Organic & Biomolecular Chemistry

PAPER

View Article Online
View Journal | View Issue

Published on 01 April 2016. Downloaded by University of Rochester on 15/06/2016 14:40:11.

Cite this: *Org. Biomol. Chem.*, 2016, **14**, 5803

# Side-chain-to-tail cyclization of ribosomally derived peptides promoted by aryl and alkyl amino-functionalized unnatural amino acids†

John R. Frost,‡ Zhijie Wu,‡ Yick Chong Lam,‡ Andrew E. Owens and Rudi Fasan*

Received 24th January 2016,
Accepted 1st April 2016

DOI: 10.1039/c6ob00192k

www.rsc.org/obc

A strategy for the production of side-chain-to-tail cyclic peptides from ribosomally derived polypeptide precursors is reported. Two genetically encodable unnatural amino acids, bearing either an aryl or alkyl amino group, were investigated for their efficiency toward promoting the formation of medium to large-sized peptide macrocycles *via* intein-mediated side-chain-to-C-terminus cyclization. While only partial cyclization was observed with precursor proteins containing *para*-amino-phenylalanine, efficient peptide macrocyclization could be achieved using *O*-2-aminoethyl-tyrosine as the reactive moiety. Conveniently, the latter was generated upon quantitative, post-translational reduction of the azido-containing counterpart, *O*-2-azidoethyl-tyrosine, directly in *E. coli* cells. This methodology could be successfully applied for the production of a 12 mer cyclic peptide with enhanced binding affinity for the model target protein streptavidin as compared to the acyclic counterpart ($K_D$: 5.1 μM *vs.* 22.4 μM), thus demonstrating its utility toward the creation and investigation of novel, functional macrocyclic peptides.

## Introduction

Owing to their peculiar structural and conformational properties, macrocyclic peptides provide attractive scaffolds for targeting complex biomolecular interfaces.[1–5] The occurrence of macrocyclic backbones amongst biologically active peptides[6] found in nature further motivates the increasing interest toward this structural class.[1–5] As illustrated by a growing body of studies, macrocyclization can help overcome important limitations of linear peptides such as limited proteolytic stability[7–9] and membrane permeability.[10–14] The conformational rigidification imparted by macrocyclization on the peptide structure can also aid in pre-organizing the molecule into a bioactive conformation, resulting in enhanced selectivity and binding affinity for the target biomolecule.[15–18]

While a variety of synthetic methods have been investigated for the synthesis of cyclic peptides,[19] the ability to cyclize ribosomally derived polypeptides can offer key advantages toward the synthesis, combinatorial diversification, and functional evaluation of collections of peptide macrocycles.[20,21] Indeed, the application of these strategies has enabled the successful identification of cyclopeptide inhibitors for a variety of target proteins and enzymes.[22–29]

Our group has reported methods to rapidly generate structurally diverse Macrocyclic Organo Peptide Hybrids (MOrPHs), *via* a dual ligation strategy involving arbitrary non-peptidic linkers and ribosomally produced intein-fusion proteins bearing an unnatural amino acid equipped with a bioorthogonal functional group.[9,30–33] More recently, complementary methodologies have been developed to enable the biosynthesis of cyclic and bicyclic peptides of arbitrary sequence in living bacterial cells.[34–36] As illustrated in these and other studies,[18,37] the type of backbone connectivity and intramolecular linkage can play a significant role in affecting the functional properties (*e.g.*, protein binding affinity) of the resulting macrocyclic peptide. Accordingly, we have been interested in exploring alternative approaches for creating peptide macrocycles through the post-translational processing of ribosomally derived polypeptides.

Here, we report the development of a new strategy to achieve this goal, which involves the side-chain-to-tail cyclization of intein-fused polypeptides by means of amino-functionalized unnatural amino acids. We also demonstrate how this cyclization strategy led to the development of a cyclic peptide with improved binding affinity toward a model target protein.

## Results and discussion

### Peptide macrocyclization strategy

Scheme 1 illustrates the general strategy investigated in the present study. This approach involves the use of a recombinant biosynthetic precursor in which an arbitrary target sequence is

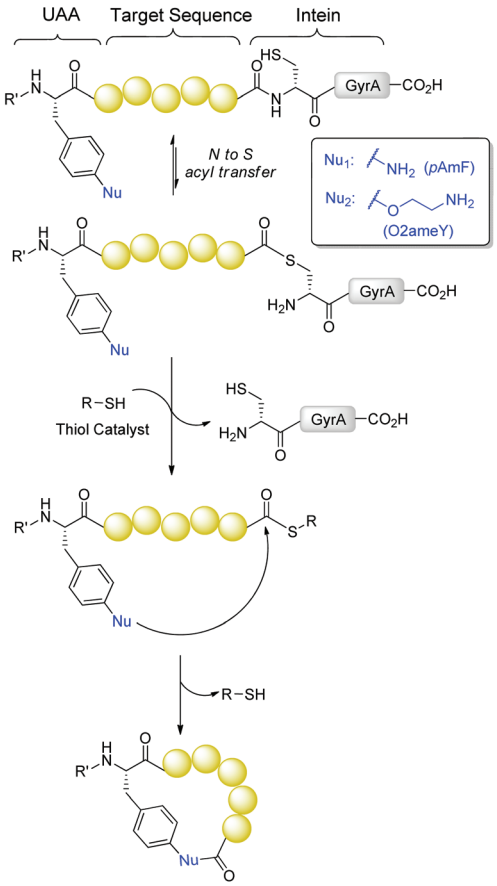
*Department of Chemistry, University of Rochester, Rochester, NY 14627, USA. E-mail: rfasan@ur.rochester.edu*

†Electronic supplementary information (ESI) available: Synthetic procedures and spectral data for O2azeY and O2ameY; additional chromatograms, MS spectra, and experimental data. See DOI: 10.1039/c6ob00192k

‡These authors contributed equally to this work.

Published on 01 April 2016. Downloaded by University of Rochester on 15/06/2016 14:40:11.

framed between an amino-functionalized unnatural amino acid and a C-terminal intein. In the latter, the conserved C-terminal asparagine residue (Asn198 in GyrA intein) and cysteine residue at the intein + 1 position are mutated (N198A) and removed, respectively, to allow for the formation of a transient thioester linkage at the junction between the peptide target sequence and intein while preventing splicing of the intein itself. We envisioned that thiol-induced cleavage of the intein moiety on this precursor polypeptide would form a reactive C-terminal thioester intermediate, which would then undergo aminolysis by action of the side-chain amino group from the unnatural amino acid upstream of the target sequence to yield the desired side-chain-to-tail cyclopeptide product (Scheme 1). Extending upon our previous studies on the cyclization of intein-fused polypeptides with amino–thiol containing amino acids,[35] this strategy would enable the formation of macrocyclic peptides featuring alternative side-chain-to-tail connectivities, thus broadening the repertoire of peptide macrocycles accessible *via* posttranslational elaboration of ribosomally produced polypeptides.




**Scheme 1** Synthesis of macrocyclic peptides *via* thiol-catalyzed cyclization of biosynthetic precursor proteins containing amino-functionalized unnatural amino acids. UAA: unnatural amino acid, GyrA: GyrA mini-intein from *Mycobacterium xenopi*. $Nu_1$ corresponds to *p*-amino-phenylalanine (*p*AmF) and $Nu_2$ corresponds to *O*-2-aminoethyl-tyrosine (O2ameY).

### *para*-Amino-phenylalanine-mediated cyclization

The realization of the reaction scheme outlined above was anticipated to require the choice of an amino-functionalized unnatural amino acid with suitable structural and reactivity properties. First, it should exhibit the desired nucleophilic reactivity in aqueous media and under conditions in which the cyclization through other nucleophilic amino acids (*e.g.*, lysine) is not favorable. Secondly, it should be amenable to ribosomal incorporation into the precursor polypeptide through convenient methods such as amber stop codon suppression[38] with orthogonal aminoacyl-tRNA synthetases. In light of these considerations, we selected *p*-amino-phenylalanine (*p*AmF) (Scheme 1, $Nu_1$) as a promising candidate for mediating the peptide cyclization process. Because of its significantly lower $pK_a$ compared to lysine (4.5 *vs.* 10.5 for Lys ε-$NH_2$), the aryl amino group in *p*AmF is expected to remain unprotonated and to possess higher nucleophilic reactivity in aqueous buffer at near-neutral pH. Indeed, aniline and derivatives thereof have been successfully applied as nucleophilic catalysts for promoting condensation reactions under physiological conditions.[39,40] In addition, an orthogonal tRNA/tRNA amino acyl synthetase (AARS) pair was previously made available to enable the incorporation of *p*AmF into recombinant proteins in response to an amber stop codon.[41]

To investigate the feasibility of *p*AmF-mediated peptide cyclization, a model precursor polypeptide was initially tested, which comprises a N-terminal chitin-binding domain (CBD) followed by *p*AmF, a 5 mer target sequence (TGSGT), and the GyrA mini-intein from *Myobacterium xenopi*[42] (Table 1, entry 1). This protein construct was expressed in BL21(*DE3*) *E. coli* cells and purified by nickel-affinity chromatography using a polyhistidine tag fused to the C-terminus of the GyrA intein. The purified protein was then incubated with either benzyl mercaptan or thiophenol in phosphate buffer at pH 7.0. Tris(2-carboxyethyl)phosphine (TCEP) was also added to the reaction mixture to maintain the thiols in the protein and reagents in reduced form. As shown in Fig. 1, MALDI-TOF mass spectrometry (MS) analysis of the reaction in the presence of benzyl mercaptan showed the formation of only the acyclic

**Table 1** Biosynthetic precursors investigated in this study. CBD: chitin binding domain, OpgY: *O*-propargyl-tyrosine. MeaF: 3-(2-mercapto-ethyl)amino-phenylalanine

| Entry | Construct Name | UAA | Target sequence |
|---|---|---|---|
| 1 | CBD-5-mer(*p*AmF) | *p*AmF | TGSGT |
| 2 | CBD-6-mer(*p*AmF) | *p*AmF | TGSYGT |
| 3 | CBD-7-mer(*p*AmF) | *p*AmF | TGSEYGT |
| 4 | CBD-8-mer(*p*AmF) | *p*AmF | TGSAEYGT |
| 5 | CBD-10-mer(*p*AmF) | *p*AmF | TGSKLAEYGT |
| 6 | CBD-10-mer(OpgY) | OpgY | TGSKLAEYGT |
| 7 | 5-mer(O2ameY) | O2ameY | TGSGT |
| 8 | 6-mer(O2ameY) | O2ameY | TGSYGT |
| 9 | 8-mer(O2ameY) | O2ameY | TGSAEYGT |
| 10 | Strep-11-mer(O2ameY) | O2ameY | FTNVHPQFANA |
| 11 | Strep-11-mer(OpgY) | OpgY | FTNVHPQFANA |
| 12 | Strep-11-mer(MeaF) | MeaF | FTNVHPQFANA |


This journal is © The Royal Society of Chemistry 2016

Published on 01 April 2016. Downloaded by University of Rochester on 15/06/2016 14:40:11.

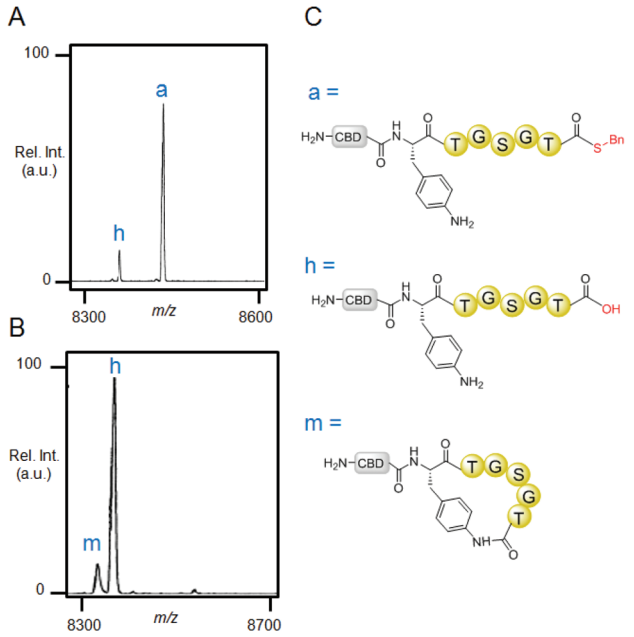



**Fig. 1** *p*AmF-mediated cyclization of 5 mer peptide sequence. (A–B) MALDI-TOF MS spectra of small MW products from reaction of CBD-5-mer(*p*AmF) (100 μM) with 10 mM thiol benzyl mercaptan (A) or thiophenol (B), and TCEP (10 mM) after 24 hours. 'm': macrocycle, 'h': hydrolysis product. $[m + H]^+$ calc: 8345.2, obs: 8345.4; $[h + H]^+$ calc: 8363.2, obs: 8363.6; $[a + H]^+$ calc: 8469.4, obs: 8469.0 (C) schematic structures of the macrocyclic product (m), thioester intermediate (a), and hydrolysis byproduct (h).

thioester intermediate ('a') together with a smaller amount of the hydrolysis byproduct ('h'). In contrast, the reaction in the presence of thiophenol was found to result in formation of the desired macrocyclic peptide product ('m', Fig. 1), along with the linear byproduct resulting from hydrolysis of the thiophenol thioester ('h'). Due to the presence of a large N-terminal tag (CBD protein) and identical target sequence, it is reasonable to assume that the cyclic and acyclic species generated in these experiments have similar ionization properties. Accordingly, the ratio of macrocycle : hydrolysis product in the thiophenol catalyzed reaction was estimated to be approximately 1 : 5. Furthermore, no significant changes in product distribution were observed over a 24 hour period.

To gain insight into the effect of the length of the target peptide sequence on the efficiency of the macrocyclization reaction, four additional biosynthetic precursors with 6 to 10 amino acid-long target sequences were generated (Table 1, entries 2–5). As for the 5 mer peptide, these sequences of arbitrary composition were previously found to have no particular propensity either toward cyclization or against it,[31,32] thus providing ideal model structures for testing the functionality of the present method. After purification, the corresponding proteins were incubated with thiophenol and TCEP as described above. Importantly, MALDI-TOF MS analysis showed the formation of the desired macrocyclic peptide for all the reactions, thus demonstrating that *p*AmF-mediated macrocyclization can be achieved across target peptide sequences of variable length (5–10 mer) and composition (Fig. 2). As observed for the 5 mer target sequence, however, these reactions also led to the accumulation of the hydrolysis byproduct in approximately 3 : 1 (6 mer) to 2 : 1 (7 mer to 10 mer) excess over the macrocycle. Since the 10 mer sequence also contains a lysine residue, this construct also provided an opportunity to probe the chemoselectivity of the cyclization reaction. Accordingly, a control protein construct (Table 1, entry 6) was prepared by incorporating *O*-propargyl-tyrosine (OpgY)[43] instead of *p*AmF upstream of the 10 mer target sequence. OpgY lacks a nucleophilic side-chain and is thus not expected to react with the C-terminal thioester. Incubation of the CBD-10-mer (OpgY) construct with thiophenol resulted only in the formation of hydrolysis product (Fig. S1†), thus confirming that lysine does not mediate the cyclization process under the applied conditions.

Although the results above supported the feasibility of the general strategy outlined in Scheme 1, they also showed that the desired thioester aminolysis reaction mediated by *p*AmF does not effectively outcompete hydrolysis of the thioester intermediate. This phenomenon can be attributed to the limited nucleophilicity of aryl amino group of *p*AmF in the context of the present reaction. The noticeable increase in the *p*AmF-mediated macrocyclization efficiency observed with the longer, 7 mer to 10 mer target sequences (Fig. 1 and 2) also suggests that unfavourable conformational constraints could be another factor contributing to the limited cyclization efficiency of the *p*AmF-containing proteins in the context of smaller rings (*i.e.*, for 5 mer to 6 mer target sequences).

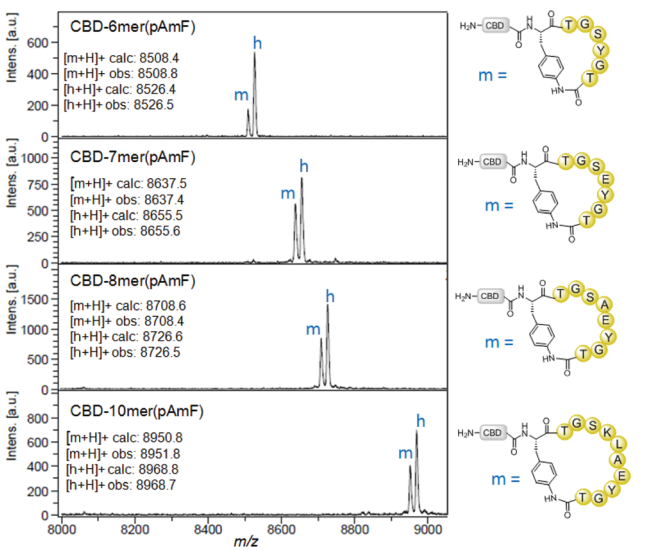



**Fig. 2** *p*AmF-mediated cyclization of 6 mer to 10 mer peptide sequences. MALDI-TOF MS spectra for the reactions with the indicated precursor proteins after incubation with 10 mM thiophenol and 10 mM TCEP (24 hours). The calculated and observed *m*/*z* values for the macrocyclic (m) and hydrolyzed (h) products are shown.

Published on 01 April 2016. Downloaded by University of Rochester on 15/06/2016 14:40:11.

### Identification of *O*-2-azidoethyl-tyrosine-tRNA synthetase

The considerations above prompted the design of an alternative amino-functionalized unnatural amino acid capable of mediating macrocyclization with higher efficiency. To this end, we identified *O*-2-aminoethyl-tyrosine (O2ameY, Scheme 1) as a promising candidate. Compared to *p*AmF, the side-chain alkyl amino functionality in O2ameY was expected to be more sterically accessible for nucleophilic attack at the C-terminal thioester group (Scheme 1), thus favouring cyclization over hydrolysis of the thioester intermediate. This hypothesis was also supported by our previous observations in the context of peptide cyclization promoted by amino–thiol unnatural amino acids.[35] Furthermore, the estimated p$K_a$ of the amino group in O2ameY (~9) was deemed suitable for mediating the desired cyclization process under near-neutral conditions.

Although a wide variety of tyrosine- and phenylalanine-derivatives have been successfully incorporated into proteins using engineered variants of the tyrosyl-tRNA synthetase from *Methanococcus jannaschii* or the pyrrolysyl-tRNA synthetase from *Methanosarcina* sp.,[38,44,45] the incorporation of unnatural amino acids containing polar side-chains have been very challenging due to the hydrophobic nature of the active sites in these enzymes.[46] Accordingly, finding a AARS variant for the efficient recognition and incorporation of O2ameY into the precursor protein was expected to be very difficult. To overcome this problem, we envisioned an alternative approach that entails the incorporation of the unnatural amino acid *O*-2-azidoethyl-tyrosine (O2azeY), followed by unmasking of the desired alkyl amino functionality *via* a Staudinger reduction of the azide group by means of the phosphine-based reducing agent (*i.e.*, TCEP) added during the cyclization reaction.

To identify a suitable AARS for incorporation of O2azeY, we screened a panel of *M. jannaschii* tyrosyl-tRNA synthetase variants previously engineered for the incorporation of *para*-substituted tyrosine and phenylalanine derivatives such as 3-aminotyrosine (3AmY),[47] *p*-azidophenylalanine (AzF),[48] *p*-aminophenylalanine (*p*AmF),[49] *O*-propargyltyrosine (OpgY),[43] *O*-(2-bromoethyl)-tyrosine (O2beY),[36] and *p*-acetylphenylalanine (pAcF)[50] (Fig. 3). The ability of these AARSs to incorporate O2azeY in response to an amber stop codon was measured by means of a Yellow Fluorescent Protein (YFP) reporter protein containing a TAG codon within the N-terminal region of the protein sequence.[35] In this assay, the suppression efficiency and fidelity of each synthetase is measured based on the relative expression levels of the reporter YFP protein in the presence and in the absence of O2azeY, respectively. As shown in Fig. 3, both O2beY-RS and pAcF-RS were found to efficiently incorporate O2azeY into the reporter YFP protein. Notably, the protein yield obtained with pAcF-RS in the presence of O2azeY is comparable to that observed in the presence of pAcF, for which this enzyme variant was selected. Despite the structural differences between pAcF and O2azeY, this result is in line with the 'polyspecificity' often exhibited by engineered AARS enzymes toward unnatural amino acids other than those involved in the functional selection process.[51] As expected,

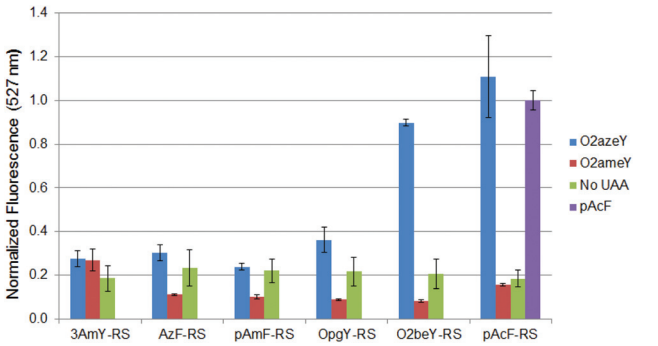



**Fig. 3** Screening of engineered aminoacyl-tRNA synthetases for ribosomal incorporation of O2azeY. The graph reports the relative fluorescence measured in the YFP reporter assay for the panel of AARSs in the presence of O2azeY, O2ameY, and no unnatural amino acid. Data are normalized to the fluorescence value obtained with pAcF-RS in the presence of pAcF.

none of the tested AARS variants were able to efficiently incorporate O2ameY (Fig. 3), likely reflecting the incompatibility of its polar side-chain with the hydrophobic environment surrounding the *para* position of the tyrosine substrate in the *M. jannaschii* tyrosyl-tRNA synthetase structure.[52] Because of its superior performance in the YFP assay, pAcF-RS was selected for preparation of the O2azeY-containing proteins.

To our surprise, the purified O2azeY-containing YFP protein did not show the expected molecular weight but rather a mass that is consistent with the quantitative reduction of the side-chain azido group of O2azeY to the corresponding amine derivative, to effectively yield a O2ameY-containing protein. Since pAcF-RS is unable to incorporate O2ameY directly (Fig. 3), it can be derived that the '*in vivo*' conversion of O2azeY to O2ameY must occur at the post-translational level, that is after O2azeY is charged onto the suppressor tRNA and incorporated into the protein by ribosomal synthesis. Partial reduction of azido-containing groups in living cells has been previously reported[53] but not in the context of compounds like O2azeY.[54] Importantly, our serendipitous discovery that O2azeY is efficiently converted to O2ameY during protein expression in *E. coli* cells provides a streamlined approach to obtain O2ameY-containing recombinant proteins for the present and other applications.

### O2ameY-mediated macrocyclization

Having acquired the capability of producing O2ameY-containing proteins in an efficient manner, we proceeded to investigate the feasibility and functionality of the envisioned O2ameY-mediated macrocyclization strategy (Scheme 1). To this end, we expressed and isolated the model 5 mer construct CBD-5-mer(O2ameY), in which O2ameY is placed upstream of the target sequence TGSGT. To identify optimal conditions for macrocyclization, this precursor protein was incubated under reducing (TCEP) conditions in the presence of either benzyl mercaptan or thiophenol and at varying pH (7.5, 8.2, and 9.0), (Fig. 4). The products of these reactions were analysed by

 This journal is © The Royal Society of Chemistry 2016

Published on 01 April 2016. Downloaded by University of Rochester on 15/06/2016 14:40:11.

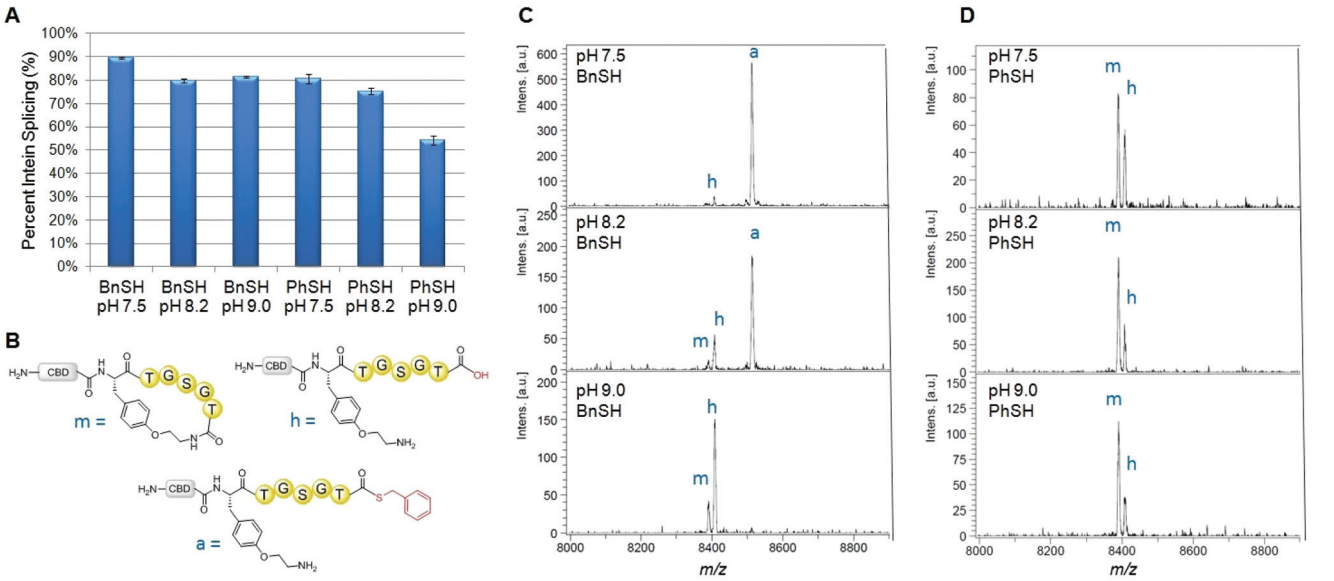



**Fig. 4** O2ameY-mediated peptide macrocyclization. (a) Extent of intein splicing as measured by SDS-PAGE gel densitometry for CBD-5-mer (O2ameY) after 24 h with benzyl mercaptan (BnSH) or thiophenol (PhSH) at varying pH. (b) Representative structures of 'm', 'h', and 'a'. MALDI-TOF mass spectra of CBD-5-mer(O2ameY) after 24 h with (c) benzyl mercaptan and (d) thiophenol at pH 7.5, 8.2, and 9.0. 'm': macrocycle, 'h': acyclic hydrolysis product, 'a': acyclic thioester product.

MALDI-TOF MS and the extent of intein cleavage was measured by SDS-PAGE followed by densitometric analysis of the gel bands corresponding to full-length protein and cleaved GyrA intein (Fig. S2 and S3†).[35]

As summarized in Fig. 4A, a high degree of intein splicing (75–90%) was observed across most of the conditions tested, consistent with the ability of both thiol nucleophiles to induce intein cleavage *via* transthioesterification. Despite the generally higher levels of intein cleavage in the presence of benzyl mercaptan, in particular at pH 7.5, these reactions mainly resulted in the formation of the acyclic thioester intermediate ('a') along with small quantities of the hydrolysis byproduct ('h'). At higher pH (9.0) however, the acyclic intermediate is fully converted to desired macrocyclic product ('m') along with the hydrolysis byproduct (Fig. 4C) in an approximately 1 : 4 ratio.

In stark contrast, the thiophenol-catalyzed reactions were found to yield the desired macrocyclic peptide as the major product across all the tested pH values (Fig. 4D). As the pH is raised from pH 7.5 to 9.0, macrocyclization becomes more favourable over hydrolysis, as reflected by an increase of the macrocycle ('m') : hydrolysis product ('h') ratio from 3 : 2 to 7 : 3. Interestingly, the acyclic thiophenol thioester intermediate is not observed in these reactions, even at earlier time points (data not shown), indicating that, once formed, this species is rapidly converted to the macrocyclic product *via* O2ameY-mediated aminolysis or is hydrolyzed. The higher efficiency of the macrocyclization process at higher pH can be rationalized considering that a larger fraction of O2ameY side-chain amine is deprotonated at more alkaline conditions and thus becomes available for nucleophilic attack to the thiophenol thioester group. On the other hand, the better results obtained with PhSH *vs*. BnSH is consistent with the higher reactivity of thiophenol thioesters as compared to the benzyl mercaptan thioester counterparts.[55] On the basis of these experiments, we established that macrocyclization of the O2ameY-containing constructs occurs most efficiently at pH 8.2 with the addition of thiophenol as the catalyst, as these conditions provide an optimal combination of high intein cleavage (75%) with high conversion to the desired macrocyclic product (70–80%). These results also clearly demonstrated the superiority of the O2ameY-based cyclization strategy over that based on *p*AmF-containing constructs described earlier.

To assess the value of the O2ameY-based approach toward generating cyclic peptides of low to medium MW (600–1000 Da), 5 mer to 8 mer target peptide sequences lacking a N-terminal tail were investigated (Table 1, entries 7–9). Following expression and purification, the corresponding precursor proteins were subjected to the optimized reaction conditions (10 mM PhSH, 10 mM TCEP, pH 8.2), followed by LC-MS analysis. Notably, all of these reactions led to the desired small-MW peptide macrocycle as the only product, whose cyclic structure was further corroborated by the characteristic MS/MS fragmentation profile (Fig. 5). Based on the measured extent of intein cleavage and the lack of observation of the thioester intermediate or hydrolysis byproduct, the macrocycle conversion yield of these reactions was estimated to be around 65–80%. To gain further insight into the kinetics of these reactions, the cyclization of the 5-mer(O2ameY) construct was monitored over time (Fig. 6). These studies revealed that the precursor protein had undergone 50% intein cleavage after 6 hours and reached a maximum of 80% cleavage after 18–24 hours (Fig. S4†). The

Published on 01 April 2016. Downloaded by University of Rochester on 15/06/2016 14:40:11.

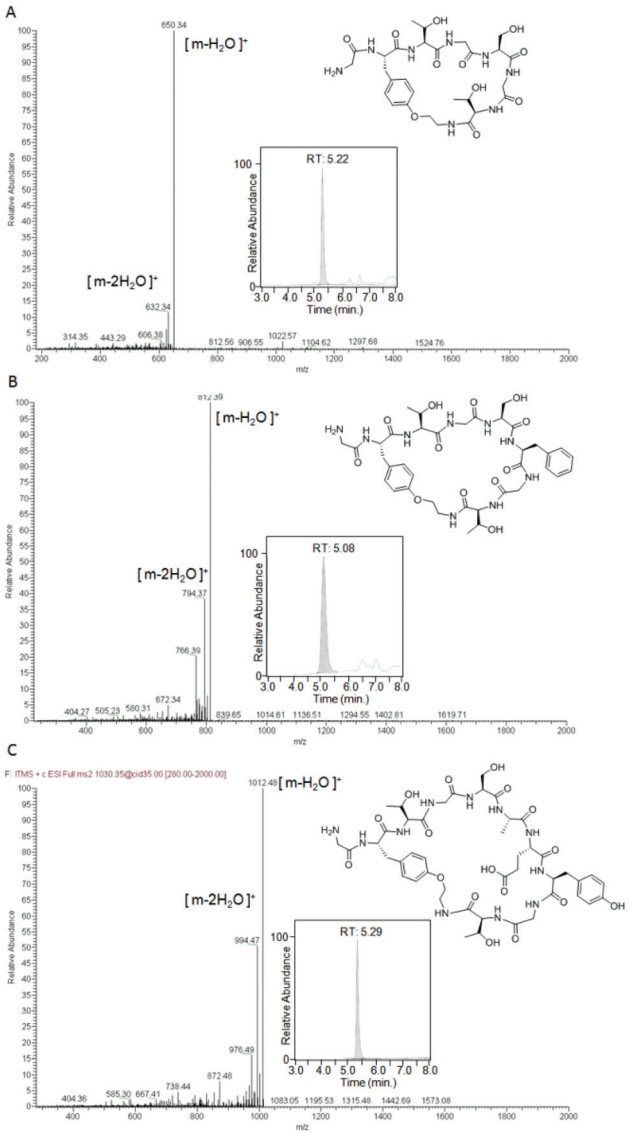



Fig. 5 Chemical structures, LC-MS ion extract chromatograms (inserts), and MS/MS spectra for macrocyclic peptide products obtained *via* O2ameY-mediated cyclization of 5-mer(O2ameY), 6-mer(O2ameY), and 8-mer(O2ameY) precursor proteins.

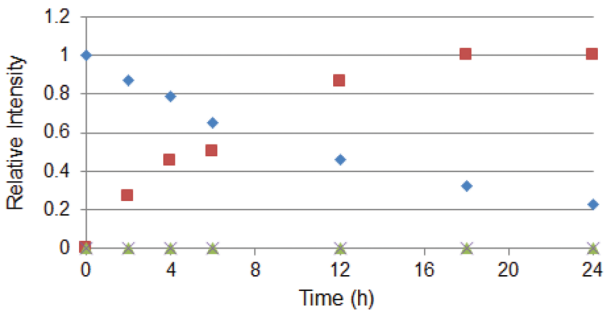



Fig. 6 Time course experiment for cyclization of 5-mer(O2ameY) construct. The graph plots the % of intein cleavage (♦) and the relative amount of macrocyclic product (■), thioester intermediate (▲), and hydrolysis byproduct (×) over time as measured by LC-MS (Fig. S4†).

incomplete cleavage even at extended reaction times suggests that a fraction of the precursor protein carries a non-functional GyrA intein, which is unable to undergo the thiol-induced transthioesterification, possibly due to partial misfolding during the expression and/or purification step. In concomitance with the cleavage of the precursor protein, the macrocyclic product was found to progressively accumulate over time. Interestingly, no accumulation of the thioester intermediate was observed at any of the time points. Altogether, these results indicate that the rate-limiting step of the overall reaction is the thiophenol-induced transthioesterification step and that all of the precursor protein capable of reacting with the thiol catalyst undergoes O2ameY-mediated cyclization. As shown by the plot of Fig. 6, about 50% and 80% of the total macrocyclic product is formed after 4.5 and 12 hours, respectively.

### Synthesis of streptavidin-binding cyclic peptide

To demonstrate the utility of this approach toward generating functional peptide macrocycles, a biosynthetic precursor encompassing a histidine–proline–glutamine (HPQ) motif known to bind streptavidin[56–58] was generated (Table 1, entry 10). Specifically, this protein construct contained an N-terminal FLAG affinity tag (MDYKDDDYK), followed by a linker sequence (GSSG), and a 11 mer target sequence (FTNVHPQFANA) flanked by O2ameY and GyrA. Subjecting the precursor protein Strep-11-mer(O2ameY) to the standard reaction conditions led to the production of the desired macrocyclic peptide **1**, as confirmed by LC-MS and MALDI-TOF MS (Fig. 7A). In the reaction only a minimal fraction (<5%) of the hydrolysis byproduct was observed ('h', Fig. 7A), demonstrating the efficiency of the macrocyclization reaction also in context of a 11 mer target sequence and in the presence of alternative residues (*i.e.*, Ala instead of Thr) at the 'I − 1' position preceding the intein. Based on the observed extent of intein cleavage (63%), the macrocycle yield for this transformation was estimated to be ~60%. In contrast, sufficient amounts of the *p*AmF-cyclized peptide for streptavidin binding studies could not be isolated, primarily due to inefficient cyclization in the presence of this unnatural amino acid as observed in the context of the model peptide sequences (Fig. 2).

In order to evaluate the effect of cyclization on the streptavidin binding affinity of the peptide, the control linear peptide FLAG-(OpgY)-FTNVHPQFANA-$CO_2H$ (**2**) was also prepared. The latter carries the non-nucleophilic *O*-propargyl-tyrosine (OpgY) in place of O2ameY and was produced by thiophenol-induced cleavage of the corresponding precursor protein (Table 1, entry 11) followed by thioester hydrolysis under alkaline conditions (pH 8.0). The successful isolation of this peptide in the linear form further demonstrated the excellent chemoselectivity of the O2ameY-mediated cyclization process, as none of the several nucleophilic residues present in the target sequence and FLAG tag caused cyclization of the peptide under the applied reaction conditions.

The dissociation constants ($K_D$) for the cyclic and linear peptide were determined by means of an immunoassay, in

 This journal is © The Royal Society of Chemistry 2016

Published on 01 April 2016. Downloaded by University of Rochester on 15/06/2016 14:40:11.

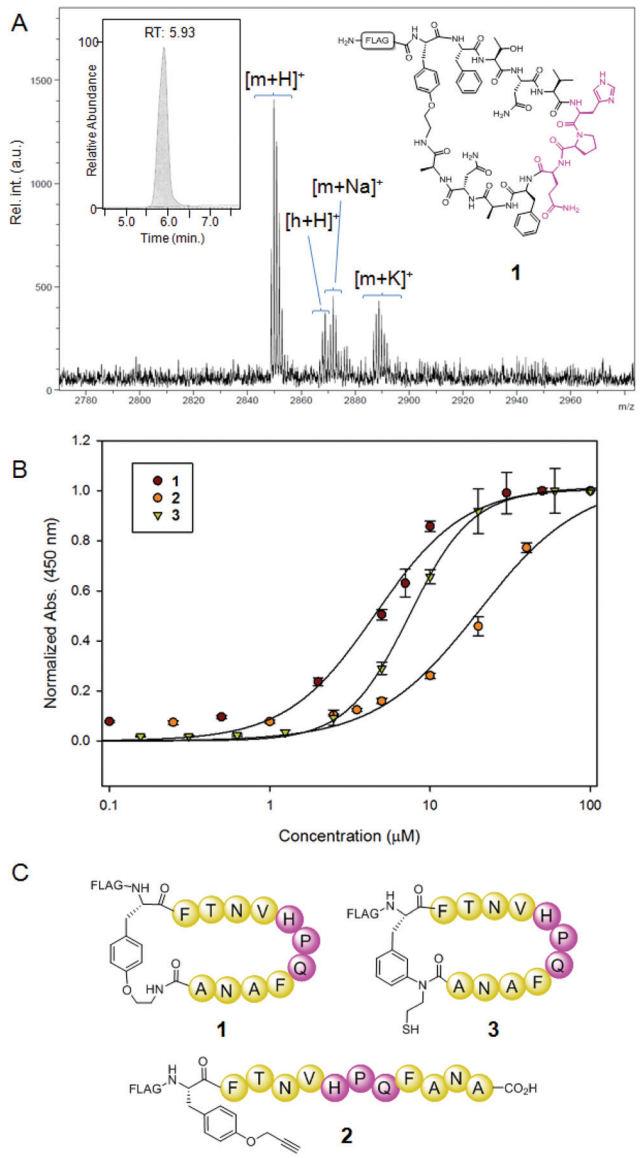



**Fig. 7** Streptavidin-binding peptides. (A) Chemical structure, LCMS ion extract chromatogram (insert), and MALDI-TOF MS spectrum of macrocycle **1** obtained *via* cyclization of Strep-11-mer(O2ameY). $[m + H]^+$ calc.: 2849.1; obs.: 2849.8. (B–C) Streptavidin binding curves and schematic structure for O2ameY-cyclized peptide **1** ($K_D$: 5.1 μM ± 0.4), MeaF-cyclized peptide **3** ($K_D$ = 7.7 μM ± 0.6), and the linear peptide **2** ($K_D$ = 22.4 μM ± 1.8).

which the FLAG-tagged peptides were incubated on streptavidin-coated plates at varying concentrations between 0.05–40 μM. The amount of the streptavidin-bound peptide was then measured based on the corresponding colorimetric signal (450 nm), after incubation with a horseradish peroxidase (HRP) conjugated anti-FLAG antibody followed by addition of the HRP substrate *o*-phenylinediamine. In this assay, the O2ameY-containing macrocycle exhibited a $K_D$ of 5.1 μM, whereas the linear peptide exhibited a $K_D$ of 22.4 μM (Fig. 7B). The 4-fold lower $K_D$ for the cyclic peptide compared to its linear counterpart clearly evidenced the beneficial effect of the macrocyclic backbone toward improving the binding affinity for the target protein. A direct comparison was also made with the cyclopeptide **3** (Fig. 7C), which comprises an identical target sequence as **1** (FTNVHPQFANA) but features a different side-chain-to-tail linkage as obtained *via* cyclization by means of 3-(2-mercapto-ethyl)amino-phenylalanine (MeaF).[35] In this assay, the MeaF-cyclized peptide **3** exhibited an approximately 1.5-fold higher $K_D$ (7.7 μM) than the O2ameY-cyclized peptide **1** (Fig. 7B). Since the peptide–streptavidin interaction is primarily driven by the HPQ motif[35] and is not expected to directly involve the intramolecular linkage, the differential binding affinity of these compounds nicely illustrates the subtle effects of the side-chain-to-tail linkage in modulating the protein binding properties of these macrocyclic peptides.

## Conclusion

We have developed a new methodology to achieve the formation of macrocyclic peptides *via* an intramolecular ligation of amino-functionalized unnatural amino acids to the C-terminus of a genetically encoded peptide. Using this approach, 5- to 11-amino acid-long peptides could be efficiently cyclized to generate medium- to large-membered rings (23–44 atoms). In these compounds, the N-terminal tail can be readily altered to include either a single residue or various affinity tags to facilitate the immobilization, isolation, and/or detection of these macrocyles for characterization purposes. Importantly, this strategy could be applied to generate a cyclic peptide with improved streptavidin-binding affinity, highlighting the effect of macrocyclization and the type of side-chain-to-tail connectivity toward modulating this functional property.

Another important result of this study concerns the discovery that the azido-containing unnatural amino acid O2azeY is consistently and quantitatively reduced to the amine-containing residue O2ameY in living *E. coli* cells. While the exact mechanism of this process was not investigated, it is most likely dependent upon the reducing intracellular environment provided by the bacterial host. As exemplified by the present study, this *in vivo* azide reduction process could be exploited to produce proteins containing O2ameY as well as other amino-functionalized unnatural amino acids, a currently challenging task using commonly adopted aminoacyl-tRNA synthetases for amber stop codon suppression. These results are also insightful in that they provide a cautionary note about the often assumed but clearly not universal, bioorthogonal nature of azido-based functionalities in biological systems.

## Experimental procedures

### General

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

Published on 01 April 2016. Downloaded by University of Rochester on 15/06/2016 14:40:11.

230–400 mesh. $^{1}H$ and $^{13}C$ 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 unnatural amino acids

Synthetic procedures and spectral data corresponding to the preparation and characterization of O2azeY and O2ameY are reported in the ESI.† OpgY was synthesized as described previously.[32]

### Plasmid constructs

The biosynthetic precursors described in Table 1 were expressed using pET22-based plasmids and the cloning procedures for preparation of these constructs were reported previously.[32,36] The plasmids for expression of the amber stop codon suppression system were derived from pEVOL[59] and their preparation was described previously.[35,36]

### Protein expression and purification

Proteins were expressed in *E. coli* BL21(DE3) cells cotransformed with the pET22-based plasmid for expression of the protein precursor and the pEVOL-based vector for expression of the appropriate AARS (*p*AmF-RS, pAzF-RS, 3AmY-RS, pAcF-RS, OpgY-RS, O2beY-RS). After overnight growth, cells were used to inoculate M9 media supplemented with ampicillin (50 mg $L^{-1}$), chloramphenicol (34 mg $L^{-1}$) and glycerol (1%). At an $OD_{600}$ of 0.6, the culture was supplemented with the appropriate unnatural amino acid (*p*AmF, O2azeY, or OpgY) at a final concentration of 2 mM and expression of the AARS was induced by addition of L-arabinose (0.06%). After one hour of shaking at 30 °C, protein expression was induced by addition of 0.25 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Cultures were grown for an additional 12 hours at 27 °C and harvested by centrifugation at 3400*g*. Cells were resuspended in Tris buffer (50 mM; pH 7.4) containing NaCl (150 mM) and imidazole (20 mM) and lysed by sonication. Protein was purified from the cell lysate by Ni-NTA affinity chromatography using Tris buffer (50 mM; pH 7.4; NaCl 150 mM) with 20 mM imidazole for washing and 300 mM imidazole for elution. The recovered protein was exchanged into potassium phosphate buffer (50 mM, NaCl 150 mM, pH 7.4), concentrated, and stored at −80 °C.

### Macrocyclization reactions

Reactions were carried out by incubating the precursor protein (100 μM) in potassium phosphate buffer (50 mM, NaCl 150 mM) in the presence of TCEP (10 mM) and thiophenol or benzyl mercaptan (10 mM). The pH of the reaction was adjusted to the specified value, and the reaction was gently mixed for up to 24 hours at room temperature. Protein splicing of the CBD-fused substrates was monitored and quantified by SDS-PAGE and densitometric analysis of the gel bands corresponding to the full-length protein (24–31 kDa) and spliced GyrA intein (~22 kDa) using ImageJ software. The low MW products (2–9 kDa) were analyzed by MALDI-TOF MS. The products of the reactions with the biosynthetic precursors without N-terminal CBD were analysed by LC-MS. The extent of intein cleavage was determined based on the deconvoluted MS spectra for the full-length protein and spliced GyrA intein, whereas the relative amount of the macrocyclic product, hydrolysis product, and acyclic thiophenol thioester intermediate was estimated based on the corresponding extracted-ion chromatograms. Each reaction was carried out at least in duplicate.

### Fluorescence assay for AARS screening

*E. coli* BL21(DE3) cells were co-transformed with a pET22-based plasmid encoding for MetGly(amber stop)YFP-$His_6$ and a pEVOL-based plasmid encoding for an appropriate aminoacyl-tRNA synthetase and cognate amber suppressor tRNA. Cells were then grown in LB media containing ampicillin (50 mg $L^{-1}$) and chloramphenicol (26 mg $L^{-1}$) at 37 °C overnight. The overnight cultures were used to inoculate 96-deep well plates containing M9 minimal media. At an $OD_{600}$ of 0.6, cell cultures were induced by adding arabinose (0.06%), IPTG (0.2 mM), and the appropriate unnatural amino acid (final concentration of 1 mM for L-isomer). After overnight growth at 27 °C, the cell cultures were diluted (1 : 1) with phosphate buffer (50 mM, 150 mM NaCl, pH 7.5) and fluorescence intensity ($\lambda_{ex}$ = 514 nm; $\lambda_{em}$ = 527 nm) was determined using a Tecan Infinite 1000 plate reader. Cell cultures containing no unnatural amino acid were included as negative controls. Each sample was measured in triplicate. The fluorescence values were normalized to that obtained for cells expressing pAcF-RS in the presence of pAcF.

### Streptavidin binding studies

Cyclic peptide **1** was prepared according to the macrocyclization conditions described above. Linear peptide **2** was produced by thiophenol-induced cleavage of the corresponding precursor protein (Table 1, entry 11) followed by thioester hydrolysis under alkaline conditions (pH 9.0). Following completion of the reactions (24 hours), the pH was adjusted to 7.0 and the solution was dialyzed against water overnight with a 500 Da cut-off membrane. The reaction was then filtered through a 10 kDa cut-off centrifugal membrane (Millipore) at 3600 *g*. The filtrate was lyophilized and the peptide was resuspended in Tris buffer (50 mM, 150 mM NaCl, pH 7.4). A solution of each peptide (100 μL), was added to NeutrAvidin-coated 96-well plates (Pierce) at various concentrations (0.1–40 μM) in triplicate. After incubation at room temperature for 2 hours, the wells were then washed three times with 200 μL of Tris buffer (50 mM, 150 mM NaCl, pH 7.4) containing 0.5% Tween 20. A solution of anti-FLAG-antibody-HRP conjugate (100 μL, 1 : 2500 dilution in Tris buffer) was added to each well and incubated for 1 hour. The wells were washed three times with 200 μL Tris buffer containing 0.5% Tween 20. Finally, 100 μL of SigmaFast OPD solution was added to each well, and the absorbance at 450 nm was measured after 20 minutes.

 This journal is © The Royal Society of Chemistry 2016

Published on 01 April 2016. Downloaded by University of Rochester on 15/06/2016 14:40:11.

$K_D$ values were calculated with SigmaPlot *via* fitting of the dose–response curves using a 1 : 1 binding model.

## Acknowledgements

This work was supported by the U.S. National Institutes of Health (grant R21 CA187502). MS instrumentation was supported by the U.S. National Science Foundation (grants CHE-0840410 and CHE-0946653).

## Notes and references

1 E. Marsault and M. L. Peterson, *J. Med. Chem.*, 2011, **54**, 1961.
2 T. A. Cardote and A. Ciulli, *ChemMedChem*, 2016, DOI: 10.1002/cmdc.201500450, in press.
3 E. A. Villar, D. Beglov, S. Chennamadhavuni, J. A. Porco, Jr., D. Kozakov, S. Vajda and A. Whitty, *Nat. Chem. Biol.*, 2014, **10**, 723.
4 N. Bionda and R. Fasan, *ChemBioChem*, 2015, **16**, 2011.
5 A. K. Yudin, *Chem. Sci.*, 2015, **6**, 30.
6 E. M. Nolan and C. T. Walsh, *ChemBioChem*, 2009, **10**, 34.
7 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 and B. A. Burkett, *J. Med. Chem.*, 2000, **43**, 1271.
8 D. Wang, W. Liao and P. S. Arora, *Angew. Chem., Int. Ed.*, 2005, **44**, 6525.
9 J. M. Smith, J. R. Frost and R. Fasan, *Chem. Commun.*, 2014, **50**, 5027.
10 L. D. Walensky, A. L. Kung, I. Escher, T. J. Malia, S. Barbuto, R. D. Wright, G. Wagner, G. L. Verdine and S. J. Korsmeyer, *Science*, 2004, **305**, 1466.
11 T. Rezai, J. E. Bock, M. V. Zhou, C. Kalyanaraman, R. S. Lokey and M. P. Jacobson, *J. Am. Chem. Soc.*, 2006, **128**, 14073.
12 T. Rezai, B. Yu, G. L. Millhauser, M. P. Jacobson and R. S. Lokey, *J. Am. Chem. Soc.*, 2006, **128**, 2510.
13 J. E. Bock, J. Gavenonis and J. A. Kritzer, *ACS Chem. Biol.*, 2013, **8**, 488.
14 Z. Qian, T. Liu, Y. Y. Liu, R. Briesewitz, A. M. Barrios, S. M. Jhiang and D. Pei, *ACS Chem. Biol.*, 2013, **8**, 423.
15 J. Rizo and L. M. Gierasch, *Annu. Rev. Biochem.*, 1992, **61**, 387.
16 R. L. Dias, R. Fasan, K. Moehle, A. Renard, D. Obrecht and J. A. Robinson, *J. Am. Chem. Soc.*, 2006, **128**, 2726.
17 D. Diana, A. Basile, L. De Rosa, R. Di Stasi, S. Auriemma, C. Arra, C. Pedone, M. C. Turco, R. Fattorusso and L. D. D'Andrea, *J. Biol. Chem.*, 2011, **286**, 41680.
18 J. S. Quartararo, P. Wu and J. A. Kritzer, *ChemBioChem*, 2012, **13**, 1490.
19 C. J. White and A. K. Yudin, *Nat. Chem.*, 2011, **3**, 509.
20 J. M. Smith, J. R. Frost and R. Fasan, *J. Org. Chem.*, 2013, **78**, 3525.
21 J. R. Frost, J. M. Smith and R. Fasan, *Curr. Opin. Struct. Biol.*, 2013, **23**, 571.
22 K. T. O'Neil, R. H. Hoess, S. A. Jackson, N. S. Ramachandran, S. A. Mousa and W. F. DeGrado, *Proteins*, 1992, **14**, 509.
23 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 and W. J. Dower, *Science*, 1996, 458.
24 A. Tavassoli and S. J. Benkovic, *Angew. Chem., Int. Ed.*, 2005, **44**, 2760.
25 A. Tavassoli, Q. Lu, J. Gam, H. Pan, S. J. Benkovic and S. N. Cohen, *ACS Chem. Biol.*, 2008, **3**, 757.
26 J. Morimoto, Y. Hayashi and H. Suga, *Angew. Chem., Int. Ed.*, 2012, **51**, 3423.
27 C. Heinis, T. Rutherford, S. Freund and G. Winter, *Nat. Chem. Biol.*, 2009, **5**, 502.
28 T. Kawakami, T. Ishizawa, T. Fujino, P. C. Reid, H. Suga and H. Murakami, *ACS Chem. Biol.*, 2013, **8**(6), 1205–1214.
29 V. Baeriswyl, H. Rapley, L. Pollaro, C. Stace, D. Teufel, E. Walker, S. Chen, G. Winter, J. Tite and C. Heinis, *ChemMedChem*, 2012, **7**, 1173.
30 J. M. Smith, F. Vitali, S. A. Archer and R. Fasan, *Angew. Chem., Int. Ed.*, 2011, **50**, 5075.
31 M. Satyanarayana, F. Vitali, J. R. Frost and R. Fasan, *Chem. Commun.*, 2012, **48**, 1461.
32 J. R. Frost, F. Vitali, N. T. Jacob, M. D. Brown and R. Fasan, *ChemBioChem*, 2013, **14**, 147.
33 J. M. Smith, N. C. Hill, P. J. Krasniak and R. Fasan, *Org. Biomol. Chem.*, 2014, **12**, 1135.
34 N. Bionda and R. Fasan, *ChemBioChem*, 2015, **16**, 2011.
35 J. R. Frost, N. T. Jacob, L. J. Papa, A. E. Owens and R. Fasan, *ACS Chem. Biol.*, 2015, **10**, 1805.
36 N. Bionda, A. L. Cryan and R. Fasan, *ACS Chem. Biol.*, 2014, **9**, 2008.
37 S. Chen, J. Morales-Sanfrutos, A. Angelini, B. Cutting and C. Heinis, *ChemBioChem*, 2012, **13**, 1032.
38 C. C. Liu and P. G. Schultz, *Annu. Rev. Biochem.*, 2010, **79**, 413.
39 A. Dirksen, T. M. Hackeng and P. E. Dawson, *Angew. Chem., Int. Ed.*, 2006, **45**, 7581.
40 P. Crisalli and E. T. Kool, *J. Org. Chem.*, 2013, **78**, 1184.
41 R. A. Mehl, J. C. Anderson, S. W. Santoro, L. Wang, A. B. Martin, D. S. King, D. M. Horn and P. G. Schultz, *J. Am. Chem. Soc.*, 2003, **125**, 935.
42 A. Telenti, M. Southworth, F. Alcaide, S. Daugelat, W. R. Jacobs and F. B. Perler, *J. Bacteriol.*, 1997, **179**, 6378.
43 A. Deiters and P. G. Schultz, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 1521.
44 K. Lang and J. W. Chin, *Chem. Rev.*, 2014, **114**, 4764.
45 A. Dumas, L. Lercher, C. D. Spicer and B. G. Davis, *Chem. Sci.*, 2015, **6**, 50.
46 D. P. Nguyen, T. Elliott, M. Holt, T. W. Muir and J. W. Chin, *J. Am. Chem. Soc.*, 2011, **133**, 11418.
47 M. R. Seyedsayamdost, J. Xie, C. T. Chan, P. G. Schultz and J. Stubbe, *J. Am. Chem. Soc.*, 2007, **129**, 15060.
48 J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang and P. G. Schultz, *J. Am. Chem. Soc.*, 2002, **124**, 9026.

Published on 01 April 2016. Downloaded by University of Rochester on 15/06/2016 14:40:11.

49 S. W. Santoro, L. Wang, B. Herberich, D. S. King and P. G. Schultz, *Nat. Biotechnol.*, 2002, **20**, 1044.
50 L. Wang, Z. Zhang, A. Brock and P. G. Schultz, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 56.
51 D. D. Young, T. S. Young, M. Jahnz, I. Ahmad, G. Spraggon and P. G. Schultz, *Biochemistry*, 2011, **50**, 1894.
52 T. Kobayashi, O. Nureki, R. Ishitani, A. Yaremchuk, M. Tukalo, S. Cusack, K. Sakamoto and S. Yokoyama, *Nat. Struct. Biol.*, 2003, **10**, 425.
53 A. Chatterjee, S. B. Sun, J. L. Furman, H. Xiao and P. G. Schultz, *Biochemistry*, 2013, **52**, 1828.
54 A. Tuley, Y. S. Wang, X. Fang, Y. Kurra, Y. H. Rezenom and W. R. Liu, *Chem. Commun.*, 2014, **50**, 2673.
55 P. E. Dawson, T. W. Muir, I. Clarklewis and S. B. H. Kent, *Science*, 1994, **266**, 776.
56 E. A. Bayer, H. Ben-Hur and M. Wilchek, *Methods Enzymol.*, 1990, **184**, 80.
57 B. A. Katz, *Biochemistry*, 1995, **34**, 15421.
58 L. B. Giebel, R. T. Cass, D. L. Milligan, D. C. Young, R. Arze and C. R. Johnson, *Biochemistry*, 1995, **34**, 15430.
59 T. S. Young, I. Ahmad, J. A. Yin and P. G. Schultz, *J. Mol. Biol.*, 2010, **395**, 361.


This journal is © The Royal Society of Chemistry 2016

Electronic Supplementary Material (ESI) for Organic & Biomolecular Chemistry.
This journal is © The Royal Society of Chemistry 2016

# Side-chain-to-tail cyclization of ribosomally derived peptides promoted by aryl and alkyl amino-functionalized unnatural amino acids

John R. Frost[§], Zhijie Wu[§], Yick Chong Lam[§], Andrew E. Owens and Rudi Fasan*

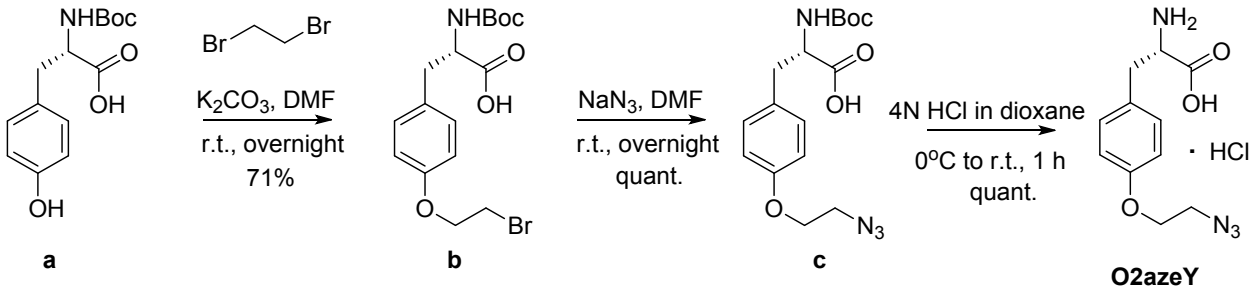
*Department of Chemistry, University of Rochester, Hutchinson Hall,* Rochester NY, 14627

[§]These authors contributed equally to this work

* Corresponding author: rfasan@ur.rochester.edu

**Table of Contents**

| | |
|---|---|
| Synthetic Procedures | S2-S4 |
| NMR spectra | S5-S6 |
| Supplementary Figures S1-S4 | S7-S10 |




**Scheme S1.** Synthesis of O-2-azidoethyl-tyrosine (O2azeY).

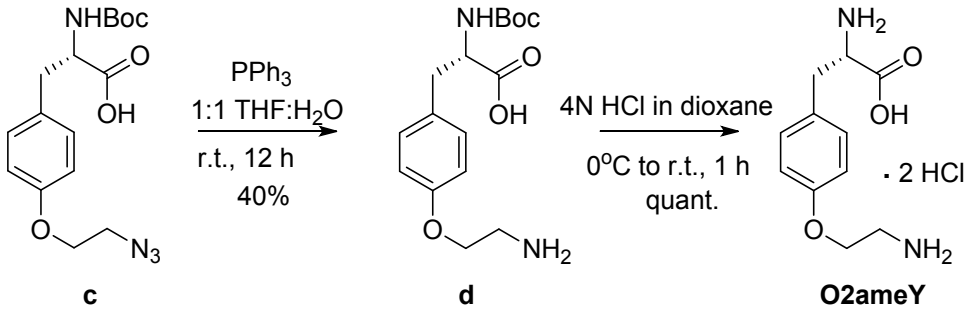
***Synthesis of O-2-azidoethyl-tyrosine, O2azeY***

Commercially available (tert-butoxycarbonyl)-L-tyrosine, (2 grams, 7.109 mmol, 1 eq) was added to a dry 100 mL round bottom flask under argon. The compound was dissolved in 20 mL anhydrous DMF and 1.23 mL of 1,2 dibromoethane (2.671 grams, 14.218 mmol, 2 eq) was added, followed by addition of $K_2CO_3$ (2.947 grams, 21.327 mmol, 3 eq ). The reaction was stirred at room temperature for 12 hours. Following completion the solution was filtered by vacuum filtration to remove the solid, and the filtrate was transferred to a separatory funnel and washed 1x 40 mL of $H_2O$. The aqueous layer was acidified to pH 4 and was extracted with 100 mL of EtOAc. The organic layer was dried by $MgSO_4$ and the volatiles were removed *in vacuo*. The resulting mixture was purified by flash column chromatography using a gradient of 15:4:1 to 10:4:1 hexane: ethyl acetate: acetic acid to afford N-Boc-O-(2-bromoethyl)-tyrosine (**b**) as a yellow oil (1.95 grams, 5.022 mmol, 70.6% yield). $^1$H NMR (400 MHz, 25 °C, $CDCl_3$): δ 6.99 (d, J = 8.4 Hz, 2H), 6.76 (d, J = 8.8 Hz, 2H), 5.15 (d, J = 8.4 Hz, 1H), 4.53 (d, J = 7.2 Hz, 1H), 4.36 (q, J = 4.8, 2H), 3.43 (t, J = 5.6 Hz, 2H), 3.02 (t, J = 6 Hz, 2H), 1.41 (s, 9H).

N-Boc-O-(2-bromoethyl)-tyrosine (0.8 grams, 2.06 mmol, 1 equiv.) was dissolved in 17 mL of anhydrous DMF in a 100 mL round bottom flask under argon. Sodium azide (0.147g, 2.26 mmol, 1.1 equiv.) was added. The reaction was heated at 50 °C, and stirred for 12 hours. The reaction mixture was then transferred to the separatory funnel and combined with 60 mL of deionized water. The resulting mixture was acidified to pH 4 and extracted with 2 x 50 mL of EtOAc. The organic layer was dried over $MgSO_4$ and the volatiles were removed *in vacuo* to yield N-Boc-O-(2-azidoethyl)-tyrosine (**c**) as a yellow oil (720 mg, 2.06 mmol, >99% yield) $^1$H NMR (400 MHz, 25 °C, $CDCl_3$): δ 6.99 (d, J = 8.4 Hz, 2H), 6.78 (d, J = 8.4 Hz, 2H), 4.97 (d, J =

7.6 Hz, 1H), 4.55 (d, J = 6.8 Hz, 1H), 4.24 (t, J = 4.8, 2H), 3.45 (q, J = 5.2 Hz, 2H), 3.01 (t, J = 6 Hz, 2H), 1.42 (s, 9H).

N-Boc-O-(2-azidoethyl)-tyrosine (720 mg, 2.06 mmol, 1 eq) was dissolved in a solution of 4N HCl in dioxane (20 mL). The reaction was stirred at room temperature for 4 hours until completion, as monitored by TLC. The completed reaction was dried under reduced pressure to yield the hydrochloride salt of O-2-azidoethyl-tyrosine as a white solid (590 mg, quantitative). $^{1}$H NMR (500 MHz, 25°C, $CD_3OD$): δ 7.10 (d, J=8.7 Hz, 2H), 6.80 (d, J=8.6 Hz, 2H), 4.35 (t, J=5.28 Hz, 2H), 4.28 (dd, J=7.3, 6.2 Hz, 1H), 3.6-3.5 (m, 2H), 3.19 (dd, J=14.50, 6.26 Hz, 1H), 3.11 (dd, 14.5, 7.5 Hz, 1H).$^{13}$C NMR (126 MHz, 25°C, $CD_3OD$) δ 170.1, 158.5, 131.5, 125.5, 116.9, 65.9, 55.4, 50.6, 36.7. MS (ESI) for $C_{11}H_{14}N_4O_3$ $[M+H]^+$ calc: 251.25; $[M+H]^+$ obs: 251.1.




**c** **d** **O2ameY**

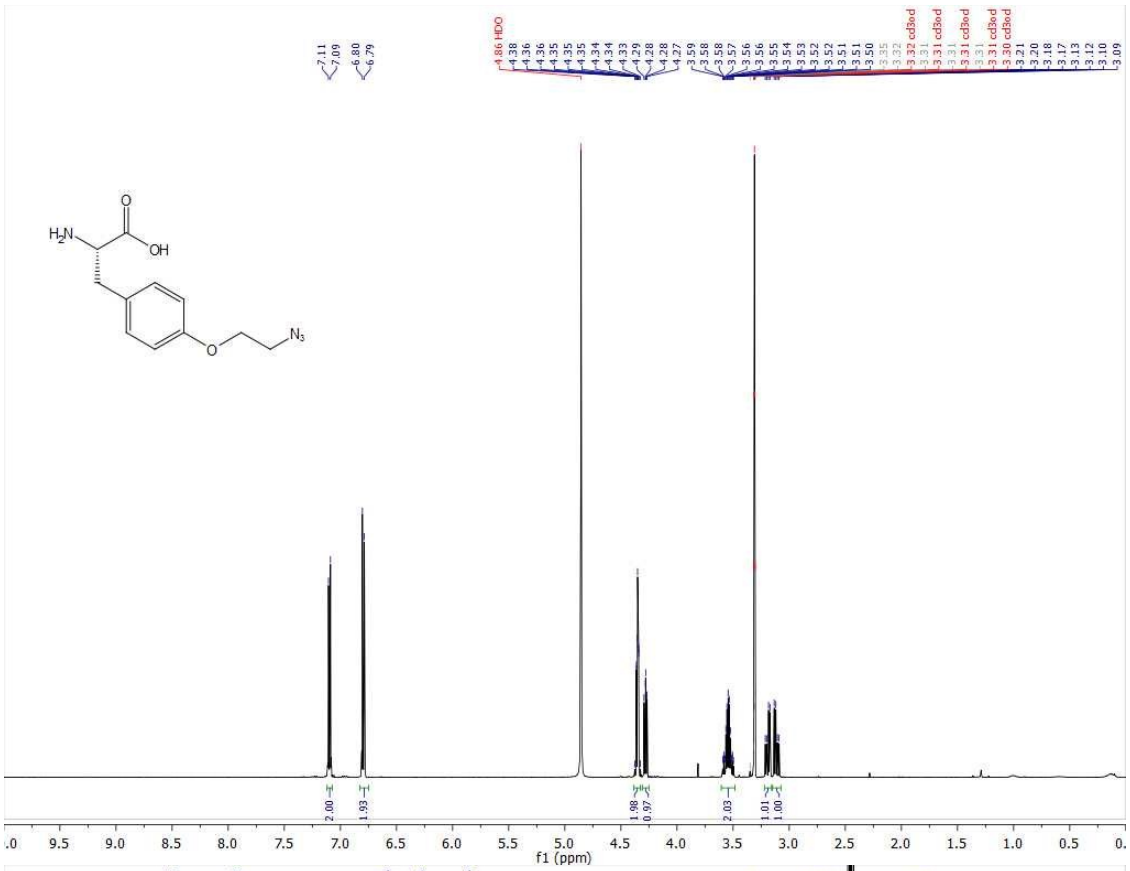
**Scheme S2.** Synthesis of O-2-aminoethyl-tyrosine (O2ameY).

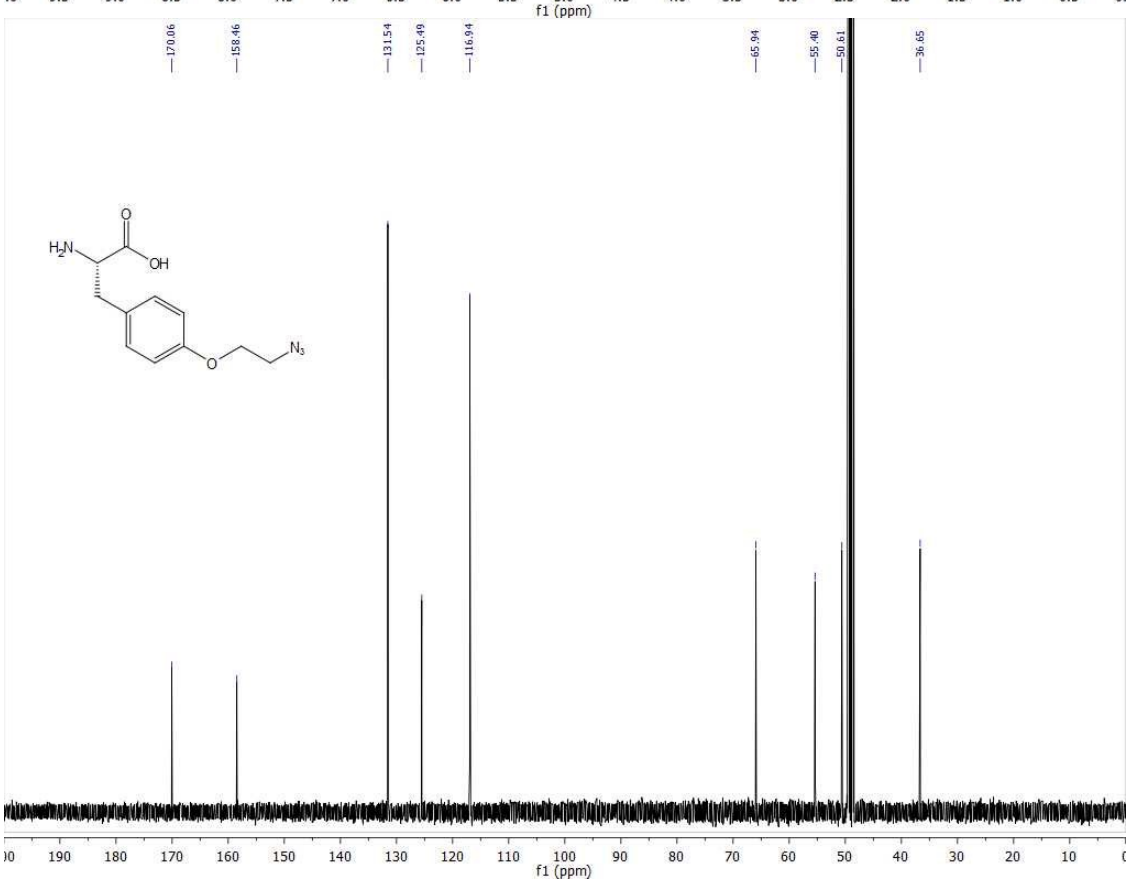
***Synthesis of O-2-aminoethyl-tyrosine, O2ameY***

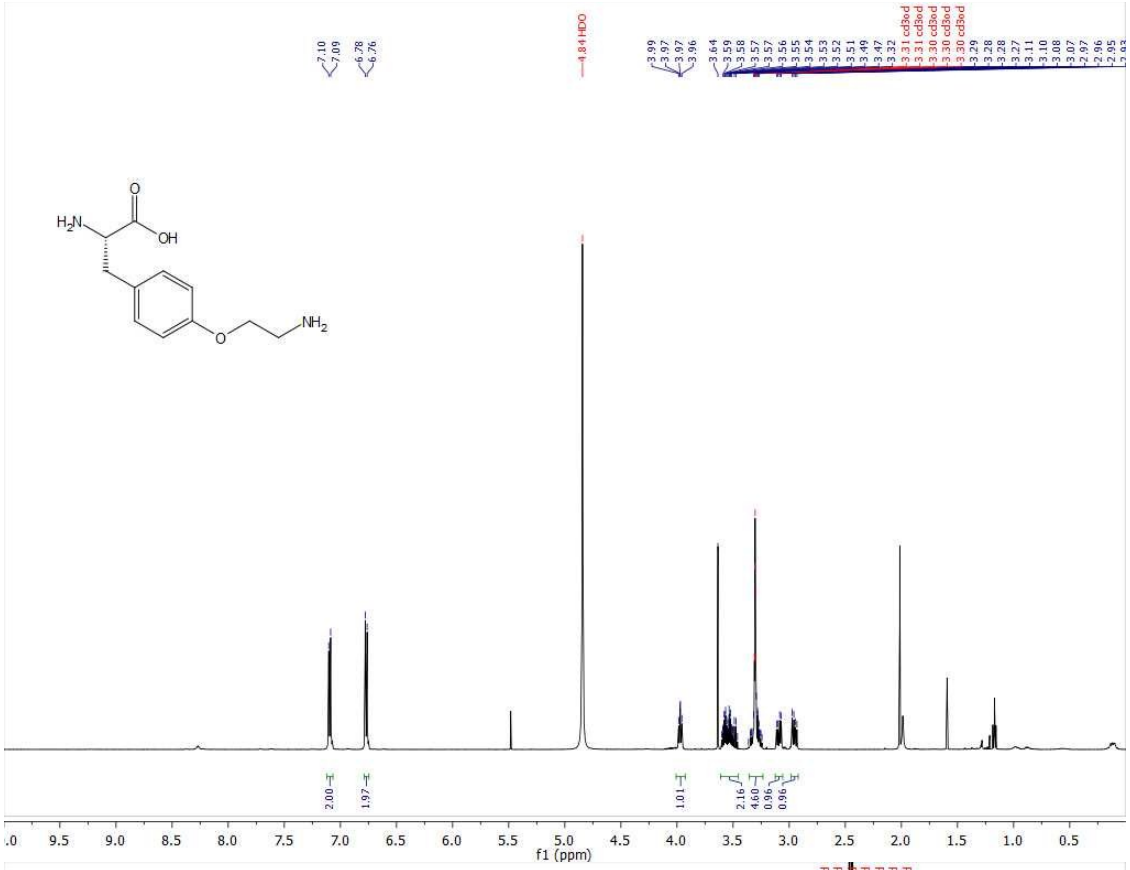
N-Boc-O-(2-azidoethyl)-tyrosine (**c**), (0.9 grams, 2.565 mmol, 1 eq) was dissolved in 25mL of THF in a 100 mL round-bottom flask and the solution was cooled to 0°C under argon. Triphenylphosphine (0.8 grams, 3.079 mmol, 1.2 eq) was added and the reaction was stirred at 0°C for 5 hours. 25 mL of de-ionized $H_2O$ was then added and the reaction was stirred for 16 hours at ambient temperature and then heated at 50 °C for an additional 3 hours. The solvent was removed *in vacuo* and the reaction mixture was extracted with EtOAc 3 x 50 mL and dried with $MgSO_4$. The volatiles were removed under reduced pressure to afford the crude product, which was purified on silica gel using 10:9:1 hexane: ethyl acetate: acetic acid as the solvent. The purified N-Boc-O-(2-aminoethyl)-tyrosine (**4**) was isolated as a yellow oil (0.367grams, 1.131

mmol, 44% yield). $^{1}$H NMR (500 MHz, 25 °C, $CDCl_3$): δ 7.07 (d, J = 8 Hz, 2H), 6.78 (d, J = 8.5 Hz, 2H), 5.11 (s, 1H), 4.21 (d, J = 6.5 Hz, 1H), 3.85 (s, 2H), 3.32 (s, 2H), 2.93 (dd, J = 8 Hz, 2H), 1.42 (s, 9H). MS (ESI) for $C_{16}H_{24}N_2O_5Na$ $[M+Na]^+$ calc: 347.38; $[M+Na]^+$ obs: 347.51.

N-Boc-O-(2-aminoethyl)-tyrosine (**d**) (160 mg, 0.493 mmol, 1 eq) was dissolved in a solution of 4 N HCl in dioxane (3 mL) at 0 °C. The reaction was stirred at room temperature for 1 hour. Following completion, the solvent was removed *in vacuo* and afforded the dihydrochloride salt of O-2-aminoethyl-tyrosine (O2ameY) as a yellow oil (146 mg, quantitative). $^{1}$H NMR (500 MHz, 25°C, $CD_3OD$): δ 7.10 (d, J=8.7 Hz, 2H), 6.79 (d, J= 8.3Hz, 2H), 3.97 (t, J=7.2 Hz, 1H), 3.60-3.45 (m, 2H), 3.34-3.24 (m, 2H), 3.09 (dd, J=13.6, 6.6 Hz, 1H), 2.95 (dd, J=13.9, 7.6 Hz, 1H). $^{13}$C NMR (126 MHz, 25°C, $CD_3OD$): δ 169.9, 158.3, 131.6, 126.1, 116.8, 61.2, 56.1, 43.0, 37.9. MS (ESI) for $C_{11}H_{17}N_2O_3$ $[M+H]^+$ calc: 225.15; $[M+H]^+$ obs: 225.47.


7.11
7.09
6.80
6.79
4.86 HDO
4.38
4.36
4.35
4.34
4.33
4.29
4.28
4.27
3.59
3.58
3.57
3.56
3.55
3.54
3.53
3.52
3.51
3.50
3.35
3.32
3.32 cd3od
3.31
3.31 cd3od
3.31
3.31 cd3od
3.31
3.31 cd3od
3.30 cd3od
3.21
3.20
3.18
3.17
3.13
3.12
3.10
3.09
2.00
1.93
1.98
0.97
2.03
1.01
1.00
f1 (ppm)


170.06
158.46
131.54
125.49
116.94
65.94
55.40
50.61
36.65
f1 (ppm)


7.10
7.09
6.78
6.76
4.84 HDO
3.99
3.97
3.97
3.96
3.64
3.59
3.58
3.57
3.57
3.56
3.55
3.54
3.53
3.52
3.51
3.49
3.47
3.32
3.31 cd3od
3.31 cd3od
3.30 cd3od
3.30 cd3od
3.30 cd3od
3.29
3.28
3.28
3.27
3.11
3.10
3.08
3.07
2.97
2.96
2.95
2.93
O
H2N
OH
NH2
O
2.00
1.97
1.01
2.16
4.60
0.96
0.96
f1 (ppm)

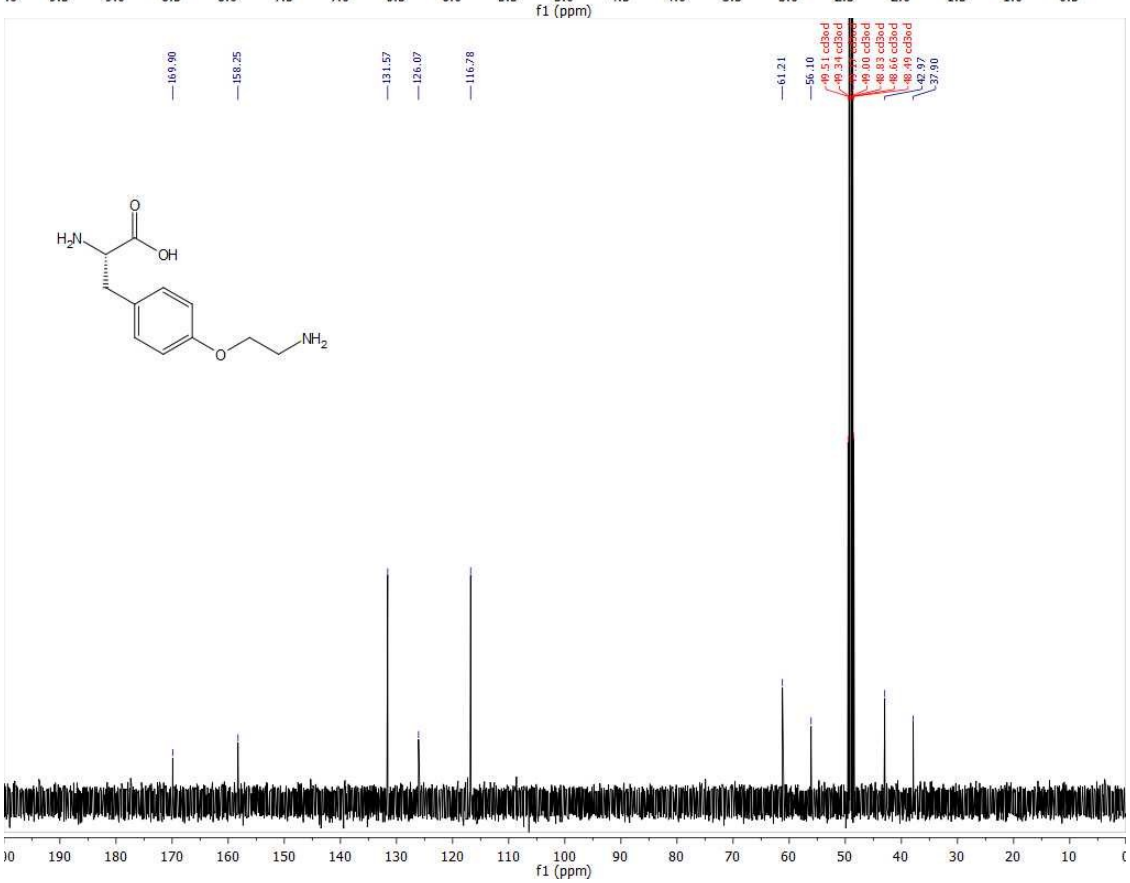

169.90
158.25
131.57
126.07
116.78
61.21
56.10
49.51 cd3od
49.34 cd3od
49.00 cd3od
48.83 cd3od
48.66 cd3od
48.49 cd3od
42.97
37.90
O
H2N
OH
NH2
O
f1 (ppm)

**Figure S1.** MALDI-TOF MS spectrum of precursor protein CBD-10mer(OpgY) after incubation with thiophenol under standard conditions. Only the acyclic product ('a') is observed indicating the lack of reactivity of the lysine comprised within the target peptide sequence toward inducing side-chain-to-tail cyclization under the applied conditions.

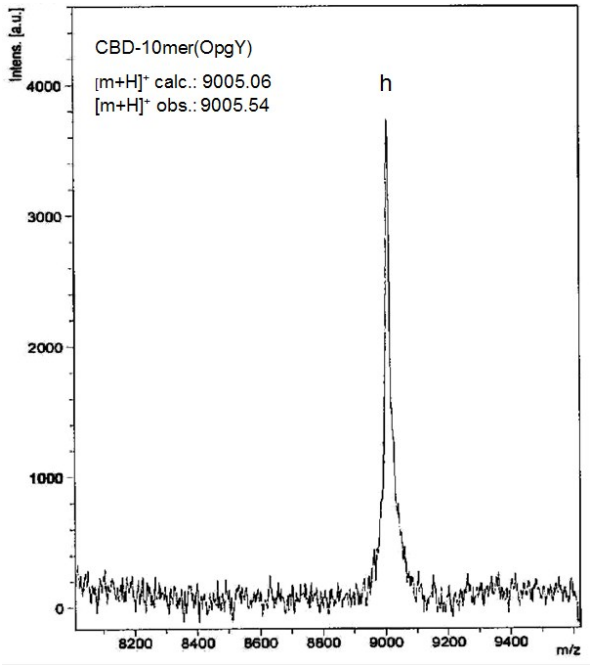


**Figure S2.** SDS-PAGE protein gels (left column) and densitometric analysis (graphs on right column) for benzyl mercaptan-induced cyclization reactions at pH 7.5, 8.2, and 9.0 at various time points. FL = full-length protein, GyrA = cleaved GyrA intein, CBD = CBD = CBD-fused macrocycle and acyclic product.

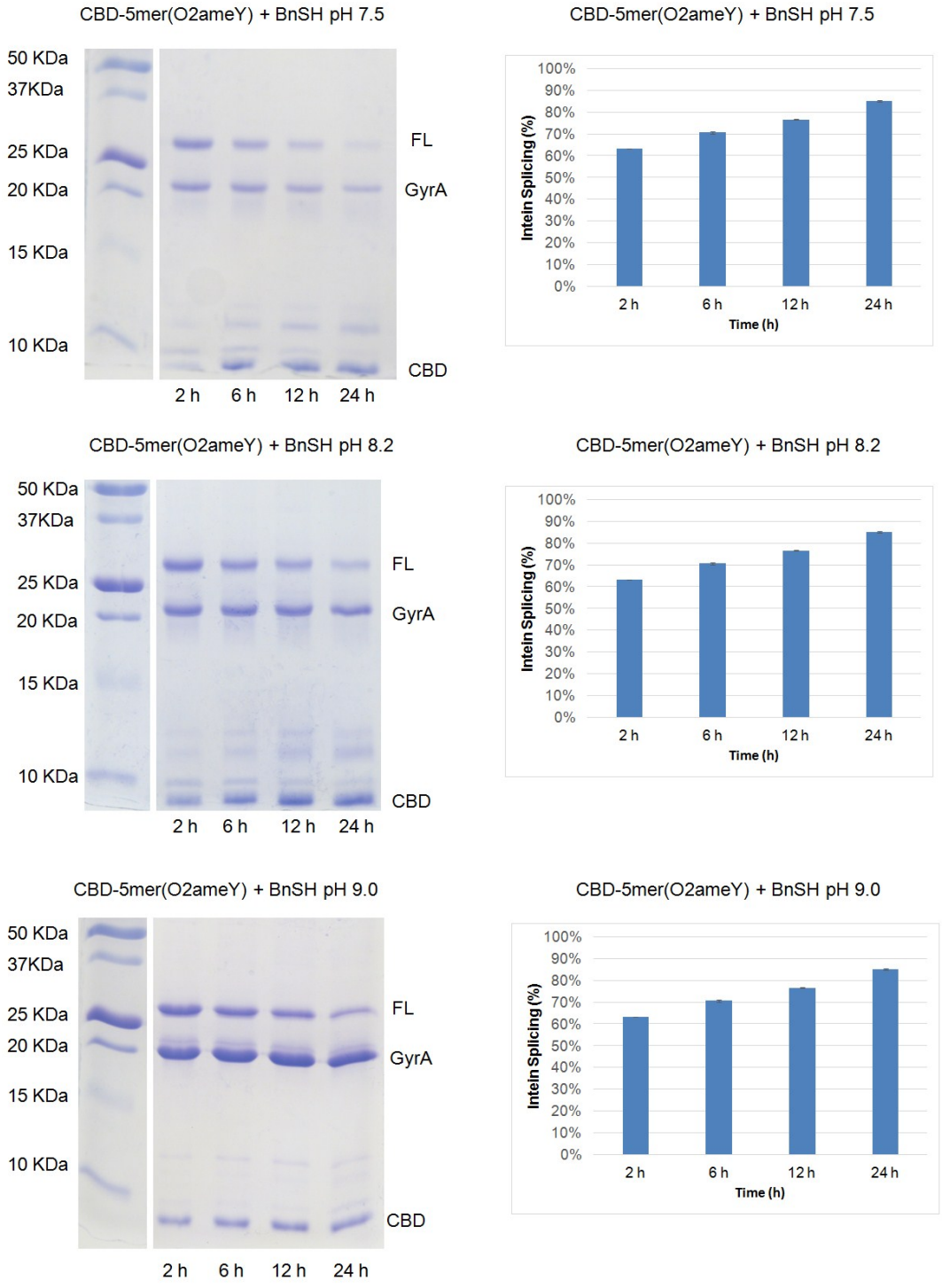


**Figure S3.** SDS-PAGE protein gels (left column) and densitometric analysis (graphs on right column) for thiophenol-induced cyclization reactions at pH 7.5, 8.2, and 9.0 at various time points. FL = full-length protein, GyrA = cleaved GyrA intein, CBD = CBD-fused macrocycle and acyclic product.

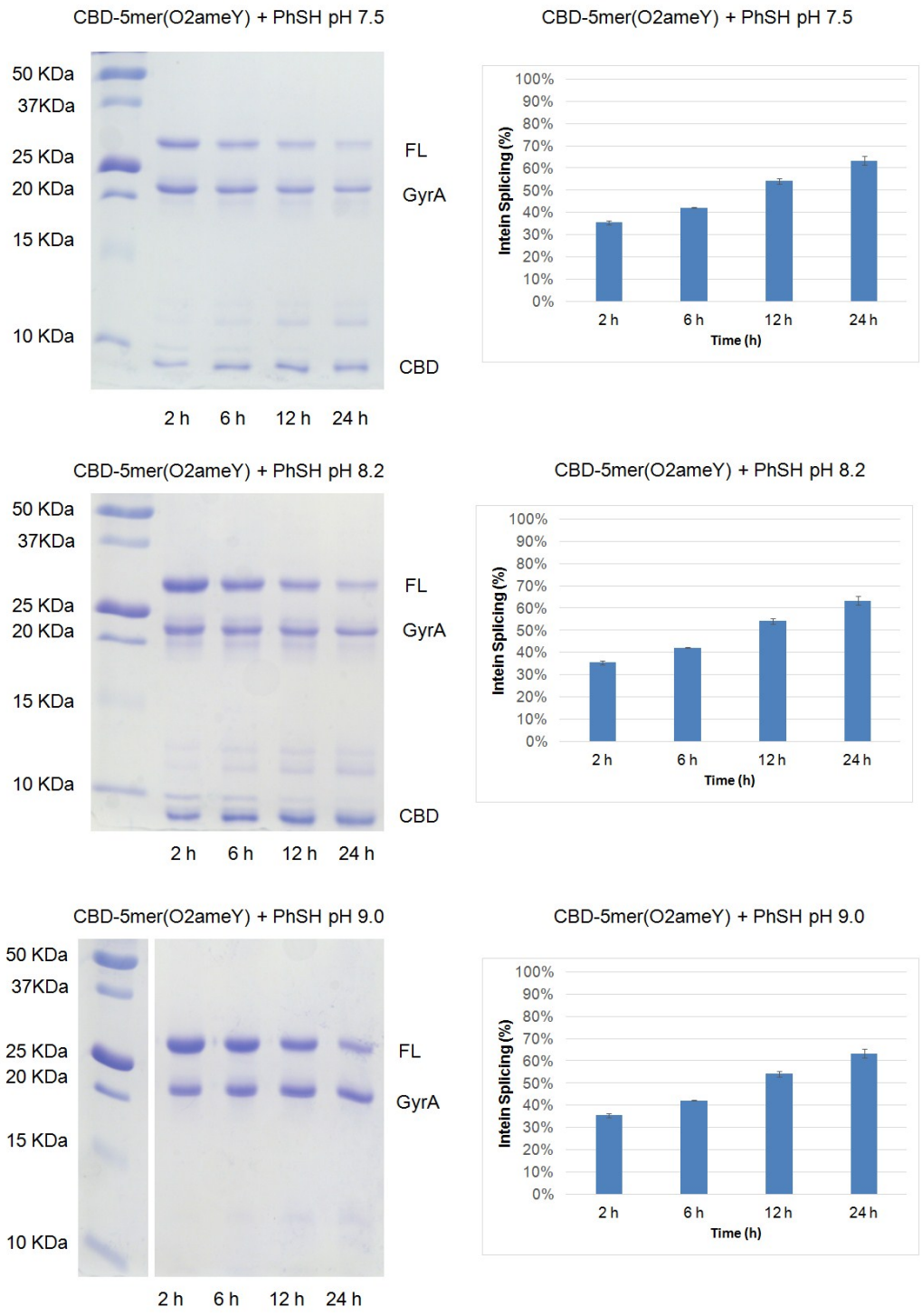


**Figure S4.** Representative LC-MS deconvolution spectra for large MW products from the cyclization reaction with 5mer(O2ameY) at different time points (2, 6, 12, 18, and 24 hours). FL = full-length protein, GyrA = cleaved GyrA intein.

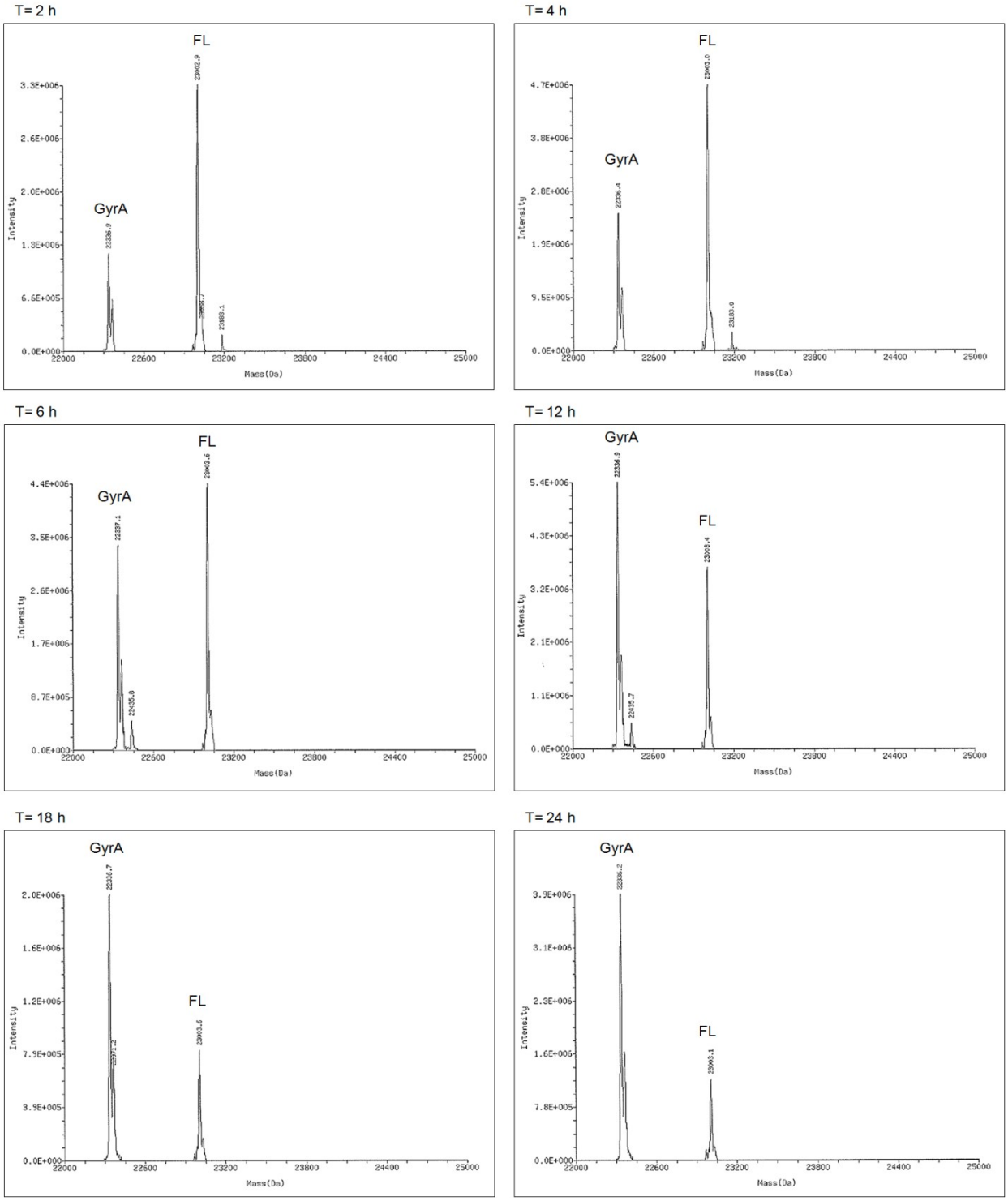
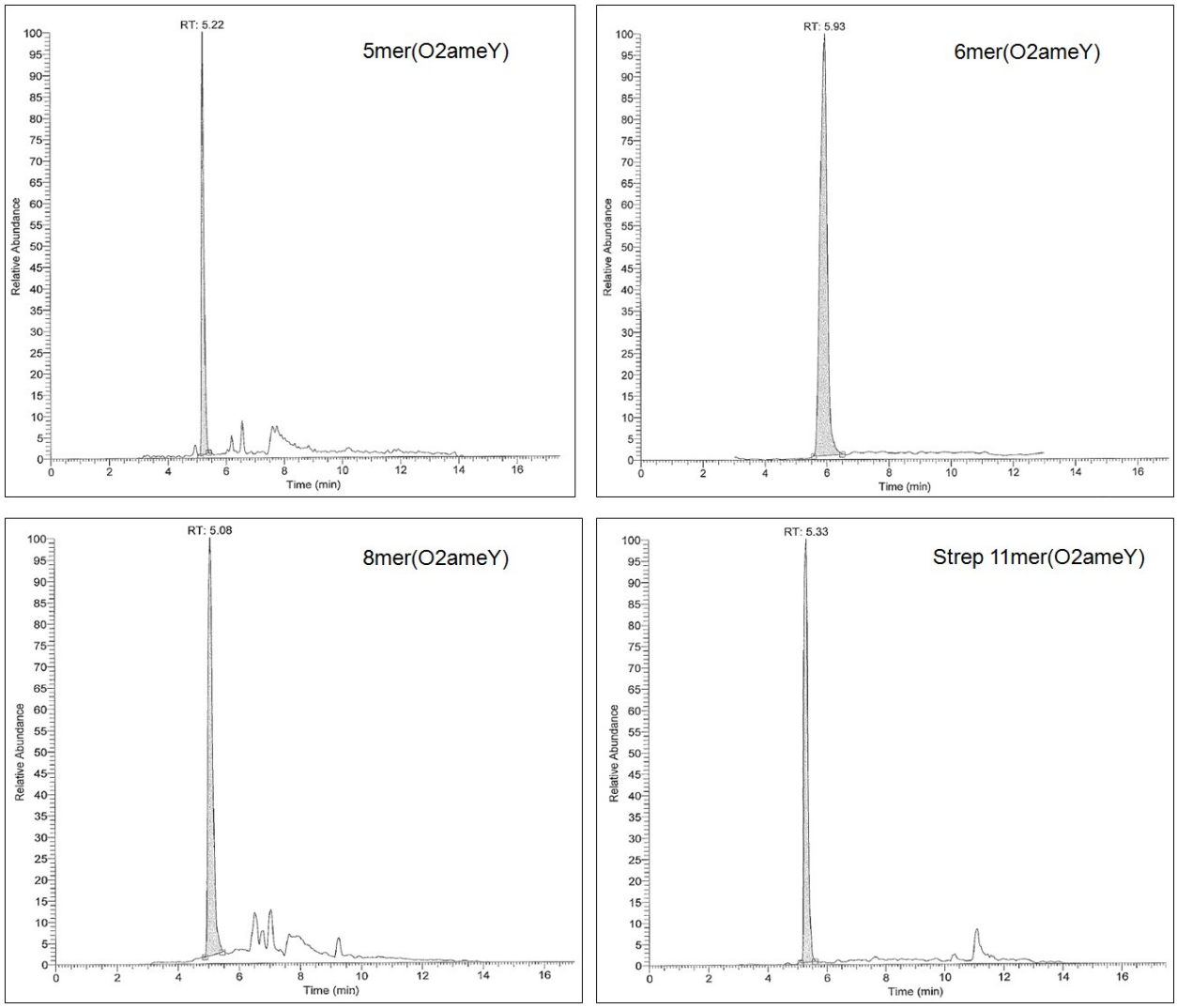



**Figure S5.** LC-MS ion extract chromatograms corresponding to O2ameY-containing macrocycles obtained via benzylmercaptan-induced cyclization of precursor proteins 5mer(OameY), 6mer(OameY), and 8mer(OameY), and Strep11mer(OameY).