ChemBioChem

Research Article
doi.org/10.1002/cbic.202400589

Chemistry Europe
European Chemical Societies Publishing

www.chembiochem.org

Phosphorylation of the HMGN1 Nucleosome Binding Domain Decreases Helicity and Interactions with the Acidic Patch

Dina Iebed[+],[a] Tobias Gökler[+],[a] Hugo van Ingen,[b] and Anne C. Conibear*[a]

Intrinsically disordered proteins are abundant in the nucleus and are prime sites for posttranslational modifications that modulate transcriptional regulation. Lacking a defined three-dimensional structure, intrinsically disordered proteins populate an ensemble of several conformational states, which are dynamic and often altered by posttranslational modifications, or by binding to interaction partners. Although there is growing appreciation for the role that intrinsically disordered regions have in regulating protein-protein interactions, we still have a poor understanding of how to determine conformational population shifts, their causes under various conditions, and how to represent and model conformational ensembles. Here, we study the effects of serine phosphorylation in the nucleosome-binding domain of an intrinsically disordered protein – HMGN1 – using NMR spectroscopy, circular dichroism and modelling of protein complexes. We show that phosphorylation induces local conformational changes in the peptide backbone and decreases the helical propensity of the nucleosome binding domain. Modelling studies using AlphaFold3 suggest that phosphorylation disrupts the interface between HMGN1 and the nucleosome acidic patch, but that the models over-predict helicity in comparison to experimental data. These studies help us to build a picture of how posttranslational modifications might shift the conformational populations of disordered regions, alter access to histones, and regulate chromatin compaction.

Introduction

Intrinsically disordered proteins (IDPs) or regions (IDRs) are abundant in the nucleus and have key roles in regulating chromatin packing and gene transcription.[1–3] Among the most abundant nucleosomal proteins are the high mobility group (HMG) protein families, HMGA, HMGB, and HMGN, which are highly charged, intrinsically disordered, and extensively post-translationally modified.[4,5] Their interactions with the nucleosome and with DNA are thought to regulate how chromatin compacts, and how transcription factors and enzymes that modify histones gain access to nucleosomes.[6,7] Lacking defined tertiary structures and binding pockets, and with multiple interaction partners, it is difficult to characterize the roles of the HMG proteins. The HMG protein family also represents typical IDPs, for which we still have a poor understanding of how to characterize and visualize their multiple conformational states. It is therefore also challenging to evaluate how posttranslational modifications (PTMs) might alter these populations and affect binding to their interaction partners.[8–11]

Within the HMG family, the nucleosome-binding HMGN proteins (HMGN1–5, Supplementary Data S1) are represented by HMGN1 in Figure 1. HMGN1 comprises 100 residues and has a split nuclear localization signal (NLS), a nucleosome binding domain (NBD) and a regulatory/chromatin-unfolding domain (CHUD/RD), all of which are intrinsically disordered.[4,12,13] Positively charged (5×Arg, 21×Lys) and negatively charged (14×Glu) amino acid residues are enriched in these domains and comprise 40% of the HMGN1 sequence. Ser and Thr as potential phosphorylation sites are located within or at the border of the domains, potentially increasing and/or extending the charged character of the protein. Several PTMs throughout the sequence have been reported: phosphorylation of serine, acetylation of lysine, as well as serine ADP-ribosylation.[12,14–16] However, it is not known what role many of these PTMs have, individually or in combination, in the biological role of HMGN1. Furthermore, little is known about the conformational states of HMGN1, and how PTMs might alter the population of these states, thereby regulating binding to the nucleosome and other proteins, and access to histones.

Elucidating the role of PTMs in the conformational states and interactions of IDRs such as HMGN1 faces several challenges. Lacking a defined structure, IDPs/IDRs are highly dynamic and may yield multiple dissimilar conformations in response to transient physiological changes in the protein environment or the installation of PTMs, making structural analysis by crystallography or cryo-electron microscopy

[a] D. Iebed,[+] T. Gökler,[+] A. C. Conibear
Institute of Applied Synthetic Chemistry, TU Wien, Getreidemarkt 9, 1060 Vienna, Austria
Phone: +43-1-58801-163723
E-mail: anne.conibear@tuwien.ac.at

[b] H. van Ingen
Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands

[+] Equal contributions.

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cbic.202400589

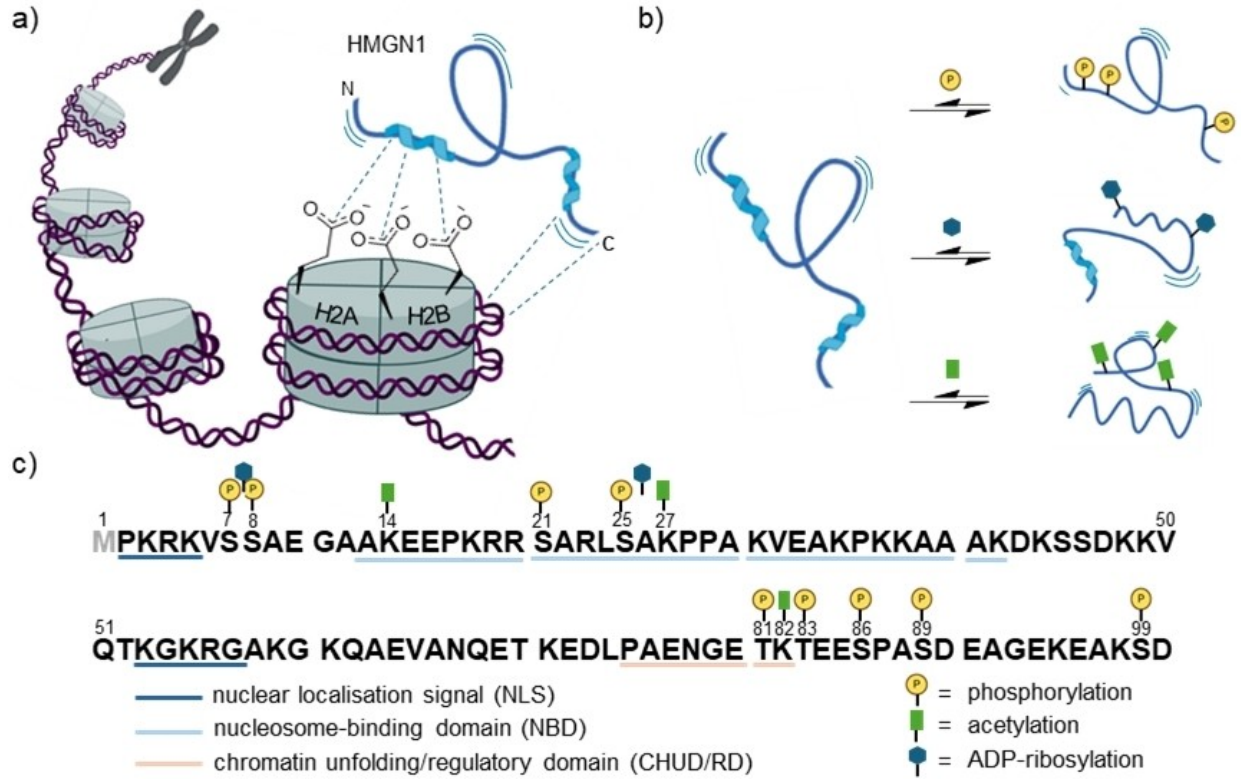

Figure 1. Schematic structure of HMGN1, the protein domains and their posttranslational modifications involved in protein dynamics and interactions. (a) According to the proposed binding mode, the nucleosome-binding domain (NBD) of HMGN1 interacts with the acidic patch of the nucleosome subunits H2A and H2B (side chains of glutamic acid residues shown as line structures), and the C-terminus binds to DNA. This interaction potentially regulates chromatin compaction and transcriptional activation. (b) Posttranslational modifications might induce conformational changes and intramolecular rearrangements, which can reshape the protein structure and dynamics, potentially yielding various subpopulations of HMGN1. (c) The amino acids of the HMGN1 protein sequence are represented by their one-letter codes. The first amino acid (M, in grey) is usually removed from the mature protein. Posttranslational modifications of HMGN1 as annotated in the UniProt entry (P05114) are marked at their position in the sequence (serine phosphorylation = yellow circle; lysine acetylation = green rectangle; and ADP ribosylation = blue hexagon). Domains of HMGN1 are underlined in the protein sequence (NLS = nuclear localization signal, NBD = nucleosome-binding domain, CHUD/RD = chromatin unfolding domain/regulatory domain). Nucleosome structures generated in BioRender.

impractical.[17,18] In contrast, NMR spectroscopy shows sufficient length- and timescale constraints (10^{-12}–10^{-9} s and m) to capture and differentiate between conformational intermediates, making this method ideally suited to the study of dynamic proteins and IDRs.[19,20] Solid phase peptide synthesis, often combined with protein ligations provides access to site-specifically modified protein variants,[21,22] yet is labour-intensive and possibilities for isotope-labelling for NMR spectroscopy are limited by the high cost of isotope-labelled SPPS building blocks. On the other hand, recombinant expression can yield isotope-labelled samples for NMR, but native PTMs, and especially combinations of PTMs, are difficult to install site-specifically. Semi-synthetic approaches, in which a synthetic, unlabelled protein segment (bearing the PTM) is ligated to a recombinant segment, provide an attractive combination, however a drawback of this approach is that local changes around the PTM site are not observable because the synthetic segment is unlabelled.[23] A detailed study of the local conformational changes and interactions around a posttranslationally modified residue is therefore typically only feasible with synthetic peptide segments bearing site-specific PTM(s) and acquiring NMR spectra at natural isotope abundance. This local information can then be integrated with experimental data from other techniques such as cross-linking mass spectrometry (XL–MS), circular dichroism, and modelling data in a hybrid approach to understand the biological roles and interactions of IDRs.

The structural and biological consequences of PTMs of HMGN1 have been studied using several approaches, *in vitro* and in cells, and indicate that serine phosphorylation within the NBD is involved in dissociation of HMGN1 from the nucleosome. Early studies on the phosphorylation of HMGN1 identified phosphorylation sites within the NBD (Ser21 and Ser25) and adjacent to the NLS (Ser7), and found that phosphorylated HMGN1 in general was enriched in the cytoplasm compared to the nucleus, suggesting that phosphorylation promotes translocation out of the nucleus.[24] Glutamic acid mutants of HMGN1 as phosphomimetics bearing a negative charge at these positions were also shown to have decreased binding to nucleosomes. In contrast, decreased nuclear import and increased 14.3.3 protein binding were only observed when the phosphate group was present.[25,26] In previous work from our group, we introduced site-specific phosphorylation and acetylation PTMs in the N-terminal NLS and the C-terminal domains of

HMGN1 using protein semi-synthesis and segmental isotope labelling. Using these constructs, we showed that N-terminal phosphorylation at Ser7 causes conformational changes in the NBD (residues 13–26), shown by chemical shift perturbations in the NMR spectra.[27] Studies on HMGN1 site-specifically phosphorylated within the NBD are however still missing. It is not known how phosphorylation might alter the local conformations or binding interactions of the HMGN1 NBD. Several studies have been carried out on HMGN2, which has overall 48% sequence identity to HMGN1 but is almost completely conserved within the NBD sequence (Supplementary Data S1). The architecture of the HMGN2-nucleosome complex was explored, employing NMR spectroscopy of methyl-labelled nucleosomes.[28] Based on this study, it was concluded that HMGN2 acts as a staple between the core histones and DNA *via* interaction of the NBD with the nucleosome acidic patch (as shown for HMGN1 in Figure 1) and the C-terminal region with the DNA at the entry/exit point of the nucleosome.[28] A glutamic acid mimic of serine phosphorylation did not bind to nucleosomes.[28] The model of the HMGN2-nucleosome interaction was then expanded using data from cellular XL–MS experiments, supporting an extended conformation of HMGN2, and showing interactions at the nucleosome acidic patch and either the histone H3 tail or the linker DNA.[29]

In this study, we investigate the effects of phosphorylation of serine at two sites, pSer21 and pSer25, within the NBD (residues 13–42) of HMGN1, comparing their local conformational changes and helical propensity with models and experimental structural biology data. The NBD of HMGN1 is highly conserved amongst the HMGN1 variants (Supplementary Data S1), and has been shown to act as an independent functional domain with the same activity as the full-length protein,[30] so we use it as a synthetically-accessible segment that is still feasible to study by 2D NMR spectroscopy at natural isotope abundance. As experimental evidence for the structure of the bound state of HMGN1 is still lacking, the effect of the phosphorylations on the interactions of the HMGN1 NBD with the acidic patch of the nucleosome is then explored using AlphaFold3 models, suggesting that these phosphorylations alter the helical architecture of the HMGN1 NBD and its electrostatic interactions with the acidic patch of the histone subunit H2A and H2B. Although the modelling results in general support previous results showing a decreased binding to the histone acidic patch (for HMGN2),[28] we also critically evaluate the modelling data in comparison with experimental data. These results provide a new structural and mechanistic understanding of the way that PTMs could regulate the biological role of HMGN1 in binding to nucleosomes and regulating chromatin folding and histone modifications. Furthermore, these studies expand our understanding of how PTMs might alter conformational populations of IDPs, and how we can integrate and evaluate information from structural biology experiments and predictive modelling tools that have recently become available.

Results and Discussion

To investigate the role of phosphorylation of serine residues 21 and 25 within the NBD of HMGN1 and to explore local conformational changes in this region, we synthesised the NBD (HMGN1_13–42) by solid phase peptide synthesis (SPPS), allowing us to install site-specific PTMs and yielding four variants (Figure 2): **HMGN1_NBD_unmod**; **HMGN1_NBD_S21**; **HMGN1_NBD_pS25**; and **HMGN1_NBD_pS21pS25**. To mimic their context in the protein chain, and to minimize any effects of charged termini, the N– and C–termini were acetylated and amidated, respectively. The synthesis proceeded smoothly *via* a combination of automated and manual SPPS, and minimal β-elimination of the phosphate was observed when performing the coupling and deprotection steps manually at room temperature. After purification by HPLC, the four variants were obtained in high purity and yielding 8–12 mg of each variant, from the synthesis at 0.033 mmol scale.

Recent advances in AI modelling tools like AlphaFold3, RoseTTAFold and others, have revolutionised how we predict the structure of a given protein from its sequence, based on training on proteins in structural databases.[31–33] Their strengths are in predicting the structures of well-folded, globular proteins.[11] Although they typically predict conformations of IDRs with low confidence and/or accuracy, the ability to predict the structures of full protein sequences (including flexible termini and loops) has increased our appreciation of how abundant IDRs are in the proteome.[32,34,35] Forming loops and linkers between domains or the N- and C-termini of many proteins, IDRs are typically cut off from X-ray crystallography and cryo-EM samples to facilitate handling or crystallisation, or are not visible because they are too dynamic.[32,36] The recently-released AlphaFold3 server,[31] now comes with the possibility to include some PTMs, including phosphorylation, but so far there are few benchmarks to evaluate how well it handles these PTMs, partly because experimental data on the structures of intrinsically disordered phosphoproteins are very limited. To model the conformations predicted for the HMGN1 NBD variants, we submitted each of the four NBD variants (HMGN1 residues 13–42) to the AlphaFold3 server and compared their modelled structures within the set of five output structures as well as between the variants. In almost all five predicted models of all four variants, an α-helix is predicted with surprisingly high confidence ($90 > \text{plDDT} > 70$) for the residues 18–25 around the phosphorylation sites (Figure 3). In the modelled structures, the remainder of the NBD is either random coil, or has short partially helical segments. Whereas the unmodified variant forms a distinct 'kink' at residues 23–25 in all models, the modified variants are overall more extended, suggesting that phosphorylation increases the random coil nature of the NBD.

HMGN1 is highly charged, with the potential for both intra- and intermolecular electrostatic interactions and salt bridges due to the high proportion of arginine and lysine residues, especially in the NBD. The helical conformation of residues 18–25 predicted in the models places the three arginine residues Arg19, Arg20 and Arg23 on the same face of the helix with Arg19 and Arg23 in a stacked *i*, $i+4$ arrangement and forming

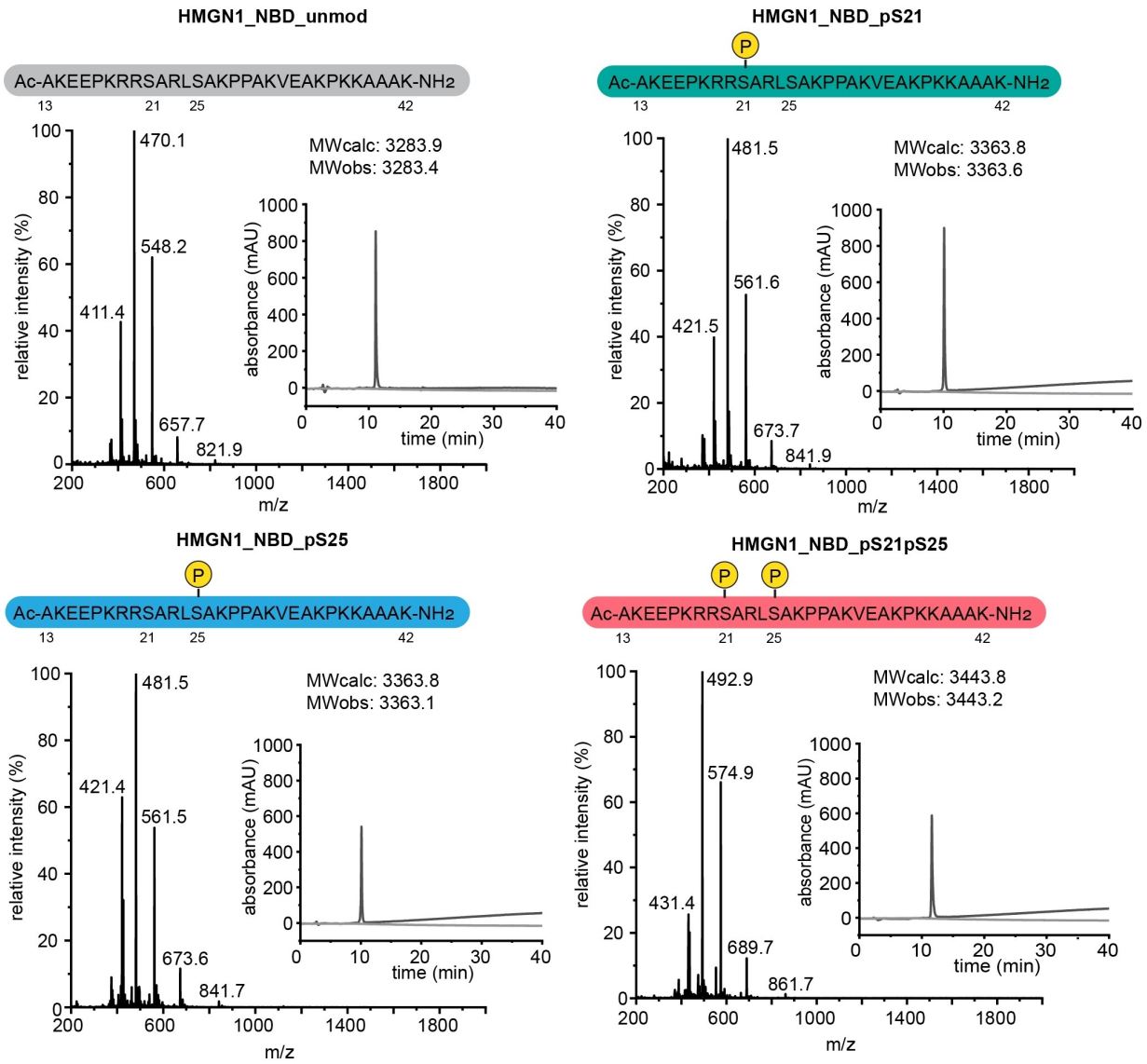

Figure 2. HMGN1 nucleosome binding domain (NBD, residues 13–42) variants synthesised by solid phase peptide synthesis. The sequences of the peptides are shown in one-letter amino acid codes and peptides are acetylated at the N-terminus and amidated at the C-terminus. Phosphoserine posttranslational modifications are shown as yellow circles and their residue positions marked. The ESI-MS trace with charge series marked, and analytical HPLC chromatograms (inset, absorbance at 214 nm in black, absorbance at 280 nm in grey) are shown.

a positively-charged face that could potentially interface with the histone acidic patch *via* bifurcated salt bridge formation. The serine or phosphoserine residues 21 and 25 project from the opposite face (Figure 3). Phosphorylation of both Ser21 and Ser25 would likely lead to electrostatic repulsion of the negative charges, potentially distorting the helix. On the other hand, phosphorylation of Ser21 alone introduces the possibility of salt bridge formation between the phosphate group and the side-chain of Lys18 (3.0 Å, Figure 3b), potentially stabilising the helical conformation; a higher helical content is seen in all predicted models of **HMGN1_NBD_pS21** than in **HMGN1_NBD_pS25** (Figure 3c, and Supplementary Data S2). In studies to develop parameters for modelling phosphoserine and phosphothreonine, it was found that glutamic acid mutants have less pronounced effects on partial secondary structure than the phosphorylated residues, especially for smaller peptides.[34] The synthesis and conformational evaluation of the true phosphorylated residues is therefore important to learn how well the models correlate with experimental data.

To determine whether the helical conformations and intramolecular interactions predicted in the modelled structures correlate with experimental data, the conformation of the NBD variants was then determined using circular dichroism (CD) spectroscopy. Concentrations of the four peptides were normalised, based on the peak height in the analytical HPLC trace. As shown in Figure 4a, all four NBD variants show a typical 'random coil' CD spectrum for the NBD, in keeping with our previous studies on full-length HMGN1.[27] At the same concentrations, the intensity of the peaks at 196 nm is greatest for the doubly-phosphorylated **HMGN1_NBD_pS21pS25** and

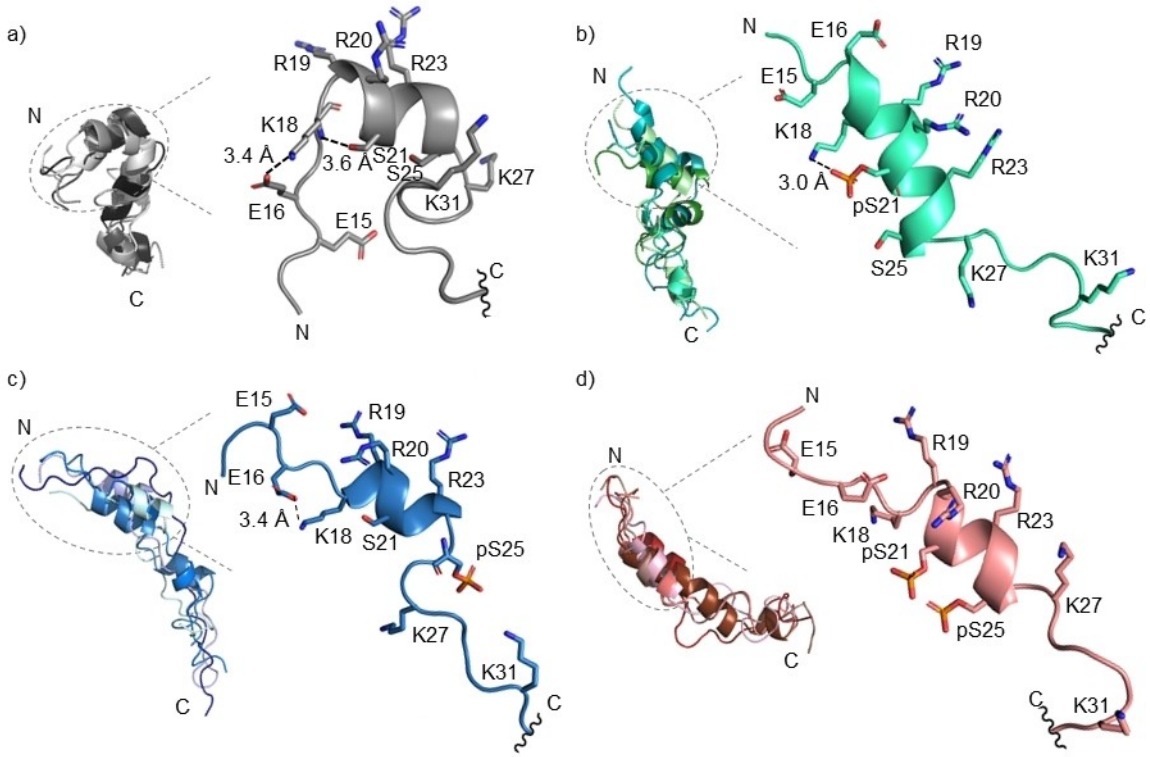

Figure 3. AlphaFold3[31] models of HMGN1 NBD variants. For each variant, the five predicted models are shown on the left, fitted pairwise to the top ranked prediction (model 0, darkest colour) and shown in cartoon representation. For the top ranked prediction (model 0) of each variant, a segment is shown with key residues around the phosphorylation site(s) labelled and shown in stick representation, coloured by atom type. The number of residues assigned as helical for each model is summarised in Supplementary Data S2.

HMGN1_NBD_pS25 variants, suggesting that these two variants are relatively more 'disordered' than the **HMGN1_NBD_pS21** and **HMGN1_NBD_unmod** variants. For the CD experiments, we focused on comparison between the differently modified variants, however, tools such as DichroIDP that include databases of IDPs have the potential to categorise and quantify the conformations of IDPs.[37] The phosphorylation at Ser25 seems to have the most influence on the conformation, as additional phosphorylation at Ser21 (**HMGN1_NBD_pS21pS25**) does not increase the intensity of the peak at 196 nm (blue and pink data series). The (partial) helical conformations indicated by the models, therefore, are not reflected in the experimental CD spectra, but the trend that **HMGN1_NBD_pS25** and **HMGN1_NBD_pS21pS25** are 'more disordered' than the **HMGN1_NBD_pS21** and **HMGN1_NBD_unmod** variants is consistent with the modelled structures and number of helical residues (Supplementary data S2).

To determine whether phosphorylation affects the overall helical propensity of the HMGN1 NBD variants, each of the variants was titrated with the helix-inducing solvent trifluoroethanol (TFE, 0–80 %) and CD spectra were acquired. The minima of the CD curves of all variants shifted towards higher wavelengths with increasing % TFE and the minima at 222 nm increased in intensity, indicating increased helical content (Supplementary Data S3). None of the variants, however, formed a purely helical structure with clear minima at 205 and 222 nm, even in 80 % TFE, showing that the HMGN1 NBD largely retains its disordered nature. This is also in keeping with expectations, based on the typical sequence composition of IDRs, with a high content of prolines and charged residues.[9,38] In this study, we used the same buffer as for the NMR experiments to be able to compare the data, however alternative buffers that include crowding agents and additional components[39,40] to more closely mimic the cell nucleus might be relevant for future studies. The intensity of the minimum at 222 nm was plotted against %TFE (Figure 4b) and suggests that both phosphorylations, individually and together, decrease the helical propensity of the HMGN1 NBD, with phosphorylation at Ser21 having the biggest influence (**HMGN1_NBD_pS21**, green data series). Although the differences in the order of the variants are small when comparing Figure 4a and b, one possible rationalization for the apparent change in order could be that the effects of phosphorylation on helicity might be the result of different types of interactions (e.g. electrostatic, hydrogen-bonding) and increasing amounts of TFE could promote these to different degrees. The data series of the doubly-modified variant **HMGN1_NBD_pS21pS25** closely overlaps with that of the **HMGN1_NBD_pS21** variant, suggesting that the effects of the two phosphorylations are not additive. The unmodified variant shows the highest helical propensity, but is still far from a classical helical CD spectrum, even at 80 % TFE.

To investigate the local and residue-specific changes associated with HMGN1 NBD phosphorylation, the conforma-

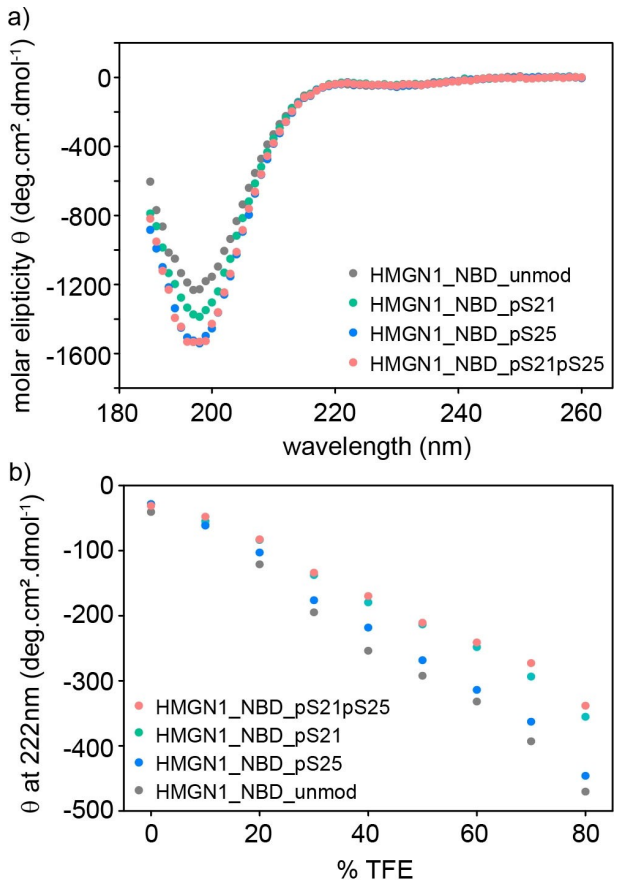

Figure 4. Circular dichroism (CD) data for HMGN1 variants. (a) CD spectra in 20 mM potassium phosphate buffer, pH 6.5 at 20 °C, and 20 μM peptide concentration. Spectra are the average of five scans, with baseline spectra (buffer) subtracted.; (b) intensity of the CD spectrum minimum at 222 nm, representing helicity, with increasing concentrations of trifluoroethanol (TFE).

tional features of the four HMGN1 variants **HMGN1_NBD_unmod**, **HMGN1_NBD_pS21**, **HMGN1_NBD_pS25**, and **HMGN1_NBD_pS21pS25** were investigated using NMR spectroscopy, at natural isotope abundance. Spectra (^{1}H, ^{1}H-^{1}H TOCSY, ^{1}H-^{1}H NOESY, ^{1}H-^{13}C HSQC and ^{1}H-^{15}N HSQC) were acquired in 20 mM potassium phosphate buffer, pH 6.5, on 600 MHz or 900 MHz NMR spectrometers to assign the backbone resonances and determine through-space NOE interactions and conformational features (Figure 5 and Supplementary Data S5). Initial analysis of the spectra confirmed the disordered nature of the peptides, with poor dispersion of the amide resonances in the proton dimension, and in agreement with the CD data and previous studies on HMGN1.[27] The phosphoserine residues were easily identified by their characteristic downfield shifts upon phosphorylation (Figure 5c and Supplementary Data S5).[41] The poor dispersion and low sequence diversity made assignment by the sequential assignment method challenging, however, all backbone Hα, Cα and NH resonances could be assigned from the higher-resolution spectra acquired at 900 MHz and by comparison of the four variants (Supplementary Data S4). The secondary Hα and Cα chemical shifts (Figure 5a and b) were calculated to determine whether any of the variants showed secondary structure propensity – indicated by series of positive or negative (>~0.1 ppm for Hα, >~1 ppm for Cα) secondary shifts. The Hα secondary shifts are all <0.06 ppm, with the exception of the phosphorylated serine residues. The random coil shifts for phosphoserine were taken from those determined previously for random coil peptides of posttranslationally modified residues.[41] The Cα secondary shifts are also small (<0.6 ppm), but show more differences between the variants than the Hα secondary shifts, especially in the first half of the sequence, suggesting that phosphorylation does cause some local perturbations within this region. Interestingly, the Cα secondary shifts are mostly positive within the first half of the sequence (typically indicating α-helical tendencies), and mostly negative in the second half (typically indicating β-sheet tendencies).[42] This is consistent with the structures predicted by AlphaFold3, at least in terms of higher α-helical content in the N-terminal half of the NBD. Although the Cα shifts are small in magnitude and therefore it is difficult to draw strong conclusions from them about the secondary structural features, they might represent a promising sensitive experimental parameter for indications of partial secondary structure and subtle conformational changes. Chemical shift perturbations (CSPs) were calculated for the amide chemical shifts (Figure 5c), and show that the phosphorylations have mostly local effects on the neighbouring residues 18–27.

Predictions of the secondary structure content and order parameter (S^2) were made from the HN, NH, Hα and Cα chemical shifts using TALOS-N.[43,44] All variants and residues are predicted to be coil or loop structures and the probabilities of helices (p_H, Figure 5d) are very low, suggesting that the experimental data does not support the extent of helicity predicted in the AlphaFold3 models (Figure 3). This is in keeping with the CD data, indicating that AlphaFold3 over-predicts the helicity of these disordered HMGN1 variants. The predicted S^2 values (most <0.5) also suggest that the NBD is dynamic. Nevertheless, the trend and the location of the helices are supported by the TALOS–N data; Helical content is highest for the unmodified variant, and variations in helicity upon phosphorylation are located in the residues 18–27, suggesting local conformational effects upon phosphorylation.

As expected for IDRs, analysis of the ^{1}H-^{1}H NOESY spectrum did not reveal any long-range interactions. The significant signal overlap made assignment of many of the NOE signals ambiguous and so it was not feasible to obtain sufficient distance restraints to calculate meaningful ensembles of NMR structures. Hα-Hδ NOEs and Cβ, Cγ chemical shifts suggested that all proline residues are in the *trans*-conformation. Predictions of the ^{3}JHNHα coupling constants[45] (Supplementary Data S6) gave values between 5.4 and 7.5 Hz, and fit with the coupling constants that could be measured in the ^{1}H spectra, further supporting the disordered nature of the NBD, with φ angles of 30–90°. Despite the lack of long-range NOEs or strong indicators of secondary structure, HN–HN sequential NOE interactions were observed between Leu24HN-pSer25HN and pSer25NH-Ala26HN for the **HMGN1_NBD_pS25** and **HMGN1_NBD_pS21pS25** variants, but not for the **HMGN1_NBD_unmod** and **HMGN1_NBD_pS21** variants, suggesting that phosphoryla-

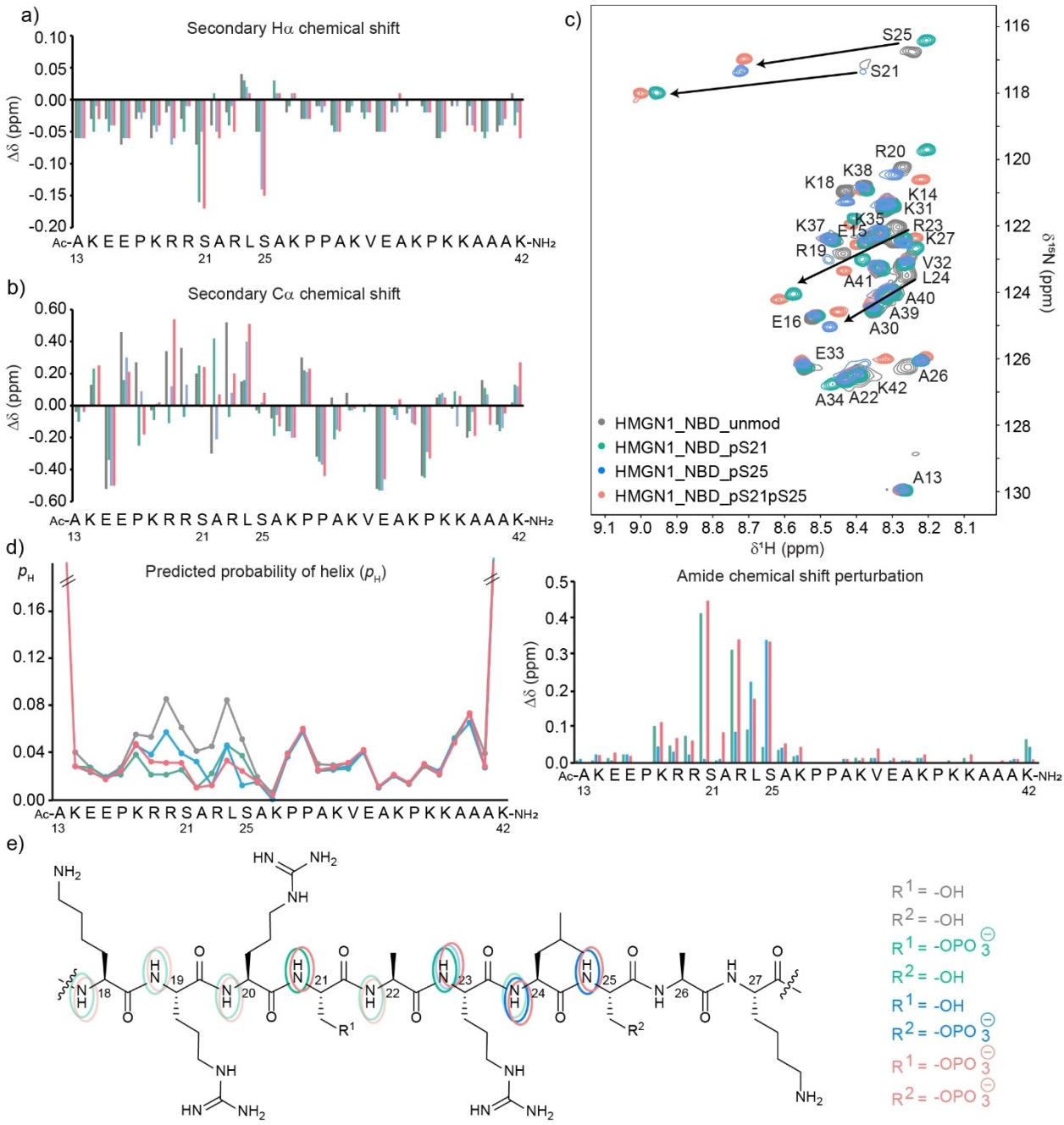

Figure 5. NMR data of HMGN1 NBD variants. (a) Secondary Hα chemical shifts; (b) Secondary Cα chemical shifts plotted against the primary sequence (Full spectra in Supplementary Data S5); (c) Section of the ^{1}H-^{15} N HSQC spectrum showing the backbone amide region and resonances that shift significantly upon phosphorylation of either Ser21, Ser25, or both (black arrows) and chemical shift perturbations (CSPs, Δδ) of the amide resonances; (d) TALOS-N[43] predictions of probability of helix (p_H) for each of the four variants. Terminal residues are by default given a probability of 0.33. (e) Section of the NBD (residues 18–27) shown as a line structure and indicating resonances in the ^{1}H-^{15}N HSQC spectrum that shift to a small (faint ovals) or large (bold ovals) extent upon phosphorylation of either Ser21, Ser25, or both.

tion of Ser25 causes a slight shift in the backbone conformation, in keeping with the predicted structure in Figure 3c.

To investigate the interactions of the HMGN1 NBD variants with the acidic patch of the nucleosome,[28] we took advantage of the possibility to model protein complexes in AlphaFold3. Experimentally, it has been shown that phosphorylation (or mutation to glutamic acid) of Ser21 and Ser25 abolishes binding to the nucleosome,[25,28] however, a model of the hypothetical complex with the modified variants might help to rationalise the mechanism. The histone dimer H2A–H2B contains eight negatively-charged residues (H2A: Glu55, Glu60, Glu63, Asp89, Glu90 and Glu 91; H2B: Glu102 und Glu110) that comprise the acidic patch, and mimics the nucleosome in terms of its interactions with HMGN2.[28,46] We first modelled the H2A–H2B complex, which forms the expected core histone fold (Figure 6a). Each of the HMGN1 NBD variants was then

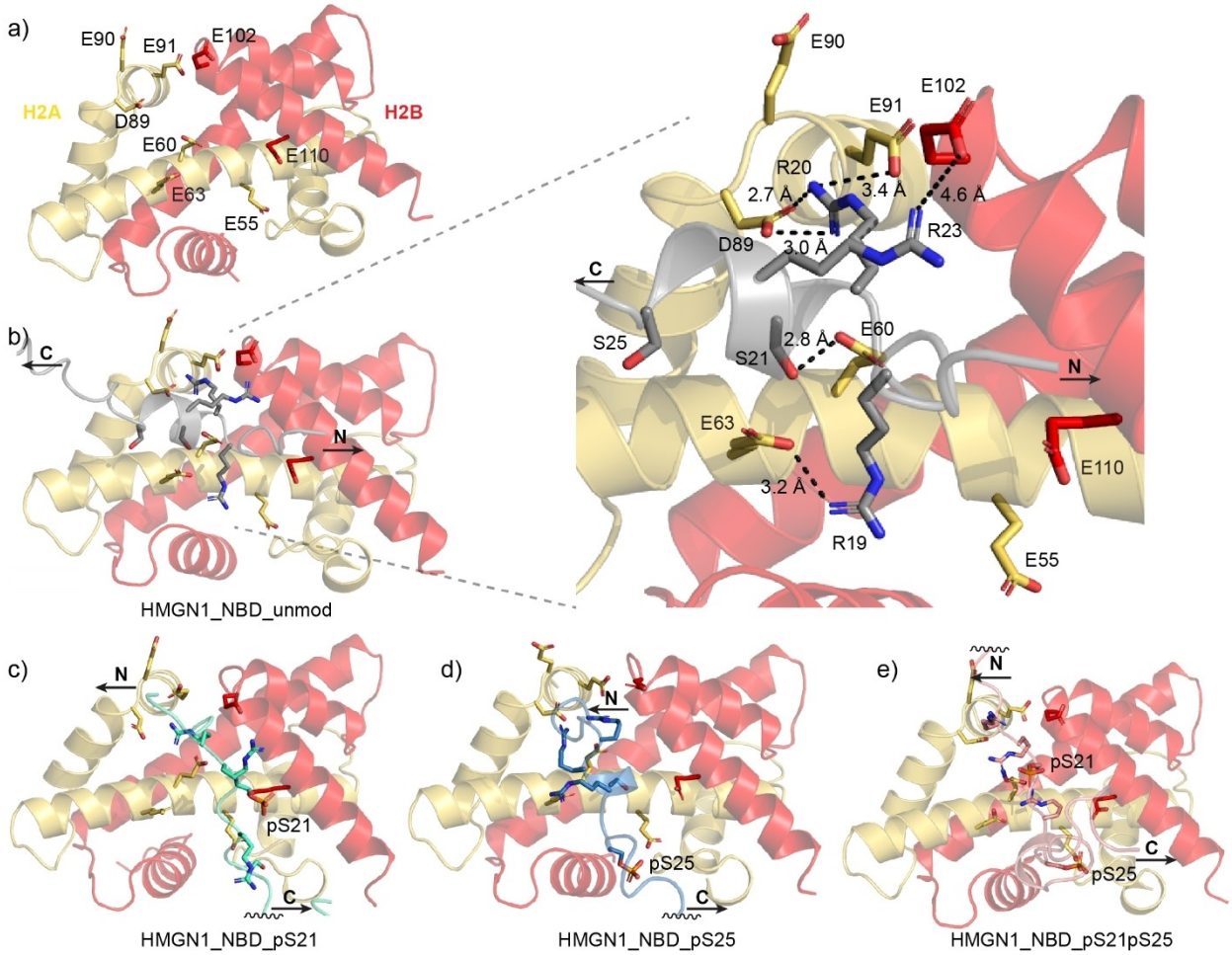

Figure 6. Top-ranked AlphaFold3[31] models of histones H2A (yellow) and H2B (red), forming the histone core fold and the acidic patch, shown in cartoon representation. (a) The histone tails are removed for clarity and the acidic patch residues are labelled and shown in stick representation and coloured by atom type. (b–e) Models of the H2A/H2B dimer with each of the HMGN1 NBD variants with residues 19–25 shown in stick representation and coloured by atom type. Orientations of the HMGN1 NBD peptides are marked by N (N-terminal) and C (C-terminal) and phosphorylated serines are labelled. The inset to panel (b) shows the interactions of unmodified HMGN1 NBD with the acidic patch residues, with potential hydrogen bonding and salt bridge interaction distances marked by dashed lines and distances.

modelled as a ternary complex with H2A–H2B, so that the intra- and intermolecular interactions between all residues in the complex are modeled simultaneously. As shown in Figure 6b–e, all four variants were predicted to interact with the face of the H2A–H2B dimer that forms the acidic patch, although this gives no information about affinities. None of the HMGN1 NBD variants caused a significant alteration of the H2A/H2B histone core fold. In all variants, the arginine residues Arg19, Arg20 and Arg23 orient towards the acidic patch. In the unmodified variant **HMGN1_NBD_unmod**, Ser21 faces towards the acidic patch face of the nucleosome and could potentially form hydrogen bond interactions with Glu60 (Figure 6b inset), in addition to salt bridges formed between the arginine residues and the acidic patch. In previous modelling and mutational experiments on HMGN2, only the side chains of Arg22 (Arg19 in HMGN1) and Arg26 (Arg23 in HMGN1) were involved in interactions with the nucleosome, while Arg23 (Arg20 in HMGN1) faces towards the solvent.[28,29]

The key hydrogen-bonding interactions between the NBD and nucleosome are disrupted upon phosphorylation (Figure 6c–e). In the phosphorylated variants **HMGN1_NBD_pS21**, **HMGN1_NBD_S25** and **HMGN1_NBD_pS21pS25**, the partial helical conformation of residues 19–25 is also disrupted in many models (Supplementary Data S2), presumably because the negatively-charged phosphoserines orient away from the negatively-charged acidic patch. Interestingly, all of the phosphorylated variants interact with the H2A-H2B dimer in the opposite orientation (N→C, left to right in Figure 6c to e) to that of the unmodified variant (C→N, left to right in Figure 6b), which would prevent a functional interaction of the C-terminal CHUD with the DNA at the entry/exit point or with the histone H1 binding site. These results overall agree with the literature in that phosphorylation of the NBD disrupts and causes a decrease in binding to the nucleosome, however the partial helical conformation of HMGN1 predicted makes the structures more compact than the random coil extended structures generated based on NMR and XL–MS data.[28,29]

Conclusions

The synthesis of all four variants of the HMGN1 NBD site-specifically modified by serine phosphorylation enables us to precisely study the effects of phosphorylation on the conformation and potential interactions of this IDR. Taken together, our results show that phosphorylation increases the disorder and decreases the helicity of the NBD. Although phosphorylation does not induce any overall conformational change in the NBD, local conformational changes are seen in the neighbouring residues, as shown by the NOE interactions and chemical shift perturbations. These small and local changes might nevertheless influence the conformational populations sampled by the HMGN1 NBD, and the energy barriers between conformations, thereby destabilising the partial helix. Models of the HMGN1 NBD with the nucleosome acidic patch suggests that electrostatic repulsion between the phosphate group(s) and the acidic patch is the main driving force for the lack of nucleosome binding of the phosphorylated variants, as intrinsic conformational changes seem insufficient to cause a lack of binding. Further studies underway in our group will investigate the effects of phosphorylation on the other domains of HMGN1, and the effects of PTMs and combinations of PTMs on nucleosome interactions in binding studies.

In addition to insights into HMGN conformation and interactions, our results increase the examples and understanding of how models of protein segments (with PTMs) correlate with experimental structural biology data and with newly available models of protein-protein interactions. For this case, we see that the helicity predicted in the models of the HMGN1 NBD is reflected in the experimental data, but to a much smaller extent – as only slight changes in the CD spectra, secondary Cα chemical shifts and helical probability. Using the standard visual representation of cartoon helices, sheets and coils used by modelling approaches and visualisation software, it is difficult to portray conformational ensembles and partial conformations.[11] In contrast, the experimental data give us a more nuanced view of the extent of structural features, but the changes are small for IDRs compared to proteins with defined structures. As AlphaFold and other modelling approaches are trained on larger, structured proteins in the structural databases, our results here suggest that they over-predict the helicity of IDRs, especially of small protein segments. Models of IDRs therefore need to be interpreted with caution, and further studies comparing predicted models of IDRs with experimental data are needed. The data presented here demonstrate some of the structural tools and parameters that can be used to determine partial conformations and conformational changes, but also highlight the ongoing need for sensitive experimental methods to study the conformational populations of IDRs and their changes upon posttranslational modification, as well as the need for better ways to represent these conformational nuances visually.

Materials and Methods

Solid Phase Peptide Synthesis

HMGN1 NBD variants **HMGN1_NBD_unmod**, **HMGN1_NBD_pS21**, **HMGN1_NBD_pS25**, and **HMGN1_NBD_pS21pS25** were synthesised by a combination of automated and manual 9-fluorenylmethoxycarbonyl (Fmoc) based solid phase peptide synthesis (SPPS) on Rink amide resin (loading: 0.57 mmol/g) at 0.033 mmol scale. Residues 22–42 were synthesised at 0.1 mmol scale on a Gyros Chorus automated peptide synthesizer using 5 eq. of amino acid, 5 eq. of diisopropylcarbodiimide (DIC) and 10 eq. of OxymaPure with coupling at 90 °C for 4 min and double couplings for Lys, Val, residues after Pro, and after 20 amino acids. Fmoc deprotection was carried out with 20% piperidine in dimethylformamide (DMF, 2×1 min at 90 °C). After residue 22, the resin was washed, dried and split. Each of the variants was continued manually at 0.033 mmol scale in 10 mL syringes equipped with a polypropylene frit. Residues were double-coupled at 2.5 eq. for 40 min at room temperature, using [benzotriazol-1-yloxy(dimethylamino)methylidene]-dimethylazanium hexafluorophosphate (HBTU, 2.4 eq., 0.5 M in DMF) as activator and diisopropylethylamine (DIPEA, 5 eq.) as base. Phosphoserine residues were introduced *via* Fmoc-Ser(PO-(OBzl)OH)-OH (Fluorochem) and coupled overnight. Fmoc deprotection was carried out with 20% piperidine in DMF 2×5 min at room temperature. After Fmoc deprotection, the N-terminal residue was acetylated using acetic anhydride/DIPEA/DMF 1:2:17 for 2×5 min. Progress of the synthesis was monitored by small-scale test cleavages and analysed by LC–MS. Peptide resins were then washed with DCM and dried under vacuum before cleavage with trifluoroacetic acid (TFA)/triisopropylsilane(TIPS)/H_2O 95:2.5:2.5 for 3 h, followed by precipitation with diethylether and extraction into water, then lyophilisation.

Crude peptides were dissolved in water, filtered and purified using RP-HPLC on a preparative scale C18 column (Phenomenex) on a Waters Autopurify system using a 2.5%/min gradient of acetonitrile (ACN) in water with 0.1% formic acid and a flow rate of 30 mL/min. Fractions were analysed by electrospray ionization mass spectrometry (ESI-MS) and analytical RP-HPLC. If necessary, peptides were repurified on a Knauer HPLC system using a semi-preparative C18 column (Phenomenex) using a 2%/min gradient of ACN in water with 0.05% TFA and a flow rate of 10 mL/min. The cleanest fractions were combined and 8–12 mg of each variant was obtained. The purified peptides were dissolved in 0.1 M HCl and lyophilised twice to exchange TFA salts for HCl salts, and divided into several aliquots.

CD Spectroscopy

Peptide stock solutions in water were quantified using a Nanodrop instrument and their UV absorbance at 214 nm, using an extinction coefficient of 82 534 $mol^{-1} cm^{-1}$, calculated from the absorbance of the amide bonds.[47] Stocks were then normalised to give equal intensity peaks at 214 nm in analytical

HPLC spectra. Circular dichroism spectroscopy was performed in 20 mM H_2KO_4P buffer, pH 6.5, with 0.01% NaN_3 at a protein concentration of 20 μM using a JASCO J-815/150S spectropolarimeter at 20 °C and a path length of 2 mm. Scans were recorded in the range 190 to 260 nm with 1 nm resolution and a scan rate of 50 nm/min and the spectra are the average of five scans after signal correction *via* subtraction of the spectrum of the relevant buffer.

CD spectra of the peptides in increasing concentrations of trifluoroethanol (TFE) were acquired by diluting a 10-fold stock of the peptide with the relevant amounts of 10×buffer, TFE and water to make nine samples of each variant corresponding to TFE concentrations of 0, 10, 20, 30, 40, 50, 60, 70 and 80% TFE. Spectra with the corresponding TFE concentration were subtracted from each of the respective peptide spectra.

NMR Data Acquisition, Assignment and Analysis

Peptide samples (2 mg) were prepared in 20 mM potassium phosphate buffer, pH 6.5, with 10% D_2O, 10 μM DSS, and 0.01% NaN_3. NMR data were acquired on a Bruker Avance III 600 MHz spectrometer with liquid-N_2 cooled cryoprobe or a NEO 900 MHz spectrometer with TCI cryoprobe at the Utrecht NMR Centre with samples at 298 K. For each sample, ^{1}H, ^{1}H-^{1}H TOCSY (isotropic mixing time 60 ms), ^{1}H-^{1}H NOESY (mixing time 200 ms), ^{1}H-^{13}C HMQC and ^{1}H-^{15}N HSQC spectra were acquired. Spectra were processed in TopSpin 4.1.4, including phasing, baseline correction and calibration on the DSS signal (^{1}H at 0 ppm). ^{13}C and ^{15}N dimensions were calibrated indirectly from the ^{1}H dimension, according to IUPAC recommendations using a ratio of 0.251449530 for ^{13}C and 0.101329118 for ^{15}N.[48,49] Assignment of the spectra was carried out in CCPN v3.2.1,[50] using the sequential assignment method and secondary chemical shifts were calculated by subtracting the random coil chemical shifts of the respective amino acid from the observed chemical shift.[51] Random coil chemical shifts of phosphorylated serine were used for residues pS21 and pS25, for the modified variants.[41]

Chemical shift perturbation (CSP) values for NH chemical shifts were calculated using the formula:[52]

$$\Delta\delta\ (\text{ppm}) = \sqrt{\frac{(wN\ \times \Delta\delta N)^2 + (wH\ \times \Delta\delta H)^2}{2}}$$

where the weighting values are wN = 0.158 and wH = 1, and $\Delta\delta N$ and $\Delta\delta H$ indicate the difference in chemical shift (ppm) between HMGN1_NBD_unmod and the various phosphorylated variants.

AlphaFold3 Models

The AlphaFold3[31] server was used to predict the structures of the four HMGN1 NBD variants (HMGN1 residues 13–42), using the default settings. The five models predicted were fitted by matching atoms in Pymol[53] 2.5.2 to the top ranked prediction (model 0) for visualization. For models of the nucleosome acidic patch, canonical histone H2 A (Uniprot: P84051) and H2B (Uniprot: P02283) sequences from *Drosophilia melanogaster (Dm.)*, as used to map the electrostatic potential of the acidic patch,[46] were modelled as a complex, first alone, and then with each of the four HMGN1 NBD variants. In the latter case, the sequences of all three proteins (H2 A, H2B and the relevant HMGN1 variant) were submitted as input so that inter- and intramolecular interactions within the complex were modelled simultaneously. The five predicted models were fitted to the top ranked prediction (model 0) and the histone tails were removed for visualization in Pymol.

Supporting Information Summary

Supporting information is included, showing the sequences of HMGN1–5, number of helical residues in the AlphaFold3 model structures, TALOS-N data, ^{1}H and ^{13}C NMR spectra, NMR chemical shifts, predicted ^{3}JHNHA coupling constants, and TFE titration curves.

Acknowledgements

This research was funded in whole or in part by the Austrian Science Fund (FWF) [grant DOI 10.55776/P36101]. For open access purposes, the author has applied a CC BY public copyright license to any author accepted manuscript version arising from this submission. CD spectroscopy was supported by an Innovative Project (RAKI-MINT) granted by Vienna University of Technology to Astrid R. Mach-Aigner (Institute of Bioprocess Technologies, TU Wien). NMR spectroscopy was carried out at the Utrecht NMR Centre and was supported by iNEXT-Discovery, grant number 871037, funded by the Horizon 2020 program of the European Commission. We thank Dr. Andrei Gurinov, NMR Facility Manager, for assistance.

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Keywords: Intrinsically disordered proteins · NMR spectroscopy · Phosphorylation · Posttranslational modifications · Solid phase peptide synthesis

[1] K. Luger, M. L. Dechassa, D. J. Tremethick, *Nat. Rev. Mol. Cell Biol.* **2012**, *13* (7), 436–447. DOI: 10.1038/nrm3382.
[2] R. K. McGinty, S. Tan, *Chem. Rev.* **2015**, *115* (6), 2255–2273. DOI: 10.1021/cr500373h.

[3] S. Baldi, P. Korber, P. B. Becker, *Nat. Struct. Mol. Biol.* **2020**, *27* (2), 109–118. DOI: 10.1038/s41594-019-0368-x.
[4] G. H. Goodwin, C. Sanders, E. W. Johns, *Eur. J. Biochem.* **1973**, *38* (1), 14–19. DOI: 10.1111/j.1432-1033.1973.tb03026.x.
[5] M. Bustin, R. Reeves, *Prog. Nucl. Acid Res. Mol. Biol.* **1996**, *54*, 35–10. DOI: 10.1016/s0079-6603(08)60360-8.
[6] R. Reeves, *Biochim. Biophys. Acta.* **2010**, *1799* (1–2), 3–14. DOI: 10.1016/j.bbagrm.2009.09.001.
[7] C. L. van Emmerik, H. van Ingen, *Prog. Nucl. Magn. Reson. Spectrosc.* **2019**, *110*, 1–19. DOI: 10.1016/j.pnmrs.2019.01.002.
[8] B. Kohl, X. Zhong, C. Herrmann, R. Stoll, *Nucleic Acids Res.* **2019**, *47* (22), 11906–11920. DOI: 10.1093/nar/gkz614.
[9] M. H. Cho, J. O. Wrabl, J. Taylor, V. J. Hilser *Proc. Natl. Acad. Sci. USA* **2020**, *117* (38), 23606–23616. DOI: 10.1073/pnas.1921473117.
[10] D. Moses, K. Guadalupe, F. Yu, E. Flores, A. R. Perez, R. McAnelly, N. M. Shamoon, G. Kaur, E. Cuevas-Zepeda, A. D. Merg, et al., *Nat. Struct. Mol. Biol.* **2024**, *31*, 283–292, DOI: 10.1038/s41594-023-01148-8.
[11] V. Agarwal, A. C. McShan, *Nat. Chem. Biol.* **2024**, *20*, 950–959, DOI: 10.1038/s41589-024-01638-w.
[12] Y. Postnikov, M. Bustin, *Biochim. Biophys. Acta.* **2010**, *1799* (1–2), 62–68. DOI: 10.1016/j.bbagrm.2009.11.016.
[13] R. Nanduri, T. Furusawa, M. Bustin, *Int. J. Mol. Sci.* **2020**, *21 (2)*, DOI: 10.3390/ijms21020449.
[14] M. Bergel, J. E. Herrera, B. J. Thatcher, M. Prymakowska-Bosak, A. Vassilev, Y. Nakatani, B. Martin, M. Bustin, *J. Biol. Chem.* **2000**, *275* (15), 11514–11520. DOI: 10.1074/jbc.275.15.11514.
[15] Q. Zhang, Y. Wang, *Biochim. Biophys. Acta.* **2010**, *1799* (1–2), 28–36. 10.1016/j.bbagrm.2009.11.009.
[16] J. J. Bonfiglio, P. Fontana, Q. Zhang, T. Colby, I. Gibbs-Seymour, I. Atanassov, E. Bartlett, R. Zaja, I. Ahel, I. Matic, *Mol. Cell* **2017**, *65* (5), 932–940.e936. DOI: 10.1016/j.molcel.2017.01.003.
[17] A. Kumar, V. Narayanan, A. Sekhar, *Biochemistry* **2020**, *59* (1), 57–73. DOI: 10.1021/acs.biochem.9b00827.
[18] D. Moses, F. Yu, G. M. Ginell, N. M. Shamoon, P. S. Koenig, A. S. Holehouse, S. Sukenik, *J. Phys. Chem. Lett.* **2020**, *11* (23), 10131–10136. DOI: 10.1021/acs.jpclett.0c02822.
[19] R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, et al., *Chem. Rev.* **2014**, *114* (13), 6589–6631. DOI: 10.1021/cr400525m.
[20] C. C. Hsu, M. J. Buehler, A. Tarakanova, *Sci. Rep.* **2020**, *10* (1), 2068. DOI: 10.1038/s41598-020-58868-w.
[21] A. C. Conibear, *Nat. Chem. Rev.* **2020**, *4* (12), 674–695. DOI: 10.1038/s41570-020-00223-8.
[22] R. E. Thompson, T. W. Muir, *Chem. Rev.* **2020**, *120* (6), 3051–3126. DOI: 10.1021/acs.chemrev.9b00450.
[23] D. P. Vogl, A. C. Conibear, C. F. W. Becker, *RSC Chem. Biol.* **2021**, *2* (5), 1441–1461. DOI: 10.1039/d1cb00045d.
[24] D. F. Louie, K. K. Gloor, S. C. Galasinski, K. A. Resing, N. G. Ahn, *Protein Sci.* **2000**, *9* (1), 170–179. DOI: 10.1110/ps.9.1.170.
[25] M. Prymakowska-Bosak, T. Misteli, J. E. Herrera, H. Shirakawa, Y. Birger, S. Garfield, M. Bustin, *Mol. Cell. Biol.* **2001**, *21* (15), 5169–5178. DOI: 10.1128/MCB.21.15.5169-5178.2001.
[26] M. Prymakowska-Bosak, R. Hock, F. Catez, J. H. Lim, Y. Birger, H. Shirakawa, K. Lee, M. Bustin, *Mol. Cell. Biol.* **2002**, *22* (19), 6809–6819. DOI: 10.1128/mcb.22.19.6809-6819.2002.
[27] G. Niederacher, D. Urwin, Y. Dijkwel, D. J. Tremethick, K. J. Rosengren, C. F. W. Becker, A. C. Conibear, *RSC Chem. Biol.* **2021**, *2* (2), 537–550. DOI: 10.1039/d0cb00175a.
[28] H. Kato, H. van Ingen, B. R. Zhou, H. Feng, M. Bustin, L. E. Kay, Y. Bai *Proc. Natl. Acad. Sci. USA* **2011**, *108* (30), 12283–12288. DOI: 10.1073/pnas.1105848108.
[29] D. Fasci, H. van Ingen, R. A. Scheltema, A. J. R. Heck, *Mol. Cell. Proteomics* **2018**, *17* (10), 2018–2033. DOI: https://doi.org/10.1074/mcp.RA118.000924.
[30] M. P. Crippa, P. J. Alfonso, M. Bustin, *J. Mol. Biol.* **1992**, *228* (2), 442–449. DOI: 10.1016/0022-2836(92)90833-6.
[31] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, et al., *Nature* **2024**, *630*, 493–500, DOI: 10.1038/s41586-024-07487-w.
[32] T. R. Alderson, I. Pritisanac, D. Kolaric, A. M. Moses, J. D. Forman-Kay *Proc. Natl. Acad. Sci. USA* **2023**, *120* (44), e2304302120. DOI: 10.1073/pnas.2304302120.
[33] D. T. Jones, J. M. Thornton, *Nat. Methods* **2022**, *19* (1), 15–20. DOI: 10.1038/s41592-021-01365-3.
[34] E. T. Usher, M. J. Fossat, A. S. Holehouse, *bioRxiv* **2024**, DOI: 10.1101/2024.06.10.598315.
[35] J. Gavalda-Garcia, B. Dixit, A. Diaz, A. Ghysels, W. F. Vranken, *bioRxiv* **2024**, DOI: 10.1101/2024.07.17.603933.
[36] I. Bludau, S. Willems, W. F. Zeng, M. T. Strauss, F. M. Hansen, M. C. Tanzer, O. Karayel, B. A. Schulman, M. Mann, *PLoS Biol.* **2022**, *20* (5), e3001636. DOI: 10.1371/journal.pbio.3001636.
[37] A. J. Miles, E. D. Drew, B. A. Wallace, *Commun. Biol.* **2023**, *6* (1), 823. DOI: 10.1038/s42003-023-05178-2.
[38] E. W. Martin, A. S. Holehouse, C. R. Grace, A. Hughes, R. V. Pappu, T. Mittag, *J. Am. Chem. Soc.* **2016**, *138*, 15323–15335, DOI: 10.1021/jacs.6b10272.
[39] R. Hancock, Y. Hadj-Sahraoui, *PLoS One* **2009**, *4* (10), e7560. DOI: 10.1371/journal.pone.0007560.
[40] C. M. Davis, J. Deutsch, M. Gruebele, *Protein Sci.* **2020**, *29* (4), 1060–1068. DOI: 10.1002/pro.3833.
[41] A. C. Conibear, K. J. Rosengren, C. F. W. Becker, H. Kaehlig, *J. Biomol. NMR* **2019**, *73* (10–11), 587–599. DOI: 10.1007/s10858-019-00270-4.
[42] D. C. Dalgarno, B. A. Levine, R. J. Williams, *Biosci. Rep.* **1983**, *3* (5), 443–452. DOI: 10.1007/bf01121955.
[43] Y. Shen, A. Bax, *J. Biomol. NMR* **2013**, *56* (3), 227–241. DOI: 10.1007/s10858-013-9741-y.
[44] M. V. Berjanskii, D. S. Wishart, *J. Am. Chem. Soc.* **2005**, *127* (43), 14970–14971. DOI: 10.1021/ja054842f.
[45] Y. Shen, J. Roche, A. Grishaev, A. Bax, *Protein Sci.* **2018**, *27* (1), 146–158. DOI: 10.1002/pro.3292.
[46] H. Zhang, J. Eerland, V. Horn, R. Schellevis, H. van Ingen, *Sci. Rep.* **2021**, *11* (1), 23013. DOI: 10.1038/s41598-021-02436-3.
[47] A. C. Conibear, N. L. Daly, D. J. Craik, *Biopolymers* **2012**, *98* (6), 518–524. DOI: 10.1002/bip.22121.
[48] D. S. Wishart, C. G. Bigam, J. Yao, F. Abildgaard, H. J. Dyson, E. Oldfield, J. L. Markley, B. D. Sykes, *J. Biomol. NMR* **1995**, *6* (2), 135–140.
[49] R. K. Harris, E. D. Becker, S. M. Cabral de Menezes, P. Granger, R. E. Hoffman, K. W. Zilm, *Pure Appl. Chem.* **2008**, *80* (1), 59–84. DOI: 10.1351/pac200880010059.
[50] S. P. Skinner, R. H. Fogh, W. Boucher, T. J. Ragan, L. G. Mureddu, G. W. Vuister, *J. Biomol. NMR* **2016**, *66* (2), 111–124. DOI: 10.1007/s10858-016-0060-y.
[51] D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges, B. D. Sykes, *J. Biomol. NMR* **1995**, *5* (1), 67–81.
[52] M. P. Williamson, *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16. DOI: 10.1016/j.pnmrs.2013.02.001.
[53] L. L. C. Schrodinger, *The pymol molecular graphics system, version 2.5.* **2015**, .

Manuscript received: July 12, 2024
Revised manuscript received: August 26, 2024
Accepted manuscript online: August 26, 2024
Version of record online: October 23, 2024