Organic & Biomolecular Chemistry

Comparison of Strategies for Non-perturbing Labeling of α-Synuclein to Study Amyloidogenesis

Journal:	*Organic & Biomolecular Chemistry*
Manuscript ID	OB-ART-11-2015-002329.R1
Article Type:	Paper
Date Submitted by the Author:	10-Dec-2015

Complete List of Authors:	Haney, Conor; University of Pennsylvania, Chemistry Wissner, Rebecca; University of Pennsylvania, Chemistry Warner, John; University of Pennsylvania, Chemistry Wang, Yanxin; University of Pennsylvania, Chemistry Ferrie, John; University of Pennsylvania, Chemistry Covell, Dustin; University of Pennsylvania, Perelman School of Medicine, Center for Neurodegenerative Disease Research Karpowicz, Richard; University of Pennsylvania, Perelman School of Medicine, Center for Neurodegenerative Disease Research Lee, Virginia; University of Pennsylvania, Perelman School of Medicine, Center for Neurodegenerative Disease Research Petersson, E. James; University of Pennsylvania, Chemistry

SCHOLARONE™
Manuscripts

Organic & Biomolecular Chemistry

ARTICLE

Comparison of Strategies for Non-perturbing Labeling of α-Synuclein to Study Amyloidogenesis

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Conor M. Haney,[a,†] Rebecca F. Wissner,[a,†] John B. Warner,[a] Yanxin J. Wang,[a] John J. Ferrie,[a] Dustin Covell,[b] Richard J. Karpowicz,[b] Virginia M.-Y. Lee,[b] and E. James Petersson[a,*]

Characterization of the amyloidogenic Parkinson's Disease protein α-synuclein (αS) has proven difficult due to its structural plasticity. Here, we present a number of complementary methods to site-specifically introduce fluorescent probes to examine αS fibril formation and cellular uptake. By using various combinations of conventional Cys modification, amber codon suppression, transferase mediated N-terminal modification, and native chemical ligation, several variants of singly- and doubly-labeled αS were produced. We validated the nonperturbative nature of the label by a combination of in vitro aggregation kinetics measurements and imaging of the resulting fibrils. The labeled αS can then be used to monitor conformational changes during fibril formation or cellular uptake of αS fibrils in models of disease propagation.

Introduction

Our understanding of the dynamics of protein folding, protein-protein interactions, cellular trafficking, and the effects of protein post-translational modifications can be facilitated by labeling proteins with appropriate reporters. Two classes of proteins of particular current interest are intrinsically disordered proteins and amyloidogenic proteins. These proteins, which are refractory to traditional structural characterization techniques, can be labeled with fluorophores so that methods such as Förster resonance energy transfer (FRET) or fluorescence quenching can be used to monitor their conformations *in vitro* and *in vivo*.[1, 2] To do so, one must find a way to selectively introduce multiple site-specific labels that can effectively report on protein dynamics without being disruptive to the native structure of the protein. Here, we describe combinations of genetic, chemoenzymatic, and semi-synthetic means of site-specifically labeling the Parkinson's disease related protein α-synuclein (αS) for studies of protein misfolding and aggregate propagation *in cellulo*.

In order to frame our discussion of how the following labeling techniques may be used together, we must first introduce them individually. The most time-tested method is Cys modification, which is typically done through reaction with an alkyl halide or maleimide label.[3] However, Cys modification will only be site-specific if the protein contains no other Cys residues or if reaction conditions can be used to suppress the reactivity of the undesired sites.

The amber codon suppression method developed by Schultz and co-workers can be used to site-specifically incorporate an unnatural amino acid (Uaa) into a protein at an amber (UAG) stop codon.[4] Efficient incorporation into recombinant proteins is typically accomplished using evolved orthogonal tRNA/aminoacyl tRNA synthetase (aaRS) pairs to generate the requisite Uaa-tRNA for ribosomal incorporation of the Uaa. The Uaa can either feature an intrinsic label such as a fluorophore, or a bioorthogonal reactive handle for further modification such as an azide. The Uaa mutagenesis method has been expanded by Chin, Liu, and others to enable the incorporation of multiple Uaas, albeit with limited yields.[5-12]

Single-residue chemoenzymatic modification can be achieved using the *E. coli* aminoacyl transferase (AaT). AaT is an enzyme that transfers Leu, Phe, Met, or hydrophobic Uaas to proteins containing an N-terminal Lys or Arg.[13] AaT uses either aminoacyl tRNA or a minimalist aminoacyl adenosine a Uaa donor.[14] Importantly, one can control the timing of AaT modification to introduce a second label following co-translational introduction of a primary modification.

Finally, one of the most versatile and general methods for protein modification is native chemical ligation (NCL).[15] Like AaT modification, NCL requires only a single amino acid substitution (Cys) at the site of ligation, and through the use of β-thiol Cys analogs and desulfurization, even this single amino acid motif can be "erased." Although one can conceivably introduce all desired modifications through NCL, due to its labor-intensive nature, it is not always the optimal technique.

[a.] *Department of Chemistry, University of Pennsylvania, 213 South 34th Street, Philadelphia, PA 19104, USA. Email: ejpetersson@sas.upenn.edu; Tel: 1-215-746-2221*
[b.] *Center for Neurodegenerative Disease Research, University of Pennsylvania, 3600 Spruce Street, Philadelphia, PA 19104*
† These authors contributed equally to this work.
Electronic Supplementary Information (ESI) available: Generation of plasmids, procedures and characterization for the synthesis of labeling reagents, MALDI MS and HPLC analysis of protein labeling, data label validation experiments, fluorescence measurements during aggregation, and additional cellular data. See DOI: 10.1039/x0xx00000x

Please do not adjust margins

The ideal synthesis of a site-specific, multiply labeled protein is likely to involve a judicious combination of the techniques outlined above. Cys modification, amber codon suppression, AaT modification, and NCL are techniques that can be combined for protein multifunctionalization using Uaas. While we stress these are general techniques, we will demonstrate all of them in syntheses of αS.

αS is an intrinsically disordered 140 amino acid protein, aggregates of which are among the hallmarks of Parkinson's disease.[16] Despite considerable effort, there is still little understanding of the aggregation process of αS or its structure within fibrils. Prior NMR and EPR studies have attempted to define the fibril core of αS, but have not achieved sufficient resolution to arrive at a consensus tertiary structure.[17-20] Recent work employing cryo-electron diffraction has provided insight into the fibril structure of short peptides derived from the αS core, though it remains to be seen if this method can be employed to define full fibril structures.[21]

Fluorescence spectroscopy has previously been employed to study the misfolding and fibrillization of αS *in vitro* and *in vivo*. Macromolecular labeling has relied on the corporation of fluorescent protein fusions or immuno-tags to study fibrillization *in vitro*.[22] Imaging of fibrillization in living cells was accomplished by inserting the CysCysProGlyCysCys motif at the C-terminus of αS and labeling using biarsenical fluorophore reagents (i.e., FlAsH).[23] However, these techniques require significant alteration of the αS sequence, which can be disruptive to the folding/misfolding processes of αS. As a result, our laboratory and others have focused on αS labeling through modification of single amino acid residues.

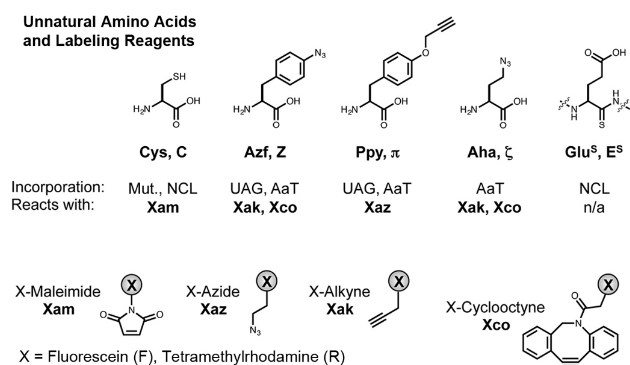

Fig. 1 Methods for Amino Acid Incorporation and Modification. Left: Cysteine (Cys, C), azidophenylalanine (Azf, Z), propargyl tyrosine (Ppy, π), azidohomoalanine (Aha, ζ), and thioglutamate (Glu^S, E^S) are incorporated by at least one of the following methods: simple mutation (Mut.), native chemical ligation (NCL), amber suppression (UAG), or aminoacyl transferase (AaT) and may be modified with the reagents shown.

Cys-maleimide fluorophore labeling was combined with Trp fluorescence to generate a FRET pair to probe αS conformational dynamics just prior to aggregation.[24] Other groups have prepared doubly labeled αS through stochastic labeling of two Cys residues, leading to product mixtures which can be difficult to separate.[25, 26] Using a combination of Cys labeling and pH-dependent N-terminal α-amine labeling, it was shown that a conformational change precedes fibril formation and that fibril nucleation is concentration dependent.[27] Lashuel, Brik, and Pratt have used both semi-synthesis and total synthesis through NCL to incorporate authentic post-translational modifications at many sites in αS, including ubiquitination and phosphorylation.[28-30] Pielak and coworkers have used amber suppression to incorporate ^{19}F NMR probes into αS for folding studies.[31]

Our laboratory has focused on the development of methods to incorporate minimally perturbing probes into αS. In prior work, we have demonstrated the co-translational incorporation of spectroscopic probes by amber codon suppression including *p*-cyanophenylalanine (Cnf) and acridonylalanine.[32, 33] Double labels, such as the fluorophore/quencher pair of fluorescein and a thioamide have been incorporated by semisynthetic means.[33] We have also combined NCL and amber codon suppression to attach Cnf/thioamide pairs.[32] We have further combined these methods with AaT modification to enable the traceless incorporation of N-terminal modifications into αS.[32] While these combinations of methods have allowed us to multiply label αS, due to the labor-intensive nature of the peptide synthesis in NCL, we have continued to examine other combinations of techniques that can be used in a straightforward manner to generate large quantities of protein for *in vitro* and cellular aggregation studies. Here, we compare combinations of a number of labeling methods to generate homogeneous, multiply labeled versions of αS.

Results and Discussion

In using combinations of labeling methods, it is important to consider whether any given pair of methods satisfies the criteria for orthogonality and for sequential use. We deem methods to be orthogonal if one type of reaction can be carried out in the presence of a functional group for the other type of reaction; we do not require that they be used in one pot. Some reactions are inherently sequential, such as the co-translational incorporation of an intrinsic spectroscopic probe and the post-translational attachment of a second probe. Other pairs of reactions could be carried out in either order, but we have identified a preferred sequence. Techniques requiring click chemistry modification and Cys modification fall into this category. Finally, some combinations are only practical in a certain order. We have generated at least one example of every labeling combination, and for several examples, we have explored multiple strategies. Of course, before generating any doubly labeled constructs, we assessed the corresponding singly labeled constructs. Purified, labeled proteins were characterized by whole protein matrix-assisted laser desorption ionization mass spectrometry (MALDI MS) as well as trypsin digestion followed by MALDI MS to confirm the extent and specificity of labeling.

Single Modifications.

Cys Modification. While several types of reactions are available for Cys modification, including radical-mediated thiol-

This journal is © The Royal Society of Chemistry 20xx

ene coupling, photochemically-controlled reactions, and elimination/addition chemistry; simple S_N2 or Michael additions are by far the most common Cys reactions, with a great number of reagents commercially available.[3] In addition, we expected either reaction with the Cys nucleophile to be compatible with the other labeling reactions considered here. Since αS has no endogenous Cys residues, we easily generated two singly modified αS constructs by expressing αS-C_9 (**1a**) and αS-C_{114} (**2a**), labeled them in a semi-pure state with fluorescein-maleimide (Fam) to generate αS-C^{Fam}_9 (**1b**) and αS-C^{Fam}_{114} (**2b**), and purified the products by high performance liquid chromatography (HPLC). Analyses of the crude labeling reactions indicated that the yields of αS-C^{Fam}_9 (**1b**) and αS-C^{Fam}_{114} (**2b**) were quantitative (see ESI). While MS analysis is not strictly quantitative without the use of isotopic standards, the simplicity of this method makes it ideal for *in situ* analysis of reaction progress. After purification, whole protein and trypsin digestion MALDI MS confirmed complete and site-specific labeling in the products (Fig. 2). Trypsin digestion MALDI MS also revealed that ring-opening hydrolysis of the maleimide unit (corresponding to a mass increase of + 18 Da in the observed fragment) occurred during the labeling and purification procedures.

Uaa Mutagenesis. For Uaa modification, we have previously made use of two types of Uaas: those with bioorthogonal reactive handles (i.e., azides and alkynes) and intrinsic probes such as the fluorescent amino acid acridonylalanine.[33, 34] While intrinisic probes are valuable, they are often limited in their functionality by their need to remain compatible with the translational machinery. Thus, several RS/tRNA pairs derived from the *Methanococcus jannaschii* (*Mj*) TyrRS/tRNA have been selected for Uaas amenable to post-translational modification through click chemistry reactions, including pXFRS for both azidophenylalanine (Azf, Z) and propargyltyrosine (Ppy, π). Selective post-translational reactions with Azf and Ppy can be used to derivatize them with virtually any type of label. Using an αS plasmid containing a TAG codon at position 94 with the pDULE2-pXF plasmid to incorporate Azf, αS-Z_{94} (**3a**) was expressed and labeled with fluorescein-dibenzocyclooctyne (Fco) to give αS-Z^{Fco}_{94} (**3b**). MALDI MS analysis of the crude labeling mixture showed that the labeling yield was incomplete (data not shown). Aryl azides such as Azf are subject to both chemical reduction and photodegradation that could prevent complete labeling.[35] Although light exposure is easily limited by working in the semi-dark, some Azf reduction can occur in cells or during handling of the proteins. Unfortunately, Azf-containing peptide fragments are ionized to *p*-aminophenylalanine species during MALDI MS analysis, so it is difficult to directly determine the extent of Azf reduction, and we infer it from the yields of reactions with alkynes. Suspecting that the problem lay in the kinetics of the cyclooctyne reaction, we subjected αS-Z_{94} (**3a**) to Cu-catalyzed cycloaddition reactions with fluorescein-alkyne (Fak) to give αS-Z^{Fak}_{94} (**3c**). We observed slightly higher modification yields (by MALDI MS) than for strain-promoted cycloaddition with Fco, but the labeling reaction did not go to completion (data not shown).

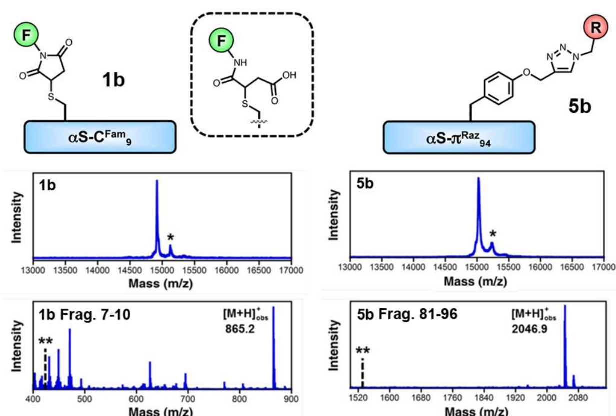

Fig. 2 Single Labeling of αS Using Cys Modification or Uaa Mutagenesis. Left: Whole protein and trypsin digestion αS-C^{Fam}_9 (**1b**) MALDI MS demonstrate labelling and subsequent maleimide hydrolysis (inset). MALDI MS data show that the purified protein is homogeneously labeled. Calcd m/z for whole protein with maleimide hydrolysis $[M+H]^+$: 14924, Obs 14925. Calcd m/z for trypsin fragment 7-10 $[M+H]^+$: 865.3, Obs 865.2. Right: Whole protein and trypsin digestion αS-π^{Raz}_{94} (**5b**) MALDI MS data show that the purified protein is homogeneously labeled. Calcd m/z for whole protein $[M+H]^+$: 15029, Obs 15026. Calcd m/z for trypsin fragment $[M+H]^+$ 81-96: 2047.0, Obs 2046.9. The asterisk (*) indicates a matrix adduct observed in all whole protein MALDI spectra of αS (+207 Da). The double asterisk (**) corresponds to the expected mass of each unmodified trypsin fragment.

Unlike azides, which can be reacted with alkynes through either strain-promoted cyclizations or Cu-catalyzed cyclizations, terminal alkynes like Ppy must be reacted with azides using Cu catalysis. We expressed αS-π_{39} (**4a**) and αS-π_{94} (**5a**) using the αS plasmids and the pDULE2-pXF plasmid, and derivatized the Ppy residues with tetramethylrhodamine-azide (Raz) to give αS-π^{Raz}_{39} (**4b**) and αS-π^{Raz}_{94} (**5b**, Fig. 2). When the proteins were analyzed by MALDI MS to confirm the extent and specificity of labeling, we observed near quantitative labeling at both sites (see ESI). We believe that the lower labeling efficiency of Azf relative to Ppy under otherwise identical Cu-catalyzed conditions is the result of reduction of some of the Azf residues.

AaT Modification. We have previously shown that AaT is capable of modifying protein N-termini with Uaas using semi-synthetic aminoacyl adenosine donors.[14] Our laboratory and others have also used RSs to generate aminoacyl tRNA substrates for AaT *in situ*.[14, 36, 37] An αS construct with an N-terminal Lys was generated by inserting a 10 His tag with a Factor Xa proteolysis site at the protein N-terminus, deleting Met1, and making the mutation Asp2Lys (His_{10}-α$S_{2\text{-}140}$-K_2, **S1a**). Following Ni-NTA affinity purification, the His_{10} tag was cleaved with Factor Xa and the N-terminal Lys was revealed. This construct, α$S_{2\text{-}140}$-K_2 (**S1b**), could then be modified with a variety of Uaas using an RS/AaT protocol, including Azf, Ppy, or azidohomoalanine (Aha, ζ). For example, Aha was transferred to α$S_{2\text{-}140}$-K_2 (**S1b**) and subsequently modified with Fco or dibenzylcyclooctyne tetramethylrhodamine (Rco) using strain-promoted azide-alkyne cycloaddition to give α$S_{2\text{-}140}$-$K_2\zeta^{Fco}_N$ (**S1d**) or α$S_{2\text{-}140}$-$K_2\zeta^{Rco}_N$ (**S1e**), respectively (see ESI).

Multiple Modifications.

NCL then Cys Modification. We have previously shown that we can label αS with spectroscopic probes at either the C-terminus or the N-terminus through NCL. While there are a great many modifications possible, here we restrict ourselves to thioamide fluorescence quenching probes since there is no equivalent way of installing these backbone modifications through any of the other labeling methods. We examined two modifications, Asp^S2 and Glu^S137, where we denote the presence of the thioamide by adding the superscript "S" to the one or three letter code for the corresponding natural amino acid (Note that we have previously used a prime symbol, but we have recently changed this convention to avoid confusion with prime symbols in protease nomenclature). To insert the Asp^S2 modification, we expressed the truncated construct His_{10}-$αS_{9\text{-}140}$-C_9 (**6a**), cleaved the N-terminal His_{10} tag with Factor Xa, and ligated the resulting $αS_{9\text{-}140}$-C_9 (**6b**) fragment to the synthetic thioester peptide $αS_{1\text{-}8}$-D^S_2-SR (**6c**), where SR is methyl 3-mercaptopropionate, in quantitative yield (MALDI analysis, see ESI). We have previously described insertion of the Glu^S137 modification, where we expressed the intein fusion construct $αS_{1\text{-}113}$-$IntH_6$, cleaved the C-terminal intein with mercaptoethanesulfonate (MESNA), and ligated the resulting thioester fragment $αS_{1\text{-}113}$-SR to the synthetic peptide $αS_{114\text{-}140}$-$C_{114}E^S_{137}$.[33] The crude proteins, αS-$D^S_2C_9$ (**6d**) and αS-$C_{114}E^S_{137}$, were taken forward to the second labeling step where they were incubated with Fam. The doubly labeled proteins αS-$D^S_2C^{Fam}_9$ (**6e**) and αS-$C^{Fam}_{114}E^S_{137}$ were then purified and characterized by MALDI MS.

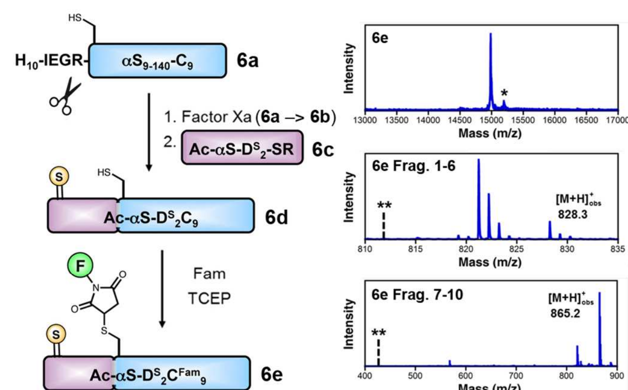

Fig. 3 NCL and Cys Modification to Generate Ac-αS-$D^S_2C^{Fam}_9$ (**6e**). His_{10}-IEGR-$αS_{9\text{-}140}$-C_9 (**6a**) was cleaved using Factor Xa to give $αS_{9\text{-}140}$-C_9 (**6b**), which was subjected to NCL with Ac-αS-D^S_2-SR (**6c**) to generate Ac-αS-$D^S_2C_9$ (**6d**). Following the ligation reaction, Ac-αS-$D^S_2C_9$ (**6d**) was labeled with Fam to yield the doubly-labeled protein Ac-αS-$D^S_2C^{Fam}_9$ (**6e**). Whole protein and trypsin digestion MALDI MS data show that the purified protein is homogeneously labelled (again, with maleimide hydrolysis). Top MALDI MS: Calcd m/z for whole protein $[M+H]^+$: 14980, Obs 14981. Middle MALDI MS: Calcd m/z for trypsin fragment 1-6 $[M+H]^+$: 828.3, Obs 828.3. Bottom MALDI MS: Calcd m/z for trypsin fragment 7-10 $[M+H]^+$: 865.3, Obs 865.2. The asterisk (*) indicates a matrix adduct observed in all whole protein MALDI spectra of αS (+207 Da). The double asterisk (**) corresponds to the expected mass of each unmodified trypsin fragment.

Uaa Mutagenesis then AaT. The examples of amber suppression and AaT modification considered thus far have involved bioorthogonal chemistry. Because one can control the timing of the AaT modification, azide/alkyne cycloadditions can easily be used for both modification steps when combining AaT and amber suppression. To develop procedures for these modifications, we generated two constructs by introducing amber codons into the His_{10}-$αS_{2\text{-}140}$-K_2 construct at either Tyr39 or Phe94. Azf was incorporated during expression in *E. coli* cells using the pDULE2-pXF plasmid, and then the process of Azf labeling with Rco, Factor Xa proteolysis, Aha transfer with AaT, and N-terminal Aha labeling with Fco was performed to produce $αS_{2\text{-}140}$-$K_2ζ^{Fco}_NZ^{Rco}_{39}$ (**S2e**) and $αS_{2\text{-}140}$-$K_2ζ^{Fco}_NZ^{Rco}_{94}$ (**S3e**, see ESI). While we were able to produce doubly labeled protein, since labeling in this way is labor-intensive and limited in scope (to placing one label at the N-terminus), we ultimately chose to focus on combining Cys and Uaa labeling.

Cys Modification and Uaa Mutagenesis. Cys modification and Uaa mutagenesis are perhaps the most versatile and obvious combination of methods for doubly labeling a protein. In spite of this, there are relatively few examples of their use together.[38-40] For both methods, many substitutions are possible. To produce doubly-labeled proteins for fluorescence experiments, we began with Cys modification of a Ppy-labeled protein. TAG mutations were introduced into the Cys mutant plasmids at positions 39 and 94, then αS-$C_9π_{94}$ (**7a**) and αS-$π_{39}C_{114}$ (**8a**) were expressed using the standard amber suppression protocols with pDULE2-pXF. We performed the Cys modification with Fam first, giving αS-$C^{Fam}_9π_{94}$ (**7b**) and αS-$π^{Raz}_{39}C_{114}$ (**8b**), and then the Cu-catalyzed addition of Raz to Ppy. We found that both reactions occurred in high yields to give αS-$C^{Fam}_9π^{Raz}_{94}$ (**7c**) and αS-$π^{Raz}_{39}C^{Fam}_{114}$ (**8c**), respectively (see ESI).

We have recently reported the use of inteins as purification tags for the recombinant expression of proteins containing Uaas.[41] We expressed an αS construct with a His_6-tagged Mxe GyrA intein fusion, αS-$C_9π_{39}$-$MxeH_6$ (**9a**), using standard amber suppression protocols with pDULE2-pXF and isolated the intein-tagged protein from cell lysate by Ni-NTA affinity chromatography. The intein was cleaved by addition of β-mercaptoethanol (BME), and semi-crude αS-$C_9π_{39}$ (**9b**) was then labeled in successive steps with Fam, giving αS-$C^{Fam}_9π_{39}$ (**9c**), followed by Cu-catalyzed Raz labeling of Ppy to give αS-$C^{Fam}_9π^{Raz}_{39}$ (**9d**). We found that both reactions utilizing semi-crude αS-$C_9π_{39}$ occurred in high yields. Here, the intein can be used to improve protein throughput by facilitating isolation of the desired protein and may improve expression yield. We have observed that several Uaa-containing αS constructs are isolated in higher yields when expressed as the C-terminal intein fusion, which may be a result of improved solubility or differences in the expression protocol.[41, 42] Leaving the intein intact for the Fam labeling step can aid in purification after the first labeling step (see ESI). However, inconsistent Raz labeling was observed in the presence of the intein, presumably due to binding of copper to the His residues. Therefore, we advise cleavage with BME prior to labeling the alkyne.

This journal is © The Royal Society of Chemistry 20xx

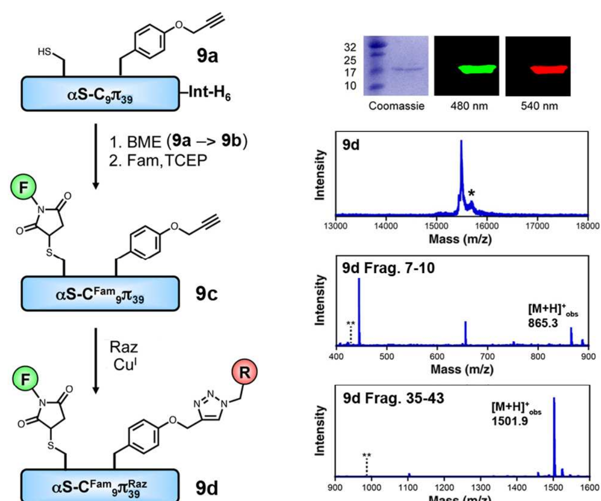

Fig. 4 Uaa Mutagenesis and Cys Modification Using Intein Handle to Generate αS-$C^{Fam}{}_{9}\pi^{Raz}{}_{39}$ (**9d**). Semi-pure αS-$C_9\pi_{39}$ (**9b**) was reacted first with Fam and then with Raz using Cu-catalyzed conditions. Whole protein and trypsin digestion MALDI MS data show that the purified protein is homogeneously labeled. Top MALDI MS: Calcd m/z for whole protein $[M+H]^+$: 15454, Obs 15473. Middle MALDI MS: Calcd m/z for trypsin fragment 7-10 $[M+H]^+$: 865.3, Obs: 865.3. Bottom MALDI MS: Calcd m/z for trypsin fragment 35-43$[M+H]^+$: 1501.7, Obs: 1501.9. The asterisk (*) indicates a matrix adduct observed in all whole protein MALDI spectra of αS (+207 Da). The double asterisk (**) corresponds to the expected mass of each unmodified trypsin fragment.

Assessment and Applications of Labeled Proteins.

Assessment of Label Positions. Making useful multiply labeled proteins involves more than just the ability to label a protein effectively, it is also important to determine whether the locations of those labels are perturbing to native protein folding and function. We have evaluated our labeled constructs using the following metrics: 1) aggregation kinetics as determined by a method that is not sensitive to the presence of the fluorescent label, 2) labeled protein incorporation percentage in fibrils, and 3) fibril morphology as determined by transmission electron microscopy (TEM). We have performed these tests on singly labeled constructs, added at 5% of the total protein concentration in a WT αS background. In all cases, the performance of the labeled construct was quantitatively compared to that of WT αS. For measurements of aggregation kinetics, the extent of fibrillization was determined using a UV/Vis absorbance-based Congo Red (CR) binding assay, which should be insensitive to the presence of our fluorophore (unlike a fluorescence-based thioflavin T assay). In gel fluorescence measurements were made on resolubilized fibrils after aggregation was complete to ensure that the labeled protein was incorporated into fibrils stoichiometrically (see ESI, Fig. S20). For TEM imaging, we assessed the widths and length distributions of mature fibrils (48 h aggregation). Data for αS-$C^{Fam}{}_{9}$ (**1b**), αS-$C^{Fam}{}_{114}$ (**2b**), and αS-$\pi^{Raz}{}_{94}$ (**5b**) are shown in Figure 5 and in ESI (Figs. S33, S34; Table S4). No label position examined dramatically disrupted fibril formation kinetics, incorporation percentage, or morphology.

Since our αS constructs are fluorescently labeled, we were able to record changes in fluorescence during the course of aggregation. For all constructs singly labeled with fluorescein, a decrease in the fluorescence intensity was found to follow the same kinetics as the CR absorbance changes (see ESI). These data imply that our fluorescently-labeled αS constructs were incorporated into coaggregates with the WT αS, and no separate populations of labeled αS fibrils formed. Thus, one could use the labeled αS at a low percentage as a method for monitoring fibril formation, either through steady state or fluorescence polarization measurements.

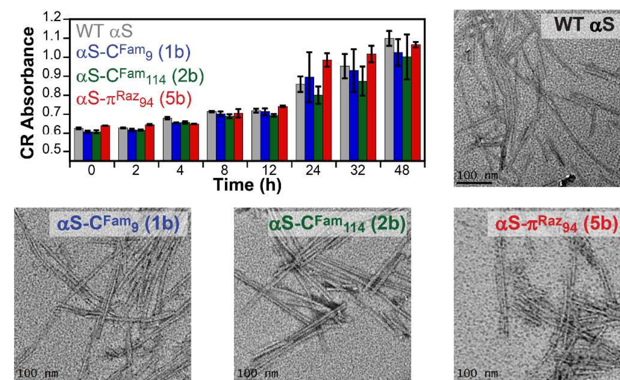

Fig. 5 Aggregation Analysis of Singly Labeled αS Constucts. Fibrils formed from either 100% WT αS or a 5% mixture of αS-$C^{Fam}{}_{9}$ (**1b**), αS-$C^{Fam}{}_{114}$ (**2b**), or αS-$\pi^{Raz}{}_{94}$ (**5b**) with WT αS. Top Left: *In vitro* aggregation kinetics monitored by the ratio of Congo Red (CR) absorbance ratio (540nm/480 nm). Primary data can be found in ESI. Transmission electron micrograph (TEM) images of fibrils (48 h aggregation time) deposited on copper grids and negatively stained with ammonium molybdate.

Applications of Labeled Proteins. Constructs either singly or multiply labeled at validated positions can be used in a number of applications. Of particular interest to our laboratory are experiments in which conformational changes during aggregation are monitored using FRET between pairs of fluorescent labels, and experiments in which singly labeled αS is used to monitor the seeding of aggregation in live cells. FRET experiments can help to elucidate the mechanism by which aggregation occurs, which is in itself a complex protein biophysics problem and may help to guide the development of drugs that block or reverse aggregation. However, FRET experiments performed *in vitro* do not allow one to fully understand the pathology of Parkinson's disease *in vivo*. Recent studies by Lee, Trojanowksi, and coworkers have shown that intrastriatal injection of pre-formed fibrils (PFFs) of αS was sufficient to trigger Parkinson's-like formation of Lewy body aggregates in mice.[43] Moreover, their laboratory has shown that similar incubation with PFFs can be used to initiate aggregation in cultured cells, including mouse neurons.[44, 45] Here, we demonstrate that one can use this method to form coaggregates of site-specifically labeled αS with endogenous mouse αS in order to study the mechanisms of uptake and fibril seeding in the propagation of Parkinson's disease pathology.

We have used FRET measurements to study the change in the distance between residues 9 and 94 upon fibrillization. (Fig. 6) To initiate aggregation for FRET studies, αS proteins were shaken in buffer according to standard protocols. Although we test the levels of label perturbation using 5% labeled protein (above), we carry out FRET studies using 1% labeled protein with 99% unlabeled background to further minimize any chances of cross-talk between labeled proteins. Thus, our FRET measurements should report exclusively on intramolecular conformational changes in αS. In order to properly interpret FRET measurements in complex mixtures, it is essential that one perform certain control experiments using singly labeled constructs. Indeed, the fluorescence of the donor only construct shows that there is substantial quenching of fluorescein emission that is independent of FRET (see ESI Figs. S25, S30). Nonetheless, we are able to correct for this quenching by using time correlated single photon counting (TCSPC) measurements to determine the fluorescence lifetimes of the doubly labeled species (τ) and the donor only species (τ_0). It should be noted that unlike steady state fluorescence, fluorescence lifetimes are independent of concentration, and thus serve as a better measure of FRET.

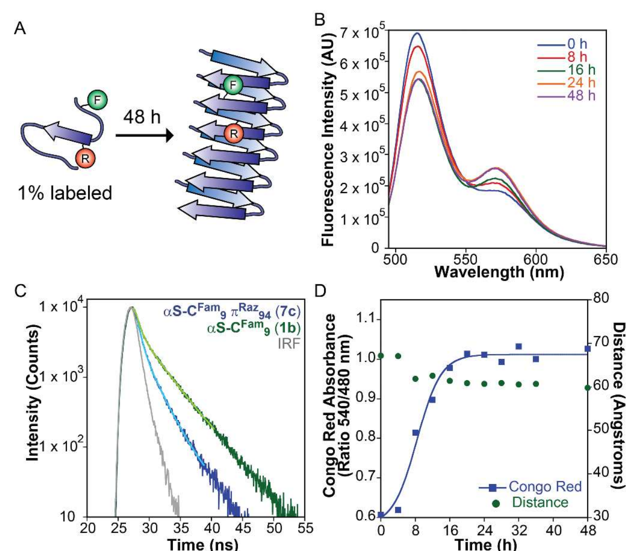

Fig. 6 FRET Studies of αS Aggregation. Top Left: Aggregation experiments were performed with 1% αS-$C^{Fam}_9\pi^{Raz}_{94}$ (**7c**) and 99% WT αS. Top Right: Fluorescence emission from αS-$C^{Fam}_9\pi^{Raz}_{94}$ (**7c**) at various times of aggregation. Bottom Left: Fluorescence lifetimes determined from TCSPC data for αS-$C^{Fam}_9\pi^{Raz}_{94}$ (**7c**, τ) and αS-C^{Fam}_9 (**1b**, τ_0) at t = 48 h are used to determine FRET efficiency ($E_{FRET} = 1 - \tau/\tau_0$). Bottom Right: Interchromophore distance calculated from E_{FRET} and Congo Red (CR) absorbance ratio at 540 and 480 nm at various times of aggregation.

We use lifetime measurements to calculate FRET efficiency ($E_{FRET} = 1 - \tau/\tau_0$) and determine an interchromophore separation using Förster theory. For αS-$C^{Fam}_9\pi^{Raz}_{94}$ (**7c**), we observe a small increase in E_{FRET}, corresponding to a change in the average separation of the labeled residues from 67 Å to 58 Å. Similar aggregation experiments were performed with the αS-$D^S_2C^{Fam}_9$ (**6e**) construct, but no significant differences in Fam quenching were observed between the monomer and fibril states (see ESI, Fig. S30-S32, Table S3). Making a large number of such FRET or PET measurements using a library of labeled constructs will allow one to map out the tertiary structure of each monomer unit within an αS fibril.

Singly labeled αS constructs can be used to study the biological mechanisms by which pathological αS is taken up and trafficked. To demonstrate this, CD-1 hippocampal neurons were incubated with fibrils containing 25% αS-C^{Fam}_{114} (**2b**) for 6 h. The cells were then fixed and stained against an intracellular marker of lysosomes, LAMP1, and DAPI to identify cell nuclei. Imaging of these samples at 100x magnification showed internalization of αS-C^{Fam}_{114} (**2b**) PFFs in similar quantities to unlabeled controls, stained with the monoclonal antibody (mAb) Syn204, and in good agreement with a previous study (Fig. 7 and ESI Fig. S35).[46] Additionally, the number of cells in which αS-C^{Fam}_{114} (**2b**) co-localized with LAMP1 was equivalent to control experiments using 100% WT αS, suggesting that the two species are processed similarly (see ESI Fig. S35).

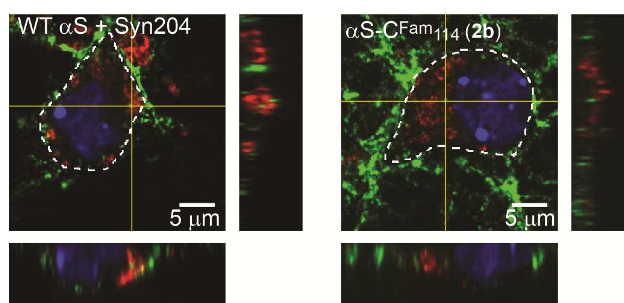

Fig. 7 Neuronal Uptake of Labeled αS. Left: Imaging of WT PFFs (Syn204 αS antibody, green), lysosomes (LAMP1 antibody, red) and nuclear staining (DAPI, blue) in the cell body of a primary non-Tg hippocampal neuron shows an internalized particle of exogenously added WT PFFs. Right: Imaging of cells transduced with fibrils made from 25% αS-C^{Fam}_{114} (**2b**) using fluorescein fluorescence (green), along with staining of lysosomes (LAMP1, red), and nuclei (DAPI, blue) shows that internalization is not disturbed by the incorporation of a single label at position 114. In both images, an internalized αS aggregate is identified in the crosshairs. The white outline is included to guide the eye in defining the cell body. Slabs at right and below the main images (XY plane) are scans in the YZ and XZ planes, respectively.

When transduced onto CD-1 primary hippocampal neurons for 7d, PFFs made from 5% or 25% αS-C^{Fam}_{114} showed a potency of >80% that of comparable WT αS fibrils for inducing misfolding of the endogenously expressed αS. Potency was determined by comparing the levels of staining of insoluble αS with phosphorylation at Ser129 (using mAb 81a), a marker of αS pathology (Fig. 8 and ESI Figs. S36, S37).[44, 47] Taken together, these data demonstrate that the introduction of a fluorophore at position 114 in αS does not significantly perturb its ability to seed aggregation in living cells. Furthermore, with a validated system in hand using fluorescently labeled PFFs, work on understanding the biological mechanisms underpinning aggregate uptake, transport/degradation, and recruitment of endogenous protein can now more rapidly be pursued via live cell imaging.

 This journal is © The Royal Society of Chemistry 20xx

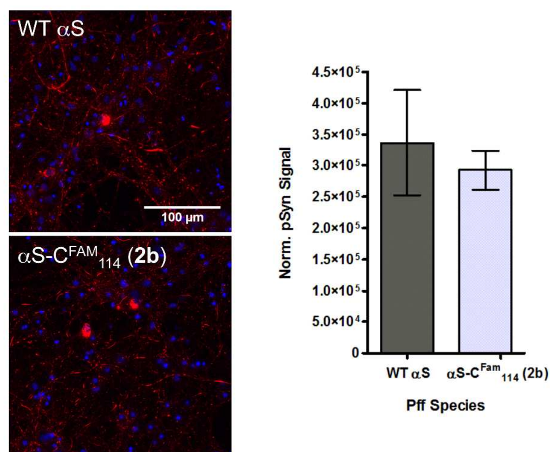

Fig. 8 Determination of Aggregation Seeding Capacity of Internalized 25% αSC$^{Fam}_{114}$ Fibrils. Left: Treatment of non-transgenic mouse primary hippocampal neurons with WT αS or 25% αS-C$^{FAM}_{114}$ PFFs (0.5 μg/coverslip) results in progressive accumulation of Lewy body-like insoluble, phosphorylated αS (pSyn) protein, as detected by immunostaining with phosphorylated Ser129 specific (81A, red) antibody. Nuclei were stained with DAPI (blue). Right: Quantitation of coverslips treated with WT αS PFFs or 25% αS-C$^{FAM}_{114}$ PFFs reveals a similar capacity for seeding pSyn aggregates. Error bars = SEM.

Conclusions

Since αS and other amyloidogenic proteins are generally not amenable to traditional structural studies, many laboratories are interested in functionalizing these proteins in order to study the aggregation process *in vitro* and *in vivo*. Here, we have employed several strategies to multifunctionalize αS with fluorophores using various combinations of conventional Cys modifications, amber codon suppression, AaT mediated N-terminal modification, and NCL. The combinations of techniques considered here fulfill the requirements of being either mutually orthogonal or able to be carried out sequentially.

Our findings are as follows. Cys modification can be used in combination with amber codon suppression, and this pair of techniques is applicable to αS, but may not be feasible in proteins with multiple Cys residues. Since AaT modification takes place post-translationally, when one uses it in combination with amber suppression, one can use similar types of bioorthogonal reactions in both labeling steps and control the position of the labels by reaction sequence. We find that when using azide/alkyne cycloaddition reactions for labeling after unnatural amino acid mutagenesis, it is best to incorporate the alkyne moiety during *in vivo* biosynthesis as we observe incomplete reactivity of azides incorporated co-translationally. For post-translational modifications with AaT, either the azide or the alkyne moiety may be transferred chemoenzymatically and then modified. The combination of AaT and Uaa mutagenesis, which has some attractive features such as tolerance to native Cys residues, was ultimately found to involve too many steps to be practical. NCL is the most versatile labeling strategy, but also the most laborious, so we reserve it for incorporation of thioamide probes that cannot currently be introduced biosynthetically or post-translationally. By borrowing the use of inteins from NCL methods, we have developed an efficient method of protein handling for rounds of Cys labeling and Cu-catalyzed click labeling of Ppy incorporated by unnatural mutagenesis. This allows us to rapidly generate doubly labeled αS, and our extensive construct validation assays can be used to show that these modification techniques do not alter the structure or conformational dynamics of αS.

A variety of *in vitro* and *in vivo* experiments can be performed using fluorescently-labeled αS in order to better understand the role of this important protein in Parkinson's Disease. Here, we have shown that site-specifically labeled αS can be used to monitor conformational changes during fibril formation or neuronal uptake of proteins in models of disease propagation. For such studies, it is important that one validate the labeled constructs as having near WT behavior. We have done so using a combination of measurements of fibrillization rates and TEM imaging of the mature fibrils. We will continue to build a library of proteins with well-tolerated label positions and use those constructs to gain a greater understanding of the mechanisms of αS aggregation as well as routes to potential therapeutics.

Acknowledgements

This work was supported by funding from the University of Pennsylvania, as well as grants from the National Institutes of Health (NIH NS081033 to EJP and NS053488-08 to VMYL). Instruments supported by the National Science Foundation and NIH include: MALDI MS (NSF MRI-0820996) and NMR (NIH RR-022442). CMH is supported by an Age Related Neurodegenerative Disease Training Grant fellowship (NIH T32AG000255).

Notes and references

1. A. C. M. Ferreon, C. R. Moran, Y. Gambin and A. A. Deniz, in *Methods in Enzymology*, ed. G. W. Nils, Academic Press, 2010, vol. Volume 472, pp. 179-204.
2. E. Haas, in *Intrinsically Disordered Protein Analysis*, eds. V. N. Uversky and A. K. Dunker, Humana Press, 2012, vol. 895, pp. 467-498.
3. J. M. Chalker, G. J. Bernardes, Y. A. Lin and B. G. Davis, *Chemistry, an Asian journal*, 2009, **4**, 630-640.
4. L. Wang and P. G. Schultz, *Angewandte Chemie-International Edition*, 2005, **44**, 34-66.
5. H. Neumann, K. Wang, L. Davis, M. Garcia-Alai and J. W. Chin, *Nature*, 2010, **464**, 441-444.
6. K. Wang, A. Sachdeva, D. J. Cox, N. W. Wilf, K. Lang, S. Wallace, R. A. Mehl and J. W. Chin, *Nature chemistry*, 2014, **6**, 393-403.
7. Y. Huang, W. K. Russell, W. Wan, P. J. Pai, D. H. Russell and W. S. Liu, *Molecular Biosystems*, 2010, **6**, 683-686.

This journal is © The Royal Society of Chemistry 20xx

Please do not adjust margins

8. W. Wan, Y. Huang, Z. Wang, W. K. Russell, P. J. Pai, D. H. Russell and W. R. Liu, *Angewandte Chemie (International ed.in English)*, 2010, **49**, 3211-3214.
9. S. Chen, N. E. Fahmi, L. Wang, C. Bhattacharya, S. J. Benkovic and S. M. Hecht, *J. Am. Chem. Soc.*, 2013, **135**, 12924-12927.
10. D. B. F. Johnson, J. Xu, Z. Shen, J. K. Takimoto, M. D. Schultz, R. J. Schmitz, Z. Xiang, J. R. Ecker, S. P. Briggs and L. Wang, *Nat Chem Biol*, 2011, **7**, 779-786.
11. H. Xiao, A. Chatterjee, S.-h. Choi, K. M. Bajjuri, S. C. Sinha and P. G. Schultz, *Angewandte Chemie International Edition*, 2013, **52**, 14080-14083.
12. I. Nikic, T. Plass, O. Schraidt, J. Szymanski, J. A. G. Briggs, C. Schultz and E. A. Lemke, *Angewandte Chemie-International Edition*, 2014, **53**, 2245-2249.
13. A. Varshavsky, *Genes to Cells*, 1997, **2**, 13-28.
14. A. M. Wagner, M. W. Fegley, J. B. Warner, C. L. J. Grindley, N. P. Marotta and E. J. Petersson, *J. Am. Chem. Soc.*, 2011, **133**, 15139-15147.
15. P. E. Dawson, T. W. Muir, I. Clarklewis and S. B. H. Kent, *Science*, 1994, **266**, 776-779.
16. M. G. Spillantini, M. L. Schmidt, V. M. Y. Lee, J. Q. Trojanowski, R. Jakes and M. Goedert, *Nature*, 1997, **388**, 839-840.
17. A. Der-Sarkissian, C. C. Jao, J. Chen and R. Langen, *Journal of Biological Chemistry*, 2003, **278**, 37530-37535.
18. M. Chen, M. Margittai, J. Chen and R. Langen, *Journal of Biological Chemistry*, 2007, **282**, 24970-24979.
19. M. Vilar, H. T. Chou, T. Luhrs, S. K. Maji, D. Riek-Loher, R. Verel, G. Manning, H. Stahlberg and R. Riek, *Proceedings of the National Academy of Sciences of the United States of America*, 2008, **105**, 8637-8642.
20. L. R. Lemkau, G. Comellas, K. D. Kloepper, W. S. Woods, J. M. George and C. M. Rienstra, *Journal of Biological Chemistry*, 2012, **287**, 11526-11532.
21. J. A. Rodriguez, M. I. Ivanova, M. R. Sawaya, D. Cascio, F. E. Reyes, D. Shi, S. Sangwan, E. L. Guenther, L. M. Johnson, M. Zhang, L. Jiang, M. A. Arbing, B. L. Nannenga, J. Hattne, J. Whitelegge, A. S. Brewster, M. Messerschmidt, B. Boutet, N. K. Sauter, T. Gonen and D. S. Eisenberg, *Nature*, 2015, **525**, 486-+.
22. T. J. van Ham, A. Esposito, J. R. Kumita, S. T. D. Hsu, G. S. K. Schierle, C. F. Kaminsk, C. M. Dobson, E. A. A. Nollen and C. W. Bertoncini, *Journal of Molecular Biology*, 2010, **395**, 627-642.
23. M. J. Roberti, T. M. Jovin and E. Jares-Erijman, *PLoS ONE*, 2011, **6**, e23338.
24. A. Grupi and E. Haas, *Journal of Molecular Biology*, 2011, **411**, 234-247.
25. A. C. Ferreon, M. M. Moosa, Y. Gambin and A. A. Deniz, *Proceedings of the National Academy of Sciences of the United States of America*, 2012, **109**, 17826-17831.
26. A. J. Trexler and E. Rhoades, *Molecular neurobiology*, 2013, **47**, 622-631.
27. S. Nath, J. Meuvis, J. Hendrix, S. A. Carl and Y. Engelborghs, *Biophysical journal*, 2010, **98**, 1302-1311.
28. A. W. Schmid, B. Fauvet, M. Moniatte and H. A. Lashuel, *Molecular & Cellular Proteomics : MCP*, 2013, **12**, 3543-3558.
29. M. Hejjaoui, M. Haj-Yahya, K. S. A. Kumar, A. Brik and H. A. Lashuel, *Angewandte Chemie-International Edition*, 2011, **50**, 405-409.
30. N. P. Marotta, Y. H. Lin, Y. E. Lewis, M. R. Ambroso, B. W. Zaro, M. T. Roth, D. B. Arnold, R. Langen and M. R. Pratt, *Nature Chemistry*, 2015, **7**, 913-920.
31. C. Li, E. A. Lutz, K. M. Slade, R. A. S. Ruf, G.-F. Wang and G. J. Pielak, *Biochemistry*, 2009, **48**, 8578-8584.
32. R. F. Wissner, S. Batjargal, C. M. Fadzen and E. J. Petersson, *J. Am. Chem. Soc.*, 2013, **135**, 6529-6540.
33. L. C. Speight, A. K. Muthusamy, J. M. Goldberg, J. B. Warner, R. F. Wissner, T. S. Willi, B. F. Woodman, R. A. Mehl and E. J. Petersson, *J. Am. Chem. Soc.*, 2013, **135**, 18806-18814.
34. J. M. Goldberg, R. F. Wissner, A. M. Klein and E. J. Petersson, *Chemical Communications*, 2012, **48**, 1550-1552.
35. S. Nehring, N. Budisa and B. Wiltschi, *PLoS ONE*, 2012, **7**, e31992.
36. T. Tanaka, A. M. Wagner, J. B. Warner, Y. X. J. Wang and E. J. Petersson, *Angewandte Chemie-International Edition*, 2013, **52**, 6210-6213.
37. K. Watanabe, Y. Toh, K. Suto, Y. Shimizu, N. Oka, T. Wada and K. Tomita, *Nature*, 2007, **449**, 867-U865.
38. D. P. Nguyen, T. Elliott, M. Holt, T. W. Muir and J. W. Chin, *J. Am. Chem. Soc.*, 2011, **133**, 11418-11421.
39. E. M. Brustad, E. A. Lemke, P. G. Schultz and A. A. Deniz, *J. Am. Chem. Soc.*, 2008, **130**, 17664-+.
40. M.-H. Seo, T.-S. Lee, E. Kim, Y. L. Cho, H.-S. Park, T.-Y. Yoon and H.-S. Kim, *Analytical Chemistry*, 2011, **83**, 8849-8854.
41. S. Batjargal, C. R. Walters and E. J. Petersson, *J. Am. Chem. Soc.*, 2015, **137**, 1734-1737.
42. R. F. Wissner, S. Batjargal, C. M. Fadzen and E. J. Petersson, *J. Am. Chem. Soc.*, 2013, **135**, 6529-6540.
43. K. C. Luk, V. Kehm, J. Carroll, B. Zhang, P. O'Brien, J. Q. Trojanowski and V. M. Y. Lee, *Science*, 2012, **338**, 949-953.
44. L. A. Volpicelli-Daley, K. C. Luk, T. P. Patel, S. A. Tanik, D. M. Riddle, A. Stieber, D. F. Meaney, J. Q. Trojanowski and V. M. Y. Lee, *Neuron*, 2011, **72**, 57-71.
45. K. C. Luk, C. Song, P. O'Brien, A. Stieber, J. R. Branch, K. R. Brunden, J. Q. Trojanowski and V. M. Y. Lee, *Proceedings of the National Academy of Sciences of the United States of America*, 2009, **106**, 20051-20056.
46. H. T. Tran, C. H. Y. Chung, M. Iba, B. Zhang, J. Q. Trojanowski, K. C. Luk and V. M. Y. Lee, *Cell Reports*, 2014, **7**, 2054-2065.
47. E. A. Waxman and B. I. Giasson, *J. Neuropathol. Exp. Neurol.*, 2008, **67**, 402-416.

This journal is © The Royal Society of Chemistry 20xx