Redefining the bacteriophage mv4 site-specific recombination system and the sequence specificity of its attB and core- attP sites

Kevin Debatisse, Pierre Lopez, Maryse Poli, Philippe Rousseau, Manuel Campos, Michèle Coddeville, Muriel Cocaign-bousquet, Pascal Le Bourgeois

HAL Id: hal-04666060
https://hal.inrae.fr/hal-04666060v1

Submitted on 1 Aug 2024

RESEARCH ARTICLE

WILEY

Redefining the bacteriophage mv4 site-specific recombination system and the sequence specificity of its *attB* and core-*attP* sites

Kevin Debatisse[1] | Pierre Lopez[1] | Maryse Poli[1] | Philippe Rousseau[2,3] | Manuel Campos[2,3] | Michèle Coddeville[2,3] | Muriel Cocaign-Bousquet[1] | Pascal Le Bourgeois[1,3]

[1]TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France

[2]CBI, LMGM, Université de Toulouse, CNRS, Toulouse, France

[3]Université Toulouse III - Paul Sabatier, Toulouse, France

Correspondence
Pascal Le Bourgeois, TBI-INSA, 135 avenue de Rangueil, 31077 Toulouse Cedex 09, France.
Email: pascal.lebourgeois@univ-tlse3.fr

Present address
Kevin Debatisse, Unité Plasticité du Génome Bactérien, Institut Pasteur, UMR 3525, Paris, France

Abstract

Through their involvement in the integration and excision of a large number of mobile genetic elements, such as phages and integrative and conjugative elements (ICEs), site-specific recombination systems based on heterobivalent tyrosine recombinases play a major role in genome dynamics and evolution. However, despite hundreds of these systems having been identified in genome databases, very few have been described in detail, with none from phages that infect *Bacillota* (formerly *Firmicutes*). In this study, we reanalyzed the recombination module of *Lactobacillus delbrueckii* subsp. *bulgaricus* phage mv4, previously considered atypical compared with classical systems. Our results reveal that mv4 integrase is a 369 aa protein with all the structural hallmarks of recombinases from the Tn*916* family and that it cooperatively interacts with its recombination sites. Using randomized DNA libraries, NGS sequencing, and other molecular approaches, we show that the 21-bp core-*attP* and *attB* sites have structural similarities to classical systems only if considering the nucleotide degeneracy, with two 7-bp inverted regions corresponding to ^mv4^Int core-binding sites surrounding a 7-bp strand-exchange region. We also examined the different compositional constraints in the core-binding regions, which define the sequence space of permissible recombination sites.

KEYWORDS
attB, *attP*, bacteriophage, heterobivalent tyrosine recombinases, integrase, *Lactobacillus delbrueckii*, protein–DNA interaction, site-specific recombination

1 | INTRODUCTION

Temperate bacteriophages are obligate parasites of bacterial cells with both lytic and lysogenic lifecycles. During infection, temperate phages typically initiate the lytic cycle, hijacking the host-cell machinery to replicate their DNA, assemble viral particles, and disseminate in the environment by lysing the host. However, depending on various factors (Casjens & Hendrix, 2015; Erez et al., 2017; Silpe & Bassler, 2019;

Zeng et al., 2010), temperate phages can enter the lysogenic cycle and incorporate their genome at a specific site within the host chromosome, referred to as a prophage. On some occasions, prophage DNA excises from the bacterial chromosome and reactivates the lytic cycle (Gandon, 2016). Prophages are present in nearly half of bacterial genomes (Touchon et al., 2016) and have a significant impact on bacterial genome dynamics (Feiner et al., 2015; Menouni et al., 2015), accounting for up to 20% of bacterial DNA (Casjens, 2003).

The integration and excision of phage DNA are mediated by phage-encoded proteins known as integrases. The most prominent subfamily among these integrases is heterobivalent tyrosine recombinases (HYRs), also referred to as "arm-binding domain-containing YRs" owing to their interaction with two distinct types of binding sites (arm-type and core-type, see below). However, an increasing number of serine integrases, which belong to a phylogenetically and mechanistically unrelated family of site-specific recombinases, have also been identified (Grindley et al., 2006; Rutherford & Van Duyne, 2014). Studied for over 50 years (Van Duyne & Landy, 2024), the recombination system of the coliphage *Lambdavirus lambda* represents the founding member of the HYR's subfamily, even if it was recently observed that its sequence differs from most YRs with specific modifications in its catalytic domain (Smyshlyaev et al., 2021).

HYRs catalyze unidirectional recombination between an approximately 250-bp site located on the phage DNA (*attP*), and a 20- to 30-bp site located on the bacterial chromosome (*attB*). *attP* sites have a complex structure (Figure 1a), which is required for the formation of a nucleoprotein complex called the intasome (Better et al., 1982). They contain several integrase binding sites, four to six arm-binding sites (P sites), and two core-binding sites (C and C′) arranged as inverted repeats (IR) on either side of the region where the strand exchange occurs, the overlap (O) sequence (Van Duyne & Landy, 2024). Intasome formation also requires additional binding sites for host factors, such as IHF and Fis in G*ammaproteobacteria*, which are located between the arm- and core-binding sites. *attB* sites are simpler (Figure 1a), consisting of two IR core-binding sites (B and B′) on either side of an overlap sequence identical to that of the *attP* site (Van Duyne & Landy, 2024). *attP* x *attB* recombination results in the creation of *attL* and *attR* sites bordering the prophage. The reverse reaction, which leads to prophage excision by recombination between the *attL* and *attR* sites (Figure 1a), restores the initial *attB* and *attP* sites and requires an additional phage-encoded protein, a recombination directionality factor (RDF, also named excisionase or Xis), which interacts with the integrase (Lewis & Hatfull, 2001).

Although hundreds of HYR systems have been identified in bacterial genomes (Smyshlyaev et al., 2021), with each integrase associated with specific *attP* and *attB* sites, only very few recombination sites have ever been fully characterized (Figure S1), namely, for *Haemophilus influenzae* phage HP1 (Esposito et al., 2001), *Escherichia coli* phages KplE1 and P2 (Panis et al., 2007; Sylwan et al., 2010), *Salmonella enterica* sv Typhimurium phage P22 (Smith-Mungo et al., 1994), *Mycobacterium tuberculosis* phage L5 (Lewis & Hatfull, 2003), *Bacteroidetes thetaiotaomicron* ICE CTnDOT (Wood et al., 2010), and *Enterococcus faecalis* ICE Tn*916* (Lu & Churchward, 1994). Each of these systems has a complex

attP site similar to phage *lambda*'s, but the number of binding sites, their spacing and orientation, and the intasome structure all differ and appear to be characteristic of each Int/*attP* pair (Esposito et al., 2001; Mattis et al., 2008; Panis et al., 2010; Peña et al., 2000). This underscores the importance of characterizing additional Int/*attP*/*attB* recombination modules to assess the biological diversity of HYR's systems.

The temperate bacteriophage mv4 (Mata et al., 1986) integrates its DNA at the 3′-end of the tRNA[Ser](CGA) locus of the *Lactobacillus delbrueckii* subsp. *bulgaricus* (hereafter abbreviated as *L. bulgaricus*) chromosome through site-specific recombination (Dupont et al., 1995). Several studies have shown that its recombination module differs from those of classical HYR systems (Auvray et al., 1997; Auvray, Coddeville, Espagno, & Ritzenthaler, 1999; Auvray, Coddeville, Ordonez, & Ritzenthaler, 1999; Coddeville et al., 2014), suggesting it operates through an alternative mechanism. In this study, we reinvestigated the organization of both *attB* and *attP* sites, through NGS of in vitro recombination experiments with various randomized DNA libraries and DNA-binding assays. With the aim of reprogramming the [mv4]Int/*attP* system to target alternative DNA sites, we also investigated which parts of the *attB* and core-*attP* sequence spaces are amenable to recombination.

2 | RESULTS

2.1 | [mv4]Int is a 369 amino acid tyrosine integrase

In a previous study of mv4 phage's integration module, [mv4]Int was described as a 427 amino acid (aa) protein with structural similarities to [λ]Int integrase (Dupont et al., 1995), despite lacking the structurally important D/E residue in the YR catalytic pocket (Figure 1b). However, the structure predicted by AlphaFold (Jumper et al., 2021) for the [mv4]Int(427aa) sequence differed from classical HYRs such as [λ]Int (Figure 1c), lacking a structured arm-binding domain and the β2 and β3 strands surrounding the catalytic pocket's lysine residue (Wojciak et al., 2002). The *int* gene from the pMC1 plasmid (Dupont et al., 1995) was then resequenced, revealing five single nucleotide deletions and two inversions (Figure S2a) compared with the initially published sequence. These nucleotide modifications were further validated for other [mv4]*int* gene-containing plasmids, such as pA3int and pET-Int (Auvray, Coddeville, Espagno, & Ritzenthaler, 1999), as well as for the PCR amplicon of the mv4 prophage *int* region (data not shown). These nucleotide changes had a profound impact on the [mv4]Int amino acid sequence, which was truncated to 369 aa, with further differences from the published sequence in 53 amino acids. The revised protein aligns identically or very closely to prophage integrases identified within the genomes of multiple *L. bulgaricus* strains (Figure S2b), and comparison of the catalytic domain (residues 172 to 369) with those of other HYRs confirmed the conservation of all the residues in the catalytic pocket (Figure 1b). The new structure also has a well-structured arm-binding domain and the canonical antiparallel β2-β3-strands (red arrows, Figure 1c), along with two additional antiparallel β-strands (blue arrows, Figure 1c) positioned between the β2-strand and the catalytic lysine, a
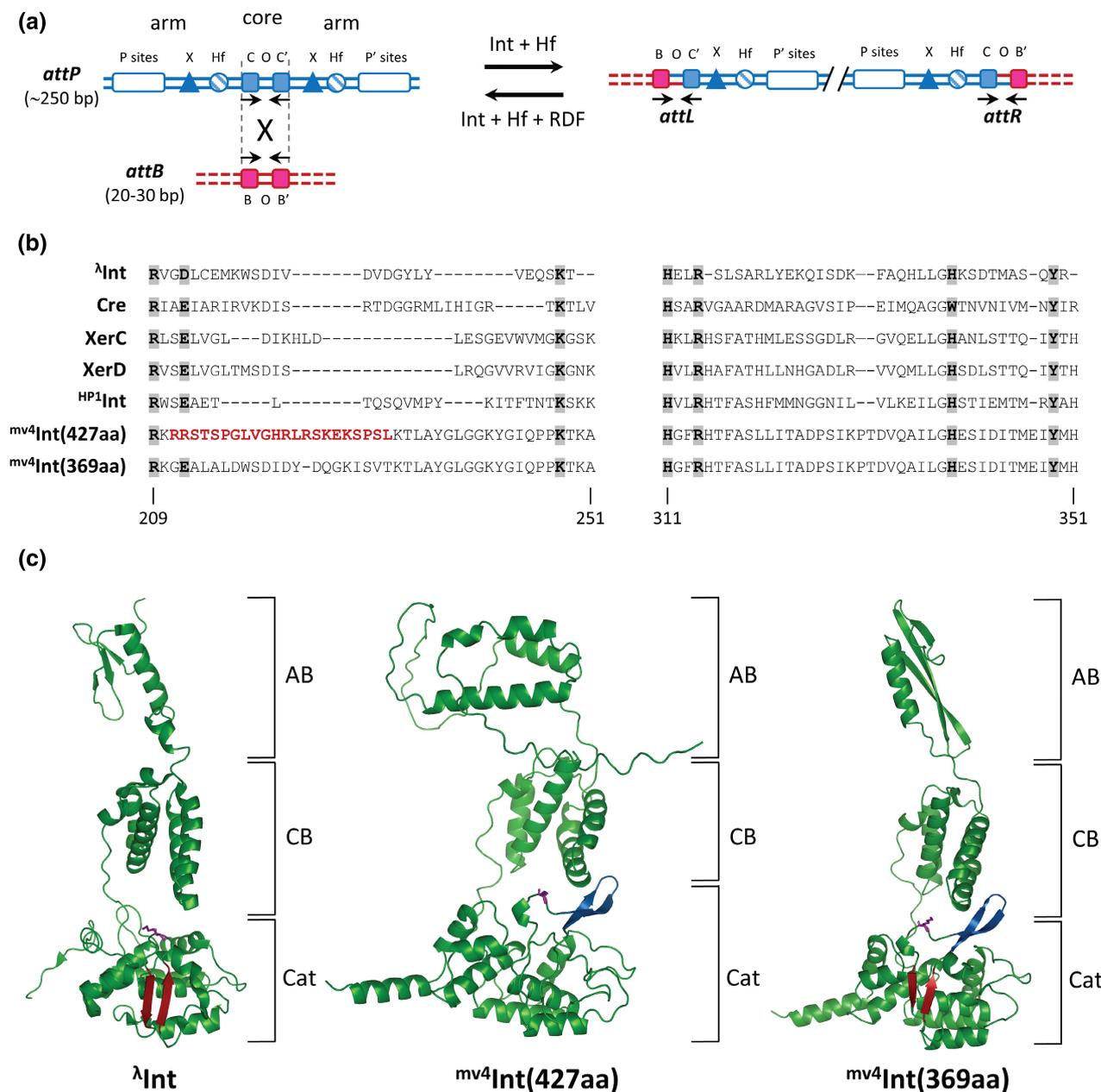
FIGURE 1 (a) General organization of HYR's recombination sites. Arm-binding and core-binding sites for the integrases (Int) are represented by empty or filled rectangles, respectively. Host factors binding sites (Hf) are represented by shaded circles. RDF/Excisionase binding sites (X) are shown by triangles. Horizontal black arrows indicate the binding-sites orientation forming the inverted-repeats. (b) ^{λ}Int and ^{mv4}Int structure comparison. The ^{λ}Int structure (PDB 1Z1G) was taken from Biswas et al. (2005), and ^{mv4}Int(427aa) or ^{mv4}Int(369aa) structures were predicted by AlphaFold2 (Jumper et al., 2021). The canonical antiparallel β2- and β3-strands surrounding the catalytic K residue (purple) are indicated in red. The β-hairpin inserted between β2- and β3-strands is highlighted in blue. AB, arm-binding domain; CB, core-binding domain; Cat, catalytic domain. (c) Alignment of catalytic domains of several YR. ^{λ}Int, Cre, XerC, XerD, ^{HP1}Int, ^{mv4}Int(427aa), and ^{mv4}Int(369aa). Numbers correspond to amino acid sequence of ^{mv4}Int(369aa). The K residue was manually adjusted according to Nunes-Düby et al. alignment (Nunes-Duby et al., 1998). The incorrect amino acids of ^{mv4}Int(427aa) are shown in red, and the seven conserved residues of the catalytic domain are highlighted in gray.

structural hallmark of the ^{Tn916}Int subfamily (Smyshlyaev et al., 2021), which in ^{Tn1549}Int, is required for DNA cleavage and strand-exchange reaction (Rubio-Cosials et al., 2018).

^{mv4}Int(369aa) was subsequently overproduced, purified (Figure S3a), and used for in vitro recombination assays between the mv4 *attP* site located on a supercoiled plasmid (pMC1, Table S2) and

a fluorescent 290-bp PCR product derived from the *L. bulgaricus* tRNA^{SER} region (Dupont et al., 1995). During experimental optimization of this in vitro assay with wild-type (WT) or modified *attP/attB* sites, a decrease in recombination activity was observed when using purified ^{mv4}Int instead of ^{mv4}Int-overexpressing *E. coli* cell extract, particularly for the modified sites (Figure S3b, middle vs left gel). However, addition

of crude extract from *E. coli* BL21 strain or *Lactococcus lactis* (data not shown) increased recombination activity by factors of 2–3 for WT sites and 25–30 for modified sites (Figure S3b, right). Both crude extracts retained their efficacy after heat, DNase, or RNase treatments, but not after proteinase K treatment, indicating that the "stimulating factor" was a heat-resistant protein present in *E. coli* and *L. lactis* cells.

Since all HYR's host factors described to date are nucleoid-associated proteins, and that HU protein, an ubiquitous protein found in all prokaryotic phyla (Grove, 2011), is the only nucleoid-associated protein common to *E. coli* and *L. lactis*, we overproduced and purified HU protein from *L. lactis* (Figure S3a). Addition of purified HU to the reaction reproduced the stimulation observed with the crude extract in a dose-dependent manner (Figure S3c), demonstrating that HU is not strictly speaking a host factor for ^{mv4}Int-mediated recombination, such as IHF or Fis for $^\lambda$Int, but acts as a facilitator for recombination, particularly in the context of modified recombination sites. Finally, we experimentally demonstrated that, as predicted for classical tyrosine recombinases, no ^{mv4}Int-mediated recombination was observed (Figure S3d) when two important residues (Y349F or K248A) in the YR catalytic site were modified.
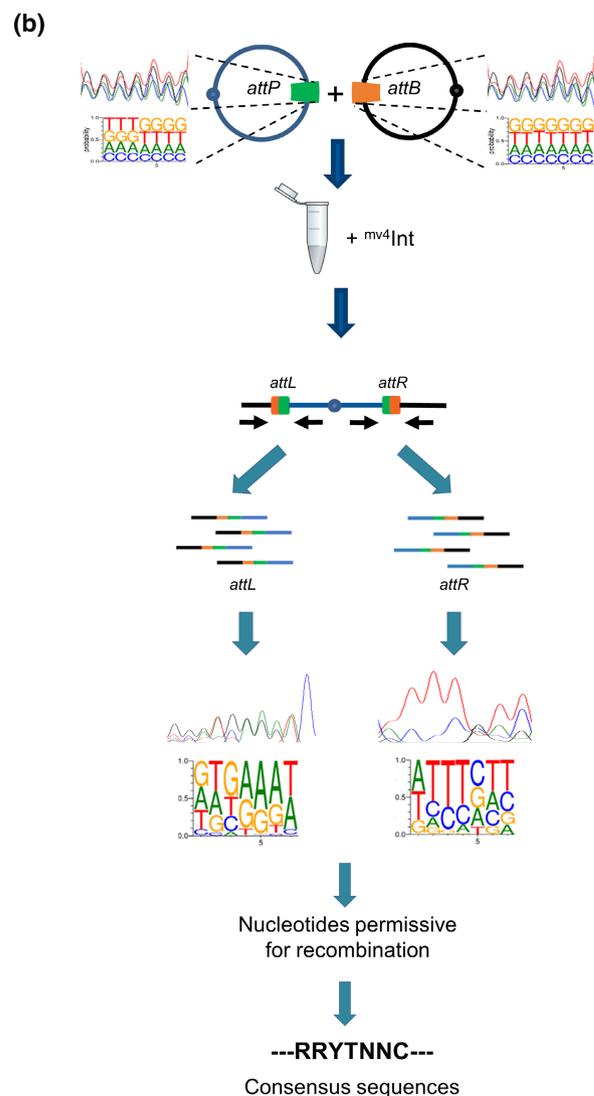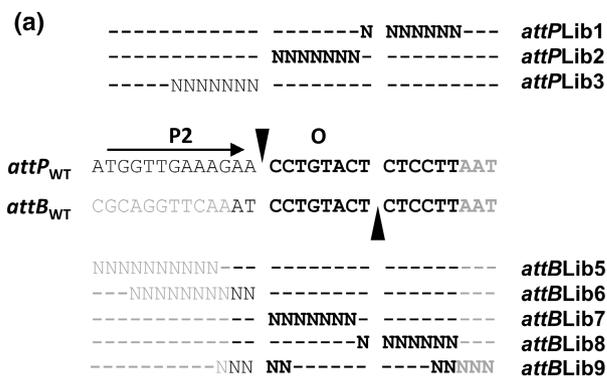
2.2 | Rationale for the use of random DNA libraries

The mv4 *attP* core region and *attB* site were comprehensively reevaluated using random DNA libraries (Escudero et al., 2016; Miele et al., 2022; Sheren et al., 2007), under the hypothesis that DNA positions in these core regions should have constrained nucleotide compositions. DNA libraries for the *attB* site or for the core region of the *attP* site, with 7–10 randomized positions (Figure 2a), were used for in vitro recombination experiments with the corresponding native partner site (*attP*$_{WT}$ or *attB*$_{WT}$) or cognate library (*attP*Lib1 x *attB*Lib8, or *attP*Lib2 x *attB*Lib7). After recombination, the populations of the *attL* and *attR* sites were amplified by PCR and sequenced. Assuming that only nucleotides permissive to recombination are retrieved in the recombined sites, this should reveal the constraints exerted on the nucleotide composition at each of the evaluated positions (Figure 2b).

2.3 | The mv4 *attB* site is 21 bp long

We first checked the in vitro recombination activity of the published minimal site (Auvray, Coddeville, Ordonez, & Ritzenthaler, 1999). However, when a DNA fragment containing the published 16 bp minimal *attB* site surrounded by nucleotides different from those in the *L. bulgaricus* chromosome (Figure 3a) was tested against the plasmid-borne *attP*$_{WT}$ site, no recombination was observed (Figure 3b), implying that the 16-bp fragment did not contain the

FIGURE 2 General strategy to characterize core-*attP* and *attB* sites. (a) Localization of random nucleotides in core-*attP* or *attB* sites. Nucleotides outside the published minimal *attB* site (Auvray, Coddeville, Ordonez, & Ritzenthaler, 1999) are indicated in gray, and sequence identical between core-*attP* and *attB* are in bold. The published P2 arm-type binding site and strand-exchange (O) region (Coddeville et al., 2014) are indicated by horizontal and vertical black arrows, respectively. Random nucleotides are represented by "N" and nucleotides identical to the native sequences are symbolized by dashes. (b) Use and analyses of randomized libraries (see text). The size of the letters in the sequence logos indicates the frequency of each nucleotide at every position. The consensus sequence indicated is arbitrary.

entire *attB* site. In contrast, recombination product was recovered with a 23-bp fragment containing additional *L. bulgaricus* nucleotides on the left of the *attB* site (Figure 3a,b).

Reasoning that site-specific recombination sites are constrained in their nucleotide composition because of their interactions with integrase, the boundaries of the *attB* site were determined by performing in vitro recombination experiments between the *attP*~WT~ site and a DNA library with five randomized positions overlapping the two boundaries of the published *attB* site (*attB*Lib9, Figure 2a). The constraints exerted on the left and right of the *attB* site were, respectively, deduced from the levels of nucleotide degeneracy of the *attL* and *attR* sites upon recombination (Figure 3c). Among the five positions tested on the right of *attB*, only the first two were constrained (Figure 3c), confirming that the *attB* site ends with the sequence 5′-CTCCTT-3′, as previously published (Auvray, Coddeville,

Ordonez, & Ritzenthaler, 1999). In contrast, all five positions on the left of the *attB* site were compositionally constrained (Figure 3c), even those previously identified as outside the minimal *attB* site (Auvray, Coddeville, Ordonez, & Ritzenthaler, 1999). The left boundary of the *attB* site was then determined by recombination between the *attP*~WT~ site and two additional *attB* random libraries (*attB*Lib5 and *attB*Lib6, Figure 2a), each containing ten randomized nucleotides overlapping the left end of the 23-bp fragment permissive for recombination. The levels of nucleotide degeneracy after recombination were similar with both libraries (Figure 3d,e) and therefore independent of the position of the randomized window across the recombination site. In both cases, all four nucleotides were found at every position up to the second G (pos4 Figure 3d and pos7 Figure 3e), implying that the mv4 *attB* site is 21 bp long, with the sequence 5′-TTCAAATCCTGTACTCTCCTT-3′.
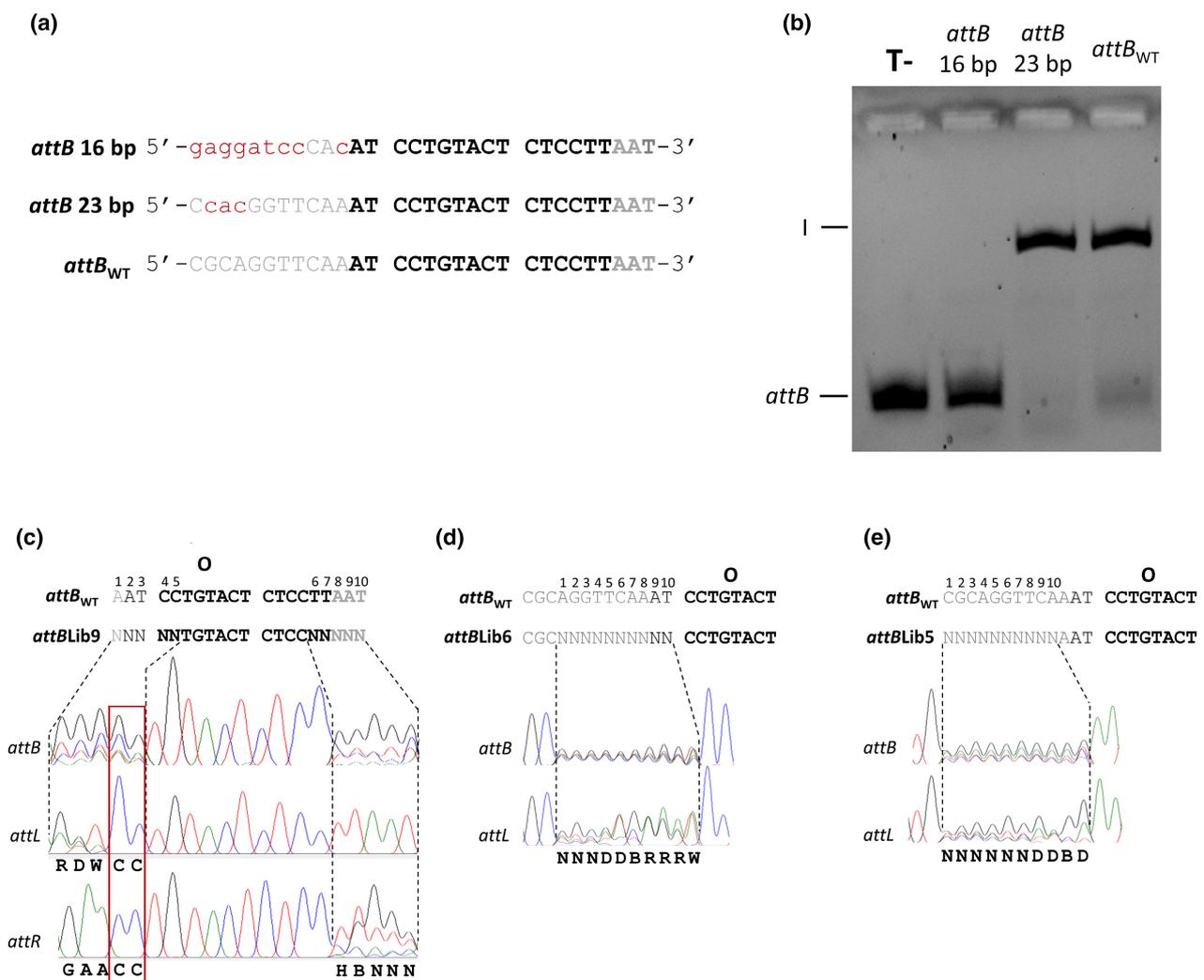
FIGURE 3 In vitro characterization of the minimal *attB* size. (a) DNA sequences used for the *attP*~WT~/*attB* in vitro recombination assay. *attB* site is described in Figure 2a. Nucleotides differing from the *L. bulgaricus* chromosomal sequence are indicated in red lowercase letters. (b) *attB* size-dependent effect on the recombination reaction. The size is indicated above each lane. The fluorescent *attB* fragment and the linear recombination product (I) are indicated in Figure S3b. T-, control (reaction without ^mv4^Int). (c) Chromatogram of *attB*, *attL*, and *attR* regions from *attB*Lib9×*attP*~WT~ in vitro recombination (*n*=2). IUPAC code for degenerate sequence, R=A/G, D=A/G/T, W=A/T, H=A/C/T, B=C/G/T. (d) Chromatogram of *attB*, and *attL* regions from *attB*Lib6×*attP*~WT~ recombination (*n*=3). (e) Chromatogram of *attB*, and *attL* regions from *attB*Lib5×*attP*~WT~ recombination (*n*=2).

2.4 | Recombination specificity depends on a 7-bp overlap region

The core-*attP* and *attB* sites from HYR's systems characterized to date generally include 7-bp overlap regions where strand exchanges occur (Craig & Nash, 1983; Hauser & Scocca, 1992; Kolot & Yagil, 1994; Malanowska et al., 2006; Peña et al., 1996; Smith-Mungo et al., 1994). Site-specific recombination rarely depends on the presence of any particular nucleotide, but the sequences of the overlap regions must be identical (Bauer et al., 1985; Hoess et al., 1986; Kolot et al., 2015; McLeod et al., 1986; Weisberg et al., 1983), although heteroduplexes can be functional in some cases (Lee & Saito, 1998; Missirlis et al., 2006; Sheren et al., 2007).

Since the mv4 system seems different from other described systems, with an 8-bp overlap region (Coddeville et al., 2014), we reassessed the size and constraints governing its overlap region using randomized core-*attP* or *attB* libraries, a method previously used to locate the overlap region in integrons (Escudero et al., 2016). In these experiments, should a randomized position in the library belong to the overlap region, the *attL* and *attR* sites should both be occupied by the nucleotide in the cognate WT site after recombination (Figure 4a). Conversely, if the randomized position is located outside the overlap region, the *attL* and *attR* sites should be occupied by different nucleotides after recombination, in one case the wild-type nucleotide and in the other one of the nucleotides permissive for recombination in the library (Figure 4a). The previously determined left border (Coddeville et al., 2014) was validated through *attB*Lib9×*attP*$_{WT}$ recombination, which yielded only the WT sequence at positions 4 and 5 of both the *attL* and the *attR* site (red box, Figure 3c). In contrast, different *attL* and *attR* sites were obtained on examination of the right border with *attP*Lib1×*attB*$_{WT}$ or *attB*Lib8×*attP*$_{WT}$ recombinations (Figure 4b), implying that position 1 in these random libraries was likely not in the overlap region. These results indicate that the core-*attP* and *attB* overlap regions are 7 bp in length, as observed in other HYR's systems.

Recombination specificity was determined using in vitro recombination assays of the WT *attB* or *attP* site with the cognate DNA libraries of random nucleotides (*attB*Lib7×*attP*$_{WT}$, or *attP*Lib2×*attB*$_{WT}$, Figure 2a). The only sequence recovered among the 16,384 (4^7) theoretical combinations was the WT *attP* (Figure S4a) or *attB* (Figure S4b) sequence. In addition, recombination of the randomized overlap region (*attB*Lib7, Figure 2a) against three different *attP* sites with different overlap sequences yielded only the tested *attP* sequence (Figure S4d), although recombination on sequences containing heteroduplexes cannot totally be excluded (Figure S4d, left). These results demonstrate that identical overlap regions in the *attP* and *attB* sites are required for a complete recombination reaction. Finally, a recombination assay on two randomized overlap regions (*attB*Lib7×*attP*Lib2) revealed no nucleotide compositional constraint in the mv4 *attB*/*attP* overlap regions (Figure S4c), implying that ^{mv4}Int can catalyze recombination with any nucleotide combination in the overlap regions, provided the *attP* and *attB* sequences are identical. However, as already observed for model

YRs such as $^\lambda$Int (Nunes-Düby et al., 1997), Flp (Turan et al., 2010; Umlauf & Cox, 1988), and Cre (Sheren et al., 2007), the recombination efficiency varied strongly with the particular nucleotide sequence in the overlap region: the amounts of recombination product observed in in vitro fluorescence assays with *attB* sites adapted to each *attP* overlap region (Figure S4e) ranged from WT-comparable (*attP*$_1$/*attB*$_1$ and *attP*$_2$/*attB*$_2$ pairs) to near the detection limit of the method (*attP*$_3$/*attB*$_3$).

2.5 | Consensus sequences indicate that mv4 *attB* and core-*attP* sites are similar in structure to classical HYR's recombination sites

The compositional bias observed in the randomized DNA libraries used to determine the left border of the *attB* site (Figure 3c–e) suggests that ^{mv4}Int is adaptable to nucleotide variations in its recombination sites, with the degenerate sequence 5′-DDBRRRW-3′ being permissive to in vitro recombination (Figure 3d). A similar investigation of the compositional bias on the right of the core-*attP* and *attB* sites (Figure 4b) showed likewise that the sequence 5′-DYYYVHB-3′ was compatible with recombination. To better characterize this flexibility, in vitro recombination experiments were performed with WT *attP* or *attB* sites against corresponding randomized DNA libraries (*attB*Lib6, *attB*Lib8, *attP*Lib3, and *attP*Lib1, Figure 2a), with the libraries and corresponding *attL* or *attR* recombination products analyzed by NGS. The assembled *attB* and core-*attP* libraries contained 99.5% (16,312) to 100% (16,384) of the theoretically expected number of heptamers (Table S4) and had almost uniform nucleotide frequency at each randomized position (see sequence logos in Figure S5a). In contrast, only 33% (5459) to 72% (11,781) of the heptamers were retrieved in the *attL* or *attR* populations after recombination (Table S4), confirming that only a subset of sequences allowed recombination.

Since all heptamers permissive for recombination are unlikely to have the same recombination efficiency, we compared the counts for each heptamer before and after recombination (i.e. the enrichment ratio) to estimate their potential to form a productive recombination complex with ^{mv4}Int. As expected, the frequency distribution was highly skewed, with NGS read counts ranging from 1 to 69,525 (Table S4) and the number of counts per heptamer decreasing sharply from the most efficient to the least (Figure S5b). For instance, only 2.8% to 7.5% of the most efficient heptamers accounted for 50% of the total read counts (Table S4), whereas 31.5% to 52.4% of the least efficient occurred fewer than 10 times (data not shown). Furthermore, hundreds of heptamers were detected in just one of the two independent experiments performed and were thus excluded from subsequent analyses (Table S4).

We normalized the nucleotide prevalence at each randomized position through mutual information (MI) analysis (Shannon, 1948) (Figure S5c), and estimated the minimum read count and enrichment ratio permitting to minimize noise in the resulting sequence logos. Setting the threshold at 10 for both parameters allowed nearly half of the read counts for each library to be retained, except for
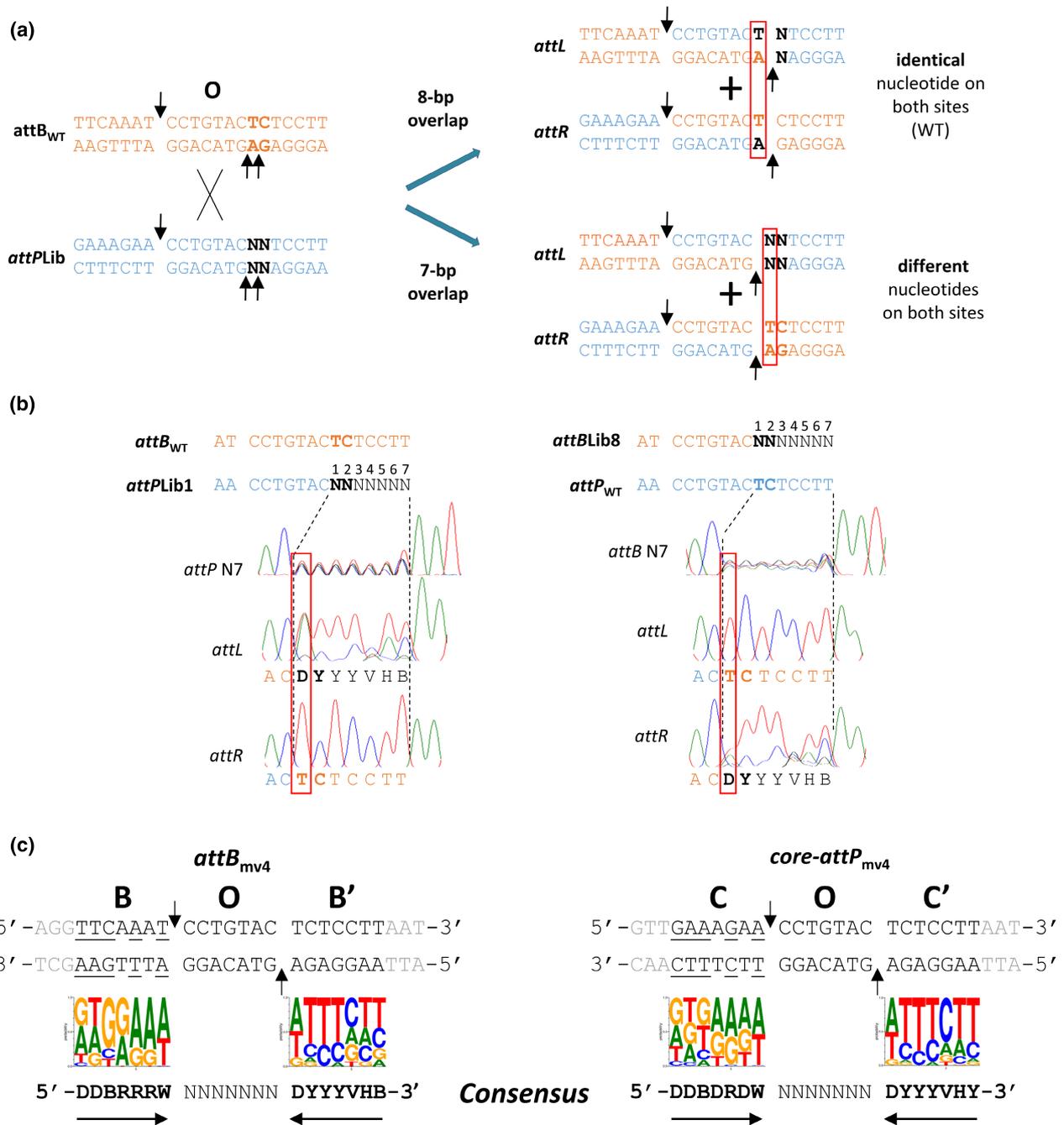
FIGURE 4 Characterization of the strand-exchange (overlap) region. (a) Expected sequence of recombination products depending on the cleavage site position. For an 8-bp overlap region, both *attL* and *attR* should contain the WT nucleotide (T) at the first position (boxed in red) of the randomized region. For a 7-bp region, sequence at the first position should be different between *attL* and *attR*, with all permissive nucleotides present at one of the recombined sites. (b) Chromatograms of *attP*, *attB*, *attL*, and *attR* regions from *attB*$_{WT}$×*attP*Lib1 recombination ($n=3$), or *attB*, *attL*, *attR* regions from *attP*$_{WT}$×*attB*Lib8 recombination ($n=3$). (c) Sequences of *attB* and core-*attP* sites, with sequence logos from NGS sequencing revealing the nucleotide degeneracy of sites permissive for recombination. Only sequencing reads present at least 10 times with a tenfold enrichment ("informative" datasets, see text) were considered for the analysis. Nucleotides outside *attB* or core-*attP* sites are indicated in gray, and those differences between *attB* and core-*attP* are underlined. IUPAC code for degenerate sequence: D = A/G/T, B = C/G/T, R = A/G, W = A/T, Y = C/T, V = A/C/G, H = A/C/T, B = C/G/T.

the *attL* site in the *attB*$_{WT}$×*attP*Lib3 recombination experiment, where a threshold of 6 for the enrichment ratio was chosen. The resulting "informative" datasets used to describe the nucleotide sequence space amenable to recombination consisted in 117 heptamers for *attB*Lib6×*attP*$_{WT}$ recombination, 613 for *attB*Lib8×*attP*$_{WT}$

recombination, 1025 for *attB*$_{WT}$×*attP*Lib3 recombination, and 231 for *attB*$_{WT}$×*attP*Lib1 recombination. The sequence logos generated from these datasets (Figure 4c) are consistent with the consensus sequences determined by Sanger sequencing. Interestingly, the DNA motifs surrounding the overlap region in the consensus sequences

exhibited complementarity, with purine enrichment on the left and pyrimidines on the right (Figure 4c).

Thus, in contrast with a previous study, which concluded that there might be an arm-type site adjacent to the overlap region (Coddeville et al., 2014), our analysis strongly supports a classical organization for the 21 bp *attB* and core-*attP* sites, similar to those of other HYR's recombination systems, with imperfect inverted repeat patterns flanking the strand-exchange region (Figure 4c), defining BOB′ for the *attB* and COC′ structures for the core-*attP* sites, in accordance to the *lambda* phage terminology.

2.6 | Identifying the mv4Int core- and arm-binding regions of the *attP* and *attB* sites

In classical HYR's systems, the core-binding sites of the integrase correspond to DNA sequences at either side of the overlap region (Sarkar et al., 2001; Sylwan et al., 2010; Wood et al., 2010). We performed different gel shift assays to demonstrate that the newly identified B, C, and B′/C′ regions correspond to mv4Int core-binding sites.

In the model *lambda* phage recombination system, it has been demonstrated that the integrase core-binding domain is only able to bind to the core-*attP* (COC′) site if its arm-binding domain is bound to the *attP* arm sites (Sarkar et al., 2001). We thus first compared the mv4Int binding to a labeled 35 bp dsDNA fragment containing the 21 bp COC′ core-*attP* region (Table S3c), with and without unlabeled dsDNA containing the previously characterized P′1P′2 arm sites (Table S3c) (Auvray, Coddeville, Espagno, & Ritzenthaler, 1999). As in *lambda*, the mv4Int/COC′ complex was unstable in the absence of the P′1P′2 fragment, resulting in faint bands and a strong background smear (Figure 5a). However, mv4Int binding to the core-*attP* sequence was strongly stabilized upon addition of a 28-bp dsDNA fragment containing the P′1P′2 sequences, leading to the formation of three complexes (Figure 5a). The direct effect of P′1P′2 sites on the interaction of mv4Int with the core-binding sites was revealed by a shift in mobility of complexes C2* and C3* (Figure 5a) when a 40-bp dsDNA fragment containing the P′1P′2 sites was added instead of the 28-bp fragment. Interestingly, stable binding of mv4Int to a labeled P′1P′2 fragment was also observed when unlabeled COC′ fragment was added (Figure S6a), indicating that the AB and CB domains reciprocally inhibit themselves for their binding at the *attP* site. This suggests that the mechanism leading to the intasome formation of mv4 could differ from that of *lambda* (Sarkar et al., 2001) or CTnDOT (Wood et al., 2010).

Based on studies of other YR systems, our data suggest that complex C1 corresponds to a single mv4Int monomer bound to one core-binding site, while complex C2 corresponds to a monomer bound to both an arm and a core-binding site, and complex C3 corresponds to an mv4Int dimer simultaneously bound to arm- and core-binding sites. In the presence of an unlabeled P′1P′2 fragment, mv4Int stably bound with the same efficiency to both core-*attP* and *attB* sites in a dose-dependent manner (Figure 5b). Since the left- and right half-sites of

the *attB* and core-*attP* sites are not perfect inverted repeats, a feature probably related to asymmetric binding of YRs to their cognate recombination sites (Nolivos et al., 2010; Sylwan et al., 2010), gel shift assays were conducted using labeled 35-bp dsDNA fragments containing individual half-sites (B′/C′, B, or C) (Table S3c). The C1, C2, and C3 binding complexes observed for the *attB*/core-*attP* sites also formed with the B′/C′ half-site arm (Figure S6b), albeit at a lower efficiency, especially for complex C3 (the putative dimer bound to both arm- and core-binding sites). Moreover, mv4Int bound to the B or C half-sites with an affinity significantly lower than to the B′/C′ half-site, with only complex C1 being detectable (Figure S6b). Thus, mv4Int appears to efficiently bind to the core-*attP* and *attB* sites in a cooperative manner, with a preference for the conserved right half-site (B′/C′), which most likely promotes binding of a second mv4Int monomer at the left half-site (B/C).

Two previous studies have reported conflicting findings on the size, location, and orientation of the arm-binding P sites (Auvray, Coddeville, Espagno, & Ritzenthaler, 1999; Coddeville et al., 2014). The first study proposed a 9-bp consensus site with a conserved 6 bp 5′-AACTAG-3′ sequence, with P1, P2, P′1, and P′2 in direct repeats and P′3 in reverse orientation (Auvray, Coddeville, Espagno, & Ritzenthaler, 1999), whereas the second suggested that the P sites are 11-bp direct repeats, with P2 adjacent to the overlap region (Coddeville et al., 2014). The P sites were characterized in this study by assessing mv4Int binding stabilization to the WT core-*attP* (COC′) site using several combinations of arm sites (Table S3c). Among the eleven tested combinations, only those containing either adjacent P1P2 or P′1P′2 pairs effectively stabilized mv4Int binding to the COC′ site (Figure 5c), demonstrating that the P′3 sequence is unlikely to serve as an arm-binding site and confirming its unclear role in integrative and excisive recombination (Auvray, Coddeville, Espagno, & Ritzenthaler, 1999; Coddeville et al., 2014). The size of the arm-binding sites was also evaluated by conducting binding experiments with 11-, 9-, or 6-bp P-site pairs (Table S3c), with complex C3 only stabilized with the two 11-bp sites in direct repeat (Figure 5d). These results indicate that the mv4 *attP* site is 215 bp in length and contains two pairs of adjacent arm-binding sites in direct repeat, one pair (P1-P2) located at the end of the left arm, and the other (P′1-P′2) at the end of the right arm (Figure 5e).

2.7 | Exploring the sequence space permissive to recombination at the core-binding regions

The observed degeneracy of mv4 *attB* and core-*attP* sites led us to hypothesize that mv4Int could potentially catalyze site-specific recombination on alternative target sites, provided their composition matches the established consensus sequence (Figure 4c). Recombination experiments were thus carried out on different *attB* sites modified to varying extents in the overlap and core-binding regions, and these modified *attB* sites were tested against either the *attP*WT site or adapted *attP* sites with an overlap region identical to the modified *attB* (Figure 6).
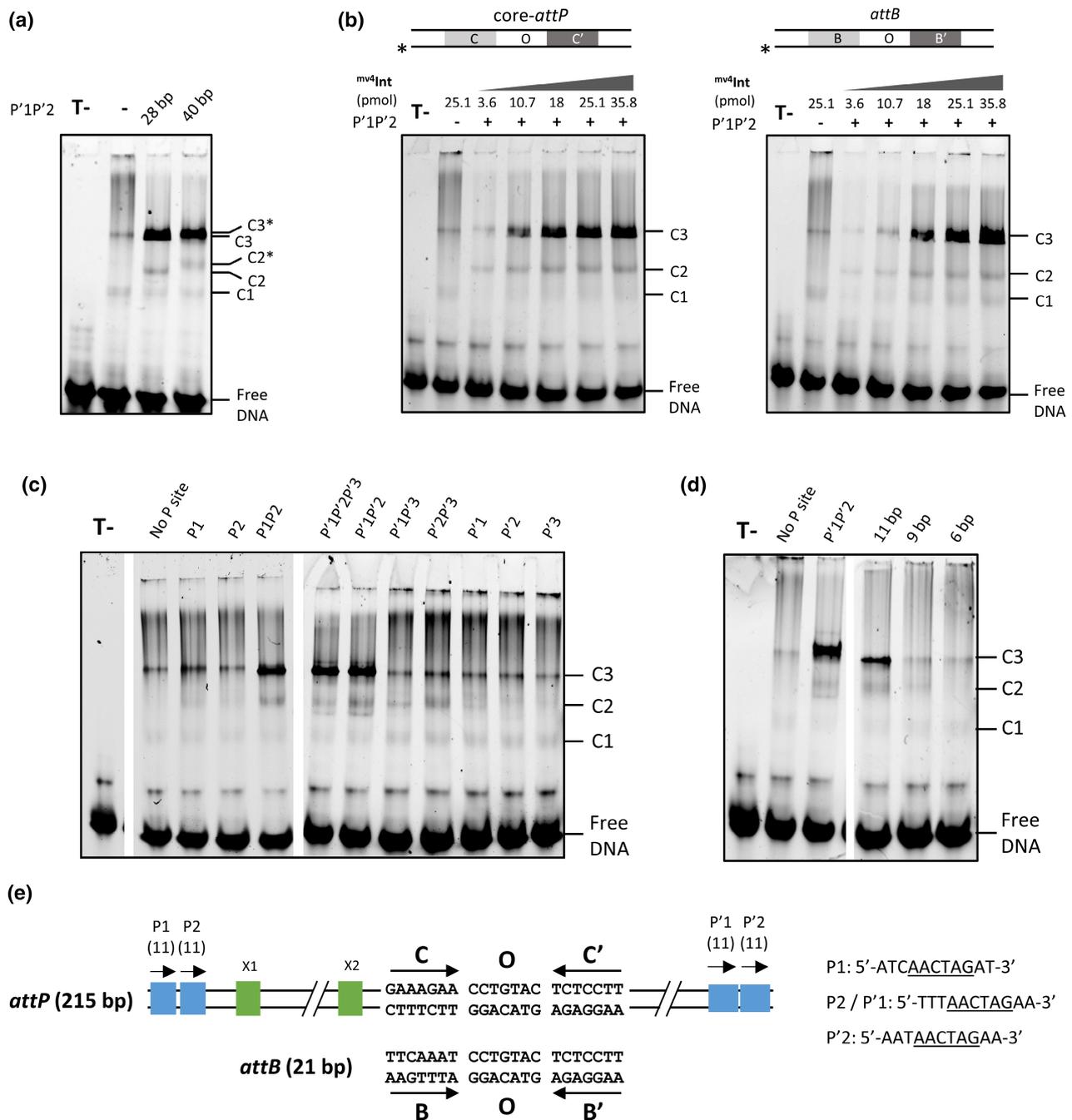
FIGURE 5 Characterization of ^{mv4}Int binding sites. (a) ^{mv4}Int binding stabilization to a 35-bp DNA fragment containing the core-*attP* region by different P′1P′2-containing unlabeled dsDNA fragments. EMSA reactions were performed as described in Section 4.7 with 25 pmol of ^{mv4}Int. Presence/absence and size of arm sites are indicated above the gel. T-, control reactions (no ^{mv4}Int, no arm site). Positions of the free DNA probes and the C1, C2/C2*, and C3/C3* ^{mv4}Int/DNA complexes are indicated. (b) Titration of a 35-bp DNA fragment containing either the core-*attP* (COC′) or *attB* (BOB′) site by increasing concentrations of ^{mv4}Int. The DNA substrate, ^{mv4}Int concentrations in picomole, and presence/absence of P′1P′2 unlabeled DNA fragment are shown at the top. Free DNA and ^{mv4}Int/DNA complexes are indicated. The asterisk locates the 3′-labeled oligonucleotide. (c) Effect of 40-bp unlabeled dsDNA fragments containing different combination of arm sites on the ^{mv4}Int binding stabilization to the COC′ 35-bp core-*attP* fragment. Reactions and legends are identical to (a). (d) Size dependence of the P′1P′2 site on the ^{mv4}Int binding stabilization to the 35-bp core-*attP* dsDNA. Reactions and legends are identical to (a). (e) New genetic organization of mv4 *attP* and *attB* recombination sites.

As expected, modifying the B or B′ hemisite by substituting one or two nucleotides with others from the consensus sequence had no discernible effect on in vitro recombination (*attP₄/B₄* to *attP₆/B₆*, Figure 6). However, effective recombination still occurred in many cases when the substituting nucleotide did not belong to the consensus sequence (*attP*$_{WT}$/*B₇* to *attP*$_{WT}$/*B₁₀*, Figure 6), and recombination was only inhibited in the presence of at least two "unfavorable" nucleotides (*attP₂/B₁₄* to *attP*$_{WT}$/attB 16-bp, Figure 6) and for the

FIGURE 6 In vitro recombination of modified core-*attP* and *attB* sites. Nucleotides differing from the WT sites are represented by bold lowercase letters. Nucleotides present or absent from the sequence logos are indicated in green or red, respectively. Nucleotides underrepresented in the sequence logos are indicated in orange. +, recombination to a level similar to WT sites; +/−, recombination at low level; −, no recombination. Asterisk indicates in vivo recombination experiments described in Auvray, Coddeville, Ordonez, and Ritzenthaler (1999).

	B →	O	← B′	*In vitro* fluorescent assay
attP$_{WT}$/*B*$_{WT}$	TTCAAAT	CCTGTAC	TCTCCTT	+
attP$_4$/*B*$_4$	TTCAAAT	CC**cc**T**tg**	**a**CTCCTT	+
attP$_5$/*B*$_5$	TTCAAAT	CC**cc**T**tg**	TCTCC**a**T	+
attP$_6$/*B*$_6$	TTCAAAT	CC**ccct**C	TCTCC**gc**	+
attP$_{WT}$/*B*$_7$	TTCA**c**AT	CCTGTAC	TCTCCTT	+
attP$_{WT}$/*B*$_8$	TTCA**t**AT	CCTGTAC	TCTCCTT	+
attP$_{WT}$/*B*$_9$	TTCAAA**c**	CCTGTAC	TCTCCTT	+
attP$_{WT}$/*B*$_{10}$	TTCAAA**g**	CCTGTAC	TCTCCTT	+
attP$_1$/*B*$_{13}$	TTC**ggga**	**tt**T**cgg**C	**g**CTCC**ac**	−
attP$_2$/*B*$_{14}$	**g**T**t**AAAT	**tgcca**AC	**g**CT**ta**TT	−
attP$_{11}$/*B*$_{11}$	TTCAAAT	CC**ct**TA**t**	**c**CTCC**g**T	−
attP$_{12}$/*B*$_{12}$	TTCAAAT	CCT**c**T**ct**	**c**C**g**CC**ac**	−
attP$_{WT}$/*attB* 16-bp	**cc**CA**c**AT	CCTGTAC	TCTCCTT	−
attP$_{WT}$/*B*$_{min(99)}$	**agagg**AT	CCTGTAC	TCTCCTT	+*
attP$_{WT}$/*B3*$_{(99)}$	**ccggg**gT	CCTGTAC	TCTCCTT	+/−*
attP$_{WT}$/*B4*$_{(99)}$	**aggtcga**	CCTGTAC	TCTCCTT	−*

attP$_1$/*B*$_{13}$ pair, which was nonfunctional despite having just one underrepresented nucleotide from the sequence conservation analysis. These results strongly suggest that this simple model is insufficient to confidently predict which modified hemisite site is fully permissive to recombination, because the consensus sequence displayed by the sequence logos only indicates which nucleotides are the most frequent at each position, and reveals nothing about other intrinsic constraints, such as interdependencies between nucleotides or incompatible nucleotide combinations. Nevertheless, this model may explain the incorrect minimum length previously attributed to the *attB* site (Auvray, Coddeville, Ordonez, & Ritzenthaler, 1999), in that the reported 16-bp site was fortuitously functional in recombination because the 21-bp modified *attB* site contained one non-consensus-sequence nucleotide among the five that were modified (*attP*$_{WT}$/*B*$_{min(99)}$, Figure 6), whereas the shorter nonfunctional derivatives (*attP*$_{WT}$/*B*$_{3(99)}$ and *attP*$_{WT}$/*B*$_{4(99)}$, Figure 6) contained two adjacent "unfavorable" nucleotides (Auvray, Coddeville, Ordonez, & Ritzenthaler, 1999).

To elucidate the intrinsic constraints governing the selection of DNA motifs within the sequence space defined by the sequence logos for the reprogramming of ^{mv4}Int/*attP* toward alternative DNA sites, we first ranked the heptamers from the B, B′, C, and C′ randomized libraries based on their enrichment ratios (supplementary File S2). The heptamer ranks tend to correlate with the recombination efficiency of the modified sites, with sites that were functional in in vitro fluorescent assays (Figure 6) generally containing heptamers found in datasets previously defined as "informative," while nonfunctional sites contained heptamers either with an enrichment ratio <1 or not found in recombined site populations (supplementary File S2). In addition, comparing top-ranked and WT hemisites revealed conserved nucleotides that are potentially important for interactions with mv4 integrase, such as the four central nucleotides (5′-CTCC-3′) for the B′ and C′ hemisites, and two or three nucleotides for the B and C motifs (Table S5). This asymmetrical conservation may be related to the enhanced binding capacity of B′/C′ sites, since complexes C2 and C3 were exclusively observed on these hemisites (Figure S6b).

There were some inconsistencies however, such as the B′ motif of the nonfunctional *attB*$_{13}$ site (Figure 6), which ranked 4th with an enrichment ratio of 28.5, and the C′ motif of the functional *attP*$_6$/*B*$_6$ pair (Figure 6), which ranked 482nd with an enrichment ratio of 5.05. We thus investigated whether additional constraints, such as nucleotide interdependencies, might also shape the recombination sites. We calculated Cramer's V values (Cramér, 1946) for each pair of position at each hemisite (Table S6), because Cramer's V quantifies

the strength of an association between two variables, with values above 0.2 conventionally interpreted as indicating a significant association (Cramér, 1946). In the case of DNA motifs, Cramer's V provides insights into the weight of a given nucleotide's influence on the nucleotide composition in its vicinity. The associations for the five positions proximal to the overlap region of the *attB* and core-*attP* sites were moderate to strong (position 3 to 7 for the left region, and 1 to 5 for the right region, Table S6), with the most significant associations found within the central triplets. Altogether, these results underline the multiple levels of constraints acting on the core-binding regions of *attB* and core-*attP* sites. Establishing definitive criteria for predicting which DNA sequences will undergo efficient [mv4]Int-mediated recombination will therefore be challenging and will likely require a more comprehensive mechanistic understanding of the recombination reaction.

3 | DISCUSSION

This study provides a comprehensive characterization of the [mv4]Int/*attP*/*attB* system, the first HYR system of a temperate phage infecting a bacterial species in the *Bacillota* (formerly *Firmicutes*) phylum to be fully characterized in this way, and the second, after the HYR system of mycophage L5 (Peña et al., 2000), outside of *Gammaproteobacteria*. [mv4]Int has recently been phylogenetically classified as belonging to the largest HYR family, the [Tn916]Int subfamily, specifically within the unstudied streptococcal phage T12 cluster (Smyshlyaev et al., 2021). Since the [mv4]Int/*attP*/*attB* system is the first such system to have been characterized in this group, which includes many *Bacillota*-infecting phages (Smyshlyaev et al., 2021), we propose considering [mv4]Int as paradigmatic of this group, which could be renamed the mv4 cluster.

The predicted structure of the corrected protein sequence has a typical HYR organization, with the conserved β-hairpin insertion characteristic of the [Tn916]Int subgroup. Structure-based alignment of the β-hairpin region from several YRs (Figure 7a) reveals subtle variations in primary sequence and in the length of the β-strands (4 to 6 amino acids) and the loop (1 to 5 amino acids). These variations create a "sharp" and a "smooth" β-hairpin (Figure 7b), which may differ in their interactions with recombination sites. [mv4]Int and its relatives seem to contain longer β-strands (6-aa) separated by just a single amino acid, producing a sharper structure than [T12]Int's or those of other integrases from the [Tn916]Int subgroup such as [Tn1549]Int and [CTn6]Int. Whether this structural difference has any functional consequence in the role of the β-hairpin requires further experimental investigation.

The results of this study also confirm that [mv4]Int does not require species-specific host factors to recombine WT *attP* and *attB* sites, but demonstrate that HU protein, while not mandatory, does significantly enhance recombination between *attP/B* sites, particularly when modified. A similar stimulation effect has been observed for Tn*916* excision in *E. coli* (Connolly et al., 2002). HU frequently acts as a cofactor for serine recombinases, such as Tn*3*/γδ or Hin

(Alonso et al., 1995; Johnson et al., 1986; Petit et al., 1995; Rowland et al., 2002), but rarely for HYRs, which typically use IHF (Goodman & Scocca, 1989; Panis et al., 2007; Smith-Mungo et al., 1994; Yu & Haggård-Ljungquist, 1993) or analogs such as mIHF for [L5]Int (Pedulla et al., 1996) and BHFa for [CTnDOT]Int (Ringwald & Gardner, 2015). HU and IHF belong to the same DNA bending family, with IHF exclusively found in *Proteobacteria* (Dey et al., 2017; Kamashev et al., 2017), but differ in their substrate preference and bending ability (Swinger & Rice, 2004). It is worth noting that, except for [L5]Int, all the members of the [Tn916]Int subgroup that have been studied to date (Alvarez et al., 1998; Raynal et al., 1998; Roberts & Mullany, 2009) are functional in different bacterial species, confirming their independence of specific host factors, as demonstrated here for [mv4]Int. We hypothesize that the β-hairpin found in the [Tn916]Int subgroup, through its bending activity (Rubio-Cosials et al., 2018), may shape the *attP* site to facilitate intasome formation with the assistance of HU, whereas HYRs that lack the β-hairpin may require more specific host factors with strong bending activity, such as IHF or BHFa, to form the intasome.

The use of in vitro recombination experiments involving randomized DNA libraries, PCR amplification of recombination products, and subsequent Sanger sequencing or NGS analysis has proven to be an effective approach in characterizing core-*attP* and *attB* sites in this system. This methodology appears well-suited for the investigation of recombination sites and has successfully been employed in other recombination systems (Miele et al., 2022; Missirlis et al., 2006; Sheren et al., 2007; Walker et al., 2023). Here, it revealed that the mv4 *attB* site is longer than previously described, with a size of 21 bp more consistent with other heterobivalent YRs systems, which range in length from 18 bp for HP1 phage (Hauser & Scocca, 1992) to 29 bp for mycophage L5 (Peña et al., 1996). This approach also unequivocally showed that the overlap region is 7 bp long, the canonical length for bacteriophage HYR systems, and revealed that all nucleotide combinations are compatible with recombination provided the *attB* and *attP* are identical. However, when several overlap regions were tested in classical in vitro fluorescent assays, the recombination efficiency varied substantially between sequences (Figure S4e), pointing to a certain amount of compositional bias. This result deserves further investigation through NGS analysis of the recombination products from the two *attB* × *attP* libraries with randomized overlap regions.

Randomized DNA libraries allowed us to map out the sequence space permissive to recombination of the core-binding regions of both recombination sites, leading to nearly identical consensus sequences for the *attB* and core-*attP* sites, and showing that the structure of mv4's recombination sites is typical of other YR systems: BOB′ in the *attB* site and COC′ in the core-*attP* site, with the genuine core-binding sites for [mv4]Int being B, C, and B′/C′ hemisites. In addition, the degeneracy observed at the hemisites indicates that the mv4 system is somewhat tolerant of nucleotide variations in its recombination sites. This kind of flexibility has recently been observed in two other YR systems in *Vibrio cholerae*: in XerCD/*dif* chromosome dimer resolution system, where XerD

FIGURE 7 Comparison of the β2-β3 region of several YR systems. (a) Structure-based alignment of the β2-β3 region. Limits of β-strands were determined from X-ray data or Alphafold2 predictions. YR is classified according to Smyshlyaev et al. (2021). The β-strands present in all YR are named according to Aihara et al. (2003) and boxed in black. β-strands of the additional β-hairpin are boxed in red. The catalytic K residue (corresponding to K235 for $^{\lambda}$Int) is indicated by an asterisk. No β-hairpin could be observed for ^{pSAM2}Int. (b) Details of the β2-β3 region revealing the two types of β-hairpins. Upper, "sharp" hairpins with 1 or 2 amino acids on the loop; Lower, "smooth" hairpins with 3 to 5 amino acids on the loop. β2-β3 strands are in purple. β-strands from the β-hairpin are in red, and loop in cyan. The catalytic K residue is in yellow.

recombinase can accommodate nucleotide modifications at its *dif* site and is even more flexible in the presence of the TLC φ phage protein, *xafT* (Miele et al., 2022), and in integron recombination system, where integrase can recombine various genomic sites with minimal sequence requirements (Loot et al., 2024). Since most existing studies of HYRs were conducted prior to the widespread availability of NGS sequencing, applying this approach to reassess the nucleotide tolerance of other model heterobivalent integrases should prove worthwhile.

Finally, our exploration of consensus sequences, coupled with classical gel-based recombination assays, shed light on the intricate interplay of various compositional constraints that has governed the adaptation of recombination sites to their cognate integrase. These additive constraints make recombination efficiencies difficult to predict when trying to reprogram ^{mv4}Int-mediated recombination toward alternative sites. Note also that our recombination experiments involved only one randomized core-binding site at a time, either left or right, thereby potentially missing additive effects between the

two regions. Nonetheless, we are considering making use of [mv4]Int's recombination site flexibility to develop a programmable genetic tool by adapting the *attP* site to novel DNA targets. This will require further developments, such as enhancing the recombination efficiency by testing *attP* sites with highly ranked core-binding motifs. Additionally, conducting in vivo recombination experiments on bacterial species of biotechnological interest, using sites with randomized core-*attP* regions, will help identify integration sites that could serve as future "landing pads," since recombination reactions appear to be less demanding in vivo than in vitro in some recombination systems (Sheren et al., 2007; Walker et al., 2023). In summary, this study underscores the diversity and complexity of recombination systems in temperate bacteriophages. Our results provide valuable insights into the structure and functioning of the [mv4]Int/*attP*/*attB* recombination module, expanding our understanding of the mechanisms underpinning site-specific recombination and the determinants of biological diversity in these systems.

4 | EXPERIMENTAL PROCEDURES

4.1 | Strains, plasmids, oligonucleotides, and media

Strains, plasmids, and oligonucleotides used in this study are listed in Tables S1–S3, respectively. *E. coli* strain MET961 was constructed by replacing the *glgB* gene of strain NEB 5-alpha (New England Biolabs) with the *glgB*::Kan-pWV01*repA* region from *E. coli* EC1000 (Leenhouts et al., 1996) using the Datsenko and Wanner method (Datsenko & Wanner, 2000). *E. coli* strains were grown at 37°C in Lysogenic Broth (Bertani, 1951). Antibiotics were used at the following concentration: carbenicillin, $100\,\mu g.mL^{-1}$; chloramphenicol, $12.5\,\mu g.mL^{-1}$; erythromycin, $150\,\mu g.mL^{-1}$; kanamycin, $50\,\mu g.mL^{-1}$.

4.2 | DNA procedures

Standard techniques were used for DNA manipulation and cloning. Polymerase chain reaction (PCR) was performed with Q5-HF polymerase (New England Biolabs, USA) or with CloneAmp Hifi polymerase (Takara Bio, Japan), according to the manufacturer's instructions. PCR products were purified using the QIAquick PCR purification kit (Qiagen, Germany). Plasmid DNA was extracted using QIAprep Spin Miniprep kit (Qiagen, Germany) or Nucleobond Xtra Midi (Macherey-Nagel, Germany), and each construction was verified by Sanger sequencing (Mix2seq, Eurofins Genomics, Germany). Plasmids were obtained from assembly of PCR-amplified DNA fragments, obtained from pMC1, pBS*attB*, or pCC1FOS™ templates, with the relevant oligonucleotides (Table S3a). Cloning was performed by Gibson assembly (Gibson et al., 2009) with NEBuilder HIFI DNA Assembly (New England Biolabs, USA) or blunt-end cloning with T4 PNK and T4 DNA ligase (New England Biolabs, USA). Random DNA libraries were constructed as follows: randomized oligonucleotides (109 bp in length for *attB* libraries; 184 bp in length for core-*attP* libraries, Table S3b)

were obtained by chemical synthesis (Integrated DNA Technologies, USA). A 10-cycle PCR was conducted to generate double-stranded DNA using primers attBLib-F / attBLib-R pair for *attB* and attPLib-F / attPLib-R pair for *attP* (Table S3b). Each PCR-product population was gel-extracted and separately cloned by Gibson assembly, either into pCC1FOS™ plasmid (Epicenter, USA) for *attB* libraries, or plasmid pMET359 (Table S2) for *attP* libraries. Clones were propagated in *E. coli* strain EPI300 under chloramphenicol selection for *attB* libraries, or in strain MET961 under carbenicillin selection for *attP* libraries. Number of clones obtained for each randomized library was $9.4.10^4$ for *attP*Lib1, $1.1.10^5$ for *attP*Lib2, $2.9.10^5$ for *attP*Lib3, $2.4.10^6$ for *attB*Lib5, $1.3.10^6$ for *attB*Lib6, $1.3.10^5$ for *attB*Lib7, $1.1.10^5$ for *attB*Lib8, and $8.4.10^5$ for *attB*Lib9.

4.3 | Protein purification

For [mv4]Int purification, the *E. coli* BL21 (DE3) strain containing pMET332 plasmid (Table S2) was grown in LB at 42°C up to an OD_{600} of 0.6. Integrase gene expression was induced by addition of 0.1 mM of IPTG, and the culture was incubated at 22°C for 3 h. Cells were recovered by centrifugation, resuspended in buffer A (50 mM Tris pH 8, 500 mM NaCl, 20 mM imidazole, 10% glycerol) supplemented with 1 mg/mL lysozyme and one tablet of SIGMAFAST Protease Inhibitor Cocktail Tablets EDTA-Free (Merck KGaA, Germany), and disrupted by sonication (10 cycles of 30 sec at 40% intensity in ice, with 45 sec pause between each cycle). The lysate was cleared by centrifugation (20,000 g, 4°C, 20 min). [mv4]Int was purified on nickel-nitrilotriacetic acid affinity resin (1 mL His-trap HP, GE Healthcare, USA). Column equilibration was performed by injecting 10 column volumes of buffer A. After equilibration, the lysate was injected and unbound proteins were washed using 10 column volumes of buffer A. [mv4]Int was eluted using a gradient from 0 to 30% of buffer B (50 mM Tris pH 8, 500 mM NaCl, 500 mM imidazole, 10% glycerol). Eluted fractions were then injected in a gel filtration column (HiLoad 16/60 Superdex 200, GE Healthcare, USA). This column was equilibrated using 2 column volumes of buffer C (50 mM Tris pH 8, 500 mM NaCl, 10% glycerol, 1 mM DTT, 1 mM EDTA) and the fractions containing [mv4]Int were injected and eluted using the same buffer. Eluted fractions containing [mv4]Int were then twofold diluted in buffer D (50 mM Tris pH 8, 10% glycerol, 1 mM DTT, 1 mM EDTA). Fractions containing [mv4]Int were injected on heparin column (1 mL HiTrap Heparin HP, GE Healthcare, USA) equilibrated with 10 column volumes of buffer E (50 mM Tris pH 8, 250 mM NaCl, 20% glycerol, 1 mM DTT, 1 mM EDTA), and unbound proteins were removed using 10 column volumes of buffer E. [mv4]Int was eluted using a gradient from 0 to 100% of buffer F (50 mM Tris pH 8, 1 M NaCl, 20% glycerol, 1 mM DTT, 1 mM EDTA). The different purification steps were monitored by SDS-PAGE (Figure S3a). Purified integrase was aliquoted, snap-frozen in liquid N2, and stored at −80°C in buffer containing 50 mM Tris pH 8, 500 mM NaCl, 20% glycerol, 1 mM DTT, and 1 mM EDTA.

For 6xHis-HU[Lla] purification, the *E. coli* BL21 (DE3) strain containing pMET369 plasmid (Table S2) was grown in LB at 37°C up to

an OD_{600} of 0.6. Expression of the *hup* gene was induced by addition of 0.4 mM of IPTG and the culture was incubated at 22°C for 3 h. Cells were recovered by centrifugation, resuspended in buffer A, and disrupted by sonication as indicated above. The lysate was cleared by centrifugation (16,000 g, 4°C, 15 min), and the recovered supernatant was heated at 70°C for 30 min and cleared by centrifugation (16,000 g, 4°C, 15 min). The 6xHis-HU$_{Lla}$ protein was purified on nickel-nitrilotriacetic acid affinity resin as described for [mv4]Int, except than elution was performed with a gradient from 0 to 70% of buffer B. Buffer of the eluted fractions containing 6xHis-HU$_{Lla}$ was replaced by buffer G (50 mM Tris pH 8, 200 mM NaCl) in Amicon® Ultra—3 kD Centrifugal Filters (Merck KGaA, Germany) and resulting solution was injected in a heparin column. Unbound proteins were removed using 10 column volumes of buffer E, and 6xHis-HU$_{Lla}$ protein was eluted using a gradient from 0 to 100% of buffer F. The different purification steps were monitored by SDS-PAGE (Figure S3a). After purification, the buffer of purified HU was exchanged with buffer H (50 mM Tris pH 8, 150 mM NaCl) using Amicon® Ultra—3 kD Centrifugal Filters (Merck KGaA, Germany). The resulting samples were aliquoted, snap-frozen in liquid N2, and stored at –80°C.

4.4 | In vitro fluorescent assay

Reaction mixtures (20 μL) contained 0.08 pmol (300 ng) of supercoiled plasmid carrying the *attP* site, 0.08 pmol (15 ng) of linear fluorescent (Cy3), 290-bp *attB* fragment obtained by PCR amplification from relevant plasmid templates (Table S2) with LbbulgattB-F / Cy3-New-attB-R primers pair, 7.2 pmol (300 ng) of [mv4]Int, and 40 μg of a crude extract from *E. coli* BL21(DE3) heated at 95°C for 10 min, in TENDP 1X buffer (25 mM Tris pH 7.5, 1 mM EDTA, 150 mM NaCl, 1 mM DTT, and 10% PEG8000). The reaction was incubated at 42°C either 1 h 30 or 16 h, depending on the presence or absence of the heated *E. coli* crude extract, respectively, and was stopped by addition of 0.1% SDS. Samples were analyzed by electrophoresis in 0.8% agarose gels. Fluorescence was revealed using the ChemidocMP imaging system (Bio-Rad, USA).

4.5 | In vitro recombination assay of random DNA libraries

The reaction (20 μL) was performed with 0.08 pmol of *attB*LibX plasmid (X being a number from 5 to 9, Table S3b) or *attP*LibX plasmid (X being a number from 1 to 3, Table S3b), 0.08 pmol of pBS*attB* or pMC1 plasmid (Dupont et al., 1995), 7.2 pmol of [mv4]Int and 40 μg of crude extract from *E. coli* BL21(DE3) heated at 95°C for 10 min, in TENDP 1X, and incubated 1 h 30 at 42°C. The different regions *attB*, *attP*, *attL*, and *attR* were amplified by a 30-cycle PCR using SeqLibattB-F / SeqLibattB-R primers pair for *attB* (size of the PCR fragment: 184 bp); SeqLibattR-F / SeqLibattL-R primers for *attP* (size of the PCR fragment: 194 bp); SeqLibattB-F / SeqLibattL-R primers for *attL* (size of the PCR fragments: 196 bp for randomized

B region, and 301 bp for randomized C′ region); and SeqLibattR-F / seqLibattP-R primers for *attR* (size of the PCR fragments: 182 bp for randomized B′ region, and 264 bp for randomized C region). PCR products were purified and either analyzed by Sanger sequencing (Mix2seq, Eurofins Genomics, Germany) or NGS (Illumina, NGSelect Amplicon, 2×150 bp, Eurofins Genomics, Germany).

4.6 | Bioinformatic analyses

For NGS analyses, reads from the two fastq files were paired using CLC Genomics Workbench 7 (Qiagen, Germany). Paired and unpaired data were uploaded and analyzed on the public server at usegalaxy.org (Afgan et al., 2018). To reduce cost of analyses, *attB*, *attP*, *attL*, and *attR* PCR-amplified DNAs were mixed before NGS sequencing, and each sample was sorted by length using "Filter Fasta" program (Galaxy version 2.3). Then, unwanted sequences in each group (*attB*, *attP*, *attL*, and *attR*) were removed with "Bowtie2" program (Langmead & Salzberg, 2012), using the expected sequence for each group as the reference genome. For each group, "Fasta Width formatter" program (Galaxy version 1.0.1) was used to get each fasta sequence on a single line. Sequences containing ambiguous bases were removed by using "seqtk_seq" program (Galaxy version 1.3.3). For each group, forward and reverse sequences were separated in two different files using "Filter Fasta" program. Reverse sequences were reverse-complemented using "seqtk_seq" program, and merged with the forward sequences using "merge.files" program (Schloss et al., 2009). To eliminate any contaminant sequence, a final verification was performed in each group using "Filter Fasta" program, by keeping only reads that perfectly matched the expected sequence, and heptamers corresponding to the randomized region of DNA libraries were extracted from each sequence using the "Trim sequences" program (Galaxy version 1.0.2), and their occurrence calculated using the "Wordcount" program (Rice et al., 2000). Only heptamers present in the two biological replicates were considered for further statistical analyses, such as mutual information (Shannon, 1948) and Cramer's V (Cramér, 1946) analyses, with occurrence values corresponding to the mean of the two replicates.

4.7 | Electrophoretic mobility shift assay

HPLC-purified 5′ Cy3 end-labeled synthetic oligonucleotides were obtained by chemical synthesis (Eurofins Genomics, Germany). Labeled dsDNA substrates were prepared by hybridization of 200 pmol of labeled and 300 pmol of unlabeled complementary oligonucleotides (Table S3c) in 10 mM Tris pH 7.5, 50 mM NaCl, and incubating the samples at 95°C for 5 min in a thermal cycler (Bio-Rad, USA) and decreasing the temperature of 1.5°C/min until it reaches 25°C. Binding reactions (20 μL) were performed with 0.87 pmol of labeled dsDNA and 4.48 pmol of unlabeled dsDNA in buffer containing 25 mM Tris pH 8, 75 mM NaCl, 10% glycerol, 0.5 mM DTT, 0.5 mM EDTA, 1 μg polydIdC (Merck KGaA,

Germany), 0.1 mg/mL BSA. The protein was added, the reaction was performed at room temperature for 20 min, and samples were loaded onto a non-denaturing 7.5% polyacrylamide gel (Mini-PROTEAN TGX, Bio-Rad, USA). Gels were run at 4°C, 75 V for 2 h. Fluorescence was revealed using the ChemidocMP imaging system (Bio-Rad, USA).

AUTHOR CONTRIBUTIONS

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST STATEMENT

KDE, MCB, and PLB are authors of patent n°WO2023237453A1. The other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this study.

DATA AVAILABILITY STATEMENT

The NGS data underlying this article are available in the ArrayExpress database at EMBL-EBI under accession number E-MTAB-13571. The GenBank accession number for the corrected nucleotide sequence of ^{mv4}Int is OR873432.

ORCID

Kevin Debatisse https://orcid.org/0009-0002-7898-4479
Pierre Lopez https://orcid.org/0000-0002-5064-3173
Philippe Rousseau https://orcid.org/0000-0002-6376-6527
Manuel Campos https://orcid.org/0000-0001-9751-4531
Michèle Coddeville https://orcid.org/0009-0006-9602-0963
Muriel Cocaign-Bousquet https://orcid.org/0000-0003-2033-9901
Pascal Le Bourgeois https://orcid.org/0000-0003-2039-1953

REFERENCES

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M. et al. (2018) The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, 46, W537–W544.

Aihara, H., Kwon, H.J., Nunes-Düby, S.E., Landy, A. & Ellenberger, T. (2003) A conformational switch controls the DNA cleavage activity of lambda integrase. *Molecular Cell*, 12, 187–198.

Alonso, J.C., Weise, F. & Rojo, F. (1995) The Bacillus subtilis histone-like protein Hbsu is required for DNA resolution and DNA inversion mediated by the beta recombinase of plasmid pSM19035. *The Journal of Biological Chemistry*, 270, 2938–2945.

Alvarez, M.A., Herrero, M. & Suárez, J.E. (1998) The site-specific recombination system of the *lactobacillus* species bacteriophage A2 integrates in gram-positive and gram-negative bacteria. *Virology*, 250, 185–193.

Auvray, F., Coddeville, M., Espagno, G. & Ritzenthaler, P. (1999) Integrative recombination of *lactobacillus delbrueckii* bacteriophage mv4: functional analysis of the reaction and structure of the *attP* site. *Molecular & General Genetics*, 262, 355–366.

Auvray, F., Coddeville, M., Ordonez, R.C. & Ritzenthaler, P. (1999) Unusual structure of the *attB* site of the site-specific recombination system of *lactobacillus delbrueckii* bacteriophage mv4. *Journal of Bacteriology*, 181, 7385–7389.

Auvray, F., Coddeville, M., Ritzenthaler, P. & Dupont, L. (1997) Plasmid integration in a wide range of bacteria mediated by the integrase of *lactobacillus delbrueckii* bacteriophage mv4. *Journal of Bacteriology*, 179, 1837–1845.

Bauer, C.E., Gardner, J.F. & Gumport, R.I. (1985) Extent of sequence homology required for bacteriophage lambda site-specific recombination. *Journal of Molecular Biology*, 181, 187–197.

Bertani, G. (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *Journal of Bacteriology*, 62, 293–300.

Better, M., Lu, C., Williams, R.C. & Echols, H. (1982) Site-specific DNA condensation and pairing mediated by the int protein of bacteriophage lambda. *Proceedings of the National Academy of Sciences of the United States of America*, 79, 5837–5841.

Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A. & Ellenberger, T. (2005) A structural basis for allosteric control of DNA recombination by λ integrase. *Nature*, 435, 1059–1066.

Casjens, S. (2003) Prophages and bacterial genomics: what have we learned so far? *Molecular Microbiology*, 49, 277–300.

Casjens, S.R. & Hendrix, R.W. (2015) Bacteriophage lambda: early pioneer and still relevant. *Virology*, 479–480, 310–330.

Coddeville, M., Spinella, J.F., Cassart, P., Girault, G., Daveran-Mingot, M.L., Le Bourgeois, P. et al. (2014) Bacteriophage mv4 site-specific recombination: the central role of the P2 ^{mv4}Int-binding site. *Journal of Virology*, 88, 1839–1842.

Connolly, K.M., Iwahara, M. & Clubb, R.T. (2002) Xis protein binding to the left arm stimulates excision of conjugative transposon Tn*916*. *Journal of Bacteriology*, 184, 2088–2099.

Craig, N.L. & Nash, H.A. (1983) The mechanism of phage lambda site-specific recombination: site-specific breakage of DNA by Int topoisomerase. *Cell*, 35, 795–803.

Cramér, H. (1946) *Mathematical methods of statistics*. Princeton University Press, Princeton, New Jersey.

Datsenko, K.A. & Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 6640–6645.

Dey, D., Nagaraja, V. & Ramakumar, S. (2017) Structural and evolutionary analyses reveal determinants of DNA binding specificities of nucleoid-associated proteins HU and IHF. *Molecular Phylogenetics and Evolution*, 107, 356–366.

Dupont, L., Boizet-Bonhoure, B., Coddeville, M., Auvray, F. & Ritzenthaler, P. (1995) Characterization of genetic elements

required for site-specific integration of *lactobacillus delbrueckii* subsp. *bulgaricus* bacteriophage mv4 and construction of an integration-proficient vector for *lactobacillus plantarum*. *Journal of Bacteriology*, 177, 586–595.

Erez, Z., Steinberger-Levy, I., Shamir, M., Doron, S., Stokar-Avihail, A., Peleg, Y. et al. (2017) Communication between viruses guides lysis-lysogeny decisions. *Nature*, 541, 488–493.

Escudero, J.A., Loot, C., Parissi, V., Nivina, A., Bouchier, C. & Mazel, D. (2016) Unmasking the ancestral activity of integron integrases reveals a smooth evolutionary transition during functional innovation. *Nature Communications*, 7, 10937.

Esposito, D., Thrower, J.S. & Scocca, J.J. (2001) Protein and DNA requirements of the bacteriophage HP1 recombination system: a model for intasome formation. *Nucleic Acids Research*, 29, 3955–3964.

Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I. & Herskovits, A.A. (2015) A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nature Reviews. Microbiology*, 13, 641–650.

Gandon, S. (2016) Why Be temperate: lessons from bacteriophage λ. *Trends in Microbiology*, 24, 356–365.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A. & Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6, 343–345.

Goodman, S.D. & Scocca, J.J. (1989) Nucleotide sequence and expression of the gene for the site-specific integration protein from bacteriophage HP1 of *Haemophilus influenzae*. *Journal of Bacteriology*, 171, 4232–4240.

Grindley, N.D.F., Whiteson, K.L. & Rice, P.A. (2006) Mechanisms of site-specific recombination. *Annual Review of Biochemistry*, 75, 567–605.

Grove, A. (2011) Functional evolution of bacterial histone-like HU proteins. *Current Issues in Molecular Biology*, 13, 1–12.

Hauser, M.A. & Scocca, J.J. (1992) Site-specific integration of the *Haemophilus influenzae* bacteriophage HP1. Identification of the points of recombinational strand exchange and the limits of the host attachment site. *The Journal of Biological Chemistry*, 267, 6859–6864.

Hoess, R.H., Wierzbicki, A. & Abremski, K. (1986) The role of the *loxP* spacer region in P1 site-specific recombination. *Nucleic Acids Research*, 14, 2287–2300.

Johnson, R.C., Bruist, M.F. & Simon, M.I. (1986) Host protein requirements for in vitro site-specific DNA inversion. *Cell*, 46, 531–539.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O. et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583–589.

Kamashev, D., Agapova, Y., Rastorguev, S., Talyzina, A.A., Boyko, K.M., Korzhenevskiy, D.A. et al. (2017) Comparison of histone-like HU protein DNA-binding properties and HU/IHF protein sequence alignment. *PLoS One*, 12, e0188037.

Kolot, M., Malchin, N., Elias, A., Gritsenko, N. & Yagil, E. (2015) Site promiscuity of coliphage HK022 integrase as a tool for gene therapy. *Gene Therapy*, 22, 521–527.

Kolot, M. & Yagil, E. (1994) Position and direction of strand exchange in bacteriophage HK022 integration. *Molecular & General Genetics*, 245, 623–627.

Langmead, B. & Salzberg, S.L. (2012) Fast gapped-read alignment with bowtie 2. *Nature Methods*, 9, 357–359.

Lee, G. & Saito, I. (1998) Role of nucleotide sequences of *loxP* spacer region in Cre-mediated recombination. *Gene*, 216, 55–65.

Leenhouts, K., Buist, G., Bolhuis, A., ten Berge, A., Kiel, J., Mierau, I. et al. (1996) A general system for generating unlabelled gene replacements in bacterial chromosomes. *Molecular & General Genetics*, 253, 217–224.

Lewis, J.A. & Hatfull, G.F. (2001) Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Research*, 29, 2205–2216.

Lewis, J.A. & Hatfull, G.F. (2003) Control of directionality in L5 integrase-mediated site-specific recombination. *Journal of Molecular Biology*, 326, 805–821.

Loot, C., Millot, G.A., Richard, E., Littner, E., Vit, C., Lemoine, F. et al. (2024) Integron cassettes integrate into bacterial genomes via widespread non-classical attG sites. *Nature Microbiology*, 9, 228–240.

Lu, F. & Churchward, G. (1994) Conjugative transposition: Tn916 integrase contains two independent DNA binding domains that recognize different DNA sequences. *The EMBO Journal*, 13, 1541–1548.

Malanowska, K., Salyers, A.A. & Gardner, J.F. (2006) Characterization of a conjugative transposon integrase, IntDOT. *Molecular Microbiology*, 60, 1228–1240.

Mata, M., Trautwetter, A., Luthaud, G. & Ritzenthaler, P. (1986) Thirteen virulent and temperate bacteriophages of *lactobacillus bulgaricus* and *lactobacillus lactis* belong to a single DNA homology group. *Applied and Environmental Microbiology*, 52, 812–818.

Mattis, A.N., Gumport, R.I. & Gardner, J.F. (2008) Purification and characterization of bacteriophage P22 Xis protein. *Journal of Bacteriology*, 190, 5781–5796.

McLeod, M., Craft, S. & Broach, J.R. (1986) Identification of the cross-over site during FLP-mediated recombination in the *Saccharomyces cerevisiae* plasmid 2 microns circle. *Molecular and Cellular Biology*, 6, 3357–3367.

Menouni, R., Hutinet, G., Petit, M.-A. & Ansaldi, M. (2015) Bacterial genome remodeling through bacteriophage recombination. *FEMS Microbiology Letters*, 362, 1–10.

Miele, S., Provan, J.I., Vergne, J., Possoz, C., Ochsenbein, F. & Barre, F.-X. (2022) The Xer activation factor of TLCΦ expands the possibilities for Xer recombination. *Nucleic Acids Research*, 50, 6368–6383.

Missirlis, P.I., Smailus, D.E. & Holt, R.A. (2006) A high-throughput screen identifying sequence and promiscuity characteristics of the *loxP* spacer region in Cre-mediated recombination. *BMC Genomics*, 7, 1–13.

Nolivos, S., Pages, C., Rousseau, P., Le Bourgeois, P. & Cornet, F. (2010) Are two better than one? Analysis of an FtsK/Xer recombination system that uses a single recombinase. *Nucleic Acids Research*, 38, 6477–6489.

Nunes-Duby, S.E., Kwon, H.J., Tirumalai, R.S., Ellenberger, T. & Landy, A. (1998) Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Research*, 26, 391–406.

Nunes-Düby, S.E., Yu, D. & Landy, A. (1997) Sensing homology at the strand-swapping step in lambda excisive recombination. *Journal of Molecular Biology*, 272, 493–508.

Panis, G., Duverger, Y., Champ, S. & Ansaldi, M. (2010) Protein binding sites involved in the assembly of the KplE1 prophage intasome. *Virology*, 404, 41–50.

Panis, G., Méjean, V. & Ansaldi, M. (2007) Control and regulation of KplE1 prophage site-specific recombination: a new recombination module analyzed. *The Journal of Biological Chemistry*, 282, 21798–21809.

Pedulla, M.L., Lee, M.H., Lever, D.C. & Hatfull, G.F. (1996) A novel host factor for integration of mycobacteriophage L5. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 15411–15416.

Peña, C.E., Stoner, J.E. & Hatfull, G.F. (1996) Positions of strand exchange in mycobacteriophage L5 integration and characterization of the *attB* site. *Journal of Bacteriology*, 178, 5533–5536.

Peña, C.E.A., Kahlenberg, J.M. & Hatfull, G.F. (2000) Assembly and activation of site-specific recombination complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 7760–7765.

Petit, M.A., Ehrlich, D. & Jannière, L. (1995) pAM beta 1 resolvase has an atypical recombination site and requires a histone-like protein HU. *Molecular Microbiology*, 18, 271–282.

Raynal, A., Tuphile, K., Gerbaud, C., Luther, T., Guérineau, M. & Pernodet, J.-L. (1998) Structure of the chromosomal insertion site for pSAM2: functional analysis in *Escherichia coli*. *Molecular Microbiology*, 28, 333–342.

Rice, P., Longden, I. & Bleasby, A. (2000) EMBOSS: the European molecular biology open software suite. *Trends in Genetics*, 16, 276–277.

Ringwald, K. & Gardner, J. (2015) The *Bacteroides thetaiotaomicron* protein *Bacteroides* host factor a participates in integration of the integrative conjugative element CTnDOT into the chromosome. *Journal of Bacteriology*, 197, 1339–1349.

Roberts, A.P. & Mullany, P. (2009) A modular master on the move: the Tn*916* family of mobile genetic elements. *Trends in Microbiology*, 17, 251–258.

Rowland, S.-J., Stark, W.M. & Boocock, M.R. (2002) Sin recombinase from *Staphylococcus aureus*: synaptic complex architecture and transposon targeting. *Molecular Microbiology*, 44, 607–619.

Rubio-Cosials, A., Schulz, E.C., Lambertsen, L., Smyshlyaev, G., Rojas-Cordova, C., Forslund, K. et al. (2018) Transposase-DNA complex structures reveal mechanisms for conjugative transposition of antibiotic resistance. *Cell*, 173, 208–220.e20.

Rutherford, K. & Van Duyne, G.D. (2014) The ins and outs of serine integrase site-specific recombination. *Current Opinion in Structural Biology*, 24, 125–131.

Sarkar, D., Radman-Livaja, M. & Landy, A. (2001) The small DNA binding domain of lambda integrase is a context-sensitive modulator of recombinase functions. *The EMBO Journal*, 20, 1203–1212.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B. et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75, 7537–7541.

Shannon, C.E. (1948) A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–423.

Sheren, J., Langer, S.J. & Leinwand, L.A. (2007) A randomized library approach to identifying functional *lox* site domains for the Cre recombinase. *Nucleic Acids Research*, 35, 5464–5473.

Silpe, J.E. & Bassler, B.L. (2019) A host-produced quorum-sensing autoinducer controls a phage lysis-Lysogeny decision. *Cell*, 176, 268–280.e13.

Smith-Mungo, L., Chan, I.T. & Landy, A. (1994) Structure of the P22 *att* site. Conservation and divergence in the lambda motif of recombinogenic complexes. *The Journal of Biological Chemistry*, 269, 20798–20805.

Smyshlyaev, G., Bateman, A. & Barabas, O. (2021) Sequence analysis of tyrosine recombinases allows annotation of mobile genetic elements in prokaryotic genomes. *Molecular Systems Biology*, 17, e9880.

Swinger, K.K. & Rice, P.A. (2004) IHF and HU: flexible architects of bent DNA. *Current Opinion in Structural Biology*, 14, 28–35.

Sylwan, L., Frumerie, C. & Haggård-Ljungquist, E. (2010) Identification of bases required for P2 integrase core binding and recombination. *Virology*, 404, 240–245.

Touchon, M., Bernheim, A. & Rocha, E.P. (2016) Genetic and life-history traits associated with the distribution of prophages in bacteria. *The ISME Journal*, 10, 2744–2754.

Turan, S., Kuehle, J., Schambach, A., Baum, C. & Bode, J. (2010) Multiplexing RMCE: versatile extensions of the Flp-recombinase-mediated cassette-exchange technology. *Journal of Molecular Biology*, 402, 52–69.

Umlauf, S.W. & Cox, M.M. (1988) The functional significance of DNA sequence structure in a site-specific genetic recombination reaction. *The EMBO Journal*, 7, 1845–1852.

Van Duyne, G.D. & Landy, A. (2024) Bacteriophage lambda site-specific recombination. *Molecular Microbiology*, 121, 895–911. Available from: https://doi.org/10.1111/mmi.15241

Walker, M.W.G., Klompe, S.E., Zhang, D.J. & Sternberg, S.H. (2023) Novel molecular requirements for CRISPR RNA-guided transposition. *Nucleic Acids Research*, 51, 4519–4535.

Weisberg, R.A., Enquist, L.W., Foeller, C. & Landy, A. (1983) Role for DNA homology in site-specific recombination. The isolation and characterization of a site affinity mutant of coliphage lambda. *Journal of Molecular Biology*, 170, 319–342.

Wojciak, J.M., Sarkar, D., Landy, A. & Clubb, R.T. (2002) Arm-site binding by lambda-integrase: solution structure and functional characterization of its amino-terminal domain. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 3434–3439.

Wood, M.M., DiChiara, J.M., Yoneji, S. & Gardner, J.F. (2010) CTnDOT integrase interactions with attachment site DNA and control of directionality of the recombination reaction. *Journal of Bacteriology*, 192, 3934–3943.

Yu, A. & Haggård-Ljungquist, E. (1993) Characterization of the binding sites of two proteins involved in the bacteriophage P2 site-specific recombination system. *Journal of Bacteriology*, 175, 1239–1249.

Zeng, L., Skinner, S.O., Zong, C., Sippy, J., Feiss, M. & Golding, I. (2010) Decision making at a subcellular level determines the outcome of bacteriophage infection. *Cell*, 141, 682–691.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.