Engineered biosynthesis of cyclotides

Thomas N. G. Handley , Hyon-Xhi Tan , Malcolm T. Rutledge , Hans Henning Brewitz , Joel D. A. Tyndall , Torsten Kleffmann , Margi I. Butler , Russell T. M. Poulter & Sigurd M. Wilbanks

View supplementary material

Published online: 26 Aug 2020.

Submit your article to this journal

Article views: 166

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Engineered biosynthesis of cyclotides

Thomas N. G. Handley [a], Hyon-Xhi Tan [a], Malcolm T. Rutledge [a],
Hans Henning Brewitz[a], Joel D. A. Tyndall [b], Torsten Kleffmann[a], Margi I. Butler [a],
Russell T. M. Poulter [a] and Sigurd M. Wilbanks [a]

[a]Department of Biochemistry, University of Otago, Dunedin, New Zealand; [b]School of Pharmacy, University of Otago, Dunedin, New Zealand

ABSTRACT

A system based on cyanobacterial split inteins, SICLOPPs (Split Intein Circular Ligation of Proteins and Peptides), has been used to synthesise a small natively cyclic plant protein, kalata B1, and cyclised versions of the natively linear therapeutic peptides ziconotide and leconotide. The cyclic versions of these naturally linear peptides include linker sequences between their native termini to allow the correct tertiary structure to form. The native structure of each includes three disulphide bonds characteristic of knottins. Cyclic permutations of leconotide yielded different proportions of correctly spliced product, identifying an optimal splice site and revealing the influence of residues around the splice junction. The rate of splicing was manipulated to facilitate affinity purification prior to intein-mediated removal of the affinity tag.

Introduction

Recombinant expression in *Escherichia coli* and other tractable hosts provides alternative means of producing useful variants of natural products, often with higher yields and easier purification than is achievable from the natural source. However, the development of bio-synthetic methods is required for production of peptides with unusual features such as a cyclic backbone. Split inteins, derived from cyanobacteria, allow cyclisation of recombinant polypeptides (Scott et al. 1999). We have characterised conditions that enhance their application to several therapeutic knottins, a group of peptidic bio-active compounds with complex cross-linked structures. This work aims to increase the appeal of using split inteins to cyclise peptides as a potential route to the generation of bioactive peptides for therapeutic applications.

Knottins are widely distributed peptides with a diverse range of useful biological activities (Wang et al. 2008; Vlieghe et al. 2010; Craik et al. 2013). Knottins contain a disulphide knot motif, in which two disulphide bonds and the adjacent peptide backbones form a ring penetrated by a third disulphide bond (Figure 1A) (Kolmar 2008). This yields a structure both highly cross-linked and topologically constrained. Cyclotides, a sub-class of
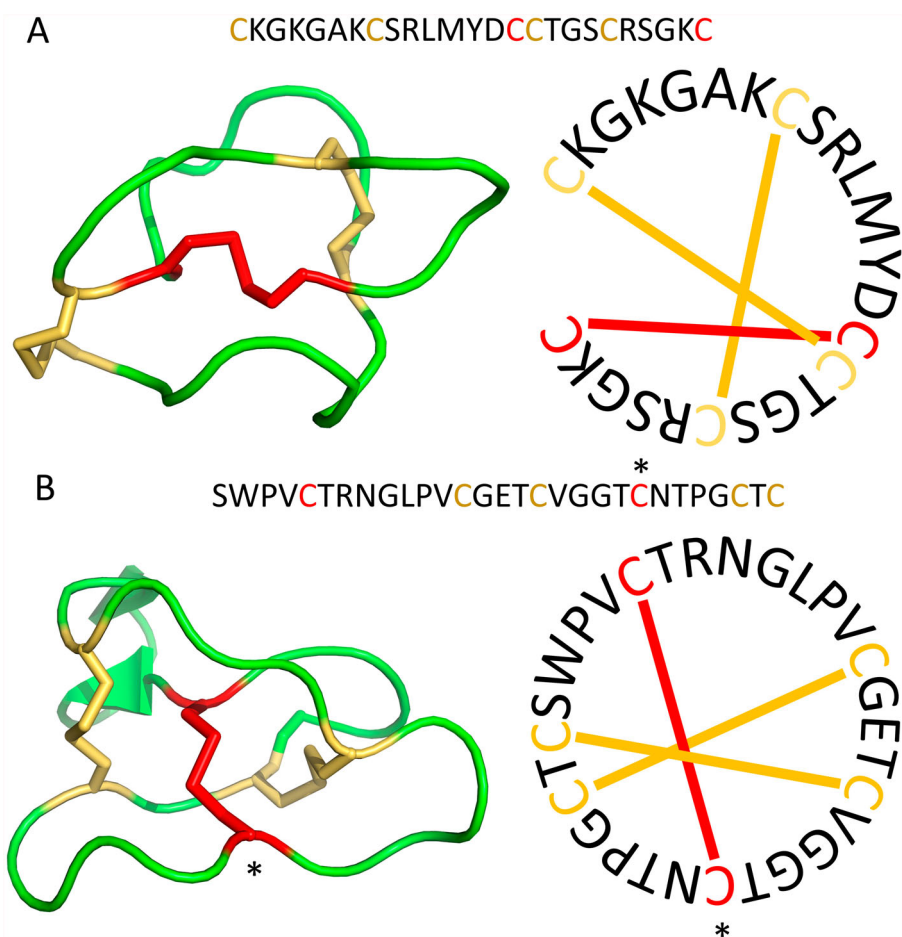
Figure 1. Structures and sequences of representative knottins and cyclotides. **A,** The sequence and structure of MVIIA (ω-conotoxin; Olivera et al. 1987). Disulphide bonds completing the ring of the knot motif are shown in yellow, the disulphide which penetrates the ring is shown in red. PDB: 1DW4. **B,** The sequence and structure of kalata B1. Colours as in panel A. Cysteine 3 is indicated with an *. PDB: 1KAL.

circularised knottins, have a disulphide knot and a continuous cyclic peptide backbone (Figure 1B) (Saether et al. 1995; Jagadish and Camarero 2010). The highly constrained structure and small size of cyclotides confers thermal stability and resistance to proteolytic degradation, making knottins and cyclotides attractive for use as bioactive peptides (Pallaghy et al. 1994; Craik et al. 2001; Li et al. 2015; Craik and Du 2017).

Different approaches have been applied to achieve practical synthesis of cyclotides for use as therapeutics. Naturally occurring cyclotides are formed intra-cellularly through a sequence-specific asparaginyl endopeptidase (Gillon et al. 2008; Harris et al. 2015; Shafee et al. 2015). Co-expression of asparaginyl endopeptidases allows production of cyclic peptides in non-native hosts such as tobacco (Poon et al. 2018), but the application to unnatural sequences has not been explored. Chemical synthesis of cyclotides requires potentially extensive and expensive optimisation of each target sequence due to the difficulties in synthesising both a thiol-rich molecule and the circular peptide backbone

(Fang et al. 2018). Biosynthesis of cyclotides is an appealing strategy as it has potential for application to the production of many different peptide sequences from a standard system with only minor optimisations. One such system is SICLOPPs (Split Intein Circular Ligation Of Proteins and Peptides) (Scott et al. 1999). This approach has been applied to cyclotides (Jagadish et al. 2013). SICLOPPs exploits a protein-splicing element, the split intein, to join the termini of the target peptide backbone into a circular molecule.

Inteins are protein elements that excise themselves from flanking protein sequences while simultaneously joining the flanking sequences with a peptide bond (Perler et al. 1994). Completion of the intein splicing requires a nucleophile provided by the side chain of a cysteine, serine or threonine in the "+1" position (Figure 2A) (Perler 2002). This residue remains in the joined product. Split inteins are derivatives of inteins in which the intein is split into two subunits which are translated at the separate termini of polypeptides; *in vivo* assembly of the entire intein from the two subunits restores activity, resulting in covalent fusion of the flanking peptides by a peptide bond, termed "splicing" (Figure 2B) (Wu et al. 1998). In SICLOPPs, the two split intein subunits are placed at the termini of a single polypeptide so that the intramolecular splicing reaction results in cyclisation of the target peptide (Figure 2C) (Scott et al. 1999). SICLOPPs has been used to generate libraries of cyclic peptides *in vivo* (Tavassoli and Benkovic 2004) and cysteine-rich peptides in *E. coli* including cyclotide MCoTI (Jagadish et al. 2013), sunflower trypsin inhibitor SFTI-1 (Li et al. 2016), θ-defensin RTD-1 (Bi et al. 2018), as well as a myriad of other applications (Tavassoli 2017).

In the present study, we applied the SICLOPPs system by purifying the full-length precursor molecule followed by *in vitro* splicing. This application allows the ready purification of cyclotides. SICLOPPs was applied to kalata B1 (a native cyclotide), as well as knottins ziconotide and leconotide (conotoxins MVIIA of *Conus magus* and CVID of *C. catus,* respectively) (Olivera et al. 1987; Lewis et al. 2000). These peptides were selected to test the capacity of SICLOPPs to generate cyclic peptide variants of naturally occurring linear peptides. Kalata B1 was the first cyclotide to be isolated (Gran 1973). Women of the Lulua people (in what is now the Democratic Republic of the Congo) used a 'kalata-kalata' decoction from leaves of *Oldenlandia affinis* to induce uterine contractions during childbirth. The active agent is a cyclic peptide – kalata B1, with 29 amino acids and three disulphide bonds forming a disulphide knot (Sletten and Gran 1973; Saether et al. 1995). Kalata B1 and other cyclotides have been used as scaffolds in drug design (Craik and Du 2017). Loop sequences can be replaced with a specific epitope to introduce a desired activity.

MVIIA and CVID are both ω-conotoxins identified in the venom of marine cone snails. Both peptides block the transmission of pain stimuli by inhibiting the activity of specific voltage-gated calcium channels (McGivern 2007; Write et al. 2000). Both MVIIA and CVID are knottins, but are linear rather than circular, limiting their potential application in a medicinal context as they are rapidly degraded *in vivo*. MVIIA is clinically administered intrathecally, under the brand name Prialt®. One means to improve the *in vivo* longevity of MVIIA and CVID would be by cyclising them. Cyclic variants of the linear α-conotoxin MII can be chemically synthesised that retain the native activity of the conopeptide (Clark et al. 2005). In that study, cyclisation with a linker peptide of 5, 6 or 7 residues conferred greater stability in serum and on exposure to enzymatic attack.
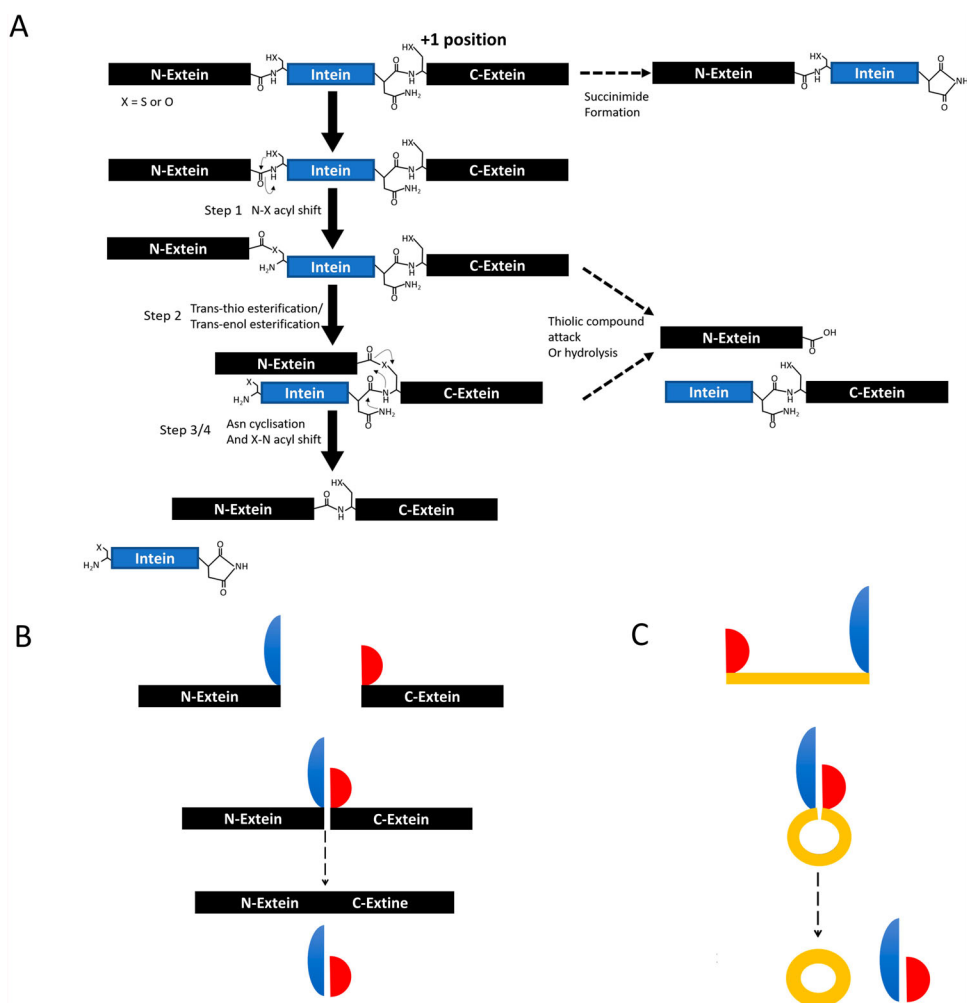
Figure 2. Inteins, split inteins and SICLOPPs. **A,** Mechanism of intein excision from the host molecule. The precursor molecule comprises two exteins (black) flanking the intein (blue). Intein excision enables joining of exteins by a peptide bond. Atomic structures of key catalytic residues are diagrammed. Step 1 – An N-S acyl rearrangement shifts the N-extein to the side chain of a cysteine residue at the 1 position, generating a linear thioester intermediate. Succinimide formation at the C-terminal junction is an abortive side reaction. Step 2 – A trans-esterification reaction cleaves the N terminal splice junction, transferring the N-extein to the side chain of the C-extein +1 cysteine to yield a branched intermediate. Step 3 – The C-terminal asparagine undergoes cyclisation, cleaving the downstream splice junction and releasing the excised intein. Step 4 - An S-N acyl rearrangement reversing step 1 to form a peptide bond joining the exteins. Both step 2 and 4 are sensitive to competition from thiolic compounds and nucleophiles, facilitating cleavage without ligation. Some naturally occurring inteins include serine or threonine in place of one or both cysteines at the 1 and +1 positions. **B,** Split intein-mediated peptide ligation. Two complementary half-inteins (N-intein shown in blue, C-intein in red) are synthesised in separate polypeptide chains, each joined to an extein (black). Reconstitution of the full intein from the halves allows splicing to yield the mature host peptide joined by a peptide bond through the mechanism shown in panel A. **C,** SICLOPPS cyclisation. In the precursor the separate extein are replaced by a single peptide of interest (yellow) that is flanked by the split intein subunits (coloured as in panel B). Splicing by the mechanism shown in the previous panels yields a cyclic peptide.

Kimura et al. (2006) produced kalata B1 via partial biosynthesis using native chemical ligation, a technique combining the first steps of the intein-mediated reaction with a non-enzymatic cyclisation. Thereafter, split inteins were used for biosynthesis and cyclisation in *E. coli* of the small, disulphide rich, knottin MCoTI, a natively cyclic peptide trypsin inhibitor originally isolated from dormant seeds of gấc melons (*Momordica cochinchinensis*) (Jagadish et al. 2013). The cyclotide was purified from cell lysate by affinity for trypsin agarose. A method incorporating a more general purification method would be of use. In both these reports, the intein-cyclised knottins were shown to spontaneously and efficiently fold into their native conformations.

Using the SICLOPPs system, we constructed split intein – cyclotide/knottin chimaerae (which we term cycloteins) that auto-process to generate cyclotides of kalata B1, MVIIA and CVID in *E. coli*. Splicing and cyclisation were confirmed by SDS-PAGE and intact mass spectrometry (MS). We further showed that engineered split inteins incorporating an affinity tag can be purified in the precursor form and induced to splice *in vitro*, producing cyclotides. We demonstrated the accumulation of splicing intermediates as well as the cyclic product, with relative yields dependent on growth conditions and reaction conditions.

Methods

Preparation of expression plasmids for cycloteins

An expression vector was constructed in the pET24a(+) vector (Novagen, USA) incorporating sequences encoding the C-intein, an α-complementation peptide, and the N-intein followed by a C-terminal hexahistidine tag. The α-complementation peptide was flanked by Mfe I and Afl II restriction sites. Coding sequences for different cyclotides were inserted between the C- and N-intein sequences as Mfe I/Afl II fragments, using blue/white screening in the presence of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside. Cyclotide coding sequences were prepared by chemical synthesis (Genscript, USA) and ligated into the Mfe I/AflII sites of the expression vector (Figure 3A). Plasmids encoded either both parts of the DnaE split intein from *Synechocystis* sp. 6803 (Wu et al. 1998), or the carboxyl-terminal part of that intein (SspDnaE-c) with the amino-terminal region of its homologue from *Nostoc punctiforme* PCC 73102 (NpuDnaE-n). The *N. punctiforme* PCC 73102 split intein has been shown to splice more quickly than the *Synechocystis* sp. 6803 intein (Zettler et al. 2009; Carvajal-Vallejos et al. 2012). The nucleotide sequences of the expression vectors were confirmed by Sanger sequencing of purified plasmids (https://gas.otago.ac.nz), prior to their use in transformation of *E. coli* expression strains.

Expression and detection of cycloteins

Expression of cycloteins was in *E. coli* strain BL21(DE3) (Novagen). Overnight cultures were inoculated from either single colonies of primary transformants or glycerol stocks made from overnight cultures. Expression cultures were inoculated at a 1:100 dilution and grown with shaking at 250 rpm in lysogeny broth (10 g/L Bacto™ tryptone, 5 g/L yeast extract, 10 g/L NaCl, autoclaved for 15 min at 15 psi.) at 37°C with agitation at
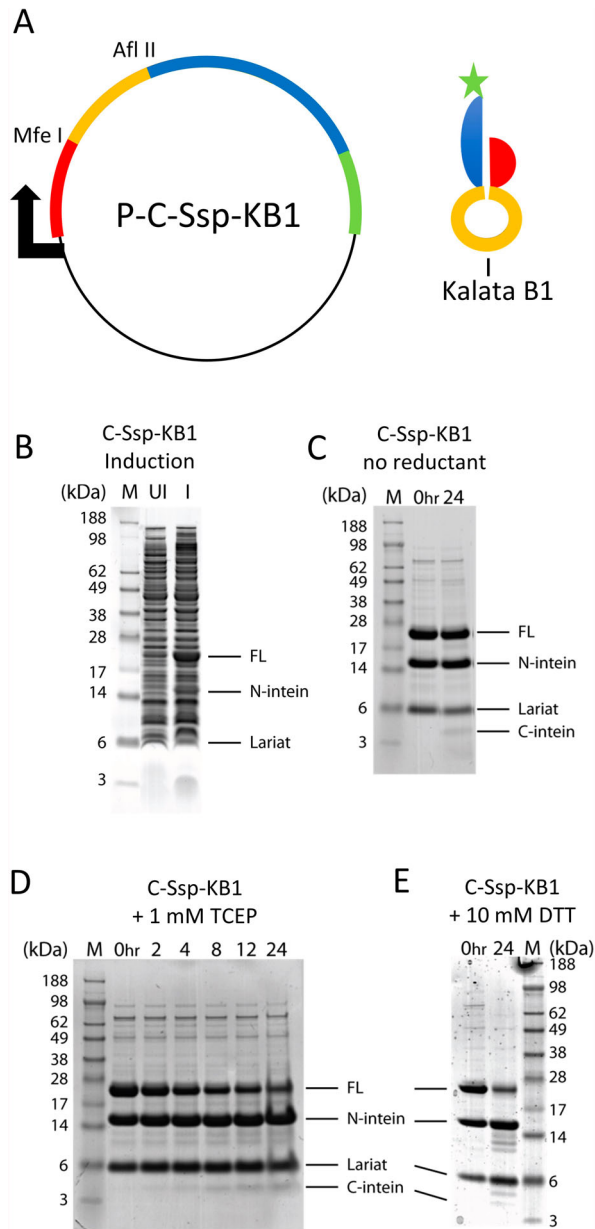
Figure 3. SICLOPPs mediated synthesis of kalata B1. **A,** A diagrammatic representation of a SICLOPPs genetic construct, P-C-Ssp-KB1 and the expressed protein C-Ssp-KB1, containing kalata B1 as the peptide to be cyclised (yellow), and N- and C-inteins (coloured as in Figure 2). The green bar and star mark the positions of the hexahistidine tag used for IMAC isolation of this peptide. **B,** Expression of SICLOPPs kalata B1. Cell lysates of the uninduced sample (UI) and induced sample (I) were analysed by SDS-PAGE. Bands of interest are identified at right; FL, full-length kalata B1 – cyclotein/SICLOPPs construct (23 kDa); N-intein (15 kDa); and lariat, comprising the C-intein and extein side-product derived from the branched intermediate shown in Figure 2A (6 kDa). M contains a molecular weight marker with masses in kDa indicated to the side. **C–E,** *In vitro* splicing analysis of affinity-purified C-Ssp-KB1. Splicing was analysed in the absence (panel C) and the presence of reducing agents TCEP (panel D) or DTT (panel E). Incubation times (in hours) are shown above lanes. Markers and bands of interest are as in panel B with addition of C-intein (4 kDa).

250 rpm. When they reached an OD_{600} of 0.4, cultures were transferred to the desired expression temperature and the inducer isopropyl-β-D-1-thiogalactopyranoside was added to a final concentration of 1 mM. After further incubation with agitation, OD_{600} was measured, cells were pelleted by centrifugation and stored frozen.

For SDS-PAGE analysis, cell pellets from 1 mL of culture were resuspended in 100 μL of SDS-loading buffer for each unit of OD_{600}, to give a consistent concentration of protein. Samples of 20 μL of resuspended cells were boiled and separated on NuPage gels (Invitrogen) and visualised according to manufacturer's instructions.

Enrichment of split intein constructs

Immobilised metal affinity chromatography (IMAC) was employed to enrich polypeptides and complexes that included the hexahistidine tag adjacent to the carboxyl terminus of the N-intein. Cell pellets were resuspended in lysis buffer (300 mM NaCl, 50 mM Na_2HPO_4/ NaH_2PO_4 pH 7) at an OD_{600} of 10–20, and sonicated repeatedly for 15 s at 80% power until 80% reduction in OD_{600} was achieved, or until sonication resulted in no further reduction (Branson Ultrasonics, Danbury, CT, USA). Intact cells and debris were pelleted by centrifugation (118,000 m s^{-2}, 600 s) and the supernatant fraction was applied directly to Ni-NTA agarose (Qiagen). Weakly-bound proteins were removed with wash buffer (300 mM NaCl, 50 mM Na_2HPO_4/NaH_2PO_4, 10 mM imidazole pH 7) before elution of fusion proteins with elution buffer (300 mM NaCl, 50 mM Na_2HPO_4/NaH_2PO_4, 150 mM imidazole pH 7). When indicated, the buffer was exchanged by extensive dialysis (at least three changes of 100-fold dilution). Enriched samples were either stored frozen or incubated at indicated temperatures to promote splicing and circularisation, before analysis by SDS-PAGE.

Tryptic digestion

Samples of cyclotein that had been enriched by IMAC and incubated overnight were digested with trypsin according to Shevchenko (Shevchenko et al. 2006). In brief, protein samples were reduced, alkylated and digested with trypsin in ammonium bicarbonate buffer in a ratio of 1 μg of protease per 10 μg of protein at 37°C for 15 h. The reaction was stopped with 1 μl of 10% (v/v) trifluoroacetic acid and then dried using a centrifugal vacuum concentrator and stored at −20°C.

Mass spectrometry analyses

Samples were co-crystallised with α-cyano-4 hydroxy-cinnamic acid and analysed on a matrix-assisted laser desorption and ionisation (MALDI) tandem Time-of-Flight (TOF/ TOF) Analyzer (4800 MALDI-TOF/TOF, AB Sciex, MA). MS and MS/MS default calibration settings were updated on eight calibration spots for each operation mode. All MS spectra were acquired in positive-ion mode with 1000 laser pulses per sample spot and inspected for peaks with a predicted $[M + H]^+$. Peptide ions of interest were subjected to collision-induced dissociation (CID) MS/MS analysis. CID fragment ion spectra were acquired with up to 4000 laser pulses per selected precursor using the 2 kV positive-ion mode and air as the collision gas at a pressure of 1×10^{-6} torr.

To identify the site of cyclisation, the cyclic form of kalata B1 was first linearised by tryptic digestion resulting in the cleavage of a single tryptic cleavage site between arginine and asparagine, which is different to the site of cyclisation. The linearised form was then subjected to CID MS/MS and fragment spectra were searched against the target sequences using an in-house Mascot search engine (http://www.matrixscience.com). To identify cyclic kalata B1 the sequence of the tryptically linearised form was used: NGLPVCGETCVGGTCNTPGCTCSWPVCTR. To identify linear kalata B1 after trypsinolysis the following two additional peptide sequences were included: NGLPVCGETCVGGT and CNTPGCTCSWPVCTR. A modified kalata B1 cyclotein was also analysed using the same procedure. This modified cyclotein sequence was predicted to produce the following peptide NGLPVCGETCVGEYCFNIGCTCSWPVCTR following cyclisation and tryptic cleavage. Without cyclisation, tryptic digestion would yield the sequences NGLPVCGETCVGEY and CFNIGCTCSWPVCTR. The search was restricted to tryptic peptides with precursor mass tolerance of 75 ppm and the maximum fragment mass error of 0.4 Da. Carboxyamidomethyl cysteine, oxidised methionine, pyroglutamic acid (E, Q) and the cysteine modifications sulfenic acid, sulfinic acid, sulfonic acid and cystine formation with free cysteine were included as variable modifications.

Results

Recombinant biosynthesis of cyclic kalata B1

To explore the ability of the SICLOPPs system to generate small cyclic peptides, kalata B1 was selected as the target peptide since this was previously produced by native chemical ligation (Kimura et al. 2006). An expression vector was constructed with sequences encoding the kalata B1 flanked by sequences encoding the C-intein and N-intein of the DnaE split-intein from *Synechococcus* sp. PCC6803 (Figure 3A). Split intein protein requires a cysteine in the "+1" position splicing (Figure 2A). Inspection of the kalata B1 crystal structure (PDB: 1KAL) identified a cysteine exposed on the surface of the molecule (Cysteine 3 -Figure 1B) that was selected to act as the nucleophile in the +1 position as it was expected to be accessible for splicing by the split intein in near-native conformations of the kalata B1 peptide.

Expression and splicing of this cyclotein were expected to yield novel molecules corresponding to the C-intein, the cyclotide and the N-intein. Expression in *E. coli* BL21(DE3) cells led to the accumulation of species that migrated on SDS-PAGE as expected for full-length cyclotein polypeptide (23 kDa), and the N-intein, generated after intein splicing (16 kDa, Figure 3B). The observed bands suggested that this SICLOPPs construct could carry out splicing and cyclisation of the target peptide *in vivo*. The C-intein and cyclised kalata B1 (~4 and 3 kDa respectively) were not resolved in these crude cell lysates, likely because of their small size.

To characterise the reaction products, proteins containing the N-intein were purified by IMAC. This protocol was expected to recover full-length cyclotein polypeptides containing the hexahistidine tag (Figure 3A), as well as the free N-intein peptide. Since the two intein moieties are tightly (non-covalently) bound, purification of the N-intein can also purify any C-intein and lariat peptide. The lariat peptide corresponds to the intermediate

step in which the N-intein has been cleaved but the cyclotide and C-intein remain covalently joined.

To investigate whether the purified full-length polypeptide was able to splice *in vitro*, the effect of incubation of purified samples was analysed (Figure 3C). The formation of three disulphide bonds within the native kalata B1 will sequester the catalytic cysteine; to avoid the resulting inhibition and permit more efficient splicing, reductants were added to the reaction mixture. Intensity of the C-intein band increased in the presence of TCEP (Figure 3D) or DTT (Figure 3E) at 22°C. Unidentified bands formed in the presence of DTT, suggesting the generation of stable adducts; as a result, DTT was not used in further experiments. A band corresponding to the expected position of kalata B1 was not observed; this is likely due to the small size and amino acid composition of kalata B1, which do not support adequate staining with Coomassie dye.

The effect of pH on splicing was investigated (Figure 4A), using the yield of free C-intein to quantify intein splicing. The yield of free C-intein, relative to full-length polypeptide, was improved by reduction of pH to 5.5 (Figure 4A). Below this pH, there is a decline in the abundance of all species, suggesting impaired solubility. Splicing also increased with increasing temperature (Figure 4B). However, incubation at 37°C led to the generation of additional, unexpected product bands. Incubation at 22°C yielded an acceptable splicing rate with limited heterogeneity of products.

To detect the generation of cyclic kalata B1 during incubation, products of *in vitro* splicing reactions (22°C, pH 5.5 and with the addition of 10 mM TCEP) were analysed by MALDI-TOF mass spectrometry. Masses consistent with mature cyclic kalata B1 were observed (Figure 5). Peak clusters were observed for isotope envelopes of cyclic and linear kalata B1 and disulphide bond variants (i.e. reduced and oxidised forms of kalata B1). The majority of kalata B1 was detected in the reduced form. A peptide with the correct mass to be cyclic kalata B1 was present (singly charged ion at m/z 2897.03). The presence of the reduced linear form of kalata B1 (singly charged ion at m/z 2915.04) suggested that abortive side reactions were occurring. The difference in mass between the reduced cyclic kalata B1 and reduced linear kalata B1 is 18 Da, corresponding to the incorporation of water during hydrolysis of the lariat, potentially competing with the C–N acyl shift (Figure 2A). CID MS/MS on the tryptically linearised form was used to identify the splice site (Supplementary Figure 1). This analysis revealed a y-ion series that bridges the splice junction of cyclic kalata B1, confirming an intact peptide bond at the splice junction between threonine and cysteine and thus the generation of cyclic kalata B1 with this SICLOPPs system.

SICLOPPS splicing of native kalata B1 generated significant amounts of unwanted linear peptides. The kalata B1 sequence around the splice junction was altered to closely resemble the native splice junction of the SspDnaE split intein, testing the hypothesis that native flanking sequences permit more efficient splicing and a higher proportion of cyclic peptide. Mass spectrometry reveals masses corresponding to the cyclic modified kalata B1 and the C-intein (Figure 6A). There is more cyclic relative to linear peptide formation in the modified kalata B1, although there is still a significant amount of the linear sequence. Figure 6A indicates the mass of the cyclic and linear peptides differs by 18 Da. To confirm the interpretation of these
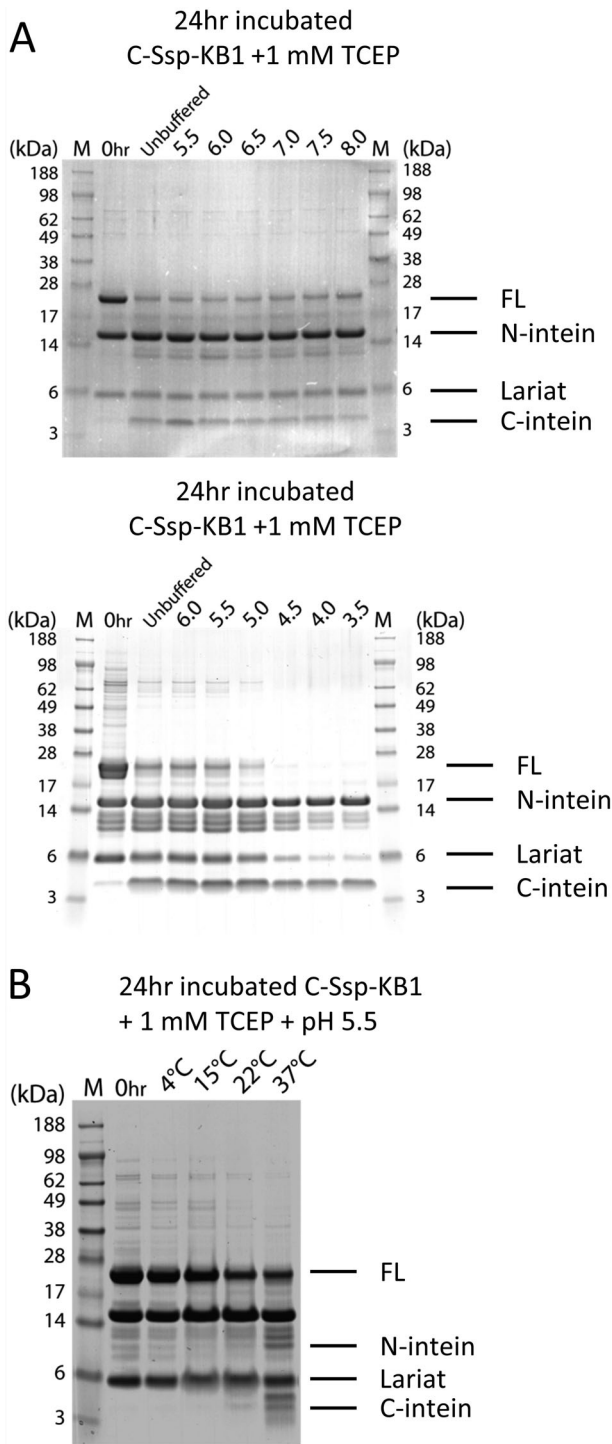
Figure 4. Effects of pH and temperature on SICLOPPs mediated synthesis of kalata B1. **A,** Affinity-purified precursors were incubated 24 h at the pH indicated above each lane (over a range of 3.5–8). **B,** Affinity-purified precursors were incubated 24 h at pH5.5 and the temperature indicated above each lane. In both panels markers and bands of interest are indicated as in Figure 3.
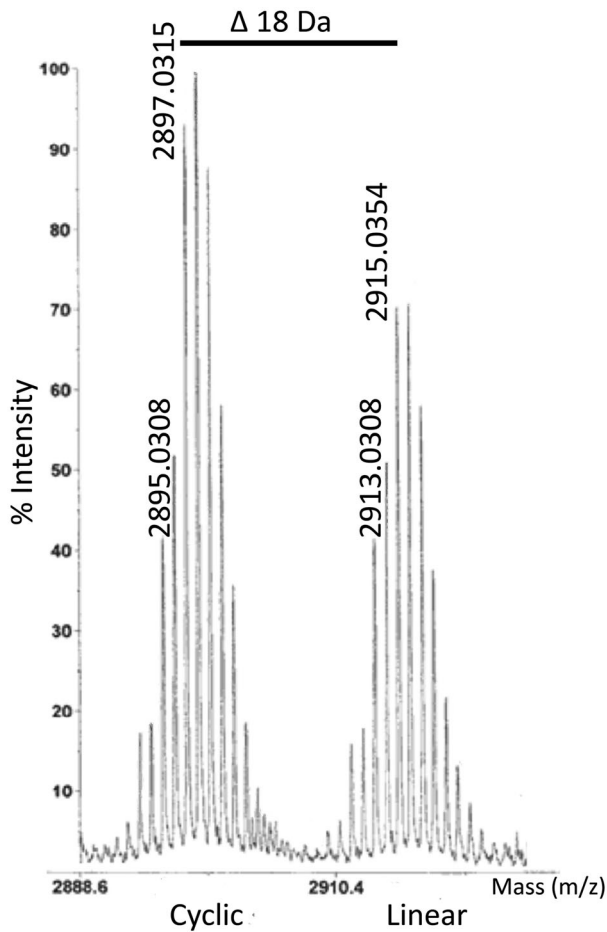
Figure 5. Mass-spectrometry of C-Ssp-KB1 splice products. A sample of C-Ssp-KB1 was incubated for 24 h at 22°C with 1 mM TCEP and the products were analysed by MALDI-TOF. Two peak clusters of oxidised and fully reduced singly charged molecular ions of intact kalata B1 were observed. The signal at m/z 2894.03 of the first peak cluster corresponds to the intact cyclic form of kalata B1 still containing one disulphide bond. The fully reduced intact cyclic form of kalata B1 is 2 Da bigger and shows a signal at m/z 2897.03. The two peaks at m/z 2913.03 and 2915.04 of the second peak cluster correspond to the oxidised and fully reduced linear form of kalata B1 each being 18 Da bigger than the corresponding cyclic molecules.

peaks, the preparation was treated with trypsin (Figure 6B). During a 30-minute trypsin digest, the relative quantity of the cyclic form decreased while a linear form increased as the cleavage of the cyclic kalata B1 at a single site generated a linear form with a mass increase of 18 Da. The presence of the splice junction in the trypsin-generated linear peptide was confirmed by CID MS/MS mass spectrometry (Supplementary Figure 2). While the modified kalata B1 cyclotein gave some increase in the amount of cyclic form, there was still a significant amount of the linear form produced.
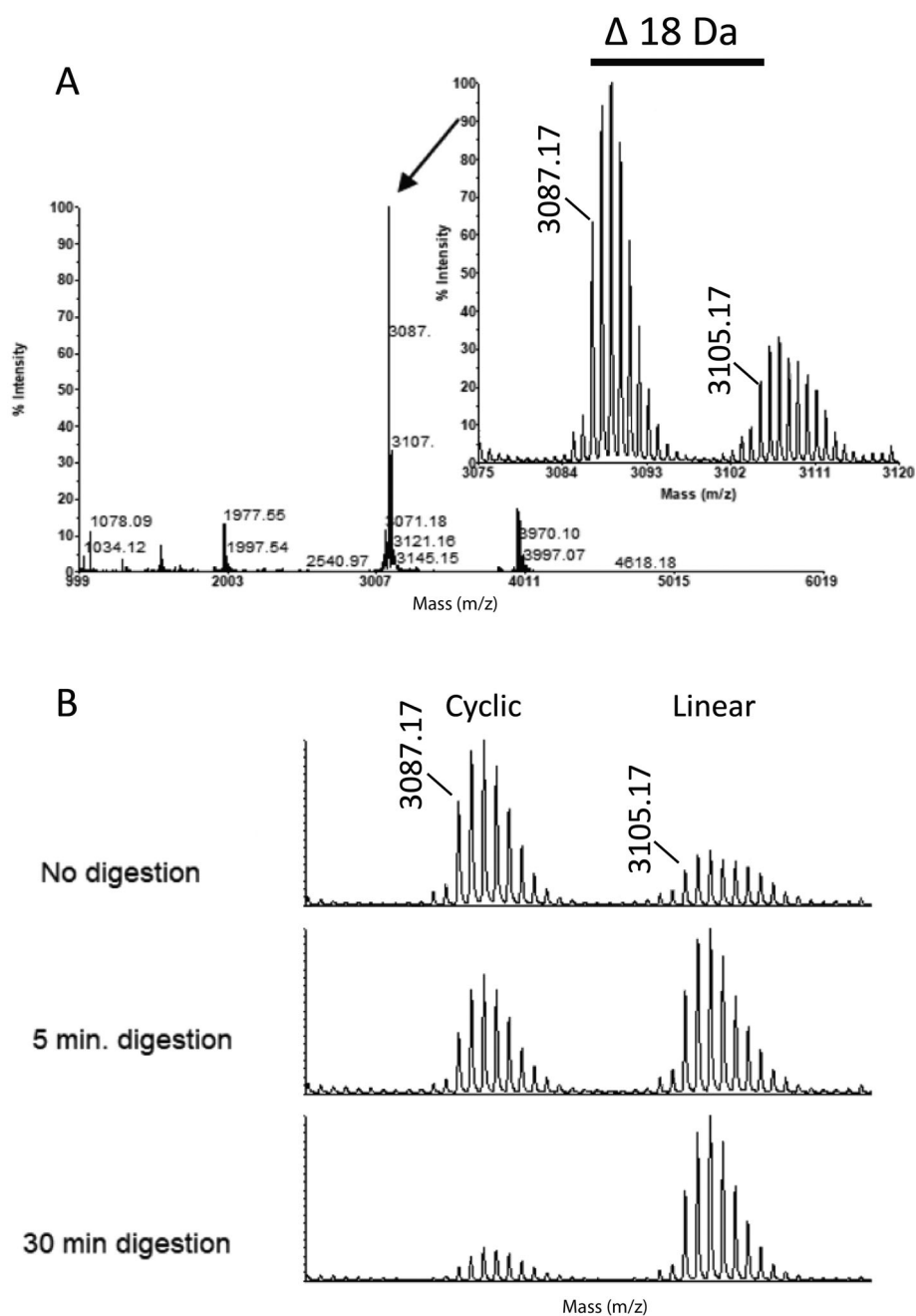
Figure 6. Mass-spectrometry of modified kalata B1. **A,** Modified kalata B1 was generated using identical conditions identified for native kalata B1, 24 h at 22°C with 1 mM TCEP and the products were analysed by MALDI-TOF. Two peak clusters of singly charged molecular ions of intact modified kalata B1 were observed. The signal at m/z 3087.17 of the first peak cluster corresponds to the intact cyclic form of modified kalata B1 containing all three disulphide bonds. The peak at m/z 3105.17 of the second peak cluster corresponds to the linear form of modified kalata B1 containing all three disulphide bonds. It is 18 Da bigger than the corresponding cyclic molecule. **B,** Relative peak intensities of the cyclic and linear forms of modified kalata B1 at different time points during the trypsin digestion.

Generation of cyclic variants of a naturally linear knottin, MVIIA

Next, we assessed SICLOPPs-mediated cyclisation of a natively linear knottin, the ω-conotoxin MVIIA. To bridge the distance between the termini of MVIIA (∼10 Å in PDB: 1DW4) and preserve the native tertiary structure of MVIIA when cyclised, two different sequences were incorporated: TSNG or TSNGL. Translation of fusions at either the amino or carboxyl terminus of MVIIA (Figure 7A) were expressed in *E. coli* BL21 (DE3) (Figure 7B), the hexahistidine-tagged proteins isolated with IMAC, and splicing *in vitro* was determined (Figure 7C).

The MVIIA SICLOPPs constructs showed efficient induction of the full-length cyclotein and gave satisfactory homogeneity of peptide following IMAC purification (Figure 7B,C). Splicing was detectable both as a reduction in full-length cyclotein and the presence of bands corresponding to the C-intein. Cyclised MVIIA peptides generated by splicing of the carboxyl-terminal fusion with the TSNG linker (C-Ssp-Cp C_4) were detected by intact MS (Figure 7D). However linear MVIIA peptide was also detected, suggesting non-enzymatic hydrolysis.

Generation of a cyclic variant of CVID using permutations of the cyclotein splice junction

The amino acids around the intein splice junction affect the splicing reaction (Cheriyan et al. 2013). In work exploring the capacity of native chemical ligation to generate cyclic peptides of kalata B1 from different linear permutations, it was apparent that each of the potential splice junctions had significantly different splicing capacities (Kimura et al. 2006). We used permutations to assess the capacity of each of the six cysteine residues forming the knottin motif in CVID to serve as the +1 cysteine in the splicing reaction of SICLOPPS cycloteins (Figure 8A). CVID is a linear peptide; to permit cyclisation without distortion of the native structure, a five amino acid linker peptide (FNIGL) was included to bridge the native termini of the target peptide. Splicing capacities *in vivo* were explored by SDS-PAGE, where the splicing efficiency of each of the permuted SICLOPPs variants was evaluated by the relative intensity of pre-splicing product bands (full-length cyclotein) to the post splicing product bands (N-intein and lariat) in the induced total cell lysate (Figure 8B). Permuted to the +1 position, the cysteine adjacent to the introduced linker, position 27, supports a considerable yield of free N-intein as do positions 15 and 20, while positions 1 and 16 show much less. Position 8 accumulates neither full-length polypeptide nor free N-intein or C-intein. To exploit faster splicing reported for the NpuDnaE split intein from *Nostoc punctiforme* (Iwai et al. 2006), a hybrid cyclotein system was constructed consisting of the SspDnaE C-intein, the permuted CVID sequence and the NpuDnaE N-intein. The hybrid split intein showed higher *in vivo* splicing efficiency, demonstrated by reduced full-length cyclotein and increased amounts of N-intein and C-intein as compared to the SspDnaE system (Figure 8C). The intensity of the lariat band is similar in the two systems. This suggests that, as expected, the hybrid system supports greater *in vivo* splicing and in consequence reduced full-length cyclotein in the cell lysate. The hybrid split intein at position 27 demonstrates the most efficient *in vivo* splicing, with negligible full-length cyclotein or lariat and strong N- and C-intein bands (Figure 8C). While five of the six
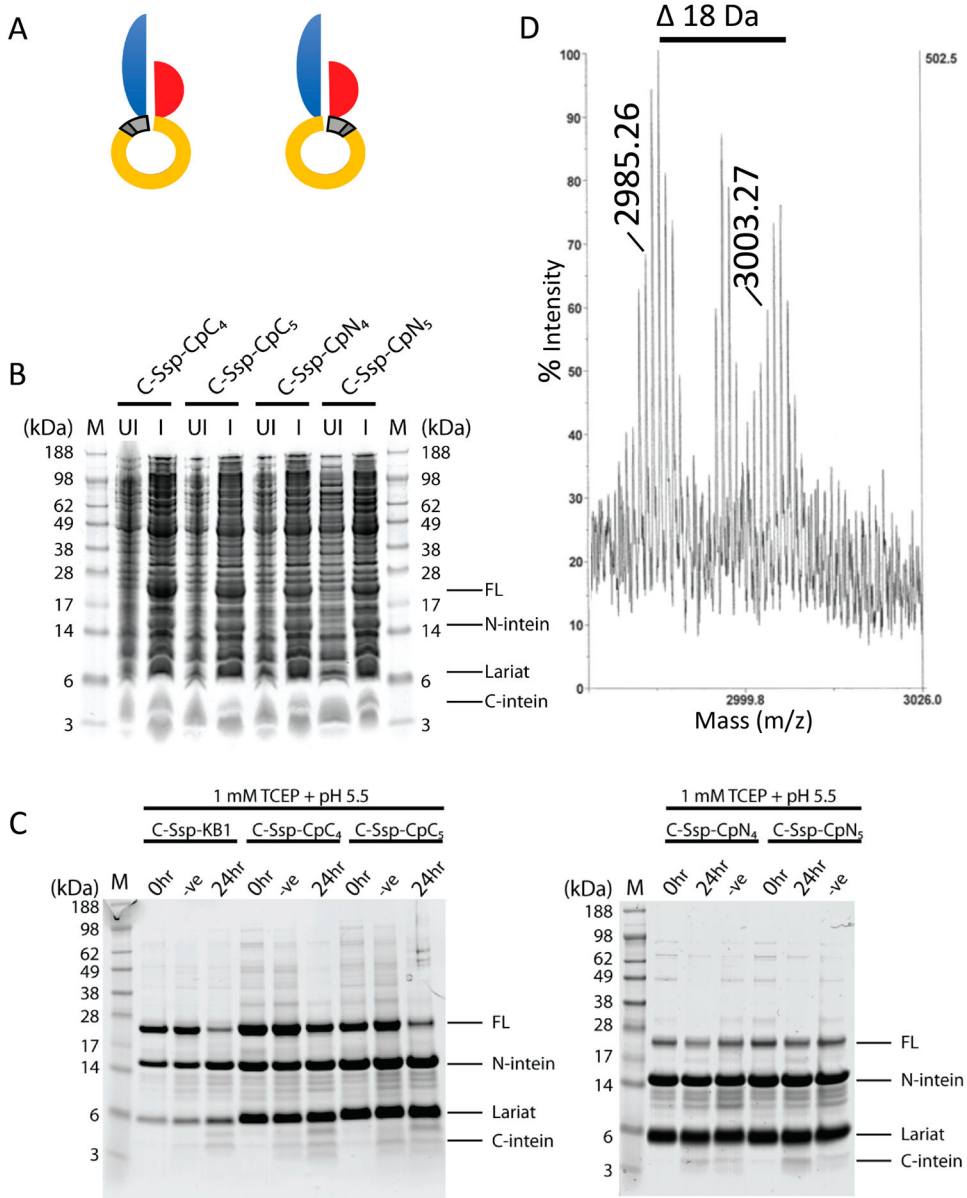
Figure 7. Evaluation of SICLOPPS MVIIA constructs. **A,** Diagrammatic representations of the MVIIA SICLOPPs constructs with position of linker sequence indicated. The linker (in gray) is indicated adjacent to the N-intein (blue) as found in C-Ssp-CpN$_4$ (for linker TSNG) and C-Ssp-CpN$_5$ (linker TSNGL) and adjacent to the C-intein (red) as found in C-Ssp-CpC$_4$ (linker TSNG) and C-Ssp-CpC$_5$ (linker TSNGL). **B,** Expression of the MVIIA SICLOPPs constructs with indication of the positions of the full-length peptide (FL) and splicing fragments (N-intein, Lariat and C-intein) all constructs induced at 37°C for 4 h. Uninduced (UI) and Induced lysates (I) of each construct (designated above as in Figure 3) were analysed by SDS-PAGE. Markers and bands of interest are indicated as in Figure 3. **C,** *In vitro* splicing of C-Ssp-KB1 (same chimera as in Figure 3), C-Ssp-CpC$_4$, C-Ssp-CpC$_5$, C-Ssp-CpN$_4$ and C-Ssp-CpN$_5$. All samples were incubated at 22°C with (24 h) or without (−ve) 1 mM TCEP at pH 5.5. Markers and bands of interest are indicated as in Figure 3. **D,** The Products of C-Ssp-CpC$_4$ were analysed by intact mass. The family of monoisotopic mass predicted for cyclic (centred on 2985.2561) and linear species (3003.2671) are indicated.
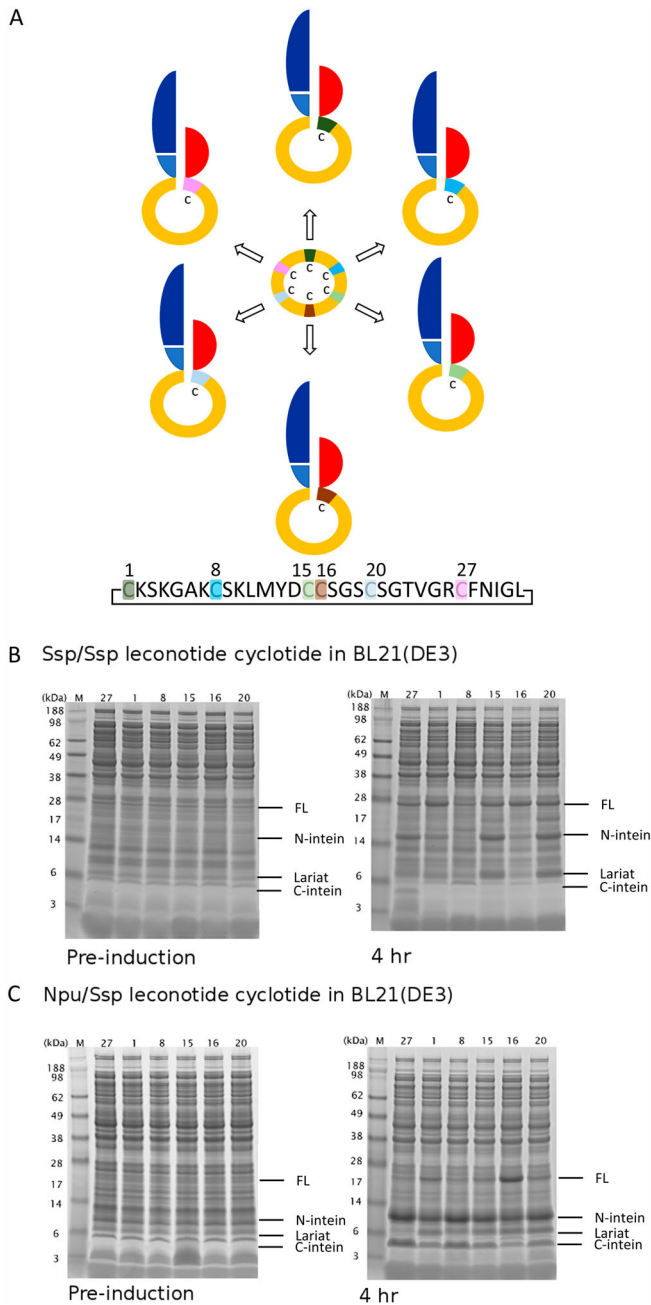
Figure 8. Permutations of CVID in SICLOPPS. **A,** A diagrammatic representation of the permutations placing each cysteine residue at the +1 cysteine, essential for the splicing reaction. The sequence of CVID is shown with the different cysteines indicated. **B,** The permutated variants of CVID in the Ssp/Ssp SICLOPPs were analysed by SDS-PAGE of cell lysates following induction at 37°C for 4 h in *E. coli* BL21. The permuted version analysed in each lane is indicated by the position of the cysteine in native, linear CVID that is placed in the +1 position. Markers and bands of interest are indicated as in Figure 3. **C,** The permutated variants of CVID in the Npu/Ssp SICLOPPs were analysed by SDS-PAGE of cell lysates following induction at 37°C for 4 h in *E. coli* BL21. The permuted version analysed, markers and bands of interest are indicated as in panel B.

permutations showed the same relative capacity in the two systems, position 8 supported more splicing in the hybrid system.

The same permutations were expressed in the Origami *E. coli* strain (Novagen) to assess the effect of the oxidising cytosol, with no significant difference relative to products observed in BL21(DE3) (Supplementary Figure 3).

Discussion

This work assessed whether SICLOPPS could facilitate production, purification and *in vitro* splicing to produce a variety of cyclotides with minimal optimisation. We have established that a precursor cyclotein can be purified using a hexahistidine-tag and that following IMAC purification, the cyclotein is able to splice efficiently *in vitro*. In addition, we have shown that SICLOPPs can mediate the cyclisation of a diversity of small, disulphide rich peptides. Demonstration that SICLOPPs can be used to cyclise disulphide rich peptides illustrates its wide range of potential applications and appeal for further development.

The occurrence of splicing was initially indicated by conversion of precursor bands to product bands in SDS-PAGE and the production of cyclic peptides in the reaction mixtures was confirmed by MALDI-TOF mass spectrometry. The predicted ions were observed both for cyclotides and for linear forms. Cleavage by trypsin at the single tryptic cleavage site (C-terminally of the arginine residue) converted the cyclic form of kalata B1 to a peptide that is 18 Da heavier, linear, and permuted relative to the gene sequence. Sequencing of the tryptic product of kalata B1 by CID-MS/MS yielded both y-series and b-series ions consistent with the expected mature cyclic form of kalata B1 and spanning the joined amino- and carboxyl-termini of the precursor.

Knottins display a compact, very stable tertiary structure stabilised by three intra-chain disulphide bonds. Formation of these disulphides competes with the requirement for a free thiol nucleophile at the +1 position of the active site for splicing to occur. If this cysteine adjacent to the asparagine at the carboxy terminus of the C-intein is sequestered in a disulphide bond, then splicing is inhibited. To optimise yield, splicing was analysed *in vitro* in the presence and absence of reducing agents. The SICLOPPs constructs carrying kalata B1 showed increased splicing *in vitro* in the presence of reducing agents (Figure 3), confirming that splicing occurred more efficiently if the cyclotide disulphide bonds were reduced and suggesting a means of modulating the reaction. In contrast to the effect of high concentrations of reductant *in vitro*, expression in the mildly oxidising cytosol of Origami *E. coli* strain showed no effect.

The splicing reaction that yielded native kalata B1 also yielded a significant fraction of an unwanted linear form. It is likely that this linear form arises by cleavage of the C-intein from the kalata B1 without the prior cyclisation. We hypothesised that this competing reaction might be prominent in the SspDnaE intein because the cyclisation reaction is slow. To accelerate cyclisation and explore this hypothesis, the native kalata B1 sequence was modified such that the splice junctions more closely resembled the extein sequences of the SspDnaE system. Relative speed of cyclisation and competing hydrolysis do indeed appear to influence the yield of mature cyclotide.

To determine if the *in vitro* splicing system could be applied to other sequences and to circularise naturally linear knottins, cycloteins were constructed that encoded sequences

corresponding to MVIIA and CVID. These linear knottins are conopeptide toxins produced by marine cone shells which have potential as analgesics. In both cases, a linker of 4 or 5 amino acids was added to join the amino- and carboxyl-termini without distorting the native structures. Constructs showed efficient expression in bacteria and evidence of splicing the following purification. The presence of splicing was confirmed for one construct using intact mass analysis.

Splicing efficiency is sensitive to the identity of amino acids around the splice junction of SspDnaE and NpuDnaE split inteins (Kimura et al. 2006; Jagadish and Camarero 2010). For split-intein-mediated cyclisation, a cyclotein precursor might use as the splice junction nucleophile any of the six cysteines of the knottin. To determine whether its six cysteine residues differ in ability to fulfil this role, six cycloteins were constructed that encoded permutations of CVID. Considerable variation in splicing yield was observed among these CVID precursor permutations; ligation at position 27 showed the most efficient splicing of the constructs investigated. There was also variation in the level of lariat observed, with position 27 generating the least lariat. One permutation, position 8, showed negligible splicing in the SspE DnaE system. This variation reflects the influences of local sequence composition and tertiary structure on splicing efficiency. Design of cycloteins must take account of positioning of a nucleophilic thiol at the splice junction.

In order to achieve more rapid splicing, a hybrid Npu/SspDnaE intein system was used together with the six CVID permutations. The hybrid split intein showed more efficient splicing than the Ssp/SspDnaE system. Not only did the Npu/SspDnaE system accumulate less full-length peptide, indicating faster splicing, it also generated less lariat in some permutations. The permutation that showed highest efficiency in the SspDnaE system (ligation at position 27) showed no full-length cyclotein and negligible lariat formation, together with abundant N- and C-intein fragments, suggesting high-efficiency splicing. In contrast, the least efficient SspDnaE/CVID permutation (ligation at position 8) showed some splicing in the Npu/SspDnaE system. SICLOPPS can cyclise a range of cyclotides, although interplay of intein and target sequence affect efficiency.

While the expression of full-length cycloteins can be readily induced, we observed two inefficiencies to the use of this system to make pure cyclotides. Firstly, a proportion of the cyclotein is spliced *in vivo* and the resulting cyclotide is therefore not readily purified. Secondly, a proportion of the *in vitro* product is a linear rather than a cyclic form. These two inefficiencies are interconnected. Rapidly splicing inteins such as Npu DnaE produce a large amount of *in vivo* splicing, in contrast slow splicing inteins such as SspDnaE allow the purification of a larger yield of un-spliced cyclotein but they produce more unwanted linear peptides. In the production of cyclic knottins, we enhanced yield by manipulating sequence permutation, pH, temperature and redox state. Further characterisation is needed to make the system fully conditional and capable of producing pure cyclic peptides via the efficient purification and *in vitro* splicing demonstrated here. Such a conditional system would offer scalable biosynthetic production of a wider range of useful cyclic peptides with potential as therapeutic agents.

Authorial contributions

TNGH contributed to molecular cloning, production and analysis of cycloteins and contributed to manuscript preparation. H-XT contributed to molecular cloning, production

and analysis of kalata B1 and ziconotide cycloteins. HB & MTR created expression plasmids for, produced, and analysed permuted leconotide cycloteins. MIB designed and produced genetic constructs. TK collected and analysed MS data. JDAT designed linkers for circularisation of leconotide and ziconotide. RTMP & SMW contributed to experimental design, data interpretation, project direction and manuscript preparation.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Thomas N. G. Handley 🆔 http://orcid.org/0000-0001-5604-0202
Hyon-Xhi Tan 🆔 http://orcid.org/0000-0003-1064-3934
Malcolm T. Rutledge 🆔 http://orcid.org/0000-0002-7180-0879
Joel D. A. Tyndall 🆔 http://orcid.org/0000-0003-0783-1635
Margi I. Butler 🆔 http://orcid.org/0000-0001-6863-3961
Russell T. M. Poulter 🆔 http://orcid.org/0000-0001-5848-646X
Sigurd M. Wilbanks 🆔 http://orcid.org/0000-0001-7565-8913

References

Bi T, Li Y, Shekhtman A, Camarero JA. 2018. In-cell production of a genetically-encoded library based on the θ- defensin RTD-1 using a bacterial expression system. Bioorganic & Medicinal Chemistry. 26(6):1212–1219.

Carvajal-Vallejos P, Pallisse R, Mootz HD, Schmidt SR. 2012. Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources. Journal of Biological Chemistry. 287:28686–28696.

Cheriyan M, Pedamallu CS, Tori K, Perler F. 2013. Faster protein splicing with the *Nostoc punctiforme* DnaE intein using non-native extein residues. Journal of Biological Chemistry. 288 (9):6202–6211.

Clark RJ, Fischer H, Dempster L, Daly NL, Rosengren KJ, Nevin ST, Meunier FA, Adams DJ, Craik DJ. 2005. Engineering stable peptide toxins by means of backbone cyclization: stabilization of the alpha-conotoxin MII. Proceedings of the National Academy of Sciences USA. 102(39):13767–13772.

Craik DJ, Daly NL, Waine C. 2001. The cysteine knot motif in toxins and the implications for drug design. Toxicon: Official Journal of the International Society on Toxinology. 39(1):43–60.

Craik DJ, Du J. 2017. Cyclotides as drug design scaffolds. Current Opinion in Chemical Biology. 38:8–16.

Craik DJ, Fairlie DP, Liras S, Price D. 2013. The future of peptide-based drugs. Chemical Biology and Drug Design. 81(1):136–147.

Fang GM, Chen XX, Yan QQ, Zhu LJ, Li NN, Yu HZ, Meng XM. 2018. Discovery structure and chemical synthesis of disulphide-rich peptide toxins and their analogs. Chinese Chemical Letters. 29:1033–1042.

Gillon ADS, Jennings CV, Guarino RF, Craik DJ, Anderson MA. 2008. Biosynthesis of circular proteins in plants. The Plant Journal. 53(3):505–515.

Gran L. 1973. On the effect of a polypeptide isolated from "kalata-kalata" (*Oldenlandia affinis* DC) on the oestrogen dominated uterus. Acta Pharmacologica et Toxicologica. 33(5):400–408.

Harris KS, Durek T, Kaas Q, Poth AG, Gilding EK, Colan BF, Saska I, Daly NL, van der Weerden NL, Craik DJ, Anderson MA. 2015. Efficient backbone cyclisation of linear peptides by a recombinant asparaginyl endopeptidase. Nature Communications. 18(6):10199.

Iwai H, Züger S, Jin J, Tam PH. 2006. Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. FEBS Letters. 580(7):1853–1858.

Jagadish K, Borra R, Lacey V, Majumder S, Shekhtman A, Wang L, Camerero JA. 2013. Expression of fluorescent cyclotides using protein trans-splicing for easy monitoring of cyclotide-protein interactions. Angewandte Chemie-International Edition. 52:3126–3131.

Jagadish K, Camarero JA. 2010. Cyclotides, a promising molecular scaffold for peptide-based therapeutics. Biopolymers. 94(5):611–616.

Kimura RH, Tran A-T, Camarero JA. 2006. Biosynthesis of the cyclotide kalata B1 by using protein splicing. Biosynthesis. 45:973–976.

Kolmar H. 2008. Alternative binding proteins: biological activity and therapeutic potential of cysteine-knot miniproteins. The FEBS Journal. 275(11):2684–2690.

Lewis RJ, Nielsen KJ, Craik DJ, Loughnan ML, Adams DA, Sharpe IA, Luchian T, Adams DJ, Bond T, Thomas L, et al. 2000. Novel omega-conotoxins from *Conus catus* discriminate among neuronal calcium channel subtypes. Journal of Biological Chemistry. 275(45):35335–35344.

Li Y, Aboye T, Breindel L, Shekhtman A, Camarero JA. 2016. Efficient recombinant expression of SFTI-1 in bacterial cells using intein-mediated protein trans-splicing. Biopolymers. 106(6):818–824.

Li Y, Bi T, Camarero JA. 2015. Chemical and biological production of cyclotides. Advances in Botanical Research. 76:271–303.

McGivern JG. 2007. Ziconotide: a review of its pharmacology and use in the treatment of pain. Neuropsychiatric Disease and Treatment. 3(1):69–85.

Olivera BM, Cruz LJ, de Santos V, LeCheminant GW, Griffin D, Zeikus R, McIntosh M, Galyean R, Varga J, Gray WR, River J. 1987. Neuronal calcium channel antagonists. Discrimination between calcium channel subtypes using ω-conotoxin from *Conus magus* venom. Biochemistry. 26:2086–2090.

Pallaghy P, Nielsen KJ, Craik DJ, Norton RS. 1994. A common structural motif incorporating a cysteine knot and a triple-stranded beta-sheet in toxic and inhibitory polypeptides. Protein Science. 3(10):1833–1839.

Perler FB. 2002. Inbase: the intein Database. Nucleic Acids Research. 30(1):383–384.

Perler FB, Davis EO, Dean GE, Gimble FS, Jack WE, Neff N, Noren CJ, Thorner J, Belfort M. 1994. Protein splicing elements: inteins and exteins- a definition of terms and recommended nomenclature. Nucleic Acids Research. 22(7):1125–1127.

Poon S, Harris KS, Jackson MA, McCorkelle OC, Gilding EK, Durek T, van der Weerden NL, Craik DJ, Anderson MA. 2018. Co-expression of a cyclizing asparaginyl endopeptidase enables the efficient production of cyclic peptides in planta. Journal of Experimental Botany. 69(3):633–641.

Saether O, Craik DJ, Campbell ID, Sletten K, Juul J, Norman DG. 1995. Elucidation of the primary and three-dimensional structure of the uterotonic polypeptide kalata B1. Biochemistry. 34(13):4147–4158.

Scott CP, Abel-Santos E, Wall M, Wahnon DC, Benkovic SJ. 1999. Production of cyclic peptides and proteins *in vivo*. Proceedings of the National Academy of Sciences U.S.A. 96(24):13638–13643.

Shafee T, Harris KS, Anderson MA. 2015. Biosynthesis of cyclotides. Advances in Botanical Research. 76:227–269.

Shevchenko A, Tomas H, Havlis J, Olsen LV, Mann M. 2006. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nature Protocols. 1(6):2856–2860.

Sletten K, Gran L. 1973. Some molecular properties of kalata-peptide B1. Meddelelser Norsk Farmasoytisk Sleskap. 35:69–82.

Tavassoli A. 2017. SICLOPPS cyclic peptide libraries in drug discovery. Current Opinion in Chemical Biology. 38:30–35.

Tavassoli A, Benkovic SJ. 2004. Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in *E. coli*. Nature Protocols. 2(5):1126–1133.

Vlieghe P, Lisowski V, Martinez J, Khrestchatsky M. 2010. Synthetic therapeutic peptides: science and market. Drug Discovery Today. 15(1–2):40–56.

Wang CK, Kass Q, Chiche L, Craik DJ. 2008. Cybase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering. Nucleic Acids Research. 36: D206–D210.

Write CE, Robertson AD, Whorlow SL, Angus JA. 2000. Cardiovascular and autonomic effects of ω-conotoxins MVIIA and CVID in conscious rabbits and isolated tissue assays. British Journal of Pharmacology. 10:1325–1336.

Wu H, Hu Z, Liu X. 1998. Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803. Proceedings of the National Academy of Sciences U.S.A. 95:9226–9231.

Zettler J, Schütz V, Mootz HD. 2009. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. FEBS Letters. 583:909–914.