ACS Synthetic Biology
Cite This: ACS Synth. Biol. 2019, 8, 16–24
Letter
pubs.acs.org/synthbio

Comparison of Integrases Identifies Bxb1-GA Mutant as the Most Efficient Site-Specific Integrase System in Mammalian Cells

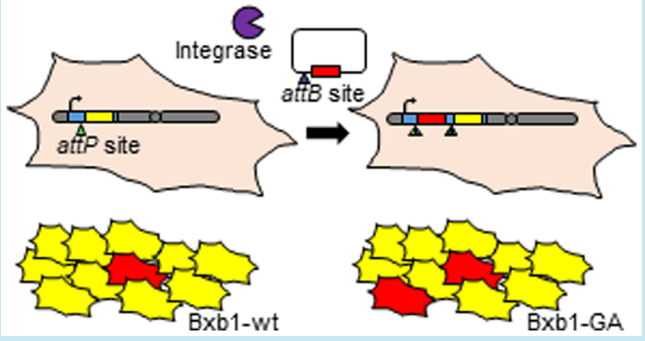
Barbara Jusiak,[†,‡] Kalpana Jagtap,[‡] Leonid Gaidukov,[‡] Xavier Duportet,[‡] Kalpanie Bandara,[§] Jianlin Chu,[§] Lin Zhang,[§] Ron Weiss,*[,‡] and Timothy K. Lu*[,†,‡]

[†]Research Laboratory of Electronics, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

[‡]Synthetic Biology Center, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

[§]Cell Line Development, Biotherapeutics Pharmaceutical Sciences, Pfizer Inc., Andover, Massachusetts 01810, United States

ABSTRACT: Phage-derived integrases can catalyze irreversible, site-specific integration of transgenic payloads into a chromosomal locus, resulting in mammalian cells that stably express transgenes or circuits of interest. Previous studies have demonstrated high-efficiency integration by the Bxb1 integrase in mammalian cells. Here, we show that a point mutation (Bxb1-GA) in Bxb1 target sites significantly increases Bxb1-mediated integration efficiency at the *Rosa26* locus in Chinese hamster ovary cells, resulting in the highest integration efficiency reported with a site-specific integrase in mammalian cells. Bxb1-GA point mutant sites do not cross-react with Bxb1 wild-type sites, enabling their use in applications that require orthogonal pairs of target sites. In comparison, we test the efficiency and orthogonality of ϕC31 and Wβ integrases, and show that Wβ has an integration efficiency between those of Bxb1-GA and wild-type Bxb1. Our data present a toolbox of integrases for inserting payloads such as gene circuits or therapeutic transgenes into mammalian cell lines.



KEYWORDS: *integrase, landing pad, integration efficiency, CHO cell*

Engineering mammalian cells with stably expressed transgenes is vital for research, cell therapy, and production of recombinant therapeutic proteins.[1,2] Targeted integration allows the user to insert the transgene or genetic circuit into a locus that favors stable, long-term expression, such as the *Rosa26* locus in mice[3] or the *AAVS1* locus in human cells.[4] Moreover, when comparing different versions of a transgene or circuit, it is vital to insert them into the same chromosomal context, in order to avoid strong positional effects that may arise from differences in local chromatin structure and the interference from the neighboring genomic environment. The performance of a gene circuit may also depend on its copy number, a property that can be controlled more easily with targeted integration than with random integration—another advantage of targeted integration methods.

Transgene integration into a chromosomal locus of interest can be achieved by using serine integrases. Discovered in bacteriophages, these enzymes integrate the prophage into its bacterial host's genome using DNA target sequences known as *attP* and *attB* sites (attachment-phage and attachment-bacterium, respectively).[5] Serine integrases recognize short DNA sequences (<50 bp), do not require accessory proteins or supercoiled DNA templates, and catalyze irreversible integration unless provided with another phage-encoded protein called the recombinase directionality factor.[5] These attributes make serine integrases attractive tools for stable, site-specific integration of transgenic payloads into the genomes of heterologous host cells, including mammalian cells[6−9] and yeast.[10,11]

Based on serine integrases, Duportet *et al.* previously presented an improved system for fast and reliable engineering of mammalian cell lines by inserting a landing pad (LP) cassette into a predefined genomic locus that supports stable, high-level transgene expression.[12] Using LP cell lines, Duportet *et al.* observed ∼10% integration efficiency in human embryonic kidney (HEK293) cells and ∼11% integration efficiency in Chinese hamster ovary (CHO) cells[12] with the integrase derived from the Bxb1 phage of *Mycobacterium smegmatis*.[7,13]

Several serine integrases besides Bxb1 can catalyze site-specific integration in mammalian host cells, including integrases from the *Streptomyces* phage ϕC31 (refs 6, 14, and 15), the *Streptomyces parvulus* phage R4 (ref 16), the *Lactococcus lactis* phage TP901-1 (ref 17), the *Streptomyces*

Received: February 26, 2018
Published: January 4, 2019

DOI: 10.1021/acssynbio.8b00089
ACS Synth. Biol. 2019, 8, 16–24

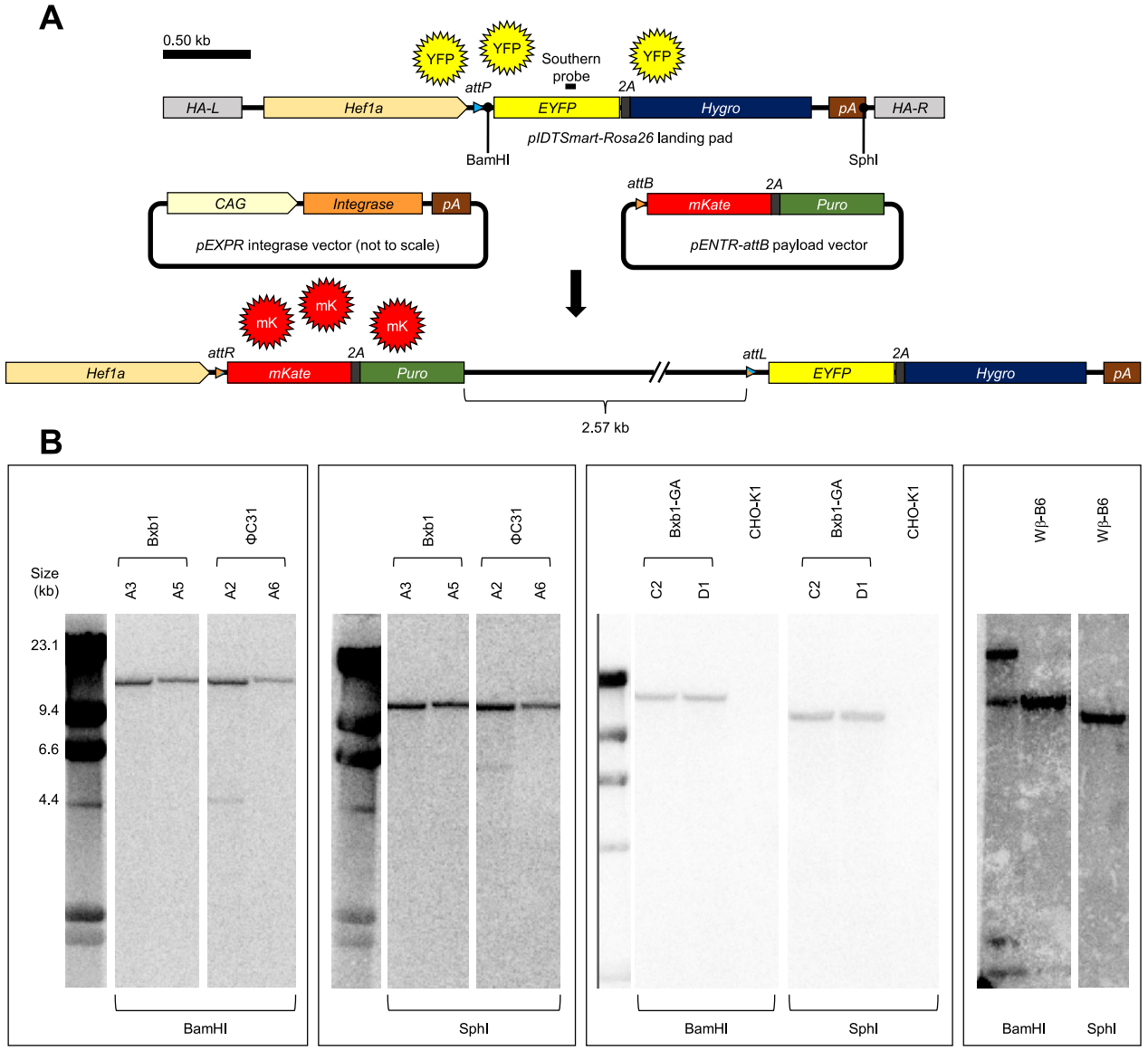


Figure 1. Construction of *Rosa26* LP cell lines. (A) Schematic of the LP and the payload used to assess integration efficiency. The LP (top) constitutively expresses EYFP and a Hygromycin resistance marker, and is targeted to the CHO *Rosa26* locus with homology arms (HA-L and HA-R). The *attP* site enables payload integration between the constitutive *Hef1a* promoter and the *EYFP-2A-Hygro* coding region. The *pENTR* payload includes a promoterless *mKate-2A-Puro* coding region downstream of an *attB* site. Successful integration (bottom) switches the cell from expressing EYFP and Hygro to expressing the red fluorescent protein mKate and Puro. (B) Southern blot of genomic DNA digested separately with *Bam*HI and with SphI indicates that all cell lines tested, except Rosa-ϕC31-A2, have the LP integrated only in the *Rosa26* locus. The positions of the Southern probe and of *Bam*HI and SphI restriction sites are indicated in part A. Southern blots performed on different days are separated by boxes. mK, mKate fluorescent protein; pA, polyadenylation sequence.

phage ϕBT1 (ref 18), and the *Bacillus anthracis* phage Wβ (ref 9). Of these, Bxb1 and ϕC31 have been reported to perform most efficiently in mammalian cell lines,[8,9] but it is difficult to compare the results of different studies due to differences in assay design, the host cell line used, and the position of the integrase target sites in the host cell genome. For example, Xu *et al.* studied intramolecular recombination between arrays of *attP* and *attB* sites on the same chromosome,[9] a method that may result in different integration frequencies than site-specific integration between a chromosomal *attP* site and an extrachromosomal payload *attB* site. Yamaguchi *et al.* tested integration efficiency using a human artificial chromosome (HAC) equipped with five recombinase target sites; however, that study did not measure absolute integration efficiency, but rather fold increase in the number of drug-resistant colonies in the presence of HAC, payload, and integrase vs in absence of either HAC or integrase.[8] Most studies used the rise of drug-resistant colonies after site-specific integration of a drug resistance marker as a measure of payload integration efficiency in mammalian chromosomes.[6,8,9] To our knowledge, only Duportet *et al.* used fluorescent protein expression normalized for transfection efficiency as a readout of integration efficiency.[12] Hence, additional studies are needed to identify the best-performing site-specific integrase in mammalian cells.

Apart from integration efficiency, another key attribute of serine integrases is their specificity for their target sites—each integrase recognizes a unique *attP*/*attB* site pair. The specificity of serine integrases can be reprogrammed by

DOI: 10.1021/acssynbio.8b00089
ACS Synth. Biol. 2019, 8, 16−24

mutating the central dinucleotide of their *attP*/*attB* target sites. The central dinucleotide is where crossover occurs between the *attP* and *attB* sites during recombination, and the sequences of the central dinucleotides in *attP* and *attB* must match in order for recombination to occur.[19,20] Introducing matching point mutations into the central dinucleotides of an *attP*/*attB* site pair results in orthogonal sites that react with each other but not with the wild-type sites[19,20]—a useful property for engineering a single cell line with multiple landing pads encoding different *attP* target sites.

Here, we use the LP method of Duportet *et al.* to assay the efficiency and specificity of Bxb1, ϕC31, and Wβ integrases in the same chromosomal context, the *Rosa26* locus in CHO cells. We find that mutating the central dinucleotide of Bxb1 target sites from GT to GA nearly doubles integration efficiency at the LP locus we tested, resulting in the highest integration efficiency reported for a site-specific integrase in mammalian cells. We also find that the Wβ integrase,[9] which has not been characterized extensively, has efficiency in between those of Bxb1-GA and wild-type Bxb1 in CHO cells, albeit Wβ has a higher rate of off-target payload integration compared with other integrases tested. All integrases tested have very limited crosstalk, making them suitable in applications such as Recombinase-Mediated Cassette Exchange (RMCE), which was demonstrated recently for the Bxb1 wild-type/GA site pair.[21] Our work presents a new and improved tool for high-efficiency integration of transgenic payloads into mammalian cells.

In our adaptation of the landing pad system of Duportet *et al.*,[12] we were able to test the efficiency of multiple integrases in the same chromosomal context without the need for time-consuming drug selection. Using CRISPR-Cas9, we targeted the putative *Rosa26* locus in CHO-K1 cells with LP vectors that constitutively express enhanced yellow fluorescent protein (EYFP) and a hygromycin resistance marker (Hygro) under the control of the strong mammalian *Hef1a* promoter, with an *attP* site between the promoter and the EYFP coding sequence (Figure 1A). To test integrase performance, we used payload vectors with an *attB* site upstream of a promoterless transgene encoding the red fluorescent protein mKate and a puromycin resistance marker (Puro). When the payload integrates into the LP, the cell switches from expressing EYFP and Hygro to mKate and Puro (Figure 1A). To coexpress a fluorescent protein and a resistance marker from a single promoter, we used a 2A sequence—a virus-derived polypeptide that causes ribosome skipping, allowing two proteins to be expressed in equal amounts from a single transcript.[22]

To construct LP cell lines, we cotransfected adherent CHO-K1 cells with the LP vectors and *pX330-t1*, which coexpresses Cas9 and a gRNA against the putative *Rosa26* locus.[23] Following two-week selection with hygromycin, single EYFP + cells were sorted in 96-well plates and clonally expanded. Six or seven clones of each *attP* type were tested for LP integration in the *Rosa26* locus by genomic PCR with on-target primers, and Bxb1-GA, and Wβ clones that tested positive with on-target primers were also tested for off-target LP insertion using genomic PCR with off-target primers. Two putative on-target clones of each *attP* type were analyzed by Southern blot (Figure 1B). Subsequent work used the following clonal LP cell lines: Rosa-Bxb1-A5 and A3; Rosa-Bxb1-GA-C2 and D1; Rosa-Wβ-B6; and Rosa-ϕC31-A6 (Figure 1B).

To measure integration efficiency, we cotransfected the LP cells with the following vectors: (a) the *attB-mKate* payload; (b) the *pEXPR-CAG-integrase* expression vector (+int condition) or *pKan-seq-3x*, a negative-control vector that does not express any proteins (-int condition); and (c) *pHef1a-EBFP2*, which constitutively expresses enhanced blue fluorescent protein 2 (EBFP2), used as transiently expressed transfection control. We measured EBFP2 and mKate fluorescence 3 days post-transfection, without drug selection. The ratio of *EBFP2* to *attB-mKate* was the same in each transfection, allowing us to use EBFP fluorescence to normalize integration efficiency for transfection efficiency. The percentage of EBFP2+ cells that are also mKate+ is the integration efficiency normalized for transfection efficiency.

Unexpectedly, although previous studies suggested that the central dinucleotide does not affect integration efficiency,[19] the Bxb1-GA point mutant *attP*/*attB* site pair shows significantly higher integration efficiency compared with Bxb1 wild-type *attP*/*attB* sites. Two independently derived Bxb1 wild-type cell lines show 10.1% and 9.4% integration efficiency (n = 6 each), similar to previous results.[12] In contrast, two independently derived Bxb1-GA cell lines exhibit more than double the normalized integration efficiency of the wild-type Bxb1 lines, 24.0% and 18.7% (n = 6 each) (Figure 2A). The difference in integration efficiency between wild-type and GA point mutant LP cell lines is statistically significant according to the pairwise exact permutation test (p = 0.0173). To our knowledge, this is the highest integration efficiency reported to date with a site-specific integrase in mammalian cells. Note that the actual number of integrant cells recovered 3 days post-transfection varies depending on transfection efficiency. For example, for the Bxb1-GA-C2 line, if transfection efficiency is 50% (as commonly seen with the Neon system), integrant cells constitute approximately 12% of the entire cell population, or 1.2×10^5 cells per million total cells.

In our landing pad design, recombination between *attB* and *attP* sites creates an *attR* site, which is part of the 5′ UTR of the *mKate* payload transcript. Although the GA point mutation does not create or abolish an ATG site upstream of the EYPF coding region, it is possible that the mutation has an unforeseen effect on the stability or translation efficiency of the payload transcript, and that this effect, rather than increased integration efficiency, is the reason for higher percentage of mKate+ cells in *LP-Bxb1-GA* vs *LP-Bxb1-wt* cell lines. To test this possibility, we built two vectors, *pRecon-Bxb1-wt* and *pRecon-Bxb1-GA*, which encode what we expect the LP locus to look like after successful integration: *Hef1a* promoter followed by *attR* site (wild-type or GA) and *mKate* coding region (Figure 1A). We transfected these vectors transiently into wild-type CHO-K1 cells along with constitutively expressed *pHef1a-EBFP* as transfection control, and assayed fluorescent protein expression by flow cytometry 3 days post-transfection. The average mKate+/EBFP+ ratio was 94.6% in *pRecon-Bxb1-wt* transfected cells and 107.2% in *pRecon-Bxb1-GA* cells (n = 5 each; the percentage is >100% in GA cells because mKate+ cells outnumber EBFP+ cells). The difference in the mKate+/EBFP+ ratio between *pRecon-Bxb1-wt* and *pRecon-Bxb1-GA* samples is not statistically significant (p = 0.254, exact permutation test), suggesting that the difference in payload expression between Bxb1-wt and Bxb1-GA LP cell lines is not due to the effect of the GA point mutation on payload expression.

To compare the efficiency of Bxb1 with other integrases in the same chromosomal context, we tested LP cell lines with *attP* sites for the ϕC31 and Wβ integrases, both of which can

DOI: 10.1021/acssynbio.8b00089
ACS Synth. Biol. 2019, 8, 16−24

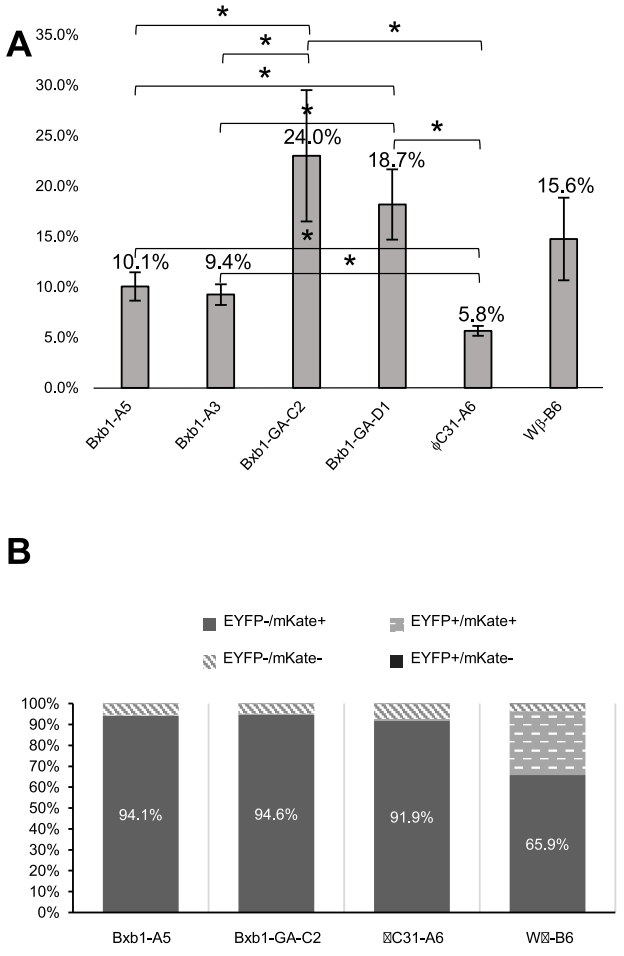


Figure 2. Of the tested integrases, Bxb1-GA is the most efficient site-specific integration system in CHO-K1 cells. (A) Integration efficiency 3 days post-transfection with payload, EBFP expression vector, and integrase expression vector. *Y*-axis shows % of EBFP+ cells (transfection control) that are also mKate+ (indicating payload integration). All cells received the same ratio of EBFP control vector to payload vector, allowing the EBFP vector to serve as transfection control. Error bars indicate standard deviation. Statistically significant differences ($p < 0.05$) are indicated with asterisks. Bxb1-GA shows significantly higher integration efficiency than wild-type Bxb1. (B) Stable integration of the payload after 12-day puromycin selection. The desired cells are EYFP-/mKate+, indicating disruption of the EYFP+ LP by the mKate+ Puro+ payload (see Figure 1A). Site-specific integration efficiency is >90% in Bxb1, Bxb1-GA, and ϕC31 cell lines; in contrast, ∼ 30% of Wβ cells are EYFP+/mKate+, suggesting off-target payload integration.

function in heterologous host cells.[6,9,14] Wβ catalyzes payload integration with an efficiency that is in between Bxb1-GA and wild-type Bxb1 (15.6%, $n = 3$) (Figure 2A). ϕC31 exhibits the lowest integration efficiency in our system, at 5.8% ($n = 6$) (Figure 2A). With all integrases tested, we observed very few to no mKate+ cells (0.00%−0.02%, $n = 3$ for each integrase) in the -int condition, indicating that payload insertion requires the integrase (data not shown).

Developing stable transgenic cell lines quickly and reliably is often the rate-limiting step in mammalian synthetic biology. Using previously described protocols,[12] we established stable payload-expressing cell pools by starting puromycin selection 3 days after transfection with *attB*-payload plus integrase. After 12 days of selection, we sorted for EYFP-/mKate+ cells, where mKate+ indicated payload integration, and EYFP- indicated disruption of the LP by site-specific payload insertion (Figure 1A). For Bxb1-A5, Bxb1-GA-C2, and ϕC31, >90% of cells that survived Puromycin selection were EYFP−/mKate+, indicating highly efficient payload integration; only 0.6%−0.7% of surviving cells were EYFP+/mKate+, suggesting off-target payload insertion (Figure 2B). In contrast, Wβ-transfected cells had a lower frequency of desired EYFP−/mKate+ cells (65.9%) and a much higher frequency of EYFP+/mKate+ cells (30.6%) (Figure 2B), suggesting payload integration into off-target sites, possibly a pseudo-*attP* site in a transcriptionally active locus. Note that we are counting only off-target events that result in payload expression, presumably due to the presence of a pseudo-*attP* site near an active endogenous promoter; therefore, the frequency of off-target events will vary between different cell lines that have distinct patterns of endogenous promoter activity. Also, cell lines from different organisms vary in their distribution of pseudo-*attP* sites, resulting in different frequencies of observed off-target integration. LP cells transfected with *attB*-payload plus the negative-control empty vector all died under puromycin selection (data not shown), indicating that payload insertion requires the integrase. Site-specific payload integration into the LP *attP* site was confirmed by genomic PCR for Bxb1 and Bxb1-GA *att* sites (Figure 3A). Both Bxb1-wt and Bxb1-GA LP cell lines are hemizygous for the LP at the *Rosa26* locus, suggesting that differences in LP copy number are not the reason for the differential integration efficiency of Bxb1-wt vs Bxb1-GA (Figure 3B).

For applications that require more than one *attP*/*attB*/integrase combination per cell, such as Recombinase-Mediated Cassette Exchange (RMCE),[21,24] it is important to avoid crosstalk between integrases and their target sites by using orthogonal integrases or *attB*/*attP* site pairs. Orthogonal integrases and target sites are also vital for constructing memory devices and other sophisticated synthetic circuits.[25,26] We tested the degree of crosstalk between our integrases and their target sites by transfecting Bxb1-A5, Bxb1-GA-C2, ϕC31, and Wβ LP cells with mismatched sets of *attB*-payloads and integrase expression vectors (Figure 4). Off-target integration was assayed by the fraction of EBFP2+ cells that were also mKate+, 3 days post-transfection without drug selection, as described above ($n = 3$ for each *attP*/*attB*/integrase combination).

We found limited to no crosstalk for all our mismatched combinations. The highest off-target integration frequency, observed for Bxb1-GA *attP* + Bxb1-GA *attB* + Wβ integrase, was 0.12%, which is 48-fold lower than the least efficient on-target integrase, ϕC31 (Figure 4). All other off-target combinations had normalized integration efficiency <0.10% (Figure 4). Consistent with previous *in vitro* studies, which showed that integrases are highly sensitive to mismatches in the central *attP*/*attB* dinucleotide,[19,20] we observed only 0.06% integration both for Bxb1 wild-type *attP* + GA *attB* and for GA *attP* + wild-type *attB* (Figure 4). We did not test combinations where all three elements were mismatched, such as Bxb1-*attP* + Wβ-*attB* + ϕC31 integrase, reasoning that such combinations are even less likely to show any activity than the single-mismatch combinations that we tested. We conclude that all the integrases are highly orthogonal to each other.

This high level of specificity is consistent with the mechanism of action of serine integrases elucidated through

DOI: 10.1021/acssynbio.8b00089
ACS Synth. Biol. 2019, 8, 16−24

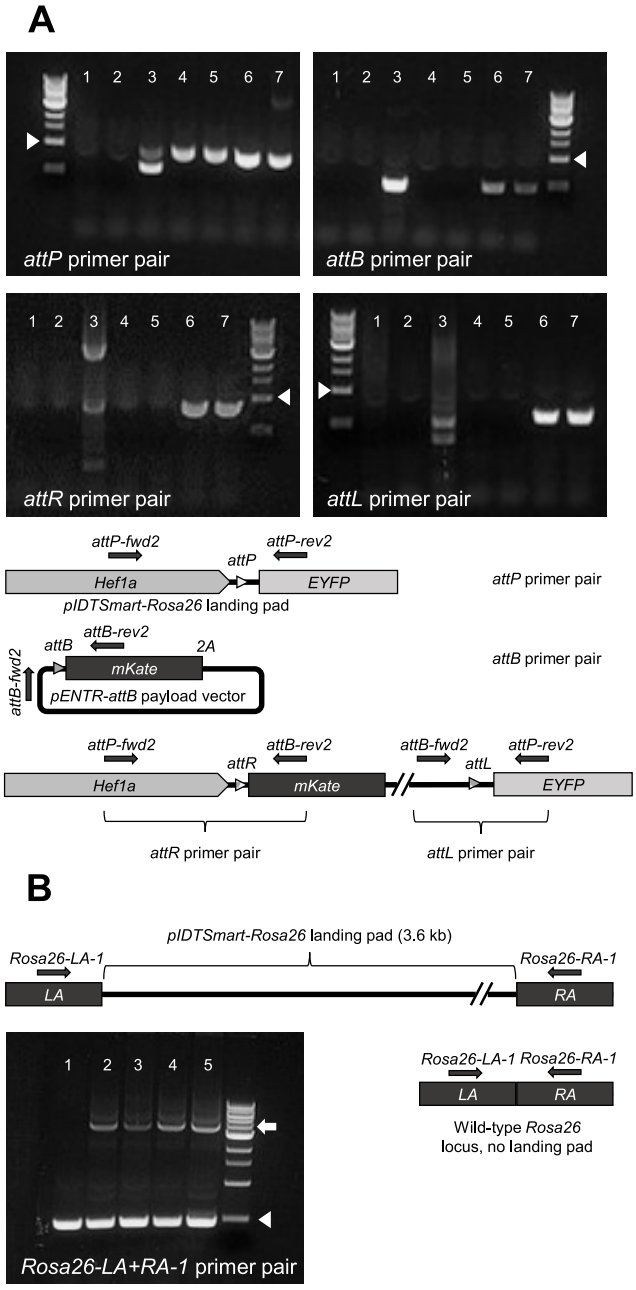


Figure 3. PCR validation of LP cell lines. (A) PCR confirms site-specific payload integration into the LP. From left to right, the lanes in each panel are (1) no-DNA negative control; (2) wild-type CHO genomic DNA; (3) *pENTR-attB* plasmid; (4) parental (preintegration) *Bxb1-attP-A5* genomic DNA; (5) parental *Bxb1-GA-C2* genomic DNA; (6) genomic DNA from Puro-selected, mKate+/YFP− sorted *Bxb1-A5-mKate* pooled integrant cells; (7) genomic DNA from selected and sorted *Bxb1-GA-C2-mKate* pooled integrant cells. White arrowheads indicate 1 kb. Bottom panel shows where the genomic primers align. As expected, the *attR* and *attL* primer pairs give strong products only with the integrant DNA, but not with the parental LP DNA or payload plasmid DNA. The integrant DNA gives bands with *attP* and *attB* primer pairs, likely as a result of incomplete integration into the LP (thus preserving an intact LP in some cells) and off-target integration of the payload vector, respectively. (B) Each LP cell line is hemizygous for the landing pad. The gel lanes are 1, wild-type CHO genomic DNA; 2, *Bxb1-attP-A5* DNA; 3, *Bxb1-attP-A3*; 4, *Bxb1-GA-attP-C2*; 5, *Bxb1-GA-attP-D1*. The primer pair, *Rosa26-LA+RA-1*, gives a 4.1 kb product in the presence of the LP and a 0.5 kb product in the absence of the LP. White arrowhead marks 0.5 kb; white arrow marks 4 kb. In all LP cell lines, we see both long (LP) and short (no LP) bands. In a cell line homozygous for the LP, we would expect to see the long band only. LA and RA, left and right homology arms, respectively.

Figure 3. continued

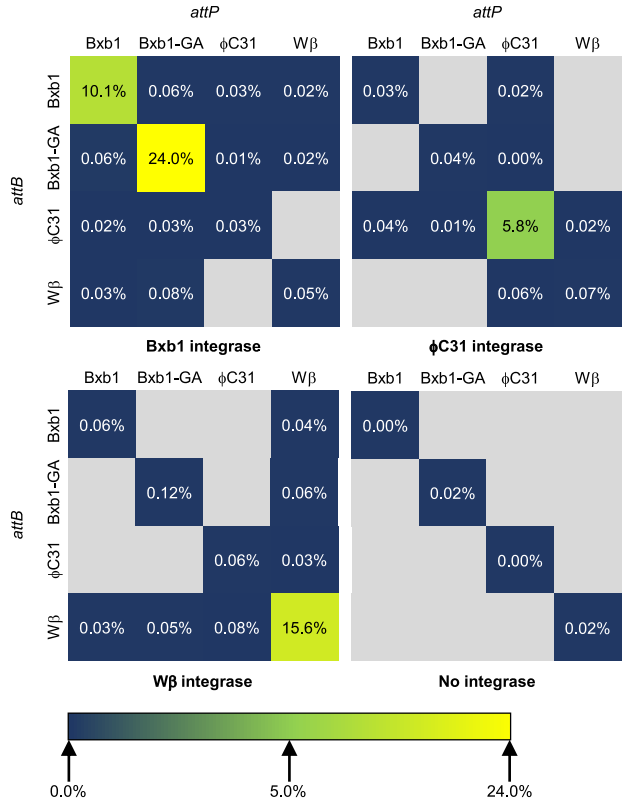


Figure 4. Integrases have minimal cross-talk with each other. The color scale represents integration efficiency 3 days post-transfection with payload, transfection control, and integrase expression vector (or negative control vector). Integration efficiency was calculated as the % of EBFP+ cells (transfection control) that are also mKate+ (integration marker). Triple mismatches, e.g., Bxb1-*attP* + Bxb1-GA-*attB* + ϕC31 integrase, were not tested, and are indicated in gray. The cell lines used were Bxb1-A5, Bxb1-GA-C2, ϕC31-A6, and Wβ-B6.

in vitro studies.[19,20] Integrase proteins bind to their *attP* and *attB* sites, and then cleave the *attP* and *attB* sites simultaneously, exposing the central dinucleotides of the sites as single-stranded overhangs. Rotation of the integrases causes the single-stranded *attP* and *attB* overhangs to line up with each other (reviewed in ref 5). If the central dinucleotides of the original *attP* and *attB* sites are the same, the integrases ligate the newly aligned DNA strands together, thus completing the process. If the central dinucleotides are mismatched, they do not undergo ligation, and the integrases continue to rotate until the original *attP* and *attB* sites are restored.[19,20]

This mechanism of integrase action explains the extremely limited crosstalk between Bxb1 wild-type and GA point mutant sites, but it does not explain the significantly higher efficiency of Bxb1-GA sites relative to wild-type sites. Previous studies of serine integrases *in vitro* and in bacteria did not observe a difference in recombination efficiency between wild-type and point mutant site pairs.[19,20] A recent *in vitro* study of Bxb1

DOI: 10.1021/acssynbio.8b00089
ACS Synth. Biol. 2019, 8, 16−24

Table 1. Oligos Used To Build *attP*/*attB* Sites in LP and Payload Vectors, To Amplify Integrase Coding Regions for Cloning into the Expression Vector, and To Build *pRecon-Bxb1-wt* and *GA* Vectors

oligo name	sequence
Bxb1-Rosa-attP-top	TCGACGTGGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCCAT
Bxb1-Rosa-attP-bottom	CTAGATGGGTTTGTACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCACG
Bxb1-GA-Rosa-attP-top	TCGACGTGGTTTGTCTGGTCAACCACCGCGGACTCAGTGGTGTACGGTACAAACCCAT
Bxb1-GA-Rosa-attP-bottom	CTAGATGGGTTTGTACCGTACACCACTGAGTCCGCGGTGGTTGACCAGACAAACCACG
phiC31-Rosa-attP-top	TCGACGGAGTAGTGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTAGGGTCT
phiC31-Rosa-attP-bottom	CTAGAGACCCTACGCCCCCAACTGAGAGAACTCAAAGGTTACCCCAGTTGGGGCACTACTCCG
Wbeta-Rosa-attP-top	TCGACTAGTTTTAAAGTTGGTTATTAGTTACTGTGATATTTATCACGGTACCCAATAACCAATGAAT
Wbeta-Rosa-attP-bottom	CTAGATTCATTGGTTATTGGGTACCGTGATAAATATCACAGTAACTAATAACCAACTTTAAAACTAG
Bxb1-GA-pENTR-attB-top	CGGCCGGCTTGTCGACGACGGCGGACTCCGTCGTCAGGATCATCCGGCCGCCAC
Bxb1-GA-pENTR-attB-bottom	CATGGTGGCGGCCGGATGATCCTGACGACGGAGTCCGCCGTCGTCGACAAGCCGGCCGGTAC
phiC31-pENTR-attB-top	CGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCGCCAC
phiC31-pENTR-attB-bottom	CATGGTGGCGGAGTACGCGCCCGGGGAGCCCAAGGGCACGCCCTGGCACCCGCACCGGTAC
Wbeta-pENTR-attB-top	CGGAAGGTAGCGTCAACGATAGGTGTAACTGTCGTGTTTGTAACGGTACTTCCAACAGCTGGCGCCGCCAC
Wbeta-pENTR-attB-bottom	CATGGTGGCGGCGCCAGCTGTTGGAAGTACCGTTACAAACACGACAGTTACACCTATCGTTGACGCTACCTTCCGGTAC
pEXPR-phiC31-fwd	TGGCAAAGAATTCGACCCCAAGTTTGTACAAAAAAGCAGGAGGCCTGCCACCATGGACACCTACGCCG
pEXPR-phiC31-rev	GGAGTGAATTCACGCGTACCACTTTGTACAAGAAAGCTGGGTATTAAACGGGCCCTCTAGATGC
pEXPR-Wbeta-fwd	TGGCAAAGAATTCGACCCCAAGTTTGTACAAAAAAGCAGGAGGCCTGCCACCATGAAATACGCAGTCTACGTCAGAG
pEXPR-Wbeta-mN-rev	GGAGTGAATTCACGCGTACCACTTTGTACAAGAAAGCTGGGTATCACAGGCTGAAGGTGTATTCGA
Recon-attR-wt-top	TTGTCTGGTCAACCACCGCGGTCTCCGTCGTCAGGATCATCCGGCCGCCACCATGGTGTCTAAGGGCGAAGAGCTGATTA
Recon-attR-wt-bottom	TAATCAGCTCTTCGCCCTTAGACACCATGGTGGCGGCCGGATGATCCTGACGACGGAGACCGCGGTGGTTGACCAGACAA
Recon-attR-GA-top	TTGTCTGGTCAACCACCGCGGACTCCGTCGTCAGGATCATCCGGCCGCCACCATGGTGTCTAAGGGCGAAGAGCTGATTA
Recon-attR-GA-bottom	TAATCAGCTCTTCGCCCTTAGACACCATGGTGGCGGCCGGATGATCCTGACGACGGAGTCCGCGGTGGTTGACCAGACAA
Recon-Hef1a-fwd	CAAAAAAGCAGGCTGGATCCGGTACCTGCTGACAGGTCCCTTGACA
Recon-Hef1a-wt-rev	TGGCGGCCGGATGATCCTGACGACGGAGACCGCGGTGGTTGACCA
Recon-Hef1a-GA-rev	TGGCGGCCGGATGATCCTGACGACGGAGTCCGCGGTGGTTGACCA

integrase suggests that the central dinucleotide affects the kinetics of recombination.[27] The stability of base-pairing interaction between cleaved *attP* and *attB* sites influences the rate of ligation: the Bxb1-GC point mutant site pair undergoes ligation more rapidly than the Bxb1-AT point mutant site pair.[27] However, we do not expect this phenomenon to affect integration efficiency in our study, since the two dinucleotide sequences we used, the wild-type GT and the GA point mutant, are predicted to form equally stable base-pairing interactions.

To our knowledge, this is the first report that mutations in the central dinucleotide of *attP* and *attB* sites affect integration efficiency. We also observed that ϕC31-CC and ϕC31-GT point mutant site pairs strongly reduce integration efficiency compared with wild-type ϕC31-TT *attP*/*attB* sites in transiently transfected CHO cells (our unpublished data). The mechanism by which the central dinucleotide sequence influences integration efficiency warrants further studies.

LP cell lines present a convenient way to stably integrate transgenes and genetic circuits into mammalian cells for research and engineering purposes.[12] This platform depends on the availability of highly efficient, site-specific integrases, yet previous studies suggest that many integrases that work well in heterologous bacterial hosts perform poorly in mammalian cells.[9] The Bxb1 integrase performs well in mammalian cells,[9,12] but there is a need for additional high-performing integrases that are orthogonal to Bxb1. Here we identify Bxb1-GA as the most efficient site-specific integration system reported in mammalian cells to date, and show that Bxb1-GA point mutant sites are orthogonal to wild-type Bxb1 sites, enabling them to be used in applications such as Bxb1-based RMCE.[21] Highly efficient and orthogonal integrase target sites may be used in applications such as multi-LP cell lines for the expression of multiple large payloads[23] or synthetic memory circuits.[25,26]

METHODS

Landing Pad Site Identification and Vector Construction. We identified a putative *Rosa26* locus in CHO-K1 cells and targeted it with an LP vector, *pIDTSmart-Rosa-Hef1a-attP(Bxb1)-EYFP-2A-Hygro* (ref 23). We designed a gRNA against the *Rosa26* locus and cloned it into *pX330-U6-chimeric_BB-CBh-hSpCas9*, a gift from Feng Zhang[28] (Addgene plasmid #42230), to make the vector *pX330-t1* (ref 23).

DOI: 10.1021/acssynbio.8b00089
ACS Synth. Biol. 2019, 8, 16−24

The payload vector *pENTR-attB(Bxb1)-mKate-2A-Puro* was previously described.[12] We designed single-stranded DNA oligos that, when annealed together, form double-stranded Bxb1-GA, ϕC31, and Wβ *attP* and *attB* sites with restriction enzyme overhangs. These sites were cloned into the LP vector digested with *Sal*I and *Xba*I, and into the payload vector digested with *Kpn*I and *Nco*I, in order to replace the original Bxb1 *attP*/*attB* sites.

To build *pRecon-Bxb1-wt* and *pRecon-Bxb1-GA*, the *pENTR-attB(Bxb1)-mKate-2A-Puro* vector was digested with *Kpn*I + *Nco*I. The vector backbone was used in Gibson assembly with the *Hef1a* promoter PCR-amplified from *pIDTSmart-Rosa-Hef1a-attP(Bxb1)-EYFP-2A-Hygro* with *Recon-Hef1a-fwd + rev* primers and a double-stranded oligo encoding the Bxb1 *attR* site, formed by annealing *pRecon-attR-top* and *bottom* single-stranded oligos together.

The *pEXPR-CAG-Bxb1* vector, which constitutively expresses mammalian codon-optimized Bxb1 integrase, was previously described.[12] *CAG* is a strong, constitutive, synthetic promoter used in mammalian cells.[29] To build *pEXPR-CAG-ϕC31* and *pEXPR-CAG-Wβ*, mammalian codon-optimized integrase coding regions were amplified from *pVAX1-CMV-ϕC31* (ref 8) and *pCAG-coWβ-IRES-Zeo* (ref 9), respectively, and cloned by Gibson assembly[30] into *pEXPR-CAG-Bxb1* backbone digested with *Bsr*GI (all restriction enzymes were from New England Biolabs). Table 1 lists the oligonucleotides encoding *attB* and *attP* sites and the primers used to amplify integrase coding regions, all ordered from Integrated DNA Technologies.

CHO Cell Culture. Adherent CHO-K1 cells were maintained in F-12K media (ATCC) with 10% fetal bovine serum (FBS) (Sigma-Aldrich), 1% HyClone nonessential amino acids (NEAA) (GE Healthcare Life Sciences), and 1% penicillin/streptomycin (Pen/Strep) (Gibco by Life Sciences). Cells were grown in a humidified 37 °C incubator with 5% CO_2 and passaged every 3−4 days. Drug selection with 400 μg/mL hygromycin (InvivoGen) or 8 μg/mL puromycin (Gibco by Life Sciences) began 3 days post-transfection. Fresh media were added daily for the first 3 days of selection, and every second day thereafter. Cells were passaged as needed during selection.

Transfection and Flow Cytometry. All transfections were performed with the Neon electroporation system (Invitrogen) following manufacturer's instructions using 10 μL Neon tips and 1.0×10^5 cells/reaction with the following settings: 1560 V, 5 ms, 10 pulses. To make LP cell lines, CHO-K1 adherent cells were cotransfected with 50−100 ng of supercoiled LP vector plus 25 ng of *pX330-t1* Cas9/gRNA vector.[23] To test integration efficiency, LP cells were transfected in triplicate in 24-well format with 500 ng of *pENTR* payload vector, 300 ng of *pEXPR* integrase vector or negative control vector *pKan-seq-3x* (ref 31), and 300 ng of transfection control vector, *pEXPR-Hef1a-EBFP2* (ref 12). For *pRecon* experiments, we used 500 ng of *pRecon* and 300 ng of *Hef1a-EBFP* per well. Media were changed 1 day after electroporation. Flow cytometry was performed 3 days post-transfection on a BD LSRFortessa running FACSDiva software at the Swanson Flow Cytometry Core at the Koch Institute for Integrative Cancer Research at MIT. Flow cytometry data were analyzed using CytoFlow software (bpteague.github.io/cytoflow).

Fluorescence-Activated Cell Sorting and LP Cell Clonal Expansion. Cells were sorted on a BD FACS Aria1 sorter running FACSDiva software at the Swanson Flow Cytometry Core. For payload integration, EYFP−/mKate+ cells were sorted in bulk into a 5-ml tube, cultured under standard conditions, and frozen. For stable LP integration, single EYFP+ cells were seeded into a 96-well plate prepared with F12-K media with 20% FBS, NEAA, and Pen/Strep. Starting 4 days postsort, wells were checked every second day for the presence of a single colony, and cells received fresh 20% FBS media every second day. Twenty-four surviving clones for each LP type were expanded into a 24-well plate, and tested for EYFP expression by flow cytometry and for LP integration by PCR.

Genomic PCR. Genomic DNA was prepared using the DNeasy Blood & Tissue Kit (Qiagen). Site-specific LP integration in the *Rosa26* locus was tested by PCR with *pLPhef1a-BGH-for* (CCCGTTTAAACCCGCTGATCAGC) and *Rosa-outRA-0.4k-rev* (CATCCGACCTTAACAACCTGCATTCC). Off-target integrations were tested with *pIDTSmart-for* (GTAAAACGACGGCCAGTTTATCTAGTCAGC) and *AT-Hef1a-rev1* (TGGCCCGCATTTACAAGACT) with 2x Pfu Ultra Master Mix (Agilent Technologies). To avoid false positives, we used 5 ng of template DNA/25 μL of reaction and the following cycling conditions: (1) 95 °C, 5 min; (2) 95 °C, 20 s; (3) 55 °C, 20 s; (4) 72 °C, 30 s; (5) go to (2) 29× more; (6) 72 °C, 3 min; (7) 4 °C, hold.

To confirm *attB*-payload integration into *attP*-LP, we used the following primers: *attP-LP-fwd2* (TGATGACCTGCTGCGACGCT), *attP-LP-rev2* (GTAGCGGGCGAAGCACTGC), *attB-payload-fwd2* (CCTACTCAGGAGAGCGTTCACCGA), and *attB-payload-rev2* (CCAGGATGTCGAAGGCGAAGG). *attP-LP-fwd2 + attB-payload-rev2* are the *attR* primer pair, and *attB-payload-fwd2 + attP-LP-rev2* are the *attL* primer pair. We used Phusion polymerase (NEB) with 3% DMSO and with the following cycling conditions: (1) 98 °C, 30 s; (2) 98 °C, 10 s; (3) 68 °C, 15 s; (4) 72 °C, 45 s for *attP-LP-fwd2*, 20 s for *attB-payload-fwd2*; (5) go to (2) 29×; (6) 72 °C, 10 min; (7) 10 °C, hold. To test whether LP is homo- or hemizygous, we did PCR with primers flanking the LP: *Rosa26-LA-1* (TCGTGATCTGCAAGTCGAGG) and Rosa26-RA-1 (GCTCTGTCTAGGGATTGGGTGA). We used Phusion polymerase with 3% DMSO and the same cycling conditions as above, except annealing temperature is 64 °C, extension time is 2 min, and number of cycles is 35.

Southern Blot. Genomic DNA from LP clones that tested positive by PCR was quantified with a Qubit fluorimeter and the Qubit dsDNA BR kit (Invitrogen). The Southern probe (GGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAG) aligns in the EYFP coding region of the LP. Probes were labeled with dCTP, α-^{32}P (PerkinElmer) using the Prime-It RmT Random Primer Labeling Kit (Stratagene), and free nucleotides were removed using Quick Spin Columns for radiolabeled DNA purification Sephadex G-50 (Roche). DNA was digested with *Sph*I and with *Bam*HI.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: timlu@mit.edu.
*E-mail: rweiss@mit.edu.

ORCID

Timothy K. Lu: 0000-0002-3918-8923

DOI: 10.1021/acssynbio.8b00089
ACS Synth. Biol. 2019, 8, 16−24

Author Contributions

B.J. and K.J. did the experiments. B.J., L.Z., R.W., and T.K.L. analyzed the data. B.J. wrote the manuscript, and R.W. and T.K.L. edited it. L.G. identified the CHO *Rosa26* locus and designed the strategy to target it with CRISPR/Cas9. X.D. characterized integrase crosstalk. K.B. and J.C. did Southern blots. All authors approved the final manuscript.

Funding

This work was supported by the CHO2.0 Pfizer-MIT Alliance and the 1P50GM098792 grant from the National Institutes of Health to T.K.L. and R.W.

Notes

The authors declare the following competing financial interest(s): This work has been funded by Pfizer, Inc. which has a financial interest in the outcome of this work. Several authors (K.B., J.C. and L.Z.) are Pfizer, Inc. employees.

■ ACKNOWLEDGMENTS

We thank Na Li, Selamawit Waka, and Cong Liu for help with CHO cell culture; Mervelina Saturno-Condon, Michele Griffin, and Michael Jennings for help with cell sorting; Brian Teague for assistance with CytoFlow software; Dr. William Brown for the *pCAG-coWβ-IRES-Zeo* vector; and Dr. Tetsuya Ohbayashi for the *pVAX1-CMV-ϕC31* vector. Statistical support was provided by data science specialist Steven Worthington at the Institute for Quantitative Social Science, Harvard University.

■ REFERENCES

(1) Xie, M., Haellman, V., and Fussenegger, M. (2016) Synthetic biology-application-oriented cell engineering. *Curr. Opin. Biotechnol. 40*, 139−48.

(2) Cacciatore, J. J., Chasin, L. A., and Leonard, E. F. (2010) Gene amplification and vector engineering to achieve rapid and high-level therapeutic protein production using the Dhfr-based CHO cell selection system. *Biotechnol. Adv. 28* (6), 673−81.

(3) Zambrowicz, B. P., Imamoto, A., Fiering, S., Herzenberg, L. A., Kerr, W. G., and Soriano, P. (1997) Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. *Proc. Natl. Acad. Sci. U. S. A. 94* (8), 3789−94.

(4) Smith, J. R., Maguire, S., Davis, L. A., Alexander, M., Yang, F., Chandran, S., ffrench-Constant, C., and Pedersen, R. A. (2008) Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration. *Stem Cells 26* (2), 496−504.

(5) Fogg, P. C., Colloms, S., Rosser, S., Stark, M., and Smith, M. C. (2014) New applications for phage integrases. *J. Mol. Biol. 426* (15), 2703−16.

(6) Thyagarajan, B., Olivares, E. C., Hollis, R. P., Ginsburg, D. S., and Calos, M. P. (2001) Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol. Cell. Biol. 21* (12), 3926−34.

(7) Russell, J. P., Chang, D. W., Tretiakova, A., and Padidam, M. (2006) Phage Bxb1 integrase mediates highly efficient site-specific recombination in mammalian cells. *BioTechniques 40* (4), 460−464.

(8) Yamaguchi, S., Kazuki, Y., Nakayama, Y., Nanba, E., Oshimura, M., and Ohbayashi, T. (2011) A method for producing transgenic cells using a multi-integrase system on a human artificial chromosome vector. *PLoS One 6* (2), e17267.

(9) Xu, Z., Thomas, L., Davies, B., Chalmers, R., Smith, M., and Brown, W. (2013) Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. *BMC Biotechnol. 13*, 87.

(10) Thomson, J. G., and Ow, D. W. (2006) Site-specific recombination systems for the genetic manipulation of eukaryotic genomes. *Genesis 44* (10), 465−76.

(11) Xu, Z., and Brown, W. R. (2016) Comparison and optimization of ten phage encoded serine integrases for genome engineering in Saccharomyces cerevisiae. *BMC Biotechnol. 16*, 13.

(12) Duportet, X., Wroblewska, L., Guye, P., Li, Y., Eyquem, J., Rieders, J., Rimchala, T., Batt, G., and Weiss, R. (2014) A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res. 42* (21), 13440−51.

(13) Keravala, A., Groth, A. C., Jarrahian, S., Thyagarajan, B., Hoyt, J. J., Kirby, P. J., and Calos, M. P. (2006) A diversity of serine phage integrases mediate site-specific recombination in mammalian cells. *Mol. Genet. Genomics 276* (2), 135−46.

(14) Groth, A. C., Olivares, E. C., Thyagarajan, B., and Calos, M. P. (2000) A phage integrase directs efficient site-specific integration in human cells. *Proc. Natl. Acad. Sci. U. S. A. 97* (11), 5995−6000.

(15) Andreas, S., Schwenk, F., Kuter-Luks, B., Faust, N., and Kuhn, R. (2002) Enhanced efficiency through nuclear localization signal fusion on phage PhiC31-integrase: activity comparison with Cre and FLPe recombinase in mammalian cells. *Nucleic Acids Res. 30* (11), 2299−306.

(16) Olivares, E. C., Hollis, R. P., and Calos, M. P. (2001) Phage R4 integrase mediates site-specific integration in human cells. *Gene 278* (1−2), 167−76.

(17) Stoll, S. M., Ginsburg, D. S., and Calos, M. P. (2002) Phage TP901−1 site-specific integrase functions in human cells. *J. Bacteriol. 184* (13), 3657−63.

(18) Xu, Z., Lee, N. C., Dafhnis-Calas, F., Malla, S., Smith, M. C., and Brown, W. R. (2008) Site-specific recombination in Schizosaccharomyces pombe and systematic assembly of a 400kb transgene array in mammalian cells using the integrase of Streptomyces phage phiBT1. *Nucleic Acids Res. 36* (1), e9.

(19) Ghosh, P., Kim, A. I., and Hatfull, G. F. (2003) The orientation of mycobacteriophage Bxb1 integration is solely dependent on the central dinucleotide of attP and attB. *Mol. Cell 12* (5), 1101−11.

(20) Smith, M. C., Till, R., and Smith, M. C. (2004) Switching the polarity of a bacteriophage integration system. *Mol. Microbiol. 51* (6), 1719−28.

(21) Inniss, M. C., Bandara, K., Jusiak, B., Lu, T. K., Weiss, R., Wroblewska, L., and Zhang, L. (2017) A novel Bxb1 integrase RMCE system for high fidelity site-specific integration of mAb expression cassette in CHO cells. *Biotechnol. Bioeng. 114*, 1837.

(22) Donnelly, M. L., Hughes, L. E., Luke, G., Mendoza, H., ten Dam, E., Gani, D., and Ryan, M. D. (2001) The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. *J. Gen. Virol. 82* (Pt 5), 1027−41.

(23) Gaidukov, L., Wroblewska, L., Teague, B., Nelson, T., Zhang, X., Liu, Y., Jagtap, K., Mamo, S., Tseng, W. A., Lowe, A., Das, J., Bandara, K., Baijuraj, S., Summers, N. M., Lu, T. K., Zhang, L., and Weiss, R. (2018) A multi-landing pad DNA integration platform for mammalian cell engineering. *Nucleic Acids Res. 46* (8), 4072−4086.

(24) Turan, S., Zehe, C., Kuehle, J., Qiao, J., and Bode, J. (2013) Recombinase-mediated cassette exchange (RMCE) - a rapidly-expanding toolbox for targeted genomic modifications. *Gene 515* (1), 1−27.

(25) Siuti, P., Yazbek, J., and Lu, T. K. (2013) Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol. 31* (5), 448−52.

(26) 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* (6297), aad8559.

(27) Keenholtz, R. A., Grindley, N. D., Hatfull, G. F., and Marko, J. F. (2016) Crossover-site sequence and DNA torsional stress control strand interchanges by the Bxb1 site-specific serine recombinase. *Nucleic Acids Res. 44* (18), 8921−8932.

(28) Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science 339* (6121), 819−23.

(29) Miyazaki, J., Takaki, S., Araki, K., Tashiro, F., Tominaga, A., Takatsu, K., and Yamamura, K. (1989) Expression vector system

DOI: 10.1021/acssynbio.8b00089
ACS Synth. Biol. 2019, 8, 16−24

based on the chicken beta-actin promoter directs efficient production of interleukin-5. *Gene 79* (2), 269−77.
(30) Gibson, D. G. (2011) Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol. 498*, 349−61.
(31) Guye, P., Li, Y., Wroblewska, L., Duportet, X., and Weiss, R. (2013) Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acids Res. 41* (16), e156.

DOI: 10.1021/acssynbio.8b00089
ACS Synth. Biol. 2019, 8, 16−24