ChemComm

Cite this: *Chem. Commun.*, 2012, **48**, 1461–1463

www.rsc.org/chemcomm

COMMUNICATION

Diverse organo-peptide macrocycles *via* a fast and catalyst-free oxime/intein-mediated dual ligation†‡

Maragani Satyanarayana, Francesca Vitali, John R. Frost and Rudi Fasan*

Received 15th June 2011, Accepted 11th August 2011
DOI: 10.1039/c1cc13533c

Macrocyclic Organo-Peptide Hybrids (MOrPHs) can be prepared from genetically encoded polypeptides *via* a chemoselective and catalyst-free reaction between a trifunctional oxyamino/amino-thiol synthetic precursor and an intein-fusion protein incorporating a bioorthogonal keto group.

Macrocyclic peptides and peptide-based structures have attracted significant interest as a source of chemical probes and therapeutic agents.[1] While peptides and peptidomimetics in rigidified configurations can be prepared synthetically,[2] genetic encoding offers the advantage to couple the creation of vast chemical libraries (10^7–10^{10}) with ultrahigh-throughput screening methods.[3–5] Notable approaches involve the introduction of disulfide bridges within randomized peptide sequences[3] or formation of cyclic peptides *via* split intein-mediated cyclization, but the range of building blocks available to assemble these structures remains inherently limited compared to synthetic methods.[4] Alternatively, ribosomal peptides have been constrained through the use of cysteine- or amine-reactive cross-linking agents but these methods rely on non-directional and non-bioorthogonal reactions which limits the choice of the cross-linking scaffolds and it may lead to multiple undesired products.[5] To overcome these major limitations, we have undertaken efforts toward implementing general methods for chemoselectively embedding variable synthetic scaffolds within ribosomal peptides to generate macrocycles with a hybrid peptidic/non-peptidic backbone, referred to as Macrocyclic Organo-Peptide Hybrids or MOrPHs.[6] Here, we report an efficient strategy for MOrPH synthesis which exploits a highly chemoselective, bioorthogonal, and catalyst-free tandem reaction between a trifunctional oxyamino/amino-thiol synthetic precursor (SP) and genetically encoded biosynthetic precursors (BPs) incorporating a keto group (Fig. 1A).

Based on our recent investigations,[6] we envisioned that a suitable biosynthetic precursor for MOrPH construction could be generated by framing a target peptide sequence ('TS') between the unnatural amino acid *para*-acetylphenylalanine (pAcF)[7] and an intein (species 'a' in Fig. 1A). This protein would display two functional groups with orthogonal reactivity, namely the keto group of pAcF at the N-terminus of the target sequence and the reactive thioester bond transiently formed at its C-terminus *via* intein-catalyzed N,S-acyl transfer (species 'b'). Macrocyclization could then be achieved *via* a synthetic precursor equipped with (i) an oxyamino group to form a stable oxime linkage with a pAcF side chain, and (ii) an amino-thiol moiety to coordinate an intein-mediated ligation and concomitant excision of the intein from the biosynthetic precursor.

To test this approach, we prepared a first set of six biosynthetic precursors with target sequences spanning 4, 5, 6, 8, 10, or 12 amino acids (CBD4(pAcF) to CBD12(pAcF), Table S1 (ESI‡)).

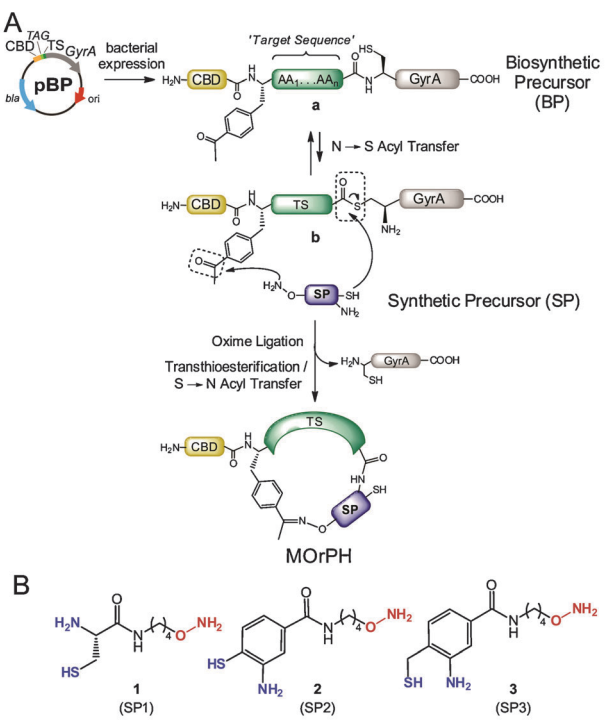


Fig. 1 (A) Synthesis of Macrocyclic Organo-Peptide Hybrids *via* oxime/intein-mediated tandem ligation. CBD: Chitin Binding Domain. TAG: amber stop codon. TS: Target Sequence. GyrA: intein GyrA from *Mycobacterium xenopi*. (B) Oxyamine/amino-thiol synthetic precursors. Their synthesis is described in Schemes S2–S5 of ESI.‡

Department of Chemistry, University of Rochester, Rochester NY, USA. E-mail: fasan@chem.rochester.edu; Fax: +1 585-276-0205; Tel: +1 585-275-3504

† This article is part of the ChemComm. 'Emerging Investigators 2012' themed issue.

‡ Electronic supplementary information (ESI) available: Experimental details, synthetic procedures, characterization data, and supplemental figures and tables. See DOI: 10.1039/c1cc13533c

A Chitin Binding Domain (CBD, 71 amino acids) was introduced as a N-terminal tail to mimic the molecular arrangement in a typical display system, where a variable amino acid sequence is tethered to a viral or cellular protein. pAcF was synthesized in three steps from 4-methyl-acetophenone *via* an optimized version of a described procedure[7] (overall yield: 86%, Scheme S1 (ESI‡)). The six precursor proteins were expressed in *E. coli* cells co-tranformed with a plasmid encoding for the precursor protein (pBP) and one containing a $tRNA_{CUA}$/pAcF-tRNA synthetase pair (pEVOL-pAcF)[7] for suppressing the amber stop codon (TAG) located upstream of the target sequence with pAcF.

Hydrazino, hydrazido, and oxyamino groups were all viable choices to mediate coupling of the synthetic precursor to the side-chain ketone in the precursor proteins. In preliminary experiments, a GyrA-fused protein was incubated with methylhydrazine, phenylhydrazide or methoxyamine at 50 mM in phosphate buffer (pH 7.5). Methoxyamine was unable to induce intein splicing at detectable levels even after overnight incubation at room temperature which suggested that oxyamino-containing SPs would not react with the thioester bond through the nucleophilic $-ONH_2$ group, providing the desired orthogonality between these functional groups. Oxime linkages also exhibit considerably higher hydrolytic stability compared to hydrazones.[8]

Based on work in the area of Native Chemical Ligation (NCL) and its extensions,[9,10] we suspected that the pK_a, nucleophilicity, and spatial arrangement of the amino and thiol groups could influence considerably the ability of the SP amino-thiol moiety to undergo the desired transthioesterification and S→N acyl transfer reactions. Accordingly, three different trifunctional SPs were synthesized (Fig. 1B, Schemes S2–S5 (ESI‡)). Given the reactivity of intein-fusion proteins toward peptides with N-terminal cysteines in Expressed Protein Ligation (EPL),[11,12] SP1 (**1**) was prepared based on this amino acid, which features a 1,2-amino-thiol connectivity and an alkyl thiol with a pK_a of ~8.5. SP2 (**2**) was designed to carry a more acidic thiol (pK_a of *o*-amino-thiophenol ≈ 6.6)[13] and present the same 1,2-amino-thiol connectivity as **1**, but rigidified through installation onto an aromatic ring. The third design, SP3 (**3**), integrates a benzylic thiol (pK_a ≈ 9.5)[14] in a semi-rigid configuration and 1,3-arrangement with respect to the amino group.

Next, we tested the feasibility of the strategy outlined in Fig. 1A by performing reactions where each protein construct (100 μM) was exposed to **1**, **2**, or **3** (15 mM) in phosphate buffer at pH 7.5. Tris(2-carboxyethyl)phosphine (TCEP, 20 mM) was added to the reaction mixtures to maintain the thiol groups in the reactants in reduced form. Splicing of the protein constructs over time was quantified by SDS-PAGE and densitometric analysis of the gel bands corresponding to the full-length protein (31 kDa) and the splicing fragments, GyrA (22 kDa) and the CBD-linked products (8 kDa) (Fig. S1, ESI‡). These experiments revealed that **1** and **2** were poorly efficient in promoting splicing of the GyrA intein across all the protein constructs (Fig. 2A). After 5 hours at room temperature, **1**- and **2**-induced splicing of the protein precursors ranged from 5 to 15%, which was comparable to that of the negative controls with no synthetic precursor. In addition, MALDI-TOF analysis showed no trace of the desired macrocycles, while observed species corresponded to the hydrolysis product CBD-(pAcF)-(target sequence)-COOH and the acyclic SP-bound product (Fig. S2–S3, ESI‡). In stark contrast, the reactions with **3** exhibited a considerably larger extent of protein splicing (50–80%) after 5 hours (Fig. 2A and Fig. S1 (ESI‡)) and quantitative splicing (90–100%) after overnight. Remarkably, MALDI-TOF analysis showed *the formation of the desired CBD-tethered hybrid macrocycle as the only product from these reactions* (Fig. 2B). Also notable was that **3**-induced macrocyclization occurred with almost equally high efficiency across all the different target sequence lengths, including the short 4mer (TGST) and the considerably longer 12mer sequence (TGSWGKLAEYGT). Furthermore, whereas the biosynthetic precursors undergo slow hydrolysis (5–8% after 5 hours, Fig. 2A), the observation of no or minimal amounts of hydrolyzed product ('h') in the presence of **3** (Fig. 2B) indicated that the macrocyclization largely outcompetes the undesired hydrolysis process, even in the context of the extended 10- and 12-amino acid target sequences.

To establish whether the *o*-amino-benzyl-thiol moiety of **3** was capable of undergoing the desired S→N acyl transfer after transthioesterification, the reaction mixtures were treated with iodoacetamide (20 mM) for 2 hours followed by MALDI-TOF analysis. Such treatment led to complete disappearance of the $[M + H]^+$ species corresponding to the MOrPHs and to the appearance of a species with an m/z of +58 (Fig. S4 (ESI‡)), which is consistent with the addition of an acetamido moiety to these molecules. These tests evidenced the accessibility of the benzylic thiol in the macrocycles to attack by the alkylating agent, supporting the occurrence of the amide-forming intramolecular rearrangement.

Given the identical alkoxyamino group and linker connecting this group to the amino-thiol moiety, **1**, **2**, and **3** are expected to

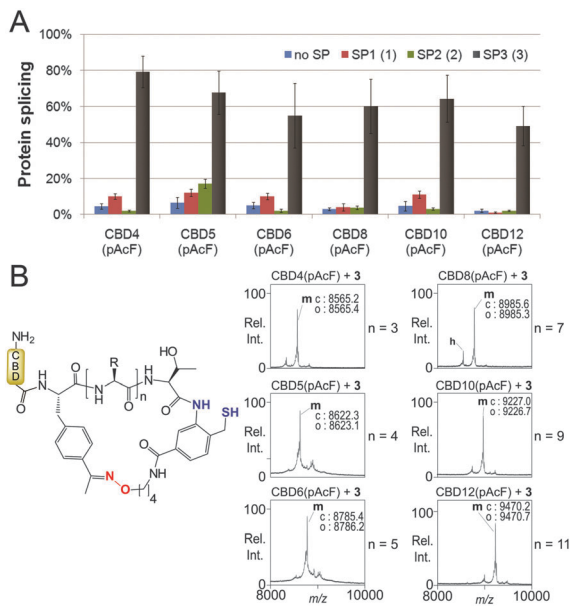


Fig. 2 (A) Percentage of precursor protein splicing in the presence of no SP, **1**, **2**, or **3** after 5 hours as determined by SDS-PAGE. Error bars are from triplicate experiments. Protein constructs are described in Table S1 (ESI‡). (B) MALDI-TOF spectra of MOrPHs obtained from reaction of **3** with precursor proteins CBD4(pAcF) to CBD12(pAcF). Calculated (c) and observed (o) m/z values corresponding to the $[M + H]^+$ adduct of the macrocyclic product ('m') are indicated.

possess the same reactivity toward oxime formation. Thus, their differential performance in promoting MOrPH formation (Fig. 2, Fig. S2, S3 (ESI‡)) can be attributed to the intein splicing properties of their amino-thiol moieties. The poor performance of the cysteine-based **1** can be rationalized considering the stringency of the applied conditions (SP at 15 mM, no thiol catalyst added, short incubation time) compared to EPL protocols, which typically involve high concentrations of thiol catalysts (up to 200 mM) as well as longer reaction times.[11] More surprising was the inefficiency of **2** to induce intein splicing given that thiophenol and related aromatic thiols, including 4-aminothiophenol, are effective catalysts for NCL reactions.[10] We conclude that the *ortho* amino group drastically reduces the nucleophilicity of the neighboring thiol in the context of intein splicing, possibly due to steric effects and/or unfavorable hydrogen bonding interactions with the protein. By comparison, the MOrPH-forming ability of **3** stems from the superior intein splicing properties of its 2-amino-benzylthiol, a structure which has never been described in the context of thioester- or intein-mediated ligations.[15] Clearly, such a structure preserves the nucleophilicity of the benzylic thiol while placing the amino group at a viable distance for acyl transfer *via* a six-membered ring intermediate.

To investigate the possibility of diversifying the macrocycle structures by varying their genetically encoded moiety, we constructed two biosynthetic precursor libraries with randomized 5mer and 8mer target sequences, namely CBD-(pAcF)-X_4T-GyrA and CBD-(pAcF)-X_7T-GyrA, where X corresponds to a fully randomized position (NNK codon). About 5000 recombinants from each library were pooled together and expressed in *E. coli*. SDS-PAGE revealed only small amounts of premature splicing during expression (<15–20%). For both libraries, **3** induced more than 35% and 60% splicing of the full-length proteins after 5 hours and 16 hours, respectively. To establish the occurrence of macrocyclization, 18 randomly chosen recombinants from each library were isolated and characterized. Remarkably, all the recombinants from the 5mer BP library and all but one of the 18 recombinants from the 8mer BP library yielded the desired hybrid macrocycle (Tables S2 and S3, ESI‡). For only 2/18 of the 5mer BPs and 1/18 of the 8mer BPs a small amount of acyclic product (15–25%) was observed. Notably, the majority of the 5mer and 8mer BP variants (63% and 58%, respectively) underwent more than 50% splicing after overnight incubation at room temperature (Fig. 3). Most importantly, these experiments proved the functionality of the method across largely divergent target sequences and demonstrated its versatility in generating diversified MOrPH structures.

In summary, we have developed an efficient method to construct Macrocyclic Organo-Peptide Hybrids *via* a dual oxime/intein-mediated ligation. The chemoselectivity, bioorthogonality and catalyst-free nature of this strategy and its demonstrated efficiency in the context of precursor target sequences of varying length and randomized composition hold promise toward exploiting it to generate diversified MOrPHs tethered to a viral/cellular surface of a display system. Efforts are ongoing to investigate this approach toward the isolation of MOrPH-based ligands for selective protein recognition.

This work was supported by startup funds from the University of Rochester. MS instrumentation was supported by National Science Foundation grant CHE-0840410. We thank Peter G. Schultz for kindly providing pEVOL_pAcF vector.

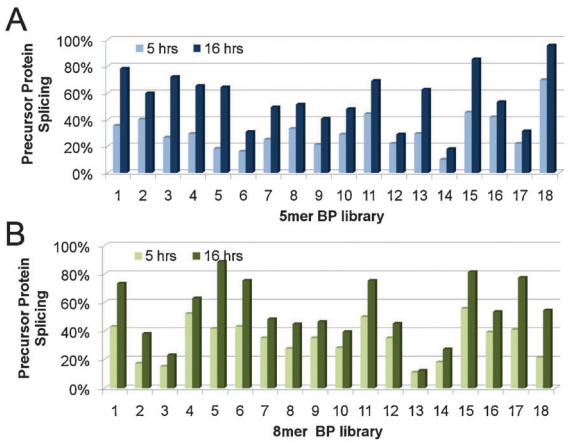


Fig. 3 Extent of **3**-induced protein splicing for 18 variants from the library of precursor proteins with randomized 5mer (A) and 8mer (B) target sequences. See also Tables S2 and S3 in ESI.‡

Notes and references

1 E. M. Driggers, S. P. Hale, J. Lee and N. K. Terrett, *Nat. Rev. Drug Discovery*, 2008, **7**, 608–624.
2 J. M. Humphrey and A. R. Chamberlin, *Chem. Rev.*, 1997, **97**, 2243–2266; P. Li, P. P. Roller and J. Xu, *Curr. Org. Chem.*, 2002, **6**, 411–440; R. Fasan, R. L. Dias, K. Moehle, O. Zerbe, D. Obrecht, P. R. Mittl, M. G. Grutter and J. A. Robinson, *ChemBioChem*, 2006, **7**, 515–526; G. T. Bourne, J. L. Nielson, J. F. Coughlan, P. Darwen, M. R. Campitelli, D. A. Horton, A. Rhumann, S. G. Love, T. T. Tran and M. L. Smythe, *Methods Mol. Biol.*, 2005, **298**, 151–165; V. S. Fluxa and J. L. Reymond, *Bioorg. Med. Chem.*, 2009, **17**, 1018–1025.
3 K. T. O'Neil, R. H. Hoess, S. A. Jackson, N. S. Ramachandran, S. A. Mousa and W. F. DeGrado, *Proteins: Struct., Funct., Genet.*, 1992, **14**, 509–515; W. L. DeLano, M. H. Ultsch, A. M. de Vos and J. A. Wells, *Science*, 2000, **287**, 1279–1283.
4 C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnon and S. J. Benkovic, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 13638–13643.
5 S. W. Millward, S. Fiacco, R. J. Austin and R. W. Roberts, *ACS Chem. Biol.*, 2007, **2**, 625–634; T. Kawakami, A. Ohta, M. Ohuchi, H. Ashigai, H. Murakami and H. Suga, *Nat. Chem. Biol.*, 2009, **5**, 888–890; C. Heinis, T. Rutherford, S. Freund and G. Winter, *Nat. Chem. Biol.*, 2009, **5**, 502–507.
6 J. M. Smith, F. Vitali, S. A. Archer and R. Fasan, *Angew. Chem., Int. Ed.*, 2011, **50**, 5075–5080.
7 L. Wang, Z. Zhang, A. Brock and P. G. Schultz, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 56–61.
8 J. Kalia and R. T. Raines, *Angew. Chem., Int. Ed.*, 2008, **47**, 7523–7526.
9 P. E. Dawson, T. W. Muir, I. Clark-Lewis and S. B. Kent, *Science*, 1994, **266**, 776–779; J. Offer and P. E. Dawson, *Org. Lett.*, 2000, **2**, 23–26; D. L. J. Clive, S. Hisaindee and D. M. Coltart, *J. Org. Chem.*, 2003, **68**, 9247–9254; G. Chen, J. D. Warren, J. H. Chen, B. Wu, Q. Wan and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2006, **128**, 7460–7462.
10 L. E. Canne, S. J. Bark and S. B. H. Kent, *J. Am. Chem. Soc.*, 1996, **118**, 5891–5896.
11 T. W. Muir, D. Sondhi and P. A. Cole, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 6705–6710; T. C. Evans, J. Benner and M. Q. Xu, *Protein Sci.*, 1998, **7**, 2256–2264.
12 J. A. Camarero and T. W. Muir, *J. Am. Chem. Soc.*, 1999, **121**, 5597–5598.
13 J. P. Danehy and C. J. Noel, *J. Am. Chem. Soc.*, 1960, **82**, 2511–2515.
14 M. M. Kreevoy, E. T. Harper, R. E. Duvall, H. S. Wilgus and L. T. Ditsch, *J. Am. Chem. Soc.*, 1960, **82**, 4899–4902.
15 S. Chattopadhaya, F. B. Abu Bakar and S. Q. Yao, *Methods Enzymol.*, 2009, **462**, 195–223.

Electronic Supplementary Material (ESI) for Chemical Communications

Diverse Organo-Peptide Macrocycles via a Fast and Catalyst-Free Oxime/Intein-Mediated Dual Ligation

Maragani Satyanarayana, Francesca Vitali, John R. Frost, and Rudi Fasan*

Departments of Chemistry, University of Rochester, Rochester, New York 14627

*fasan@chem.rochester.edu

Figure S1. Representative SDS-PAGE protein gel illustrating the differential amount of precursor protein splicing (CBD5(pAcF)) observed at 2 and 5 hours after addition of the synthetic precursor **1** (lanes 2-3), **2** (lanes 4-5), or **3** (lanes 6-7).

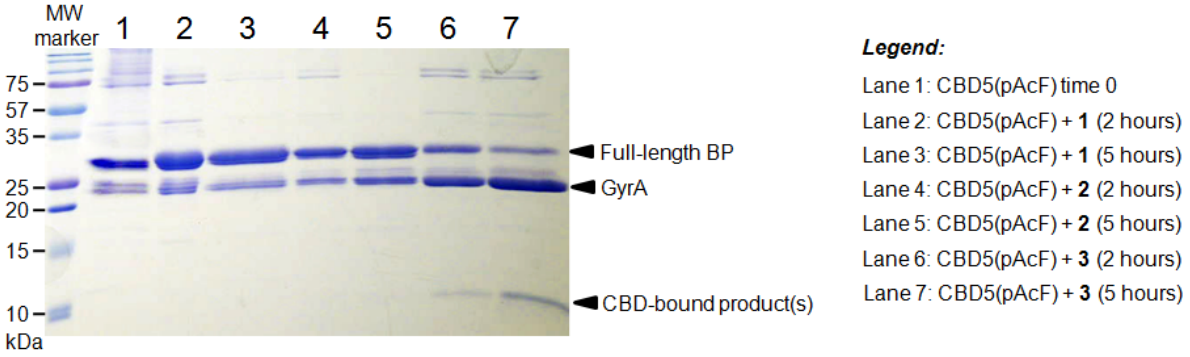


Figure S2. MALDI-TOF spectra of the small molecular weight products observed in the reactions (5 hours) between synthetic precursor **1** and biosynthetic precursor CBD4(pAcF) to CBD12(pAcF). Peaks corresponding to the fragment produced by spontaneous hydrolysis of the biosynthetic precursor ('h') and to the acyclic SP-containing product are indicated ('a').

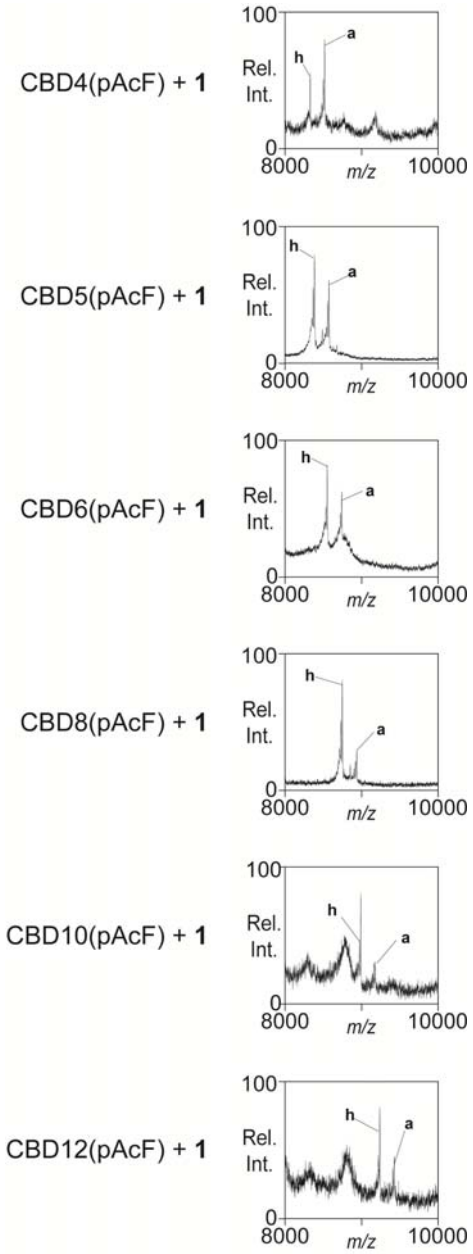


Figure S3. MALDI-TOF spectra of the small molecular weight products observed in the reactions (5 hours) between synthetic precursor **2** and biosynthetic precursor CBD4(pAcF) to CBD12(pAcF). Peaks corresponding to the fragment produced by spontaneous hydrolysis of the biosynthetic precursor ('h') and to the acyclic SP-containing product are indicated ('a').

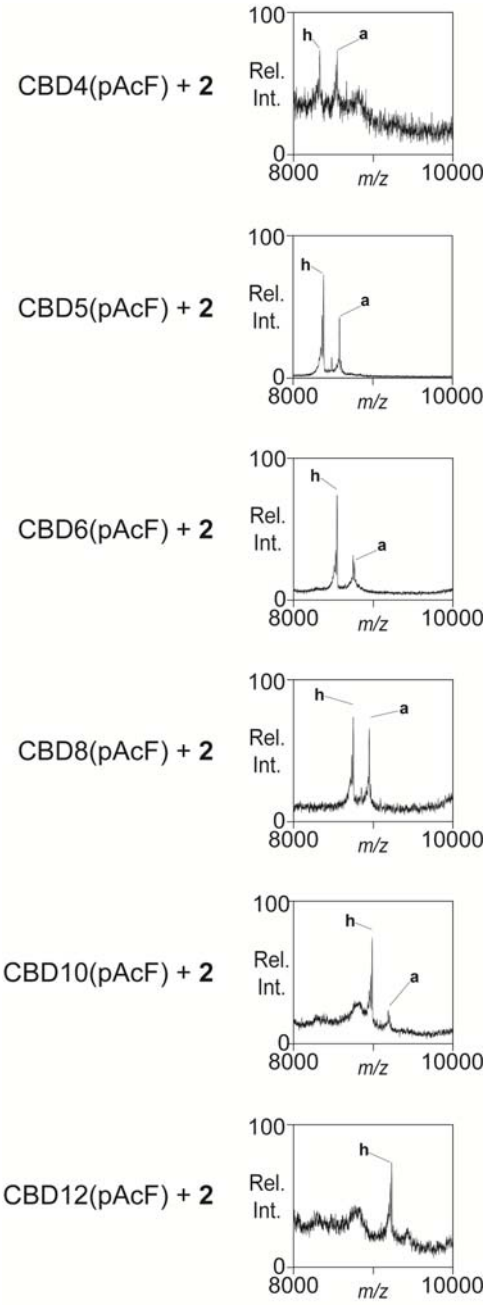


Figure S4. MALDI-TOF spectra of the MOrPHs obtained from the reactions between **3** and CBD5(pAcF) or CBD8(pAcF) before and after treatment with the thiol-alkylating agent iodoacetamide (ICH_2CONH_2). 'm' corresponds to the MOrPH product.

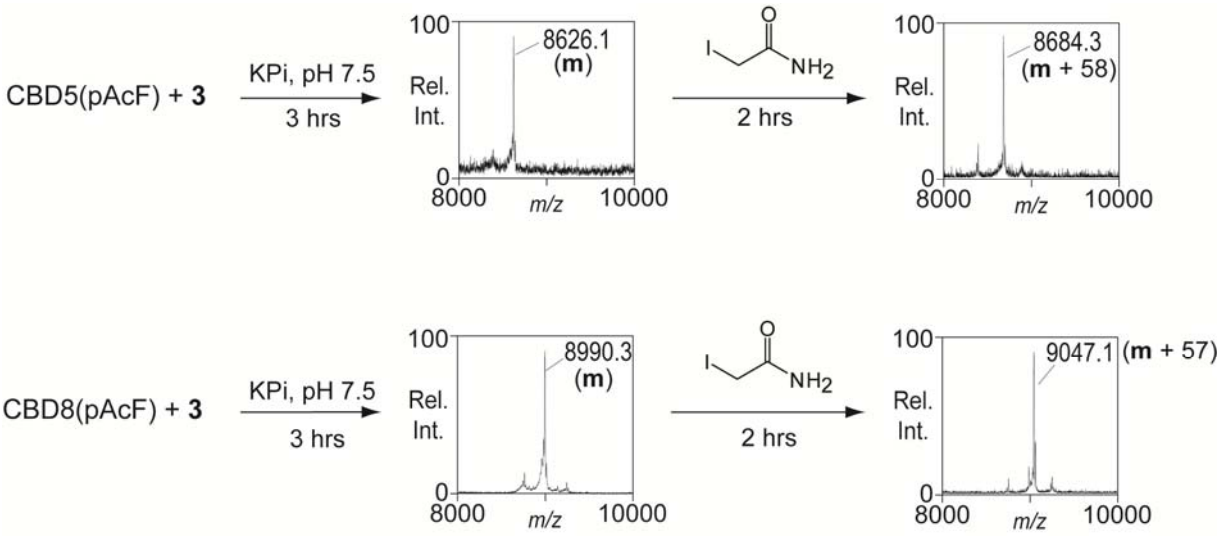


Table S1. Name and composition of the biosynthetic precursors investigated in this study. CBD correspond to the 71-amino acid Chitin Binding Domain (CBD) of chitinase A1 from *Bacillus circulans*[1]. GyrA corresponds to the GyrA intein from *Mycobacterium xenopi*[2].

Name	**Composition**
CBD4(pAcF)	CBD(pAcF)TGST-GyrA
CBD5(pAcF)	CBD(pAcF)TGSGT-GyrA
CBD6(pAcF)	CBD(pAcF)TGSYGT-GyrA
CBD8(pAcF)	CBD(pAcF)TGSAEYGT-GyrA
CBD10(pAcF)	CBD(pAcF)TGSKLAEYGT-GyrA
CBD12(pAcF)	CBD(pAcF)TGSWGKLAEYGT-GyrA

Table S2. Calculated and observed molecular weights of the reaction products between **3** and 18 members of the 5mer biosynthetic precursor library (CBD-(pAcF)-X_4T-GyrA).

5NNK	**CBD(pAcF)X_4T-SBn** observed *m/z*		**MORPH** calculated *m/z*	**MORPH** observed m/z
mutant 1	8680	Macrocycle (m):	8807	8813
		Acyclic (a) :	8825	Not Observed
mutant 2	8553	Macrocycle (m):	8680	8687
		Acyclic (a) :	8698	Not Observed
mutant 3	8683	Macrocycle (m):	8810	8815
		Acyclic (a) :	8828	Not Observed
mutant 4	8747	Macrocycle (m):	8874	8879
		Acyclic (a) :	8892	Not Observed
mutant 5	8637	Macrocycle (m):	8764	8764
		Acyclic (a) :	8782	Not Observed
mutant 6	8681	Macrocycle (m):	8808	8809
		Acyclic (a) :	8826	Not Observed
mutant 7	8612	Macrocycle (m):	8739	8740
		Acyclic (a) :	8757	Not Observed
mutant 8	8659	Macrocycle (m):	8786	8787
		Acyclic (a) :	8804	Not Observed
mutant 9	8674	Macrocycle (m):	8801	8803
		Acyclic (a) :	8819	Not Observed
mutant 10	8564	Macrocycle (m):	8691	8693
		Acyclic (a) :	8709	Not Observed
mutant 11	8554	Macrocycle (m):	8681	8685
		Acyclic (a) :	8699	Not Observed
mutant 12	8646	Macrocycle (m):	8773	8776 (~75%)[a]
		Acyclic (a) :	8791	8794 (~25%)[a]
mutant 13	8535	Macrocycle (m):	8662	8663
		Acyclic (a) :	8680	Not Observed
mutant 14	8610	Macrocycle (m):	8737	8740 (~70%)[a]
		Acyclic (a) :	8755	8757 (~30%)[a]
mutant 15	8678	Macrocycle (m):	8805	8806
		Acyclic (a) :	8823	Not Observed
mutant 16	8662	Macrocycle (m):	8789	8792
		Acyclic (a) :	8807	Not Observed
mutant 17	8606	Macrocycle (m):	8733	8736
		Acyclic (a) :	8751	Not Observed
mutant 18	8733	Macrocycle (m):	8860	8862
		Acyclic (a) :	8878	Not Observed

[a] Estimated based on peak intensity.

Table S3. Calculated and observed molecular weights of the reaction products between **3** and 18 members of the 8mer biosynthetic precursor library (CBD-(pAcF)-X_7T-GyrA).

8NNK	**CBD(pAcF)X_7T-SBn** observed *m/z*		**MORPH** calculated *m/z*	**MORPH** observed m/z
mutant 1	8831	Macrocycle (m): Acyclic (a) :	8958 8976	8960 Not Observed
mutant 2	8875	Macrocycle (m): Acyclic (a) :	9002 9020	9003 Not Observed
mutant 3	9072	Macrocycle (m): Acyclic (a) :	9199 9217	9200 Not Observed
mutant 4	8969	Macrocycle (m): Acyclic (a) :	9096 9114	9099 Not Observed
mutant 5	8967	Macrocycle (m): Acyclic (a) :	9094 9112	9094 Not Observed
mutant 6	8703 (='h')	Macrocycle (m): Acyclic (a) :	8936 8954	Not Observed 8954
mutant 7	9061	Macrocycle (m): Acyclic (a) :	9188 9206	9188 (~80%)[a] 9204 (~20%)[a]
mutant 8	9005	Macrocycle (m): Acyclic (a) :	9132 9150	9135 Not Observed
mutant 9	8935	Macrocycle (m): Acyclic (a) :	9062 9080	9060 Not Observed
mutant 10	8934	Macrocycle (m): Acyclic (a) :	9061 9079	9062 Not Observed
mutant 11	8829	Macrocycle (m): Acyclic (a) :	8956 8974	8957 Not Observed
mutant 12	8648	Macrocycle (m): Acyclic (a) :	8775 8793	8778 Not Observed
mutant 13	8842	Macrocycle (m): Acyclic (a) :	8969 8987	8666 Not Observed
mutant 14	9040	Macrocycle (m): Acyclic (a) :	9167 9185	9168 Not Observed
mutant 15	8841	Macrocycle (m): Acyclic (a) :	8968 8986	8972 Not Observed
mutant 16	9130	Macrocycle (m): Acyclic (a) :	9257 9275	9260 Not Observed
mutant 17	8971	Macrocycle (m): Acyclic (a) :	9098 9116	9100 Not Observed
mutant 18	9078	Macrocycle (m): Acyclic (a) :	9205 9223	9207 Not Observed

[a] Estimated based on peak intensity.

Materials, Methods, and Experimental Procedures

Reagents and Analytical Methods. Chemical reagents, substrates, and solvents were purchased from Sigma-Aldrich, Acros Organics, and Fluka. Silica gel chromatography purifications were carried out using AMD Silica Gel 60 230-400 mesh. ^{1}H NMR spectra and ^{13}C NMR spectra were recorded on Bruker Avance spectrometers using solvent peaks as reference. Data for ^{1}H NMR are reported in the conventional form: chemical shift (δ ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad), coupling constant (Hz), integration. Data for ^{13}C NMR are reported in the terms of chemical shift (δ ppm). LC-MS analyses were performed on a Thermo Scientific LTQ Velos ESI/ion-trap mass spectrometer coupled to an Accela U-HPLC. Small molecule samples were diluted in acetonitrile/water mixtures and analyzed using a Thermo Scientific Hypersil GOLD C_{18} column (50 x 2.1 mm, 1.9 μm) at a flow rate of 0.5 mL/min with a gradient of 5% to 85% acetonitrile (+ 0.1% formic acid) in water (+ 0.1% formic acid). MALDI-TOF spectra were acquired on a Bruker Autoflex III MALDI-TOF spectrometer using a stainless steel MALDI plate and sinapinic acid as matrix.

Cloning and plasmid construction. Plasmids for the expression of the Chitin Binding Domain-containing biosynthetic precursors are derived from vector pET22b(+) (Novagen) and contain a gene encoding for CBD of chitinase A1 from *Bacillus circulans*[1], followed by an amber stop codon TAG, followed by the target sequence of variable length (from 4 to 12 amino acids), followed by the gene encoding for GyrA intein from *Mycobacterium xenopi*[2]. The 5mer and 8mer libraries of biosynthetic precursors were prepared by PCR amplifying randomized target sequences fused to the GyrA intein using *4mer_for*: 5'-CTGCGCCATGGCTAGNNKNNKNNKNNKACCTGCA-TCACGGGAGATGC-3' and

8mer_for: 5'-CTGCGCCATGGCTAGNNKNNKNNKNNKNNKNNKNNKACCTGCATCAC-GGGAGATGC-3' as forward primers, respectively, *T7_term_rev*: 5'-GCTAGTTATTGCTCAGCGG-3' as reverse primer, and pCBD5 as template. The PCR products (0.75 Kbp) were digested with *Nco* I and *Xho* I and cloned into *Nco* I/*Xho* I cassette of pCBD5 vector. In all the pCBD vectors, the genes encoding for the biosynthetic precursor are under the control of an IPTG-inducible T7 promoter. DNA sequencing was performed at the Functional Genomics Center of the University of Rochester. The plasmid pEVOL_pAcF encoding for the engineered $tRNA_{CUA}$ (*Mj*tRNACUA) and aminoacyl-tRNA synthetase (*Mj*TyrRS) from *Methanococcus jannaschii* for amber codon suppression with *para*-acetyl-phenylalanine (pAcF) was provided by the Schultz group and it has been described previously.[3]

Protein Expression and Purification. To express the pAcF-containing biosynthetic precursors, each of the pCBD vectors was co-transformed with pEVOL_pAcF into BL21(*DE3*) *E. coli* cells. Overnight cultures were grown in Luria-Bertani (LB) media containing ampicillin (50 mg L^{-1}) and chloramphenicol (34 mg L^{-1}) and used to inoculate a 0.4 L M9 media containing ampicillin (50 mg L^{-1}) and chloramphenicol (34 mg L^{-1}) and supplemented with 1% glycerol. Cultures were grown at 37° C until an OD_{600} of 0.6 was reached, at which point protein expression was induced by adding 0.05% L-arabinose, 0.25 mM IPTG, and 1 mM pAcF. Cultures were grown for an additional 16 hours at 25° C and harvested by centrifugation at 4,000 rpm. Frozen cell pellets were resuspended in 50 mM Tris, 300 mM NaCl, 20 mM imidazole buffer (pH 7.4) and the cells lysed through sonication. After centrifugation at 14,000 rpm, the cell lysate was loaded onto a Ni-NTA affinity column and the protein eluted with 50 mM Tris, 150 mM NaCl, 300 mM imidazole (pH 7.4). After pooling and concentration of the protein-containing fractions, the

buffer was exchanged with potassium phosphate 50 mM, NaCl 150 mM buffer (pH 7.5) and aliquots of the protein solutions stored at -80°C. Protein concentration was determined using the extinction coefficient at 280 nm (ε_{280}) calculated based on the protein primary sequence. Typical expression yields for the pAcF-containing biosynthetic precursors were 25-35 mg L^{-1} culture. The identity of the isolated proteins was confirmed by MALDI-TOF.

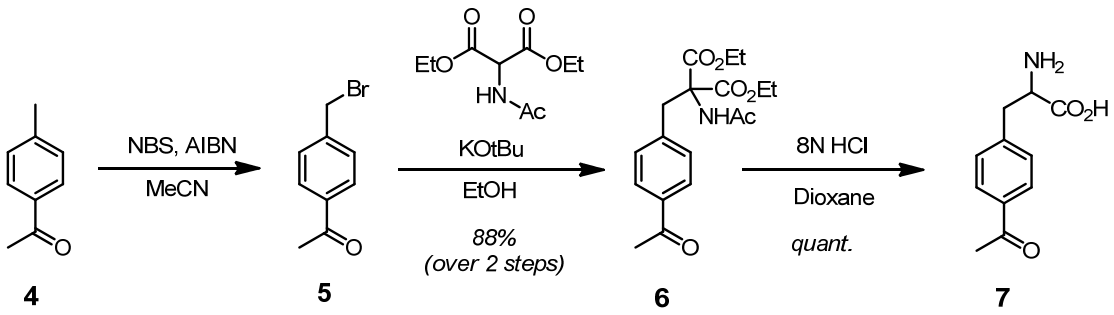
Macrocyclization reactions. Reactions were carried out at 20 μL scale by adding synthetic precursor **1**, **2**, or **3** (final concentration: 15 mM) to a solution of protein precursor (100 μM) in potassium phosphate buffer (50 mM, NaCl 150 mM, pH 7.5) in the presence of TCEP (final concentration: 20 mM). For SDS-PAGE analyses, 5 μl of the reaction mixture were removed at the indicated time point(s), diluted in DTT-free 4x loading buffer, and analysed 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 time zero and at the time of the analysis. MALDI-TOF analyses of the small molecular weight products (8-10 kDa) of the reactions were carry out on a Bruker Autoflex III MALDI-TOF spectrometer. Prior to analysis, protein samples were diluted in 50% acetonitrile in H_2O (0.1% TFA) and this solution mixed with a sinapinic acid solution (10 mg/mL in 50% acetonitrile in H_2O with 0.1% TFA). The samples were analyzed using reflectron positive (RP) mode and calibration using small molecular weight (2-15 kDa) protein standards.

5mer and 8mer BP library expression and reaction analysis. From the transformation plates, 18 randomly chosen colonies were used to inoculate 5 mL of fresh LB medium containing

ampicillin (50 mg L^{-1}). The respective pBP plasmid were extracted from overnight cultures and used to transform BL21(DE3) cells containing pEVOL_pAcF. The corresponding precursor proteins were produced and isolated from 50 mL-cultures according to the protocols provided above. Macrocyclization reactions were carried out in phosphate buffer (50 mM, NaCl 150 mM, pH 7.5) by mixing the precursor protein (100 μM) with compound **3** (15 mM) in the presence of TCEP (20 mM). The molecular weight of the expected macrocyclic product from each reaction was calculated based on the molecular weight corresponding to the benzyl thioester obtained from parallel reactions with the same protein and benzyl mercaptan (4 hours, room temperature). The reliability of this method was validated by DNA sequencing of 5 variants from each library.

Synthetic Procedures

Scheme S1. *Synthesis of p-acetyl-phenylalanine (pAcF).*



Para-acetylphenylalanine (**7**) was synthesized through an optimized version of a reported protocol[4].

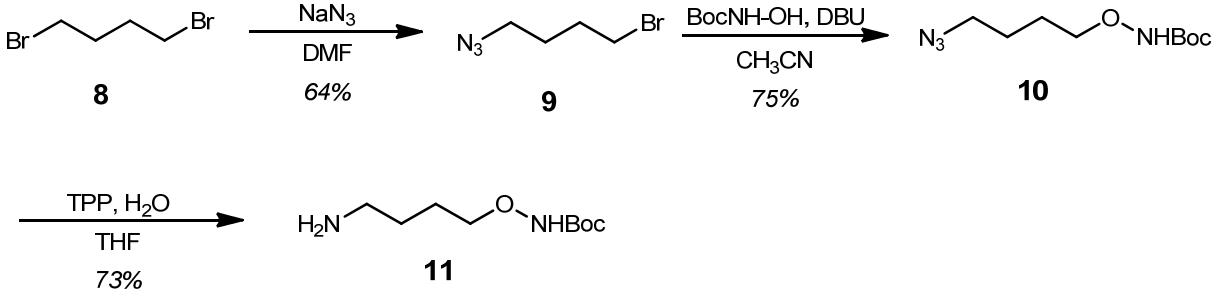
4-bromomethyl-acetophenone (**5**). 4-methyl acetophenone (**4**) (10 mL, 74.5 mmol) was dissolved in 80 mL anhydrous acetonitrile in a dry flask under argon. *N*-bromosuccinimide (NBS) (14.6 g, 82 mmol), freshly re-crystallized from water, was added to the solution of **4** followed by addition of azo-*bis*-isobutyronitrile (AIBN) (1.23 g, 7.49 mmol). The reaction mixture was heated at reflux for 1.5 hour and then cooled to room temperature. Volatiles were removed *in vacuo* and the resulting oil was re-dissolved in 500 mL dichloromethane and washed once with 1 M HCl, twice with a saturated solution of $NaHCO_3$, and once with a saturated solution of NaCl. The organic layer was dried over $MgSO_4$ and filtered. Volatiles were removed to yield **5** as a yellow oil (16.9 g) which was carried on to the next step without further purification.

Diethyl 2-acetamido-2-(4-acetylbenzyl)malonate (**6**). Compound **5** (4.3 g, 20.2 mmol) was dissolved in 200 mL anhydrous ethanol in a dry flask under argon. The solution was added with diethylacetamidomalonate (4.82 g, 27.2 mmol) followed by 1 M potassium *tert*-butoxide in *tert*-butanol (24.25 mL, 24.2 mmol). The reaction mixture was stirred at reflux for 24 hours and then

cooled to room temperature. The reaction was concentrated *in vacuo* to about 20 mL and then diluted with 50 mL cold diethylether. The off white precipitate was collected by filtration and the procedure was repeated on the filtrate solution. The combined solids were dried *in vacuo* to yield **6** as a white solid (6.24 g, 88.3%).

***p*-acetylphenylalanine** (**7**). Compound **6** (1.8 g, 5.2 mmol) was dissolved in 8 N HCl in dioxane and the reaction mixture was heated to reflux for 8 hour and then cooled to room temperature. Volatiles were removed *in vacuo* yielding *p*-acetylphenylalanine (**7**) as a light brown solid (1.25 g, quant.) ^{1}H NMR (CD_3OD, 400MHz): δ 2.60 (s, 3H), 3.27 (m, 1H), 3.4 (m, 1H), 4.33 (dd, J = 7.2 Hz, 1H), 7.45 (d, J = 8 Hz, 2H), 7.99 (d, J = 8 Hz, 2H). ^{13}C NMR (100 MHz, CD_3OD): δ 26.7, 37.2, 54.7, 130.1, 130.9, 137.8, 141.4, 170.9, 200.0. MS (ESI) calcd for $C_{11}H_{13}NO_3$ $[M+H]^+$: *m/z* 208.1, found: 208.1

Scheme S2. *Synthesis of Boc protected alkoxyamino linker.*



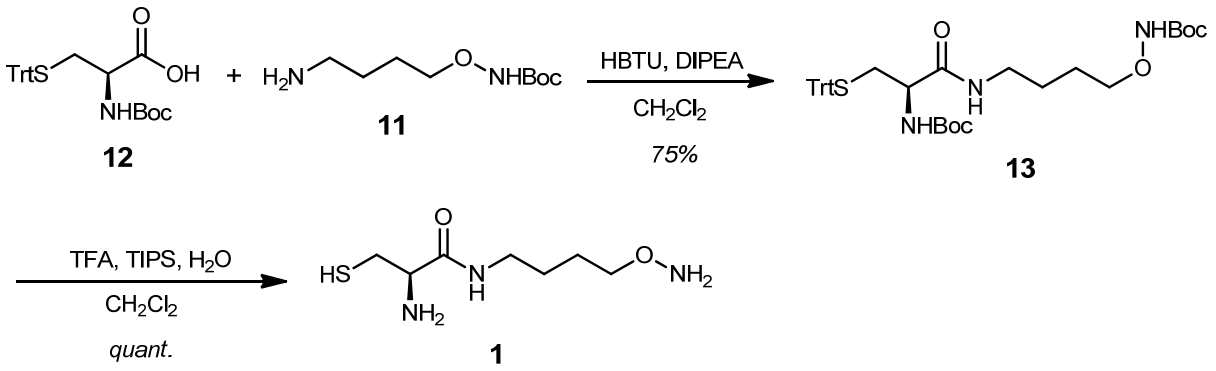
1-azido-4-bromobutane (9): To a solution of 1,4-dibromobutane (**8**) (8.0 g, 37.03 mmol), in DMF (60 mL), sodium azide (2.16 g, 33.33 mmol) was added and the mixture was stirred overnight at 50 °C. Ice was then added to the mixture followed by extraction with ethyl acetate. The organic layer was washed with water, then brine, and then dried over Na_2SO_4. The crude residue obtained after solvent evaporation was purified by silica gel flash chromatography (100%

hexanes) to give **9** (4.2 g, 64%) as a clear oil. ^{1}H NMR (400 MHz, $CDCl_3$): δ 1.68-1.77 (m, 2H), 1.88-1.96 (m, 2H), 3.30 (t, *J* = 6.6 Hz, 2H), 3.39 (t, *J* = 6.6 Hz, 2H); ^{13}C NMR (100 MHz, $CDCl_3$): δ 27.4, 29.7, 32.9, 50.5.

***tert*-butyl-4-azidobutoxycarbamate (10)**: 1-azido-4-bromobutane (**9**) (2.22 g, 12.47 mmol) was dissolved in acetonitrile (40 ml), and *N*-Boc-hydroxylamine (2.48 g, 18.7 mmol) and DBU (3.4 ml, 24.94 mmol) added to the solution under argon atmosphere at room temperature. The resulting mixture was heated to 50°C and stirred for 24 h. Water (100 ml) was then added, and the product extracted ethyl acetate. The organic layer was washed with brine, dried (Na_2SO_4), and concentrated under reduced pressure. The resulting oil was purified by silica gel flash chromatography with hexanes : ethyl acetate (95:5) to yield **10** as a clear oil (2.1 g, 75 % yield). ^{1}H NMR (400 MHz, $CDCl_3$): δ 1.44 (s, 9H), 1.64-1.71 (m, 4H), 3.26-3.31 (m, 2H), 3.81-3.85 (m, 2H); ^{13}C NMR (100 MHz, $CDCl_3$): δ 25.1, 25.5, 27.8, 28.1, 28.2, 51.0, 75.9, 81.6, 156.99.

***tert*-butyl-4-aminobutoxycarbamate (11)**: To a solution of *t*-butyl-4-azidobutoxycarbamate **10** (2.0 g, 8.695 mmol) in THF (40 mL), triphenylphosphine (2.96 g, 11.3 mmol) and water (1 mL) were added. The reaction mixture was stirred at room temperature for 12 hours. After removing THF under reduced pressure, the residue was purified by silica gel flash chromatography using dichloromethane : methanol (85:15) to afford **11** as a pale yellow liquid (1.29 g, 73 % yield). ^{1}H NMR ($CDCl_3$, 400 MHz): δ 1.21-1.37 (br, 2H), 1.45 (s, 9H), 1.46-1.55 (m, 2H), 1.60-1.68 (m, 2H), 2.70 (t, *J* = 6.57 Hz, 2H), 3.84 (t, *J* = 6.57 Hz, 2H); ^{13}C NMR (100 MHz, $CDCl_3$): δ 25.3, 27.9, 28.1, 28.2, 29.8, 45.6, 76.3, 81.2, 157.0 .

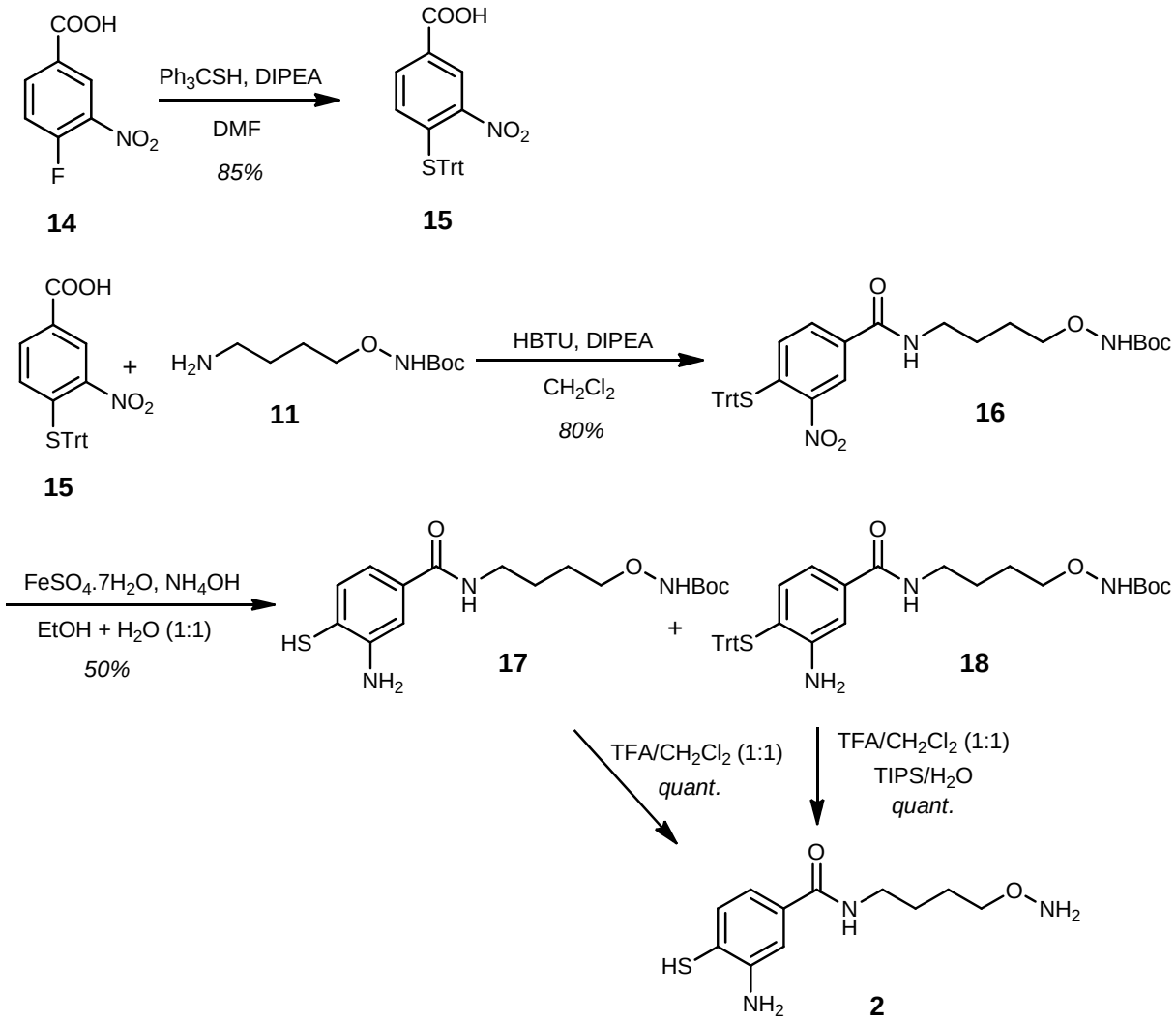
Scheme S3. *Synthesis of Synthetic Precursor SP1* (**1**).



(*R*)-2-amino-N-(4-(aminooxy)butyl)-3-mercaptopropanamide (1): *S*-trityl-*N*-Boc-cysteine (1.0 g, 2.16 mmol) was dissolved in dichloromethane and *t*-butyl-4-aminobutoxycarbamate **11** (0.484g, 2.37 mmol), *O*-benzotriazole-*N*,*N*,*N*',*N*'-tetramethyl-uronium-hexafluoro-phosphate (HBTU) (1.22 g), and diisopropylethyl amine (0.937 mL, 5.39 mmol) were added to the solution at room temperature under argon. The mixture was stirred for 3 hours. The reaction mixture was diluted with water and extracted with dichloromethane. The organic layer was washed with brine and evaporated under reduced pressure. The residue was purified by silica gel flash chromatography using a gradient of 30 to 35% ethyl acetate in hexanes to give **13** (0.53g, 75% yield). ^{1}H NMR ($CDCl_3$, 400 MHz): δ 1.38 (s, 9H), 1.44 (s, 9H), 1.55-1.58 (m, 4H), 2.48-2.55 (m, 1H), 2.60-2.70 (m, 1H), 3.17-3.26 (m, 2H), 3.77-3.88 (m, 3H), 4.82-4.90 (m, 1H), 6.25-6.32 (m, 1H), 7.17-7.30 (m, 9H), 7.35-7.50 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ 25.3, 26.1, 28.0, 28.2, 28.3, 28.5, 34.0, 38.6, 39.1, 67.0, 76.2, 80.2, 81.5, 126.8, 128.0, 128.3, 129.5, 129.6, 129.89, 144.4, 155.3, 157.0, 165.7, 170.4. Compound **13** (1.0 g, 1.55 mmol) was dissolved in 50% TFA in dichloromethane (16 mL) in the presence of triisopropylsilane (0.64 mL, 3.0 mmol) and water (0.5 mL) at 0 ^{o}C. The reaction mixture was stirred for an hour in ice and completion of the deprotection was monitored by thin layer chromatography. Volatiles were then removed by evaporation and the residue washed with hexanes to yield **1**. ^{1}H NMR (400 MHz, CD_3OD):

δ=1.76-1.60 (m, 4H), 2.92 (dd, *J*= 14.8 Hz, 6.8 Hz, 1H), 3.01 (dd, *J*= 14.4 Hz, 5.2 Hz, 1H), 3.21-3.33 (m, 2H), 3.98 (t, *J*= 6.4 Hz, 1H), 4.04 (t, *J*= 5.6 Hz, 2H). ^{13}C NMR (100 MHz, CD_3OD): δ = 24.70, 25.13, 37.48, 38.69, 54.73, 74.39, 167.03. MS (ESI) Calcd. for $C_7H_{17}N_3O_2S$ $[M+H]^+$: *m/z* 208.3, found: 208.4.

Scheme S4. *Synthesis of Synthetic Precursor SP2* (**2**).



3-nitro-4-(tritylthio)benzoic acid (15): 4-fluoro-3-nitro-benzoic acid (**14**) (2.0 g, 10.81 mmol) and triphenylmethylmercaptan (3.87 g, 14.05 mmol) were dissolved in dry DMF (50 mL) under

argon. To this solution diisopropylethylamine (DIPEA) (3.94 mL, 22.68 mL) was added dropwise and the reaction mixture was stirred for 36 h at room temperature. The reaction mixture was diluted with ice-cold water and extracted with ethyl acetate. After evaporation of the solvent under reduced pressure, 4-thiotrityl-3-nitro-benzoic acid **15** was obtained by crystallization from methanol (4.0 g, 85% yield). ^{1}H NMR (DMSO-D6, 400 MHz): δ 7.01 (d, *J* = 7.0 Hz, 1H), 7.21-7.34 (m, 15H), 7.57 (d, *J* = 8.0 Hz, 1H), 8.34 (bd, *J* = 1.9 Hz, 1H); ^{13}C NMR (100 MHz, $CDCl_3$): δ 71.7, 125.92, 128.0, 128.5, 129.9, 131.4, 132.1, 139.4, 142.54, 145.16, 147.5, 165.3. MS (ESI) Calcd. for $C_{26}H_{19}NO_4S$ $[M+H]^+$: *m/z* 442.5, found: 442.6.

***tert*-butyl-4-(3-nitro-4-(tritylthio)benzamido)butoxycarbamate (16)**: 3-nitro-4-(tritylthio)-benzoic acid (**15**) (1.0 g, 2.26 mmol) was dissolved in dichloromethane (30 mL). To this solution *t*-butyl-4-aminobutoxycarbamate **11** (0.508 g, 2.5 mmol), HBTU (1.28 g, 3.39 mmol), and DIPEA (0.98 mL, 5.65 mmol) were added under argon. The reaction mixture was stirred for 3 hours at room temperature. The reaction mixture was diluted with dichloromethane (30 mL) and washed with water and brine. After evaporation of the organic layer, the residue was purified by silica gel flash chromatography using 35% ethyl acetate in hexanes to give **16** in 70% yield. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 1.37 (s, 9H), 1.61-1.73 (m, 4H), 3.41 (q, *J* = 6.0, 11.8 Hz, 1H), 3.84 (t, *J* = 5.4 Hz, 2H), 7.01 (d, *J* = 8.6 Hz, 1H), 7.10 - 7.28 (m, 9H), 7.34-7.39 (m, 5H), 7.43 (d, *J* = 1.8 Hz, 1H), 7.46 (s, 1H), 8.38 (d, *J* = 2.0 Hz, 1H); ^{13}C NMR (75 MHz, $CDCl_3$): δ 25.5, 26.3, 28.1, 38.6, 39.7, 71.7, 81.7, 123.8, 127.5, 128.1, 130.1, 131.0, 131.3, 139.5, 142.6, 146.9, 157.2, 164.9, 165.7. MS (ESI) Calcd. for $C_{35}H_{37}N_3O_6S$ $[M+H]^+$: *m/z* 628.7, found: 628.4.

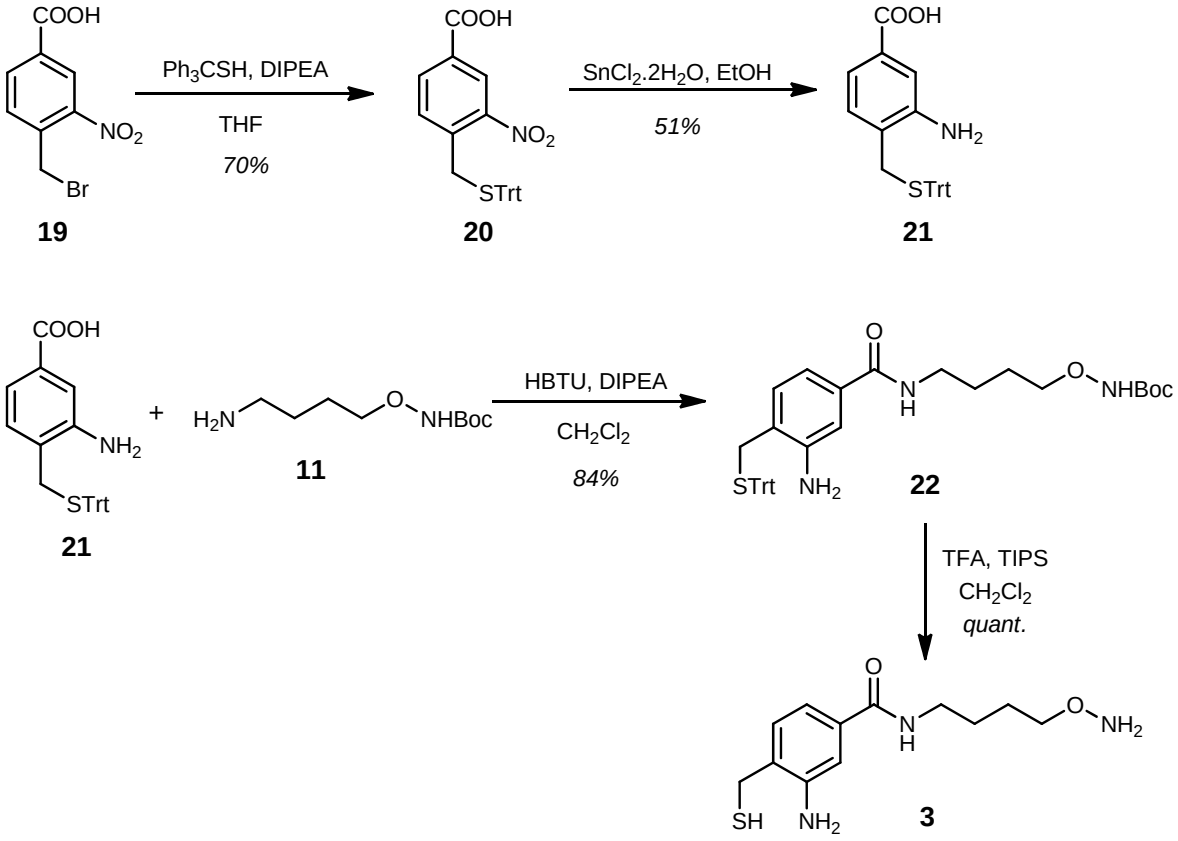
***tert*-butyl-4-(3-amino-4-mercaptobenzamido)butoxycarbamate (17) and *tert*-butyl-4-(3-amino-4-(tritylthio)benzamido)butoxycarbamate (18)**: *t*-butyl-4-(3-nitro-4-(tritylthio)benzamido)-butoxycarbamate **16** (0.69 g, 1.10 mmol) and $FeSO_4.7H_2O$ (3.06 g, 11.0

mmol) were dissolved in 32 mL ethanol : water (1:1). The mixture was heated to 78°C for 30 minutes, and then added with ammonia (3.2 mL). The mixture was stirred at 78°C for two hours, then diluted with water (30 mL), and extracted with ethyl acetate (3 x 30 mL). The combined organic layers were washed with brine and evaporated under reduced pressure. The residue, which contained a mixture of **17** as disulfide dimer and **18** was further purified by silica gel flash chromatography. **17** was eluted with 35% ethyl acetate in hexanes, while **18** was eluted with 80% ethyl acetate in hexanes. Spectral data for **17**: ^{1}H NMR ($CDCl_3$, 400 MHz): δ 1.44 (S, 9H), 1.64-1.66 (m, 4H), 3.34 (m, 2H), 3.80 (m, 2H), 6.85 (m, 1H), 7.10 (m, 1H), 7.15 (m, 1H); ^{13}C NMR (75 MHz, $CDCl_3$): 25.1, 25.6, 27.1, 39.1, 75.3, 80.6, 113.5, 114.7, 120.4, 135.7, 137.0, 149.3, 157.7, 168.7. MS (ESI) Calcd. for disulfide dimer $C_{32}H_{48}N_6O_8S_2$ $[M+H]^+$: *m/z* 709.9, found: 709.7. Spectral data for **18**: ^{1}H NMR ($CDCl_3$, 400 MHz): δ 1.42 (S, 9H), 1.58-1.70 (m, 4H), 3.34-3.41 (m, 2H), 3.80-3.85 (m, 2H), 6.53-6.58 (m, 1H), 6.66-6.71 (m, 1H), 6.91-6.97 (m, 2h), 7.12-7.22 (m, 9H), 7.28-7.35 (m, 5H), 7.50 (s, 1H). MS (ESI) Calcd. for $C_{35}H_{39}N_3O_4S$ $[M+H]^+$: *m/z* 598.7, found: 598.7.

3-amino-*N*-(4-(aminooxy)butyl)-4-mercaptobenzamide (2): Compound **2** was prepared from deprotection of **17** and **18** under similar conditions. Compound **17** (0.22 g, 0.619 mmol) was dissolved in dichloromethane (4 mL) and the mixture placed at 0°C under argon atmosphere. 4 mL TFA were added to the mixture dropwise followed by stirring for an hour. Completion of the deprotection reaction was monitored by thin layer chromatography (20% methanol in dichloromethane). Volatile solvents were removed under reduced pressure first followed by evaporation under high vacuum to yield **2** as a disulfide dimer. MS (ESI) Calcd. for $C_{22}H_{32}N_6O_4S_2$ $(M+H)^+$: *m/z* 509.6, found: 509.7. Compound **18** (0.1 g, 0.167 mmol) was dissolved in 50% TFA in dichloromethane (6 mL) containing triisopropylsilane (69 μL, 0.335

mmol) and water (50 μL). The reaction mixture was stirred for an hour in ice and completion of the deprotection was monitored by thin layer chromatography. Volatiles were then removed by evaporation and the residue washed with hexanes to yield **2** as a disulfide dimer. ^{1}H NMR (CD_3OD, 500 MHz): δ 1.5-1.55 (m, 4H), 3.26 (t, *J*= 5.95 Hz, 2H), 3.59 (t, *J*= 5.5 Hz, 2H), 6.70 (d, *J*=8 Hz, 1H), 7.01 (d, *J*= 8 Hz, 1H), 7.07 (s, 1H). ^{13}C NMR (125 MHz, CD_3OD): δ= 25.36, 25.57, 39.24, 75.09, 113.45, 114.87, 120.45, 135.56, 137.19, 149.33, 168.56. MS (ESI) Calcd. for $C_{22}H_{32}N_6O_4S_2$ $(M+H)^+$: *m/z* 509.6, found: 509.4.

Scheme S5. *Synthesis of Synthetic Precursor SP3* (**3**).



3-nitro-4-((tritylthio)methyl)benzoic acid (20): 4-bromomethyl-3-nitrobenzoic acid **19** (2 g, 7.69 mmol) was dissolved in tetrahydrofuran (40 mL) and the solution added with trityl

mercaptan (2.34 g, 8.46 mmol) and diisopropyl ethylamine (2.8 mL, 16.14 mmol) at room temperature. The reaction mixture was stirred for 36 hours and then was quenched with a saturated solution of ammonium chloride (4 mL) followed by extraction with ethyl acetate. The organic layer was washed with brine and the solvent evaporated under reduced pressure. The residue was purified by flash chromatography using 7.5% methanol in dichloromethane to yield **20** (2.5 g, 70% yield). ^{1}H NMR ($CDCl_3$, 400 MHz): δ 3.80 (s, 2H), 6.80-6.90 (m, 1H), 7.15-7.50 (m, 15 H), 7.95-8.05 (m, 1H), 8.55 (s, 1H); ^{13}C NMR (100 MHz, $CDCl_3$): δ 33.8, 68.1, 126.1, 126.5, 127.0, 128.1, 129.0, 129.6, 129.7, 130.0, 132.93, 133.5, 133.6, 139.1, 139.3, 144.0, 148.5, 163.7. MS (ESI) calcd. for $C_{27}H_{21}NO_4S$ $(M+H)^+$: *m/z* 456.1 found : 456.4

3-amino-4-((tritylthio)methyl)benzoic acid (21): 3-nitro-4-((tritylthio)methyl)benzoic acid **20** (1.86 g, 3.428 mmol) was dissolved in methanol (40 mL) and the solution added with $SnCl_2.2H_2O$ (3.85 g, 17.14 mmol). The reaction mixture was heated to 65 ^{o}C and stirred for 2 hours. The solvent was evaporated and the residue added with saturated sodium bicarbonate solution to reach pH 6 followed by extraction with ethyl acetate. The organic layer was washed with brine and evaporated under reduced pressure. The residue purified by flash chromatography using 35% ethyl acetate in hexanes to yield **21** (0.746 g, 51% yield). ^{1}H NMR ($CDCl_3$, 400 MHz): δ 3.34 (s, 2H), 7.11-7.13 (m, 1H), 7.28-7.45 (m, 12 H), 7.55-7.59 (m, 5H). MS (ESI) calcd. for $C_{27}H_{23}NO_2S$ $[M+H]^+$: *m/z* 426.5 found : 426.6.

***tert*-butyl-4-(3-amino-4-((tritylthio)methyl)benzamido)butoxycarbamate (22)** 3-amino-4-((tritylthio)methyl)benzoic acid (**21**) (0.372 g, 0.875 mmol) was dissolved in dichloromethane and the solution added with *t*-butyl-4-aminobutoxycarbamate **11** (0.196 g, 0.962 mmol), HBTU (0.497 g, 1.31 mmol), and DIPEA (0.38 mL, 2.187 mmol) under argon. The reaction mixture was stirred at room temperature. After 3 hours, the reaction mixture was diluted with water and

extracted with dichloromethane (3 x 30 mL). The combined organic layers were concentrated under reduced pressure and the residue was purified by flash column chromatography in 35% ethyl acetate in hexanes to give **22** (0.45g, 84% yield). ^{1}H NMR ($CDCl_3$, 400 MHz): δ 1.44 (s, 9H), 1.65-1.75 (m, 4H), 3.40-3.48 (q, *J* = 6.5, 12.5 Hz, 2H), 3.64 (s, 2H), 3.85 (t, *J* = 5.8 Hz, 2H), 6.34 (t, *J* = 5.9 Hz, 1H), 6.95-7.04 (m, 3H), 7.20-7.26 (m, 5H), 7.28-7.33 (m, 5H), 7.47-5.72 (m, 5H); ^{13}C NMR (75 MHz, $CDCl_3$): δ 25.2, 26.1, 28.3, 33.3, 39.5, 67.2, 81.7, 114.7, 116.6, 123.5, 125.9-131.4(m), 135.2, 144.4, 145.4, 156.8, 167.7. MS (ESI) calcd. for $C_{36}H_{41}N_3O_4S$ $[M+H]^+$: *m/z* 612.8 found : 612.6.

3-amino-*N*-(4-(aminooxy)butyl)-4-(mercaptomethyl)benzamide (**3**). *tert*-butyl-4-(3-amino-4-((tritylthio)methyl)benzamido)butoxycarbamate **22** (47 mg, 0.076 mmol), triisopropylsilane (31 μL, 0.152 mmol) and water (0.1 mL) were dissolved in 50% TFA in dicholoromethane (3 mL) at 0^{o}C under argon. The reaction mixture was stirred for an hour in ice and completion of the deprotection was monitored by thin layer chromatography. Volatiles were then removed by evaporation and the residue washed with hexanes (3 x 20 mL) to yield **3**. NMR (D6-DMSO, 500 MHz): δ= 1.57-1.66 (m, 4H), 3.26 (dt, *J*= 6.45 Hz, 6.15 Hz, 2H), 3.81 (s, 2H), 3.89 (t, *J*= 6.1 Hz, 2H), 6.79 (dd, *J*= 7.75 Hz, 1.14 Hz, 1H), 7.06 (d, *J*= 7.85, 1H), 7.13 (d, *J*= 1.45, 1H), 8.25 (t, *J*= 5.9 Hz, 1H). ^{13}C NMR (125 MHz, D6-DMSO): δ= 25.18, 25.91, 30.43, 74.43, 114.39, 114.62, 122.70, 131.42, 135.47, 146.75, 167.18. MS (ESI) calcd. for $C_{12}H_{19}N_3O_2S$ $C_{11}H_{13}NO_3$ $[M+H]^+$: 270.1, found: 270.2.

References:

(1) Watanabe, T.; Ito, Y.; Yamada, T.; Hashimoto, M.; Sekine, S.; Tanaka, H. *J. Bacteriol.* **1994**, *176*, 4465.

(2) Telenti, A.; Southworth, M.; Alcaide, F.; Daugelat, S.; Jacobs, W. R., Jr.; Perler, F. B. *J. Bacteriol.* **1997**, *179*, 6378.

(3) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. *J. Mol. Biol.* **2010**, *395*, 361.

(4) Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 56; Pieck, J. C.; Kuch, D.; Grolle, F.; Linne, U.; Haas, C.; Carell, T. *J. Am. Chem. Soc.* **2006**, *128*, 1404.