Protein Synthesis

DOI: 10.1002/anie.201204144

Sequential α-Ketoacid-Hydroxylamine (KAHA) Ligations: Synthesis of C-Terminal Variants of the Modifier Protein UFM1**

*Ayodele O. Ogunkoya, Vijaya R. Pattabiraman, and Jeffrey W. Bode**

The role of conjugation and deconjugation of modifier proteins, such as ubiquitin and SUMO, to their targets has rapidly emerged as one of the most exciting areas of molecular biology and drug discovery.[1] The development of new therapeutics perturbing these regulatory pathways requires access to labeled and derivatized modifier proteins and a means to prepare their conjugates.[2] In some cases, the modifier protein, the target protein, and the specific ligases can be obtained from natural sources but research is increasingly turning to synthetic methods to prepare them and study their biological roles.[2,3] In this regard, the native chemical ligation (NCL) of C-terminal modifier protein thioesters to a 5-thiolysine residue in a protein stands out for its promise to impact this area.[4] The unnatural 5-thiolysine residue can be incorporated into a target protein by total chemical synthesis,[5] semi-synthesis from expressed protein segments[6] or by ribosomal incorporation.[7] It has already been used to attach ubiquitin to α-synuclein[6] and SUMO[7] and to prepare di-,[4,8] tri-,[5] and tetra-ubiquitins.[9]

The execution of synthetic and semi-synthetic approaches to preparing modifier protein–protein conjugates requires C-terminal derivatives of modifier proteins. Access to such molecules is currently limited to ubiquitin,[4,5,8,9,10,11a] SUMO, Nedd8, and ISG15[11a] prepared either synthetically, semi-synthetically by NCL of an expressed protein fragment or with the use of their specific E1 enzymes[11a] to provide thioesters for NCL. The extension of this technique to the preparation and incorporation of other modifier proteins including FUB1,[12] MUB,[13] and UFM1[14] would require either the expression of intein-forming fusion proteins followed by capture with exogenous thiols,[15] the use of specific E1 enzymes, or the total chemical synthesis of C-terminal thioesters. The size of these modifier proteins (70–140 residues) generally precludes linear solid-phase synthesis of the thioesters and many are arguably outside the range of two-segment ligations. A three- or four-segment ligation strategy is required.

We now report the chemical synthesis of C-terminal variants of ubiquitin-fold modifier 1 (UFM1) by multiple segment ligations using the α-ketoacid-hydroxylamine (KAHA) ligation with 5-oxaproline (Opr). All of the necessary peptide segments for ligation can be easily prepared by standard Fmoc solid-phase peptide synthesis (SPPS). The N-terminal Fmoc-protected Opr peptides are stable to resin cleavage, purification, oxidation, and ligation but can be rapidly and cleanly removed from the otherwise unprotected ligation product. This work is the first alternative to native chemical ligation for sequential chemoselective ligations that afford native amide bonds.[16]

The key to any sequential ligation strategy (Figure 1) is the identification of a suitable protecting group for the N or the C terminus of a ligating segment. Protection strategies for sequential NCL include N-terminal protection of cysteine groups as a thiazolidine and their liberation with hydroxylamine and C-terminal thioesters prepared post-ligation from rearrangement.[17] The major limitation of sequential NCL is the need for multiple cysteine residues that must be desulfurized to avoid interference in protein function. In contrast, KAHA-ligation with Opr results in the formation of α-homoserine (Hse, T[§]), which may serve as an innocuous substitute for other amino acids.[18]

The primary goal of our study was to identify a protecting group that fulfilled the following criteria: 1) orthogonal to the

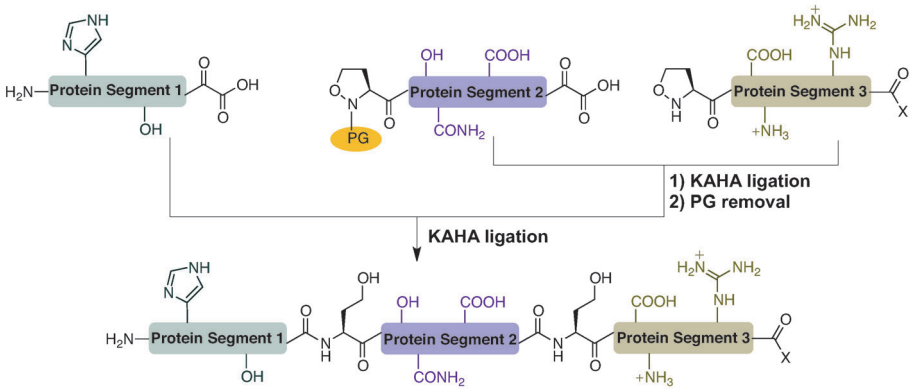


Figure 1. Sequential KAHA ligation. PG = Protecting group, X = C-terminal variants.

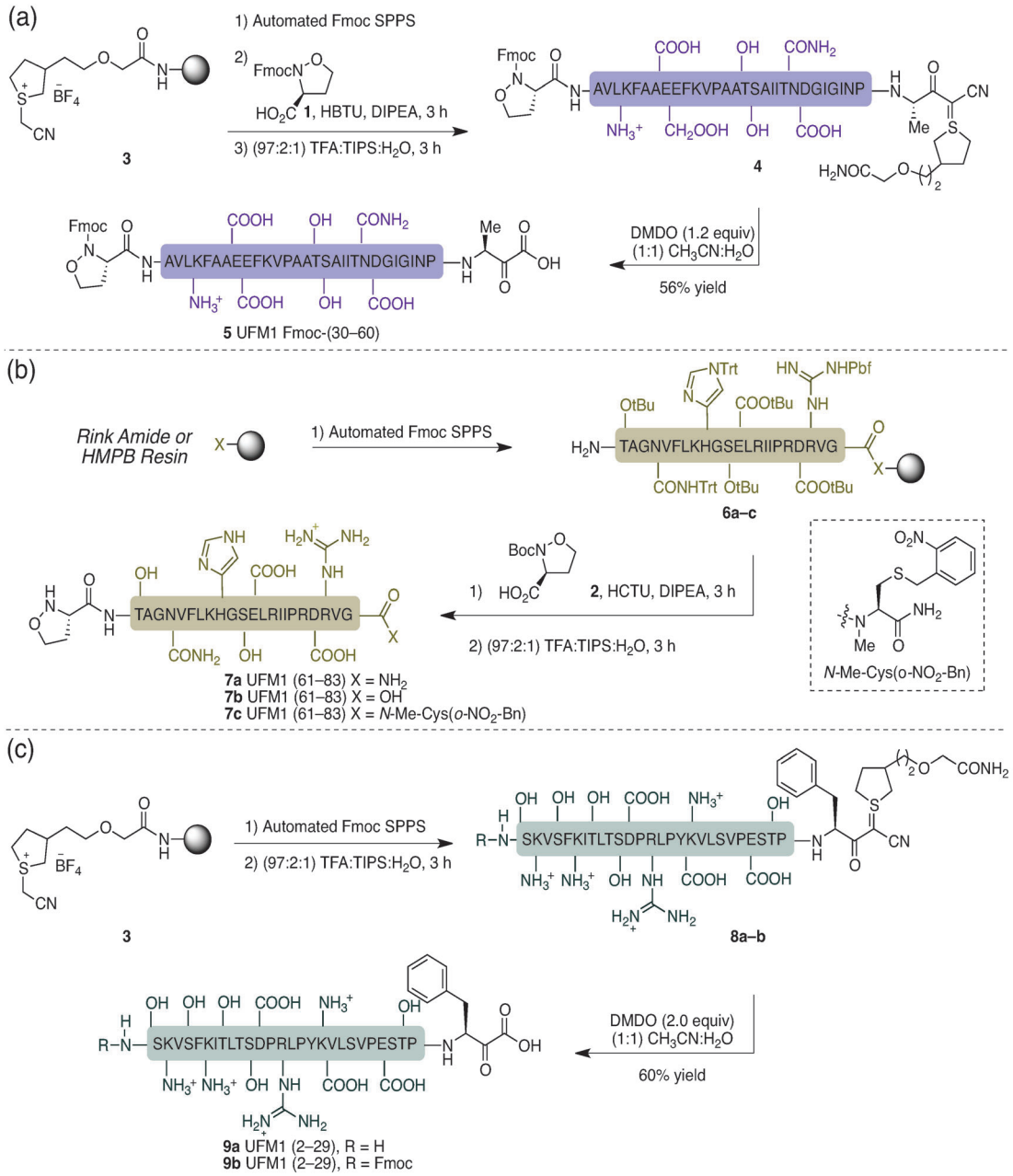
[*] A. O. Ogunkoya, Dr. V. R. Pattabiraman, Prof. Dr. J. W. Bode
Laboratorium für Organische Chemie, Department of Chemistry and Applied Biosciences, ETH Zürich
Wolfgang Pauli Strasse 10, 8093 Zürich (Switzerland)
E-mail: bode@org.chem.ethz.ch
Homepage: http://www.bode.ethz.ch

[**] This work was supported by the Swiss National Science Foundation (200021_131957) and an ETH Fellowship to V.R.P. (ETH FEL 06-10-1). We are grateful to Dr. Xiangyang Zhang and Rolf Häfliger of the LOC Mass Spectrometry Service for analyses.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201204144.

acid-labile protecting groups typically employed in SPPS, 2) resistant to oxidative conditions required to transform cyanosulfurylides (SY) to α-ketoacids (KA), and 3) stability during the ligation and purification of unprotected peptide segments. The use of groups cleaved by thiols, reduction, and metal catalysts such as Ns, Troc, and Alloc, were deemed incompatible with unprotected peptides or only partially fulfilled the outlined criteria. A survey of base-labile protecting groups suggested that they may not be compatible with unprotected functional groups or could be difficult to deprotect post-ligation. Despite these concerns—and lacking a better option—we elected to examine the use of Fmoc-based protection for sequential KAHA ligations with Opr. We therefore prepared enantiopure (*S*)-*N*-Fmoc-Opr **1** (see Supporting Information), using modified literature precedents.[18,19]

As part of our interest in synthetic and semi-synthetic studies of new modifier proteins, we sought to provide access to three C-terminal variants of UFM1: 1) the C-terminal acid, which could be conjugated to a target protein with an appropriate E1 ligase; 2) a suitably protected C-terminal thioester, which in combination with a target protein containing a 5-thiolysine, could be conjugated by native chemical ligation; 3) a C-terminal amide, which would serve as an



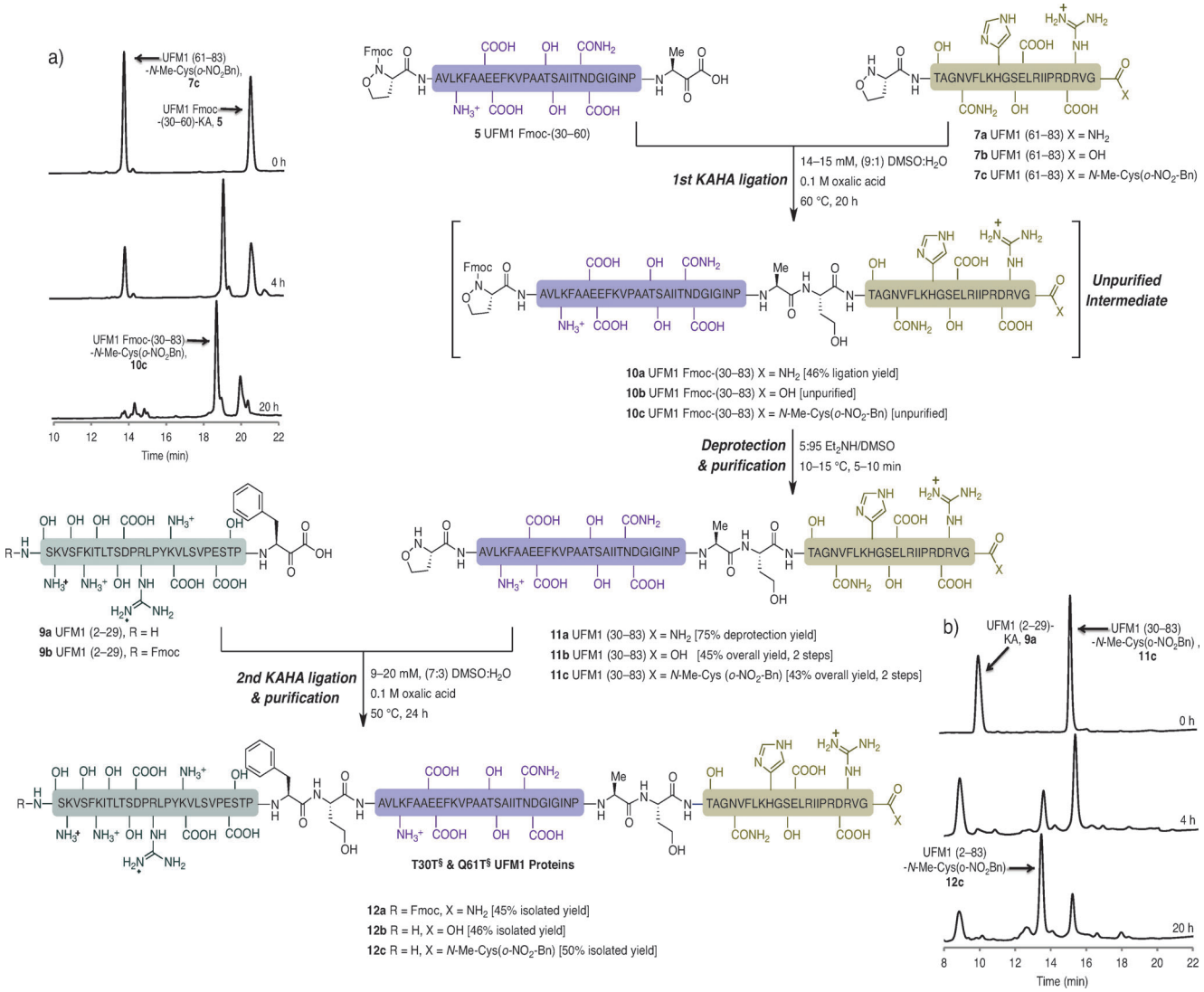
Scheme 1. Preparation of UFM1 peptide segments: a) Synthesis of UFM1 Fmoc-(30–60) **5**, b) UFM1 (61–83) **7a**, **7b**, and **7c**, c) and UFM1 (2–29) **9a** and **9b**. DIPEA = *N*,*N*-diisopropylethylamine, TFA = trifluoroacetic acid, TIPS = triisopropylsilane, Fmoc = 9-fluorenylmethoxycarbonyl, Trt = triphenylmethyl, Pbf = 2,2,4,6,7-pentamethyldihydro-1-benzofurane-5-sulfonyl, HCTU is a uronium-type coupling reagent.

important negative control in biochemical conjugation studies. The sequential ligation strategy with modular protein segments would allow the efficient preparation of C-terminal protein variants with minimum synthetic effort. In general, the preparation of C-terminal variants of proteins as opposed to the N-terminal modifications is arduous by linear SPPS, as it requires individual solid-phase synthesis of each of the C-terminal protein variant.[20]

Our sequential strategy for synthesis of UFM1 (2–83) variants required the preparation of three segments (Scheme 1). We selected ligation sites at Phe29-Thr30 and between Ala60-Gln61. This strategy would give synthetic C-terminal modified UFM1 with two mutations to the native sequence: T30T§ and Q61T§. It also allowed us to assess the viability of the KAHA ligation with N-terminal Opr segments at Ala and Phe-α-KA; our prior studies examined Leu, Tyr, and Glu.

Preparation of Fmoc-Opr-(31–60)-α-KA **5**, the foundation of this sequential strategy, was accomplished by using our previously established procedure for the synthesis of peptide α-KA (Scheme 1a). Cyanosulfurylide linker on Rink amide resin **3** was used to prepare UFM1 (31–60) by automated SPPS.[21] Fmoc-Opr **1** was manually coupled with HBTU. Side-chain deprotection and cleavage of the peptide from the resin gave **4** without any side products. Unprotected SY **4** was oxidized with dimethyldioxirane (DMDO) and purified by preparative HPLC to provide analytically pure α-KA **5**, leaving the N-terminal Opr30 unaffected. The Fmoc-protected Opr30 proved to be stable throughout the oxidation, purification and handling.

UFM1 (61–83) segments with three distinct C-termini and an unprotected N-terminal Opr61 (**7a–c**) (Scheme 1b) were prepared by standard methods.[18] The C-terminal amide segment **7a** and C-terminal acid segment **7b** were prepared by Fmoc SPPS on Rink amide and HMPB resins, respectively. Thioester precursor **7c** was prepared from Rink amide resin pre-loaded with Fmoc-*N*-Me-Cys(*o*-NO_2-Bn)-OH.[9] In each case, automated SPPS was used to extend the peptides to



Scheme 2. Synthesis of UFM1 proteins by sequential KAHA ligations: a) HPLC monitoring of the ligation between UFM1 Fmoc-(30–60)-KA **5** and (61–83)-*N*-Me-Cys(*o*-NO_2Bn) **7c**, b) HPLC monitoring of ligation between UFM1 (2–29)-KA **9a** and (30–83)-*N*-Me-Cys(*o*-NO_2Bn) **11c**.

residue 62 (**6a–c**) and (*S*)-*N*-Boc-Opr61 **2** was introduced by manual coupling with HCTU. Acid-promoted cleavage from the resin followed by purification gave unprotected peptides **7a–c**.

The N-terminal segments, UFM1 (2–29)-α-KA **9a** and **9b** (Scheme 1c), were prepared by SPPS on **3** followed by acidic resin cleavage, oxidation with DMDO and purification by HPLC. This procedure afforded the Phe α-KA with either a free N-terminus (**9a**) or an Fmoc-labeled N-terminus (**9b**).

Our studies on the assembly of the segments began with the ligation of α-KA **5** and Opr **7a**, using a small excess of the α-KA segment (1:1.3–2.0 equiv, respectively). In our previous report of the Opr KAHA ligation, we found that higher proportions of water accelerate the ligation.[18] Unfortunately, due to peptide solubility, it was difficult to perform ligations with acceptable concentrations of peptides (10–20 mM) containing higher than 10% water in DMSO (v/v). Nonetheless, with 10% water in DMSO and 0.1M oxalic acid at 60°C, the ligation (14–15 mM in peptides) proceeded smoothly within 20 h to give **10a** (Scheme 2a). Importantly, we observed no loss of Fmoc from the protected Opr on segment **5**. Purification of the ligation mixture by HPLC afforded the ligated product in 46% yield. Unlike our previous work and the second ligation (see below) we observed a small amount of an isomeric side-product (ca. 10%) corresponding to the same mass as the desired product.[22] This isomeric product was easily separated from the desired ligation product by preparative HPLC.

We were pleased that the Fmoc-protected Opr30 was stable to the ligation and purification conditions; the remaining milestone was its effective and high-yielding deprotection. Fmoc removal is typically conducted on fully protected peptides and a protocol for the rapid and high-yielding deprotection of otherwise unprotected ligation products required some development. The choice of base and reaction temperature proved essential to avoid a side product that was detected under some conditions. After careful optimization, we determined that treating peptide **10a** at 10–15°C with 5% Et_2NH in DMSO for 5–10 min cleanly gave the desired product **11a**. For development studies, we opted to purify the deprotected peptide by HPLC prior to examining the conditions for the final ligation. In practice, and in our subsequent syntheses of C-terminally modified UFM1 proteins, we performed the Fmoc-deprotection immediately after the ligation in the same reaction vessel. For example, the ligation mixture of **5** and **7b** to afford **10b** was cooled to 10–15°C and treated with a 10% solution of Et_2NH in DMSO (final concentration ca. 5% Et_2NH) for 10 min. The unprotected ligation product **11b** was purified and isolated by preparative HPLC in 45% overall yield. This procedure reduces the number of manipulations needed for sequential ligations and consequently improves the overall yields. We have adopted this in situ deprotection in all of our subsequent studies on the synthesis of UFM1 and other protein targets.

The final ligations required for the synthesis of UFM1 variants was initially examined with ligation product **11a** and Fmoc-labeled peptide **9b**, as this allowed for facile monitoring due to the Fmoc-chromophore. The ligation was performed by using 1.0 equiv **11a** and 1.4 equiv **9b** at 20 mM in 30% water in DMSO with 0.1M oxalic acid at 50°C. It was essentially complete within 24 h and provided **12a** in 45% isolated yield with no observable isomeric product (Scheme 2b). Qualitatively, this ligation was more efficient than the first and proceeded at lower temperature, despite the fact that a bulkier α-KA (Phe vs. Ala) was used for the ligation. We attribute the improved reactivity at lower temperature to the enhanced solubility of the peptides and increased proportion of water in the ligation mixture.

The sequential ligation procedures were readily extended to the synthesis of C-terminal UFM1 carboxylic acid **12b** and masked thioester **12c**. In all the cases, the first and second ligation, as well as the in situ Fmoc-deprotection proceeded under similar conditions and yields as described for the synthesis of UFM1 variant **12a**. Only minor variations in the reaction conditions due to solubility and reaction scale were made. For example, the second ligation of **9a** with **11b** to give **12b** was performed at 9 mM in 25% water in DMSO but was essentially finished after 24 h at 50°C.

The purity, sequence identity and exact mass of the final proteins were confirmed by analytical HPLC, high-resolution mass spectrometry (FTMS), MS/MS analysis and SDS-PAGE (Figure 2b). Also of importance for the synthetic UFM1 variants is their tertiary structure, which has a key role in its biological function. NMR studies of UFM1 protein have shown that it contains multiple β-strands and α-helical regions.[23] CD spectra recorded at 23°C for the synthetic UFM1 variants **12a**, **12b** and **12c** in 10 mM sodium phosphate buffer with 100 mM NaCl at pH 6.0 indicated the presence of folded proteins with β-strand and α-helical motifs (Figure 2a).

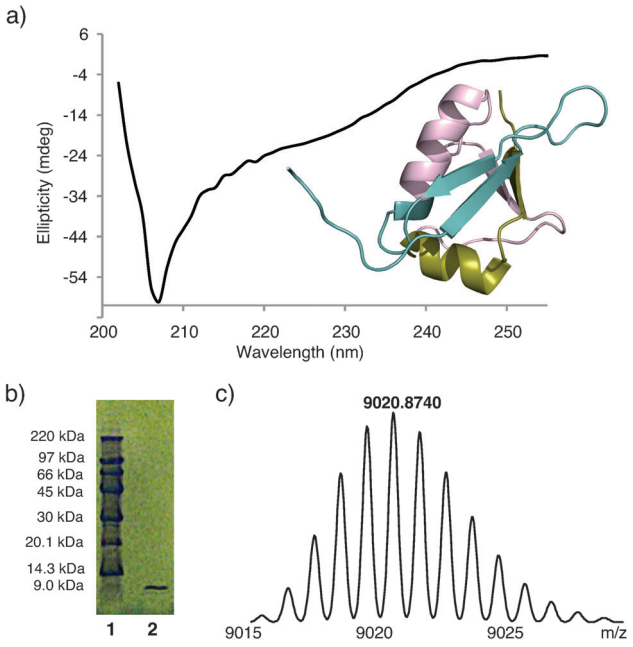


Figure 2. a) CD spectrum of UFM1 (2–83)-*N*-Me-Cys(*o*-NO_2-Bn) **12c** at 23°C; inset: solution NMR structure of UFM1.[23] b) SDS-PAGE of protein ladder (14.3–220 kDa, lane 1) and pure **12c** (9.0 kDa, lane 2). c) Deconvoluted ESI-FTMS of **12c**.

In conclusion, we have shown that Fmoc-protected 5-oxaproline is suitable for sequential KAHA ligation for protein synthesis. The development of this strategy made possible the facile synthesis of three C-terminal variants of UFM1, an important but under-studied modifier protein of contemporary interest. These are the largest proteins with native amide bonds prepared to date by segment ligations using a method other than native chemical ligation. These studies will both improve access to synthetic modifier proteins for biochemical studies and provide a new approach to iterative segment assembly for chemical protein synthesis.

Received: May 28, 2012
Published online: August 22, 2012

Keywords: amides · chemoselectivity · ligation reactions · protein synthesis · proteins

[1] a) M. Hochstrasser, *Nature* **2009**, *458*, 422 – 429; b) E. R. Strieter, D. A. Korasick, *Chem. Biol.* **2012**, *7*, 52 – 63.
[2] T. Fekner, X. Li, M. K. Chan, *ChemBioChem* **2011**, *12*, 21 – 33.
[3] P. Siman, A. Brik, *Org. Biomol. Chem.* **2012**, DOI: 10.1039/C2OB25149C.
[4] a) K. S. Ajish Kumar, M. Haj-Yahta, D. Olschewski, H. A. Lashuel, A. Brik, *Angew. Chem.* **2009**, *121*, 8234 – 8238; *Angew. Chem. Int. Ed.* **2009**, *48*, 8090 – 8094; b) R. Yang, K. K. Pasunooti, F. Li, X. W. Liu, C. F. Liu, *J. Am. Chem. Soc.* **2009**, *131*, 13592 – 13593; c) F. El Qualid, R. Merkx, R. Ekkebus, D. S. Hameed, J. J. Smit, A. de Jong, H. Hilkmann, T. K. Sixma, H. Ovaa, *Angew. Chem.* **2010**, *122*, 10347 – 10351; *Angew. Chem. Int. Ed.* **2010**, *49*, 10149 – 10153.
[5] S. N. Bavikar, L. Spasser, M. Haj-Yahya, S. V. Karthikeyan, T. Moyal, K. S. A. Kumar, A. Brik, *Angew. Chem.* **2012**, *124*, 782 – 787; *Angew. Chem. Int. Ed.* **2012**, *51*, 758 – 763.
[6] M. Hejjaoui, M. Haj-Yahya, K. S. A. Kumar, A. Brik, H. A. Lashuel, *Angew. Chem.* **2011**, *123*, 425 – 429; *Angew. Chem. Int. Ed.* **2011**, *50*, 405 – 409.
[7] S. Virdee, P. B. Kapadnis, T. Elliott, K. Lang, J. Madrzak, D. P. Nguyen, L. Riechmann, J. W. Chin, *J. Am. Chem. Soc.* **2011**, *133*, 10708 – 10711.
[8] K. S. A. Kumar, L. Spasser, L. A. Erlich, S. N. Bavikar, A. Brik, *Angew. Chem.* **2010**, *122*, 9312 – 9317; *Angew. Chem. Int. Ed.* **2010**, *49*, 9126 – 9131.
[9] K. S. A. Kumar, S. N. Bavikar, L. Spasser, T Moyal, S. Ohayon, A. Brik, *Angew. Chem.* **2011**, *123*, 6261 – 6265; *Angew. Chem. Int. Ed.* **2011**, *50*, 6137 – 6141.
[10] L. A. Erlich, K. S. A. Kumar, M. Haj-Yahya, P. E. Dawson, A. Brik, *Org. Biomol. Chem.* **2010**, *8*, 2392 – 2396.
[11] a) P. P. Geurink, F. El Qualid, A. Jonker, D. S. Hameed, H. Ovaa, *ChemBioChem* **2012**, *13*, 293 – 297; b) O. N. Burchak, M. Jaquinod, C. Cottin, L. Mugherli, K. Iwai, F. Chatelain, M. Y. Balakirev, *ChemBioChem* **2006**, *7*, 1667 – 1669.
[12] K. Kas, E. Schoenmakers, W. Van de Ven, G. Weber, M. Nordenskjold, L. Michiels, J. Merregaert, C. Larsson, *Genomics* **1993**, *17*, 387 – 392.
[13] B. P. Downes, S. A. Saracco, S. S. Lee, D. N. Crowell, R. D. Vierstra, *J. Biol. Chem.* **2006**, *281*, 27145 – 27157.
[14] M. Komatsu, T. Chiba, K. Tatsumi, S. Lemura, I. Tanida, N. Okazaki, T. Ueno, E. Kominami, T. Natsume, K. Tanaka, *EMBO J.* **2004**, *23*, 1977 – 1986.
[15] a) T. W. Muir, *Annu. Rev. Biochem.* **2003**, *72*, 297 – 303; b) R. M. Hofmann, *Curr. Opin. Biotechnol.* **2002**, *13*, 297 – 303; c) 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.
[16] a) I. E. Valverde, F. Lecaille, G. Lalmanach, V. Aucagne, A. F. Delmas, *Angew. Chem.* **2012**, *124*, 742 – 746; *Angew. Chem. Int. Ed.* **2012**, *51*, 718 – 722; b) Z. Miao, J. P. Tan, *J. Am. Chem. Soc.* **2000**, *122*, 4253 – 4260.
[17] a) M. Villain, J. Vizzavona, K. Rose, *Chem. Biol.* **2001**, *8*, 673 – 679; b) S. Ueda, M. Fujita, H. Tamamura, N. Fujii, A. Otaka, *ChemBioChem* **2005**, *6*, 1983 – 1986; c) G. Fang, Y. Li, F. Shen, Y. Huang, J. Li, Y. Lin, H. Cui, L. Lui, *Angew. Chem.* **2011**, *123*, 7787 – 7791; *Angew. Chem. Int. Ed.* **2011**, *50*, 7645 – 7649; d) N. Ollivier, J. Vicogne, A. Vallin, H. Drobecq, R. Desmet, O. El Mahdi, B. Leclercq, G. Goormachtigh, V. Fafeur, O. Melnyk, *Angew. Chem.* **2012**, *124*, 213 – 217; *Angew. Chem. Int. Ed.* **2012**, *51*, 209 – 213.
[18] V. R. Pattabiraman, A. O. Ogunkoya, J. W. Bode, *Angew. Chem.* **2012**, *124*, 5204 – 5208; *Angew. Chem. Int. Ed.* **2012**, *51*, 5114 – 5118.
[19] A. Vasella, R. Voeffray, *J. Chem. Soc. Chem. Commun.* **1981**, 97 – 98.
[20] J. Alsina, F. Albericio, *Biopolymers* **2003**, *71*, 454 – 477.
[21] a) L. Ju, J. W. Bode, *Org. Biomol. Chem.* **2009**, *7*, 2259 – 2264; b) L. Ju, A. R. Lippert, J. W. Bode, *J. Am. Chem. Soc.* **2008**, *130*, 4253 – 4255.
[22] J. Wu, J. Ruiz-Rodriguez, J. M. Cornstock, J. Z. Dong, J. W. Bode, *Chem. Sci.* **2011**, *2*, 1976 – 1979.
[23] H. Sasakawa, E. Sakata, Y. Yamaguchi, M. Komatsu, K. Tatsumi, E. Kominami, K. Tanaka, K. Kato, *Biochem. Biophys. Res. Commun.* **2006**, *343*, 21 – 26.

