Serine Integrases

Citation for published version:

Digital Object Identifier (DOI):
10.1021/acssynbio.7b00308

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
ACS Synthetic Biology

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Serine Integrases: Advancing Synthetic Biology

Christine A. Merrick,*†‡ Jia Zhao,‡ and Susan J. Rosser†

†School of Biological Sciences, University of Edinburgh, Roger Land Building, Alexander Crum Brown Road, Edinburgh EH9 3FF, U.K.
‡Novo Nordisk (China) Pharmaceuticals Co., Ltd., Lei Shing Hong Center, Guangshunnan Avenue, Beijing 100102, China

ABSTRACT: Serine integrases catalyze precise rearrangement of DNA through site-specific recombination of small sequences of DNA called attachment (att) sites. Unlike other site-specific recombinases, the recombination reaction driven by serine integrases is highly directional and can only be reversed in the presence of an accessory protein called a recombination directionality factor (RDF). The ability to control reaction directionality has led to the development of serine integrases as tools for controlled rearrangement and modification of DNA in synthetic biology, gene therapy, and biotechnology. This review discusses recent advances in serine integrase technologies focusing on their applications in genome engineering, DNA assembly, and logic and data storage devices.

KEYWORDS: serine integrase, recombination directionality factor (RDF), DNA assembly, genome engineering, logic devices, genetic data storage

DNA manipulation and modification are at the core of synthetic biology. Advances in the capacity to engineer DNA with increasingly complex properties are accelerating a diverse range of medical and biotechnological applications. In recent years, the principles of homologous recombination have been applied to DNA assembly1 (for example, in construction of the first synthetic genome)2 and to targeting specific changes in genomic DNA (for example, in nuclease-based genome editing with clustered regularly interspaced short palindromic repeat (CRISPR)-associated nucleases like Cas9 and Cpf1).3–8 However, another group of DNA modifying enzymes—the site-specific recombinases—are becoming established as vital tools for genetic modification both in vivo and in vitro. Unlike homologous recombination, site-specific recombination requires no extensive DNA sequence homology, no synthesis or degradation of DNA, and has no reliance on endogenous repair pathways or cofactors.

Serine integrases are a subfamily of serine site-specific recombinases, which are evolutionarily and mechanistically distinct from tyrosine site-specific recombinases.9 Serine recombinases make double strand breaks in DNA forming covalent 5′-phosphoserine bonds with the DNA backbone before religation, whereas tyrosine recombinases cleave single strands forming covalent 3′-phosphothioerysine bonds with the DNA backbone and rejoin the strands via a Holliday junction-like intermediate state. In comparison to tyrosine recombinases, recombination by serine integrases has the advantage of simplicity, normally only requiring the integrase protein and small att sites (<50 bp). This DNA recombination is also highly directional, reversed only in the presence of a single accessory protein called a recombination directionality factor (RDF), and requires no cofactors to bind or bend DNA. In comparison, the tyrosine integrase A Int requires the host-encoded integration host factor (IHF) protein to bend DNA for integration, and additional excisionase (Xis) and Fis proteins, encoded by the phage and host respectively, for DNA excision.10 Serine integrases can catalyze recombination between att sites on linear or circular DNA substrates and, depending on the position and orientation of the att sites, integrate, excise or invert sections of DNA (Figure 1).11 In addition, serine integrases are active in a range of organisms including bacteria, yeast, flies, zebrafish, frogs, plants, and cell lines including mouse, rabbit, chicken, bovine, and human, reviewed elsewhere.12,13 The reader is referred to related reviews on the tyrosine and serine recombinase families,9 their development as tools,11,13 serine integrase structures and mechanisms,14,15 and their application in heterologous systems.12,13 This paper reviews established and emerging serine integrase technologies currently being applied to genome engineering, in vitro DNA assembly, and in the development of genetic logic and data storage devices.

SERINE INTEGRASE SITE-SPECIFIC RECOMBINATION

Serine integrases are encoded by temperate bacteriophages and catalyze their integration into bacterial genomes through recombination of attP (phage) and attB (bacteria) attachment sites, generating attL (left) and attR (right) sites. Upon lysogeny, the prophage is excised from the host chromosome by the serine integrase and its cognate RDF. Importantly, recombination of attP and attB is highly directional, it is not reversible in the presence of the serine integrase alone. The
sequences of attP and attB are distinct from each other, and from attL and attR sites. Based on the crystal structure of the C-terminal domain of LI integrase bound to half of an attP site, Rutherford et al. proposed that the integrase adopts different spatial arrangements when bound to different att sites.16 From this, they formed a structure-based model for understanding why integrases preferentially form complexes with attP and attB sites, and prohibit attL × attR recombination.14–16

To date, over 4000 putative serine integrases have been identified,17 14 of which have been studied biochemically.14 Mechanistically, serine integrase dimers bind to attP and attB sites that are brought together to form a tetrameric synaptic complex (Figure 2). Serine residues cleave the att sites on either side of their central dinucleotide generating half sites with 3’ 2 base pair overhangs. Next, integrase subunits rotate in relation to each other, effectively swapping half sites.16,19 Finally, if the 2 bp overhangs on the swapped half sites are complementary, they will religate to make attL and attR sites. Pairs of attP and attB sites with nonpalindromic central dinucleotide sequences that do not share inverted complementarity (TT, CT, GT, CA, CC, and TC) are considered to be orthogonal. This means that in the presence of one serine integrase, as many as six pairs of these orthogonal att sites can recombine (attPTT will specifically recombine with attBTT, attBTC will specifically recombine with attBCT, and so on).20,21 Such orthogonality can also be applied to attL and attR sites.20

RDFs are phage-encoded proteins that bind integrase proteins, stimulate recombination of attL and attR sites, and inhibit recombination of attP and attB sites.16,22–24 Each integrase subunit interacts with an RDF to trigger recombination of attL and attR sites, and it is hypothesized that this interaction induces conformational changes in the integrase that inhibits attB × attP recombination.16,25,26 Recently, these protein interactions and stoichiometries were optimized through fusion of phiC31 and Bxb1 integrases with their cognate RDFs, gp3 and gp47.27 Among the nine RDFs identified to date, there is considerable variation in peptide length, sequence conservation, and location within their phage genome.25–24,28–33

**Figure 1.** Examples of DNA rearrangement mediated by serine integrases and RDFs. In the presence of a serine integrase, attP and attB sites recombine to make attL and attR sites. In the presence of a cognate RDF, this reaction is reversed. For integration and excision of DNA, att sites are in the same orientation. For inversion of DNA, att sites are in opposite orientations. (a) DNA integration mediated by attP × attB recombination, and excision by attL × attR recombination. (b) Integration mediated by attL × attR recombination, and excision by attP × attB recombination. (c) Recombination between att sites in linear DNA. (d) DNA inversion through attP × attB recombination, and reversed by attL × attR recombination.

**Figure 2.** Mechanism of site-specific recombination by serine integrases. (i) A serine integrase binds its cognate attP and attB sites as a dimer before forming a protein tetramer in which the att sites are brought together. (ii) The four serine integrase protomers are activated to cleave the att sites generating half sites with 3’ overhangs, 2 base pairs in length. At this point each DNA half site is attached to a protomer via a covalent 5’ phosphoserine linkage. (iii) Two of the serine integrase subunits rotate in relation to the others swapping attP and attB half-sites. (iv) If the 3’ overhangs of the half-sites are complementary, they will religate forming attL and attR sites. The C-terminal domains of serine integrase protomers are pale gray, attP half-sites are pale pink, and attB half-sites are dark pink.

**GENOME ENGINEERING WITH SERINE INTEGRASES**

In the 30 years since serine integrases were first used to engineer streptomycese genomes at innate attB sites,34,35 these enzymes have been shown to be active in a number of organisms and have been engineered for enhanced in vivo activity.17 For example, enhanced phiC31 integrase activity was achieved in Chinese hamster ovary (CHO) cells with the addition of a C-terminal nuclear localization signal,16 and a 3’ polyadenylation signal has been adopted to stabilize phiC31 integrase mRNA injected into Drosophila embryos.36 Chromosomal insertion of transgenes using serine integrases offers a number of advantages over alternative genome engineering methods. In comparison to homologous recombination-based methods including those that employ homing endonucleases, transcription activator-like effector nucleases and CRISPR-Cas9, integration is mediated by only one enzyme without relying on host factors for DNA integration, and serine integrases do not degrade DNA or need to be engineered to bind a specific DNA sequence.5 Unlike transposons and retroviruses, integration can be targeted to a specific locus known to have minimal positional effects on transgene expression.38 Transgenes integrated by serine integrases though attP × attB recombination cannot be inverted or remobilized without the presence of a cognate RDF. Genome engineering approaches that use serine integrases can be broadly grouped into those that integrate...
Figure 3. Serine integrase-mediated genome engineering strategies. (a) Genomic integration of DNA at a single att site with a serine integrase. A donor vector carries a gene of interest, a marker gene such as an antibiotic resistance gene, and an attP site. An attP site is present in the genome. The attB site in the donor vector recombines with an attP site in the genome in the presence of the integrase. This integrates the whole vector into this genomic locus where it is flanked by attR and attL sites. Integration of the vector can be detected by selection for the marker gene. (b) Jump-In GripTite Targeted Integration. The R4 integrase attP site and a blasticidin resistance gene (bsr) are preintegrated in HEK293 genomic DNA at a known genomic locus. In this example, an donor vector carrying a gene of interest driven by Promoter 1, an R4 attB site, and Promoter 2, is integrated into the genome through attP × attB recombination in the presence of R4 integrase, expressed from a second vector. Once integrated, Promoter 2 drives expression of a blasticidin resistance gene enabling selection for integration events. (c) Recombinase-mediated cassette exchange (RMCE) with a serine integrase. A donor vector carries a gene of interest between two attP sites. A marker gene, such as one encoding a fluorescent protein, is present between two attP sites. Integrate mediates two attP × attB recombination reactions that result in exchange of the marker gene in the genome with the gene of interest. This exchange stops expression of the marker gene enabling screening for integration events. Once integrated, the gene of interest is flanked with two att sites. (d) Dual integrase cassette exchange (DICE). A DNA cassette carrying genes for neomycin resistance (neo) and green fluorescent protein (GFP) flanked by attP sites for phiC31 and Bxb1 integrases is inserted into the genome of embryonic stem cells or induced pluripotent stem cells by homologous recombination. A donor vector carries a puromycin resistance gene (puroR), a red fluorescent protein (mCherry), and a gene of interest flanked by phiC31 and Bxb1 attB sites. In the presence of phiC31 and Bxb1 integrases, expressed from separate expression vectors, the attP and attB sites recombine to remove the neo-GFP cassette from the genome, replacing it with the puroR-mCherry-GOI cassette flanked by phiC31 attR and Bxb1 attL sites. GOI stands for gene of interest, R4 att sites are orange, phiC31 att sites are pink, Bxb1 att sites are green, attP half-sites are pale, and attB half-sites are dark.
different genomic loci have been made and characterized for positional effects on transgene expression, and integration of a large vector (133 kb) has been achieved in the presence of phiC31 integrase expressed from injected mRNA. For mammalian cells, the Jump-In Targeted Integration System from ThermoFisher Scientific offers CHO-K1 and human embryonic kidney (HEK) 293 cells with preintegrated attP sites for R4 integrase at known genomic loci. A donor vector, carrying a complementary attB site and a gene of interest, is integrated into the genome in the presence of R4 integrase, expressed from a second vector. GripTite is a version of Jump-In in which the expression vector carries a promoter to activate an antibiotic resistance gene located in the target genome next to the attP site (Figure 3b). A benefit of landing pads is that the genomic loci is known prior to insertion of transgenes. However, as is the case for pseudo sites, a strategy is required to detect integration, often requiring incorporation of extra DNA for selection, and an additional strategy may also be required to excise excess donor vector DNA from the genome.

RMCE is a targeted genome integration strategy originally established with tyrosine recombinases in which two recombination events mediate integration of a desired transgene and excision of the donor backbone (Figure 3c). In RMCE, a target site consisting of a selection marker flanked by two noncomplementary att sites is inserted at a genomic locus through transposition or homologous recombination. A vector containing the gene of interest flanked with att sites complementary to those in the target site undergoes two recombination reactions with the target site in the presence of the recombinase. Serine integrases are ideal candidates to mediate RMCE because recombination of attP and attB sites is highly directional, the att sites are small, and transgenes can be integrated from linearized or supercoiled DNA. In comparison experiments, Bxb1 and phiBT1 integrases were the most efficient serine integrases for RMCE in human HT1080 cells and Saccharomyces cerevisiae, respectively. The Drosophila research community has used a phiC31 integrase-mediated RMCE approach extensively to study transgene expression from defined genomic loci.

DNA ASSEMBLY BY SITE-SPECIFIC RECOMBINATION

Since the introduction of BioBricks 15 years ago, innovative DNA assembly strategies have been fundamental to progress in synthetic biology including construction of the genome of the first synthetic organism, metabolic engineering of microbes for production of high-value compounds, establishment of DNA registries for the distribution of reusable DNA parts, and cloning of repetitive DNA sequences. Current DNA assembly methods can be grouped into three categories: endonuclease-mediated assembly methods including BioBrick-based methods, Golden Gate-based Assembly, homology-based methods including sequence and ligation independent cloning (SLIC), Gibson Assembly, DNA Assembly in yeast, and PCR-based methods like circular polymerase extension cloning (CPEC) and PaperClip. Site-specific recombination-based methods including Gateway system, site-specific recombination-based tandem assembly (SSTRA) and serine integrase recombinational assembly (SIRA). The main complication of endonuclease-based methods is the requirement to remove incompatible restriction sites from DNA parts prior to assembly, meaning that often DNA parts must be synthesized, or precloned and mutated. Limitations of in vitro homology-based methods are associated with the size of DNA parts: both SLIC and Gibson Assembly use exonuclease activity to generate single-strand complementary overhangs on DNA parts meaning they have the potential to degrade small DNA parts; and PCR-based methods are limited to DNA parts that are short enough to be amplified in vitro with a DNA polymerase. Although these methods form no defined scar sequences, mutation events can occur at junctions between DNA parts, particularly in repetitive sequences. Also, once assembled, the junctions between DNA parts do not allow individual sections of DNA to be easily exchanged. Site-specific recombination-based DNA assembly methods recombine att sites that flank DNA parts. Gateway Cloning uses the tyrosine integrase from phage λ, together with an accessory protein, IHF, to clone DNA parts flanked with orthogonal attB sites into donor vectors containing complementary attP sites, through attP × attB recombination. This produces entry vectors in which DNA parts are flanked with attL sites. From this point, the DNA part can be transferred into expression vectors containing complementary attR sites through recombination mediated by λ integrase, IHF and an excisionase protein that reverses the direction of the recombination reaction, enabling attL × attR recombination. Additional orthogonal att sites have enabled Gateway assembly of up to five DNA parts in a single reaction and, due to the controllable directionality of the recombination reaction, it is possible to modify a Gateway construct postassembly. However, pre-cloning of DNA parts.
into donor and entry vectors is time-consuming, and the att sites to which IHF must bind and bend are large—the Gateway attP site is 200 bp.\textsuperscript{20} Compared to those recognized by tyrosine integrases, att sites for serine integrases are small (<50 bp), and require no accessory proteins to bind or bend them for recombination. It is therefore possible to recombine serine integrase att sites on linear pieces of DNA, eliminating the need to preclone DNA parts into entry vectors. Zhang et al. developed SSTRA which uses linearized DNA parts flanked with attB site and an attP site oriented in the same direction and mutated to recombine orthogonally.\textsuperscript{88,90} SSTRA has been used to assemble a functional carotenoid biosynthetic pathway, and a 56 kb construct of DNA sites in the same orientation.\textsuperscript{88,90}

This facilitates targeted postassembly modification of an assembled construct for replacement and addition of more DNA parts (Figure 4b). A key feature of SIRA is that its design principles can be applied to other serine integrases that have the potential to be used in combination for DNA assembly and modification in vitro or in vivo.\textsuperscript{20}
While it has been demonstrated that SSTRA and SIRA are reliable methods for cloning genes and assembling metabolic pathways, both create constructs interspersed with \textit{att} sites deeming them unsuitable for applications such as directed evolution of proteins. It is also worth noting that low levels of recombination have been observed between orthogonal \textit{att} sites under selective pressure, and that it is possible that \textit{att} sites within constructs could affect gene expression levels.\cite{4,5}

## SERINE INTEGRASE LOGIC AND MEMORY DEVICES

Much work in synthetic biology has been inspired by the goal to create biological systems that can record specific input signals, perform computations, and respond with outputs that can be applied in biotechnological, environmental and therapeutic contexts. To this end, a number of biological components including logic and memory devices have been developed based on transcriptional modules, RNA-based post-transcriptional regulation, and protein interactions (reviewed elsewhere). However, in order to maintain a steady state, these systems rely on levels of transcription that could impose considerable resource burden on the host cell by redirecting nucleic acids, amino acids and other metabolites away from native cellular processes. This could cause evolutionary counter-selection of such systems, or render them subject to changes in cellular metabolism resulting in spontaneous alterations in device outputs. As applications of these biological computation systems become more complex, so too does the requirement for robust, scalable components that are genetically stable and cause minimal resource burden.

Recombinase-based logic and memory devices respond to inputs by rearranging DNA. This means that the presence of an input can be recorded as a permanent, heritable DNA rearrangement from a single pulse of recombinase expression.

---

**Figure 5.** Serine integrase logic and memory devices. (a) The AND gate based on flipping a promoter and a gene. Int 1 (phiC31 integrase) and Int 2 (Bxb1 integrase) are expressed from promoters that are activated by input signals A and B, respectively. After both input signals have been present, the expressed integrases recognize and recombine their cognate sites and flip the promoter and the \textit{gfp} gene to the same orientation, resulting in the expression of the reporter protein, \textit{GFP}.\cite{93} (b) The framework of recombinase-based state machines (RSMs). A RSM requires a DNA register consisting of orthogonal and interleaving pairs of \textit{att} sites, and their corresponding integrases which are expressed in response to input signals. Three chemical signals are used to independently trigger expression of integrases: Input A for Int 1 (Bxb1 integrase); Input B for Int 2 (TP901 integrase); Input 3 for Int 3 (A118 integrase). The integrases rearrange the DNA register by inverting and excising sections of DNA. The table shows that each possible temporal order of input signals results in a DNA register with a unique sequence.\cite{21} (c) The rewritable recombinase addressable data (RAD) module. Int 1 (Bxb1 integrase) recombines \textit{attP} and \textit{attB} sites in opposite orientations to generate \textit{attL} and \textit{attR} sites. This inverts the intervening constitutive promoter, switching on expression of \textit{RFP} and switching off expression of \textit{GFP}. In the presence of the RDF (gp47), Bxb1 integrase recombines the \textit{attL} and \textit{attR} sites to generate \textit{attB} and \textit{attP} sites. This restores the constitutive promoter to the original state, switching on expression of \textit{GFP} and switching off expression of \textit{RFP}.\cite{96} The \textit{att} sites are blue for Bxb1 integrase; pink for phiC31 integrase; green for TP901 integrase; and purple for A118 integrase. Darker shades indicate \textit{attB} half sites, and lighter shades indicate \textit{attP} half sites. Different shapes are used to distinguish between \textit{att} sites with orthogonal dinucleotides.
and therefore causes minimal resource burden. Such systems have been described using tyrosine recombinases such as FimE invertase which irreversibly flips the orientation of specified DNA sequences, and combinations of tyrosine and serine recombinases. A system recently developed in mammalian cells, Boolean logic and arithmetic through DNA excision (BLADE), uses tyrosine recombinases to excise transcription terminators located upstream of genes, and serine recombinases to invert genes downstream of a stationary promoter, analogous to BUF (buffer) gates. Using an array of four of these BUF gates, Weinberg et al. built a field-programmable, read-only memory circuit that can be programmed with combinations of four recombinases to act as any of 16 2-input Boolean logic gates.

In earlier work, Siuti et al. elegantly demonstrated the potential of using simply phiC31 and Bxb1 serine integrases to carry out logic functions in E. coli by constructing all 16 two-input Boolean logic gates in DNA with promoters, terminators and genes that could be inverted by just attP × attB recombination (Figure 5a). The AND gate maintained a stable output memory for over 90 generations after input signals had been removed, and this output could be detected by PCR after cell death. Furthermore, this study produced a suite of two-bit digital-to-analog converters in which two GFP expression cassettes could be flipped by two integrases, respectively, to be expressed from different constitutive promoters of different strengths. Different combinations of two inducers (digital input) triggered expression of either, both or no integrases, and therefore produced four different levels of GFP expression ranging from low to high (analog output).

Independently, the Endy group developed permanent, amplifying AND, NAND, OR, NOR, XOR, and XNOR transcription-based gates using two serine integrases. These used the inversion or excision functions of TP901 and Bxb1 inteinase of these two serine inteinases to rearrange the orienta and downstream of regulatory DNA sequences, thereby controlling the flow of RNA polymerase along DNA and thus amplifying expression of a reporter gene. In recognition of the fact that a small change in input signal can trigger an all-or-nothing effect on recombination and therefore output, the group endeavored to compare gate output levels (GFP expression) to increasing levels of input signals (inducers of inteinase expression) to find a combination of input levels that would provide a threshold above which GFP expression was “on” and below which it was “off”. By producing binary outputs (i.e., expression of GFP is either switched on or off) in response to low or high inputs, these transcription gates act as analog-to-digital converters. One such transcription-based AND gate was later shown to enable a bacterial whole-cell biosensor to execute a signal-processing algorithm in clinical samples demonstrating a new approach to medical diagnosis.

In recent years, a number of studies have contributed to increasing the information storage capacity of integrase-based logic and memory devices. Using a bioinformatic approach, Yang et al. mined 4105 putative large serine-type phage inteinases from prophage genomes and then, using att sites for 11 inteinases that shared a maximum of 60% sequence identity, built an 11-bit array consisting of 11 distinct segments of DNA that could be flipped in orientation. This DNA array was 2 kb in size and capable of storing 1.375 bytes of data. In addition to recording the presence of biological inputs, memory devices with overlapping switch modules for different inteinases can record the temporal order of these input signals. This was first realized when the inversion systems of fim and hin tyrosine and serine recombinases were integrated to create a heritable sequential memory switch with as many as four states.

More recently, the Lu group created recomb inase-based state machines (RSMs) which use the inversion and excision functions of multiple input-driven serine inteinases to rearrange DNA registers composed of orthogonal and interleaving pairs of att sites (Figure 5b). In this framework, DNA registers are rearranged into completely different conformations depending on the temporal order of input signals. Using TP901, Bxb1, and A118 inteinases, and multiple orthogonal pairs of att sites, Roquet et al. demonstrated that the information capacity of this framework could be scaled to that of a 16-state RSM and, by incorporating regulatory elements and fluorescent protein genes, built functional 16-state gene-regulatory RSMs (GRSMs).

An intrinsic issue of logic and memory devices that rely on the switching of states is noise generated by heterogeneity in biological populations. Taking this into account, Hsiao et al. developed a two-input, serine inteinase-based temporal logic gate containing att sites for Bxb1 and TP901–1 integrates, which was integrated into the chromosome of a strain of E. coli. By analyzing the heterogeneity of individual cellular states generated in response to different sequences of chemical inputs, applying a stochastic model of single-cell trajectories, and amalgamating these into population-level distributions, this group was able to determine the order, timing and duration of input signals. A mechanistic model of this two-input temporal logic gate has been developed.

The serine-inteisate logic and memory devices discussed up to this point rearrange DNA to produce stable, heritable, permanent outputs. While such devices are ideal for recording one-off historical events such as the presence of an environmental signal, rewritable devices are being developed to enable counting of real-time, dynamic biological processes such as cell divisions. The ability to stringently control the direction of recombination by serine inteinases with RDFs makes them ideal mediators to set and reset the orientation of DNA in a controlled manner, thus increasing the data storage capacity of memory devices and enabling reuse of logic gates. The Endy group applied serine inteinases and their cognate RDFs to data storage by making a rewritable recombinase addressable data (RAD) module (Figure 5c). In this module, the orientation of a DNA sequence flanked by attP and attB sites was flipped by Bxb1 inteinase setting the module to “state 1” in which the DNA sequence is flanked by attL and attR sites. Upon induction of the cognate RDF, gp47, the DNA sequence was reset to its original orientation, state “0”, by attL × attR recombination. A mechanistic model of a rewritable RAD module that uses phiC31 inteinase and its cognate RDF, gp3, has been described. It has been proposed that an N-bit counter built from 2N RAD modules could count up to 2N states, limited only by the number of known orthogonal serine inteinases with cognate RDFs.

In contrast to recombinase-based memory devices, a number of recent studies have adopted CRISPR-Cas9 systems to record the presence of biological signals. Temporal recording in arrays by CRISPR expansion (TRACE) uses CRISPR as a biological tape to record cellular events over time. In this system, detection of a biological signal triggers an increase in the copy number of a reporter plasmid. The increase in the presence of this plasmid is recorded as increased incorporation of corresponding spacers into CRISPR arrays across a population of cells. However, data storage capacity is limited by the slow...
rate of spacer incorporation (~1 in 15 arrays incorporated a new spacer each day) and an analytical model is required to interpret CRISPR sequences across the population of cells. Mammalian synthetic cellular recorders integrating biological events (mSCRIBE) from the Lu lab,108 and homing CRISPR barcodes from the Church lab,109 use self-targeting guide RNAs (sgRNAs)108 (or homing guide RNAs100), which direct a Cas9 endonuclease to cleave loci that encode the sgRNAs. This induces nonhomologous end joining (NHEJ), which introduces mutations to the sgRNAs. Over time, these accumulating mutations form a record of analog data, such as fluctuations in biological signals, and computational metrics are applied to determine the duration and magnitude of the input signal, or to trace cell lineage. Limitations of these approaches are associated with in situ sequencing for analysis of a mixed population of cells and the restricted diversity that can be generated in the guide RNAs. Some mutations in these systems occur with a higher frequency than others,108 and NHEJ can repair DNA cleavage without introducing mutations at all. It is also not known if Cas9 would have any off-target activity, or if integrated arrays of nuclease targets would affect genome stability.

Compared to transcription-based biological logic and memory systems including those that use CRISPR-Cas9 components, serine integrase-based devices have a simple composition comprising of DNA, integrases and RDFs. As serine integrases and RDFs are active in a range of cell-types and do not rely on endogenous DNA repair systems for DNA modification, these devices could be made portable and utilized in multiple organisms. Data stored in DNA can be reliably retained throughout cell divisions, transferred across species, and easily detected after cell death. Importantly, these devices have the potential to detect transient signals and to record data in otherwise inaccessible environments.

PROSPECTS FOR SERINE INTEGRASES

The capacity of serine integrases to recombine DNA both in vitro and in a broad spectrum of in vivo systems, and to record the presence of biological signals and carry out logic functions, means they have the future potential to be applied in self-programming, decision-making biological machines. These machines could be applied in disease diagnosis, production and delivery of bespoke therapeutics, remote environment monitoring, universal bioremediation systems, and self-evolving microbial factories for biosynthesis of biomaterials and valuable compounds.

Taking into account some of the serine integrase-based systems discussed in this review, one could consider how such a machines could take shape. The RSM described above uses a DNA register composed of two pairs of overlapping and orthogonal att sites for three different integrases, that produces 16 different DNA sequence outputs in response to 16 different temporal input sequences (Figure 5b). If four such DNA registers, each responsive to three different integrases, were implemented side by side, this would generate a 12-integrase RSM with 65 536 potential temporal input sequences and the same number of DNA sequence outputs. Given that 11 serine integrases have been identified that recombine their own att sites without showing cross-reactivity with att sites of other integrases,17 it is not unreasonable to imagine this and larger RSMs will be constructed in the near future. In the long term, synthetic genomes could be constructed in a RSM-based structure to enable rapid evolution in a random manner similar to SCRAMbLE66,67 (all integrases triggered at once), or in a targeted, finely tuned manner (in response to a distinct temporal sequence of inputs). It is also realistic to expect that RDFs will become key components of RSM-like systems for rewritable logic and counting functions, and for incorporation of additional DNA sequences. This would utilize processes not dissimilar to targeted postassembly modification with SIRA (targeting pairs of att sites with different central dinucleotides) (Figure 4b), DICE (targeting two att sites for different integrases) (Figure 3d), and DNA integration into a single att site (Figure 3a).

Progress toward realization of such multifunctional systems is largely restrained by one key bottleneck: the current collection of serine integrase-related tools is small. To the best of our knowledge, there are currently 14 biochemically characterized serine integrases,14 nine RDFs have been identified,22−24,28−33 and each integrase has a potential set of six orthogonal att sites.20 The first steps toward expanding this collection of tools would be to characterize the 410S recently identified putative large serine-type phage integrases, and to develop bioinformatic solutions to facilitate mining of phage genomes for potential RDFs (this is currently a challenge because known RDFS differ considerably in sequence, size and genomic location). Ideally, each integrase and RDF characterized should be assessed for cross-reactivity with att sites corresponding to other integrases and RDFs. Logical progression from integrase and RDF discovery would be generation of more integrase-RDF chimeras to increase the stringency of attL × attR recombination reactions.27

There are a number of reasons in favor of expanding the repertoire of available att sites. First, the more orthogonal att sites an integrase could target, the more scalable its applications would be. Second, it has been noted that some att sites could have an effect on mRNA stability or transcription initiation rates of reporter genes.33,93,94 Finally, as previously mentioned, selective pressure can result in low levels of cross-reactivity between orthogonal att sites with different central dinucleotide sequences.20 Generation of additional orthogonal att sites would likely entail engineering of serine integrases to recognize and bind to new DNA sequences. Farraggio et al. created phiBT1-phiC31 and phiC31-TG1 hybrid integrases by joining their N-terminal and C-terminal portions, and these were active in E. coli on att sites that were hybrids between those of the donor enzymes.116 It is currently not known if recombination of these att sites could be reversed in the presence of an RDF. Several “hyperactive” small serine recombinase mutants that act independently of their native recombination regulatory systems have been engineered to recognize non-native DNA sequences through replacement of their C-terminal DNA-binding domains with DNA-binding zinc finger domains, or complexes of dead Cas9 and guide RNAs.111−114 In contrast, the C-terminal domains of serine integrases are much larger than those of small serine recombinases, and consist of multiple structural domains that are involved in regulating att site recognition and recombination directionality.16,115,116 Given that deletion of the coiled-coil motif from the C-terminal of LI integrase renders DNA integration by attP × attB recombination negligible, and facilitates excision through attP × attB, attL × attR, attP × attP, and attB × attB recombination,16 it is plausible that replacing serine integrase C-terminal domains with alternative DNA binding strategies would compromise their functionality.
In conclusion, serine integrases, and their cognate RDFs and att sites, provide reliable DNA recombination systems that can be used in a number of applications including in vivo genomic integration of DNA in a range of organisms, and in vitro assembly and targeted modification of DNA constructs. These recombination systems have also been applied in low-burden, heritable DNA logic and memory devices that can store information permanently or in a rewritable fashion. As many serine integrases work independently on distinct att sites, there is potential to scale up their applications by using multiple recombination systems together. This is currently limited to the relatively small number of biochemically characterized serine integrases, RDFs and att sites available. Future advances in serine integrase technologies would benefit greatly from development of new strategies for identification and characterization of these components, and greater understanding of how they interact and regulate recombination.

■ AUTHOR INFORMATION

Corresponding Author
E-mail: christine.merrick@ed.ac.uk. Tel: 0131 650 5416.

ORCID
Christine A. Merrick: 0000-0002-6464-1224

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Funding was provided by the UK Engineering and Physical Sciences Research Council (EP/H019154/1 and EP/K034359/1).

■ ABBREVIATIONS

RDF, recombination directionality factor; CRISPR, clustered regularly interspaced short palindromic repeat; RMCE, recombinase-mediated cassette exchange; CHO, Chinese hamster ovary; HEK, human embryonic kidney; DICE, Dual recombinase-mediated cassette exchange; SCRaMbLE, Synthetic Chromosome Rearrangement and Modification by LoxPsym-mediated Evolution; SLIC, sequence and ligation independent cloning; CPEC, circular polymerase extension cloning; STRA, site-specific recombination-based tandem assembly; SIRA, serine integrase recombinational assembly; BLADE, Boolean logic and arithmetic through DNA excision; RAD, recombinase addressable data; RSM, recombinase-based state machines; GRSM, gene-regulatory recombinase-based state machines; TRACE, temporal recording in arrays by CRISPR expansion; mSCRIBE, mammalian synthetic cellular recorders integrating biological events.; stgRNA, self-targeting guide RNAs; NHEJ, non-homologous end joining

■ REFERENCES


DOI: 10.1021/acssynbio.7b00308

ACS Synth. Biol. XXX, XXX, XXX–XXX


