bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Control of ϕC31 integrase-mediated site-specific recombination by protein trans splicing

Femi J. Olorunniji*[1], Makeba Lawson-Williams[1], Arlene L. McPherson[2], Jane E. Paget[3], W. Marshall Stark[2], and Susan J. Rosser[3].

[1]School of Pharmacy and Biomolecular Sciences, Faculty of Science, Liverpool John Moores University, James Parsons Building, Byrom Street, Liverpool L3 3AF, U.K.

[2]Institute of Molecular, Cell and Systems Biology, University of Glasgow, Bower Building, Glasgow G12 8QQ, U.K.

[3]UK Centre for Mammalian Synthetic Biology at the Institute of Quantitative Biology, Biochemistry, and Biotechnology, SynthSys, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3JD, U.K.; Institute for Bioengineering, University of Edinburgh, Faraday Building, The King's Buildings, Edinburgh, 2 EH9 3DW, UK.

Correspondence: Femi J. Olorunniji, F.J.Olorunniji@ljmu.ac.uk; +44-151-231-2116

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

ABSTRACT

Serine integrases are emerging as core tools in synthetic biology and have applications in biotechnology and genome engineering. We have designed a split-intein serine integrase-based system for rapid regulation of site-specific recombination events *in vivo*. The ϕC31 integrase was split into two extein domains, and intein sequences (*Npu* DnaE^{N} and *Ssp* DnaE^{C}) were attached to the two termini to be fused. Expression of these two components followed by post-translational protein *trans*-splicing in *E. coli* generated a fully functional ϕC31 integrase. Protein splicing is necessary for recombination activity; no activity was observed when the ϕC31 integrase N- and C-terminal extein domains without the intein sequences were co-expressed, nor when a key intein catalytic residue was mutated. As a proof of principle, we used a bistable switch based on an invertible promoter reporter system to demonstrate a potential application of the split intein-regulated site-specific recombination system. We used *araC* and *tet* inducible promoters to regulate the expression of the two parts of the split recombinase. Inversion of a DNA segment containing a constitutive promoter, catalyzed by trans-spliced integrase, switches between RFP and GFP expression only when both inducible promoters are ON. We used the same split inteins to regulate the reconstitution of a split integrase-RDF fusion that efficiently catalyzed the reverse *attR* x *attL* recombination, demonstrating that our split-intein regulated recombination system can function as a reversible AND gate in which the forward reaction is catalyzed by the integrase, and the reverse reaction by the integrase-RDF fusion. The split-intein integrase is a potentially versatile, regulatable component for building synthetic genetic circuits and devices.

Keywords:

Site-specific Recombination, Serine Integrases, Split Inteins, Logic Gates, Synthetic Biology.

INTRODUCTION

It has recently become possible to create computational and memory systems in cells[1,2,3] allowing us to foresee many new ways to enhance the applications of living organisms[4,5]. Engineered cells could act as powerful biosensors with applications in health, environmental, and industrial processes. As well as sensing components, it is necessary to process the received information and express it as outputs in the form of specific biological responses[3,6]. However, the genetic switches and logic gates that have been constructed to date are based on a limited repertoire of biological component types[7], and there is a need for new systems that can be used to implement more elaborate and robust devices.

DNA site-specific recombination has been much exploited for rapid DNA assembly, and to build genetic switches and memory devices[4,8,9,10]. In a typical module, two recombination

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

sites flank a promoter sequence. Expression of the recombinase promotes inversion of the orientation of the promoter sequence, thus switching between expression of two genes which are divergently transcribed from the module (Figure 1A). One group of site-specific recombinases known as the serine integrases is especially suited for the construction of switching devices, particularly because these enzymes promote very efficient and highly directional recombination[11]. For such modules to be useful, fully integrated components of the cell, activity of the recombinase must be tightly regulated, so that switching occurs only when other cellular conditions are fulfilled. Here we demonstrate a powerful new approach to regulation of serine integrase activity, in which the enzyme itself is assembled by intein-mediated fusion of two precursor components.

Inteins are naturally occurring autocatalytic systems that catalyse protein splicing reactions to generate active proteins from precursor polypeptides[12]. Synthetic “split inteins” have been developed to carry out protein splicing *in trans*, covalently joining two proteins together for a wide range of biotechnological applications[13,14] (Figure 1B). For example, Schaerli *et al.*[15] split T7 polymerase into two parts and fused each part to split-intein sequences so that the two parts were covalently joined together by *trans*-splicing. Expression of each of the two individual parts was placed under the control of an inducible promoter to allow conditional expression of integrase activity.

A critical constraint on the construction of split-intein logic gates is the need to introduce an intein nucleophilic residue (typically Cys, Ser, or Thr) by mutation of the sequence of the target protein adjacent to the junction between the two ‘exteins’. Such changes can potentially lead to reduction or loss of enzyme activity. However, inteins have been engineered that tolerate variations of these flanking residues, thereby minimizing the number of changes that need to be made in the target protein[16,17].

Conditional site-specific recombinase activation by assembly of split protein fragments has been achieved; for example, for the popularly used tyrosine recombinase Cre[18-22]. Wang *et al.*[22] used a Cre split-intein system to reconstitute functional recombinase in transgenic mice. However, there are no reports to date of analogous systems using split serine integrases. The transposase TnpX (distantly related to the serine integrases) was split into two parts and some DNA-binding activity was reconstituted when both parts were present, but no recombination (transposition) activity was observed[23].

With the increasing importance of serine integrases as tools in synthetic biology, methods for signal-induced post-translational regulation of integrase activity are becoming very desirable. Here, we report post-translational activation of site-specific recombination by reconstitution of functional ϕC31 integrase using split intein-catalysed reactions.

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

RESULTS AND DISCUSSION

Design of split intein serine integrase

We designed our split-intein system for conditional expression of ϕC31 integrase, a prototype serine integrase, based on previously reported systems[15]. Since the natural DnaE split inteins used in this work require an invariant active site cysteine to remain in the spliced product protein[24-26], their use for split integrase reconstitution depends upon the identification of a short region of the protein sequence where insertion of a cysteine does not disrupt activity. In addition, for *trans*-splicing to be essential, the two extein components of the spliced integrase must not associate non-covalently to reconstitute a functional enzyme. We started by analysing the domain structure of serine integrases to determine where to split the protein, since a functional split intein could require introduction of mutations that would remain in the spliced protein product, if suitably placed natural Cys, Ser, or Thr residues were unavailable. Based on sequence alignments of serine integrases (Figure 2A) and published crystal structures[27], we split ϕC31 integrase at the non-conserved region of the recombinase domain between the β9 and αI domains; this region includes a 10-12 residue loop which is absent in some related serine integrases[28] (Figure 2A). We therefore predicted that introduction of the required intein nucleophilic residue and any flanking residues would not have a deleterious effect on integrase activity. Furthermore, Lucet *et al.*[23] found that when the related large serine recombinase TnpX (from *Clostridium perfringens*) was split between the β9 and αI domains, the two fragments complemented each other to restore DNA binding (but not recombination) activity.

We then attached two well-characterised split-intein components to the ϕC31 integrase sequences; *Npu* DnaE$^{\mathbf{N}}$, **I**$^{\mathbf{N}}$ (102 amino acids) from *Nostoc punctiforme* DnaE, and *Ssp* DnaE$^{\mathbf{C}}$, **I**$^{\mathbf{C}}$ (36 amino acids) from *Synechocystis sp*. DnaE. The *Npu* DnaE$^{\mathbf{N}}$ variant contains a L22S change and the *Ssp* DnaE$^{\mathbf{C}}$ variant has a P21R mutation[24]. This pair was chosen based on previous reports of their optimal activities at 37 °C in *E. coli* [15,24,29]. *Npu* DnaE$^{\mathbf{N}}$ (**I**$^{\mathbf{N}}$) was fused to the C-terminus of ϕC31 Integrase (Integrase-**E**$^{\mathbf{N}}$), and *Ssp* DnaE$^{\mathbf{C}}$ (**I**$^{\mathbf{C}}$) was fused to the N-terminus of ϕC31 Integrase (Integrase-**E**$^{\mathbf{C}}$) (Figure 2B). Also, residues 308-310 of Integrase- **E**$^{\mathbf{N}}$ were changed from EGY to EY, since these sequences exist naturally at the extein-intein junction of *Npu* DnaE$^{\mathbf{N}}$, and Integrase-**E**$^{\mathbf{N}}$ residues 311-313 were changed from RIQ to CFN, sequences found at the intein-extein junction of *Ssp* DnaE$^{\mathbf{C}}$ [24,29]. It is known that these flanking residues are involved in enhancing the splicing efficiencies of this pair of split inteins[25,29]. In the reconstituted *trans*-spliced integrase (ϕC31.Int$^{\boldsymbol{\alpha}}$), the natural 6-residue sequence at positions 308 to 313, EGYRIQ, is replaced with the 5-residue sequence EYCFN. We made these changes to maximize the efficiency of the splicing reaction in order to optimize activity in *E. coli*. Hence the reconstituted *trans*-spliced integrase, ϕC31.Int$^{\boldsymbol{\alpha}}$ is one amino acid residue shorter than the wild-type ϕC31 integrase. Since these changes are

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

in a non-conserved region (see above), we predicted that they would not substantially affect recombination activity.

Intein-mediated reconstitution of functional ϕC31 integrase

To assay *in vivo* recombination activity of our split-intein ϕC31 integrase, we used a well-characterized colour-based *galk* assay[30-33]. The *attP* and *attB* recombination sites (Supporting Information, Table S1) on the substrate plasmid pϕC31-delPB are in a direct repeat orientation, so that recombination between them causes deletion of the *galK* gene. Pale-coloured (*galK*−) colonies on the indicator plates indicate recombination proficiency, whereas red (*galK*+) colonies indicate incomplete or zero recombination. The coding sequences for Integrase-$\mathbf{E^N}$-$\mathbf{I^N}$ and $\mathbf{I^C}$-Integrase-$\mathbf{E^C}$ (and also appropriate control proteins; see Figure 3C) were cloned into separate low-level expression vectors[30], as illustrated in Figure 3B. An *E. coli* strain containing the test substrate was transformed with these plasmids, and recombination activity was assessed by colony colour and by gel electrophoresis analysis of plasmid DNA recovered from the cells (Figure 3C).

Neither Integrase-$\mathbf{E^N}$-$\mathbf{I^N}$ (Figure 3C, lane 3) nor $\mathbf{I^C}$-Integrase-$\mathbf{E^C}$ (Figure 3C, lane 4) on their own were able to catalyse *attP* x *attB* recombination. Furthermore, co-expression of the N-extein (Integrase-$\mathbf{E^N}$) and C-extein (Integrase-$\mathbf{E^C}$) components of the integrase (lane 5) gave no recombination. This important control shows that the intein-less precursor proteins Integrase-$\mathbf{E^N}$ and Integrase-$\mathbf{E^C}$ do not complement each other by non-covalent association to give recombination activity. This contrasts with reported split tyrosine recombinases, the components of which associate to reconstitute recombination activity without protein splicing[18-20]. When the intein-tagged integrase exteins Integrase-$\mathbf{E^N}$-$\mathbf{I^N}$ and $\mathbf{I^C}$-Integrase-$\mathbf{E^C}$ were co-expressed, reconstitution of integrase recombination activity was observed (lane 6). The DNA analysis shown in the lower panel of Figure 3C suggests that recombination by the trans-spliced integrase (ϕC31.Int$^{\alpha}$) has proceeded to over 80%.

Our results show that the changes from the wild-type integrase sequence that had to be introduced at the splice site are compatible with recombination activity. As many other serine integrases have similarly non-conserved, variable lengths of amino acid sequence in the region of the protein between β9 and αI (Figure 2A), we predict that these enzymes could also be engineered to create active split-intein systems.

***Trans*-splicing is required for integrase recombination activity**

The results shown in Figure 3C show that the *Npu* DnaE and *Ssp* DnaE intein moieties are required for reconstitution of recombination activity, but these experiments do not unambiguously establish the requirement for *trans*-splicing. Some split inteins are known to associate tightly via non-covalent interactions[34], and this can lead to reconstitution of the split protein activity, without splicing[35]. To test whether this applies in our split integrase system,

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

we mutated the nucleophilic cysteine residue and rate-enhancing flanking residues at the active sites of the two intein moieties[24] to render them catalytically inactive. We changed the junction residues "EY" in the Integrase-$\mathbf{E^N}$ moiety to "GA" and the Cys residue in the *Npu* DnaE$^{\mathbf{N}}$ moiety of Integrase-$\mathbf{E^N}$-$\mathbf{I^N}$ to "A" to give the mutated version Integrase-$\mathbf{E^N}$-$\mathbf{I^{N*}}$ (Figure 2C). Similarly, the residues "CFN" in the *Ssp* DnaE$^{\mathbf{C}}$ moiety of $\mathbf{I^C}$-Integrase-$\mathbf{E^C}$ were changed to "ASA", to derive the mutated version $\mathbf{I^{C*}}$-Integrase-$\mathbf{E^C}$ (Figure 2C). No recombination activity was observed when active Integrase-$\mathbf{E^N}$-$\mathbf{I^N}$ was co-expressed with inactive $\mathbf{I^{C*}}$-Integrase-$\mathbf{E^C}$ (Figure 3C, lane 7), nor when active $\mathbf{I^C}$-Integrase-$\mathbf{E^C}$ was co-expressed with inactive Integrase-$\mathbf{E^N}$-$\mathbf{I^{N*}}$ (Figure 3C, lane 8), showing that reconstitution of recombination activity requires the catalytic activities of both split intein fragments. In this aspect, serine integrases differ from integrases of the tyrosine recombinase family where non-covalent association is sufficient to allow reconstitution of activity from split protein fragments[18-20].

The *attP* x *attB* *in vivo* deletion reactions were slower in cultures where recombination required post-translational trans-splicing reactions to reconstitute functional integrase polypeptides, when compared to recombination by native integrase (see Supporting Information, Figure S1). It is likely that the rate of trans-splicing between *Npu* DnaE$^{\mathbf{N}}$ and *Ssp* DnaE$^{\mathbf{C}}$ is significantly slower *in vivo* in contrast to the fast *in vitro* splicing reaction rate[29].

The post-translational control system described in this report is tightly regulated because significant expression of both extein components will be necessary to generate enough spliced integrase for any recombination reaction to happen. This offers an advantage over regulation at the transcriptional level, where 'leaky' promoters often result in unreliable output signals. The faster kinetics of intein-catalysed protein trans-splicing would enable rapid control of gene expression, in contrast to the slower method of conditional gene expression by means of transcriptional control.

Split intein-regulated integrase-catalysed inversion system

To demonstrate the potential application of our split intein ϕC31 integrase in conditional expression of recombination activity, we designed an invertible genetic system based on an inversion susbstrate plasmid, pϕC31-invPB. Recombination between *attP* and *attB* sites in pϕC31-invPB inverts the orientation of a constitutive promoter sequence (Biobrick J23104), thereby switching expression from RFP to GFP (Figure 4A). The *E. coli* strain DS941/Z1/pϕC31-invPB was co-transformed with two vectors, each expressing one of the two split-intein integrase fragments. Expression of Integrase-$\mathbf{E^N}$-$\mathbf{I^N}$ was placed under the control of the *araC* promoter, and expression of $\mathbf{I^C}$-Integrase-$\mathbf{E^C}$ was placed under the control of the P*tet* promoter (Figure 4B). Co-expression of the split-intein fragments (and thus recombination) is dependent on the presence of both of the inducers arabinose and anhydrotetracycline (aTc) (Figure 4C). No recombination (GFP expression) was observed

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

unless both split-intein components were expressed (when aTC and arabinose were added to the growth medium; column IV.

The split-intein regulated serine integrase can be deployed as an effective tool for transient inversion of genetic regulatory modules (Figure 4). The ability to reconstitute integrase activity from inactive protein fragments using split inteins could allow precise control of site-specific recombination activity, either by using promoters that respond to different chemical signals or by specific activation of the intein splicing reaction. In this work we have demonstrated a split-integrase system that could be engineered to function as components of a logic gate. This could be achieved by means of light-sensitive protein domains that regulate intein splicing[36,37] or small molecule ligands that act directly on the split inteins[38] to control the protein-protein association step that precedes protein trans-splicing. Others have demonstrated the use of pH changes[39] and temperature[40] as tools for regulating intein functions. These technologies can enhance capacity for building orthogonal logic gate components for conditional gene expression regulation and genome engineering applications, and add to the existing tools for programmable cellular functions[41,42].

Intein-mediated assembly of integrase-RDF fusion recombinase and catalysis of *attR* x *attL* recombination

We recently showed that recombination between *attR* x *attL* sites can be catalysed efficiently by artificial proteins in which the recombination directionality factor (RDF) is fused to the integrase using a short peptide linker. The fusion recombinase ϕC31.integrase-gp3 catalysed *attR* x *attL* recombination efficiently[30] (Figure 5A). This system allows more predictable regulation of integrase-catalysed conversion of *attP*/*attB* to *attR*/*attL* and vice versa, with potential applications in building genetic switches and recombinase-based counters.

To demonstrate the use of our split-intein regulated system in *attR* x *attL* recombination, we split ϕC31.integrase-gp3 between the β9 and αI domains of the integrase (Figure 5B), at the same position as described above for ϕC31 integrase itself. To test for recombination activity, we used a test substrate (pϕC31-delRL) similar to the *attP* x *attB* substrate (pϕC31-delPB) but with the *attP* and *attB* sites replaced by *attR* and *attL* respectively (Supporting Information, Table S1). The sites are arranged in the head-to-tail orientation such that deletion of the *galK* gene leads to a reaction product that is smaller in size than the starting substrate (Figure 5C). As expected, neither integrase (Figure 5D, lane 2) nor co-expression of the Integrase-$\mathbf{E^N}$-$\mathbf{I^N}$ and $\mathbf{I^C}$-Integrase-$\mathbf{E^C}$ (lane 4) were able to catalyse *attR* x *attL* recombination. Co-expression of the intein-tagged integrase N-extein (Integrase-$\mathbf{E^N}$-$\mathbf{I^N}$) and C-extein-gp3 ($\mathbf{I^C}$-Integrase-$\mathbf{E^C}$-gp3) results in efficient *attR* x *attL* recombination (lane 5). The activity of the trans-spliced fusion recombinase (ϕC31.Int^{α}-gp3) was comparable to that observed from native ϕC31.integrase-gp3 (lane 3).

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

These experiments demonstrate that our split-intein regulated recombination system can be used to toggle between two states in which the forward reaction is catalyzed by the serine integrase, and the reverse reaction by the integrase-RDF fusion. In a practical application, reconstitution of the integrase and the integrase-RDF fusion could be regulated by two pairs of orthogonal split inteins. Several split inteins have been described in the literature[43,44]. Further characterization of their properties would make them available as orthogonal functional parts for regulation of serine integrase-based applications. More recently, mutations at residues within the GEP loop of the *Npu* DnaE split intein have been identified that affect the specificity, reaction rates, and tolerance to changes in the flanking residues[16,17]. Use of these intein variants with broader specificities would require minimum changes to the extein sequence of the target integrase, thereby reducing the potential risk of introducing deleterious mutations. Such promiscuous split inteins that have more flexible requirements for flanking residues should make it easier to find new split positions in a wider range of target proteins.

Conclusion

Serine integrases promote efficient directional DNA site-specific recombination, and thus have major potential applications in genome engineering and metabolic pathway engineering[9,11,45-48]. They have also been incorporated into the design of cellular state machines and biocomputing devices[4,8,11]. Here we have demonstrated a novel method for regulation of serine integrase activity by making activity dependent on post-translational *trans*-splicing of two integrase extein components using split-intein technology. Since *in vitro* protein splicing by some split inteins has very fast kinetics[24,44], this technology can be developed for the expression and purification of serine integrases that are toxic to the expression host[12]. This could significantly increase the range of integrase specificities available for *in vitro* applications such as genome engineering and DNA fragment assembly.

MATERIALS AND METHODS

Plasmids and DNA

The DNA sequences of all protein constructs and plasmid substrates used in this study are shown in Supporting Information (Table S2). The codon-optimized ϕC31 integrase sequence was derived from pFM141[30]. Codon-optimized sequences of NpuN, a 102-amino acid residue split intein from *Nostoc punctiforme* DnaE, and SspC, a 36-amino acid residue split intein from *Synechocystis sp*. DnaE, were from GeneArt (Invitrogen). Plasmids for constitutive expression of the N-terminal extein-intein fusion protein (Integrase-Ext^{N}-I^{N}) in *E. coli* were made by inserting protein-coding DNA sequences between NdeI and Acc65I sites

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

in pMS140, a low-level expression vector with a pMB1 origin of replication[30]. Plasmids for constitutive expression of the C-terminal intein-extein fusion protein (**I^{C}**-Integrase-**E^{C}**) were made in a similar way by inserting the coding sequence between NdeI and Acc65I sites in pEK76, which has a p15a (pACYC184) origin[30,49]. Plasmids for *tet*-inducible expression of the split integrase components were made by cloning the protein-coding DNA sequences between SacI and SalI sites in pTet[50] while those for *ara*-inducible expression were made by cloning the protein-coding sequences between SacI and SalI sites in pBAD[51].

The plasmid substrate for assessing recombination (deletion) by native ϕC31 integrase or *trans*-spliced ϕC31 integrase (pϕC31-delPB) was described in Olorunniji *et al.*[30]. The test substrate contains a *galK* gene flanked by *attP* and *attB* sites arranged in direct repeat (head-to-tail orientation) resulting in deletion of the *galK* gene upon integrase-catalysed recombination. The plasmid substrate for assessing recombination (inversion) activity (pϕC31-invPB) was made by cloning the invertible promoter device shown in Figure 3A in a pSC101 origin plasmid with kanamycin resistance selection.

Full sequences of all plasmids are shown in Supporting Information (Table S2).

Recombination analysis

Assays for integrase-mediated recombination were carried out in *E. coli* strain DS941[30]. DS941 was first transformed with the test substrate plasmid (pϕC31-delPB or pϕC31-invPB). The substrate plasmid-containing strains were then either transformed with an integrase-expressing plasmid (pFM141), or co-transformed with two plasmids, each expressing part of the split integrase. The transformants were grown on selective MacConkey-galactose indicator plates (MacConkey agar base (Difco) supplemented with 1% galactose); kanamycin (50 μg/ml) was included to select for the substrate plasmid and the recombination product, while ampicillin (50 μg/ml) and/or chloramphenicol (25 μg/ml) were included to select for the integrase expression plasmids. Integrase-mediated deletion of the *galK* gene results in pale-coloured (*galK*$^{-}$) colonies on the indicator plates, whereas red (*galK*$^{+}$) colonies suggest lack of recombination. In some cases, colonies in which low-level recombination has occurred maintain a red colour. To determine the extent of recombination more accurately, plasmid DNA recovered from the cells was examined. L-broth (1 ml) was added to the surface of each plate and the cells were suspended using a plate spreader. An aliquot of this suspension was used to inoculate L-broth (1 in 1000 dilution), and the culture was incubated overnight at 37 °C with kanamycin selection for the substrate and recombinant plasmids. Plasmid DNA was prepared using a Qiagen miniprep kit, and analysed by 1.2% agarose gel electrophoresis.

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Inducible expression of integrase-intein fusions for *trans*-splicing

In vivo expression of the integrase-intein fusion proteins under the control of arabinose and tet inducible promoters was carried out in strain DS941/Z1 (S.D. Colloms, unpublished), which consitutively expresses TetR, required for regulation of the pTet promoter. The DS941/Z1 strain was made competent by a standard calcium chloride method[52]. The cells were transformed with the plasmid substrate pϕC31-invPB, then cultured for 90 minutes and selected on L-agar plates containing kanamycin (50 µg/ml). A single colony was picked and grown in kanamycin-containing L-broth (5 ml) to make a stationary phase overnight culture. An aliquot of the stationary phase culture was then diluted into L-broth containing kanamycin and grown to mid-log phase. These cells were made 'chemically competent' (as above) and transformed with the two vectors containing the coding sequences of the intein-integrase fragments. The plasmid vector pFEM148 (ampicillin selection) expresses Integrase-$\mathbf{E^N}$-$\mathbf{I^N}$ under the control of the pTet promoter. The second plasmid pFEM149 (chloramphenicol selection) expresses $\mathbf{I^C}$-Integrase-$\mathbf{E^C}$ under the control of the pBAD promoter.

The transformant cells were cultured for 90 minutes, and selected on L-agar plates containing kanamycin (50 µg/ml), ampicillin (100 µg/ml), and chloramphenicol (25 µg/ml). To carry out the recombination assays, a culture from a single colony was grown overnight in L-broth in the presence of the three antibiotics, to stationary phase. The culture was diluted further (1:100) and grown to mid-log phase (about 90 minutes), after which expression of the split-intein-integrase fragments was induced by the addition of anhydrotetracycline, aTc (0.1 µg/ml), arabinose, Ara (0.2% w/v), or both, as shown in Figure 3C. Glucose was added to 0.4% concentration to the cultures where it is required to turn off expression of the arabinose promoter. The induced cultures were grown for 24 hours at 37 °C, after which the cultures were left at room temperature for 8 hours. Next, aliquots of each culture (50 µl) were diluted with 950 µl phage buffer (10 mM Tris, pH 7.5, 10 mM $MgCl_2$, 68 mM NaCl), and GFP expression was measured by means of fluorescence.

Fluorescence measurements

Fluorescence measurements were carried out on a Typhoon FLA 9500 fluorimager (GE Healthcare). Aliquots of the diluted cultures (200 µl) were added to a 96-well plate, and the fluorescence of the expressed proteins were measured (GFP: excitation, 485 nm; emission, 520 nm and RFP: excitation, 532 nm; emission, 575 nm). To determine cell density for each sample, 50 µl aliquots were diluted to 1000 µl, and the spectrophotometric absorbance was read at 600 nm. The GFP-independent background signals of the cells were determined by measuring the fluorescence of strain DS941/Z1 strain (containing the test substrate but without the split-intein-integrase expression vectors). The background fluorescence was subtracted from the values measured for samples in the different treatment groups, after normalization using the cell density measurements.

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

REFERENCES

(1) Anderson, J.C., Voigt, C.A., and Arkin, A.P. (2007) Environmental signal integration by a modular AND gate. *Mol. Syst. Biol. 3*, 133.

(2) Moon, T.S., Lou, C., Tamsir, A., Stanton, B.C., and Voigt, C.A. (2012) Genetic programs constructed from layered logic gates in single cells. *Nature 491*, 249–253.

(3) Glasscock, C.J., Lucks, J.B., and DeLisa, M.P. (2016) Engineered Protein Machines: Emergent Tools for Synthetic Biology. *Cell Chemical Biol. 23*, 45-56.

(4) Bonnet, J., Subsoontorn, P., and Endy, D. (2012) Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci. USA 109*, 8884-8889.

(5) Bonnet, J., Yin, P., Ortiz, M.E., Subsoontorn, P. and Endy, D. (2013) Amplifying genetic logic gates. *Science 340,* 599–603.

(6) Brophy, J.A.N. and Voigt, C.A. (2014) Principles of genetic circuit design. *Nat. Methods 11*, 508-520.

(7) Wang, B., Kitney, R.I., Joly, N. and Buck, M. (2011) Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nat. Commun. 2*, 508.

(8) Yang, L., Nielsen, A.A.K., Fernandez-Rodriguez, J., McClune, C.J., Laub, M.T., Lu, T.K. and Voigt, C.A. (2014) Permanent genetic memory with >1-byte capacity. *Nat. Methods 11*, 1261-1266.

(9) Colloms, S.D., Merrick, C.A., Olorunniji, F.J., Stark, W.M., Smith, M.C.M., Osbourn, A., Keasling, J.D., and Rosser,S.J. (2014) Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Res. 42*, e23.

(10) Olorunniji, F.J., Merrick, C., Rosser, S.J., Smith, M.C.M., Stark, W.M., and Colloms, S.D. (2017a) Multipart DNA assembly using site-specific recombinases from the large serine integrase family. Nikolai Eroshenko (ed.), *Site-specific Recombinases*, Methods in Molecular Biology, Springer Science & Business Media New York, pp. 303-324.

(11) Stark, W.M. (2017) Making serine integrases work for us. *Curr. Opin. Microbiol. 38*, 130-136.

(12) Shah, N.H. and Muir, T.W. (2014) Inteins: nature's gift to protein chemists. *Chem. Sci. 5*, 446-441.

(13) Li, Y. (2015) Split-inteins and their bioapplications. *Biotechnol. Lett. 37*, 2121–2137.

(14) Siu, K-H. and Chen, W. (2017) Control of the Yeast Mating Pathway by Reconstitution of Functional α-Factor Using Split Intein-Catalyzed Reactions. *ACS Syn. Biol. 6*, 1453-1460.

(15) Schaerli, Y., Gili, M., Isalan, M. (2014) A split intein T7 RNA polymerase for transcriptional AND-logic. *Nucleic Acids Res.* 42, 12322-12328.

(16) Stevens, A.J., Brown, Z.Z., Shah, N.H., Sekar, G., Cowburn, D., and Muir, T.W. (2016) Design of a split intein with exceptional protein splicing activity. *J. Am. Chem. Soc. 138*, 2162-2165.

(17) Stevens, A.J., Sekar, G., Shah, N.H., Mostafavi, A.Z., Cowburn, D., and Muir, T.W. (2017) A promiscuous split intein with expanded protein engineering applications. *Proc Natl Acad Sci U S A. 114*, 8538-8543.

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

(18) Gu, W., Schneider, J.W., Condorelli, G., Kaushal, S., Mahdavi, V., and Nadal-Ginard, B. (1993) Interaction of myogenic factors and the retinoblastoma protein mediates muscle-cell commitment and differentiation. *Cell 72*, 309–324.

(19) Kellendonk, C., Tronche, F., Monaghan, A.P., Angrand, P.O., Stewart, F., and Schütz, G. (1996) Regulation of Cre recombinase activity by the synthetic steroid RU 486. *Nucleic Acids Res. 2*4, 1404–1411.

(20) Han, X.Z., Han, F.Y., Ren, X.S., Si, J., Li, C.Q., and Song, H.Y. (2013) *Ssp* DnaE split-intein mediated split-Cre reconstitution in tobacco. *Plant Cell Tissue Organ Cult 113*, 529–542.

(21) Rajaee, M. and Ow, D.W. (2017) A new location to split Cre recombinase for protein fragment complementation. *Plant Biotechnology J. 15*, 1420-1428.

(22) Wang, P., Chen, T., Sakurai, K., Han, B.X., He, Z., Feng, G., and Wang, F. (2012) Intersectional Cre driver lines generated using split-intein mediated split-Cre reconstitution. *Sci. Rep. 2*, 497.

(23) Lucet, I.S., Tynan, F.E., Adams, V., Rossjohn, J., Lyras, D., and Rood, J.I. (2005) Identification of the structural and functional domains of the large serine recombinase TnpX from Clostridium perfringens*. J. Biol. Chem. 280*, 2503-2511.

(24) Lockless, S.W. and Muir, T.W. (2009) Traceless protein splicing utilizing evolved split inteins. *Proc. Natl. Acad. Sci. U.S.A. 106*, 10999–11004.

(25) Amitai, G., Callahan, B.P., Stanger, M.J., Belfort, G., and Belfort, M. (2009) Modulation of intein activity by its neighboring extein substrates. *Proc Natl Acad Sci U S A. 106*, 11005-11010.

(26) Volkmann, G. and Mootz, H.D. (2013) Recent progress in intein research: from mechanism to directed evolution and applications. *Cell Mol Life Sci. 70*, 1185-206.

(27) Van Duyne, G.D. and Rutherford, K. (2013) Large serine recombinase domain structure and attachment site binding. *Crit. Rev. Biochem. Mol. Biol. 48*, 476-491.

(28) Rutherford, K., Yuan, P., Perry, K., Sharp, R., and Van Duyne, G.D. (2013) Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Res. 41*, 8341-8356.

(29) Iwai, H., Zuger, S., Jin, J., and Tam, P.H. (2006) Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. *FEBS Lett. 580*, 1853–1858.

(30) Olorunniji, F.J., McPherson, A.L., Rosser, S.J., Smith, M.C.M., Colloms, S.D., Stark, W.M. (2017b) Control of serine integrase recombination directionality by fusion with the directionality factor. *Nucleic Acids Res. 45*, 8635-8645.

(31) Olorunniji, F.J., He, J., Wenwieser, S.V.C.T., Boocock, M.R., and Stark, W.M. (2008) Synapsis and catalysis by activated Tn*3* resolvase mutants. *Nucleic Acids Res*. *36*, 7181-7191.

(32) Arnold, P.H., Blake, D.G., Grindley, N.D.F., Boocock, M.R., and Stark, W.M. (1999) Mutants of Tn3 resolvase which do not require accessory binding sites for recombination activity. *EMBO J. 18*, 1407–1414.

(33) Burke, M.E., Arnold, P.H., He, J., Wenwieser, S.V.C.T., Rowland, S.J., Boocock, M.R., and Stark, W.M. (2004) Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation. *Mol. Microbiol. 51*, 937–948.

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

(34) Schwartz, E.C., Saez, L., Young, M.W., and Muir, T.W. (2007) Post-translational enzyme activation in an animal via optimized conditional protein splicing. *Nat. Chem. Biol. 3*, 50-54.

(35) Shi, J. and Muir, T.W. (2005) Development of a tandem protein trans-splicing system based on native and engineered split inteins. *J. Am. Chem. Soc. 127*, 6198-6206.

(36) Han, T. Chen, Q., and Haiyan Liu, H. (2017) Engineered Photoactivatable Genetic Switches Based on the Bacterium Phage T7 RNA Polymerase. *ACS Synth. Biol. 6*, 357−366

(37) Jones, D.C., Mistry, I.N., and Tavassoli, A. (2016) Post-translational control of protein function with light using LOV-intein fusion protein. *Mol. BioSyst. 12*, 1388-1393.

(38) Buskirk, A.R., Ong, Y.C., Gartner, Z.J., and Liu, D.R. (2004) Directed evolution of ligand dependence: Small-molecule-activated protein splicing. *Proc. Natl. Acad. Sci. U.S.A. 101*, 10505-10510.

(39) Wood, D.W., Wu, W., Belfort, G., Derbyshire, V., and Belfort, M. (1999) A genetic system yields self-cleaving inteins for bioseparations. *Nat. Biotechnol. 17*, 889-892.

(40) Zeidler, M.P., Tan, C., Bellaiche, Y., Cherry, S., Häder, S., Gayko, U., Perrimon, N. (2004) Temperature-sensitive control of protein activity by conditionally splicing inteins. *Nat. Biotechnol. 22*, 871-876.

(41) Maranhao, A.C. and Ellington, A.D. (2013) Endowing cells with logic and memory. *Nat. Biotechnol. 31*, 413-415.

(42) Roquet, N., Soleimany, A.P., Ferris, A.C., Aaronson, S., and Lu, T.K. (2016) Synthetic recombinase-based state machines in living cells. *Science 353*, aad8559.

(43) Dassa, B., London, N., Stoddard, B.L., Schueler-Furman, O., and Pietrokovoski, S. (2009) Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. *Nucleic Acids Res. 45*, 2560-2573.

(44) Carvajal-Vallejos, P., Pallissé, R., Mootz. R.D., and Schmidt, S.R. (2012) Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources. *J. Biol. Chem.* 287, 28686–28696.

(45) Brown,W.R.A., Lee,N.C.O., Xu,Z. and Smith,M.C.M. (2011) Serine recombinases as tools for genome engineering. *Methods 53*, 372-379.

(46) Fogg, P.C.M., Colloms, S.D., Rosser, S.J., Stark, W.M., and Smith, M.C.M. (2014) New applications for phage integrases. *J. Mol. Biol. 426*, 2703-2716.

(47) Smith, M.C.M. (2015) Phage-encoded serine integrases and other large serine recombinases. In Mobile DNA III (ed. Craig, N.L., Chandler, M., Gellert, M., Lambowitz, A., Rice, P.A. and Sandmeyer, S.B., ASM Press, Washington DC, USA), pp. 253-272.

(48) Olorunniji, F.J., Rosser, S.J. and Stark, W.M. (2016) Site-specific recombinases: molecular machines for the genetic revolution. *Biochem. J. 473*, 673-684.

(49) Proudfoot, C.M., McPherson, A.L., Kolb, A.F., and Stark, W.M. (2011) Zinc finger recombinases with adaptable DNA sequence specificity. *PLoS ONE 6*, e19537.

(50) Lutz, R., and Bujard, H. (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. N*ucleic Acids Res. 25*, 1203−1210.

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

(51) Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol. 177*, 4121-4130.

(52) Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd ed. Plainview, N.Y.: Cold Spring Harbor Laboratory Press.

SUPPORTING INFORMATION:

Table S1: Sequences of ϕC31 integrase *att* sites.

Figure S1: Comparison of *in vivo* recombination activity of native ϕC31 integrase with the reconstituted trans-spliced ϕC31 integrase.

ACKNOWLEDGEMENTS

We thank Dr Sean Colloms and Dr Steve Kane for the gift of the plasmid template (pSWITCH) used in constructing the invertible promoter assay substrate.

FUNDING

This work was supported by the Biotechnology and Biological Sciences Research Council [grant numbers BB/003356/1, BB/M018040/1 and BB/M018229/1], and the Medical Research Council Proximity [grant number MC_PC_16077].

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

FIGURE 1

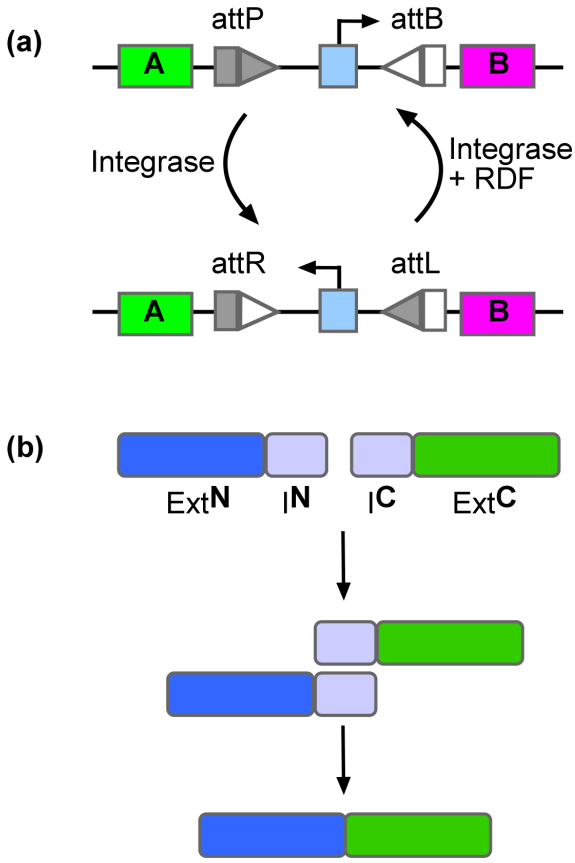


Figure 1: Building blocks for split intein-mediated trans-splicing of ϕC31 integrase. (**A**) Serine integrase-catalysed inversion of the orientation of a promoter sequence functions as a toggle switch for the expression of two genes. Integrase catalysed recombination between *attP* and *attB* sites (arranged in inverted orientation) results in switching the direction of the promoter, and hence changes the gene expression status of the module. Recombination of *attR* and *attL* sites by integrase and the recombination directionality factor (RDF), restores the module to its original state. (**B**) Split-intein catalyzed protein trans-splicing. Non-covalent association of the N-terminal intein ($\mathbf{I^N}$) and the C-terminal intein ($\mathbf{I^C}$) is followed by a trans-splicing reaction that covalently joins the N-terminal extein ($\mathbf{E^N}$) and C-terminal extein ($\mathbf{E^C}$).

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

FIGURE 2

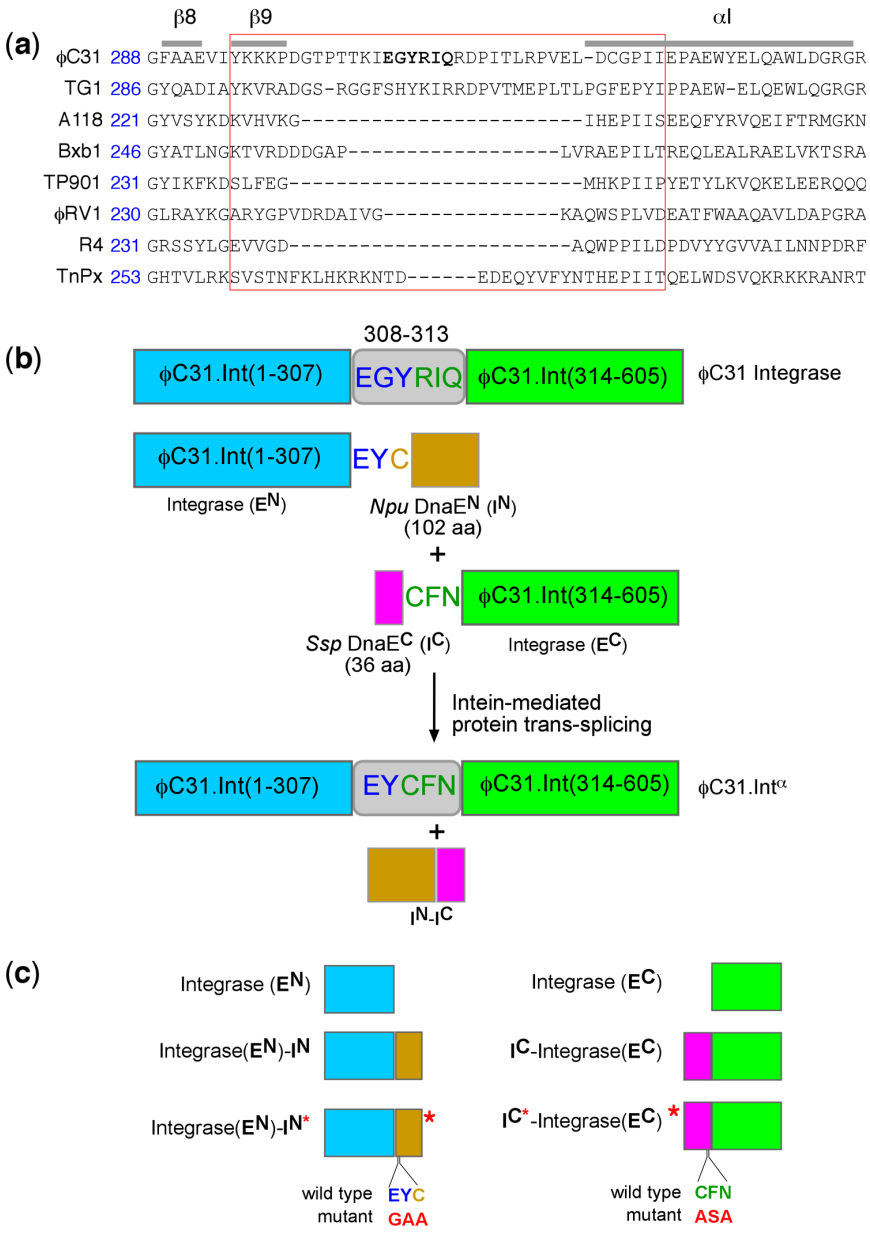


Figure 2: Design of split *ϕ*C31 integrase. (**A**) Sequence alignment of *ϕ*C31 integrase with related members of the 'large serine recombinase' protein family based on published structures and similar alignments[27]. Amino acid sequences in the loop region where *ϕ*C31 integrase is split are highlighted with a red box. (**B**) Structures of the split integrase-intein constructs. Changes made to the integrase sequence at the junctions where the fragments are fused to the split inteins are shown. The functional domains are shown as rectangular boxes: Integrase-$\mathbf{E^N}$ (blue), *Npu* DnaE^N (orange), *Ssp* DnaE^C (pink), and Integrase-$\mathbf{E^C}$ (green). Intein-catalysed protein *trans*-splicing generates an active version of the integrase in which residues at positions 308-312 of wild-type integrase (EGYRIQ) are replaced with EYCFN. The *trans*-spliced integrase is thus one amino acid shorter than the native integrase. (**C**) Variants of the split *ϕ*C31 integrase designed to probe the requirements for reconstitution of integrase activity. (i) Integrase-$\mathbf{E^N}$-$\mathbf{I^N}$; Integrase-$\mathbf{E^N}$ (residues 1-307) fused to the 102-amino acid residue *Npu* DnaE^N ($\mathbf{I^N}$) using the two-residue EY and the N-terminal Cys residue from *Npu* DnaE^N. (ii) Integrase-$\mathbf{E^N}$; *ϕ*C31 integrase (1-307). (iii) Integrase-$\mathbf{E^N}$-$\mathbf{I^N}$*; *ϕ*C31 integrase (1-307) fused to *Npu* DnaE^N as described in (i) but with the EY residue at the splice site changed to GA to inactivate splicing activity. (iv) $\mathbf{I^C}$-Integrase-$\mathbf{E^C}$; *ϕ*C31 integrase C-extein (residues 314-605) fused directly to the *Ssp* DnaE^C ($\mathbf{I^C}$). *Ssp* DnaE^C nucleophilic cysteine residue and the flanking residues required for splicing are highlighted (CFN). (v) Integrase-$\mathbf{E^C}$; *ϕ*C31 integrase C-extein (residues 314-605). (vi) $\mathbf{I^{C*}}$-Integrase-$\mathbf{E^C}$; *ϕ*C31 integrase C-extein fused to *Ssp* DnaE^C as described in (i) but with the CFN residue at the splice site changed to ASA to inactivate splicing activity.

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

FIGURE 3

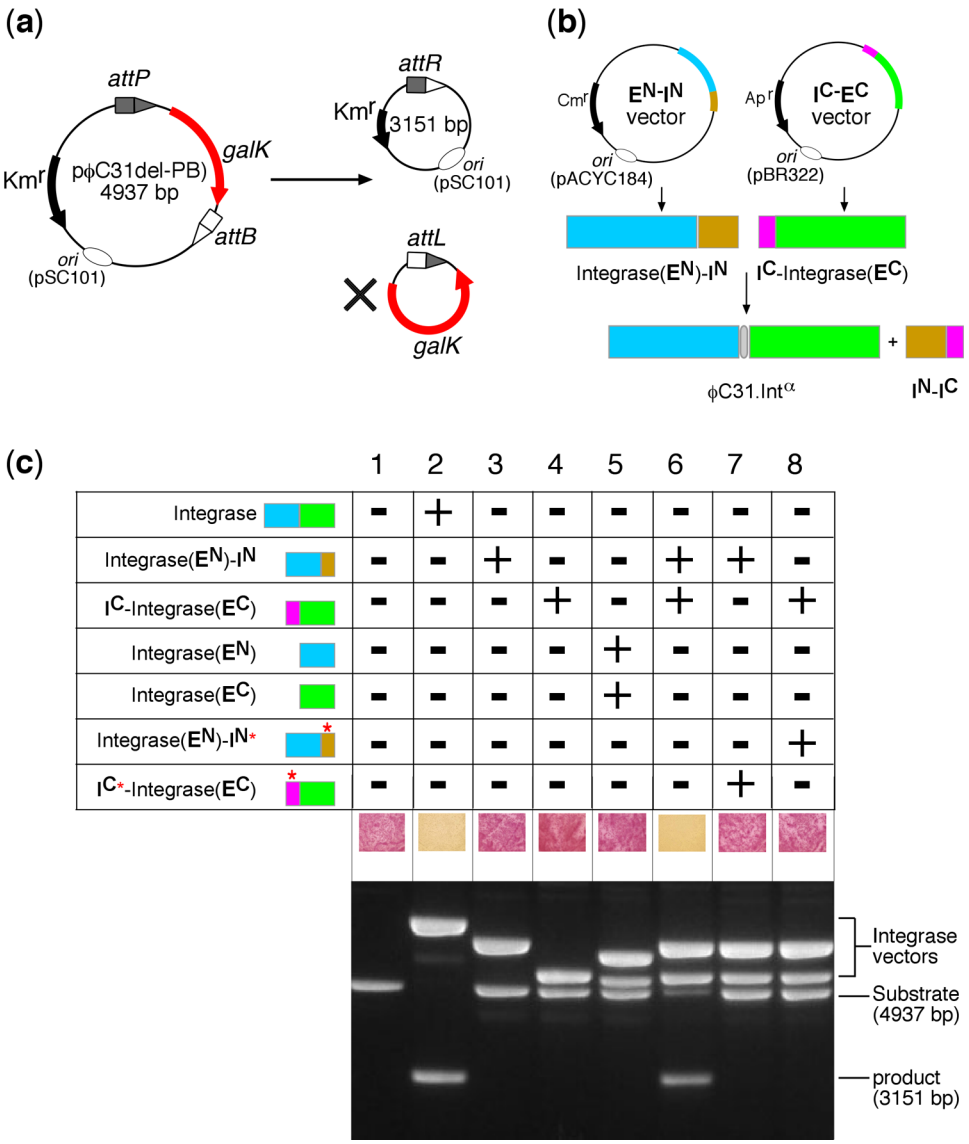


Figure 3: *In vivo* recombination activity of reconstituted split-intein ϕC31 integrase. (**A**) *In vivo* recombination assay using a deletion substrate (pϕC31-delPB) carrying the *galK* gene[30]. The recombination product circle carrying the *galK* DNA has no origin of replication, and is lost during subsequent cell divisions. Inability of the cells to metabolize galactose leads to a change in colony colour on MacConkey galactose medium. (**B**) Constitutive expression of precursor polypeptides. The open reading frame of Integrase-E^N-I^N was expressed from a pACYC184-based vector, and I^C-Integrase-E^C was expressed from a pBR322-based vector[30]. Expression of the two polypeptides followed by intein-catalysed protein *trans*-splicing reconstitutes ϕC31.Int^{α}. (**C**) Assay of recombination activities (*galK* colour assay) of integrase constructs on *attP* × *attB* deletion substrates (pϕC31-delPB). DNA analysis of recombination activities of integrase constructs on *attP* × *attB* deletion substrates (pϕC31-delPB). In these assays, cells containing the substrate plasmid (pϕC31-delPB) were transformed with the expression vectors indicated and grown for 20 hours in selection media. For analysis of *in vivo* recombination products, plasmid DNA was recovered from cells[30] and separated by means of 1.2% agarose gel electrophoresis. (1) Substrate only (blank control); (2) FEM141: ϕC31 Integrase; (3) FEM136: Integrase-E^N-I^N; (4) FEM137: I^C-Integrase-E^C; (5) FEM155 (Integrase-E^N) and FEM157 (Integrase-E^C); (6) FEM136 (Integrase-E^N-I^N) and FEM137 (I^C-Integrase-E^C); (7) FEM136 (Integrase-E^N-I^N) and FEM162 (I^{C*}-Integrase-E^C); (8) FEM137 (I^C-Integrase-E^C) and FEM161 (Integrase-E^N-I^{N*}).

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

FIGURE 4

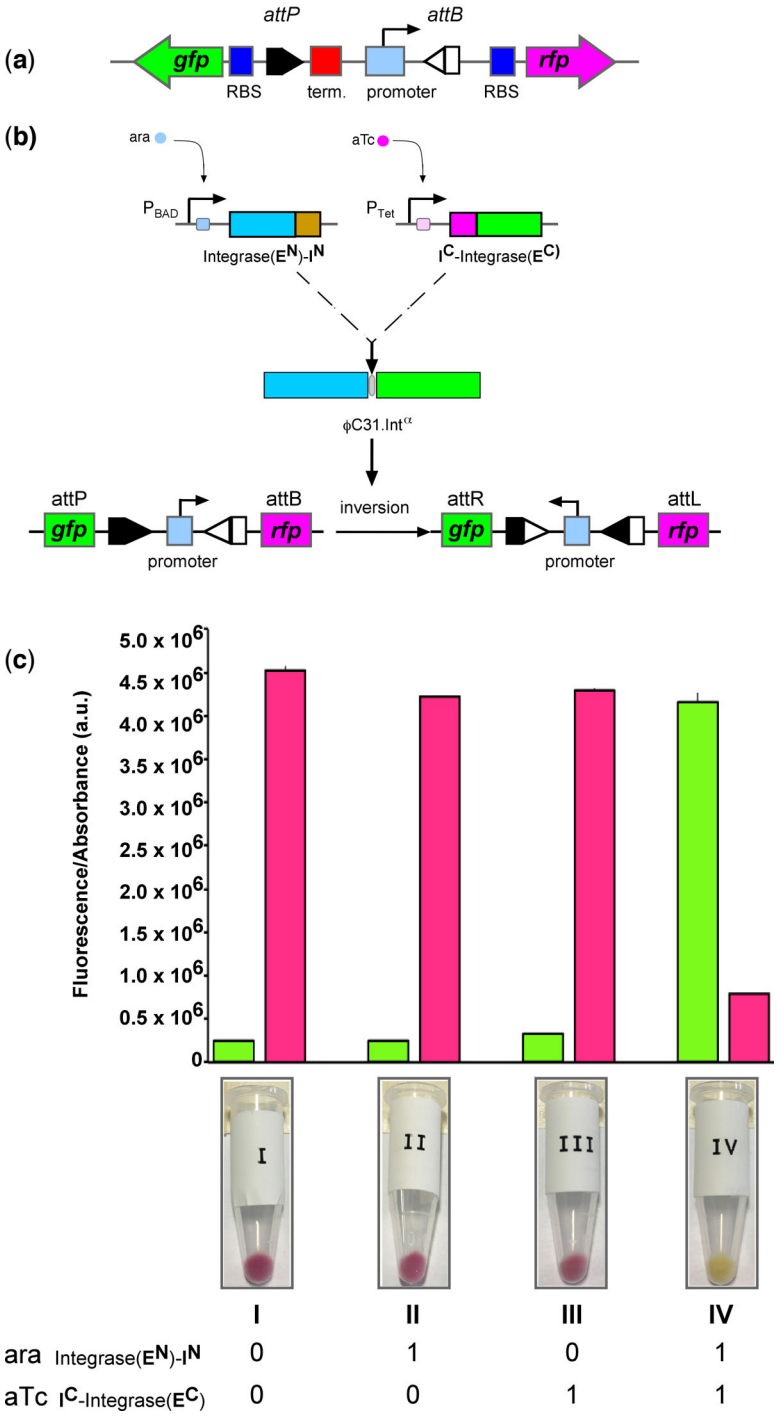


Figure 4: Controlling the function of an invertible promoter system using split intein-regulated ϕC31 integrase activity. (**A**) Design of a recombinase-operated switch using an invertible promoter reporter system. The constitutive promoter sequence (cyan rectangle; Biobrick J23104) is flanked by ϕC31 integrase *attP* and *attB* sites (grey arrows) that are arranged in inverted orientation. In its default state, the promoter constitutively drives the expression of a red fluorescent protein (*rfp*) gene (pink arrow). The T1 terminator sequence (red squares) located immediately upstream of the promoter prevents transcriptional read-through to the green fluorescent protein (*gfp*) gene (green arrow). Upon integrase-catalysed site-specific recombination, the orientation of the promoter is reversed to allow the expression of GFP, and block RFP production. A Biobrick ribosomal binding site, RBS, (RBS_B0034, blue rectangles), is positioned 5' of the *rfp* and *gfp* genes to drive optimal translation of the synthesized mRNAs. (**B**) Conditional expression of ϕC31 integrase fragments and split intein-mediated reconstitution of activity. The two split-intein integrase fragments are expressed from two inducible promoter vectors. Arabinose (ara) induces expression of Integrase-$\mathbf{E^N}$-$\mathbf{I^N}$ under the control of the pBAD

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

promoter (pFEM149), while anhydrotetracycline (aTC) induces expression of **I**^{C}-Integrase-**E**^{C} under the control of the pTet promoter (pFEM148). Post-translational *trans*-splicing of the split-intein integrase fragments generates the functional reconstituted integrase (ϕC31.Int$^{\alpha}$). Catalysis of *attP* x *attB* inversion by ϕC31.Int$^{\alpha}$ results in reversal of the orientation of the promoter. (**C**) Validation of the split-intein ϕC31 integrase recombinational AND-gate. Fluorescence measurements indicating the expression of GFP and RFP in cells induced with aTc and/or ara. Cells were grown at 37 °C for 24 hours in the presence of aTC (0.1 µg/ml) and ara (0.2%). Glucose was added to 0.4% concentration to treatments A and C to turn off expression of the arabinose promoter. Each bar represents mean and standard deviation of four determinations. Below the bar chart are images of Eppendorf tubes showing the pellet obtained after centrifugation of 5 ml of bacterial culture.

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

FIGURE 5

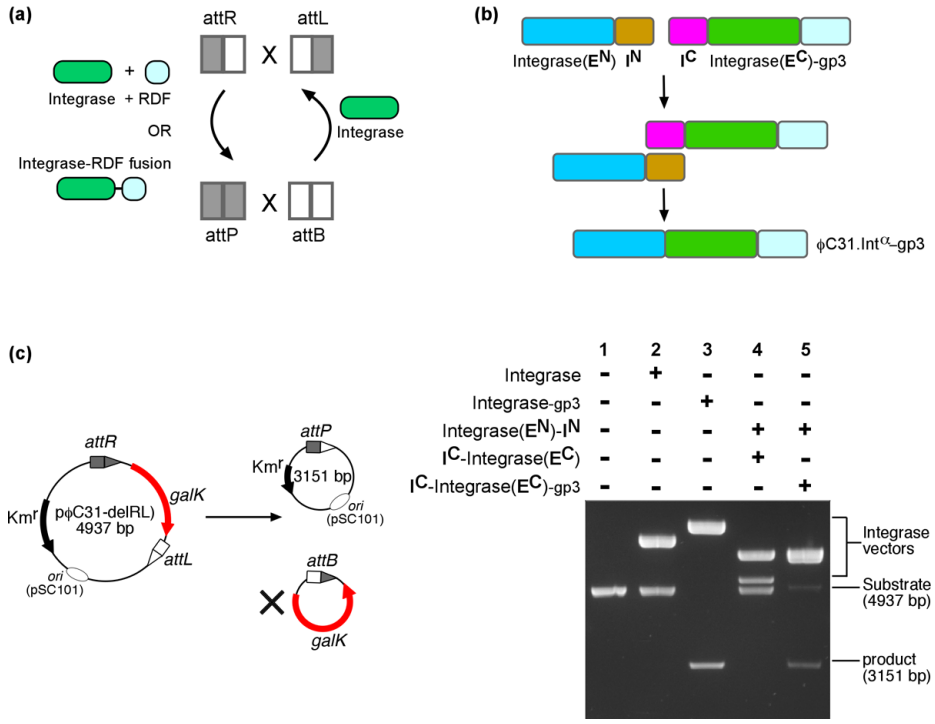


Figure 5: Intein-mediated assembly of ϕC31 integrase-RDF fusion recombinase and catalysis of *attR* x *attL* recombination. (**A**) Catalysis of *attR* x *attL* recombination by integrase + RDF, or integrase-RDF fusion protein (see text for details). The RDF (Recombination Directionality Factor) for ϕC31 integrase is gp3. (**B**) Split-intein catalyzed assembly of integrase-RDF fusion recombinase. Non-covalent association of Integrase-$\mathbf{E^N}$-$\mathbf{I^N}$ and I^C-Integrase-$\mathbf{E^C}$-gp3 via the intein domains ($\mathbf{I^N}$ and $\mathbf{I^C}$) is followed by a trans-splicing reaction that gives the fusion recombinase ϕC31.Int^{α}-gp3. (**C**) Assay of recombination activities (*galK* colour assay) of integrase constructs on *attR* × *attL* deletion substrates (pϕC31-delRL). The test substrate is identical to pϕC31-delPB (Figure 3A), except that the *attP* and *attB* sites are replaced by *attR* and *attL* respectively. The *in vivo* recombination assay and analysis are as described in Figure 4C. (1) Substrate only (blank control); (2) FEM141: ϕC31.Integrase; (3) FEM33: ϕC31.Integrase-gp3 fusion; (4) FEM136 (Integrase-$\mathbf{E^N}$-$\mathbf{I^N}$) and FEM137 ($\mathbf{I^C}$-Integrase-$\mathbf{E^C}$); (5) FEM136 (Integrase-$\mathbf{E^N}$-$\mathbf{I^N}$) and FEM188 ($\mathbf{I^C}$-Integrase-$\mathbf{E^C}$-gp3).

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

SUPPORTING INFORMATION

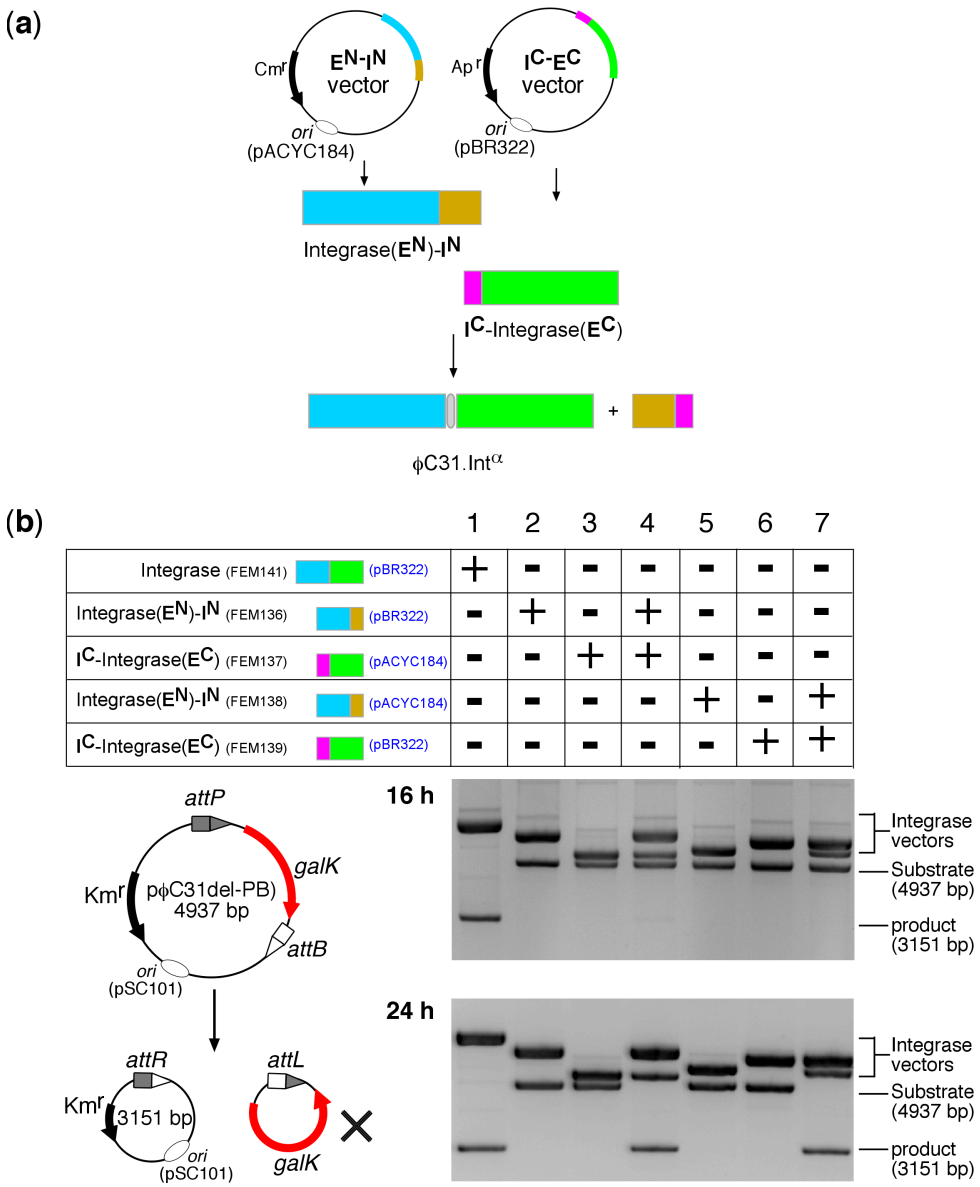


Figure S1:

Comparison of *in vivo* recombination activity of native ϕC31 integrase with reconstituted trans-spliced ϕC31 integrase. (A) The precursor polypeptides for the split-intein integrase were consitutively expressed from vectors with either a pACYC184 or pBR322 origin of replication (Proudfoot *et al.*, 2011). For each precursor, both types of vector were tested, to confirm that the expression vector did not have any serious effect on the outcome of the experiment (see part B). In the example shown, the $\mathbf{E^N}$ –$\mathbf{I^N}$ vector is expressed in pACY184, and the $\mathbf{I^C}$-$\mathbf{E^C}$ construct placed in pBR322 vector. (**B**) Assay of recombination activities (*galK* deletion assay) of integrase constructs on an *attP* × *attB* substrate (pϕC31-delPB). The *in vivo* recombination assay and analysis are as described in Figure 4C. Cells containing pϕC31-delPB were transformed with the expression vectors indicated and grown for 16 hours (top panel, **16 h**) or for 24 hours (bottom panel, **24 h**) on selective plates. For analysis of *in vivo* recombination products, plasmid DNA was recovered from cells (Olorunniji *et al.*, 2017) and separated by means of 1.2% agarose gel electrophoresis.

bioRxiv preprint doi: https://doi.org/10.1101/540872; this version posted February 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Table S1

Sequences of ΦC31 integrase att sites. The sequences are shown in the head-to-tail orientation that results in deletion of the intervening sequences upon recombination. The central 2-bp overlap sequences of the att sites are highlighted in bold text.

ΦC31 (*attP*)
AGTAGTGCCCCAACTGGGGTAACCT**TT**GAGTTCTCTCAGTTGGGGGCGTA
TCATCACGGGGTTGACCCCATTGGA**AA**CTCAAGAGAGTCAACCCCCGCAT

ΦC31 (*attB*)
GCGGTGCGGGTGCCAGGGCGTGCCC**TT**GGGCTCCCCGGGCGCGTACTCC
CGCCACGCCCACGGTCCCGCACGGG**AA**CCCGAGGGGCCCGCGCATGAGG

ΦC31 (attR)
AGTGCCCCAACTGGGGTAACCT**TT**GGGCTCCCCGGGCGCGTACTCCACC
TCACGGGGTTGACCCCATTGGA**AA**CCCGAGGGGCCCGCGCATGAGGTGG

ΦC31 (*attL*)
GCGGTGCGGGTGCCAGGGCGTGCCC**TT**GAGTTCTCTCAGTTGGGGGCGTA
CGCCACGCCCACGGTCCCGCACGGG**AA**CTCAAGAGAGTCAACCCCCGCAT