JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Communication
pubs.acs.org/JACS

Inteins as Traceless Purification Tags for Unnatural Amino Acid Proteins

Solongo Batjargal, Christopher R. Walters, and E. James Petersson*

Department of Chemistry, University of Pennsylvania, 231 South 34th Street, Philadelphia, Pennsylvania 19104, United States

Supporting Information

ABSTRACT: The expression of proteins containing unnatural amino acids through suppression of a stop codon can be limited by truncation due to competition with release factors. When the site of incorporation is near the C-terminus, it may not be feasible to separate the full-length unnatural amino acid protein from the truncated form. We report a simple, traceless procedure that allows one to isolate the desired protein using a C-terminal intein fusion.

Genetic incorporation of unnatural amino acids (Uaas) into proteins has provided the ability to probe and control protein structure and function with extraordinary precision.[1,2] To accomplish site-specific incorporation, an orthogonal tRNA is selectively charged with a Uaa by its aminoacyl-tRNA synthetase (RS) and directs the insertion of the Uaa into the nascent polypeptide chain in response to nonsense (i.e., premature stop) or quadruplet codons during translation.[3,4] The "suppressor" tRNA has an anticodon that recognizes the nonsense codon; for example, $tRNA_{CUA}$ pairs with the UAG stop codon. One of the major limitations of this method is the competition of $tRNA_{CUA}$ with release factor 1 (RF1) at the UAG codon, leading to protein truncation. Purification of the full-length protein can be very difficult, if not impossible, when the UAG codon is located near the C-terminus. The use of C-terminal tags such as His_6 can alter the native properties of some proteins, and cleavage of the tags by proteases may not always be specific. Other efforts to solve the "truncation problem" have centered on nontrivial manipulations of the *E. coli* translation system, including genetic deletion of RF1 or reassignment of all of the UAG stop codons in the bacterium.[5,6] These manipulations may lead to compromised cell health and lower levels of protein production.

Here, we report a simple and efficient method of isolating full-length proteins containing a Uaa in a traceless manner by exploiting the self-excising property of inteins. Engineered inteins have been leveraged for the preparation of protein thioesters, polypeptide cyclization, and the regulation of protein functions.[7] In our own laboratory, we have used inteins to prepare Uaa-containing protein fragments for ligation to synthetic peptides to label proteins with both side chain and backbone modifications.[8] This line of investigation led us to develop a traceless intein tagging method for full-length Uaa proteins (Figure 1).

The C-terminal His-tagged Mxe GyrA intein provides an affinity purification tool that, by virtue of its placement, can exist only in full-length Uaa protein **2** and not in truncated protein **1**. Subsequent removal of the intein tag (**3**) in a β-mercaptoethanol (βME) triggered process yields the Uaa-labeled protein with a C-terminal carboxylate (**7**) via intermediates **4** and **5**. We demonstrate the generality of this approach with three test proteins: an intrinsically disordered protein, α-synuclein (αS), a small Ca^{2+} binding protein, calmodulin (CaM), and a larger globular protein, maltose binding protein (MBP). In addition, we examine a broad spectrum of Uaas including the spectroscopic probes *p*-cyanophenylalanine (Cnf or F*)[9,10] and acridon-2-ylalanine (Acd or δ),[11] the "click" chemistry handles propargyltyrosine (Ppy or π) and *p*-azidophenylalanine (Azf or Z),[12] and the photo-cross-linker *p*-benzoyl-phenylalanine (Bzf or B).[13] This technique should be of general use to the Uaa mutagenesis community as it provides a straightforward solution to the truncation problem.

We began our investigation with the expression of αS (140 amino acids) containing Cnf at position 136, fused to Mxe GyrA intein followed by a C-terminal His_6 tag (αS-F*$_{136}$-MxeH_6, **2a**). To incorporate Cnf, we transformed BL21(DE3) cells with a pTXB1-αS-TAG_{136}-MxeH_6 plasmid and a plasmid encoding a *Methanocaldococcus jannaschii* (*Mj*) mutant TyrRS/$tRNA_{CUA}$ pair previously selected for Cnf incorporation (pDULE2-pXF).[14] After lysing the cells, both truncation product αS_{1-135} (**1a**) and mature fusion protein **2a** were observed. Only the mature protein with the Uaa was obtained from Ni nitrilotriacetic acid (Ni-NTA) purification by elution with imidazole, leaving the truncation product in the flow-through (Figure 1). Matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (MS) and polyacrylamide gel electrophoresis (PAGE) analyses of the Ni-NTA purified αS-F*$_{136}$-MxeH_6 **2a** showed no evidence of the truncation product.

Next, we carried out an in-depth analysis of hydrolysis of the fusion protein by βME. A study carried out by New England Biolabs reported that dithiothreitol (DTT) induced faster intein cleavage than βME.[15] However, Kent and co-workers showed that the conversion rate from peptide-thioesters (−COSR) to peptide-carboxylates ($-CO_2H$) with DTT was significantly slower in comparison to βME in 6 M guanidinium.[16] To test βME as a traceless cleaving reagent, we treated αS-F*$_{136}$-MxeH_6 **2a** with 200 mM βME in HEPES buffer at pH 7.5 at room temperature (RT, 22 °C) and monitored the cleavage periodically using SDS PAGE and MALDI MS. The fusion protein was >90% cleaved to give the

Received: October 7, 2014
Published: January 27, 2015

DOI: 10.1021/ja5103019
J. Am. Chem. Soc. 2015, 137, 1734−1737

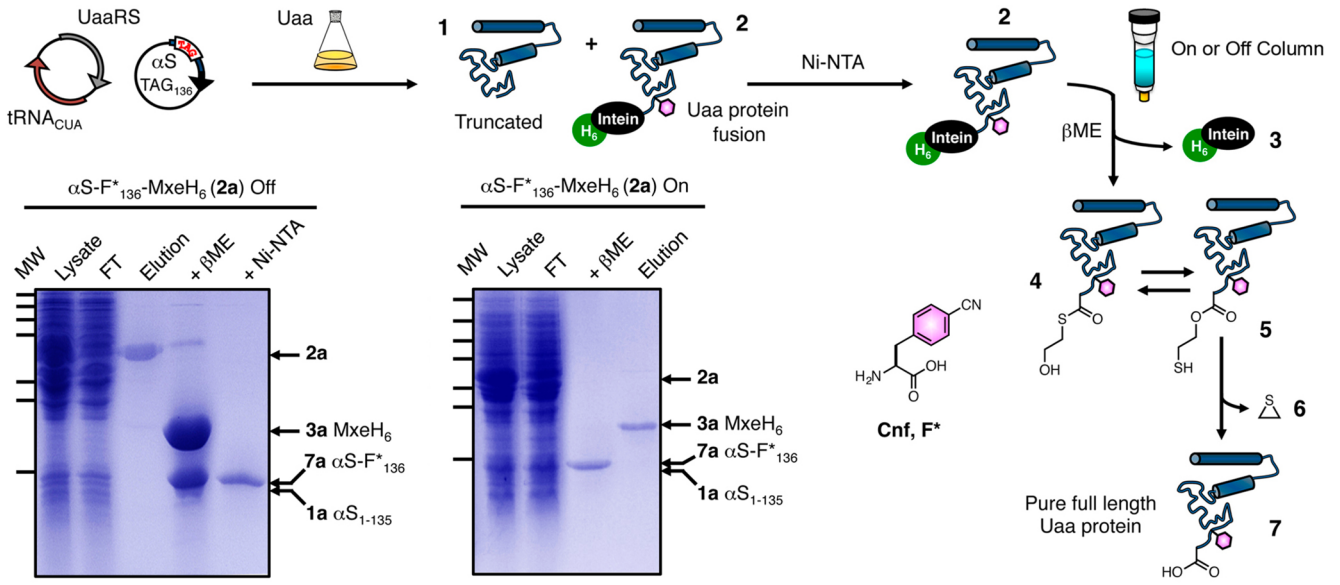

Figure 1. Purification of unnatural amino acid proteins using a C-terminal intein tag. Unnatural amino acid (Uaa; e.g., *p*-cyanophenylalanine, Cnf) proteins are expressed using a plasmid encoding the protein of interest with a TAG stop codon (e.g., α-synuclein mutant at position 136, αS-TAG$_{136}$) and a plasmid encoding an evolved tRNA synthetase (UaaRS). The full length intein fusion protein containing the Uaa (**2**) is purified away from the protein truncated at the stop codon (**1**) using Ni-NTA resin. Hydrolysis resulting from β-mercaptoethanol (βME) treatment generates purified Uaa protein **7** via intermediates **4** and **5**. Intein cleavage with βME can be carried out on or off of the Ni-NTA column. Inset: A sample expression and purification of αS-Cnf$_{136}$ with either off- or on-column hydrolysis. Off: Crude cell lysate, flow through from Ni-NTA column (FT), elution of Ni-NTA column with 300 mM imidazole, cleavage of eluted intein fusion with 200 mM βME for 24 h, and removal of intein and uncleaved fusion protein by incubation with Ni-NTA beads. On: Crude cell lysate, FT, elution of Ni-NTA column with 200 mM βME after incubation for 24 h, and elution of cleaved intein from Ni-NTA column with 300 mM imidazole: **1a**, truncated αS$_{1-135}$; **2a**, Uaa protein fusion αS-F*$_{136}$-MxeH$_6$; **3a**, cleaved MxeH$_6$ intein; **7a**, pure full length Uaa protein αS-F*$_{136}$, molecular weight (MW) markers 17, 25, 30, 46, 58, 80, and 175 kDa.

protein−CO_2H after 8 h incubation. The intermediate product(s), presumably a mixture of αS-F*$_{136}$ βME-thioester (**4a**) and βME-oxoester (**5a**), were observed at the 2 h time point by MALDI MS. This mixture disappeared over time to form C-terminal carboxylate **7a**, expelling ethylene sulfide (**6**). Similar analyses with DTT resulted in slightly faster splicing than βME, but the formation of the desired C-terminal carboxylate was significantly slower with DTT (see Supporting Information (SI) Figure S1). MALDI MS analyses of full-length and trypsin-digested protein were used to ensure the completion of carboxylate formation and the fidelity of Cnf insertion.

To optimize intein hydrolysis with βME, we incubated αS-F*$_{136}$-MxeH$_6$ **2a** while varying conditions including temperature, βME concentration, and pH (see SI, Figures S2−S4). As expected, we observed a faster splicing rate as we increased the temperature or βME concentration (E_a = 10.6 kcal·mol^{-1}, k = 5.79 × 10^{-4} M^{-1} s^{-1}, see SI). In contrast, the splicing reactivity was not significantly influenced by pH, probably due to a balance of acid- and base-catalyzed steps in transthioesterification, S−O rearrangement, and hydrolysis. On the basis of these analyses, we carried out the rest of the hydrolysis reactions with 200 mM βME at pH 7.5, 22 °C; conditions under which hydrolysis of **2a** was >90% complete in 8 h, giving **7a** in yields of 6 mg per L of cell culture (see SI, Figure S5). However, hydrolysis conditions, including the βME concentration, may need to be optimized for a protein of interest depending on factors such as stability and the identity of the C-terminal residue forming the intein thioester.

Instead of eluting the protein from the Ni-affinity column (off-column), hydrolysis can be carried out when the protein is still bound to the column (on-column). After removing cellular proteins by washing, resin-bound αS-F*$_{136}$-MxeH$_6$ **2a** was either retained on the beads or eluted from the column with imidazole. The on- and off-column approaches were both hydrolyzed in >95% yield (Figure 1). For on-column hydrolysis, βME treatment inherently led to elution of pure product. For off-column hydrolysis, incubation with Ni-NTA resin allowed for easy removal of cleaved intein **3a** and unhydrolyzed fusion protein **2a**. It should be noted that some reduction of the Ni^{2+} resin can occur with on-column hydrolysis, and the resin should be regenerated with fresh $NiSO_4$ if discoloration is observed.

It is important that the use of the intein fusion does not disrupt the native fold of the protein of interest. Since αS is a disordered protein, we chose to evaluate the impact of the intein on folding using CaM. Previous attempts to express CaM^F-F*$_{139}$ (**7b**) using standard stop codon suppression methods resulted in ∼70% truncation (see SI, Figure S6). In contrast, when we expressed CaM^F-F*$_{139}$-MxeH$_6$ (**2b**) and carried out off-column hydrolysis, we obtained **7b** in 8 mg/L yield (see SI, Figure S7). Through the use of circular dichroism spectroscopy and peptide binding assays, we confirmed that CaM^F-F*$_{139}$ structure was not perturbed by fusion to the intein (see SI, Figure S8). However, it must be noted that the use of high concentrations of βME may compromise the folding or activity of some proteins that rely on key, labile disulfide bonds. A separate activity assay should always be used.

To test the viability of the intein tag strategy for larger Uaas that are incorporated less efficiently, we incorporated Acd in CaM and Bzf in αS and triggered hydrolysis with βME to efficiently isolate CaM-δ_{139}-CO_2H (**7c**, Figure 2) or αS-B$_{136}$-

DOI: 10.1021/ja5103019
J. Am. Chem. Soc. 2015, 137, 1734−1737

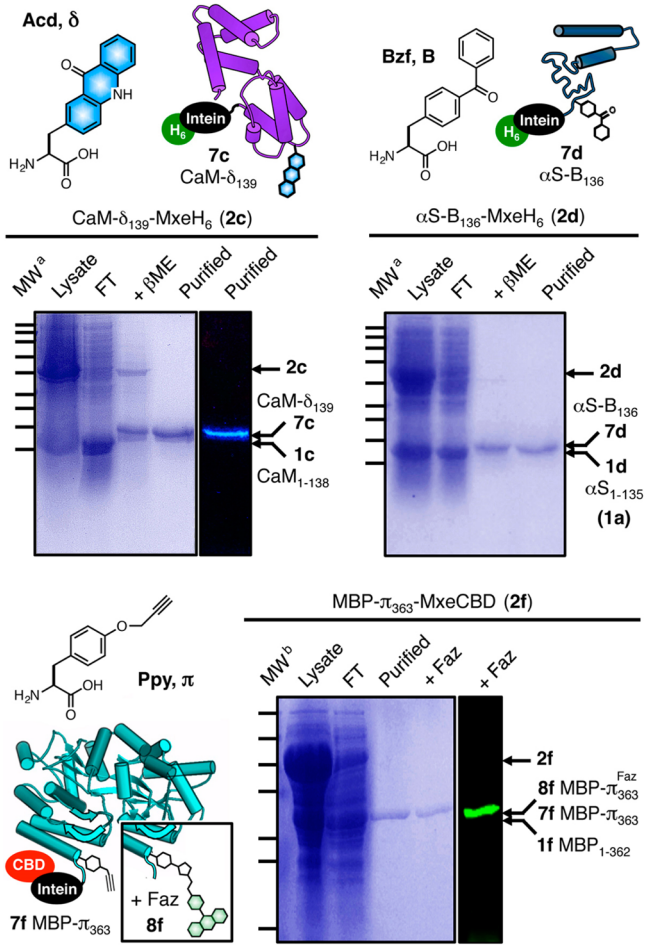

Figure 2. Application of C-terminal intein tag strategy with various amino acids, proteins, and affinity tags. Top: expression and purification of acridon-2-ylalanine (Acd, δ) mutant of calmodulin (CaM-δ_{139}), or *p*-benzoylphenylalanine (Bzf, B) mutant of α-synuclein (αS-B_{136}) with on-column hydrolysis. Crude cell lysate, flow through from Ni-NTA column (FT), elution of protein after cleavage with 200 mM βME for 16 h, protein purified by fast protein liquid chromatography (FPLC). Top Left: **1c**, truncated CaM_{1-138}; **2c**, Uaa protein fusion CaM-δ_{139}-$MxeH_6$; **7c**, pure full length Uaa protein CaM-δ_{139}, molecular weight (MW) markers 11, 17, 22, 25, 32, 46, 58, 75, and 100 kDa. Top Right: **1d** (**1a**), truncated αS_{1-135}; **2d**, Uaa protein fusion αS-B_{136}-$MxeH_6$; **7d**, Pure full length Uaa protein αS-B_{136}, MW markers 11, 17, 22, 25, 32, 46, 58, 75, and 100 kDa. Bottom: expression and purification of propargyl tyrosine (Ppy, π) mutant of maltose binding protein (MBP-π_{363}) with on-column hydrolysis and labeling with fluorescent azide reagent Rhodamine 110 azide (Faz). Crude cell lysate, FT, elution of protein after cleavage with 200 mM βME for 16 h, protein labeled with Faz using Cu catalysis and purified by FPLC: **1f**, truncated MBP_{1-362}; **2f**, Uaa protein fusion MBP-π_{363}-$MxeH_6$; **7f**, pure full length Uaa protein MBP-π_{363}; **8f**, fluorescently labeled protein MBP-π^{Faz}_{363}, MW markers 17, 22, 25, 30, 46, 58, and 80 kDa.

CO_2H (**7d**, Figure 2) in 3−10 mg/L yields in spite of substantial truncation (≥50%). Although the crude on-column hydrolysis products ("+ βME" lanes in Figure 2) were relatively clean, additional fast protein liquid chromatography (FPLC) purification should be used to purify proteins for biophysical assays, as is typical of most Ni-NTA purifications. We also tested the compatibility of our method with azide amino acids that may be subject to reduction by βME. Azf was incorporated into αS as a "click" chemistry handle for reactions with strained cyclooctynes. After on-column hydrolysis of αS-Z_{136}-$MxeH_6$ (**2e**), the incubation of αS-Z_{136}-CO_2H (**7e**) with tetramethylrhodamine dibenzocyclooctyne (Rco) resulted in labeling of αS to give αS-Z^{Rco}_{136}-CO_2H (**8e**, SI, Figure S9). Again, at least one round of additional purification is typically necessary to remove unreacted dye.

We also wished to ensure that our methods were applicable to larger proteins, alternate immobilization tags, and other inteins. We tested an MBP (367 amino acids) fusion with a C-terminal Mxe GyrA intein and chitin binding domain (CBD). We introduced a UAG codon at position 363, four residues from the normal MBP C-terminus, and expressed the construct with Ppy. Isolation of the full-length protein MBP-π_{363}-MxeCBD (**2f**) by incubation with chitin beads and subsequent treatment with βME resulted in the liberation of MBP-π_{363}-CO_2H (**7f**) in an 8 mg/L yield. After additional purification via FPLC, **7f** was treated with a fluorescent azide compound (Rhodamine 110 azide, Faz) under Cu-catalyzed "click" conditions to give MBP-π^{Faz}_{363}-CO_2H (**8f**). In separate experiments, we generated fusions of αS and CaM with DnaE inteins (Ava, Mcht, and Npu) and found that expression levels were comparable to Mxe GyrA fusions and that off-column hydrolysis rates were comparable to or slower than Mxe GyrA (see SI, Figures S7 and S10).

In conclusion, we have developed a traceless and efficient strategy for separating desired Uaa proteins from similarly sized truncation byproducts. We have demonstrated that a variety of proteins, Uaas, inteins, and affinity tags can be used in our method. Since Mxe GyrA intein expression systems are already commercially available, the strategy described here can easily be adopted by the biochemical community and should be applicable to a wide array of proteins expressed in *E. coli*. Our intein tags should also be useful in combination with expression systems with reduced RF-1 activity where some truncation may still take place.[4−6] We expect that our method will expand the applications of suppression-based Uaa methodology for labeling near the C-terminus, including protein therapeutics or studies of protein structure and function.

■ ASSOCIATED CONTENT

Supporting Information

All experimental procedures and analysis of protein expression and labeling reactions. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*ejpetersson@sas.upenn.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by funding from the National Institutes of Health (NIH NS081033 to E.J.P.). MALDI MS instrumentation was supported by the National Science Foundation (NSF MRI-0820996). The authors thank Prof. Tom Muir for the gifts of several intein plasmids and Itthipol Sungwienwong for acridon-2-ylalanine. S.B. thanks the Parkinson's Disease Foundation for a summer predoctoral fellowship. C.R.W. thanks the NIH for funding through the Structural Biology and Molecular Biophysics Training Program (T32 GM008275).

DOI: 10.1021/ja5103019
J. Am. Chem. Soc. 2015, 137, 1734−1737

REFERENCES

(1) Wang, Q.; Parrish, A. R.; Wang, L. *Chem. Biol.* **2009**, *16*, 323−336.
(2) Liu, C. C.; Schultz, P. G. *Annu. Rev. Biochem.* **2010**, *79*, 413−444.
(3) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498−500.
(4) Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. *Nature* **2010**, *464*, 441−444.
(5) Johnson, D. B. F.; Xu, J.; Shen, Z.; Takimoto, J. K.; Schultz, M. D.; Schmitz, R. J.; Xiang, Z.; Ecker, J. R.; Briggs, S. P.; Wang, L. *Nat. Chem. Biol.* **2011**, *7*, 779−786.
(6) Lajoie, M. J.; Rovner, A. J.; Goodman, D. B.; Aerni, H.-R.; Haimovich, A. D.; Kuznetsov, G.; Mercer, J. A.; Wang, H. H.; Carr, P. A.; Mosberg, J. A.; Rohland, N.; Schultz, P. G.; Jacobson, J. M.; Rinehart, J.; Church, G. M.; Isaacs, F. J. *Science* **2013**, *342*, 357−360.
(7) Vila-Perelló, M.; Muir, T. W. *Cell* **2010**, *143*, 191−200.
(8) Wissner, R. F.; Batjargal, S.; Fadzen, C. M.; Petersson, E. J. *J. Am. Chem. Soc.* **2013**, *135*, 6529−6540.
(9) Schultz, K. C.; Supekova, L.; Ryu, Y.; Xie, J.; Perera, R.; Schultz, P. G. *J. Am. Chem. Soc.* **2006**, *128*, 13984−13985.
(10) Miyake-Stoner, S. J.; Miller, A. M.; Hammill, J. T.; Peeler, J. C.; Hess, K. R.; Mehl, R. A.; Brewer, S. H. *Biochemistry* **2009**, *48*, 5953−5962.
(11) Speight, L. C.; Muthusamy, A. K.; Goldberg, J. M.; Warner, J. B.; Wissner, R. F.; Willi, T. S.; Woodman, B. F.; Mehl, R. A.; Petersson, E. J. *J. Am. Chem. Soc.* **2013**, *135*, 18806−18814.
(12) Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, J. C.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 11782−11783.
(13) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11020−11024.
(14) Taskent-Sezgin, H.; Chung, J.; Patsalo, V.; Miyake-Stoner, S. J.; Miller, A. M.; Brewer, S. H.; Mehl, R. A.; Green, D. F.; Raleigh, D. P.; Carrico, I. *Biochemistry* **2009**, *48*, 9040−9046.
(15) Chong, S.; Mersha, F. B.; Comb, D. G.; Scott, M. E.; Landry, D.; Vence, L. M.; Perler, F. B.; Benner, J.; Kucera, R. B.; Hirvonen, C. A.; Pelletier, J. J.; Paulus, H.; Xu, M.-Q. *Gene* **1997**, *192*, 271−281.
(16) Gates, Z. P.; Stephan, J. R.; Lee, D. J.; Kent, S. B. H. *Chem. Commun.* **2013**, *49*, 786−788.

DOI: 10.1021/ja5103019
J. Am. Chem. Soc. 2015, 137, 1734−1737