

# Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in *E. coli*

Ali Tavassoli<sup>1</sup> & Stephen J Benkovic<sup>2</sup>

<sup>1</sup>School of Chemistry, University of Southampton, Southampton SO17 1BJ, UK. <sup>2</sup>Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. Correspondence should be addressed to S.J.B. (sjb1@psu.edu).

Published online 3 May 2007; doi:10.1038/nprot.2007.152

Recent advances in chemical biology and the advantages presented by *in vivo* screening have highlighted the need for a robust and flexible biologically synthesized small-molecule library. Herein we describe a method for the biosynthesis of cyclic peptide libraries of up to  $10^8$  members in *Escherichia coli* using split-intein circular ligation of peptides and proteins (SICLOPPS). The method utilizes split-intein chemistry to cyclize randomized peptide sequences. The cyclic peptide library can potentially be of any size and the peptide itself may contain unlimited random residues. However, the library size is limited by the transformation efficiency of *E. coli* and random residues are generally limited to five, but additional amino acids can be used in the cyclic peptide backbone, varying the structure and ring size of the cyclic peptide. SICLOPPS libraries have been combined with a bacterial reverse two-hybrid system in our labs and used in the identification of inhibitors of several protein–protein interactions. This protocol is expected to take around 3–4 weeks to implement.

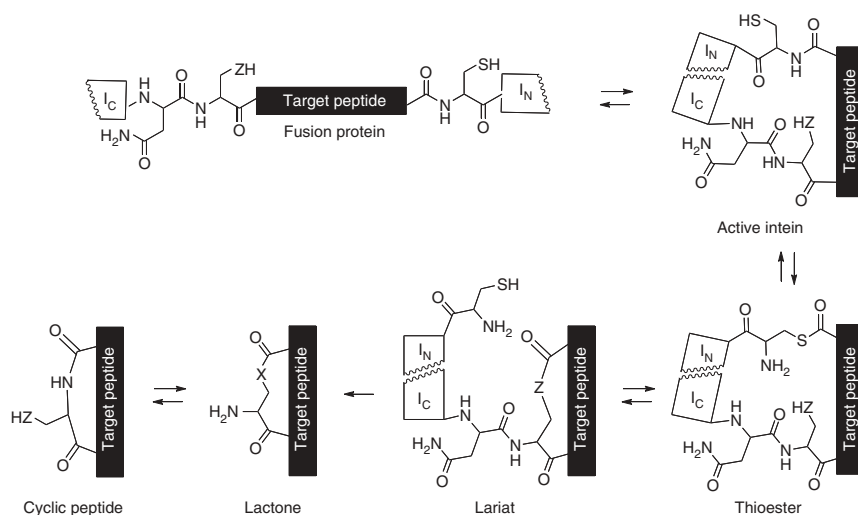
## INTRODUCTION

The recent advances in chemical biology and functional genomics have given rise to new methods and opportunities for drug discovery, and in turn new methods for the generation of small-molecule libraries<sup>1–5</sup>. Combinatorial chemistry, the method of choice for assembling chemically synthesized libraries, has also undergone rapid development over the last few years<sup>6–8</sup>. Yet, despite providing almost unlimited functional group diversity, chemically synthesized libraries of small molecules still lack a suitable, straightforward method for their decoding. Split and pool protocols allow for the chemical synthesis of libraries with up to  $\sim 10^5$  members<sup>9,10</sup>; however, the paths to identifying the active members of such libraries are often laborious and complicated<sup>11</sup>. In contrast, biologically synthesized libraries of small molecules are often several-fold larger in size and allow for very straightforward identification of the active members<sup>12–14</sup>. When combined with an *in vivo* screen, biological libraries become part of a powerful, rapid and facile method for the screening of a large number of compounds against a chosen target<sup>4,14–16</sup>.

Biological libraries are principally polypeptides, typically embedded within, or fused to larger molecules to overcome the host cell's catabolic machinery. Alternatively, intracellular stability is achieved by constraining the ends of the peptide with non-covalent and covalent interactions, with varying results<sup>17,18</sup>. We have pursued intracellular backbone cyclization to generate bio-stable peptide libraries. This procedure, termed SICLOPPS<sup>19,20</sup>, utilizes the *Synechocystis* sp PCC6803 DnaE split intein<sup>21</sup>. By rearranging the order of the elements of the intein, an active *cis*-intein (I<sub>C</sub>:target peptide:I<sub>N</sub>) is yielded that upon splicing results in cyclization of the target protein/peptide sequence (Fig. 1).

Any target can be incorporated into the SICLOPPS vector without limitations on sequence identity by engineering restriction sites into the C-terminal (I<sub>C</sub>) and N-terminal inteins (I<sub>N</sub>). Cyclic peptide characterization and protein purification is aided by a chitin-binding domain (CBD) fused to the C terminus of I<sub>N</sub> (Fig. 2).

To utilize SICLOPPS in the biosynthesis of cyclic peptide libraries<sup>19</sup>, degenerate oligonucleotides encoding those peptides need to be introduced between the I<sub>C</sub> and I<sub>N</sub> genes, while making sure that the correct reading frame is maintained throughout. A PCR-based technique was developed in which the random oligonucleotides of the library are incorporated into the forward primer (library primer; see Fig. 3) between the region that binds the 3' end of I<sub>C</sub> and the 5' end of I<sub>N</sub>. The reverse primer anneals to the

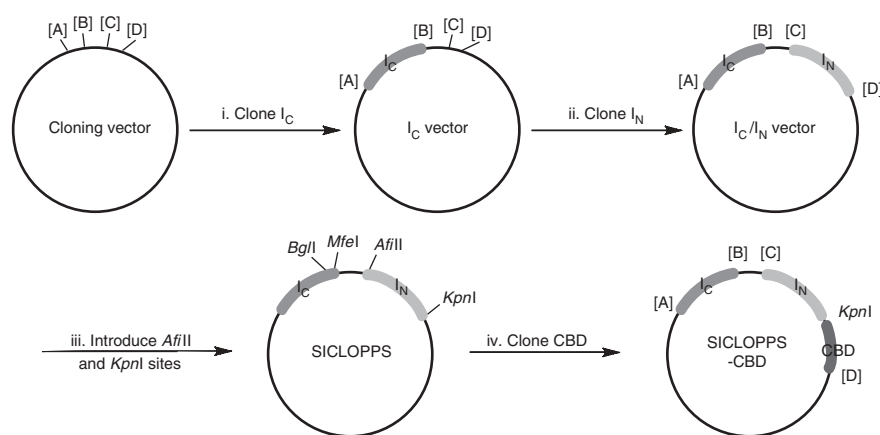


**Figure 1** | SICLOPPS mechanism. The expressed fusion protein folds to form an active intein. An N-to-S acyl shift at the target N-terminal intein junction produces a thioester, which undergoes transesterification with a side-chain nucleophile (serine or cysteine, X = O or S) at the C-terminal intein junction to form a lariat intermediate. An asparagine side chain liberates the cyclic product as a lactone, and an X-to-N acyl shift generates the thermodynamically favored lactam product *in vivo*.



chitin-CBD, amplifying the IN-CBD fusion gene (CBDr primer; see Fig. 3). The variable segment was encoded in the form NNS where N represents any of the four DNA bases (A, C, G or T) and S represents C or G. The NNS sequence generates 32 codons and encodes all 20 amino acids while eliminating the ochre (UAA) and opal (UGA) stop codons from the library. The intein chemistry requires the first amino acid to be a nucleophilic cysteine or serine. There are no limits on the number of amino acids in the target peptide, allowing cyclic peptides of various sizes. Typically, five variable amino acids are used because the theoretical number of library members at DNA levels (34 million) is within the number of transformants that can be readily achieved.

After the first PCR, half of the amplified DNA fragments were found to contain mismatches in the random nucleotide region, probably owing to the sequence complexity of the library. A second PCR using a “zipper” primer corresponding to the 3' end of  $I_C$  was therefore used to ensure the annealing of all DNA sequences to their complementary strand. These pieces of DNA were incorporated into the SICLOPPS plasmid using standard techniques, thus generating the desired library (Fig. 3).



**Figure 2** | Construction of SICLOPPS vectors. The cloning instructions have been kept general so that any expression vector may be used in the construction of the SICLOPPS plasmid as long as it contains four contiguous restriction sites [A]–[D]. i, the  $I_C$  domain is cloned between restriction sites [A] and [B]; ii, the  $I_N$  domain is cloned into the  $I_C$  vector between restriction sites [C] and [D]; iii, introduction of an *AfIII* restriction site at the 5' end of the  $I_N$  gene allows peptide sequences targeted for cyclization to be cloned between the *BglII/MfeI* and *AfIII* sites. Introduction of the *KpnI* at the 3' end of  $I_N$  allows the introduction of affinity tags. iv, an affinity tag (CBD) is cloned into the SICLOPPS vector between the *KpnI* and [D] restriction sites.

We have demonstrated the flexibility of the procedure toward peptide length and composition, making it a versatile method for the intracellular generation of a library of stable cyclic peptides. SICLOPPS has been successfully combined with a reverse two-hybrid system in our laboratories to yield a method for the genetic selection of inhibitors of protein–protein interactions<sup>15,16</sup>.

## MATERIALS

### REAGENTS

- ElectroMAX DH5 $\alpha$ -E (Invitrogen, # 11319-019)
- *Synechocystis* sp PCC6803 genomic DNA (ATCC, # 27184)
- Any expression vector containing four contiguous restriction sites corresponding to [A], [B], [C] and [D] in the PCR primers (in our laboratories we use the pARCBD-p plasmid vector)<sup>20</sup>
- pTYB1 plasmid vector (New England Biolabs, # N6701S)
- PCR primers can be synthesized “in house” if a DNA synthesizer is available, otherwise they are ordered (IDT in the USA, Sigma in the UK). The oligonucleotides below are listed 5' to 3':
  - $I_C$ f primer: [restriction site A]—ATG GTT AAA GTT ATC GGT CGT CGT TCC C
  - $I_C$ r primer: [restriction site B]—ATT GTG CGC AAT CGC CCC AT
  - $I_N$ f primer: [restriction site C]—GCC TCA GTT TTG GC
  - $I_N$ r primer: [restriction site D]—TTA TTT AAT AGT CCC AGC GTC
  - $I_N$  *AfIII*f primer: [restriction site C]—TGC TTA AGT TTT GGC ACC
  - $I_N$  *AfIII*r primer: GGT GCC AAA ACT TAA GCA—[restriction site C]
  - $I_N$  *KpnI*f primer: TAC TTG ACG CTG GTA CCA TTA AAT AA—[restriction site D]
  - $I_N$  *KpnI*r primer: [restriction site D]—TTA TTT AAT GGT ACC AGC GTC AAG TA
  - CBDf primer: GGG GTA CCA TTA AAA CGA CAA ATC CTG GTG TA
  - CBDr primer: [restriction site D]—TCA TTG AAG CTG CCA CAA GG
  - C+5 library primer: G GAA TTC GCC AAT GGG GCG ATC GCC CAC AAT TGC NNS NNS NNS NNS NNS TGC TTA AGT TTT GGC
  - Zipper primer: GGA ATT CGC CAA TGG GGC GAT CGC C
- Unless otherwise stated, all reagents are bought from Fisher Scientific
- Lauria-Bertani (LB) broth: 10 g bacto-tryptone, 5 g bacto-yeast extract, 10 g NaCl and 1 liter water. Autoclaved<sup>22</sup> **▲ CRITICAL** It is important to autoclave the denoted reagents to prevent contamination.
- Deionized water. Autoclaved

- 10% (v/v) glycerol solution: 10 ml glycerol and 90 ml deionized water. Autoclaved
- 10 $\times$  thermophilic DNA polymerase buffer and *Taq* DNA polymerase (Promega, # M3171)
- dNTP mix (dATP, dCTP, dGTP and dTTP, 10 mM) (Promega, # U1511)
- QiaQuick PCR purification, Plasmid midi kit and QiaQuick gel extraction kits (QIAGEN, # 28104, # 27104 and # 28704)
- Agarose, molecular biology grade (CAS, # 9012-36-6)
- Ethidium bromide (CAS, # 1239-45-8) **! CAUTION** Ethidium bromide is a known carcinogen and should be handled and disposed of with care.
- Appropriate 10 $\times$  restriction enzyme buffer, 100 $\times$  BSA and restriction enzymes corresponding to restriction sites [A], [B], [C] and [D] in the PCR primers (New England Biolabs)
- Shrimp alkaline phosphatase (Promega, # M8201)
- T4 DNA ligase and 10 $\times$  ligation buffer (Promega, # M1801)
- PhiX174 DNA/*HaeII* and Lambda DNA/*HinDIII* markers (Promega, # G1761 and # G1711)
- Quick change site-directed mutagenesis kit (Stratagene)
- 0.2% (w/v) arabinose solution (CAS, # 87-72-9)
- Chitin affinity resin (New England Biolabs, # S6651S)
- Chitin buffer: 20 mM Tris · HCl, pH 7.8, 1 mM TCEP and 0.5 mM NaCl
- Sodium dodecyl sulfate (SDS) gel-loading buffer: 100 mM Tris · HCl, pH 6.8, 200 mM dithiothreitol, 4% (w/v) SDS, 0.2% (w/v) bromophenol blue and 20% (v/v) glycerol
- SDS-PAGE gel: 15% resolving gel and 4% stacking gel (Bio-Rad Laboratories, # 161-1103EDU)
- 1 $\times$  SDS-PAGE Tris-glycine electrophoresis buffer: 5 $\times$  solution: 15.1 g Tris base, 94 g glycine and 5 g SDS made up to 1 liter with deionized water
- SDS-PAGE stain solution: 0.25 g Coomassie brilliant blue R250 (CAS, # 6104-58-1) in 45 ml methanol, 45 ml water and 10 ml glacial acetic acid
- SDS-PAGE de-stain solution: 45 ml methanol, 45 ml water and 10 ml glacial acetic acid

# PROTOCOL

## EQUIPMENT

- Thermal cycler programmed with the amplification protocol
- Incubator/shaker at 37 °C
- Electroporator
- Protein and nucleic acid electrophoresis equipment
- Gel imaging equipment
- UV transilluminator
- UV spectrophotometer
- 0.2 cm gene pulser cuvettes (Bio-Rad Laboratories, # 165-2083)
- Membrane filter paper, 0.025 µm pore size, 25 mm diameter (Millipore, # VSWP02500)

## REAGENT SETUP

- 1× Tris-acetate EDTA buffer** 50× solution: 242 g Tris base, 57.1 ml glacial acetic acid and 100 ml EDTA, pH 8.0, make up to 1 liter final volume with water.
- SOC medium** 20 g bacto-tryptone, 5 g bacto-yeast extract and 0.5 g NaCl. Add 10 ml of 25 mM KCl. Adjust to pH 7.0 with 5 N NaOH. Dilute to 1 liter with distilled water. Autoclave. Add 2 M MgCl<sub>2</sub> and 20 ml of 1 M glucose before use.
- LB agar plates** 15 g bacto-agar added to LB broth. Autoclave. Allow the agar to cool to 60 °C and add the appropriate antibiotic. Dispense in 100 × 15 mm Petri dishes.
- Library agar plates** LB agar (as above) dispensed in 245 × 245 × 25 bio-assay dish (NUNC, # 240835).

## PROCEDURE

### Cloning of I<sub>C</sub> and I<sub>N</sub> genes ● TIMING 2 days

1| To clone the I<sub>C</sub> and I<sub>N</sub> gene inserts, set up two PCRs containing the following:

Volume	Component	Final concentration
5.0 µl	10× PCR buffer	1×
2.5 µl	MgCl <sub>2</sub> (25 mM)	1.25 mM
5.0 µl	10× dNTP mix	200 µM
5.0 µl	I <sub>C</sub> f or I <sub>N</sub> f primer (10 µM)	1 µM
5.0 µl	I <sub>C</sub> r or I <sub>N</sub> r primer (10 µM)	1 µM
1.0 µl	<i>Synechocystis</i> sp PCC6803 genomic DNA	—
25.5 µl	ddH <sub>2</sub> O	—
1 µl	<i>Taq</i> DNA polymerase (5 U µl <sup>-1</sup> )	0.1 U µl <sup>-1</sup>

2| Run a PCR with the following program:

Cycle	Denaturation	Annealing	Extension
1	94 °C for 5 min	55 °C for 1 min	72 °C for 1.5 min
2–26	94 °C for 1 min	55 °C for 1 min	72 °C for 1.5 min
27	—	—	72 °C for 10 min

3| Run an agarose gel to verify the presence of amplified DNA of the correct size in the reaction mixture<sup>22</sup>.

**! CAUTION** Ethidium bromide is a carcinogen; therefore, the gel should be handled with care and disposed off accordingly.

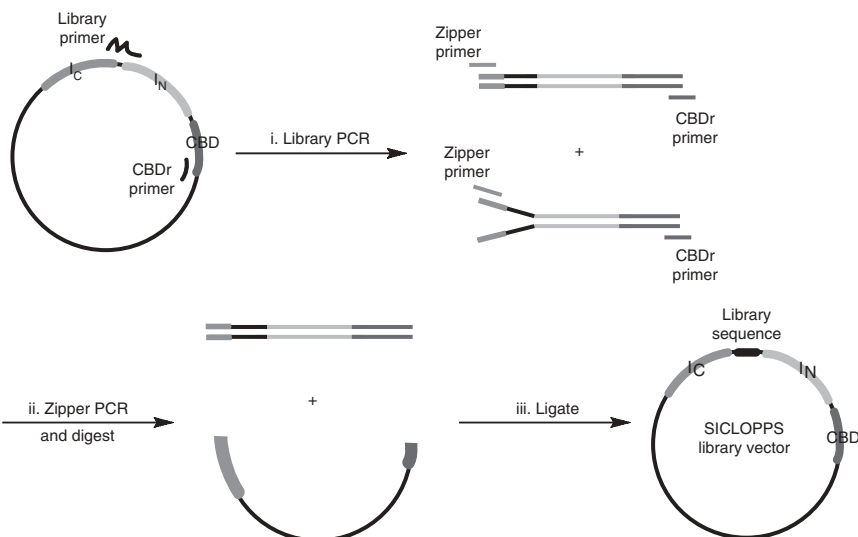
4| Purify the product of each PCR with the QiaQuick PCR kit (as instructed by the manufacturer).

■ **PAUSE POINT** The PCR product should be stored overnight at –20 °C.

5| Digest the PCR products with the corresponding restriction enzymes ([A] and [B] for I<sub>C</sub>; [C] and [D] for I<sub>N</sub>) according to the manufacturer's protocol (typically, mix 5 µl of 10× digestion buffer with 5 µl of BSA, 25 µl of purified PCR product, 2 µl of restriction enzyme and 13 µl water. Incubate for 4 h at 37 °C).

Recover and purify the digested PCR products with the QiaQuick PCR kit.

■ **PAUSE POINT** The DNA should be stored overnight at –20 °C.



**Figure 3** | Construction of the library vectors. i, a library primer encoding the 5' end of I<sub>C</sub>, a degenerate sequence and the 3' end of I<sub>N</sub> is used to amplify an I<sub>N</sub>-CBD fragment; ii, the resulting CR product was found to contain mismatched pairs, and so is subjected to a second round of amplification with a zipper primer. The amplified fragments and the SICLOPPS vector are digested with the corresponding restriction enzymes. iii, ligation of the vector with the library insert results in the SICLOPPS library plasmids.

**Ligation and electroporation** ● **TIMING** ~ 12 days

- 6| Digest the vector with the chosen restriction enzymes [A] and [B] corresponding to the sites on the  $I_C$  gene PCR products above. Again, refer to the manufacturer's instructions corresponding to the chosen restriction enzymes.
- 7| Add 1  $\mu$ l of shrimp alkaline phosphatase to the digested vector (to dephosphorylate the termini). Incubated the mixture at 37 °C for 1 h. Inactivate the enzyme by incubating at 65 °C for 30 min.
- 8| Separate the digested vector backbone from the digested fragment on an agarose gel in 1 $\times$  Tris-acetate EDTA (TAE) buffer as outlined above. Identify the vector backbone (by size) under a UV transilluminator and cut out of the gel with a blade.
- 9| Purify the excised band using a QiaQuick gel extraction kit as instructed by the manufacturer. Determine the DNA concentration of the vector and  $I_C$  gene insert using a UV spectrophotometer.
  - **PAUSE POINT** The DNA should be stored overnight at -20 °C.
- 10| Electrocompetent DH5 $\alpha$ -E cells can be prepared as described in the literature<sup>22,23</sup> or bought (Invitrogen).
- 11| Set up the following ligation mixtures, one for the actual ligation, the other, without insert, as a control:

Volume	Component	Final concentration
1.0 $\mu$ l	10 $\times$ ligation buffer	1 $\times$
? $\mu$ l	Digested $I_C$ fragment, 100 ng or ddH <sub>2</sub> O as control	10 ng $\mu$ l <sup>-1</sup>
? $\mu$ l	Digested, dephosphorylated vector, 100 ng	10 ng $\mu$ l <sup>-1</sup>
Complement to 10 $\mu$ l total volume	ddH <sub>2</sub> O	—
1 $\mu$ l	T4 DNA ligase (1–3 U $\mu$ l <sup>-1</sup> )	0.1–0.3 U $\mu$ l <sup>-1</sup>

- 12| Incubate both reactions at room temperature (25 °C) for 4 h. Heat-inactivate the T4 ligase by incubating at 65 °C for 30 min.
  - 13| Add 3  $\mu$ l of the ligation mixture to 100  $\mu$ l of electrocompetent cells. Transfer the mixture to an electroporation cuvette. Apply electric pulse with an electroporator as instructed by the manufacturer. Immediately add 900  $\mu$ l of SOC medium and transfer to a 10 ml culture tube. Repeat for the negative control. Allow the cells to recover in an incubator/shaker for 1 h at 37 °C.
  - 14| Plate 100  $\mu$ l of the solution on LB agar plates containing the appropriate antibiotic.
    - **PAUSE POINT** Incubate overnight at 37 °C.
  - 15| Pick eight random colonies and inoculate them into separate culture tubes containing 3 ml of LB broth and the appropriate antibiotic for the marker on the vector.
    - **PAUSE POINT** Grow to stationary phase overnight.
- ? **TROUBLESHOOTING**

- 16| Purify the plasmids with the QiaGen plasmid midi kit.
- 17| Identify those *E. coli* colonies containing the plasmids and the IC gene insert by PCR. Use the  $I_C^f$  and  $I_C^r$  primers to amplify the plasmid solutions obtained in Step 16 following the same PCR protocol described in Step 1 (see above). Run an agarose gel to identify the plasmids containing the IC gene (look for amplified gene products on the gel).
- 18| Confirm the sequence integrity of the  $I_C$  gene in the vector by DNA sequencing.
  - **PAUSE POINT** If an "in house" DNA sequencing facility is not available, this process may take several days.
  - ▲ **CRITICAL STEP** Despite the delay it may cause in the completion of this protocol, we strongly advise against proceeding further without checking the integrity of the inserted sequence. All the following sequencing steps should be treated in a similar manner.
- 19| Insert the  $I_N$  gene into the  $I_C$ -containing vector between digestion sites [C] and [D] following Steps 6–18 of the procedure outlined above for the insertion of the  $I_C$  gene.
- 20| After insertion into the vector, confirm the integrity of the  $I_N$  gene by DNA sequencing.
- 21| Use the manufacturer's instructions to introduce an *Afl*III restriction site in the above vector with the Quick Change site-directed mutagenesis kit using the  $I_N$  Aflf and  $I_N$  Aflr primers.
- 22| Use the Quick Change site mutagenesis kit with the  $I_N$  Kpnf and  $I_N$  Kpnr primers to introduce a *Kpn*I restriction site in the above vector.
- 23| Confirm the vector identity by DNA sequencing.



## PROTOCOL

**24|** Introduce a CBD (or any other chosen affinity tag) into the SICLOPPS plasmid using a similar protocol (Steps 6–18) as that outlined for the cloning and introduction of the  $I_C$  gene. Insert the PCR product (CBDf, CDBr primers, pTYB1 template) into the SICLOPPS plasmid between the *KpnI* and [D] digestion sites.

**25|** Confirm the sequence identity of the SICLOPPS-CBD vector by DNA sequencing.

### SICLOPPS library construction ● TIMING 4 days

**26|** The library primer used here codes for a cysteine + 5 random amino acid, hexapeptide library. You can design your own SICLOPPS cyclic peptide library by changing the sequence of the library primer. Remember to include a cysteine or serine at amino-acid position 1 to act as the nucleophile to initiate intein processing (see **Fig. 1**).

**27|** Set up the following PCR mixture in a 200  $\mu$ l PCR tube:

Volume	Component	Final concentration
5.0 $\mu$ l	10 $\times$ PCR buffer	1 $\times$
2.5 $\mu$ l	MgCl <sub>2</sub> (25 mM)	1.25 mM
5.0 $\mu$ l	10 $\times$ dNTP mix	200 $\mu$ M
5.0 $\mu$ l	C + 5 library primer (stock solution conc.: 10 $\mu$ M)	1 $\mu$ M
5.0 $\mu$ l	CBDr primer (stock solution conc.: 10 $\mu$ M)	1 $\mu$ M
—	SICLOPPS-CDB plasmid (PCR template)	100 ng
25.5 $\mu$ l	ddH <sub>2</sub> O	—
1 $\mu$ l	<i>Taq</i> DNA polymerase (5 U $\mu$ l <sup>-1</sup> )	0.1 U $\mu$ l <sup>-1</sup>

**28|** Run a PCR with the following program:

Cycle	Denaturation	Annealing	Extension
1	94 °C for 5 min	65 °C for 2 min	72 °C for 1.5 min
2–31	94 °C for 1 min	65 °C for 1 min	72 °C for 1.5 min
32	—	—	72 °C for 10 min

**▲ CRITICAL STEP** Do not use lower annealing temperatures, as the resulting constructs have been found to contain frameshifts downstream of the library sequence.

**29|** Purify the PCR products with the QiaQuick PCR purification kit. Determine the DNA concentration of the purified PCR product with a UV spectrophotometer.

**30|** Set up and run a second PCR with a “zipper” primer and the above PCR product as template to ensure the annealing of all library members to their corresponding strands (see INTRODUCTION and **Fig. 3**), as follows:

Volume	Component	Final concentration
5.0 $\mu$ l	10 $\times$ PCR buffer	1 $\times$
2.5 $\mu$ l	MgCl <sub>2</sub> (25 mM)	1.25 mM
5.0 $\mu$ l	10 $\times$ dNTP mix	200 $\mu$ M
5.0 $\mu$ l	Zipper primer (stock solution conc.: 10 $\mu$ M)	1 $\mu$ M
5.0 $\mu$ l	CBDr primer (stock solution conc.: 10 $\mu$ M)	1 $\mu$ M
—	Library fragment (PCR template)	100 ng
25.5 $\mu$ l	ddH <sub>2</sub> O	—
1 $\mu$ l	<i>Taq</i> DNA polymerase (5 U $\mu$ l <sup>-1</sup> )	0.1 U $\mu$ l <sup>-1</sup>

Cycle	Denaturation	Annealing	Extension
1	94 °C for 5 min	65 °C for 2 min	72 °C for 1.5 min
2–26	94 °C for 1 min	65 °C for 1 min	72 °C for 1.5 min
27	—	—	72 °C for 10 min

**31|** Digest the fragment with *BglI* (or *MfeI*) and [D] following the protocol supplied by the manufacturer of the restriction enzymes.

**32|** Digest the SICLOPPS-CBD vector with *BglI* (or *MfeI*) and [D]. Treat with shrimp alkaline phosphatase and purify the linear vector backbone on an agarose gel in 1 $\times$  TAE buffer as above.



▲ **CRITICAL STEP** We have found that small changes in the vector to insert ratio in the ligation mixture have large effects on transformation efficiency and therefore the size of the library. Therefore, optimal conditions need to be determined for every plasmid/insert pair.

**33|** Combine 100, 150, 200, 250 and 300 ng of digested library insert with 50, 85 and 125 ng of digested and dephosphorylated vector to give a total of 15 ligation reactions. To each reaction, add 1  $\mu\text{l}$  of  $10\times$  ligation buffer and water to a final volume of 9  $\mu\text{l}$ . Prepare control ligations for each vector concentration (do not add insert). Add 1  $\mu\text{l}$  of T4 DNA ligase to each reaction and controls.

■ **PAUSE POINT** Incubate at room temperature overnight.

**34|** In the morning, inactivate the enzyme by incubating at 65 °C for 30 min.

**35|** Dialyze each ligation mixture by carefully placing the reaction mixture as a drop (with a micropipette) in the center of a Millipore membrane filter in a beaker of 50 ml deionized water for an hour. Carefully retrieve the drop with a micropipette.

**36|** Transform the DNA from each ligation mixture and control reaction into DH5 $\alpha$ -E cells as described above (Steps 13 and 14). After recovery of the cells, plate  $10^{-4}$  and  $10^{-6}$  dilutions for every ligation condition and controls.

■ **PAUSE POINT** Incubate overnight at 37 °C.

**37|** Subtract the number of colonies on the control plates from those on the library plates to determine the best ligation condition (giving the most colonies). Repeat the ligation with the optimal ratio of vector to insert.

**38|** Dialyze the DNA and transform the library into DH5 $\alpha$ -E as in Step 35.

**39|** After recovery at 37 °C for an hour, plate  $10^{-6}$  dilution on an agar plate containing the antibiotic marker corresponding to the SICLOPPS plasmid. Plate the rest of the cells onto a library agar plate (the reference plate whose bacterial colonies will be used for further applications) containing the same antibiotic marker.

■ **PAUSE POINT** Incubate both plates overnight at 37 °C.

#### ? TROUBLESHOOTING

**40|** Determine library size by counting the number of colonies on the  $10^{-6}$  dilution plate (or the dilution plate containing the smallest number of colonies).

**41|** Scrape the colonies off the library plate into 10 ml LB with a glass scraper. Pellet the cells in a clinical centrifuge (1300g, 10 min). Resuspend the cells in 2 ml LB. The SICLOPPS library is now in hand.

■ **PAUSE POINT** To store the library for future use, place 0.5 ml aliquots in 1.7 ml Eppendorf tubes. Freeze in liquid nitrogen and store at  $-78$  °C.

**42|** (OPTIONAL) If you require the library for a biological screen or a genetic selection, purify the plasmids from 500  $\mu\text{l}$  of the resuspended cells with a Qiagen plasmid midi kit.

**43|** (OPTIONAL) Transform the above purified plasmids into the reporter strain. Initiate cyclic peptide production by adding the appropriate inducer (arabinose, IPTG) according to the promoter on your chosen vector. DNA sequencing of the corresponding SICLOPPS plasmid readily identifies active cyclic peptides.

#### Purification and analysis of individual library members ● TIMING 2 days

**44|** Inoculate 100 ml of LB broth with cells containing the SICLOPPS plasmid of interest. Incubate at 37 °C with agitation until OD<sub>600</sub> of 0.6 is reached. Induce SICLOPPS with the appropriate compound (IPTG, arabinose, etc.) for the promoter/repressor in your vector. Continue to incubate at 37 °C for 2 h.

**45|** Pellet the cells by centrifugation (12,000g; 6,000 r.p.m. in a Sorvall evolution RC centrifuge, Sorvell SLC-6000 rotor) for 10 min. Resuspend the pellet in 4 ml of chitin buffer and lyse the cells by sonication. Pellet the insoluble debris by centrifugation (85,000g; 16,000 r.p.m. in a Sorvall evolution RC centrifuge, Sorvell SLC-6000 rotor) for 30 min at 4 °C.

**46|** Add the soluble fraction to a column containing 200  $\mu\text{l}$  of chitin beads and wash the column with 5 ml of chitin buffer.

■ **PAUSE POINT** Leave the inteins process on the chitin column overnight at room temperature.

**47|** Remove around five beads of resin from the column and suspend in a tube with 100  $\mu\text{l}$  of SDS-PAGE loading buffer. Incubate in a boiling water bath for 10 min.

**48|** Centrifuge the samples (110,000g; 18,000 r.p.m. in a Sorvall evolution RC centrifuge, Sorvell SLC-6000 rotor) for 10 min and load 2  $\mu\text{l}$  onto the SDS-PAGE gel. Run with Tris-glycine buffer at 250 V until the tracking dye is 5 mm from the bottom of the gel. Remove the gel and cover with stain solution, carefully heat the gel and stain in a microwave oven for 2–3 min until the gel is stained.

## PROTOCOL

**! CAUTION** Take care not to let the stain solution boil over when using the microwave. Also the stain solution will be extremely hot, so handle with care.

**49|** Remove the stain solution and cover the gel with de-stain solution and agitate until protein bands are visible. As some SICLOPPS libraries may produce peptides that are too small to be visible on the Tris/glycine gel, the processed intein fragments ( $I_N = 14$  kDa,  $I_C = 4$  kDa) serve as suitable markers of cyclization (**Fig. 4**).

**50|** (OPTIONAL) To isolate cyclic peptides, repeat Steps 44 and 45 using 1 liter of LB broth, inoculating with 10 ml of overnight culture of the colony containing the SICLOPPS plasmid of interest. After the second centrifugation (having resuspended the cells in 40 ml of chitin buffer), extract the soluble fraction with 40 ml of n-BuOH. Separate the organic phase and remove the solvent *in vacuo*. The pure peptide obtained is further analyzed or purified by HPLC.

### ● TIMING

This procedure is expected to take around 3–4 weeks to implement, depending on the time taken for DNA sequencing.

### ? TROUBLESHOOTING

**Problem:** The ligation/transformation step is not working.

**Solution:** Optimize the insert/vector ratio for the ligation step as outlined in Steps 32 and 33. Also make sure the T4 DNA ligase buffer is at room temperature and that all its components are fully dissolved in solution.

**Problem:** Library size is much lower than expected.

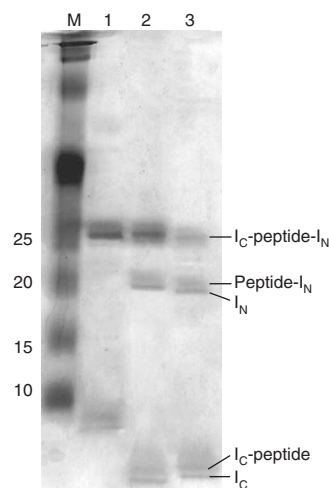
**Solution:** As library size depends on transformation efficiency, ensure that steps are taken to maximize this process<sup>22</sup>, such as using maximum efficiency electrocompetent cells, and pre-warming the SOC medium to 37 °C and adding it rapidly after the electric pulse. Be sure to optimize the ligation step between the library insert and SICLOPPS vector as outlined in Steps 32 and 33.

### ANTICIPATED RESULTS

In our hands, the above procedure yields cyclic peptide libraries of between  $10^7$  and  $10^8$  transformants. It should be noted that despite the relative promiscuity of the *Synechocystis* sp PCC6803 DnaE intein, not all peptide sequences are cyclized. The cyclization efficiency of a library is sequence dependent and can be probed by MALDI mass spectrometry or SDS-PAGE analysis of random individual members.

As well as peptides, SICLOPPS has also been used to cyclize the *E. coli* enzyme dihydrofolate reductase (DHFR)<sup>20</sup>. DHFR had been previously circularly permuted<sup>24–26</sup> and cyclized via a disulfide bond<sup>27</sup>. The cyclic DHFR produced by SICLOPPS was shown to be resistant to proteolysis and had steady-state kinetic parameters, and substrate, cofactor and methotrexate dissociation constants that were indistinguishable from those of the wild-type enzyme at 25 °C. Activity assays conducted after preincubation of wild-type and cyclic DHFR at 65 °C showed an improvement in the thermostability of the cyclic enzyme.

Despite lower transformation efficiencies, application of this split-intein strategy in eukaryotes is of interest. To this end, SICLOPPS has been ported for use in mammalian cells<sup>28</sup> and efforts are underway in our laboratories to port this system into *Saccharomyces cerevisiae*.



**Figure 4 |** SDS-PAGE data. Lane 1 shows an unprocessed intein, lanes 2 and 3 show processed inteins, randomly selected from a C + 5 peptide library.

**COMPETING INTERESTS STATEMENT** The authors declare no competing financial interests.

Published online at <http://www.natureprotocols.com>

Rights and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>

1. Boldt, G.E., Dickerson, T.J. & Janda, K.D. Emerging chemical and biological approaches for the preparation of discovery libraries. *Drug Discov. Today* **11**, 143–148 (2006).
2. Borghouts, C., Kunz, C. & Groner, B. Peptide aptamers: recent developments for cancer therapy. *Expert Opin. Biol. Ther.* **5**, 783–797 (2005).
3. Chang, K., Elledge, S.J. & Hannon, G.J. Lessons from nature: microRNA-based shRNA libraries. *Nat. Methods* **3**, 707–714 (2006).

4. Matsumoto, S., Akashi, H. & Taira, K. Screening and determination of gene function using randomized ribozyme and siRNA libraries. *Handb. Exp. Pharmacol.* 197–221 (2006).
5. Rothe, A., Hosse, R.J. & Power, B.E. Ribosome display for improved biotherapeutic molecules. *Expert Opin. Biol. Ther.* **6**, 177–187 (2006).
6. Baldwin, J.J. Design, synthesis and use of binary encoded synthetic chemical libraries. *Mol. Divers* **2**, 81–88 (1996).
7. Edwards, P.J. & Morrell, A.I. Solid-phase compound library synthesis in drug design and development. *Curr. Opin. Drug Discov. Dev.* **5**, 594–605 (2002).
8. Huwe, C.M. Synthetic library design. *Drug Discov. Today* **11**, 763–767 (2006).
9. Tan, D.S. Diversity-oriented synthesis: exploring the intersections between chemistry and biology. *Nat. Chem. Biol.* **1**, 74–84 (2005).

10. Webb, T.R. Current directions in the evolution of compound libraries. *Curr. Opin. Drug Discov. Dev.* **8**, 303–308 (2005).
11. Irwin, J.J. How good is your screening library? *Curr. Opin. Chem. Biol.* **10**, 352–356 (2006).
12. Falciani, C., Lozzi, L., Pini, A. & Bracci, L. Bioactive peptides from libraries. *Chem. Biol.* **12**, 417–426 (2005).
13. Ulrich, H. DNA and RNA aptamers as modulators of protein function. *Med. Chem.* **1**, 199–208 (2005).
14. Uttamchandani, M., Wang, J. & Yao, S.Q. Protein and small molecule microarrays: powerful tools for high-throughput proteomics. *Mol. Biosyst.* **2**, 58–68 (2006).
15. Horswill, A.R., Savinov, S.N. & Benkovic, S.J. A systematic method for identifying small-molecule modulators of protein–protein interactions. *Proc. Natl. Acad. Sci. USA* **101**, 15591–15596 (2004).
16. Tavassoli, A. & Benkovic, S.J. Genetically selected cyclic-peptide inhibitors of AICAR transformylase homodimerization. *Angew. Chem. Int. Ed. Engl.* **44**, 2760–2763 (2005).
17. Gururaja, T.L., Narasimhamurthy, S., Payan, D.G. & Anderson, D.C. A novel artificial loop scaffold for the noncovalent constraint of peptides. *Chem. Biol.* **7**, 515–527 (2000).
18. Jermutus, L., Honegger, A., Schwesinger, F., Hanes, J. & Pluckthun, A. Tailoring *in vitro* evolution for protein affinity or stability. *Proc. Natl. Acad. Sci. USA* **98**, 75–80 (2001).
19. Scott, C.P., Abel-Santos, E., Jones, A.D. & Benkovic, S.J. Structural requirements for the biosynthesis of backbone cyclic peptide libraries. *Chem. Biol.* **8**, 801–815 (2001).
20. Scott, C.P., Abel-Santos, E., Wall, M., Wahnou, D.C. & Benkovic, S.J. Production of cyclic peptides and proteins *in vivo*. *Proc. Natl. Acad. Sci. USA* **96**, 13638–13643 (1999).
21. Wu, H., Hu, Z. & Liu, X.Q. Protein trans-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. USA* **95**, 9226–9231 (1998).
22. Sambrook, J. & Russell, D.W. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001).
23. Ausubel, F.M. *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology* (New York, 2002).
24. Buchwalder, A., Szadkowski, H. & Kirschner, K. A fully active variant of dihydrofolate reductase with a circularly permuted sequence. *Biochemistry* **31**, 1621–1630 (1992).
25. Uversky, V.N. *et al.* Circularly permuted dihydrofolate reductase possesses all the properties of the molten globule state, but can resume functional tertiary structure by interaction with its ligands. *Protein Sci.* **5**, 1844–1851 (1996).
26. Iwakura, M. & Nakamura, T. Effects of the length of a glycine linker connecting the N- and C-termini of a circularly permuted dihydrofolate reductase. *Protein Eng.* **11**, 707–713 (1998).
27. Iwakura, M. & Honda, S. Stability and reversibility of thermal denaturation are greatly improved by limiting terminal flexibility of *Escherichia coli* dihydrofolate reductase. *J. Biochem. (Tokyo)* **119**, 414–420 (1996).
28. Kinsella, T.M. *et al.* Retrovirally delivered random cyclic peptide libraries yield inhibitors of interleukin-4 signaling in human B cells. *J. Biol. Chem.* **277**, 37512–37518 (2002).