Solid-Phase Chemical Ligation

DOI: 10.1002/ange.201206428

Towards the Simplification of Protein Synthesis: Iterative Solid-Supported Ligations with Concomitant Purifications**

Vincent Aucagne, Ibai E. Valverde, Philippe Marceau, Mathieu Galibert, Nabil Dendane, and Agnès F. Delmas**

Chemical synthesis is a complementary alternative to recombinant approaches for the production of small (40–80 amino acids) to medium-sized (60–200 amino acids) proteins and analogues. Protein assembly through chemoselective coupling of unprotected peptide segments, that is chemical ligation,[1] has become the state-of-the-art strategy thanks to the pioneering work of Kemp and co-workers,[2] and the development of native chemical ligation (NCL) by Kent and co-workers.[3,4] More recently, mechanistically unrelated amide-forming ligations,[5] and peptidomimetic techniques[6] have been developed.

The ligation of two peptides is inherently limited to the synthesis of relatively small proteins. To expand the range of targets accessible by chemical synthesis to larger and more-complex proteins, sophisticated, iterative or convergent synthetic strategies are required.[7] Thus, ligations of more than three segments, and numerous nontrivial purification steps, would be required; the handling of the ligation product intermediates is often tricky, owing to aggregation or limited solubility.

Peptide segments are typically obtained through solid-phase peptide synthesis (SPPS), which is generally limited to the synthesis of peptides containing up to 40–50 amino acids.[8] Decades of continuous efforts in optimizing coupling reagents and protective groups have eliminated most prominent side reactions, but incomplete amide couplings are still a significant limitation in Fmoc-based SPPS. Post-coupling acetylations result in the capping of unreacted amine groups and the consequent N-truncated peptides constitute most of the by-products formed in peptide synthesis. We reasoned that the assembly of a protein by iterative solid-supported ligations of unpurified peptide segments in the opposite direction to SPPS, that is the N-to-C direction, should display self-purification[9] features, as all truncated impurities would be left unreacted (Figure 1b). This approach has never been exploited until the present work, despite the fact that a wide variety of related nonchromatographic purification strategies were developed during the 1980s and 1990s.[10,11] Most of these strategies rely on the post-elongation chemical tagging of the target peptide with an N-terminal linker that differentiates it from the truncated by-products (Figure 1a). After cleavage from the SPPS resin, the tagged molecules can specifically be captured on a second, hydrophilic solid support through a chemoselective reaction[10] or noncovalent interactions.[11] A simple filtration work-up separates the truncated peptides, and the subsequent cleavage of the tag results in a traceless release of the purified peptide into solution.

A few of these approaches[10c,d] have been utilized for solid-phase NCLs,[12] but the corresponding studies focus only on minimizing the chromatographic and handling steps.[13] All the examples of sequential solid-phase ligations reported so far were limited to polypeptides smaller than 100 amino acids.

We thus decided to re-examine the potential of applying this strategy to develop a self-purifying method by using contemporary ligation techniques, as we believe that new reagents or methods that allow for facile and rapid purification of the starting materials and intermediates are critical to the field of protein chemical synthesis. We disclose herein a new N-terminal linker[14] and its application to the non-chromatographic purification of peptides that are difficult to synthesize by existing methods, as well as to self-purifying N-to-C ligations. The latter approach is exemplified by an unprecedented synthesis of a 160-amino-acid triazole-containing analogue of the tandem repeat region of human MUC1[15] protein through successive N-to-C peptidomimetic triazole ligations[6b] of four segments.

We designed N_3-Esoc (2-[2-(2-azido-ethoxy)-ethyl sulfonyl]-2-ethoxycarbonyl, Scheme 1) as an N-terminal cleavable linker functionalized with an azide moiety. The versatility of this functional group makes it useful for many types of chemoselective couplings such as Cu^I-catalyzed[16] (this work) and strain-promoted[17] azide/alkyne cycloadditions (CuAAC and SPAAC, respectively) or Staudinger ligations.[5a,18] These reactions are superior to the ones employed in previous capture/release approaches, because of their better chemoselectivity, their nonreversible nature, and because of the chemical inertness of the resulting heterocyclic or amide linkages. These linkages should be compatible with a wide range of post-immobilization modifications and thus this method should extend the range of applications of solid-support capture approaches.[9,10] The Esoc linker is based on a alkylsulfonylethoxycarbonyl core,[19,10a–c,11d,f] which is stable

[*] Dr. V. Aucagne, Dr. I. E. Valverde, P. Marceau, Dr. M. Galibert, Dr. N. Dendane, Dr. A. F. Delmas
Centre de Biophysique Moléculaire, CNRS UPR 4301
Rue Charles Sadron, 45071 Orléans cedex 2 (France)
E-mail: aucagne@cnrs-orleans.fr
delmas@cnrs-orleans.fr

[**] The Région Centre council is gratefully acknowledged for financial support (FibroCat project and PhD fellowship for I.E.V.), as is ANR (JCJC08-0138, SynProt project) and la Ligue contre le cancer. We thank the CBM spectrometric platforms, and in particular Guillaume Gabant for recording most of the MALDI-TOF spectra and Hervé Meudal for the NMR spectra. We also thank Violette Senille, Dominique Lelièvre, and Dr. Chrystelle Derache for peptide synthesis.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201206428.

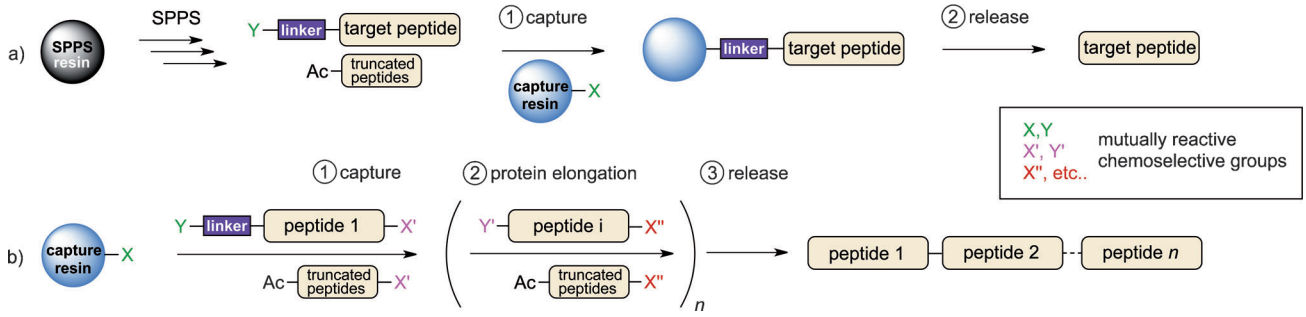
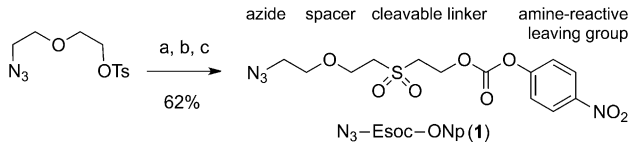


Figure 1. a) Capture and release approach for nonchromatographic purification of synthetic peptides from truncated coproducts. b) Extension to protein synthesis through multiple solid-phase ligations (this work).



Scheme 1. Synthesis of N_3-Esoc-ONp (**1**). a) $HSCH_2CH_2OH$, NaOH, MeOH/$H_2$0,79%; b) *m*CPBA, CH_2Cl_2, 86%; c) *p*-NO_2-C_6H_5OCOCl, pyridine, CH_2Cl_2, 91%.

under a wide range of reaction conditions including strong acids and neutral aqueous conditions, as well as being labile at slightly higher pH as a result of β-elimination. A short diethyleneglycol-like hydrophilic spacer confers good water-solubility properties to the N_3-Esoc-tagged peptides and provides a nonhindered environment to the azide group. Activation as a *p*-nitrophenyl carbonate results in a reagent (N_3-Esoc-ONp; **1**; see the Supporting Information for the detailed synthesis)[20] that readily reacts with N-terminal amino groups of protected solid-supported peptides under standard reaction conditions.

We synthesized N_3-Esoc-ARYG-OH and conjugated it to a water soluble alkyne by CuAAC to give **2** (Figure 2). This tetrapeptide was used as a model substrate for a preliminary evaluation of the lability of the Esoc linker under solution-phase reaction conditions. As expected, **2** was stable to acids, but was quickly and cleanly converted into H-ARYG-OH upon treatment with a basic buffer solution (< 15 min at pH 10.7). Esoc proved to be stable under neutral conditions (< 1% cleavage over 3 days at pH 7; see the Supporting Information, Figure S2).

Next, peptide N_3-Esoc-WPPAHGVTSAPDTRPAPGSTA-OH was used to evaluate the lability of an Esoc linker attached to a solid support. The peptide was attached by CuAAC onto a set of polyethylene glycol based hydro-

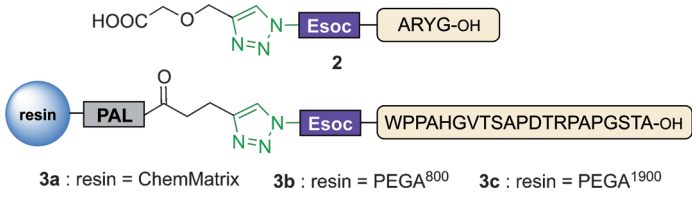


Figure 2. Model peptides used to evaluate the stability of the Esoc linker in solution (**2**) and under solid-supported conditions (**3a–c**).

compatible resins (ChemMatrix,[21] $PEGA^{800}$, and $PEGA^{1900}$[22]) to give **3a–c** (Figure 2). An additional acid-labile 5-[3,5-dimethoxy-4-(aminomethyl)phenoxy]pentanoic acid (PAL)[23] linker was inserted; such a double-linker strategy greatly facilitates reaction monitoring.[24] The base-mediated cleavages of Esoc for **3a–c** were significantly slower than those observed for **2** (see the Supporting Information, Figure S3); this result is presumably due to polymeric supports.[25] *N*-Cyclohexyl-3-aminopropanesulfonic acid (CAPS) pH 11.7 buffer was found to be very effective for cleavage of the Esoc linker; more than 90% release of the expected peptide (**S9**, see the Supporting Information) was observed in all cases upon 30 minute incubation at room temperature.

With a clean release procedure established, we applied our strategy to difficult peptide syntheses. The tests were conducted using $PEGA^{800}$ resin equipped with a reporter PAL linker. The potential and limitation of a CuAAC/N_3-Esoc capture/release purification method is illustrated with its application to three diverse biologically-relevant peptides that are known to be difficult-to-obtain sequences that lead to numerous truncated peptides (see the Supporting Information). The standard protocol was applied to the D enantiomer of the chicken beta-defensin (AvBD2; 36 amino acids), with acetamidomethyl (Acm) protecting groups on its six cysteine thiol side-chains, for biological evaluation.[26]

The solid-supported synthesis of the native L-AvBD2 proved to be quite challenging,[27] and we had to rigorously optimize the elongation conditions, to include three pseudoproline dipeptides together with a polar solid support.[28] Unfortunately, D-pseudoproline or the equivalent building blocks are not readily available so far, and therefore standard protective groups were used for the synthesis of D-AvBD2-$(Acm)_6$. This gave a moderate 40% elongation yield[29] (Figure 3, red RP-HPLC trace). This peptide was tagged with N_3-Esoc before resin cleavage; the chromatogram of this N_3-Esoc-tagged peptide showed a shifted peak (Figure 3, blue trace) and peaks corresponding to seven different truncated peptides. This result highlights the potential of capture/release methods for challenging purifications: a classical HPLC purification could have led to low-purity material with a misleadingly clean HPLC trace. CuAAC-mediated capture was undertaken under deoxygenated reaction conditions and using an excess amount of tris(3-hydroxypropyltriazolylmethyl)amine (THPTA) ligand to prevent oxidation of histidine,

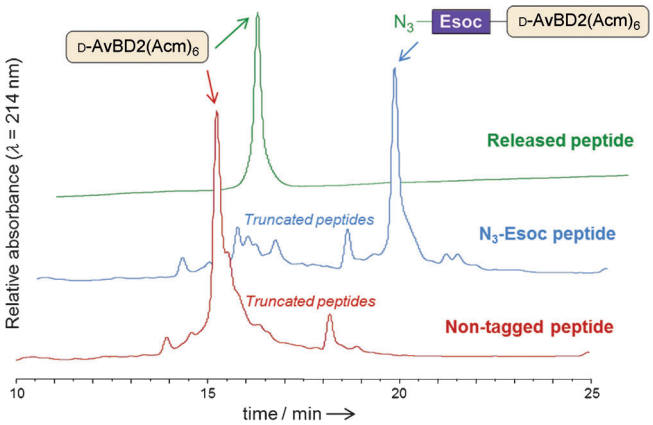


Figure 3. Capture and release purification of D-AvBD2(Acm)$_6$. Red trace: RP-HPLC chromatogram of the crude peptide. Blue trace: N_3-Esoc-tagged peptide. This trace shows a right shift of the major peak compared to the crude peptide (red trace). Green trace: peptide purified by resin capture followed by tag cleavage.

methionine, and cysteine side chains.[30] Such standard precautions were used for all other CuAAC reactions. CAPS-mediated release resulted in a peptide of high purity (green trace, Figure 3) and the yield was good (ca. 30% overall yield, > 95% purity).[31]

Having demonstrated the robustness and efficiency of N_3-Esoc as an N-terminal purification auxiliary, it was further exploited as a linker in iterative solid-supported N-to-C ligations, to demonstrate the concomitant purification potential of this strategy. This potential was illustrated by using our recently published peptidomimetic triazole ligation[6b] based on CuAAC and on the amide-bond-mimicking properties of triazoles. This method, involving the assembly of azido alkyne peptide segments equipped with alkyne-masking silyl groups, is particularly well suited for N-to-C ligations.[32]

The human MUC1 mucin was chosen as a suitable synthetic target.[15] The extracellular domain of this transmembrane protein possesses a tandem repeat domain composed of a 20 amino acid sequence (STAPPAHGVTSAPDTRPAPG), repeated between 30 and 150 times and heavily *O*-glycosylated. Such tandem repeat domains are benchmark examples to test an iterative ligation process, as the syntheses of all the segments require only one SPPS elongation. The forty-residue segments **4** and **5** (Figure 4) were synthesized following a previously described backbone-amide-linker (BAL) strategy. The Fmoc SPPS elongation proceeded in an excellent yield of greater than 90%,[29] but as many as eleven different truncated by-products were identified (see the Supporting Information, Figure S16). Attachment of **5** to the resin yielded **6**, which was cleanly desilylated upon treatment with $AgNO_3$ in water.[32b] Protein elongation was accomplished through three successive ligations by using a slight excess amount of **4** (1.5 equiv). A capping step, involving a CuAAC with a large excess amount of a water-soluble azide, was carried out before each desilylation, a procedure reminiscent to SPPS acetylation. Esoc cleavage led to the expected 160 residue tris(triazolo) polypeptide **7**, which was shown to be > 80% pure by HPLC ; this result is

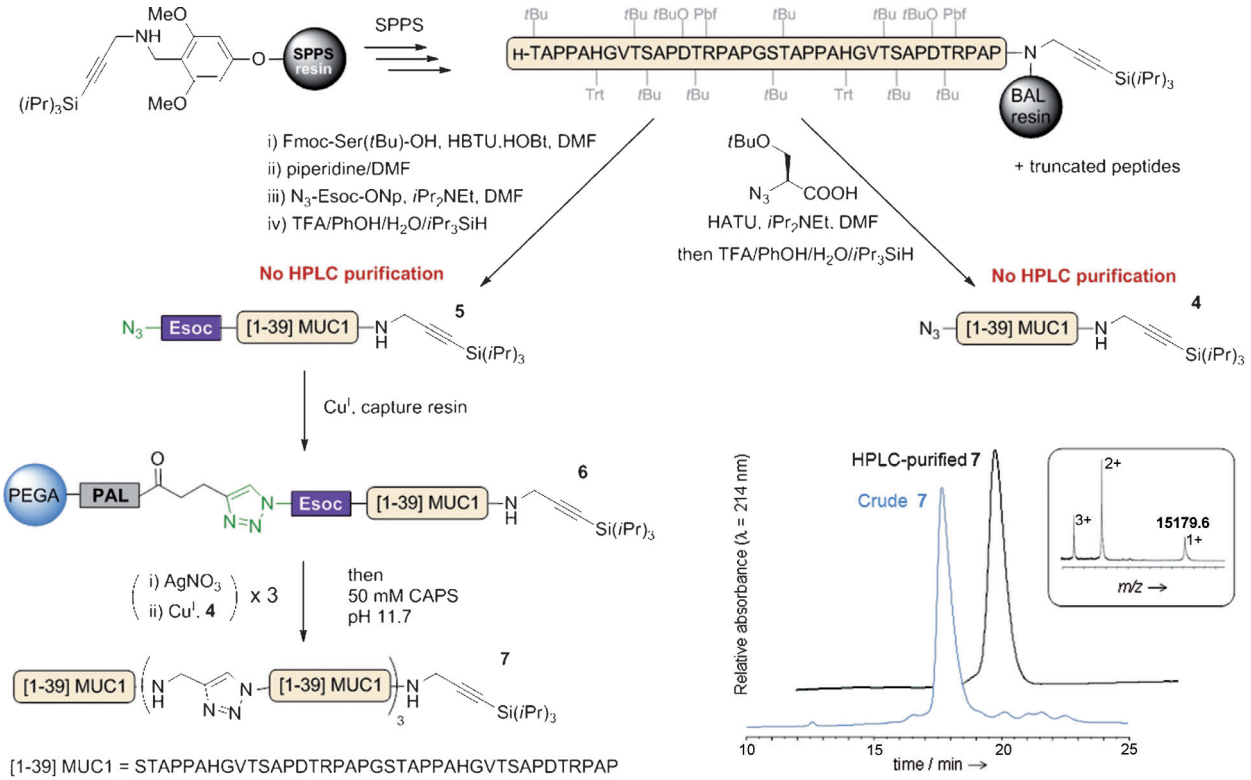


Figure 4. Synthesis of a 160 residue MUC1 fragment through iterative solid-supported ligations with concomitant purifications. DMF = *N*,*N*′-dimethylformamide, Fmoc = 9-fluorenylmethyloxycarbonyl, HBTU = *O*-(1*H*-benzotriazoyl-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium hexafluorophosphate, HOBt = 1-hydroxybenzotriazole.

impressive considering that none of the segments have been purified by chromatography. A single HPLC purification afforded **7** in good purity and excellent overall yield (ca. 60% based on **5**).

To our knowledge, this report constitutes the first example of a solid-phase chemical ligation of such a long polypeptide from more than three segments. The outstanding purity of the crude target polypeptide highlights the self-purification feature of successive solid-supported N-to-C ligations.

In conclusion, we introduced N_3-Esoc-ONp **1** as a new reagent for the temporary tagging of peptides at their N-terminus and demonstrated its potential in capture/release purifications. We extended this concept to a promising self-purifying N-to-C iterative triazole ligation strategy, exemplified by the synthesis of a polypeptide containing 160 residues. Solid-phase coupling of four peptide segments led to high purity material without any chromatographic purification steps. We believe that our reinspection of an early concept in the light of contemporary chemical ligation methods could have a major impact in regard to the simplification of the total synthesis of proteins.

Received: August 9, 2012
Published online: October 10, 2012

Keywords: chemical ligation · click chemistry · proteins · solid-phase synthesis · synthetic methods

[1] M. Schnölzer, S. Kent, *Science* **1992**, *256*, 221 – 225.
[2] a) D. Kemp, N. Galakatos, *J. Org. Chem.* **1986**, *51*, 1821 – 1829; b) N. Fotouhi, N. Galakatos, D. Kemp, *J. Org. Chem.* **1989**, *54*, 2803 – 2817.
[3] P. Dawson, T. Muir, I. Clark-Lewis, S. Kent, *Science* **1994**, *266*, 776 – 779.
[4] a) C. P. R. Hackenberger, D. Schwarzer, *Angew. Chem.* **2008**, *120*, 10182 – 10228; *Angew. Chem. Int. Ed.* **2008**, *47*, 10030 – 10074; b) S. B. H. Kent, *Chem. Soc. Rev.* **2009**, *38*, 338 – 351.
[5] a) B. L. Nilsson, L. L. Kiessling, R. T. Raines, *Org. Lett.* **2000**, *2*, 1939 – 1941; b) V. R. Pattabiraman, A. O. Ogunkoya, J. W. Bode, *Angew. Chem.* **2012**, *124*, 5204 – 5208; *Angew. Chem. Int. Ed.* **2012**, *51*, 5114 – 5118; c) I. Pusterla, J. W. Bode, *Angew. Chem.* **2012**, *124*, 528 – 531; *Angew. Chem. Int. Ed.* **2012**, *51*, 513 – 516.
[6] a) N. Assem, A. Natarajan, A. K. Yudin, *J. Am. Chem. Soc.* **2010**, *132*, 10986 – 10987; b) 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.
[7] See for example: a) S. Liu, B. L. Pentelute, S. B. H. Kent, *Angew. Chem.* **2012**, *124*, 1017 – 1023; *Angew. Chem. Int. Ed.* **2012**, *51*, 993 – 999; b) 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; c) J. A. Brailsford, S. J. Danishefsky, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 7196 – 7201.
[8] a) F. Albericio, *Curr. Opin. Chem. Biol.* **2004**, *8*, 211 – 221; b) M. Amblard, J. A. Fehrentz, J. Martinez, G. Subra, *Mol. Biotechnol.* **2006**, *33*, 239 – 254.
[9] The term "self-purification" has been used for another non-chromatographic purification concept: F. Mende, M. Beisswenger, O. Seitz, *J. Am. Chem. Soc.* **2010**, *132*, 11110 – 11118.
[10] a) S. Funakoshi, H. Fukuda, N. Fujii, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 6981 – 6985; b) G. del Solar, F. Albericio, R. Eritja, M. Espinosa, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 5178 – 5182; c) L. E. Canne, R. L. Winston, S. B. H. Kent, *Tetrahedron Lett.* **1997**, *38*, 3361 – 3364; d) M. Villain, J. Vizzavona, K. Rose, *Chem. Biol.* **2001**, *8*, 673 – 679; e) J. Vizzavona, M. Villain, K. Rose, *Tetrahedron Lett.* **2002**, *43*, 8693 – 8696; f) T. Hara, A. Tainosho, K. Nakamura, T. Sato, T. Kawakami, S. Aimoto, *J. Pept. Sci.* **2009**, *15*, 369 – 376.
[11] For selected examples, see a) D. E. Krieger, B. W. Erickson, R. B. Merrifield, *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 3160 – 3164; b) R. Merrifield, A. Bach, *J. Org. Chem.* **1978**, *43*, 4808 – 4816; c) A. Brown, S. Irving, R. Ramage, *Tetrahedron Lett.* **1993**, *34*, 7129 – 7132; d) S. Funakoshi, H. Fukuda, N. Fujii, *J. Chromatogr.* **1993**, *638*, 21 – 27; e) P. Mascagni, H. L. Ball, G. Bertolini, *Anal. Chim. Acta* **1997**, *352*, 375 – 385; f) P. C. de Visser, M. van Helden, D. V. Filippov, G. A. van der Marel, J. W. Drijfhout, J. H. van Boom, D. Noort, H. S. Overkleeft, *Tetrahedron Lett.* **2003**, *44*, 9013 – 9016.
[12] a) L. E. Canne, P. Botti, R. J. Simon, Y. J. Chen, E. A. Dennis, S. B. H. Kent, *J. Am. Chem. Soc.* **1999**, *121*, 8720 – 8727; b) P. Botti, H. Gaertner, S. Manganiello, M. Villain, U.S. Patent 0113563A1, **2005**.
[13] a) A. Brik, E. Keinan, P. E. Dawson, *J. Org. Chem.* **2000**, *65*, 3829 – 3835; b) G. J. Cotton, T. W. Muir, *Chem. Biol.* **2000**, *7*, 253 – 261; c) D. Bang, S. B. H. Kent, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5014 – 5019; d) E. C. B. Johnson, T. Durek, S. B. N. Kent, *Angew. Chem.* **2006**, *118*, 3361 – 3365; *Angew. Chem. Int. Ed.* **2006**, *45*, 3283 – 3287.
[14] V. Aucagne, A. F. Delmas, International patent WO2011/058188, **2011**.
[15] S. Gendler, J. Taylor-Papadimitriou, T. Duhig, J. Rothbard, J. Burchell, *J. Biol. Chem.* **1988**, *263*, 12820 – 12823.
[16] a) C. W. Tornøe, M. Meldal in *Peptides: The wave of the future* (Eds.: M. Lebl, R. A. Houghten), Kluwer Academic Publishers, Dordrecht **2001**, pp. 263 – 264; b) C. W. Tornøe, C. Christensen, M. Meldal, *J. Org. Chem.* **2002**, *67*, 3057 – 3064; c) V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem.* **2002**, *114*, 2708 – 2711; *Angew. Chem. Int. Ed.* **2002**, *41*, 2596 – 2599.
[17] N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J. Am. Chem. Soc.* **2004**, *126*, 15046 – 15047.
[18] E. Saxon, C. R. Bertozzi, *Science* **2000**, *287*, 2007 – 2010.
[19] G. Tesser, I. Balvertgeers, *Int. J. Pept. Protein Res.* **1975**, *7*, 295 – 305.
[20] N_3-Esoc-ONp **1** is stable for years at 4 °C under air atmosphere.
[21] F. García-Martín, M. Quintanar-Audelo, Y. García-Ramos, L. J. Cruz, C. Gravel, R. Furic, S. Côté, J. Tulla-Puche, F. Albericio, *J. Comb. Chem.* **2006**, *8*, 213 – 220.
[22] M. Meldal, *Tetrahedron Lett.* **1992**, *33*, 3077 – 3080.
[23] F. Albericio, R. Vanabel, G. Barany, *Int. J. Pept. Protein Res.* **1990**, *35*, 284 – 286.
[24] S. Mezzato, M. Schaffrath, C. Unverzagt, *Angew. Chem.* **2005**, *117*, 1677 – 1681; *Angew. Chem. Int. Ed.* **2005**, *44*, 1650 – 1654.
[25] For a review on the effect of the solid support on kinetics rates: D. Walsh, D. Q. Wu, Y. T. Chang, *Curr. Opin. Chem. Biol.* **2003**, *7*, 353 – 361.
[26] To evaluate the importance of the three-dimensional structure imposed by disulfide formation on biological activity.
[27] C. Derache, H. Meudal, V. Aucagne, K. J. Mark, M. Cadène, A. F. Delmas, A.-C. Lalmanach, C. Landon, *J. Biol. Chem.* **2012**, *287*, 7746 – 7755.
[28] a) G. A. Cremer, H. Tariq, A. F. Delmas, *J. Pept. Sci.* **2006**, *12*, 437 – 442; b) F. García-Martín, P. White, R. Steinauer, S. Côté, J. Tulla-Puche, F. Albericio, *Biopolymers* **2006**, *84*, 566 – 575.
[29] Based on the titration of the first and last Fmoc-group deprotection (UV, 301 nm $\varepsilon = 7800\ cm^{-1}M^{-1}$).
[30] C. Gaulier, A. Hospital, B. Légeret, A. F. Delmas, V. Aucagne, F. Cisnetti, A. Gautier, *Chem. Commun.* **2012**, *48*, 4005 – 4007, and references therein.

[31] This capture/release procedure was applied to two other targets, the [31–53] fragment of human mitogaligin and the leader peptide sequence of *E. coli*'s McJa (36 amino acids), with excellent results in terms of recovery and enrichment in the target peptides (see the Supporting Information, Figures S4–S11).

[32] a) V. Aucagne, D. A. Leigh, *Org. Lett.* **2006**, *8*, 4505 – 4507; b) I. E. Valverde, A. F. Delmas, V. Aucagne, *Tetrahedron* **2009**, *65*, 7597 – 7602.

