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‡)).

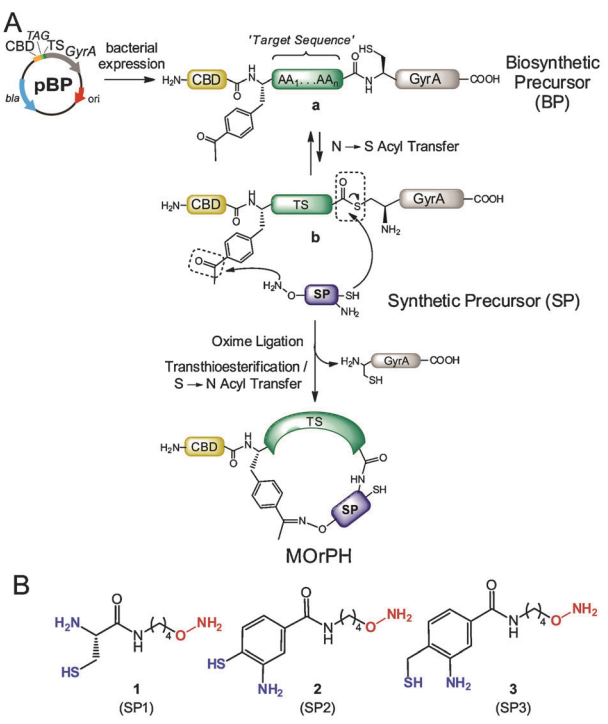

Fig. 1 (A) Synthesis of Macrocyclic Organo-Peptide Hybrids *via* oxime/intein-mediated tandem ligation. CBD: Chitin Binding Domain. TAG: amber stop codon. TS: Target Sequence. GyrA: intein GyrA from *Mycobacterium xenopi*. (B) Oxyamine/amino-thiol synthetic precursors. Their synthesis is described in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

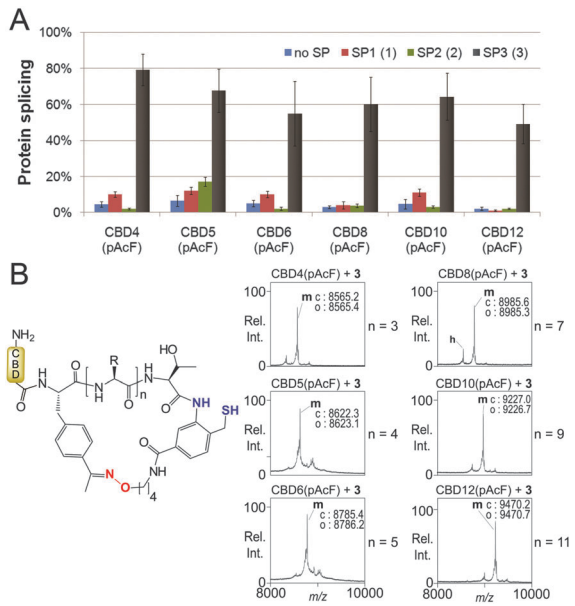

Fig. 2 (A) Percentage of precursor protein splicing in the presence of no SP, **1**, **2**, or **3** after 5 hours as determined by SDS-PAGE. Error bars are from triplicate experiments. Protein constructs are described in Table S1 (ESI‡). (B) MALDI-TOF spectra of MOrPHs obtained from reaction of **3** 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.

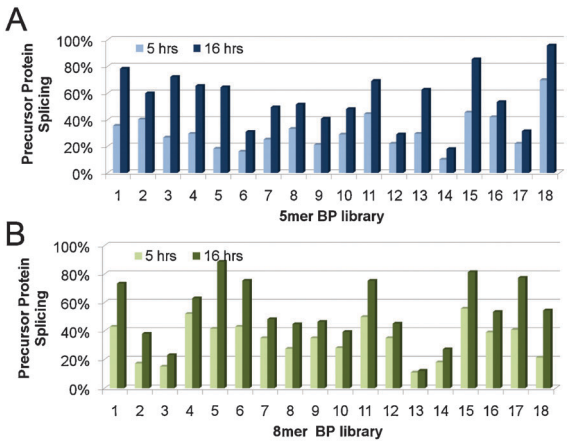

Fig. 3 Extent of **3**-induced protein splicing for 18 variants from the library of precursor proteins with randomized 5mer (A) and 8mer (B) target sequences. See also Tables S2 and S3 in ESI.‡

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.