

RESEARCH ARTICLE

Open Access

An intein with genetically selectable markers provides a new approach to internally label proteins with GFP

Richard Ramsden, Luther Arms, Trisha N Davis and Eric GD Muller*

Abstract

Background: Inteins are proteins that catalyze their own removal from within larger precursor proteins. In the process they splice the flanking protein sequences, termed the N- and C-terminal exteins. Large inteins frequently have a homing endonuclease that is involved in maintaining the intein in the host. Splicing and nuclease activity are independent and distinct domains in the folded structure. We show here that other biochemical activities can be incorporated into an intein in place of the endonuclease without affecting splicing and that these activities can provide genetic selection for the intein. We have coupled such a genetically marked intein with GFP as the N-terminal extein to create a cassette to introduce GFP within the interior of a targeted protein.

Results: The *Pch* PRP8 mini-intein of *Penicillium chrysogenum* was modified to include: 1) aminoglycoside phosphotransferase; 2) imidazoleglycerol-phosphate dehydratase, His5 from *S. pombe*; 3) hygromycin B phosphotransferase; and 4) the transcriptional activator LexA-VP16. The proteins were inserted at the site of the lost endonuclease. When expressed in *E. coli*, all of the modified inteins spliced at high efficiency. Splicing efficiency was also greater than 96% when expressed from a plasmid in *S. cerevisiae*. In addition the inteins conferred either G418 or hygromycin resistance, or histidine or leucine prototrophy, depending on the inserted marker and the yeast genetic background. DNA encoding the marked inteins coupled to GFP as the N-terminal extein was PCR amplified with ends homologous to an internal site in the yeast calmodulin gene *CMD1*. The DNA was transformed into yeast and integrants obtained by direct selection for the intein's marker. The His5-marked intein yielded a fully functional calmodulin that was tagged with GFP within its central linker.

Conclusions: Inteins continue to show their flexibility as tools in molecular biology. The *Pch* PRP8 intein can successfully tolerate a variety of genetic markers and still retain high splicing efficiency. We have shown that a genetically marked intein can be used to insert GFP in one-step within a target protein *in vivo*.

Background

Inteins are small proteins naturally found either embedded within a larger precursor protein (cis-acting inteins), or split between two pro-proteins (trans-acting inteins). They have the unique ability to self-catalyze their excision from the pro-protein(s) and in the process form a peptide bond between the flanking amino acid sequences that reside at the N- and C-terminal boundaries with the intein [1,2]. These flanking regions adjacent to the intein are referred to as exteins, analogous to the terminology for introns and exons. An intein acts

autonomously, splicing together the exteins without the need of a co-factor or an assisting complex. Since splicing is post-translational, for cis-acting inteins a single open reading frame driven by a single promoter expresses two distinct mature protein products: the spliced exteins and the intein.

The first intein described, *Sce* VMA1, was found in a subunit of the vacuolar H⁺-adenosine triphosphatase complex of the budding yeast *Saccharomyces cerevisiae* [3,4]. Currently over 500 inteins are known, found in over 200 species of microorganisms in over 70 different proteins [5]. Present in all three phylogenetic domains of life, inteins in Eukaryotes are predominantly found in fungi and unicellular alga [5].

* Correspondence: emuller@uw.edu
Department of Biochemistry, University of Washington, Seattle, WA, USA

A homing endonuclease is often present within cis-acting inteins. The splicing and endonuclease domains are structurally and functionally distinct, as shown both by the crystal structure of *Scd* VMA1 [6] and the conservation of splicing following the deletion of the homing endonuclease [1,7]. In addition minimal or mini-inteins lacking the homing endonuclease occur naturally or have been engineered to be as small as 143 amino acids [8,9].

The use of inteins in molecular biology, biotechnology and drug discovery has recently been reviewed [10-12]. The broad spectrum of applications for inteins is in part due to the minimal constraints placed on the sequence of the exteins. The only absolute requirement is the presence of cysteine, serine or threonine at the downstream splice junction in the C-terminal extein. There are no known size constraints. Splicing efficiency is reduced if the extein domains are unfolded or have poor solubility, but *in vitro* even this can usually be overcome with detergents or refolding regimes [13]. The immediate flanking sequence of the intein can affect splicing efficiency to varying degrees [14], but a variety of sequences are tolerated. Thus the inherent tolerance of inteins allows them to be adapted to splice proteins of diverse character.

In this study the flexibility of inteins has been exploited to investigate the potential for inteins to direct *in vivo* gene modification, specifically the tagging of an internal region of a protein with GFP. The microscopy group of the Yeast Resource Center is focused on using Förster resonance energy transfer, FRET, to study the organization of large protein complexes *in vivo* [15-18]. A limiting step for FRET studies is identifying regions within proteins that can accept a GFP-related fluorescent protein tag. As a first step towards overcoming this limitation, we engineered genetic markers within the *Pch* PRP8 intein that provided a way to internally tag proteins in one step in yeast. The *Pch* PRP8 intein was chosen for its robust, temperature-independent splicing activity [9,19] and the engineered inteins continued to display very high splicing efficiency.

Results

Pch PRP8 intein splices with inserts at the empty endonuclease site

During the course of their initial characterization of the *Pch* PRP8 mini-intein Elleuche *et al.* [19] aligned the sequence of the intein with PRP8 inteins from other species. The position of a presumptive lost homing endonuclease was identified from a gap in the alignment. This position, designated “E”, later was shown to tolerate a deletion of six amino acids without loss of splicing activity [9]. In addition the intein could be split in half at this position and the resulting two

independent proteins, when expressed in *E. coli*, had active *trans*-splicing activity [20].

Given the tolerance of the “E” position for change, this site in the *Pch* PRP8 intein was chosen for the insertion of genes encoding genetically selectable enzymes. These enzymes were: 1) aminoglycoside phosphotransferase; 2) imidazoleglycerol-phosphate dehydratase from *S. pombe*; and 3) hygromycin B phosphotransferase. They confer G418/kanamycin resistance, histidine prototrophy and hygromycin resistance in the appropriate yeast genetic background, respectively. In addition the gene encoding the transcriptional activator LexA-VP16 [21] was inserted. These proteins share in common a modest size (216, 269, 280 and 341 amino acids for His5, G418^R, LexA-VP16 and Hyg^R, respectively) all below the typical size of homing endonucleases (350-450 aa). Perhaps more importantly the N- and C-terminal ends of each protein are proximal to each other [22-25]. A schematic diagram of these constructs is shown in Figure 1.

We used the experimental system of Elleuche *et al.* [19] to test whether the *Pch* PRP8 intein with inserted markers could splice. In their system GST is the N-terminal extein and a 6 × His-tag is the C-terminal extein. The embedded intein is expressed in *E. coli* and splicing results in the production of GST-6 × His and the released intein.

We confirmed the robust splicing activity of *Pch* PRP8 intein (Figure 2, lane 2). GST-6 × His and the released intein are clearly visible in SDS-PAGE, and the identity of spliced GST-6 × His was confirmed by Western blot analysis with anti-6 × His and anti-GST antibodies. The addition of cloning sites flanked by flexible protein linkers at position E did not impair splicing since the free inteins and the GST-6 × His were clearly visible in SDS-PAGE with no evidence of unspliced products (Figure 2, lane 3).

The modified inteins with the 4 different selection marker cassettes also showed highly efficient splicing

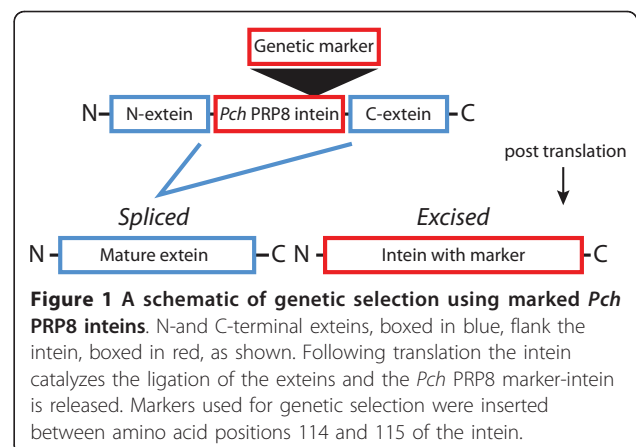
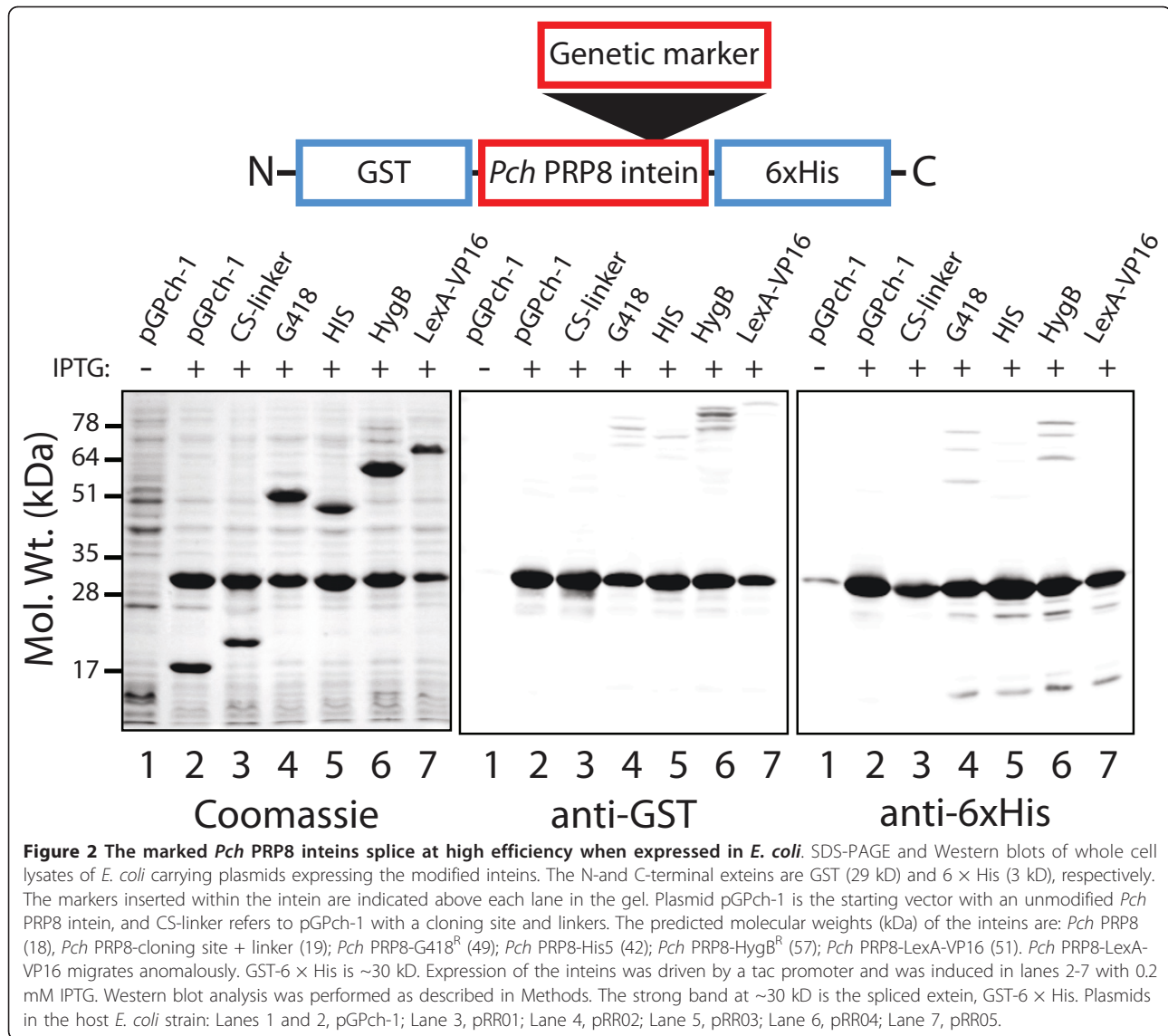


Figure 1 A schematic of genetic selection using marked *Pch* PRP8 inteins. N- and C-terminal exteins, boxed in blue, flank the intein, boxed in red, as shown. Following translation the intein catalyzes the ligation of the exteins and the *Pch* PRP8 marker-intein is released. Markers used for genetic selection were inserted between amino acid positions 114 and 115 of the intein.



(Figure 2, lanes 4-7). Based on the ratio of spliced products to larger, unspliced proteins in the Western blots, the splicing efficiencies ranged from ~88% for the *Pch* PRP8 Hyg^R-intein to almost 100% for the His5- and LexA-VP16-inteins. Thus the *Pch* PRP8 intein tolerated the insertion of four different proteins at the vacant endonuclease site, maintaining nearly complete excision and splicing activity in this system.

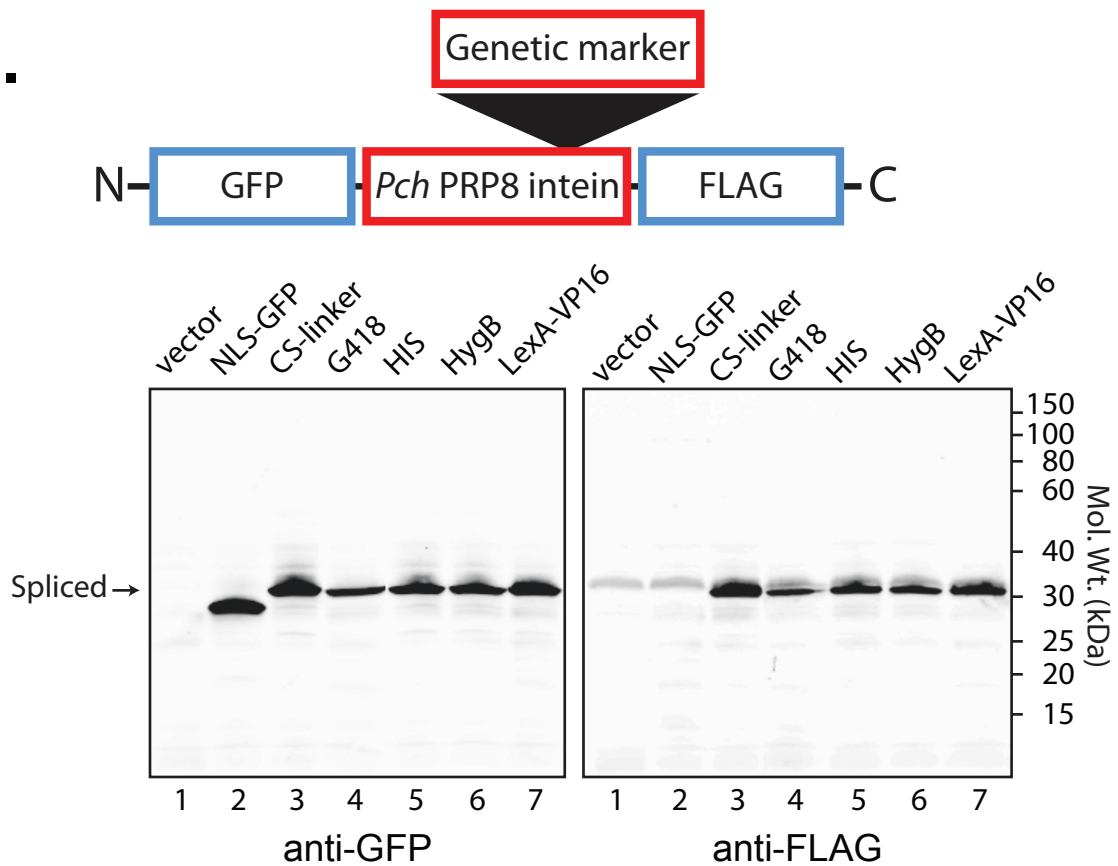
The marked *Pch* PRP8 inteins splice in yeast and confer selectable genetic traits

To determine whether the modified inteins would function in yeast we constructed a new panel of yeast *CEN4/URA3* plasmids that contained the marked inteins expressed under the control of the *GAL1* promoter. For these plasmids the N-terminal extein

was GFP containing a nuclear localization signal (NLS) and the C-terminal extein was the FLAG peptide, DYKDDDDK. Just as in the previous *E. coli* plasmids, the extein/intein junction encodes five N- and four C-terminal extein residues of the *P. chrysogenum* Prp8 protein. The spliced product is NLS-GFP-FLAG (33 kDa) and the released intein carries the selectable genetic marker or a transcription factor. NLS-GFP provided a simple way to follow expression and proper folding of the extein.

In agreement with the results in *E. coli*, all of the modified inteins when expressed in yeast displayed high efficiency splicing (Figure 3). Based on the limits of detection in the anti-GFP Westerns, we estimate that the efficiency of splicing is greater than 96% for the modified inteins.

A.



B.

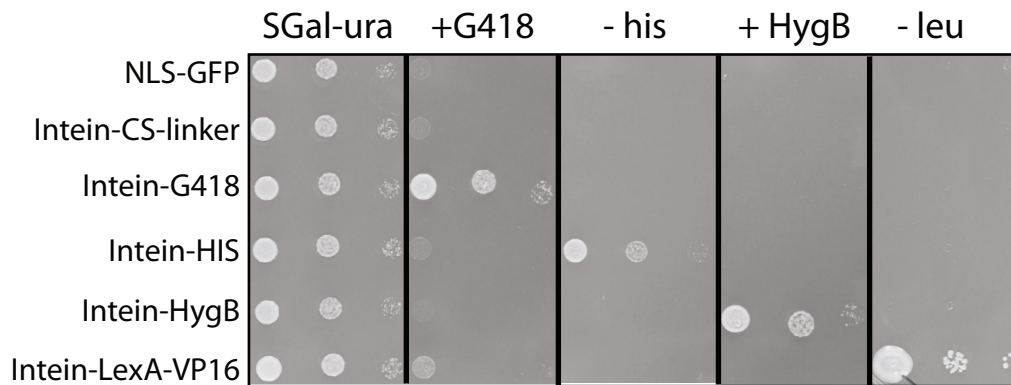


Figure 3 Marked *Pch* PRP8 inteins splice and allow genetic selection in yeast. The inteins described in Figure 2, but with different exteins, were expressed in yeast under the control of the *GAL1* promoter. For expression in yeast, the N-extein was GFP with a nuclear localization signal, and the C-extein was a FLAG epitope tag. The spliced extein product is NLS-GFP-FLAG (33 kD). The molecular weights of the *Pch* marker-inteins are given in Figure 2. However in the anti-GFP and anti-FLAG Westerns there is no clear evidence for unspliced or partially spliced products of higher molecular weight than the spliced product, such as seen in Figure 2. **A.** Western blots showing the high efficiency of splicing in strain YL (2,4) LU expressing the modified inteins (and controls) from the following plasmids: Lane 1, pBM258; Lane 2, pRR26; Lane 3, pRR21; Lane 4, pRR22; Lane 5, pRR23; Lane 6, pRR24; Lane 7, pRR25. Analysis was performed as described in Methods. The genetic markers of the inteins and antibodies used for Western blots are indicated above and below the gels, respectively. Note that in the anti-FLAG western a background band was apparent in all lanes, including the protein extracts from the vector alone and NLS-GFP expression in lanes 1 and 2. **B.** Five-fold serial dilutions of the various yeast strains were prepared and spotted onto SGal-ura medium, SGal-ura medium lacking histidine or leucine, or SGal-ura supplemented with antibiotic as indicated. The plasmid backbone confers uracil prototrophy.

Concomitant with splicing the intein is released from the precursor protein. The inteins did not inhibit growth and provided resistance to G418 or hygromycin in cells expressing the *Pch* PRP8 G418^R-intein or the HygB^R-intein, respectively (Figure 3B). Likewise the *Pch* PRP8 His5-intein was able to complement the mutation in *HIS3* in this genetic background (Figure 3B). Finally the *Pch* PRP8 LexA-VP16-intein was able to drive expression of a *LEU2* gene that had 2 *lexA* operator sequences within its promoter and confer selectable leucine prototrophy to the yeast strain. Thus in all cases each marker remained functional within the context of the excised intein.

Direct integration of the marked inteins at the *CMD1* locus

Having established that the intein remained active and the embedded markers functional in yeast, we next examined whether the marked inteins could provide a method to introduce GFP within the coding sequence of a targeted gene. For our test we chose *CMD1*, the gene encoding yeast calmodulin [26]. Calmodulin contains two globular EF-hand calcium-binding domains separated by a flexible linker. The design of the experiment is presented in Figure 4. Briefly the GFP-*Pch* PRP8 marker-intein cassette was PCR amplified using oligonucleotide primers with ends homologous to the center of the

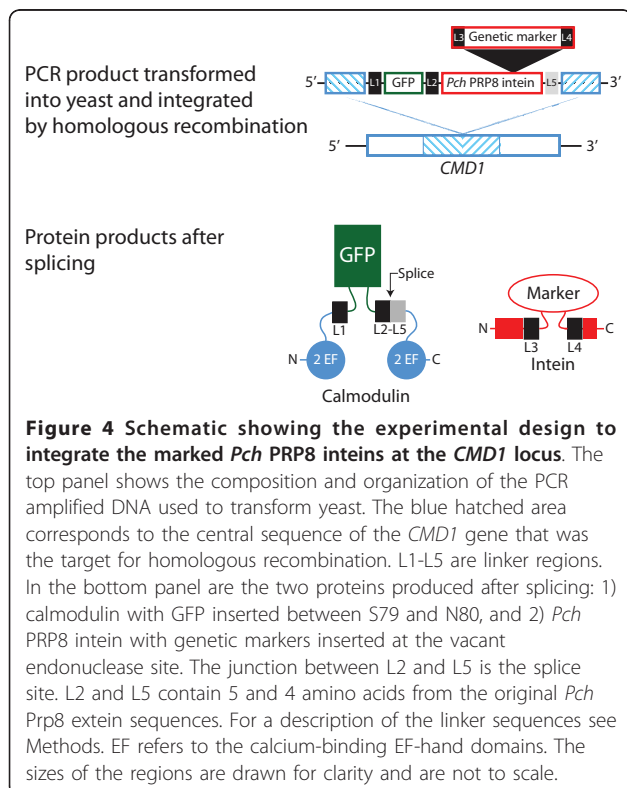
CMD1 gene encoding the central linker region. Transformation and direct selection for the markers within the inteins isolates cells in which the *Pch* PRP8 GFP-intein cassettes were inserted by homologous recombination between amino acids 79 and 80 of calmodulin.

Transformation with DNA from the *Pch* PRP8 G418^R-, His5- and HygB^R-inteins produced transformants with the selected auxotrophic or antibiotic resistance, while the *Pch* PRP8 LexA-VP16-intein did not yield transformants. Since the level of expression of the intein is dependent on the level of *CMD1* expression we conclude that the LexA-VP16 was not at a level to fully activate *lexAop₆-LEU2* gene expression in strain YL [6,4] LU. However in agreement with the previous results from yeast episomal expression, the recovery of resistant or prototrophic cells from the other *Pch* PRP8 marked-inteins again showed that the intein domain did not impair the activity of the selection markers.

In our experimental design the intein was initially transformed into a diploid strain so that transformants could be recovered even if the integration of the intein impaired the expression or activity of calmodulin. Diploid transformants were sporulated and the tetrads dissected to yield four haploids, two with the intein insertion and two without. The results of the tetrad analysis are shown in Figure 5A. The *Pch* PRP8 His5-intein *CMD1* insertions grew at the same rate as wild-type, but both the *Pch* PRP8 G418^R- and HygB^R-inteins impaired growth.

Associated with the impaired growth was evidence of inefficient splicing (Figure 5B). The *Pch* PRP8 His5-intein was effectively spliced from calmodulin, producing Cmd1-GFP at high efficiency (lane 3). On the other hand, the G418^R-intein (lanes 4,5) and HygB^R-intein (not shown) were not. Different isolates showed varying levels of accumulation of unspliced product (lanes 4 and 5). The slow growth and phenotypic variation of the G418^R and Hyg^R isolates precluded further analysis and in any case meant that these inteins were not suitable for tagging Cmd1.

Calmodulin concentrates at sites of cell growth and at the spindle pole body in yeast [27-29]. The wild-type localization pattern was seen for Cmd1 with the internal GFP-His5-intein (Figure 5C). In agreement with the published results, unbudded cells had calmodulin in puncta at the cell membrane. In newly budded cells calmodulin formed a patch at the bud tip that dispersed as the bud grew. Preceding cytokinesis calmodulin was concentrated at the bud neck. Localization at the pair of spindle pole bodies is best seen in the medium budded cell in Figure 5. For comparison the online Yeast Resource Center Public Image Repository [29] contains over 2000 images of yeast showing the localization of *CMD1* tagged with different fluorescent proteins.



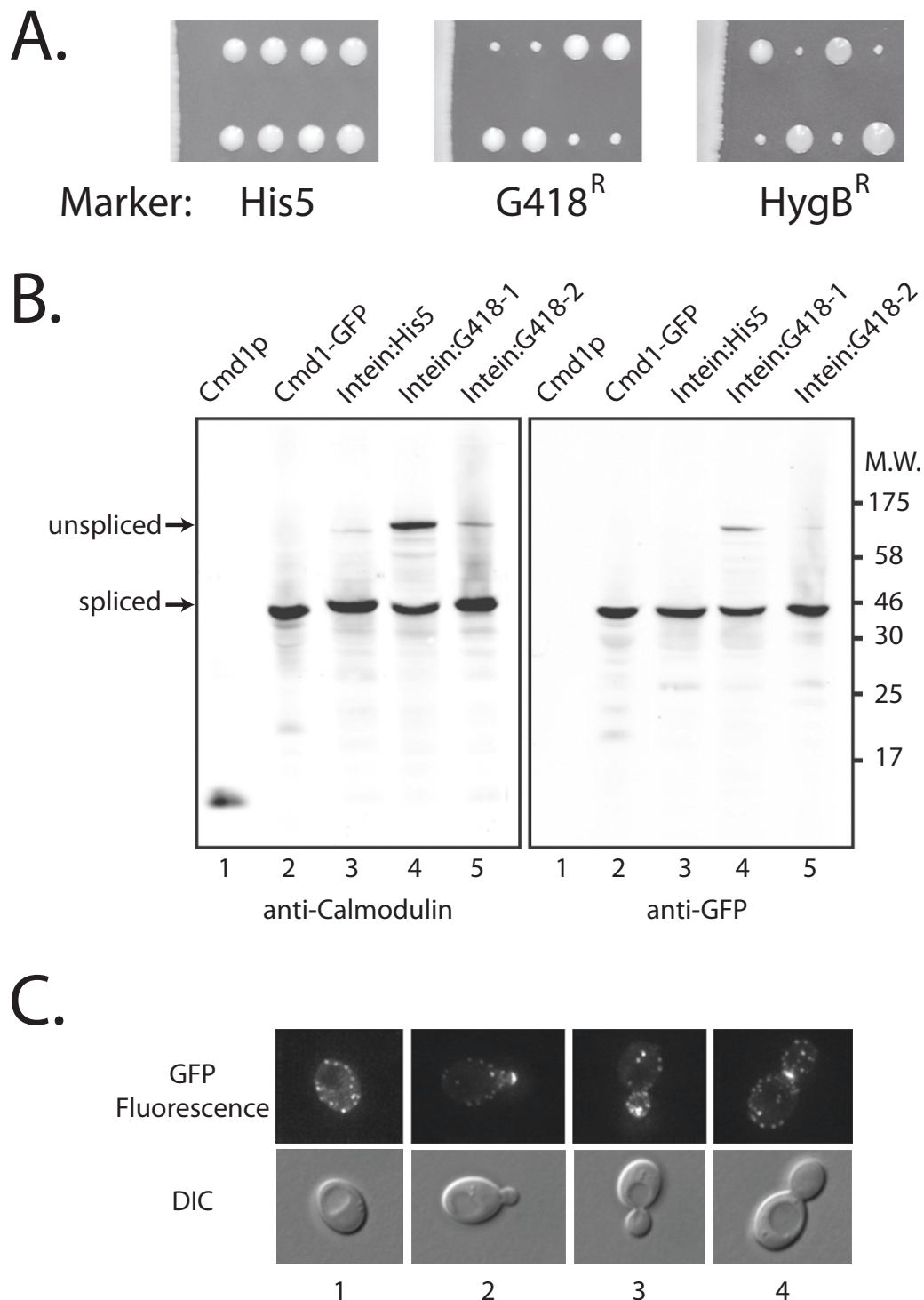


Figure 5 Integration of the marked *Pch* PRP8 inteins at *CMD1* locus. A. Tetrad analysis of diploids that are heterozygous for the insertion of the marked inteins at the *CMD1* locus (*CMD1/CMD1::GFP-Pch PRP8 marker-intein*). Individual haploid spores were grown on non-selective YPD medium. Segregation of the mutant alleles was 2:2 and the slow growing colonies contain *CMD1* with the G418^R- and HygB^R-intein insertions as determined by plating the haploids on selective medium (not shown). **B.** Western blot analysis of protein extracts from haploids containing the *Pch* PRP8 marked-inteins at the *CMD1* locus. Lane 1, 100 ng purified calmodulin; Lane 2 LAY18-2C expressing Cmd1-GFP (constructed using the classic PCR method [41]); Lane 3, extract from a haploid *CMD1::GFP-Pch PRP8-HIS5-intein* isolate; Lanes 4, 5 extracts from two separate haploid *CMD1::GFP-Pch PRP8-G418^R-intein* isolates. **C.** Fluorescence microscopy of the haploid *CMD1::GFP-Pch PRP8-HIS5-intein* isolate. Numbers 1-4 refer to images showing unbudged, small budded, medium budded or large budded cells, respectively.

Discussion

Inteins are widely used in biotechnology to manipulate proteins. The capacity of inteins to break and form peptide bonds has been exploited to develop a range of methodologies from a routine procedure to purify proteins, to more advanced semi-synthetic applications that assemble proteins for NMR, crystallography and mechanistic studies. These uses and others have been recently reviewed [10,11,30].

The use of inteins as genetic tools is still an emerging area of research. Temperature-sensitive inteins [31-33], light-activated inteins [34], and inteins controlled by small molecules [35-37], provide a way to control protein activity by controlling when a protein that has been interrupted by an intein is spliced together to yield its mature form. Conditional protein assembly by inteins *in vivo* is a promising method to generate temperature-sensitive mutants, and target when and where a protein is functional in a cell.

Here we add a new role for inteins in genetic manipulations, as a genetic marker with the unique property of self-excision. The engineered intein marks the expression of the protein into which it is embedded and yet after splicing is not a part of the mature protein. We demonstrate in yeast the use of a *Pch* PRP8 His5-marked intein to deliver, in one step and *in vivo*, GFP within a targeted protein without disturbing the function of the protein. In principle, the method could be extended to other protein tags, and adapted to other organisms.

We created a family of genetically marked inteins based on the PRP8 intein of *Penicillium chrysogenum*. PRP8 inteins have an ancient heritage and are widespread [19,38-40]. However they are not found in *S. cerevisiae*. The *Pch* PRP8 intein and the native yeast *VMA1* intein are sufficiently divergent (21% sequence identity) to preclude homologous recombination between the two during the transformation step. For comparison the *S. cerevisiae* *HIS3* and the *S. pombe* *his5⁺* genes share 58% sequence identity and yet do not recombine during integrative transformation [41]. The sequence divergence, the fact that the *Pch* PRP8 intein is well characterized [9,19,20,39], splices at high efficiency across a broad range of temperatures [19] and lacks a DNA binding domain led to its choice in our studies.

The primary sequence of the *Pch* PRP8 mini-intein shares sequence similarity with other inteins [20]. Thus its three-dimensional structure likely conforms to the canonical Hedgehog/Intein (HINT) protein scaffold [42-48]. The HINT module has a flat horseshoe shape with a series of β -strands forming the framework. The active-site is centered at the interior base or "toe" of the horseshoe. Endonucleases if present in the intein are a

separate domain located away from the active site region at the "heel" of the horseshoe.

Here we show that the *Pch* PRP8 intein, and presumably other inteins, have a remarkable capacity to accept foreign protein domains at the region of the structure where the homing endonuclease resides in full inteins. High splicing efficiency was observed for all of the inserted single domain proteins that were used for genetic markers. We attribute the success to several factors. An alignment of the *Pch* PRP8 intein with the *Synechocystis* sp. DnaB and the *Mycobacterium xenopi* GyrA inteins shows that the insertion site for our genetic markers is located in disordered loops in the crystal structures of both of these inteins [42,44]. Thus the region of insertion is inherently flexible and unstructured. In addition, one shared feature of the proteins used for markers is that their N- and C-termini are proximal [22-25]. The exit of the inserted protein domain out of the intein and its return occurs nearby in space and would tend to minimize strain in the structure of the intein. Since the majority of single domain proteins have some region of their two termini within 5 Å [49,50] and termini are consistently found at the surface of the protein [51], we predict that many single domain proteins could be inserted into inteins without detriment to splicing efficiency.

Three enzymes and a transcription factor were used as genetic markers to modify the *Pch* PRP8 intein. In pilot experiments all four modified inteins showed high splicing efficiency, both in *E. coli* and yeast, and in yeast the proteins within the excised inteins were active and enabled growth under conditions of selection. However when these inteins were used to deliver GFP within calmodulin only the strain tagged with the *Pch* PRP8 His5-intein grew normally. Under selective conditions transformation with the *Pch* PRP8 LexA-VP16-intein did not yield any transformants, and the G418^R- and Hyg^R-inteins yielded cells with poor growth and diminished splicing efficiency. The disruption of calmodulin function was surprising, given the pilot results, and is not understood. It is possible that the unspliced products which accumulate interfere with growth. We are actively investigating ways to boost the level of stable protein and the efficiency of splicing through changes in codon usage and the composition of the linkers between the various domains. In addition, given the success of the *Pch* PRP8 His5-intein, we are constructing other inteins with metabolic markers that may require lower levels of expression.

Our intein based method to insert the gene encoding GFP within an ORF offers several advantages over other PCR-based methods available in yeast. Other methods [52-54] introduce a gene with its own promoter that

must be first be removed before the GFP is correctly in frame. Thus a second recombination event is required, either catalyzed by the induction of Cre-recombinase or by a rare natural recombination event that is isolated by genetic selection, to remove the genetic marker with its associated promoter. In addition the Cre-recombinase based system leaves behind within the ORF a peptide sequence coded for by the remaining *loxP* site, a sequence that can disrupt activity (unpublished). The introduced gene also temporarily interrupts the expression of the target protein. With the intein-based method integration and tagging occur in one step. Since the marked inteins are inserted in frame, expression of the target is only disrupted during recombination. Given these advantages marked inteins show great promise and complement other available methods.

Conclusions

We have shown that the *Pch* PRP8 intein can incorporate a variety of selectable genetic markers and still exhibit high splicing activity. This led to the development of a simple method to insert GFP within the interior of a protein in yeast. This novel use of an intein has the potential to provide a way to add, substitute, or delete domains within targeted proteins in a single step in yeast. In addition it may find application in other systems as well, such as recombineering, a method that is also based on genetic manipulation through recombination.

Refinement of this method is underway at the Yeast Resource Center. Primary areas to investigate will be the effect of codon optimization, the utility of other selection markers, and cassette development for other useful tags such as the FRET partners YFP and CFP.

Methods

Strains and media

KGY315 [55] (*MATa/MAT α* , *ade2-1oc/ade2-1oc*, *ADE3/ade3 Δ 100*, *can1-100/can1-100*, *CYH2^s/cyh2^r*, *his3-11,15/his3-11,15*, *leu2-3,112/leu2-3,112*, *trp1-1/trp1-1*, *ura3-1/ura3-1*) was derived from W303 [56]. YL(2, 4)LU (*MATa*, *his3*, *trp1*, *leu2::lexAop₂-LEU2*, *ura3::lexAop₄-SPO13TATA-URA3*) is a reporter strain for LexA-VP16 expression [21]. YL(6, 4)LU is the same as YL(2, 4)LU except that there are 6 *lexAop* sequences before *LEU2*. LAY18-2C was derived from KGY315 in which the C-terminus of *CMD1* was tagged with GFP as described at the Yeast Resource Center web site http://depts.washington.edu/yeastrc/pages/fm_3.html. *Escherichia coli* strain Tuner™(DE3) (EMB, San Diego, CA) was used for expression and XL10-Gold® (Stratagene, La Jolla, CA) for routine cloning.

Yeast were grown in either synthetic dextrose minimal media (SD) or synthetic galactose minimal media with

appropriate supplements, or yeast peptone dextrose-rich medium (YPD) with added antibiotics [17]. *E. coli* was grown in NZCYM medium with antibiotics.

Plasmid construction

Plasmid pGPch-1 (gift of Stefanie Pöggeler) [19] was the source of the *Pch* PRP8 intein and contains an N-terminal extein of glutathione S-transferase (GST) and a C-terminal extein of a 6 × His epitope tag. pRR01 is derived from pGPch-1. The ΔE sequence [9], 5'-GGCGCCGATGATTCGGCT-3', coding for GAD DSA, was replaced with 5'-GGCGCCGCAG GTGCAG GTGC AGGTGCAGCT AGCGCAGCAGGTACCG CGGG TGCTGGTGCT GGTGCTTCGG CT-3' (NheI and KpnI sites underlined), to code for the protein sequence GAAGAGAGAASAAGTAGAGAGASA. Six codons within the intein sequence were optimized for expression in yeast by site-directed mutagenesis using the QuikChange® protocol (Stratagene, La Jolla, CA): Leu2 (CTC to TTG); Gly5 (GGG to GGA); Leu8 (CTC to TTG); Arg10 (CGA to AGA); Arg46 (CGC to AGA); and Arg49 (CGC to AGA). Numbering is relative to Cys (in the sequence CLAK) at position 1.

Plasmids pRR02-05 were derived from pRR1. The *G418^R* gene, the *Hyg^R* gene, the *his5⁺* gene of *Schizosaccharomyces pombe* and the *lexA-VP16* fusion were PCR amplified from pBS7, pBS4, and pDH5 (YRC, University of Washington) [57] and pBS LexA::VP16-SV40 (gift of Tzumin Lee, Janelia Farm Research Campus) [58], respectively. The ORFs were amplified with NheI and KpnI ends and inserted at the NheI/KpnI junction in pRR01. All amplifications for cloning were carried out using Vent_R® DNA polymerase (New England Biolabs, Beverly, MA).

Plasmids pRR11-15 and pRR21-25 include the intein sequence from the corresponding marked inteins in the pRR02-05 series. However the N-extein is either GFP or GFP with a nuclear localization signal, NLS-GFP, and the C-extein is a FLAG® epitope. In addition protein linker sequences have been inserted at the junctions between ORFs. These linkers, L2-L5 in Figure 4, encode the following protein sequences, with adjacent sequences in parentheses: L2, (...LYK)-GAGAGAGA-GAG-(FWEKACLAK...); L3, (...KGA)-AGAGAGAA-(SG...); L4, AGAGAGA-(SAQ...); and L5, (SGFE)-GSG. The underlined sequences are from the N- and C-terminal exteins of *P. chrysogenum* Prp8.

As described in Table 1 these constructs were cloned into 3 base plasmids: pGEX-4T-1 (Amersham Bioscience, Europe GmbH, Freiburg, Germany); pET-Duet modified to contain only one promoter (EMD Biosciences, San Diego, CA), and pBM258 [59]. The complete DNA sequences are deposited in Genbank [60].

Table 1 Plasmids

Name	Base plasmid	Promoter	Origin/marker	Extein[intein]extein
pGPch-1	pGEX-4T-1	tac	ori/Amp	GST[<i>Pch</i> PRP8]HIS ₆
pRR01	pGEX-4T-1	tac	ori/Amp	GST[<i>Pch</i> PRP8 linker]HIS ₆
pRR02	pGEX-4T-1	tac	ori/Amp	GST[<i>Pch</i> PRP8 G418 ^R]HIS ₆
pRR03	pGEX-4T-1	tac	ori/Amp	GST[<i>Pch</i> PRP8 HIS5]HIS ₆
pRR04	pGEX-4T-1	tac	ori/Amp	GST[<i>Pch</i> PRP8 HygB ^R]HIS ₆
pRR05	pGEX-4T-1	tac	ori/Amp	GST[<i>Pch</i> PRP8 LexA-VP16]HIS ₆
pRR11	pET	T7	ori/Amp	GFP[<i>Pch</i> PRP8 linker]FLAG
pRR12	pET	T7	ori/Amp	GFP[<i>Pch</i> PRP8 G418 ^R]FLAG
pRR13	pET	T7	ori/Amp	GFP[<i>Pch</i> PRP8 HIS5]FLAG
pRR14	pET	T7	ori/Amp	GFP[<i>Pch</i> PRP8 HygB ^R]FLAG
pRR15	pET	T7	ori/Amp	GFP[<i>Pch</i> PRP8 LexA-VP16]FLAG
pRR21	pBM258	Gal	cen/ura3	NLS-GFP[<i>Pch</i> PRP8 linker]FLAG
pRR22	pBM258	Gal	cen/ura3	NLS-GFP[<i>Pch</i> PRP8 G418 ^R]FLAG
pRR23	pBM258	Gal	cen/ura3	NLS-GFP[<i>Pch</i> PRP8 HIS5]FLAG
pRR24	pBM258	Gal	cen/ura3	NLS-GFP[<i>Pch</i> PRP8 HygB ^R]FLAG
pRR25	pBM258	Gal	cen/ura3	NLS-GFP[<i>Pch</i> PRP8 LexA-VP16]FLAG
pRR26	pBM258	Gal	cen/ura3	NLS-GFP (no intein)

All plasmids were constructed in this study except pGPch-1 [19]. In the table the promoter that drives expression of the extein-intein is indicated.

Episomal expression of inteins in *E. coli* and yeast

Expression in *E. coli* was performed in Tuner™ (DE3) with pLysSRARE to express rare tRNAs. Bacteria were grown to early log phase in NZCYM medium supplemented with ampicillin (100 µg/ml) and chloramphenicol (34 µg/ml) then induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 17 h at 20°. Culture densities were measured with a Klett-Summers Colorimeter.

For plasmid-based expression in yeast from the *GAL1* promoter, the strain YL (2,4) LU was used with various plasmids based on pBM258. Cells were grown overnight at 30° in S-medium -ura + 2% raffinose. When cells reached a turbidity of 20 Klett units, galactose was added to a final concentration of 2% and growth was continued to 100 Klett units.

Western blot analysis and the estimation of splicing efficiency

Equivalent amounts of cells were pelleted, precipitated with 10% TCA [61], resuspended in sample buffer and separated by SDS-PAGE. Three identical gels were prepared: one directly stained by Coomassie® Brilliant Blue for detection of total protein and two transferred to Immobilon™ (Millipore, Billerica, MA) for Western blot analysis. Antibodies were either directed against GST (Santa Cruz Biotechnology, Santa Cruz, CA), Penta-His (Qiagen, Valencia, CA), GFP (Roche Applied Science, Indianapolis, IN) or the FLAG epitope (anti-FLAG® M2, Sigma-Alrich, St. Louis, MO). Antibodies were detected using Alexa Fluor 680 anti-IgG secondary (Invitrogen, Carlsbad, CA) and visualized on an Odyssey

Infrared Imaging System (LI-COR Biosciences, Lincoln, NE). The Odyssey software was used to quantify band intensity. To estimate the limits of detection for the anti-GFP antibody, Western blot analysis was performed on serial dilutions of extracts from strain YL (2,4) LU expressing NLS-GFP from plasmid pRR26.

Genomic integration of marked inteins into the *CMD1* locus

The intein cassettes were amplified from the pRR11-15 plasmids using the Expand Long Template PCR System (Roche, Indianapolis, IN). Primers included a 40 base extension homologous to the target *CMD1* gene, a short linker sequence encoding L1 and L5 in Figure 4, and ~ 23 bp homologous to the ends of the intein sequence. The sequences of the primers were: 5'-TAGT-GAATTTTTGGCTCTGATGTCTCGTCAACTCAAATCA GGTTCAGGT AGTAAAGGAG AAGAACTTTT CACTG-3' and 5'-ATACTTTAAAAGCTTCTAG-TAGTTCTTGTTTCAGAGTCATT TCCAGATCC TT CGAAACCA CTGTTGTGCA G-3' in which bold sequence is homologous to the intein cassette and underlined is homologous to the *CMD1* site of insertion. The amplification and integration protocols are described at the Yeast Resource Center web site http://depts.washington.edu/yeastrc/pages/fm_3.html. KGY315 was the host except in the case of the transformation of the LexA-VP16-intein, which was transformed into YL [6,4] LU.

Fluorescence microscopy was performed on a Delta-Vision® imaging system (Applied Precision, Issaquah, WA) as previously reported [17]. Protein extraction and

Western blot analysis were performed as described above using anti-GFP and anti-calmodulin antibodies [62]. The correct integration of the inteins was confirmed by four different PCR reactions using the genomic DNA as template and primer pairs from these regions: 5' UTR of *CMD1* as forward primer (f.p.) × GFP as reverse primer (r.p.); GFP (f.p.) × C-terminal intein (r.p.); C-terminal intein (f.p.) × 3' UTR of *CMD1*(r.p.); and 5' UTR of *CMD1* (f.p.) × 3' UTR of *CMD1*(r.p.).

Acknowledgements

We would like to thank Stefanie Pöggeler for the gift of pGPch-1 and Tzumin Lee for the gift of pBS LexA-VP16 SV40. We would like to thank Simon Diep and Jing Huang for strain YL(6,4)LU and plasmids. This work is supported by the National Center for Research Resources of the National Institutes of Health by a grant to T.N.D. entitled 'Comprehensive Biology: Exploiting the Yeast Genome', P41 RR011823.

Authors' contributions

RR carried out the plasmid constructions, molecular genetic studies and Western blot analysis, drafted the manuscript and conceived of the project. LA carried out Western blot analysis, molecular genetic studies and fluorescence microscopy. TD helped to direct the study and draft the manuscript. EM co-ordinated and directed the study, carried out fluorescence microscopy and drafted the manuscript. All authors participated in the design of the study. All authors read and approved the final manuscript.

Received: 7 March 2011 Accepted: 27 June 2011
Published: 27 June 2011

References

- Paulus H: Protein splicing and related forms of protein autoprocessing. *Annu Rev Biochem* 2000, **69**:447-496.
- Perler FB: Protein splicing mechanisms and applications. *IUBMB Life* 2005, **57**(7):469-476.
- Hirata R, Ohsumi Y, Nakano A, Kawasaki H, Suzuki K, Anraku Y: Molecular structure of a gene, *VMA1*, encoding the catalytic subunit of H (+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *J Biol Chem* 1990, **265**(12):6726-6733.
- Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebel M, Stevens TH: Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* 1990, **250**(4981):651-657.
- Perler FB: InBase: the Intein Database. *Nucleic Acids Res* 2002, **30**(1):383-384.
- Duan X, Gimble FS, Quijoco FA: Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell* 1997, **89**(4):555-564.
- Chong S, Shao Y, Paulus H, Benner J, Perler FB, Xu MQ: Protein splicing involving the *Saccharomyces cerevisiae* VMA intein. The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system. *J Biol Chem* 1996, **271**(36):22159-22168.
- Derbyshire V, Wood DW, Wu W, Dansereau JT, Dalgaard JZ, Belfort M: Genetic definition of a protein-splicing domain: functional mini-inteins support structure predictions and a model for intein evolution. *Proc Natl Acad Sci USA* 1997, **94**(21):11466-11471.
- Elleuche S, Doring K, Pöggeler S: Minimization of a eukaryotic mini-intein. *Biochem Biophys Res Commun* 2008, **366**(1):239-243.
- Cheriyian M, Perler FB: Protein splicing: A versatile tool for drug discovery. *Adv Drug Deliv Rev* 2009, **61**(11):899-907.
- Elleuche S, Pöggeler S: Inteins, valuable genetic elements in molecular biology and biotechnology. *Appl Microbiol Biotechnol* 2010, **87**(2):479-489.
- Vila-Perello M, Muir TW: Biological applications of protein splicing. *Cell* 2010, **143**(2):191-200.
- Muralidharan V, Muir TW: Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nature methods* 2006, **3**(6):429-438.
- Amitai G, Callahan BP, Stanger MJ, Belfort G, Belfort M: Modulation of intein activity by its neighboring extein substrates. *Proc Natl Acad Sci USA* 2009, **106**(27):11005-11010.
- Kollman JM, Zelter A, Muller EG, Fox B, Rice LM, Davis TN, Agard DA: The structure of the gamma-tubulin small complex: implications of its architecture and flexibility for microtubule nucleation. *Mol Biol Cell* 2008, **19**(1):207-215.
- Mc Intyre J, Muller EG, Weitzer S, Snyderman BE, Davis TN, Uhlmann F: In vivo analysis of cohesin architecture using FRET in the budding yeast *Saccharomyces cerevisiae*. *EMBO J* 2007, **26**(16):3783-3793.
- Muller EG, Snyderman BE, Novik I, Hailey DW, Gestaut DR, Niemann CA, O'Toole ET, Giddings TH, Sundin BA, Davis TN: The organization of the core proteins of the yeast spindle pole body. *Mol Biol Cell* 2005, **16**(7):3341-3352.
- Mathieson EM, Suda Y, Nickas M, Snyderman B, Davis TN, Muller EG, Neiman AM: Vesicle docking to the spindle pole body is necessary to recruit the exocyst during membrane formation in *Saccharomyces cerevisiae*. *Molecular biology of the cell* 2010, **21**(21):3693-3707.
- Elleuche S, Nolting N, Pöggeler S: Protein splicing of PRP8 mini-inteins from species of the genus *Penicillium*. *Appl Microbiol Biotechnol* 2006, **72**(5):959-967.
- Elleuche S, Pöggeler S: Trans-splicing of an artificially split fungal mini-intein. *Biochem Biophys Res Commun* 2007, **355**(3):830-834.
- Huang J, Schreiber SL: A yeast genetic system for selecting small molecule inhibitors of protein-protein interactions in nanodroplets. *Proc Natl Acad Sci USA* 1997, **94**(25):13396-13401.
- Nurizzo D, Shewry SC, Perlin MH, Brown SA, Dholakia JN, Fuchs RL, Deva T, Baker EN, Smith CA: The crystal structure of aminoglycoside-3'-phosphotransferase-IIa, an enzyme responsible for antibiotic resistance. *J Mol Biol* 2003, **327**(2):491-506.
- Glynn SE, Baker PJ, Sedelnikova SE, Davies CL, Eadsforth TC, Levy CW, Rodgers HF, Blackburn GM, Hawkes TR, Viner R, Rice DW: Structure and mechanism of imidazoleglycerol-phosphate dehydratase. *Structure* 2005, **13**(12):1809-1817.
- Stogios PJ, Shakya T, Evdokimova E, Savchenko A, Wright GD: Structure and function of APH(4)-Ia, a hygromycin B resistance enzyme. *The Journal of biological chemistry* 2011, **286**(3):1966-1975.
- Fogh RH, Otteleben G, Ruterjans H, Schnarr M, Boelens R, Kaptein R: Solution structure of the LexA repressor DNA binding domain determined by 1 H NMR spectroscopy. *The EMBO journal* 1994, **13**(17):3936-3944.
- Davis TN, Urdea MS, Masiarz FR, Thorne J: Isolation of the yeast calmodulin gene: calmodulin is an essential protein. *Cell* 1986, **47**(3):423-431.
- Brockerhoff SE, Davis TN: Calmodulin concentrates at regions of cell growth in *Saccharomyces cerevisiae*. *J Cell Biol* 1992, **118**(3):619-629.
- Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK: Global analysis of protein localization in budding yeast. *Nature* 2003, **425**(6959):686-691.
- Riffle M, Davis TN: The Yeast Resource Center Public Image Repository: A large database of fluorescence microscopy images. *BMC bioinformatics* 2010, **11**:263.
- Flavell RR, Muir TW: Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology. *Acc Chem Res* 2009, **42**(1):107-116.
- Adam E, Perler FB: Development of a positive genetic selection system for inhibition of protein splicing using mycobacterial inteins in *Escherichia coli* DNA gyrase subunit A. *J Mol Microbiol Biotechnol* 2002, **4**(5):479-487.
- Zeidler MP, Tan C, Bellaiche Y, Cherry S, Hader S, Gayko U, Perrimon N: Temperature-sensitive control of protein activity by conditionally splicing inteins. *Nat Biotechnol* 2004, **22**(7):871-876.
- Tan G, Chen M, Foote C, Tan C: Temperature-sensitive mutations made easy: generating conditional mutations by using temperature-sensitive inteins that function within different temperature ranges. *Genetics* 2009, **183**(1):13-22.
- Tyszkiewicz AB, Muir TW: Activation of protein splicing with light in yeast. *Nat Methods* 2008, **5**(4):303-305.

35. Buskirk AR, Ong YC, Gartner ZJ, Liu DR: **Directed evolution of ligand dependence: small-molecule-activated protein splicing.** *Proc Natl Acad Sci USA* 2004, **101**(29):10505-10510.
36. Mootz HD, Blum ES, Tyszkiewicz AB, Muir TW: **Conditional protein splicing: a new tool to control protein structure and function in vitro and in vivo.** *J Am Chem Soc* 2003, **125**(35):10561-10569.
37. Mootz HD, Muir TW: **Protein splicing triggered by a small molecule.** *J Am Chem Soc* 2002, **124**(31):9044-9045.
38. Theodoro RC, Volkmann G, Liu XQ, Bagagli E: **PRP8 intein in Ajellomycetaceae family pathogens: Sequence analysis, splicing evaluation and homing endonuclease activity.** *Fungal Genet Biol* 2010.
39. Elleuche S, Pelikan C, Nolting N, Poggeler S: **Inteins and introns within the prp8 -gene of four Eupenicillium species.** *J Basic Microbiol* 2009, **49**(1):52-57.
40. Butler MI, Gray J, Goodwin TJ, Poulter RT: **The distribution and evolutionary history of the PRP8 intein.** *BMC Evol Biol* 2006, **6**:42.
41. Wach A, Brachat A, Alberti-Segui C, Rebischung C, Philippson P: **Heterologous HIS3 marker and GFP reporter modules for PCR-targeting in Saccharomyces cerevisiae.** *Yeast* 1997, **13**(11):1065-1075.
42. Ding Y, Xu MQ, Ghosh I, Chen X, Ferrandon S, Lesage G, Rao Z: **Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing.** *The Journal of biological chemistry* 2003, **278**(40):39133-39142.
43. Hall TM, Porter JA, Young KE, Koonin EV, Beachy PA, Leahy DJ: **Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins.** *Cell* 1997, **91**(1):85-97.
44. Klabunde T, Sharma S, Telenti A, Jacobs WR, Sacchetti JC: **Crystal structure of GyrA intein from Mycobacterium xenopi reveals structural basis of protein splicing.** *Nature structural biology* 1998, **5**(1):31-36.
45. Koonin EV: **A protein splice-junction motif in hedgehog family proteins.** *Trends Biochem Sci* 1995, **20**(4):141-142.
46. Perler FB: **Protein splicing of inteins and hedgehog autoproteolysis: structure, function, and evolution.** *Cell* 1998, **92**(1):1-4.
47. Sun P, Ye S, Ferrandon S, Evans TC, Xu MQ, Rao Z: **Crystal structures of an intein from the split dnaE gene of Synechocystis sp. PCC6803 reveal the catalytic model without the penultimate histidine and the mechanism of zinc ion inhibition of protein splicing.** *Journal of molecular biology* 2005, **353**(5):1093-1105.
48. Van Roey P, Pereira B, Li Z, Hiraga K, Belfort M, Derbyshire V: **Crystallographic and mutational studies of Mycobacterium tuberculosis recA mini-inteins suggest a pivotal role for a highly conserved aspartate residue.** *Journal of molecular biology* 2007, **367**(1):162-173.
49. Krishna MM, Englander SW: **The N-terminal to C-terminal motif in protein folding and function.** *Proc Natl Acad Sci USA* 2005, **102**(4):1053-1058.
50. Mounce BC, Kurt N, Ellison PA, Cavagnero S: **Nonrandom distribution of intramolecular contacts in native single-domain proteins.** *Proteins* 2009, **75**(2):404-412.
51. Jacob E, Unger R: **A tale of two tails: why are terminal residues of proteins exposed?** *Bioinformatics* 2007, **23**(2):e225-230.
52. Prein B, Natter K, Kohlwein SD: **A novel strategy for constructing N-terminal chromosomal fusions to green fluorescent protein in the yeast Saccharomyces cerevisiae.** *FEBS Lett* 2000, **485**(1):29-34.
53. Gauss R, Trautwein M, Sommer T, Spang A: **New modules for the repeated internal and N-terminal epitope tagging of genes in Saccharomyces cerevisiae.** *Yeast* 2005, **22**(1):1-12.
54. Schneider BL, Seufert W, Steiner B, Yang QH, Futcher AB: **Use of polymerase chain reaction epitope tagging for protein tagging in Saccharomyces cerevisiae.** *Yeast* 1995, **11**(13):1265-1274.
55. Greenland KB, Ding H, Costanzo M, Boone C, Davis TN: **Identification of Saccharomyces cerevisiae Spindle Pole Body Remodeling Factors.** *PLoS One* 2010, **5**(11):e15426.
56. Wallis JW, Chretien G, Brodsky G, Rolfe M, Rothstein R: **A hyper-recombination mutation in S. cerevisiae identifies a novel eukaryotic topoisomerase.** *Cell* 1989, **58**(2):409-419.
57. [<http://depts.washington.edu/yeastrc/>].
58. Lai SL, Lee T: **Genetic mosaic with dual binary transcriptional systems in Drosophila.** *Nat Neurosci* 2006, **9**(5):703-709.
59. Johnston M, Davis RW: **Sequences that regulate the divergent GAL1-GAL10 promoter in Saccharomyces cerevisiae.** *Mol Cell Biol* 1984, **4**(8):1440-1448.
60. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW: **GenBank.** *Nucleic Acids Res* 2010.
61. Wright AP, Bruns M, Hartley BS: **Extraction and rapid inactivation of proteins from Saccharomyces cerevisiae by trichloroacetic acid precipitation.** *Yeast* 1989, **5**(1):51-53.
62. Moser MJ, Lee SY, Klevit RE, Davis TN: **Ca²⁺ binding to calmodulin and its role in Schizosaccharomyces pombe as revealed by mutagenesis and NMR spectroscopy.** *J Biol Chem* 1995, **270**(35):20643-20652.

doi:10.1186/1472-6750-11-71

Cite this article as: Ramsden et al.: An intein with genetically selectable markers provides a new approach to internally label proteins with GFP. *BMC Biotechnology* 2011 **11**:71.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

