Protein Ligation

DOI: 10.1002/anie.201407014

Traceless Preparation of C-Terminal α-Ketoacids for Chemical Protein Synthesis by α-Ketoacid–Hydroxylamine Ligation: Synthesis of SUMO2/3**

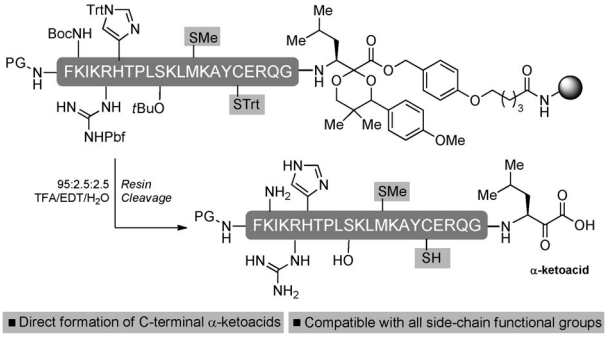
*Thomas G. Wucherpfennig, Vijaya R. Pattabiraman, Felix R. P. Limberg, Javier Ruiz-Rodríguez, and Jeffrey W. Bode**

Abstract: *A novel protecting group for enantiopure α-ketoacids delivers C-terminal peptide α-ketoacids directly upon resin cleavage and allows the inclusion of all canonical amino acids, including cysteine and methionine. By using this approach, SUMO2 and SUMO3 proteins were prepared by KAHA ligation with 5-oxaproline. The synthetic proteins containing homoserine residues were recognized by and conjugated to RanGAP1 by SUMOylation enzymes.*

The α-ketoacid–hydroxylamine ligation (KAHA ligation) makes possible the chemical synthesis of proteins from unprotected peptide segments.[1] It offers an alternative to the well-established native chemical ligation (NCL) developed by Kent and co-workers.[2] The KAHA ligation does not require a cysteine residue and permits ligation under aqueous acidic conditions, which are well-suited for solubilizing peptide segments. KAHA ligation is compatible with unprotected functional groups in the peptide side chain and has been used for the synthesis of proteins including Pup, cspA, and UFM1 by two and three-segment ligation strategies.[3]

The best implementation of this ligation uses a 5-oxaproline residue as the N-terminus of one segment, which leads to an innocuous but noncanonical homoserine residue ($T^{\S}$) at the ligation site.[4] The initial product of the KAHA ligation with 5-oxaproline is an ester, which readily rearranges to an amide in basic buffers.[5] A variety of α-ketoacids are suitable for the C-terminus of the other peptide segment, including leucine, phenylalanine, tyrosine, arginine, glutamic acid, and alanine. The preparation of the α-ketoacid requires oxidation of a cyanosulfurylide.[6] This oxidation is fast and reliable, but is not compatible with peptide segments containing cysteine or methionine residues; tryptophan can also be problematic. It also introduces an additional synthetic step after solid-phase peptide synthesis (SPPS) of the segment.

Herein, we introduce a protected form of leucine α-ketoacid that is suitable for obtaining the corresponding enantiopure peptide α-ketoacids directly upon cleavage from the resin (Scheme 1). No post-SPPS manipulations are needed and this method is compatible with all unprotected side chains, including cysteine, methionine, and tryptophan. In a continuation of our interest in the synthesis and study of modifier proteins,[7] we prepared SUMO3 by a three-segment ligation strategy, and SUMO2 with a pro-peptide by a four-segment strategy by using this new approach. This work complements the recent elegant synthesis of SUMO1 by Melnyk and co-workers using NCL.[8] In addition to documenting the new peptide α-ketoacid synthesis, this work also demonstrates 1) the viability of the KAHA ligation in the presence of unprotected cysteine residues; 2) ligation with valine α-ketoacids, and 3) confirmation that synthetic SUMO2 and SUMO3 with multiple homoserine residues are biochemically active.



Scheme 1. A protecting group for the traceless preparation of peptide α-ketoacids by Fmoc solid-phase peptide synthesis. Boc = *tert*-butoxycarbonyl, Pbf = 2,2,4,6,7-pentamethyl-2,3-dihydro-1-benzofuran-5-sulfonyl, PG = protecting group, TFA = trifluoroacetic acid, Trt = triphenylmethyl.

A traceless protecting group for the synthesis of C-terminal peptide α-ketoacids must meet the following strict criteria: 1) It must be stable to the reagents and basic conditions of the SPPS; 2) it must be removed under acidic

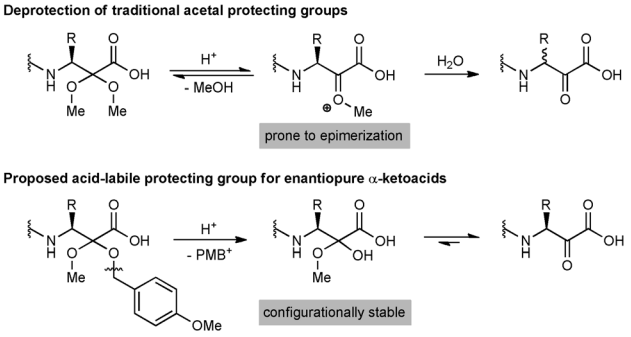
[*] T. G. Wucherpfennig, Dr. V. R. Pattabiraman, F. R. P. Limberg, Dr. J. Ruiz-Rodríguez, Prof. Dr. J. W. Bode
Laboratorium für Organische Chemie
Department of Chemistry and Applied Biosciences
ETH Zürich, 8093 Zürich (Switzerland)
E-mail: bode@org.chem.ethz.ch
Homepage: http://www.bode.ethz.ch
Prof. Dr. J. W. Bode
Institute of Transformative Bio-Molecules (WPI-ITbM)
Nagoya University, Chikusa, Nagoya 464-8602 (Japan)
Homepage: http://www.itbm.nagoya-u.ac.jp

[**] This work was supported by an ETH Research Grant (ETH-43 13–2) and the Swiss National Science Foundation (200020_150073). We are grateful to the mass spectrometry service of the Laboratorium für Organische Chemie at ETH Zürich, the Functional Genomics Center Zürich for analyses, and Michele Mari for the preparation of a SUMO3 segment.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201407014.

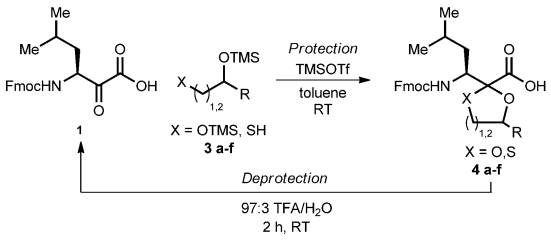
resin-cleavage conditions to directly afford the side-chain unprotected, C-terminal peptide α-ketoacid; 3) it must undergo deprotection through a mechanism that does not lead to epimerization, and 4) it must be readily prepared on sufficient scale and coupled to a resin by a suitable linker.[9] For initial development of protected α-ketoacids, we elected to focus on protections of Fmoc-leucine α-ketoacid **1**, as leucine is by far the most prevalent amino acid in proteins[10] and contains no side-chain functionality that could interfere with the α-ketoacid protection.[11]

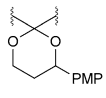
At the outset of our studies we were particularly concerned with the issue of epimerization. The simplest protection—masking the keto group as an acyclic dimethyl acetal—was feasible, but in some cases resulted in partial epimerization upon cleavage (Scheme 2). We postulated that

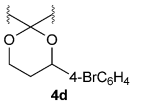
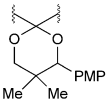
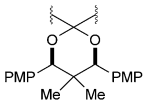


Scheme 2. Design of an α-ketoacid protecting group that is removable without epimerization.

Table 1: Evaluation of protecting groups for α-ketoacids.[a]



Protecting group	 **2**	 **4a**	 **4b**	 **4c**
protection yield	20%[b]	<5%	<5%	<5%
deprotection	yes	yes	no	yes
epimerization[c]	yes	–	–	–

Protecting group	 **4d**	 **4e**	 **4f**
protection yield	29%	28%[d] 66%[e]	N.R.
deprotection	no	yes	–
epimerization[c]	–	no	–

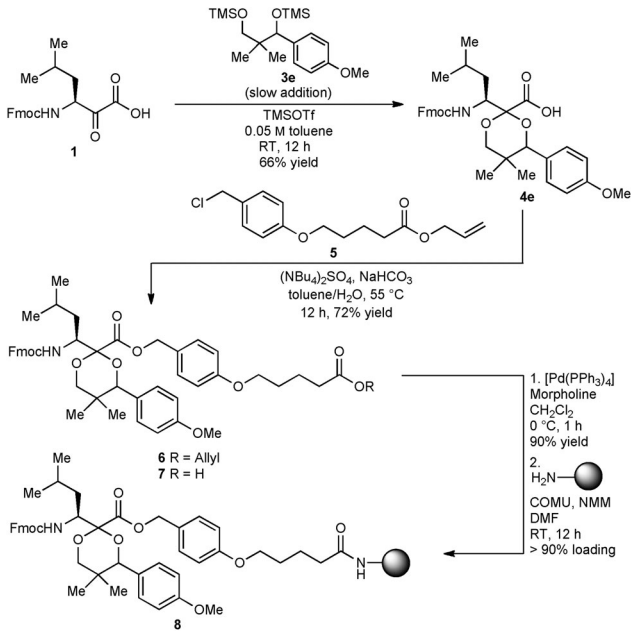
[a] See Supporting Information for a full list of protecting groups investigated. [b] Prepared from **1**, trimethylorthoformate, and sulfuric acid. [c] Determined by HPLC after ligation of the deprotected α-ketoacid. [d] Initial, unoptimized yield. [e] Yield after optimization. PMP = 4-methoxyphenyl.

a protecting group that could be removed by a mechanism involving the formation of a hemiacetal rather than an oxonium ion could overcome this problem.[12] Despite many attempts, we could not prepare the simplest implementation of this design with acyclic mono- or di-*para*-methoxybenzyl acetals. This prompted us to focus on a series of 1,2- and 1,3-diols that would give cyclic acetals (Table 1).

In early experiments, we found that 1,2-diol derivative **3a** could protect leucine α-ketoacid **1** to give **4a**, but only in poor yield.[13] Hydroxythiol **3b** also gave the product **4b** in low yield. Unlike the oxo derivative **4a**, the deprotection of **4b** was not successful under the desired conditions, which we attribute to the higher stability of the O,S-acetal compared to the O,O-acetal.[14] The use of 1,3-diol derivative **3c** did not offer any advantage. Changing the electronic properties of the aromatic ring to a less electron rich system by introducing a 4-bromo substituent led to increased yields of **4d**, but the deprotection was unfeasible. Attempts to derivatize **4d** by Cu- or Pd-catalyzed reactions were unsuccessful.[15] The *gem*-dimethyl-1,3-diol derivative **3e** provided the protected leucine α-ketoacid **4e** in better yield and allowed facile, epimerization-free deprotection to return the α-ketoacid. To avoid the formation of diastereomers arising from the racemic protecting group, we attempted protection using *meso*-1,3-diol **3f**, but the protection step was not successful. Based on these results, we selected *gem*-dimethyl compound **3e** for further optimization of the protection step (see the Supporting Information). Careful tuning of the reaction conditions and monitoring of by-product formation led to an optimized procedure, whereby the initial yield was increased from 28% to 66%. Although the protected leucine α-ketoacid **4e** is formed as a mixture of diastereomers because of the racemic protecting group **3e**, we elected to proceed with it as it is removed in a traceless manner during the resin cleavage step to give the enantiopure C-terminal leucine α-ketoacid. The easy availability of both the leucine α-ketoacid **1** and the racemic protecting group **3e**, along with the ability to perform the protection reaction on a decagram scale renders this procedure suitable to produce enough protected leucine α-ketoacid for routine use in protein synthesis.

Protected leucine α-ketoacid **4e** was coupled to a Wang-type linker[16] by alkylation with benzylic chloride **5**, followed by deprotection of the allyl ester to afford the free carboxylic acid **7** (Scheme 3). Loading **7** onto standard amino-functionalized polystyrene or a polyethylene glycol based solid support gives a stable, storable resin that readily delivers the C-terminal peptide α-ketoacid upon acidic cleavage of the resin. A critical aspect in using the leucine α-ketoacid protection strategy in Fmoc-SPPS was establishing a cleavage cocktail for liberating the fully unprotected peptide α-ketoacid segment without reduction or other harm to the α-ketoacid group.[17]

Certain commonly used scavengers, such as ethanethiol or thiophenol led to thioacetal adducts or inefficient scavenging. Fortunately, the non-malodorous thiol compound 2,2′-(ethylenedioxy)diethanethiol (DODT)[18] or 1,2-ethanedithiol



Scheme 3. Synthesis and immobilization of protected leucine α-ketoacid **4e**. COMU = 1-[(1-(cyano-2-ethoxy-2-oxoethylideneaminooxy) dimethylaminomorpholino)] uronium hexafluorophosphate, NMM = *N*-methylmorpholine, Tf = trifluoromethanesulfonyl, TMS = trimethylsilyl.

(EDT) act as effective scavengers to give the fully unprotected peptide α-ketoacids that can be purified by HPLC and stored as a lyophilized powder for several months.

To demonstrate the utility of the protected leucine α-ketoacid for protein synthesis we used it for the preparation of the modifier proteins SUMO2 and SUMO3. For SUMO3, we selected ligation sites at Val29-Gln30 and Leu52-Ser53, which would give two mutations when using the 5-oxaproline for the ligation: Q30T[§] and S53T[§]. This strategy reduces the synthesis of SUMO3 to three segments, each with about 30 residues. For the synthesis of the slightly longer SUMO2 with a pro-peptide by a four-segment strategy we chose ligation sites at Val30-Gln31, Leu53-Ser54, and Leu76-Glu77, which would introduce three homoserine residues in the final protein.

The segments were assembled into SUMO2/3 protein by three or four segment ligations as shown in Scheme 4A. For the preparation of SUMO3 (53–92) **12** and SUMO2 (54–76) **10**, the use of Fmoc-Asp(O*t*Bu)-N(Dmb)Gly-OH during SPPS proved critical for avoiding aspartimide side products. For the longer SUMO2 with the pro-peptide, we chose a four-segment ligation strategy that required bifunctional α-ketoacid **10**—prepared using the protected leucine α-ketoacid resin—and oxaproline segment **9.** The ligation proceeded smoothly to give the ester product, and the O to N acyl shift at pH 9.5 followed by Fmoc cleavage gave SUMO2 (54–95) **11** after HPLC purification.

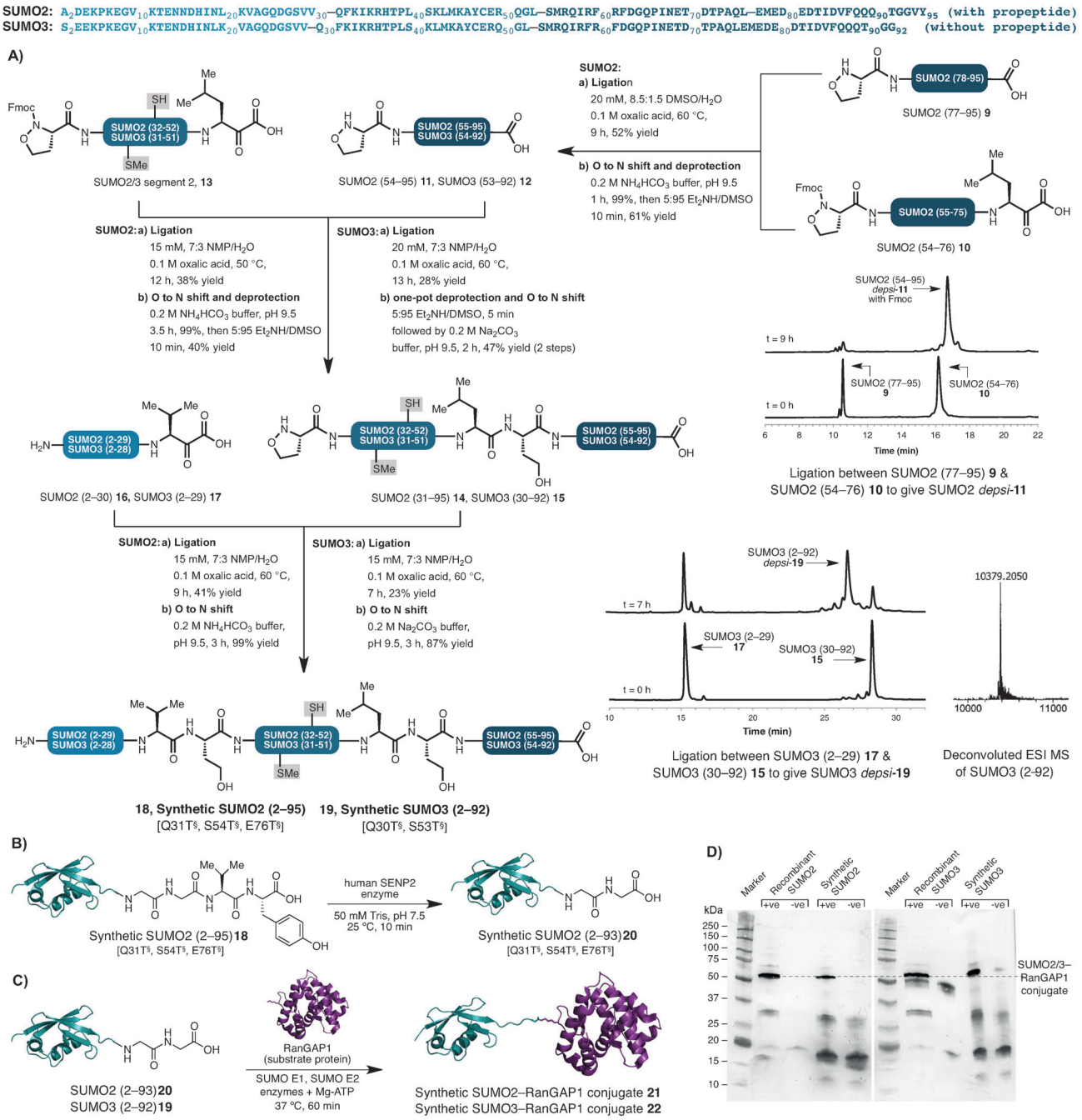
The middle regions of SUMO2 and SUMO3 share an identical sequence containing cysteine and methionine residues, which was prepared as Fmoc-protected peptide α-ketoacid **13** using our newly developed resin **8**. Ligation with the corresponding oxaproline segments of SUMO2 (54–95) **11** or SUMO3 (53–92) **12** proceeded readily to give the ligation products SUMO2 (31–95) *depsi*-**14** and SUMO3 (30–92) *depsi*-**15**, respectively. The esters were easily converted into the amide products either by a one-pot Fmoc cleavage and O to N acyl shift or by a two-step sequence. These examples are the first published reports of the KAHA ligation in the presence of an unprotected cysteine residue. While performing these ligations, we observed that distilled and degassed NMP/H_2O often delivered better results than DMSO/H_2O mixtures, presumably because of oxidation of the cysteine and methionine residues in DMSO.

The C-terminal valine α-ketoacids SUMO2 (2–30) **16** and SUMO3 (2–29) **17** were prepared uneventfully by oxidation of the corresponding cyanosulfurylides, which were synthesized using our previously reported linker.[6] The final ligations with C-terminal valine α-ketoacids SUMO2 (2–30) **16** and SUMO3 (2–29) **17** were complete within 10 h at 60 °C, despite the use of the more sterically demanding valine residue. The target proteins SUMO2 (2–95) **18** and SUMO3 (2–92) **19** were obtained after O to N acyl shifts to convert the esters into the amides. The ability to use valine α-ketoacids in the KAHA ligation increases the number of disconnection sites suitable for chemical protein synthesis.

To verify that the two or three homoserine residues in synthetic SUMO3 **19** and SUMO2 **18** do not disturb the protein structure or function, we initiated biochemical studies. The SUMO protease SENP2, unlike many other proteases, requires the characteristic SUMO fold to perform its proteolytic function.[19] To evaluate whether SUMO2 (2–95) **18** with the pro-peptide would be recognized and processed by SENP2, the synthetic protein **18** was incubated in pH 7.5 Tris buffer to allow folding, followed by the addition of human SENP2 protease. Within minutes, the proSUMO2 (2–95) **18** was completely converted into the active SUMO2 (2–93) **20** to reveal the C-terminal Gly-Gly motif (Scheme 4B). This demonstrates that the synthetic SUMO2 protein **18** with three homoserine residues adopts the characteristic SUMO fold similar to the wild-type protein.

We performed in vitro SUMOylation reactions to further evaluate whether SUMO activation and conjugation enzymes will recognize SENP2-processed synthetic SUMO2 (2–93) **20** or synthetic SUMO3 (2–92) **19** and SUMOylate a protein of interest, RanGAP1 (Scheme 4C).[20] Separating the products of the enzymatic reaction by SDS-PAGE followed by Western blotting using a SUMO2/3 antibody revealed that the synthetic SUMO2/3 proteins were readily accepted by the SUMOylation enzymes to generate synthetic SUMO2-RanGAP1 conjugate **21** and SUMO3-RanGAP1 conjugate **22** (Scheme 4D). This result further demonstrates that the synthetic SUMO2/3 proteins containing the homoserine mutations[21] exhibit full activity comparable to the natural SUMO2/3.

In summary, we have developed a protecting group for α-ketoacids that allows the inclusion of all canonical amino acids, including cysteine, methionine, and tryptophan in SPPS and delivers the C-terminal peptide α-ketoacid directly upon cleavage of the resin. We successfully utilized the newly developed resin for the synthesis of the α-ketoacid segments necessary for the preparation of SUMO2/3 proteins by



Scheme 4. A) Synthesis of SUMO2 and SUMO3 proteins by sequential KAHA ligations. B) Biochemical processing of synthetic SUMO2 **18** with pro-peptide to its active form by the SENP2 enzyme. C) In vitro conjugation of synthetic SUMO2/3 to RanGAP1 using SUMOylation enzymes. D) Western blot analysis of SUMOylation reactions with synthetic SUMO proteins using a SUMO2/3 antibody. The SUMO2/3-RanGAP1 conjugate bands appear at about 50 kDa on the Western blot. NMP = *N*-methylpyrrolidine.

sequential KAHA ligations. Additionally, we provide biochemical evidence that the homoserine residues introduced by the KAHA ligation with 5-oxaproline in SUMO2/3 do not affect the in vitro recognition and processing by the SUMOylation machinery.

Received: July 8, 2014
Published online: September 22, 2014

Keywords: amides · ligation · protecting groups · protein modification · protein synthesis

[1] a) J. W. Bode, R. M. Fox, K. D. Baucom, *Angew. Chem. Int. Ed.* **2006**, *45*, 1248–1252; *Angew. Chem.* **2006**, *118*, 1270–1274; b) J. Wu, J. Ruiz-Rodriguez, J. M. Comstock, J. Z. Dong, J. W. Bode, *Chem. Sci.* **2011**, *2*, 1976–1979.
[2] a) P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776–779; b) S. B. H. Kent, *Chem. Soc. Rev.*

2009, *38*, 338 – 351; c) P. E. Dawson, *Isr. J. Chem.* **2011**, *51*, 862 – 867; d) T. Moyal, H. P. Hemantha, P. Siman, M. Refua, A. Brik, *Chem. Sci.* **2013**, *4*, 2496 – 2501; e) J. S. Zheng, S. Tang, Y. K. Qi, Z. P. Wang, L. Liu, *Nat. Protoc.* **2013**, *8*, 2483 – 2495; f) L. Raibaut, N. Ollivier, O. Melnyk, *Chem. Soc. Rev.* **2012**, *41*, 7001 – 7015; g) H. P. Hemantha, N. Narendra, V. V. Sureshbabu, *Tetrahedron* **2012**, *68*, 9491 – 9537; h) C. P. R. Hackenberger, D. Schwarzer, *Angew. Chem. Int. Ed.* **2008**, *47*, 10030 – 10074; *Angew. Chem.* **2008**, *120*, 10182 – 10228.

[3] A. O. Ogunkoya, V. R. Pattabiraman, J. W. Bode, *Angew. Chem. Int. Ed.* **2012**, *51*, 9693 – 9697; *Angew. Chem.* **2012**, *124*, 9831 – 9835.

[4] V. R. Pattabiraman, A. O. Ogunkoya, J. W. Bode, *Angew. Chem. Int. Ed.* **2012**, *51*, 5114 – 5118; *Angew. Chem.* **2012**, *124*, 5204 – 5208.

[5] T. G. Wucherpfennig, F. Rohrbacher, V. R. Pattabiraman, J. W. Bode, *Angew. Chem. Int. Ed.* **2014**, DOI: 10.1002/anie.201406097; *Angew. Chem.* **2014**, DOI: 10.1002/ange.201406097.

[6] 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.

[7] a) M. Hochstrasser, *Nature* **2009**, *458*, 422 – 429; b) R. T. Hay, *Mol. Cell* **2005**, *18*, 1 – 12; c) A. Flotho, F. Melchior, *Annu. Rev. Biochem.* **2013**, *82*, 357 – 385; d) R. Geiss-Friedlander, F. Melchior, *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 947 – 956.

[8] E. Boll, H. Drobecq, N. Ollivier, L. Raibaut, R. Desmet, J. Vicogne, O. Melnyk, *Chem. Sci.* **2014**, *5*, 2017 – 2022.

[9] a) C. D. Chang, J. Meienhofer, *Int. J. Pept. Protein Res.* **1978**, *11*, 246 – 249; b) I. Coin, M. Beyermann, M. Bienert, *Nat. Protoc.* **2007**, *2*, 3247 – 3256.

[10] B. L. Nilsson, M. B. Soellner, R. T. Raines in *Annual Review of Biophysics and Biomolecular Structure*, *Vol. 34*, Annual Reviews, Palo Alto, **2005**, pp. 91 – 118.

[11] The preparation of other peptide α-ketoacids are underway and will be reported in due course.

[12] E. H. Cordes, H. G. Bull, *Chem. Rev.* **1973**, *74*, 581 – 603.

[13] T. Tsunoda, M. Suzuki, R. Noyori, *Tetrahedron Lett.* **1980**, *21*, 1357 – 1358.

[14] a) D. P. N. Satchell, R. S. Satchell, *Chem. Soc. Rev.* **1990**, *19*, 55 – 81; b) J. L. Jensen, W. P. Jencks, *J. Am. Chem. Soc.* **1979**, *101*, 1476 – 1488.

[15] a) C. W. Cheung, S. L. Buchwald, *Org. Lett.* **2013**, *15*, 3998 – 4001; b) R. A. Altman, A. Shafir, A. Choi, P. A. Lichtor, S. L. Buchwald, *J. Org. Chem.* **2008**, *73*, 284 – 286.

[16] S. S. Wang, *J. Am. Chem. Soc.* **1973**, *95*, 1328 – 1333.

[17] D. S. King, C. G. Fields, G. B. Fields, *Int. J. Pept. Protein Res.* **1990**, *36*, 255 – 266.

[18] A. Teixeira, W. E. Benckhuijsen, P. E. de Koning, A. R. Valentijn, J. W. Drijfhout, *Protein Pept. Lett.* **2002**, *9*, 379 – 385.

[19] a) D. Reverter, C. D. Lima, *Structure* **2004**, *12*, 1519 – 1531; b) D. Reverter, C. D. Lima, *Nat. Struct. Mol. Biol.* **2006**, *13*, 1060 – 1068.

[20] A. Werner, M. C. Moutty, U. Moller, F. Melchior, *Methods Mol. Biol.* **2009**, *497*, 187 – 199.

[21] L. Baumann, A. G. Beck-Sickinger, *Angew. Chem. Int. Ed.* **2013**, *52*, 9550 – 9553; *Angew. Chem.* **2013**, *125*, 9729 – 9732.

