A CON-based NMR assignment strategy for Pro-rich intrinsically disordered proteins with low signal dispersion: The C-terminal domain of histone H1.0 as a case study

Belén Chaves-Arquero,^a David Pantoja-Uceda,^a Alicia Roque,^b Inmaculada Ponte,^b Pedro Suau,^b M. Angeles Jiménez^{a*}

^a Departamento de Química-Física Biológica, Instituto de Química Física Rocasolano (IQFR-CSIC), Serrano 119, 28006-Madrid, Spain

^b Departamento de Bioquímica y Biología Molecular, Facultad de Biociencias, Universidad Autónoma de Barcelona (UAB), Bellaterra, Barcelona, Spain

*Corresponding author: majimenez@iqfr.csic.es

Abstract

The C-terminal domain of histone H1.0 (C-H1.0) is involved in DNA binding and is a main determinant of the chromatin condensing properties of histone H1.0. Phosphorylation at the (S/T)-P-X-(K/R) motifs affects DNA binding and is crucial for regulation of C-H1.0 function. Since C-H1.0 is an intrinsically disordered domain, solution NMR is an excellent approach to characterize the effect of phosphorylation on the structural and dynamic properties of C-H1.0. However, its very repetitive, low-amino acid-diverse and Pro-rich sequence, together with the low signal dispersion observed at the ¹H-¹⁵N HSQC spectra of both non- and tri-phosphorylated C-H1.0 preclude the use of standard ¹H-detected assignment strategies. We have achieved an essentially complete assignment of the heavy backbone atoms (¹⁵N, ¹³C' and ¹³C_α), as well as ¹H^N and ¹³C_β nuclei, of non- and tri-phosphorylated C-H1.0 by applying a novel ¹³C-detected CON-based strategy. No C-H1.0 region with a clear secondary structure tendency was detected by chemical shift analyses, confirming at residue level that C-H1.0 is disordered in aqueous solution. Phosphorylation only affected the chemical shifts of phosphorylated Thr's, and their adjacent residues. Heteronuclear {¹H}-¹⁵N NOEs were also essentially equal in the non- and tri-phosphorylated states. Hence, structural tendencies and dynamic properties of C-H1.0 free in aqueous solution are unmodified by phosphorylation. We propose that the assignment strategy used for C-H1.0, which is based on the acquisition of only a few 3D spectra, is an excellent choice for short-lived intrinsically disordered proteins with repetitive sequences.

Keywords: Intrinsically disordered domain, IDP, phosphorylation, histone, NMR assignment strategy

INTRODUCTION

H1 linker histones bind to linker DNA regions on the surface of the nucleosome, and participate in the maintenance of chromatin higher-order structure and in gene regulation. In mammals, histone H1 is a multigene family composed of 7 somatic subtypes H1.1-H1.5, H1.0 and H1x.¹ Metazoan H1 contains three distinct domains: a short N-terminal domain (20-35 residues), a central globular domain (about 80 residues), and a long C-terminal domain (approximately 100 residues). Terminal domains are variable in length, amino acid sequence and post-translational modifications, and thus they determine subtype specificity.^{2,3} The C-terminal domain is the primary determinant of H1 binding to chromatin *in vitro* and *in vivo*, and also is responsible for the preferential binding of H1 to scaffold-associated regions.⁴⁻⁶

There is increasing evidence that H1 interacts with proteins of different functional categories, including pre-mRNA splicing, core histone chaperones and transcription-associated proteins.⁷ It has also been described that the C-terminal domain interacts with specific proteins such as the apoptotic nuclease DFF40⁸ and prothymosin alpha.⁹ In particular, the study of H1.0 interactome has revealed that that one-third of the H1.0-dependent interaction proteins found in the nucleolus were mediated by the C-terminal domain.¹⁰

H1 terminal domains are enriched in disorder-promoting amino acids, including lysine, serine, proline and alanine. Peptides derived from the basic portion of the N-terminal domain of subtypes H1.0 and H1.4, close to the globular domain, have been studied in aqueous solution and in presence of 2,2,2 trifluoroethanol (TFE) by circular dichroism (CD) and NMR.^{11,12} The results revealed that while both peptides were disordered in aqueous solution, helical populations of 40-50% were detected in the presence of 90% TFE. A further study of the H1.0-derived by Fourier Transform Infrared Spectroscopy (FTIR) showed that DNA induced an amount of α -helical structure equivalent to that observed in TFE.¹¹ Analysis of a peptide derived from the first portion of the C-terminal domain of H1.0, residues 99-121, provided similar results by NMR in presence of TFE and by FTIR in presence of DNA.^{13,14} FTIR studies on the C-terminal domain of H1.0 (C-H1.0) showed that this domain was mostly disordered when free in aqueous solution, but acquired significant amounts of secondary structure when bound to DNA.¹⁵ Thus, based on the structural information available up to now, H1 terminal domains can be classified as intrinsically disordered regions with coupled binding and folding. Further

evidence of the folding of C-H1.0 upon binding to DNA and to reconstituted chromatin has been provided by FRET studies.¹⁶⁻¹⁸

Cell-cycle dependent phosphorylation is the most studied post-translational modification of histone H1, as it may be involved in the regulation of chromatin dynamics.¹⁹ H1 is phosphorylated by cyclin-dependent kinases (CDK) at the (S/T)-P-X-(K/R) consensus sequence, which are mainly located in the C-terminal domain. It has been described that phosphorylation of C-H1.0 by CDK2 triggers a conformational change, which results in significant changes in the secondary structure of the domain when bound to DNA.²⁰ Additional evidences of this conformational change in chromatin have been provided by FTIR and FRET.^{21,22} Interestingly, some results suggest that the conformational change induced by phosphorylation is associated with *cis/trans* isomerization of the proline adjacent to the phosphorylated residues.²² NMR analyses of a 23-mer C-H1.0-derived peptide, which contains a CDK consensus sequence (TPKK), showed that the proline adjacent to the phosphorylatable Thr was predominantly in *trans* conformation in the non-phosphorylated state.¹³ In the presence of 90 % of TFE, the TPKK sequence adopts a type I β -turn conformation, a σ -turn conformation or a combination of both, in fast equilibrium with unfolded states.

All the structural data about the whole C-H1.0 either free or DNA-bound has been obtained by techniques (CD, FTIR and FRET), which do not provide information at atomic level. In this context, we decided to use solution NMR to characterise the structural and dynamic properties of the full-length C-H1.0, and to examine the effect of phosphorylation. Considering that C-H1.0 is an intrinsically disordered domain, which has a very repetitive sequence composed of only a few types of amino acids, and contains a large number of Pro residues (see Fig. 1a-b), the NMR assignment task was not expected to be easy. But, we did not anticipate to have so low signal dispersion as found at the 2D ¹H-⁵N HSQC of C-H1.0 (Fig. 1c). The low signal dispersion and high signal overlap observed in that spectrum, together with the repetitive sequence impeded the application of standard assignment strategies starting from the amide protons (see below). Therefore, we examined the alternative assignment strategies, which have been proposed in the last years. The most successful ones are the ¹³C-detected CON-based approaches, which correlate two consecutive CON groups in a protein.²³⁻²⁶ Some of them are based on the acquisition of several 5D and/or 4D experiments recorded using non-uniform sampling.^{23,25,26} These protocols speed up the assignment process providing backbone and aliphatic chemical shifts, but they require long acquisition

times. As a result, they are not applicable when proteins are short-lived, because of degradation/stability problems. This was the case of C-H1.0 and pT-C-H1.0, which we found to get degraded in a few days. Hence, we searched for other assignment strategies, and succeeded by using a ^{13}C -detected CON-based approach, which uses only a few 3D spectra, which reduces the total acquisition time in relation with high-dimensionality approaches. This assignment strategy is described below, and proposed as an excellent alternative strategy in the case of IDP's with repetitive, low-amino acid diverse, and Pro-rich sequences and, in particular, if they are short-lived and signal dispersion at 2D ^1H - ^{15}N HSQC is low, and signal overlap high.

The heavy atoms of the backbone (^{15}N , $^{13}\text{C}'$, and $^{13}\text{C}_\alpha$), and the $^1\text{H}^\text{N}$ and $^{13}\text{C}_\beta$ nuclei were fully assigned for C-H1.0 in aqueous solution in its non-phosphorylated state as well as tri-phosphorylated, denoted as pT-C-H1.0. Analyses of these chemical shifts indicated the absence of any secondary structure tendency in C-H1.0, either non-phosphorylated or tri-phosphorylated, when free in solution. Furthermore, dynamic characterisation showed that both C-H1.0 and pT-C-H1.0 have a very high flexibility on the fast time scale.

MATERIALS AND METHODS

Expression and purification of the C-terminal domain of H1.0 (C-H1.0)

The coding sequence of the C-terminal domain of mouse H1.0 was cloned in the pQE-60 vector (Qiagen), with a 6xHis-tag at the C-terminus, as previously described.⁶ The recombinant expression vector pCTH1.0 was transformed into *E.coli* M15 (Qiagen). The isotopically labeled C-H1.0 was prepared by Marley et al.'s method.²⁷ Briefly, cells were grown to an $\text{OD}_{600\text{ nm}}$ of 0.8 in rich medium (Luria-Bertani), centrifuged at 5,000 g for 15 min and re-suspended in minimal medium M9 salts. After a second step of centrifugation the cell mass obtained in 2L of rich medium was finally re-suspended in 1L of minimal medium (M9) prepared with ^{13}C -D-glucose (Cambridge Isotope Laboratories Inc.) and $^{15}\text{NH}_4\text{Cl}$ (Cambridge Isotope Laboratories Inc.). Following a short period for growth recovery, protein expression was induced with 1 mM IPTG, allowing expression to proceed for 4 h at 37°C. Cells were then harvested and, if necessary, stored at -80°C. Next, cells were lysed in the lysis buffer (0.05 M NaH_2PO_4 , 0.75 M NaCl, 0.02 M imidazol) plus 4 M guanidine hydrochloride, pH 8.0 for 15 min at room temperature. The extract was centrifuged at 20,000 g for 25 min. The supernatants were loaded on a HiTrap chelating HP column (GE Healthcare) equilibrated with lysis buffer. The column was then washed in three steps with lysis buffer containing

increasing amounts of imidazol: 40, 60 and 80 mM. Finally, the proteins were eluted with 250 mM imidazol in lysis buffer and desalted by gel filtration through Sephadex G-25 (Sigma-Aldrich).

Preparation of the phosphorylated C-terminal domain of H1.0 (pT-C-H1.0)

The isotopically labeled C-H1.0 was phosphorylated *in vitro* using CDK2-cyclin A2 kinase (Sigma-Aldrich) as previously described.²⁰ Briefly, the phosphorylation reaction was carried out in 50 mM Tris-HCl, 10 mM MgCl₂, 1 mM EGTA, 20 mM dithiotreitol, pH 7.5, plus 200 μ M ATP and 1U of CDK2-cyclin A per 5 μ g of C-H1.0. The mixture was incubated at 30°C for 1 h and the reaction buffer was eliminated by gel filtration on a HiTrap desalting column (GE Healthcare). The extent of phosphorylation was evaluated by MALDI-TOF mass spectrometry.

NMR Experiments

All NMR experiments were collected at 298 K on a Bruker Avance spectrometer, operating at a ¹H frequency of 800.1 MHz equipped with a TCI cryoprobe.

¹³C and ¹⁵N uniformly labeled C-H1.0 and pT-C-H1.0 samples were prepared at approximately 1 mM protein concentration in H₂O/D₂O 9:1 v/v at pH 5.5. The set of collected experiments for each protein consisted of four ¹³C-detected experiments, i.e. 2D CON, and 3D hacacoNcaNCO, hacaCONcaNCO and CBCANCO, and two ¹H-detected experiments, i.e. 2D ¹H-¹⁵N HSQC and 3D HNCO. In the case of pT-C-H1.0 3D HncacoNH and hNcacoNH were also recorded. In the case of 3D spectra, the less sensitive ¹³C-detected experiments were recorded using linear sampling, and the most sensitive ¹H-detected using non-uniform sampling (NUS).

The 2D ¹³C-detected CON experiment was acquired with 512 and 256 complex points in the ¹³C' direct dimension and ¹⁵N indirect dimension. The carriers of ¹³C' and ¹⁵N dimensions were set at 172.5 ppm and 127.5 ppm, respectively, and the spectral widths were 10 and 27 ppm in the ¹³C' and ¹⁵N dimensions, respectively. The total measurement time was 1.8h.

The 3D ¹³C-detected spectra were acquired with 512 complex points in the ¹³C' direct dimension (f3), 24 complex points in the indirect dimension ¹⁵N (f2) attached directly to ¹³C', 24 complex points in the indirect dimension ¹³C' or ¹⁵N (f1) and 48 complex point in ¹³C β /¹³C α (f1) dimension for CBCANCO. The carriers of the ¹³C' and ¹⁵N dimension were set at 172.5 ppm and 127.5 ppm, respectively. The spectral widths were 10 and 27 ppm in the ¹³C' and ¹⁵N dimensions, respectively, or 39 ppm in the ¹³C dimension of

CBCANCO. The total measurement time was 30 hours (16 scans) or 19 hours (8 scans) for CBCANCO experiment.

For the 2D ^1H - ^{15}N HSQC spectrum 1K and 256 complex data points were acquired for direct and indirect dimension, respectively, 4 scans were accumulated and the total experimental time was 41 minutes.

The 3D ^1HN -detected HNCO experiments were acquired with 1K complex points in the direct dimension (f3) and 24 complex points in the ^{15}N dimension (f2) and 64 complex points in the ^{13}C dimension (f1), and 8 scans. The total measurement time was 8.5h. The 3D HncacoNH and hNcacoNH experiments were acquired with non-uniform sampling (NUS). They were acquired using a 310-complex point sampling schedule with 8 scans per FID in 3h and 40min each one. The maximum increment in the NUS schedule is 48 or 32 for the ^{15}N attached directly to ^1H acquired in the direct dimension (f2) and ^{15}N or ^1H (f1) dimensions, respectively. ^{15}N (f2) is centered in 122 ppm with the spectral width being 1297.33 Hz, and ^{15}N and ^1H (f1) are centered in 122 ppm and 7.17 ppm with the spectral width being 1297.33 Hz and 3880.5 Hz, respectively. The directly observed ^1H dimension has a spectral width of 8802.817 Hz (centered at 4.75 ppm) with 1024 complex points.

$\{^1\text{H}\}$ - ^{15}N NOE experiments for C-H1.0 and pT-C-H1.0 were acquired in an interleaved manner with and without proton saturation during an overall recycling delay of 10 s to ensure the maximal development of NOEs before acquisition and to allow solvent relaxation, thus avoiding transfer of saturation to the most exposed amide protons of the protein between scans.²⁸ They were acquired with 1K and 128 complex data point were acquired for direct and indirect dimension, respectively, 8 scans were accumulated and the total experimental time was 14 hours.

For all the experiments, the resulting matrix was zero filled to double the number of original points in all dimensions and shifted squared sine-bell apodization functions were applied in all dimensions prior to Fourier transformation.

Spectra were processed using either TOPSPIN v2.1 pl6 (Bruker, Inc) or NMRpipe²⁹ and the istHMS reconstruction method was used to process NUS data.³⁰ Finally, they were analyzed with the programs SPARKY (T. Goddard and D.G. Kneller, SPARKY 3, University of California, San Francisco, USA) and/or NMRview.³¹

Chemical shift perturbation (CSP) were obtained by applying the following equation:

$$\text{CSP}=[(\delta_{\text{N}}^{\text{pT-C-H1.0}}-\delta_{\text{N}}^{\text{C-H1.0}})^2+((\delta_{\text{C}}^{\text{pT-C-H1.0}}-\delta_{\text{C}}^{\text{C-H1.0}})\text{A})^2]^{1/2}$$

where the scaling factor A , which is the ratio between ^{15}N and ^{13}C spectral widths ($A = \text{SW}_{\text{N}}/\text{SW}_{\text{C}}$), is equal to 0.37.³²

RESULTS AND DISCUSSION

^1H -detected amide-based assignment strategies are not applicable to C-H1.0

For our solution NMR study, we used a C-H1.0 construct, whose overall secondary structure has been previously examined by FTIR.²⁰ This construct, denoted as C-H1.0, corresponds to residues 98-193 of mouse histone H1.0 (Fig. 1a), and contains an additional N-terminal Met residue, two extra amino acids added by the restriction enzyme used for cloning, and a C-terminal 6xHis-tag, which are required for expression and purification of the protein. In terms of simplicity, herein we are going to number the C-H1.0 sequence from 1 (the N-terminal Met) to 105 (Fig. 1a).

The sequence of C-H1.0 contains only 12 types of residues (Fig. 1b). It lacks Gly residues and, excluding the six His from the histidine-tag, the only aromatic residue among them is a Phe (F10). Also, three out of the 12 types represent about a 65 % of the sequence; they are Lys (38.1 %; 40 out of 105), Ala (16.2 %; 17 out of 105), and Pro (11.4 %; 12 out of 105). As a consequence, many stretches of residues are repeated along the sequence, i.e. one 6-residue stretch ($^{21}\text{ATPKKA}^{26}$ & $^{55}\text{ATPKKA}^{60}$; Fig. 1A), one 5-residue segment ($^{28}\text{KPKKA}^{32}$ & $^{75}\text{KPKKA}^{79}$), and two 4-residue stretches ($^{39}\text{KKPK}^{42}$ & $^{39}\text{KKPK}^{42}$, and $^{40}\text{KPKA}^{43}$ & $^{83}\text{KPKA}^{86}$) are each found twice in the C-H1.0 sequence; besides the 6- and 5-residue stretches share the 4-residue PKKA sub-sequence, which is four times in the sequence ($^{23}\text{PKKA}^{26}$, $^{29}\text{PKKA}^{32}$, $^{57}\text{PKKA}^{60}$ and $^{76}\text{PKKA}^{79}$). Because of these repetitions, the residues of the same type, in particular the very numerous Lys, have very similar local chemical environments, which translates into their nuclei (particularly, $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ carbons) having very similar chemical shifts, so that their cross-peaks appear at the same or very close places in the NMR spectra. Furthermore, C-H1.0 is an intrinsically disorder protein, which shows very little dispersion in the $^1\text{H}^\text{N}$ nuclei; all amide protons appear in 0.51 ppm (Fig. 1c). This range for amide protons is about a 25 % narrower than that typically observed in prototype IDPs, such as α -synuclein (about a 1.0 ppm; 140 residues, BMRB-6968). In brief, the ^1H -detected conventional methods based on the triple resonance experiments,^{33,34} in which assignment starts from the amide groups ($^1\text{H}^\text{N}/^{15}\text{N}$) are not suitable for C-H1.0 due to the severely overlapped signals.

As a first alternative, we intended to apply a ^1H -detected methodology using the experiment HNcacoNH ,³⁵⁻³⁸ which correlates the $^1\text{H}^\text{N}$ and ^{15}N chemical shifts of two

consecutive amide moieties without involving other nuclei. However, the main drawback of these methods is that the connection path is broken at Pro residues. In our case, the sequence of C-H1.0 contains 12 Pro, so that the amide connection path splits into 13 segments, whose length is between 3-7 residues, except for the N- and C-terminal segments, which are 18 residues long. This sums up to the overlapping signals problem in the ^1H - ^{15}N HSQC and the high redundant sequence. On the whole, using this approach we are only able to assign around a 67% of the amide HN groups of C-H1.0.

C-H1.0 assignment was achieved by a novel ^{13}C -detected CON-based strategy

Considering the failure to achieve a full or at least almost full assignment using the available ^1H -detected methodologies, we decided to explore alternative strategies based on ^{13}C -detected experiments. Although they are less sensitive than proton detection, we can benefit of its advantages relative to the wide chemical shift range of carbonyl carbons ($^{13}\text{C}'$), as well as the direct observation of cross-peaks for Pro residues. The higher dispersion of $^{13}\text{C}'$ - ^{15}N cross-peaks relative to the $^1\text{H}^{\text{N}}$ - ^{15}N cross-peaks can be appreciated by comparison of the 2D CON spectrum of C-H1.0 (Fig. 1d) with the 2D ^1H - ^{15}N HSQC spectrum acquired using the same C-H1.0 sample (Fig. 1c).

In the 2D CON spectrum, the number of cross-peaks that we can count approximately coincides with those expected from the C-H1.0 sequence, and we observe signals for the Pro residues (see lower part of Fig. 1d).

The assignment strategies proposed up to now starting from 2D CON spectra consist in the use of ^{13}C -detected triple resonance experiments,³⁹⁻⁴² equivalent to those starting from the amide protons in the standard ^1H -detected strategy.³³ Thus, for sequential assignment they make use of nuclei with poorer dispersion ($^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$ or $^1\text{H}_\alpha$), which may adversely affect the assignment process, particularly in the case of repetitive sequences, such as C-H1.0. To overcome these difficulties some authors proposed the use of ^1H -detected^{43,44} or ^{13}C -detected^{23,26,45-48} high-dimensionality experiments, 4D or 5D. However, their application to C-H1.0 was problematic because of the sequence peculiarities of C-H1.0 (a sequence with low-amino acid diversity, many repeats, and a relatively high number of Pro residues; see above), together with its very low long-term stability. Therefore, we decided to explore the use of a simpler alternative.

Our proposal is based on the acquisition of a single 2D CON spectrum plus two 3D experiments: hacacoNcaNCO and hacaCONcaNCO, to directly correlate two consecutive CO-N groups in a protein, i.e. without mediation of the other nuclei.²⁴ Main advantage of these experiments is that the connection path is not interrupted at Pro

residues, since the experiments use the H_α protons as the starting point of the magnetization transfer pathway. This is very important in Pro-rich proteins, such as C-H1.0, which contains 12 Pro (Fig. 1a-b). To get additional information to determine the type of spin system, we collected a 3D CBCACON.³⁹ Thus, using only a set of four NMR experiments, i.e. 2D CON, 3D hacacoNcaNCO, 3D hacaCONcaNCO, and 3D CBCACON, we achieved an essentially complete assignment of the backbone ^{15}N , $^{13}\text{C}_\alpha$, and $^{13}\text{C}'$ nuclei as well as the $^{13}\text{C}_\beta$ carbons of C-H1.0. The only unassigned residues correspond to K97, and the C-terminal cloning-tag (residues 98-105), as well as the $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ carbons of K52 and K96. These ^{13}C and ^{15}N chemical shifts have been deposited in the BioMagResBank database (<http://www.bmrb.wisc.edu>) under BMRB accession number 27537.

Tri-phosphorylated C-H1.0 was assigned by the ^{13}C -detected CON-based strategy

From the biological point of view, it is important to examine the structural and dynamic properties not only of C-H1.0, but also how they are affected by phosphorylation. Therefore, we proceeded to acquire NMR spectra of pT-C-H1.0, in which the Thr residues of the three phosphorylation-motifs present in its sequence ($^{22}\text{TPKK}^{25}$, $^{44}\text{TPVK}^{47}$, and $^{56}\text{TPKK}^{59}$; Fig. 1a) are phosphorylated. As in the case of non-phosphorylated C-H1.0, the 2D ^1H - ^{15}N HSQC spectrum of tri-phosphorylated CH1.0 (denoted as pT-C-H1.0) showed low dispersion and high signal overlap, but most of the expected cross-peaks were present in the 2D CON of pT-C-H1.0 (Fig. 2). It was not possible to unambiguously assign the cross-peaks in this spectrum by comparison to the previously assigned 2D CON spectrum of C-H1.0. Although the differences in chemical shift between C-H1.0 and pT-C-H1.0 resulted to be small (see below), they suffice to impede the assignment of the phosphorylated protein just by comparison of CON spectra. Therefore, we applied the CON-based strategy used for the non-phosphorylated protein (see above) to pT-C-H1.0, as a new test case for the validity of the method. As in the case of C-H1.0, an almost complete assignment of ^{15}N , $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$ and $^{13}\text{C}'$ nuclei was obtained by analyses of 3D hacacoNcaNCO, 3D hacaCONcaNCO, and 3D CBCACON acquired for pT-C-H1.0. In this case, the unassigned residues correspond to segment 100-105 of the C-terminal cloning-tag, the $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ carbons of S89, K91, and S94, the $^{13}\text{C}'$ carbon of R5, and the ^{15}N nitrogen of R6. These chemical shifts have been deposited in the BioMagResBank database (<http://www.bmrb.wisc.edu>) under BMRB accession number 27538.

Chemical shift-based structural characterisation of C-H1.0 and pT-C-H1.0

Once assigned C-H1.0 and pT-C-H1.0, we proceeded to see whether any region in these domains show a tendency towards some secondary structure (α -helical or β -sheet). It is well established that the regions without any secondary structure tendency show chemical shifts very close to reference values, whereas those with some tendency to be structured differ from reference values. The controversy in this point relates to the best reference values to be used. We are going to use those reported by Poulsen's group,^{49,50} because they are generally considered as the most appropriate for intrinsically disordered proteins. Among the assigned chemical shifts, those of $^{13}\text{C}_\alpha$ carbons are probably the most reliable as indicators of secondary structure. Therefore, we calculated the chemical shift deviations of $^{13}\text{C}_\alpha$ carbons ($\Delta\delta_{\text{C}_\alpha} = \delta_{\text{C}_\alpha}^{\text{observed}} - \delta_{\text{C}_\alpha}^{\text{reference}}$, ppm) of C-H1.0, using the Poulsen's reference values, as implemented at Poulsen's server (https://spin.niddk.nih.gov/bax/nmrserver/Poulsen_rc_CS/), and plot them as a function of residue number (Fig. 3a). The $\Delta\delta_{\text{C}_\alpha}$ were all within the random coil range ($|\Delta\delta_{\text{C}_\alpha}| \leq 0.4$ ppm), and no region could be pointed out as having even the slightest secondary structure tendency (Fig. 3a). This conclusion was confirmed by examination of the chemical shift deviations for $^{13}\text{C}_\beta$ and $^{13}\text{C}'$ nuclei, in which no secondary structure tendency was detectable (Suppl. Figure S1). To check whether the use of other reference values could affect this result, we calculated the CSI (chemical shift index) as implemented at Wishart's server (<http://csi3.wishartlab.com/cgi-bin/index.php>).⁵¹ The CSI uses a different set of random coil values, and takes into account $^{13}\text{C}_\beta$, $^1\text{H}^{\text{N}}$, and ^{15}N , apart from the $^{13}\text{C}_\alpha$ and $^{13}\text{C}'$, and gives a value of +1 for helical residues, -1 for sheet residues, and zero for random coil residues. In C-H1.0, CSI was zero for all the residues, which indicates a fully unstructured polypeptide chain.

Next, we compare the chemical shifts of tri-phosphorylated pT-C-H1.0 relative to non-phosphorylated C-H1.0. Fig. 3b shows the chemical shift perturbation (CSP), as defined in Materials and Methods, as function of residue number. The only significant differences are found for the three Thr residues, whose hydroxyl group is phosphorylated, and for their adjacent residues. Thus, it seems that phosphorylation has only local effects. Circular dichroism (CD) data of C-H1.0 and pT-C-H1.0 also indicate that phosphorylation does not induce any conformational change (see Suppl. Figure S2). Nevertheless, we decided to check whether the chemical shift deviations for $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, and $^{13}\text{C}'$, as well as the CSI by themselves might show some secondary structure tendency, but the problem is that there is no appropriate chemical shift reference for phosphorylated Thr. The only reference values available for pT are those measured in a

short peptide of sequence Ac-GGpTGG-NH₂,⁵² and they can be inadequate for Pro-preceding pT's. Using the Thr reference for pT, the profiles of $\Delta\delta_{C\alpha}$, $\Delta\delta_{C\beta}$, and $\Delta\delta_C$ (Suppl. Figure S1) and a CSI value of zero for all the residues indicate that pT-C-H1.0 has no detectable secondary structure tendency, as in C-H1.0. The only significant differences placed around the pT residues. As deduced from the CSP profile (Fig. 3B), phosphorylation has only local effects.

We have also examined whether phosphorylation might affect the *cis/trans* isomerization equilibrium of the X-Pro bonds. To that end, we have analyzed the chemical shifts for the ¹³C_β carbons of the 12 Pro residues. All of them are in the range characteristic of *trans* X-Pro bonds⁵³ in both the non-phosphorylated C-H1.0 and the tri-phosphorylated pT-C-H1.0. To get further confirmation of this result, we have run the PROMEGA program⁵⁴, which uses the chemical shifts of several nuclei (¹⁵N, ¹³C', ¹³C_α and ¹³C_β) from the Pro residue and its preceding and following residues, and takes into account amino acid sequence. The PROMEGA output indicates that all X-Pro bonds are *trans* in C-H1.0 and pT-C-H1.0, as deduced only on the basis of Pro ¹³C_β chemical shifts. No minor signals attributable to low-populated *cis* species were detected.

Dynamics characterization of C-H1.0 and pT-C-H1.0

To obtain information of protein dynamics in the fast time scale (ps-ns), we collected heteronuclear {¹H}-¹⁵N NOE experiments for C-H1.0 and pT-C-H1.0. To assign the signals at these spectra we need to extend the assignment to the ¹H^N protons. This was successfully done by analyses of the corresponding high-resolution 3D HNCO spectrum, which is very sensitive, and correlates the already assigned ¹³C', and ¹⁵N to the ¹H^N proton. To solve a few remaining ambiguities we used 3D HncacoNH and hNcacoNH experiments acquired using NUS. Out of the 92 ¹H^N protons present in C-H1.0 and pT-C-H1.0 (105 residues in total minus the N-terminal residue and minus the 12 Pro residues, which lack amide protons), we could assign 66 in C-H1.0, and 68 in pT-C-H1.0.

Once assigned these ¹H^N amide protons, we could measure the heteronuclear {¹H}-¹⁵N NOE for 55 residues of C-H1.0 and for 56 residues in the tri-phosphorylated pT-C-H1.0. Those of the other residues, even if assigned, could not be measured reliably because of the high level of overlap. These NOEs, which have been deposited in the BioMagResBank database (<http://www.bmrb.wisc.edu>) under BMRB accession numbers 27537 and 27538, are plotted as a function of sequence in Fig. 3c. The N-terminal residues, which show negative values, are slightly more flexible than the rest of

the protein in both C-H1.0 and pT-C-H1.0. Excluding the N-terminal residues, the average values of these NOEs are 0.01 ± 0.11 for C-H1.0 and 0.07 ± 0.12 for pT-C-H1.0, which indicates that both proteins have a very high flexibility on the fast time scale (picoseconds to nanoseconds). No significant differences can be appreciated between the non- and the tri-phosphorylated forms. Thus, flexibility in C-H1.0 is not affected by phosphorylation, whether this result can be extended to other IDPs or not is an open question.

CONCLUSIONS

The C-terminal domain of histone H1.0 (C-H1.0) plays a regulatory role in the functionality of histone H1.0, which is regulated by phosphorylation of Thr at motifs $^{118}\text{TPKK}^{121}$, $^{140}\text{TPVK}^{143}$ and $^{152}\text{TPKK}^{155}$. That this domain is intrinsically disordered hampers its structural characterisation. NMR is probably the best technique to get structural information about IDPs, but it requires spectral assignment, which is more complex in IDPs than in globular proteins, because of signal dispersion decrease and signal overlap increase. In the case of C-H1.0, NMR assignment is still more complicated than in prototype IDPs because of its very repetitive sequence, its large amount of Pro residues, and a signal dispersion of only 0.5 ppm for the amide protons, which leads to strong signal overlap at the ^1H - ^{15}N HSQC. This impedes the application of ^1H -detected assignment strategies starting at this experiment. Therefore, we applied a ^{13}C -detected assignment strategy based on the analyses of a set of only four NMR experiments, i.e. 2D CON, 3D hacacoNcaNCO, 3D hacaCONcaNCO, and 3D CBCACON. This strategy allow us to successfully assign the backbone ^{15}N , $^{13}\text{C}_\alpha$, and $^{13}\text{C}'$ nuclei as well as the $^{13}\text{C}_\beta$ carbons of C-H1.0 and its tri-phosphorylated form pT-C-H1.0. We propose this strategy as a best choice for NMR assignment in Pro-rich IDPs with repetitive sequences, which usually show very poor signal dispersion and a strong signal overlap in the ^1H - ^{15}N HSQC spectra.

The availability of the assignment has allowed us to perform the first structural and dynamics characterisation of the non- and tri-phosphorylated C-H1.0, free in aqueous solution. Analyses of chemical shifts and hetenonuclear $\{^1\text{H}\}$ - ^{15}N -NOEs indicated that the two forms have a similar structural and dynamics behaviour; both lack of any region with detectable secondary structure tendency, and are very flexible. The effect of phosphorylation is local and affects only the chemical shifts of the phosphorylated Thr residues (T22, T44 and T56), and their neighbor residues (preceding and following). Regarding the *cis/trans* Pro isomerism, all X-Pro bonds are *trans* in both C-H1.0 and

pT-C-H1.0 free in solution, so Thr phosphorylation does not affect the isomer state of the following Pro. To have the assignment of C-H1.0 and pT-C-H1.0 opens the way to further characterisation of these domains to understand the molecular bases of their biological relevance, such as DNA binding or interactions with other proteins. Of particular interest would be the characterisation of the secondary structure induced by DNA-binding, which in the phosphorylated forms might include *cis/trans* isomerization of prolines adjacent to phosphorylated residues. The application of the ^{13}C -detected CON-based assignment strategy described here, as well as the availability of the C-H1.0 chemical shifts would provide further details to the recently published NMR study of the interaction between ProT α and human histone H1.0,⁵⁵ in which ProT α was fully characterised but H1.0 was not assigned.

ACKNOWLEDGMENTS

This work has been supported projects CTQ2014-52633-P and CTQ2017-84371-P from Spanish MINECO and MCIU, respectively (both co-financed by European FEDER funds). BB was a recipient of pre-doctoral FPI scholarship BES-2015-073383 from Spanish MINECO. The NMR experiments were performed in the “Manuel Rico” NMR laboratory (LMR) of the Spanish National Research Council (CSIC).

Fig. 1. (a) Sequence of C-H1.0 including the cloning-tag residues (in italics). Phosphorylation sequences are underlined. Numbers above the sequence correspond to full-length H1.0, and below to the construct used in this work. (b) Bar plot showing the total number of each residue type in the C-H1.0 sequence. The absent residues are not included in the plot. (c) 2D ^1H - ^{15}N HSQC and (d) 2D CON spectra recorded for C-H1.0 at 1 mM concentration in $\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1 v/v at pH 5.5 and 25° C. The vertical y-axis of the upper part of the 2D CON is identical to that in the 2D ^1H - ^{15}N HSQC.

with the one letter code for amino acids and the residue number in the primary sequence according to the ^{15}N nucleus (see Fig. 1a; notice that each cross-peak corresponds to ^{15}N of residue i , and $^{13}\text{C}'$ of residue $i - 1$). Some crowd regions are zoomed for better visualization.

Fig. 3. (a) $\Delta\delta_{C\alpha}$ values ($\Delta\delta_{C\alpha} = \Delta\delta_{C\alpha}^{\text{observed}} - \Delta\delta_{C\alpha}^{\text{reference}}$, ppm) of C-H1.0 plotted as a function of residue number. Reference random coil values were taken from Poulsen (see Methods). (b) Chemical shifts perturbation (CSP) due to phosphorylation (see Methods). The horizontal line is the averaged CSP plus the standard deviation. (c) $\{^1\text{H}\}-^{15}\text{N}$ heteronuclear NOEs for C-H1.0 (circles) and pT-C-H1.0 (triangles) in $\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1 v/v at pH 5.5 and 25° C. Grey vertical bars indicate the positions of the phosphorylated Thr residues.

REFERENCES

1. Talbert, P.B. et al. A unified phylogeny-based nomenclature for histone variants. *Epigenetics Chromatin* **5**, 7 (2012).
2. Roque, A., Ponte, I. & Suau, P. Interplay between histone H1 structure and function. *Biochim Biophys Acta* **1859**, 444-54 (2016).
3. Roque, A., Ponte, I. & Suau, P. Post-translational modifications of the intrinsically disordered terminal domains of histone H1: effects on secondary structure and chromatin dynamics. *Chromosoma* **126**, 83-91 (2017).
4. Th'ng, J.P., Sung, R., Ye, M. & Hendzel, M.J. H1 family histones in the nucleus. Control of binding and localization by the C-terminal domain. *J Biol Chem* **280**, 27809-14 (2005).
5. Hendzel, M.J., Lever, M.A., Crawford, E. & Th'ng, J.P. The C-terminal domain is the primary determinant of histone H1 binding to chromatin in vivo. *J Biol Chem* **279**, 20028-34 (2004).
6. Roque, A., Orrego, M., Ponte, I. & Suau, P. The preferential binding of histone H1 to DNA scaffold-associated regions is determined by its C-terminal domain. *Nucleic Acids Res* **32**, 6111-9 (2004).
7. Kalashnikova, A.A., Rogge, R.A. & Hansen, J.C. Linker histone H1 and protein-protein interactions. *Biochim Biophys Acta* **1859**, 455-61 (2016).
8. Widlak, P. et al. The histone H1 C-terminal domain binds to the apoptotic nuclease, DNA fragmentation factor (DFF40/CAD) and stimulates DNA cleavage. *Biochemistry* **44**, 7871-8 (2005).
9. Zakharova, N.I. et al. [Prothymosin alpha interacts with C-terminal domain of histone H1 and dissociates p53-histone H1 complex]. *Mol Biol (Mosk)* **45**, 679-88 (2011).
10. Kalashnikova, A.A. et al. Linker histone H1.0 interacts with an extensive network of proteins found in the nucleolus. *Nucleic Acids Res* **41**, 4026-35 (2013).
11. Vila, R. et al. DNA-induced alpha-helical structure in the NH2-terminal domain of histone H1. *J Biol Chem* **276**, 46429-35 (2001).
12. Vila, R., Ponte, I., Jimenez, M.A., Rico, M. & Suau, P. An inducible helix-Gly-Gly-helix motif in the N-terminal domain of histone H1e: a CD and NMR study. *Protein Sci* **11**, 214-20 (2002).
13. Vila, R., Ponte, I., Jimenez, M.A., Rico, M. & Suau, P. A helix-turn motif in the C-terminal domain of histone H1. *Protein Sci* **9**, 627-36 (2000).
14. Vila, R., Ponte, I., Collado, M., Arrondo, J.L. & Suau, P. Induction of secondary structure in a COOH-terminal peptide of histone H1 by interaction with the DNA: an infrared spectroscopy study. *J Biol Chem* **276**, 30898-903 (2001).
15. Roque, A., Iloro, I., Ponte, I., Arrondo, J.L. & Suau, P. DNA-induced secondary structure of the carboxyl-terminal domain of histone H1. *J Biol Chem* **280**, 32141-7 (2005).
16. Caterino, T.L., Fang, H. & Hayes, J.J. Nucleosome linker DNA contacts and induces specific folding of the intrinsically disordered H1 carboxyl-terminal domain. *Mol Cell Biol* **31**, 2341-8 (2011).
17. Fang, H., Clark, D.J. & Hayes, J.J. DNA and nucleosomes direct distinct folding of a linker histone H1 C-terminal domain. *Nucleic Acids Res* **40**, 1475-84 (2012).

18. Fang, H., Wei, S., Lee, T.H. & Hayes, J.J. Chromatin structure-dependent conformations of the H1 CTD. *Nucleic Acids Res* **44**, 9131-9141 (2016).
19. Liao, R. & Mizzen, C.A. Interphase H1 phosphorylation: Regulation and functions in chromatin. *Biochim Biophys Acta* **1859**, 476-85 (2016).
20. Roque, A., Ponte, I., Arrondo, J.L. & Suau, P. Phosphorylation of the carboxy-terminal domain of histone H1: effects on secondary structure and DNA condensation. *Nucleic Acids Res* **36**, 4719-26 (2008).
21. Lopez, R. et al. Linker histone partial phosphorylation: effects on secondary structure and chromatin condensation. *Nucleic Acids Res* **43**, 4463-76 (2015).
22. Raghuram, N. et al. Pin1 promotes histone H1 dephosphorylation and stabilizes its binding to chromatin. *J Cell Biol* **203**, 57-71 (2013).
23. Novacek, J. et al. 5D ¹³C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. *J Biomol NMR* **50**, 1-11 (2011).
24. Pantoja-Uceda, D. & Santoro, J. New ¹³C-detected experiments for the assignment of intrinsically disordered proteins. *J Biomol NMR* **59**, 43-50 (2014).
25. Bermel, W. et al. High-dimensionality ¹³C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. *J Biomol NMR* **57**, 353-61 (2013).
26. Dziekanski, P., Grudziak, K., Jarvoll, P., Kozminski, W. & Zawadzka-Kazimierzczuk, A. (¹³)C-detected NMR experiments for automatic resonance assignment of IDPs and multiple-fixing SMFT processing. *J Biomol NMR* **62**, 179-90 (2015).
27. Marley, J., Lu, M. & Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* **20**, 71-5 (2001).
28. Renner, C., Schleicher, M., Moroder, L. & Holak, T.A. Practical aspects of the 2D ¹⁵N-[¹H]-NOE experiment. *J Biomol NMR* **23**, 23-33 (2002).
29. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-93 (1995).
30. Hyberts, S.G., Frueh, D.P., Arthanari, H. & Wagner, G. FM reconstruction of non-uniformly sampled protein NMR data at higher dimensions and optimization by distillation. *J Biomol NMR* **45**, 283-94 (2009).
31. Johnson, B.A. & Blevins, R.A. NMR View: A computer program for the visualization and analysis of NMR data. *J Biomol NMR* **4**, 603-14 (1994).
32. Williamson, M.P. Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc* **73**, 1-16 (2013).
33. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Mag Res Sp* **34**, 93-158 (1999).
34. Permi, P. & Annala, A. Coherence transfer in proteins. *Prog Nucl Mag Res Sp* **44**, 97-137 (2004).
35. Grzesiek, S., Anglister, J., Ren, H. & Bax, A. C-13 Line Narrowing by H-2 Decoupling in H-2/C-13/N-15-Enriched Proteins - Application to Triple-Resonance 4d J-Connectivity of Sequential Amides. *J Am Chem Soc* **115**, 4369-4370 (1993).
36. Panchal, S.C., Bhavesh, N.S. & Hosur, R.V. Improved 3D triple resonance experiments, HNN and HN(C)N, for H-N and N-15 sequential correlations in

- (C-13, N-15) labeled proteins: Application to unfolded proteins. *J Biomol NMR* **20**, 135-147 (2001).
37. Pantoja-Uceda, D. & Santoro, J. Aliasing in reduced dimensionality NMR spectra: (3,2)D HNHA and (4,2)D HN(COCA)NH experiments as examples. *J Biomol NMR* **45**, 351-6 (2009).
 38. Sun, Z.Y., Frueh, D.P., Selenko, P., Hoch, J.C. & Wagner, G. Fast assignment of ¹⁵N-HSQC peaks using high-resolution 3D HNcocaNH experiments with non-uniform sampling. *J Biomol NMR* **33**, 43-50 (2005).
 39. Bermel, W. et al. Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. *Angew Chem Int Ed Engl* **44**, 3089-92 (2005).
 40. Bermel, W. et al. H-start for exclusively heteronuclear NMR spectroscopy: the case of intrinsically disordered proteins. *J Magn Reson* **198**, 275-81 (2009).
 41. Mantylahti, S., Aitio, O., Hellman, M. & Permi, P. HA-detected experiments for the backbone assignment of intrinsically disordered proteins. *J Biomol NMR* **47**, 171-81 (2010).
 42. Mantylahti, S., Hellman, M. & Permi, P. Extension of the HA-detection based approach: (HCA)CON(CA)H and (HCA)NCO(CA)H experiments for the main-chain assignment of intrinsically disordered proteins. *J Biomol NMR* **49**, 99-109 (2011).
 43. Zerko, S. et al. Five and four dimensional experiments for robust backbone resonance assignment of large intrinsically disordered proteins: application to Tau3x protein. *J Biomol NMR* **65**, 193-203 (2016).
 44. Piai, A. et al. "CON-CON" assignment strategy for highly flexible intrinsically disordered proteins. *J Biomol NMR* **60**, 209-18 (2014).
 45. Bermel, W. et al. Speeding up sequence specific assignment of IDPs. *J Biomol NMR* **53**, 293-301 (2012).
 46. Motackova, V. et al. Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments. *J Biomol NMR* **48**, 169-77 (2010).
 47. Novacek, J., Haba, N.Y., Chill, J.H., Zidek, L. & Sklenar, V. 4D non-uniformly sampled HCBCACON and (1)J(NCalpha)-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. *J Biomol NMR* **53**, 139-48 (2012).
 48. Novacek, J., Janda, L., Dopitova, R., Zidek, L. & Sklenar, V. Efficient protocol for backbone and side-chain assignments of large, intrinsically disordered proteins: transient secondary structure analysis of 49.2 kDa microtubule associated protein 2c. *J Biomol NMR* **56**, 291-301 (2013).
 49. Kjaergaard, M. & Poulsen, F.M. Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J Biomol NMR* **50**, 157-65 (2011).
 50. Kjaergaard, M., Brander, S. & Poulsen, F.M. Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *J Biomol NMR* **49**, 139-49 (2011).
 51. Hafsa, N.E., Arndt, D. & Wishart, D.S. CSI 3.0: a web server for identifying secondary and super-secondary structure in proteins using NMR chemical shifts. *Nucleic Acids Res* **43**, W370-7 (2015).

52. Bienkiewicz, E.A. & Lumb, K.J. Random-coil chemical shifts of phosphorylated amino acids. *J Biomol NMR* **15**, 203-6 (1999).
53. Schubert, M., Labudde, D., Oschkinat, H. & Schmieder, P. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ¹³C chemical shift statistics. *J Biomol NMR* **24**, 149-54 (2002).
54. Shen, Y. & Bax, A. Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts. *J Biomol NMR* **46**, 199-204 (2010).
55. Borgia, A. et al. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **555**, 61-66 (2018).