scientific reports

Check for updates

OPEN

Variable orthogonality of serine integrase interactions within the ϕC31 family

Alasdair I. MacDonald[1], Aron Baksh[2], Alexandria Holland[2], Heewhan Shin[3], Phoebe A. Rice[3], W. Marshall Stark[1] & Femi J. Olorunniji[2]✉

Serine integrases are phage- (or mobile element-) encoded enzymes that catalyse site-specific recombination reactions between a short DNA sequence on the phage genome (*attP*) and a corresponding host genome sequence (*attB*), thereby integrating the phage DNA into the host genome. Each integrase has its unique pair of *attP* and *attB* sites, a feature that allows them to be used as orthogonal tools for genome modification applications. In the presence of a second protein, the Recombination Directionality Factor (RDF), integrase catalyses the reverse excisive reaction, generating new recombination sites, *attR* and *attL*. In addition to promoting *attR* x *attL* reaction, the RDF inhibits *attP* x *attB* recombination. This feature makes the directionality of integrase reactions programmable, allowing them to be useful for building synthetic biology devices. In this report, we describe the degree of orthogonality of both integrative and excisive reactions for three related integrases (ϕC31, ϕBT1, and TG1) and their RDFs. Among these, TG1 integrase is the most active, showing near complete recombination in both *attP* x *attB* and *attR* x *attL* reactions, and the most directional in the presence of its RDF. Our findings show that there is varying orthogonality among these three integrases – RDF pairs. ϕC31 integrase was the least selective, with all three RDFs activating it for *attR* x *attL* recombination. Similarly, ϕC31 RDF was the least effective among the three RDFs in promoting the excisive activities of the integrases, including its cognate ϕC31 integrase. ϕBT1 and TG1 RDFs were noticeably more effective than ϕC31 RDF at inhibiting *attP* x *attB* recombination by their respective integrases, making them more suitable for building reversible genetic switches. AlphaFold-Multimer predicts very similar structural interactions between each cognate integrase – RDF pair. The binding surface on the RDF is much more conserved than the binding surface on the integrase, an indication that specificity is determined more by the integrase than the RDF. Overall, the observed weak integrase/RDF orthogonality across the three enzymes emphasizes the need for identifying and characterizing more integrase – RDF pairs. Additionally, the ability of a particular integrase's preferred reaction direction to be controlled to varying degrees by non-cognate RDFs provides a path to tunable, non-binary genetic switches.

Keywords Large serine recombinases, Serine integrase, Site-specific recombination, Alphafold multimer.

Large serine recombinases (LSRs) or serine integrases catalyse site-specific recombination reactions between short DNA sequences on temperate phages (*attP*) and equivalent sequences on the genome of their bacterial hosts (*attB*)[1,2]. The reaction results in the generation of new sequences called *attR* and *attL* flanking the prophage integrated into the host genome. In the reverse reaction, another phage-encoded protein called the recombination directionality factor (RDF) binds to the integrase and switches its specificity, allowing it to recombine *attR* and *attL*, regenerating *attP* and *attB* sites and excising the prophage DNA from the host genome (Fig. 1). The RDF also inhibits further *attP* x *attB* recombination. Although the presence or absence of the RDF determines which pairs of DNA sites the integrase can synapse and subsequently recombine, the details of how it functions are poorly understood. These two reactions are part of the natural lysis/lysogeny cycle of temperate phages. The mechanism of catalysis of recombination by serine integrases follows the pathway established for the serine recombinase family, in which the two recombining DNA sites are held together by a tetramer of recombinase protein subunits in a synaptic complex (Fig. 1). The recombination steps involving DNA cleavage, subunit

[1]School of Molecular Biosciences, University of Glasgow, Bower Building, Glasgow G12 8QQ, UK. [2]School of Pharmacy and Biomolecular Sciences, Faculty of Science, Liverpool John Moores University, James Parsons Building, Byrom Street, L3 3AF Liverpool, UK. [3]Department of Biochemistry and Molecular Biology, The University of Chicago, 60637 Chicago, IL, USA. ✉email: F.J.Olorunniji@ljmu.ac.uk

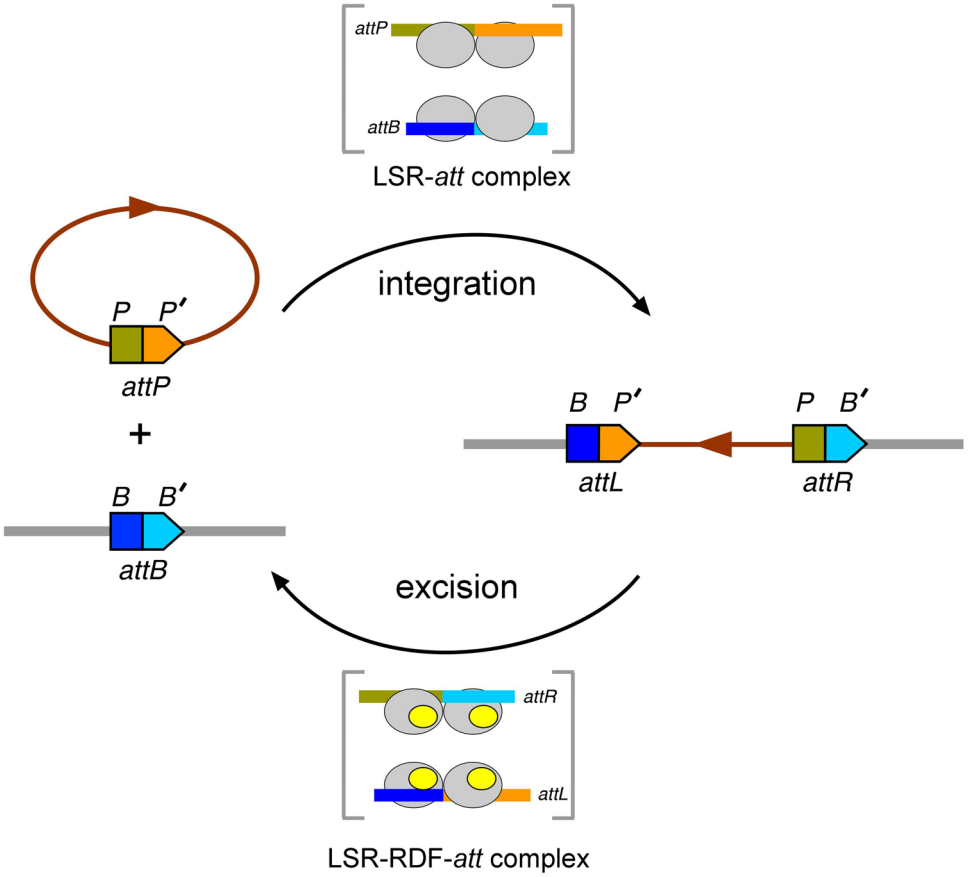


Figure 1. Integrative and excisive recombination reactions catalysed by large serine recombinases (LSRs). LSRs integrate DNA bearing an *attP* site into a genomic location that harbours an *attB* site. In the reverse (excision) reaction, recombination directionality factor (RDF, yellow ovals) binds to the LSR (grey ovals) and modifies its specificity to catalyse *attR* x *attL* recombination. In both reactions, synapsis of *att* sites and catalysis of DNA strand exchange happen within a synaptic complex involving a tetramer of the recombinase.

rotation, and DNA religation occur within the synaptic complex, ensuring tight regulation of the chemical events and conformational changes involved[3,4].

Due to the unidirectional nature of the integration reaction and the abundance of serine integrases, each with its own unique sequence specificity, this group of enzymes offer the potential for developing powerful genome editing tools[1,5]. Serine integrases carry out complete DNA cutting and rejoining in the recombination reaction without leaving any broken ends behind, and hence do not rely on endogenous host systems to fix broken DNA ends. Serine integrases have been adapted for a wide range of applications, including insertion of foreign DNA into specific sequences in the genomes of cells, plants and animals and the construction of genetic logic gates[1,2,6–12]. Some tools, such as the SIRA method for assembly of replicons, require a panel of serine integrases with orthogonal sequence specificity. The ability of an RDF to trigger reversal of a particular integrative recombination event renders serine integrases even more versatile as genetic tools.

ϕC31 integrase (605 amino acid residues) is the prototype serine integrase and was the first to be fully characterised in vivo and in vitro[13]. It is derived from the *Streptomyces* phage ϕC31, and it is used extensively for genetic manipulations in bacteria and several eukaryotic systems. TG1 integrase and ϕBT1 integrase are two other integrases derived from *Streptomyces*; their properties and applications have been reported in the literature[14–17]. The RDFs for all three of these integrases have been identified, and the activities for ϕC31 and ϕBT1 have been characterized both in vivo and in vitro[18,19].

These three related integrases and their RDFs show significant sequence similarities, yet they have different *att* site sequence specificities. Hence, they are ideal candidates for analysis of the structural and biochemical basis for integrase-*att* site recognition and catalysis. They also have potential as orthogonal tools for synthetic biology applications requiring multiple serine integrases.

TG1 integrase (619 amino acid residue) is derived from the TG1 actinophage isolated from *Streptomycetes*. The integrase gene, its recombination *attB* and *attP* sites, and its activity in *E. coli* were first described by Morita et al.[15]. The minimal *att* site requirements and in vitro activities were also reported by the same group[16].

Similarly, ϕBT1 integrase (594 amino acid residue) is derived from a phage from *Streptomyces coelicolor*[20], and its activity in vivo and in vitro as well as the minimal *attB* and *attP* sequences were established by Zhang et al.[14].

Zhang et al.[19]. found that the RDFs for ϕBT1 and ϕC31 integrases were fully exchangeable in in vitro recombination reactions, despite the two integrases sharing just 26% amino acid sequence identity. It was reasoned that this could be due to the 85% similarity between the two RDFs. In contrast, TG1-RDF shows 60% and 62% similarity to ϕBT1-RDF and ϕC31-RDF, respectively. To date, the effects of TG1-RDF on ϕBT1 and ϕC31 integrases have not been reported.

Since the three RDFs have regions of highly conserved sequences, they may cross-react with each other's integrases, as has been reported for ϕC31 and ϕBT1 RDFs[19]. Any significant cross-reactivity across these integrases will have implications for their simultaneous (or concurrent) use in in vivo and in vitro applications. However, a full analysis of the degree of cross-reactivity across these related integrase/RDF pairs has not been reported.

To clarify the extent to which the three integrases and their RDFs could be used together in synthetic biology applications, we investigated the orthogonality of the integrases and their RDFs in recombination reactions. The findings could help shed some light on the structural basis for integrase-RDF recognition and orthogonality. We also investigated whether the nature of the integrase-RDF specificities changes when the integrase and RDF are covalently joined as integrase-RDF fusions[21]. We anticipated that the findings would lay the foundation for further studies aimed at understanding the factors that determine specificity of integrase-RDF interactions.

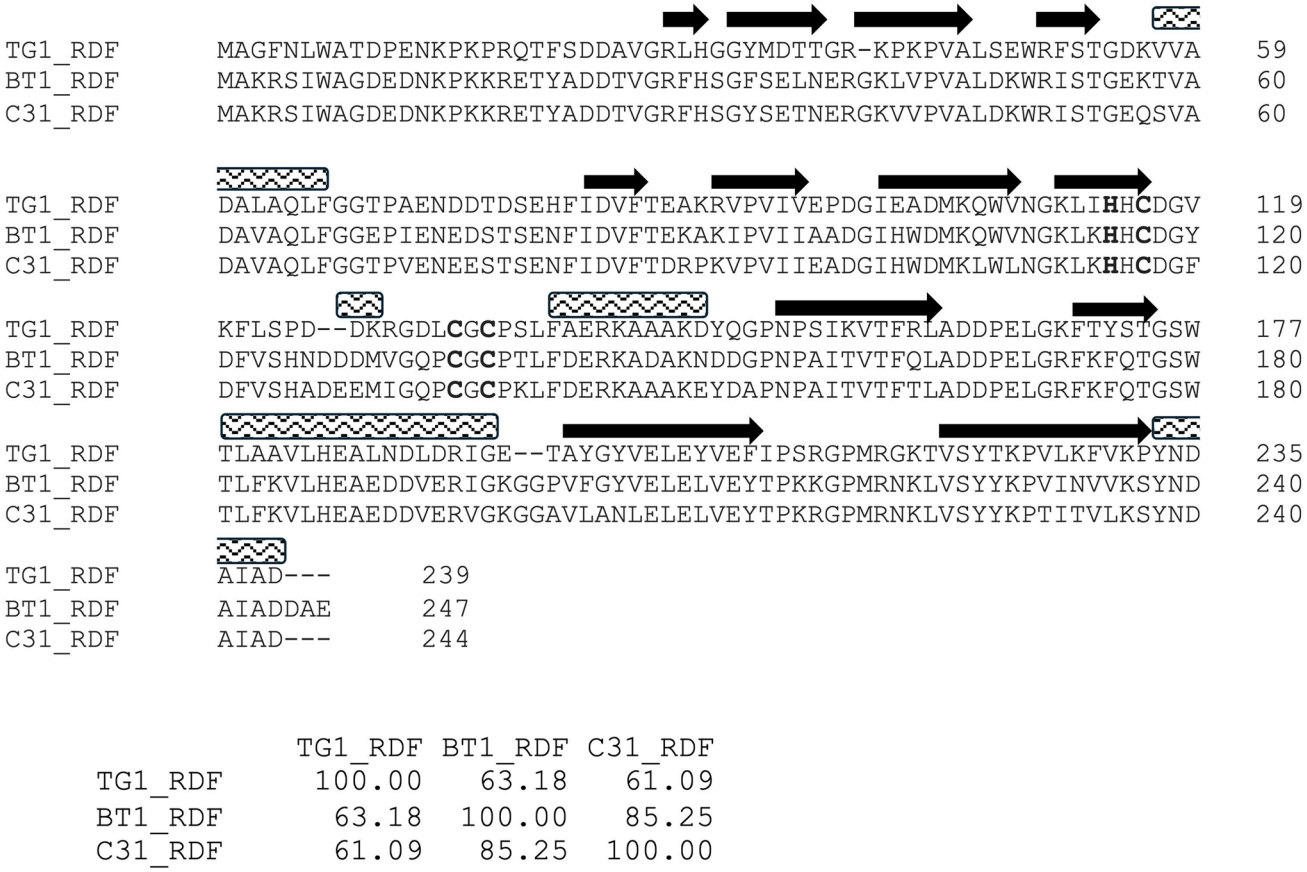
Results

Sequence alignment and predicted structures of ϕC31, ϕBT1, and TG1 integrases and their recombination directionality factors (RDFs)

Sequence alignments (Figs. 2 and 3) show that the three RDFs are more closely related than the corresponding RDF-binding domains on the integrases. TG1 integrase is only 24–25% identical to ϕBT1 and ϕC31 integrases, which are about 43% identical to one another. Similarly, among the RDFs, TG1 RDF is the outlier, with 61–63%

	TG1_Int	BT1_Int	C31_Int
TG1_Int	100.00	25.32	24.08
BT1_Int	25.32	100.00	43.39
C31_Int	24.08	43.39	100.00

Figure 2. Sequence alignment of the second DNA-binding domains (DBD2s; also known as zinc-binding domains or ZD) of ϕC31, ϕBT1, and TG1 integrases. Multiple sequence alignment was generated using Clustal Omega[22] and adjusted based on predicted structures. The four predicted Zn^{2+} binding residues are highlighted in bold. Residues highlighted in bold have the putative Zn^{2+} binding side chains. For ϕC31 and TG1, 5 rather than the expected 4 are highlighted due to ambiguity in the predicted structures. Bold and underlined DPDD residues form an insertion in TG1 integrase that interacts with its RDF. Arrows, beta strands; rectangles, helices. The coiled coil (CC) rectangles are shaded grey and labelled as shown. A matrix of pairwise percent identities of the DBD2 is shown below the alignment (https://www.ebi.ac.uk/jdispatcher/msa).



	TG1_RDF	BT1_RDF	C31_RDF
TG1_RDF	100.00	63.18	61.09
BT1_RDF	63.18	100.00	85.25
C31_RDF	61.09	85.25	100.00

Figure 3. Sequence alignments of ϕC31, ϕBT1, and TG1 Recombination Directionality Factors (RDFs). Multiple sequence alignment was generated using Clustal Omega[22] and adjusted based on predicted structures. The four predicted Zn^{2+} binding residues are highlighted in bold. Arrows = beta strands; rectangles = helices. A matrix of pairwise percent identities of the RDFs is shown below the alignment (https://www.ebi.ac.uk/jdispatcher/msa).

identity to the other two, which are 85% identical to one another. These observations suggest likely RDF cross-reactivity between ϕBT1 and ϕC31 integrases.

To understand the nature of integrase-RDF interactions, we used AlphaFold2-multimer to model the structures of the three integrases in complex with their cognate RDFs. As expected, the predicted structures of the two DNA-binding domains are very similar to the experimental structure of the DNA binding domains of LI-Int bound to half an *att*P site[23], which was used to model binding of our integrases to DNA. (Fig. 4). All three complex models are quite similar and predict that the RDF uses a set of loops to clamp onto a hinge region between the integrases' second DNA-binding domain (DBD2; sometimes called the zinc-binding domain or ZD) and the coiled coil that is inserted within it. (Fig. 4). These models are in good general agreement with prior experimental data[24–26]. The coiled coil is known to mediate synaptic contacts between paired *att* sites[18,23,24,26–28]. Our models suggest that RDF binding partially but not fully restrains the mobility of the coiled coil, but further structural work is needed to fully understand how such partial restraint controls reaction directionality.

Recombination activities of ϕC31, ϕBT1, and TG1 integrases

In vivo, all three integrases catalysed *attP* x *attB* recombination to near completion, and as expected did not act on their *attR* x *attL* substrate in the absence of the RDF (Fig. 5), showing the strict directionality as well as efficiency of the integration reactions. A similar pattern was observed in vitro, with recombination being generally efficient (Fig. 6). However, in vitro recombination did not go to completion after 2 h, with ϕC31, ϕBT1, and TG1 integrases converting 76%, 54%, and 92% of the substrate, respectively. We cannot differentiate from these data whether the lower amount of in vitro product for ϕC31 and ϕBT1 integrases can be ascribed to an intrinsically lower initial reaction rate or instability of the protein over 2 h under the conditions used. The higher completeness of recombination observed in vivo could be due to the continuous expression of the proteins over the 16-hour growth period. The high activity of TG1 integrase is particularly noticeable, outperforming ϕC31 integrase, a recombinase that has been used in several in vitro and in vivo applications. Overall, the activities of the three integrases are consistent with our findings reported in an earlier study where we compared the activities of 10 different integrases[29].

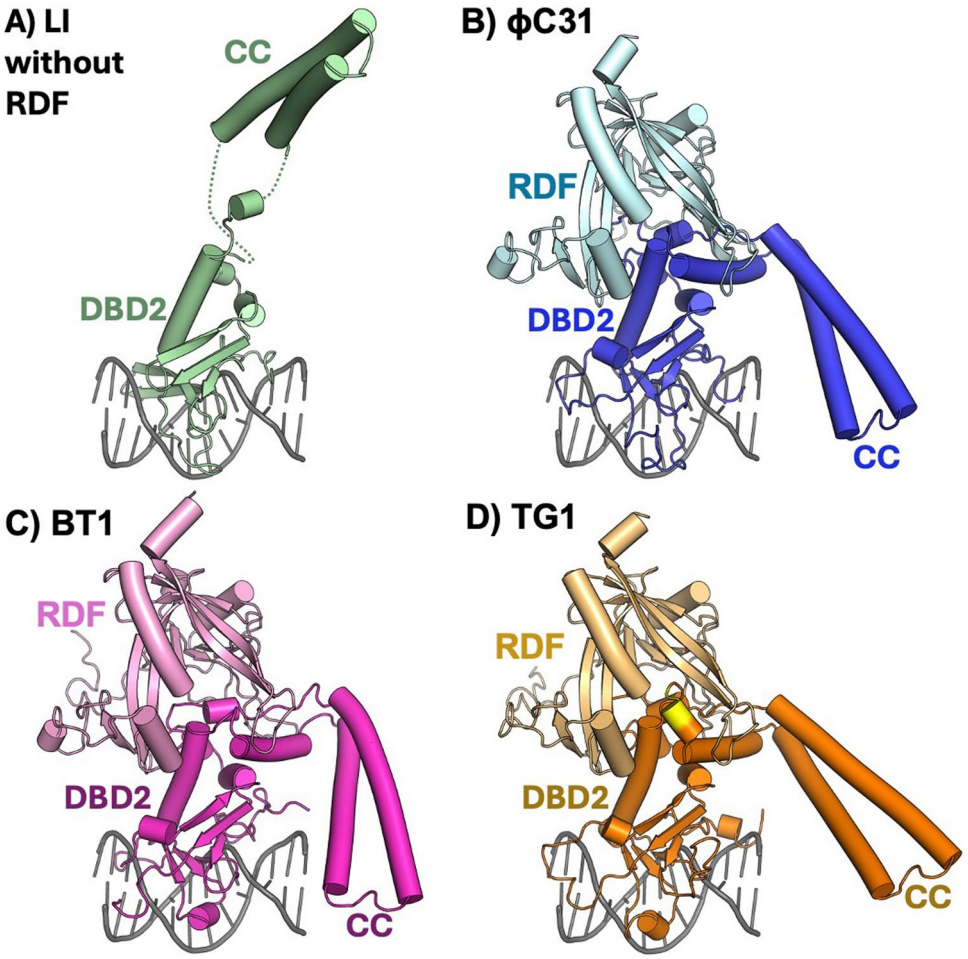


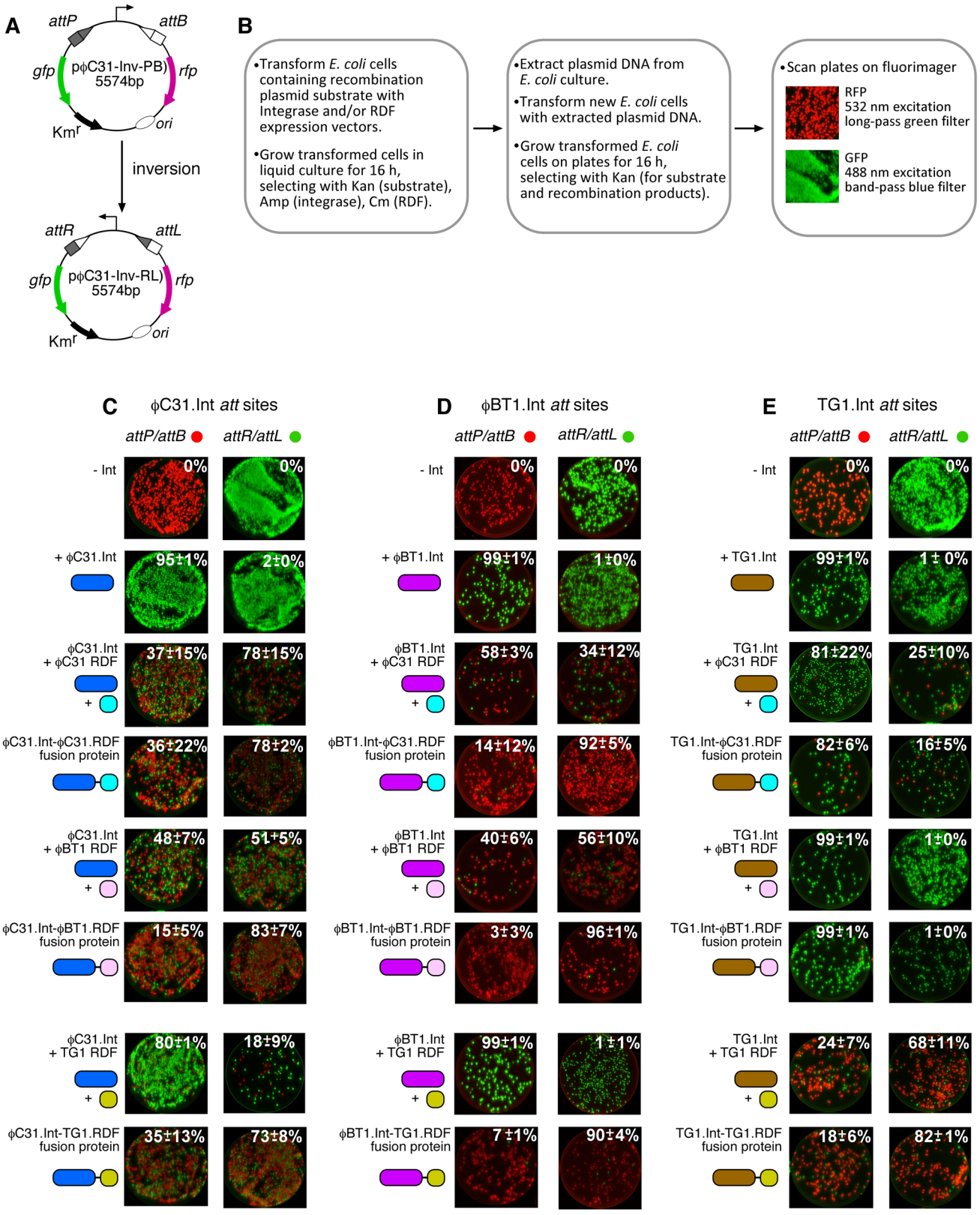
Figure 4. AlphaFold2-multimer structures of serine integrases and their RDFs. (**A**) The experimentally determined structure of LI integrase bound to A118 integrase *attP* in the absence of the RDF. Only the second DNA binding domain (DBD2; also known as zinc-binding domain or ZD) and the DNA it contacts are shown (4kis.pdb[23]). (**B**-**D**): AlphaFold2-multimer - predicted models for DBD2 of ϕC31 (blue), ϕBT1 (magenta), and TG1 (brown) integrases with their cognate RDFs (paler shades). The DNA from the crystal structure 4kis is shown with each model to guide the eye. Flexible C-terminal extensions of the integrases were removed for clarity. For TG1, the DPDD insertion is highlighted in yellow (see alignment in Fig. 2).

Effects of RDFs on the recombination activities of ϕC31, ϕBT1, and TG1 integrases

To study the specificity of integrase-RDF interactions across the three integrases, we studied the activities of each integrase in the presence of the three different RDFs both in vivo and in vitro. We did this in two ways: First, by using the integrase and the RDF as separate proteins, and secondly by constructing integrase-RDF fusions[21] to account for effects due to differential binding affinities of integrases for non-cognate RDFs. Use of fusions also avoids potential effects on activity due to differences in expression levels of the integrase and RDF proteins. In addition to activating *attR* x *attL* recombination, RDFs inhibit recombination of *attP* x *attB* by their respective integrases[18,30,31]. To see if there is a correlation between the degree of RDF-mediated activation of excisive recombination and inhibition of integrative recombination, we also investigated the inhibition of *attP* x *attB* recombination by the cognate RDF for the three integrases.

In vivo, the ϕBT1 RDF, when fused to ϕBT1 integrase (and to a lesser extent ϕC31 integrase), was the most effective in both activating excisive recombination and inhibiting integrative recombination (Fig. 5). Furthermore, plotting the reaction endpoints for all of the pairwise tests shown in Fig. 5 gives a strong anti-correlation between the endpoints of the *attP* x *attB* and the *attR* x *attL* reactions: the data plotted in Fig. 7a have a correlation coefficient of -0.99. This confirms that equilibrium was reached in these in vivo assays, and that the effectiveness of a particular RDF in promoting *attR* x *attL* reactions directly correlates with its effectiveness in inhibiting *attP* x *attB* reactions: each pair tested reached its particular "set point" regardless of the starting conditions.

In vitro, in the presence of their cognate RDFs, ϕC31, ϕBT1, and TG1 integrases recombined 76%, 29%, and 94% respectively of their *attR* x *attL* substrates. As for *attP* x *attB* recombination, TG1 integrase is the most active among the three, giving near complete conversion of the substrate plasmid (Fig. 6). The overall pattern emerging from this analysis is that TG1 integrase is the most active in both integrative and excisive reactions. Incomplete



reactions in vitro could be due to the factors discussed above for the *in vitro attP* x *attB* reactions, as well as weak directionality for the *attR* x *attL* reaction: that is, the equilibrium constant for the *attR* x *attL* reaction in the presence of RDF may not lie as far in favour of products vs. substrates as it does for the *attP* x *attB* reaction in the absence of RDF. This is supported in relation to the ϕC31 integrase by the conversion of 35–44% of the *attP* x *attB* substrate to product in the presence of the RDF (Fig. 6). In contrast, the in vitro data for ϕBT1 integrase in the presence of its RDF suggests that it simply did not reach equilibrium under the conditions used.

Figure 5. In vivo recombination reactions of ϕC31, ϕBT1, and TG1 integrases. (**A**) Scheme illustrating the in vivo intramolecular recombination (inversion) assay. In its default state, the promoter constitutively drives the expression of a red fluorescent protein (*rfp*) gene (pink arrow). A terminator sequence upstream of the promoter inhibits transcriptional read-through to the green fluorescent protein (*gfp*) gene (green arrow). Upon integrase-catalysed site-specific inversion reaction, the orientation of the promoter is flipped to allow the expression of GFP, and block RFP production. (**B**) Summary of the protocol for in vivo recombination reactions using constitutive integrase and RDF expression vectors. (**C**) Recombination activities of ϕC31 integrase in the presence of ϕC31-RDF, ϕBT1-RDF, and TG1-RDF; and as integrase-RDF fusions. In *attP*/*attB* reactions, cells start expressing RFP and produce GFP upon recombination. The extent of recombination is indicated as percentage of cells expressing GFP as outlined in **B** above. In each case, the value shown is the average and standard deviation of at least three experiments. The reverse applies in reactions where the starting substrates are attR x attL. (**D**) Recombination activities of ϕBT1 integrase in the presence of ϕC31-RDF, ϕBT1-RDF, and TG1-RDF; and as integrase-RDF fusions. (**E**) Recombination activities of ϕTG1 integrase in the presence of ϕC31-RDF, ϕBT1-RDF, and TG1-RDF; and as integrase-RDF fusions. In panels **C**, **D**, and **E**, integrases are depicted as long ovals, and RDFs as short ovals. Each integrase and its cognate RDF are colour-coded to highlight expected orthogonal interactions.

Plotting the reaction extent for each in vitro experiment (Fig. 7b) highlights additional aspects of these reactions. Unlike the in vivo inversion assays, the in vitro assays monitor deletion of a plasmid segment. Therefore, the expected equilibrium of a given reaction is more complicated to predict. While the forward reaction depends on intramolecular synapsis of two *att* sites within the same plasmid, the reverse reaction requires intermolecular synapsis between *att* sites on separate DNA circles that may have diffused away from one another. If formation of the intermolecular synapse is too difficult under the conditions used, the endpoints would be expected to lie on the axes, as they do for ϕBT1 and TG1. In contrast, if the barrier to intermolecular synapsis is not significantly different from the barrier to intramolecular synapsis, the reaction endpoints would be expected to lie on the diagonal, similar to what is seen in Fig. 7a. That is indeed approximately the case for ϕC31, indicating that ϕC31 integrase may form intermolecular synaptic complexes more readily than ϕBT1 and TG1 integrases do.

In vitro, ϕBT1 and TG1 RDFs were strikingly effective at inhibiting *attP* x *attB* recombination by their respective integrases, giving less than 5% activity in both cases. Complete inhibition of *attP* x *attB* reaction by the RDF is a key feature necessary for use of integrase-RDF pairs in applications where integrases are used as binary genetic switches. TG1 integrase will be particularly suitable for building such devices since it shows near complete integrative and excisive activities. Figure 7a also shows that different integrase – RDF pairs could be used in applications where a tunable switch is required (e.g. a promoter inversion reaction that is partially rather than fully biased toward one outcome).

The effect of covalent integrase-RDF linkage

We used integrase-RDF fusions[21] to further investigate the specificity of integrase/RDF interactions across the three integrases and their RDFs. Among the three integrases, ϕC31 integrase showed the least orthogonal behaviour, responding to *attR* x *attL* activation and *attP* x *attB* inhibition by all three RDFs in both in vivo and in vitro reactions. In all cases, ϕC31-RDF was not as effective as the other two at regulating the activities of the integrase (Figs. 5 and 6).

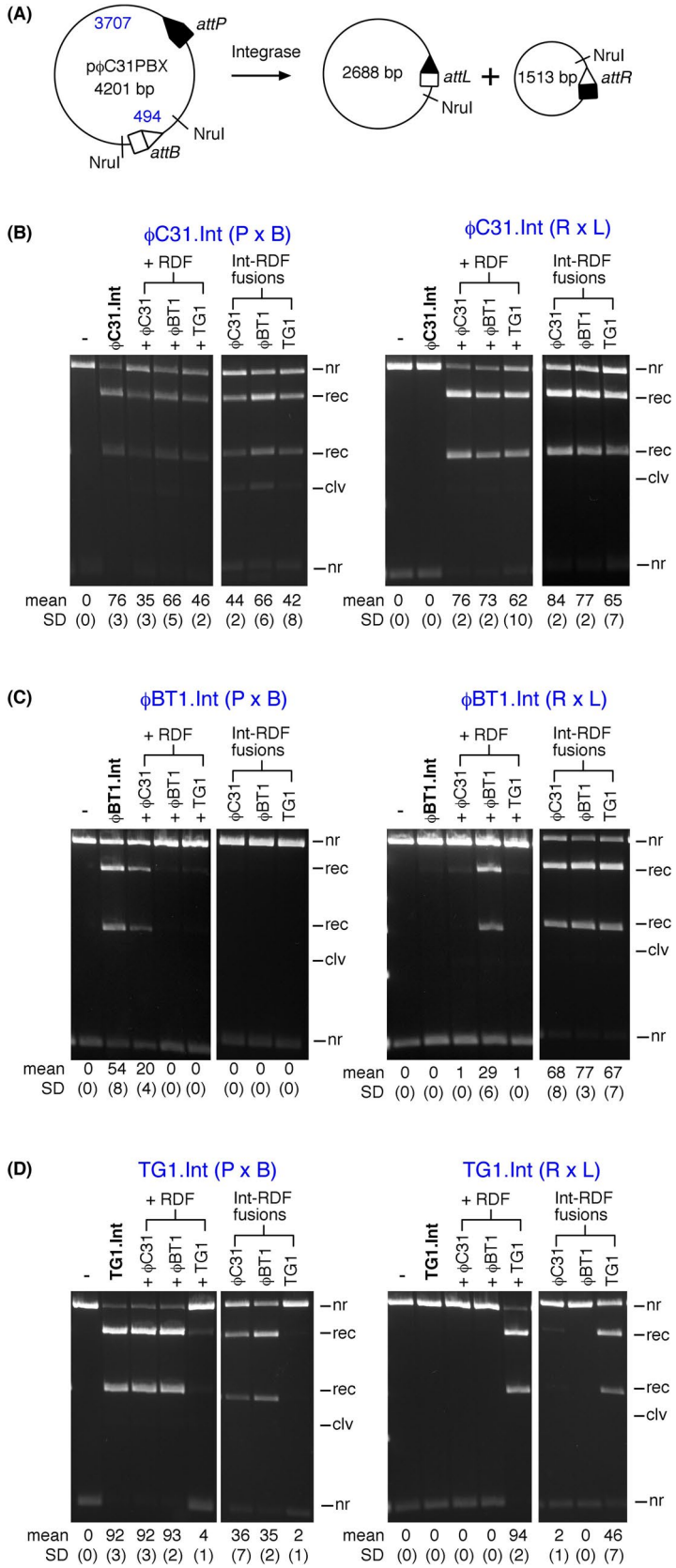
As expected, ϕBT1 integrase prefers its cognate RDF in regulation of *attR* x *attL* and *attP* x *attB* recombination (Fig. 6). However, there is a noticeable difference in the interaction of ϕBT1 integrase with the RDFs when the two proteins are supplied separately and when they are fused together. ϕBT1 integrase had limited affinity for ϕC31-RDF and TG1-RDF when the RDFs were used as separate proteins. However, when the non-cognate RDFs were fused to ϕBT1 integrase, they were more effective in activation of *attR* x *attL* recombination and inhibition of *attP* x *attB* recombination (Fig. 6).

In contrast to ϕC31 integrase and ϕBT1 integrase, TG1 integrase showed a high degree of selectivity for its cognate RDF, and insignificant effects on its activity by ϕC31-RDF and ϕBT1-RDF, either supplied as separate proteins or when fused to the integrase. This is especially noticeable in in vitro reactions, the exception being ϕC31-RDF showing *attR* x *attL* activation when fused to TG1 integrase (Fig. 6).

Discussion

In vivo and *in vitro attP* x *attB* activities of ϕC31, ϕBT1 and TG1 integrases

Activities of the three integrases described in this work have been reported in a previous in vitro study in which the properties of 10 integrases were compared[29]. Several earlier reports have noted that Bxb1 integrase is the most active integrase characterised in vitro[32–35]. Among the three integrases studied here, TG1 Integrase is the most active in recombining its *attP* and *attB* sites. Despite the high degree of sequence similarity, the recombination activity of TG1 integrase is noticeably higher than that of the other two integrases. Since there are no obvious differences in the sequences of the integrases around the catalytic residues, it is likely that the observed differences are due to other factors, including the fit of the *att* sites for each integrase – while evolutionary pressure on the phage may select an optimal *attP* sequence for integrase action to attain lysogeny, evolutionary pressure on the bacterial host may have the opposite effect on the sequence of *attB*[27,36–38]. Presuming the fitness of each integrase for its natural biological roles, these differences in *E. coli* and in vitro might 'accidentally' reflect the ability of each integrase to adapt to unnatural conditions by integrating at non-cognate *attB* sites. It is also possible that the integrases have unknown factors in their natural contexts that stimulate their activity, but which are absent in these assays.



Efficiency and specificity of RDF-dependent integrase activities

It is not surprising that there is a degree of cross-reactivity between the integrases and their RDFs. Among the three integrases, ϕC31 integrase was the least selective, with all three RDFs being able to activate it for *attR* x *attL* recombination. It is not clear why ϕC31 integrase is less selective than TG1 and ϕBT1 integrases, but these findings suggest a limitation in its use in the presence of the other two integrases.

◀**Figure 6**. In vitro recombination reactions of ϕC31, ϕBT1, and TG1 integrases. (**A**) Scheme illustrating the in vitro intramolecular recombination assay (substrate plasmid pϕC31PBX for ϕC31 integrase is illustrated; the substrates for ϕBT1 and TG1 integrases are of the same design). The plasmid substrates are named after the integrase (ϕC31) and the *att* sites recombining (PBX; *attP* X *attB* resolution reaction). Upon recombination, the plasmid substrate gives two circular products in which the *attR* and *attL* sites are separated. For the reverse reaction, the starting substrate plasmid has *attP* and *attB* sites replaced by *attR* X *attL* sites, respectively, with recombination giving *attP* and *attB* sites on separate circular plasmid products. (**B**) Recombination activities of ϕC31 integrase in the presence of ϕC31-RDF, ϕBT1-RDF, and TG1-RDF. Reactions were incubated for 2 h in the reaction buffer described in Materials and Methods. Reaction products were digested with the restriction endonuclease NruI prior to 1.2% agarose gel electrophoresis. In reactions where the integrase and RDF are added as separate proteins, the final concentration of both proteins were 200 nM. When the reactions were carried out using integrase-RDF fusion, the final concentration was 200 nM. The bands on the gel are labeled *nr* (non-recombinant, i.e. substrate), *rec* (recombination product). The mean extent of recombination and standard deviation (%) from quantitation of triplicate experiments are given below each lane. (**C**) Recombination activities of ϕBT1 integrase in the presence of ϕC31-RDF, ϕBT1-RDF, and TG1-RDF. Reaction conditions, gel electrophoresis, data acquisition and analyses areas described above in (A). (**D**) Recombination activities of TG1 integrase in the presence of ϕC31-RDF, ϕBT1-RDF, and TG1-RDF. Reaction conditions, gel electrophoresis, data acquisition and analyses are as described in (A).

In contrast to ϕC31 integrase, TGI integrase is highly selective and interacts with its RDF, gp25, to promote orthogonal clean switching reactions. For applications requiring control of directionality by the RDF, clean switching integrase-RDF pairs are essential. The findings here suggest that TG1 integrase would be suitable for such systems.

The affinity of ϕBT1 integrase for non-cognate RDFs (of TG1 and ϕC31) is significantly increased by covalent attachment of the RDF. When used as separate proteins, ϕBT1 integrase did not interact significantly with these non-cognate RDFs. However, covalent attachment resulted in the ability of TG1 and ϕC31 RDFs to activate *attR* x *attL* recombination and inhibit *attP* x *attB* recombination.

Overall, the observed weak integrase/RDF orthogonality among these three enzymes emphasizes the need for identifying more integrases with known RDFs. To date, only a handful of active integrase-RDF pairs have been characterised, in vivo and/or in vitro. These systems include Bxb1[31], ϕRV1[30], ϕC31[18], ϕBT1[19], TP901[39], A118[37], SPBc[40], ϕJoe[41], and CD1231[42]. In addition, the RDF for TG1 integrase (gp25) was identified by Zhang et al.[19]. , but its activity has not been demonstrated in vivo or in vitro.

The availability of a larger set of integrase-RDF pairs, with known clean switching activities and orthogonal to each other, will facilitate the use of serine integrases in designing genetic circuits and other regulated systems in synthetic biology. Additionally, the plot in Fig. 7a shows that tunability could be added to existing genetic circuits by using related but not strictly cognate RDFs for a given integrase: in that way, the set point of a particular inversion switch could be changed simply by changing the RDF.

Our findings also provide an opportunity for identifying the protein-protein interactions among this family of integrase/RDF pairs that determine specificity for the integrase. For example, the strict specificity of TG1 integrase for its RDF in contrast to the broad tolerance of ϕC31 integrase to all three RDFs could be a starting point for identification of protein features and interactions that determine *attR* x *attL* activation, *attP* x *attB* inhibition, and selectivity for the cognate integrase.

AlphaFold2 multimer predicts a model for integrase/RDF interaction

The predicted interactions of each of the three integrases with their cognate RDFs are very similar. However, the sequence within the binding surface of the RDF is much more conserved than that of the integrase (Figs. 2 and 3). This could be an indication that specificity is determined more by the integrase than the RDF. The significant exception is seen in TG1 integrase, which has a small insertion that could allow it to make a more stable or more rigid interaction with the RDF (Fig. 4). This insertion might explain the enhanced ability of TG1 integrase to discriminate among RDFs (Figs. 5 and 6). Investigating the effects of deleting this element from TG1 integrase or inserting it into ϕC31 or ϕBT1 integrases could provide some insights useful in the design and understanding of future genetic switches.

Methods

In vivo recombination reactions

Plasmids for constitutive expression of integrases, RDFs, and integrase–RDF fusion proteins in *E. coli* were made as described in Olorunniji et al.[21]. Recombination reactions (inversion) with ϕC31 integrase were carried out using the plasmid substrates, pϕC31-invPB and pϕC31-invRL, as described in Olorunniji et al.[43]. Similar plasmid substrates for ϕBT1 and TG1 integrases were made by cloning the corresponding *att* sites into pϕC31-invPB or pϕC31-invRL. Recombination activity was measured using the invertible promoter reporter system (see Fig. 5). To assay ϕC31 integrase *attP* x *attB* recombination, *E. coli* DS941 cells containing the pϕC31-invPB substrate were transformed with the vector plasmid expressing ϕC31integrase. The extent of *attP* x *attB* recombination was monitored by counting the number of colonies expressing either RFP or GFP. TG1 and ϕBT1 integrases were assayed the same way using their corresponding in vivo substrates. Scanning of *E. coli* cell fluorescence was carried out using a Typhoon FLA 9500 fluorimager (GE Healthcare) as described in Olorunniji et al.[43]. Briefly, fluorescence of the expressed proteins was measured (GFP: excitation, 488 nm, band-pass blue filter; RFP: excitation, 532 nm, long-pass green filter).

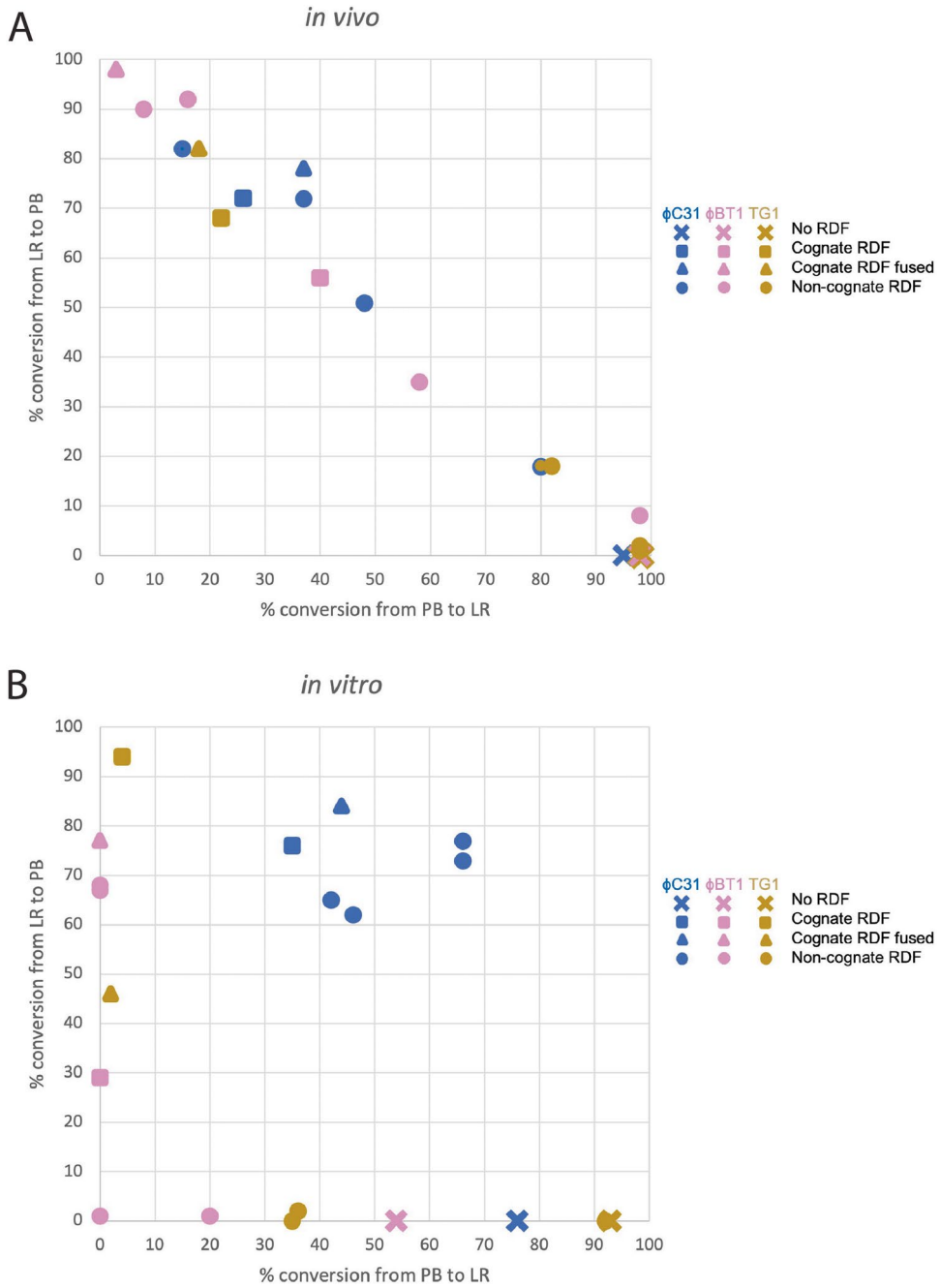


Figure 7. Correlation of extent of recombination in attP x attB and attR x attL reactions in the presence of the recombination directionality factor. (**A**) In vivo recombination, showing data taken from Fig. 5. (**B**) In vitro recombination, showing data taken from Fig. 6. The colour scheme is the same as in Figs. 5 and 6: blue, ϕC31 integrase; pink, ϕBT1 integrase, gold, TG1 integrase. A triangle denotes a cognate RDF fused to the integrase; square, a cognate RDF as a separate protein; circle, a non-cognate RDF; and X, no RDF.

Expression and purification of serine integrases

The three integrases were expressed in *E. coli* BL21(DE3)pLysS and purified as described in Olorunniji et al.[21]. Briefly, the expression strain for each integrase was grown at 37 °C in 2x YT-broth to A_{600} of 0.6 to 0.8. Cultures were cooled to 20 °C and integrase expression was induced with 0.75 mM IPTG, after which the cultures were grown for 16 h at 20 °C. The proteins were purified by nickel affinity chromatography and bound proteins were eluted with an imidazole gradient buffer system. Fractions were collected and SDS-PAGE was used to determine peaks corresponding to the proteins of interest. Fractions containing the integrases were dialysed against Protein Dilution Buffer, PDB (25 mM Tris–HCl (pH 7.5), 1 mM DTT, 1 M NaCl and 50% glycerol), and stored at − 20 °C. Dilutions of each integrase for in vitro recombination reactions were made into the same buffer.

In vitro recombination reactions

Substrate plasmids for the assay of intramolecular activities of the three integrases in vitro are described in Abioye et al.[29]. The plasmid substrate for each integrase is named according to the integrase *att* sites. Figure 6 shows the substrate for ϕC31 integrase in which pϕC31PBX carries ϕC31 *attP* and *attB* sites. The *att* sites are arranged in a 'head to tail' orientation leading to resolution of the substrate plasmid into two smaller product plasmids upon recombination. In vitro recombination of supercoiled plasmid substrates and analysis of recombination products were carried out as reported in Abioye et al.[27]. Typically, recombination reactions were carried out by adding integrase (2 µM, 5 µl) to a 30 µl solution containing the plasmid substrate (25 µg/ml), 50 mM Tris-HCl (pH 7.5), 100 µg/ml BSA, 5 mM spermidine, and 0.1 mM EDTA. Samples were incubated at 30 °C for 2 h, after which the reactions were stopped by heating at 80 °C for 10 min. The samples were cooled and treated with NruI (New England Biolabs) to facilitate analysis of recombination products. Following the digest, samples were treated with SDS and protease K before reaction products were separated by agarose gel electrophoresis[27,29].

AlphaFold2-based protein structure prediction

Three-dimensional (3D) protein structures were generated using the colabfold implementation AlphaFold2-multimer; version 1.5.2 with default parameters[44–46]. The structures were viewed and manipulated using PyMol (https://pymol.org) (Fig. 4). The models shown were predicted using full-length integrase and RDF sequences. Nearly identical interactions between DBD2 and the RDF were predicted for each pair when the integrase sequences were truncated to include only DBD2 and the coiled coil.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 17 April 2024; Accepted: 23 October 2024
Published online: 01 November 2024

References

1. Olorunniji, F. J., Rosser, S. J. & Stark, W. M. Site-specific recombinases: molecular machines for the genetic revolution. *Biochem. J.***473**, 673–684 (2016).
2. Stark, W. M. Making serine integrases work for us. *Curr. Opin. Microbiol.***38**, 130–136 (2017).
3. Rice, P. P.A. Serine resolvases. *Microbiol. Spectr.***3**, MDNA3–0045 (2015).
4. Smith, M. C. M. Phage-encoded serine integrases and other large serine recombinases. *Microbiol. Spectr.***3**, 3.4.06 (2015).
5. Fogg, P. C. M., Colloms, S., Rosser, S., Stark, M. & Smith, M. C. M. New Applications for Phage integrases. *J. Mol. Biol.***426**, 2703–2716 (2014).
6. Colloms, S. D. *et al.* Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Res.***42**, e23 (2014).
7. Muroi, T. *et al.* TG1 integrase-based system for site-specific gene integration into bacterial genomes. *Appl. Microbiol. Biotechnol.***97**, 4039–4048 (2013).
8. Blanch-Asensio, A. *et al.* STRAIGHT-IN enables high-throughput targeting of large DNA payloads in human pluripotent stem cells. *Cell. Rep. Methods*. **2**, 100300 (2022).
9. Yarnall, M. T. N. *et al.* Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. *Nat. Biotechnol.***41**, 500–512 (2023).
10. Ginsburg, D. S. & Calos, M. P. Site-specific integration with ϕC31 integrase for prolonged expression of therapeutic genes. in Advances in Genetics **54**, 179–187 (Elsevier, (2005).
11. Herisse, M. *et al.* The ΦBT1 large serine recombinase catalyzes DNA integration at pseudo-attb sites in the genus Nocardia. *PeerJ.* **6**, e4784 (2018).
12. Zhang, M., Yang, C., Tasan, I. & Zhao, H. Expanding the potential of mammalian genome Engineering *via* targeted DNA integration. *ACS Synth. Biol.***10**, 429–446 (2021).
13. Thorpe, H. M. & Smith, M. C. M. *In vitro* site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc. Natl. Acad. Sci.***95**, 5505–5510 (1998).
14. Zhang, L., Ou, X., Zhao, G. & Ding, X. Highly efficient in vitro site-specific recombination system based on streptomyces phage phiBT1 integrase. *J. Bacteriol.***190**, 6392–6397 (2008).
15. Morita, K. *et al.* The site-specific recombination system of actinophage TG1. *FEMS Microbiol. Lett.***297**, 234–240 (2009).
16. Morita, K. *et al.* In vitro characterization of the site-specific recombination system based on actinophage TG1 integrase. *Mol. Genet. Genomics MGG.* **282**, 607–616 (2009).
17. Du, D. *et al.* Genome engineering and direct cloning of antibiotic gene clusters via phage ϕBT1 integrase-mediated site-specific recombination in Streptomyces. *Sci. Rep.***5**, 8740 (2015).
18. Khaleel, T., Younger, E., McEwan, A. R., Varghese, A. S. & Smith, M. C. M. A phage protein that binds φC31 integrase to switch its directionality. *Mol. Microbiol.***80**, 1450–1463 (2011).
19. Zhang, L., Zhu, B., Dai, R., Zhao, G. & Ding, X. Control of directionality in Streptomyces Phage φBT1 integrase-mediated site-specific recombination. *PLoS ONE.* **8**, e80434 (2013).
20. Gregory, M. A., Till, R. & Smith, M. C. M. Integration site for *Streptomyces* Phage φBT1 and development of site-specific integrating vectors. *J. Bacteriol.***185**, 5320–5323 (2003).
21. Olorunniji, F. J. *et al.* Control of serine integrase recombination directionality by fusion with the directionality factor. *Nucleic Acids Res.***45**, 8635–8645 (2017).
22. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.***7**, 539 (2011).
23. Rutherford, K., Yuan, P., Perry, K., Sharp, R. & Van Duyne, G. D. Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Res.***41**, 8341–8356 (2013).
24. Mandali, S., Gupta, K., Dawson, A. R., Van Duyne, G. D. & Johnson, R. C. Control of recombination directionality by the Listeria phage A118 protein Gp44 and the coiled-Coil Motif of its serine integrase. *J. Bacteriol.***199**, e00019–e00017 (2017).
25. Abe, K., Takahashi, T. & Sato, T. Extreme C-terminal element of SprA serine integrase is a potential component of the 'molecular toggle switch' which controls the recombination and its directionality. *Mol. Microbiol.***115**, 1110–1121 (2021).
26. Fogg, P. C. M. *et al.* Recombination directionality factor gp3 binds ϕC31 integrase via the zinc domain, potentially affecting the trajectory of the coiled-coil motif. *Nucleic Acids Res.***46**, 1308–1320 (2018).

27. McEwan, A. R., Rowley, P. A. & Smith, M. C. M. DNA binding and synapsis by the large C-terminal domain of C31 integrase. *Nucleic Acids Res.***37**, 4764–4773 (2009).
28. Chen, Y. W., Su, B. Y., Van Duyne, G. D., Fogg, P. & Fan, H. F. The influence of coiled-coil motif of serine recombinase toward the directionality regulation. *Biophys. J.***122**, 4656–4669 (2023).
29. Abioye, J. *et al.* High fidelity one-pot DNA assembly using orthogonal serine integrases. *Biotechnol. J.***18**, 2200411 (2023).
30. Bibb, L. A., Hancox, M. I. & Hatfull, G. F. Integration and excision by the large serine recombinase phiRv1 integrase. *Mol. Microbiol.***55**, 1896–1910 (2005).
31. Ghosh, P., Wasil, L. R. & Hatfull, G. F. Control of phage Bxb1 excision by a novel recombination directionality factor. *PLoS Biol.***4**, e186 (2006).
32. Russell, J. P., Chang, D. W., Tretiakova, A. & Padidam, M. Phage Bxb1 integrase mediates highly efficient site-specific recombination in mammalian cells. *BioTechniques*. **40**, 460–464 (2006).
33. Xu, Z. *et al.* 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 (2013).
34. Duportet, X. *et al.* A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res.***42**, 13440–13451 (2014).
35. Jusiak, B. *et al.* Comparison of Integrases identifies Bxb1-GA mutant as the most efficient site-specific integrase system in mammalian cells. *ACS Synth. Biol.***8**, 16–24 (2019).
36. Ghosh, P., Pannunzio, N. R. & Hatfull, G. F. Synapsis in Phage Bxb1 integration: selection mechanism for the correct pair of Recombination sites. *J. Mol. Biol.***349**, 331–348 (2005).
37. Mandali, S., Dhar, G., Avliyakulov, N. K., Haykinson, M. J. & Johnson, R. C. The site-specific integration reaction of Listeria phage A118 integrase, a serine recombinase. *Mob. DNA*. **4**, 2 (2013).
38. Adams, V., Lucet, I. S., Lyras, D. & Rood, J. I. DNA binding properties of TnpX indicate that different synapses are formed in the excision and integration of the Tn *4451* family. *Mol. Microbiol.***53**, 1195–1207 (2004).
39. Breüner, A., Brøndsted, L. & Hammer, K. Novel Organization of Genes Involved in Prophage Excision identified in the Temperate Lactococcal bacteriophage TP901-1. *J. Bacteriol.***181**, 7291–7297 (1999).
40. Abe, K., Takamatsu, T. & Sato, T. Mechanism of bacterial gene rearrangement: SprA-catalyzed precise DNA recombination and its directionality control by SprB ensure the gene rearrangement and stable expression of spsM during sporulation in Bacillus subtilis. *Nucleic Acids Res.***45**, 6669–6683 (2017).
41. Fogg, P. C. M., Haley, J. A., Stark, W. M. & Smith, M. C. M. Genome integration and excision by a New Streptomyces Bacteriophage, ϕJoe. *Appl. Environ. Microbiol.***83**, e02767–e02716 (2017).
42. Serrano, M. *et al.* A recombination directionality factor controls the cell type-specific activation of σK and the Fidelity of Spore Development in Clostridium difficile. *PLOS Genet.***12**, e1006312 (2016).
43. Olorunniji, F. J. *et al.* Control of ϕC31 integrase-mediated site-specific recombination by protein trans-splicing. *Nucleic Acids Res.***47**, 11452–11460 (2019).
44. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods*. **19**, 679–682 (2022).
45. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**, 583–589 (2021).
46. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. Preprint at. https://doi.org/10.1101/2021.10.04.463034 (2021).

Author contributions

F.J.O, P.R, and W.M.S.: Conceptualization, Methodology, Writing, Funding acquisition.; A.M., A.B., A.H., H.S.: Methodology, Investigation, Software; All authors reviewed the manuscript.

Funding

This work was supported by the UKRI/BBSRC grant BB/003356/1 to WMS; and collaborative grants NSF/BIO 2223480 and UKRI/BBSRC BB/X012085/1 to FJO and PAR.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-77570-9.

Correspondence and requests for materials should be addressed to F.J.O.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.