SUPPORTING INFORMATION

FLEXITau: Quantifying Post-translational Modifications of Tau Protein *in vitro* and in Human Disease

Waltraud Mair[1,2], Jan Muntel[3,4], Katharina Tepper[5,6], Shaojun Tang[3,4], Jacek Biernat[5,6], William W. Seeley[7,8], Kenneth S Kosik[9], Eckhard Mandelkow[5,6], Hanno Steen[3,4], Judith A Steen[1,2]

[1]Boston Children's Hospital, F. M. Kirby Neurobiology Center, Boston, MA, [2]Harvard Medical School, Neurology, Boston, MA, [3]Boston Children's Hospital, Pathology, Boston, MA, [4]Harvard Medical School, Pathology, Boston, MA, [5]DZNE, German Center for Neurodegenerative Diseases, Bonn, Germany, [6]CAESAR, Center of Advanced European Studies and Research, Bonn, Germany, [7]Memory and Aging Center, Department of Neurology, University of California, San Francisco, CA, [8]Department of Pathology, University of California, San Francisco, CA, [9]University of California Santa Barbara, Neuroscience Research Institute, Santa Barbara, CA

Running title: FLEXITau - A quantitative assay for tau modifications

Note: This supplement contains Details to Experimental Procedures, 5 Supplementary Figures, and legends for 4 Supplementary Tables.

EXPERIMENTAL PROCEDURES

Cells and viruses

Sf9 cells were obtained from Invitrogen (San Diego, CA) and grown at 27 °C in monolayer culture Grace's medium (Life Technologies, Gaithersburg, MD) supplemented with 10% fetal bovine serum, 50 µg/ml Gentamycin, and 2.5 µg/ml Amphotericin. Sapphire™ baculovirus DNA was obtained from Orbigen/Biozol (Eching, Germany), and pVL1392 was from Invitrogen.

Baculovirus construction

The hTau40 cDNA, the longest tau isoform in human CNS (2N4R), was excised from the bacterial expression vector pNG2 [1] with XbaI and BamHI and inserted into the baculovirus transfer vector pVL1392. For the construction of tau containing baculovirus vectors, Sapphire™ baculovirus DNA was used for homologous recombination with pVLhtau40 plasmid in Sf9 cells.

Sf9-tau protein preparation and purification

Phosphorylated Sf9-tau ("P-tau") was purified as described before[2]. Briefly, Sf9-cells were infected with recombinant virus at a MOI of 1-5, typically in six T150 cell culture flasks containing 75% confluent Sf9 cells. Cells were incubated for three days at 27 °C and collected directly in lysis buffer (50 mM Tris HCl pH 7.4, 500 mM NaCl, 10% glycerol, 1% Nonidet-P40, 5 mM dithiothreitol (DTT), 10 mM ethylene glycol tetra-acetic acid (EGTA), 20 mM NaF, 1 mM orthovanadate, 5 µM microcystin, 10 µg/ml each of protease inhibitors leupeptin, aprotinin, and pepstatin). For the generation of "PP-tau" (higher phosphorylated Sf9 P-tau), Sf9 cells were treated for 1h with 0.2 µM okadaic acid (OA, a phosphatase inhibitor, Enzo-Lifescience) 1 hour prior to harvesting. Lysates were boiled in a water bath at 100 °C for 10 min and cell debris was removed by centrifugation for 15 min at 16,000 x g. The supernatant containing soluble tau protein was concentrated in Millipore Amicon Ultra-4 Centrifugal filter units (MWCO 3 kDa). The concentrated material was applied to a size exclusion column Superdex G200 (GE Healthcare) and eluted with PBS Buffer (pH 7.4; 1 mM DTT), collecting 1 ml fractions. A second purification step was performed, using anion exchange chromatography on a MonoQ HR 16/10 column (GE Healthcare). For this purpose the tau-containing fractions of the G200-column were pooled and dialyzed against buffer A (100 mM MES (pH 6.8), 2 mM DTT, 1 mM NaEGTA, 1 mM MgSO4, 0.1 mM PMSF), before loading onto the MonoQ column. Tau protein was eluted by a three step salt gradient (Buffer A supplemented with 1 M NaCl was used to create salt gradient steps of 0-0.2 M, 0.2-0.3 M and 0.3 -1 M NaCl). To generate dephosphorylated Sf9-tau ("deP-tau") 30 µg purified P-tau protein was incubated with 10 U of alkaline phosphatase (FastAP, Invitrogen) for ~16 hours at 37 °C. The enzyme was removed afterwards by precipitation (5 mM DTT, 0.5 M NaCl), followed by centrifugation and dialysis to PBS. Protein amounts were estimated by a bicinchoninic acid test (BCA, Sigma). Samples were additionally analyzed by SDS-PAGE to verify purity and protein degradation (supplementary Fig. S5).

Preparation of human Tau40 from *E. coli*

Expression and purification of human Tau40 from *E. coli* cells was carried out as described[28]. Note that human Tau40 is purified differently to Sf9-tau, as it does not carry the negative charges of phosphates. Human Tau40 was first purified by cation exchange chromatography (SP sepharose; GE healthcare), and then by size exclusion chromatography (G200).

SDS-PAGE and silver staining

Samples were boiled 5 min at 98 °C in 2x Laemmli buffer and separated by SDS-PAGE (4-12% Bis-Tris, NuPage, Invitrogen) at 120 V. Gels were stained with colloidal blue (Nuvex, Invitrogen). Silver staining was performed by fixing the gels in 30% ethanol / 10% acidic acid solution, cross-linking the proteins (in 0.5% glutarealdehyde) and staining in 0.1% $AgNO_3$ solution, followed by development with 2.5% Na_2CO_3 / 1% formaldehyde until the protein marker was visible.

Extraction of AD-tau

Human AD brain tissue was obtained from the Human Brain and Spinal Fluid Resource Center, VA West Los Angeles Healthcare Center, Los Angeles, and the Neurodegenerative Disease Brain Bank, University of California, San Francisco (supplementary Table S4). AD cases had advanced disease, meeting NIA-Reagan criteria for high likelihood AD[29]. Tissue blocks representing the angular gyrus (1-3 g) were dissected from frozen brain slabs and shipped overnight to Boston Children's Hospital on dry ice. For generating insoluble tau fractions, sarkosyl fractionation was performed. Briefly, while still frozen, 0.3 g sections were homogenized in 5 volumes 25 mM Tris-HCl buffer, pH 7.4, containing 150 mM NaCl, 10 mM ethylene diamine tetraacetic acid (EDTA), 10 mM EGTA, 1 mM DTT, 10 mM nicotinamide, 2 µM trichostatin A, phosphatase inhibitor cocktail (Sigma), protease inhibitor cocktail (Roche). Crude brain homogenates were then clarified by centrifugation at 11,000 x g for 30 min at 4 °C. Lysate was incubated with sarkosyl (1% final concentration) for 60 min at 4 °C and ultracentrifuged at 100,000 x g for 2 h at 4 °C. The supernatant was transferred to a new tube (sarkosyl soluble fraction). The pellet was air-dried, washed twice with 50 µl ddH2O and solubilized in Tris buffer containing 1% SDS, 10 mM nicotinamide, 2 µM trichostatin A, and phosphatase and protease inhibitor cocktail (0.3 µl buffer per mg wet weight of the starting material). Solubilized pellets were used as the sarkosyl insoluble tau fraction. All samples were stored at -80 °C until use.

Preparation of heavy tau standard

Full-length human 2N4R (GI:294862262) was subcloned into the previously generated pEU-E01-TEV-N1-AQUA vector[25]. After verification by DNA sequencing (Molecular Genetics Core Facility, Children's Hospital Boston), tau was *in vitro* transcribed and translated in a cell-free wheat germ expression (WGE) system according to the manufacturer's protocols (Cell Free Sciences, Wheat Germ Expression H Kit-NA). Expression was carried out in the presence of isotope labeled lysine K8 (13C6 15N2), arginine R10 (13C6 15N4) and aspartate D5 (13C4 15N1). The triple labeling strategy allowed us to minimize co-expressed light tau standard that could lead to a bias in quantification of endogenous tau. Heavy tau standard was batch-purified using Ni-Sepharose beads (Ni-Sepharose High Performance resin, GE Healthcare). Briefly, after a prewash in binding buffer (20 mM phosphate buffer, pH 7.5, 500 mM NaCl, 10 mM imidazole) beads were incubated with WGE (ratio 1:4) for 1h rotating head-over-head at 4 ˚C for binding. After removal of the unbound fraction, beads were washed once with 50 µl and 3 times with 500 µl wash buffer (20 mM phosphate buffer, pH 7.5, 500 mM NaCl, 10 mM imidazole). Elution of tau was carried out in three steps (50 µl binding buffer with 100/300/500 mM imidazole, respectively). Purification was verified by SDS-PAGE and western blot analysis (data not shown). Pooled eluates were stored at -20 ˚C in 50 µl aliquots until further use.

Sample preparation for MS

Heavy tau standard was dephosphorylated by incubation with lambda protein phosphatase (New England Biolabs) for 30 min at 30 °C at 300 rpm. Digestion was performed using 1 µg of Sf9-tau or 50 µg of AD sarkosyl insoluble pellet. For spikes, dephosphorylated standard was added prior to digest. Protein mixtures were reduced with 50 mM DTT (20 min, 56 °C), alkylated with 1% acrylamide (30 min, RT), diluted with 8 M urea and digested using FASP[3] (FASP Protein Digestion Kit, Expedeon) following the manufacturer's protocol. Briefly, samples were digested overnight 37 °C with 2 ng/µl trypsin (sequencing grade modified trypsin, Promega) in 50 mM ammonium bicarbonate (ABC) or 4 ng/µl LysC (endoproteinase LysC sequencing grade, Roche) in LysC buffer (0.1 M Tris, pH 9.2, 1 mM EDTA). Eluted peptides were acidified, desalted using C18 microspin tips (Nest Group) and dried under vacuum. Peptides were reconstituted in sample buffer (5% formic acid, 5% acetonitrile) containing 10 fmol/µl non-labeled FLEX-peptide (TENLYFQGDISR, synthesized by Sigma Life Science, quantified via amino acid analysis of Molecular Biology Core Facilities, Dana Farber Cancer Institute, Boston) and indexed retention time (iRT) peptides (Biognosys)[4].

LC-MS/MS measurements

To identify the most sensitive and selective transitions (pair of peptide and their fragment ion masses), we performed high-resolution liquid chromatography tandem MS (LC-MS/MS) of purified, digested tau standard in DDA mode and generated a collection of experimentally detected peptides and their fragment

ions (spectral library). Multiple measurements of up to 400 fmol of tau were performed using two different instruments platforms: First, a quadrupole Orbitrap tandem mass spectrometer (Q Exactive, Thermo Fisher Scientific) was used to maximize the number of peptide identifications. Samples were analyzed on a Q Exactive hyphenated with a micro-autosampler AS2 and a nanoflow HPLC pump (both Eksigent), using the trap-elute chip system (cHiPLC nanoflex, Eksigent). Peptides were first loaded onto the trap-chip (200 µm x 75 µm, ChromXP C18-CL 3 µm 120 A, Nano cHiPLC Eksigent) and then separated using an analytical column-chip (75 µm x 15 cm, ChromXP C18-CL 3 µm 120 A, Nano cHiPLC Eksigent) by a linear 30 min gradient from 95% buffer A (0.1% (v/v) formic acid in HPLC-H_2O) and 5% buffer B (0.2% (v/v) formic acid in acetonitrile) to 35% buffer B. A full mass spectrum with resolution of 70,000 (relative to an *m/z* of 200) was acquired in a mass range of 300-1500 m/z (AGC target 3 x 10^6, maximum injection time 20 ms). The 10 most intense ions were selected for fragmentation via higher-energy c-trap dissociation (HCD, resolution 17,500, AGC target 2 x 10^5, maximum injection time 250 ms, isolation window 1.6 m/z, normalized collision energy 27%). The dynamic exclusion time was set to 20 s and unassigned/singly charged ions were not selected.

In addition, the purified tau standard was analyzed on a Sciex Triple TOF 5600 to generate fragmentation spectra comparable to the employed SRM instrument (Sciex QTRAP 5500), using the same LC setup as described above. The Triple TOF was operated in data-dependent TOP30 mode with following settings: MS1 mass range 350-1300 Th with 175 ms accumulation time; MS2 mass range 100-1500 Th with 25 ms accumulation time and following MS2 selection criteria: UNIT resolution, intensity threshold 8 cts; charge states 2-5. To identify sites of modifications on tau, Sf9-tau digests were analyzed on the Q Exactive applying the settings described above, replacing the chip-system with an in-house packed C18 analytical column (Magic C18 particles, 3 µm, 200 Å, Michrom Bioresource). After initial measurements, an inclusion list containing all identified phosphorylated tau peptides was created using Skyline. For final measurements, following settings were used in order to increase peptide identification: AGC target 5 x 10^6, maximum injection time 120 ms, MS/MS resolution 35,000, AGC target 2 x 10^5, maximum injection time 200 ms, isolation window 2 m/z. The dynamic exclusion time was set to 4 s and unassigned and charge state 1 and >5 ions were rejected. The inclusion list was turned on allowing picking others if idle.

LC-MS/MS data processing

Q Exactive raw files were converted into mgf data format using ProteoWizard[5]. The spectra were centroided and filtered using ms2preproc to select the 6 most intense peaks in a 30 Th window[6]. MS/MS spectra from mgf or wiff files were assigned to peptides and corresponding proteins using ProteinPilot™ Software 4.5 Beta (Paragon Algorithm 4.5.0.0. 1575, Sciex). The following settings were applied: sample type 'SILAC (Lys+8, Arg+10, Asp+5)', instrument: 'Orbi MS (1-3ppm)', 'Orbi MS/MS' and 'TripleTOF 5600' respectively; 'Urea denaturation'; 'rapid' search mode. Spectra were searched against a custom database containing wheat germ proteins and the human 2N4R Tau sequence tagged with the FLEX peptide. For the mapping of Sf9-tau PTMs, raw files were converted and processed in ProteinPilot as described above except for following search parameters: 'thorough' search mode, 'phosphorylation emphasis', 'acetylation emphasis', 'ID focus on biological modifications' and using Homo Sapiens database (downloaded from uniprot.org on 11/01/2011). Note that ProteinPilot doesn't allow the user to pick mass tolerances and number of missed cleavages. All MS/MS spectra of identified post-translationally modified peptides were subjected to manual verification.

SRM assay development

To generate a spectral library from the generated LC-MS/MS datasets, xml files were extracted from ProteinPilot and loaded into Skyline[7] using cut off score of 0.5. A FASTA file containing the 2N4R tau protein sequence tagged with the FLEX-peptide was imported, using a wheat germ protein database as a background proteome. Filter settings for tryptic peptides (Trypsin/P KR|-) were as follows: a maximum of 2 missed cleavages, a peptide length between 5 and 40 amino acids, a maximum of 3 variable structural modifications (cysteine propionamidation, serine/threonine phosphorylation, methionine oxidation and asparagine/glutamine deamidation) and a maximum of 1 neutral loss. The spectra were used to confirm identities, extract the optimal fragment ions for SRM analysis and obtain retention times. An initial transition list for each tryptic and LysC samples were generated choosing 8 most intense product ions from the library spectrum, considering only y ions with charges 1 and 2 (from precursor ions with charges

2, 3 and 4) from ion 3 to last ion -1. The transition lists were validated and optimized after SRM measurements, as described below.

SRM measurements and data processing

The SRM assay using the transition lists developed above was tested using the tau standard. Measurements were performed on a triple quadrupole mass spectrometer (5500 QTRAP, Sciex) using the same LC trap-elute chip setup as described above. Initial measurements for optimization of transitions were done using a scheduled SRM mode, a retention time window of 7 min, a gradient of 30 min and a maximum of 250 transitions per method. Resulting SRM data were analyzed and manually validated in Skyline. Transition groups corresponding to the targeted peptides were evaluated based on specific parameters (in order of importance): co-elution of light and heavy peptides; rank correlation between the SRM relative intensities and the intensities obtained in the MS/MS spectra; and consistence among technical and biological replicates. Using these criteria, the transition lists were reduced from 8 to 4-5 most intense product ions per peptide.

To assess linearity of product ion signals and to determine detection limit of the assay, a dilution series of heavy tau standard was performed. Absolute quantification of the standard was carried out using SRM relative intensities between the heavy FLEX-peptide and its light counterpart FLEX-peptide, as described previously[8]. Samples containing different amounts of heavy tau spanning four orders of magnitude (from 0.8 – 800 fmol) were prepared and measured in triplicates using the optimized transition list.

Final FLEXITau measurements of mixed peptide samples were performed using the validated transition lists (supplementary Table S1), a retention time window of 5 min and a total scan time of 1.2 s, which ensured a dwell time over 20 ms per transition. To achieve the desired concentration range (as discussed in 'Results and Discussion') we initially collected data from a 1:10 dilution of the samples and adjusted the injection amount appropriately. Blank runs between SRM measurements ensured minimal sample carry over, and three replicate injections were measured per sample (MS injection on separate days).

SRM data were analyzed and manually curated in Skyline. Peptide transitions were re-evaluated for variability, similarity between y-ion ratios, elution times, and interfering signals by manual analysis. For quantification, the 3 highest intense transitions were used. For each individual run, ratios of L/H peak intensities were normalized using the average of the 3 peptides with highest ratio. The same normalization factor was used for the LysC sample as calculated for the trypsin-digested sample. To assess quantification precision of technical and biological replicates, averages of normalized L/H ratio as well as light peak area of each peptide were calculated from the triplicate measurements for each biological replicate, and subsequently the average and %CV from the three biological replicates was calculated for each species (Fig. 3B and 3C). For the calculation of the modification extent of each peptide, first the average of technical replicates (normalized L/H ratio) was taken, followed by normalization of each sample by the average of all control samples. The modification extent for each biological replicate was then calculated by subtracting this value from 1. Negative values were transformed to zero and averages of the three biological replicates were calculated. FLEXITau data was expressed as mean +/- STDEV of normalized L/H ratio of biological replicates and analyzed by the Student's t test (two-sided) between two groups (Fig. 3D). Statistical significance was accepted at the $p<0.05$ level.

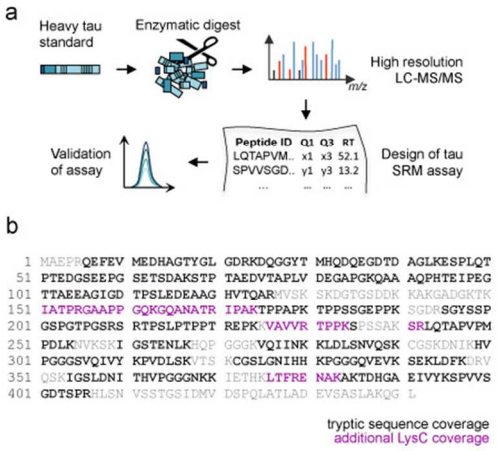
Calculation of site-occupancy

Individual site occupancies of the mapped phosphorylation sites were calculated for each biological replicate separately using the equations listed in Fig. 4A. Then the average for each tau species was calculated (Fig. 4B). Next we used a recursive approach to calculate the polynomial probability distribution of observing a specific number of phosphorylations per tau molecule (supplementary Fig. S4). The input data consisted of site occupancies X_{ij} for a total number of N = 22 sites for each biological replicate (three replicates each). The probability p_j of a site j (j = 1,2,...,N) being modified from a total of $r=3$ replicates (i=1,2,3) was calculated as

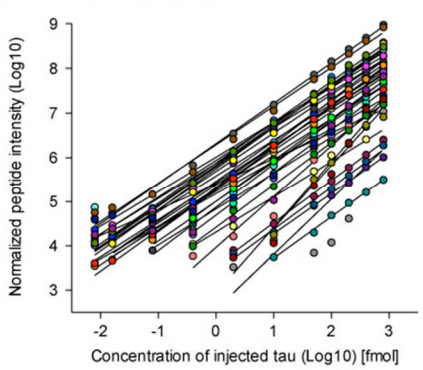
$$p_j = \frac{1}{r}\sum_{i=1}^{r} X_{ij}$$

Given D={p_1,p_2,...,p_N} the entire list of probability values for all *N* sites, P(1|p_m) = p_m the probability of observing site m in a particular tau species, and P(0|p_m) = 1 - p_m the probability of not observing site *m,* the probability of seeing *k* sites to be modified in that particular species out of all *N* sites was calculated as $P(k \mid D) = p_1(k-1 \mid D-p_1) + (1-p_1) * P(k \mid D-p_1)$ with

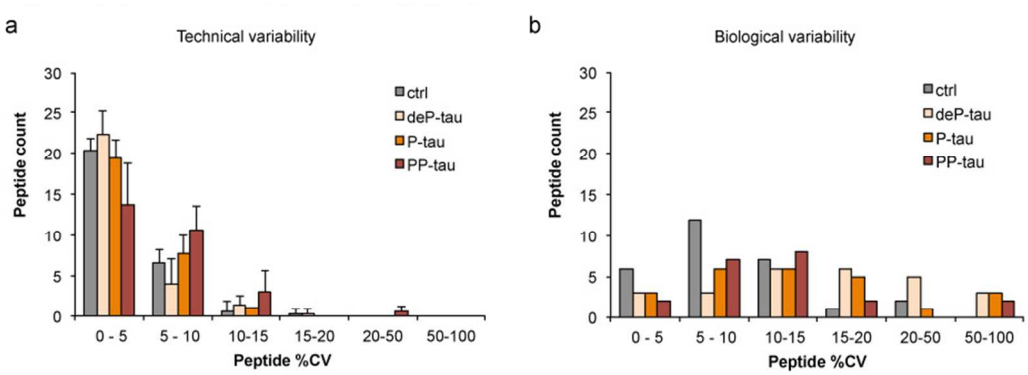
$$P(0 \mid D) = \prod_{i=1}^{17}(1-p_i)$$



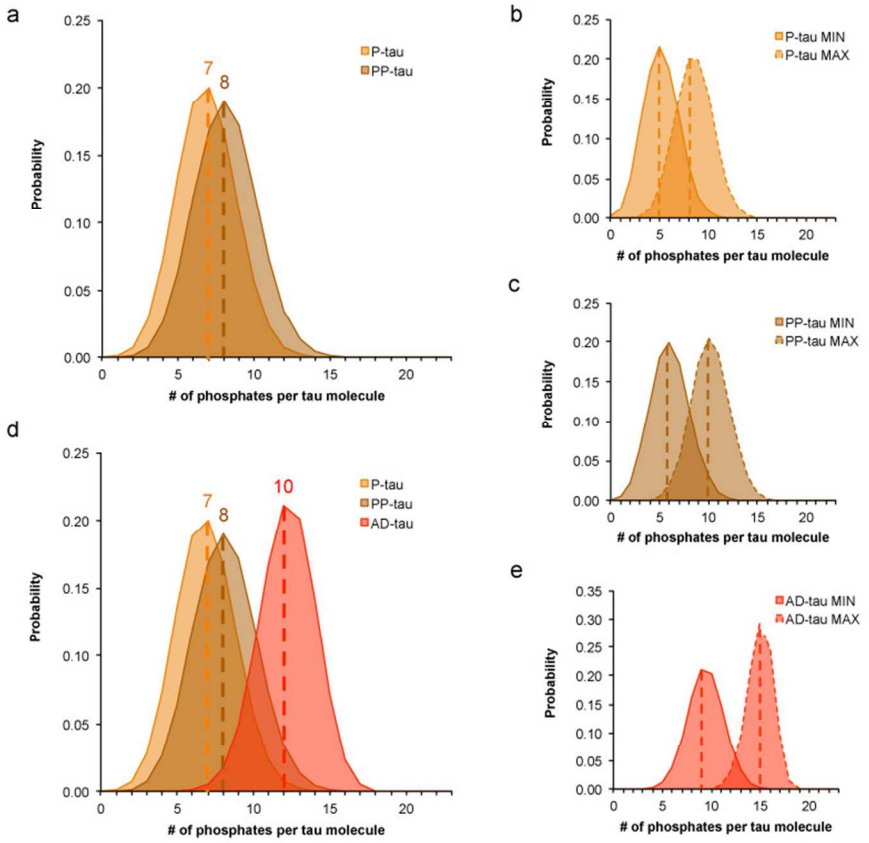
Supplementary Figure S1 FLEXITau SRM assay development and sequence coverage. (**A**) For the development of the tau SRM assay, an enzymatic digest of His-tag purified heavy tau is analyzed by high resolution LC-MS/MS and used to generate a transition list. (**B**) Sequence coverage of quantifiable tau peptides by SRM is shown in bold black for trypsin, additional coverage using LysC is shown in purple (tau sequence of human 2N4R).



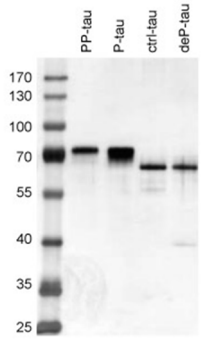
Supplementary Figure S2 Detection limit and linearity of the peptide quantification. A dilution series of expressed heavy tau standard was performed across 5 orders of magnitude. Expressed heavy tau standard was digested by trypsin and LysC and heavy peptides were measured using the developed SRM assay. iRT standard peptides (Biognosys) were used to normalize for run-to-run variability. Shown is normalized peptide abundance of all quantified tau peptides plotted against the amount of tau injected (mean value of triplicate injections).



Supplementary Figure S3 Distribution of peptide %CV. To assess variability of peptide quantification, three independent preparations of each tau species (biological replicates) were subjected to the FLEXITau workflow and analyzed in triplicates (technical replicates). %CV distributions of normalized L/H ratio of technical (**A**) and biological (**B**) replicates are shown. For the technical replicates, the average across the 3 biological batches is shown for each bin (mean +/- stdev).



Supplementary Figure S4 Frequency distribution of number of phosphates per tau molecule. Joined occupancies (more than 1 site per quantified occupancy) were assumed to be either exclusively occupied (MIN) or jointly occupied (MAX), and average distribution was determined by measuring the mean of MIN and MAX values for joined sites. Maximum likelihood estimate is indicated by vertical line. For details of calculation, see 'Methods'. (**A**) average for Sf9 P-tau and PP-tau, (**B**) and (**C**) MIN/MAX distributions for P-tau and PP-tau, respectively. (**D**) average for Sf9 P-tau and PP-tau, and AD-tau, (**E**) MIN/MAX for AD-tau.



Supplementary Figure S5 Purity of Sf9-tau preparation. To assess purity and degradation of prepared tau species, samples were analyzed by SDS-PAGE (200 ng protein/lane), followed by silver staining. Both phosphorylated species (P-tau and PP-tau) show a shift towards higher molecular weight due to the phosphates. Dephosphorylated tau (deP-tau) runs at a similar molecular weight as the recombinant non-phosphorylated *E. coli* tau. Shown is one representative experimental batch.

SUPPLEMENTARY TABLES

Supplementary Table S1. Transition list of FLEXITau. (A) Trypsin, (B) LysC. RT, retention time; DP, declustering potential; CE, collision energy

Supplementary Table S2. Summary of phosphorylated tau peptide species and phosphorylation sites detected in deP-tau, P-tau and PP-tau. Two phosphorylation sites detected simultaneously on a peptide are indicated by '+', e.g. 212 + 214. Ambiguous assignment of sites is indicated by '/', e.g. 68/69. AA, amino acid; z, charge; MW, Molecular Weight (monoisotopic mass of the ion fragmented in this analysis)

Supplementary Table S3. Summary of site occupancies for the quantified tau phosphorylation sites, in %. Shown are values for each biological replicate, the average and standard deviation (stdev).

Supplementary Table S4. Characteristics of patient material. UCSF, Neurodegenerative Disease Brain Bank, University of California, San Francisco; HBSFRC, Human Brain and Spinal Fluid Resource Center, West Los Angeles; PMI, postmortem interval; BA, Brodmann Area

SUPPLEMENTARY REFERENCES

1. Biernat, J., Gustke, N., Drewes, G., Mandelkow, E.M. & Mandelkow, E. Phosphorylation of Ser262 strongly reduces binding of tau to microtubules: distinction between PHF-like immunoreactivity and microtubule binding. *Neuron* **11**, 153-163 (1993).
2. Tepper, K.*, et al.* Oligomer formation of tau protein hyperphosphorylated in cells. *J. Biol. Chem.* **289**, 34389-34407 (2014).
3. Wisniewski, J.R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359-362 (2009).
4. Escher, C.*, et al.* Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **12**, 1111-1121 (2012).
5. Kessner, D., Chambers, M., Burke, R., Agus, D. & Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **24**, 2534-2536 (2008).
6. Renard, B.Y.*, et al.* When less can yield more - Computational preprocessing of MS/MS spectra for peptide identification. *Proteomics* **9**, 4978-4984 (2009).
7. MacLean, B.*, et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-968 (2010).
8. Singh, S., Springer, M., Steen, J., Kirschner, M.W. & Steen, H. FLEXIQuant: a novel tool for the absolute quantification of proteins, and the simultaneous identification and quantification of potentially modified peptides. *J. Proteome Res.* **8**, 2201-2210 (2009).