Proline-Directed Random-Coil Chemical Shift Values as a Tool for the NMR Assignment of the Tau Phosphorylation Sites

Guy Lippens,^{*[a]} Jean-Michel Wieruszeski,^[a] Arnaud Leroy,^[a, b] Caroline Smet,^[a] Alain Sillen,^[a] Luc Buée,^[c] and Isabelle Landrieu^[a]

NMR spectroscopy of the full-length neuronal Tau protein has proved to be difficult due to the length of the protein and the unfavorable amino acid composition. We show that the randomcoil chemical shift values and their dependence on the presence of a proline residue in the $(i + 1)$ position can successfully be exploited to assign all proline-directed phosphorylation sites. This is a first step toward the study of the phosphorylation of Tau by NMR spectroscopy.

Introduction

The neuronal protein Tau in its aggregated state has long since been a marker for Alzheimer's disease.^[1] Its occurrence in paired helical fragments is mostly associated with hyperphosphorylation, although it is not yet clear which phenomenon, phosphorylation or aggregation, comes first.[2] Still, the aggregation of Tau is remarkable as the protein itself is highly soluble despite its apparent lack of structure. Structural studies on full-length Tau are not as numerous as one might expect for a protein of this physiological and medical importance, and no study of the atomic detail has been reported thus far. The lack of stable tertiary structure has hampered crystallization, and the sheer length (441 amino acids for the longest isoform) and unfavorable amino acid composition of the protein tend to preclude detailed NMR studies.

Our interest in Tau stems from a report that it is a potential substrate for the cell-cycle-linked prolyl cis/trans isomerase Pin1.^[3] This latter protein, comprising two distinct domains^[4] carrying the catalytic activity and a WW domain that typically mediates protein - protein interactions, recognizes many epitopes that are generated during mitosis and interact with the mitosis-specific MPM2 antibody.[5] As Tau in its hyperphosphorylated and aggregated form contains several epitopes for the MPM2 antibody, the authors looked at the Pin1 - Tau interaction and found, amazingly, that a single site (phospho-Thr231/ Pro232) was essential and sufficient for interaction between both partners.[3]

The structural and functional roles of this interaction, however, were not immediately clear. We showed that the interaction between the Pin1 WW domain and a Tau peptide centered around this very phospho-Thr231/Pro232 motif was mainly limited to the dipeptide, $[6]$ and the same limited recognition motif was concluded from an X-ray crystal structure of the fulllength Pin1 protein in which the WW domain was complexed to a peptide from the C-terminal domain of polymerase $I^{[7]}$ Very recently, we identified the phospho-Thr212/Pro213 motif as a second interaction motif, and there probably are other interaction sites between the 17 Thr/Pro or Ser/Pro proline-directed phosphorylation sites.[8]

During those studies, we and others have looked at the proline conformation, as it is related to the functional questions that surround Pin1. Indeed, because Pin1 stimulates dephosphorylation of Tau and CDC25 by the PP2A phosphatase and this phosphatase only recognizes the phospho-Thr/Pro motif in the trans conformation, $[9]$ it is tempting to conclude that Pin1 recognizes such a motif in the cis conformation, exerts its catalytic activity by raising the speed of the conformational transition, and thereby prepares a substrate for PP2A. However, the role of the WW domain is less clear in this scenario because both direct NMR spectroscopy observations^[6] and modelling^[7] showed that this latter domain cannot accommodate a Pro $_{ci}$. Still, for Pin1 acting as a conformational switch during the cell cycle, it would be reassuring to find some substrate with a Pro residue at least partly in the cis conformation. In the Thr212 peptide, we determined that none of the three proline residues

EIEMBIOCHEM

in the peptide was predominantly in the cis conformation, with a maximum of 12% for the Pro216/Thr217 peptide bond.^[8] The conformational preference for the trans configuration is even stronger in the Thr231 peptide and is independent of the phosphorylation state of this latter residue.[10] Since the protein context could exert such a local conformational strain that our conclusions derived from peptide studies cannot be extrapolated to the protein, we have set out to study the full-length protein by triple-resonance NMR spectroscopy. We report here our first results in the assignment process, whereby the length and unusual amino acid composition of Tau have forced us to abandon the conventional assignment strategy. The main aim of our work at this moment is to assign the potential phosphorylation sites, although all other assignments will probably be useful for future NMR studies directed at other aspects of the intriguing biological role of Tau.

Results

The 1D proton spectrum of Tau immediately shows why NMR studies on full-length Tau have not yet been successful. Indeed, chemical dispersion is limited to a mere 0.8 ppm, a fact confirming immediately the macroscopic unfolded nature of the protein (Figure 1). Moreover, if one looks at the amino acid composition, where five amino acids make up half of the sequence, the designation of "polymer" rather than protein makes sense.

The situation is somewhat better when one considers the nitrogen dispersion. Indeed, here we find a normal dispersion of 23 ppm, even though the core of the spectrum $(120 - 126$ ppm) remains very crowded. Also in this spectrum, we can note the narrow line width that allows us to record up to 128 complex points in the nitrogen dimension. For a protein of over 50 kD, this again indicates a great amount of local flexibility leading to an effective lengthening of the nitrogen $T₂$ times.

Our initial efforts to assign the protein used the now-familiar strategy of recording complementary triple-resonance experiments and exploiting them in a pairwise fashion.^[11] However, whether we looked at the HNCA/HN(CO)CA, HNCO/HN(CA)CO, or CBCANH/CBCA(CO)NH spectra, we found many lines with four to eight residues confined to the same ${}^{1}H_{N}{}^{15}N$ cross-peak. Considering that one wants to correlate the $Ca/C\beta$ /CO information of a given residue to the same information for the residue that precedes it on those spectra that use the carbonyl relay, the limited success of this approach for the assignment of Tau is obvious.

Still, because we had transformed the different 3D spectra in a pseudo-1D mode, starting from a semiautomatic peak picking of the high-resolution HSQC spectrum of Figure 1 (to yield 299 peaks), and had used a routine developed in house that allows individual thresholds to be set by hand for each of the 299 13C traces that result after the Fourier transformation in the third dimension (B. Odaert, G. Lippens, unpublished results), the peak picking in the third carbon dimension yielded a list of carbon frequencies per correlation peak in the ¹H/¹⁵N HSQC spectrum. By comparing the HNCA spectrum, which contains information about $Ca(i)$ and $Ca(i - 1)$, with the HN(CO)CA spectrum, which

Figure 1. a) 1D proton spectrum and b) 2D 1 H, 15 N HSQC spectrum of the uniformly ¹⁵N-labeled Tau protein.

only contains the latter $Ca(i - 1)$ information, we could distinguish the C α (*i*) and C α (*i* $-$ 1) frequencies without ambiguity. A similar procedure could be followed for the C β (*i*) and C β (*i* $-$ 1) frequencies.

As stated above, the extreme spectral degeneration obviously made these lists almost useless for the usual assignment strategy. However, a simple histogram listing the occurrence of each chemical shift value led to some interesting observations (Figure 2). First, we noted that certain chemical shift values only occurred as $C\alpha(i)$ values, whereas they could not be found in the list of the $Ca(i-1)$ values. Those carbon frequencies therefore indicate residues that precede a Pro residue, as the lack of an amide proton on the following residue and thus a correlation peak in the initial ¹H/¹⁵N HSQC spectrum makes their observation in the HN(CO)CA spectrum impossible. Moreover, we noted equally a well-populated region of Ca peaks between 64 -63 ppm; these peaks are only present as $Ca(i-1)$ correlations

FULL PAPERS

Figure 2. a) Graph indicating the number of times a given 13 C frequency is found as a C α (i) (blue) or C α (i $-$ 1) (red) value in the spectrum of the Tau protein. The distinction was based on comparison of the HNCA and HN(CO)CA spectra. b) Graph of the $C\beta$ chemical shift values, based on the CBCANH and CBCAcoNH spectra. Those resonances that were assigned to a $C(\beta(i))$ position are in blue, whereas those that correspond to a C β (i $-$ 1) position are indicated in red.

and not as C α (i) values. These are the C α chemical shift values of the proline residues themselves, and therefore the initial ¹H/¹⁵N correlation peaks that carry them have to originate from residues that follow a proline residue. In conclusion, whereas the Pro residues are rather a nuisance in the general assignment strategy because they interrupt the connection chain, here they immediately gave us a starting point for the assignment process. If one further considers that there are 44 Pro residues in Tau, this observation looks like a promising step for the assignment.

Another observation from the histogram (Figure 2) is that the $13C$ chemical shifts of the different residues are probably not very far from their random-coil values.^[12] Just considering the Gly residues, for example, we see that the C α chemical shifts are all within 0.5 ppm of their averaged value of 45.33 ppm. For Ala, the situation is barely better, with a similar spread of 0.5 ppm around the averaged value of 18.79 ppm for the methyl C β carbon atom. Again, those that deviate from this rule are the resonances of those residues that precede a Pro residue; for those, we expect a sequence-dependent chemical shift of 2 ppm to the high field.^[12]

This combined information about the direct neighborhood of a Pro residue and the minimal deviation of random-coil chemical shift values can be exploited in a simple graphical manner. We retransformed all 3D spectra, this time as full 3D matrices, and extracted selectively ¹³C planes according to the values suggested by the histogram. Rather than taking individual planes, however, we opted for sums of those planes within \pm 0.5 ppm of the random-coil value, as suggested by the histograms (Figure 2). An example of this strategy is given for the Ala/Pro motifs. There are five of them in Tau, and if our reasoning is correct, we should be able to identify them directly from the extraction of the carbon planes between $50 - 51$ ppm in the HNCA spectrum, with the planes between $17.8 - 18.4$ ppm in the CBCANH spectrum as an independent control. The resulting planes superimposed on the initial high-resolution ¹H/¹⁵N HSQC spectrum indeed show five peaks of major intensity (Figure 3). The superposition of the slice between $17.8 - 18.4$ ppm from the CBCANH spectrum confirms that we have identified the Ala/Pro motifs (Figure 4 a).

Figure 3. 2D¹H,¹⁵N slice from 50-51 ppm extracted from the HNCA spectrum (green) superposed on the original HSQC spectrum (red). At the top is the proton projection of the green spectrum.

HEMIBIO CHEMI

Figure 4. a) 2D ¹H,¹⁵N slice from 17.8 – 18.4 ppm extracted from the CBCANH spectrum (black) superposed on the original HSQC spectrum (red) and the slice from 50 – 51 ppm from the HNCA spectrum (green). b) The slice from 44.5 - 45.5 ppm from the HN(CO)CA spectrum (blue) defines the unique (G)A₈₄(P) Ala residue. c) The superposition of the slice from 52.0 - 53.0 ppm from the CBCA(CO)NH spectrum (blue) defines the unique AA₁₅₈P pattern. d) The slice from 69.5 - 70.5 ppm from the CBCA(CO)NH spectrum (blue) defines the two (T)A(P) residues at positions 77 and 246 in the primary sequence of Tau.

Further information about those five Ala/Pro motifs can be gained by considering that one is preceded by a Gly residue, one by another Ala, one by a Pro, and two by Thr residues. Building on the random-coil chemical shift values, the first can be identified by the intersection of the former five peaks and the planes from the HN(CO)CA spectrum between $45 - 46$ ppm, that contain all residues that follow a Gly (Figure 4b). Similar procedures can be applied for the Ala/Ala/Pro motif, where superposition of the planes between 18.7 - 19.7 ppm on the former mask leads to the unambiguous assignment of this peak at 8.37 ppm for proton and 125.1 ppm for nitrogen (Figure 4 c). The Pro/Ala/Pro can be assigned on the basis that it belongs to those residues that follow a Pro, characterized by a $Ca(i-1)$

chemical shift value around 63 ppm. Finally, the two Propreceding Ala residues that follow a Thr can be assigned on the basis of the characteristic C β chemical shift value of the latter, which is centered around 70 ppm (Figure 4 d). However, a sequence-specific assignment is not possible at this moment, as our method does not allow us to go further than the $i-1$ position.

When we look at an enlargement of these superimposed spectra (for example, Figure 4 c), two reasons for the difficulties experienced with the traditional list-based assignment programs become clear. The first one is evidently the poor resolution in the ¹⁵N dimension of the 3D spectra. However, one should consider that with a 500 μ s increment corresponding to the 2000 Hz window, one cannot get more than 47 points in the 22.9 ms evolution time for the $15N-13C$ antiphase magnetization to build up. Increasing this delay would lead to a rapid loss of magnetization, where both classical dipolar mechanisms and exchange with water protons would constructively cooperate to kill the magnetization. The second reason, however, is less evident for well-behaved proteins but becomes crucial for such polymers as Tau. Indeed, small differences in experimental conditions (due to slight differences in duty cycle and/or homo- or heteronuclear decoupling) lead to small shifts in the resonance peaks. This is not a problem at all for well-separated peaks, where one can immediately identify the same peak in the different spectra. Whereas the same is true for Tau upon using the human eye as a shape detector, a program will have more difficulty in recognizing which green and red peaks of Figure 4 c coincide.

We were especially interested in the assignment of the Thr/ and Ser/Pro motifs of Tau because of their definition as phosphoepitopes, as the identification of these motifs is a first step toward their conformational characterization. In the C α histogram (Figure 2 a), the Thr residues followed by a Pro are probably represented by the blue peak centered around 59.8 ppm, which is the random-coil value for a Thr followed by a Pro. $[12]$ Independent confirmation of the Thr identification comes again from the superposition of the slice between $69 - 71$ ppm of the CBCANH spectrum, where we see that at least the three peaks with (1 H, 15N) chemical shift values of (8.30, 122.18), (8.18, 122.74), and (8.21, 123.90) ppm are most probably not Thr residues (Figure 5). After closer inspection of the averaged chemical shift tables,^[12] they could well be Val residues preceding a Pro residue because their C α frquency is close to that of the Thr, whereas the $\mathsf{C}\beta$ value is readily distinguishable. The averaged ¹⁵N chemical shift of 121 ppm for a Val residue rather than the value of 115 ppm for Thr confirms this hypothesis.

Sequence-specific assignment of all ten Thr/Pro motifs proved somewhat trickier then the Ala/Pro case that we detailed above, because of their number (ten instead of five) and their lesser spread in the spectrum. Still, Thr153 could readily be identified as the peak at (8.27, 117.0) ppm, as it is the only Thr followed by a Pro and preceded by an Ala residue. Similarly, the single (G)T(P) Thr residue at position 205 can be assigned to (8.11, 115.6) ppm, and the (P)T(P) Thr residue at position 217 can be assigned to (8.39, 118.1) ppm. With the exception of the (K)T(P) and R(T)P motifs, where we encounter the same ambiguity as for the double (T)A(P) motif (see above), we list the assignment of all Thr residues preceding a Pro in Table 1. Similarly, we assigned all Ser/ Pro motifs and list their chemical shift values in Table 2.

Whereas Table 1 demonstrates that the strategy based on the random-coil chemical shift values, and more specifically on those of Pro-preceding residues, can be succesfully used to assign a majority of the residues that precede a proline, one should keep in mind that the success is based on the validity of the initial assumption that the protein can be considered as a contiguous collection of randomly structured small peptides, and as such may not be applicable to the assignment of any protein. Conversely, the success that we have had with the assignment (and that was confirmed by mapping of the assignment with HSQC spectra on small peptides (C. Smet, A. Sillen, G. Lippens,

FULL PAPERS

Figure 5. Definition of the Thr residues followed by a Pro on the basis of their C α frequency centered around 59.8 ppm (green) and C β frequency around 69.8 ppm (black). The three green correlations at frequencies of 8.2 $-$ 8.3 ppm (1 H) and 122 $-$ 124 ppm (¹⁵N) are Val residues.

unpublished results) does confirm the random-coil character of the Tau protein.

Currently, we are studying the phosphorylation of Tau induced by incubation of the protein with a cell extract of CDK5/p25 overexpressing neuroblastoma cells.^[13] The assignment of the Thr/ and Ser/Pro motifs reported in this paper will be crucial in validating the different phosphorylation sites of Tau and might

G. Lippens et al.

Table 2. Assignment of the different Ser residues followed by a Pro in the Tau sequence. Only the (K)S(P) is subject to ambiguity.

equally allow the potentially sequential nature of the phosphorylation of this protein to be detected for the first time.

Experimental Section

Expression of the recombinant Tau protein: A bacterial culture of Tau44 - pET15b-transformed Escherichia coli BL21(DE3) (Invitrogen) was grown at 37°C. The Tau44 - pET15b encoded the longest human Tau isoform (441 amino acid residues) under the control of a T7 promoter with a sequence encoding an N-terminal 6His tag (sequence numbering in the text does not take into account this 6His tag). When the optical density at 600 nm reached a value around 0.6, Tau production was induced by adding isopropyl-ß-Dthiogalactopyranoside (IPTG; 0.4 mm), and the growth was continued for 3 h before collecting the cells by centrifugation. In order to label the Tau protein with $15N$ and $13C$ stable isotopes, the culture was grown in a M9 minimal medium, with 15 NH₄Cl (1 g L⁻¹)and 13 C glucose (2 g L⁻¹; Cambridge Isotopes, Cambridge, MA) as the sole nitrogen and carbon sources.

Purification of the recombinant Tau protein: The cell pellet was resuspended in the extraction buffer (300 mm NaCl, 0.25 mm 1,4dithiothreitol (DTT), 1% Triton X100, 10 mm imidazole, 50 mm sodium phosphate, pH 7.4) complemented with a protease inhibitor cocktail (Complete, ethylenediaminetetraacetate (EDTA) free, Roche). The cell lysis was performed by sonication after addition of lysozyme and warming to 75° C for 15 min. The soluble extract was isolated by centrifugation and loaded on a Hitrap affinity column charged with $Ni²⁺$ ions (Amersham Pharmacia) and equilibrated in 300 mm NaCl, 5 mm imidazole, and 50 mm sodium phosphate (pH 8). Unbound proteins were washed away by 300 mm NaCl, 20 mm imidazole, and 50 mm sodium phosphate (pH 8), and the protein of interest was eluted by 300 mm NaCl, 200 mm imidazole, and 50 mm sodium phosphate (pH 7). The NMR spectroscopy sample was prepared by buffer exchange against a deuterated 50 mm tris(hydroxymethyl)aminomethane (Tris) buffer containing 100 mm NaCl (pH 6.8). The final NMR spectroscopy sample contained 200 μ M Tau protein, as estimated by absorption measurements at 280 nm.

NMR spectroscopy: All NMR spectra were recorded at 20°C on a 600 MHz Bruker DMX spectrometer (Bruker, Karlsruhe, Germany) equipped with a cryogenic triple-resonance probe head with actively shielded z gradient. The spectra were recorded with the pulse

sequences as described in the literature.^[14] Spectral windows were 7800 Hz for proton and 2000 Hz for nitrogen, centered at 4.8 ppm and 115.2 ppm, respectively, and sampled with 1000 and 47 complex points. The 30.7 ppm Ca window in the HNCA and HN(CO)Ca spectra was centered at 54.5 ppm and was sampled with 31 complex points. Similarly, the 20 ppm CO window was centered at 175.4 ppm and was sampled with 21 complex points. Finally, the 70.5 ppm combined C β and C α window in the CBCANH and CBCAcoNH spectra was centered at 40.2 ppm and was sampled with 72 complex points. All spectra were processed in the SNARF program (F. Van Hoesel, Groningen, The Netherlands), with 2000 complex points in the proton dimension and 512 complex points in the indirect nitrogen and carbon dimensions. Spectral referencing was done with respect to DSS in the proton dimension and by using the indirect referencing method^[15] for the nitrogen and carbon dimensions.

Acknowledgements

The 600 MHz facility used in this study was funded by the Région Nord - Pas de Calais (France), the CNRS, and the Institut Pasteur de Lille. Part of this work was funded by a grant of the Génopole of Lille and by a European Training and Mobility Grant.

Keywords: NMR spectroscopy \cdot phosphorylation \cdot proline \cdot protein structures \cdot Tau protein

- [1] A. Delacourte, L. Buée, Int. Rev. Cytol. 1997, 171, 167 224.
- [2] A. Schneider, J. Biernat, M. von Bergen, E.-M. Mandelkow, E. Mandelkow, Biochemistry 1999, 38, 3549 - 3558.
- [3] P.-J. Lu, G. Wulf, X. Z. Zhou, P. Davies, K. P. Lu, Nature 1999, 399, 784 788.
- [4] R. Ranganathan, K. P. Lu, T. Hunter, J. Noel, Cell 1997, 89, 875 886.
- [5] M. B. Yaffe, M. Schutkowski, M. Shen, X. Z. Zhou, P. T. Stukenberg, J. U. Rahfeld, J. Xu, J. Kuang, M. W. Kirschner, G. Fischer, L. C. Cantley, K. P. Lu, Science 1997, 278, 1957 - 1960.
- [6] R. Wintjens, J.-M. Wieruszeski, P. Rousselot-Pailley, H. Drobecq, G. Lippens, I. Landrieu, J. Biol. Chem. 2001, 276, 25 150 - 25 156.
- [7] M. A. Verdecia, M. E. Bowman, K. P. Lu, T. Hunter, J. P. Noël, Nat. Struct. Biol. 2000, $7,639-643$.
- [8] C. Smet, A.-V. Sambo, J.-M. Wieruszeski, A. Leroy, I. Landrieu, L. Buée, G. Lippens, Biochemistry 2003, submitted.
- [9] X. Z. Zhou, O. Kops, A. Werner, P. J. Lu, M. Shen, G. Stoller, G. Kullertz, M. Stark, G. Fischer, K. P. Lu, Mol. Cell 2000, 6, 873 - 883.
- [10] N. L. Daly, R. Hoffmann, L. Otvos, Jr., D. J. Craik, Biochemistry 2000, 39, $9039 - 9046.$
- [11] M. Ikura, L. E. Kay, A. Bax, Biochemistry 1990, 29, 4659 4667.
- [12] D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges, B. D. Sykes, J. Biomol. NMR 1995, 5, 67 - 81.
- [13] M. Hamdane, A.-V. Sambo, P. Delobel, S. Begard, A. Violleau, A. Delacourte, P. Bertrand, J. Benavides, L. Buée, J. Biol. Chem. 2003, 278, 34026 - 34034.
- [14] S. Grzesiek, A. Bax, J. S. Hu, J. Kaufman, I. Palmer, S. J. Stahl, N. Tjandra, P. T. Wingfield, Protein Sci. 1997, 6, 1248 - 1263.
- [15] D. S. Wishart, C. G. Bigam, J. Yao, F. Abildgaard, H. J. Dyson, E. Oldfield, J. L. Markley, B. D. Sykes, J. Biomol. NMR 1995, 6, 135 - 140.

Received: September 5, 2003 [F 763]