfacial ion pairs in three-dimensional structures of HMGB1-DNA complexes available in Protein Data Bank (e.g., 4 ion pairs for A-box in 1CKT; 4 ion pairs for B-box in 2GZK). Interpretation of the parameter *a* is not straightforward because both counterion accumulation and coion exclusion can significantly contribute to charge neutralization of proteins.[42–44] From the determined parameters and Eq. (4), the equilibrium constant K_{ai} for autoinhibition at 150 mM KCl was determined to be $(8.3 \pm 1.4) \times 10^{2}$. As shown in Figure 1(C), the two-state model for autoinhibition can adequately explain the salt concentration-dependent $K_{d,app}$ data for the full-length HMGB1 protein.

Conformational equilibrium of dynamic autoinhibition

We used NMR spectroscopy to gain further information about salt concentration dependence of autoinhibition and investigated the salt concentration dependence of conformational equilibrium of HMGB1. With typical cryogenic NMR probes, experiments on samples at a salt concentration greater than 300 mM are difficult because their high conductivity makes it impossible to achieve an optimal impedance matching of the ^{1}H RF circuit, diminishing the sensitivity in NMR detection. To circumvent this situation, we used a coaxial tube with the protein solution in a thinner inner tube (the inner diameter 3.2 mm) and D_2O in the outer tube. A small sample diameter is known to help resolve the problem on NMR experiments for high conductivity samples.[45] This sample configuration utilizing the coaxial tubes allowed us to achieve an optimal impedance matching of the cryoprobe ^{1}H RF circuit, even for samples containing 900 mM KCl.

Owing to this approach which permits a wide range of ionic strength in the NMR experiments, we were able to observe the salt-driven modulation of the HMGB1 autoinhibition to the full extent. Figure 2(A) shows overlaid ^{1}H-^{15}N TROSY spectra recorded for the full-length HMGB1 protein (red) and the Δ30 variant (blue) at various concentrations of KCl between 40 and 900 mM. The ^{1}H-^{15}N spectra of the same backbone NH groups in these proteins were remarkably different at 40 mM KCl, suggesting that the full-length protein undergoes the autoinhibition involving the D/E repeats at the C-terminal region. As the KCl concentration increased, the ^{1}H-^{15}N cross peaks of the two proteins became increasingly closer. The full-length HMGB1 protein and the Δ30 variant at 900 mM KCl exhibited virtually identical ^{1}H-^{15}N spectra for the same residues. These results suggest that the autoinhibition of HMGB1 is electrostatically driven, occurs in the fast exchange regime on the NMR chemical shift timescale, and is completely disrupted at KCl concentrations $\geq$ 700 mM.

We also compared the spectra recorded for the DNA-bound states of the full-length HMGB1 protein and the Δ30 variant. Unfortunately, many signals in the spectra for the complexes with the cisplatin-modified 20-bp DNA duplex were severely broadened, perhaps due to competition between A-box and B-box for the single cisplatin-modification site on the DNA, which may occur in the intermediate exchange regime. However, the complexes with the corresponding unmodified 20-bp DNA duplex at 100 mM KCl exhibited ^{1}H-^{15}N correlation spectra of reasonable quality (Figure 2(C)). As previously demonstrated for a DNA complex of the isolated A-box domain,[46] the full-length HMGB1 and Δ30 variant proteins are likely to change their location on the 20-bp DNA duplex in the fast exchange regime due to the lack of any high-affinity site in this DNA. In contrast to the case for the free proteins at 100 mM KCl (see Figure 2(A)), the spectrum recorded for the DNA complex of the full-length HMGB1 protein was similar to the spectrum recorded for the DNA complex of the Δ30 variant (Figure 2(C)). This similarity between the NMR spectra for the two DNA complexes suggests that the intramolecular interactions involving the D/E repeats are largely (but not completely) disrupted when HMGB1 is bound to DNA. However, chemical shifts of some residues are somewhat different between these complexes, suggesting that some intramolecular interactions of the D/E repeats remain in the DNA-bound state of HMGB1. In fact, molecular dynamics simulations indicate that some intramolecular interactions between the D/E repeats and other IDRs remain while the interactions between the D/E repeats and A-box/B-box are disrupted in the DNA-bound state (see below). Our data collectively show the alteration in the intramolecular interactions upon HMGB1′s binding to DNA.

Electrostatic fuzzy interactions in autoinhibition

Figure 3(A) shows {^{1}H-}^{15}N nuclear Overhauser effects (NOE) measured for the full-length HMGB1

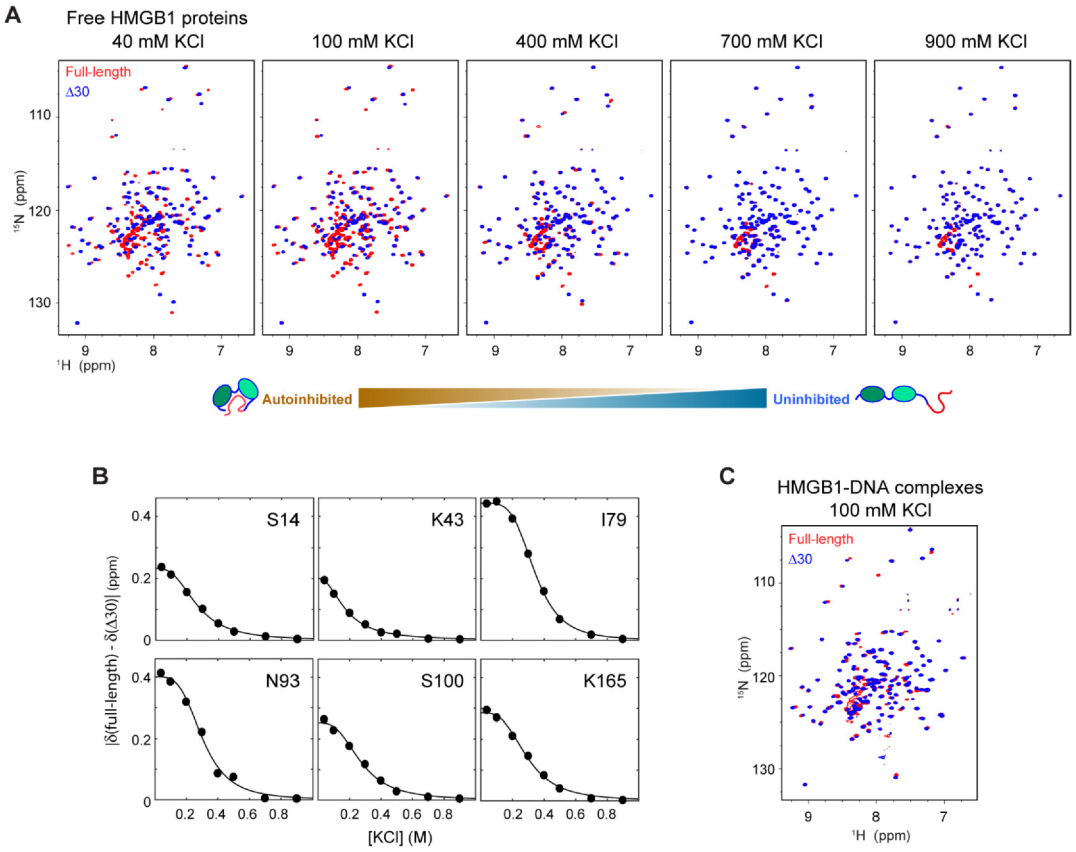

Figure 2. Salt concentration-dependent equilibrium between the autoinhibited and uninhibited states of HMGB1. (**A**) Overlaid ^{1}H-^{15}N TROSY spectra recorded for the full-length HMGB1 protein (red) and the Δ30 variant (blue) at 40, 100, 400, 700, and 900 mM KCl. The spectra for the Δ30 variant are laid over the corresponding spectra for the full-length HMGB1. At high concentrations of KCl, the two proteins exhibited virtually identical resonances, reflecting a shift of the equilibrium toward the uninhibited state. (**B**) Chemical shift differences ($\Delta\delta$) between the full-length HMGB1 and the Δ30 variant measured at various concentrations of KCl. The solid lines represent the best-fit curves (see the main text). (**C**) Overlaid ^{1}H-^{15}N TROSY spectra recorded for the proteins bound to 20-bp DNA. Note that the spectra of the full-length and Δ30 variant proteins are more similar in the DNA-bound state than in the free state at 100 mM KCl.

protein at 100 and 700 mM KCl. Heteronuclear NOE is useful to identify regions that are conformationally mobile on a ps-ns timescale: a smaller heteronuclear NOE qualitatively suggest higher mobility.[47] Many residues in the IDRs of HMGB1 exhibited heteronuclear NOE smaller than 0.5 at both 100 and 700 mM KCl, suggesting that the IDRs are flexible both in the uninhibited and autoinhibited states. For these residues, the heteronuclear NOEs at 700 mM KCl were considerably smaller than at 100 mM KCl. The decrease in heteronuclear NOE was particularly remarkable for the D/E repeats. Although severe overlaps of the NMR signals from the D/E repeats precluded us from determining heteronuclear NOE values for these residues except D214 and D215, many signals from the D/E repeats in the sub-spectra recorded with ^{1}H saturation were negative at 700 mM KCl and positive at 100 mM KCl (Figure 3(B)). These results suggest that the D/E repeats and other IDRs are more restricted in the autoinhibited state than in the uninhibited state.

To obtain more quantitative information into the electrostatic nature of the autoinhibition, we analyzed the NMR chemical shift differences between the full-length HMGB1 protein and the Δ30 variant at various concentrations of KCl. The populations of the autoinhibited state and the uninhibited state are given by $K_{ai}/(1 + K_{ai})$ and $1/(1 + K_{ai})$, respectively. At all ionic strengths, the full-length HMGB1 protein exhibited only a single ^{1}H-^{15}N cross peak for each NH group (Figure 2), indicating that the conformational exchange between the autoinhibited and uninhibited states occurs in the fast exchange regime. The chemical shift of the full-length HMGB1 protein can be represented by $\delta_a K_{ai}/(1 + K_{ai}) + \delta_u/(1 + K_{ai})$, where δ_a and δ_u represent the chemical shifts of the autoinhibited and uninhibited states, respectively. Since the Δ30 variant should always exhibit δ_u due

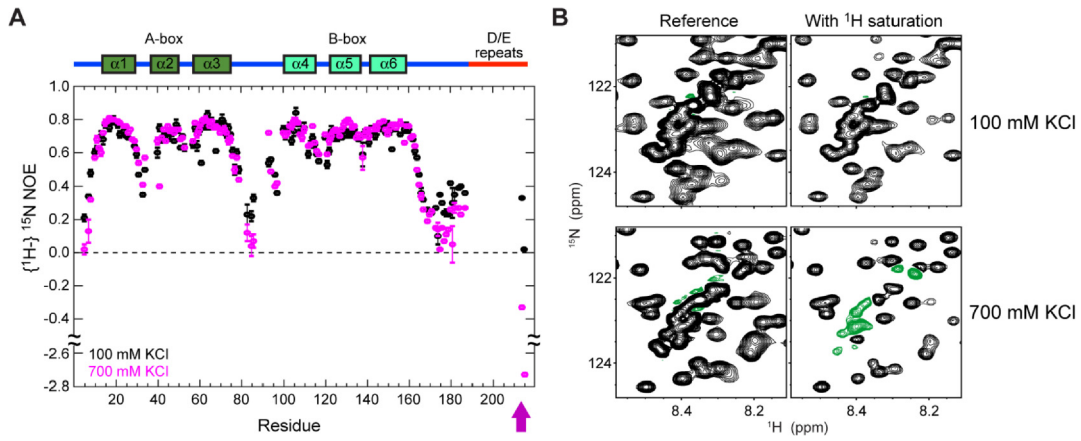

Figure 3. The IDRs of HMGB1 are mobile in both states but are more restricted in the autoinhibited state than in the uninhibited state. (**A**) Heteronuclear $\{^1H\text{-}\}^{15}N$ NOE measured for the full-length HMGB1 protein at 100 mM and 700 mM KCl. Note that many residues in IDRs exhibit smaller heteronuclear NOEs at 700 mM KCl than at 100 mM KCl. This was particularly obvious for the C-terminal two residues, which are indicated by an arrow. NMR signals from many residues in the D/E repeats could not be assigned due to severe overlaps. (**B**) Part of the subspectra recorded for measuring the heteronuclear NOE. Positive and negative contours are shown in black and green, respectively. Severely overlapped signals around ^1H 8.3–8.5 ppm and ^{15}N 122–124 ppm are from the D/E repeats. Some of these signals in the ^1H-saturation spectra were negative at 700 mM KCl but were positive at 100 mM KCl, suggesting that the D/E repeats are less mobile in the autoinhibited state than in the uninhibited state.

to the lack of the D/E repeats, the chemical shift difference ($\Delta\delta$) between the full-length HMGB1 protein and the Δ30 variant is given by:

$$\Delta\delta = \Delta\delta_{max} K_{ai}/(1 + K_{ai}), \quad (5)$$

where $\Delta\delta_{max}$ represents $\delta_a - \delta_u$. Using Eq. 4 and 5, we carried out nonlinear least-squares fitting to the experimental $\Delta\delta$ data for individual NH groups of HMGB1 at KCl concentrations of 40, 100, 200, 300, 400, 500, 700, and 900 mM. The parameters $\Delta\delta_{max}$, *a*, and *b* (see Eq. (4)) were optimized in these fitting calculations. Examples of the fitting are shown in Figure 2(B). Using Eq. (4), the value of the equilibrium constant K_{ai} at 150 mM KCl, a physiological ionic strength, was calculated from the determined values of the parameters *a* and *b*. Figure 4 shows the values of K_{ai} and the population of the autoinhibited state at 150 mM KCl for individual residues.

If the free HMGB1 protein undergoes a simple exchange process between the autoinhibited state and the uninhibited state, the equilibrium constant K_{ai} should be virtually identical no matter which residues are analyzed. However, our data show a large variety of K_{ai} values for different regions of HMGB1. The IDR between A-box and the B-box and the IDR between B-box and the D/E repeats showed large K_{ai} values, indicating that the D/E repeats interact more strongly with these two IDRs than with the other regions. Relatively large values of $\Delta\delta_{max}$ also implicated that the D/E repeats preferentially interact with these IDRs. Comparison of the ^{13}C NMR chemical shifts for the full-length HMGB1 and Δ30 variant proteins suggests that the intramolecular interactions involving the D/E repeats do not induce formation of any secondary structures within the IDRs (see Figure S1 in Supplemental Information). Our data shown in Figures 2–4 clearly indicate the presence of the dynamic interactions between the D/E repeats and the IDRs in the autoinhibited state. The folded regions of A-box and B-box appear to interact with the D/E repeats, but the K_{ai} data suggest that their interactions with the D/E repeats are weaker. For example, the population of the autoinhibited state for A-box at 150 mM KCl was calculated to be ~ 0.6, while that of the IDR between A-box and B-box was calculated to be higher than 0.9 (Figure 4).

Interestingly, the range of the K_{ai} values from the NMR data was remarkably smaller than the corresponding K_{ai} value from the salt concentration-dependent binding affinity data. It should be noted that the NMR data reflect the microscopic binding equilibrium of the D/E repeats at each intramolecular interaction site, whereas the salt-concentration dependent binding affinity data reflect the macroscopic equilibrium. The discrepancy between the microscopic and macroscopic K_{ai} constants can be explained as follows. The autoinhibition may occur when the D/E repeats mask any crucial part for DNA-binding. Fuzzy interactions of the D/E repeats can dynamically mask various parts of HMGB1, but not simultaneously. While the occupancy of the D/E repeats at each intramolecular interaction sites is relatively small, the probability for the D/E repeats to occupy any crucial part for DNA-binding can be high for the ensemble of the fuzzy interactions. In other words, the multivalent nature of the fuzzy interactions of the D/E repeats may

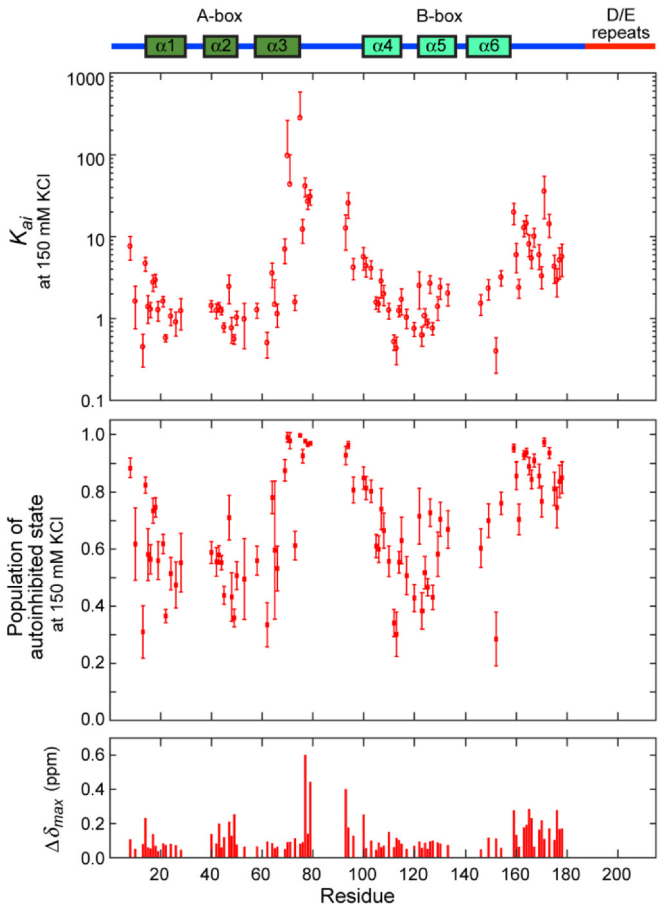

Figure 4. The microscopic equilibrium constants for intramolecular interactions with the D/E repeats are diverse and smaller than the macroscopic equilibrium constant for autoinhibition. Shown here are the K_{ai} constants, the populations of the autoinhibited state, and $\Delta\delta_{max}$ determined using the salt concentration-dependent differences in NMR chemical shifts between the full-length and Δ30 variant HMGB1 proteins (see Figure 2). For each residue, using Eqs. 4 and 5, the parameters $\Delta\delta_{max}$, *a*, and *b* were optimized to fit to the salt concentration-dependence data. Some examples of the best-fit curves are shown in Figure 2(B). The shown K_{ai} constants were calculated using [KCl] = 0.15 M (a physiological ionic strength) along with Eq. (4) and the determined values of the parameters *a* and *b*. These K_{ai} constants correspond to the microscopic equilibrium constants for intramolecular interactions with the D/E repeats. The population of the autoinhibited state was calculated as $K_{ai}/(1 + K_{ai})$.

create strong autoinhibition against the specific interaction with DNA at a macroscopic level.

Autoinhibition dynamics observed in molecular dynamics simulations

In order to gain additional microscopic insight into the fuzzy interactions of the D/E repeats, we used computational approaches as well. Although others have conducted all-atom or coarse-grained molecular dynamics simulations of HMGB1,[48–50] these studies do not provide us with adequate information regarding the autoinhibition process. We performed all-atom and coarse-grained molecular dynamics simulations for HMGB1 and examined the intramolecular interactions between the D/E repeats and other parts of the molecule. Both types of simulations show a wide variety of relative positionings for the two DNA-binding domains, the D/E repeats, and other IDRs of HMGB1. To better understand the preference of the D/E repeats in the fuzzy interactions with the other regions, we analyzed intramolecular interaction energies of the D/E repeats for the structure ensembles obtained from the atomistic and coarse-grained simulations.

Figure 5(A) shows two-dimensional maps that compare the interaction energies of the D/E repeats with the DNA-binding domains and with the IDR regions. As indicated, four major states were identified in the two-dimensional map for the atomistic simulations of the wild-type HMGB1 protein. For each state, a representative snapshot is shown in Figure 5(B). In states 1 and 2, the D/E repeats interact primarily with the IDRs only, whereas in states 3 and 4, the D/E repeats dynamically interact with both the IDRs and the

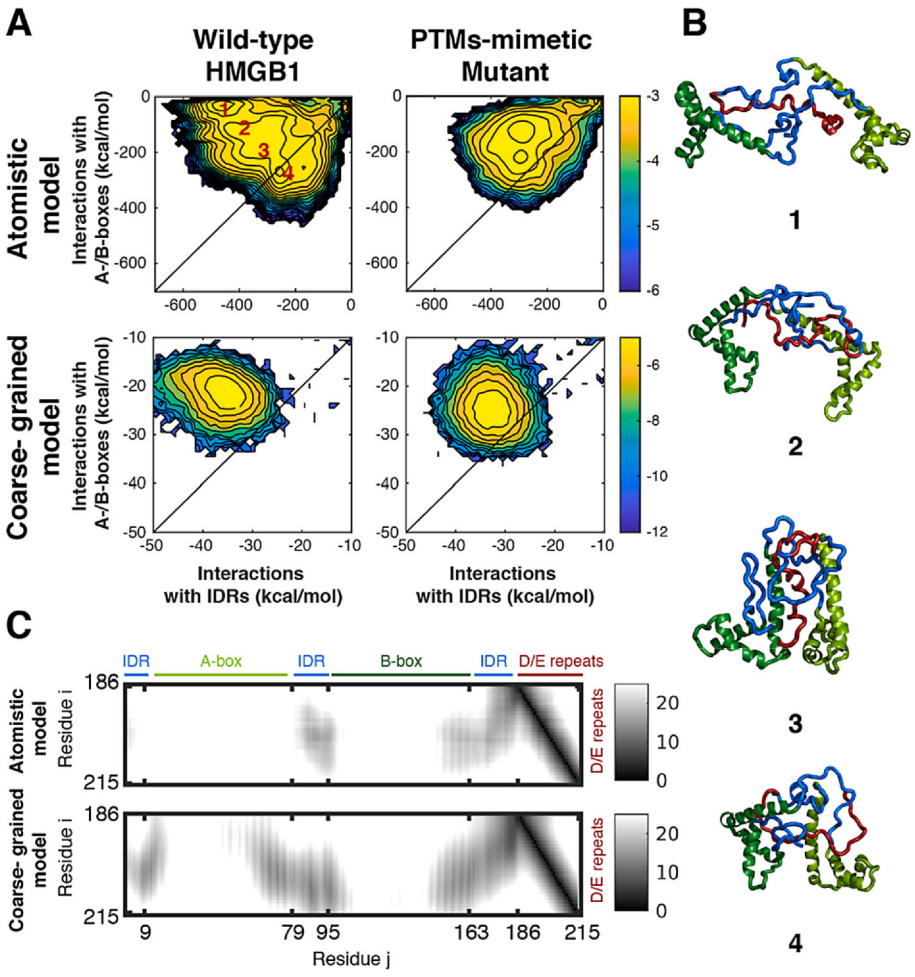

Figure 5. Fuzzy interactions of the D/E repeats with other regions of the HMGB1 protein observed in atomistic and coarse-grained molecular dynamics simulations. (**A**) Two-dimensional maps of the interaction energy of the D/E repeats are plotted for both models. The conformational analysis of the autoinhibited state is shown for wild-type and PTMs-mimetic mutant of HMGB1. The strength of the interactions between the D/E repeats and the A-/B-boxes is plotted versus the strength of the interactions between the D/E repeats and the IDRs in HMGB1. These maps indicate that the D/E repeats interact more strongly with the IDRs than with the two boxes. The plotted energy refers to non-bonded energy between the selected residues. In the atomistic model, the non-bonded energies include both columbic and van der Waals interactions. In the coarse-grained model, the non-bonded energies include only columbic interactions. In all the energy calculations, we excluded the residues that were mutated in the PTMs-mimetic mutant so that the comparison between the wild-type and PTMs-mimetic mutant will be more direct. The color of the 2D map represents the logarithm of the population of each bin in the map. The atomistic and coarse-grained maps were obtained for salt concentrations of 150 and 30 mM, respectively. (**B**) Representative snapshots of full-length HMGB1 for the states 1–4 indicated in the 2D map obtained for the wild-type HMGB1 protein using the atomistic simulations. These snapshots illustrate the interactions between the D/E repeats (colored red) and the IDRs (colored blue). A-box and B-box are shown in dark and light green, respectively. (**C**) Matrices of the distances between each residues of the D/E repeats and the residues of HMGB1 obtained from conformational ensemble sampled using atomistic (top) and coarse-grained (bottom) simulations. The distances are indicated by the grey scale bar in Å.

DNA-binding domains. For states 1–3, the energy for the interactions with the IDRs is significantly larger than that for the interactions with the DNA-binding domains. This trend of stronger interactions with the IDRs than with the DNA-binding domains is also seen in the two-dimensional maps for the coarse-grained simulations. Thus, both atomistic and coarse-grained simulations are consistent with the experimental data showing that the D/E repeats prefer the IDRs over the folded region of HMGB1. Our data illustrate that the autoinhibited state is heterogeneous with a large ensemble of fuzzy interactions between the D/E repeats and other regions of HMGB1.

To further analyze the modes of interactions between the D/E repeats and other regions of HMGB1, we measured the distances between

each residue of the D/E repeats and the residues of HMGB1. Figure 5(C) shows the mean of these distances from ensembles of HMGB1 sampled in the atomistic and coarse-grained simulations. It is evident that the D/E repeats interact extensively with all three major IDRs of HMGB1, both in the atomistic and the coarse-grained simulations. Figure 5(C) shows that the D/E repeats also interact with some residues of A- and B-boxes. The atomistic simulations suggest that the interaction of the D/E repeats is more significant with B-box than with the A-box. In the coarse-grained simulations, the interactions of the D/E repeats with A- and B-boxes are more extensive compared to the atomistic simulations. This can be either due to the more extensive sampling of the coarse-grained simulations or alternatively due to their limited accuracy. The interactions between the D/E repeats and the three IDRs as well as the DNA-binding domains are in accord with the experimental results (Figure 4).

To gain some insight about the molecular origin of autoinhibition, we simulated the Δ 30 variant at atomistic resolution for a total of 9 μs. In the conformational ensemble of Δ 30 HMGB1, the distance between the center of mass of the DNA-recognition helices of the A- and B-boxes is 49 ± 12 Å. The corresponding distance in full-length HMGB1 is 43 ± 14 Å. The distance between these helices in a complex between HMGB1 and DNA (PDB ID 5ZDZ) is 57 Å, suggesting that the autoinhibition may originate from trapping the conformation in a more compact state as a consequence of the interactions between the D/E repeats and other regions of the HMGB1, particularly its IDRs.

To gain further insight into the autoinhibition against DNA-binding, we performed coarse-grained simulations of HMGB1 in the presence of nonspecific B-DNA. Through these simulations, we examined how the D/E repeats and DNA compete for the DNA-binding domains (i.e., A-box and B-box) of HMGB1. Our analysis of interaction energies for A-box and B-box clearly showed that their intramolecular interactions with the D/E repeats and intermolecular interactions with DNA are mutually exclusive (Figure 6(A) and (D)). When HMGB1 binds to DNA, the intramolecular interactions between the D/E repeats and the DNA-binding domains are largely disrupted. Interestingly, the D/E repeats strengthen intramolecular interactions with other IDRs upon HMGB1's binding to DNA (this observation is more pronounced for A-box than for B-box, Figure 6(B) and (E)). The association of the D/E repeats with the IDRs seems to be significantly coupled to the disruption of the interactions between the D/E repeats and the DNA-binding domains (Figure 6(C) and (F)). These computational results suggest that the release of the autoinhibition occurs through alteration in the fuzzy interactions of the D/E repeats in a dynamic

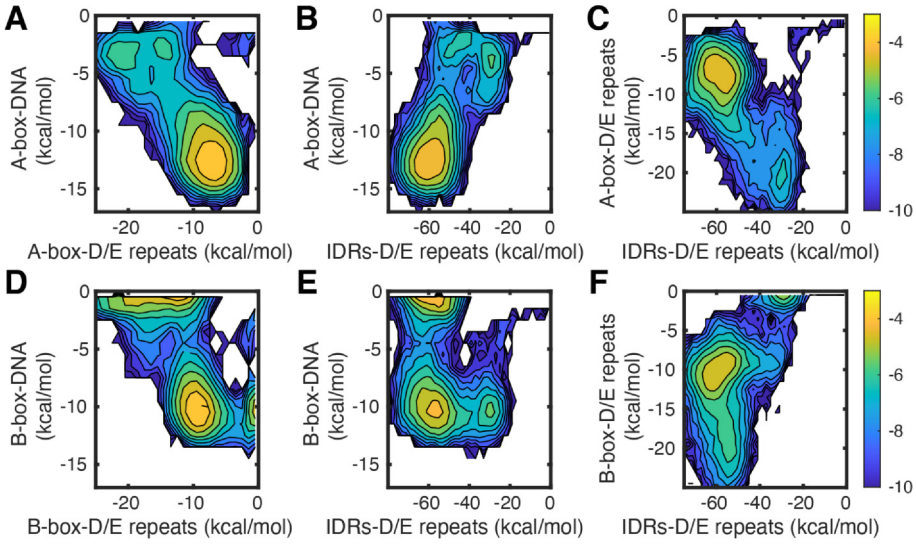

Figure 6. Interplay between the interactions of the D/E repeats with the IDRs and the A- and B-boxes of HMGB1 when bound to DNA. Simulations of the coarse-grained model at low salt concentrations are projected along several reaction coordinates to elucidate the binding mechanism of HMGB1 and the nature of the autoinhibition. The sampled trajectories are projected two-dimensionally along three order parameters: the interaction between the DNA-binding domains with DNA (y-axis of panels A, B, D, E), the intramolecular interactions between the DNA-binding domain and the D/E repeats (x-axis of panels A, D and y-axis of panels C, F), and the intramolecular interactions between the IDRs and the D/E repeats (x-axis of panels B, C, E, F). Panels A-C and panels D-F correspond to results for the A- and B-box, respectively.

ensemble rather than through complete disruption of all the intramolecular interactions. The remaining intramolecular interactions can explain why the NMR spectra for the protein-DNA complexes of the full-length HMGB1 and Δ30 variant proteins are somewhat different (Figure 2(C)).

Impact of PTM-mimetic mutations

Although HMGB1 molecules normally reside in the cell nuclei, some cells (e.g., macrophage) actively secrete HMGB1 into extracellular space, where HMGB1 acts as a cytokine.[51,52] The HMGB1 export from the nuclei involves the acetylation of particular lysine residues, including K180, K182, K183, K184, and K185.[53] Phosphorylation of S181 also occurs in the relocation process.[54] Since the region of the residues 180–185 is immediately adjacent to the D/E repeats (residues 186–215), it is reasonable to speculate that the acetylation and phosphorylation may make a significant impact on the autoinhibition of HMGB1. To examine this possibility, we prepared an HMGB1 variant in which the residues 180–185 are mutated from KSKKKK to QEQQQQ to mimic the post-translational modifications (PTMs). K → Q and S → E mutations are commonly used to mimic acetyl-lysine and phospho-serine side chains, respectively.[55,56] Using this variant, we investigated the extent to which the naturally occurring PTMs can impact HMGB1′s binding affinities and conformational equilibrium.

As shown in Figure 7(A), the PTMs-mimetic mutant exhibited a ~ 3-fold higher affinity for the cisplatin-modified DNA than the wild-type HMGB1 protein at physiological ionic strength. Considering that the mutations reduce the overall charge by 6, the higher affinity of the PTMs-mimetic mutant may be surprising. In fact, when the ionic strength was high enough to disrupt autoinhibition, the PTMs-mimetic mutant exhibited lower affinity than the wild-type HMGB1 protein. The high affinity of the PTMs-mimetic mutant could be explained by a shift in conformational equilibrium of autoinhibition. As shown in Figure 7(B), some NMR signals from the mutant protein were located between the signals from the wild-type HMGB1 protein and those from the Δ30 variant, implicating that the conformational equilibrium of this mutant might be shifted slightly toward the uninhibited state. The molecular dynamics simulations indicate that the PTMs-mimetic mutations weakened the interactions of the D/E repeats with the IDRs while the interactions with the DNA-binding domains remained similar to those in the wild-type HMGB1 protein. These results, which support that the autoinhibition mostly occurs via interactions between the D/E repeats and the linker region that tether the DNA-binding domains, were observed in both the atomistic and coarse-grained simulations, as shown in Figure 5(A).

Discussion

Electrostatic interactions are crucial for protein-DNA association processes.[40,57] It seems reasonable that various DNA-binding proteins utilize competitive electrostatic interactions as a mechanism for autoinhibition against DNA binding. Our experimental data on HMGB1 show that the competitive electrostatic interactions for autoinhibition are diminished when the KCl concentration is increased far beyond physiological ionic strength. The salt concentration-dependent $K_{d,app}$ data can be explained by the two-state model involving the autoinhibited state and the uninhibited state at a macroscopic level (Figure 1). At higher ionic strengths, the equilibrium between these two states is shifted toward the uninhibited state due to weaker electrostatic interactions between the D/E repeats and other regions of HMGB1. At lower ionic strengths, the equilibrium between the two states is shifted toward the autoinhibited state, and consequently, the apparent affinity of the Δ30 variant lacking the D/E repeats becomes far stronger than that of the full-length HMGB1. Our NMR and computational data suggest that the underlying processes are more dynamic and fuzzier at a microscopic level (Figures 2–6). These processes appear to involve various states in which the D/E repeats dynamically interact with various parts of HMGB1 with a strong preference for the positively charged IDRs. To a lesser extent, the D/E repeats also interact with the folded regions of A-box and B-box. Thus, the macroscopic autoinhibited state seems to be an ensemble of various states in which the D/E repeats are associated with different segments of HMGB1 through electrostatic fuzzy interactions.

Intramolecular fuzzy interactions of IDRs have been proposed to play important roles in long-range organization of different functional parts and in signal sensing.[1,58] In the current case, the multivalent nature of fuzzy interactions seems to enhance the macroscopic autoinhibition effect. The fuzzy interactions may also create a larger number of microscopic states, which may provide an entropic advantage. HMGB1′s binding to DNA appears to disrupt many of the fuzzy interactions involving the D/E repeats (Figure 2(C)). This disruption should further increase the mobility of the IDRs and may contribute to the entropic term of the binding free energy. Furthermore, when HMGB1 encounters DNA, the dynamic autoinhibited state might facilitate the transition first to the uninhibited state and then to the DNA-bound state through relatively low energy barriers. Upon binding to DNA, the D/E repeats may compensate the loss of interactions with A- and B-boxes by having more extensive interactions with the IDRs of HMGB1 (Figure 6). Some of the microscopic states in the autoinhibition process may also promote recognition of target molecules other than DNA. Since HMGB1 is known to interact with various nuclear

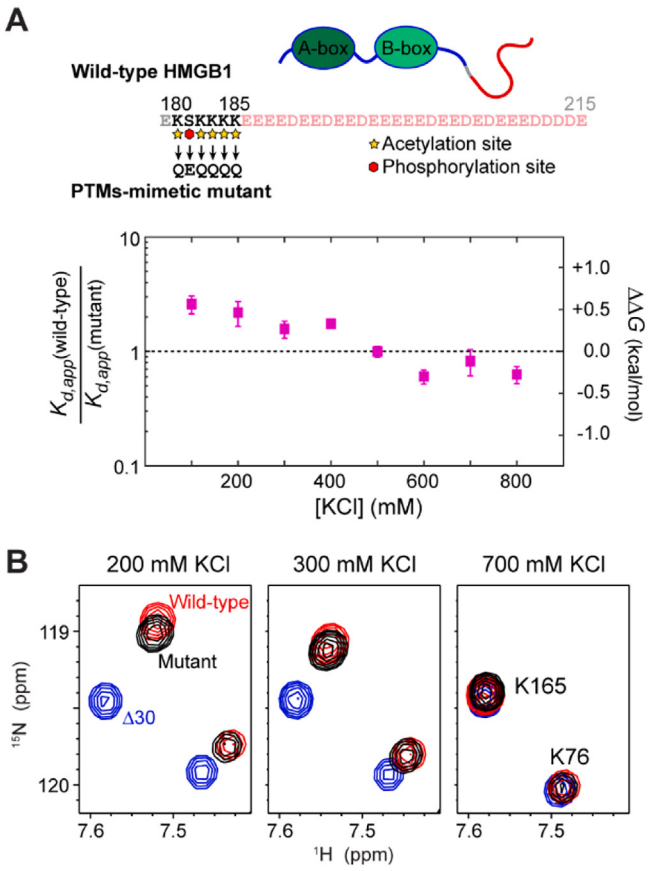

Figure 7. The impact of PTM-mimetic mutations near the D/E repeats of HMGB1. (**A**) Ratio of the apparent dissociation constants ($K_{d,app}$) for the complexes with the cisplatin-modified 20-bp DNA at various concentrations of KCl. (**B**) NMR signals from backbone NH groups of K165 and K76 in the wild-type full-length HMGB1 (red), the PTMs-mimetic mutant HMGB1 (black) and the Δ30 variant (blue) at 200, 300, and 700 mM KCl. Note that the signals from the mutant are located between the signals from the wild-type full-length HMGB1 protein and those from the Δ30 variant at low ionic strengths.

and extracellular proteins,[28,52] the larger number of microscopic states created by fuzzy interactions of IDRs might assist HMGB1 in recognizing different molecules.

The D/E repeats are required for the HMGB1 protein to facilitate chromatin remodeling and displacement of histone H1 from linker DNA.[29,30] When nuclear HMGB1 assists various gene-regulatory proteins and DNA-repair/recombination enzymes, HMGB1 typically acts on them only transiently and does not serve as a stable subunit of multi-subunit protein-DNA complexes.[26,28] In fact, fluorescence imaging studies showed that HMGB1 rapidly diffuses in the cell nuclei.[59,60] It seems likely that the short-lived nature of the interactions with DNA is important for HMGB1′s dynamic action as a DNA chaperone. We should also point out that some HMG-box type DNA-binding proteins do not possess D/E repeats. For example, HMGB4, a paralogue of HMGB1, does not possess D/E repeats. The function of these proteins lacking D/E repeats might require stable association with DNA rather than short-lived interactions. The dynamic autoinhibition via the D/E repeats might mobilize HMGB1 so that it can efficiently serve as a DNA chaperone.

Our data on the PTMs-mimetic mutant HMGB1 protein suggest that the PTMs involved in the export of HMGB1 from the cell nuclei increase the DNA-binding affinity (Figure 7). This might be somewhat counterintuitive because the higher affinity may help sequester HMGB1 on the genomic DNA. However, the region of the PTMs (the residues 180–185) corresponds to one of the two nuclear localization sequences (NLSs) of HMGB1.[28] Since basic amino-acid residues are crucial for NLSs,[61] neutralization of lysine side chains through acetylation should have negative impacts on nuclear localizations. In fact, K → Q mutations in this region were shown to cause a decrease in nuclear HMGB1 and an increase in cytoplasmic HMGB1.[53] Our data (Figure 7A) show that the apparent affinity of HMGB1 for the cisplatin-modified DNA at a physiological ionic strength becomes ~ 3-fold stronger by the PTMs-mimetic mutations. This impact is rather moderate, compared to other systems. For example, S46/T55

phosphorylation in the p53 protein and S282/S285 phosphorylation in the Ets-1 protein reduce the binding affinity for their target DNA by a factor of ~ 10 and ~ 100, respectively.[4,15] Autoinhibition via D/E repeats might be generally less sensitive to PTMs because interactions involving D/E repeats are fuzzy and can occur at multiple distinct sites. Studies on other proteins containing D/E repeats would provide further insight into this possibility.

In conclusion, our experimental and computational investigations have provided thermodynamic and structural dynamic insights into the electrostatically driven autoinhibition of HMGB1. The strongly negatively charged D/E repeats at the C-terminal region of HMGB1 can interact with various parts of the HMGB1 protein with strong preference for the positively charged IDRs over the folded regions. Our NMR data show that the D/E repeats and other parts of HMGB1 undergo fuzzy interactions, each of which is weaker than expected from the macroscopic autoinhibitory effect. This discrepancy may suggest that the multivalent nature of the fuzzy interactions enables strong autoinhibition at a macroscopic level despite relatively weak intramolecular interaction at each site. The mutations mimicking the acetylation and phosphorylation relevant to the export from the nuclei moderately modulate HMGB1′s affinity for DNA, possibly through a shift in the autoinhibition equilibrium.

Materials and methods

Preparation of proteins

A synthetic gene of human HMGB1 (215 residues) was sub-cloned into the NcoI/HindIII sites of the pET28a vector. This is different from the plasmid used in our previous study[62] and was capable of producing a larger amount of HMGB1. The plasmid for expression of the Δ30 variant (residues 1–185) was obtained using a Quick-Change Lightning mutagenesis kit (Agilent Technologies) by introducing two stop codons at the positions of residue 186 and 187 in the full-length HMGB1 plasmid. The plasmid for expression of the K180Q/S181E/K182Q/K183Q/K184Q/K185Q mutant of HMGB1 was produced using a Quick-Change Lightning kit with a single pair of PCR primers. The sequences of the plasmids were confirmed through Sanger sequencing. Proteins were expressed in *Escherichia coli* BL21 (DE3) cells cultured in minimal media, where ^{13}C glucose and ^{15}N ammonium chloride were used as the sole carbon and nitrogen sources to label the proteins with ^{13}C and ^{15}N isotopes. Harvested cells were lysed by sonication at 4 °C in a buffer of 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 100 mM NaCl, 1% Triton X-100, and 5% glycerol. To avoid proteolysis, Roche cOmplete protease inhibitor cocktail was used (1 tablet per 50 ml of the sonication buffer). The supernatant was loaded onto an SP-FF cation exchange column (GE Healthcare) equilibrated with 20 mM sodium phosphate (pH 6.0), 1 mM EDTA, and 100 mM NaCl. The protein was eluted with a gradient of 100–2000 mM NaCl. For the full-length wild-type and mutant HMGB1 proteins, each protein was loaded onto a Resource-Q anion-exchange column (GE Healthcare) and eluted with a gradient of 0–1500 mM NaCl in 50 mM Tris-HCl (pH 7.5) and 1 mM EDTA. For the Δ30 variant, this purification step was replaced with cation-exchange chromatography using a Resource S cation-exchange column (GE Healthcare) using the same gradient. Each protein was concentrated to ~ 10 mL and further purified through size-exclusion chromatography using an S-100 column (GE Heathcare) equilibrated with a buffer of 100 mM ammonium acetate (pH 7). The purified proteins were lyophilized and stored at −20 °C until use.

Preparation of DNA

Chemically synthesized DNA strands were purchased from Integrated DNA Technologies, Inc. Each strand was purified by Mono-Q anion-exchange chromatography. For preparing cisplatin-modification of DNA, an activated cisplatin solution was made by reaction with 1.97 mol eq of $AgNO_3$ in distilled, deionized water on a shaker at room temperature for 16 h in the dark followed by centrifugation to remove the AgCl precipitate.[63] The 20-mer DNA strand containing single GpG site (5′-CTCTGGACCTTCCTTTCTTC-3′; denoted GG20) was platinated by addition of the activated cisplatin solution (1.5 eq) to a 70 μM solution of the oligonucleotide. The reaction mixture was incubated at 37 °C for 8 h. The product was purified by Resource-Q column with a gradient of 25–45% buffer containing 50 mM sodium phosphate (pH 6.0) and 1500 mM NaCl. After annealing of complimentary strands, each duplex was purified by Resource-Q. The site-specific cisplatin modification in the DNA duplex was confirmed by NMR. For fluorescence studies, a fluorescein amidite (FAM) was attached to the 5′-terminus of GG20.

Fluorescence-based assays of binding of HMGB1 to cisplatin-modified DNA

The affinities of the full-length HMGB1, Δ30 variant, and PTMs-mimetic mutant proteins for the cisplatin-modified DNA were determined by protein titration. The completely reduced form of the protein was obtained by incubating with 5 mM dithiothreitol (DTT) for 1 h. In protein-titration experiment, the fluorescence anisotropy of 1 nM FAM-labeled cisplatin-modified DNA was measured at an excitation wavelength of 490 nm and an emission wavelength of 521 nm using an ISS PC-1 spectrofluorometer in a buffer of 10 mM potassium phosphate (pH 7.5), 1 mM DTT and various concentrations of KCl (100–900 mM) at

25 °C. The dissociation constant K_d was determined from the anisotropy data using MATLAB (MathWorks, Inc), as previously described.[64] Nonlinear least-squares fitting calculations for the salt concentration-dependent K_d data were also performed with MATLAB software.

NMR chemical shift assignment

The NMR spectra for resonance assignment were recorded at 25C using a Bruker Avance III NMR spectrometer equipped with a cryogenic probe operated at the ^{1}H frequency of 600 MHz. Although the resonance assignment data for the full-length HMGB1 protein and for the Δ30 variant were available at Biomolecular Magnetic Resonance Bank (accession codes 15,502 and 7408, respectively), our NMR experiments showed somewhat different chemical shifts for some residues, presumably due to different experimental conditions. Using 0.4 mM ^{13}C/^{15}N-labeled proteins in a buffer of 10 mM potassium phosphate (pH 7.5), 5 mM deuterated DTT, and 100 mM KCl, and 5% D_2O, we recorded 3D ^{15}N-edited NOESY spectra and TROSY-versions of 3D HNCO, HNCA, HN(CO)CA, HNCACB, CBCA (CO)NH spectra[65] for backbone ^{1}H/^{13}C/^{15}N resonance assignment of the full-length HMGB1 protein and the Δ30 variant. NMR-Pipe[66] was used for the data processing and NMRFAM-SPARKY[67] was used for the spectral analysis.

NMR experiments on salt dependence of autoinhibition equilibrium

Salt titration samples were made of 270 μL solutions with 0.3 mM ^{15}N-labelled full-length protein or Δ 30 variant, in a buffer containing 40–900 mM KCl, 10 mM potassium phosphate (pH 7.5), 5 mM deuterated DTT and 1 mM sodium 2,2-dimethyl 2-silapentane-5-sulfonate (DSS). 8 different concentrations of KCl (40, 100, 200, 300, 400, 500, 700, and 900 mM) were used. For each, a ~ 270 ul solution was sealed in an inner tube (the tube inner diameter 3.2 mm) with a 100 μL of D_2O in the outer tube of Shigemi coaxial NMR tubes (the outer diameter 5.0 mm). The coaxial tubes were used to achieve optimal impedance matching for the cryogenic probe on high ionic-strength samples (see above). For each sample, 1D ^{1}H and ^{1}H-^{15}N TROSY spectra were recorded at 25 °C. NMR chemical shifts were referenced to the ^{1}H signal from DSS as the internal reference. For each backbone NH group, a unified chemical shift difference $\Delta\delta = [(\delta_H^{FL} - \delta_H^{\Delta 30})^2 + 0.25(\delta_N^{FL} - \delta_N^{\Delta 30})^2]^{1/2}$, where δ_H and δ_N represent ^{1}H and ^{15}N chemical shifts, was calculated from the resonances of the full-length (FL) and Δ30 proteins at each concentration of KCl. Nonlinear least-squares fitting to the $\Delta\delta$ data using Eqs. (4) and (5) was performed using MATLAB software. Heteronuclear {^{1}H-} ^{15}N NOEs for backbone NH groups of the full-length HMGB1 protein at 100 and 700 mM KCl were measured at the ^{1}H frequency of 800 MHz with a 5 s period for ^{1}H saturation.

Molecular dynamics simulations

Atomistic model. We performed atomistic MD simulations for HMGB1 and its Δ30 variant using GROMACS (v. 2020).[68] The force field parameters for the protein, SPC water, and ions were derived from the AMBER99SB-ILDN force field.[69] All structures were placed in a dodecahedral box, and solvated. Sodium and Chloride ions were added to a concentration of 0.125 M, with slight adjustments to neutralize the overall charge. All structures were subject to minimization and NVT and NPT equilibration. The initial atomic coordinates were taken from three of the NMR models available in PDB entry 2YRQ, using HMGB1 residues 2–166, (models 1, 16, and 20). PyMOL was used to generate the missing residues (167–215), in a mostly helical conformation. The same modelled region was used for all three starting structures. Each structure was simulated for 1000 ns, in two separate runs, resulting in a total of six simulations. To mimic the acetylated and phosphorylated forms, K180, K182, K183, K184, and K185 were replaced with glutamine residues and S181 was replaced with glutamic acid. The changes were done in COOT.[70] A total of six productions runs, each 1000 ns, were run for the modified structures.

Coarse-grained model. The dynamics of HMGB1 and its binding to DNA were studied using coarse-grained molecular dynamics (CG-MD) simulations. Each residue was represented by a single bead at the position of its Cα atom. The DNA was modeled with three beads per nucleotide, representing phosphate, sugar, and base.

The force-field applied in our simulations used a native-topology based model that includes a Lennard-Jones potential to reward native contacts and a repulsive potential to penalize non-native contacts.[71–73] The positively charged residues of the protein (Lys, Arg) were assigned a point charge of (+1e) and the negatively charged residues (Asp, Glu) as well as the phosphate beads of the DNA backbone were assigned a negative charge of ($-1e$). The electrostatic potential between charged beads q_i, q_j was modeled by the Debye-Hückel interaction, which accounts for the ionic strength of a solute immersed in aqueous solution.[74] The explicit form of the force field is reported elsewhere.[75]

The structure of HMGB1 used in this study was based on the first conformation in the NMR structure from PDB ID 2YRQ. For wild-type HMGB1, the missing disordered residues 167–215 were modeled as an elongated chain, and more realistic conformations were obtained during

simulation. In the CG-MD, the interactions between the D/E repeats and the other domains of HMGB1 are modeled by electrostatic interactions only. All other interactions within HMGB1 are modeled as repulsive (i.e., excluded volume), unless a contact is defined in the selected NMR structure using the CSU program.[76] These native contacts are modeled in the CG model using the Lennard-Jones potential. Two additional variants were designed for this study: one with a neutral tail, in which the charges from the negatively charged residues in position 186–215 were removed, and a mutant mimicking acetylation of lysine residues 180,182–185 and phosphorylation of S181 (shown in bold in the sequence shown in Figure 7(A)). In this variant, the lysine residues were neutralized and a negative charge was assigned to the serine.

The dynamics of the three constructs of HMGB1 in isolation and in presence of DNA were simulated using the Langevin equation.[77] The simulation temperature was set to 0.4 (reduced units), which is lower than the folding temperatures of HMGB1. The dielectric constant was 80, and the salt concentration was varied as mentioned throughout the main text. For each system, we preformed 5 simulations consisting of 10^7 MD steps. When DNA was included, DNA was modeled as a linear double-stranded B-DNA molecule with length 100 base-pairs. Trajectory frames were saved every 1000 steps. Periodic boundary conditions were not used in our model.

CRediT authorship contribution statement

Xi Wang: Investigation, Formal analysis, Writing - original draft. **Harry M. Greenblatt:** Investigation, Formal analysis. **Lavi S. Bigman:** Investigation, Formal analysis. **Binhan Yu:** Investigation, Formal analysis. **Channing C. Pletka:** Resources. **Yaakov Levy:** Conceptualization, Investigation, Formal analysis, Funding acquisition, Writing - review & editing. **Junji Iwahara:** Conceptualization, Investigation, Formal analysis, Funding acquisition, Writing - original draft, Writing - review & editing.

Acknowledgments

This work was supported by Grant MCB-2026805 from National Science Foundation (to J.I.), Grant R35-GM130326 from National Institutes of Health (to J.I.), and by a grant from the United States – Israel Binational Science Foundation (BSF), Jerusalem, Israel (to Y.L.). We thank Ross Luu for preliminary experiments; and Tianzhi Wang for maintenance of the NMR equipment at the University of Texas Medical Branch.

Conflict of interests

The authors declare no competing financial interests.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2021.167122.

Received 26 March 2021;
Accepted 22 June 2021;
Available online 25 June 2021

Keywords:
conformational equilibrium;
molecular dynamics;
NMR;
protein-DNA interactions;
thermodynamics

1 In some literature on HMGB1, a residue-numbering scheme starting from the actual amino-terminal glycine (i.e., G1-K2-...) is used, whereas in others, a scheme starting from the initial methionine in the gene (i.e., M1-G2-K3-...) is used. In this paper, we use the latter.

Abbreviations:
IDR, intrinsically disordered region; NLS, nuclear localization sequence; PTM, post-translational modification

References

1. van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., Kim, P.M., Kriwacki, R.W., Oldfield, C.J., Pappu, R.V., Tompa, P., Uversky, V.N., Wright, P.E., Babu, M.M., (2014). Classification of intrinsically disordered regions and proteins. *Chem. Rev.*, **114**, 6589–6631.
2. Liu, J., Perumal, N.B., Oldfield, C.J., Su, E.W., Uversky, V. N., Dunker, A.K., (2006). Intrinsic disorder in transcription factors. *Biochemistry*, **45**, 6873–6888.
3. Vuzman, D., Levy, Y., (2012). Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Mol. BioSyst.*, **8**, 47–57.
4. Desjardins, G., Meeker, C.A., Bhachech, N., Currie, S.L., Okon, M., Graves, B.J., McIntosh, L.P., (2014). Synergy of aromatic residues and phosphoserines within the intrinsically disordered DNA-binding inhibitory elements of the Ets-1 transcription factor. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 11019–11024.
5. He, F., Borcherds, W., Song, T., Wei, X., Das, M., Chen, L., Daughdrill, G.W., Chen, J., (2019). Interaction between p53 N terminus and core domain regulates specific and nonspecific DNA binding. *Proc. Natl. Acad. Sci. U. S. A.*, **116**, 8859–8868.
6. Katan-Khaykovich, Y., Shaul, Y., (2001). Nuclear import and DNA-binding activity of RFX1. Evidence for an autoinhibitory mechanism. *Eur. J. Biochem.*, **268**, 3108–3116.

7. Knapp, S., Muller, S., Digilio, G., Bonaldi, T., Bianchi, M.E., Musco, G., (2004). The long acidic tail of high mobility group box 1 (HMGB1) protein forms an extended and flexible structure that interacts with specific residues within and between the HMG boxes. *Biochemistry*, **43**, 11992–11997.
8. Krois, A.S., Dyson, H.J., Wright, P.E., (2018). Long-range regulation of p53 DNA binding by its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. U. S. A.*, **115**, E11302–E11310.
9. Lee, S., Miller, M., Shuman, J.D., Johnson, P.F., (2010). CCAAT/Enhancer-binding protein beta DNA binding is auto-inhibited by multiple elements that also mediate association with p300/CREB-binding protein (CBP). *J. Biol. Chem.*, **285**, 21399–21410.
10. Liu, Y., Matthews, K.S., Bondos, S.E., (2008). Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the Drosophila hox protein ultrabithorax. *J. Biol. Chem.*, **283**, 20874–20887.
11. Perez-Borrajero, C., Lin, C.S., Okon, M., Scheu, K., Graves, B.J., Murphy, M.E.P., McIntosh, L.P., (2019). The biophysical basis for phosphorylation-enhanced DNA-binding autoinhibition of the ETS1 transcription factor. *J. Mol. Biol.*, **431**, 593–614.
12. Pufall, M.A., Lee, G.M., Nelson, M.L., Kang, H.S., Velyvis, A., Kay, L.E., McIntosh, L.P., Graves, B.J., (2005). Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science*, **309**, 142–145.
13. Stott, K., Watson, M., Bostock, M.J., Mortensen, S.A., Travers, A., Grasser, K.D., Thomas, J.O., (2014). Structural insights into the mechanism of negative regulation of single-box high mobility group proteins by the acidic tail domain. *J. Biol. Chem.*, **289**, 29817–29826.
14. Stott, K., Watson, M., Howe, F.S., Grossmann, J.G., Thomas, J.O., (2010). Tail-mediated collapse of HMGB1 is dynamic and occurs via differential binding of the acidic tail to the A and B domains. *J. Mol. Biol.*, **403**, 706–722.
15. Sun, X., Dyson, H.J., Wright, P.E., (2021). A phosphorylation-dependent switch in the disordered p53 transactivation domain regulates DNA binding. *Proc. Natl. Acad. Sci. U. S. A.*,, 118.
16. S. Ueshima, K. Nagata, M. Okuwaki, Internal associations of the acidic region of upstream binding factor control its nucleolar localization, Mol. Cell. Biol. 2017, 37.
17. Wang, F., Marshall, C.B., Yamamoto, K., Li, G.Y., Plevin, M.J., You, H., Mak, T.W., Ikura, M., (2008). Biochemical and structural characterization of an intramolecular interaction in FOXO3a and its binding with p53. *J. Mol. Biol.*, **384**, 590–603.
18. Watson, M., Stott, K., Thomas, J.O., (2007). Mapping intramolecular interactions between domains in HMGB1 using a tail-truncation approach. *J. Mol. Biol.*, **374**, 1286–1297.
19. Wiebe, M.S., Nowling, T.K., Rizzino, A., (2003). Identification of novel domains within Sox-2 and Sox-11 involved in autoinhibition of DNA binding and partnership specificity. *J. Biol. Chem.*, **278**, 17901–17911.
20. Pufall, M.A., Graves, B.J., (2002). Autoinhibitory domains: modular effectors of cellular regulation. *Annu. Rev. Cell Dev. Biol.*, **18**, 421–462.
21. Chou, C.C., Wang, A.H., (2015). Structural D/E-rich repeats play multiple roles especially in gene regulation through DNA/RNA mimicry. *Mol. BioSyst.*, **11**, 2144–2151.
22. Wang, H.-C., Chou, C.-C., Hsu, K.-C., Lee, C.-H., Wang, A. H.J., (2019). New paradigm of functional regulation by DNA mimic proteins: Recent updates. *IUBMB Life*, **71**, 539–548.
23. Finke, J.M., Jennings, P.A., Lee, J.C., Onuchic, J.N., Winkler, J.R., (2007). Equilibrium unfolding of the poly (glutamic acid)20 helix. *Biopolymers*, **86**, 193–211.
24. Olander, D.S., Holtzer, A., (1968). The stability of the polyglutamic acid alpha helix. *J. Am. Chem. Soc.*, **90**, 4549–4560.
25. Lu, H., Wang, J., Bai, Y., Lang, J.W., Liu, S., Lin, Y., Cheng, J., (2011). Ionic polypeptides with unusual helical stability. *Nat. Commun.*, **2**, 206.
26. Agresti, A., Bianchi, M.E., (2003). HMGB proteins and gene expression. *Curr. Opin. Genet. Dev.*, **13**, 170–178.
27. Mandke, P., Vasquez, K.M., (2019). Interactions of high mobility group box protein 1 (HMGB1) with nucleic acids: Implications in DNA repair and immune responses. *DNA Repair (Amst)*, **83**, 102701
28. Stros, M., (2010). HMGB proteins: interactions with DNA and chromatin. *BBA*, **1799**, 101–113.
29. Bonaldi, T., Langst, G., Strohner, R., Becker, P.B., Bianchi, M.E., (2002). The DNA chaperone HMGB1 facilitates ACF/CHRAC-dependent nucleosome sliding. *EMBO J.*, **21**, 6865–6873.
30. Cato, L., Stott, K., Watson, M., Thomas, J.O., (2008). The interaction of HMGB1 and linker histones occurs through their acidic and basic tails. *J. Mol. Biol.*, **384**, 1262–1272.
31. Thomas, J.O., Stott, K., (2012). H1 and HMGB1: modulators of chromatin structure. *Biochem. Soc. Trans.*, **40**, 341–346.
32. Bianchi, M.E., Beltrame, M., Paonessa, G., (1989). Specific recognition of cruciform DNA by nuclear protein HMG1. *Science*, **243**, 1056–1059.
33. Lee, K.B., Thomas, J.O., (2000). The effect of the acidic tail on the DNA-binding properties of the HMG1,2 class of proteins: insights from tail switching and tail removal. *J. Mol. Biol.*, **304**, 135–149.
34. Lorenz, M., Hillisch, A., Payet, D., Buttinelli, M., Travers, A., Diekmann, S., (1999). DNA bending induced by high mobility group proteins studied by fluorescence resonance energy transfer. *Biochemistry*, **38**, 12150–12158.
35. Musumeci, D., Bucci, E.M., Roviello, G.N., Sapio, R., Valente, M., Moccia, M., Bianchi, M.E., Pedone, C., (2011). DNA-based strategies for blocking HMGB1 cytokine activity: design, synthesis and preliminary in vitro/in vivo assays of DNA and DNA-like duplexes. *Mol. BioSyst.*, **7**, 1742–1752.
36. Amato, J., Cerofolini, L., Brancaccio, D., Giuntini, S., Iaccarino, N., Zizza, P., Iachettini, S., Biroccio, A., Novellino, E., Rosato, A., Fragai, M., Luchinat, C., Randazzo, A., Pagano, B., (2019). Insights into telomeric G-quadruplex DNA recognition by HMGB1 protein. *Nucleic Acids Res.*, **47**, 9950–9966.
37. Amato, J., Madanayake, T.W., Iaccarino, N., Novellino, E., Randazzo, A., Hurley, L.H., Pagano, B., (2018). HMGB1 binds to the KRAS promoter G-quadruplex: a new player in oncogene transcriptional regulation?. *Chem. Commun. (Camb.)*, **54**, 9442–9445.
38. Jung, Y., Lippard, S.J., (2003). Nature of full-length HMGB1 binding to cisplatin-modified DNA. *Biochemistry*, **42**, 2664–2671.
39. Mitkova, E., Ugrinova, I., Pashev, I.G., Pasheva, E.A., (2005). The inhibitory effect of HMGB-1 protein on the

repair of cisplatin-damaged DNA is accomplished through the acidic domain. *Biochemistry*, **44**, 5893–5898.
40. Privalov, P.L., Dragan, A.I., Crane-Robinson, C., (2011). Interpreting protein/DNA interactions: distinguishing specific from non-specific and electrostatic from non-electrostatic components. *Nucleic Acids Res.*, **39**, 2483–2491.
41. Record Jr., M.T., Lohman, M.L., De Haseth, P., (1976). Ion effects on ligand-nucleic acid interactions. *J. Mol. Biol.*, **107**, 145–158.
42. Record Jr., M.T., Zhang, W., Anderson, C.F., (1998). Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: a practical guide to recognizing and interpreting polyelectrolyte effects, Hofmeister effects, and osmotic effects of salts. *Adv. Protein Chem.*, **51**, 281–353.
43. Yu, B., Pletka, C.C., Iwahara, J., (2021). Quantifying and visualizing weak interactions between anions and proteins. *Proc. Natl. Acad. Sci. U. S. A.*, **118**, e2015879118
44. Yu, B., Iwahara, J., (2021). Experimental approaches for investigating ion atmospheres around nucleic acids and proteins. *Comput. Struct. Biotechnol. J.*, **19**, 2279–2285.
45. Voehler, M.W., Collier, G., Young, J.K., Stone, M.P., Germann, M.W., (2006). Performance of cryogenic probes as a function of ionic strength and sample tube geometry. *J. Magn. Reson.*, **183**, 102–109.
46. Iwahara, J., Schwieters, C.D., Clore, G.M., (2004). Characterization of nonspecific protein-DNA interactions by ^{1}H paramagnetic relaxation enhancement. *J. Am. Chem. Soc.*, **126**, 12800–12808.
47. Kay, L.E., Torchia, D.A., Bax, A., (1989). Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, **28**, 8972–8979.
48. Anggayasti, W.L., Ogino, K., Yamamoto, E., Helmerhorst, E., Yasuoka, K., Mancera, R.L., (2020). The acidic tail of HMGB1 regulates its secondary structure and conformational flexibility: A circular dichroism and molecular dynamics simulation study. *Comput. Struct. Biotechnol. J.*, **18**, 1160–1172.
49. Kanada, R., Terakawa, T., Kenzaki, H., Takada, S., (2019). Nucleosome crowding in chromatin slows the diffusion but can promote target search of proteins. *Biophys. J .*, **116**, 2285–2295.
50. Meng, X.Y., Li, B., Liu, S., Kang, H., Zhao, L., Zhou, R., (2016). EGCG in green tea induces aggregation of HMGB1 protein through large conformational changes with polarized charge redistribution. *Sci. Rep.*, **6**, 22128.
51. Malarkey, C.S., Churchill, M.E., (2012). The high mobility group box: the ultimate utility player of a cell. *Trends Biochem. Sci.*, **37**, 553–562.
52. Yang, H., Wang, H., Czura, C.J., Tracey, K.J., (2005). The cytokine activity of HMGB1. *J. Leukoc. Biol.*, **78**, 1–8.
53. Bonaldi, T., Talamo, F., Scaffidi, P., Ferrera, D., Porto, A., Bachi, A., Rubartelli, A., Agresti, A., Bianchi, M.E., (2003). Monocytic cells hyperacetylate chromatin protein HMGB1 to redirect it towards secretion. *EMBO J.*, **22**, 5551–5560.
54. Youn, J.H., Shin, J.S., (2006). Nucleocytoplasmic shuttling of HMGB1 is regulated by phosphorylation that redirects it toward secretion. *J. Immunol.*, **177**, 7889–7897.
55. Chen, Z., Cole, P.A., (2015). Synthetic approaches to protein phosphorylation. *Curr. Opin. Chem. Biol.*, **28**, 115–122.
56. Kamieniarz, K., Schneider, R., (2009). Tools to tackle protein acetylation. *Chem. Biol.*, **16**, 1027–1029.
57. Yu, B., Pettitt, B.M., Iwahara, J., (2020). Dynamics of ionic interactions at protein-nucleic acid interfaces. *Acc. Chem. Res.*, **53**, 1802–1810.
58. Arbesu, M., Iruela, G., Fuentes, H., Teixeira, J.M.C., Pons, M., (2018). Intramolecular fuzzy interactions involving intrinsically disordered domains. *Front Mol. Biosci.*, **5**, 39.
59. Agresti, A., Scaffidi, P., Riva, A., Caiolfa, V.R., Bianchi, M. E., (2005). GR and HMGB1 interact only within chromatin and influence each other's residence time. *Mol. Cell*, **18**, 109–121.
60. Scaffidi, P., Misteli, T., Bianchi, M.E., (2002). Release of chromatin protein HMGB1 by necrotic cells triggers inflammation. *Nature*, **418**, 191–195.
61. Dingwall, C., Laskey, R.A., (1991). Nuclear targeting sequences–a consensus?. *Trends Biochem. Sci.*, **16**, 478–481.
62. Zandarashvili, L., Sahu, D., Lee, K., Lee, Y.S., Singh, P., Rajarathnam, K., Iwahara, J., (2013). Real-time kinetics of high-mobility group box 1 (HMGB1) oxidation in extracellular fluids studied by in situ protein NMR spectroscopy. *J. Biol. Chem.*, **288**, 11621–11627.
63. Takahara, P.M., Frederick, C.A., Lippard, S.J., (1996). Crystal structure of the anticancer drug cisplatin bound to duplex DNA. *J. Am. Chem. Soc.*, **118**, 12309–12321.
64. Zandarashvili, L., Nguyen, D., Anderson, K.M., White, M. A., Gorenstein, D.G., Iwahara, J., (2015). Entropic enhancement of protein-DNA affinity by oxygen-to-sulfur substitution in DNA phosphate. *Biophys. J .*, **109**, 1026–1037.
65. Cavanagh, J., Fairbrother, J.W., Palmer, G.A., Rance, M., Skelton, J.N., (2007). Protein NMR Spectroscopy: Principles and Practice. Elsevier Academic Press, Second edit.
66. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., Bax, A., (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
67. Lee, W., Tonelli, M., Markley, J.L., (2015). NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, **31**, 1325–1327.
68. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., Lindahl, E., (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1–2**, 19–25.
69. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., Shaw, D.E., (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, **78**, 1950–1958.
70. Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
71. Clementi, C., Nymeyer, H., Onuchic, J.N., (2000). Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.*, **298**, 937–953.
72. Noel, J.K., Levi, M., Raghunathan, M., Lammert, H., Hayes, R.L., Onuchic, J.N., Whitford, P.C., (2016).

SMOG 2: A versatile software package for generating structure-based models. *PLoS Comput. Biol.*, **12**, e1004794

73. Noel, J.K., Whitford, P.C., Sanbonmatsu, K.Y., Onuchic, J. N., (2010). SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res.*, **38**, W657–W661.
74. Azia, A., Levy, Y., (2009). Nonnative electrostatic interactions can modulate protein folding: molecular dynamics with a grain of salt. *J. Mol. Biol.*, **393**, 527–542.
75. Bigman, L.S., Levy, Y., (2020). Protein diffusion on charged biopolymers: DNA versus microtubule. *Biophys. J.*, **118**, 3008–3018.
76. Sobolev, V., Sorokine, A., Prilusky, J., Abola, E.E., Edelman, M., (1999). Automated analysis of interatomic contacts in proteins. *Bioinformatics*, **15**, 327–332.
77. Givaty, O., Levy, Y., (2009). Protein sliding along DNA: dynamics and structural characterization. *J. Mol. Biol.*, **385**, 1087–1097.