Prediction of Nucleating Sequences from Amyloidogenic Propensities of Tau-Related Peptides[†]

Federico A. Rojas Quijano,[§] Dana Morrow,[§] Barry M. Wise,[‡] Francesco L. Brancia,[||] and Warren J. Goux*[,§]

Department of Chemistry, The University of Texas at Dallas, P.O. Box 830688 Richardson, Texas 75083-0688, Shimadzu Research Laboratory (Europe), Wharfside Trafford Wharf Road, Manchester, M17 1GP, U.K., and Eigenvector Research, Inc., P.O. Box 561, Manson, WA 98831.

Received November 1, 2005; Revised Manuscript Received February 8, 2006

ABSTRACT: Physical properties, including amyloid morphology, FTIR and CD spectra, enhancement of Congo red absorbance, polymerization rate, critical monomer concentration, free energy of stabilization, hydrophobicity, and the partition coefficient between soluble and amyloid states, were measured for the tau-related peptide Ac-VQIVYK amide (AcPHF6) and its single site mutants Ac-VQIVXK amide (X ≠ Cys). Transmission electron microscopy showed that 15 out of the 19 peptides formed amyloid in buffer, with morphologies ranging from straight and twisted filaments to sheets and rolled sheets. Using principal component analysis (PCA), measured properties were treated in a comprehensive manner, and scores along the most significant principal components were used to define individual amino acid amyloidogenic propensities. Quantitative structure−activity modeling (QSAM) showed that residues with greater size and hydrophobicity made the largest contributions to the propensity of peptides to form amyloid. Using individual amino acid propensities, sequences within tau with high amyloid-forming potential were estimated and found to include ^{226}VAVVR^{230} in the proline-rich region, ^{275}VQIINK^{280} (PHF6*) and ^{306}VQIVYK^{311} (PHF6) within the microtubule binding region, and ^{392}IVYK^{395} in the C-tail region of the protein. The results suggest that regions outside the microtubule-binding region may play important roles in tau aggregation kinetics or paired helical filament structure.

The deposition of insoluble protein deposits are characteristic of over 100 human diseases, including familial amyloidosis, type II diabetes, and numerous neurodegenerative diseases, including transmissible spongiform encephalopathies, Alzheimer's disease (AD[1]) and Parkinson's disease *(1*−*6)*. These deposits, named amyloid because they, like the polysaccharide amylose, bind iodine, display a yellow-green birefringence when stained with Congo red (CR) and a characteristic fluorescence when stained with thioflavin dyes *(2, 7*−*10)*. In addition, amyloid is frequently characterized by its cross-β X-ray diffraction pattern and by its infrared (IR) or circular dichroism (CD) spectra, all of which suggest the presence of a high relative abundance of β-structure *(10*−*16)*. Although it was long believed that all amyloid existed as 3−15 nm wide straight unbranched fibers, more recent evidence, from our laboratory and others, suggests that amyloid may exist as twisted filaments, flat and rolled sheets, or spherical or annular particles *(16*−*23)*. Hence, diverse morphodologies exist for amyloid, and a wide variety of methodologies have been used in the past to characterize it.

Destabilization of the native conformation brought about by mutations or denaturing conditions of the media has been shown to induce globular proteins, some not normally associated with amyloid disease pathology, to form amyloid *(24*−*29)*. This finding, and the lack of sequence homology among proteins associated with amyloid, has led some to suggest that the ability to form amyloid is a generic trait of all proteins, independent of their sequence *(24, 30)*. However, it has been shown that certain defined sequences or "hot spots" nucleate amyloid formation in globular and nonglobular proteins *(3, 31*−*43)*. We refer to such sequences as "nucleating sequences" because they generate centers that can help seed amyloid and regulate its rate of formation *(26, 27, 40, 43)*. Studies of short amyloidogenic peptides homologous to these sequences suggest that amphiphilic motifs or π-stacking interactions of aromatic side chains play a crucial role in amyloid formation *(3, 16, 32, 37*−*54)*.

Tau is a microtubule-associated protein that regulates microtubule stability, neurite growth, and other microtubule-dependent functions. The human protein exists as six

[†] W.J.G. acknowledges The University of Texas at Dallas for their financial support.

* To whom correspondence should be addressed. Tel: 972-883-2660. Fax: 972-883-2925. E-mail: wgoux@utdallas.edu.

[‡] Eigenvector Research, Inc.

[§] The University of Texas at Dallas.

[||] Shimadzu Research Laboratory (Europe).

[1] Abbreviations: PHFs, paired helical filaments; NFTs, neurofibrillary tangles; FTDP-17, frontotemporal dementias with Parkinsonism linked to chromosome 17; AD, Alzheimer's disease; MTBR, microtubule binding region of tau protein; R2, R3, or R4, the second, third, or fourth repeat motif of the tau microtubule binding region; 3R, 4R, tau protein or peptides containing three or four repeat regions; HFIP, hexafluoro2-propanol; MOPS, 3-[*N*-morpholino]propanesulfonic acid; ThS, thioflavin S; TFA, trifluoroacetic acid; Ac, acetyl group; CR, Congo red; RTs, retention times; TEM, transmission electron microscopy; MALDI-TOF MS, matrix assisted laser desorption/ionization mass spectrometry; ESI, electrospray ionization; PCA, principal component analysis; PLS, partial least squares; QSAM, quantitative structure−activity modeling; LVs, latent variables; PHF6, ^{306}VQIVYK^{311}; PHF6*, ^{275}VQIINK^{280}.

10.1021/bi052226q CCC: $33.50
Published on Web 03/16/2006

isoforms, consisting of an acidic *N*-terminal with up to two 29-amino acid inserts, a proline-rich region, and a basic microtubule-binding region (MTBR) containing either three or four tandem 31 (or 32) amino acid pseudo-repeats (R1−R4) *(55−59)*. In individuals with AD, phosphorylation of the protein leads to its dissociation from microtubules and results in its aggregation into paired helical filaments (PHFs), which make up neurofibrillary tangles (NFTs), a hallmark of AD pathology *(60, 61)*. Although recombinant tau and tau isolated from microtubule preparations are largely unstructured, aggregated forms of the protein display physical characteristics of amyloid, including a cross-β X-ray diffraction pattern *(62−64)*. Nucleating sequences ^{275}VQIINK^{280}(PHF6*) in R2 and ^{306}VQIVYK^{311}(PHF6) and ^{314}DLSKVTS^{320} in R3 have been identified, and it has been shown that one or more of these sequences aid in filament formation *(35, 56, 65)*.

The goal of the present study was to formulate a model in which the chemical properties of amino acid residues substituted at a single site in the PHF6 structure (VQIVXK, where X is the substituted residue) could be correlated to the propensity of each peptide mutant to form amyloid. Principal component analysis (PCA) was used to formulate a characteristic response for each peptide by treating a set of measured parameters in a comprehensive manner. Quantitative structure−activity modeling (QSAM), an approach similar to that used to correlate structural parameters to biological activity *(66−69)*, showed that the size and hydrophobicity of the residue side chains made the greatest contribution to propensity. Individual amino acid propensities were extrapolated from the model and used to estimate the amyloidogenic propensities of sequences within the tau protein capable of amyloid formation. The model predicted that nucleating sequences occur throughout the protein but that sequences having the greatest amyloid-forming potential occur in the proline-rich region preceding the MTBR, within the MTBR, and in the *C*-terminal region.

MATERIALS AND METHODS

Peptide Synthesis. *N*-acetyl peptide amides with the sequence Ac-VQIVXK-NH_2, where X is one of 19 naturally occurring L-amino acids (omitting Cys to prevent complications due to disulfide bond formation), were prepared by solid-phase peptide synthetic methods according to previously published procedures and purified by reverse-phase HPLC using a water−acetonitrile gradient *(16, 70)*. Purity was checked using ESI-MS or MALDI-TOF MS. Peptide mixtures were prepared in a similar fashion, replacing X with a mixture of amino acids that yielded equimolar concentrations of peptides *(71)*. A total of 18 amino acids were used in the mixture (leucine is isobaric to isoleucine and, therefore, was left out). Desalting of the mixture was carried out in a C18 Strata cartridge (Phenomenex, Torrance, CA) equilibrated with water. A 10 mg sample in 150 μL of HFIP was loaded onto the cartridge, and the cartridge was rinsed with water containing 0.1% v/v TFA and the sample eluted with 10 mL of 50% acetonitrile/50% water containing 0.1% v/v TFA.

Transmission Electron Microscopy (TEM). Lyophilized peptides (1 mg) dissolved in 1 mL of 5 mM MOPS containing 0.01% NaN_3 at pH 7.2 (buffer A) or in buffer A containing 0.15 M NaCl were incubated at room temperature for at least a week. Samples were loaded onto carbon-coated Formvar copper grids (200 mesh) and stained with 2% w/w uranyl acetate. A JEOL 1200 EX scope interfaced to a digital camera was used to visualize samples.

Fourier Transforms Infrared (FTIR) Spectroscopy. FTIR spectra of peptide samples (10 mg/mL in D_2O incubated for 2 days) were acquired as described previously *(16)*. Spectra, displayed as 100-*T*%, were smoothed and baseline corrected and then fit by Gaussian functions using nonlinear deconvolution software (Peakfit, Systat, Richmond, CA). The ratio of the area of amide I bands between 1600 and 1640 cm^{-1} to the total area between 1600 and 1700 cm^{-1} was used as an indicator of the fraction of β-sheets present *(72−75)*.

Far-UV CD Spectroscopy. Peptides were dissolved in buffer A (∼1 mg/mL), and the concentrations were determined by absorbance using published molar extinction coefficients *(16, 76, 77)*. Stock solutions were diluted 1:4 with buffer and left to age at room temperature for at least 4 days. Far-UV CD spectra were acquired as previously described *(16)*, using a 0.2-mm-path cylindrical cell (Starna, 31-Q-0.2) in an AVIV 202 CD spectrophotometer. The CD results used in PCA modeling were expressed as the ratio of mean residue ellipicity at 215 nm to that at 192 nm ($[\Theta]_{Mr}$(215 nm)/$[\Theta]_{Mr}$(192 nm)).

Congo Red (CR) Binding. Samples (∼1 mg/mL) in 20 mM MOPS buffer containing 0.01% NaN_3 at pH 7.2 (buffer B) were left to incubate at room temperature for at least one week. The samples containing 5−60 μM peptide and 5 μM CR were prepared by diluting the stock solution. The absorbance at 540 nm was measured as a function of peptide concentration, and the 540 nm absorbance of a 35 μM peptide sample was used as an indicator of amyloid formation *(24, 32, 47, 63, 78, 79)*.

Kinetic Analysis. Polymerization kinetics was measured by monitoring the increase in thioflavin S (ThS) fluorescence at 490 nm (440 nm excitation) in a manner similar to that used to follow the aggregation of other amyloid-forming peptides and proteins *(10, 16, 80−82)*. For each peptide, a 1 mg/mL sample stock solution was filtered through a Millipore Ultrafree-MC 100,000 NMWL filter unit (Billerica, MA) by centrifuging for 10 min at 14 500 rpm, using an Eppendorf MiniSpin plus benchtop centrifuge (Westbury, NY). Immediately after centrifugation, three 20 μL aliquots of this solution were pipetted into wells of a 96-well clear-bottom ELISA plate containing 160 μL of buffer B, 20 μL of 1 mM ThS (final concentration of 100 μM) and a 1/8 inch PTFE grade 2 polished mixing bead (Orange Products, Allentown, PA). Fluorescence data were collected in kinetic mode, using a Molecular Devices SPECTRA Max Gemini XPS spectrofluorimeter (Sunnyvale, CA), with measurements taken at either 30 s or 10 min intervals, mixing between readings. A blank containing 180 μL of buffer and 20 μL of 1 mM ThS was recorded in triplicate. Kinetics data, corrected for the blank, were fit using a nonlinear least-squares algorithm to a Gompertz growth curve *(83)* (eq 1)

$$y = ae^{-e^{\left(\frac{(t-t_i)}{b}\right)}} \quad (1)$$

where y is defined as the fluorescence signal at time t, t_i corresponds to the inflection point of the curve, that is, the

time of maximum growth rate, a is the maximum fluorescence observed for a given sample, and $b = 1/k$ where k is the rate constant of aggregation, in units of s^{-1}. Lag times were calculated as $t_i - b$. Standard errors in fitted parameters were estimated assuming a 10% error in the sum of the squares.

Critical Monomer Concentration (CC). Experiments were carried out in the same fashion as the kinetic analysis, with the following modifications. Samples of 25−100 μM peptide in buffer B containing 100 μM ThS were prepared in triplicate and allowed to reach equilibrium. The fluorescence was plotted as a function of peptide concentration and then fitted using a linear regression. The x intercept of the curve was taken as the CC required to initiate polymerization.

Denaturation Studies. A preincubated (>2 days) peptide stock solution (20 μL of 1 mg/mL) was added to a series of solutions containing 100 μM ThS in 100−160 μL of buffer B. The solutions were incubated for 1−3 days at room temperature in a covered 96-well microplate. HFIP was added (0−60 μL), and the plate was incubated at room temperature for 2−4 h prior to fluorescence measurements. Following the subtraction of a blank containing only ThS, the fraction of the monomeric (unfolded) peptide (f_u) was calculated using published procedures *(84)*. The $\Delta G°$ of unfolding (stabilization) was calculated according to

$$\Delta G^{\circ}_{unf}(\text{cal/mol}) = -RT \ln (K_{eq}) \qquad (2)$$

where $K_{eq} = f_u/(1 - f_u)$. Curves were then fitted to a linear regression, where their y intercepts were taken as the free energies required to disaggregate the amyloid without the addition of any denaturant ($\Delta G^{\circ}_{unf}(H_2O)$).

HPLC Retention Times. The retention times for all peptides (~1 mg/mL in 50% acetonitrile/50% water (v/v) or 50% HFIP/50% water (v/v)) were determined on a 250 × 4.6 mm C18 column (Alltech Adsorbosil, Nicholasville, KY) using a linear gradient (0−100%) of water and acetonitrile (each containing 0.1% TFA) over 30 min at a flow rate 1 mL/min.

Synthesis of O-Methylisourea (OMIU). The preparation of OMIU was carried out according to published protocols *(85)* using urea (1 g; Fluka, Buchs SG, Switzerland) or ^{15}N-urea (1 g; Cambridge Isotope Laboratories, Andover, MA). The identity of the product was confirmed by ^{1}H NMR (270 MHz); ppm: 2.35 (s, methyl group in the aromatic ring), 3.94 (s, methoxy group), 7.5 (dd, para aromatic ring). The tosylate form of the compound was converted to its chloride form by passing the material down a DOWEX-1 ion exchange column (2.5 (ID) × 8 cm at 4 mL/min with DI water).

Guanidination Modification of the Peptides. A sample of the peptide mixture (5.3 mg) was suspended in 1.5 mL of buffer B, and the solution was left to incubate at room temperature for 4−5 days to allow for amyloid formation. The sample was then transferred to an 11 × 34 mm polyallomer centrifuge tube (Beckman, Palo Alto, CA) and centrifuged at 55 000 rpm for 30 min using a TL-100 Tabletop Ultracentrifuge (Beckman) equipped with a TLS-55 swinging bucket rotor. (The volume of the sample had to be adjusted to 2.2 mL with buffer.) The supernatant and pellet fractions were modified with OMIU using a slight modification of published procedures *(86)*. Lyophilized samples were dissolved in 300 μL of 50% HFIP/50% water and stirred for 2 h to allow for the formation of the monomeric species in solution. Ten milligrams of ^{14}N-OMIU was then added to the supernatant fraction, and 10 mg of ^{15}N-OMIU was added to the pellet fraction. To both the supernatant and the pellet fractions was added 50 μL of diisopropylethylamine (DIEA), and the solutions were stirred overnight. The HFIP and DIEA were then evaporated using a nitrogen stream. Finally, the fractions were recombined and lyophilized prior to MS analysis. The peaks for each peptide in the LC-MALDI spectra were integrated, and the ratio of the area of the peak for the peptide containing ^{15}N-homoarginine to the area for the peptide containing ^{14}N-homoarginine was used in the calculation of the partition coefficient.

LC-MALDI Analysis of Peptide Mixtures. Components of differentially labeled peptide mixtures were prepared using 0.05% (v/v) TFA in 5% acetonitrile and separated by reverse-phase capillary HPLC using a C18 Pepmap 100 column (3 μm), as previously described *(86)*. The fractions were eluted with a flow rate of 5 μL/min using a linear gradient (5−80%) of water and acetonitrile (each containing 0.05% v/v TFA) delivered to an online AccuSpot LC-MALDI spotting robot system (Shimadzu Corp., Kyoto, Japan) and mixed with a saturated solution of a cyano-4-hydroxycinnamic acid prepared in 50% v/v acetonitrile acidified with 0.1% v/v TFA. MALDI-TOF mass spectra were acquired on an AXIMA CFR instrument (Shimadzu Biotech, Manchester, U.K.), as described *(86)*.

Principal Component Analysis (PCA). For PCA, the data were arranged in a 19 × 10 matrix, where rows represent each of the peptides studied, and the columns are the measured variables, including the CD ($[\Theta]_{Mr}$(215 nm)/$[\Theta]_{Mr}$(198 nm)), the absorbance of Congo red at 540 nm and 35 μM peptide (A540), the FTIR intensity ratio A(1600−1640 cm^{-1})/A(1600−1700 cm^{-1}), the measured rate constant for amyloid aggregation (k), $\Delta G^{\circ}_{unf}(H_2O)$ (DelG), the critical monomer concentration (*CC*), the log of the lag time (*LT*), the HPLC retention time (*RT*), coded results for the TEM (*TEM*), and the partition coefficient of the peptide extrapolated from the LC-MALDI data (*LCMS*) (eq 3).

$$\begin{array}{c|cccccccccc} \underline{AA} & \underline{CD} & \underline{A540} & \underline{IR} & \underline{k} & \underline{DelG} & \underline{CC} & \underline{LT} & \underline{RT} & \underline{TEM} & \underline{LCMS} \\ A & 0.21 & 7.48 & 3.98 & 0.62 & 2984 & 4.58 & 3.50 & 14.6 & 1 & 0.0 \\ D & 0.26 & 8.25 & 8.50 & ND & ND & ND & ND & 13.9 & 1 & 0.0 \\ E & 0.25 & 7.73 & 0.79 & ND & ND & ND & ND & 14.3 & 1 & 0.56 \\ F & -0.27 & 12.82 & 1.98 & 5.65 & 1820 & 2.90 & 2.00 & 17.7 & 1 & 3.70 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & 0 & 0.0 \end{array} \qquad (3)$$

Here AA represents the one letter amino acid descriptor code X in the sequence Ac-QVIVXK-NH_2, and ND indicates that no data could be obtained because the peptide did not form the amyloid as detected by that particular measurement. Because ND is not really missing data but is real data that could not be measured in the experimental time period, it was assigned values that were physically reasonable yet set apart from values observed for other mutants. For example, the ND value for rate constant k was assigned a value of 0 because the rate of aggregation was so slow that it could not be observed in a reasonable time frame, whereas the ND value for $\Delta G^{\circ}_{unf}(H_2O)$ was assigned a value of −2000

cal/mol because a negative $\Delta G°$ implies that equilibrium favors a nonaggregated state. Other assigned values for ND were 100 μM for CC and 100 000 s for LT. TEM results were coded as 1, indicating the presence of amyloid, or as 0, indicating that no amyloid of any morphology was observed. Data were mean-centered and autoscaled to unit variance and analyzed by PCA *(87)*. Score plots were constructed from the two most significant components, and the projection onto PC1 was taken as the amyloidogenic propensity of amino acid residue X.

Amyloidogenic Propensities of Tau Sequences. The sequence of full-length human tau protein was searched for motifs that conformed to a polar residue followed by a series of hydrophobic residues. The amyloidogenic propensity of the sequence was defined as the sum of absolute values of individual residue PC1 scores.

QSAM. In QSAM, a set of structural variables, **X**, are related to measured properties, **Y**, according to

$$\mathbf{Y} = \mathbf{XB} + \mathbf{F} \tag{4}$$

where matrix **B** is composed of the partial least-square (PLS) regression coefficients, and **F** is the error matrix following the projection of **X** and **Y** onto latent variables (LVs) *(66−69)*. Structural descriptors for the mutant amino acids, z_1-z_5, were used to model **X** and were taken from the data of Sandberg *(68)*. Scores along PC1 calculated from PCA were used for **Y**, the representation of measured observables. The correlation coefficient (R^2) between the predicted and actual values of y_i was used as a measure of fit of **X** to **Y** in the set of samples upon which the model was built (calibration set). Full leave-one-out cross validation was used to determine the optimum number of LVs. The sum of the residual squares for each block of missing data is defined as the predictive residual sum of the squares (PRESS) and is related to Q^2, the cross validation equivalent of R^2 *(69)*.

PCA and PLS modeling were carried out using MATLAB (Mathworks, Inc., Natick, MA) along with PLS_Toolbox (EigenVector Research, Inc., Manson, WA).

RESULTS

Characterization by TEM. AcPHF6 (Ac-VQIVYK amide) and single-site mutants (Ac-VQIVXK amide, where X ≠ Cys) were studied by TEM. The peptides were incubated at least 7 days in 5 mM MOPS at pH 7.2 (0.01% NaN_3) buffer alone, or in buffer containing 0.15 M NaCl, prior to microscopy. Micrographs for some of the peptides are shown in Figure 1 and morphologies observed under both sets of incubation conditions are summarized in Table 1. In the absence of NaCl, amyloid formation was observed with all peptides except for the His, Gly, Lys, and Arg mutants. Abundant straight filaments were observed for AcPHF6 and the Asp, Phe, Met, Val, and Trp mutants, whereas less abundant filaments were observed for the Ala, Pro, Ser, and Thr mutants. Rolled sheets were observed for the Glu, Ile, and Leu mutants. Twisted filaments were observed for the Asn, Gln, and Thr peptides, which were similar in width but had longer half-periodicities than PHFs formed from tau protein ($d \approx 20$ nm, $p \approx 80$ nm) *(56, 60)*. In the presence of NaCl, all peptides except for AcPHF6 and the Thr mutant displayed morphologies that differed from those observed for the same peptides in its absence. For example, twisted

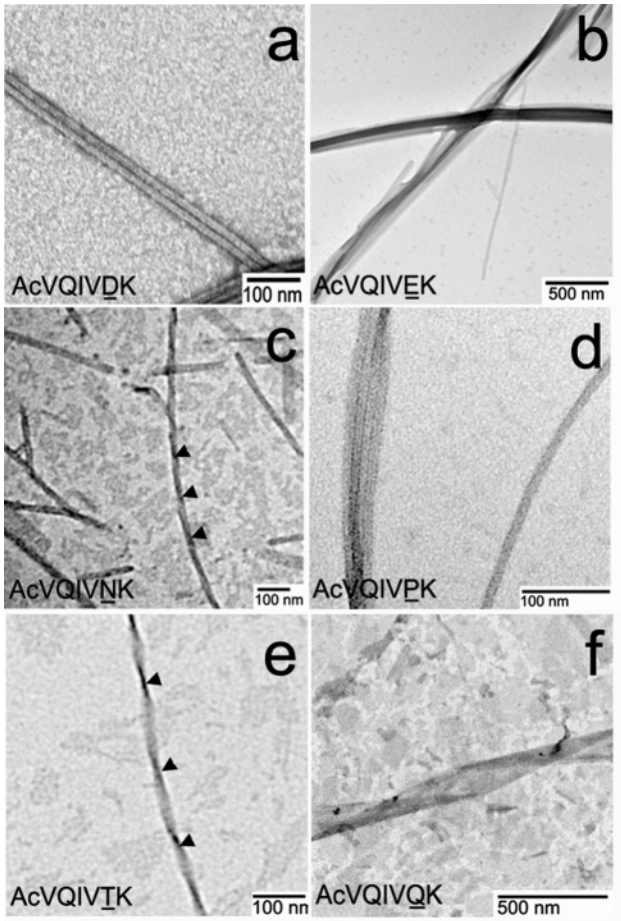

FIGURE 1: TEMs of amyloid formed from AcPHF6 and single-site mutants. Samples were prepared from aged peptides (~1 mg/mL) in 5 mM MOPS at pH 7.2 and negatively stained with 2% uranyl acetate. In the absence of added NaCl, AcVQIVDK and AcVQIVPK amides formed laterally aligned filaments ((a) and (d)), AcVQIVEK amide formed rolled sheets (b), and AcVQIVNK amide formed twisted fibers (c). In the presence of added 0.15 M NaCl, AcVQIVLK amide formed twisted fibers (e) and AcVQIVQK amide formed rolled sheets (f). A summary of amyloid morphologies formed from these and other mutants is given in Table 1.

filaments were observed for the Asn mutant in the absence of NaCl, whereas rolled sheets were observed in its presence. No amyloid was observed for the Asp, Glu, His, Lys, or Arg mutants in the presence of NaCl. Twisted filaments with morphologies slightly different from those of PHFs were observed for the Leu, Thr, and Trp mutants under these conditions.

FTIR of Peptides. The amide I region of the IR spectra are shown for four of the mutant peptides in Figure 2 along with best-fit Gaussian bands obtained by spectral deconvolution. Bands characteristic of β-sheet secondary structures (1600−1640 cm^{-1}), were observed for all of the peptides studied. In addition, some of the peptides showed substantial intensities between 1640 and 1660 cm^{-1}, characteristic of random-coiled or helical structures. The ratio of integrated intensities of the bands between 1600 and 1640 cm^{-1} relative to the total integrated intensity was used to estimate the fraction of β-structure present in each of the peptides, and these values are summarized in Table 2 ($A_{(1600-1640)}/A_{total}$). On the basis of these data, the Asp and Lys mutants are predicted to have less than 50% β-structure, whereas

Table 1: TEM Results for AcVYIVXK Peptides

mutant	buffer[a]	buffer + NaCl
Ala	sparse straight filaments ($d = 13 \pm 2$)	abundant straight filaments ($d = 12 \pm 2$)
Asp	abundant straight filaments ($d = 34 \pm 19$)	none
Glu	rolled sheets	none
Phe	abundant straight filaments ($d = 12 \pm 1$)	abundant straight filaments (d= 19 ± 3)
Gly	none	sparse straight filaments ($d = 29 \pm 6$)
His	none	none
Ile	rolled sheets	sparse straight filaments ($d = 24 \pm 4$)
Lys	none	none
Leu	rolled sheets	sparse twisted filaments ($d = 20 \pm 4$, $p = 129 \pm 19$)
Met	abundant straight filaments ($d = 27 \pm 12$)	sheets
Asn	twisted filaments ($d = 21 \pm 3$, $p = 131 \pm 21$)	rolled sheets
Pro	sparse straight filaments ($d = 8 \pm 2$)	rolled sheets
Gln	twisted filaments ($d = 27 \pm 10$, $p = 189 \pm 18$)	rolled sheets
Arg	none	none
Ser	sparse straight filaments ($d = 19 \pm 3$)	abundant straight filaments ($d = 15 \pm 3$)
Thr	straight and twisted filaments ($d = 27 \pm 4$)[b]	twisted filaments ($d = 45 \pm 8$, $p = 276 \pm 52$)
Val	abundant straight filaments ($d = 15 \pm 2$)	sparse twisted filaments ($d = 22 \pm 6$, $p = 156 \pm 27$)
Trp	abundant straight filaments ($d = 19 \pm 4$)	abundant straight and twisted ($d = 20 \pm 5$, $p = 156 \pm 25$)
Tyr (AcPHF6)	abundant straight filaments ($d = 8 \pm 1$)	abundant straight filaments ($d = 17 \pm 4$)

[a] Diameter (d) and periodicity (p) measurements given in nm. The number of measurements ranged from 10 to 32. [b] The periodicity was not measured because only one crossover was observed for all filaments.

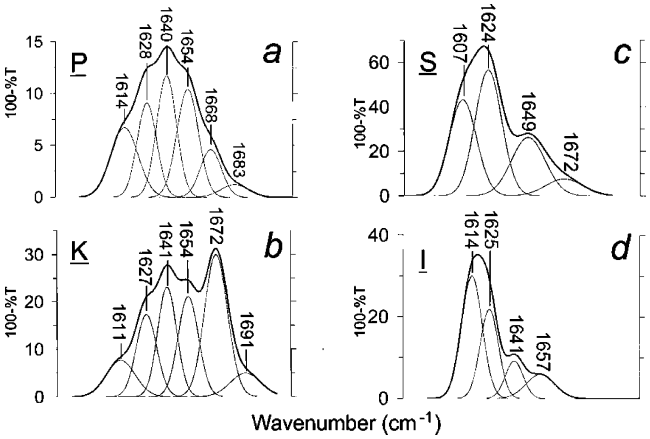

FIGURE 2: IR spectra in the region of the amide I bands for amides AcVQIVPK (a), AcVQIVKK (b), AcVQIVSK (c), and AcVQIVIK (d). The spectra were acquired for the peptides dissolved in D_2O. The band appearing between 1668 and 1672 cm^{-1} arises from the TFA in the sample. Spectra were fit using Gaussian bandwidths.

AcPHF6 and the Ala, Glu, Phe, His, Ile, Leu, Met, Gln, Thr, Val, and Trp mutants are predicted to have greater than 75% β-structure. Because of the presence of a high-frequency band (>1680 cm^{-1}), the Lys, Pro, Asp, Arg, Gln, and Met mutant peptides possess a measurable fraction of antiparallel β-sheet structure (*13*, *72*−*75*).

Far-UV CD of Peptides. CD has been used in the past to estimate the fraction of secondary structure in proteins (*88*, *89*). Similar methods have not been used for peptides because bands arising from the aromatic side chains can make a substantial but variable contribution to the observed ellipticity in the far-UV region (*90*, *91*). The CD spectra for AcPHF6 and some of the mutant peptides in 5 mM MOPS at pH 7.2 are shown in Figure 3a. Spectra with a strong positive band below 200 nm and a negative band at approximately 215 nm, similar to those observed for AcPHF6 and the Val, Phe, and Trp mutants, are characteristic of peptides with β-sheet conformations in solution (*90*). However, spectra observed for the remaining mutants in the Figure display a negative band extending from about 220 nm to below 200 nm and are typical of peptides in a random-coil conformation.

We used the ratio $(\Theta_{MR}(215\text{ nm}))/(\Theta_{MR}(192\text{ nm}))$ as a descriptor of the observed CD spectrum and this data is summarized in Table 2. Peptides with a CD spectrum characteristic of β-sheets have a negative ratio, whereas the ratio is positive for those peptides that have a CD typical of a random coil. AcPHF6 and the Phe, Ile, Met, Gln, Val, and Trp mutant peptides had negative values of $(\Theta_{MR}(215\text{ nm}))/(\Theta_{MR}(192\text{ nm}))$, suggesting that they have a large fraction of β-structure in solution. Although the sign of $(\Theta_{MR}(215\text{ nm}))/(\Theta_{MR}(192\text{ nm}))$ is indicative of the type of CD spectra, its magnitude is not meant to be a quantitative estimate of the fraction of β-structure present. For example, a larger negative ratio is calculated for the Gln mutant (−1.68) compared to that for the Val mutant (−0.66), although the absolute intensities of the 215 and 192 nm bands appeared larger in the case of the Val mutant, suggesting that a larger fraction of this mutant assumes a β-sheet configuration.

In HFIP, peptides that form amyloid in aqueous solution become soluble and no longer bind to CR or thioflavin dyes, and spectroscopic studies suggest that they are induced into random-coil or α-helical conformations (*92*−*94*). Figure 3b shows the CD spectra of some of the peptides in 100% HFIP. AcPHF6 and the Trp, Phe, and Val mutants, all of which displayed β-sheet type spectra in aqueous buffer, displayed negative shoulders between 217 and 222 nm and strong negative bands below 200 nm in HFIP, characteristic of peptides in random-coil conformations.

Congo Red Binding. Figure 4a illustrates the changes in the 540 nm absorption band of CR with increasing concentration of AcPHF6. Changes in the 540 nm absorbance with increasing peptide concentration, at a constant CR concentration (5 μM), are shown for some of the peptides in Figure 4b. Compared to the Asn, Asp, and Lys mutants, greater changes in absorbance with increasing peptide concentration were observed for AcPHF6 and the Met and Val mutants. These changes likely reflect the greater amounts of amyloid contained in the stock solutions (∼1 mM) from which these samples were prepared (*24*, *32*, *47*, *78*, *79*). We used the absorbance at fixed peptide and CR concentrations (35 μM peptide and 5 μM CR) as a semiquantitative measure of the amount of amyloid present in the remaining peptides, and the results are summarized in Table 2. With the exception of the Gln mutant, the same mutants associated with high CR absorbance values (>12) were also found to have negative values of $(\Theta_{MR}(215\text{ nm}))/(\Theta_{MR}(192\text{ nm}))$.

Kinetics of Amyloid Aggregation. The enhancement of the fluorescence at 490 nm, due to the binding of ThS to amyloid, was followed as a function of time for all peptides. However, only AcPHF6 and single-site mutants containing Ala, Met, Val, Phe, Ile, Trp, Leu, Gln, and Asn showed an

Table 2: Parameters Measured for Mutants of AcPHF6[a]

mutant	Θ_{MR}(215 nm)/ Θ_{MR}(192 nm)	$A_{540} \times 10^2$ (35 μM)	$A_{(1600-1640)}$/ A_{total}	$k_1 \times 10^3$ (s^{-1})	*CC* (μM)	lag time (s)	$\Delta G^\circ_{unf}(H_2O)$ (cal/mol)	retention time (min)	LC-MS partition coefficients
Ala	0.23	7.48	0.81	0.62 ± 0.01	4.6 ± 2.2	3037 ± 147	2984 ± 380	14.6	0.00
Asp	0.30	8.25	0.50	ND	ND	ND	ND	13.9	0.00
Glu	0.41	7.73	0.84	ND	ND	ND	ND	14.3	0.57
Phe	−0.36	12.82	0.77	5.6 ± 0.2	2.9 ± 1.0	99 ± 10	1820 ± 269	17.7	3.72
Gly	0.35	7.79	0.60	ND	ND	ND	ND	14.0	0.00
His	0.31	8.18	0.81	ND	ND	ND	ND	8.9	0.00
Ile	−1.31	11.01	0.75	2.9 ± 0.1	3.9 ± 1.8	55 ± 24	75 ± 92	16.7	4.93
Lys	0.36	7.56	0.45	ND	ND	ND	ND	8.6	0.00
Leu	0.85	13.81	0.85	3.3 ± 0.1	3.6 ± 1.2	383 ± 25	809 ± 77	17.7	[b]
Met	−1.22	17.71	0.90	2.0 ± 0.1	1.3 ± 0.1	101 ± 32	672 ± 115	16.6	0.83
Asn	0.46	8.92	0.67	2.0 ± 0.3	3.3 ± 1	10 ± 135	763 ± 454	13.4	1.03
Pro	0.50	8.57	0.66	ND	ND	ND	ND	14.3	0.00
Gln	−1.68	8.39	0.88	0.13 ± 0.01	4.8 ± 2.5	24032 ± 1500	661 ± 128	13.7	0.00
Arg	0.31	6.85	0.69	ND	ND	ND	ND	16.6	0.00
Ser	0.86	8.47	0.70	ND	ND	ND	ND	13.9	0.00
Thr	0.34	8.22	0.80	ND	ND	ND	ND	14.2	0.00
Val	−0.66	14.95	0.79	5.8 ± 0.2	2.5 ± 0.6	106 ± 12	−61 ± 57	15.5	3.69
Trp	−0.42	15.93	0.86	1.9 ± 0.1	4.4 ± 1.2	102 ± 48	904 ± 272	17.3	1.29
Tyr (AcPHF6)	−1.90	12.08	0.85	6.5 ± 0.5	3.1 ± 1.4	1 ± 23	2352 ± 122	14.9	11.31

[a] ND is real data that could not be measured under the time constraints of the experiment. For the following variables, ND was assigned the following values: *k*, 0.00; *CC*, 100 μM; lag time, 100 000 s; $\Delta G^\circ_{unf}(H_2O)$, −2000 cal/mol. [b] The Leu mutant was not present in the sample mixture because it is isobaric to the Ile mutant.

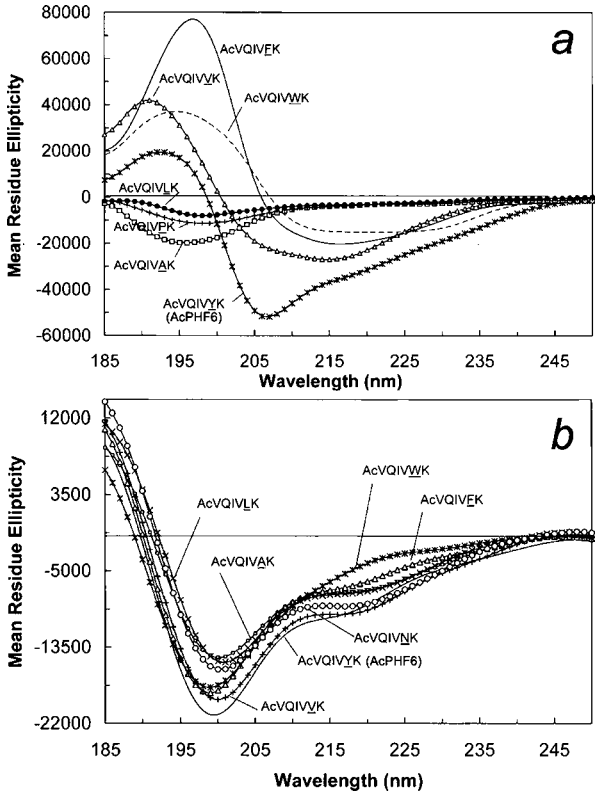

FIGURE 3: CD spectra of AcPHF6 and some single-site mutants dissolved in 5 mM MOPS at pH 7.2 (a) and peptides in 100% HFIP (b). The peptides were equilibrated in the solvent for at least 4 days prior to acquisition.

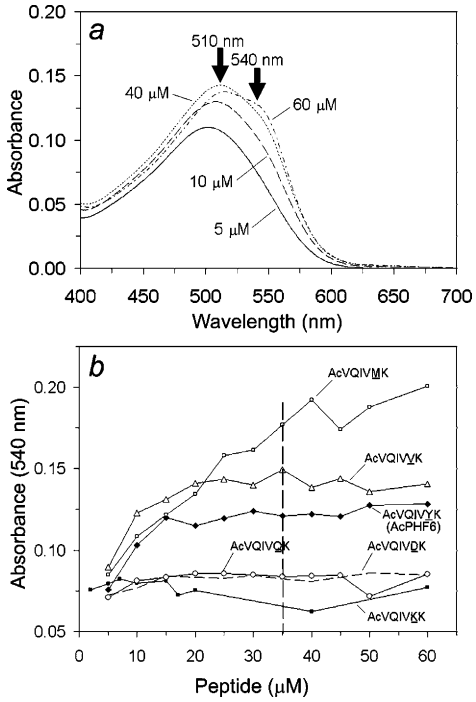

FIGURE 4: UV−Vis spectra of 5 μM Congo red in the presence of 5−60 μM AcPHF6 dissolved in 5 mM MOPS at pH 7.2 (a). Absorbance at 540 nm for AcPHF6 and some single-site mutants as a function of increasing peptide concentration (b). The absorbance at 35 μM was used to quantify the amount of amyloid present.

increase in ThS fluorescence (Figure 5a and b). In general, aggregation of these peptides appeared to follow a nucleation−elongation mechanism with characteristic lag times (*80*, *81*). The data were fit to Gompertz growth curves (*83*), and the kinetic rate constants and lag times extrapolated from the fit are summarized in Table 2. Mutants containing Tyr, Val, and Phe showed the fastest rates of aggregation, whereas the Ala and Gln mutants showed the slowest aggregation rates and longest lag times.

Critical Monomer Concentration (CC) of Peptides. The critical monomer concentration required to trigger aggregation was extrapolated from a plot of ThS fluorescence as a function of peptide concentration (Figure 5c), and the data is summarized in Table 2. The Met mutant peptide had the lowest CC (1.3 μM), whereas the mutants containing Ala, Phe, Ile, Leu, Asn, Gln, Val, Trp, and Tyr had CCs within

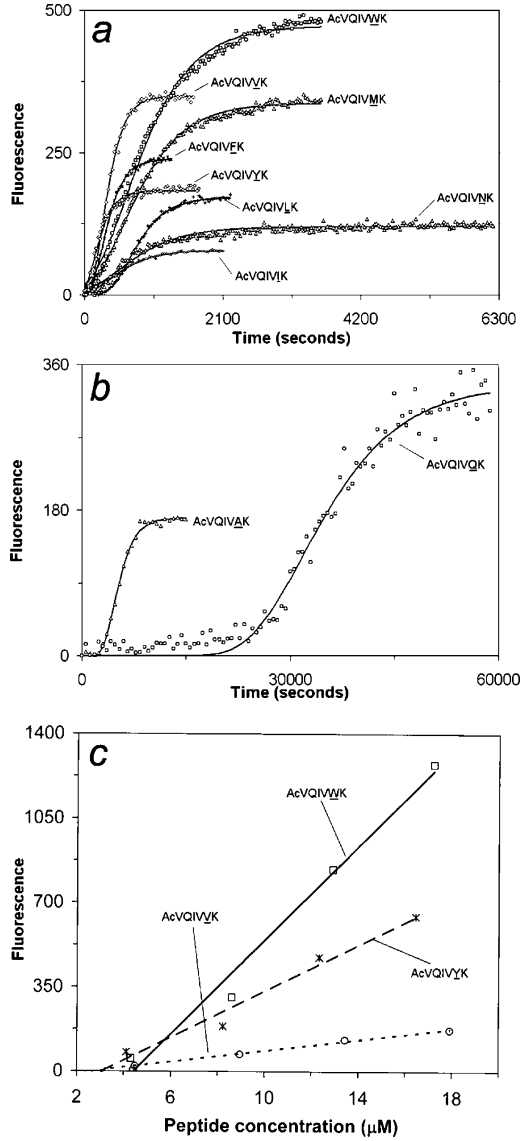

FIGURE 5: Kinetics of assembly (21 °C) of AcPHF6 and some single-site mutants ((a) and (b)). The peptides were dissolved (~ 100 μM) in 20 mM MOPS at pH 7.2. The aggregation was monitored by following the fluorescence of ThS upon binding to amyloid. Curves connecting the data were fitted using a Gompertz growth curve (eq 1). Critical monomer concentrations were calculated by linear extrapolation of the equilibrium ThS fluorescence (c).

error, with values ranging from 2.5 to 4.8 μM. Data for the remaining mutants were not determined owing to the absence of ThS binding at the peptide concentrations studied.

Free Energy of Stabilization (Unfolding) of Amyloid Aggregates. CD spectra of the peptides in HFIP suggest that in this solvent they exist primarily in random-coil conformations (Figure 3b). Because of its ability to convert peptides with a β-sheet conformation in buffer to a random-coil conformation, we used HFIP as a denaturant. Using the ThS fluorescence at the beginning and end of an HFIP titration to represent values for the aggregated and monomeric forms of the peptides, the fraction unfolded, and the values for free energy of unfolding (ΔG°_{unf}) were calculated and are shown in Figure 6a and b for AcPHF6 and for the Leu and Phe mutants. The free energy of unfolding in buffer $\Delta G^\circ_{unf}(H_2O)$, a measure of the free energy of stabilization of the insoluble amyloid state, was obtained by extrapolating

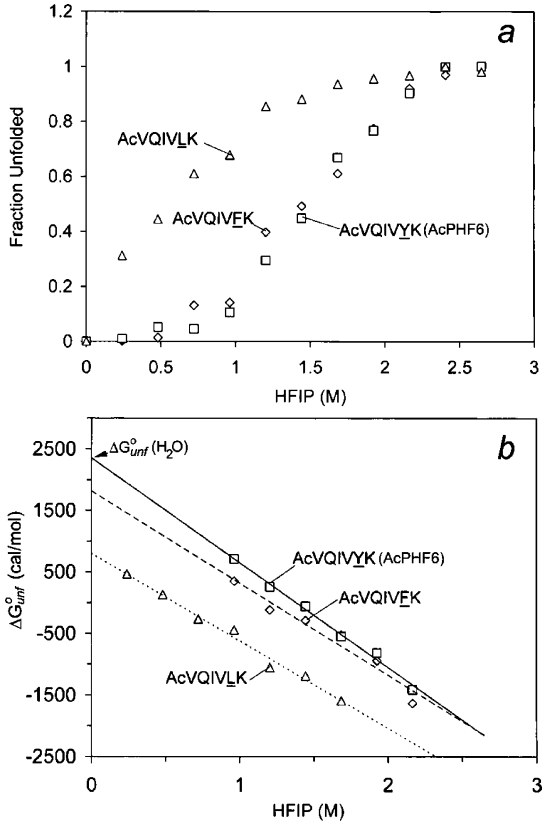

FIGURE 6: Unfolding of peptide aggregates by HFIP as followed by the change in equilibrium ThS fluorescence. The fraction of unfolded peptide was calculated according to ref 84 (a) and $\Delta G^\circ_{unf}(H_2O)$ was found by extrapolating to zero HFIP concentration and ΔG°_{unf} was calculated from the unfolded fraction (eq 2) (b).

the ΔG°_{unf} data to zero concentration of HFIP. The data for all peptides studied is summarized in Table 2. The data show relatively large free energy of stabilization values for AcPHF6 and the Ala and Phe mutants. However, the amyloid forms of the Ile and Val mutant peptides are nearly in equilibrium with their monomeric forms ($\Delta G^\circ_{unf}(H_2O) \approx 0$).

HPLC Retention Times. Reverse-phase HPLC RTs have been used as a measure of hydrophobicity (*95*). Table 2 summarizes RTs at low pH (~2) for the mutant peptides studied. The longest RTs were obtained for the Phe, Leu, and Trp mutants, whereas the shortest RTs were obtained for the His and Lys mutants.

LC-MALDI Analysis. We used the GlaD method (*86*) to measure the propensity of a peptide to form amyloid by measuring its partitioning between soluble and insoluble forms in a solution containing a mixture of 18 of the 19 single-site mutant peptides. (The Leu mutant was not included because it is isobaric to the peptide containing Ile.) Following centrifugation, the peptides present in the amyloid pellet were labeled with $^{15}N_2$-OMIU, whereas the peptides in the supernatant were reacted with unlabeled OMIU. Figure 7a shows the labeled peptide ions observed in the MALDI-TOF spectrum of the unfractionated mixture after the fractions were modified and recombined. Ions produced by derivatization with the light version of OMIU were observed for all peptides except the Val and Asn mutants, where overlap with other ion signals in the spectrum prevented their observation. To circumvent this problem, the resolution of all components was obtained following chromatographic

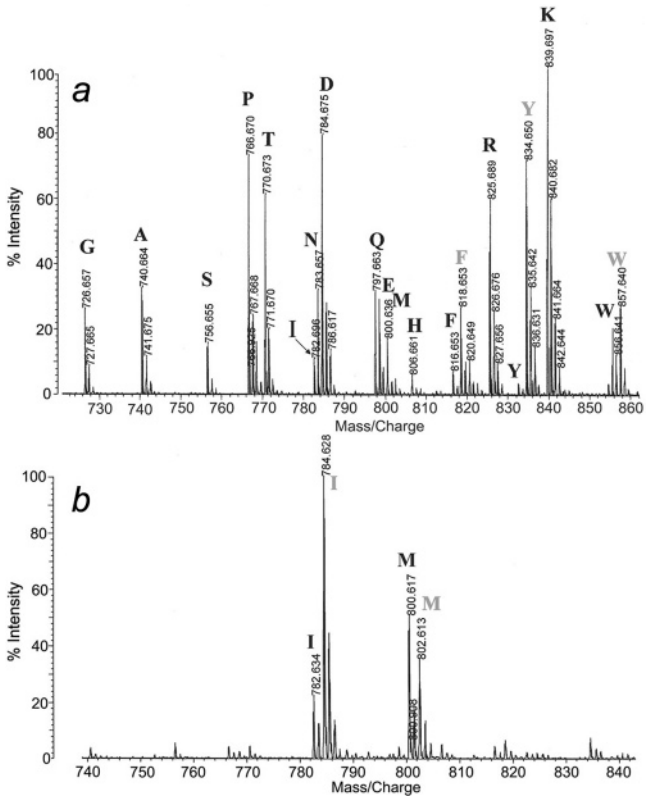

FIGURE 7: MALDI-TOF spectra of the mixture containing single-site mutants of AcPHF6 (a). The amyloid formed in the mixture was isolated by centrifugation and labeled with ^{15}N-OMIU (light abbreviations) and the supernatant was labeled with ^{14}N-OMIU (bold abbreviations). The two fractions were recombined prior to MALDI analysis. The signal of the Lys mutant corresponds to the doubly modified peptide because it contains two Lys residues at positions 5 and 6 in the sequence. MALDI spectrum containing a single chromatographically purified fraction (b) Abbreviations represent single amino acids substituted at position 5 of AcPHF6.

separation. Figure 7b shows the MALDI-TOF spectrum of the fraction containing the Ile and Met mutants. The peptides present in the amyloid fraction, which were labeled with the heavy isotopes, included AcPHF6 and the Glu, Phe, Ile, Asn, Met, Val, and Trp mutants. Partition coefficients were calculated from peak intensities and are summarized in Table 2. AcPHF6 was at least 2−3 times more abundant in the amyloid fraction compared to that in other peptides. Of those present, the least abundant peptide in the amyloid fraction was the Glu mutant.

Principal Component Analysis. When considered cumulatively, data in Tables 1 and 2 suggest that different peptides have different morphologies and spectroscopic, thermodynamic, and kinetic properties. We used these measured properties as a basis for PCA of the data. The data was arranged in a 19 × 10 matrix (19 samples, 10 variables) and 2 significant principal components (PCs) were found to capture 71% of the variance of the data. A scores plot of the data is shown in Figure 8a. Peptides that possess many of the properties characteristic of amyloid, such as fibrous morphology and binding of ThS and CR, were found to have positive PC1 scores, whereas mutants that had weaker interactions with CR or that did not bind ThS (Pro, Thr, Ser, Glu, Asp, Arg, Gly, His, and Lys) had negative PC1 scores. Within the group of peptides with positive PC1 scores, those peptides closer to the origin (the Ala, Gln, and Asn mutants)

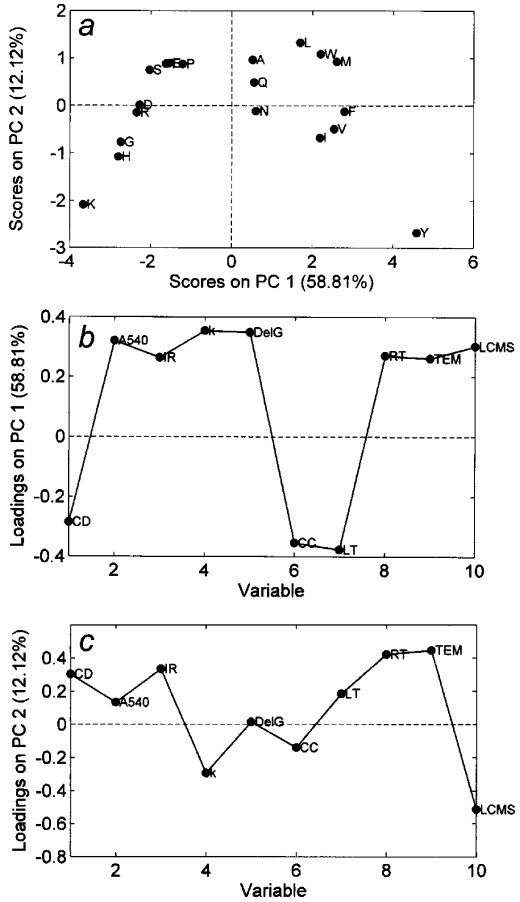

FIGURE 8: Results of PCA for AcPHF6 and single-site mutants. Scores plot along the largest two PCs (a). Loadings plots for PC1 (b) and PC2 (c). Data in Table 1 and 2 served as a basis for the analysis.

were found to possess a few but not all of the properties characteristic of peptides with positive PC1 scores. For example, the Ala mutant was found to bind ThS and was observed by TEM to have filaments but had a positive value of $(\Theta_{MR}(215\text{ nm}))/(\Theta_{MR}(192\text{ nm}))$ (0.23). The peptides in the lower left quadrant of the plot (Lys, His, Gly, and Arg mutants) did not form filaments observable by TEM, whereas all of the peptides in the lower right quadrant of the plot (Ile, Val, Phe, and Tyr mutants) were found to form amyloid in the prepared mixture of peptides analyzed by MALDI-TOF.

The loadings plots provide information about the variables that may be important for the distribution seen in the scores plot. The plot for PC1 (Figure 8b) shows that $(\Theta_{MR}(215\text{ nm}))/(\Theta_{MR}(192\text{ nm}))$, CC, and the lag time make a negative contribution to PC1. This makes physical sense because the smaller or more negative values for these parameters are associated with amyloid formation. The remaining variables make approximately equal positive contributions, suggesting that no single variable is more important for separating the peptides along PC1. The loadings plot for PC2 (Figure 8c) shows that retention time and the results of the TEM and MALDI-TOF are most important in separating peptides along PC2.

Amyloidogenic Propensities of Individual Amino Acid Residues. We used the PC1 scores to define the propensity of individual amino acids to form amyloid at the single mutation site in the AcPHF6 sequence, and these propensities

Table 3: Amyloidogenic Propensities of Individual Amino Acid Residues[a]

mutant	PC1 propensity
Ala	0.52
Asp	−2.28
Glu	−1.56
Phe	2.80
Gly	−2.74
His	−2.81
Ile	2.19
Lys	−3.67
Leu	1.70
Met	2.61
Asn	0.60
Pro	−1.22
Gln	0.55
Arg	−2.35
Ser	−2.03
Thr	−1.63
Val	2.54
Trp	2.20
Tyr	4.59
(AcPHF6)	

[a] Derived from PCA analysis of AcVQIVXK amide where X is one of 19 L-amino acid residues.

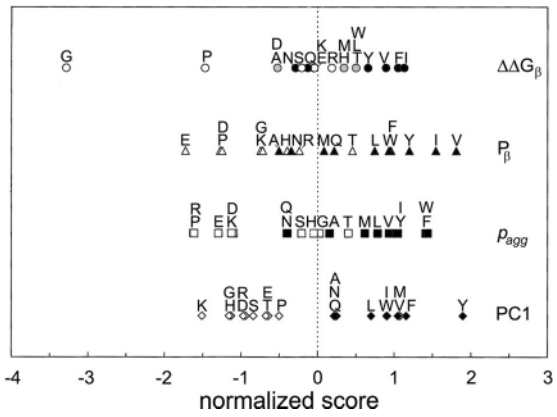

FIGURE 9: Comparison of the amyloidogenic propensities of amino acid residues predicted from this study (PC1) with those based on thermodynamic β-sheet propensities ($\Delta\Delta G_{\beta-\text{sheet}}$) *(97)*, β-sheet propensities based on statistical abundance (P_β) *(96)*, and amyloidogenic propensities based on rates of amyloid formation (p_{agg}) *(33, 34)*. The values have been mean-centered and normalized to unit variance. Filled symbols are included to aid in the comparison of propensities determined from other methods with residues with positive PC1 scores. Gray symbols in the plot for $\Delta\Delta G_{\beta-\text{sheet}}$ represent an overalap of filled and unfilled symbols.

are listed in Table 3. As a comparison, we show amyloidogenic propensities (scaled to unit variance) derived by fitting aggregation rate data from human acylphosphatase (p_{agg}) (*33*, *34*), β-sheet propensities based on frequencies of occurrence of individual amino acid residues in crystalline proteins (P_β) (*96*), and thermodynamic β-sheet propensities derived from changes in the stability of zinc-finger protein ($\Delta\Delta G_{\beta-\text{sheet}}$) (*97*) in Figure 9. According to the PC1 scale, AcPHF6 (Tyr mutant) is the most amyloidogenic peptide followed by the Phe, Met, Val, Ile, Trp, and Leu mutants. The least amyloidogenic peptides according to this scale are the Lys, His, and Gly mutants. Although the order within the group of more amyloidogenic peptides differs in comparison to PC1, the p_{agg} value also predicts the amino acids with the greatest amyloidogenic propensities to be Trp, Phe, Tyr, Ile, Val, Leu, and Met. The P_β scale is similar to p_{agg},

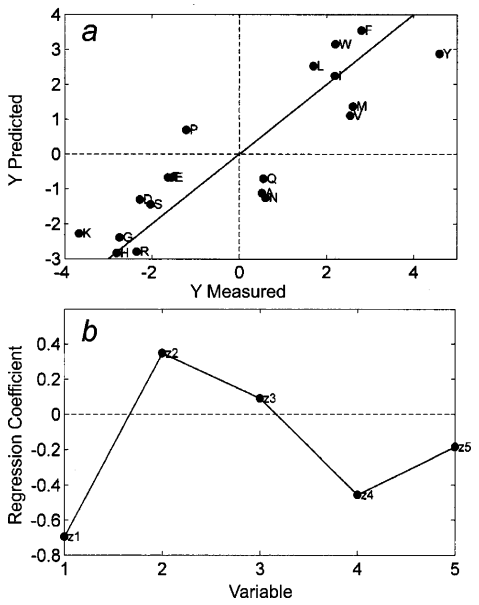

FIGURE 10: PC1 scores predicted from PLS vs PC1 scores obtained by PCA modeling (a). The scores were predicted from a PLS model constructed from three LVs ($R^2 = 0.76$ and $Q^2 = 0.56$). Amino acid descriptors (z_1−z_5) for the single-site mutation served as a basis for the chemical structural aspect of the model *(68)*. Regression coefficients for amino acid descriptor variables obtained from the model (b).

differing only in that Thr replaces Met in the group of seven most amyloidogenic residues. The $\Delta\Delta G_{\beta-\text{sheet}}$ scale correctly classifies Ile, Phe, Val, and Tyr among the most amyloidogenic residues but misclassifies the His, Arg, Asn, Gln, Thr, and Ala mutants.

Interpretation of Propensities in Terms of Structural Properties. Although PCA suggests which observables are important for differentiating peptides, it does not provide a structural interpretation of the data. We attempted to correlate observable properties, as reflected in PC1 scores, to structures using PLS modeling, where each peptide was represented by a 5-parameter *z* code for the single amino acid residue at position 5 of the AcPHF6 structure (because the peptide structures differ by only one amino acid residue) (*66*−*69*). The *z* scores are based on 26 physical and chemical properties, including hydrophobicity, steric parameters, heat of formation, α-polarizability, NMR parameters, HOMO and LUMO energies, and side-chain formal charge (*68*). To a large extent, z_1 corresponds to hydrophobicity, z_2 to steric constraints, z_3 to electronic properties, z_4 to heat of formation and electronegativity, and z_5 to hardness of the HOMO and LUMO energies. Using the model, structural data (**X**) and observables (**Y**) are projected onto a common set of LVs, similar to principal components, and the scores are related to one another through a set of regression coefficients (*69*). We found that a 3 LV model was able to account for 85.3% of the variance in the **X** block and 76.0% of the variance in the **Y** block and gave reasonably good modeling and predictive ability ($R^2 = 0.76$, $Q^2 = 0.56$). Figure 10a shows a plot of the PC1 scores predicted from structural parameters to those based on measured properties. The regression coefficients of the variables obtained from the fitting are plotted in Figure 10b. A structural interpretation of the regression coefficients would suggest that large (z_2), hydrophobic (z_1) amino acids substituted at position 5 in our

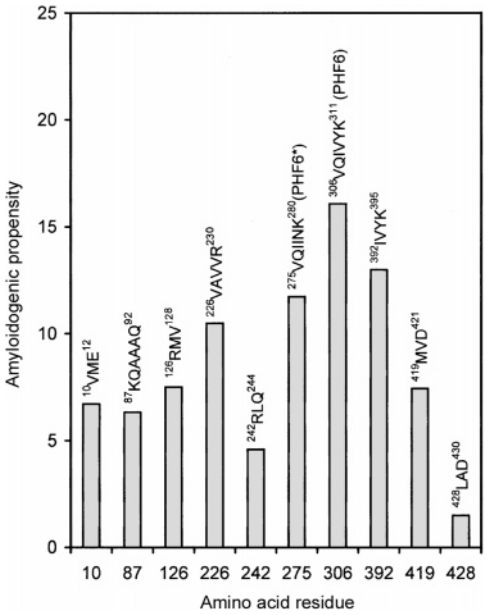

FIGURE 11: Nucleating sequences in tau predicted from individual amino acid propensities. The sequences were predicted assuming a polar (or charged) terminal residue followed by a hydrophobic tail.

peptides favor amyloid formation. The negative contribution of hydrophobicity (z_1) seen in the plot may be accounted for because greater negative values of z_1 correspond to more hydrophobic residues (*68*). Amino acids that have negative values for z_1 and would thus be expected, on the basis of the model, to support amyloid formation include Leu, Ile, Met, Phe, Pro, Trp, Tyr, and Val. On the basis of the signs of the regression coefficients, AcPHF6 and the peptides with Phe and Trp mutants were predicted to be particularly amyloidogenic.

Prediction of Nucleating Sequences in Tau from Amyloidogenic Propensities. There are two types of structural themes that have been proposed to initiate amyloid formation in model peptide systems: a sequence of consecutive hydrophobic residues followed by a charged or polar residue, as in PHF6 and PHF6* (*43*, *49*), or alternatively, a sequence in which hydrophobic and polar residues alternate (*45*−*48*). Both sequences are amphiphilic, the first end-to-end and the second, above and below the plane of the peptide chain. We scanned the sequence of tau and identified sequences fitting the first of these two structural motifs because it most closely fits the AcPHF6 peptides studied. Once such themes were identified, the sum of the absolute values of their PC1 scores was used to quantify the amyloidogenic propensity of the sequence. Figure 11 shows the identified sequences and their corresponding propensities. The most amyloidogenic sequences include ^{226}VAVVR^{230} preceding R1, ^{275}VQIINK^{280} (PHF6*) in R2, PHF6 in R3, and ^{392}IVYK^{395} in the C-tail region of the protein. Other sequences with a lower probability of seeding amyloid formation according to our model include ^{10}VME^{12}, ^{87}KQAAAQ^{92}, and ^{126}RMV^{128} in the *N*-terminal region of the protein and ^{419}MVD^{421} in the C-tail.

DISCUSSION

We observed that 15 of the 19 peptides studied formed amyloid in buffer, and the morphologies of the amyloid observed varied greatly between peptides, ranging from filaments and twisted filaments to rolled sheets. The morphologies observed for each peptide strongly depended on the characteristics of the medium, with only two peptides (AcPHF6 and the Thr mutant) displaying identical morphologies in the presence and absence of added NaCl. Only the Arg, Lys, and His mutants did not form amyloid under either set of conditions. In some cases, similar morphologies were observed for peptides that shared similar chemical properties. Mutants that contained residues with aromatic side chains (AcPHF6, Trp, and Phe) all formed abundant straight filaments in the presence and absence of added salt. The structurally similar Leu and Ile mutants or the Asn and Gln mutants also formed similar morphologies in buffer or in buffer containing NaCl. An exception is observed in the case of the Glu and Asp mutants, which are structurally similar but in buffer form amyloid with two different morphologies (rolled sheets and straight filaments). Several of the mutants we studied here, including the Asn, Gln, Thr, Trp, and Val mutants, formed twisted filaments of similar widths but longer periodicities than those seen in PHFs.

Amyloid has traditionally been defined as insoluble proteineous deposits that are birefringent upon CR staining, display a cross-β X-ray fiber diffraction pattern, and has an unbranched fibrous morphology (*2*, *21*). However, our results and those of others have shown that some peptides and proteins aggregate into amyloid displaying a range of morphologies, including straight and twisted fibers, spheres, ribbons, and sheets, and their morphologies depend on extrinsic factors, including salt concentration, pH, and temperature (*16*−*22*, *31*, *45*, *46*, *48*, *52*, *53*, *98*). Some workers have claimed that the binding of thioflavin dyes or spectral parameters observed in the CD or IR spectra can be correlated to the presence of amyloid aggregates or to the rate of amyloid formation (*10*, *13*, *46*, *49*, *52*−*54*, *94*). To further complicate the criteria for amyloid formation, it has been found that many proteins that are soluble and do not form amyloid bind CR (*79*). Some of the peptides we studied possessed some but not all of the characteristics typical of amyloid. For example, the Gln mutant formed twisted filaments or rolled sheets in the absence and presence of added salt, had a CD and IR spectra characteristic of amyloid (a negative (Θ_{MR}(215 nm))/(Θ_{MR}(192 nm)) and 88% β-sheet by IR) but did not bind CR, showed relatively slow aggregation kinetics, and was not seen to partition in the amyloid fraction in a mixture of peptides. Hence, different methods provide complimentary criteria for the presence of amyloid, or, in other words, the definition of amyloid is dependent on the methods used to define it. Our approach to this problem was to treat a variety of measurements in a comprehensive manner. Scores obtained from PCA were used to define the amyloidogenic tendency of different peptides and because the peptides differed by only a single amino acid, the amyloidogenic propensity of different residues. The distance in the PCA scores plot between the Ala, Asn, and Gln mutants and the more amyloidogenic Leu mutant is nearly equivalent to their distance to the less amyloidogenic Pro mutant. Hence, the scores should be viewed as a continuous scale upon which correlations to structure can be based.

One of the most surprising findings was that the Pro mutant formed either straight filaments or rolled sheets. Because of the lack of an amino hydrogen and the steric

constraints of the ring, Pro has long been considered a β-sheet breaker, and scanning mutagenesis using either Pro or Ala substitutions has been used to search for nuclei in amyloid-forming peptides and proteins (*35*, *42*, *49*, *99*). In our study of AcPHF6 peptides, the Ala mutant was found to be one of the more amyloidogenic peptides. Hence, our results suggest that the validity of using scanning mutagenesis as a tool for finding amyloidogenic nuclei may depend on the sequence searched, where the residue is substituted in the sequence, and the criteria used to assess amyloid.

The loadings plot for PC1 (Figure 8b) suggests that all of the observable properties contribute nearly equally to the scores along PC1 and, hence, to the amyloidogenic propensity of each peptide. This suggests that the propensity of each of the peptides to form amyloid cannot be evaluated in terms of any single property. In the past, CD and TEM have been used to characterize peptides capable of forming amyloid and to screen peptides from combinatorial mixtures (*52*−*54*). Our results suggest that not all peptides giving rise to fibers in electron micrographs should be considered amyloidogenic. Furthermore, the CD spectrum obtained from the mixture is a weighted average, and the β-sheet spectra of some amyloidogenic peptides in a mixture may be obscured by the random-coil spectra of nonamyloidogenic peptides occurring in greater abundance. A more reliable method for screening may be the separation of amyloid-forming peptides by centrifugation and identification by using the GlaD method.

Recently, Dobson and co-workers introduced an amyloidogenic propensity scale for individual amino acids based on phenomenological expressions that relates a set of intrinsic factors, including hydrophobicity, charge, β-sheet propensity, α-helix propensity, and the rates of aggregation of single-site mutants of human muscle acylphosphatase (AcP) and a few other proteins for which data were available (*33*, *34*). Because the fitting was carried out on intact proteins, the intrinsic propensities (p_{agg}) are sequence-independent and do not necessarily reflect propensities within nucleating sequences. In contrast, the PC1 propensities are specific for a single position in the AcPHF6 sequence but, unlike p_{agg}, reflect a combination of properties characteristic of amyloid formation. Despite these differences, residues with the greatest propensities, according to either scale, are residues with aromatic or large hydrophobic side chains. Both scales also show that positively charged residues, such as His, Arg, and Lys, have low amyloidogenic propensities. In AcPHF6 mutant peptides, the presence of these residues gives rises to two sequential positively charged residues (at position 5 and 6). Others have previously noted that model peptides containing two consecutive positively charged residues fail to form amyloid, and the tau ΔK280 mutant, in which one of two consecutive Lys residues is omitted, shows a faster aggregation rate (*49*, *56*, *62*, *100*). In other cases, notable differences occur between the p_{agg} and the PC1 scales. For example, according to the p_{agg} scale, Pro is the least amyloidogenic residue because of its negative contribution of β-sheet propensity in the phenomenological expressions used to construct the model. In contrast, Pro has a higher relative PC1 score because the Pro mutant was observed to form filaments, whereas the His, Arg, and Lys mutants did not form filaments under any of the conditions investigated. Other differences include propensities for Thr, which has a p_{agg} nearly equivalent to Met, and Gln and Asn, which have lower p_{agg} values than their comparable PC1 scores.

The P_β scale was formulated to reflect β-sheet propensity on the basis of the frequency of occurrence of residues in β-sheet secondary structure in crystalline globular proteins (*96*). The $\Delta\Delta G_{\beta-\text{sheet}}$ propensity scale is the thermodynamic stability of β-sheet structure in zinc-finger proteins, and the differences between this scale and the others likely reflect the unique β-structure of this protein (*97*). For example, the Arg and His mutants of AcPHF6 do not form amyloid, but these residues are relatively well tolerated in zinc-finger protein. Both the P_β and the $\Delta\Delta G_{\beta-\text{sheet}}$ scales differ from PC1 and p_{agg} in de-emphasizing the contribution of aromatic amino acids to amyloid formation. Hence, Val and Ile, two residues with large hydrophobic nonaromatic side chains, are calculated to be the most amyloidogenic. The positive contribution of aromatic residues to amyloid formation has been previously noted in other peptides (*3*, *38*−*41*). Our QSAM calculations suggest that, at least in the case of the AcPHF6 mutants, hydrophobicity is the most significant structural property determining amyoidogenicity, followed by size. Electronic structure, which might give rise to π-stacking contributions, was seen to play a less significant role. Others have found that amyloidogenic propensity, as reflected by changes in aggregation rates between mutant proteins, arises from an equal contribution of hydrophobicity, β-propensity, π-stacking, and charge (*101*).

We have used individual amino acid PC1 propensities in sequences containing a hydrophilic (or charged) headgroup and a hydrophobic tail, similar to the PHF6 and PHF6* motifs, to predict nucleating sequences in tau. Our results suggest that a number of sequences throughout the protein have the potential of acting as amyloidogenic seeds, including ^{10}VME^{12}, ^{87}KQAAAQ^{92}, and ^{126}RMV^{128} in the *N*-terminal region, ^{226}VAVVR^{230} in the proline-rich region, PHF6* and PHF6 in the repeat region, and ^{392}IVYK^{395} and ^{419}MVD^{421} in the C-tail region of the protein. Nucleating sequences with a theme of alternating residues of high and low amyloidogenic propensities have been observed to seed amyloid formation in peptides (*45*−*48*) and are present in tau, for example, ^{337}VEVK^{340} and ^{343}KLDFK^{347} in the R4 region. However, because peptides following this motif were not characterized in our study, their amyloidogenic propensities may not be directly compared to those following the PHF6 motif.

Early work showed that peptide fragments composed of three of the four pseudorepeats of tau were tightly bound to pronase-treated PHFs termed the PHF core (*102*−*105*). It was subsequently shown that tau constructs containing either the 3R or 4R forms of the MTBR could form twisted filaments with PHF-like morphology (*56*, *106*). The nucleating fragments PHF6 and PHF6* were then isolated from the smallest peptide fragment of the MTBR (PHF43) able to seed full-length tau into filaments resembling bona fide PHFs (*35*, *56*). However, there is both direct and indirect evidence that suggests the presence of sequences other than PHF6 and PHF6* aid in the nucleation of amyloid or regulate tau aggregation kinetics. This evidence and its relationship to our predicted nucleating sequences include the following: (1) The intensities of the bands attributable to β-structure in the CD and FTIR spectra of the aggregated form of the 120-residue repeat domain are proportional to a much greater

fraction of the protein than can be attributed to the two hexapeptide nucleating sequences, suggesting that other sequences contribute to β-sheet structure (*56*). Recent NMR evidence suggests that sequences containing PHF6* and PHF6 at the beginning of R2 and R3 and the sequence ^{336}QVEVKSEKLD^{345} at the beginning of R4 are in rapid exchange with β-structure in aggregated forms of these tau constructs (*107*). In agreement with these observations, PHF6* and PHF6 are predicted by our algorithm to be highly amyloidogenic, whereas the later sequence containing portions of ^{337}VEVK^{340} and ^{343}KLDFK^{347} are predicted to be amyloidogenic on the basis of the motif of alternating residues of high and low propensities. (2) The C-tail residues 409−438 and 383−438 also form filaments, suggesting that nucleation sequences are present in this region, and NMR studies suggest that ^{230}Arg and ^{393}Val show restricted mobility in the aggregated form of full-length tau (*108*, *109*). In addition, immunological evidence (with Tau-66) suggests that the region containing ^{226}VAVVR^{230} interacts with PHF6 (*110*). These studies are consistent with the potential role that our predicted amyloidogenic sequences ^{392}IVYK^{395}, ^{419}MVD^{421}, and ^{226}VAVVR^{230} play in aggregation. (3) Immunological analysis (with Alz50 and MC1) suggests a pathological conformation where the *N*-terminal is folded over the repeat domain, and spectroscopic evidence suggests that ^{18}Tyr interacts with residues in the region of PHF6 (*111*−*113*). Mutants in which residues 2−18 are deleted aggregate more slowly (*114*), consistent with the important role the *N*-terminal region plays in aggregation. Our prediction of the nucleation sequence ^{10}VME^{12} is consistent with these experimental observations. (4) Full-length (4R) tau constructs missing the two *N*-terminal inserts (residues 46−103 of full-length tau) aggregate more slowly (*115*), suggesting the presence of amyloidogenic regions in this segment of the protein. Our work predicts the amyloidogenic sequence ^{87}KQAAAQ^{92} in this region, consistent with these observations. In summary, NMR studies of the tau MTBR confirm the structural role which PHF6 and PHF6* and ^{336}QVEVKSEKLD^{345} play in the formation of PHFs, whereas NMR, fluorescence, and immunological studies confirm that sequences outside the MTBR in the *N*-terminal region (residues 2−18), the proline-rich region (^{230}Arg), and in the C-tail (^{393}Val) play either a kinetic or a structural role in tau polymerization. In many cases, residues or sequences in tau implicated by these studies as being important in PHF formation contain sequences predicted by our algorithm to have high amyloid-forming potential. These sequences may either seed amyloid in and of themselves or interact with other amyloidogenic regions of the protein such as the MTBR to form partially folded intermediates.

Pawar et al. (*33*) have used amyloidogenic propensities derived from their kinetic based model to predict changes in aggregation rates caused by missense of point mutations in tau associated with pathology, leading to increased NFT accumulation. Although these workers make the argument that changing a residue from a less amyloidogenic one to a more amyloidogenic one would increase the aggregation rate, we believe that such changes without the increase in length (and hence the propensity) of an existing amphiphilic sequence would have only a marginal effect. Using this strategy, N279K, which breaks up the amphiphilic PHF6* sequence, would be expected to have a decreased aggregation rate, whereas S305N and ΔK280, which extend the motifs in PHF6 and PHF6*, would be expected to show increased aggregation rates. Other missense mutations are predicted from our calculations to have only a small effect on aggregation rate. Experimentally, G272V, N279K, and V337M have been shown to have slightly greater aggregation rates, whereas P301L and ΔK280 have been shown to have much greater rates of aggregation compared to that of wild-type tau (*100*). However, the accumulation of PHFs characteristic of these pathologies is a result of increased tau aggregation rates, the decreased affinity that some of the tau mutants have for microtubules leading to increased unbound tau concentrations, and the changes in isoform expression linked to the mutation (*116*). Hence, it is not surprising that neither our results nor those of Pawar et al. are able to quantitatively predict the increased aggregation rates associated with these pathologies.

REFERENCES

1. McParland, V. J., Kalverda, A. P., Homans, S. W., and Radford, S. E. (2002) Structural properties of an amyloid precursor of beta-(2)-microglobulin, *Nat. Struct. Biol. 9*, 326−331.
2. Lansbury, P. T. (1992) In pursuit of the molecular structure of amyloid plaque: New technology provides unexpected and critical information, *Biochemistry 31*, 6865−6870.
3. Azriel, R., and Gazit, E. (2001) Analysis of the minimal amyloid-forming fragment of the islet amyloid polypeptide, *J. Biol. Chem. 276*, 34156−34161.
4. Lansbury, P. T. (1996) A reductionist view of Alzheimer's disease, *Acc. Chem. Res.* 29, 317−321.
5. Kelly, J. W. (1996) Alternative conformations of amyloidogenic proteins govern their behavior, *Curr. Opin. Struct. Biol. 6*, 11−17.
6. Prusiner, S. B. (1997) Prion diseases and the BSE crisis, *Science 278*, 245−250.
7. Kelly, J. W. (1997) Amyloid fibril formation and protein misassembly: a structural quest for insights into amyloid and prion diseases, *Structure 5*, 595−600.
8. Kirschner, D. A., and Inouye, H. (2005) Alzheimer's β-amyloid: Insights into fibril formation and structure from Congo red binding, in *Subcellular Biochemistry* (Harris, R., and Fahrenholz, F., Eds.), Vol. 38, pp 203−224, Kluwer Academic/Plenum, New York.
9. Sipe, J. D., and Cohen, A. S. (2000) Review: history of the amyloid fibril, *J. Struct. Biol. 130*, 88−98.
10. Krebs, M. R. H., Bromley, E. H. C., and Donald, A. M. (2005) The binding of thioflavin-T to amyloid fibrils: localization and implications, *J. Struct. Biol. 149*, 30−37.
11. Sunde, M., and Blake, C. C. F. (1997) The structure of amyloid fibrils by electron microscopy and X-ray diffraction, *Adv. Protein Chem. 50*, 123−159.
12. LeVine, H., III (1993) Thioflavine T interaction with synthetic Alzheimer's disease β-amyloid peptides: detection of amyloid aggregation in solution, *Protein Sci. 2*, 404−410.
13. Zurdo, J., Guijarro, J. I., Jimenez, J. L., Saibil, H. R., and Dobson, C. M. (2001) Dependence on solution conditions of aggregation and amyloid formation by an SH3 domain, *J. Mol. Biol. 311*, 325−340.
14. MacPhee, C. E. and Dobson, C. M. (2000) Chemical dissection and reassembly of amyloid fibrils formed by a peptide fragment of transthyretin, *J. Mol. Biol. 297*, 1203−1215.
15. Goux, W. J. (2002) The conformations of filamentous and soluble tau associated with Alzheimer paired helical filaments, *Biochemistry 41*, 13798−13806.
16. Goux, W. J., Kopplin, L., Nguyen, A. D., Leak, K., Rutkofsky, M., Shanmuganandam, V. D., Sharma, D., Inouye, H., and Kirschner, D. A. (2004) The formation of straight and twisted filaments from short tau peptides, *J. Biol. Chem. 279*, 26868−26875.
17. Kad, N. M., Myers, S. L., Smith, D. P., Smith, D. A., Radford, S. E., and Thomson, N. H. (2003) Hierarchical assembly of β2-microglobulin amyloid *in vitro* revealed by atomic force microscopy, *J. Mol. Biol. 330*, 785−797.

18. Lashuel, H. A., Hartley, D. M., Petre, B. M., Wall, J. S., Simon, M. N., Walz, T., and Lansbury, P. T., Jr. (2003) Mixtures of wild-type and a pathogenic (E22G) form of $A^{\beta}40$ *in vitro* accumulate protofibrils, including amyloid pores, *J. Mol. Biol. 332*, 795−808.
19. Hoshi, M., Sato, M., Matsumoto, S., Noguchi, A., Yasutake, K., Yoshida, N. and Sato, K. (2003) Spherical aggregates of β-amyloid (amylospheroid) show high neurotoxicity and activate tau protein kinase I/glycogen synthase kinase-3β, *Proc. Natl. Acad. Sci. U.S.A. 100*, 6370−6375.
20. Ishimaru, D., Andrade, L. R., Teixeira, L. S. P., Quesado, P. A., Maiolino, L. M., Lopez, P. M., Cordeiro, Y., Costa, L. T., Heckl, W. M., Weissmuller, G., Foguel, D., and Silva, J. L. (2003) Fibrillar aggregates of the tumor suppressor p53 core domain, *Biochemistry 42*, 9022−9027.
21. Rochet, J. C., Lansbury, P. T., Jr. (2000) Amyloid fibrillogenesis: themes and variations, *Curr. Opin. Struct. Biol. 10*, 60−68.
22. Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L, Zurdo, J., Taddei, N., Ramponi, G., Dobson, C. M., and Stefani, M. (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases, *Nature 416*, 507−511.
23. Quist, A., Doudevski, I., Lin, H., Azimova, R., Ng, D., Frangione, B., Kagan, B., Ghiso, J., and Lal, R. (2005) Amyloid ion channels: a common structural link for protein-misfolding disease, *Proc. Natl. Acad. Sci. U.S.A. 102*, 10427−10432.
24. Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D. and Dobson, C. M. (1998) Amyloid fibril formation by an SH3 domain, *Proc. Natl. Acad. Sci. U.S.A. 95*, 4224−4228.
25. Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. N., Dobson, C. M., Radford, S. E., Blake, C. C. R., and Pepys, M. B. (1997) Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis, *Nature 385*, 787−793.
26. Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani, M., Ramponi, G., and Dobson, C. M. (2002) Kinetic partitioning of protein folding and aggregation, *Nat. Struct. Biol. 9*, 137−143.
27. Liu, R., McAllister, C., Lyubchenko, Y., and Sierks, M. R. (2004) Residues 17−20 and 30−35 of beta-amyloid play critical roles in aggregation, *J. Neurosci. Res. 75*, 162−171.
28. Ramirez-Alvarado, M., Merkel, J. S., and Regan, L. (2000) A systematic exploration of the influence of the protein stability on amyloid formation *in vitro*, *Proc. Natl. Acad. Sci. U.S.A. 97*, 8979−8984.
29. Smith, D. P., Jones, S., Serpell, L. C., Sunde, M., and Radford, S. E. (2003) A systematic investigation into the effect of protein destabilization on beta 2-microglobulin amyloid formation, *J. Mol. Biol. 330*, 943−954
30. Dobson, C. M. (1999) Protein misfolding, evolution and disease, *Trends Biochem. Sci. 24*, 329−332.
31. Nilsson, M. R., and Dobson, C. M. (2003) In vitro characterization of lactoferrin aggregation and amyloid formation, *Biochemistry 42*, 375−382.
32. Cottingham, M. G., Hollinshead, M. S., and Vaux, D. J. T. (2002) Amyloid fibril formation by a synthetic peptide from a region of human acetylcholinesterase that is homologous to the Alzheimer's amyloid-β peptide, *Biochemistry 41*, 13539−13547.
33. Pawar, A. P., DuBay, K. F., Zurdo, J., Chiti, F., Vendruscolo, M., and Dobson, C. M. (2005) Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases, *J. Mol. Biol. 350*, 379−392.
34. DuBay, K. F., Pawar, A. P., Chiti, F., Zurdo, J., Dobson, C. M., and Vendruscolo, M. (2004) Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains, *J. Mol. Biol. 341*, 1317−1326.
35. von Bergen, M., Friedhoff, P., Biernat, J., Heberle, J., Mandelkow, E. M., and Mandelkow, E. (2000) Assembly of τ protein into Alzheimer paired helical filaments depends on a local sequence motif (^{306}VQIVYK^{311}) forming β structure. *Proc. Natl. Acad. Sci. U.S.A. 97*, 5129−5134.
36. Gamblin, T. C. (2005) Potential structure/function relationships of predicted secondary structural elements of tau, *Biochim. Biophys. Acta 1739*, 140−149.
37. Reches, M., Porat, Y., and Gazit, E. (2002) Amyloid fibril formation by pentapeptide and tetrapeptide fragments of human calcitonin, *J. Biol. Chem. 277*, 35475−35480.
38. Mazor, Y., Gilead, S., Benhar, I., and Gazit, E. (2002) Identification and characterization of a novel molecular-recognition and self-assembly domain within the islet amyloid polypeptide, *J. Mol. Biol. 322*, 1013−1024.
39. Cherney, I., Rockah, L., Levy-Nissenbaum, O., Gophna, U., Ron, E. Z., Gazit, E.. (2005) The formation of *Escherichia coli* curli amyloid fibrils is mediated by prion-like peptide repeats, *J. Mol. Biol. 352*, 245−252.
40. Tjernberg, L., Hosia, W., Bark, N., Thyberg, J., and Johansson, J. (2002) Charge attraction and β propensity are necessary for amyloid fibril formation from tetrapeptide, *J. Biol. Chem. 277*, 43243−43246.
41. Porat, Y., Mazor, Y., Efrat, S., Gazit, E. (2004) Inhibition of islet amyloid polypeptide fibril formation: a potential role for heteroaromatic interactions, *Biochemistry 43*, 14454−14462.
42. Scrocchi, L. A., Ha, K., Chen, Y., Wu, L., Wang, F., and Fraser, P. E. (2003) Identification of minimal peptide sequences in the (8−20) domain of human islet amyloid polypeptide involved in fibrillogenesis, *J. Struct. Biol. 141*, 218−227.
43. Ventura, S., Zurdo, J. Narayanan, S., Parreno, M., Mangues, R., Reif, B., Chiti, F., Giannoni, E., Dobson, C. M., Aviles, F. X., and Serrano, L. (2004) Short amino acid stretches can mediate amyloid formation in globular proteins: The Src homology 3 (SH3) case, *Proc. Natl. Acad. Sci. U.S.A. 101*, 7258−7263.
44. Vauthey, S., Santoso, S., Gong, H., Watson, N., and Zhang, S. (2002) Molecular self-assembly of surfactant-like peptides to form nanotubes and nanovesicles, *Proc. Natl. Acad. Sci. U.S.A. 99*, 5355−5360.
45. Caplan, M. R., Moore, P. N., Zhang, S., Kamm, R. D., and Lauffenburger, D. A. (2000) Self-assembly of a β-sheet protein governed by relief of electrostatic repulsion relative to van der Waals attraction, *Biomacromolecules 1*, 627−631.
46. Kammerer, R. A., Kostrewa, D., Zurdo, J., Detken, A., Garcia-Echeverria, C., Green, J. D., Muller, S. A., Meier, B. H., Winkler, F. K., Dobson, C. M., and Steinmetz, M. O. (2004) Exploring amyloid formation by a *de novo* design, *Proc. Natl. Acad. Sci. U.S.A. 101*, 4435−4440.
47. West, M. W., Wang, W., Patterson, J., Mancias, J. D., Beasley, J. R., and Hect, M. H. (1999) *De novo* amyloid proteins from designed combinatorial libraries, *Proc. Natl. Acad. Sci. U.S.A. 96*, 11211−11216.
48. Zhang, S., and Rich, A. (1997) Direct conversion of an oligopeptide from a beta-sheet to an alpha-helix: a model for amyloid formation, *Proc. Natl. Acad. Sci. U.S.A. 94*, 23−28.
49. Blondelle, S. E., Forood, B., Houghten, R. A., and Perez-Paya, E. (1997) Polyalanine-based peptides as models for self-associated β-pleated-sheet, *Biochemistry 36*, 8393−8400.
50. Soreghan, B., Kosmoski, J., and Glabe, C. (1994) Surfactant properties of Alzheimer's A beta peptides and the mechanism of amyloid aggregation, *J. Biol. Chem. 269*, 28551−28554.
51. Lopez de la Paz, M., Lacroix, E., Ramirez-Alverado, M., and Serrano, L. (2001) Computer-aided design of β-sheet peptides, *J. Mol. Biol. 312*, 229−246.
52. Lopez de la Paz, M., Goldie, K., Zurdo, J., Lacroix, E., Dobson, C. M., Hoenger, A., and Serrano, L. (2002) *De novo* designed peptide-based amyloid fibrils, *Proc. Natl. Acad. Sci. U.S.A. 99*, 16052−16057.
53. Lopez de la Paz, M., and Serrano, L. (2004) Sequence determinants of amyloid fibril formation, *Proc. Natl. Acad. Sci. U.S.A. 101*, 87−92.
54. Pastor, M. T., Lopez de la Paz, M., Lacroix, E., Serrano, L., and Perez-Paya, E. (2002) Combinatorial approaches: A new tool to search for highly structured β-hairpin peptides, *Proc. Natl. Acad. Sci. U.S.A. 99*, 614−619.
55. Kosik, K. S., and Greenberg, S. M. (1994) Tau proteins and Alzheimer disease, in *Alzheimer Disease* (Terry, R. D., Katzman, R., and Bick, K. L., Eds.), pp 335−344, Raven Press, New York.
56. von Bergen, M., Sarghorn, S., Biernat, J., Mandelkow, E. M., Mandelkow, E. (2005) Tau aggregation is driven by a transition from random coil to β-sheet structure, *Biochim. Biophys. Acta 1739*, 158−166.
57. Mandelkow, E. M., Schweers, O., Drewes, G., Biernat, J., Gustke, N., Trinczek, B., and Mandelkow, E. (1996) Structure, microtubule interactions, and phosphorylation of tau protein, *Ann. N.Y. Acad. Sci. 777*, 96−106.
58. Friedhoff, P., von Bergen, M., Mandelkow, E. M., and Mandelkow E. (2000) Structure of tau protein and assembly into paired helical filaments, *Biochim. Biophys. Acta 1502*, 122−132.
59. Mandelkow, E. M., Mandelkow, E. (1998) Tau in Alzheimer's disease, *Trends Cell Biol. 8*, 425−427.

60. Feany, M. B., and Dickson, D. W. (1996) Neurodegenerative disorders with extensive tau pathology: A comparative study and review, *Ann. Neurol. 40*, 139−148.
61. Dickson, D. W., Crystal, H. A., Bevona, C., Honer, W., Vincent, I., and Davies, P. (1995) Correlations of synaptic and pathological markers with cognition of the elderly, *Neurobiol. Aging 16*, 285−298.
62. von Bergen, M., Barghorn, S., Li, L., Marx, A., Biernat, J., Mandelkow, E.-M., and Mandelkow, E. (2001) Mutations of tau protein in frontotemporal dementia promote aggregation of paired helical filaments by enhancing local β-structure, *J. Biol. Chem. 276*, 48165−48174.
63. Kirschner, D. A., Abraham, C., and Selkoe, D. J. (1986) X-ray diffraction from intraneuronal paired helical filaments and extraneuronal amyloid fibers in Alzheimer disease indicates cross-β conformation, *Proc. Natl. Acad. Sci. U.S.A. 83*, 503−507.
64. Berriman, J., Serpell, L. C., Oberg, K. A., Fink, A. L., Goedert, M., and Crowther, R. A. (2003) Tau filaments from human brain and from in vitro assembly of recombinant protein show cross-beta structure, *Proc. Natl. Acad. Sci. U.S.A. 100*, 9034−9038.
65. Abraha, A., Ghoshal, N., Gamblin, T. C., Cryns, V., Berry, R. W., Kuret, J., and Binder, L. I. (2000) C-terminal inhibition of tau assembly in vitro and in Alzheimer's disease, *J. Cell Sci. 113*, 3737−3745.
66. Jonsson, J., Eriksson, L., Hellberg, S., Sjöström, M., and Wold, S. (1989) Multivariate parametrization of 55 coded and noncoded amino acids, *Quant. Struct.-Act. Relat. 8*, 204−209.
67. Hellberg, S., Sjöström, M., Skagerberg, B., and Wold, S. (1986) Peptide quantitative structure−activity relationships, a multivariate approach, *J. Med. Chem. 30*, 1126−1135.
68. Sandberg, M., Eriksson, L., Jonsson, J., Sjöström, M., and Wold, S. (1998) New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids, *J. Med. Chem. 41*, 2481−2491.
69. Wold, S., Sjöström, M., and Eriksson, L. (2001) PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst. 58*, 109−130.
70. Choma, C. T., Lear, J. D., Nelson, M. J., Dutton, P. L., Robertson, D. E., and DeGrado, W. F. (1994) Design of a heme-binding four-helix bundle, *J. Am. Chem. Soc. 116*, 856−865.
71. Ostresh, J. M., Winkle, J. H., Hamashin, V. T., Houghten, R. A. (1994) Peptide libraries: determination of relative reaction rates of protected amino acids in competitive couplings, *Biopolymers 34*, 1681−1689.
72. Haris, P. I., Chapman, D. (1995) The conformational-analysis of peptides using Fourier transform IR spectroscopy, *Biopolymers 37*, 251−263.
73. Bandekar, J. (1992) Amide modes and protein conformation, *Biochim. Biophys. Acta 1120*, 123−143.
74. Venyaminov, S. Y., Kalnin, N. N. (1990) Quantitative IR spectrophotometry of peptide compounds in water (H_2O) solutions. 2. Amide absorption-bands of polypeptides and fibrous proteins in alpha-coil, beta-coil, and random coil conformations, *Biopolymers 30*, 1259−1271.
75. Zurdo, J., Guijarro, J. I., and Dobson, C. M. (2001) Preparation and characterization of purified amyloid fibrils, *J. Am. Chem. Soc. 123*, 8141−8142.
76. Edelhock, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins, *Biochemistry 6*, 1948−1954.
77. Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data, *Anal. Biochem. 182*, 319−326.
78. Khurana, R., Uversky, V. N., Nielsen, L., and Fink, A. L. (2001) Is Congo red an amyloid-specific dye? *J. Biol. Chem. 276*, 22715−22721.
79. Klunk, W. E., Jacob, R. F., and Mason, R. P. (1999) Quantifying amyloid β-peptide (Aβ) aggregation using the Congo red-Aβ (CR-Aβ) spectroscopic assay, *Anal. Biochem. 266*, 66−76.
80. Friedhoff, P., Schneider, A., Mandelkow, E. M., and Mandelkow, E. (1998) Rapid assembly of Alzheimer-like paired helical filaments from microtubule-associated protein tau monitored by fluorescence in solution, *Biochemistry* 37, 10223−10230.
81. Friedhoff, P., von Bergen, M., Mandelkow, E. M., and Mandelkow, E. (1998) A nucleated assembly mechanism of Alzheimer paired helical filaments, *Proc. Natl. Acad. Sci. U.S.A.* 95, 15712−15717.
82. LeVine, H., III (1999) Quantification of beta-sheet amyloid fibril structures with thioflavin T, *Methods Enzymol. 309*, 274−284.
83. Necula, M., and Kuret, J. (2004) Pseudophosphorylation and glycation of tau protein enhance but do not trigger fibrillization *in vitro*, *J. Biol. Chem. 279*, 49694−49703.
84. Pace, C. N., Shirley, B. A., and Thomson, J. A. (1989) in *Protein Structure: A Practical Approach* (Creighton, T. E., Ed.), pp 311−330, Oxford Press, New York.
85. Kennelly, P. J., Timkovich, R., and Cusanovich, M. A. (1981) Lysine to ^{13}C-labeled homoarginine conversion in microbial cytochromes c: electron transport rates and ^{13}C nuclear magnetic resonance spectroscopy, *J. Mol. Biol. 145*, 583−602.
86. Brancia, F. L., Montgomery, H., Tanaka, K., and Kumashiro, S. (2004) Guanidino labeling derivatization strategy for global characterization of peptide mixtures by liquid chromatography matrix-assisted laser desorption/ionization mass spectrometry, *Anal. Chem. 76*, 2748−2755.
87. Sharaf, M. A., Illman, D. L., and Kowalski, B. R. (1986) Chemometrics, in *Chemical Analysis: A series of monographs on Analtyical Chemistry and its Applications*, (Elving, P. J., Winefordner, J. D., Kolthoff, I. M., Eds.), Vol. 82, pp 179−295, John Wiley & Sons, New York.
88. Sreerama, N., Venyaminov, S. Y., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis, *Anal. Biochem. 287*, 243−251.
89. Sreerama, N., Venyaminov, S. Y., and Woody, R. W. (2001) Analysis of protein circular dichroism spectra based on the tertiary structure classification, *Anal. Biochem. 299*, 271−274.
90. Woody, R. W. (1978) Aromatic side-chain contributions to the far ultraviolet circular dichroism of peptides and proteins, *Biopolymers 17*, 1451−1467.
91. Goux, W. J., and Hooker, T. M., Jr. (1980) Chiroptical properties of proteins. 1. Near-ultraviolet circular dichroism of ribonuclease S, *J. Am. Chem. Soc. 102*, 7080−7087.
92. Hirota-Nakaoka, N., Hasegawa, K., Naiki, H., and Goto, Y. (2003) Dissolution of beta2-microglobulin amyloid fibrils by dimethylsulfoxide, *J. Biochem. (Tokyo) 134*, 159−164.
93. Higham, C. E., Jaikaran, E. T., Fraser, P. E., Gross, M., and Clark, A. (2000) Preparation of synthetic human islet amyloid polypeptide (IAPP) in a stable conformation to enable study of conversion to amyloid-like fibrils, *FEBS Lett. 17*, 55−60.
94. Kayed, R., Bernhagen, J., Green field, N., Sweimeh, K., Brunner, H., Voelter, W., and Sapurniotu, A. (1999) Conformational transitions of islet amyloid polypeptide (IAPP) in amyloid formation *in vitro*, *J. Mol. Biol. 287*, 781−796.
95. Mathis, C. A., Wang, Y. M., Holt, D. P., Huang, G. F., Debnath, M. L., and Klunk, W. E. (2003) The synthesis and evaluation of C-11-labeled-6-substituted 2-arylbenzothiazoles as amyloid imaging agents, *J. Med. Chem. 46*, 2740−2754.
96. Chou, P. Y., and Fasman, G. D. (1978) Prediction of the secondary structure of proteins from their amino acid sequence, *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45−148.
97. Kim, C. A., and Berg, J. M. (1993) Thermodynamic β-sheet propensities measured using a zinc-finger host peptide, *Nature 362*, 267−270.
98. Petkova, A. T., Leapman, R. D., Guo, Z., Yau, W. M., Mattson, M. P., and Tycko, R. (2005) Self-propagating, molecular-level polymorphism in Alzheimer's beta-amyloid fibrils, *Science 307*, 262−265.
99. Wood, S. J., Setzel, Martin, J. D., and Hurle, M. R. (1995) Prolines and amyloidogenicity in fragments of the Alzheimer's peptide β/A4, *Biochemistry 34*, 724−730.
100. Barghorn, S., Zheng-Fischhofer, Q., Ackmann, M., Biernat, J., von Bergen, M., Mandelkow, E. M., and Mandelkow, E. (2000) Structure, microtubule interactions, and paired helical filament aggregation by tau mutants of frontotemporal dementias, *Biochemistry 39*, 11714−11721.
101. Tartaglia, G. G., Cavalli, A., Riccardo, P., and Caflisch, A. (2004) The role of aromaticity, exposed surface, and dipole moment in determining protein aggregation rates, *Protein Sci. 13*, 1939−1941.
102. Goedert., M., Wischik, C. M., Crowther, R. A., Walker, J. E., and Klug, A. (1988) Cloning and sequencing of the cDNA encoding a core protein of the paired helical filament of Alzheimer disease: Identification as the microtubule-associated protein tau, *Proc. Natl. Acad. Sci. U.S.A. 85*, 4051−4055.
103. Wischik, C. M., Novak, M., Thøgersen, H. C., Edwards, P. C., Runswick, M. J., Jakes, R., Walker, J. E., Milstein, C., Roth, M., and Klug, A. (1988) Isolation of a fragment of tau derived from

the core of the paired helical filament of Alzheimer disease, *Proc. Natl. Acad. Sci. U.S.A. 85*, 4506−4510.
104. Wischik, C. M., Novak, M., Edwards, P. C., Klug, A., Tichelaar, W., and Crowther, R. A. (1988) Structural characterization of the core of the paired helical filament of Alzheimer disease, *Proc. Natl. Acad. Sci. U.S.A. 85*, 4884−4888.
105. Jakes, R., Novak, M., Davison, M., and Wischik, C. M. (1991) Identification of 3- and 4-repeat tau isoforms within the PHF in Alzheimer's disease, *EMBO J.* 10, 2725−2729.
106. Wille, H. Drewes, G., Biernat, J., Mandelkow, E. M., Mandelkow, E. (1992) Alzheimer-like paired helical filaments and antiparallel dimmers formed from microtubule-associated protein tau in vitro, *J. Cell Biol. 118*, 573−584.
107. Mukrasch, M. D., Biernat, J., von Bergen, M., Griesinger, C, Mandelkow, E., and Zweckstetter, M. (2005) Sites of tau important for aggregation populate β-structure and bind to microtubules and polyanions, *J. Biol. Chem. 280*, 24978−24986.
108. Sillen, A., Leroy, A., Wieruszeski, J. M., Loyens, A., Beauvillain, J. C., Buée, L, Landrieu, I., and Lippens, G. (2005) Regions of tau implicated in the paired helical fragment core as defined by NMR, *ChemBioChem 6*, 1849−1856.
109. Yanagawa, H., Chung, S.-H., Ogawa, Y., Sato, K., Shibata-Seki, T., Masai, J., and Ishiguro, K. (1998) Protein Anatomy: C-tail region of human tau protein as a crucial structural element in Alzheimer's paired helical filament formation in vitro, *Biochemistry 37*, 1979−1988.
110. Ghoshal, N., Garcia-Sierra, F., Wuu, J., Leurgans, S., Bennett, D. A., Berry, R. W., and Binder, L. I. (2002) Tau conformational changes correspond to impairments of episodic memory in mild cognitive impairment and Alzheimer's disease, *Exp. Neurol. 177*, 475−493.
111. Jicha, G. A., Bowser, R., Kazam, I. G., and Davies, P. (1997) Alz-50 and MC-1, a new monoclonal antibody raised to paired helical filaments, recognize conformational epitopes on recombinant tau, *J. Neurosci. Res. 48*, 128−132.
112. Jicha, G. A., Berenfeld, B., and Davies, P. (1999) Sequence requirements for formation of conformational variants of tau similar to those found in Alzheimer's disease, *J. Neurosci. Res. 55*, 713−723.
113. Li, L., von Bergen, M., Mandelkow, E. M., and Mandelkow, E. (2002) Sutructure, stability, and aggregation of paired helical filaments from tau protein and FTDP-17 mutants probed by tryptophan scanning mutagenesis, *J. Biol. Chem. 277*, 41390−41400.
114. Gamblin, T. C., Berry, R. W., and Binder, L. I. (2003) Tau polymerization: Role of the amino terminus, *Biochemistry 42*, 2252−2257.
115. King, M. E., Gamblin, T. C., Kuret, J., and Binder, L. I. (2000) Differential assembly of human tau isoforms in the presence of arachidonic acid, *J. Neruochem. 74*, 1749−1757.
116. Goedert, M., Ghetti, B., and Spillantini, M. G. (2000) Tau gene mutations in frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17). Their relevance for understanding the neurodegenerative process, *Ann. N.Y. Acad. Sci. 920*, 74−83.

BI052226Q