Expressed Protein Ligation at Methionine: N-Terminal Attachment of Homocysteine, Ligation, and Masking**
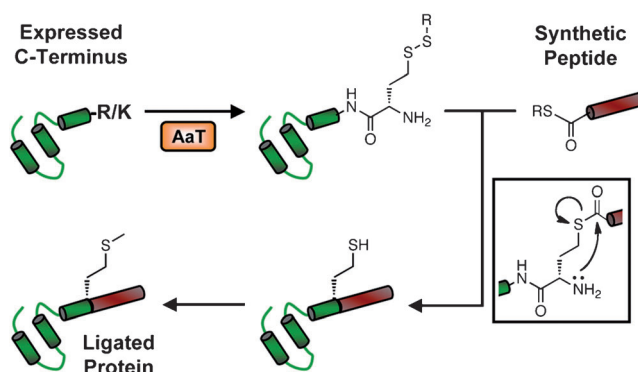
*Tomohiro Tanaka, Anne M. Wagner, John B. Warner, Yanxin J. Wang, and E. James Petersson**

The synthesis of chemically modified proteins is a valuable tool for the study of their function.[1] One of the most enabling protein synthesis technologies is native chemical ligation (NCL), in which a peptide thioester is reacted with a peptide bearing an N-terminal Cys to form a native amide bond.[2] In many cases, one of these two fragments can be expressed in *E. coli* and then ligated to a smaller synthetic peptide so that a large semi-synthetic protein can be made with a minimum of peptide synthesis. This sort of expressed protein ligation (EPL) can be implemented in two ways.[3] For a synthetic C-terminus, the N-terminal region can be expressed as a fusion to an intein and then converted to a thioester for ligation. Likewise, a protein can be expressed with an N-terminal Cys and ligated to a synthetic thioester. EPL is an extremely efficient way of making large quantities of synthetic proteins. However, there are limitations to the method, most notably, the need for Cys at the site of ligation.

Since many proteins do not have Cys residues at convenient locations or lack Cys entirely, several research groups have developed methods for "erasing" the ligation site. The most common strategies are alkylation of Cys to make Lys or Glu analogs and desulfurization to convert Cys to Ala.[4] Danishefsky and others have developed strategies using synthetic Cys analogs that are converted to other amino acids such as Leu, Phe, Val, Thr, and Lys by desulfurization.[5] Dawson and Muir have used nitrobenzyl Cys surrogates that can be removed by photolysis after ligation.[6] Raines and Bertozzi have demonstrated "traceless" ligations based on Staudinger reductions and Bode has developed a novel ligation using ketoacid/alkoxyamine partners.[7] Homocysteine (Hcs) has been used for ligation and then converted to Met by selective methylation.[8] While chemically elegant, most of these strategies are still limited in that they require the ligation handles to be incorporated synthetically. Therefore, these handles can be placed in a synthetic peptide to be ligated to an intein-derived thioester, but cannot be applied to an expressed C-terminal fragment for ligation to a synthetic thioester.

One way of functionalizing protein N-termini makes use of the aminoacyl transferase (AaT) enzyme from *E. coli*, which transfers Phe, Leu, or Met from an aminoacyl tRNA (Zaa-tRNA) to the N-terminus of a protein.[9,10] Our group has recently shown that AaT does not require full-length tRNAs, and can use aminoacyl adenosine (Zaa-A) substrates to transfer a variety of amino acids.[11] Previous work from Tirrell and Sisido demonstrated that AaT could transfer unnatural amino acids from Zaa-tRNAs produced either through semi-synthesis or by the activity of a modified aminoacyl tRNA synthetase (RS).[12] These two strategies are complementary. Our Zaa-A donor substrates are easily synthesized and allow us to explore the inherent substrate specificity of AaT without the necessity of identifying a novel RS. Once a useful substrate is identified, higher yields can be achieved with the catalytic regeneration of Zaa-tRNA by an appropriate RS. We wished to determine whether AaT could be used to functionalize proteins with erasable Cys analogs for ligation. Varshavsky and others have shown that AaT recognizes N-terminal Arg or Lys residues regardless of the adjacent sequence.[10,13] Therefore, we expected that any transferable analogs we identified would be applicable to a wide variety of protein targets. Here, we show that AaT can deliver disulfide-protected Hcs to the N-terminus of a protein under mild conditions. After reduction, Hcs can be used in a ligation reaction and then converted to Met by alkylation (Figure 1). We demonstrate the utility of this approach by
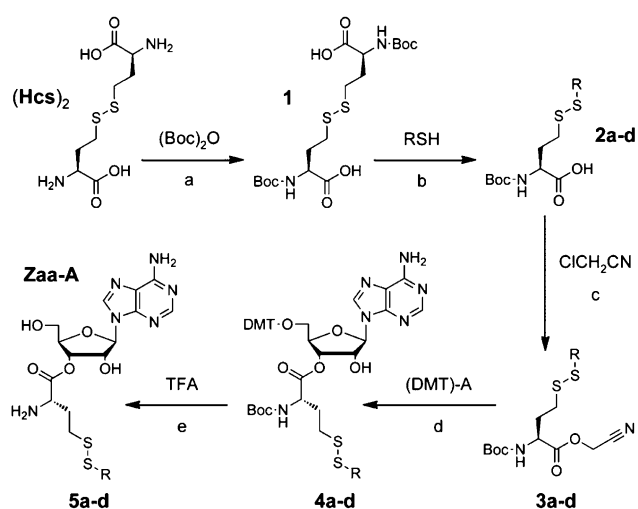
[*] Dr. T. Tanaka, A. M. Wagner, J. B. Warner, Y. J. Wang,
Prof. E. J. Petersson
Department of Chemistry, University of Pennsylvania
231 South 34th Street, Philadelphia, PA 19104-6323 (USA)
E-mail: ejpetersson@sas.upenn.edu
Homepage: http://ejpweb1.chem.upenn.edu/

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201302065.

Figure 1. Expressed protein ligation at methionine. An expressed protein bearing an N-terminal Arg or Lys (green) is functionalized with a protected Hcs analog from an aminoacyl adenosine donor, under catalysis by AaT transferase. A synthetic peptide thioester (red) is ligated to the modified protein under reducing conditions that allow deprotection of Hcs. After ligation, Hcs is converted to Met.

Scheme 1. Synthesis of protected homocysteine donor molecules. Reagents and conditions: a) (Boc)$_2$O, DIPEA, THF; b) RSH, NaOH, I$_2$, THF/H$_2$O (1:1); c) ClCH$_2$CN, DIPEA, THF; d) (DMT)-A, DIPEA, NBu$_4$Ac, THF; e) TFA, TIPSH, THF. **2a**, **3a**, **4a**, **5a** R = Me; **2b**, **3b**, **4b**, **5b** R = iPr; **2c**, **3c**, **4c**, **5c** R = tBu; **2d**, **3d**, **4d**, **5d** R = Ph. Yields reported in Supporting Information. Boc = tert-butyloxycarbonyl, DIPEA = N,N-diisopropylethylamine, THF = tetrahydrofuran, (DMT)-A = 5′-O-dimethoxytrityl adenosine, NBu$_4$Ac = N,N,N,N-tetrabutylam-monium acetate, TFA = trifluoroacetic acid, TIPSH = triisopropylsilane.

synthesizing α-synuclein, a protein whose aggregation contributes to Parkinson's disease pathology.[14]

Previous studies had shown that Cys-tRNA is not a substrate for AaT and our own preliminary experiments showed that Hcs was not transferred (data not shown).[13,15] Therefore, we prepared Hcs analogs masked as aliphatic or aromatic disulfides in order to mimic the endogenous AaT substrates Met, Leu, and Phe. Disulfide protecting groups (MeS, iPrS, tBuS, and PhS) were used so that they could be removed in situ under the reducing conditions of the subsequent NCL reaction.

AaT donor molecules were synthesized using a simple protocol beginning with Boc protection of oxidized Hcs dimer to give **1**, followed by disulfide exchange to form protected Hcs analogs **2a–d** (Scheme 1) These were activated as cyanomethyl esters **3a–d**, allowing for acylation of the 2′ and 3′ hydroxy groups of 5′-O-dimethoxytrityl-adenosine, (DMT)-A, without amidation of its unprotected exocyclic amine. TFA treatment cleaved both the DMT and the Boc groups to give Zaa-A donor molecules **5a–d**. Overall yields of the DMT-protected adenosyl analogs (**4a–d**) were ca. 40% from (Hcs)$_2$. However, several byproducts of the TFA deprotection reaction were observed, requiring HPLC purification of **5a–d**.

Once the Zaa-A donor molecules were purified, they were tested in transfer reactions to a reporter peptide LysAlaAcm, where Acm is 7-aminocoumarin. We compared yields of ZaaLysAlaAcm by HPLC analysis after 4 h reactions at pH 8.0 at 37°C (Scheme 2). Under these conditions, transfer of Phe is quantitative and Met proceeds in 75% yield. We first tested the transfer of disulfide-protected compounds using thiomethyl Cys (Csm), a Met isostere, prepared through a route similar to Scheme 1. Csm was transferred in 21%
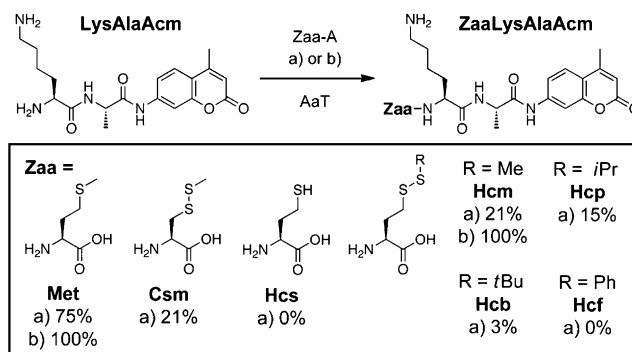
Scheme 2. Amino acid transfer to reporter peptide. Reagents and conditions: a) 1 mM Zaa-A, 0.1 mg mL^{-1} AaT; or b) 1 mM Zaa, 2 mg mL^{-1} E. coli total tRNA, 0.1 mg mL^{-1} Met*RS, 0.1 mg mL^{-1} AaT; with 100 μM LysAlaAcm in 50 mM HEPES, 150 mM KCl, 10 mM MgCl$_2$, pH 8.0, 37°C, 4 h. Yields determined by HPLC analysis.

yield. While this was lower than Met, it established that we could transfer disulfide-protected analogs without substantial side reactions. Hcs donors **5a–d** were tested in a similar fashion. Hcm was transferred efficiently from **5a**, with somewhat lower yields for Hcp and Hcb. No transfer of Hcf was observed from donor **5d**. In order to maintain the disulfide-protected analogs during the transfer reaction, we eliminated β-mercaptoethanol (BME) from the AaT purification process and the reaction buffer. This increased the yields of HcmLysAlaAcm (**6**) to 37%. Once Hcm was identified as an optimal substrate, we assessed the possibility of generating Hcm-tRNA in situ using an RS. Free Hcm was synthesized by TFA deprotection of Boc-Hcm. Wild-type E. coli MetRS and an L$_{13}$G mutant (Met*RS) previously used by the Tirrell group were tested with Hcm, ATP, and E. coli total tRNA.[16] Met*RS was more active than MetRS, giving full transfer at ≤ 0.1 mg mL^{-1} loading. This is consistent with Met*RS structural analysis, where longer sidechains are accommodated by the expanded active site.[16]

We then took HcmLysAlaAcm (**6**) on to a model ligation reaction with Ac-MetAspValPhe-SR (**7**) (Scheme 3). HPLC and MALDI MS analyses showed that disulfide cleavage was
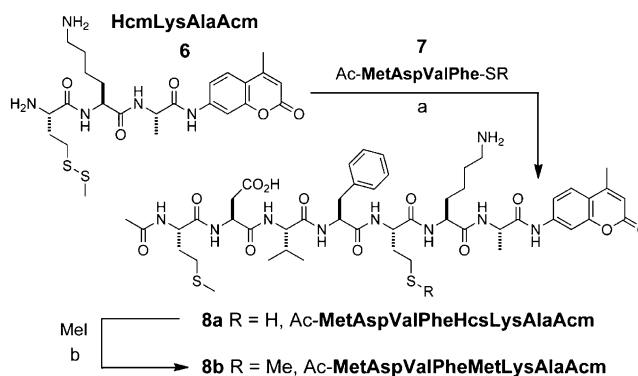
Scheme 3. Ligation and alkylation of Hcs-functionalized peptide. Reagents and conditions: a) 1.0 mM **6**, 3.3 mM **7**, 100 mM sodium phosphate dibasic, 20 mM TCEP, 0.5% PhSH, 0.5% BnSH, pH 7.2, room temperature, 24 h, 100%; b) 50 mM MeI, 100 mM sodium bicarbonate, pH 8.6, room temperature, 5 min, 89%.

completed almost immediately and that the subsequent ligation reached 100% completion after 24 h. After ligation, the product peptide (**8a**) was treated with 1000 equiv MeI at pH 8.6 to form **8b**. After 5 min, 89% conversion of Hcs_5 to Met_5 was observed by HPLC and MALDI MS analyses (Supporting Information). Unreacted **8a** (11%) could be removed after disulfide formation. Oxidation of Met_1, was observed (4%), but no undesired alkylation of Met or Lys was seen. This is consistent with previous uses of Hcs as a ligation handle, where subsequent alkylation is surprisingly selective, given that the pK_a of Hcs (8.9) is close to the pK_a of Lys (9.5).[17]

After successful completion of the model peptide ligation, we wished to test the Hcs transfer/ligation strategy with a full-sized protein. One protein that we have previously studied which has the requisite MetLys motif is α-synuclein (αS). Several recent studies have shown that the N-terminal sequence of αS can have important consequences for its folding and self-association.[18] However, these studies profoundly disagree on what those consequences are. Therefore, the preparation of N-terminally modified αS could be useful to addressing these questions and serve as a valuable demonstration of our method.

The αS NCL protocol is shown in Figure 2. We chose Met_5Lys_6 as the point of disconnection and prepared an $αS_{6-140}$ plasmid with an N-terminal His_{10} tag and a Factor Xa proteolysis site. The precursor protein, His_{Tag}-$αS_{6-140}$ (**9**), was isolated by Ni-affinity chromatography and cleaved with Factor Xa to give $αS_{6-140}$ (**10**). We incubated **10** with AaT, Met*RS, tRNA, Hcm, and ATP. Transfer of Hcm was complete after 1 h, within the limits of quantitation by MALDI MS (Supporting Information). Complete transfer is extremely valuable for large proteins, since separation of an

Hcm-modified protein such as **11** from unmodified **10** is not generally feasible. Modified αS **11** was isolated from AaT, Met*RS, and tRNA by boiling and FPLC purification. Then **11** was incubated with thioester AcMcmMetAspValPhe-SR (**S13**), which corresponds to $αS_{1-4}$ labeled at the N-terminus with 7-methoxycoumarinylalanine (Mcm). Overnight incubation at 37°C gave **12a** in high yield from **11**. No unligated starting materials were observed, though some oxidized **12a** was observed (Supporting Information). Purified **12a** was quantitatively converted to **13a** by treatment with aqueous MeI, as determined by analyses using Ellman's reagent and MALDI MS (Supporting Information). In general, we expect that the NCL reaction will not be 100% efficient, but we do not view this as a severe limitation, provided that the NCL product can be purified away from the starting materials and methylation is quantitative and selective.

A similar ligation was carried out with Ac-MetAspVal-Phe-SR (**7**), yielding *N*-Ac-αS after methylation (**13b**). The ligation product **13b** was subjected to trypsin digest and MALDI MS analysis to confirm that methylation occurred with high yield and few side reactions. Observation of the Ac-MetAspValPheMetLys fragment (Ac-$αS_{1-6}$) with only trace Ac-MetAspValPheHcsLys contamination confirmed that methylation proceeded in >95% yield within the limitations of quantitation by MS. Furthermore, MALDI MS analysis of the rest of the tryptic fragments demonstrated that there were no substantial alkylation side reactions, even on His_{50} in the $αS_{46-58}$ fragment (Supporting Information). Such selectivity is rewarding, but probably not completely general; for any novel protein, similar analysis by digestion and MS should be performed. It should also be noted that protein targets containing unprotected Cys residues would also be methylated on the free thiols. However, more conventional NCL strategies using ligation at Cys are probably viable in those cases.

Application of our system to targets other than αS may require mild denaturation of the protein in order to access the N-terminal amino acids. Therefore, we have tested AaT activity in the presence of low concentrations of denaturants and detergents. We find that AaT activity has a Gdn·HCl IC_{50} of 0.4 M and that full activity could be maintained in 5% v/v Triton X-100. While the N-termini of some proteins may be buried in important structural interactions, these concentrations should allow one to partially unfold many proteins to access the terminus for Hcm transfer and ligation.

While our experiments begin to remove the sequence limitations for EPL, they certainly do not eliminate them entirely. Essentially, one can now use XxxMetArg or XxxMetLys as a disconnection point in a retrosynthetic analysis of a potential protein target. The utility of our approach depends on how frequently this motif occurs. To address this, we analyzed the protein sequences in the PDB and found that out of the 60325 proteins surveyed, 5405 possible N-terminal ligation sites (<40 amino acids from the N-terminus) were identified in 5156 unique proteins. These numbers increased to 31259 ligation sites and 20701 proteins when we considered proteins with the ligation motif in the center of the protein where one might use multiple ligations to synthesize the full-length protein. We chose to restrict our
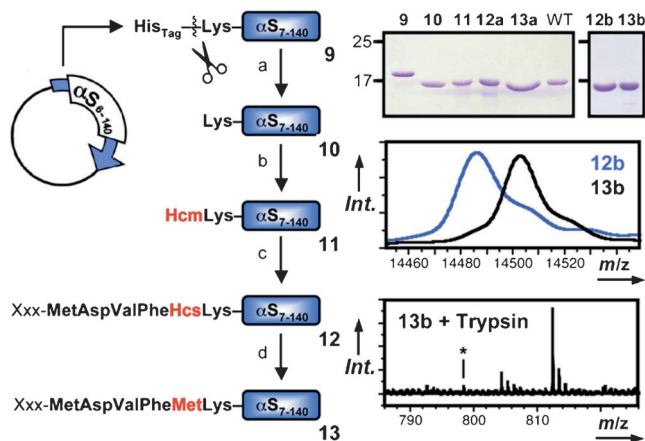
Figure 2. α-Synuclein model ligation. Functionalization of $αS_{6-140}$ by a) cleavage of His tag at Factor Xa site, b) attachment of Hcm by AaT-catalyzed modification, c) ligation of an N-terminal thioester peptide (**7** or **S13**), and d) conversion of Hcs to Met by methylation. **12a**, **13a** Xxx = AcMcm; **12b**, **13b** Xxx = Ac. Top right: Gel image showing product of each step, WT indicates full-length αS. Middle right: MALDI MS analysis of full length Ac-αS before (**12b**) and after (**13b**) methylation. Bottom right: Trypsinized fragment corresponding to Ac-$αS_{1-6}$ (812.4 *m/z*) confirms successful methylation. Asterisk indicates the expected mass of unmethylated Ac-$αS_{1-6}Hcs_5$.

search to the PDB since these sequences represent proteins with some degree of prior characterization typical of a protein target chosen for synthesis by NCL.

While it is clear that many proteins could be potential synthesis targets using AaT, we wish to eventually make our methods as broadly useful as possible. Therefore, it is worth noting that there exist other transferases with different N-terminal specificity, such as the BpT transferase from *V. vulnificus*, which transfers Leu to an N-terminal Asp or Glu.[19] Our preliminary experiments indicate that BpT can transfer Leu from small-molecule donors in vitro, but our current yields are too low (< 10 % after 4 h) to be useful. In situ aminoacylations using Hcp, Hcb, or Hcm with BpT and either Met*RS or *E. coli* LeuRS have not produced significant transfer yields. Efforts are underway to improve transfer of Hcs analogs. We are also working to mutate AaT to alter the substrate specificity for both Zaa and the N-terminal residue.

Prior to our work, the only way to avoid Cys at a ligation site in the expressed protein fragment during EPL was to desulfurize to form Ala or mask Cys by forming a non-native amino acid. Here, we have shown that a protein N-terminus can be functionalized with Hcs under mild conditions. After ligation, Hcs can be efficiently and selectively converted to Met, erasing the ligation site. Our work begins to remove the N-terminal Cys requirement for ligation by allowing one to use the MetArg or MetLys motif as a point of disconnection in protein semi-synthesis.

[1] C. P. R. Hackenberger, D. Schwarzer, *Angew. Chem.* **2008**, *120*, 10182–10228; *Angew. Chem. Int. Ed.* **2008**, *47*, 10030–10074.

[2] S. B. H. Kent, *Chem. Soc. Rev.* **2009**, *38*, 338–351.

[3] V. Muralidharan, T. W. Muir, *Nat. Methods* **2006**, *3*, 429–438.

[4] a) K. Okazaki, H. Yamada, T. Imoto, *Anal. Biochem.* **1985**, *149*, 516–520; b) L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, *123*, 526–533; c) C. E. Hopkins, G. Hernandez, J. P. Lee, D. R. Tolan, *Arch. Biochem. Biophys.* **2005**, *443*, 1–10.

[5] a) D. Crich, A. Banerjee, *J. Am. Chem. Soc.* **2007**, *129*, 10064–10065; b) J. Chen, Q. Wan, Y. Yuan, J. L. Zhu, S. J. Danishefsky, *Angew. Chem.* **2008**, *120*, 8649–8652; *Angew. Chem. Int. Ed.* **2008**, *47*, 8521–8524; c) C. Haase, H. Rohde, O. Seitz, *Angew.*

Chem. **2008**, *120*, 6912–6915; *Angew. Chem. Int. Ed.* **2008**, *47*, 6807–6810; d) R. L. Yang, K. K. Pasunooti, F. P. Li, X. W. Liu, C. F. Liu, *J. Am. Chem. Soc.* **2009**, *131*, 13592–13593; e) J. Chen, P. Wang, J. L. Zhu, Q. Wan, S. J. Danishefsky, *Tetrahedron* **2010**, *66*, 2277–2283; f) Z. Harpaz, P. Siman, K. S. A. Kumar, A. Brik, *ChemBioChem* **2010**, *11*, 1232–1235; g) S. D. Townsend, Z. P. Tan, S. W. Dong, S. Y. Shang, J. A. Brailsford, S. J. Danishefsky, *J. Am. Chem. Soc.* **2012**, *134*, 3912–3916.

[6] a) J. Offer, C. N. C. Boddy, P. E. Dawson, *J. Am. Chem. Soc.* **2002**, *124*, 4642–4646; b) C. Chatterjee, R. K. McGinty, J. P. Pellois, T. W. Muir, *Angew. Chem.* **2007**, *119*, 2872–2876; *Angew. Chem. Int. Ed.* **2007**, *46*, 2814–2818.

[7] a) E. Saxon, J. I. Armstrong, C. R. Bertozzi, *Org. Lett.* **2000**, *2*, 2141–2143; b) B. L. Nilsson, L. L. Kiessling, R. T. Raines, *Org. Lett.* **2001**, *3*, 9–12; c) J. W. Bode, R. M. Fox, K. D. Baucom, *Angew. Chem.* **2006**, *118*, 1270–1274; *Angew. Chem. Int. Ed.* **2006**, *45*, 1248–1252.

[8] a) J. P. Tam, Q. T. Yu, *Biopolymers* **1998**, *46*, 319–327; b) K. Pachamuthu, R. R. Schmidt, *Synlett* **2003**, 659–662; c) A. Saporito, D. Marasco, A. Chambery, P. Botti, S. M. Monti, C. Pedone, M. Ruvo, *Biopolymers* **2006**, *83*, 508–518; d) B. Aussedat, B. Fasching, E. Johnston, N. Sane, P. Nagorny, S. J. Danishefsky, *J. Am. Chem. Soc.* **2012**, *134*, 3532–3541.

[9] A. Kaji, G. D. Novelli, H. Kaji, *Biochem. Biophys. Res. Commun.* **1963**, *10*, 406–409.

[10] A. Varshavsky, *Nat. Struct. Mol. Biol.* **2008**, *15*, 1238–1240.

[11] A. M. Wagner, M. W. Fegley, J. B. Warner, C. L. J. Grindley, N. P. Marotta, E. J. Petersson, *J. Am. Chem. Soc.* **2011**, *133*, 15139–15147.

[12] a) M. Taki, M. Sisido, *Biopolymers* **2007**, *88*, 263–271; b) R. E. Connor, K. Piatkov, A. Varshavsky, D. A. Tirrell, *ChemBioChem* **2008**, *9*, 366–369.

[13] M. J. Leibowitz, *J. Biol. Chem.* **1971**, *246*, 5207–5212.

[14] P. K. Auluck, G. Caraveo, S. Lindquist, *Annu. Rev. Cell Dev. Biol.* **2010**, *26*, 211–233.

[15] R. C. Scarpulla, C. E. Deutch, R. L. Soffer, *Biochem. Biophys. Res. Commun.* **1976**, *71*, 584–589.

[16] A. J. Link, M. K. S. Vink, N. J. Agard, J. A. Prescher, C. R. Bertozzi, D. A. Tirrell, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10180–10185.

[17] M. Trujillo, R. Radi, *Arch. Biochem. Biophys.* **2002**, *397*, 91–98.

[18] a) K. Vamvaca, M. J. Volles, P. T. Lansbury, *J. Mol. Biol.* **2009**, *389*, 413–424; b) T. Bartels, L. S. Ahlstrom, A. Leftin, F. Kamp, C. Haass, M. F. Brown, K. Beyer, *Biophys. J.* **2010**, *99*, 2116–2124; c) T. Bartels, J. G. Choi, D. J. Selkoe, *Nature* **2011**, *477*, 107–110; d) K. Vamvaca, P. T. Lansbury, L. Stefanis, *J. Neurochem.* **2011**, *119*, 389–397; e) B. Fauvet, M. B. Fares, F. Samuel, I. Dikiy, A. Tandon, D. Eliezer, H. A. Lashuel, *J. Biol. Chem.* **2012**, *287*, 28243–28262.

[19] E. Graciet, R. G. Hu, K. Piatkov, J. H. Rhee, E. M. Schwarz, A. Varshavsky, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3078–3083.