Effect of phosphorylation in the structural behavior of peptides derived from the intrinsically disordered C-terminal domain of Histone H1.0

Belén Chaves-Arquero,^[a] José M. Pérez-Cañadillas,^[a] and M. Angeles Jiménez^{[a]*}

Departamento de Química-Física Biológica

Instituto de Química Física Rocasolano (IQFR-CSIC)

Serrano 119, 28006-Madrid, Spain

*E-mail: majimenez@iqfr.csic.es

Abstract: To investigate the structural impact of phosphorylation in human H1.0 C-terminal domain, we performed NMR structural studies of model peptides containing a single phosphorylation site: T¹¹⁸-H1.0 (T¹¹⁸PKK motif) and T¹⁴⁰-H1.0 (T¹⁴⁰PVK motif). Both model peptides are mainly disordered in aqueous solution in their non-phosphorylated and phosphorylated forms, but become structured in the presence of trifluoroethanol (TFE). The peptides T¹¹⁸-H1.0 and pT¹¹⁸-H1.0 contain two helical regions: a long amphipathic α -helix spanning residues 104-115 and a short $\alpha/3_{10}$ helix (residues 119-123), which are almost perpendicular in T¹¹⁸-H1.0, but their orientation is poorly defined in pT¹¹⁸-H1.0. Peptides T¹⁴⁰-H1.0 and pT¹⁴⁰-H1.0 form very similar α -helices between residues 141-147. The TPKK and TPVK motifs show the same backbone conformation, but differ in side-chain contacts; Thr and pThr side chains interact with the i+2 Lys side chain in the TPKK motif, and with the i+3 Lys side chain in the TPVK motif. The pT phosphate group in pT¹¹⁸-H1.0 and pT¹⁴⁰-H1.0 has pK_a values below the intrinsic one that can be explained by non-specific charge-charge interactions with nearby Lys. The non-polar Val in the TPVK motif accounts for the pT¹⁴⁰ pK_a being closer to the intrinsic than the pT¹¹⁸ pK_a. Altogether, these results validate minimalist strategies using model peptides that can provide structural details difficult to get in short-lived intrinsically disordered proteins (IDPs) and domains.

Introduction

H1 linker histones bind to DNA regions on the nucleosome surface, are thought to be responsible for chromatin condensation and play a regulatory role in transcription.^[1] Eukaryotic linker histones comprises three domains, of which the middle one is globular, and the N- and the C-terminal ones are intrinsically disordered in their free states.^{[2] [3] [4]} Based on FTIR and CD studies, these N- and C-terminal domains (among them the C-terminal domain of the H1.0 subtype, C-H1.0)^[3] were reported to acquire some secondary structure, including turns, α -helices and β -conformations, in the presence of trifluoroethanol (TFE) and upon DNA-binding.^[3-4] The multiple subtypes of mammalian H1 histones differ in chromatin affinity, genomic localization, expression pattern and post-translational modifications (PTMs).^[5] Their differences in chromatin-binding affinities seems to correlate to their C-terminal domains sequence diversity^[6] and in particular to their net positive charge and the PTMs like Lys acetylation and

We designed two model peptides containing a single phosphorylation site of the three present in C-H1.0 (Figure 1) and a single Pro residue, to avoid multiple *cis/trans* isomerism equilibria that would complicate NMR analysis. Peptide T¹⁴⁰-H1.0 (Ac-KATPVKKAKKK-NH₂) was selected because it is the longest sequence containing the T¹⁴⁰PVK motif and no additional Pro residue (Figure 1), and peptide T¹¹⁸-H1.0 (Ac-SVAFKKTKEVKKVATPKKAAK-NH₂) as representative of the two TPKK phosphorylation motifs contained in C-H1.0 (T¹¹⁸PKK and T¹⁵²PKK; Figure 1). These peptides and their phosphorylated counterparts, pT¹¹⁸-H1.0 (Ac-SVAFKKTKEVKKVA(pT)PKKAAK-NH₂) and pT¹⁴⁰-H1.0 (Ac-KA(pT)PVKKAKKK-NH₂), have their N- and C-termini, respectively, acetylated and amidated to avoid charged-end effects.

Effect of phosphorylation on Pro *cis/trans* isomerism

Since phosphorylation occurs at Pro-preceding Thr residues (T¹¹⁸PKK and T¹⁴⁰PVK motifs; Figure 1), we examined the impact of this PTM in the populations of Pro *cis/trans* isomers. The Pro *cis/trans* isomerism usually occurs in a slow time scale, allowing to observe two distinct sets of NMR signals for the *cis/trans* isomers in Pro-containing peptides. This proved to be the case for the T¹¹⁸-H1.0 peptide in aqueous solution (Figure S1), and for T¹⁴⁰-H1.0 and pT¹⁴⁰-H1.0 peptides in both aqueous solution and in 90 % TFE (Figure S2). In contrast, a single set of NMR signals was observed for peptide T¹¹⁸-H1.0 in 90 % TFE and for pT¹¹⁸-H1.0 in both aqueous solution and 90 % TFE. In all the cases the major (or unique) set of signals correspond to the *trans* X-Pro isomer, characterized by $\Delta\delta_{C\beta\gamma}$ values ($\Delta\delta_{C\beta\gamma} = \delta_{C\beta} - \delta_{C\gamma}$, ppm) in the 4.0-4.6 ppm range^[12] (Tables S1-S8). We fully assigned ¹H and ¹³C chemical shifts of the *trans*-species (Materials and Methods) of the four peptides in aqueous solution and in 90 % TFE (Tables S1-S8). Concerning the *cis* species, residues K¹³⁸ to V¹⁴³ were assigned for peptides T¹⁴⁰-H1.0 and pT¹⁴⁰-H1.0 (Tables S5-S8). Table 1 lists the percentages of *cis* species obtained from the intensity ratios of equivalent *cis* and *trans* cross-peaks. These results show that the populations of the *cis* X-Pro isomers are lower than those reported for other peptides in the literature (7-8 % for AP in model tetrapeptides,^[13]), except for pT¹⁴⁰-H1.0 (Table 1). Interestingly, phosphorylation causes different effects on the peptides: slightly increasing the population of the *cis* X-Pro isomer in the pT¹⁴⁰-H1.0 and apparently decreasing it in the case of the pT¹¹⁸-H1.0 peptide. These slight differences in the *cis/trans* percentages might be biologically important by leading to differential interaction modes with DNA or any other biologically relevant partner like PPlases.

Since the conformational equilibrium is strongly shifted toward the *trans* X-Pro isomers, we will refer exclusively to this from here on, except if the *cis* species are explicitly mentioned

Table 1. Percentages of *cis* species estimated from the intensity ratios of equivalent *cis* and *trans* cross-peaks.

Peptide	Conditions	
	H ₂ O, 5°C	90 % TFE, 25°C
T ¹¹⁸ -H1.0	3.2 ± 0.2 % ^[a] ^[b]	Not detected
pT ¹¹⁸ -H1.0	Not detected	Not detected
T ¹⁴⁰ -H1.0	5 ± 2 % ^[a]	0.9 ± 0.2 % ^[a]
pT ¹⁴⁰ -H1.0	11 ± 4 % ^[a]	10 ± 5 % ^[a]

[a] Errors are standard deviations between all the intensity ratios. [b] Two minor observed signals tentatively assigned to T¹¹⁸ and K¹²⁰ (Figure S1A).

Structural behavior in aqueous solution

Once assigned the NMR spectra of the non-phosphorylated and phosphorylated peptides (see Materials & Methods; Tables S1-S8), we examined the ¹H_α and ¹³C_α conformational shifts ($\Delta\delta_{H\alpha} = \delta_{H\alpha}^{observed} - \delta_{H\alpha}^{RC}$, ppm; and $\Delta\delta_{C\alpha} = \delta_{C\alpha}^{observed} - \delta_{C\alpha}^{RC}$, ppm) in aqueous solution. In general, all the peptides show $\Delta\delta_{H\alpha}$ and $\Delta\delta_{C\alpha}$ values within the random coil range ($|\Delta\delta_{H\alpha}| \leq 0.05$ ppm and $|\Delta\delta_{C\alpha}| \leq 0.4$ ppm) (Figures 2A&B, S3A and S4A) and both phospho-peptides pT¹¹⁸-H1.0 and pT¹⁴⁰-H1.0 show almost identical to their non-phosphorylated counterparts. The only residues displaying $\Delta\delta_{H\alpha}$ and $\Delta\delta_{C\alpha}$ values large in magnitude are T¹¹⁸, pT¹¹⁸, T¹⁴⁰ and pT¹⁴⁰, which are explained as being Pro-preceding residues (positive $\Delta\delta_{H\alpha}$ in the range 0.15-0.30 ppm, and negative $\Delta\delta_{C\alpha}$ ranging from -2.5 to -1.7 ppm;^[14]). When comparing the phosphorylated with the non-phosphorylated versions of the same sequence the main, though very small, differences are located around pT (residue 118 in the pair T¹¹⁸-H1.0 pT¹¹⁸-H1.0, and 140 in the pair T¹⁴⁰-H1.0 and pT¹⁴⁰-H1.0), so that they might be ascribed to the effect of the pT residue on the chemical shifts of their neighbours. In any case, the absence of non-sequential NOEs further shows that the peptides are mainly in random coil in aqueous solution.

In agreement with the NMR data, the CD spectra of all peptides in aqueous solution show a typical random coil strong minimum at about 195 nm (Figure 3) and no other secondary structure features,^[15] confirming that they are predominantly disordered in water.

These results are consistent with data from a non-phosphorylated peptide, which encompasses residues 99-121 of C-H1.0 (Figure 1) and contains the T¹¹⁸PKK phosphorylation site,^[16] as well as from the full-length C-H1.0,^[3, 10] which were shown to be disordered in aqueous solution. Furthermore, we compared the ¹³C_α and ¹³C_β chemical shifts in the non-phosphorylated peptides with those in the non-phosphorylated C-H1.0 domain,^[10] and the equivalent values in phosphorylated peptides with those in the tri-phosphorylated pT-C-H1.0.^[10] The averaged differences $\Delta\delta$ ($\Delta\delta = \delta_{peptide} - \delta_{full-length\ domain}$, ppm) in ¹³C_α and ¹³C_β are quite small (between -0.1 and +0.1 ppm), within the range of experimental errors showing that the isolated peptides are reliable models to mimic the conformational behaviour of H1.0 C-terminal domain.

Figure 2. $\Delta\delta_{H\alpha}$ conformational shifts as a function of peptide sequence: (A) T^{118} -H1.0 (black bars) and pT^{118} -H1.0 (grey striped bars) in aqueous solution 5°C; (B) T^{140} -H1.0 (black bars) and pT^{140} -H1.0 (grey striped bars) in aqueous solution 5°C; (C) T^{118} -H1.0 (black bars) and pT^{118} -H1.0 (grey striped bars) in 90 % TFE at 25°C; and (D) T^{140} -H1.0 (black bars) and pT^{140} -H1.0 (grey striped bars) in 90 % TFE at 25°C. In all panels, pH 5.5 and the two dashed lines indicate the random coil range (RC). An * indicates a non-observed value. Residues belonging to the phosphorylation motif are in bold.

Table 2. $[\theta]^{222nm}$, averaged $\Delta\delta$ values and α -helix populations estimated from $[\theta]^{222nm}$, $\Delta\delta_{H\alpha}$ and $\Delta\delta_{C\alpha}$ (see Methods) for peptides T^{118} -H1.0, pT^{118} -H1.0, T^{140} -H1.0 and pT^{140} -H1.0 in aqueous solution and in 90 % TFE at pH 5.5 at 25°C, and for the full. N- and C-terminal residues were excluded to calculate the averaged $\Delta\delta$ values.

Peptide	Conditions	$[\theta]^{222nm}$, deg.cm ² .dmol ⁻¹	% helix ^[b] from $[\theta]^{222nm}$	Helix length ^[b]	$\Delta\delta_{H\alpha}$, ppm	% α -helix ^[b] from $\Delta\delta_{H\alpha}$	$\Delta\delta_{C\alpha}$, ppm	% α -helix ^[a] from $\Delta\delta_{C\alpha}$	Averaged % α -helix ^[b,c]
T^{118} -H1.0	H ₂ O	-69.9	8	105-115	-0.05 ^[a]	13 ^[a]	+0.12 ^[a]	4 ^[a]	9 ± 5 ^[a]
	90 % TFE	-11229.6	37		-0.29	75	+3.35	100 ^[a]	87 ± 13
pT^{118} -H1.0	H ₂ O	86.4	7	105-115	-0.05 ^[a]	13 ^[a]	+0.15 ^[a]	5 ^[a]	9 ± 4 ^[a]
	90 % TFE	-10988.9	36		0.27	66	+2.92	95 ^[a]	81 ± 14
T^{140} -H1.0	H ₂ O	-915.8	10	141-147	-0.06 ^[a]	16	+0.21 ^[a]	7 ^[a]	12 ± 5 ^[a]
	90 % TFE	-7137.4	26		-0.15	39	+1.58	51 ^[a]	45 ± 6
pT^{140} -H1.0	H ₂ O	1175.9	5	141-147	-0.06 ^[a]	16	+0.28 ^[a]	9 ^[a]	13 ± 4 ^[a]
	90 % TFE	-8168.4	29		-0.16	40	+1.80	58 ^[a]	49 ± 9

[a] Values measured at 5°C. [b] Notice that CD-estimated helix percentages are an average for all the peptide residues, whereas NMR-estimated helix percentages are for the residues within the helix. [c] Reported errors are standard deviations for the mean of the percentages obtained from $\Delta\delta_{H\alpha}$ and $\Delta\delta_{C\alpha}$ values.

Structural behavior in the presence of trifluoroethanol (TFE)

The use of co-solvents like TFE can unmask low-populated conformations in aqueous solution.^[17] Previous data showed that full-length C-H1.0 and a related non-phosphorylated peptide, which contains the T^{118} PKK motif (residues 99-121; Figure 1), increase their helical content in the presence of TFE.^[3]

^{16]} This prompted us to examine if our C-H1.0 model peptides would behave likewise and, more interestingly, to investigate the effect of phosphorylation in the presence of TFE.

The addition of TFE causes an increase of structure on the four peptides according to the CD spectra, with a progressive conversion of the 195 nm minima into a maximum at about 197 nm. At 90 % TFE (Figure 3), the peptides show minima at positions close to those characteristic of helices (208 nm and 222 nm).^[15] To better characterise these structures, we proceeded to study the four peptides in 90 % TFE by NMR.

Figure 3. CD spectra of peptides T¹¹⁸-H1.0 (A), pT¹¹⁸-H1.0 (B), T¹⁴⁰-H1.0 (C) and pT¹⁴⁰-H1.0 (D) in aqueous solution (dotted line) and in 90 % TFE (black line) at pH 5.5 and 25 °C.

The profiles of $\Delta\delta_{H\alpha}$ and $\Delta\delta_{C\alpha}$ values in 90 % TFE are practically identical between the corresponding non-phosphorylated and phosphorylated peptides (Figures 2C-D, S3 and S4), but with magnitudes substantially larger than in aqueous solution and outside the random coil range. Thus, the pair of peptides T¹¹⁸-H1.0 and pT¹¹⁸-H1.0 shows a stretch of negative $\Delta\delta_{H\alpha}$ and positive $\Delta\delta_{C\alpha}$ values extending residues A¹⁰⁵ to V¹¹⁶ (Figure 2C and S3B), which indicates that they form helical structures in that region.^[18] A similar pattern found for peptides T¹⁴⁰-H1.0 and pT¹⁴⁰-H1.0 evidences the presence of an helix spanning from P¹⁴¹ to K¹⁴⁷ (Figures 2D and S4B). It is worth to note that the profiles of the four peptides in aqueous solution follow the same pattern that in 90 % TFE (same signs), except for their very small magnitudes (Figure 2), which led us to propose that a low population of helical structures

(similar to those in TFE) might exist in water. This reinforces the conclusions extracted from the TFE study. Further and stronger evidence about the four peptides forming helical structures in 90 % TFE comes from the fact that they show some helix-characteristic NOEs, such as medium-range $\alpha\text{N}(i, i+2)$, $\alpha\text{N}(i, i+3)$, $\alpha\text{N}(i, i+4)$ and $\alpha\beta(i, i+3)$, and intense sequential $\text{NN}(i, i+1)$.

The helix populations can be estimated from the $\Delta\delta_{\text{H}\alpha}$ and $\Delta\delta_{\text{C}\alpha}$ averaged for the helical residues (see Materials and Methods). As seen in Table 2, the helix spanning residues 105-115 is highly populated in $\text{T}^{118}\text{-H1.0}$ and $\text{pT}^{118}\text{-H1.0}$ in 90 % TFE (81-87 % at pH 5.5 and 25 °C). The population of the helix formed by $\text{T}^{140}\text{-H1.0}$ and $\text{pT}^{140}\text{-H1.0}$ in 90 % TFE is not so high (45-49 % at pH 5.5 and 25 °C). The helix contents estimated by CD (Table 1) are lower than by NMR. The discrepancies can be explained by taking into account that CD estimates overall helix populations including all residues of the peptides, while NMR inform about specific segments (105-115 *versus* 103-124 in $\text{T}^{118}\text{-H1.0}$ and $\text{pT}^{118}\text{-H1.0}$; and 141-147 *versus* 138-141 in $\text{T}^{140}\text{-H1.0}$ and $\text{pT}^{140}\text{-H1.0}$). Another possible error source for the CD data is the uncertainty in the concentration measurements of these aromatic-less peptides.

Applying the same procedure the helix percentages present in the conformational ensemble equilibrium of these peptides in aqueous solution are very small, in the range 7-13 % at pH 5.5 and 5 °C (Table 1). Based on the $\Delta\delta_{\text{C}\alpha}$ averages (obtained from the chemical shifts reported in [10]) the helix percentages for the segments 105-115 and 141-147 in the full-length C-H1.0 and pT-C-H1.0 in aqueous solution are quite similar (4-8 % at pH 5.5 and 25 °C; Table S9) to those in the isolated peptides. Comparing once again non-phosphorylated and phosphorylated forms, it is noticeable that the differences in helix populations between the two forms in the pairs $\text{T}^{118}\text{-H1.0/pT}^{118}\text{-H1.0}$, and $\text{T}^{140}\text{-H1.0/pT}^{140}\text{-H1.0}$ are in the range 1-6 %, which is within the experimental error (3-7 %).^[19] Overall this data is consistent with the peptides, and also the full-length C-terminal domain, existing in aqueous solution as disordered ensembles, which contain residual helical populations in segments 105-115 and 141-147.

We performed structure calculations using the distance and angle restraints derived from the NMR parameters observed in 90 % TFE (see Methods and Table S9). The resulting structures are of high quality according their restrains violations and Ramachandran statistics (Table S10). The structure of $\text{T}^{118}\text{-H1.0}$ (Figures 4A-C and S5A) is well defined (backbone RMSD = 0.7 ± 0.3 Å; Table S10), and shows two regular helices: a long α -helix extending residues $\text{V}^{104}\text{-K}^{115}$, approximately coincident with that identified by qualitative analysis of $\Delta\delta_{\text{H}\alpha}$ and $\Delta\delta_{\text{C}\alpha}$ values (see above), and a very short helix spanning approximately residues $\text{P}^{119}\text{-A}^{123}$, which is classified as α or 3_{10} depending on the conformer. These two helical regions are connected by a turn/loop at residues $\text{A}^{117}\text{-T}^{118}$. The population of this second short helix would be 9 ± 4 % in 90 % TFE, just above experimental uncertainty. The angle between the helices is relatively well defined, in the range $98^\circ \pm 8^\circ$. The long α -helix 104-115 (Figure 4B) exhibits an amphipathic character, having all Lys side chains pointing towards the same helix face, except for K^{108} and K^{115} , which form salt bridges with the side chain of E^{112} . This helix might be stabilised by the $\text{K}^{108}/\text{E}^{112}$ and $\text{E}^{112}/\text{K}^{115}$ salt bridges and an interaction between F^{106} and K^{110} side chains (pi-cation and/or hydrophobic between aromatic ring and Lys long aliphatic chain). Further evidence for this later interaction comes from the deviation from random coil values of some K^{110} side chain protons,

presumably caused by the Phe aromatic ring effects. Concerning the short C-terminal helix, it involves the phosphorylation motif (T¹¹⁸PKK), and T¹¹⁸ or/and P¹¹⁹ could be acting as N-caps. The conformation of this motif might facilitate access of the cyclin-dependent kinases to the Thr hydroxyl to be phosphorylated.

Like in the non-phosphorylated peptide T¹¹⁸-H1.0, the structure of the phosphorylated pT¹¹⁸-H1.0 shows two well-defined helices (Figure 4D-F), a long α -helix at segment 104-115 (backbone RMSD= 0.6 ± 0.2 Å) and a short α_3 at residues 119-123 (backbone RMSD= 0.3 ± 0.1 Å), but their relative disposition is not defined (backbone RMSD₁₀₄₋₁₂₃= 1.9 ± 1.0 Å and Figure S5B). The long N-terminal helix in pT¹¹⁸-H1.0 presents the same stabilising interactions (K¹⁰⁸/E¹¹² and F¹⁰⁶/K¹¹⁰; Figure 4D-E) that in T¹¹⁸-H1.0. With respect to the C-terminal short helix its architecture is quite similar in phosphorylated (Figure 4F) and non-phosphorylated peptides (Figure 4C), but with the presence of a salt-bridge between phosphate group of pT¹¹⁸ and the side chain of K¹²⁰ in pT¹¹⁸-H1.0. A putative pT¹¹⁸/K¹²¹ salt-bridge can be discarded because the distance between pT¹¹⁸ and K¹²¹ side chains is not close enough.

In summary, the structures of T¹¹⁸-H1.0 and pT¹¹⁸-H1.0 show similar secondary structure elements but differ in their relative arrangement, which is relatively fixed in T¹¹⁸-H1.0, and poorly defined in pT¹¹⁸-H1.0 (Figure 4G). A plausible explanation for this result is that salt-bridge contacts between E¹¹² and K¹²⁰/K¹²¹ might contribute to maintain the orientation between the two helices in T¹¹⁸-H1.0. In the phosphorylated pT¹¹⁸-H1.0 these lysine sidechains might probably contact more favourably with pT¹¹⁸ losing the tertiary contacts between the two helical sections. This conformational difference, together with the negative charge of the phosphate group, might contribute to the lower DNA affinity of the phosphorylated full-length C-H1.0 relative to the non-phosphorylated one.^[10]

Concerning the structures of peptides T¹⁴⁰-H1.0 and pT¹⁴⁰-H1.0, they showed short and well-defined α -helices spanning residues 141-147 (Figure 5A-D; Table S10). Considering the 40 conformers, from both ensembles, the RMSD value raises only slightly to 0.6 ± 0.2 Å, demonstrating that, in this case, the structures adopted by the non-phosphorylated and the phosphorylated peptides are very similar (Figure 5E). The side chains of residues pT¹⁴⁰ and K¹⁴³ seem to interact (Figure 5D).

Given their central role, it is interesting to compare the conformations of the T¹¹⁸PKK and T¹⁴⁰PVK phosphorylation motifs. Their backbone topology is very similar in the four peptides, likely determined by the P residue. However, T and pT residues show different interactions with the i+2 and i+3 residues. Thus T¹¹⁸ and pT¹¹⁸ interact with the side chain of the i+2 Lys (K¹²⁰ in Figure 4F), but not with the i+3 Lys, and T¹⁴⁰ and pT¹⁴⁰ interacts with the side chain of the i+3 Lys (K¹⁴³ in Figure 5D), more than with the i+2 Val. These differences might be relevant for the different biological roles of the various phosphorylation motifs present in the Histone H1.0 C-terminal domain.

Figure 4. Structures of T¹¹⁸-H1.0 (A-C) and pT¹¹⁸-H1.0 (D-G) in 90 % TFE. Structural ensemble of the 20 lowest target function conformers calculated for T¹¹⁸-H1.0 (A) and pT¹¹⁸-H1.0 (D). pT¹¹⁸-H1.0 superposition is done for backbone atoms of residues 104-115. Backbone atoms are in black, and side chains in blue for K, in green for non-polar residues (A, P and V), in magenta for S and T, and in red for pT. Representative structures for T¹¹⁸-H1.0 (B) and pT¹¹⁸-H1.0 (E) with the backbone shown as a ribbon and the side chains in neon. The N-terminal long helix is in red and the short C-terminal in gold. Superposition of the 20 lowest target function conformers for the backbone atoms of residues 117-124 for T¹¹⁸-H1.0 (C) and pT¹¹⁸-H1.0 (F). (G) Superposition of the backbone atoms for helix 1 in T¹¹⁸-H1.0 (in magenta) and pT¹¹⁸-H1.0 (in black). N- and C-termini are labelled in all the panels, and side chains in panels A-B and D-E. The inset gives the pairwise RMSD values for the backbone atoms of the structural ensembles.

Figure 5. Structures of peptides T¹⁴⁰-H1.0 (A-B) and pT¹⁴⁰-H1.0 (C-D) in 90 % TFE. Structural ensemble of the 20 lowest target function conformers calculated for peptides T¹⁴⁰-H1.0 (A) and pT¹⁴⁰-H1.0 (C). Backbone atoms are in black, and side chains in blue for K, in green for non-polar residues (A, P and V), in magenta for T and in red for pT. Representative structures for T¹⁴⁰-H1.0 (B) and pT¹⁴⁰-H1.0 (D) with the backbone shown as a ribbon and the side chains in neon. (E) Superposition of the backbone atoms of the structural ensembles for T¹⁴⁰-H1.0 (in magenta) and pT¹⁴⁰-H1.0 (in black). N- and C-termini are labelled in all the panels, and side chains in panels A-D. The inset gives the pairwise RMSD values for the backbone atoms of residues 139-147 of the structural ensembles.

Effect of pH on the conformational behaviour of the phosphorylated peptides

The protonation state of ionisable side chain groups can significantly affect the stability of peptide structures.^[20] Both pT¹¹⁸-H1.0 and pT¹⁴⁰-H1.0 peptides contain a pT phosphate group presumably titrating around the physiological pH, and hence we investigate the effect of the ionization state of pT on the conformational behaviour of these peptides. In model non-structured peptides the pK_a value for protonation from di-anionic to mono-anionic of pT phosphate has been reported to be 6.30±0.07^[21] or 6.1±0.1.^[22] Therefore, we monitored the changes in chemical shifts upon pH titration in the range of 2.5 to 9.0. Apart from the pT phosphate, only the carboxylate of E¹¹² group (intrinsic pK_a is 4.25±0.05)^[23] is expected to titrate within this pH range (Figure 1). We also examined it in the non-phosphorylated T¹¹⁸-H1.0.

pH titrations were followed in series of ¹H,¹H-TOCSY and ¹H,¹³C-HSQC spectra of peptides T¹¹⁸-H1.0, pT¹¹⁸-H1.0 and pT¹⁴⁰-H1.0 in aqueous solution at 25°C (see Methods). Most of ¹H and ¹³C chemical shifts could be readily assigned by comparisons with the assignment done at pH 5.5 (Tables S1, S3 and S7). However, the H^N signals start to disappear due to solvent exchange at about pH > 7.0, and none are observed at pH 9.0. The assignments were facilitated by the fact that many signals remained almost

invariable within all the examined pH range. Indeed, the number of carbons and protons suffering a large change in chemical shift between the extreme pH values ($|\Delta\delta^{13C}| > 1$ ppm; $|\Delta\delta^1H| > 0.1$ ppm) or even a small change ($0.03 \text{ ppm} \leq |\Delta\delta^1H| \leq 0.1$ ppm) is quite small in the three peptides (Tables S11-13). As expected, the nuclei displaying the largest changes include those belonging to the residue bearing the titrating group at its side chain: $^{13}C_\gamma$ and $^{13}C_\beta$ carbons and $^1H_{\gamma 2}$ and $^1H_{\gamma 3}$ of E¹¹² in peptides T¹¹⁸-H1.0 and pT¹¹⁸-H1.0; and H^N, H_α, and H_β of pT¹¹⁸ and pT¹⁴⁰ in peptides pT¹¹⁸-H1.0 and pT¹⁴⁰-H1.0, respectively. The chemical shifts of these nuclei showed a sigmoidal dependence with pH (Figures S6-S8). The pK_a values for the E carboxylate and pT phosphate groups were obtained by fitting these sigmoidal pH-dependences to equation [7] (see Methods) (Table 3). The chemical shifts of the H^N and H_α of pT¹⁴⁰ in the *cis* species of pT¹⁴⁰-H1.0 are also pH-dependent. Their $|\Delta\delta^1H|$ are smaller than in the *trans* species (Table S12), but large enough for getting the pK_a value from their δ /pH curves. As far as we know, it is the first time that the pK_a for the phosphate group of a Pro-preceding pT residue has been determined for the *cis* species. This pK_a value is very close to that of the *trans* species (Table 3). Strikingly, all the resulting pK_a values are smaller than the intrinsic ones (Table 3), particularly in the case of peptides T¹¹⁸-H1.0 and pT¹¹⁸-H1.0. Thus, the pK_a of E¹¹² is 0.85-0.88 units below the intrinsic value for E carboxylates,^[23] and those of the pT phosphate groups are 0.66 units below in pT¹¹⁸ and 0.26 units below in pT¹⁴⁰.^[21] These differences cannot be attributed to experimental errors in pH measurements because they are about ± 0.1 units. It seems reasonable to assume that ionisable groups in these mainly random peptides should behave as those in protein surfaces, for which charge-charge interactions have been described as the main source for pK_a perturbations.^[20] Then, non-specific interactions with the large number of Lys residues present in the sequences of peptides T¹¹⁸-H1.0, pT¹¹⁸-H1.0 and pT¹⁴⁰-H1.0 can account for the decrease in the pK_a values of both E carboxylate and pT phosphate groups (Figure 1); the net charge of the peptides in the examined pH range is positive. That the pK_a values of E¹¹² are almost identical in T¹¹⁸-H1.0 and pT¹¹⁸-H1.0 (Table 2) indicates that the negatively charged phosphate group of pT¹¹⁸ has no detectable effect onto E¹¹². This is not surprising considering that they are separated by 5 residues and the large number of positively charged K residues in the vicinity of E¹¹² (Figure 1). Concerning the fact that the decrease in the pK_a of pT¹¹⁸ is quite larger than in that of pT¹⁴⁰, we can speculate that this difference might be caused by the i+2 residue, which is K in pT¹¹⁸-H1.0 and V in pT¹⁴⁰-H1.0 (Figure 1). The V¹⁴² side chain is likely creating a non-polar and less-positively charged environment around the pT¹⁴⁰ phosphate, and non-polar surroundings favoured neutral or less-charged states of ionisable groups, and consequently increase the pK_a values.^[20] In the case of pT¹⁴⁰, the increase due to the non-polar V¹⁴² side chain partially compensates the decrease due to non-specific charge-charge interactions, and as a consequence the pK_a is only slightly reduced relative to intrinsic value. On the other hand, the K¹²⁰ side chain contributes to create a positively charged surrounding for the pT¹¹⁸ phosphate, and so its pK_a is greatly reduced.

Once determined the pK_a values, we proceeded to see whether the protonation state of the phosphate group affects the conformational behaviour of peptides pT¹¹⁸-H1.0 and pT¹⁴⁰-H1.0 in aqueous solution. To that end, we compared the plots of conformational shifts for the $^1H_\alpha$ protons and $^{13}C_\alpha$ carbons obtained at pH 3.5-4.0 (mono-anionic state of phosphate group) and at pH 7.5 (di-anionic state of phosphate group) in aqueous solution at 25°C (Figure S9). Excluding pT¹¹⁸, pT¹⁴⁰ and A¹¹⁷ in pT¹¹⁸-H1.0,

the profiles of $\Delta\delta_{H\alpha}$ and $\Delta\delta_{C\alpha}$ values in the two peptides are essentially equal at low and high pH values (Figure S9). This indicates that the ionisation state of the phosphate group hardly affects the conformational behaviour of peptides pT¹¹⁸-H1.0 and pT¹⁴⁰-H1.0, which remain disordered in aqueous solution regardless the pH value.

Table 3. pK_a values for the E¹¹² carboxylate and pT phosphate groups in peptides T¹¹⁸-H1.0, pT¹¹⁸-H1.0 and pT¹⁴⁰-H1.0 obtained by fitting the pH dependence of the chemical shifts measured in aqueous solution at 25 °C to equation [9].

Peptide	Residue	Resonance ^[a]	Titration group	pK _a ^[b]	pK _a ^{RC}
T ¹¹⁸ -H1.0	E ¹¹²	H _{γ2} , H _{γ3} , C _β , C _γ	E ¹¹²	3.40±0.01	4.25±0.05
pT ¹¹⁸ -H1.0	E ¹¹²	H _{γ2} , H _{γ3} , C _β , C _γ	E ¹¹²	3.37±0.01	4.25±0.05
pT ¹¹⁸ -H1.0	pT ¹¹⁸	H ^N , H _α , H _β	pT ¹¹⁸	5.64±0.01	6.30±0.07
pT ¹⁴⁰ -H1.0	pT ¹⁴⁰	H ^N , H _α , H _β	pT ¹⁴⁰	6.04±0.01	6.30±0.07
<i>cis</i> pT ¹⁴⁰ -H1.0	pT ¹⁴⁰	H ^N , H _α	pT ¹⁴⁰	6.10±0.02	6.30±0.07

[a] Nuclei with large changes in chemical shift between low and high pH. [b] Values for joint fitting of the curves for the nuclei with large changes in chemical shift between low and high pH.

Next, we examined the distribution of residues that sense the ionisation state of E¹¹², pT¹¹⁸ and pT¹⁴⁰. Charge effects depend on the distance to the ionisable group and therefore the pH-dependency of chemical shift perturbations can provide another source of structural insights. We consider that a residue sense the ionisable group if the chemical shifts of one or more of its protons is pH dependent ($|\Delta\delta^{1H}| > 0.03$ ppm). Surprisingly the distribution of sensing residues does not correlate to the sequential proximity, as it would be expected for a fully disordered peptide (Figure 6 and Tables S11 & S13), instead it resembles some of the features of the structures in TFE. For example, the distribution of residues feeling the E¹¹² protonation state in peptide T¹¹⁸-H1.0 recalls the periodicity of an α -helix, suggesting that the low populated helical conformations present in conformational ensemble contribute to this effect (Table 2). In pT¹¹⁸-H1.0, the residues located between E¹¹² and pT¹¹⁸ seems to feel the two ionisable groups, though the $\Delta\delta^{1H}$ changes are too small to allow fitting to two pK_a values. On the other hand, it is noticeable that the effect of the phosphate ionization state distributes differently among the protons in pT-neighbour residues in pT¹¹⁸-H1.0 (TPKK motif) and in pT¹⁴⁰-H1.0 (TPVK motif; Figure 6); for example, considering the pT-following residue, which is P in the two peptides, the H_{δ3} of P¹¹⁹ shows a large $\Delta\delta^{1H}$ (0.13 ppm), but it is the only pH-dependent proton in P¹¹⁹, whereas H_{β3}, H_{δ2} and H_{δ3} of P¹⁴¹ are pH-dependent, but the magnitude of their $\Delta\delta^{1H}$ are smaller (Tables S11-S12). Thus, these differences in pH-dependence might reflect preferred local conformations in and around pT in the phosphorylated TPKK and TPVK motifs, which are different between them, and undetectable by other NMR parameters, such as the profiles of $\Delta\delta_{H\alpha}$ (Figure 2A-B) and $\Delta\delta_{C\alpha}$ (Figures S3-S4). As a cautionary note, it has to be mentioned that the sequences preceding the motifs, which is shorter in pT¹⁴⁰-H1.0, might also be affecting their conformation. But, having a native longer N-terminal sequence in pT¹⁴⁰-H1.0 implies to incorporate additional Pro residues (Figure 1), which would complicate NMR analysis due to multiple *cis/trans* equilibria.

Figure 6. Peptide sequences and residues affected by the ionisation state of E carboxylate (red) and pT phosphate (green) groups. In peptide pT¹¹⁸-H1.0, residues affected by the two groups are in magenta.

Figure 7. Effect of pH on the conformational shifts of peptides pT¹¹⁸-H1.0 (A) and pT¹⁴⁰-H1.0 (B) in 90 % TFE at 25°C. (A) and (B) panels show $\Delta\delta_{\text{H}\alpha}$ values, and (C) and (D) $\Delta\delta_{\text{C}\alpha}$. In all panels, grey, black and striped bars correspond to values at pH 4.0, 5.5 and 7.5, respectively; and the two dashed lines indicate the random coil range (RC)

Finally, we examined whether the protonation state of the phosphate group affects the structures formed by pT¹¹⁸-H1.0 and pT¹⁴⁰-H1.0 in the presence of TFE. To that end, as it was done in aqueous solution, we compared the plots of $\Delta\delta_{H\alpha}$ and $\Delta\delta_{C\alpha}$ at pH 4.0, 5.5 and 7.5 in 90 % TFE at 25 °C (Figure 7). In the case of peptide pT¹¹⁸-H1.0, the two helices formed in 90 % TFE are affected differently by the ionisation state of pT¹¹⁸: the long N-terminal helix (residues 105-115; Figure 5) is unaffected, which is not surprising since pT¹¹⁸ lies outside the helix; and the short C-terminal helix (residues 121-123) is stabilised at pH 7.5, as shown by the increase in the magnitudes of $\Delta\delta_{H\alpha}$ and $\Delta\delta_{C\alpha}$ values (Figure 7A & C). The population of this helix is almost twice at pH 7.5 relative to pH 4.0 (28 % and 15 %, respectively; Table S14). That the di-anionic state of the phosphate group placed at the N-end is more stabilising than the mono-anionic state can be accounted for by (i) the well-known stabilising effect of negative charges at the N-end of helices because of the interaction with the helix dipole and (ii) an interaction between the negatively charged pT¹¹⁸ phosphate group and the positively charged K¹²⁰/K¹²¹ side chains, which could act as a kind of stabilising N-capping motif, and becomes stronger upon increasing the negative charge of the phosphate. Based on the distances of the pT¹¹⁸/K¹²⁰ and pT¹¹⁸/K¹²¹ side chains in the structures calculated for pT¹¹⁸-H1.0 (Figure 5E-F), we could speculate that the *i,i+2* pT¹¹⁸/K¹²⁰ interaction contributes more than the *i,i+3* pT¹¹⁸/K¹²¹.

In peptide pT¹⁴⁰-H1.0 in 90 % TFE, the small increments in the magnitudes of the $\Delta\delta_{H\alpha}$ and $\Delta\delta_{C\alpha}$ values in the segment 141-147 (Figure 7B & D) indicate an increase in the α -helix population at the highest pH (from 45 % at pH 4.0 to 57 % at pH 7.5 at 25 °C; Table S13). This difference in helix population might be explained by the helix dipole (see above), and by a stabilising *i* to *i+3* interaction between the negatively charged pT¹⁴⁰ phosphate group and the positively charged K¹⁴³ side chain, which becomes more effective when the negative charge at the phosphate group increases; from anionic at pH 4.0 to di-anionic at pH 7.5. In concordance with this, the pT¹⁴⁰ and K¹⁴³ side chains are close in the structures calculated for pT¹⁴⁰-H1.0 (Figure 6).

Conclusions

We have examined the structural behaviours of two C-H1.0-derived peptides, T¹¹⁸-H1.0 and T¹⁴⁰-H1.0, which contain phosphorylation motifs TPKK and TPVK, respectively, and compared them with those in their phosphorylated counterparts (pT¹¹⁸-H1.0 and pT¹⁴⁰-H1.0). Given that Pro *cis/trans* isomerism has been hypothesized to be important for DNA-binding of H1.0,^[8] we have analysed it finding that the *cis* percentages are different in the T¹¹⁸PKK and T¹⁴⁰PVK motifs, and that, upon phosphorylation, they diminish in the T¹¹⁸PKK motif and increments in the T¹⁴⁰PVK motif. In any case, the *cis* percentages are quite low (< 10%), so that our NMR study was focussed on the structural behaviour of the major *trans* species. It should be mentioned that no information about the *cis/trans* isomerism was obtained from the NMR study of the full-length C-H1.0,^[10] whose sequence contains 12 Pro residues in total.

CD and NMR data evidenced that the four peptides in aqueous solution are mainly random coil with some residual helical tendencies, which are enhanced in the presence of 90 % TFE. These results are

similar to those on the full-length C-terminal H1.0 domain (C-H1.0),^[10] showing the usefulness of model peptides to obtain structural insights about IDPs.

In the presence of TFE, the structures of T¹¹⁸-H1.0 and pT¹¹⁸-H1.0 consist of two helical regions: a long amphipathic α -helix (residues 104-115) and a short $\alpha/3_{10}$ helix (residues 119-123). But, the orientation between the helices is well defined and approximately perpendicular only in the non-phosphorylated T¹¹⁸-H1.0. The peptides T¹⁴⁰-H1.0 and pT¹⁴⁰-H1.0 form a single α -helix (residues 141-147). These results agree with previous CD data on the full-length C-H1.0, which indicated helix formation in the presence of TFE.^[3]

The phosphorylation motifs T¹¹⁸PKK and T¹⁴⁰PVK displayed the same backbone conformation in the four peptides in the presence of TFE. However, the side-chain contacts are different, with the T¹¹⁸/pT¹¹⁸ side chains interacting with the i+2 K¹²⁰, while T¹⁴⁰/pT¹⁴⁰ do the same with the i+3 K¹⁴³. These interactions seem to act as capping motifs of the short C-terminal $\alpha/3_{10}$ helix (residues 119-123) in peptides T¹¹⁸-H1.0/pT¹¹⁸-H1.0, and of the T¹⁴⁰-H1.0/pT¹⁴⁰-H1.0 α -helix (residues 141-147). Accordingly, the stability of these helices (as estimated by helix populations) in the phosphorylated peptides in 90 % TFE is pH-dependent, increasing the stability at pH values in which the pT phosphate group is di-anionic. The structural differences between the TPKK and TPVK motifs might be biologically relevant and be related to different roles for the various phosphorylation motifs present in the C-terminal domain of Histone H1.0, which have not been discerned up to now.

Concerning the pH-dependence in aqueous solution, the fact that the pK_a values of E¹¹² carboxylate and pT phosphates are lower than the intrinsic ones (Table 3) can be explained by the stabilisation of anionic forms in the positively charged environment created by the numerous Lys side chains. Based on the non-homogeneous distribution of affected residues (which can be up to 4-5 residues distant from the ionisable groups; Figure 6) we could speculate about the existence of low populated helix-like conformations within the structural ensemble, which are mostly undetectable by other NMR parameters. We might also speculate that regulation by phosphorylation in Histone H1.0, and other Lys-rich IDPs, occurs because a negatively charged phosphate has non-specific interactions with surrounding Lys, so that local conformations present in the IDP structural ensemble are modified. Coming back to the phosphorylated TPKK and TPVK motifs, they show different patterns of nuclei and residues around pT¹¹⁸ and pT¹⁴⁰ affected by their ionisation state, what suggests that they differ in preferred local conformations in aqueous solution, and not only in the presence of TFE.

Finally, it is important to highlight that short model peptides (11 and 22 residues) are found able to reproduce the known conformational behaviour of the full-length C-H1.0 domain. Furthermore, the NMR study of the model peptides have provided us structural details, which would be very difficult to get by studying the full-length C-H1.0 domain, because of its very repetitive sequence and its short-life. In brief, minimalist approaches are effective as alternative and/or complementary strategies to get experimental data on IDPs, in particular if they show short long-term stability.

Experimental Section

Materials and peptide synthesis

The deuterated compounds [D2]-TFE (99.0 %), and D₂O (99.9 %) were from Cambridge Isotope Lab (USA). Peptides were synthesised using Fmoc (fluorenyl-9-methyloxycarbonyl) solid phase protocols and purified by reverse-phase HPLC up to 95 % or more purity by Caslo Aps (Lyngby, Denmark).

T¹¹⁸-H1.0 (Ac-SVAFKKTKEVKKVATPKKAAK-NH₂): RP-HPLC: t_R = 13.89 min; 98.35 % (linear 5-25 % B gradient in 20 min; buffer A: 0.05 % TFA in H₂O/CH₃CN 98:2; buffer B: 0.05 % TFA in H₂O/CH₃CN 1:9). HRMS: Theoretical molecular weight (MW) = 2457.07; Found [M+H]⁺ = 2457.84.

pT¹¹⁸-H1.0 (Ac-SVAFKKTKEVKKVA(pT)KKAAK-NH₂): RP-HPLC: t_R = 10.29 min; 99.38 % (linear 8-25 % B gradient in 17 min; buffer A: 0.05 % TFA in H₂O/CH₃CN 98:2; buffer B: 0.05 % TFA in H₂O/CH₃CN 1:9). HRMS: Theoretical molecular weight (MW) = 2537.05; Found [M+H]⁺ = 2538.13.

T¹⁴⁰-H1.0 (Ac-KATPVKKAKKK-NH₂): RP-HPLC: t_R = 11.46 min; 100 % (linear 0-17 % B gradient in 17 min; buffer A: 0.05 % TFA in H₂O/CH₃CN 98:2; buffer B: 0.05 % TFA in H₂O/CH₃CN 1:9). HRMS: Theoretical molecular weight (MW) = 1267.64; Found [M+H]⁺ = 1267.73.

pT¹⁴⁰-H1.0 (Ac-KA(pT)PVKKAKKK-NH₂): RP-HPLC: t_R = 9.70 min; 95.1 % (linear 0-15 % B gradient in 15 min; buffer A: 0.05 % TFA in H₂O/CH₃CN 98:2; buffer B: 0.05 % TFA in H₂O/CH₃CN 1:9). HRMS: Theoretical MW = 1347.62; Found [M+H]⁺ = 1348.20.

Circular Dichroism

All experiments were recorded in a Jasco J-810 spectropolarimeter equipped with a Peltier temperature control unit using cell path lengths of 0.1 cm. CD spectra were recorded for the peptides in aqueous solution and in different percentages of TFE at pH 5.5, 25 °C and peptide concentrations of approximately 37 μM. Since the peptides lack aromatic residues, concentrations were estimated by weighting the peptide to prepare a 370 μM stock solution.

Results were expressed as mean residue ellipticity ([θ], deg.cm².dmol⁻¹):

$$[\theta] = \frac{\theta^{observed}}{N \times l \times c} \quad \text{Eq. [1]}$$

where $\theta^{observed}$ is the observed ellipticity in degrees, N is the number of amino acids in the peptide, l is the cell path length and c is the peptide concentration.

The helix percentages were estimated from the experimental [θ] value at 222 nm ($[\theta]^{222nm}$, deg.cm².dmol⁻¹) by using equation [2]:^[24]

$$\% \text{ helix} = \frac{-[\theta]^{222nm} + 3000}{39000} \quad \text{Eq. [2]}$$

NMR experiments

Samples for NMR spectra acquisition were prepared at 0.5-1.0 mM concentration and contained sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) as internal reference for ^1H chemical shifts. The required amount of lyophilised peptide was dissolved in $\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1 v/v and the pH was measured with a glass micro-electrode, not corrected for isotopic effects, and adjusted to the desired pH by adding minimal amounts of NaOD or DCl. To prepare the samples in 90 % TFE, a sample in aqueous solution was prepared at the chosen pH, then these samples were freeze-dried and finally re-dissolved in 90 % [D2]-TFE in $\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1 v/v.

NMR spectra were recorded on a Bruker Avance-600 spectrometer operating at a proton frequency of 600.1 MHz and equipped with a cryoprobe. A methanol sample was employed to calibrate cryoprobe temperature. As formerly reported,^[25] 1D ^1H NMR and 2D phase-sensitive two-dimensional correlated spectroscopy (COSY), total correlated spectroscopy (TOCSY), nuclear Overhauser enhancement spectroscopy (NOESY), and ^{13}C natural abundance ^1H - ^{13}C heteronuclear single quantum coherence (HSQC) were recorded by standard techniques, and processed using the TOPSPIN program (Bruker Biospin, Karlsruhe, Germany). Water signal was suppressed by presaturation. ^{13}C δ -values were indirectly referenced using the IUPAC-IUB recommended $^1\text{H}/^{13}\text{C}$ (0.25144953) chemical shift ratio.^[26]

Assignment of ^1H chemical shifts for the peptides was performed by analyses of the 2D NMR spectra using the SPARKY software (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco) and following the standard sequential assignment strategy.^[27] The ^{13}C chemical shifts were assigned from the $^1\text{H}/^{13}\text{C}$ cross-peaks present in the ^1H , ^{13}C -HSQC spectra. Tables S1-S4 list the ^1H , and ^{13}C chemical shifts for the four peptides in aqueous solution and in 90 % TFE.

The $^1\text{H}_\alpha$ and $^{13}\text{C}_\alpha$ conformational shifts are defined by equations [3] and [4], respectively:

$$\Delta\delta_{\text{H}\alpha} = \delta_{\text{H}\alpha}^{\text{observed}} - \delta_{\text{H}\alpha}^{\text{RC}}, \text{ ppm} \quad \text{Eq. [3]}$$

$$\Delta\delta_{\text{C}\alpha} = \delta_{\text{C}\alpha}^{\text{observed}} - \delta_{\text{C}\alpha}^{\text{RC}}, \text{ ppm} \quad \text{Eq. [4]}$$

where the reference $\delta_{\text{H}\alpha}^{\text{RC}}$ and $\delta_{\text{C}\alpha}^{\text{RC}}$ values for the random coil (RC) state are taken from ^[14].

α -helix populations were estimated from $^1\text{H}_\alpha$ and $^{13}\text{C}_\alpha$ chemical shifts by a previously described method, which assumes a two-state α -helix /unfolded transition and applies equations [5] and [6] for $^1\text{H}_\alpha$ and $^{13}\text{C}_\alpha$, respectively.

$$\% \alpha - \text{helix} = \frac{\langle \Delta\delta_{\text{H}\alpha} \rangle}{-0.39} \times 100 \quad \text{Eq. [5]}$$

$$\% \alpha - \text{helix} = \frac{\langle \Delta\delta_{\text{C}\alpha} \rangle}{3.09} \times 100 \quad \text{Eq. [6]}$$

being $\langle \Delta\delta_{\text{H}\alpha} \rangle$ and $\langle \Delta\delta_{\text{C}\alpha} \rangle$, respectively, the $\Delta\delta_{\text{H}\alpha}$ and $\Delta\delta_{\text{C}\alpha}$ values averaged for all the helical residues, as defined by equations [7] and [8]

$$\langle \Delta\delta_{\text{H}\alpha} \rangle = \sum_i \frac{\Delta\delta_{\text{H}\alpha}^i}{n} \quad \text{Eq. [7]}$$

$$\langle \Delta\delta_{\text{C}\alpha} \rangle = \sum_i \frac{\Delta\delta_{\text{C}\alpha}^i}{n} \quad \text{Eq. [8]}$$

where the $\Delta\delta_{H\alpha}^i$ and $\Delta\delta_{C\alpha}^i$ are, respectively, the $\Delta\delta_{H\alpha}$ and $\Delta\delta_{C\alpha}$ conformational shifts for residue i (residues 3 to 13 in peptides T¹¹⁸-H1.0 and pT¹¹⁸-H1.0, and residues 4 to 10 in peptides T¹⁴⁰-H1.0 and pT¹⁴⁰-H1.0), and n the number of helical residues (11 in peptides T¹¹⁸-H1.0 and pT¹¹⁸-H1.0, and 7 in peptides T¹⁴⁰-H1.0 and pT¹⁴⁰-H1.0).

Assuming experimental errors in the measurement of ¹H and ¹³C δ -values of ± 0.01 ppm and ± 0.1 ppm, respectively, the errors in the estimated populations are ± 3 % and ± 7 %.

Structure calculation

Structures for peptides T¹¹⁸-H1.0 and pT¹¹⁸-H1.0, T¹⁴⁰-H1.0 and pT¹⁴⁰-H1.0 in 90 % TFE were calculated using the standard iterative protocol for automatic NOE assignment of the CYANA 2.1 program.^[28] This procedure carries out seven cycles of combined automated NOE assignment and structure calculation, in which 100 conformers were calculated per cycle. The experimental input data were the lists of: (1) assigned chemical shifts, (2) NOE integrated cross-peaks present in 150 ms NOESY spectra, and (3) ϕ and ψ dihedral angle restraints. The automatic integration subroutine of SPARKY software (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco) was used to integrate NOE cross-peaks.

Restraints for dihedral angles were obtained from ¹H and ¹³C chemical shifts using TALOSn webserver.^[29] The non-standard phosphorylated Thr was incorporated into the CYANA amino acid library from that built by ^[30]. For each peptide, the final structure is the ensemble of the 20 lowest target function conformers calculated at the final cycle. These ensembles were visualized and examined using the program MOLMOL.^[31]

pK_a determination

To determine the pK_a of ionisable groups (E carboxylate and pT phosphate) in peptides T¹¹⁸-H1.0, pT¹¹⁸-H1.0 and pT¹⁴⁰-H1.0, ¹H, ¹H-TOCSY and ¹H, ¹³C-HSQC spectra were acquired at 25° C for 1 mM samples in H₂O/D₂O 9:1 at increasing pH values from 2.5 up to 9.0 (approx. every 0.5 pH units). Because of fast solvent exchange, intensities of the H^N amide signals started to decrease at pH > 7.0, and became zero for all of them at pH 9.0.

The sigmoidal curves for the pH-dependence of δ -values were fitted to equation [9]^[32] by non-linear fitting using the Microcal Origin 6.0 software.

$$\delta^{observed} = \frac{\delta^1 + \delta^2 \cdot 10^{(pH-pK_a)}}{1 + 10^{(pH-pK_a)}} \quad \text{Eq. [9]}$$

where δ^1 and δ^2 are the chemical shifts for the dianionic and monoanionic states of the phosphate group and for the anionic and neutral carboxylate group, respectively.

Acknowledgements

We thank financial support from projects CTQ2014-52633-P (MINECO) and CTQ2017-84371-P (AEI/FEDER, UE). BB was a recipient of pre-doctoral FPI scholarship BES-2015-073383 from Spanish MINECO. We thank Dr I. Ponte, Dr. A. Roque and Prof P. Suau for helpful comments on the H1.0 biology. The NMR experiments were performed in the “Manuel Rico” NMR laboratory, LMR, CSIC, a node of the Spanish Large-Scale National Facility ICTS R-LRB.

Keywords: phosphorylation • peptide structure • NMR • intrinsically disordered proteins • histone

- [1] A. Roque, I. Ponte, P. Suau, *Biochim Biophys Acta* **2016**, 1859, 444-454.
- [2] A. Roque, I. Ponte, P. Suau, *Chromosoma* **2017**, 126, 83-91.
- [3] A. Roque, I. Iloro, I. Ponte, J. L. Arrondo, P. Suau, *J Biol Chem* **2005**, 280, 32141-32147.
- [4] R. Vila, I. Ponte, M. Collado, J. L. Arrondo, M. A. Jiménez, M. Rico, P. Suau, *J Biol Chem* **2001**, 276, 46429-46435.
- [5] L. Millan-Arino, A. Izquierdo-Bouldstridge, A. Jordan, *Biochim Biophys Acta* **2016**, 1859, 510-519.
- [6] X. Lu, J. C. Hansen, *Biochem Cell Biol* **2003**, 81, 173-176.
- [7] A. Roque, I. Ponte, J. L. Arrondo, P. Suau, *Nucleic Acids Res* **2008**, 36, 4719-4726.
- [8] N. Raghuram, H. Strickfaden, D. McDonald, K. Williams, H. Fang, C. Mizzen, J. J. Hayes, J. Th'ng, M. J. Hendzel, *J Cell Biol* **2013**, 203, 57-71.
- [9] R. Lopez, B. Sarg, H. Lindner, S. Bartolome, I. Ponte, P. Suau, A. Roque, *Nucleic Acids Res* **2015**, 43, 4463-4476.
- [10] B. Chaves-Arquero, D. Pantoja-Uceda, A. Roque, I. Ponte, P. Suau, M. A. Jiménez, *J Biomol NMR* **2018**, 72, 139-148.
- [11] a) A. Kumar, V. Narayanan, A. Sekhar, *Biochemistry* **2020**, 59, 57-73; b) V. Csizmok, J. D. Forman-Kay, *Curr Opin Struct Biol* **2018**, 48, 58-67; c) S. J. Humphrey, D. E. James, M. Mann, *Trends Endocrinol Metab* **2015**, 26, 676-687.
- [12] M. Schubert, D. Labudde, H. Oschkinat, P. Schmieder, *J Biomol NMR* **2002**, 24, 149-154.
- [13] W. J. Wu, D. P. Raleigh, *Biopolymers* **1998**, 45, 381-394.
- [14] D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges, B. D. Sykes, *J Biomol NMR* **1995**, 5, 67-81.
- [15] a) N. Sreerama, R. W. Woody, *Methods Enzymol* **2004**, 383, 318-351; b) W. C. Johnson, Jr., *Annu Rev Biophys Biophys Chem* **1988**, 17, 145-166.
- [16] R. Vila, I. Ponte, M. A. Jimenez, M. Rico, P. Suau, *Protein Sci* **2000**, 9, 627-636.
- [17] a) M. Vincenzi, F. A. Mercurio, M. Leone, *Curr Protein Pept Sci* **2019**, 20, 425-451; b) M. Buck, *Q Rev Biophys* **1998**, 31, 297-355.
- [18] a) D. S. Wishart, B. D. Sykes, F. M. Richards, *J Mol Biol* **1991**, 222, 311-333; b) S. Spera, A. Bax, *J Am Chem Soc* **1991**, 113, 5490-5492.
- [19] C. M. Santiveri, M. Rico, M. A. Jiménez, *J Biomol NMR* **2001**, 19, 331-345.
- [20] C. N. Pace, G. R. Grimsley, J. M. Scholtz, *J Biol Chem* **2009**, 284, 13285-13289.

- [21] a) E. A. Bienkiewicz, K. J. Lumb, *J Biomol NMR* **1999**, *15*, 203-206; b) R. Hendus-Altenburger, C. B. Fernandes, K. Bugge, M. B. A. Kunze, W. Boomsma, B. B. Kragelund, *J Biomol NMR* **2019**, *73*, 713-725; c) G. Platzner, M. Okon, L. P. McIntosh, *J Biomol NMR* **2014**, *60*, 109-129.
- [22] R. Hoffmann, I. Reichert, W. O. Wachs, M. Zeppezauer, H. R. Kalbitzer, *Int J Pept Protein Res* **1994**, *44*, 193-198.
- [23] R. L. Thurlkill, G. R. Grimsley, J. M. Scholtz, C. N. Pace, *Protein Sci* **2006**, *15*, 1214-1218.
- [24] a) C. B. Falcao, C. Pérez-Peinado, B. G. de la Torre, X. Mayol, H. Zamora-Carreras, M. A. Jiménez, G. Radis-Baptista, D. Andreu, *J Med Chem* **2015**, *58*, 8553-8563; b) Y. H. Chen, J. T. Yang, K. H. Chau, *Biochemistry* **1974**, *13*, 3350-3359.
- [25] H. Zamora-Carreras, B. Maestro, E. Strandberg, A. S. Ulrich, J. M. Sanz, M. A. Jiménez, *Chemistry* **2015**, *21*, 8076-8089.
- [26] J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright, K. Wüthrich, *J Mol Biol* **1998**, *280*, 933-952.
- [27] a) K. Wüthrich, *John Wiley & Sons* **1986**, New York; b) K. Wuthrich, M. Billeter, W. Braun, *J Mol Biol* **1984**, *180*, 715-740.
- [28] P. Güntert, *Methods Mol Biol* **2004**, *278*, 353-378.
- [29] Y. Shen, A. Bax, *J Biomol NMR* **2013**, *56*, 227-241.
- [30] J. W. Craft, Jr., G. B. Legge, *J Biomol NMR* **2005**, *33*, 15-24.
- [31] R. Koradi, M. Billeter, K. Wüthrich, *J Mol Graph* **1996**, *14*, 51-55, 29-32.
- [32] J. M. Pérez-Cañadillas, R. Campos-Olivas, J. Lacadena, A. Martinez del Pozo, J. G. Gavilanes, J. Santoro, M. Rico, M. Bruix, *Biochemistry* **1998**, *37*, 15865-15876.

Effect of phosphorylation in the structural behaviour of peptides derived from the intrinsically disordered C-terminal domain of Histone H1o

*Belén Chaves-Arquero, José M. Pérez-Cañadillas, M. Angeles Jiménez**

Spot the differences

Based on NMR data of model peptides, the TPKK and TPVK phosphorylation motifs that are present in the intrinsically disordered C-terminal domain of histone H1.0 show equal backbone conformations, both non- and phosphorylated, but different side chain interactions. Furthermore, the phospho-threonine pK_a in the TPKK motif is much lower than the intrinsic one, whereas in the TPVK is only slightly lower.