Nucleic Acids Research, 2025, **53**, gkaf050
https://doi.org/10.1093/nar/gkaf050
Advance access publication date: 5 February 2025
Nucleic Acid Enzymes

Large serine integrases utilise scavenged phage proteins as directionality cofactors

Abdulrazak Alsaleh[1], Alexandria Holland[1], Heewhan Shin[2], Tania Pena Reyes[2], Aron Baksh[1], Oluwateniola T Taiwo-Aiyerin[1], Ying Pigli[2], Phoebe A Rice[2,*], Femi J Olorunniji[1,*]

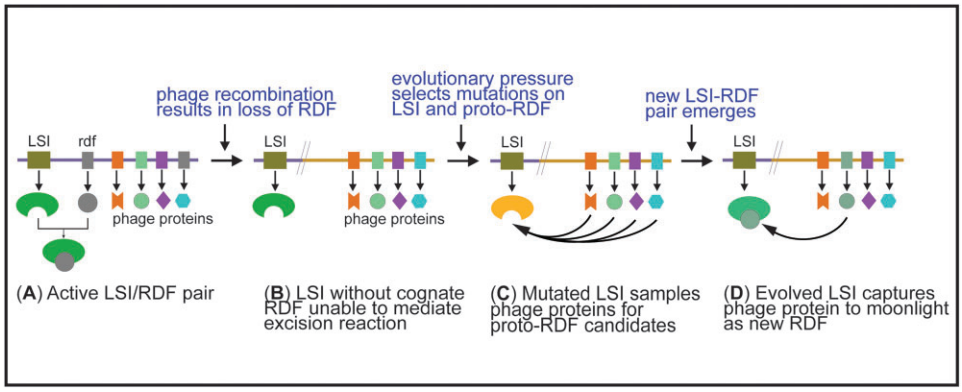
[1]School of Pharmacy & Biomolecular Sciences, Faculty of Health, Innovation, Technology, and Science, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, United Kingdom
[2]Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637, United States

[*]To whom correspondence should be addressed. Email: F.J.Olorunniji@ljmu.ac.uk
Correspondence may also be addressed to Phoebe A Rice. Email: price@uchicago.edu

Abstract

Recombination directionality factors (RDFs) for large serine integrases (LSIs) are cofactor proteins that control the directionality of recombination to favour excision over insertion. Although RDFs are predicted to bind their cognate LSIs in similar ways, there is no overall common structural theme across LSI RDFs, leading to the suggestion that some of them may be moonlighting proteins with other primary functions. To test this hypothesis, we searched for characterized proteins with structures similar to the predicted structures of known RDFs. Our search shows that the RDFs for two LSIs, TG1 integrase and Bxb1 integrase, show high similarities to a single-stranded DNA binding (SSB) protein and an editing exonuclease, respectively. We present experimental data to show that Bxb1 RDF is probably an exonuclease and TG1 RDF is a functional SSB protein. We used mutational analysis to validate the integrase-RDF interface predicted by AlphaFold2 multimer for TG1 integrase and its RDF, and establish that control of recombination directionality is mediated via protein–protein interaction at the junction of recombinase's second DNA binding domain and the base of the coiled-coil domain.

Graphical abstract



Introduction

Site-specific recombination events are frequently used for integration, excision, and inversion of mobile genetic elements in phages and prokaryotes [1–3]. Typically, the process involves cleavage of all four DNA strands in recombining duplexes, exchange of DNA ends and re-ligation to give recombinant DNA products [4]. A characteristic feature of reactions involving site-specific recombination is the formation of covalent protein—DNA intermediates, presumably to ensure that cleaved DNA ends are not lost during the strand exchange events. The identity of the nucleophilic amino acid that forms this covalent intermediate is the basis for classification of site-specific recombinases as serine or tyrosine recombinases: Tyrosine recombinases e.g. Flp, Cre, and phage lambda integrases, utilize a catalytic tyrosine, while serine recombinases, e.g. large serine integrases (LSIs), many transposon resolvases and many invertases, use serine [4]. The covalent intermediates store the energy of the broken phosphodiester bonds. Therefore, the net change in chemical bond energy for these reactions is zero, and therefore it cannot be used to drive the reaction forward [4]. Site-specific recombinase systems have evolved an array of different, sometimes complex strategies

Received: August 20, 2024. **Revised:** January 15, 2025. **Editorial Decision:** January 16, 2025. **Accepted:** January 21, 2025

to favour the forward or reverse reaction direction [5–9]. For LSIs, the balance between reaction directions is tipped by a second protein, termed a recombination directionality factor (RDF) [2, 10–12]. When only the LSI is present, the integration reaction direction is strongly favoured, whereas when the RDF is present, integration is inhibited and the reverse (excision) reaction is favoured [2].

LSIs are used by temperate phages (and some mobile genetic elements) to insert their genomes into that of their host bacteria as stably integrated prophages (Fig. 1). Integration involves recombination of short attachment sites (40–50 bp) on the phage (*attP*) and the bacterial host (*attB*) resulting in new sites *attR* and *attL* flanking the inserted prophage [2, 13]. LSIs are typically phage-encoded proteins, and they form a group of highly conserved proteins that work via a similar mechanism [2].

Recombination of *attP* and *attB* is catalysed by the integrase in a synaptic tetramer complex in which dimers of integrase bind and synapse *attP* and *attB*, followed by DNA cleavage reactions, subunit rotation to exchange ends, and ligation of the recombinant DNA to form the products (*attR* and *attL*). In the lysogenic stage of the phage life cycle, the RDF is not expressed and the forward or 'integrative' reaction is irreversible resulting in stable insertion of the phage DNA into the host's genome. The reverse 'excision' reaction happens during the lytic phase, when the RDF is expressed. The RDF modifies the preferred reaction direction of the integrase to promote *attR* × *attL* recombination. In addition to promoting excision, the RDF also inhibits integration to ensure the unidirectionality of LSI reactions either in the forward or reverse direction [10–12].

Tyrosine recombinases are also sometimes encoded by phages to function as integrases. However, they are more complicated: for example, phage lambda requires an ∼250 bp *attP* site and additional host proteins [14–17]. Furthermore, the mechanism for strand exchange by tyrosine integrases is different, as are their RDFs (often called 'xis' proteins). RDFs for tyrosine recombinases have been shown to bind both DNA and the cognate integrase. However, some perform additional roles as transcriptional regulators [18].

Although LSIs can easily be identified in the genomes of phages (or mobile genetic elements), it is not readily obvious which genes code for RDFs. RDFs often lack synteny with their cognate LSIs, and there is no sequence homology across the set of <30 known RDFs [2, 18, 19]. Although there are no published experimental RDF structures, our recent analysis of predicted structures [19] using AlphaFold2 and AlphaFold multimer [20–22] revealed no universally shared structural motifs that could explain their mechanism of action. However, they were all predicted to bind to the integrases at the same general location: the junction between the second DNA binding domain (DBD2; sometimes called the zinc-binding domain) and the coiled-coil (CC) motif of the integrase. The conserved location of this interaction aided in developing a high throughput 'virtual pulldown' approach to finding RDFs for LSIs within phage genomes and suggests a conserved mode of action [19].

In contrast to the RDFs used by tyrosine recombinases, LSI RDFs generally have no known DNA binding role during recombination [11, 12, 18], and their functions appear to be solely to bind the integrase and modify its activity from integration-catalyzing enzyme to one that promotes excision of the prophage from the host's genome. The lack of DNA binding requirement may account for the diversity of proteins used by LSIs for RDF functions. This raises the question of how phages evolve functional RDF–LSI pairs. One likely mechanism is to repurpose existing proteins to acquire a moonlighting role as RDF. In fact, it has been shown that the phage Bxb1 RDF is indeed required for phage DNA replication as well as for excision [23]. The tyrosine integrases encoded by phages HP1 and P2 use RDFs that also function as transcriptional repressors of early genes [24–27]. Moonlighting is an approach used by phages since it allows them to derive broader functions using a limited genome [28–30].

We explored the possibility that RDF proteins for LSIs could have other 'primary' functions in the phage by searching for proteins with similar structures as the predicted structures for the known RDFs. We focused on the two largest known RDFs: that for phage Bxb1 (which is also known as gp47), and that for phage TG1 (which is also known as gp25), and which is similar to the RDFs for phages ϕC31 and ϕBT1 [31]. Our findings suggest that these RDFs are different phage replication proteins moonlighting as RDFs for their cognate LSIs. We show that the TG1 RDF is a functional single-stranded DNA binding (SSB) protein, and present evidence that Bxb1 RDF shows limited exonuclease activity.

Materials and methods

Structure predictions and bio-informatics

Three-dimensional protein structures were initially predicted using the colabfold implementation AlphaFold2-multimer; version 1.5.2 with default parameters [20, 21, 32] (Fig. 6). In the structures shown in Fig. 6 the integrase sequence was truncated to include only the C-terminal segment, which includes DBD2 and the CC. The inter- and intra-molecular measures of confidence for the AlphaFold2-multimer model of the TG1 RDF-integrase complex were ipTM = 0.71 and pTM = 0.5, respectively, and for the Bxb1 model were ipTM = 0.80 and pTM = 0.65. RDF structures were later predicted with AlphaFold3 [22] in order to test addition of ions and single-stranded DNA (ssDNA) to the models (Fig. 2). The inter- and intra-molecular confidence scores for the Bxb1 model [RDF plus 1 Zn^{2+} ion, 1 Fe^{3+} ion and poly $d(T)_6$] were ipTM = 0.86, pTM = 0.91, respectively, and for the TG1 model [RDF plus 1 Zn^{2+} ion and poly $d(T)_6$] were ipTM = 0.82, pTM = 0.84. The structures were viewed and manipulated and figures made using PyMol (https://pymol.org/pymol.html). Surface electrostatic potential was calculated and visualized using the APBS PyMOL plugin [33]. The Dali server was used to find characterized proteins with structures similar to our predicted RDF structures using 'Exhaustive PDB25 search' mode to avoid redundant hits [34]. The top hits are listed in Supplementary Fig. S1. To identify putative interaction partners of these RDFs, we performed virtual pulldowns as in [19] using the RDF as bait and the other phage-encoded proteins as prey.

Integrase and RDF expression vectors for protein purification

Codon-optimized protein-coding DNA fragments expressing TG1 integrase and TG1 RDF were inserted into pET-28a(+) (Novagen), between the NdeI and XhoI sites as described previously [31, 35]. Mutants of the integrase and RDF were made by cloning sysnthetic g-block DNA (Integrated DNA Technologies) containing the desired changes into the appropriate

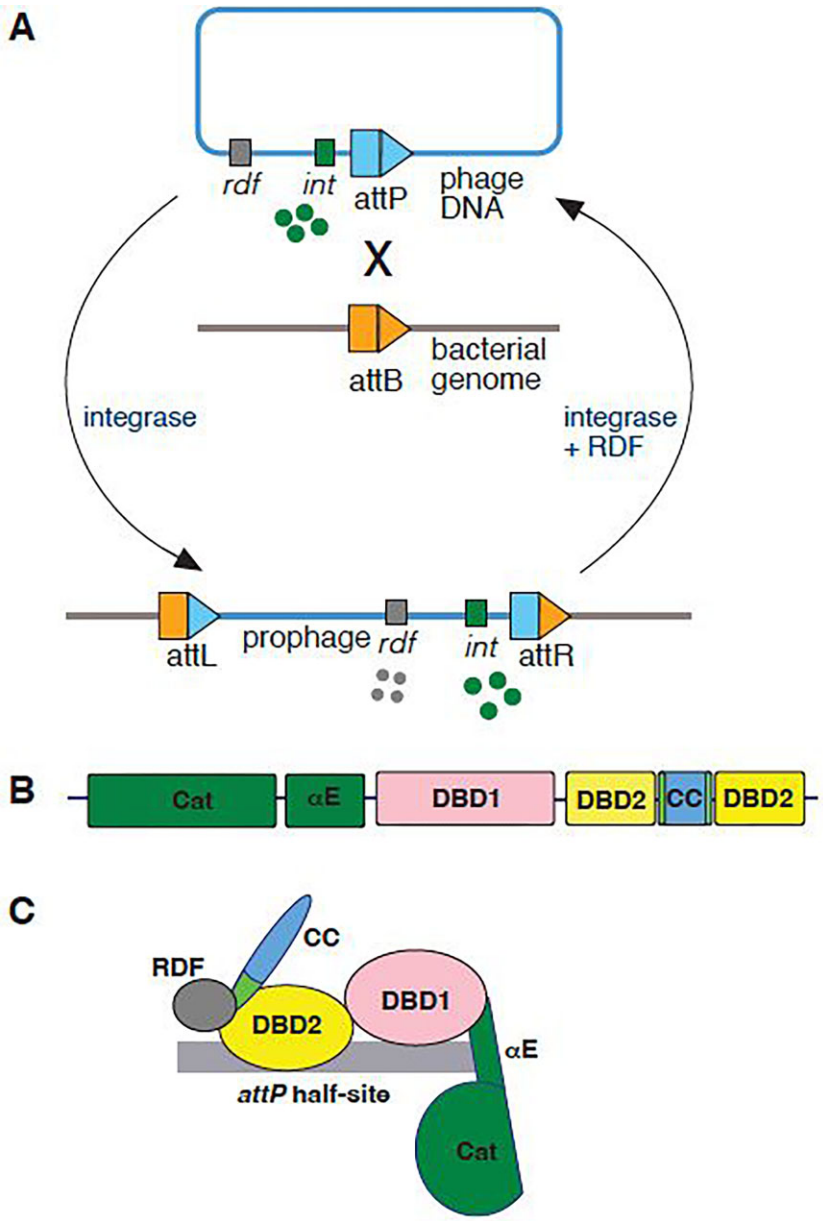


Figure 1. Phage integration and excision by LSIs. (**A**) Phage-encoded integrase (green ovals) catalyse the integration of phage DNA at *attP* site (blue arrow) into the bacterial host genome at an *attB* site (amber). The recombination reaction results in the formation of recombinant *attR* and *attL* sites flanking the integrated prophage genome. In the reverse excision reaction, the RDF (grey ovals) is expressed from the prophage genome and binds to the integrase to promote *attR* × *attL* recombination. (**B**) Domain architecture of LSIs. The catalytic domain, cat (green) and the α-E helix, αE (green) mediate catalysis and provide the subunit rotation interface. The two DNA-binding domains (DBD1 and DBD2) mediate DNA binding and sequence specificity. DBD1 also known as 'recombinase domain' or RD is shown in pink. DBD2, also known as the zinc-binding domain or ZD is shown in yellow. The CC domain (orange) is embedded within DBD2 and mediate directionality control. The DBD2-proximal region in the CC domain is shown in pale green. (**C**) Schematic illustration of serine integrase domain interactions with the recombination site (attP) and the RDF based on structure of LI integrase-attP complex [53] and AlphaFold multimer-predicted interaction of the RDF with DBD2-proximal region of the CC domain [19, 31]). Colors are matched among Figs 1, 6, and Supplementary Fig. S4.

pET-28a(+)-based expression vectors. All proteins expressed from these plasmids carry N-terminal hexahistidine tags to allow purification via nickel affinity chromatography.

Expression and purification of RDF for binding reactions

The TG1 RDF protein used for *in vitro* binding reactions was purified as follows. *Escherichia coli* [Rosetta(DE3)plysS] containing a pET vector encoding the TG1 RDF (described below) were grown at 37 °C in LB supplemented with (50 μg/ml kanamycin and 100 μM $ZnSO_4$), induced by the addition of 0.5 mM IPTG once OD_{600} reach ∼1, then grown at 20 °C overnight (∼16 h). Cell pellets were collected and resuspended in Ni-Buffer A (50 mM Phosphate, 1 M NaCl, 5% Glycerol, 1 mM TCEP, pH7.5) supplemented with complete Mini protease inhibitor cocktail (MilliporeSigma; one tablet per liter of culture). Lysozyme (200 μg/ml) was added before sonication. The sonicated sample was centrifuged at 20 000 rpm in an SS-34 rotor for 1 h at 4 °C, and the supernatant was collected and filtered. Affinity chromatography was carried out using a 5 ml HisTrap HP column (Cytiva). The column was washed with

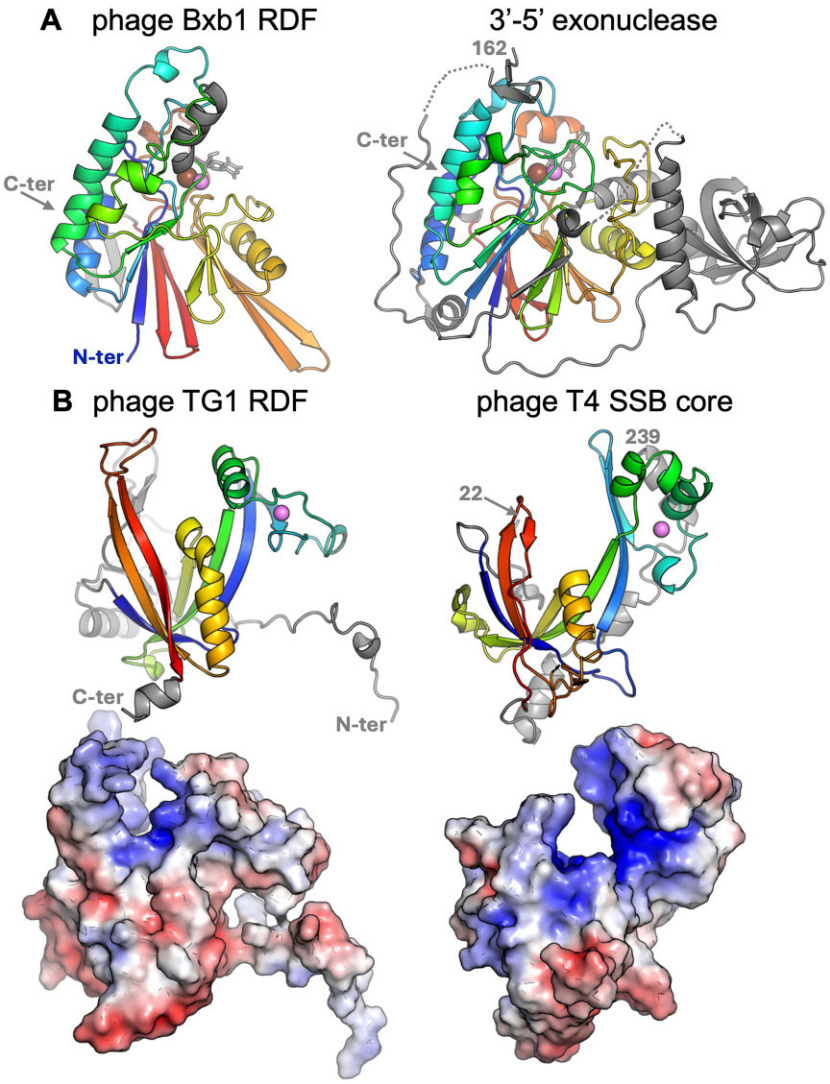


Figure 2. Structural similarities between Bxb1 and TG1 RDFs and replication proteins. (**A**) Comparison of the predicted Bxb1 RDF structure (left) to the crystal structure of the proofreading exonuclease subunit of PolD {right; PDB: 5IHE; [37]}. The related portions of each structure are shaded from N (blue) to C (red). Zinc ions are violet and iron ions brown. One nucleotide from the predicted structure of Bxb1 RDF with dT_6 is shown in gray, as is the single nucleotide seen in the active site of the exonuclease. (**B**) Top: comparison of the predicted TG1 RDF structure (left) to that of phage T4 SSB protein core {right; PDB: 1GPC; [46]}. Bottom: surface representations of each colored by electrostatic potential (blue = +5 kT/e, shading to red = −5 kT/e). N and C termini, or 1st and last residues within the structure shown, are labeled.

Ni-Buffer A, after which the sample was loaded, washed with Ni-Buffer A, and eluted with Ni-Buffer B (Ni-Buffer A + 0.5 M imidazole, pH 7.5) following a 0%–100% linear gradient over 30 min at a rate of 2 ml/min. Selected fractions were pooled and rechromatographed on the same column. Selected fractions were pooled again, then polished on HiPrep QFF ion exchange column (Cytiva) equilibrated with Q FF-Buffer A (20 mM Tris–HCl, 5% Glycerol, 1 mM TCEP, pH8). The column was then washed with 10% Q FF-Buffer B (Q FF-Buffer A + 2 M NaCl) after which the protein was eluted with Q FF -Buffer B (10%–70% gradient; flow rate 2 ml/min). Fractions were chosen after SDS-PAGE and nuclease activity assay. To assay for nuclease activity, fractions were incubated with supercoiled plasmid (pUC19) and additional Mg^{2+} followed by agarose gel electrophoresis. Purified protein was concentrated to 340 μM in 20 mM Tris, pH 8, 0.5 mM EDTA, 200 mM NaCl, 20% Glycerol, 2 mM TCEP, flash-frozen in small aliquots in liquid nitrogen and stored at −80 °C.

In vitro binding reactions

DNA substrates for binding assays were purchased from IDT. The single-stranded substrate was poly$(dT)_{50}$ with a 5′ fluorescein label (5′-/56-FAM). The 50 bp double-stranded TG1 attL substrate had a 5′ fluorescein label on the top strand, and the sequence: 5′ CAGCTCCGCGGGCAAGACCTAGCTCTTACCCAGTTGGGCGGGATAATTAA 3′. The binding assays were conducted in a total volume of 20 μl in a buffer containing 20 mM Tris, pH 8, 100 mM NaCl, 10% Glycerol, 50 ng/μl BSA (bovine serum albumin), and 1 mM TCEP. DNA substrates were present at 0.2 μM and protein at 0, 0.2, 0.4, and 0.8 μM. After 1 h incubation at room temperature, 15 μl of each sample was loaded onto a 10% 0.5 × TBE gel and run at 140 V for 2 h at 4 °C. Bands were visualized on a Chemidoc imager. Due to the predicted zinc binding site in this protein, we also tested addition of 0.5 mM $ZnSO_4$ but found that it did not change the DNA-binding ability of the protein.

In vitro plasmid substrates for intramolecular recombination

The intramolecular recombination assay used is as decribed in Abioye *et al.* [35] and Olorunniji *et al.* [36]. The plasmids for *attP* × *attB* and *attR* × *attL* recombination reactions are named pTG1-PBX and pTG1-RLX, respectively. In both substrates, the *att* sites are arranged in direct repeat or 'head to tail' orientation such that recombination results in resolution (excision) of the substrate plasmid into two separate smaller plasmid circles (Fig. 5A). The sequences of the *attP, attB, attR, and attL* sites are shown in Fig. 5B, and all the recombination substrate plasmids are available upon request. Supercoiled plasmid DNA used for *in vitro* reactions was prepared from transformed *E. coli* DS941 cells using a Qiagen miniprep kit [36]. The concentrations of DNA preps were determined by measuring absorbance at 260 nm, and the quality of each prep was verified by electrophoresis on 1.2% agarose gels.

Expression and purification of TG1 integrase and TG1 RDF for activity assays

Expression and purification of TG1 integrase and TG1 RDF, and their mutants were carried out as described in Olorunniji *et al.* [36] and Abioye *et al.* [35]. *E. coli* strain BL21(DE3)pLysS was made chemically competent and transformed with the appropriate protein expression vector. The expression strain for each protein was grown in 2x Y-broth at 37°C to mid-log phase (OD_{600}, 0.6–0.8) and cooled down rapidly to 20°C before inducing protein expression with the addition of 0.5 mM IPTG, after which the cultures were grown for a further 16 h at 20 °C. The cultures were grown with kanamycin (50 μg/ml) and chloramphenicol (50 μg/ml) added to the media. Cells were harvested by centrifugation at 4 °C, and the pellet was washed in 25 mM Tris–HCl (pH 7.5), 10 mM $MgCl_2$, and the pellet was collected by centrifugation at 4 °C. The washed pellet was resuspended in 25 ml of Buffer A [20 mM sodium phosphate (pH 7.4), 1 M NaCl, 1 mM dithiothreitol (DTT), 50 mM imidazole, and 1% (v/v) ethanol]. The suspension was cooled in ice, and the cells were lysed by sonication (Branson, SFX 150). The suspension was centrifuged for 30 min at 4 °C, 12 000 rpm, after which the supernatant was collected and filtered. Proteins were purified by nickel affinity chromatography using a 1 ml HisTrap FF pre-packed column (GE Healthcare). The column was equilibrated with the starting Buffer A, at a constant flow rate of 1 ml/min, prior to loading the protein sample, also in Buffer A. The column was washed with Buffer A (25 ml) to remove unbound proteins and the bound protein of interest was eluted with Buffer B (Buffer A, but with 500 mM imidazole), increasing in a 0%–100% linear gradient over 25 min. Purity of selected fractions was assessed by SDS-polyacrylamide gel electrophoresis and chosen fractions containing the protein of interest were dialysed against Protein Dilution Buffer [PDB; 25 mM Tris–HCl (pH 7.5), 1 mM DTT, 1 M NaCl and 50% v/v glycerol), and stored at −20 °C.

Expression and purification of Bxb1 RDF

Bxb1 RDF was expressed and purified as described earlier in Olorunniji *et al.* [36] following a procedure similar to that used for TG1 integrase and TG1 RDF as outlined above. BL21(DE3) pLysS cells were transformed with a pET-based vector expressing Bxb1 RDF. Three 500 ml cultures were grown at 37°C with shaking until mid-log phase (OD_{600}, 0.6). Protein expression was induced by adding IPTG to a final concentration of 0.5 mM. Following induction, the temperature was reduced to 18°C, and the cultures were incubated for an additional 16 h. Cells were harvested by centrifugation at 6000 × g for 30 min at 4°C. Pellets were resuspended in 25 ml resuspension buffer (50 mM sodium phosphate, pH 7.4, 500 mM NaCl, 50 mM imidazole, 1 mM DTT, and 1 mM PMSF). The resuspended cells were lysed by sonication on ice using a pulse setting (10 s on, 20 s off, for a total of 10 min). The lysate was clarified by centrifugation at 12 000 rpm (∼20 000 × g) for 1 h at 4 °C, followed by filtration through a 0. 5 μm syringe filter. The clarified lysate was loaded onto a 5 mL HisTrap HP column (Cytiva) pre-equilibrated with buffer (50 mM sodium phosphate, pH 7.4, 500 mM NaCl, 50 mM imidazole, and 1 mM DTT). The column was washed with 10% elution buffer (50 mM sodium phosphate, pH 7.4, 500 mM NaCl, 500 mM imidazole, and 1 mM DTT) and eluted using a step gradient with 50% elution buffer. Eluted fractions were collected and analysed for purity and yield, and selected fractions were dialysed against Protein Dilution Buffer (PDB; 25 mM Tris–HCl (pH 7.5), 1 mM DTT, 1 M NaCl, and 50% v/v glycerol), and stored at −20 °C.

In vitro recombination of supercoiled plasmid substrates and product analysis

Purified integrases and RDFs were stored at −20 °C and diluted to the appropriate concentrations at 0 °C just before use. Integrases and RDFs are diluted in Protein Dilution Buffer as described above. *In vitro* recombination reactions were carried out as described in Abioye *et al.* [35]. In summary, reactions were initiated by adding integrase (8 μM, 5 μl) to a 30 μl solution containing the appropriate substrate plasmid DNA (25 μg/ml), 50 mM Tris–HCl (pH 7.5), 100 μg/ml BSA, 5 mM spermidine, and 0.1 mM EDTA. For reactions involving integrase and RDF, equal volumes of integrase (16 μM) and RDF (16 μM) were mixed thoroughly and kept on ice for 15 min [35, 36], after which 5 μl of this mixture was added to the reactions. Reaction samples were incubated at 30 °C for 2 or 16 h, after which the reactions were stopped by heating at 80 °C for 10 min to denature the proteins. The samples were cooled and treated with NruI (New England Biolabs) to facilitate separation and analysis of recombination products. This was done by mixing a 30 μl aliquot of the reaction mixture with 28 μl of a buffer containing 90 mM Tris–HCl, pH 7.5, 20 mM $MgCl_2$ prior to addition of 20 units (2 μl) NruI (New England Biolabs). The restriction digests were carried out at 37°C for 2 h. Following the digest, samples were treated with SDS and protease K by adding 7.5 μl loading buffer [25 mM Tris–HCl, pH 8.2, 20%(w/v) Ficoll, 0.5% sodium dodecyl sulphate, 1 mg/ml proteinase K, and 0.25 mg/ml bromophenol blue] to the reaction sample and incubated at 50°C for 30 min. The reaction products were separated by by electrophoresis on 1.2% agarose gels in 1× TAE, then stained with SYBR safe and visualized as previously described, using a BioRad GelDoc apparatus [35, 37]. Digital images of the gels are shown in reverse contrast.

In vitro exonuclease reactions

Assay of *in vitro* exonuclease activity of Bxb1 RDF was based on a method used to characterize the 3′-5′ exonuclease activity of *Pyrococcus abyssi* DP1 [37]. The DNA substrate is a 27mer top strand labelled with a 5′ fluorescein label (5′-/56-FAM), and with four non-complementary bases at the

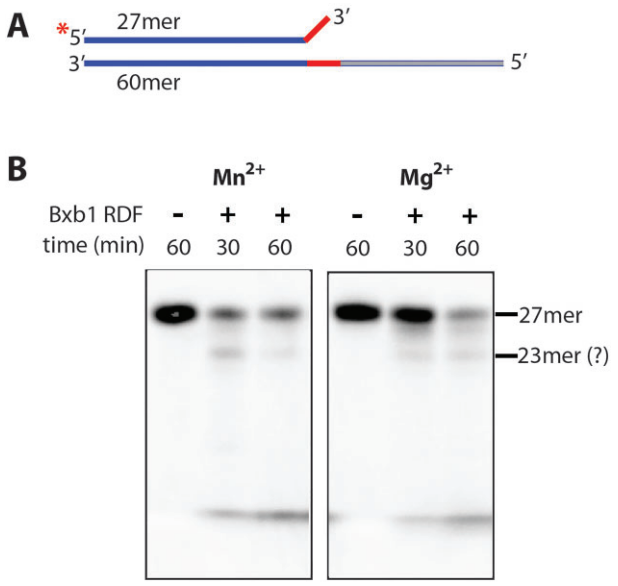


Figure 3. Assay of exonuclease activities of Bxb1 RDF. (**A**) Design of DNA substrate for 3′-5′ exonuclease activity. The DNA substrate is a 27mer top strand labelled with a 5′ fluorescein label (5′-/56-FAM), and with four non-complementary bases at the 3′ end annealed to a 60mer unlabelled bottom strand. The region of the top and bottom strands with complementary base sequences are depicted in blue, the non-complementary bases are in red, and the single-stranded region of the 60mer are shown in grey. (**B**) Reactions were carried out for 30 or 60 min in the reaction buffer described in the 'Materials and methods' section (Bxb1 RDF, 2 and 0.5 μM DNA). Reaction products were analysed on 1.2% denaturing polyacrylamide gel electrophoresis.

3′ end (5′ACGCCAGGCTTCGCCAGTCACGATACT3'), annealed to a 60mer unlabelled bottom strand (5′GCGGACTGCGATCGTACCTACGGACCTGC AGCTGACGTCGTGACTGGCGAAGCCTGGCGT3') to (Fig. 3A). Reactions were started by adding 2 μM Bxb1 RDF protein to a reaction that contains 50 mM Tris–HCl, pH 7.5, 2 mM $MnCl_2$ or $MgCl_2$, 100 mM NaCl, 15% polyethyleneglycol 4000, and 0.5 μM DNA. Assays were carried out at 37 °C for 30 or 60 min, and reactions were stopped by adding 1 volume of formamide loading buffer (80% deionised formamide, 10 mM EDTA pH 8.0, 1 mg/ml xylene cyanol, and 1 mg/ml bromophenol blue). The samples were heated at 80°C for 5 min, and reaction products were separated through a 12% 1 × TBE denaturing (7.5 M urea) polyacrylamide gel (120 V for 2.5 h; 65 °C) and imaged using a fluorescence imager (iBright FL1500, ThermoFisherScientific). Digital images of the gels are shown in reverse contrast.

Results

TG1 and Bxb1 RDFs are structurally homologous to DNA replication proteins

We recently showed that AlphaFold2-predicted structures of serine integrase RDFs reveal a wide diversity of structures with no obvious pattern to explain their role as RDFs for the otherwise highly conserved LSIs [19]. Despite this structural variation, most of the RDFs are relatively small proteins. The RDFs for SPbeta, A118, ϕRV1, Nm60, Bt24, Int10, and Int30 range between ∼7 and 9.3 kDa [19, 38–40]. In contrast, RDFs of the ϕC31 family (ϕC31, ϕBT1, and TG1, ∼27 kDa) and Bxb1 (28 kDa) are larger proteins [11, 12]; Fig. 2 in [19]. However, for all of these RDFs, models of their complexes with their respective integrases predict interactions at the same integrase CC/DBD2 junction [19, 31]. Our search for characterized proteins with structures similar to those predicted for the two known types of RDF—those for phages Bxb1 and TG1—strongly supported the moonlighting hypothesis.

Bxb1's RDF was previously noted to be required for phage DNA replication and to have sequence homology, including conservation of catalytic residues, with purple acid phosphatase-family enzymes (also refered to as calcineurin-like phosphoesterases) [11, 23]. Its predicted structure is in agreement with that observation (Fig. 2A). In our Dali search for structurally related proteins, the top hits (after a hypothetical protein) are Mre11 [41], which is an endo/exonuclease involved in double-strand break repair, and the small subunit ('DP1') of the D-family polymerase of *Pyrococcus abyssi*, which is an editing exonuclease (Supplementary Fig. S1). Editing exonucleases work in concert with DNA polymerases during replication to remove mismatched nucleotides from the 3′ end of the growing chain, and they can belong to more than one structural family [42]. Both DP1 and Mre11 have a calcineurin-like fold [37]. Slightly lower on the list is a UDP-2,3-diacylglucosamine hydrolase [43], but we feel that is an unlikely function for the Bxb1 RDF because, like Mre11 and DP1, the Bxb1 RDF lacks the lipid-binding 'cap' found within the calcineurin-like domain of the lipid hydrolyze enzyme [43]. Both Mre11 and DP1 have additional domains (with different folds) that mediate interactions with other DNA-binding proteins: for Mre11, with Rad50 and Nbs1, and for DP1, with the polymerase DP2 as well as with sliding clamp PCNA [44]. DP1 includes a domain with an OB fold and that was proposed to bind ssDNA [37]. Our AlphaFold-base virtual pulldown failed to convincely identify interaction partners for the Bxb1 RDF. Although useful for generating hypotheses, such a virtual pulldown could fail for many reasons, such as that the interaction requires cooperative binding to DNA substrate. Based on all of the considerations above, and the fact that the Bxb1 RDF is essential for phage DNA replication [23], we consider that the Bxb1 RDF could be an editing exonuclease that has evolved an as-yet-unknown mechanism for interacting with other components of the phage's replication apparatus.

The predicted fold of the RDF used by phage TG1 strongly suggests that it is a SSB protein (Fig. 2B). The Dali search for structural homologs retrieved numerous OB fold domains, which are often used to bind ssDNA [45]. The top hit was an ssDNA binding domain of BRAC2. However, this protein lacks the zinc-binding site predicted in the TG1 RDF. The first hit on the list that featured an OB fold with a zinc-binding site was gp32 from phage T4, a known SSB [46]. Although Alphafold3 predicted that ssDNA would bind in the positively charged cleft of both the TG1 RDF and phage T4 gp32, that is not shown in Fig. 2 because there are no experimental structures showing ssDNA bound to T4 gp32. In addition to the ssDNA-binding core shown in Fig. 2, gp32 has an N-terminal extension involved in multimerization (and thus cooperative ssDNA binding), and a C-terminal helicase-recruiting domain [47]. The TG1 RDF has a 27 residue N-terminal extension that is predicted to be poorly ordered. However, a virtual pulldown from the TG1 phage genome using this peptide as bait predicted that it interacts with other copies of the RDF and with a family A polymerase (Supplementary Fig. S2). The same

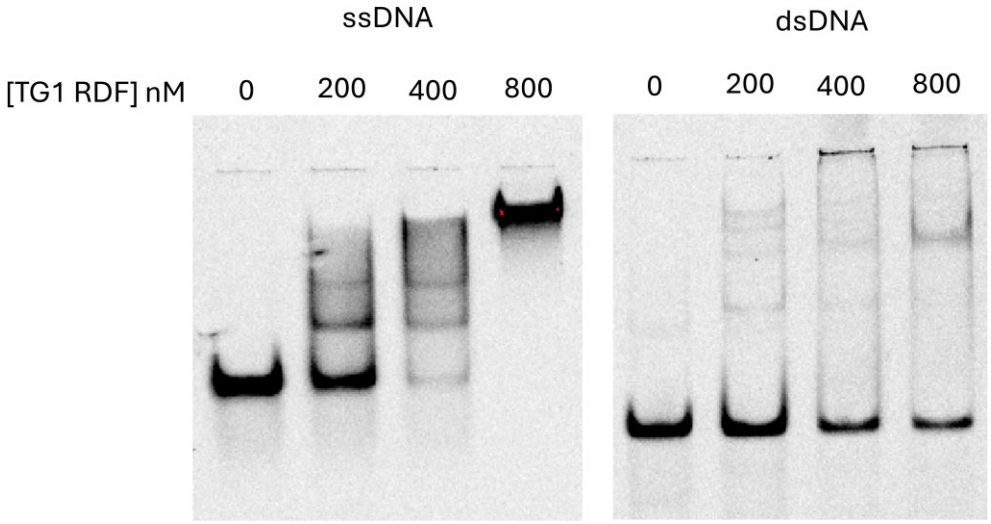


Figure 4. TG1 RDF is a SSB protein *in vitro*. Electrophoretic mobility shift assays of TG1 RDF binding to ssDNA (left) and dsDNA (right). The protein was present at 0, 200, 400, and 800 nM and the 50 nt/bp fluorescein-labeled DNA substrate was kept constant at 200 nM.

three hydrophobic residues (F4, L6, and W7) are predicted to be central to both interactions. This is conceptually similar to *E. coli* SSB, which uses a C-terminal tail to interact with numerous other proteins [48].

Furthermore, within each of these phages, the gene encoding the RDF is closer to other replication-related genes than to the integrase gene—this was previously noted for Bxb1 [18] and is also true for TG1 (Supplementary Fig. S3).

Bxb1 RDF shows limited exonuclease activity

We tested Bxb1 RDF for 3′-5′ exonuclease activity in an *in vitro* assay previously used to characterize *Pyrococcus abyssi* DP1 [37]. To test for exonuclease activity we used a substrate similar to the design used in Sauguet *et al.*, [37]. In this design, a 5′FAM-labelled 27mer DNA substrate with four non-complementary bases at the 3′ end was annealed to a 60mer unlabelled strand (Fig. 3A).

The results show that our sample of the RDF shows a limited 3′ exonuclease activity in the presence of Mn^{2+} or Mg^{2+} (Fig. 3B). In addition, we also observed liberation of the labelled 5′ end, which could indicate 5′ exonuclease activity endonuclease activity, or processive 3′ exonuclease activity. Although we cannot rule out contamination of our sample with trace amounts of other nucleases, our results agree well with our structural predictions and with the Hatfull lab's finding that this protein is required for phage DNA replication.

TG1 RDF is a phage SSB protein

Based on the similarity between the predicted structure of TG1 RDF and the experimental structure of phage T4 SSB, we tested the ability of TG1 RDF to bind ssDNA using an *in vitro* EMSA binding assay. The results show that the RDF binds ssDNA in a concentration dependent manner (Fig. 4). In contrast, the RDF does not bind double-stranded DNA.

Next, we asked if binding ssDNA could interfere with the RDF function of TG1 RDF (Fig. 5). As shown in Fig. 5C, the presence of ssDNA reduces the extent of activation of *attR* × *attL* reaction by the RDF (the ratio of non-recombinant to recombinant product increases in the presence of ssDNA). Correspondingly, ssDNA relieves the inhibition of *attP* × *attB* reaction by the RDF (Fig. 5D) (the ratio of non-recombinant to recombinant product decreases in the presence of ssDNA). In contrast, ssDNA has no effect on *attP* × *attB* recombination by the integrase when the RDF is not included in the reaction (Fig. 5E), showing that the effect of ssDNA on the reactions seen in Fig. 5C and D are due to interference with RDF function. This shows that TG1 RDF binds ssDNA using the same surface as it uses to bind the integrase.

Experimental data support the modeled RDF-integrase interfaces

The Hatfull lab has extensively characterized the Bxb1 integrase and its RDF [11, 23, 49]. Supplementary Fig. S4 shows their mutations onto the AlphaFold2 model. The majority of those mutations that disrupted RDF function (pink) map at or near the predicted protein–protein interface, and the few outliers may affect overall protein structure (e.g. two of them added a side chain to a glycine). In contrast, mutations in conserved functional motifs that mark the putative active site (green) did not disrupt RDF function and do not map to the predicted protein–protein interface.

Next, we tested the functional relevance of the integrase-RDF interface predicted by AlphaFold2-Multimer. We made specific amino acid changes on both the integrase and RDF (Fig. 6) and tested the effects of the mutations on recombination activities *in vitro* (Figs 7 and 8). These mutations were spread across the predicted interface, and were designed to alter the charge in hydrophilic patches or to remove large hydrophobic side chains that could be important in the protein–protein interactions. As described below, the results of our mutational analysis are consistent with the model shown in Fig. 6, and suggest that the portion of the interface highlighted in Fig. 6E [involving the base of the integrase's CC (pale green)] is more important than the portion highlighted in Fig. 6D.

First, we tested the activities of the mutant TG1 integrases in *attP* × *attB* reactions (which should be inhibited by the RDF) and in *attR* × *attL* reactions (which should be stimulated by the RDF) (Fig. 7). Mutant 1 was less active than WT integrase: it showed limited activity on *attP* × *attB* after 2 h (Fig. 7A, lanes 2 versus 4), but more noticeable activity after

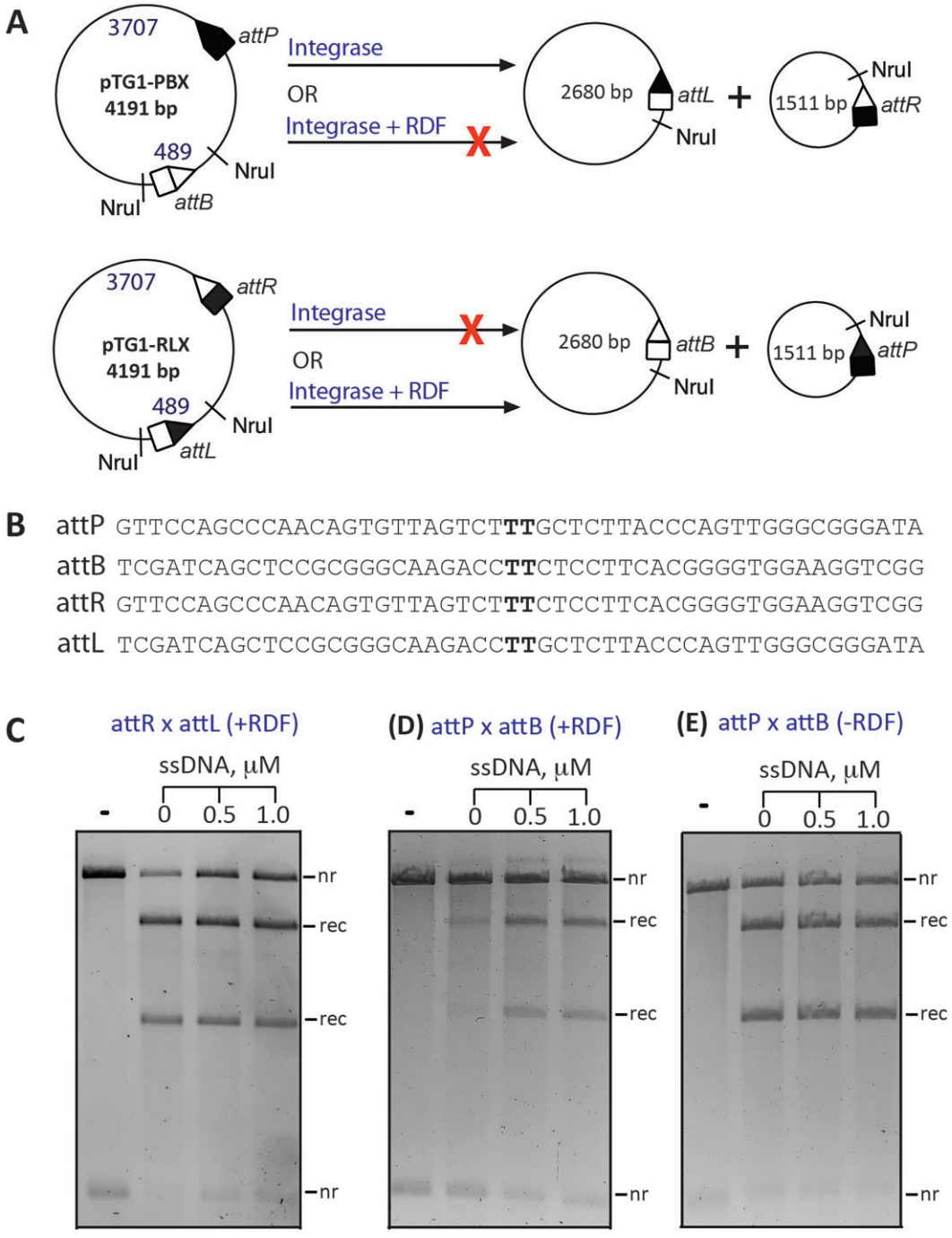


Figure 5. Single-strand DNA interferes with recombination functions of TG1 RDF *in vitro*. (**A**) Scheme depicting the *in vitro* intramolecular recombination reaction. Recombination of the *attP* × *attB* plasmid substrate (pTG1-PBX) gives two circular products in which the *attR* and *attL* sites are separated. Presence of the cognate RDF inhibits *attP* × *attB* recombination reaction. For the *attR* × *attL* recombination reaction, in the presence of the cognate RDF, the starting substrate plasmid (pTG1-RLX) is recombined to give *attP* and *attB* sites on separate circular plasmid products. (**B**) Sequences of recombination *att* sites for TG1 integrase. The effects of ssDNA on RDF function in *in vitro* recombination reactions are shown in panels (C–E). (**C**) *attR* x *attL* in the presence of RDF and ssDNA, (**D**) *attP* × *attB* in the presence of RDF and ssDNA, and (**E**) *attR* × *attL* in the presence of ssDNA, without RDF. In all cases, reactions were carried out for 2 h in the reaction buffer described in the 'Materials and methods' section (integrase, 800 nM; RDF, 800 nM), with the addition of 0, 0.5, and 1.0 μM ssDNA (polydT) as indicated. Reaction products were digested with the restriction endonuclease NruI prior to 1.2% agarose gel electrophoresis. The bands on the gel are labeled *nr* (non-recombinant, i.e. substrate), *rec* (recombination product).

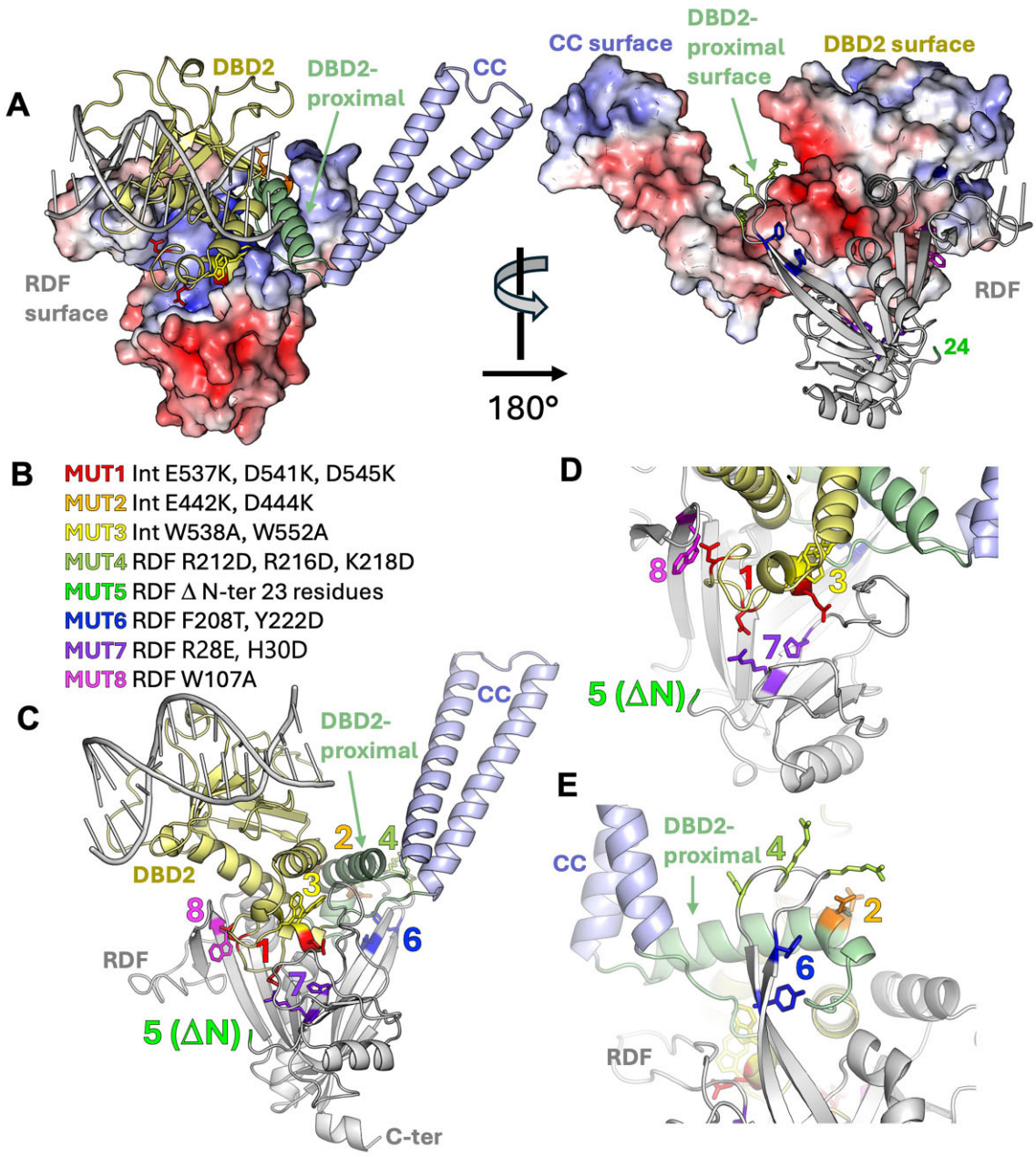


Figure 6. Integrase-RDF interface that mediates regulation of recombination. (**A**) Surface complementary of the predicted TG1 integrase–RDF interface. Left: the C-terminal domain of the integrase is shown as ribbons (DBD2, yellow; DBD2-proximal portion of the CC, pale green; CC pale blue), with DNA docked based on 4kis.pdb to guide the eye [53]. A surface representation of the RDF is shown, colored by electrostatic potential (blue = +5 kT/e, shading to red = −5 kT/e). Right: the RDF is shown as a gray ribbon, with the integrase surface colored by electrostatic potential (same scale). Side chains that were mutated are shown as colored sticks (see panel B). The C-terminus of the integrase (600–619) and the N-terminus [1–22] of the RDF are not shown as they were predicted to be disordered. (**B**) Table of mutations made. (**C**) The model shown in panel (A), with side chains mutated shown as sticks, color coded as in panel (B). (**D**, **E**): close-up views of integrase–RDF interactions and the side chains that were mutated.

16 h (Fig. 7B, lane 4). Given the overall defect of this mutant, it is unclear if there was an additional defect in RDF-mediated inhibition of its *attP* × *attB* reaction. However, despite showing some catalytic activity in the *attP* × *attB* reaction, mutant 1 completely failed to be stimulated by the RDF in an *attR* × *attL* reaction, even after 16 h, indicating a defect in communication with the RDF (Fig. 7C and D, lanes 5 and 6). Mutant 2, on the opposite side of DBD2 from mutant 1, showed near-WT activity in the *attP* × *attB* reaction (Fig. 7A, lanes 2 versus 7 and Fig. 7B lanes 2 versus 7), but was defective in inhibition by the RDF (Fig. 7B, lane 3 versus 8 and 9). Furthermore, Mutant 2 was inefficiently stimulated by the RDF in the *attR* × *attL* reaction (Fig. 7C and D, lane 3 versus 8 and 9). Finally, the interactions of mutant 3 with the RDF could not be assessed because it was almost completely lacking in integrase activity (Fig. 7B, lanes 10–12). This mutation was designed to test a more central region of the predicted interface, but the rather drastic change of two tryptophans to alanines probably destabilized the fold of DBD2.

Next we tested specific mutations in the TG1 RDF in similar assays. Mutations 4 and 6 showed the strongest effect—these RDFs completely failed to inhibit attP × attP recom-

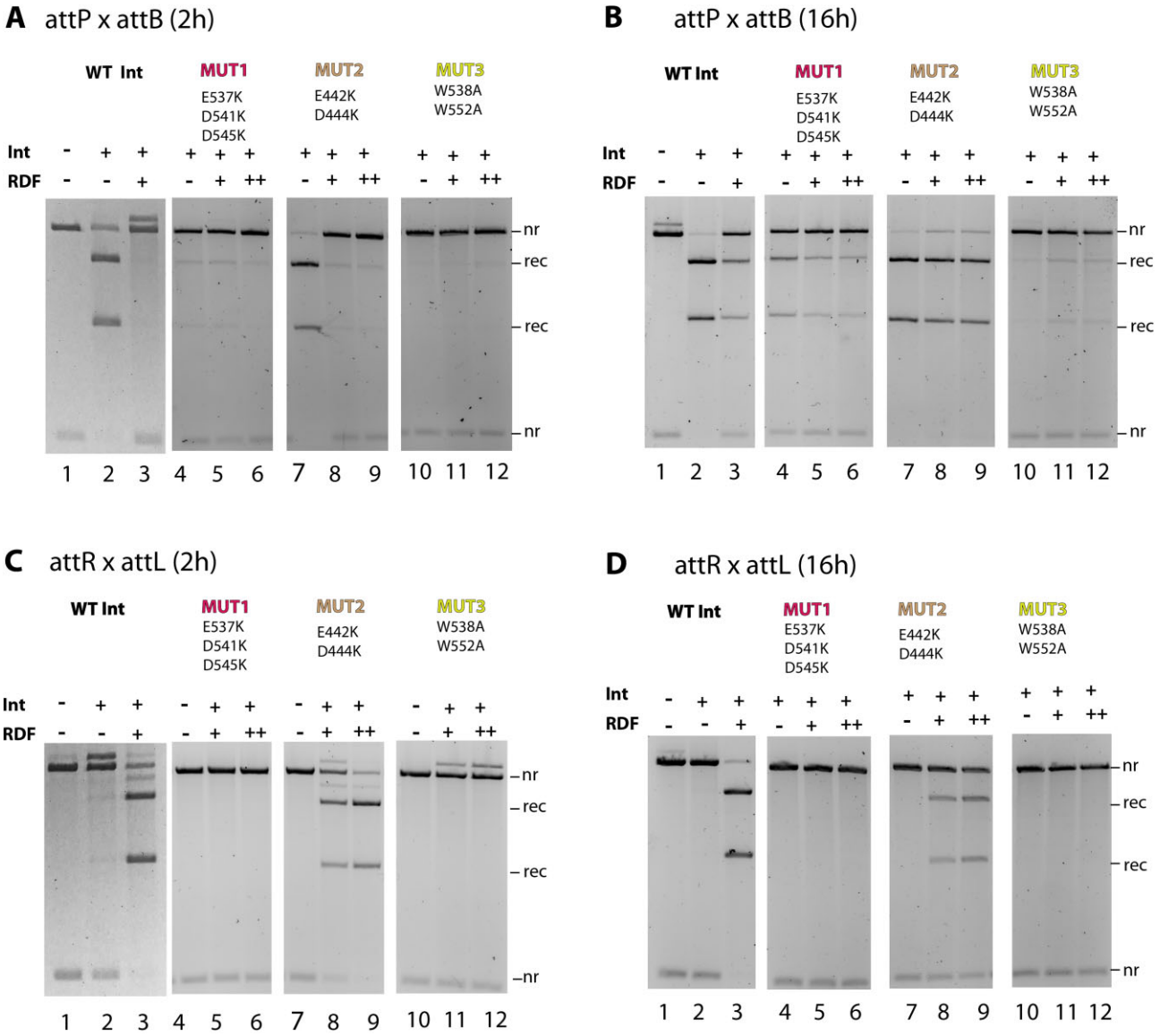


Figure 7. *In vitro* recombination activities of TG1 integrase mutants. Reactions were carried out for 2 or 16 h in the reaction buffer described in the 'Materials and methods' section. (**A**) Two hours reaction, *attP* × *attB*. Integrase was not added to reaction in lane 1. (**B**) Sixteen hours reaction, *attP* × *attB*. Integrase was not added to reaction in lane 1. (**C**) Two hours reaction, *attR* × *attL*. Integrase was not added to reaction in lane 1. (**D**) Sixteen hours reaction, *attR* × *attL*. Integrase was not added to reaction in lanes 1. Reaction products were digested with the restriction endonuclease NruI prior to 1.2% agarose gel electrophoresis. The bands on the gel are labeled *nr* (non-recombinant, i.e. substrate), *rec* (recombination product). The final integrase concentrations in all reactions were 800 nM. The final RDF concentrations were 800 nM (+) or 1600 nM (++). The bands on the gel are labelled *nr* (non-recombinant, i.e. substrate) or *rec* (recombination product). The sizes of the products of the recombination reactions are shown alongside each band on the gel image. MUT1: TG1 Int-3K (E537K, D541K, D545K). MUT2: TG1 Int-2K (E442K, D444K). MUT3: TG1 Int-2A (W538A, W552A).

bination or to stimulate the *attR* × *attL* reaction (Fig. 8A and B, lanes 3 and 4 versus lanes 5, 6, 9, and 10). Figure 6E shows that they cluster with integrase mutation 2, at a point where the RDF is predicted to bind the DBD2-proximal segment of the CC (pale green). Mutations 7 and 8 had a modest effect on RDF function (Fig. 8A and B, lanes 3 and 4 versus lanes 11–14). Figure 6D shows that these lie on the other side of the predicted interface. This result complements the findings that integrase mutation 1 at the same interface resulted in loss of RDF activity (Figs 6D and 7). Finally, RDF mutation 5 (Fig. 6B, green), deletion of the N-terminal end (23 aa residues) had no effect on function (Fig. 8A and B, lanes 3 and 4 versus 7 and 8), an unsurprising result since that region of the RDF was predicted to be disordered and not in contact with the integrase (Fig. 6C; only residue 23 is shown).

Discussion

In this study, we provide evidence that two different phage DNA replication proteins have been recruited to 'moonlight' as RDFs for the serine integrase of the phage that encodes them. We experimentally verified that the TG1 RDF is a functional SSB protein, and further virtual pulldown suggests it interacts with a phage polymerase enzyme. We showed that our sample of purified Bxb1 RDF has exonuclease activity, albeit weak. Based on the data shown here, we were unable

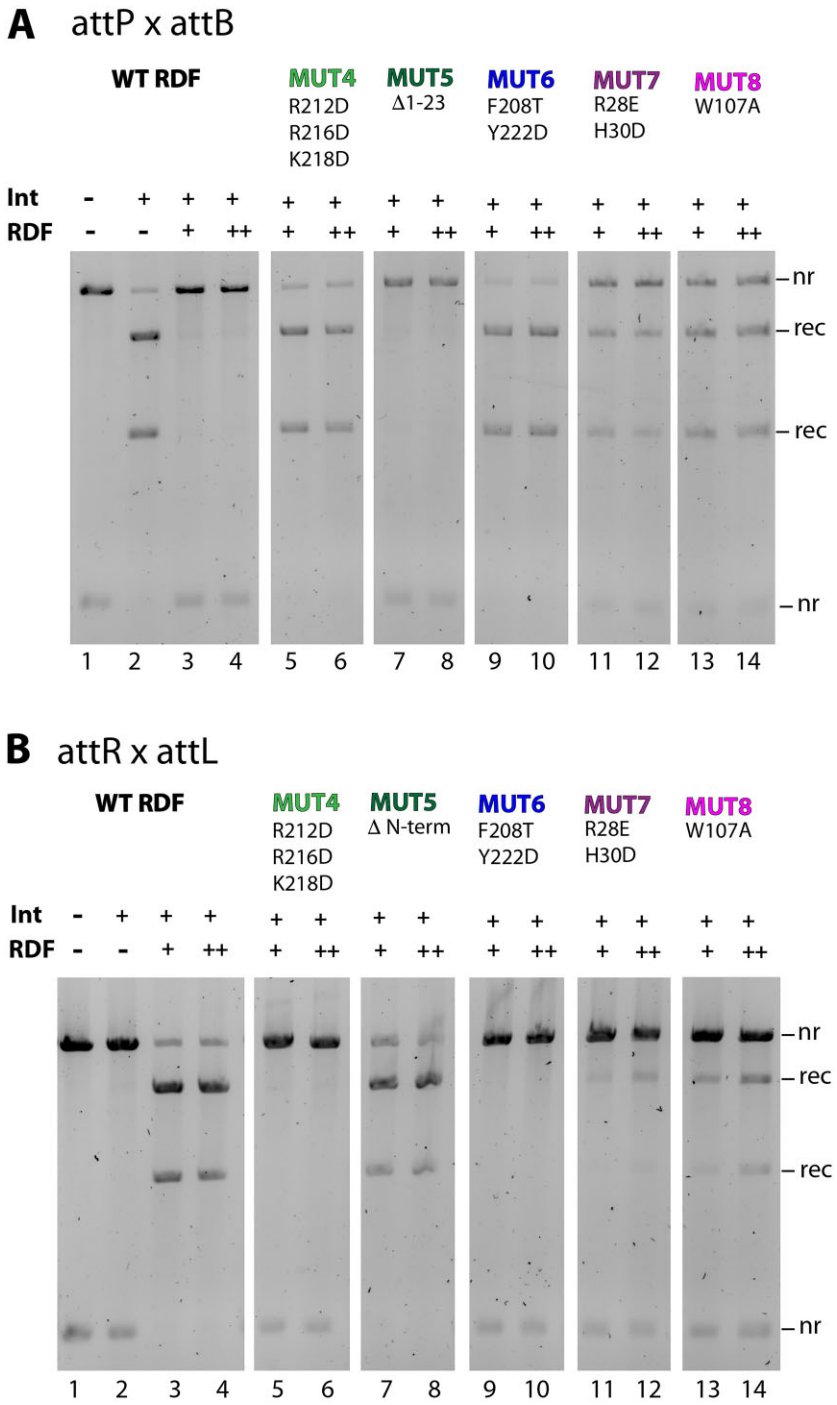


Figure 8. *In vitro* recombination activities of TG1 RDF mutants. Reactions were carried out for 2 h in the reaction buffer described in the 'Materials and methods' section. (**A**) *attP* × *attB* recombination reactions. (**B**) *attR* × *attL* recombination reactions. Reaction products were digested with the restriction endonuclease NruI prior to 1.2% agarose gel electrophoresis. TG1 integrase (800 nM) was used in all reactions. Integrase was not added to reactions in lane 1 of panels (A) and (B). The final RDF concentrations were 0 nM (-), 800 nM (+), or 1600 nM (++). The bands on the gel are labelled *nr* (non-recombinant, i.e. substrate) or *rec* (recombination product). The sizes of the products of the recombination reactions are shown alongside each band on the gel image. MUT4: TG1 RDF (R212D, R216D, K218D). MUT5: TG1 RDF (Δ23). MUT6: TG1 RDF (F208T, Y222D). MUT7: TG1 RDF (R28E H30D). MUT8: TG1 RDF (W107A).

to pinpoint the specific nature of Bxb1 RDF exonuclease activity, nor rule out endonuclease activities. As shown through our bioinformatic and structural comparison analysis above, it is likely that additional phage proteins are required for the optimization and regulation of any functionally-relevant endo- or exonuclease activity it might have. Interestingly, Mre11, one of the top hits for Bxb1 RDF in the virtual pulldown results (Supplementary Fig. S1), has both endo- and exonuclease activities [41]. It has already been shown that the Bxb1 RDF is essential for phage DNA replication [23].

Our work strongly supports and expands previous proposals suggesting that recruitment of phage proteins for RDF function is a mechanism through which both LSIs and tyrosine-family integrases mediate excision reactions [2, 11, 12, 18].

Repurposing phage proteins to perform secondary functions allows frugal use of limited genome space, and is not limited to RDFs [50]. For example, phage-inducible pathogenicity islands (PICIs) use phage proteins to moonlight as derepressors to lift repression of SapI induction. Examples of proteins used as PICI derepressors are dUTPase (SapIbov1 and SaPIbov5) [28]; Sri, a phage protein involved in blocking bacterial DNA replication (SaPI1) [28]; and DNA-single strand annealing proteins (SapI2) [51]. Although moonlighting depressors and RDFs are similar in concept, the evolutionary paths to their repurposing may be different.

Integrase proteins share an RDF-binding hotspot

The findings from this study, our previous modeling work [19], and the work of others [39, 39, 52–56] all suggest that LSIs share a common RDF-binding hotspot, even though the RDFs do not share a common structure. Although there are no published experimental structures of integrase–RDF complexes, we can now use Alphafold to make testable predictions [19]. In addition to the mutational testing of the TG1–RDF interface model described here, we previously showed that alphafold2-multimer models are largely in agreement with prior mutational data for ϕC31, SPbeta, and Bxb1 (Supplementary Fig. S2 of Shin *et al*., [19]; [11, 54, 57]. Previous work has also shown that the RDF redirects the CC, and that simply binding DBD2 is insufficient for function [39, 52, 54]. We deduce that the RDF-binding hotspot lies at the DBD2- CC junction, and includes the DBD2-proximal portion of the CC. That portion was poorly ordered in the experimental structure of an LSI's DNA binding domains with DNA [53] and is not predicted to be helical in all integrases, although we refer to it as part of the CC for historical simplicity. We suggest that contacts to DBD2 itself may provide affinity for the RDF-integrase interaction, and that contacts to the DBD2-proximal CC segment may be responsible for redirecting the trajectory of the CC.

Here we tested a detailed Alphafold2-multimer model for the TG1 RDF-integrase interface. Although mutations across the predicted interface interfered with RDF function, the strongest effects were seen for mutations in the portion of the interface where the RDF is predicted to grip the DBD2-proximal portion of the integrase's CC (Fig. 6E). Our finding that ssDNA interferes with TG1-RDF function *in vitro* (Fig. 5) provides additional support for the model, because the positively charged surface on the RDF that is expected to bind ssDNA (Fig. 2) is also used to bind the integrase (Fig. 6A). These findings support the predicted integrase-RDF interface, and suggest additional insights into how the RDF switches the directionality of the recombination reaction. We propose that the RDF partially restrains the mobility of the CC by binding to the DBD2-proximal segment of it (green in Fig. 6 and Supplementary Fig. S4). Experimental structural data are required to gain better understanding of how the interaction of the RDF with the CC domain control recombination directionality.

An intriguing feature not yet tested is the negatively charged, unstructured C-terminal tail of TG1 integrase—a feature not found in most LSIs, but shared by ϕC31 and ϕBT1 integrases, and others which use RDFs closely related to TG1's [31]. This tail may mimic ssDNA and compete with it for initial, transient binding to the RDF, which could lead to a high local concetration of RDF.

How and why do phages moonlight existing proteins to take up RDF functions?

We hypothesize that a phage LSI could adopt a new RDF when it becomes genetically separated from its previous RDF (Fig. 9). This could occur due to simultaneous activation of multiple prophages within a single host cell [58, 59]. Recombination between replicating phage genomes could result in chimeric progeny encoding one parent phage's LSI but not its cognate RDF. Such recombinant phage would be capable of infecting and lysogenizing a new host, but their further spread would be limited until they evolved to use another phage-encoded protein as an RDF. It is conceivable that most LSI RDFs could be scavenged proteins whose functions are yet to be characterized. Phages encode many small proteins of unknown functions that may also have been co-opted as RDFs. It is also likely that the original functions of some scavenged RDFs have been taken over by other phage proteins while some retain their original function.

The two previously identified RDFs studied in this work, those for Bxb1 and TG1, are both DNA replication proteins. Not only do they have the predicted structures of replication-related proteins, but the Bxb1 RDF is known to be required for phage replication as well as for excision [23] and in this work we showed that the TG1 RDF has ssDNA-binding activity. That some prophages can initiate replication before excision has been shown experimentally [60]. However, a prophage that replicates within the host chromosome but never excises is likely to be an evolutionary dead end for both parties. Therefore, the integrase may be under strong selection for variants that can use one of the phage proteins already present as an RDF. This suggests that replication proteins may have been advatageous because they were present in multiple copies at this make-or-break stage in the prophage life cycle. The existence of a binding hotspot on the integrase that is remote from its catalytic domain may facilitate such evolution.

The RDF binding hotspot may provide both advantages and disadvantages to the prophage. For example, an insufficiently selective integrase might bind constitutively expressed host proteins, triggering unregulated excision of the prophage. On the other hand, host proteins that are only expressed under conditions that induce prophage activation could be good candidates for moonlighting RDFs. In fact, such scavenging of host proteins could explain why we could not identify putative RDFs for all of the LSIs targets in our virtual pulldown study, which only considered phage-encoded proteins as pos-

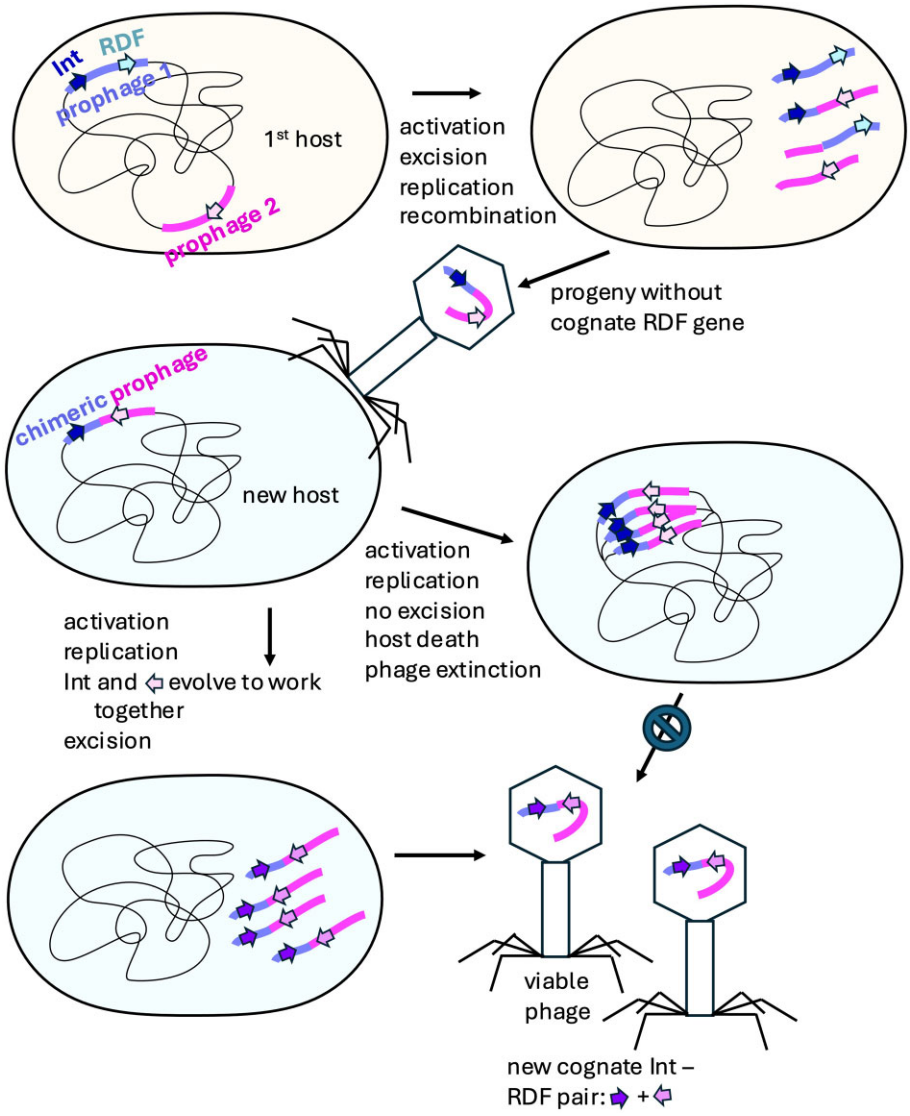


Figure 9. Proposed pathway for evolution of new RDFs. Recombination of two simultaneously activated prophages yields progeny with an integrase (Int) gene from prophage 1 but no cognate RDF gene (top). The chimeric phage can establish a new lysogen (middle left) but is unable to excise from the host chromosome until it evolves to use another phage protein (pink arrow) as an RDF. The mutations that allow the use of a new RDF could occur during the prophage state or during replication and could occur in either the integrase or the RDF (thus the resulting Int and new RDF genes are shown as purple and lavender rather than the original dark blue and pink).

sible RDFs [19]. There is also precedent for crosstalk among mobile genetic elements: The RDF for some serine integrases found in PICI-like elements (PLEs) are not encoded by the PLE but come from the phages parasitised by the PLE [61].

The existence of a hotspot on LSIs where a second protein can bind, redirect the CC, and change the preferred reaction direction has implications for the evolution of RDFs and of directionality in LSIs in general. The original LSI may have been a simpler, bidirectional enzyme that simply used its CC subdomains to stabilize the synaptic complex that holds the two recombining DNAs together. For example, the LSIs encoded by the SCC*mec* family of mobile elements do not appear to follow the directionality 'paradigm' of most characterized LSIs [62]. Fortuitous interactions of the LSI hotspot with a second protein that could redirect the CC's trajectory may have initiated the evolutionary path toward control of directionality.

Conclusion

In this report, we have shown that phage TG1 likely adapted its integrase to recognize a DNA single-stranded binding protein as a cofactor for excisive recombination. Structural models of how TG1 integrase interacts with its RDF built using AlphaFold2 and AlphaFold multimer were validated through biochemical characterization of integrase and RDF mutants. We also discuss bioinformatic analysis and some experimental evidence that the Bxb1 RDF could be an exonuclease that may have a role DNA replication protein. Future identification of the primary roles of other RDFs will rely on a highthroughput version of AlphaFold Multimer that that can process virtual pulldowns on a metagenome-wide level. It will be interesting to know the mechanism through which phages select proteins to repurpose as RDFs.

Acknowledgements

We thank Adebayo Bello for helpful comments on the manuscript.

Author Contributions: Femi Olorunniji and Phoebe Rice: Conceptualization, Methodology, Writing (review and editing), and Funding acquisition. Heewhan Shin: Methodology, Software. Abdulrazak Alsaleh, Alexandria Holland, Tania Peña Reyes, Aron Baksh, Oluwateniola Taiwo-Aiyerin, and Ying Pigli: Investigation, Methodology.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This work was supported by the Biotechnology and Biological Sciences Research Council grant (BBSRC BB/X012085/1) to F.J.O. and the National Science Foundation grant (NSF/BIO 2223480) to P.A.R. Funding to pay the Open Access publication charges for this article was provided by BBSRC BB/X012085/1.

Data availability

The data underlying this article are available in the article and in its online supplementary material.

References

1. Jayaram M, Ma C-H, Kachroo AH *et al.* An overview of Tyrosine site-specific recombination: from an flp perspective. *Microbiol Spectr* 2015;**3**:3.4.12. https://doi.org/10.1128/microbiolspec.MDNA3-0021-2014
2. Smith MCM. Phage-encoded serine integrases and other large serine recombinases. *Microbiol Spectr* 2015;**3**:3.4.06. https://doi.org/10.1128/microbiolspec.MDNA3-0059-2014
3. Olorunniji FJ, Rosser SJ, Stark WM. Site-specific recombinases: molecular machines for the Genetic Revolution. *Biochem J* 2016;**473**:673–84. https://doi.org/10.1042/BJ20151112
4. Grindley NDF, Whiteson KL, Rice PA. Mechanisms of site-specific recombination. *Annu Rev Biochem* 2006;**75**:567–605. https://doi.org/10.1146/annurev.biochem.73.011303.073908
5. Montano SP, Rowland S-J, Fuller JR *et al.* Structural basis for topological regulation of Tn3 resolvase. bioRxiv, https://doi.org/10.1101/2021.12.07.471667, 8 December 2021, preprint: not peer reviewed.
6. Rowland S, Boocock MR, Burke ME *et al.* The protein–protein interactions required for assembly of the tn *3* resolution synapse. *Mol Microbiol* 2020;**114**:952–65. https://doi.org/10.1111/mmi.14579
7. Mouw KW, Rowland S-J, Gajjar MM *et al.* Architecture of a serine recombinase-DNA regulatory complex. *Mol Cell* 2008;**30**:145–55. https://doi.org/10.1016/j.molcel.2008.02.023
8. Laxmikanthan G, Xu C, Brilot AF *et al.* Structure of a Holliday junction complex reveals mechanisms governing a highly regulated DNA transaction. *eLife* 2016;**5**:e14313. https://doi.org/10.7554/eLife.14313
9. Johnson RC. Site-specific DNA inversion by serine recombinases. *Microbiol Spectr* 2015;**3**:MDNA3–0047–2014. https://doi.org/10.1128/microbiolspec.MDNA3-0047-2014
10. Bibb LA, Hancox MI, Hatfull GF. Integration and excision by the large serine recombinase phiRv1 integrase. *Mol Microbiol* 2005;**55**:1896–910. https://doi.org/10.1111/j.1365-2958.2005.04517.x
11. Ghosh P, Wasil LR, Hatfull GF. Control of phage Bxb1 excision by a novel recombination directionality factor. *PLoS Biol* 2006;**4**:e186. https://doi.org/10.1371/journal.pbio.0040186
12. Khaleel T, Younger E, McEwan AR *et al.* A phage protein that binds φC31 integrase to switch its directionality. *Mol Microbiol* 2011;**80**:1450–63. https://doi.org/10.1111/j.1365-2958.2011.07696.x
13. Van Duyne GD, Rutherford K. Large serine recombinase domain structure and attachment site binding. *Crit Rev Biochem Mol Biol* 2013;**48**:476–91. https://doi.org/10.3109/10409238.2013.831807
14. Van Duyne GD. A structural view of cre-loxp site-specific recombination. *Annu Rev Biophys Biomol Struct* 2001;**30**:87–104. https://doi.org/10.1146/annurev.biophys.30.1.87
15. Biswas T, Aihara H, Radman-Livaja M *et al.* A structural basis for allosteric control of DNA recombination by λ integrase. *Nature* 2005;**435**:1059–66. https://doi.org/10.1038/nature03657
16. Landy A. The λ integrase site-specific recombination pathway. *Microbiol Spectr* 2015;**3**:MDNA3–0051–2014. https://doi.org/10.1128/microbiolspec.MDNA3-0051-2014
17. Van Duyne GD, Landy A. Bacteriophage lambda site-specific recombination. *Mol Microbiol* 2024;**121**:895–911. https://doi.org/10.1111/mmi.15241
18. Lewis JA, Hatfull GF. Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res* 2001;**29**:2205–16. https://doi.org/10.1093/nar/29.11.2205
19. Shin H, Holland A, Alsaleh A *et al.* Identification of cognate recombination directionality factors for large serine recombinases by virtual pulldown. bioRxiv, https://doi.org/10.1101/2024.06.11.598349, 11 June 2024, preprint: not peer reviewed.
20. Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–9. https://doi.org/10.1038/s41586-021-03819-2
21. Evans R, O'Neill M, Pritzel A *et al.* Protein complex prediction with AlphaFold-multimer. bioRxiv, https://doi.org/10.1101/2021.10.04.463034, 10 march 2022, preprint: not peer reviewed.
22. Abramson J, Adler J, Dunger J *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 2024;**630**:493–500. https://doi.org/10.1038/s41586-024-07487-w
23. Savinov A, Pan J, Ghosh P *et al.* The Bxb1 gp47 recombination directionality factor is required not only for prophage excision, but also for phage DNA replication. *Gene* 2012;**495**:42–8. https://doi.org/10.1016/j.gene.2011.12.003
24. Yu A, Haggård-Ljungquist E. The Cox protein is a modulator of directionality in bacteriophage P2 site-specific recombination. *J Bacteriol* 1993;**175**:7848–55. https://doi.org/10.1128/jb.175.24.7848-7855.1993
25. Esposito D, Scocca JJ. Identification of an HP1 phage protein required for site-specific excision. *Mol Microbiol* 1994;**13**:685–95. https://doi.org/10.1111/j.1365-2958.1994.tb00462.x
26. Saha S, Haggård-Ljungquist E, Nordström K. The cox protein of bacteriophage P2 inhibits the formation of the repressor protein and autoregulates the early operon. *EMBO J* 1987;**6**:3191–9. https://doi.org/10.1002/j.1460-2075.1987.tb02631.x
27. Esposito D, Wilson JCE, Scocca JJ. Reciprocal regulation of the early promoter region of bacteriophage HP1 by the Cox and CI proteins. *Virology* 1997;**234**:267–76. https://doi.org/10.1006/viro.1997.8646
28. Tormo-Más MÁ, Mir I, Shrestha A *et al.* Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. *Nature* 2010;**465**:779–82. https://doi.org/10.1038/nature09065

29. Leveles I, Németh V, Szabó JE *et al.* Structure and enzymatic mechanism of a moonlighting dUTPase. *Acta Crystallogr D Biol Crystallogr* 2013;**69**:2298–308. https://doi.org/10.1107/S0907444913021136
30. Singh MI, Ganesh B, Jain V. On the domains of T4 phage sliding clamp gp45: an intermolecular crosstalk governs structural stability and biological activity. *Biochim Biophys Acta Gen Subj* 2017;**1861**:3300–10. https://doi.org/10.1016/j.bbagen.2016.08.012
31. MacDonald AI, Baksh A, Holland A *et al.* Variable orthogonality of serine integrase interactions within the ϕC31 family. *Sci Rep* 2024;**14**:26280. https://doi.org/10.1038/s41598-024-77570-9
32. Mirdita M, Schütze K, Moriwaki Y *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* 2022;**19**:679–682. https://doi.org/10.1038/s41592-022-01488-1
33. Jurrus E, Engel D, Star K *et al.* Improvements to the APBS biomolecular solvation software suite biomolecular solvation software suite. *Protein Sci* 2018;**27**:112–28. https://doi.org/10.1002/pro.3280
34. Holm L, Laiho A, Törönen P *et al.* DALI shines a light on remote homologs: one hundred discoveries. *Protein Sci* 2023;**32**:e4519. https://doi.org/10.1002/pro.4519
35. Abioye J, Lawson-Williams M, Lecanda A *et al.* High fidelity one-pot DNA assembly using orthogonal serine integrases. *Biotechnol J* 2023;**18**:2200411. https://doi.org/10.1002/biot.202200411
36. Olorunniji FJ, McPherson AL, Rosser SJ *et al.* Control of serine integrase recombination directionality by fusion with the directionality factor. *Nucleic Acids Res* 2017;**45**:8635–45. https://doi.org/10.1093/nar/gkx567
37. Sauguet L, Raia P, Henneke G *et al.* Shared active site architecture between archaeal PolD and multi-subunit RNA polymerases revealed by X-ray crystallography. *Nat Commun* 2016;**7**:12227. https://doi.org/10.1038/ncomms12227
38. 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* 2021;**115**:1110–21. https://doi.org/10.1111/mmi.14654
39. Mandali S, Gupta K, Dawson AR *et al.* Control of recombination directionality by the Listeria phage A118 protein Gp44 and the coiled-coil motif of its serine integrase. *J Bacteriol* 2017;**199**:e00019-17. https://doi.org/10.1128/JB.00019-17
40. Bibb LA, Hatfull GF. Integration and excision of the mycobacterium tuberculosis prophage-like element, phiRv1. *Mol Microbiol* 2002;**45**:1515–26. https://doi.org/10.1046/j.1365-2958.2002.03130.x
41. Rotheneder M, Stakyte K, Van De Logt E *et al.* Cryo-EM structure of the Mre11-Rad50-Nbs1 complex reveals the molecular mechanism of scaffolding functions. *Mol Cell* 2023;**83**:167–85. https://doi.org/10.1016/j.molcel.2022.12.003
42. Aravind L, Koonin EV. Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res* 1998;**26**:3746–52. https://doi.org/10.1093/nar/26.16.3746
43. Okada C, Wakabayashi H, Kobayashi M *et al.* Crystal structures of the UDP-diacylglucosamine pyrophosphohydrase LpxH from Pseudomonas aeruginosa. *Sci Rep* 2016;**6**:32822. https://doi.org/10.1038/srep32822
44. Mayanagi K, Oki K, Miyazaki N *et al.* Two conformations of DNA polymerase D-PCNA-DNA, an archaeal replisome complex, revealed by cryo-electron microscopy. *BMC Biol* 2020;**18**:152. https://doi.org/10.1186/s12915-020-00889-y
45. Bianco PR. OB-fold families of genome guardians: a universal theme constructed from the small β-barrel building block. *Front Mol Biosci* 2022;**9**:784451. https://doi.org/10.3389/fmolb.2022.784451
46. Shamoo Y, Friedman AM, Parsons MR *et al.* Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA. *Nature* 1995;**376**:362–6. https://doi.org/10.1038/376362a0
47. He X, Yun M-K, Li Z *et al.* Structural and functional insights into the interaction between the bacteriophage T4 DNA processing proteins gp32 and dda. *Nucleic Acids Res* 2024;**52**:12748–62. https://doi.org/10.1093/nar/gkae910
48. Bonde NJ, Kozlov AG, Cox MM *et al.* Molecular insights into the prototypical single-stranded DNA-binding protein from *E. coli*. *Crit Rev Biochem Mol Biol* 2024;**59**:99–127. https://doi.org/10.1080/10409238.2024.2330372
49. Kim AI, Ghosh P, Aaron MA *et al.* Mycobacteriophage Bxb1 integrates into the *mycobacteriumsmegmatisgroEL1* gene. *Mol Microbiol* 2003;**50**:463–73. https://doi.org/10.1046/j.1365-2958.2003.03723.x
50. Penadés JR, Christie GE. The phage-inducible chromosomal islands: a family of highly evolved molecular parasites. *Annu Rev Virol* 2015;**2**:181–201. https://doi.org/10.1146/annurev-virology-031413-085446
51. Bowring J, Neamah MM, Donderis J *et al.* Pirating conserved phage mechanisms promotes promiscuous staphylococcal pathogenicity island transfer. *eLife*, 2017;**6**:e26487.
52. Rowley PA, Smith MCA, Younger E *et al.* A motif in the C-terminal domain of phiC31 integrase controls the directionality of recombination. *Nucleic Acids Res* 2008;**36**:3879–91. https://doi.org/10.1093/nar/gkn269
53. Rutherford K, Yuan P, Perry K *et al.* Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Res* 2013;**41**:8341–56. https://doi.org/10.1093/nar/gkt580
54. Fogg PCM, Younger E, Fernando BD *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* 2018;**46**:1308–20. https://doi.org/10.1093/nar/gkx1233
55. Mandali S, Johnson RC. Control of the serine integrase reaction: roles of the coiled-coil and helix E regions in DNA site synapsis and recombination. *J Bacteriol* 2021;**203**:e0070320. https://doi.org/10.1128/JB.00703-20
56. Chen Y-W, Su B-Y, Van Duyne GD *et al.* The influence of coiled-coil motif of serine recombinase toward the directionality regulation. *Biophys J* 2023;**122**:4656–69. https://doi.org/10.1016/j.bpj.2023.11.009
57. 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* 2017;**45**:6669–83. https://doi.org/10.1093/nar/gkx466
58. De Paepe M, Hutinet G, Son O *et al.* Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: the role of Rad52-like recombinases. *PLoS Genet* 2014;**10**:e1004181. https://doi.org/10.1371/journal.pgen.1004181
59. Dragoš A, Priyadarshini B, Hasan Z *et al.* Pervasive prophage recombination occurs during evolution of spore-forming *Bacilli*. *ISME J* 2021;**15**:1344–58. https://doi.org/10.1038/s41396-020-00854-1
60. Chen J, Quiles-Puchalt N, Chiang YN *et al.* Genome hypermobility by lateral transduction. *Science* 2018;**362**:207–12. https://doi.org/10.1126/science.aat5867
61. McKitterick AC, Seed KD. Anti-phage islands force their target phage to directly mediate island excision and spread. *Nat Commun* 2018;**9**:2348. https://doi.org/10.1038/s41467-018-04786-5
62. Misiura A, Pigli YZ, Boyle-Vavra S *et al.* Roles of two large serine recombinases in mobilizing the methicillin-resistance cassette SCCmec. *Mol Microbiol* 2013;**88**:1218–29. https://doi.org/10.1111/mmi.12253

Received: August 20, 2024. **Revised:** January 15, 2025. **Editorial Decision:** January 16, 2025. **Accepted:** January 21, 2025