Article

Chemical protein synthesis using a second generation N-acylurea linker for the preparation of peptide-thioester precursors.

Juan B. Blanco-Canosa, Brunello Nardone, Fernando Albericio, and Philip E. Dawson

J. Am. Chem. Soc., **Just Accepted Manuscript** • DOI: 10.1021/jacs.5b03504 • Publication Date (Web): 15 May 2015

Chemical protein synthesis using a second generation *N*-acylurea linker for the preparation of peptide-thioester precursors.

Juan B. Blanco-Canosa,[a] Brunello Nardone,[b] Fernando Albericio,[a,c] Philip E. Dawson[d]

[a]Institute for Research in Biomedicine (IRB Barcelona), Chemistry and Pharmacology programme, Baldiri Reixac 10, Barcelona 08028 (Spain)

[b]University of Salerno, Departments of Chemistry and Biology, Via Giovanni Paolo II 132, Fisciano 84084 (Italy)

[c]University of Barcelona, Department of Organic Chemistry, Martí i Franqués 1-11, Barcelona 08028 (Spain)

[d]The Scripps Research Institute (TSRI), Department of Chemistry 10550 N. Torrey Pines Road, La Jolla 92037 (USA)

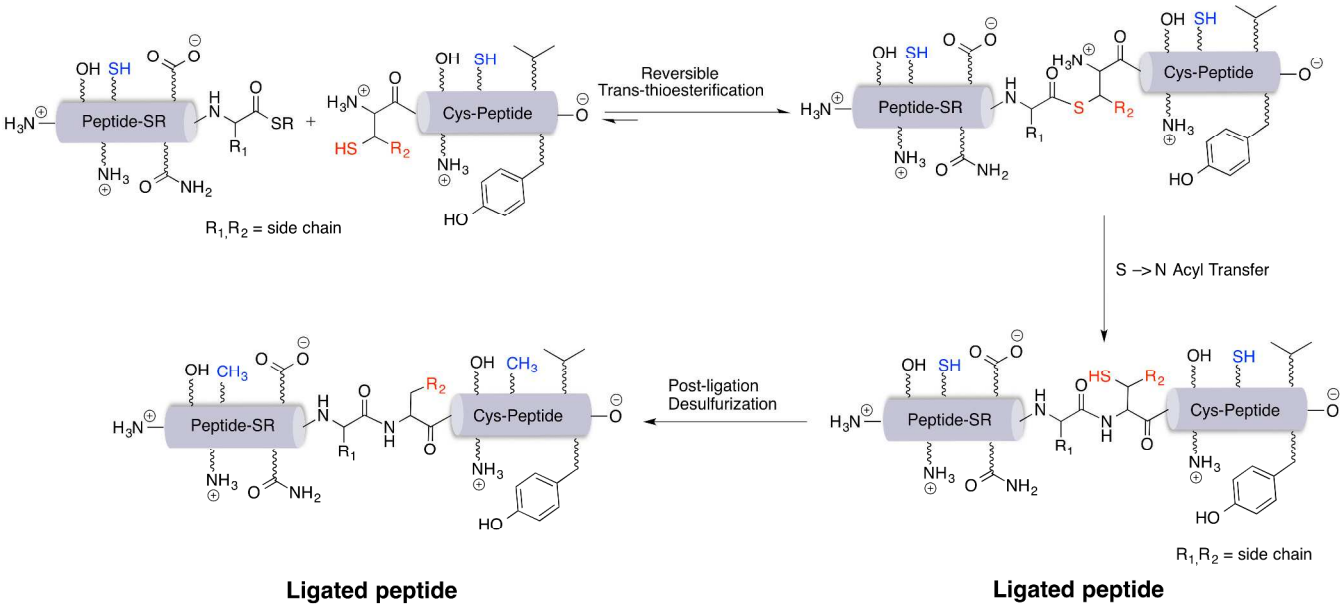
KEYWORDS. Native Chemical Ligation – thioesters – acylurea – microwave peptide synthesis.

ABSTRACT: The broad utility of Native Chemical Ligation (NCL) in protein synthesis has fostered a search for methods that enable the efficient synthesis of C-terminal peptide-thioesters, key intermediates in NCL. We have developed an *N*-acylurea (Nbz) approach for the synthesis of thioester peptide precursors that efficiently undergo thiol exchange yielding thioester peptides, and subsequently NCL reaction. However, the synthesis of some glycine rich sequences revealed limitations, such as diacylated products that can not be converted into *N*-acylurea peptides. Here, we introduce a new *N*-acylurea linker bearing an *o*-amino(methyl)aniline (MeDbz) moiety that enables in a more robust peptide chain assembly. The generality of the approach is illustrated by the synthesis of a pentaglycine sequence under different coupling conditions including microwave heating at coupling temperatures up to 90 C, affording the unique and desired *N*-acyl-*N'*-methylacylurea (MeNbz) product. Further extension of the method allowed the synthesis of all 20 natural amino acids and their NCL reactions. The kinetic analysis of the ligations using model peptides shows the MeNbz peptide rapidly converts to arylthioesters that are efficient at NCL. Finally, we show that the new MeDbz linker can be applied to the synthesis of cysteine rich proteins such the cyclotides Kalata B1 and MCoTI-II through a one cyclization/folding step in the ligation/folding buffer.

1. Introduction.

Chemical Protein Synthesis (CPS) enables the synthesis of proteins with precise control over the final synthetic product, allowing the access to a wide diversity of post-translational modifications (PTMs) including glyco-, phospho-, lipo- and cyclic proteins.[1] Native Chemical Ligation (NCL) has been shown to be the most generally applied approach for the chemoselective assembly of unprotected peptides in aqueous solution.[2] NCL relies on the selective reaction of a C-terminal peptide thioester and a N-terminal cysteine fragment. In the absence of a cysteine, the N-terminal residue can be replaced by a cysteine-surrogate with a thiol group in ß or γ-position, which can be further desulfurized. This later approach was first demonstrated for Ala and Abu ligation sites (ß and γ- position, respectively) and the generality of the strategy to other amino acids was noted,[3] suggesting a general approach for NCL at residues other than cysteine. Thus, NCL has been applied to alanine,[3] phenylalanine,[4] leucine,[5] glutamic acid,[6] aspartic acid,[7] glutamine,[8] valine,[9] proline,[10] lysine,[11] arginine,[12] threonine,[13] and likely could be extended to the remaining amino acids using a similar logic (**Scheme 1**). Furthermore, the use of N^{α}-auxiliaries enabled the extension of NCL to glycine,[14] and in principle these N^{α}-auxiliaries could be utilized for ligations with other residues. Although this ligation is very attractive, it is sterically very demanding and the utilization of N^{α}-auxiliaries is still principally constrained to *Xaa*-Gly junctions (*Xaa* = any amino acid).

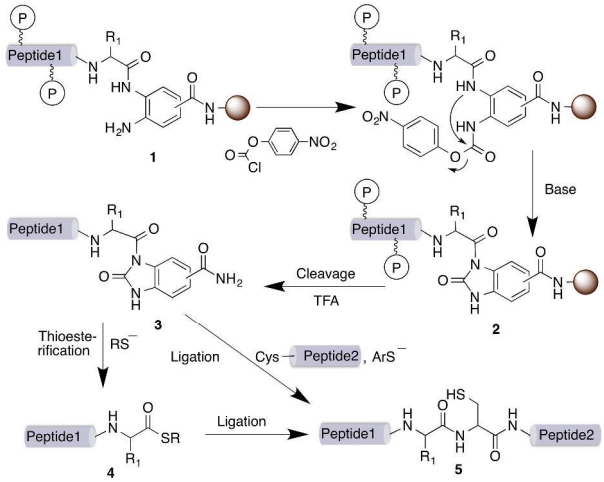
Common to all NCL family reactions is the synthesis of C-terminal peptide thioesters which can be obtained in Boc-Solid Phase Peptide Synthesis (Boc-SPPS) in a straightforward manner using mercaptoalkyl acid linkers such as mercaptopropionic or mercapto acetic acids.[15] However, the lability of the thioester bond under the basic/nucleophilic conditions used in Fmoc chemistry (20% Piperidine/DMF) fostered the development of alternative strategies. For example, activation followed by thioesterification of the C-terminal amino acid on fully



Scheme 1. Native Chemical Ligation and extended NCL using β-mercapto amino acid derivatives.

protected peptides,[16] backbone amide linker (BAL),[17] metal-assisted thiolysis of PAM/Wang resins,[18] the Kenner's sulfonamide safety-catch linker[19] and the aryl hydrazine handle.[20] More recently, a variety of different methods designed to undergo *N* to *S*[21] or *O* to *S*[22] acyl shift under acidic or during NCL conditions, or the utilization of the transfer active ester approach using terminal hydrazide peptide precursors.[23]

Previously, we disclosed a new approach based on *N*-acyl-benzimidazolinones or *N*-acylureas[24] (**Nbz**, **Scheme 2**). Starting from 3,4-diaminobenzoic acid, selective acylation of one of the amines yields an *o*-aminoanilide peptide **1**. Following chain assembling, acylation with *p*-nitrophenyl chloroformate and cyclization under basic conditions affords the *N*-acylurea protected peptide, **2**. Acidolytic cleavage of the resin and concomitant amino acid side chain deprotection yields the *N*-acylurea peptide **3**, which can be further trans-thioesterified (**4**) or utilized directly in NCL in presence of a thiol catalyst (**5**). We and others have used the *N*-acylurea method for the synthesis of proteins or protein domains. For example, we have prepared the second type 1 repeat of Thrombospondin-1 (TSR2),[25] a protein that has shown a potent anti angiogenic activity. Elsewhere, the Nbz strategy has been applied to the synthesis of ubiquitinated peptide chains and C-terminal ubiquitinated hydrazide peptides,[26] salicylaldehyde ester precursors for serine and threonine ligation,[27] cyclotides,[28] human α-synuclein,[29] combinatorial assembly of affibody molecules,[30] a phosphorylated fragment of the MDM2 E3 ubiquitin ligase,[31] the collagenous domain of the Adiponectin hormone,[32] a 44-mer histone-derived fragment,[33] the HIV-1 Rev protein,[34] the N-terminal domain of the Anthrax Lethal Factor,[35] glycoproteins,[36] and phosphorylated and acetylated fragments of the histone H2B.[37]



Scheme 2. Key intermediates in the synthesis of C-terminal peptide *N*-acylureas. P = protecting group.

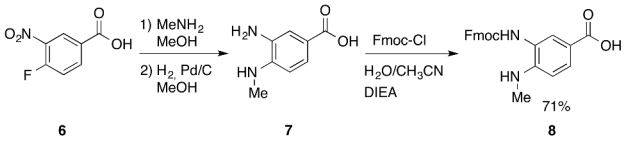
Nevertheless, it has been reported that in glycine rich sequences acylation of the unprotected *o*-aminoanilide can occur, especially in the presence of excess base.[33,38] Although reversible protection of the *o*-aminoanilide peptide with alloc group can address the problem,[33] an alternative, more powerful and elegant approach would be to furnish each amine of the 3,4-diaminoaryl linker with different reactivity. This approach would avoid extra protection and deprotection steps and also produce a single isomer of the N-acylurea product. Here we report the application to peptide synthesis, peptide ligation and cyclic protein synthesis of a second generation of *o*-aminoaniline linker, the *o*-amino(methyl)aniline (**Scheme 3**, **MeDbz**) in which a methyl group is introduced onto the *para*-amino group. Following chain elongation the linker can be transformed into a mildly activated *N*-acylurea, (*N*-acyl-*N'*-methyl-benzimidazolinone or *N*-acyl-*N'*-methylurea) (**MeNbz**) that shows a higher resistance to

acylation during SPPS compared to the first generation of the Nbz linker.

2. Results.

2.1 Synthesis of Fmoc-MeDbz (8).

Synthesis of **8**, was achieved following reported procedures[39] in 3 steps in a gram scale. Starting from 3-fluoro-4-nitrobenzoic acid, **6**, nucleophilic addition of methylamine yields the 3-nitro-4-(methylamino)benzoic acid, **7**. After reduction of the nitro group through catalytic hydrogenation over Pd/C, chemoselective protection of the primary aryl amine with Fmoc-Cl yields the expected Fmoc-3-amino-4-(methylamino)benzoic acid (**Fmoc-MeDbz**) (**8**, 71% overall yield, **Scheme 3**). Alternatively, the reduction of the nitro aromatic group can also be carried out on solid phase using $SnCl_2$.



Scheme 3. Synthesis of **Fmoc-MeDbz (8)**.

2.2 Synthesis and evaluation of the MeDbz chemoselectivity in a polyglycine peptide: Gly_5 sortase tag.

Gly_5 is a 5-mer peptide employed in the sortase labeling technology[40] that recently has found application in the sortase-assisted synthesis of peptide/protein thioesters.[41] Owing to its exclusive glycine sequence, it renders a challenging model peptide for testing the robustness of the MeDbz linker. In order to confirm previous results,[33,38] we decided to investigate the synthesis of Fmoc-Gly_5-Nbz-R (**9**) on the original Dbz linker under strong acylating conditions: 10-fold excess each of Fmoc-Gly-OH and HBTU and 20 eq of base. (Note: the originally described acylation conditions for Dbz utilized HOBt to buffer excess base). Under these conditions, significant amounts of branched peptides were observed (54%, **method A** section **4.2**, and **Figures 1a** and **S1**).

This pentaGly peptide proved to be generally challenging on the Dbz linker. For example, at 4 eq Fmoc-Gly-OH/HBTU/DIEA **Figure S2**) or under mildly acidic conditions (4 eq each Fmoc-Gly, HBTU, HOBt, and 6 eq DIEA **Figure 1b**, **method B** in **section 4.2**) some diacylated products were detected (16 and 12% diacylated products, respectively) representing 2-4% acylation per Gly residue coupled. Nevertheless, this low side-acylation percentage has not resulted in any impairment in the synthesis of Nbz peptides possessing Gly residues at the C-terminal position, such as in the case of the C-terminal GG sequence of ubiquitin-derived Nbz peptides.[26]

To control for the possibility that the over-acylation could arise from the coupling of the first glycine, the free amine was protected with alloc after introduction of the Fmoc-Dbz moiety on the resin, and removed following the pentaglycine assembling.[33] As expected, this route yielded one major compound corresponding to the desired monoacylated Fmoc-Gly_5-Nbz-Arg peptide (**Figure S3**). Conversely, the synthesis of the pentaGly following alloc removal after incorporation of the C-terminal Gly still shows some overacylated products (**Figure S4**). These results are consistent with a low level of acylation during the chain elongation steps.

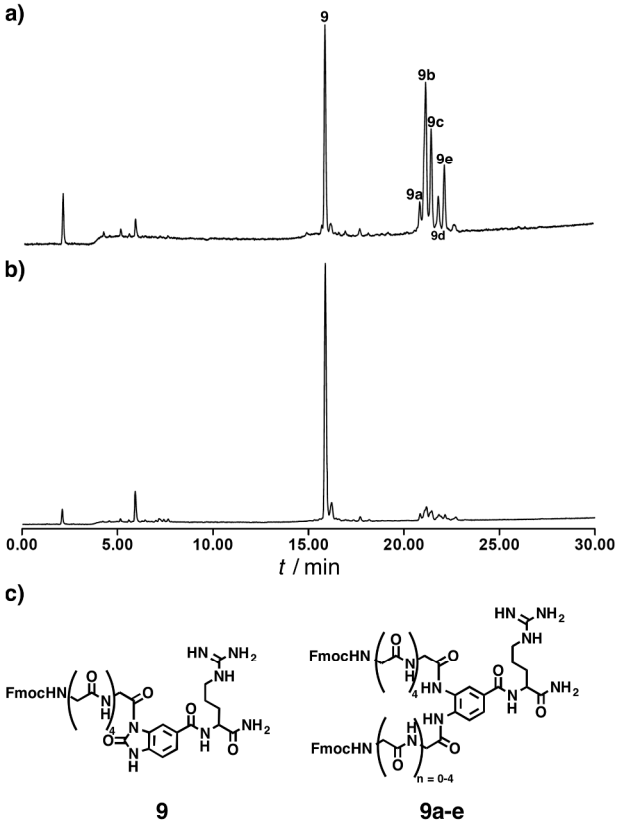


Figure 1. HPLC traces of synthetic **9**: **9a**-**9e** = diacylated products. Coupling conditions: 1st Fmoc-Gly 4 eq, HBTU 4 eq, DIEA 4 eq, 40 min; then, **a)** Fmoc-Gly 10 eq, HBTU 10 eq (0.5 M in DMF), DIEA 20 eq, 1h at rt; **b**) Fmoc-Gly 4 eq, HBTU (4 eq, 0.2 M)/HOBt (4 eq, 0.2 M), DIEA 6 eq, 1h at rt; c) structures of the peptides **9** and **9a-e** (see **Figure S1** for detailed structures).

The introduction of the *N*-Me group on the *o*-amino group of the linker should significantly decrease the sensitivity of the linker towards undesired acylation. Gratifyingly, the utilization of the MeDbz moiety and 5-fold excess of Fmoc-Gly and HBTU activation afforded the desired peptide Fmoc-Gly_5-MeNbz-R (**10**) with no traces of side products (**Figure 2a**, **method C** in **section 4.3**). Assaying harsher conditions such as a 10-fold excess of amino acid and basic conditions, i.e. 20-fold excess of DIEA (**Figure S5**), and the utilization of microwave irradiation at either 75 C or 90 C (**Figure S6** and **Figure 2b**, respectively) did not generate significant (< 1%) diacylated peptides. Together, these results show that it is very challenging to couple even an unhindered amino acid such as Fmoc-Gly due to the steric and electron withdrawing properties of the *o*-amino(methyl)anilide, which becomes deactivated by the anilide and the amide groups in *o*- and *p*-positions, respectively.[42] Critically, the *N*-methylaniline moiety is still reactive enough towards acylation with chloroformate derivatives (5 eq in CH_2Cl_2, 1h), and is able to

undergo cyclization (0.5 M in DMF, 30 min) to yield *N*-acyl-*N'*-methylurea peptides.

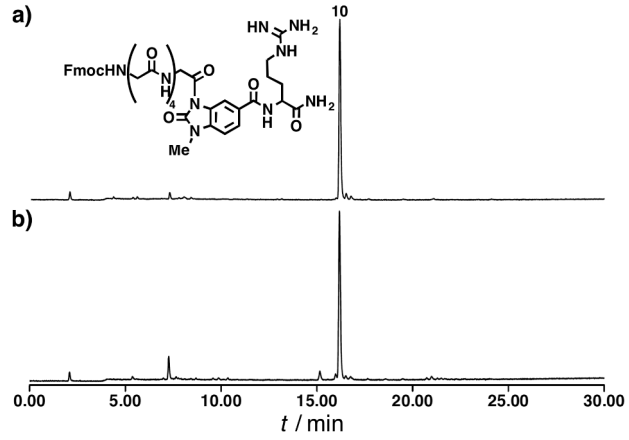
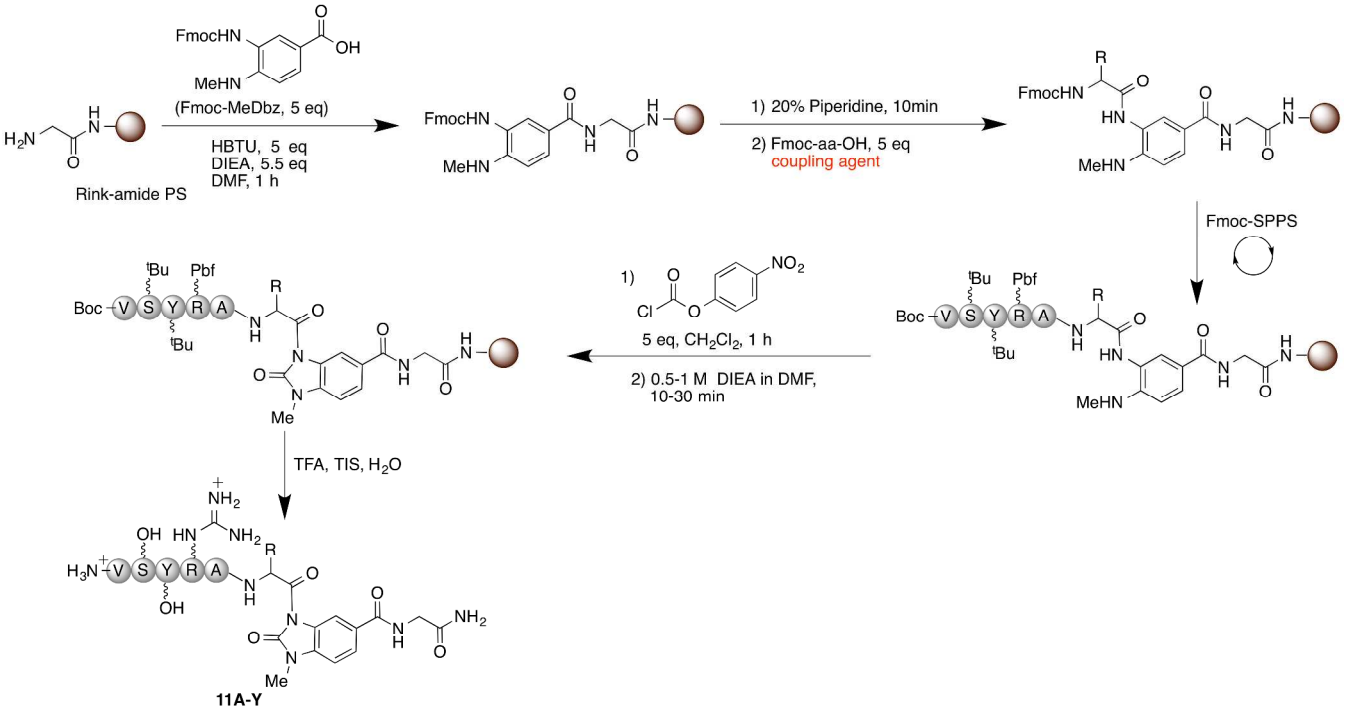


Figure 2. HPLC traces of synthetic **10**. Coupling conditions: **a**) 5 eq Fmoc-Gly, 5 eq HBTU (0.5 M in DMF), 5.5 eq DIEA, 1h at rt. **b**) 1st Fmoc-Gly as in **a**), next 10 eq Fmoc-Gly, 10 eq DIC, 10 eq Oxyma Pure, 3 min at 90 ºC following **MW1** setup (see experimental **section 4.3** for microwave irradiation characteristics).

2.3 Synthesis of VSYRA*Xaa*-MeNbz-G (11A-Y) model peptides with different C-terminal amino acids attached to the MeDbz linker.

Having confirmed the superior characteristics of the MeDbz linker, its general applicability in peptide synthesis and NCL was demonstrated through preparing a group of 20 model peptides VSYRA*Xaa*-MeNbz-G (**11A-Y**, the letter depicts the amino acid *Xaa* at the C-terminus) including all twenty C-terminal amino acids attached to the MeDbz moiety, on a polysterene resin functionalized with a Rink linker that is cleavable with TFA. In order to facilitate the introduction of the Fmoc-MeDbz-OH, a glycine spacer was coupled to the resin in the first place. Coupling of the Fmoc-MeDbz-OH moiety was accomplished using a 5-fold excess with regard to the number of mmol of resin, HBTU as activating agent (5-fold excess) and DIEA as base (5.5 fold-excess) in DMF at a final concentration around 0.33 M. Following introduction of Fmoc-MeDbz, the C-terminal residue was introduced either with HBTU or HATU. Nevertheless, the utilization of HATU, or a similar strong activating agent such as PyAOP or COMU, led to complete acylation of the amino acid with no need for double coupling (**Scheme 4**). Although not strictly necessary for all amino acids, the utilization of 5-fold eq of amino acid, HATU and basic conditions (10 eq of DIEA), at a concentration around 0.4-0.5 M afforded reliable loadings >99%. However, in order to avoid the epimerization of some racemization prone residues, such as His[43] and Cys[44] (check **section 2.4**), a neutral protocol based on DIC/HOOBt is epimerization free. A significant advantage of MeDbz linker over the Alloc-Dbz strategy is the complete acylation of Val and Ile, whereas in the alloc-Dbz approach only loadings of 47 and 33% were achieved, respectively.[33] The next amino acids were coupled under standard conditions: 5 eq of *Xaa*, 5 eq of HBTU (0.5 M in DMF) and 5.5 eq of DIEA. The last residue was introduced with N-terminal Boc protection. Following acylation with *p*-nitrophenyl chloroformate in dichloromethane, cyclization was carried out in a solution of 0.5 M or 1 M DIEA (in DMF) for 10-30 min. After cleavage, crude peptides could be recovered in yields and purities typically higher than 90% (**Table S1** and **Figure S7**).



Scheme 4. Synthetic route to peptides **11A-Y**. PS = polysterene.

The carbamylation reaction could also be carried out in other solvents including 1,2-dichloroethane, THF, α,α,α-trifluorotoluene, 2,2,2-trifluoroethanol, hexafluoroisopropanol or in DMF which is known as a better swelling solvent for peptide resins.[34] Alternatively, intramolecular acylurea formation was accomplished in a second step in solvents such as DCM or DMSO, or using bases such as *N*-Methylmorpholine, although in these cases the cyclization has a slower kinetics.[45]

2.4 Assessing the epimerization on Cys and His.

One of the major drawbacks of the Fmoc approach for the solid phase synthesis of peptide thioesters is the racemization of the C-terminal amino acid during the coupling or activation-thioesterification steps.[47] In particular, special care must be taken on His[43] and Cys.[44] Model studies on both amino acids on the MeDbz linker showed that epimerization levels were below 0.1% when the linker was loaded using carbodiimide activation in presence of additives such as HOOBt: 5 eq of Fmoc-aa-OH, 5 eq of DIC and 5 eq of HOOBt with preactivation time of 3 min. More specifically, Fmoc-Cys(Trt)-OH could be coupled under several different conditions with no detectable epimerization (**Figure S8)**: HATU/DIEA (1:1.1), DIC/Oxyma (1:1), DIC/HOOBt (1:1). However, the acylation of Fmoc-His(Trt)-OH turned out to be more troublesome, and from the coupling mixtures tested only DIC/HOOBt gave no detectable levels of racemization (**Figure S9**).[46]

2.5 Optimization of the Native Chemical Ligation conditions.

With the availability of the 20 model peptides we undertook a systematic study using the N-terminal cysteine peptide **CRAFS**. First, ligations were performed varying pH and concentration of thiol catalyst, choosing Leu (**11L**) as a case study, a residue that using previously studied mercaptoproprionyl thioesters gives a moderately fast reaction (completion in 2 h) in presence of high concentrations (200 mM) of 4-Mercaptophenylacetic acid (MPAA).[48] This allowed straightforward monitoring of the ligation kinetics. A short screening of different thiols (considering both the kinetics and water solubility) led to the choice of MPAA and (4-Hydroxy)thiophenol (MPOH) as the optimal thiol catalysts for NCL with acylurea peptides, both displaying similar kinetics. However, we found that MPOH is less expensive than MPAA, is less acidic, shares its lack of odor and is soluble up to at least 200 mM at pH 7.0. So, for the model ligation reactions we chose MPOH as the thiol catalyst.[48]

The ligation reaction is highly dependent on pH, mediating the desired conversion of the acylurea to a thioester, transthioesterication (typically rate determining for NCL), while minimizing the potential side products of epimerization and hydrolysis of the thioester. Ligations were carried out in the presence of guanidine hydrochloride as denaturing agent and TCEP to maintain a reducing environment and concentration of the MPOH additive. We undertook a study of the rate of the ligation in 3 different pH buffers: 5.8, 7.0 and 8.2 (**Figure 3a**). The results show that the optimal pH for the ligation is around 7.0. At pH 8.2, the percentage of hydrolysis increases and the ligation rate is not significantly improved. At lower pH, the rate of the ligation is significantly slower. Interestingly, the relative rate of ligation is reduced to a larger extent than that of hydrolysis resulting in an increase in the hydrolysis observed in the final product. As with the Dbz linker, if desired, the N-acylurea peptide could be fully exchanged with a soluble alkyl thiol such as MESNA for long-term storage.

Figure 3b shows the reaction progress of VSYRAL-MeNbz-G and a slight excess of CRAFS (1.5-2-fold excess) in presence of increasing concentrations of MPOH (0-200 mM). Several conclusions can be drawn from these reactions: *i*) as expected, the rate of the ligation increases with the concentration of MPOH, but *ii*) concentrations above 100 mM do not significantly speed up

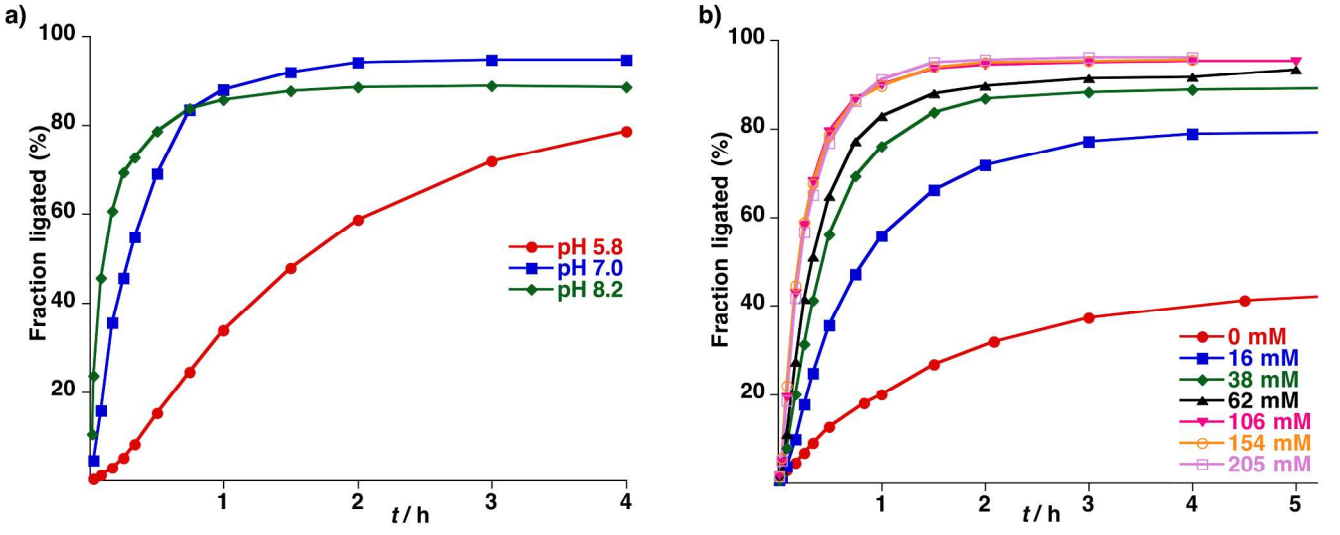


Figure 3. Kinetics of the formation of **12L**: **a**) influence of the pH, **b**) influence of the concentration of 4-Mercaptophenol.

the rate of the ligation. This suggests an optimal MPOH concentration around 100 mM, *iii*) hydrolysis is a competitive process, especially when the reaction is carried out with no catalyst or at MPOH concentrations below 100 mM. For example, in the absence of MPOH, only a ~50% of the fraction ligated could be achieved after 10 h. These observations are consistent with hydrolytic sensitivity of acylurea peptides compared to thioester peptides.[24] Therefore, attaining a rapid intermolecular MeNbz –> S transfer is important to generate the hydrolytically stable thioester peptide,[49] which is hydrolytically stable at pH 7.0 yet highly reactive towards N-terminal cysteine peptides. From these model ligation experiments, it is observed that a critical step in the ligation of C-terminal acylurea and N-terminal cysteine peptides is the intermolecular *N* to *S* acyl transfer to the thiol additive. In order to achieve a high exchange rate, thiols with a pKa around 7 (principally aryl thiols) with good water solubility (~ 100 mM) are the optimal catalyst.

2.6 Assaying the Native Chemical Ligation with VSYRA*Xaa*-MeNbz model peptides.

The ligation buffer conditions of 0.2 M sodium phosphate, 0.1 M MPOH, 6 M guanidine hydrochloride, 20 mM TCEP, pH 7.0, were selected for subsequent studies. To evaluate the general utility of the linker, an exhaustive study of the ligation between peptides **11** (**A** to **Y**) and CRAFS was performed to afford the ligated peptides **12** (**A** to **Y**). These peptides included the fastest C-terminus, glycine and the slowest C-terminus, proline as observed previously with alkyl-thioester peptides using thiophenol/benzyl mercaptan as catalysts (3:1).[15a]

The ligation rates fell into 3 different groups: *i*) Pro, which gives the slowest rate (88% conversion in 48 h, **Figure 4b**), *ii*) ß-branched amino acids such as Val, Ile that achieve near a > 95% conversion within 20 h (**Figure 4b**), and *iii*) amino acids where the ligation is complete in ≤ 5 h (**Figure 4a**). When compared to the reported kinetics of mercaptopropionyl thioesters[15a] some divergences are clearly seen, which can be explained based on the lower solubility of thiophenol (the saturating concentration in the guanidine buffer is around 30 mM) that results in a slower transthioesterification, in addition to the lower reactivity and competitive transthioesterification of the alkyl mercaptoproprionyl leaving group. This explanation is also consistent with the increased ligation displayed by preformed aryl thioesters.[48,50]

Intramolecular reactions involving the side chain of the C-terminal amino acid can affect the utility of NCL. Consistent with the observations of Dang et al,[51] Asn, Gln were ligated efficiently, with no interfering reactions of the side chains resulting in unreactive or isomeric products. The ligation with C-terminal Glu peptides, yielded only traces, (~ 2%) of the undesired γ-linked peptide at pH 7.0 (**Figure 5a**).[51] However, ligations with Asp result in rapid ligation but with significant quantities of the β-linked peptide in a ratio 1.8:1 (Asp:^{iso}Asp, **Figure 5b**) presumably as a result of the anhydride formation with the Asp side chain and subsequent thiol attack to either carbonyl group.[52,53]

Thus, the repertoire of amino acids that are viable for NCL can be extended to any given residue in the thioester fragment, with the only exception of C-terminal Asp which would require orthogonal protecting group and subsequent removal after NCL.[52]

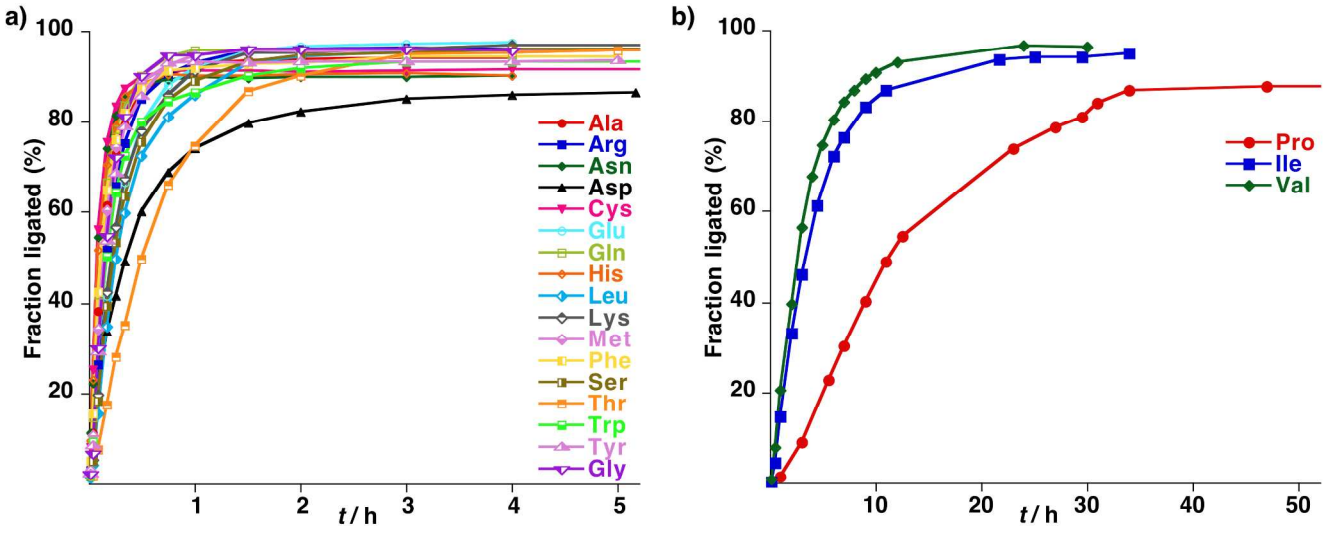


Figure 4. Kinetics of the formation **12A-Y** peptides using MPOH (100 mM) as catalyst at pH 7.0, room temperature. Note that the fraction ligated for Asp includes **12D** and **12^{iso}D**.

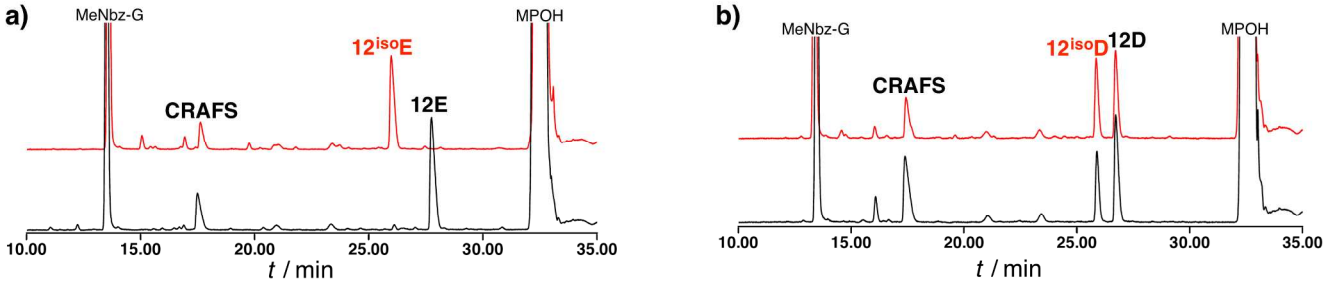


Figure 5. NCL traces of: **a) 11E** (black) and **11^{iso}E** (red), **b) 11D** (black) and **11^{iso}D** (red) peptides. Reactions were carried out at rt and 2 mM concentration of MeNbz peptides, 2.4-2.7 mM of CRAFS in 6M guanidine buffer (pH 7.0) containing 20 mM of TCEP and 100 mM MPOH.

2.7 Synthesis of cyclotides Kalata B1 and MCoTI-II.

Cyclotides are naturally occurring ribosomal proteins isolated from plants with a head-to-tail cyclic structure and an intramolecular cystine knot arrangement, also known as cyclic cystine knot (CCK).[54] Cyclotides can be classified in two major groups: Möbius if they have a *cis*-Pro configuration in loop 5, and Bracelet.[55] Another important class of cyclic proteins bearing a cystine knot motif comprises the family of squash and *Momordica cochinchinensis trypsin inhibitors* MCoTI-I/II.[56] Cyclotides have a strong resistance to proteolytic and chemical degradation and a broad spectrum of biological activities. For example, Kalata B1 is thought to have uterotonic properties and has been shown to display anti-HIV, anti-microbial and insecticidal activity, whereas MCoTI-II is one of the most potent trypsin inhibitor known. Due to the variability in the loop size and sequence, it is possible to replace some of the loops by other non native sequences, without affecting the final folding, and thus enabling the utilization of these cyclotides as scaffolds for the grafting of bioactive peptide drugs.[57]

SPPS of cyclotides represents a synthetic challenge due to the presence of 6 cysteine residues that, apart from their higher tendency to racemize during the coupling,[44] require careful conditions for the final cleavage in order to avoid undesired side-reactions such as alkylation with ^{t}Bu groups which was observed as a significant adduct in a previous synthesis of this protein using the Dbz approach[28]. The complexity of this protein made it a rigorous test of synthesis, activation and cleavage conditions for the MeDbz linker on a variety of commonly used solid supports.

2.7.1 Kalata B1 (KB1).

The linear synthesis of cyclotides can be started at any given amino acid after a cysteine residue. Although there is a Gly-Cys junction, we decide to cyclize at a ^{4}Val-^{5}Cys point to show that hindered and slow reacting amino acids are also effective cyclization points for the synthesis of cyclotides (**Figure 6a**). The reported synthesis of Kalata B1 using the Nbz linker carried out on a PEG-PS resin Tentagel showed, apart from the expected peptide, a large percentage of an adduct with an extra mass of 56 units corresponding to the *tert*-Butyl group, likely due to the alkylation of the Trp or some Cys residue.[28a] When we synthesized linear KB1 on a Tentagel resin we observe, in addition to the expected linear KB1, a product with the same extra 56 units of mass (**Figure 6b**). Changing the composition of the cleavage cocktail (TFA, PhOH, TIS, EDT, H_2O) and increasing the concentration of EDT or adding some specific ^{t}Bu scavengers such as 1,2-Dimethylindole did not decrease significantly the percentage of this by-product. The utilization of the PEG-PS resin together with pseudoproline dipeptides only led to a slight improvement of the quality of the crude peptide (**Figure 6c**).

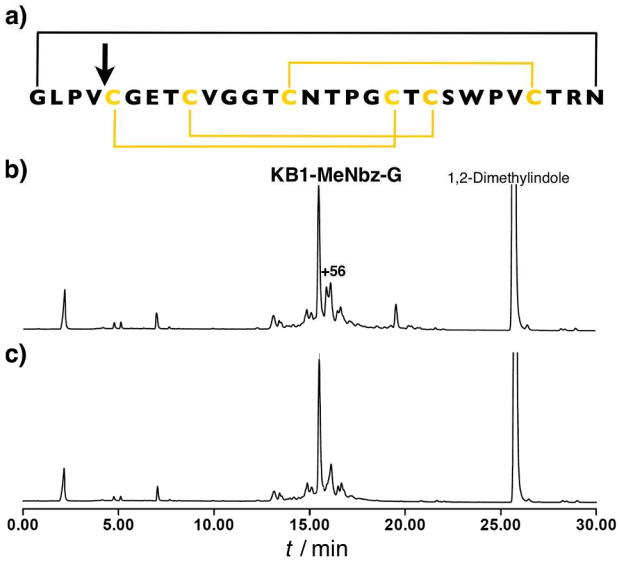


Figure 6. **a)** Amino acid sequence and disulfide knot arrangement (yellow lines) of KB1; the black arrow denotes the C-terminal residue of the linear synthesis; **b)** HPLC trace of the synthesis of linear **KB1-MeNbz-G** on Tentagel resin; **c)** HPLC trace of the synthesis of linear **KB1-MeNbz-G** on Tentagel resin using the pseudoproline dipeptide **GT**.

Another type of solid support is the fully pegylated ChemMatrix resin, which displays good swelling properties in a wide range of solvents including DMF, CH_2Cl_2 and H_2O. Synthesis on this resin in a 0.1 mmol scale using an excess of amino acid of 10-fold to ensure complete acylation led to a good crude purity (65%, **Figure 7a**, **method D** in **section 4.5**), and recovery yield (73%), but still contained significant *tert*-butylated side-products. Similar results were accomplished under microwave conditions using a coupling temperature of 75 ºC (**Figure S11**) and even when

resynthesizing a circular permutation of the sequence to facilitate a ^{8}Thr-^{9}Cys ligation site (**Figure S12**).

In order to characterize the alkylated residue (likely a Cys or Trp), we introduced a different combination of protecting groups: in the first case, the trityl for ^{20}Thr and ^{22}Ser in loops 4 and 5 that are next to ^{23}Trp (CTCSW, **Figure S13**), and in the second place the TFA stable phenylacetamidomethyl (Phacm) for the protection of the all Cys residues except the N-terminal.

Only in this later approach, which allowed the linear partially protected Kalata B1 in good analytical quality (75%) and recovery yield (77%), showed minimum amounts of *tert*-butylated peptides (**Figure 7b**), thus indicating that some of the Cys residues were acting as an internal scavenger and capturing a ^{t}Bu group.[58]

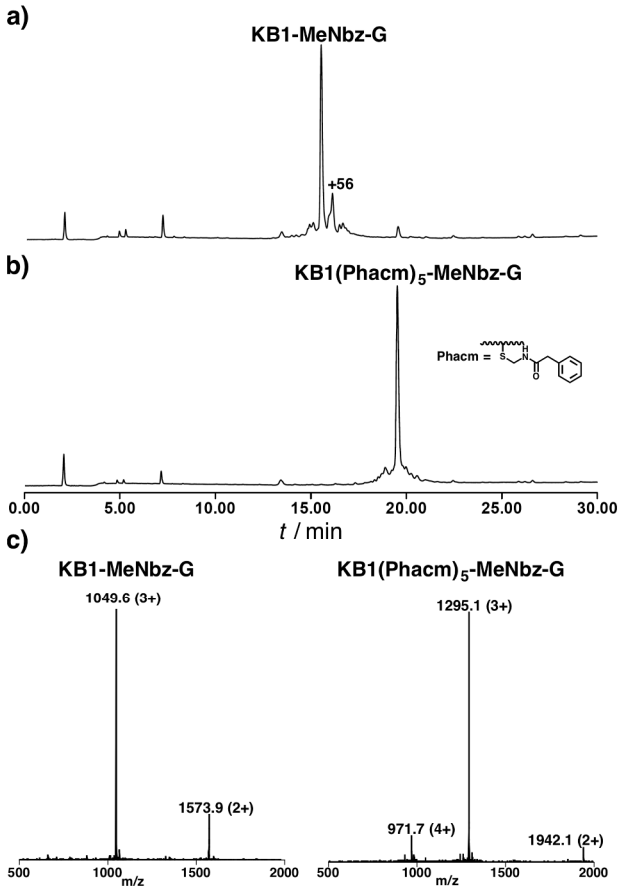


Figure 7. HPLC trace of the synthesis of linear **KB1-MeNbz-G** (**a**) and **KB1(Phacm)$_5$-MeNbz-G** (**b**) synthesized **both** on ChemMatrix resin; c) ESI-MS spectra of the main peak of the **KB1-MeNbz-G** synthesis (left), and **KB1-(Phacm)$_5$-MeNbz-G** crude (right). (Note that the linear synthesis started at Val in position 4).

Following the linear synthesis, the crude peptide was cyclized using the above conditions optimized for NCL (**Figure 8a**), and then folded in a $(NH_4)_2CO_3$:^{i}PrOH (1:1) buffer at pH 8.5 in presence of the GSH:GSSG (2 mM:0.4 mM) redox system.[59] As reported for linear cyclotides synthesized using either Boc or Fmoc chemistry,[59b,60] the cyclization and folding could be carried out in one single step in the folding buffer at pH = 7.3 - 7.5 affording the folded Kalata B1 in recovered yield of ~20% (**Figure 8b**). It is interesting to note that the cyclic folded peptide displays a more hydrophobic character than the cyclic peptide, due to that the hydrophobic residues are exposed to the solvent.[59]

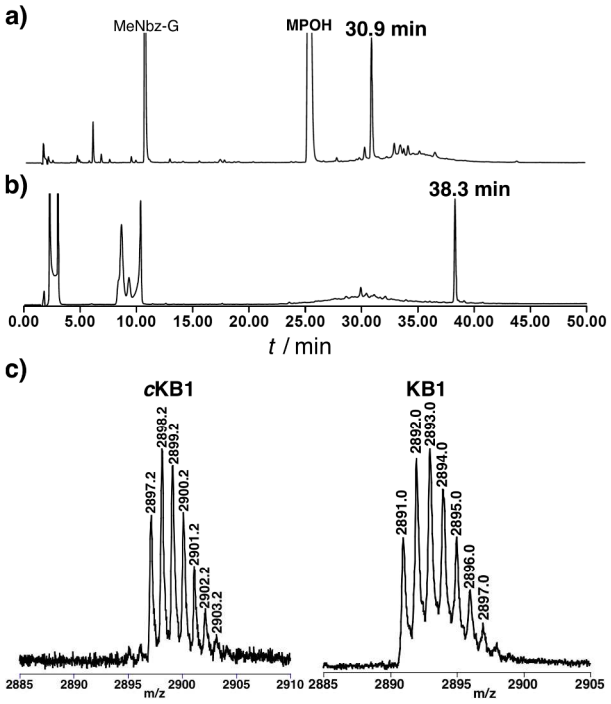


Figure 8. **a)** Cyclization of **KB1-MeNbz-G** crude synthesis; **b)** one-pot cyclization-folding of **KB1-MeNbz-G** crude synthesis; **c)** MALDI-TOF spectra of **cyclic KB1** (***c*KB1)** (left) and folded **KB1** (right).

2.7.2 MCoTI-II.

The linear synthesis of MCoTI-II was performed following the same strategy of KB1, chosen the cyclization point at the ^{28}Tyr residue (**Figure 9a**). In order to avoid aspartimide formation in the Asp-Gly sequence during the Fmoc removal, the Fmoc-(Hmb)Gly-OH amino acid was used for the temporal protection of the amide bond with the TFA labile Hmb (2-Hydroxy-4-methoxybenzyl) protecting group. The linear **MCoTI-II-MeNbz-G** peptide was recovered in an 81% yield and an analytical crude purity around 70% (**Figure 9b**). Following the same, single cyclization-folding approach for KB1, the linear precursor was correctly folded affording the MCoTI-II cyclotide in good overall yields (22%, **Figure 10a**). Notably, the folded MCoTI-II is more hydrophilic than the cyclic peptide (**Figure 10b**), indicating that in this case the hydrophilic residues have a more contact surface with the solvent.

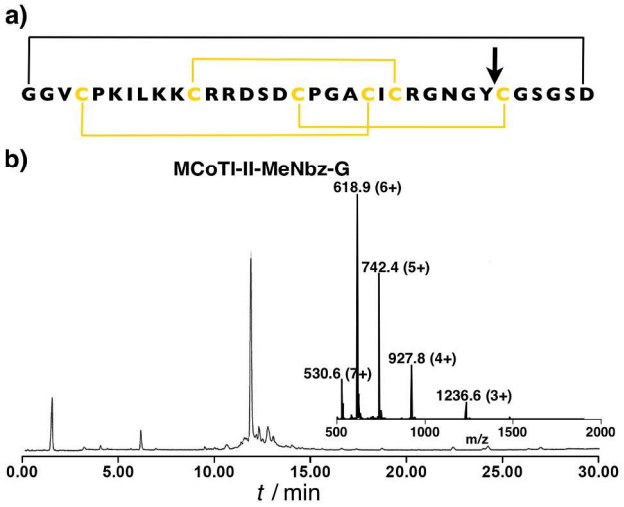


Figure 9. **a)** amino acid sequence and disulfide knot arrangement (yellow lines) of **MCoTI-II** (the black arrow denotes the C-terminal residue of the linear synthesis); **b)** HPLC trace of the synthesis of linear **MCoTI-II-MeNbz-G** on ChemMatrix resin. Inset, ESI-MS of the product corresponding to the main peak.

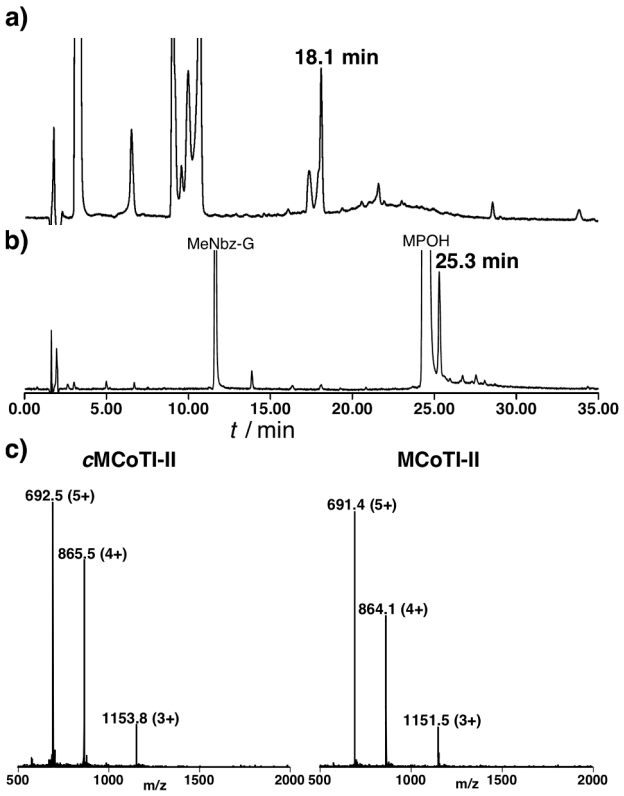


Figure 10. **a)** one-pot cyclization-folding of **MCoTI-II-MeNbz-G** crude synthesis (peak at 18.1 min is the expected **MCoTI-II** product); **b)** cyclization of **MCoTI-II-MeNbz-G** crude synthesis (peak at 25.3 min matches the cyclic **MCoTI-II** product, ***c*MCoTI-II**); c) ESI-MS spectra of ***c*MCoTI-II** (left), and folded **MCoTI-II** (right).

3. Discussion.

Efficient methods for ligating unprotected peptides in aqueous solution has greatly facilitated the chemical synthesis and semisynthesis of proteins, including proteins with PTMs, incorporation of non-natural amino acids or site-specific labeling. By enabling the production of homogeneous forms of these macromolecules, precise structure determinations, and structure-activity relationships of enzymes, substrates and ligand binding proteins have been interrogated. Native Chemical Ligation has become the reaction of choice for CPS, especially because its wide compatibility with biological macromolecules. NCL relies in the condensation between two unprotected peptides, a C-terminal thioester and another N-terminal cysteine, in aqueous buffer at neutral pH and room temperature. From a synthetic point of view, the critical element of utilizing NCL is the efficient production of C-terminal peptide-thioesters, which can be easily prepared in Boc-SPPS using mercaptoalkyl linkers. Complementary Fmoc-based methods for the synthesis of peptide-thioesters continue to be developed although it has been challenging to attain suitable approaches that are robust for chain elongation by SPPS while maintaining efficient C-terminal activation, avoiding epimerization and maintaining peptide solubility.

We hypothesized that C-terminal *N*-acylurea peptides[61] could act as precursors for thioester peptides and undergo exchange with thiols to yield a C-terminal thioester for NCL. Using *o*-aminoanilide linkers based on diaminobenzoic acid, we found that *N*-acylurea peptides undergo fast *N*-acylurea –> thiol exchange under NCL conditions, and consequently allow their utilization for CPS. This approach has been widely used for the synthesis of proteins, including proteins with PTMs. However, under strongly basic conditions and in Gly rich sequences, the acylation of the unprotected *o*-aminoanilide, can give rise to branched side products that cannot not be further converted in *N*-acylurea peptides.[33,38]

In order to analyze this side reaction in a systematic manner, we designed a pentaglycine containing peptide **9**, and under basic coupling conditions, indeed observed a high ratio of doubly-acylated products. While these products could be significantly reduced using a more acidic protocol and a lower excess of Fmoc-Gly during coupling. Nevertheless, a robust and universally applicable Dbz linker should be compatible with all coupling conditions used for SPPS while maintaining the simplicity of the Dbz linker with respect to activation with mild chloroformate derivatives and avoidance of unnecessary protecting groups.

The 3-amino-4-(methylamino)benzoic acid (MeDbz) linker was designed to enable chemoselectivity between the two aryl-amines by *N*-alkylation. Acylation at the unhindered primary aromatic amine would facilitate chain elongation, while di-acylation at the *N*-methylated position would be significantly reduced. Critical to the approach was maintaining efficient formation of the desired *N*-acylurea (MeNbz) product on resin using nitrophenyl chloroformate. Gratifyingly, the MeDbz linker showed superior performance in the synthesis of a pentaglycine sequence **10** when compared to the Dbz moiety (peptide **9**). The *o*-(methylamino)anilide functional group is significantly more resistant towards acylation during amino acid couplings, even withstand-

ing microwave coupling conditions with temperatures up to 90 C. Critically, the secondary *o*-(methylamino)anilide is sufficiently reactive towards chloroformate, and the resulting *N*-methylcarbamate undergoes efficient cyclization under basic conditions to afford the desired C-terminal MeNbz peptide. While this activation can be successfully implemented with many solvents and bases, for large peptides we suggest a two-step protocol in which the MeNbz resin is treated with nitrophenyl chloroformate (10 eq either in CH_2Cl_2 or $ClCH_2CH_2Cl$, 0.25 M) for 2 h, followed by washing and treatment with 1 M DIEA in DMF for 1 h. Since reactions on large peptides are often strongly affected by sequence dependent aggregation effects, we recommend a small test cleavage to confirm full activation of the MeDbz resin.

The high resistance towards di-acylation also facilitates the introduction of the C-terminal amino acid, which previously required mild acylation conditions that was tailored to the individual amino acid. Loading the MeDbz linker with a 5-fold excess of amino acid / HATU / excess DIEA, allows efficient incorporation of most amino acids. However, for His and Cys a more conservative protocol based on activation with DIC and HOOBt was employed in order to avoid the chance of racemization. Using these conditions we did not observe epimerization in the final peptide-MeNbz peptides or their ligation products.

NCL conditions were optimized to suppress the MeNbz hydrolysis and maximize ligation rate. We find that a pH of 7.0 and a thiol concentration (4-mercaptophenol) of 100 mM is generally suitable for peptide ligation reactions. We observed a direct dependence of the ligation rate with the concentration of the thiol additive which suggests that the *N*-acylurea –> thiol transthioesterification is rate determining when the additive is below 100 mM. These optimized conditions were further applied to the ligation of model peptides corresponding to all 20 amino acids at the C-terminal position. Similar to previous, the peptides could be grouped into three different categories based on ligation rate: *i.* 17 out of 20 gave ligations in a 5 h time frame, *ii.* the ß-branched amino acids valine and isoleucine reacted more slowly, needing ~20 h to achieve conversions above 90%, *iii.* proline, which requires around 48 h to achieve conversions near to 90%. While ligation reactions at Asp gave a significant ratio of the ß-linked isomer, likely as a consequence of anhydride formation, the analogous reaction with Glu resulted in only traces of isomeric products. Interestingly, a comparison of the kinetics between the Phe-mercaptopropionyl thioester and the Phe-*N*-methylacylurea, both at a concentration of 2 mM and using the standard ligation conditions (100 mM MPOH, 20 mM TCEP, pH 7.0, rt in 6 M guanidine buffer) yielded similar ligation rates (**Figure S19**) showing that both display similar reactivity.[62]

Since the primary use of these C-terminally activated peptides is for the synthesis of complex proteins by NCL, we synthesized two members of the cyclotide family: Kalata B1 and MCoTI-II. Cyclotides are a special class of PTM proteins found in plants with a head to tail cyclic structure and a cystine knot topology. Cyclotides own a high chemical and proteolytic stability, and some of their different loops hold a significant flexibility that allows the exchange by non native cyclotide sequences of approximately the same length. This versatility, the oral bioavailability and the stability are fundamental advantages for the grafting of small peptide motifs with potential therapeutic use, and thus why it is conceived the utilization of the cyclotides as scaffolds in drug design.

Synthesis of cyclotides has been previously accomplished, either using Boc or Fmoc chemistry, but with no satisfactory results. In Fmoc chemistry, the main problems arise from the number of cysteines, which apart from a higher tendency to racemize, have the propensity to capture ^{t}Bu cations during the cleavage step. Preliminary synthesis of KB1 carried out on a polystyrene resin gave poor analytical LC-MS profiles, with deletions in some amino acids and a high abundance of ^{t}Bu adducts. Chain elongation on a more hydrophilic polymer such as PEG-polysterene resin (Tentagel) led to an improvement of the crude purity, but a significant side-product was observed, with a mass consistent with a ^{t}Bu adduct. Shifting to a fully pegylated resin, such as ChemMatrix, the desired KB1-MeNbz-G was obtained following cleavage in an analytical purity around 65%, although the ^{t}butylated side-product was still present. Changing the protecting group scheme on ^{20}Thr and ^{22}Ser (^{t}Bu –> Trt) or swapping the cyclization point did not significantly reduce this side-product. However, the choice of the Phacm as cysteine protecting group yielded a linear crude peptide with no major side-products, which let us to surmise that likely one of the cysteine residues was scavenging the ^{t}Bu group. Nevertheless, using a one-step approach in a GSH:GSSG buffer, it was possible to perform the cyclization and folding of the linear KB1-MeNbz-G yielding the fully folded KB1 in recovered overall yields around 20%. Following the same approach, MCoTI-II-MeNbz-G was cyclized and folded in a similar yield. Interestingly, KB1 and MCoTI-II have very different HPLC elution profiles with regard to their reduced cyclic precursors: whereas folded KB1 is remarkably more hydrophobic, native MCoTI-II is more hydrophilic, reflecting the side chains that are presented towards solvent when the disulfide rich hydrophobic core is formed. These different hydrophobic characteristics of the cyclotides may be significant when designing engineered cyclotides with biomedical applications.

In summary, we have developed a second generation of the *N*-acylurea linker, the MeNbz, with significant practical advantages over the first generation Nbz linker. Importantly, MeDbz is resistant to multiple acylation cycles of Fmoc-Gly, forcing coupling conditions with HATU and excess base and, remarkably, is compatible with coupling at elevated temperatures of 90 C (microwave). While the linker is robust towards undesired acylation, the peptide-MeDbz linker can be transformed to the desired MeNbz (*N*-acylurea) in a straightforward manner using nitrophenyl chloroformate. Importantly, the resulting mildly activated peptide-MeNbz products are competent for ligation by NCL in the presence of thiol additives. The MeDbz linker represents an

important step forward towards the development of a rapid and efficient Chemical Protein Synthesis technology, because with the current fast-SPPS technologies which includes, for instance, rapid flow-based peptide synthesis[63] and microwave heating,[64] it is easy to envision that 30 to 40-mer MeNbz peptides can be synthesized in under 5 h, and then assembled by NCL, speeding enormously the synthesis of proteins by chemical means.

4. Experimental section.

4.1 Synthesis of Fmoc-MeDbz (8).

To a mixture of 4-Fluoro-3-nitrobenzoic acid (10 g, 0.054 mol) in MeOH (150 mL) was added a $MeNH_2$ solution (40% in MeOH, 40 mL). The resulting mixture was stirred at room temperature overnight, and then poured into 150 mL of water. The reaction was acidified with aqueous HCl (37%) until a yellow solid crashed out from the solution. The product was collected by filtration, washed with H_2O and finally dried out. Next, it was dissolved in MeOH (500 mL) and hydrogenated over 500 mg of Pd/C (10%) overnight. The catalyst was separated by filtration, and the solvent removed under reduced pressure in a rotary evaporator. The 3-amino-4-(methylamino)benzoic acid was re-dissolved in a H_2O/CH_3CN mixture (1:1, 100 mL) containing DIEA (9 mL, 0.052 mol) and a solution of Fmoc-Cl (12.8 g, 0.049 mol) in CH_3CN (50 mL) poured dropwise. Following completion of the reaction, the CH_3CN was removed and the product collected by filtration, washed with CH_3CN, and dried out under vacuum (15 g, 71%). ^{1}H-NMR (400 MHz, [DMSO-D6], 25ºC): 2.79 (s, 3H), 4.30-4.40 (m, 3H), 6.62 (t, J = 6.6 Hz, 1H), 7.30-7.40 (m, 2H), 7.45 (t, J = 7.4 Hz, 2H), 7.63-7.80 (m, 3H), 7.92 (d, J = 7.9 Hz, 2H), 8.73 (s, 1H), 12.13 (s 1H). ^{13}C-NMR (100 MHz, [DMSO-D6], 25ºC): 29.5, 46.7, 66.0, 108.8, 116.7, 120.1, 122.0, 125.3, 127.1, 127.7, 140.7, 143.8, 154.7, 167.3

ESI-MS: calcd. ($C_{23}H_{21}N_2O_4$, M + H) 389.4, found 389.1

4.2 Synthesis of Fmoc-Gly$_5$-Nbz-R (9).

Method A: starting with 0.05 mmol of Rink-amide functionalized resin, following coupling of Arg and Dbz, Fmoc-Gly (0.2 mmol) and HBTU (0.2 mmol) were dissolved in DMF (1 mL) and DIEA (0.035 mL, 0.2 mmol) was added. The amino acid was preactivated for 30 s, and then added to the resin. The resulting suspension was bubbled with a gentle stream of N_2, and after 40 min the resin was filtered and washed with DMF. The next Fmoc-Gly residues were incorporated using 0.5 mmol of amino acid activated with a solution of HBTU 0.5 M in DMF (0.5 mmol of HBTU) and DIEA (1.0 mmol). Coupling times were 1 h. Following last glycine, the *o*-(methylamino)anilide peptide was acylated with a solution of *p*-nitrophenyl chloroformate (0.25 mmol) in CH_2Cl_2 (1 mL) during 1 h. Then, the resin was filtered, washed with CH_2Cl_2 and treated with a solution of DIEA (0.5 M in DMF) during 15 min. After washing the resin with DMF, CH_2Cl_2 and finally dried, it was cleaved using the standard cleavage cocktail (TFA/TIS/H_2O, 95:2.5:2.5).

Method B: following exactly the same protocol as in method A until the introduction of the first glycine, then the next 4 Fmoc-Gly amino acids (0.2 mmol per coupling cycle) were coupled using a solution of HBTU/HOBt (0.2 M in DMF, 0.2 mmol HBTU, 0.2 mmol HOBt) and DIEA (0.3 mmol) during 1 h. Next, the peptide was acylated with *p*-nitrophenyl chloroformate, cyclized under basic treatment and cleaved from the resin using acydolitic conditions.

4.3 Synthesis of Fmoc-Gly$_5$-MeNbz-R (10).

Method C: starting with 0.1 mmol of Rink-amide functionalized resin, following coupling of Arg and Dbz, Fmoc-Gly (0.5 mmol) was dissolved in a HBTU solution (0.5 M, 0.5 mmol, 1 mL) and DIEA (0.096 mL, 0.55 mmol). The amino acid was preactivated for 30 s, and then added to the resin. The resulting suspension was bubbled with a gentle stream of N_2, and after 1 h the resin was filtered and washed with DMF. The next Fmoc-Gly residues were incorporated using 0.5 mmol of amino acid activated with a solution of HBTU 0.5 M in DMF (0.5 mmol of HBTU) and DIEA (0.096 mL, 0.55 mmol). Coupling times were 1 h. Following last glycine, the *o*-(methylamino)anilide peptide was acylated with a solution of *p*-nitrophenyl chloroformate (0.5 mmol) in CH_2Cl_2 (2 mL) during 60 min. Then, the resin was filtered, washed with CH_2Cl_2 and treated with a solution of DIEA (0.5 M in DMF) during 15 min. After washing the resin with DMF, CH_2Cl_2 and finally dried, it was cleaved using the standard cleavage cocktail.

Method MW1: starting with 0.1 mmol of Rink-amide resin, and using the same protocol as in method C until the introduction of the first Gly residue (Fmoc-Gly-MeDbz-Arg(Pbf)-resin), the next 4 Fmoc-Gly residues were introduced under microwave conditions using the following setup:

i. coupling: Fmoc-Gly (1.0 mmol, 2.5 mL, 0.4 M in DMF), DIC (1.0 mmol, 2.0 mL, 0.5 M in DMF) and Oxyma Pure (1.0 mmol, 1 mL, 1.0 M in DMF) were added to the resin following this order; $[\text{Fmoc-Gly, DIC, Oxyma Pure}]_{final}$ = 0.18 M. The resulting suspension was heated under microwave irradiation: first 20 s at 75 ºC, power = 170 watts; then, 3 min at 90 ºC, power = 50 watts.

ii. fmoc removal: first 30 s at 75 ºC, power= 155 watts; next, 90 s at 90 ºC, power = 30 watts.

Following last Gly, the Fmoc-Gly$_5$-MeDbz-R peptide was acylated with *p*-nitrophenyl chloroformate, cyclized under basic conditions to yield, after acydolitic cleavage, the Fmoc-Gly$_5$-MeNbz-R peptide.

4.4 Synthesis of model MeNbz peptides (11A-Y).

The following protocols (**a**-**c**) describe the incorporation on solid phase of the first amino acid after the MeDbz linker. HATU is used as a general coupling agent. Nevertheless, in order to avoid racemization in some amino acids such as His and Cys a different protocol is recommended. Moreover, it is not strictly necessary the utilization of excess of base for all amino acids, especially those that are not β-branched. Additionally, Fmoc-Gly is easily incorporated using HBTU.

The coupling conditions refer to 0.1 mmol of resin (Rink-PS, 0.69 mmol/g).

a) introduction of Gly: Fmoc-Gly (149 mg, 0.5 mmol) and HBTU (189 mg, 0.5 mmol) were dissolved in DMF (1 mL) and DIEA (0.096 mL, 0.55 mmol) was added. After 30 s, the solution was added to the resin and gently stirred under a N_2 stream for 40 min. Then, the resin was washed with DMF and the peptide elongation continued using standard conditions.

b) introduction of His and Cys: Fmoc-His/Cys(Trt)-OH (0.5 mmol) and HOOBt (82 mg, 0.5 mmol) were dissolved in DMF (1 mL) and DIC (0.077 mL, 0.50 mmol) was added. After 3 min, the solution was added to the resin and gently stirred under a N_2 stream for 90 min. Following coupling, the resin was washed with DMF and the peptide elongation continued following standard conditions.

c) coupling of the remaining amino acids: Fmoc-aa (0.5 mmol) and HATU (190 mg, 0.5 mmol) were dissolved in DMF (1 mL) and DIEA (0.191 mL, 1.1 mmol) was added. After 30 s, the solution was added to the resin and gently stirred under a N_2 stream for 90 min. After coupling, the resin was washed with DMF and the chain elongation carried on following standard conditions.

Coupling of the VSYRA sequence following the C-terminal amino acid: the Fmoc/Boc-aa (0.5 mmol) was dissolved in a HBTU solution (0.5 M in DMF, 1.0 mL) and DIEA (0.096 mL, 0.55 mmol) was added. After 30 s, the solution was added to the resin and gently stirred under a N_2 stream for 30 min. The last Val was coupled as Boc-Val. Fmoc removal was carried out with a piperidine solution (20% in DMF) for 5 min. Following last residue, the resin was washed with DMF and CH_2Cl_2, and a solution of *p*-nitrophenyl chloroformate (100 mg, 0.5 mmol) in CH_2Cl_2 (2 mL) was added to the resin and the resulting mixture shaken for 1 h. Next, the solution was removed and the resin washed with CH_2Cl_2. A solution of DIEA (0.5 M or 1 M in DMF) was added to the resin and stirred under a N_2 stream. β-branched residues (Val and Ile) were treated with DIEA 1 M for 30 min. For the remaining amino acids, the cyclization step was carried out in a solution of DIEA 0.5 M for 15-30 min.

Following acylurea formation, the resin was washed with CH_2Cl_2, dried under vacuum and cleaved using the standard cleavage solution for 1 h. Then, the TFA peptide containing solution was concentrated in a rotavapor, and precipitated over Et_2O, spun out and the solid redissolved in a H_2O/CH_3CN solution and lyophilized. Typical recovery yields of crude peptides were above 90% and analytical crude purities ranged from 90 to 95%.

4.5 Synthesis of cyclotides.

The linear synthesis of the cyclotides KB1 and MCoTI-II was carried out on a ChemMatrix solid support (0.52 mmol/g) on a 0.1 mmol scale. Following coupling of Gly and MeDbz linker, the C-terminal amino acid was introduced using the conditions reported above for each specific residue.

Manual synthesis (**method D**): next residues after the C-terminal were coupled using 1.0 mmol of Fmoc-aa preactivated during 30 s with a solution of HBTU 0.5 M (1.0 mmol of HBTU) and DIEA (0.191 mL, 1.1 mmol). Coupling times were typically 30 min. For the Fmoc removal, a solution of piperidine (20% in DMF) was added to the resin and the suspension shaken for 10 min. The N-terminal cysteine was introduced as Boc-Cys(Trt)-OH. Following last amino acid, the resin was washed with CH_2Cl_2 and a solution of *p*-nitrophenyl chloroformate (201 mg, 1.0 mmol) in CH_2Cl_2 (4 mL) was added and the resulting suspension shaken for 2 h. Following carbamylation, cyclization to yield the acylurea peptide was carried out by treatment with a basic solution of DIEA (1 M in DMF) during 1 h. Finally the peptidyl-resin was washed with CH_2Cl_2 and dried under vacuum. Cleavage with cocktail 'R' (TFA, TIS, H_2O, EDT; 92.5:2.5:2.5:2.5), with or without 1,2-Dimethylindole (2.5%), yielded the linear KB1-MeNbz-G, KB1$(Phacm)_5$-MeNbz-G and MCoTI-II-MeNbz-G cyclotides. KB1-MeNbz-G: 430 mg resin –> 133 mg crude (73%); ESI-MS: calcd. ($C_{128}H_{197}N_{39}O_{42}S_6$, M^+ average) 3146.58, found 3145.7. KB1$(Phacm)_5$-MeNbz-G: 582 mg resin –> 250 mg crude (77%); ESI-MS: calcd. ($C_{173}H_{242}N_{44}O_{47}S_6$, M^+ average) 3882.46, found 3882.5; MCoTI-II-MeNbz-G: 653 mg resin –> 266 mg crude (81%); ESI-MS: calcd. ($C_{149}H_{241}N_{51}O_{48}S_6$, M^+ average) 3707.24, found 3707.0.

Automated microwave synthesis of KB1 (**method MW2**): following manual couplings of Gly, MeDbz linker and the C-terminal Val residue, the remaining amino acids were introduced under microwave conditions using a 10-fold excess of amino acid (1.0 mmol) and HBTU/DIEA (1.0:1.2 mmol, respectively). The final concentration of Fmoc-aa and HBTU was 0.125 M, and for DIEA 0.15 M. The coupling cycles were as following:

i. coupling: a solution of Fmoc-aa (0.2 M, 5 mL), HBTU (0.5 M, 2 mL) and DIEA (1.2 M, 1 mL) were added to the resin following this order, and the resulting suspension was heated up under microwave irradiation during 5 min at 75 ºC with a power = 27 watts. Cys was coupled at 50 ºC keeping identical the rest of the parameters.

Arg double coupling: *1st coupling*, T = 25 ºC, power = 0 watts, t = 30 min; then, T = 75 ºC, power = 27 watts, t = 5 min. *2nd coupling*: T = 75 ºC, power = 27 watts, t = 5 min.

ii. fmoc removal: *1st deprotection*, t = 30 s, T = 75 ºC, power = 55 watts; *2nd deprotection*, t = 180 s, T= 75 ºC, power = 65 watts.

KB1-MeNbz-G: 452 mg resin –> 142 mg crude (73%).

Cyclization of linear cyclotides (***c*KB1** and ***c*MCoTI-II**): the crude linear peptides were dissolved in the ligation buffer (6 M guanidine, 0.2 M sodium phosphate, 50 mM MPOH, 20 mM TCEP, pH 7.0) at a final concentration around 1 mM. The progress of the reaction was monitored by LC-MS, and after completion of the intramolecular ligation, the cyclic cyclotides were purified by reverse-phase preparative HPLC. *c*KB1: 103 mg crude –> 30 mg (33%); MALDI-TOF: calcd.

($C_{117}H_{185}N_{35}O_{39}S_6$, M^+ average) 2898.33, found 2898.2. *c*MCoTI-II: 50 mg crude –> 19 mg (41%); ESI-MS: calcd. ($C_{138}H_{229}N_{47}O_{45}S_6$, M^+ average) 3458.99, found 3458.5.

Folding of cyclic-cyclotides: *c*KB1 and *c*MCoTI-II were dissolved in a $(NH_4)_2CO_3$ (0.2 M in H_2O)/^{i}PrOH solution (1:1) containing GSH:GSSG (2 mM:0.4 mM) at a final concentration around 0.2 mM and pH 8.5. The folding reaction was monitored by LC-MS, and after proper cystine formation, the solution was acidified with aqueous HCl until pH 2-3, the ^{i}PrOH removed by rotary evaporation and the folded cyclotides purified by reverse-phase. KB1: 25 mg –> 16 mg (64%); MALDI-TOF: calcd. ($C_{117}H_{179}N_{35}O_{39}S_6$, M^+ average), 2892.28 found 2892.0. MCoTI-II: 13 mg –> 10 mg (77%) ESI-MS: calcd. ($C_{138}H_{223}N_{47}O_{45}S_6$, M^+ average) 3452.95, found 3452.3.

One step cyclization-folding of the linear cyclotides: the crude linear cyclotides (KB1-MeNbz-G and MCoTI-II-MeNbz-G) were dissolved in a $(NH_4)_2CO_3$ (0.2 M in H_2O)/^{i}PrOH solution (1:1) containing GSH:GSSG (10 mM: 2.0 mM) at a final concentration around 0.5 mM and a pH 7.5. After folding, the reaction was acidified with HCl until pH 2-3, the ^{i}PrOH removed by rotary evaporation and the folded cyclotides purified by reverse-phase. KB1: 125 mg crude –> 20 mg (18%). MCoTI-II: 100 mg crude –> 20 mg (22%).

4.6 Protocol for ligation reactions at different pHs, varying concentrations of MPOH and model peptides.

A degassed aqueous solution of guanidine hydrochloride (6 M), sodium phosphate (0.2 M), TCEP (20 mM) and the MPOH concentration of the assay, was fitted to the desired pH with either an aqueous $NaOH_{(aq)}$ (concentrated) or HCl (2 M in H_2O) solution. The ligation buffer was added to a mixture of MeNbz peptides **11A-Y** and **CRAFS** to achieve a final concentration of 2 mM and 3-3.5 mM, respectively. The resulting solution was stirred at room temperature, and at the indicated times aliquots of the reaction (50 µL) were taken and quenched with a solution of hydroxylamine hydrochloride (50 µL, 1 M in H_2O, pH ~ 5.0). The mixtures were analyzed by HPLC at 220 and 280 nm and all the signals with a Tyr absorption profile were integrated at 280 nm to calculate the fraction of ligated peptide. The same ligations were run in parallel until completion without hydroxylamine quenching, and the kinetics and product distribution analyzed obtaining similar results, which supports the kinetics analysis (**Figure S10**).

ASSOCIATED CONTENT

Supporting Information. NMR characterization of **Fmoc-MeDbz**, LC-MS of the synthesis of pentaglycine peptides **9** and **10**, model peptides **11A-Y**, epimerization studies on His and Cys, ligations with model peptides (**12A-Y**), analytical LC of **KB1-MeNbz-G** synthesis, analytical LC of pure ***c*KB1** and cyclic ***c*MCoTI-II**, pure folded **KB1** and folded **MCoTI-II**, ligation kinetics of **11F** and **VSYRAF-(SCH_2CH_2CO)-G** with **CRAFS**, an expanded view of rate of ligation with model peptides **11A-Y** and **CRAFS** (Figure 4a), and a flow chart showing the synthetic route and ligation of MeNbz peptides. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

* dawson@scripps.edu

ACKNOWLEDGMENT

The present work has been funded by the National Institutes of Health R01 GM098871 (PED) Spanish Ministerio de Economía y Competitividad through the grants CTQ2012-31197 and RYC-2011-09001 (JBBC).

REFERENCES

(1) a) Dawson, P. E.; Kent, S. B. *Annu. Rev. Biochem.* **2000**, *69*, 923-960; b) Kent, S. B. *Chem. Soc. Rev.* **2009**, *38*, 338-351; c) Payne, R. J.; Wong, C. H. *Chem. Commun.* **2010**, *46*, 21-43; d) Kan, C.; Danishefsky, S. J. *Tetrahedron* **2009**, *65*, 9047-9065; e) Unverzagt, C.; Kajihara, Y. *Chem. Soc. Rev*. **2013**, *42*, 4408-4420.

(2) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266*, 776-779.

(3) Yan, L. Z.; Dawson, P. E. *J. Am. Chem. Soc*. **2001**, *123*, 526-533.

(4) Crich, D.; Banerjee, A. *J. Am. Chem. Soc.* **2007**, *129*, 10064-10065.

(5) Harpaz, Z.; Siman, P.; Kumar, K. S.; Brik, A. *Chembiochem* **2010**, *11*, 1232-1235.

(6) Cergol, K. M.; Thompson, R. E.; Malins, L. R.; Turner, P.; Payne, R. J. *Org. Lett.* **2014**, *16*, 290-293.

(7) Thompson, R. E.; Chan, B.; Radom, L.; Jolliffe, K. A.; Payne, R. J. *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 9723-9727.

(8) Siman, P.; Karthikeyan, S. V.; Brik, A. *Org. Lett.* **2012**, *14*, 1520-1523.

(9) a) Haase, C.; Rohde, H.; Seitz, O. *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 6807-6810; b) Chen, J.; Wan, Q.; Yuan, Y.; Zhu, J.; Danishefsky, S. J. *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 8521-8524.

(10) Townsend, S. D.; Tan, Z.; Dong, S.; Shang, S.; Brailsford, J. A.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2012**, *134*, 3912-3916.

(11) Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X. W.; Liu, C. F. *J. Am. Chem. Soc.* **2009**, *131*, 13592-13593.

(12) Malins, L. R.; Cergol, K. M.; Payne, R. J. *Chembiochem* **2013**, *14*, 559-563.

(13) Chen, J.; Wang, P.; Zhu, J.; Wan, Q.; Danishefsky, S. J. *Tetrahedron* **2010**, *66*, 2277-2283.

(14) a) Offer, J.; Dawson, P. E. *Org. Lett.* **2000**, *2*, 23-26; b) Offer, J.; Boddy, C. N.; Dawson, P. E. *J. Am. Chem. Soc.* **2002**, *124*, 4642-4646; c) Canne, L. E.; Bark, S. J.; Kent, S. B. H. *J. Am. Chem. Soc.* **1996**, *118*, 5891-5896; d) Offer, J. *Biopolymers* **2010**, *94*, 530-541.

(15) a) Hackeng, T. M; Griffin, J. H; Dawson, P. E. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 10068-10073; b) Camarero, J. A.; Adeva, A.; Muir, T. W. *Lett. Pep. Sci.* **1999**, *7*, 17-21; c) Hojo, H.; Aimoto, S. *Bull. Chem. Soc. Jpn.* **1991**, *64*, 111-117.

(16) a) Futaki, S.; Sogawa, K.; Maruyama, J.; Asahara, T.; Niwa, M.; Hojo, H. *Tet. Lett.* **1997**, *38*, 6237-6240; b) Mezo, A. R.; Cheng, R. P.; Imperiali, B. *J. Am. Chem. Soc.* **2001**, *123*, 3885-3891.

(17) Alsina, J.; Yokum, T. S.; Albericio, F.; Barany, G. *J. Org. Chem.* **1999**, *64*, 8761-8769.

(18) Sewing, A.; Hilvert, D. *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 3395-3396.

(19) a) Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **1999**,

121, 11684-11689; b) Ingenito, R.; Bianchi, E.; Fattori, D.; Pessi, A. *J. Am. Chem. Soc.* **1999**, *121*, 11369-11374.

(20) Camarero, J. A.; Hackel, B. J.; de Yoreo, J. J.; Mitchell, A. R. *J. Org. Chem.* **2004**, *69*, 4145-4151.

(21) a) Ollivier, N.; Dheur, J.; Mhidia, R.; Blanpain, A.; Melnyk, O. *Org Lett* **2010**, *12*, 5238-5241; b) Ollivier, N.; Vicogne, J.; Vallin, A.; Drobecq, H.; Desmet, R.; El Mahdi, O.; Leclercq, B.; Goormachtigh, G.; Fafeur, V.; Melnyk, O. *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 209-213; c) Sato, K.; Shigenaga, A.; Tsuji, K.; Tsuda, S.; Sumikawa, Y.; Sakamoto, K.; Otaka, A. *Chembiochem* **2011**, *12*, 1840-1844; d) Zheng, J. S.; Chang, H. N.; Wang, F. L.; Liu, L. *J. Am. Chem. Soc.* **2011**, *133*, 11080-11083.

(22) a) Botti, P.; Villain, M.; Manganiello, S.; Gaertner, H. *Org. Lett.* **2004**, *6*, 4861-4864; b) Warren, J. D.; Miller, J. S.; Keding, S. J.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2004**, *126*, 6576-6578.

(23) a) Fang, G. M.; Li, Y. M.; Shen, F.; Huang, Y. C.; Li, J. B.; Lin, Y.; Cui, H. K.; Liu, L. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 7645-7649; b) Wang, P.; Layfield, R.; Landon, M.; Mayer, R. J.; Ramage, R. *Tetrahedron Lett.* **1998**, 39, 8711-8714; c) Wang, P.; Shaw, K. T.; Whigham, B.; Ramage, R. *Tetrahedron Lett.* **1998**, 39, 8719-8720; d) Wang, P.; Layfield, R.; Landon, M.; Mayer, R.; Ramage, R. J. *J. Peptide Res.* **1999**, *53,* 673-677; e) Wang, P. *Tetrahedron Lett.* **2007**, *48*, 7313-7315.

(24) Blanco-Canosa, J. B.; Dawson, P. E. *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 6851-6855.

(25) Tiefenbrunn, T. K.; Blanco-Canosa, J.; Dawson, P. E. *Biopolymers* **2010**, *94*, 405-413.

(26) a) Hemantha, H. P.; Brik, A. *Bioorg. Med. Chem.* **2013**, *21*, 3411-3420; b) Hemantha, H. P.; Bavikar, S. N.; Herman-Bachinsky, Y.; Haj-Yahya, N.; Bondalapati, S.; Ciechanover, A.; Brik, A. *J. Am. Chem. Soc.* **2014**, *136*, 2665-2673.

(27) Zhang, Y.; Xu, C.; Lam, H. Y.; Lee, C. L.; Li, X. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 6657-6662.

(28) a) Gunasekera, S.; Aboye, T. L.; Madian, W. A.; El-Seedi, H. R.; Goransson, U. *Int. J. Pept. Res. Ther.* **2013**, *19*, 43-54; b) Akcan, M.; Craik, D. J. *Methods Mol. Biol.* **2013**, *1047*, 89-101.

(29) Fauvet, B.; Butterfield, S. M.; Fuks, J.; Brik, A.; Lashuel, H. A. *Chem. Commun.* **2013**, *49*, 9254-9256.

(30) Lindgren, J.; Ekblad, C.; Abrahmsen, L.; Karlstrom, A. E. *Chembiochem* **2012**, *13*, 1024-1031.

(31) Zhan, C.; Varney, K.; Yuan, W.; Zhao, L.; Lu, W. *J. Am. Chem. Soc.* **2012**, *134*, 6855-6864.

(32) Harris, P. W. R.; Brimble, M .A. *Int. J. Pept. Res. Ther.* **2012**, *18*, 63-70.

(33) Mahto, S. K.; Howard, C. J.; Shimko, J. C.; Ottesen, J. J. *Chembiochem* **2011**, *12*, 2488-2494.

(34) Siman, P.; Blatt, O.; Moyal, T.; Danieli, T.; Lebendiker, M.; Lashuel, H. A.; Friedler, A.; Brik, A. *Chembiochem* **2011**, *12*, 1097-1104.

(35) Pentelute, B. L.; Barker, A. P.; Janowiak, B. E.; Kent, S. B.; Collier, R. J. *ACS. Chem. Biol.* **2010**, *5*, 359-364.

(36) Okamoto, R.; Mandal, K.; Ling, M.; Luster, A. D.; Kajihara, Y.; Kent, S. B. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 5194-5198.

(37) Chiang, K. P.; Jensen, M .S.; McGinty, R. K.; Muir, T. W. *Chembiochem* **2009**, *10*, 2182-2187.

(38) White, P. D.; Behrendt, R. *J. Pep. Sci.* **2010**, *16*, 71-72.

(39) Kuzniewski, C. N.; Gertsch, J.; Wartmann, M.; Altmann, K. H. *Org. Lett.* **2008**, *10*, 1183-1186.

(40) Proft, T. *Biotechnol. Lett.* **2010**, *32*, 1-10.

(41) Ling, J. J.; Policarpo, R. L.; Rabideau, A. E.; Liao, X.; Pentelute, B. L. *J. Am. Chem. Soc.* **2012**, *134*, 10749-10752.

(42) Otaka, A.; Sato, K.; Ding, H.; Shigenaga, A. *Chem. Rec.* **2012**, *12*, 479-490.

(43) a) Jones, J. H.; Ramage, W. I. *J. Chem. Soc., Chem. Comm.* **1978**, 472-473; b) Isidro-Llobet, A.; Alvarez, M.; Albericio, F. *Chem. Rev.* **2009**, *109*, 2455-2504.

(44) a) Han, Y.; Albericio, F.; Barany, G. *J. Org. Chem.* **1997**, *62*, 4307-4312; b) Angell, Y. M.; Alsina, J.; Albericio, F.; Barany, G. *J. Pept. Res.* **2002**, *60*, 292-299.

(45) It is important to mention that, probably due to traces amounts of H_2O in the solvents, and under some conditions, i.e. 1 M DBU in DMF can lead to complete hydrolysis of the peptide.

(46) Valeur, E.: Bradley, M. *Chem. Soc. Rev.* **2009**, *38*, 606-631.

(47) Nagalingam, A. C.; Radford, S. E.; Warriner, S. L. *Synlett* **2007**, 2517-2520.

(48) Johnson, E. C.; Kent, S. B. *J. Am. Chem. Soc.* **2006**, *128*, 6640-6646.

(49) Jencks, W. P.; Cordes, S.; Carriuolo, J. *J. Biol. Chem.* **1960**, *235*, 3608-3614.

(50) Bang, D.; Pentelute, B. L.; Gates, Z. P.; Kent, S. B. *Org. Lett.* **2006**, *8*, 1049-1052.

(51) Dang, B.; Kubota, T.; Mandal, K.; Bezanilla, F.; Kent, S. B. *J. Am. Chem. Soc.* **2013**, *135*, 11911-11919.

(52) Villain, M.; Gaertner, H.; Botti, P. *Eur. J. Org. Chem.* **2003,** 3267-3272.

(53) Interestingly, hydrolysis was only 15%, likely due to the rapid thiolysis of the anhydride intermediate.

(54) Craik, D. J.; Daly, N. L.; Bond, T.; Waine, C. *J. Mol. Biol.* **1999**, *294*, 1327-1336.

(55) Rosengren, K. J.; Daly, N. L.; Plan, M. R.; Waine, C.; Craik, D. J. *J. Biol. Chem.* **2003**, *278*, 8606-8616.

(56) a) Polanowski, A.; Wilusz, T.; Nienartowicz, B.; Cieslar, E.; Slominska, A.; Nowak, K. *Acta Biochim. Pol.* **1980**, *27*, 371-382; b) Hernandez, J. F.; Gagnon, J.; Chiche, L.; Nguyen, T. M.; Andrieu, J. P.; Heitz, A.; Hong, T. T.; Pham, T. T.; Nguyen, D. L. *Biochemistry* **2000**, *39*, 5722-5730; c) Heitz, A.; Hernandez, J. F.; Gagnon, J.; Hong, T. T.; Pham, T. T.; Nguyen, T. M.; Nguyen, D. L.; Chiche, L. *Biochemistry* **2001**, *40*, 7973-7983.

(57) Henriques, S. T.; Craik, D. J. *Drug Discov. Today* **2010**, *15*, 57-64; b) Garcia, A. E.; Camarero, J. A. *Curr. Mol. Pharmacol.* **2010**, *3*, 153-163.

(58) Tiefenbrunn, T. K.; Dawson, P.E. *Protein Sci.* **2009,** *18*, 970-979.

(59) a) Daly, N. L.; Love, S.; Alewood, P. F.; Craik, D. J. *Biochemistry* **1999**, *38*, 10606-10614; b) Gunasekera, S.; Daly, N. L.; Clark, R. J.; Craik, D. J. *Antioxid. Redox Signal.* **2009**, *11*, 971-980.

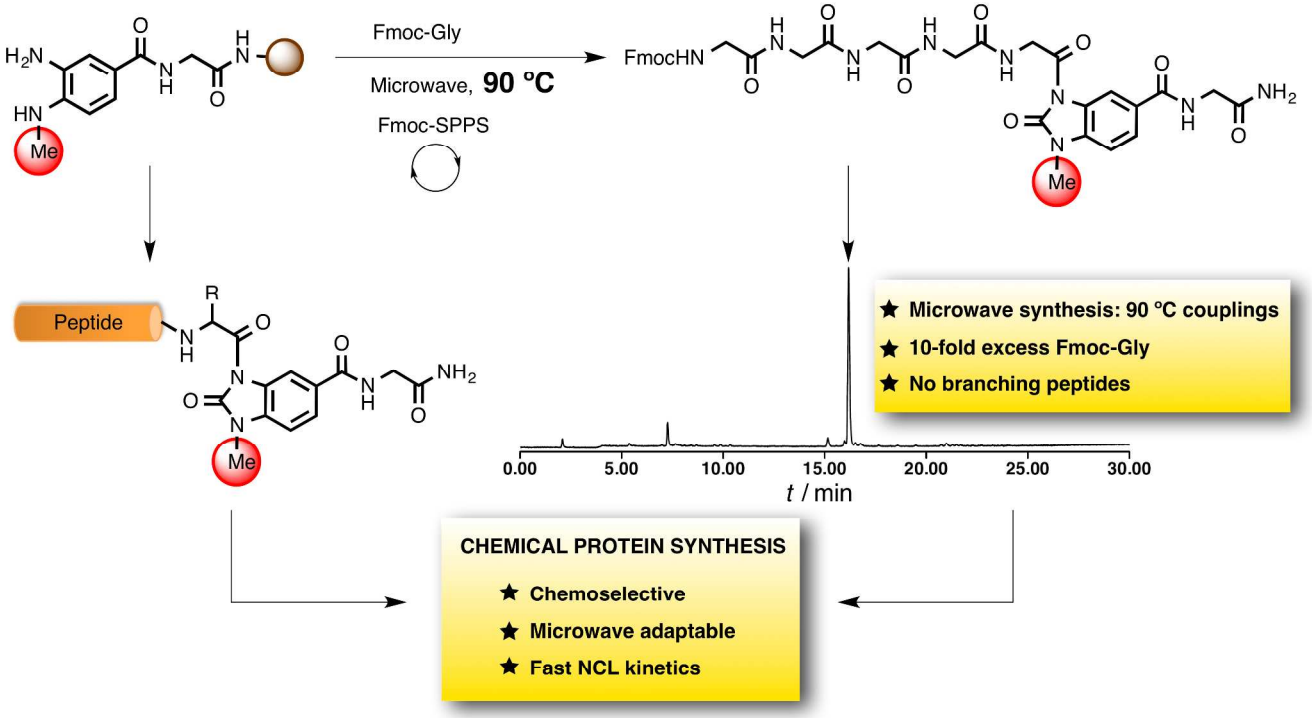
(60) a) Thongyoo, P.; Roque-Rosell, N.; Leatherbarrow, R. J.; Tate, E. W. *Org. Biomol. Chem.* **2008**, *6*, 1462-1470; b) Park, S.; Gunasekera, S.; Leta-Aboye, T.; Göransson, U. *Int. J. Pept. Protein Res.* **2010**, *16*, 167-176.

(61) a) Sola, R.; Saguer, P.; David, M. L.; Pascal, R. *J. Chem. Soc., Chem. Comm.* **1993**, 1786-1788; b) Pascal, R.; Chauvey, D.; Sola, R. *Tetrahedron Lett.* **1994**, *35*, 6291-6294.

(62) However, alkyl- and arylthioesters are more stable than *N*-acylureas which slowly hydrolyze even at acidic pH ~ 2. Despite this lower stability, its purification, isolation and handling are perfectly possible.

(63) Simon, M. D.; Heider, P. L.; Adamo, A.; Vinogradov, A. A.; Mong, S. K.; Li, X.; Berger, T.; Policarpo, R. L.; Zhang, C.; Zou, Y.; Liao, X.; Spokoyny, A. M.; Jensen, K. F.; Pentelute, B. L. *ChemBioChem* **2014**, *15*, 713-720.

(64) Collins, J. M.; Porter, K. A.; Singh, S. K.; Vanier, G. S. *Org. Lett.* **2014**, *16*, 940-943.


Fmoc-Gly
Microwave, 90 °C
Fmoc-SPPS
FmocHN
Peptide
Me
★ Microwave synthesis: 90 °C couplings
★ 10-fold excess Fmoc-Gly
★ No branching peptides
0.00
5.00
10.00
15.00
20.00
25.00
30.00
t / min
CHEMICAL PROTEIN SYNTHESIS
★ Chemoselective
★ Microwave adaptable
★ Fast NCL kinetics