Accepting its Random Coil Nature Allows a Partial NMR Assignment of the Neuronal Tau Protein

Caroline Smet,[a] Arnaud Leroy,[a, b] Alain Sillen,[a] Jean-Michel Wieruszeski,[a] Isabelle Landrieu,[a] and Guy Lippens*[a]

A combined strategy to obtain a partial NMR assignment of the neuronal Tau protein is presented. Confronted with the extreme spectral degeneracy that the spectrum of this 441 amino acid long unstructured protein presents, we have introduced a graphical procedure based on residue type-specific product planes. Combining this strategy with the search for pairwise motifs, and combining the spectra of different Tau isoforms and even of peptides derived from the native sequence, we arrive at a partial assignment that is sufficient to map the interactions of Tau with its molecular partners. The obtained assignments equally confirm the absence of regular secondary structure in the isolated protein.

Introduction

The physiological function of the neuronal Tau protein involves the dynamic stabilization of tubulin into microtubules.[1, 2] Hence it plays an important role in neuronal development and plasticity. Moreover, the same microtubule-associated Tau protein is one of the major components of the neurofibrillary paired helical fragments (PHFs) characteristic of Alzheimer's disease (AD).[3, 4] Primarily composed of Tau, these tangles accumulate intracellularly in the diseased neurons, and their presence correlates rather well with the cognitive impairments that result from the disease.[5]

Despite its enormous biological and medical relevance, little is known about the detailed structure of Tau in solution. Small-angle X-ray scattering (SAXS) and other spectroscopic techniques such as circular dichroism (CD) and infrared spectroscopy (IR) have lead to a description of the protein as a random coil or Gaussian polymer.[6, 7] However, as these techniques only give an overall view of the global structure, local elements of secondary and/or tertiary structure might well escape observation. The absence of a stable tertiary structure precludes the use of X-ray crystallography, but, as an alternative, NMR spectroscopy might well give hints about otherwise undetected sequence specific structural elements. Although this latter technique has been successfully applied in the case of several other unfolded proteins,[8–10] the literature contains no detailed report about the NMR characterization of Tau.

We have recently shown that proline residues and their distinct influence on the random coil chemical-shift values can be used to assign at least those residues in Tau that precede a proline amino acid.[11] In order to extend the assignment to all residues irrespective of their direct neighbours, we demonstrate here the utility of a novel graphical method to handle the huge information content in the triple-resonance spectra of Tau. Pushing the comparison of the spectra of full-length adult Tau and its shorter foetal isoform to the limit of a comparison with natural-abundance spectra of short peptides, allows us to proceed further in the assignment process. We demonstrate the assignment of Tau to such a level that i) the presence or absence of well-defined structural elements can be ascertained, and ii) the interaction with other molecular partners such as tubulin or aggregating agents can be studied.

Results

Conventional assignment strategies that use triple resonance experiments, correlate the Cα, CO or Cβ of a given residue with that of its immediate neighbour.[12] A typical example of this strategy is the use of HNCA and HN(CO)CA experiments, where the Cα of residue [*i*] is seen from its own [*i*] amide cross peak in the first experiment and from the amide correlation of the next residue [*i*+1] in the HN(CO)CA. When applied to the full-length Tau protein, spectral overcrowding that results from i) the sheer size of the protein (Figure 1), ii) the limited chemical-shift dispersion that results from the unfolded nature of the protein and iii) the high degeneracy in the primary sequence (Table 1) all made this strategy largely unsuccessful. The only nucleus that gives a good chemical-shift dispersion is the nitrogen, as can be seen from the HSQC spectrum (Figure 2). We have previously observed that all Glycine Cα carbons, which represent a total of 49 residues (Table 1), fall within 0.5 ppm of their random-coil value.[13] We found the same extreme de-

[a] *C. Smet, Dr. A. Leroy, Dr. A. Sillen, Dr. J.-M. Wieruszeski, Dr. I. Landrieu, Dr. G. Lippens*
CNRS—Université de Lille 2, UMR 8525, Institut Pasteur de Lille
B.P. 245, 59019 Lille Cedex (France)
Fax: (+33) 320-871-233
E-mail: guy.lippens@pasteur-lille.fr

[b] *Dr. A. Leroy*
Laboratoire de Biochimie Appliquée
Faculté de Pharmacie à Châtenay-Malabry (Paris XI)
Tour D4 2ème étage, 5 rue Jean-Baptiste Clément
92296 Chatenay-Malabry Cedex (France)

 DOI: 10.1002/cbic.200400145

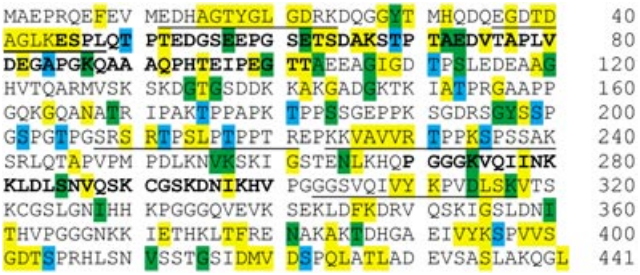

Figure 1. *Primary sequence of Tau. The amino acids that are absent from the shortest Tau352 isoform are in bold. The peptides synthesized for the peptide mapping are underlined. Residues that are unambiguously assigned are shown in yellow and those with only a twofold uncertainty in green. Light blue indicates the S/T(P) sites that were previously assigned.*[11]

Table 1. *Sequence degeneracy for the full-length Tau441 protein. Only the ten most abundant amino acids are listed.*

Amino acid	Number	Percentage
Gly	49	11.1
Ser	45	10.2
Lys	44	10.0
Pro	43	9.8
Thr	35	8.0
Ala	34	7.7
Asp	29	6.6
Glu	27	6.1
Val	27	6.1
Leu	21	4.8

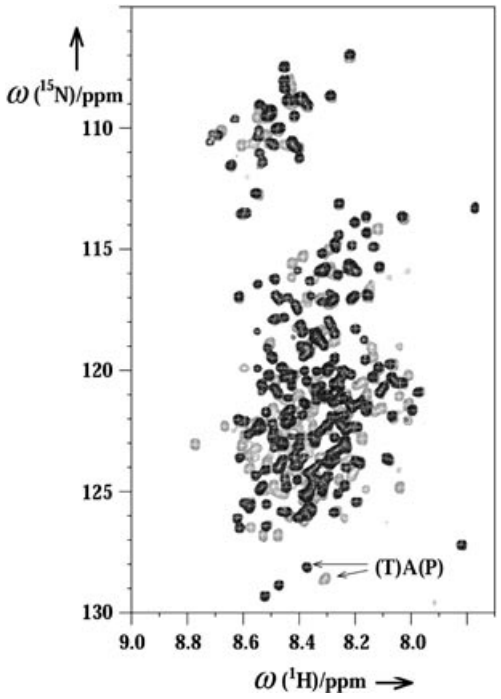

Figure 2. *^{1}H,^{15}N HSQC spectrum of the adult isoform Tau441 (light) and the smaller foetal isoform Tau352 (dark). The two (T)A(P) correlation peaks are indicated by an arrow.*

generacy for all other carbons including the carbonyl frequencies.[11]

The absence of carbon chemical-shift dispersion can, however, be used constructively once we accept the random-coil chemical-shift values from the onset on to be correct. Indeed, it means that we not only know the carbon chemical-shift values of every residue, but that we can also utilize them to identify those same residues, in a manner similar to amino acid selective labelling. As the random-coil values can be equally used to extract the nature of the residue in the [$i-1$] position, we have to deal with a pairwise degeneracy. In the following paragraphs, we will demonstrate that this degeneracy is less stringent than initially thought, but we will first illustrate our graphical strategy based on product planes and pattern searching on the concrete example of the assignment of lysine residues in Tau.

Lysine assignment

Lys is characterized by a Cα chemical-shift value of 56.2 ppm and a Cβ value of 33.1 ppm.[13] Therefore, if we extract the corresponding planes for the Cα carbon chemical shift from a HNCA spectrum and for the Cβ carbons from the CBCANH spectrum, two planes result whose intersection should contain the Lys resonances. To illustrate this observation, we show in Figure 3A the sum of all ^{1}H,^{15}N HNCA planes with carbon chemical-shift values between $(56.2-0.5 = 55.7)$ ppm and $(56.2+0.5 = 56.7)$ ppm, which allow a deviation of 0.5 ppm from the listed random-coil value based on the carbon histograms that we previously derived for Tau.[11] A similar procedure for the planes with Cβ values comprised between 32.6 and 33.6 ppm (Figure 3B) leads to the two basic planes that define the Lys residues. Rather than superimposing both planes, however, we define the Lys *product plane* as the ^{1}H,^{15}N plane resulting from a point-by-point multiplication of both Lys Cα and Cβ defining planes (Figure 3C). The result of this operation is that we only recover intensity where both carbon chemical-shift values fall within the correct interval, thereby simplifying the visual information of both parent planes. For example, the peak at (8.09−123.5) ppm, present in the Lys Cα plane, is absent from the Lys product plane, because of its different Cβ value (30.3 ppm, as extracted from the CBCANH spectrum).

Lysine residues allow an immediate experimental verification of our assignment strategy since a selective Lys labelled sample is relatively simple to produce.[14] A comparison of the Lys product plane with the experimental spectrum of the Lys labelled sample is illustrative of the limitations of our method (Figure 3C). First, the majority of the experimental lysine peaks are present in the product plane. The two peaks at (8.21, 125.4) ppm and (8.41, 124.7) ppm that violate this principle, concern Lys residues that precede a Pro and therefore undergo a 2 ppm deviation towards the high field.[13] Secondly, the peaks that do coincide between the product plane and the spectrum of the Lys labelled sample are not all of equal intensity. Because the intensity in the product plane directly results from the intensity in the HNCA and CBCANH plane, an obvious

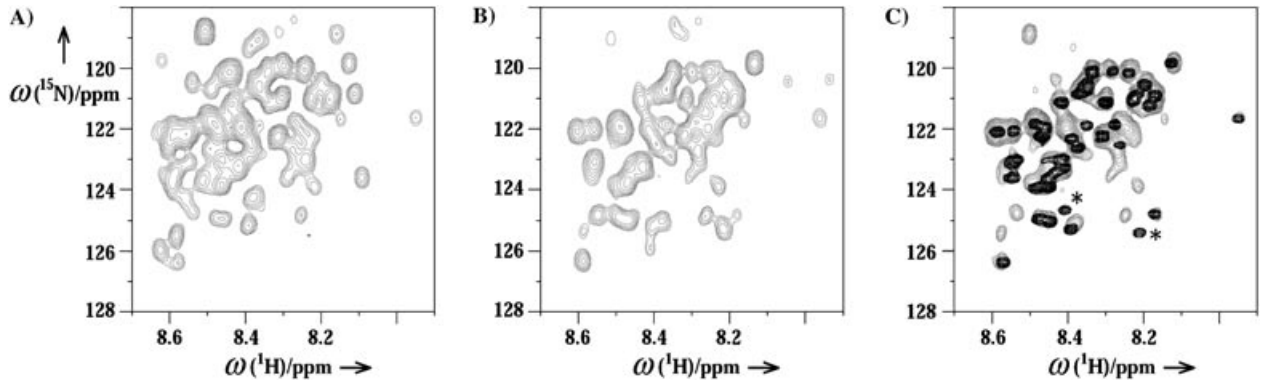

Figure 3. *A) Plane around the Lys Cα random coil chemical shift of 56.2 ppm, extracted from the Tau441 HNCA spectrum and B) around its Cβ value of 33.1 ppm, extracted from the CBCANH spectrum. C) Product plane (light) defined by the point-by-point multiplication of the planes in (A) and (B), with superimposed (dark) $^{1}H,^{15}N$ HSQC spectrum of the lysine, selectively labelled in Tau441.*

reason for the intensity variations is the less efficient transfer of magnetization in one or both experiments. A deviation of the assumed random coil chemical-shift value might be a second unrelated cause of intensity variations, as the parent planes were constructed as slices around the random-coil values, such that a deviation superior to 0.5 ppm would result in a reduced or absent intensity of the peak. A third observation is that the product plane contains more signals than just the expected Lys correlation peaks. This is the case for the strong correlation found at (8.50, 118.8) ppm, absent from the spectrum of the Lys-labelled sample. Close inspection of the underlying frequencies shows that this residue is a Ser followed by a Pro, leading to a 2 ppm upfield shift of its Cα random-coil chemical shift from 58.3 to 56.4 ppm.[11,13] Moreover, the residue is preceded by a Lys, and the efficient magnetization transfer from the Ser nitrogen to the preceding Lys Cα (and hence its Cβ) leads to some intensity at the Lys Cβ position in the CBCANH spectrum as well. Whereas specific experiments favouring the intraresidue transfer over the inter-residue one have been developed to simplify the triple-resonance spectra[15,16] and would avoid artefacts such as the one described above, the individual subspectra generated by our product-plane approach contain substantially less information than the initial HSQC spectrum; this brings manual inspection within reach. Finally, a close look at the random-coil chemical shift tables reveals that Met residues, equally, might be represented by the intersection of both planes, because the upper limits for both the Cα (55.4+0.5 = 55.9 ppm) and the Cβ (32.9+0.5 = 33.4 ppm) intersect with the Lys-defining carbon intervals. This overlap further accounts for the observed "leakage" of correlation peaks in the Lys product plane.

These latter observations may appear to cast some doubt on the utility of the product plane approach, but we claim that our method represents a true advantage when one considers that the 3D HNCOCA and CBCACONH experiments can be used in an identical way to extract product planes defining the nature of the residue in the [$i-1$] position. In Tau, there are four Pro–Lys patterns. The Pro [$i-1$] product plane can readily be constructed from the sum of planes around 63.3 ppm in the HNCOCA and 32.1 ppm in the CBCACONH spectra, and shows all amide signals from those residues that are preceded by a Pro (Figure 4A). If we now form the product plane of the initial Lys product plane of Figure 3C with the Pro [$i-1$] plane of Figure 4A, we obtain the $^{1}H,^{15}N$ subspectrum of those Lys residues that are preceded by a Pro (Figure 4B). There are four correlations as expected, demonstrating the validity of the strategy.

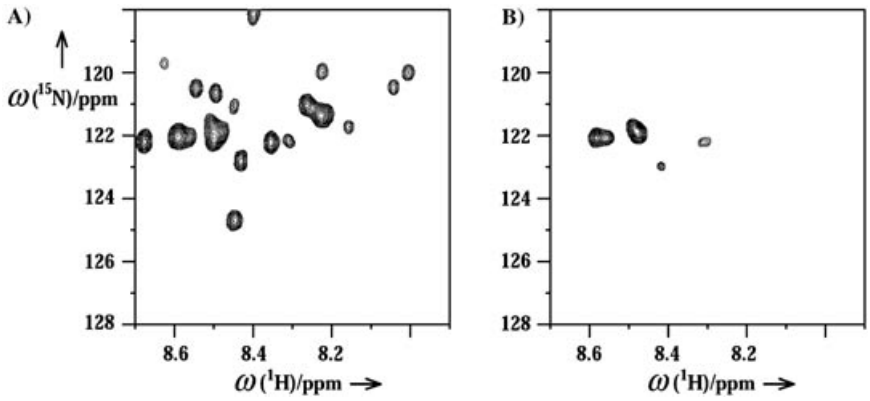

Figure 4. *A) Product plane defining those residues that have a Pro in the [i−1] position. B) The product of the Lys plane of Figure 3C and the Pro [i−1] plane of Figure 4A defines the resonance position of the four (Pro) Lys pairs in Tau.*

The described procedure does not allow us to go further in the identification of these (Pro)Lys patterns, as we do not connect more than two consecutive residues. However, despite the extreme degeneracy in amino acids for Tau, its sequence does contain 79 unique pairwise patterns, and another 49 residue pairs that occur only twice in the protein. These statistics

show that the product planes conjugated with a pairwise pattern search, can lead to an acceptable level of assignment without having to go through extensive lists of assignments.

Alternative splicing

Tau occurs in human neurons in six isoforms. The longest isoform (Tau441) is mainly present in adult brain, whereas the shortest isoform, with 352 amino acids (Tau352), is dominant in foetal brain.[17] The two proteins differ in two inserts in the N-terminal region, absent in the shortest isoform, and in the absence of the second microtubule binding repeat for the foetal form (Figure 1). Based on the presumed absence of secondary or tertiary structure, we expected the NMR spectrum of the shorter isoform to be similar to that of the longer Tau441, but with fewer cross peaks and some border effects for those residues that immediately flank the missing regions in Tau352. A comparison of the resulting HSQC spectra of both proteins (Figure 2) confirms this, with the Tau352 HSQC roughly appearing as a subspectrum of that of Tau441. This observation allows us to solve some ambiguities. We have previously reported the assignment of all 5 Ala–Pro motifs in Tau,[11] but for the two (Thr)Ala(Pro) residues, the assignment was not unique, especially as both Thr residues have degenerate Cα and Cβ chemical shifts. One of the (Thr)Ala(Pro) motifs, however, centred around Ala77 (where we consistently use the Tau441 isoform numbering), is in the N-terminal insert that distinguishes Tau352 from Tau441. In the spectrum of Tau352, the peak at (8.30, 128.5) ppm is absent (Figure 2), allowing us to assign it this time without any ambiguity to Ala77, whereas the other Ala residue at position 246 necessarily has its correlation at (8.37, 128.1) ppm. Other similar examples exist, as of the 49 patterns that occur twice in the Tau sequence, 19 are absent in the Tau352 sequence, and hence become unique. This removes the ambiguity for those 19 pairs, or for approximately another 10% of the protein sequence.

Splicing to the peptide level

Other isoforms of Tau exist in the brain,[17] and could potentially be used in a similar way as Tau352 to assist the assignment. However, we decided to push this strategy to the limit of small peptides, thereby circumventing the need to produce isotope-labelled samples. Several peptides of 13–17 amino acids (Figure 1) were synthesized by solid-phase synthesis. The peptides were solubilised at a concentration of 4 mM in the same buffer as was used for the protein spectra, and standard DQF-COSY, TOCSY and NOESY spectra were recorded for the assignment. The absence of secondary structure elements was verified by the measurement of coupling constants and analysis of the NOE pattern. Coupling constants were consistently found in the range of 6–7 Hz, and except for the sequential $H_\alpha(i)-H_N[i+1]$ contacts, no other NOE cross peaks of significant intensity were detected. Heteronuclear $^1H,^{15}N$ correlation spectra at natural abundance were recorded overnight by using a cryoprobe on a 600 MHz instrument and peaks were assigned on the basis of the proton frequencies derived from the homonuclear TOCSY spectra. The HSQC spectra of the Lys224–Lys240 peptide and of full-length Tau show that peptide resonances indeed coincide for the most part with amide signals of the full-length protein (Figure 5A). Because both Arg230 and

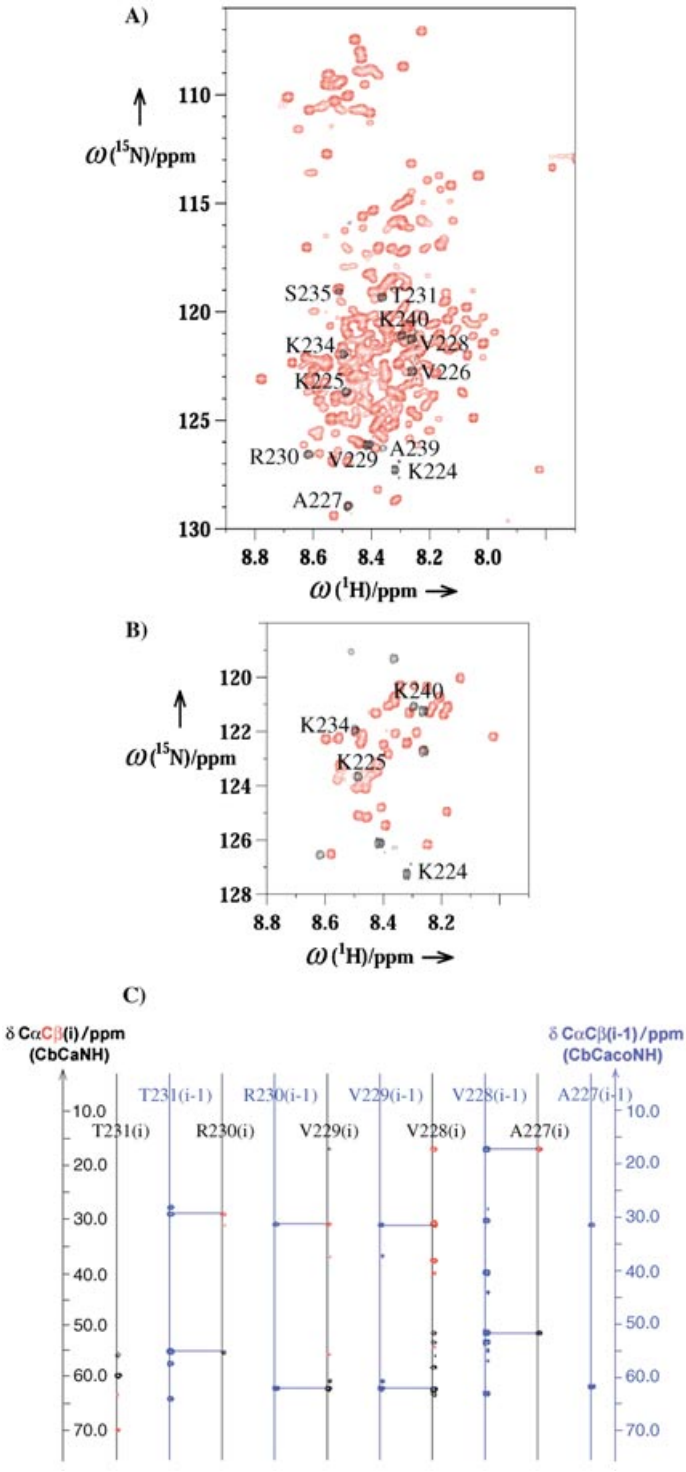

Figure 5. *A) Superposition of the uniformly or B) ^{15}N-Lys labelled Tau441 HSQC spectrum (red) with the same spectrum on the Lys224–Lys240 peptide (black). C) Strips extracted from the CBCANH (black for positive and red for negative contours) and CBCA(CO)NH (blue) spectrum of $^{15}N/^{13}C$ labelled Tau441 confirming the assignment of the Thr231–Ala227 stretch.*

Ala227 are unique (Val)Arg and (Val)Ala motifs had already been unambiguously assigned, we could immediately verify that those two amino acids give the same signal when embedded in the peptide or full-length protein. The correlation peaks of the first two residues of the peptide, Lys224 and Lys225, did not coincide with the equivalent signals in the protein, as verified on a HSQC spectrum of a Lys selectively labelled sample (Figure 5B), leading to the conclusion that border effects extend to at least two residues. Further verification of the other resonances in the crowded region of the full-length protein spectrum was done on the basis of strips extracted from the CBCANH and CBCA(CO)NH spectra on the doubly labelled Tau441 (Figure 5C). The C-terminal part of the peptide following Pro236 was characterized by resonance broadening of the amide signals, hindering the assignment of the two Ser residues.

Whereas the good agreement between peptide and protein spectra was verified for several other peptides (Figure 1), we did experience solubility problems with the Gly303–Lys317 fragment. This peptide contains the PHF_6 motif previously defined to be the nucleus of the aggregation because of its strong interaction with the proteolytic PHF_{43} fragment, which also forms fibres in solution.[18] The synthesis of the peptide had not given particular problems, but we did not succeed well in dissolving it in our standard buffer. Based on the signal intensity of the NMR signals in the supernatant, we estimate that only 200 μM peptide remains in solution, whereas the rest readily precipitates. This concentration, though not sufficient for a $^{1}H,^{15}N$ HSQC at natural abundance, did allow the recording of the classical homonuclear spectra. Based on the TOCSY and NOESY spectra, we could assign the proton frequencies of the amide protons and in Figure 6, we show the assignment of the Val309 and Tyr310 amide protons on top of the HSQC spectrum of ^{15}N labelled Tau441. Our pairwise motif search had not succeeded in unambiguously assigning both residues, as exactly the same motif is found later in the sequence (Ile393-ValTyrLys396). However, for both residues, two candidate correlation peaks had been identified in the HSQC spectrum (Figure 6). The H_N chemical-shift values of Val309 and Tyr310 in the peptide spectrum indicate that the peaks at (8.24, 126.1) ppm and (8.51, 126.4) ppm correspond to the two residues.

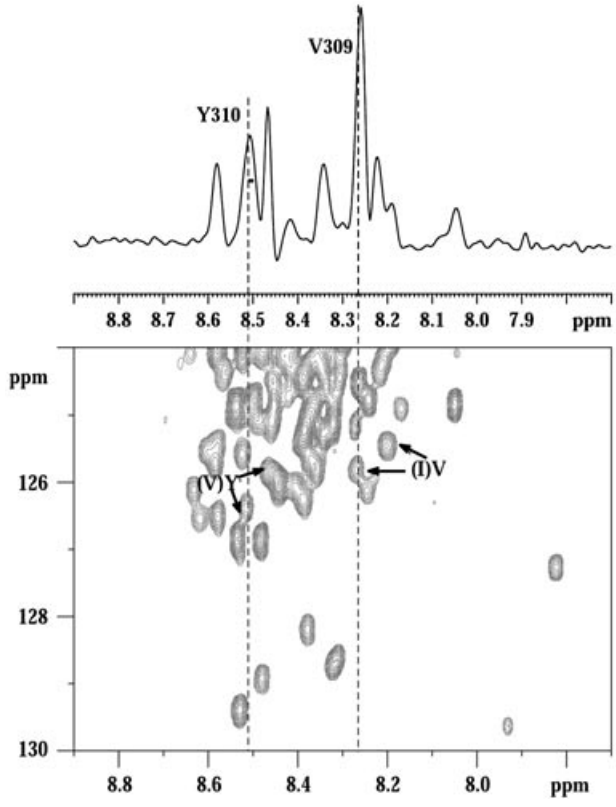

Figure 6. *A) Proton 1D spectrum of the Gly303–Lys317 peptide and B) corresponding zone of the Tau441 HSQC. Candidate peaks for both (Ile)Val and (Val)Tyr are indicated.*

Discussion

The increasing number of natively unfolded proteins[19] poses two challenges to structural biology, one fundamental—if the protein is natively unstructured, what is the meaning of a structural biology effort, at least at the level of the isolated biomolecule?—and one methodological—these proteins tend to resist crystallization efforts, and NMR spectroscopy is hampered by a reduced chemical dispersion. Despite these issues, the biological and medical importance of a large number of these proteins has motivated a substantial effort in obtaining information regarding their conformation and dynamics. An important class of natively unfolded proteins is found to be involved with neurological disorders, such as α-synuclein in Parkinson's disease and Tau in Alzheimer's disease. Whereas an NMR effort on the former α-synuclein has given a glimpse of the dynamical effect of certain disease-related mutations,[20] there is a lack of detailed information on Tau. This protein plays an important role in the dynamic process of tubulin polymerization into microtubules, with four repeat regions of some 30 amino acids, dubbed the MT binding repeats, that are directly involved in this interaction.[2,17] Besides its normal physiological role and despite the high solubility of the isolated protein, Tau also aggregates to form the PHF motif, a hallmark of AD that directly correlates to the severity of the cognitive impairments in the patients.[5]

Because of the importance of the biological and medical phenomena associated with Tau, a large effort has gone into its structural characterization. However, all macroscopic methods have concluded the absence of regular secondary structure.[6,7] Selected studies on isolated microtubule binding domains or short peptide fragments derived from these domains have suggested a tendency for them to adopt an extended β-strand like structure, although this conformation seems only adopted after the initial stages of aggregation.[18] Still, the structure of these same peptides within the full protein remains largely unknown.

Since the amide proton dispersion of Tau is limited to less than 1 ppm, it is an unambiguous confirmation that the protein is macroscopically unfolded, but this lack of proton chemical shift dispersion is simultaneously one of the main reasons

that NMR studies of Tau have been devoid of success. A second obstacle for its NMR study is the sequence degeneracy of Tau, where five amino acids make up for almost 50% of its full length (Table 1). Moreover, the ^{13}C chemical-shift values of these residues are extremely limited, leaving only the nitrogen chemical shift as element of spectral dispersion (Figure 2). Therefore, despite the fact that triple resonance experiments are easily collected (relaxation properties of the protein are very favourable despite its large size, and TROSY[21,22] techniques do not lead to any detectable signal enhancement), the traditional triple-resonance NMR strategy of connecting through the combined Cα, Cβ and CO carbon values a given residue with its next neighbour is of limited value.

Despite these discouraging considerations, the lack of carbon chemical shift dispersion and their coincidence with random-coil values can be used in a constructive way if one adopts them directly as the true values. When we first applied this to those residues that are followed by a proline, where a 2 ppm shift towards the high field is expected[13] we were able to assign many of them, although some ambiguity persisted for double of triple pairs.[11] Here, we have gone a significant step further in the assignment process of Tau441. Clearly, the basic hypothesis of our previous work—the random coil chemical-shift values apply for all residues over the whole sequence—is maintained. The direct implication is that we know the residue type directly from the Cα and Cβ values. Nevertheless, because the information content of the spectra is so overwhelming, and inspired by the colocalization process commonly employed in confocal microscopy, we introduced a visual procedure in the form of product planes, centred directly around the random-coil values. The advantage of this approach is to directly obtain subspectra of the original $^{1}H,^{15}N$ HSQC map, precluding the need to go through an entire list of peaks to find those that match the required residue type both at the level of their Cα and Cβ chemical-shift values. Potential pitfalls are that residues could be missed if their carbon chemical-shift values fall outside the specified windows, or because magnetization transfer in the Cα or Cβ directed experiment has been hampered by relaxation losses. The experimental approach of residue specific labelling does not suffer the same limitations, but residue specific labelling requires several samples and special efforts to overcome isotopic scrambling. The product plane approach can be performed for all residues starting from a single set of triple resonance experiments recorded on a unique sample, but the equivalent of spectral scrambling cannot be prevented, as we have shown with the example of Lys and Met. A less obvious advantage of the experimental residue specific labelling is that the resulting spectrum can be acquired with any resolution that the operator decides, simply by collecting more increments in the indirect dimension. The product planes, on the other hand, are limited by resolution in the indirect nitrogen dimension of the corresponding 3D experiments, which in most cases comprises substantially fewer than 128 complex points.

The main advantage of the product plane approach is the possibility to derive not only the residue type corresponding to a given correlation peak, but also the residue type of the preceding residue from the carbonyl relayed triple-resonance spectra. Whereas this has been obtained for the [*i*] and [*i*−1] residue experimentally by the double labelling of an antibody Fc fragment,[23] the approach is clearly not applicable to all amino acid pairs. The assignment of doublets, of course, is equally a major part of the traditional triple-resonance strategy, but there, the main effort is to connect the correlation peaks further to obtain stretches of amino acid residues that are then mapped onto the protein sequence. Here, we did not attempt to go further, but used the fact that Tau441 contains 79 unique pairwise patterns and another 49 residue pairs that occur only twice in the protein; this provides a starting basis for at least a partial assignment of the protein.

Comparing the spectra of the full-length protein and its fragments to divide the assignment problem is clearly limited to the case of natively unstructured proteins. We have used the Tau352 spectrum for the elimination of cross peaks in those cases in which the initial assignment was ambiguous, as shown for the two (T)A(P) cross peaks in Figure 2. Other isoforms of Tau could be used in a similar way, but we have opted to divide the full-length Tau protein into short peptides, and recompose the HSQC spectrum of Figure 2 as a superposition of the HSQC spectra of the short peptides. Central to our approach is the statement that the same amino acid stretch in the isolated peptide or in the full protein should give an identical HSQC spectrum when the individual amino acids sample the same conformational space in both contexts. That this will actually be the case was not obvious at all, as especially the nitrogen chemical shift is extremely sensitive to the time-averaged spatial environment of the amide group. Strikingly, empirical prediction methods that use the peptide or database derived chemical shift parameters still have the most problems with nitrogen chemical shifts as well.[24,25] The consequence of this extreme sensitivity is that in the case of a good agreement, we can safely say that the amino acids indeed sample conformational space in a very similar way, be they embedded in a short peptide or in the full protein. Disagreements, on the other hand, indicate structural differences between the isolated peptide and the equivalent amino acid stretch in the full-length protein, and therefore point to regions in the protein that should be examined in more detail by other NMR methods.

Peptides should be sufficiently small allowing for their straightforward chemical synthesis, purification and NMR analysis. An additional reason for a limited length is that they should not adopt a stable structure (stressing our initial hypothesis that the protein is natively unfolded). They should not be too short, however, because we expect some spectral border effects for those residues that are at the N or C terminus of the peptide and, hence, do not have the same chemical environment as their equivalents in the full-length protein. Our present work indicates that these border effects generally do not exceed two residues at both termini. Because we had previously synthesized a number of peptides of 13 to 17 amino acids containing potential phosphorylation sites,[26] we first used these and extended them with some additional peptides from the N terminus or third microtubule binding repeat

(Figure 1). NMR analysis led to the conclusion of absence of stable secondary structure elements in all of them. The second stage of the procedure consists in the recording of $^{1}H,^{15}N$ HSQC spectra on these peptides in the same buffer conditions, pH and temperature as for the full-length protein. Present cryogenic probe heads[27,28] installed on a 600 MHz spectrometer enable such spectral recordings within one night, if the peptide concentration is in the millimolar range or larger. The graphical superposition of the peptide spectra with the full-length Tau HSQC completes the procedure and the triple-resonance spectra can be used to verify rather than determine the identical assignment of the correlation peaks. The superposition of the Lys224–Lys240 peptide and Tau441 spectra (Figure 5A) indeed shows a perfect agreement for the isolated Ala227 and Arg230 correlation peaks. The strips extracted for the Thr231–Ala227 stretch at the exact positions of the peptide resonances (Figure 5C) show that the sequential walk is indeed possible and thereby confirm that the superposition is valid. However, the CBCA(CO)NH strip extracted for Thr231 shows the difficulty that we would face if we had to use the same information for the ab initio assignment of full-length Tau. We indeed observe three Cα and two Cβ peaks, but have no indication how to combine them into meaningful pairs. This situation is even worse for the strip of Val228, the $^{1}H,^{15}N$ correlation peak of which is in the crowded central region of the HSQC spectrum.

How far has the approach described here brought us, and, more importantly, what have we learned from the obtained assignments? First, we have succeeded in the assignment of a significant fraction of the residues occurring in unique patterns. Moreover, some more residues could be assigned on the basis of their unique presence or absence in the Tau352 spectrum or from the peptide mapping. Taken together with the assignments of those residues that are followed by a proline, we presently cover about 30% of the sequence without any ambiguity, and a further 10% with only two choices for the assignment (Figure 1). Overall, these assignments are distributed quite evenly over the sequence, although certain regions such as the quadruply occurring $Pro(Gly)_3$ quartet clearly have not yet been assigned unambiguously. This extent of the assignment should allow us to probe the interactions of Tau with such molecular partners as tubulin or heparin at the level of the individual residues.

As for the attempt to detect stretches of residual structure, one can question whether our approach, which is based on the validity of the random-coil chemical-shift values, can give any indication of this, as we should miss those assignments corresponding to residues that are in a locally well-defined structure and therefore escape from their respective product plane. The same objection can be made on the second approach, where we assume an identical environment of a given residue in a short unstructured peptide or full-length protein. For this latter approach, though, a discrepancy between the peptide and protein spectrum is easily detected. For the product plane strategy, we actually have one specific example where we did note a clear discrepancy: the isolated peak at (7.82, 127.2) ppm is a Leu, based on the spectrum of a Leu selective labelled sample (data not shown), but its carbon Cα and Cβ frequencies were found at 56.6 and 43.4 ppm. Because both values deviate substantially from the random-coil values of 55.1 and 42.4 ppm, the correlation peak does not show up in the Leu product plane, and we at first thought we had detected a "hotspot" of residual structure. Closer examination, however, revealed that the cross peak represents the C-terminal Leu441. As for the other resonances assigned, next to the Lys labelled sample we have produced two further selectively labelled samples with Leu and Val. The coincidence of the product planes with the HSQC spectra recorded on the respective selectively labelled samples, further demonstrates the validity of our initial assumption. An even more striking example is given by the (Ile)Ile278 doublet. Unique in the sequence and moreover present in the second repeat encoded by the alternatively spliced exon 10, this motif is part of the PHF_6* peptide that was previously shown to promote aggregation by the formation of β-structure.[29] Despite the compatibility of this peptide with an extended structure in the PHFs, the Cα and Cβ chemical-shift values of the second Ile278 are 60.9 and 38.7 ppm, which is less than 0.2 ppm from their random-coil value. The a posteriori justification of our conclusion that Tau lacks well-defined secondary-structure elements is that we have succeeded in the assignment of a substantial fraction of its NMR signals based on two methods that assume this lack of structure from the start. Subtle deviations from the random-coil carbon values, however, that do not exceed our 0.5 ppm limit do exist, and we are currently trying to interpret them in conjunction with the information content in the nitrogen chemical-shift values.[24,25] Furthermore, it should be noted that the resonances in the repeat regions are generally weaker in intensity than their counterparts in the N- or C-terminal region of the protein. A full relaxation analysis of this phenomenon is currently in progress, but it is already clear that those residues, despite their sampling of the whole conformational space that defines the "random coil" conformation, do sample this space on a slower time scale. Therefore, it will be interesting to see whether this factor of dynamics will equally contribute to the propensity of fibrilization, on top of the recently determined structural criteria that govern aggregation.[30]

In conclusion, we have demonstrated here that the basic assumption of the validity of the random coil chemical-shift values can be successfully exploited to assign, at least partially, this at first sight impossible spectrum. The introduction of product planes allows a graphical view of the different residues and can easily be applied to the preceding residues when one derives them from the carbonyl relayed triple-resonance spectra. The HSQC spectrum of the shortest Tau352 isoform is at least to first order a subspectrum of that of Tau441, although we cannot exclude that subtle conformational differences cause deviations beyond the border effects. Peptide mapping proved possible, and allowed us to ascertain that the same residues embedded in the full-length protein or in a short peptide sequence sample the in same conformational space. Importantly, our present level of coverage and the validity of the random-coil values confirm that, at the level of the individual amino acids, this important protein belongs to the

class of natively unfolded proteins. The results presented form the basis for future structural studies of disease-related mutations of Tau, and of its interactions with its diverse molecular partners.

Experimental Section

Expression and purification of the recombinant Tau protein: Expression and purification of the uniformly labelled samples was as previously described.[7] Residue specific labelling was obtained by adding ^{15}N-labeled Lysine, Leucine or Valine (100 mg) to M9 medium (1 L) supplemented with the other unlabelled amino acids (50 mg) before induction. Induction delays were kept to 2 h, in order to minimize isotopic scrambling.

Peptide synthesis and purification: Peptides were synthesized by using a classical Fmoc strategy on a Pioneer™ peptide continuous flow synthesizer with intermediate capping after each coupling step. Cleavage of peptide from the resin and side-chain-protecting groups were performed by using TFA mixtures that contained either 2.5% triisopropylsilane or 1% triisopropylsilane/2.5% ethanedithiol for peptide containing Cys or Met residues. Crude peptides were then purified by RP-HPLC on a C18 Nucleosil column by performing a linear gradient of acetonitrile. Fractions were checked both by RP-HPLC and MALDI-TOF mass spectrometry. NMR samples of peptides were prepared at a concentration of 4 mM in the same buffer that was used for recombinant Tau protein.

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. Spectral windows were 7800 Hz for proton and 2000 Hz for nitrogen, centred at 4.8 and 115.2 ppm, respectively, and sampled with 1024 and 47 complex points. All spectra were recorded with 8 scans per increment. The 30.7 ppm Cα window in the HNCA and HN(CO)Ca spectra was centred at 54.5 ppm and was sampled with 31 complex points. Similarly, the 20 ppm CO window was centred at 175.4 ppm and was sampled with 21 complex points. Finally, the 70.5 ppm combined Cβ and Cα window in the CBCANH and CBCAcoNH spectra was centred 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 2048 complex points in the proton dimension and 512 complex points in the indirect nitrogen and carbon dimensions. The final 3D spectra were treated as a three-dimensional matrix, or, alternatively, a collection of ^{1}H,^{15}N planes. The latter can be considered as two dimensional matrices $A^k[i,j]$, with i ranging from 1–2048 and j from 1–512, containing an index k corresponding to the carbon frequency of the plane. The summation plane $S[i,j]$ is defined by $S[i,j] = \Sigma\ A^k[i,j]$, where the summation is over the indices k that are within the specified carbon range. Similarly, the point-by-point multiplication plane $P[i,j]$ of two sum planes S^1 and S^2 is defined by $P[i,j] = S^1[i,j] \times S^2[i,j]$. Normalization of the intensity in the latter product planes was consistently done with respect to the Cα sum plane.

The homonuclear peptide spectra for peptide assignment and structural verification were recorded with DQFCOSY or Water flip-back NOESY and TOCSY sequences, with mixing times of 500 ms and 69 ms for the latter. ^{1}H,^{15}N HSQC spectra of the peptides at natural abundance were recorded with 256 scans per increment, and 256 points in the nitrogen dimension.

Acknowledgements

We thank E. Diesis and A. Menet for help with the peptide synthesis, Dr. L. Buée (Lille, France) for insightful discussions about Tau and Alzheimer's disease, and Dr. J. Christodoulou (Cambridge, UK) for careful reading of the manuscript. 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. A.S. is funded by a European Training and Mobility Grant (HPRN-CT-2002-00241).

Keywords: neuronal Tau protein · NMR spectroscopy · peptide mapping · protein folding · protein structures

[1] M. D. Weingarten, A. H. Lockwood, S.-Y. Wy, M. W. Kirschner, *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 1858–1862.
[2] M. Goedert, C. M. Wischik, R. A. Crowther, J. E. Walker, A. Klug, *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 4051–4055.
[3] A. Delacourte, A. Defossez, *J. Neurol. Sci.* **1986**, *76*, 173–186.
[4] S. G. Greenberg, P. Davies, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 5827–5831.
[5] P. V. Arriagada, J. H. Crowdon, E. T. Hedley-White, B. T. Hyman, *Neurology* **1992**, *42*, 631–639.
[6] D. W. Cleveland, S. Y. Hwo, M. W. Kirschner, *J. Mol. Biol.* **1977**, *116*, 227–247.
[7] O. Schweers, E. Schonbrunn-Hanebeck, A. Marx, E. Mandelkow, *J. Biol. Chem.* **1994**, *269*, 24290–24297.
[8] D. Neri, M. Billeter, G. Wider, K. Wüthrich, *Science* **1992**, *257*, 1559–1563.
[9] M. Hennig, W. Bermel, A. Spencer, C. M. Dobson, L. J. Smith, H. Schwalbe, *J. Mol. Biol.* **1999**, 288, 705–723.
[10] H. J. Dyson, P. Wright, *Adv. Protein Chem.* **2002**, 62, 311–340.
[11] G. Lippens, J.-M. Wieruszeski, A. Leroy, C. Smet, A. Sillen, L. Buée, I. Landrieu, *ChemBioChem* **2004**, *5*, 73–78.
[12] M. Ikura, L. E. Kay, A. Bax, *Biochemistry* **1990**, *29*, 4659–4667.
[13] D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges, B. D. Sykes, *J. Biomol. NMR*, **1995**, *5*, 67–81.
[14] L.-Y. Lian, D. A. Middleton, *Progr. Nucl. Magn. Reson. Spectrosc.* **2001**, *39*, 171–190.
[15] B. Brutscher, *J. Magn. Reson.* **2002**, *156*, 155–159.
[16] A. Permi, A. Annila, *Progress Nucl. Magn. Reson. Spectrosc.* **2004**, *44*, 97–137.
[17] L. Buee, T. Bussiere, V. Buee-Scherrer, A. Delacourte, P. R. Hof, *Brain Res. Rev.* **2000**, *33*, 95–130.
[18] M. von Bergen, P. Friedhoff, J. Biernat, J. Heberle, E.-M. Mandelkow, E. Mandelkow, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5129–5134.
[19] J. Dyson, P. Wright, *Curr. Opin. Struct. Biol.* **2002**, *12*, 54–60.
[20] R. Bussell Jr., D. Eliezer, *J. Biol. Chem.* **2001**, *276*, 45996–46003.
[21] K. Pervushin, R. Riek, G. Wider, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12366–12371.
[22] R. Riek, K. Pervushin, K. Wüthrich, *Trends Biochem. Sci.* **2000**, *25*, 462–468.
[23] K. Kato, C. Sautès-Fridman, W. Yamada, K. Kobayashi, S. Uchiyama, H. Kim, J. Enokizono, A. Galinha, Y. Kobayashi, W.-H. Fridman, Y. Arata, I. Shimada, *J. Mol. Biol.* **2000**, *295*, 213–224.
[24] Y. Wang, O. Jardetzky, *J. Am. Chem. Soc.* **2002**, *124*, 14075–14084.
[25] J. Meiler, *J. Biomol. NMR* **2003**, *26*, 25–37.
[26] C. Smet, A.-V. Sambo, J.-M. Wieruszeski, A. Leroy, I. Landrieu, L. Buée, G. Lippens, *Biochemistry* **2004**, *43*, 2032–2040.
[27] P. Styles, N. F. Soffe, C. A. Scott, D. A. Cragg, F. Row, D. J. White, P. C. J. White, *J. Magn. Reson.* **1984**, *60*, 397–404.
[28] P. Styles, N. Soffe, C. A. Scott, *J. Magn. Reson.* **1989**, *84*, 376–378.
[29] M. von Bergen, S. Baghorn, L. Li, A. Marx, J. Biernat, E.-M. Mandelkow, E. Mandelkow, *J. Biol. Chem.* **2001**, *276*, 48165–48174.
[30] F. Chiti, M. Stefani, N. Taddei, G. Ramponi, C. M. Dobson, *Nature* **2003**, *424*, 805–808.

Received: May 10, 2004

Early View Article
Published online on November 8, 2004

