

DOI: 10.1002/cbic.201200579

# Macrocyclization of Organo-Peptide Hybrids through a Dual Bio-orthogonal Ligation: Insights from Structure–Reactivity Studies

John R. Frost, Francesca Vitali, Nicholas T. Jacob, Micah D. Brown, and Rudi Fasan\*<sup>[a]</sup>

Macrocycles constitute an attractive structural class of molecules for targeting biomolecular interfaces with high affinity and specificity. Here, we report systematic studies aimed at exploring the scope and mechanism of a novel chemo-biosynthetic strategy for generating macrocyclic organo-peptide hybrids (MORPHs) through a dual oxime-/intein-mediated ligation reaction between a recombinant precursor protein and bifunctional, oxyamino/1,3-amino-thiol compounds. An efficient synthetic route was developed to access structurally different synthetic precursors incorporating a 2-amino-mercaptomethyl-aryl (AMA) moiety previously found to be important for macrocyclization. With these compounds, the impact of the synthetic precursor scaffold and of designed mutations within the genetically encoded precursor peptide sequence on macrocyclization efficiency was investigated. Importantly, the desired MORPHs were obtained as the only product from all the different synthetic precursors probed in this study and across peptide sequences comprising four to 15 amino acids. Systematic

mutagenesis of the “*i*–1” site at the junction between the target peptide sequence and the intein moiety revealed that the majority of the 20 amino acids are compatible with MORPH formation; this enables the identification of the most and the least favorable residues for this critical position. Furthermore, interesting trends with respect to the positional effect of conformationally constrained (Pro) and flexible (Gly) residues on the reactivity of randomized hexamer peptide sequences were observed. Finally, mechanistic investigations enabled the relative contributions of the two distinct pathways (side-chain→C-end ligation versus C-end→side-chain ligation) to the macrocyclization process to be dissected. Altogether, these studies demonstrate the versatility and robustness of the methodology to enable the synthesis and diversification of a new class of organo-peptide macrocycles and provide valuable structure–reactivity insights to inform the construction of macrocycle libraries through this chemo-biosynthetic strategy.

## Introduction

Peptide-containing macrocycles represent an attractive class of compounds for the development of chemical probes and therapeutics.<sup>[1]</sup> Growing interest in this structural class has been fueled by the identification of several bioactive cyclopeptides occurring in nature<sup>[2]</sup> as well as the recognition of the potential benefits conferred by backbone cyclization in terms of proteolytic stability,<sup>[3]</sup> cell permeability,<sup>[4]</sup> and protein-binding properties.<sup>[5]</sup> Methods for generating libraries of macrocyclic peptides can thus provide valuable sources of relevant ligand diversity, in particular toward the identification of agents that could target extended biomolecular interfaces.<sup>[1]</sup>

In this regard, generating conformationally constrained peptides from genetically encoded precursors offer several potential advantages. There is the opportunity to access vast chemical diversity ( $10^6$ – $10^{10}$  molecules) rapidly through combinatorial mutagenesis combined with the possibility of coupling library creation with powerful selection methods (e.g., phage-, yeast-, mRNA-, or ribosome-display) to isolate the member(s) that

exhibit the desired target-binding properties. A widespread approach used to restrict the conformational flexibility of ribosomal peptides is the formation of disulfide bridges between cysteine residues flanking randomized target sequences.<sup>[6]</sup> Alternatively, cysteine- or amine-reactive agents have been applied to introduce covalent crosslinks within displayed polypeptide sequences to generate macrocyclic or bicyclic peptide structures.<sup>[7]</sup> Other notable approaches in this area have involved the generation of cyclic peptides by means of split inteins,<sup>[8]</sup> through chemoenzymatic<sup>[9]</sup> or *in vitro* translation<sup>[10]</sup> methods, or by manipulating the enzymes involved in the biosynthesis of naturally occurring cyclopeptides.<sup>[11]</sup> Within these strategies, however, the pool of building blocks available for library construction is limited to the 20 natural amino acids, and thus remains considerably smaller than that potentially accessible through chemical synthesis.<sup>[12]</sup> This limitation has been in part addressed by adapting these protocols to enable the introduction of unnatural amino acids at specific positions within the peptide sequence.<sup>[7c, 10b, 11b, 13]</sup> Despite these advances, none of current methodologies allows for the straightforward incorporation of synthetic, nonproteinogenic elements of arbitrary design into genetically encoded peptide frameworks.

To provide a way to combine the advantages of genetic encoding with the versatility of chemical synthesis, our group has

[a] J. R. Frost, Dr. F. Vitali, N. T. Jacob, M. D. Brown, Prof. R. Fasan  
Department of Chemistry, University of Rochester  
Hutchison Hall, Rochester, NY 14620 (USA)  
E-mail: fasan@chem.rochester.edu

 Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/cbic.201200579>.

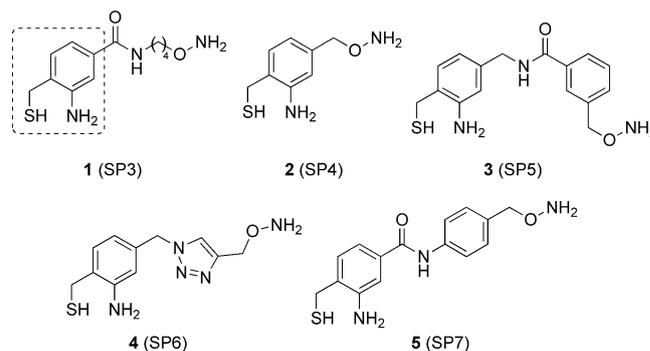
been exploring methodologies for constructing organo-peptide macrocycles (macrocyclic organo-peptide hybrids or MORPHs) through a dual ligation between bifunctional synthetic molecules (referred to as “synthetic precursors” or SPs) and recombinant intein-fused proteins (referred to as “biosynthetic precursors” or BPs) containing an unnatural amino acid.<sup>[14]</sup> In a first implementation of this concept, MORPHs were obtained through a Cu<sup>I</sup>-catalyzed alkyne/azide cycloaddition-/hydrazide-mediated ligation.<sup>[14a]</sup> More recently, we introduced a convenient, catalyst-free approach to MORPH synthesis that exploits a tandem ligation between a oxyamino/amino-thiol compound and an intein-fused protein precursor incorporating a side-chain keto group from a genetically encoded *p*-acetylphenylalanine residue.<sup>[14b]</sup> Interestingly, in this previous study, only SP3 (1, Scheme 1) was found to be a viable precursor for inducing the desired macrocyclization reaction, whereas bifunctional molecules containing a similar alkyloxyamine functionality but a different amino-thiol group (i.e., either cysteine- or 2-aminothiophenol-based moiety) failed to do so. These findings pointed at the importance of the 2-amino-mercaptomethyl-aryl (AMA) moiety in SP3 (dotted box in Scheme 1) as a critical structural element for directing macrocycle formation within this approach.

Building upon these preliminary investigations, we report here a series of systematic studies aimed at expanding and exploring in more depth the scope of this methodology towards the synthesis and diversification of this new class of hybrid organo-peptide macrocycles. A first point of interest concerned the amenability of this methodology to permit the use of structurally diverse AMA-containing synthetic precursor scaffolds as a way to modulate the architecture and topology of the final MORPH products. We then investigated the effect of amino acid substitutions within the precursor peptide sequence on the accessibility and efficiency of macrocyclization, thereby gaining structure–reactivity insights also at the level of the genetically encoded portion of the MORPHs. Finally, mechanistic studies were performed to elucidate the relative contribution of the two distinct pathways envisioned to mediate the formation of the hybrid macrocycles through this chemo-bio-synthetic strategy (Scheme 2).

## Results and Discussion

### Design and synthesis of structurally diverse AMA-based synthetic precursors

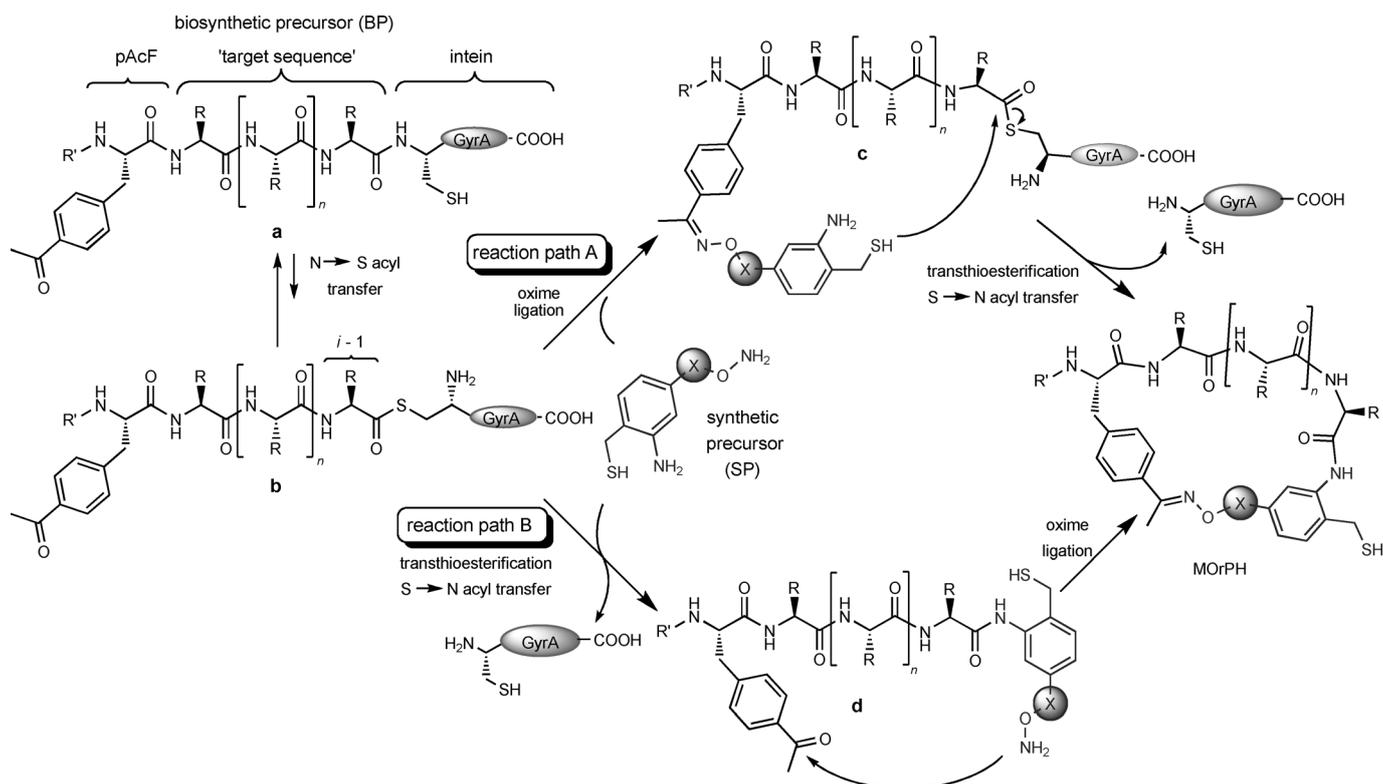
In SP3 (Scheme 1), the oxyamino functionality was connected to the 2-amino-mercaptomethyl-aryl (AMA) moiety by a flexible four-methylene linker. This design was initially useful for comparing the relative efficiency of different amino-thiol moieties for macrocyclization.<sup>[14b]</sup> The ability to use more compact and/or conformationally rigid scaffolds to direct MORPH formation would be desirable though, as these compounds could induce a greater degree of rigidification in the backbone of the resulting macrocycle. However, these features could have a potentially negative effect on the efficiency of the macrocyclization.



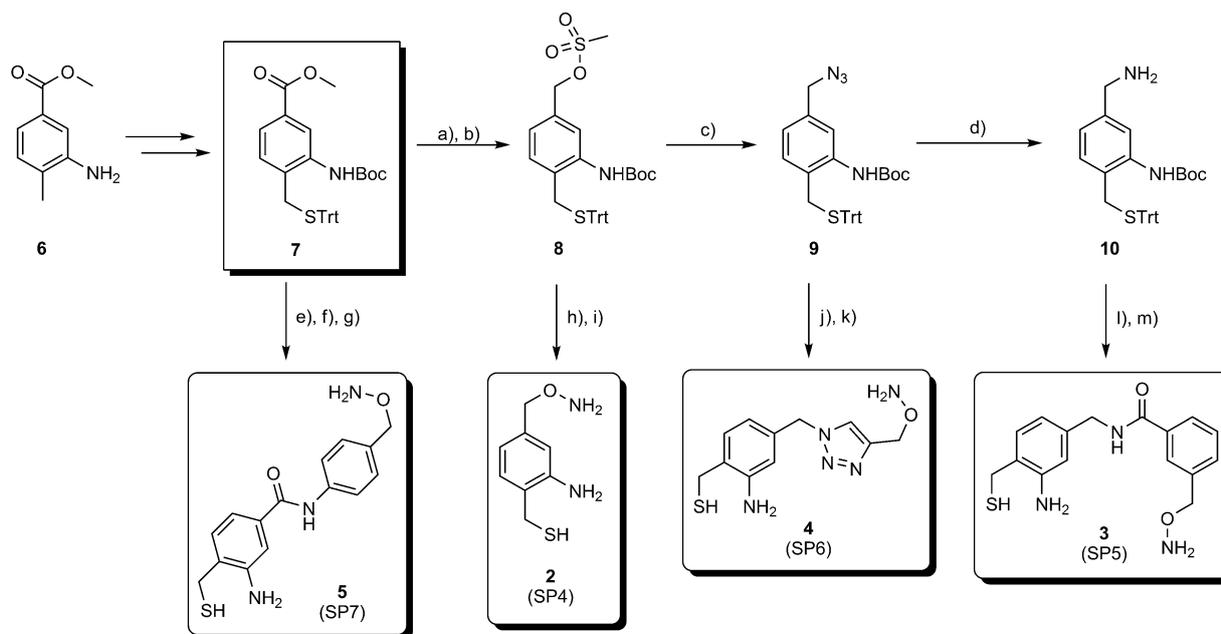
**Scheme 1.** Chemical structures of the oxyamino-/2-amino-mercaptomethyl-aryl (AMA)-based synthetic precursors described in this study. The AMA moiety is highlighted by the dotted box.

To investigate this aspect, a panel of four different AMA-based synthetic precursors were designed and synthesized (compounds 2–5, Scheme 1). SP4 (2) provides the most compact scaffold, with the oxyamine functionality being directly connected to the AMA moiety through a single methylene unit. SP5 (3), SP6 (4), and SP7 (5) present more extended structures incorporating aryl (3, 5) or triazolyl (4) linker groups. Compared to SP3, these second-generation synthetic precursors are expected to possess a lower degree of conformational freedom as given by the presence of two (SP4), four (SP6, SP7), and five (SP5) rotatable bonds separating the AMA moiety from the oxyamine oxygen atom as opposed to six in 1. These compounds were also meant to present a range of different spacing distances between the two ligation points within the molecule (i.e., the oxyamino group and the amino group in the AMA moiety, Scheme 2) as determined based on calculated energy-minimized conformations (6.5 Å (SP4), 8.5 Å (SP5), 11 Å (SP6), 12.5 Å (SP7)). These features would allow us to test the possibility of accessing macrocycles of variable ring size in addition to incorporating different synthetic moieties.

To synthesize 2–5, a different synthetic scheme (Scheme 3) was implemented from that used to prepare SP3,<sup>[14b]</sup> as the latter was found to lack both versatility and scalability. In particular, our goal was to develop an efficient, versatile, and scalable route that could give rapid access to a variety of structurally different AMA-containing synthetic precursors. The new strategy embodies these features and hinges on the preparation of the key intermediate **7**, which can be afforded in high yield (72%) from commercially available methyl 3-amino-4-methylbenzoate (**6**) through a three-step sequence involving N-Boc protection followed by photocatalyzed benzylic bromination with NBS and substitution of the benzylic bromide with tritylmercaptan. Compound **7** can be used to conveniently link the N-Boc-S-trityl-protected AMA moiety to a variety of oxyamine-containing organic scaffolds. Accordingly, basic hydrolysis of the benzoate group in this molecule followed by amide coupling with the aniline derivative **11** and final TFA deprotection yielded SP7 (**5**). To generate SP4 (**2**), **7** was converted into the mesylate derivative **8**, which was then substituted with *N*-Boc-hydroxylamine followed by TFA deprotection. To obtain SP6 (**4**), the mesylate group in **8** was substituted with sodium



**Scheme 2.** Overview of the chemo-biosynthetic method to generate MORPHs through an oxime-/AMA-mediated dual ligation. The precursor protein (BP) is composed of an N-terminal sequence ( $R'$  = chitin binding domain), a genetically encoded *p*-acetyl-phenylalanine (pAcF), a target peptide sequence of variable length ( $n = 2$ – $13$ ), and the GyrA mini-intein from *M. xenopi* (N198A variant). In the oxyamine/amino-thiol synthetic precursor (SP), X corresponds to a variable linker structure (see structures in Scheme 1). The two possible reaction pathways discussed in the text are outlined. In species “b”, the position of the  $i-1$  residue is indicated.



**Scheme 3.** Synthesis of synthetic precursors 2–5. a)  $\text{LiAlH}_4$ , THF, 95%; b) Ms-Cl, DIPEA,  $\text{CH}_2\text{Cl}_2$ , 88%; c)  $\text{NaN}_3$ , DMF, 100%; d)  $\text{LiAlH}_4$ , THF, 95%; e) LiOH, THF/ $\text{H}_2\text{O}$ , 100%; f) 11, DCC, DMAP,  $\text{CH}_2\text{Cl}_2$ , 26%; g) TIPS, TFA/ $\text{CH}_2\text{Cl}_2$ , 100%; h) HONHBoc, DBU, DMF, 89%; i) 12,  $\text{CuSO}_4$ , NaAsc,  $\text{CH}_2\text{Cl}_2$ / $\text{H}_2\text{O}$ , 72%; k) TIPS, TFA/ $\text{CH}_2\text{Cl}_2$ , 100%; l) 13, HBTU, DIPEA,  $\text{CH}_2\text{Cl}_2$ , 55%; m) TIPS, TFA/ $\text{CH}_2\text{Cl}_2$ , 100%.

azide, and the resulting product (**9**) was subjected to Cu<sup>I</sup>-catalyzed alkyne–azide cycloaddition with *N*-Boc-propargyloxamine (**12**) followed by final deprotection with TFA. Finally, the azido-containing intermediate **9** was reduced to the benzylamine counterpart **10**, which was then coupled to the benzoic acid derivative **13** followed by TFA treatment to yield SP5 (**3**).

To test the ability of this scheme to yield gram quantities of a desired synthetic precursor, the synthesis of **2** was repeated on a larger scale; this resulted in the isolation of 1.3 g of this compound in 55% overall yield. Altogether, these studies demonstrated both the versatility and the scalability of the newly developed synthetic scheme for generating oxyamine-/AMA-containing synthetic precursors.

### Macrocyclization reactions with target peptide sequences of variable length

Next, we tested the ability of compounds **2–5** to induce MORPH formation according to Scheme 2. We used a set of seven model biosynthetic precursors comprising, from the N to the C terminus, an affinity tag (chitin binding domain, CBD), the unnatural amino acid *p*-acetylphenylalanine (pAcF), and a variable target peptide sequence (vide infra) fused to *Mxe* GyrA intein (species “a”, Scheme 2). These constructs were produced in *Escherichia coli* by coexpressing an orthogonal, pAcF-specific tRNA<sub>CUA</sub>/aminoacyl-tRNA synthetase pair<sup>[15]</sup> to introduce pAcF into the protein in response to an amber stop codon. Altogether, the seven constructs comprised a set of target sequences consisting of four, five, six, eight, ten, 12, and 15 amino acids (Table 1); these enabled us to test the accessi-

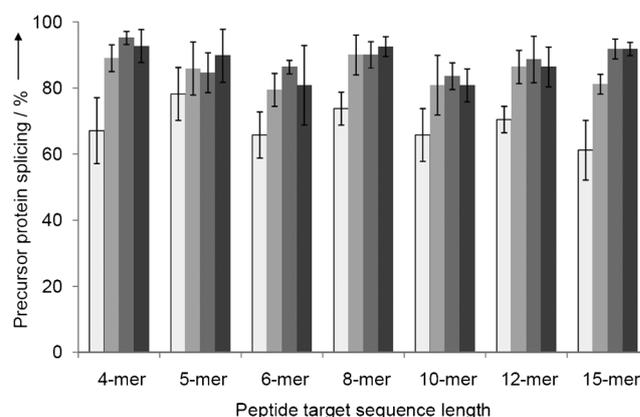
**Table 1.** Length and composition of the target peptide sequences in the biosynthetic precursors used in this study.

Protein	Length	Amino acid sequence
CBD4(pAcF)	tetramer	TGST
CBD5(pAcF)	pentamer	TGSGT
CBD6(pAcF)	hexamer	TGSYGT
CBD8(pAcF)	octamer	TGSAEYGT
CBD10(pAcF)	decamer	TGSKLAEYGT
CBD12(pAcF)	dodecamer	TGSWGKLAIEYGT
CBD15(pAcF)	pentadecamer	TGSHNRWGKLAIEYGT

bility of macrocyclization with the different synthetic precursors across peptide moieties of varying length. With the exception of the last amino acid residue prior to the intein (the “*i*–1 site”, Scheme 2), the composition of the sequences was randomly chosen and included a representative subset of the 20 natural amino acids. The N-terminal chitin binding domain, on the other hand, served the double purpose of reproducing the settings of a typical display system (in which the MORPHs would be generated as tethered to the C terminus of a carrier protein) and that of providing a convenient handle for affinity purification.

Test reactions were carried by mixing SP4–SP7 (15 mM) with each of the biosynthetic precursors (100 μM) in phosphate

buffer (pH 7.5) at room temperature and under reducing conditions (20 mM tris-(2-carboxyethyl)phosphine hydrochloride (TCEP)). These mixtures were then analyzed after 5 and 12 h by SDS-PAGE followed by gel densitometry to quantify the extent of protein splicing; MALDI-TOF mass spectrometry was used to monitor product formation. In the presence of SP4, the protein precursors showed between 61 and 78% splicing after 5 h across all the target sequence lengths (Figure 1, Figure S1 in

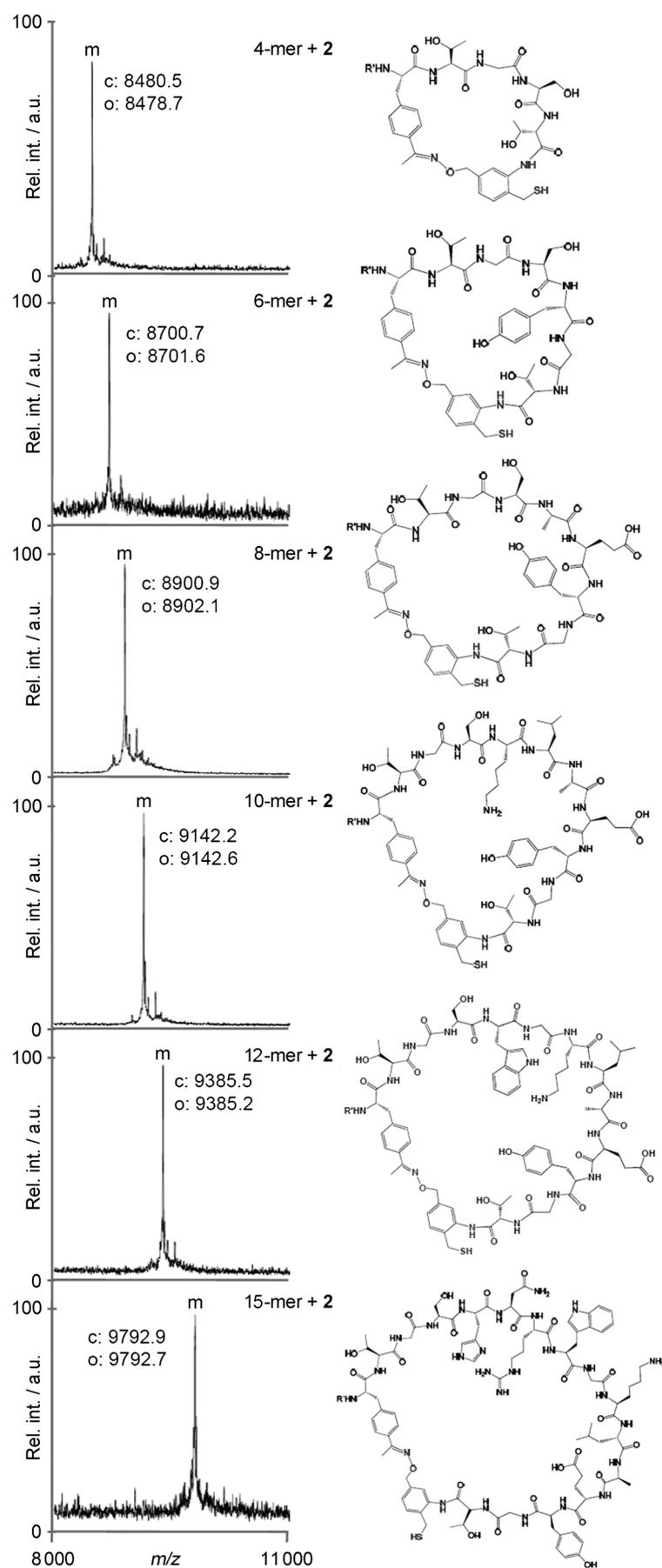


**Figure 1.** Extent of protein splicing for biosynthetic precursors CBD4(pAcF) to CBD15(pAcF) after 5 hours of incubation with compound **2** (□), **3** (▨), **4** (▩), or **5** (■), as determined by densitometric analysis of SDS-PAGE protein gel (see also Figure S1).

the Supporting Information). After 12 h, nearly quantitative splicing (88–96%) was observed for the protein constructs (Figure S1). Similarly, all the other synthetic precursors were found to induce splicing of the biosynthetic precursors at comparable or slightly higher degree than SP4 after 12 hours of incubation. However, significantly more splicing (about 80 to 95%) was observed with these compounds at the shorter time point (5 h; Figure 1). Also in these cases, the extent of precursor protein splicing did not appear to be affected by the length of the target sequence.

In addition to the desired macrocyclization, potential side reactions that would result in splicing of the protein precursors are the spontaneous hydrolysis of the intein-mediated thioester to give the linear product H<sub>2</sub>N-CBD-pAcF-(target sequence)-COOH or incomplete cyclization after SP-induced splicing due to failure of the oxime-mediated ring closure resulting in the linear intermediate “d” in Scheme 2. All these by-products can be distinguished from the desired macrocycles by their different masses. Remarkably, for all the reactions with SP4, MALDI-TOF MS analysis revealed the presence of a single peak consistent with the *m/z* ([*M*+H]<sup>+</sup> ion) of the desired MORPH after both 5 and 12 h (Figure 2). Similarly, the other synthetic precursors SP5, SP6, and SP7 were able to undergo efficient macrocyclizations across all the peptide target sequence lengths tested, as indicated by the observation of a single product with a mass consistent with that of the expected macrocycle (Figures 3 and S2–S4).

To confirm the occurrence of the desired S→N acyl transfer at the level of the AMA moiety in these products, the samples



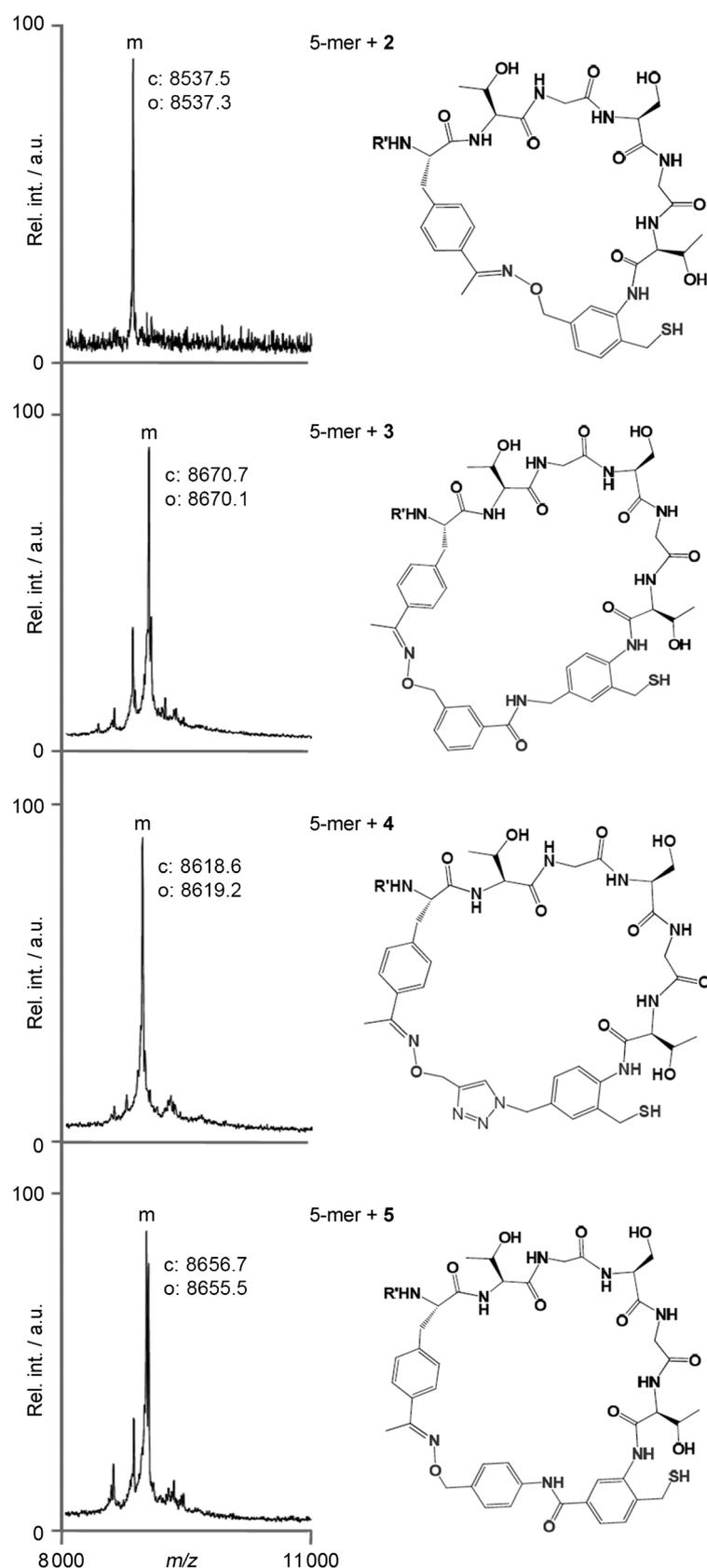
were briefly exposed to iodoacetamide (20 mM, 10 min). In each case, this led to the complete disappearance of the MORPH product peak and the appearance of a new peak with  $m/z$  of +57. This signal is consistent with the expected mass for the S-alkylated MORPH adduct, thereby indicating that the thiol group in the synthetic moiety of the macrocycle is available for reaction with the electrophilic reagent and, thus, that the desired S→N acyl transfer had occurred.

Overall, the results from the above experiments showed that the methodology is remarkably tolerant to variations in both the synthetic precursor structure and the length of the genetically encoded peptide sequence, thus allowing the construction of MORPHs of variable backbone and ring size (Figures 2 and 3).

### Synthesis and stability of MORPHs in cell lysates

In view of future applications of this method for generating and screening MORPH libraries against biological targets, it was important to establish the bio-orthogonality and stability of the macrocycles in a complex biological milieu. Owing to the presence of many metabolites, other biomolecules, and proteases that could prevent the reaction or degrade the product, a most stringent test is to perform the macrocyclization in cell lysate. Accordingly, synthetic precursor 2 was directly added to a clarified lysate solution of *E. coli* cells expressing the pentamer biosynthetic precursor CBD5(pAcF). After 2-hour, 5-hour, and overnight incubation, the reaction products were isolated from the mixture by affinity chromatography on chitin beads. This procedure would allow isolation of the desired CBD-tethered MORPH as well as of by-products resulting from hydrolysis of the intein, incomplete cyclization (C-terminal ligation but no oxime formation), and proteolytic degradation of the peptide moiety embedded in the MORPH. Importantly, MALDI-TOF MS analysis of the chitin bead eluates reproduced the results obtained in the case of the purified biosynthetic precursors, with the desired CBD-tethered MORPH occurring as the only observable species (Figure S5). These data demonstrated the excellent bio-orthogonality of the ligation reactions involved in MORPH formation as well as the suitability of using this approach to produce MORPHs in biological media.

**Figure 2.** MALDI-TOF spectra and chemical structures of MORPHs generated upon treatment of SP4 (2) with biosynthetic precursors CBD4(pAcF) to CBD15(pAcF). Calculated (c) and observed (o)  $m/z$  values corresponding to the  $[M+H]^+$  adduct of the macrocyclic product (m) are indicated. R' = chitin binding domain.

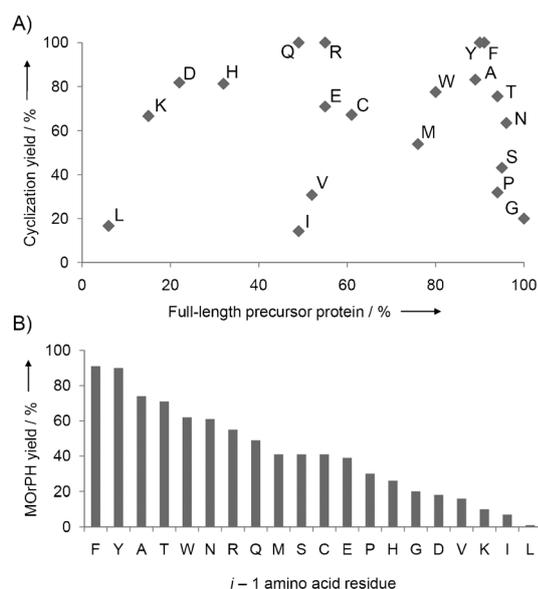


**Figure 3.** MALDI-TOF spectra and chemical structures of MORPHs generated upon treatment of biosynthetic precursor CBD5(pAcF) with SP4 (2), SP5 (3), SP6 (4), and SP7 (5). Calculated (c) and observed (o)  $m/z$  values corresponding to the  $[M+H]^+$  adduct of the macrocyclic product (m) are indicated. R' = chitin binding domain.

### Influence of the “ $i-1$ ” residue on macrocyclization

Next, we sought structure–reactivity insights into the effect of substitutions at the level of the genetically encoded moiety of the macrocycles. The amino acid residue preceding the N-terminal cysteine of an intein (the “ $i-1$ ” position) has been shown to influence the splicing behavior of these proteins.<sup>[16]</sup> Accordingly, this position was expected to play a major role in affecting the reactivity of the biosynthetic precursor toward MORPH formation. In general, the ability to mutate this position with as many residues as possible is desirable in order to increase diversity in the macrocycle libraries generated through this method. At the same time, certain substitutions could decrease the reactivity of the thioester linkage (due to steric hindrance or by disfavoring the intein-catalyzed N→S acyl transfer reaction) or favor premature splicing of the biosynthetic precursor during expression, both of which are undesirable as they would have a negative impact on the accessibility of the precursor polypeptide for the macrocyclization reaction. Knowledge of the most and least favorable mutations at this site would thus be useful during library design in order to maximize the content of the desired products.

In the panel of biosynthetic precursors described earlier (Table 1), a threonine was kept constant at the  $i-1$  site as this residue was reported to cause minimal premature splicing of GyrA fusion proteins.<sup>[16d]</sup> In order to investigate the effect of the  $i-1$  residue in a systematic manner, this position was mutated to each of the other 19 natural amino acids in the pentamer target sequence construct, CBD5-(pAcF). After 12 hours of expression in *E. coli* at 27 °C, the corresponding biosynthetic precursors were purified by Ni-affinity chromatography using a polyhistidine tag fused to the C terminus of the intein. This procedure allows the isolation of both the full-length protein and spliced intein, thus enabling the quantification of the extent of premature splicing by means of SDS-PAGE. These analyses revealed that as many as nine out of the 20 constructs ( $i-1 = T, S, G, P, A, F, W, Y,$  and  $N$ ) showed little or no premature splicing (0–20%; Figure 4A). Seven constructs ( $i-1 = C, V, I, M, E, Q,$  and  $R$ ) exhibited moderate *in vivo* splicing (25–50%), and only four variants ( $i-1 = D, H, K,$  or  $L$ ) were isolated in just 30 to 5% full-length form (Figure 4A). These results confirmed the importance of the  $i-1$  residue in affecting the reactivity and stability of the GyrA intein-catalyzed thioester toward hydrolysis; in most cases, they can be rationalized on the basis of the reactivity of the  $i-1$  residue side-chain functionality. For example, the high level of premature splicing for the constructs with D, H, and K at this site is likely caused by the corresponding side-chain group favoring the



**Figure 4.** Reactivity of the 20 biosynthetic precursor *i*-1 variants. A) Plot of the amount of full-length protein after purification from *E. coli* versus yield of SP4-induced macrocyclization after 5 hours of incubation. Errors are  $\leq 10\%$ . B) Overall yield for the MORPH-forming reactions as given by the % of full-length protein  $\times$  % of SP-induced macrocyclization yield from full-length precursor protein.

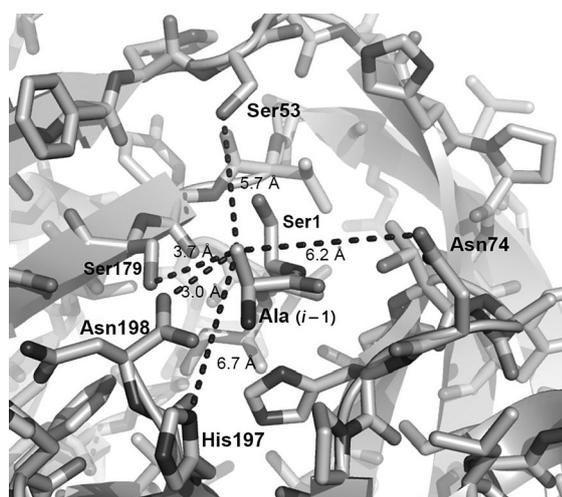
hydrolysis of the neighboring thioester through either direct nucleophilic attack or through a general base-catalyzed mechanism. This notwithstanding, these experiments showed that the majority (80%) of the *i*-1 substitutions resulted in excellent to good amounts of the full-length precursor protein (50–100%). At least for these variants, further increase in the full-length content could be achieved under less stringent expression conditions (i.e., lower temperature and/or shorter culturing time).

To test the reactivity of these 16 variants toward macrocyclization, each construct was incubated with **2** under standard reaction conditions (15 mM SP, 20 mM TCEP, RT). Interestingly, high levels of **2**-induced splicing (75–100%) after 5 h were observed for eight of these constructs (*i*-1 = T, A, F, Y, W, E, Q, and R; Figure 4A, Figure S6). At the same time point, four constructs (*i*-1 = S, C, M, N) exhibited moderate splicing (40–70%), whereas four (*i*-1 = G, V, I, P) show only little reactivity (10–30% splicing; Figure 4A, Figure S6). Importantly, for all of the variants, the desired MORPH was obtained as the only product, as illustrated by the MALDI-TOF MS spectra in Figure S7.

Since the macrocycle was the only observable product of these reactions (Figure S7), the percentage of SP-induced splicing (as determined by densitometric analysis of SDS-PAGE gels) provides a measure of the yield of macrocyclization starting from the full-length protein precursor (Figure 4A). The data above could therefore be combined to calculate overall yields (= % of full-length protein  $\times$  % MORPH yield) to define the most and least optimal amino acid substitutions at position *i*-1 for maximizing the amount of MORPH formed (Figure 4B). These analyses revealed that Phe, Tyr ( $\geq 90\%$  overall yield) and Ala,

Thr (70–75% overall yield) represent the optimal choices at this site, providing an ideal combination of low susceptibility to hydrolysis and high reactivity toward SP-induced cyclization. Following in terms of performance, there is the group consisting of Trp, Asn, Arg, and Gln (50–60% overall yield) and that comprising Met, Ser, Cys, and Glu (40% overall yield). Finally, these analyses revealed the least favorable amino acid substitutions for this position, due to their effect on dampening the reactivity of the precursor protein toward SP-induced cyclization (Gly, Pro), promoting premature splicing (Asp, Lys, His), or both (Leu, Val, Ile). Collectively, these studies indicated that 12 out of the 20 possible amino acids at the *i*-1 site are compatible with MORPH formation, thus leading to the identification of the subsets of most and least favorable residues for this position.

In light of the results, it was also interesting to compare and contrast the efficiency of AMA-mediated ligations in the context of the various *i*-1 variants here (given by the y values in Figure 4A) with that of native chemical ligation (NCL) reactions in the presence of peptide thioesters containing variable terminal amino acids.<sup>[17]</sup> In both cases, aliphatic residues (L, V, I) and proline have the most negative impact on the ligation rates. In contrast, aromatic residues at the terminal site allow for fast AMA-induced ligations but work only moderately well in the case of NCL. However, a most striking difference between the two systems concerns Gly and Thr. In NCL, thioesters with terminal Gly and Thr exhibit, respectively, among the fastest and slowest ligation rates,<sup>[17]</sup> whereas a reversion of this trend is observed with the SP-mediated ligations carried out here. Thus, other factors play a role in modulating the reactivity of the thioester bond in the latter system beyond steric effects, which appear to predominate in the context of NCL reactions. Intein GyrA has been the focus of detailed structural and mechanistic studies,<sup>[18]</sup> and, in order to rationalize these observations, we inspected the available crystal structure of the GyrA(Cys1Ser) variant. Interestingly, the  $\beta$ -carbon of the *i*-1 residue (Ala) was found to lie within 3–7 Å of other residues in the active site and adjacent protein surface (Figure 5).<sup>[18a]</sup> Thus, interactions are likely to occur between these residues that could modulate the reactivity of the neighboring thioester linkage. These interactions could be the basis of the unexpectedly higher performance of threonine compared to the other  $\beta$ -branched amino acids (Ile, Val; Figure 4), the former being the only one possessing a side chain capable of forming H-bonds (e.g., with the proximal Ser179, Figure 5). Another interesting observation concerns the optimal performance of the protein variants containing Tyr and the structurally related Phe at *i*-1, both of which show little reactivity toward hydrolysis but high reactivity toward AMA-mediated ligation. Intriguingly, in the natural protein containing the GyrA intein used here (*Mycobacterium xenopi* DNA gyrase subunit A), tyrosine occurs at the terminal position of the N-extein.<sup>[19]</sup> It is therefore tempting to speculate that this residue might have been selected during evolution because it can provide an optimal performance in the context of protein splicing, a property also exhibited in the context of the present application.

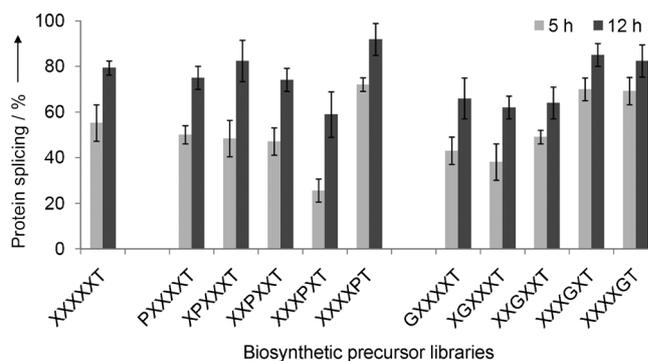


**Figure 5.** View of the GyrA(C15) active site illustrating the proximity of the  $i-1$  amino acid to other residues of the protein (from PDB ID: 1AM2).<sup>[18a]</sup> Distances (....) between the  $\alpha$ -carbon atom of Ala( $i-1$ ) and the side chains of the nearest amino acids are indicated.

### Effect of conformationally constrained/flexible residues within the target peptide sequence

Experiments were then carried out to gain further structure–reactivity insights with respect to the other positions of the peptide target sequence. Among the natural amino acids, proline and glycine possess unique conformational properties because of the restricted rotation around the N–C( $\alpha$ ) bond ( $\phi$  angle) as a result of the cyclic structure in the former and the absence of a side-chain group in the latter. Accordingly, these residues were chosen to probe the effect of altering the conformational flexibility of the genetically encoded peptide moiety on the reactivity of the precursor protein toward SP-induced macrocyclization. A “proline scan” set of biosynthetic precursor libraries was constructed by inserting proline at each of five positions in a hexamer target sequence. A threonine was kept constant at the  $i-1$  position in these constructs to eliminate biases due to the large effect of substitutions at this site on the biosynthetic precursor reactivity, as determined by the studies above. To average the sequence-dependent effect of the remaining positions not occupied by proline, these sites were fully randomized by using a comprehensive degenerate codon (NNK). A similar set of “glycine scan” libraries was constructed. For comparison purposes, a biosynthetic precursor library containing a fully randomized hexamer sequence with Thr at  $i-1$  (i.e., CBD-pAcF- $X_5$ T-GyrA, where X is any amino acid; referred to as “ $X_5$ T”) was also prepared.

About 20 000 recombinants from each of these libraries were pooled together, and the corresponding biosynthetic precursors were expressed and purified as a mixture. As expected with a Thr residue always at  $i-1$ , premature splicing was found to occur at comparably low levels for all the libraries (13% for  $X_5$ T compared to an average of 12 and 5% for the Pro and Gly scan libraries, respectively). Upon treatment with SP4, significant differences in the biosynthetic precursor reactivity were observed for some of the positions as a result of the Pro/Gly



**Figure 6.** Amount of SP4-induced protein splicing for the proline-scan and glycine-scan biosynthetic precursor libraries and the reference  $X_5$ T library. The target sequence, in which X corresponds to a fully randomized position (NNK codon), is specified.

insertion (Figure 6). Consistently, the largest effects for both sets of libraries were observed when positions “ $i-2$ ” and “ $i-3$ ” were substituted with either proline or glycine. At the  $i-2$  position, the insertion of both of these residues improved the efficiency of SP4-induced splicing (69–72%, 5 h) as compared to the reference library  $X_5$ T (55%, 5 h). The “ $i-3$ ” site also appeared to be sensitive to changes in the conformational flexibility of the corresponding residue, but in this case the introduction of proline resulted in reduced biosynthetic precursor reactivity (26% splicing, 5 h), whereas replacement with Gly increased it (70% splicing, 5 h). For all the other positions, the insertion of neither proline nor glycine altered the reactivity of the biosynthetic precursor libraries at a significant level compared to the reference library  $X_5$ T.

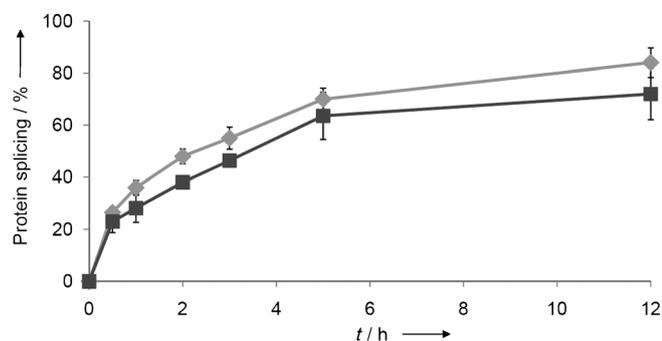
A trend emerging from these analyses is that the positional effect of the Pro/Gly substitution on the target sequence reactivity correlates with its proximity to the intein and thus to the reactive thioester bond. This result supports the critical role of the C-terminal ligation step during MORPH formation, as determined by mechanistic studies (vide infra). Based on the experimental evidence accumulated here and during our earlier investigations with biosynthetic precursors with randomized pentamer and octamer target sequences,<sup>[14b]</sup> it is reasonable to assume that most (> 90%) of the randomized target sequences that undergo C-terminal ligation lead to the desired macrocycle product. Thus, these experiments showed the possibility of accessing MORPH libraries with possibly reduced or increased peptide backbone flexibility through designed Pro/Gly mutations within the genetically encoded peptide moiety of the MORPHs. Proline, in particular, has been exploited to generate conformationally biased peptide libraries.<sup>[6e,g]</sup> The demonstrated compatibility of Pro substitutions with most of the positions of an hexamer peptide moiety suggest that this approach could provide a viable strategy to bias the conformational properties of the MORPHs, complementary to the use of different synthetic precursor scaffolds.

### Mechanism of the MORPH-forming reaction

Finally, we conducted studies to elucidate the mechanism of the macrocyclization orchestrated by the AMA-based synthetic precursors. As illustrated in Scheme 2, two possible reaction pathways to the MORPH product can be envisioned. The first would proceed through the formation of an oxime SP–BP adduct followed by intramolecular AMA-mediated cyclization (“path A”), a process reminiscent of that involved in the formation of MORPHs through CuAAC/hydrazide-mediated ligation.<sup>[14a]</sup> Alternatively, SP-induced intein splicing could occur first, followed by an intramolecular oxime ligation (“path B”).

To investigate this aspect, the products formed from the reaction between SP4 (2) and the pentamer biosynthetic precursor, CBD5(pAcF), were analyzed by MALDI-TOF MS after 30, 60, 120, and 180 min. At each time point, the only observed species corresponded to the macrocycle. Similar results were obtained for SP5 (3) and SP6 (4), whereas in the presence of SP7 (5), a certain amount of the linear acyclic intermediate CBD-(pAcF)-TGSST-5 (species “d” in Scheme 2) was also observed at shorter times (30, 60 min). However, after 180 min, only the 5-containing MORPH was observed. In general, these results are consistent with “path A”, in which a branched SP–protein adduct is formed by oxime ligation, and this is followed by macrocyclization through the SP amino-thiol moiety. At the same time, a direct attack on the intein thioester by the AMA moiety in the synthetic precursor followed by a fast oxime ligation is not excluded.

To discern the relative contributions of paths A and B to macrocyclization, the extent of protein splicing induced by SP4 was monitored over time in parallel reactions with CBD5(pAcF) and the analogous construct CBD5(OpgY), in which pAcF is replaced with *O*-propargyl-tyrosine, which lacks the side-chain ketone required for oxime ligation. This allows measurement of the rate at which the protein is spliced solely by action of the amino-thiol moiety in the synthetic precursor (i.e., the first step of path B). It emerged that the amount of spliced protein in the reaction with CBD5(OpgY) was 90–94% of that observed in the presence of CBD5(pAcF) across the various time points (Figure 7). The same experiments were repeated with SP5–SP7, and trends similar to that observed with SP4 were observed



**Figure 7.** Time course of SP4-induced splicing of precursor protein CBD5 incorporating either pAcF (◆) or the ketone-free amino acid OpgY (■). Error bars are from experiments carried out in duplicate. Results from the same experiment with synthetic precursors SP5–SP7 are provided in Figure S8.

(Figure S8); thus suggesting that all these synthetic precursors share a common mechanism. Based on these results, we concluded that, whereas both reaction pathways can participate in the MORPH-forming process, the contribution of path B predominates, accounting for  $\geq 90\%$  of the macrocyclic product formed over time.

The observation that the bimolecular ligation reaction at the C-terminal end of the target sequence is considerably faster than that at the side-chain level is reasonable given the slow kinetics of ketoxime formation at neutral pH.<sup>[20]</sup> The derived mechanism is also consistent with the larger effect of Pro/Gly substitutions close to the C terminus of the target sequence on the precursor protein reactivity toward SP4 (Figure 6) as well as with the low dependence of the rate of SP-induced protein splicing on the target sequence length for constructs with an identical  $i-1$  residue (Figure 1). Indeed, a larger dependence on this parameter would be expected if the reaction proceeded largely through path A, as observed for our MORPH synthesis through CuAAC/hydrazide-mediated ligations, which relies exclusively on a side-chain→C-end tandem ligation mechanism.<sup>[14a]</sup> Furthermore, as no acyclic intermediate “d” is observed even at the shortest time points, we deduce that the oxime ligation must be greatly accelerated when proceeding intramolecularly rather than intermolecularly. This rate acceleration, combined with the fact that both pathways remain available for macrocyclization, can thus account for the remarkably clean outcome of these reactions.

In light of these mechanistic insights, it is also possible to explain the marked difference in MORPH-forming ability between the AMA-based SPs reported here and the cysteine- or 2-amino-thiophenol-based synthetic precursors described earlier.<sup>[14b]</sup> Because path A appears to be a minor contributor in the MORPH-forming reaction, fast and efficient ligation at the C-terminal end of the target sequence becomes crucial, thereby highlighting the critical role of the AMA moiety and its superior intein-splicing properties for the success of the macrocyclization.

### Conclusions

In summary, we have developed an efficient, versatile, and scalable synthetic route to afford structurally different bifunctional precursors for the synthesis of MORPHs through a chemo-biosynthetic strategy based on a tandem oxime-/AMA-mediated ligation. With these compounds, we demonstrated that a variety of nonproteogenic structures can be efficiently incorporated into the final macrocyclic products. This feature is of particular interest given the range of different scaffolds found to be amenable to this process, and thus the important changes they are expected to impose on the overall topology and conformational properties of the resulting MORPHs, especially in the context of the shorter target sequences (i.e., tetra- to octamer). Overall, the high yields and remarkably clean outcomes of the MORPH-forming reactions across this panel of compounds and target sequence lengths of four to 15 amino acids demonstrate the high degree of versatility and modularity of the methodology for the preparation and diversification of this new class of

organo-peptide macrocycles. Furthermore, the demonstrated possibility to carry out these cyclization reactions in complex biological mixtures (i.e., cell lysate) adds to the value of this strategy for molecular discovery campaigns.

Our structure–reactivity analyses at the level of the genetically encoded moiety of these compounds showed the critical role of the “*i*–1” residue in influencing the reactivity and accessibility of the precursor proteins toward macrocyclization. Importantly, a considerable fraction (60%) of the 20 possible amino acid substitutions at this site was found to be compatible with MORPH formation. In addition, specific residues (F, Y, T, A) that can maximize the yield of the macrocyclization reaction were identified. These structure–reactivity data, together with those gathered in the context of the Pro/Gly scanning experiments, will be valuable to guide the design and construction of high-quality and, optionally, conformationally biased MORPH libraries for screening purposes.

Finally, we have shed light on the mechanism of the macrocyclization and defined the relative contribution of the two reaction pathways participating in the MORPH-forming process. These studies unveiled that the mechanism involving C-end → side-chain ligation predominates, thus highlighting the critical role of the AMA moiety in promoting MORPH formation in this method. These findings explained the structure–reactivity trends observed here as well as previous observations on other amino-thiol moieties.<sup>[14b]</sup> They also have a number of other important implications. For example, they suggest that even faster macrocyclization rates could be achieved by increasing the rate of the ligation reaction at the side-chain level. Furthermore, given the proven efficiency of the AMA group in mediating efficient ligations at the level of polypeptide thioesters under catalyst-free conditions, we envision that this moiety could prove useful in a number of other relevant applications beyond MORPH synthesis, such as protein conjugation and auxiliary-assisted peptide/protein ligation.

## Experimental Section

**Reagents and analytical methods:** Chemical reagents and solvents were purchased from Sigma–Aldrich, Acros Organics, and Fluka. Silica gel chromatography purifications were carried out by using AMD Silica Gel 60 230–400 mesh. <sup>1</sup>H and <sup>13</sup>C NMR spectra were recorded on Bruker Avance spectrometers by using solvent peaks as reference. LC–MS analyses were performed on a Thermo Scientific LTQ Velos ESI/ion-trap mass spectrometer coupled to an Accela U–HPLC. MALDI–TOF spectra were acquired on a Bruker Autoflex III MALDI–TOF spectrometer by using a stainless steel MALDI plate and sinapinic acid as matrix.

**Synthesis of *N*-Boc-*S*-trityl-3-amino-4-(mercaptomethyl)benzoate methyl ester (7):** Methyl 3-amino-4-methylbenzoate (9.7 g, 58.7 mmol) and di-*tert*-butyl dicarbonate (17 mL, 74 mmol, 1.2 equiv) were dissolved in dry THF (200 mL), and the mixture was heated under reflux for 72 h. The solvent was removed by rotary evaporation to afford a pink–white solid that was suspended in ice-cold hexanes (30 mL) and filtered to afford methyl 3-((*tert*-butoxycarbonyl)amino)-4-methylbenzoate as a white solid, (99% yield). <sup>1</sup>H NMR (CDCl<sub>3</sub>, 500 MHz): δ = 8.45 (s, 1 H), 7.69 (d, *J* = 7.9 Hz, 1 H), 7.21 (d, *J* = 7.9 Hz, 1 H), 6.29 (s, 1 H), 3.90 (s, 3 H), 2.30 (s, 3 H),

1.55 ppm (d, *J* = 11.2 Hz, 9 H); <sup>13</sup>C NMR (CDCl<sub>3</sub>, 126 MHz): δ = 166.9, 152.8, 136.4, 132.6, 130.3, 128.9, 124.9, 121.8, 80.8, 52.0, 28.3, 17.9 ppm.

This compound (6.63 g, 25 mmol) was dissolved in carbon tetrachloride (100 mL), and *N*-bromosuccinamide (4.89 g, 27.5 mmol, 1.1 equiv) was added. The reaction vessel was equipped with a reflux condenser and irradiated with UV light for 3 h, then cooled to room temperature; the mixture was filtered. The filtrate was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (100 mL), washed with saturated K<sub>2</sub>CO<sub>3</sub> (aq) and brine, then dried over anhydrous MgSO<sub>4</sub>. Volatiles were removed to afford methyl 4-(bromomethyl)-3-((*tert*-butoxycarbonyl)amino)benzoate (6.7 g, 78% yield) as an orange–white solid. <sup>1</sup>H NMR (CDCl<sub>3</sub>, 500 MHz): δ = 8.47 (s, 1 H), 7.73 (dd, *J* = 8.0, 1.7 Hz, 1 H), 7.36 (d, *J* = 8.0 Hz, 1 H), 6.75 (s, 1 H), 4.50 (s, 2 H), 3.91 (s, 3 H), 1.55 ppm (s, 9 H); <sup>13</sup>C NMR (CDCl<sub>3</sub>, 126 MHz): δ = 28.2, 29.9, 52.3, 81.3, 123.8, 125.1, 130.0, 131.5, 131.7, 136.9, 152.6, 166.2 ppm.

Methyl 4-(bromomethyl)-3-((*tert*-butoxycarbonyl)amino)benzoate (6.7 g, 19.59 mmol), tritylmercaptan (6.5 g, 23.5 mmol, 1.2 equiv) and potassium carbonate (3.25 g, 23.5 mmol, 1.2 equiv) were dissolved in dry DMF (100 mL), and the mixture was stirred under argon at room temperature for 15 h, concentrated by rotary evaporation, then dissolved in CH<sub>2</sub>Cl<sub>2</sub>. The solution was washed with ice-cold H<sub>2</sub>O, with saturated NaHCO<sub>3</sub> (aq), and finally with brine. The organic layer was dried over anhydrous MgSO<sub>4</sub> and filtered, and volatiles were removed to afford methyl 3-((*tert*-butoxycarbonyl)amino)-4-((tritylthio)methyl)benzoate (7) as a golden yellow solid (10.24 g, 97% crude yield). <sup>1</sup>H NMR (CDCl<sub>3</sub>, 500 MHz): δ = 8.41 (s, 1 H), 7.65 (d, *J* = 7.9 Hz, 1 H), 7.48 (d, *J* = 8.0 Hz, 5 H), 7.34 (t, *J* = 7.8 Hz, 6 H), 7.25 (t, *J* = 7.3 Hz, 5 H), 7.18 (d, *J* = 8.0 Hz, 1 H), 6.72 (s, 1 H), 3.88 (s, 3 H), 3.21 (s, 2 H), 1.56 ppm (d, *J* = 2.5 Hz, 9 H); <sup>13</sup>C NMR (CDCl<sub>3</sub>, 126 MHz): δ = 166.7, 152.8, 144.1, 136.9, 130.7, 129.3, 128.2, 126.9, 124.9, 123.1, 80.8, 67.4, 52.1, 34.1, 28.4 ppm. This material was carried forward without further purification.

**Synthesis of *N*-Boc-*S*-trityl-3-amino-4-(mercaptomethyl)benzyl mesylate (8):** Compound 7 (20.32 g, 48 mmol) was dissolved in anhydrous THF (400 mL), then the solution was cooled to 0 °C. A solution of LiAlH<sub>4</sub> in THF (1 M, 52.8 mL, 52.8 mmol, 1.1 equiv) was added slowly. The reaction mixture was stirred under argon at 0 °C for 3 h, the reaction was quenched by slow addition of cold H<sub>2</sub>O (3 mL) and NaOH (4 N, 1 mL) at 0 °C, then the mixture was stirred for 10 min at room temperature. The mixture was concentrated under reduced pressure, suspended in EtOAc/sat. NaHCO<sub>3</sub> (10:1, 330 mL), then filtered through celite. The filtrate was washed once with saturated NaHCO<sub>3</sub>, then with brine. The organic layer was dried with anhydrous MgSO<sub>4</sub>, and volatiles were removed to afford a yellow solid, which was purified by flash column chromatography (silica gel, hexanes/EtOAc 7:3) to afford a yellow oil (18 g, 95% yield). <sup>1</sup>H NMR (CDCl<sub>3</sub>, 500 MHz): δ = 7.78 (s, 1 H), 7.49 (d, *J* = 7.3 Hz, 5 H), 7.34 (t, *J* = 7.7 Hz, 5 H), 7.26 (t, *J* = 3.0 Hz, 5 H), 7.13 (d, *J* = 7.8 Hz, 1 H), 7.01 (d, *J* = 7.8 Hz, 1 H), 6.73 (s, 1 H), 4.63 (s, 2 H), 3.17 (s, 2 H), 1.54 ppm (s, 9 H); <sup>13</sup>C NMR (CDCl<sub>3</sub>, 126 MHz): δ = 153.1, 144.3, 141.5, 136.8, 130.9, 129.3, 128.2, 126.9, 124.5, 122.2, 120.4, 80.5, 67.2, 65.1, 33.9, 28.4 ppm.

This compound (9.3 g, 18.19 mmol) was dissolved in anhydrous CH<sub>2</sub>Cl<sub>2</sub> (100 mL), and the solution was cooled to 0 °C. Methanesulfonylchloride (1.8 mL, 23.66 mmol, 1.3 equiv) and *N,N*-diisopropylethylamine (DIPEA; 4.2 mL, 23.66 mmol, 1.3 equiv) were added, and the reaction mixture was stirred under argon at 0 °C for 2 h. The mixture was then dissolved in CH<sub>2</sub>Cl<sub>2</sub>, washed twice with saturated NaHCO<sub>3</sub> (aq), and then once with brine. The organic layer was dried over anhydrous MgSO<sub>4</sub>, and volatiles were removed to

afford **8** as a yellow solid (9.42 g, 88% yield).  $^1\text{H NMR}$  ( $\text{CDCl}_3$ , 500 MHz):  $\delta = 7.88$  (s, 1H), 7.49 (d,  $J = 7.3$  Hz, 5H), 7.34 (t,  $J = 7.7$  Hz, 5H), 7.26 (d,  $J = 14.6$  Hz, 5H), 7.16 (d,  $J = 7.8$  Hz, 1H), 7.04 (d,  $J = 9.5$  Hz, 1H), 6.75 (s, 1H), 5.18 (s, 2H), 3.17 (s, 2H), 2.90 (s, 3H), 1.54 ppm (s, 9H);  $^{13}\text{C NMR}$  ( $\text{CDCl}_3$ , 126 MHz):  $\delta = 152.8, 144.1, 137.3, 133.7, 131.3, 129.3, 128.2, 126.9, 126.3, 123.8, 121.9, 80.8, 71.3, 67.3, 38.4, 33.9, 28.4$  ppm. The material was carried forward without further purification.

**Synthesis of SP4 (2):** Compound **8** (1.06 g, 1.8 mmol) was dissolved in dry  $\text{CH}_3\text{CN}$  (18 mL) and cooled to  $0^\circ\text{C}$ . *tert*-Butyl-*N*-hydroxycarbamate (0.32 g, 2.4 mmol, 1.3 equiv) then 1,8-diazabicyclo[7.2.1]undec-7-ene (DBU; 0.37 mL, 2.4 mmol, 1.3 equiv) were slowly added. The reaction was stirred at  $0^\circ\text{C}$  for 1 h, then warmed to ambient temperature, and stirred under argon overnight. Volatiles were removed and the resulting crude mixture was dissolved in  $\text{CH}_2\text{Cl}_2$ , washed with saturated  $\text{K}_2\text{CO}_3$  (aq), and then with brine. The organic layer was dried over anhydrous  $\text{MgSO}_4$  then concentrated to afford a yellow oil, which was purified through flash chromatography (silica gel, hexanes/EtOAc 7:3) to afford a yellow oil (1.00 g, 89% yield).  $^1\text{H NMR}$  ( $\text{CDCl}_3$ , 400 MHz):  $\delta = 7.81$  (s, 1H), 7.49 (d,  $J = 7.6$  Hz, 6H), 7.32 (q,  $J = 7.6$  Hz, 6H), 7.24 (t,  $J = 7.2$  Hz, 3H), 7.13 (d,  $J = 7.6$  Hz, 2H), 7.02 (dd,  $J = 8, 1.6$  Hz, 1H), 6.74 (s, 1H), 4.79 (s, 2H), 3.17 (s, 1H), 1.54 (s, 9H), 1.46 ppm (s, 4H);  $^{13}\text{C NMR}$  ( $\text{CDCl}_3$ , 126 MHz):  $\delta = 156.6, 152.9, 144.2, 136.9, 136.3, 130.8, 129.3, 128.1, 126.9, 125.4, 124.2, 122.3, 81.6, 80.4, 77.9, 67.2, 33.9, 28.4, 27.5$  ppm; MS-ESI: calcd for  $\text{C}_{37}\text{H}_{42}\text{N}_2\text{O}_5\text{S}$ : 649.79  $[M+\text{Na}]^+$ ; found: 649.33.

The protected precursor (0.551 g, 0.88 mmol) was dissolved in anhydrous  $\text{CH}_2\text{Cl}_2$  (9 mL), and the solution was cooled to  $0^\circ\text{C}$ . Triisopropylsilane (TIPS; 0.45 mL, 2.2 mmol) was added followed by trifluoroacetic acid (TFA; 2 mL). The reaction mixture was stirred under argon at  $0^\circ\text{C}$  for 30 min, then warmed to ambient temperature and concentrated under reduced pressure to afford a solid; this was washed with ice-cold hexanes to afford **2** as an off-white solid (0.366 g, quantitative yield).  $^1\text{H NMR}$  ( $\text{CD}_3\text{OD}$ , 500 MHz):  $\delta = 7.06$  (d,  $J = 8$  Hz, 1H), 6.76 (d,  $J = 1.5$  Hz, 1H), 6.65 (dd,  $J = 8, 1.5$  Hz, 1H), 4.56 (s, 2H), 3.69 (s, 2H), 1.38 ppm (s, 1H);  $^{13}\text{C NMR}$  ( $\text{CD}_3\text{OD}$ , 126 MHz):  $\delta = 146.5, 136.6, 130.4, 126.6, 118.9, 117.3, 78.9, 25.8$  ppm; MS-ESI: calcd for disulfide  $\text{C}_{16}\text{H}_{22}\text{N}_4\text{O}_2\text{S}_2$ : 367.51  $[M+\text{H}]^+$ ; found: 367.53.

**Synthesis of *N*-Boc propargyloxamine (12):** Propargyl bromide (80% by weight in toluene; 1.6 g, 13.44 mmol) was dissolved in dry MeCN (40 mL), and the mixture was cooled to  $0^\circ\text{C}$ . *tert*-Butyl-*N*-hydroxycarbamate (2.32 g, 17.47 mmol, 1.3 equiv) and DBU (2.61 mL, 17.47 mmol, 1.3 equiv) were added. The reaction mixture was stirred for 20 min at  $0^\circ\text{C}$ , then warmed to ambient temperature, and stirred for another 1 h. Volatiles were removed under reduced pressure, and the resulting yellow oil was suspended in  $\text{CH}_2\text{Cl}_2$ , washed twice with saturated  $\text{NaHCO}_3$  (aq) and once with brine, then dried over anhydrous  $\text{MgSO}_4$ . Volatiles were removed under reduced pressure, and the resulting crude material was purified on silica gel (hexanes/EtOAc 8:1 $\rightarrow$ 7:3) to give **12** (1.5 g, 65% yield).  $^1\text{H NMR}$  ( $\text{CDCl}_3$ , 400 MHz):  $\delta = 7.39$  (s, 1H), 4.48 (d,  $J = 2$  Hz, 2H), 2.5 (s, 1H), 1.49 ppm (s, 1H);  $^{13}\text{C NMR}$  ( $\text{CDCl}_3$ , 126 MHz):  $\delta = 156.5, 82.1, 78.3, 75.6, 63.7, 28.2$  ppm.

**Synthesis of *N*-Boc-*S*-trityl-3-amino-4-(mercaptomethyl)benzyl azide (9):** Compound **8** (2.5 g, 4.24 mmol) and sodium azide (0.56 g, 8.6 mmol) were dissolved in anhydrous DMF (30 mL), and the mixture was stirred under argon at ambient temperature for 12 h. The reaction mixture was then dissolved in  $\text{CH}_2\text{Cl}_2$  (150 mL) and washed with saturated  $\text{NaHCO}_3$  (aq) and with brine. The or-

ganic layer was dried over anhydrous  $\text{MgSO}_4$ , filtered, and concentrated under reduced pressure to afford a yellow oil, which was purified on silica gel with hexanes/EtOAc (1:1) as eluent to afford **9** as a yellow oil (2.3 g, quant.).  $^1\text{H NMR}$  ( $\text{CDCl}_3$ , 400 MHz):  $\delta = 7.80$  (s, 1H), 7.50 (t,  $J = 4.38$  Hz, 6H), 7.34 (t,  $J = 7.64$  Hz, 6H), 7.25 (t,  $J = 7.28$  Hz, 3H), 7.14 (d,  $J = 7.80$  Hz, 1H), 6.94 (dd,  $J = 7.98, 1.70$  Hz, 1H), 6.76 (s, 1H), 4.28 (s, 2H), 3.17 (s, 2H), 1.55 ppm (s, 9H);  $^{13}\text{C NMR}$  ( $\text{CDCl}_3$ , 126 MHz):  $\delta = 152.9, 144.2, 137.2, 135.8, 131.2, 129.3, 128.2, 126.9, 123.2, 121.3, 80.6, 67.2, 54.4, 33.9, 28.4$  ppm; MS-ESI: calcd for  $\text{C}_{32}\text{H}_{32}\text{N}_4\text{O}_2\text{S}$ : 559.68  $[M+\text{Na}]^+$ ; found: 559.22.

**Synthesis of SP6 (4):** Compounds **9** (0.1 g, 0.186 mmol) and **12** (0.127 g, 0.745 mmol, 4 equiv) were dissolved in THF/ $\text{H}_2\text{O}$  (1:1, 6 mL).  $\text{CuSO}_4$  (0.045 g, 0.28 mmol, 1.5 equiv) and sodium ascorbate (0.147 g, 0.745 mmol, 4 equiv) were added, and the reaction mixture was stirred at room temperature for 30 min, then dissolved in  $\text{CH}_2\text{Cl}_2$  and washed twice with concentrated ammonium hydroxide, once with saturated  $\text{NaHCO}_3$  (aq), and once with brine, then dried over anhydrous  $\text{MgSO}_4$ . Volatiles were removed under reduced pressure, and the resulting material was purified on silica gel (hexanes/EtOAc 7:3) to yield protected **4** (0.094 g, 72% yield).  $^1\text{H NMR}$  ( $\text{CDCl}_3$ , 400 MHz):  $\delta = 7.77$  (brs, 1H), 7.54 (s, 1H), 7.47 (d,  $J = 4$  Hz, 6H), 7.38 (s, 1H), 7.33 (t,  $J = 8$  Hz, 6H), 7.26–7.23 (m, 3H), 7.11 (d,  $J = 8$  Hz, 1H), 6.68–6.83 (m, 1H), 6.76 (s, 1H), 5.47 (s, 2H), 4.96 (s, 2H), 3.15 (s, 2H), 1.53 (s, 9H), 1.45 ppm (s, 1H); MS-ESI: calcd for  $\text{C}_{40}\text{H}_{45}\text{N}_5\text{O}_5\text{S}$ : 730.87  $[M+\text{Na}]^+$ ; found: 730.26.

The protected precursor (0.094 g, 0.133 mmol) was deprotected with TFA in  $\text{CH}_2\text{Cl}_2$ , as described above for **2**, to afford **4** (0.065 g, quant.).  $^1\text{H NMR}$  ( $\text{CD}_3\text{OD}$ , 500 MHz):  $\delta = 8.00$  (s, 1H), 7.08 (d,  $J = 8$  Hz, 1H), 6.71 (d,  $J = 1.5$  Hz, 1H), 6.63 (dd,  $J = 8, 1.5$  Hz, 1H), 5.468 (s, 2H), 4.933 (s, 2H), 3.671 ppm (s, 2H);  $^{13}\text{C NMR}$  ( $\text{CD}_3\text{OD}$ , 126 MHz):  $\delta = 146.8, 143.4, 136.4, 131.0, 127.3, 126.0, 118.9, 116.8, 68.9, 54.9, 25.6$  ppm; MS-ESI: calcd for disulfide  $\text{C}_{22}\text{H}_{28}\text{N}_{10}\text{O}_2\text{S}_2$ : 529.66  $[M+\text{H}]^+$ ; found: 529.18.

**Synthesis of *N*-Boc-*S*-trityl-3-amino-4-(mercaptomethyl)benzyl amine (10):** Compound **9** (2.3 g, 4.29 mmol) was dissolved in dry THF (30 mL), and the mixture was cooled to  $0^\circ\text{C}$ . A solution of  $\text{LiAlH}_4$  in THF (1 M, 5.16 mL, 5.16 mmol) was added slowly. The reaction mixture was stirred at  $0^\circ\text{C}$  under argon for 2 h, the reaction was quenched by the slow addition of cold  $\text{H}_2\text{O}$  (3 mL) and  $\text{NaOH}$  (4 N, 1 mL) at  $0^\circ\text{C}$ , then the mixture was stirred for 10 min at room temperature. Volatiles were removed under reduced pressure, and the resulting material was dissolved in a mixture of ethyl acetate/sat.  $\text{NaHCO}_3$  (100:15 mL). The solution was filtered through celite. The filtrate was concentrated under reduced pressure to afford a light brown solid that was purified by flash chromatography (silica gel, hexanes/EtOAc 8:2 $\rightarrow$ 7:3) to afford **10** as a white solid (2 g, 95% yield).  $^1\text{H NMR}$  ( $\text{CDCl}_3$ , 400 MHz):  $\delta = 7.74$  (s, 1H), 7.49 (d,  $J = 7.70$  Hz, 6H), 7.34 (t,  $J = 7.71$  Hz, 6H), 7.26–7.25 (m, 4H), 7.10 (d,  $J = 8.70$  Hz, 1H), 6.39 (d,  $J = 7.63$  Hz, 1H), 6.73 (s, 1H), 3.80 (s, 2H), 3.16 (s, 2H), 1.54 ppm (s, 9H);  $^{13}\text{C NMR}$  ( $\text{CDCl}_3$ , 126 MHz):  $\delta = 153.1, 144.3, 143.9, 136.9, 130.9, 129.4, 128.2, 126.2, 122.4, 120.5, 80.4, 67.1, 46.3, 33.9, 28.43$  ppm. MS-ESI: calcd for  $\text{C}_{32}\text{H}_{34}\text{N}_2\text{O}_2\text{S}$ : 1021.47  $[2M+\text{H}]^+$ ; found: 1021.51.

**Synthesis of *tert*-butyl 3-(carboxybenzyl)oxycarbamate (13):** 3-Bromobenzyl benzoate (0.5 g, 2.18 mmol) was dissolved in dry MeCN (12 mL), and the mixture was cooled to  $0^\circ\text{C}$ . *tert*-Butyl *N*-hydroxy carbamate (0.39 g, 2.9 mmol), then DBU (0.45 mL) were slowly added. The reaction mixture was stirred at  $0^\circ\text{C}$  for 2 h and was then warmed to ambient temperature and left under argon overnight. Volatiles were removed under reduced pressure, and the resulting material was dissolved in  $\text{CH}_2\text{Cl}_2$ . This solution was

washed once with saturated  $K_2CO_3$  (aq) and once with brine, then dried over anhydrous  $MgSO_4$  and concentrated under reduced pressure to afford a white solid (0.392 g, 64% yield).  $^1H$  NMR ( $CDCl_3$ , 400 MHz):  $\delta$  = 8.06 (s, 1H), 8.02 (d,  $J$  = 7.75 Hz, 1H), 7.61 (d,  $J$  = 7.63 Hz, 1H), 7.45 (t,  $J$  = 7.68 Hz, 1H), 7.14 (s, 1H), 4.90 (s, 2H), 3.93 (s, 3H), 1.48 ppm (s, 9H).

This material (0.173 g, 0.615 mmol) was dissolved in THF (5 mL), and LiOH (1 N, 1.24 mL) was added. The reaction mixture was stirred under argon at ambient temperature overnight, then concentrated under reduced pressure, dissolved in ethyl acetate, and washed once with HCl (aq) (0.25 M) then once with brine. The organic layer was dried over anhydrous  $MgSO_4$ , filtered, then concentrated under reduced pressure to afford **13** as a white solid (0.054 g, 79% yield). MS-ESI: calcd for  $C_{13}H_{17}NO_5$ : 290.27  $[M+Na]^+$ ; found: 290.23.

**Synthesis of SP5 (3):** Compounds **10** (0.263 g, 0.98 mmol) and **13** (0.487 g, 0.98 mmol) were dissolved in dry  $CH_2Cl_2$  (8 mL). 2-(1*H*-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU; 0.558 g, 1.47 mmol, 1.5 equiv) and DIPEA (0.43 mL, 2.45 mmol, 2.5 equiv) were added. The reaction mixture was stirred at ambient temperature under argon for 3 h, then dissolved in  $CH_2Cl_2$ , and washed with water and brine. The organic layer was dried over anhydrous  $MgSO_4$  and concentrated under reduced pressure to afford a yellow oil, which was purified by flash chromatography (silica gel, hexanes/EtOAc 8:2) to afford a protected **3** as a yellow oil (0.4 g, 55% yield).  $^1H$  NMR ( $CDCl_3$ , 400 MHz):  $\delta$  = 7.80 (d,  $J$  = 7.4 Hz, 2H), 7.64 (d,  $J$  = 7.8 Hz, 1H), 7.5 (s, 1H), 7.48 (d,  $J$  = 7.5 Hz, 6H), 7.41 (t,  $J$  = 7.65 Hz, 1H), 7.33 (t,  $J$  = 7.5 Hz, 6H), 7.25 (d,  $J$  = 5.7 Hz, 3H), 7.23 (s, 1H), 7.19 (brs, 1H), 7.11 (d,  $J$  = 7.8 Hz, 1H), 7.01 (d,  $J$  = 7.8 Hz, 1H), 6.73 (s, 1H), 4.87 (s, 2H), 4.57 (d,  $J$  = 5.5, 2H), 3.16 (s, 2H), 1.53 (s, 9H), 1.44 ppm (s, 9H);  $^{13}C$  NMR ( $CDCl_3$ , 126 MHz):  $\delta$  = 166.9, 156.7, 153.0, 144.2, 138.5, 136.9, 136.4, 134.7, 131.8, 131.2, 129.3, 128.7, 128.2, 127.3, 127.1, 126.9, 123.5, 121.5, 81.9, 80.6, 77.8, 67.2, 43.9, 33.9, 28.4, 28.1 ppm; MS-ESI: calcd for  $C_{45}H_{49}N_3O_6S$ : 782.33  $[M+Na]^+$ ; found: 782.35.

The protected precursor (0.186 g, 0.244 mmol) was deprotected with TFA in  $CH_2Cl_2$ , as described for **2**, to afford **3** as a white solid (0.136 g, quant.).  $^1H$  NMR ( $CD_3OD$ , 400 MHz):  $\delta$  = 7.83 (s, 1H), 7.78 (d,  $J$  = 7.5 Hz, 1H), 7.51 (d,  $J$  = 7.5 Hz, 1H), 7.44 (t,  $J$  = 8 Hz, 1H), 6.96 (d,  $J$  = 8 Hz, 1H), 6.74 (s, 1H), 6.65 (dd,  $J$  = 8, 1.5 Hz, 1H), 5.50 (s, 1H), 4.70 (s, 2H), 4.48 (s, 2H), 3.64 ppm (s, 2H);  $^{13}C$  NMR ( $CD_3OD$ , 126 MHz):  $\delta$  = 169.9, 147.2, 140.9, 139.5, 135.9, 132.8, 132.4, 129.7, 128.1, 127.9, 121.2, 118.2, 116.3, 78.2, 44.4, 40.7 ppm; MS-ESI: calcd for disulfide  $C_{32}H_{36}N_6O_4S_2$ : 633.80  $[M+H]^+$ ; found: 633.60.

**Synthesis of tert-butyl-(4-aminobenzyl)oxycarbamate (11):** 1-(Bromomethyl)-4-nitrobenzene (0.5 g, 2.3 mmol) was dissolved in dry MeCN (12 mL) and cooled to 0 °C. *tert*-Butyl-*N*-hydroxycarbamate (0.616 g, 4.6 mmol, 2 equiv) and DBU (0.71 mL, 4.6 mmol, 2 equiv) were added. The reaction mixture was stirred at 0 °C for 2 h, then warmed to ambient temperature overnight. Volatiles were removed under reduced pressure, and the resulting material was dissolved in  $CH_2Cl_2$ , washed with saturated  $K_2CO_3$  (aq) and with brine, then dried over anhydrous  $MgSO_4$  and filtered. The volatiles were removed to yield crude *tert*-butyl-(4-nitrobenzyl)oxycarbamate (0.432 g, 70% crude yield), which was carried forward without further purification.  $^1H$  NMR ( $CDCl_3$ , 500 MHz):  $\delta$  = 8.21 (dd,  $J$  = 6.83, 1.93 Hz, 2H), 7.58 (dd,  $J$  = 6.63, 2.04 Hz, 2H), 7.41 (s, 1H), 4.97 (s, 2H), 1.48 ppm (s, 9H).

*tert*-Butyl-(4-nitrobenzyl)oxycarbamate (0.1 g, 0.37 mmol) was combined with 10% Pd/C (0.032 g). The mixture was suspended in dry methanol (4 mL), then purged with hydrogen gas for 30 min; ethyl

acetate (50 mL) was added, and the solution was filtered through celite. The filtrate was concentrated to afford the crude product as yellow oil, which was purified by flash chromatography (silica gel, hexanes/EtOAc 7:3) to afford **11** as a yellow oil (0.081 g, 92% yield).  $^1H$  NMR ( $CDCl_3$ , 400 MHz):  $\delta$  = 7.18 (d,  $J$  = 8.31 Hz, 2H), 7.05 (s, 1H), 6.67 (d,  $J$  = 8.31 Hz, 2H), 4.73 (s, 2H), 1.48 ppm (s, 9H).

**Synthesis of SP7 (5):** Compound **7** (1.6 g, 2.96 mmol) was dissolved in THF (37 mL). LiOH (1.0 M, 7.54 mL) was added, and the reaction mixture was stirred under argon at ambient temperature for 48 h. Volatiles were removed under reduced pressure, and the resulting material was dissolved in ethyl acetate, washed with HCl (0.25 M) and brine, then dried over anhydrous  $MgSO_4$ . Volatiles were again removed to yield the carboxylic acid derivative of **7** as an off-white solid (1.6 g, quant.).  $^1H$  NMR ( $CD_3OD$ , 400 MHz):  $\delta$  = 7.99 (s, 1H), 7.67 (dd,  $J$  = 7.97, 1.62 Hz, 1H), 7.43 (q,  $J$  = 3.13 Hz, 6H), 7.31 (t,  $J$  = 7.46 Hz, 6H), 7.23 (t,  $J$  = 7.31 Hz, 3H), 7.09 (d,  $J$  = 8.07 Hz, 2H), 3.33 (s, 2H), 1.49 ppm (s, 9H).

The carboxylic acid (1.03 g, 2 mmol) and **11** (0.619 g, 2.6 mmol, 1.3 equiv) were dissolved in anhydrous  $CH_2Cl_2$  (4 mL). HBTU (1.1379 g, 3 mmol, 1.5 equiv) and DIPEA (2.6 mL, 2.5 equiv) were added. The solution was heated to 30 °C and stirred under argon for 48 h.  $CH_2Cl_2$  (100 mL) was added, and the solution was washed with water and saturated  $NaHCO_3$ , and then dried over anhydrous  $MgSO_4$ . Volatiles were removed to yield a crude product, which was purified by flash chromatography (silica gel, hexanes/EtOAc 7:3) to afford fully protected **5** as a yellow oil (0.536 g, 26% yield).  $^1H$  NMR ( $CDCl_3$ , 400 MHz):  $\delta$  = 8.23 (s, 1H), 8.08 (s, 1H), 7.63 (d,  $J$  = 8.4 Hz, 2H), 7.56 (dd,  $J$  = 7.6, 1.2 Hz, 1H), 7.49 (d,  $J$  = 7.6 Hz, 6H), 7.35 (t,  $J$  = 7.6 Hz, 6H), 7.28–7.21 (m, 6H), 7.11 (s, 1H), 6.82 (s, 1H), 4.82 (s, 2H), 3.21 (s, 1H), 1.56 (s, 9H), 1.49 ppm (s, 9H);  $^{13}C$  NMR ( $CDCl_3$ , 126 MHz):  $\delta$  = 165.2, 156.6, 153.0, 143.9, 138.2, 136.6, 134.9, 131.4, 131.1, 129.7, 129.2, 128.1, 126.9, 123.3, 120.2, 81.4, 80.9, 77.8, 67.3, 33.8, 28.3, 28.1 ppm; MS-ESI: calcd for  $C_{44}H_{47}N_3O_6S$ : 768.32  $[M+Na]^+$ ; found: 768.26.

Protected **5** (0.268 g, 0.36 mmol) was deprotected with TFA in  $CH_2Cl_2$ , as described for **2**, to afford **5** (0.199 g, quantitative yield).  $^1H$  NMR ( $CD_3OD$ , 400 MHz):  $\delta$  = 7.69 (d,  $J$  = 8 Hz, 2H), 7.37 (d,  $J$  = 8 Hz, 2H), 7.28 (s, 1H), 7.22–7.20 (m, 2H), 4.67 (s, 2H), 3.76 (s, 2H), 3.33 ppm (s, 2H);  $^{13}C$  NMR ( $CD_3OD$ , 126 MHz):  $\delta$  = 169.2, 146.9, 139.8, 136.1, 134.6, 130.5, 130.0, 122.1, 117.9, 116.2, 78.6, 25.7 ppm; MS-ESI: calcd for disulfide  $C_{30}H_{32}N_6O_4S_2$ : 605.75  $[M+H]^+$ ; found: 605.57.

**Cloning and plasmid construction:** Construction of the plasmids for the expression of CBD4(pAcF) through CBD12(pAcF) was described earlier.<sup>[14b]</sup> The plasmid for CBD15(pAcF) was prepared by overlap extension PCR with CBD12(pAcF) as template. The PCR product (750 bp) was cloned into a pET22b(+) vector (Novagen) by using NdeI and XhoI restriction enzymes. The glycine and proline scan libraries were prepared in a similar manner by using pCBD6 vector as template and forward primers containing NNK codons corresponding to the randomized positions. The plasmids encoding for the *i*–1 variants were prepared by using forward primers of general sequence 5'-AGACAG GATCCG GCXXX GCATCA CGG-3', 5'-GCTAGT TATTGC TCAGCGG-3' as reverse primer, and pCBD5 vector as template. The PCR product was cloned into the BamHI/XhoI cassette of the pCBD5 vector. The plasmid constructs were confirmed by DNA sequencing (Functional Genomics Center of the University of Rochester).

**Protein expression and purification:** Proteins were expressed in *E. coli* BL21(DE3) cells co-transformed with the plasmid encoding for the biosynthetic precursor and pEVOL\_pAcF vector, which en-

codes for an orthogonal tRNA<sub>CUA</sub>, and pAcF-tRNA synthetase<sup>[15]</sup> for the incorporation of pAcF through amber stop codon suppression. Overnight cultures were grown in lysogeny broth containing ampicillin (50 mg L<sup>-1</sup>) and chloramphenicol (34 mg L<sup>-1</sup>) and were used to inoculate M9 medium (0.4 L) containing ampicillin (50 mg L<sup>-1</sup>) and chloramphenicol (34 mg L<sup>-1</sup>) and supplemented with 1% glycerol. At OD<sub>600</sub>=0.6, protein expression was induced by adding 0.05% L-arabinose, isopropyl-β-D-thiogalactopyranoside (IPTG; 0.25 mM), and pAcF (1 mM). Cultures were grown for an additional 16 h at 27 °C and harvested by centrifugation at 3399 g. Frozen cells were resuspended in Tris (50 mM), NaCl (300 mM), and imidazole buffer (20 mM; pH 7.4) and lysed by sonication. Proteins were purified through Ni-NTA affinity chromatography by using Tris (50 mM), NaCl (150 mM), and imidazole (300 mM; pH 7.4) for the elution. The protein samples were concentrated, buffer-exchanged with potassium phosphate buffer (50 mM, pH 7.5) containing NaCl (150 mM), and stored at -80 °C. The OpgY-containing biosynthetic precursors were prepared in a similar manner with only difference being that cells contained a vector encoding for a tRNA<sub>CUA</sub>/AARS pair specific for OpgY, and OpgY was added to the culture medium during protein expression.<sup>[14a]</sup> The biosynthetic precursor libraries were prepared by pooling together *E. coli* colonies from the corresponding LB plates after transformation of the DNA ligation mixture.

**Cyclization reactions:** Reactions were carried out by adding the synthetic precursor (15 mM) to a solution of protein precursor (100 μM) in potassium phosphate buffer (50 mM, NaCl 150 mM, pH 7.5) in the presence of TCEP (20 mM). For SDS-PAGE analyses, and aliquot (5 μL) of the reaction mixture was removed at the indicated time point(s), diluted in DTT-free 4× loading buffer, and analyzed on 18% polyacrylamide gels. The extent of protein splicing was measured and quantified by densitometry analysis using the NIH Image Software. The percentage of SP-induced protein splicing was calculated based on the difference between the amount of spliced protein at *t*=0 and at the time of the analysis. MALDI-TOF analyses of the low-molecular-weight products (8–10 kDa) of the reactions were carried out on a Bruker Autoflex III MALDI-TOF spectrometer. Prior to analysis, protein samples were diluted in 50% acetonitrile/H<sub>2</sub>O (0.1% TFA), and this solution was mixed with a sinapinic acid solution (10 mg mL<sup>-1</sup> in 50% acetonitrile/H<sub>2</sub>O with 0.1% TFA). The samples were analyzed in reflectron positive (RP) mode and calibration by using low-molecular-weight (2–15 kDa) protein standards.

## Acknowledgements

This work was supported by the U.S. National Science Foundation grant CHE-1112342. N.T.J. acknowledges the NSF REU program for financial support. MS instrumentation was supported by the U.S. National Science Foundation grants CHE-0840410 and CHE-0946653. We are grateful to Peter G. Schultz for the plasmid encoding for tRNA<sub>CUA</sub>/pAcF-tRNA synthetase.

**Keywords:** bio-orthogonal ligation • cyclization • intein • macrocycles • organo-peptide hybrids

- [1] a) L. A. Wessjohann, E. Ruijter, D. Garcia-Rivera, W. Brandt, *Mol. Diversity* **2005**, *9*, 171–186; b) E. M. Driggers, S. P. Hale, J. Lee, N. K. Terrett, *Nat. Rev. Drug Discovery* **2008**, *7*, 608–624; c) J. A. Robinson, S. Demarco, F. Gombert, K. Moehle, D. Obrecht, *Drug Discovery Today* **2008**, *13*, 944–951; d) E. Marsault, M. L. Peterson, *J. Med. Chem.* **2011**, *54*, 1961–2004.
- [2] a) Y. Hamada, T. Shioiri, *Chem. Rev.* **2005**, *105*, 4441–4482; b) N. H. Tan, J. Zhou, *Chem. Rev.* **2006**, *106*, 840–895.
- [3] a) T. Satoh, S. Li, T. M. Friedman, R. Wiaderkiewicz, R. Korngold, Z. Huang, *Biochem. Biophys. Res. Commun.* **1996**, *224*, 438–443; b) D. P. Fairlie, J. D. Tyndall, R. C. Reid, A. K. Wong, G. Abbenante, M. J. Scanlon, D. R. March, D. A. Bergman, C. L. Chai, B. A. Burkett, *J. Med. Chem.* **2000**, *43*, 1271–1281; c) D. Wang, W. Liao, P. S. Arora, *Angew. Chem.* **2005**, *117*, 6683–6687; *Angew. Chem. Int. Ed.* **2005**, *44*, 6525–6529.
- [4] a) L. D. Walensky, A. L. Kung, I. Escher, T. J. Malia, S. Barbuto, R. D. Wright, G. Wagner, G. L. Verdine, S. J. Korsmeyer, *Science* **2004**, *305*, 1466–1470; b) T. Rezai, B. Yu, G. L. Millhauser, M. P. Jacobson, R. S. Lokey, *J. Am. Chem. Soc.* **2006**, *128*, 2510–2511; c) O. S. Gudmundsson, D. G. Vander Velde, S. D. Jois, A. Bak, T. J. Siahaan, R. T. Borchardt, *J. Pept. Res.* **1999**, *53*, 403–413.
- [5] a) Y. Q. Tang, J. Yuan, G. Osapay, K. Osapay, D. Tran, C. J. Miller, A. J. Ouellette, M. E. Selsted, *Science* **1999**, *286*, 498–502; b) F. Al-Obeidi, A. M. Castrucci, M. E. Hadley, V. J. Hruby, *J. Med. Chem.* **1989**, *32*, 2555–2561; c) M. A. Dechantsreiter, E. Planker, B. Matha, E. Lohof, G. Holzemann, A. Jonczyk, S. L. Goodman, H. Kessler, *J. Med. Chem.* **1999**, *42*, 3033–3040; d) N. R. Graciani, K. Y. Tsang, S. L. McCutchen, J. W. Kelly, *Bioorg. Med. Chem.* **1994**, *2*, 999–1006; e) R. L. Dias, R. Fasan, K. Moehle, A. Renard, D. Obrecht, J. A. Robinson, *J. Am. Chem. Soc.* **2006**, *128*, 2726–2732; f) R. M. Cardoso, F. M. Brunel, S. Ferguson, M. Zwick, D. R. Burton, P. E. Dawson, I. A. Wilson, *J. Mol. Biol.* **2007**, *365*, 1533–1544; g) L. K. Henchey, J. R. Porter, I. Ghosh, P. S. Arora, *ChemBioChem* **2010**, *11*, 2104–2107.
- [6] a) E. Koivunen, D. A. Gay, E. Ruoslahti, *J. Biol. Chem.* **1993**, *268*, 20205–20210; b) N. C. Wrighton, F. X. Farrell, R. Chang, A. K. Kashyap, F. P. Barbone, L. S. Mulcahy, D. L. Johnson, R. W. Barrett, L. K. Jolliffe, W. J. Dower, *Science* **1996**, *273*, 458–464; c) L. Oligino, F. D. Lung, L. Sastry, J. Bigelow, T. Cao, M. Curran, T. R. Burke, Jr., S. Wang, D. Krag, P. P. Roller, C. R. King, *J. Biol. Chem.* **1997**, *272*, 29046–29052; d) W. J. Fairbrother, H. W. Christinger, A. G. Cochran, G. Fuh, C. J. Keenan, C. Quan, S. K. Shriver, J. Y. Tom, J. A. Wells, B. C. Cunningham, *Biochemistry* **1998**, *37*, 17754–17764; e) G. R. Nakamura, M. A. Starovasnik, M. E. Reynolds, H. B. Lowman, *Biochemistry* **2001**, *40*, 9828–9835; f) K. Deshayes, M. L. Schaffner, N. J. Skelton, G. R. Nakamura, S. Kadkhodayan, S. S. Sidhu, *Chem. Biol.* **2002**, *9*, 495–505; g) J. Stamos, C. Eigenbrot, G. R. Nakamura, M. E. Reynolds, J. Yin, H. B. Lowman, W. J. Fairbrother, M. A. Starovasnik, *Structure* **2004**, *12*, 1289–1301; h) C. D. Shomin, E. Restituyo, K. J. Cox, I. Ghosh, *Bioorg. Med. Chem.* **2011**, *19*, 6743–6749.
- [7] a) S. W. Millward, T. T. Takahashi, R. W. Roberts, *J. Am. Chem. Soc.* **2005**, *127*, 14142–14143; b) C. Heinis, T. Rutherford, S. Freund, G. Winter, *Nat. Chem. Biol.* **2009**, *5*, 502–507; c) Y. V. Guillen Schlippe, M. C. T. Hartman, K. Josephson, J. W. Szostak, *J. Am. Chem. Soc.* **2012**, *134*, 10469–10477; d) S. Chen, J. Morales-Sanfrutos, A. Angelini, B. Cutting, C. Heinis, *ChemBioChem* **2012**, *13*, 1032–1038.
- [8] a) C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnnon, S. J. Benkovic, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13638–13643; b) T. M. Kinsella, C. T. Ohashi, A. G. Harder, G. C. Yam, W. Li, B. Peelle, E. S. Pali, M. K. Bennett, S. M. Molineaux, D. A. Anderson, E. S. Masuda, D. G. Payan, *J. Biol. Chem.* **2002**, *277*, 37512–37518; c) J. A. Kritzer, S. Hamamichi, J. M. McCaffery, S. Santagata, T. A. Naumann, K. A. Caldwell, G. A. Caldwell, S. Lindquist, *Nat. Chem. Biol.* **2009**, *5*, 655–663; d) G. Deschuyteneer, S. Garcia, B. Michiels, B. Baudoux, H. Degand, P. Morsomme, P. Soumillon, *ACS Chem. Biol.* **2010**, *5*, 691–700.
- [9] a) R. M. Kohli, C. T. Walsh, M. D. Burkart, *Nature* **2002**, *418*, 658–661; b) T. Hamamoto, M. Sisido, T. Ohtsuki, M. Taki, *Chem. Commun.* **2011**, *47*, 9116–9118.
- [10] a) Y. Goto, A. Ohta, Y. Sako, Y. Yamagishi, H. Murakami, H. Suga, *ACS Chem. Biol.* **2008**, *3*, 120–129; b) T. Kawakami, A. Ohta, M. Ohuchi, H. Ashigai, H. Murakami, H. Suga, *Nat. Chem. Biol.* **2009**, *5*, 888–890.
- [11] a) M. S. Donia, B. J. Hathaway, S. Sudek, M. G. Haygood, M. J. Rosovitz, J. Ravel, E. W. Schmidt, *Nat. Chem. Biol.* **2006**, *2*, 729–735; b) Y. X. Shi, X. A. Yang, N. Garg, W. A. van der Donk, *J. Am. Chem. Soc.* **2011**, *133*, 2338–2341; c) S. J. Pan, A. J. Link, *J. Am. Chem. Soc.* **2011**, *133*, 5016–5023; d) T. A. Knappe, F. Manzenrieder, C. Mas-Moruno, U. Linne, F. Sasse, H. Kessler, X. Xie, M. A. Marahiel, *Angew. Chem.* **2011**, *123*, 8873–8876; *Angew. Chem. Int. Ed.* **2011**, *50*, 8714–8717.
- [12] a) Z. J. Gartner, B. N. Tse, R. Grubina, J. B. Doyon, T. M. Snyder, D. R. Liu, *Science* **2004**, *305*, 1601–1605; b) R. Fasan, R. L. Dias, K. Moehle, O.

- Zerbe, D. Obrecht, P. R. Mittl, M. G. Grutter, J. A. Robinson, *ChemBioChem* **2006**, *7*, 515–526; c) S. B. Shin, B. Yoo, L. J. Todaro, K. Kirshenbaum, *J. Am. Chem. Soc.* **2007**, *129*, 3218–3225; d) Y. U. Kwon, T. Kodadek, *Chem. Commun.* **2008**, 5704–5706; e) V. S. Fluxa, J. L. Reymond, *Bioorg. Med. Chem.* **2009**, *17*, 1018–1025; f) D. G. Rivera, O. E. Vercillo, L. A. Wessjohann, *Org. Biomol. Chem.* **2008**, *6*, 1787–1795; g) S. A. Fowler, D. M. Stacy, H. E. Blackwell, *Org. Lett.* **2008**, *10*, 2329–2332; h) A. Isidro-Llobet, T. Murillo, P. Bello, A. Cilibrizzi, J. T. Hodgkinson, W. R. Galloway, A. Bender, M. Welch, D. R. Spring, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6793–6798; i) R. Hili, V. Rai, A. K. Yudin, *J. Am. Chem. Soc.* **2010**, *132*, 2889–2891; j) S. Ghosh, L. A. Ingerman, A. G. Frye, S. J. Lee, M. R. Gagne, M. L. Waters, *Org. Lett.* **2010**, *12*, 1860–1863; k) V. Dewan, T. Liu, K. M. Chen, Z. Qian, Y. Xiao, L. Kleiman, K. V. Mahasenan, C. Li, H. Matsuo, D. Pei, K. Musier-Forsyth, *ACS Chem. Biol.* **2012**, *7*, 761–769.
- [13] a) S. W. Millward, S. Fiacco, R. J. Austin, R. W. Roberts, *ACS Chem. Biol.* **2007**, *2*, 625–634; b) T. S. Young, D. D. Young, I. Ahmad, J. M. Louis, S. J. Benkovic, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 11052–11056; c) M. D. Tianero, M. S. Donia, T. S. Young, P. G. Schultz, E. W. Schmidt, *J. Am. Chem. Soc.* **2012**, *134*, 418–425.
- [14] a) J. M. Smith, F. Vitali, S. A. Archer, R. Fasan, *Angew. Chem.* **2011**, *123*, 5181–5186; *Angew. Chem. Int. Ed.* **2011**, *50*, 5075–5080; b) M. Satyanarayana, F. Vitali, J. R. Frost, R. Fasan, *Chem. Commun.* **2012**, *48*, 1461–1463.
- [15] L. Wang, Z. Zhang, A. Brock, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 56–61.
- [16] a) S. Chong, F. B. Mersha, D. G. Comb, M. E. Scott, D. Landry, L. M. Vence, F. B. Perler, J. Benner, R. B. Kucera, C. A. Hirvonen, J. J. Pelletier, H. Paulus, M. Q. Xu, *Gene* **1997**, *192*, 271–281; b) T. T. Hoang, Y. Ma, R. J. Stern, M. R. McNeil, H. P. Schweizer, *Gene* **1999**, *237*, 361–371; c) T. C. Evans, Jr., J. Benner, M. Q. Xu, *Protein Sci.* **1998**, *7*, 2256–2264; d) M. W. Southworth, K. Amaya, T. C. Evans, M. Q. Xu, F. B. Perler, *Biotechniques* **1999**, *27*, 110–120.
- [17] T. M. Hackeng, J. H. Griffin, P. E. Dawson, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 10068–10073.
- [18] a) T. Klabunde, S. Sharma, A. Telenti, W. R. Jacobs, Jr., J. C. Sacchettini, *Nat. Struct. Biol.* **1998**, *5*, 31–36; b) A. Romanelli, A. Shekhtman, D. Cowburn, T. W. Muir, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 6397–6402; c) S. Frutos, M. Goger, B. Giovani, D. Cowburn, T. W. Muir, *Nat. Chem. Biol.* **2010**, *6*, 527–533.
- [19] A. Telenti, M. Southworth, F. Alcaide, S. Daugelat, W. R. Jacobs, Jr., F. B. Perler, *J. Bacteriol.* **1997**, *179*, 6378–6382.
- [20] W. P. Jencks, *J. Am. Chem. Soc.* **1959**, *81*, 475–481.

---

Received: September 10, 2012

Please note: Minor changes have been made to Figure 2 since this manuscript was published in *ChemBioChem* EarlyView. The Editor.

Published online on November 30, 2012

CHEM**BIO**CHEM

## Supporting Information

© Copyright Wiley-VCH Verlag GmbH & Co. KGaA, 69451 Weinheim, 2012

### **Macrocyclization of Organo-Peptide Hybrids through a Dual Bio-orthogonal Ligation: Insights from Structure–Reactivity Studies**

John R. Frost, Francesca Vitali, Nicholas T. Jacob, Micah D. Brown, and Rudi Fasan<sup>\*[a]</sup>

cbic\_201200579\_sm\_miscellaneous\_information.pdf

CHEM**BIO**CHEM

## Supporting Information

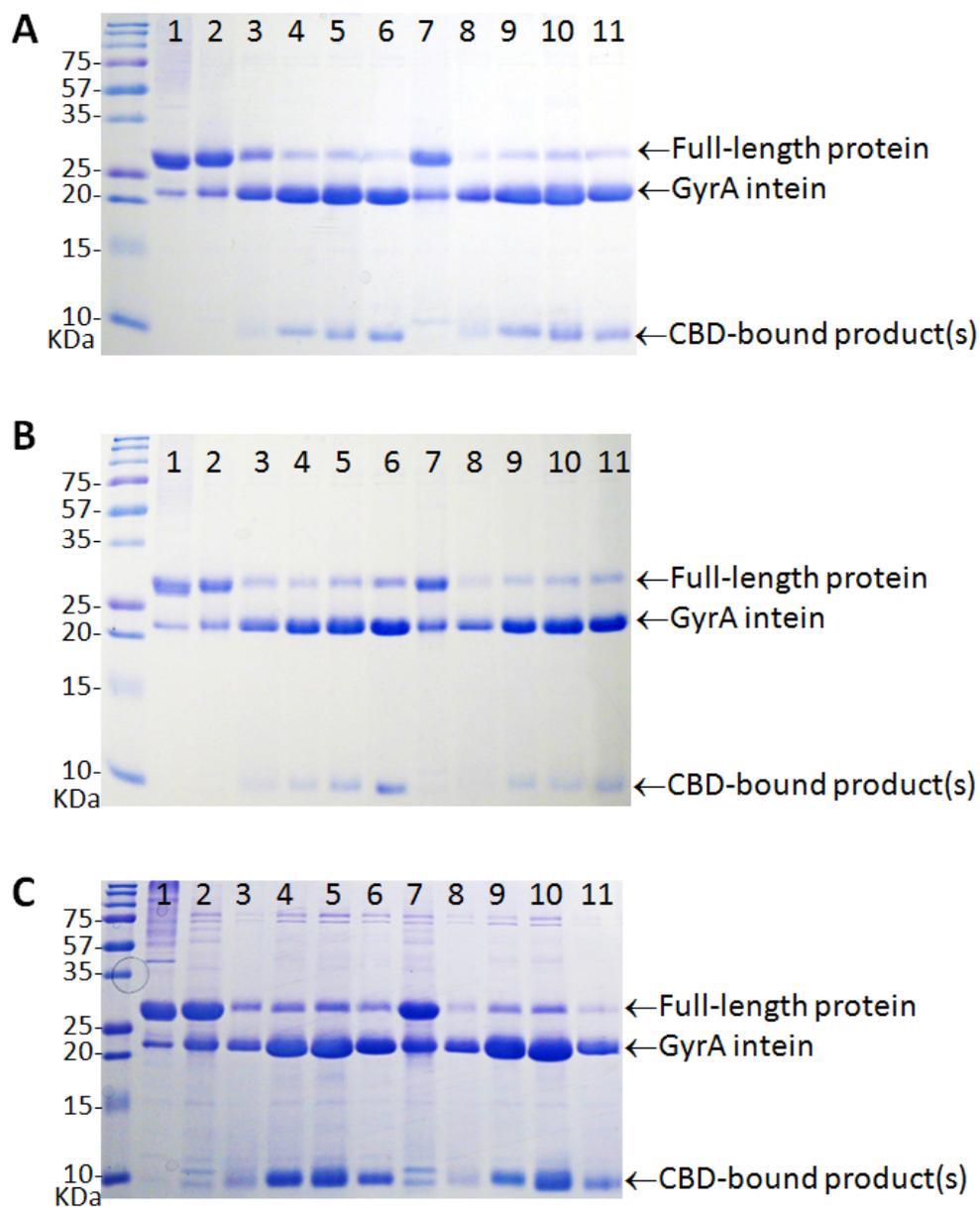
© Copyright Wiley-VCH Verlag GmbH & Co. KGaA, 69451 Weinheim, 2012

### **Macrocyclization of Organo-Peptide Hybrids through a Dual Bio-orthogonal Ligation: Insights from Structure–Reactivity Studies**

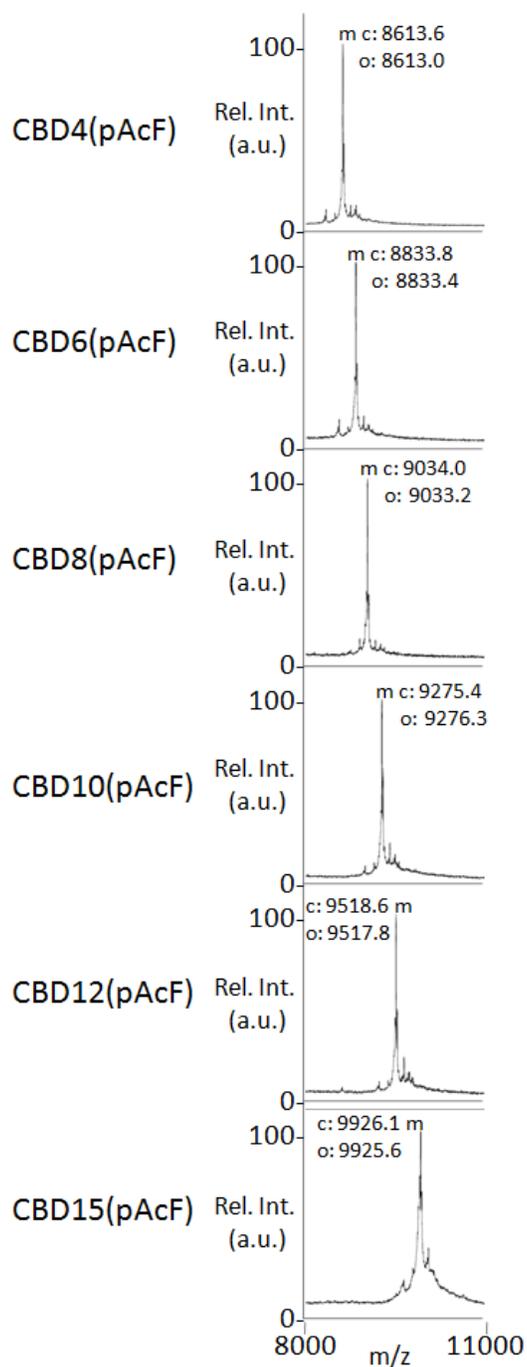
John R. Frost, Francesca Vitali, Nicholas T. Jacob, Micah D. Brown, and Rudi Fasan<sup>\*[a]</sup>

cbic\_201200579\_sm\_miscellaneous\_information.pdf

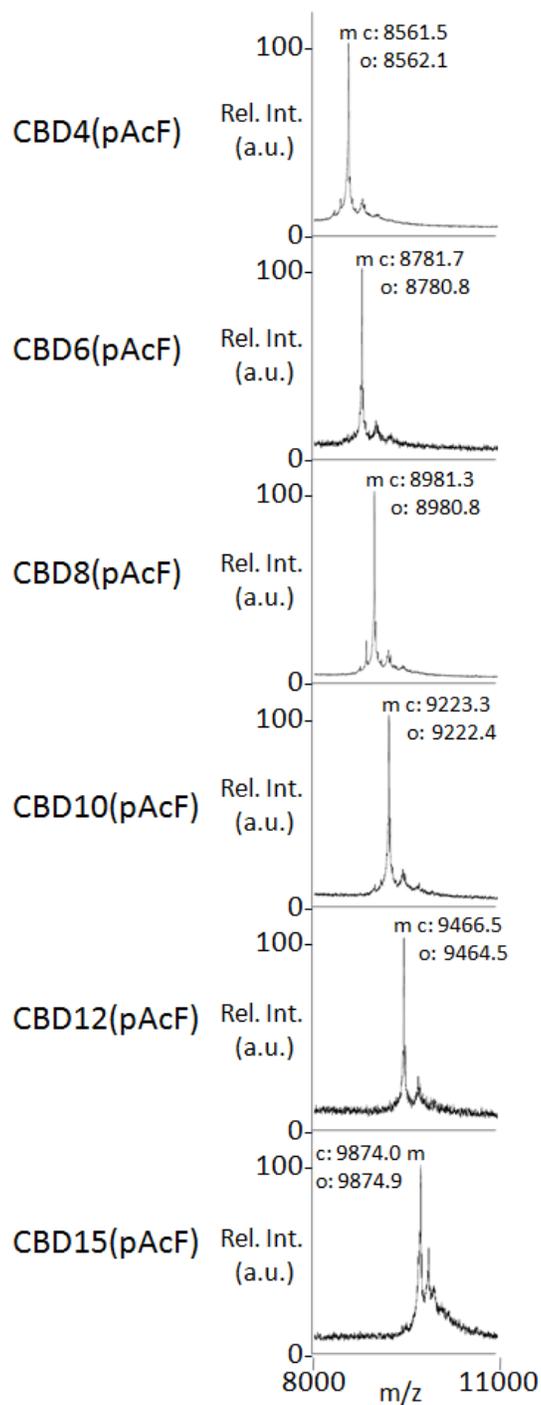
**Figure S1.** Representative SDS-PAGE protein gels for the reactions with synthetic precursors **2-5** and precursor proteins (A) CBD4(pAcF), (B) CBD5(pAcF), and (C) CBD6(pAcF) after 5 hours (lanes 3-6) and 12 hours (lanes 8-11). Lane 1: protein only (time 0); lane 2: protein only (5 h); lane 3: protein + **2** (5 h); lane 4: protein + **5** (5 h); lane 5: protein + **3** (5 h); lane 6: protein + **4** (5 h); lane 7: protein only (12 h); lane 8: protein + **2** (12 h); lane 9: protein + **5** (12 h); lane 10: protein + **3** (12 h); lane 11: protein + **4** (12 h).



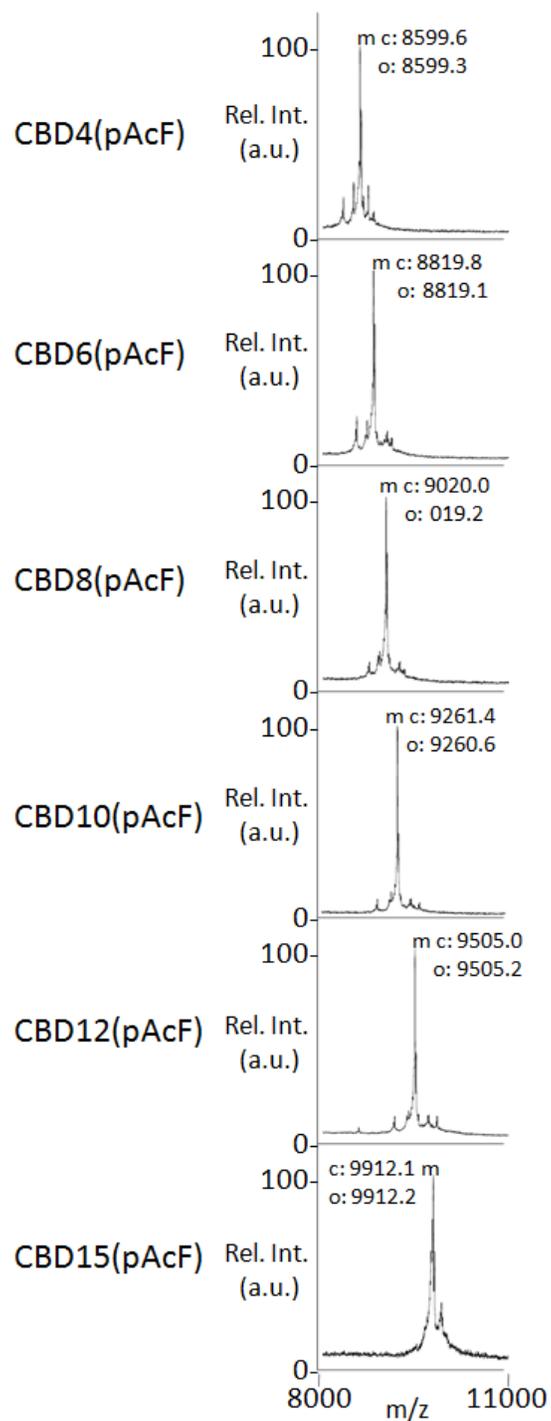
**Figure S2.** Cyclization reactions with SP5 (**3**). MALDI-TOF MS spectra of MORPHs generated upon reaction of SP5 (**3**) with the biosynthetic precursors containing 4mer, 6mer, 8mer, 10mer, and 15mer target sequences (**Table 1**). Calculated (c) and observed (o) m/z values corresponding to the  $[M + H]^+$  adduct of the macrocyclic product ('m') are indicated.



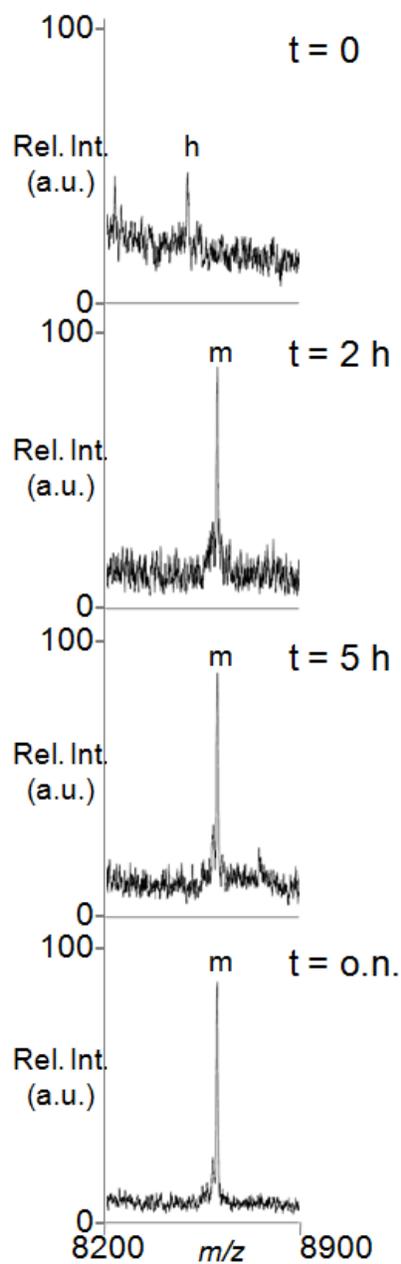
**Figure S3.** Cyclization reactions with SP6 (**4**). MALDI-TOF MS spectra of MORPHs generated upon reaction of SP6 (**4**) with the biosynthetic precursors containing 4mer, 6mer, 8mer, 10mer, and 15mer target sequences (**Table 1**). Calculated (c) and observed (o)  $m/z$  values corresponding to the  $[M + H]^+$  adduct of the macrocyclic product ('m') are indicated.



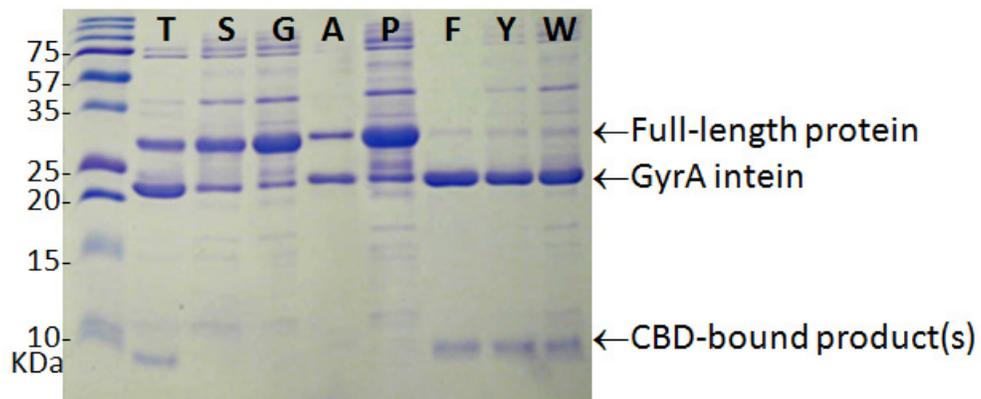
**Figure S4.** Cyclization reactions with SP7 (**5**). MALDI-TOF MS spectra of MORPHs generated upon reaction of SP7 (**5**) with the biosynthetic precursors containing 4mer, 6mer, 8mer, 10mer, and 15mer target sequences (**Table 1**). Calculated (c) and observed (o)  $m/z$  values corresponding to the  $[M + H]^+$  adduct of the macrocyclic product ('m') are indicated.



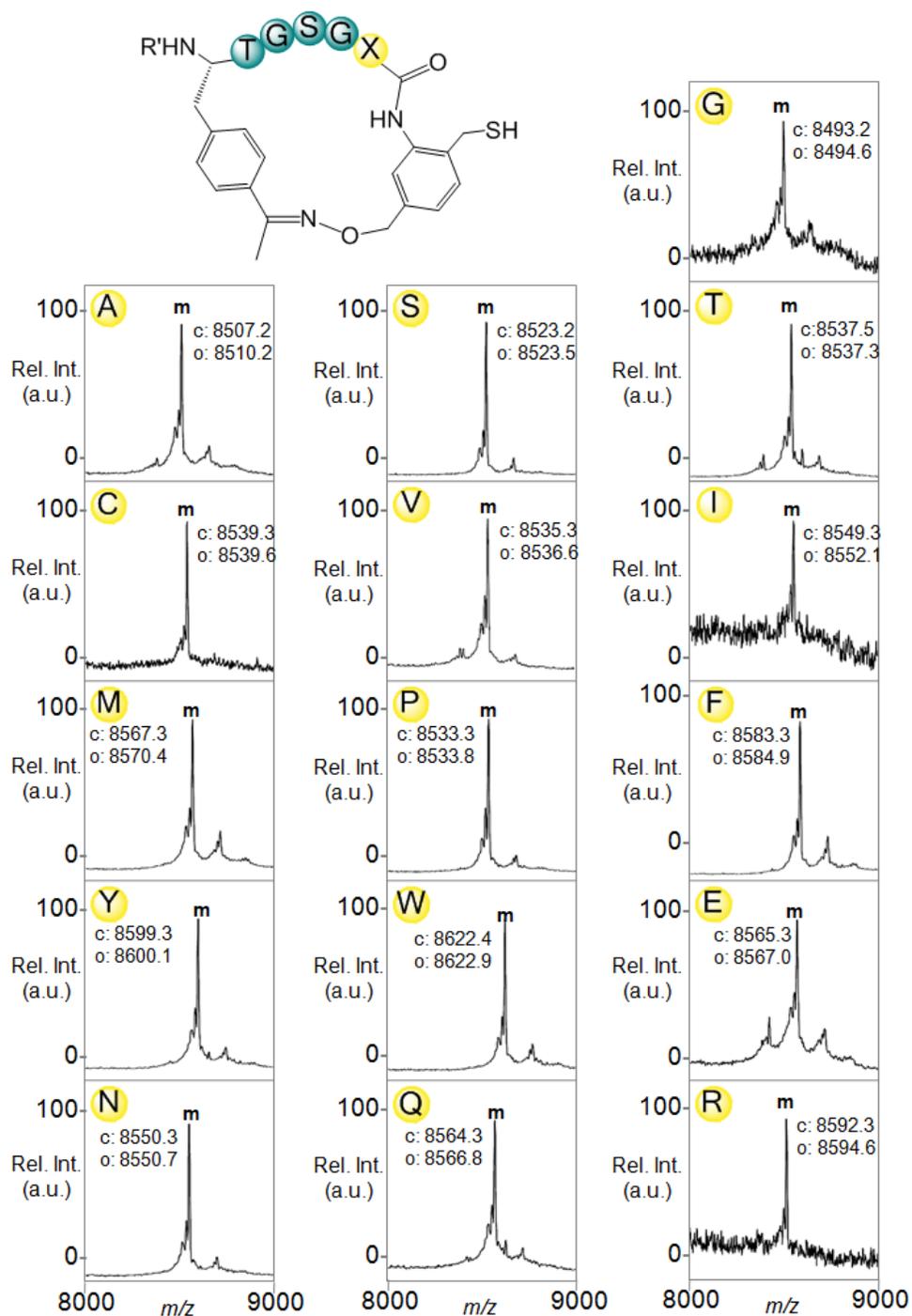
**Figure S5.** MOrPH synthesis and stability in cell lysate. MALDI-TOF MS spectra of the eluate from chitin beads before ( $t = 0$ ) and after addition of SP4 (**2**) to a cell lysate solution of *E. coli* cells expressing CBD5(pAcF). 'm' corresponds to the  $[M + H]^+$  adduct of the desired macrocyclic product, whereas 'h' indicates the  $[M + H]^+$  adduct of the small-MW product resulting from hydrolysis of the protein (i.e. CBD-pAcF-TGSGT-COOH).



**Figure S6.** I - 1 protein variants. Representative SDS-PAGE protein gel illustrating the different amount of SP4-induced splicing for the I - 1 variants of protein precursor CBD5(pAcF). The identity of the I - 1 residue is indicated on the top of each lane. Protein samples were analyzed after 5-hour incubation with SP4 (2).



**Figure S7.** Cyclization reactions with I – 1 protein variants. MALDI-TOF MS spectra of MORPHs generated upon reaction of SP4 (**2**) with biosynthetic precursor CBD5(pAcF) containing different amino acids at the I – 1 site. Calculated (c) and observed (o)  $m/z$  values corresponding to the  $[M + H]^+$  adduct of the macrocyclic product ('m') are indicated. R' = chitin binding domain.



**Figure S8.** Time course of SP-induced splicing of precursor protein CBD5 incorporating either pAcF or OpgY. Reactions with (A) SP5 (**3**), (B) SP6 (**4**), and (C) SP7 (**5**). Error bars are derived from experiments carried out in duplicate.

