Site-specific recombination by φC31 integrase and other large serine recombinases

Margaret C.M. Smith*,1, William R.A. Brown†, Andrew R. McEwan* and Paul A. Rowley*

*Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, U.K. and †Institute of Genetics, University of Nottingham, Nottingham NG7 2UH, U.K.

Abstract
Most temperate phages encode an integrase for integration and excision of the prophage. Integrases belong either to the λ Int family of tyrosine recombinases or to a subgroup of the serine recombinases, the large serine recombinases. Integration by purified serine integrases occurs efficiently in vitro in the presence of their cognate (~50 bp) phage and host attachment sites, attP and attB respectively. Serine integrases require an accessory protein, Xis, to promote excision, a reaction in which the products of the integration reaction, att and attR, recombine to regenerate attP and attB. Unlike other directional recombinases, serine integrases are not controlled by proteins occupying accessory DNA-binding sites. Instead, it is thought that different integrase conformations, induced by binding to the DNA substrates, control protein–protein interactions, which in turn determine whether recombination proceeds. The present review brings together the evidence for this model derived from the studies on φC31 integrase, Bxb1 integrase and other related proteins.

Introduction
After infection of a sensitive host cell, temperate bacteriophages ‘make a decision’ between lytic growth and lysogeny, a latent state in which phage lytic genes are repressed and the phage genome is often integrated into the host chromosome [1]. Lysogeny is reversible, and a minority of cells within a population of lysogens will spontaneously lyse, releasing progeny phage. Allan Campbell [2] first proposed that the λ prophage is integrated into the host chromosome at a specific site by recombination, a process that requires specialized ‘attachment sites’, attP and attB, in the phage and host chromosomes respectively [3]. The integrated prophage is flanked by attL and attR, which are the substrates for excisive recombination, a process that reforms attP and attB and releases the prophage.

Phage integrases within the serine recombinase family
Until the mid-1990s, all known phage integrases were related to the Escherichia coli K12 phage λ integrase. The λ Int proteins belong to the large and diverse family that use a tyrosine residue to generate a transient 3′ phosphotyrosine link to store the energy of the phosphodiester bond during recombination [3,4]. In contrast, the integrases from the Streptomyces phages φC31 and R4, the lactococcal phage TP901-1 and SpoIVCA, a recombinase that excises a prophage-like element from the Bacillus genome during sporulation, were found to be related to the resolvase/invertases and therefore belong to the serine recombinase family of site-specific recombinases [5–8]. The serine recombinases are evolutionarily and mechanistically distinct from the λ Int family as they use a serine residue as a nucleophile during catalysis [4]. Recombination by the resolvase/invertases occurs within a tetramer of recombinase subunits bound to two DNA substrates [4]. A concerted double-strand break in both substrates occurs with the concomitant formation of covalent 5′ phosphoserine links to the recombinase subunits and an overhanging 3′ dinucleotide. Strand exchange is thought to involve a mechanism in which two recombinase subunits covalently bound to half-sites from different DNA substrates are exchanged by a 180° rotation with respect to the other two subunits [9–11]. This rotation mechanism brings together the cleaved ends of the two half-sites that will comprise the products; the ends are joined and integrase is released. In the resolvase/invertases, the catalysis and strand exchange is driven by the NTD (N-terminal domain).

The sequence similarity between the phage-encoded serine integrases with the resolvase/invertases extends only so far as the catalytic NTDs, and it is clear that these proteins use the same mechanisms of DNA cleavage and DNA strand exchange as the resolvase/invertases [7,12–17]. In the place of the short ~40-amino-acid CTD (C-terminal domain) in the resolvase/invertases, SpoIVCA and the phage-derived serine integrases have much larger C-terminal domains (~300 to ~500 amino acids) and so earned a collective term, the large serine recombinases [18]. Alignments of the large serine recombinases show that, although they are a diverse group of proteins, there are common motifs in the CTDs, mostly

Key words: bacteriophage, excision, φC31 integrase, integration, protein–DNA interaction, protein–protein interaction, serine recombinase.

Abbreviations used: CTD, C-terminal domain; NTD, N-terminal domain; RDF, recombination directionality factor.

1To whom correspondence should be addressed (email maggie.smith@abdn.ac.uk).
Figure 1 | Conserved motifs in φC31 and Bxb1 integrases and in the transposase, TnpX

The horizontal black line is a linear representation of the amino acid sequence. Five conserved motifs identified either by sequence alignments or a structural prediction algorithm, COILS [34,59], are colour-coded: green, pfam0239, the resolvase NTD; purple, pfam07508; yellow, the putative zinc finger; blue, putative coiled-coil structural motif; orange, a conserved isoleucine-, leucine- and valine-rich region). The sequence co-ordinates for each motif in the three proteins are indicated. No coiled-coil motif is included on Bxb1 integrase because, although the COILS program found a candidate motif, it fell short of the statistical criteria. The white boxes summarize structure-function experiments in which protein fragments have been prepared and their \textit{in vitro} activities reported [15,39,44,60]. For each protein, specific amino acid substitutions are also indicated which have revealed structure-function information. Abbreviations: N, N-terminus; C, C-terminus; PSS, protease-sensitive site; PI, protein interaction activity.

of unknown function (Figure 1) [18]. The CTD is, however, known to be required for DNA binding and for the control of integration against excision (see below).

The incidence of the large serine recombinases appears to be enriched in phages that infect actinobacteria and in the mobile genetic elements of the Gram-positive bacteria. The conjugative transposon TnJ397 from \textit{Clostridium difficile} and the mobilizable transposon Tn4451 in \textit{Clostridium perfringens} both contain large serine recombinases, TndX and TnpX for integration and excision respectively of
the elements [19–21]. The SCC elements (staphylococcal chromosomal cassettes) that encode the clinically important methicillin-resistance gene, mec, are also integrated and excised via large serine recombinases [22].

**The default activity of phage-encoded serine integrases is integration**

Several of the phage-encoded serine integrases have been expressed in *E. coli* and purified for use in *in vitro* recombination assays and/or DNA-binding assays [13,23–28]. The integration reactions mediated by φC31, Bxb1, φRv1, φBT1 and φFCG1 integrases require a simple buffer, the substrates (attP and attB), which can be on the same or separate, supercoiled or linear molecules, and no additional energy sources. The recombination products attL and attR comprise a half-site from attP and a half-site from attB in a reciprocal format. The minimal recombination sites are generally less than 50 bp for both attP and attB [14,23,25–27,29,30]. This contrasts with the phage λ integration system in which a 240 bp attP site contains the core and arm-binding sites for λ Int and the accessory-protein-binding sites [3]. The reaction conditions for the serine integrases are also different from those of the resolvase/invertases that require supercoiled DNA with both recombination sites located on the same DNA molecule [4,31,32].

At first glance, the properties of the phage serine integrases in these *in vitro* recombination reactions resemble those for the bidirectional recombinase Cre [33]. However, all of the serine integrases are unidirectional in the absence of other proteins [13,24,25,27]. Indeed, even using a highly sensitive *in vitro* recombination assay for φC31 integrase activity, no recombination could be detected between any pair of recombination sites other than attP and attB [34]. In fact, serine integrases require an accessory protein, an Xis or RDF (recombination directionality factor), to mediate the excision reaction between attL and attR to produce attP and attB [23,35,36]. At the mechanistic level, how does integrase discriminate between the four attachment sites without resorting to the use of accessory binding sites as seen with other directional recombinases?

**Distinct Int–attP and Int–attB interactions enable synopsis during integration**

For integration to occur, integrase must recognize and bind to attP and attB and bring these sites together in a synapse before DNA cleavage, strand exchange and ligation of the products (Figure 2). φC31 and Bxb1 integrases are dimers in solution and bind to the attachment sites with one subunit contacting each half-site [30]. The bound dimers then bring attP and attB together in a synaptic complex that probably contains a tetramer of integrase subunits [15] (Figure 2). With φC31 integrase (in this case, a catalytic mutant of integrase IntS12A), the synaptic complexes are observed when integrase is provided with attP and attB [16]. With native φC31 integrase and attP/attB as substrates, there is little accumulation of synaptic complex, and it is likely that activation of DNA cleavage and strand exchange occurs very rapidly compared with the rate of formation of the synapse [34,37].

The critical stage of recombination where sites are selected for recombination is at synopsis. φC31 IntS12A, in the absence of any other proteins, cannot synapse any pair of attachment sites other than attP and attB, including, for example, two attP or attB sites, or attR with attL. The selection of sites for synopsis is not due to differences of affinity [15,30]. The favoured hypothesis is that integrase adopts distinct conformations when bound to attP and attB that enable the formation of synaptic tetramers (Figure 2) [15,16,30]. Mutations in φC31 attB can severely reduce the rate of synopsis without affecting binding affinity, and this is consistent with a mechanism in which DNA contacts induce specific conformational changes [38].

DNA recognition in the large serine recombinases occurs through the CTDs. The CTDs from Bxb1 and φC31 have
been purified as isolated domains and shown to bind specifically to their cognate attP and attB sites [15,39]. In both cases, the isolated CTDs are monomers in free solution, but the φC31 CTD binds co-operatively to attB (and attL and attR), but not to attP, indicating that integrase recognizes attB and attP differently [39]. This is perhaps not surprising given that the sequences of attP and attB are very different from each other (see, for example, [18]).

Structure–function studies have not revealed a common DNA-binding motif within the large CTDs (Figure 1). A fragment of Bxb1 integrase containing the NTD and a subdomain of the CTD having homology with the conserved ‘recombinase’ motif (pfam07508) binds with low affinity [15]. A mutation in a linker that joins the NTD and CTD in Bxb1 integrase reduces binding to attB, but not to attP, and this may be the first clue that different motifs of integrase interact with different attachment sites [15]. Moreover, the phenotype of this mutant adds further to the evidence for different integrase conformations when bound to attB and attP. Structure–function analysis with TnpX that showed this protein has two regions at the extreme C-terminus involved in DNA binding. The most C-terminal of these two DNA-binding motifs can be removed, and the protein still functions as a transposase, suggesting that this motif has an accessory function. The second DNA-binding motif is not in a conserved region, and so it is not clear whether other recombinesins have an equivalent function at this position. Although mutations in TnpX that map in the region with homology with pfam07508 can affect binding, a fragment containing this region does not bind to DNA [40].

This discussion highlights how little we understand about how the serine integrases recognize their sites, both in terms of the protein motifs that are used to contact DNA and the sequence requirements in the attachment sites themselves.

### Protein–protein interactions between integrase subunits control synapsis

The serine integrases bring attP and attB together to form a putative synaptic tetramer during the integration reaction. In solution, φC31, Bxb1 integrases and the isolated NTDs from Bxb1 and TP901-1 integrases are dimers [15,30,39,41]. The structure of the TP901-1 NTD shows that the serine integrases dimerize via a long α-helix (the E-helix) that, in the resolvase/invertase family, also mediates dimer formation [41,42].

The NTDs also mediate the interactions that bring the recombinase dimers together to form a synaptic tetramer at the crossover site [4,10,41]. Structural studies of resolvase and the NTD of TP901-1 integrase provide details of the synaptic interface at the atomic level [10,41]. The formation of the synaptic interface through the NTD–NTD interactions activates DNA cleavage and is therefore a critical stage in all serine recombinase reactions [4]. In φC31 integrase, Val129 is equivalent to residues in resolvase, Hin invertase and in TP901-1 NTD that have been shown to contribute to the NTD synaptic interface [10,37,41,43]. Substitutions in φC31 integrase can affect synapsis (V129A) and DNA cleavage (V129G), phenotypes that strongly suggest that this interface is both conserved and critical for integrase activity.

Recent data show that the CTD also has a role in the synopsis of substrates in φC31 integrase [39]. The isolated φC31 integrase CTD, a monomer in free solution, binds attP and attB and brings them together to form a putative synaptic tetramer. The CTD synapse echoes the synapse made by full-length integrase, as both complexes only form with attP and attB. Studies on hyperactive mutants of φC31 integrase provide some insights on the role of the CTD synaptic complex [34,39]. Mutants of φC31 integrase were isolated that could recombine attL×attR [34]. IntE449K is similar to native integrase in attP×attB recombination activity, but can also synapse and recombine attL×attR (and reactions that mimic excision, i.e. attL×attL and attR×attR). It seems likely that E449K disrupts a mechanism that normally prevents attL×attR synopsis. The E449K substitution in the context of the isolated CTD prevented the formation of the CTD synapse. This suggests that the CTD synapse may be required to remove a block on the formation of the NTD synaptic interface. Thus IntE449K is free to synapse attL×attR, attL×attL, attR×attR in addition to attP×attB entirely via NTD interactions. The observation that E449K partially rescues a mutation in the NTD (V129A) that reduces the rate of formation of the synaptic complex is consistent with the idea that the NTD and CTD interact to control synapsis [37].

### A coiled-coil motif in the large CTD mediates protein–protein interactions

All of the mutants that showed any hyperactivity were clustered in the region 449–485 (Figure 1), where structure predictions indicate that φC31 integrase (and several other serine integrases) may adopt a coiled coil [34]. These motifs are often used for protein–protein interaction and a construct containing this region fused to MBP (maltose-binding protein) was able to oligomerize [39]. A fragment of TnpX containing the equivalent region could also oligomerize [44] (Figure 1). In φC31 integrase, the coiled-coil motif appears to have a role in formation of the CTD synaptic complex, as substitutions that disrupt the protein–protein interaction cannot form the CTD synaptic tetramer. However, these substitutions also disrupt DNA binding and/or co-operative binding. The current model proposes that the coiled-coil motifs mediate an interaction between CTDs bound to adjacent half-sites in attB, attL and attR, whereas the coiled-coil motifs in the CTDs bound to attP are poised to nucleate interactions between subunits bound to attP and attB, leading to the formation of the CTD synaptic tetramer. This tetramer, which could be transient in the presence of the NTDs, then releases the inhibition of the synopsis via the NTDs. This model explains the co-operative/non-co-operative DNA binding by the isolated CTD and the phenotypes of amino acid substitutions leading to hyperactivity.
Integrate complexes with \( \text{attP} \) and \( \text{attB} \) are functionally symmetrical and asymmetrical with \( \text{attL} \) and \( \text{attR} \)

The hyperactive \( \phi \text{C31} \) integrase mutant \( \text{IntE449K} \) permitted a comparison of the integration and excision reactions [17,34]. As \( \text{attP} \) and \( \text{attB} \) contain imperfect inverted repeats centred around the 2 bp crossover site, they contain two P-type and two B-type arms respectively, whereas \( \text{attL} \) and \( \text{attR} \) contain one B-type and one P-type arm in a reciprocal arrangement. By manipulating the sequence at the 2 bp crossover site, the nature of the synaptic complex can be inferred by studying the products of recombination [14,17]. To date, the 2 bp crossover sequences in the substrates for every large serine recombines so far studied are non-palindromic. This is for a good reason, as it ensures that, during integration, the left arm of \( \text{attB} \) becomes joined to the right arm of \( \text{attP} \) and vice versa to form \( \text{attL} \) and \( \text{attR} \) respectively [14,17]. If the 2 bp crossover sequence is made to be palindromic, aberrant products are observed where the joining is left arm to left arm and right arm to right arm and this occurs at the same frequency as formation of the normal products. The two types of products are observed because two types of synapse are constructed that differ by the parallel or antiparallel alignment of the substrates. This can only occur if integrase dimers bound to \( \text{attP} \) and \( \text{attB} \) are functionally symmetrical. Using the same approach for the \( \text{attL} \times \text{attR} \) synopsis with the hyperactive \( \phi \text{C31} \) integrase mutant, a preferred excision synapse is inferred in which integrase subunits bound to P-type arms contact integrase subunits bound to B-type arms at the tetramer interface [34]. This implies that the integrase dimers bound to \( \text{attL} \) and \( \text{attR} \) are functionally asymmetrical.

The formation of the synaptic tetramer is therefore influenced by the half-sites to which integrase subunits are bound. It is worth emphasizing, however, that the integrase subunits do not act independently of each other in adopting a conformation that will enable a subunit–subunit interaction, as even with the hyperactive mutant, the selection of DNA substrates to synapse (and recombine) is not indiscriminate. Synapsis is only observed where the substrates collectively contain two B-type and two P-type arms. This argues that the conformation of the integrase dimer bound to the attachment site is the unit that determines whether a synaptic interface is made.

The role of Xis in excision

The serine integrases, like those that belong to the tyrosine integrase group, recruit an Xis or RDF to enable excision of the prophage [23,35,36,45]. Xis has been identified for three phages, TP901-1, \( \phi \text{Rv1} \) and Bxb1, and there is no detectable sequence conservation between them. TP901-1 Xis (encoded by \( \text{orf7} \) on the TP901-1 genome) is a 64–amino-acid basic protein whose only homologue in the database is from another lactococcal phage which also has a serine integrase [35]. \( \phi \text{Rv1 gpXis} \) is also small (73 amino acids) and basic, and many mycobacteriophage genomes encode homologues [23]. The Bxb1 RDF, gp47, is 255 amino acids in length and has a conserved domain in the protein phosphatase family [36]. Close homologues of Bxb1 gp47 are present in other mycobacteriophages, including those that carry a tyrosine integrase (e.g. L5), and it has been suggested that this protein may have a function in addition to its role as an RDF. In the \( \lambda \) Int family, the genes encoding integrase (int) and Xis (xis) are normally neighbours. In \( \phi \text{Rv1}, \text{TP901-1} \) and Bxb1, the genes encoding integrase and Xis are separated by one, eleven and five genes respectively [23,35,36]. The diversity of the Xis proteins and the random location of the Xis-encoding genes suggest that phage proteins could be recruited to control integrase directionality on a rather ad hoc basis.

Studies on the biochemical mechanism of the RDF from Bxb1 indicate that this protein controls integrase directionality by protein–protein interactions. Unlike other RDFs, Bxb1 gp47 does not bind to DNA, and the minimal sites for \( \text{attL} \times \text{attR} \) recombination are \( \sim 50 \) bp [36]. Bxb1 gp47 recognizes and binds tightly to integrase bound to \( \text{attP} \) and \( \text{attB} \), which may explain the inhibition by gp47 on integration. Although gp47 binds less strongly to integrase bound to \( \text{attL} \) and \( \text{attR} \), it stimulates formation of the integrase/\( \text{attL} \times \text{attR} \) synapse. Hatfull and colleagues have also demonstrated the same structural bias in the formation of excisive synaptic complexes as seen with the hyperactive \( \phi \text{C31} \) integrase described above [45]. The bias in the alignment of the substrates at synopsis combined with the non-palindromic nature of the 2 bp at the crossover site prevents unwanted \( \text{attL} \times \text{attL} \) or \( \text{attR} \times \text{attR} \) recombination and ensures the correct use of \( \text{attL} \times \text{attR} \) for prophage excision [45].

Behaviour of \( \phi \text{C31} \) integrase in the eukaryotic cell

The in vitro properties of \( \phi \text{C31} \) integrase prompted its application as a tool for genome engineering in heterologous systems, and its activity has been demonstrated in many eukaryotic organisms [29,46–50]. The irreversible activity of \( \phi \text{C31} \) integrase has enabled efficient integration and inversion of DNA segments, manipulations which are either inefficient or incomplete with a bidirectional recombinase such as Cre [51,52]. Precise genome engineering requires the insertion of the cognate \( \text{attP} \) and \( \text{attB} \) sites. However, it appears that attachment site recognition by integrase is sufficiently relaxed [53] to permit the use of endogenous sequences in the human and mouse genome as integration sites in a novel approach to gene therapy [54,55]. This ‘ectopic’ recombination occurs at very low frequency and integrase targets many so-called ‘pseudo-\( \text{attP} \)’ sites within the human and mouse genomes [29,52,56]. Despite these problems, it is has been shown that \( \phi \text{C31} \) integrase may be ‘improved’ by protein engineering to recognize the pseudo-sites more efficiently [57].

The behaviour of \( \phi \text{C31} \) integrase in eukaryotic cells has contributed to our mechanistic understanding of these proteins. Integrate-mediated \( \text{attP} \times \text{attB} \) recombination in chick or in human cells lines often leads to small deletions in the recombination sites [52]. It was proposed that these
deletions were the result of NHEJ (non-homologous end-joining) enzymes detecting the recombination intermediate containing integrase covalently bound to the cleaved substrates, which are then repaired. If this is the case, then it implies that the rotation mechanism that drives strand exchange may be much slower in the nuclear environment, possibly due to the presence of nucleosomes and/or different DNA topology compared with that in the prokaryotic nucleoid. With naked supercoiled substrates, C31 integrase does not cause accumulation of cleaved intermediates, but they are easily detected when both substrates are linear. This suggests that negative supercoiling does stimulate either strand exchange or joining of the products [34,35].

A second observation is the preference for one substrate (attP) over the other (attB) as the docking sites in heterologous systems [29]. The bias for this is not clear, but one possibility is that there are cellular proteins that interact with and inhibit integrase bound to attB, but not to attP, possibly due to different conformations of integrase bound to the two sites. In support of this a human protein, DAXX (death-domain-associated protein), has been shown to interact with the coiled-coil motif, which as described above has profound effects on integrase activity [34,38].

Funding

Funding for research in the M.C.M.S. and W.R.A.B. laboratories is from the Biotechnology and Biological Research Council [grant numbers BB/D007836/1 and BB/000894/1].

References


10 Li, W., Kamtekar, S., Xiong, Y., Sarkis, G.J., Grindley, N.D. and Steitz, T.A. (2005) Structure of a synaptic φX resolvase tetramer covalently linked to two cleaved DNAs. Science 309, 1210-1215


12 Crellin, P.K. and Rood, J.I. (1997) The resolve/invertase domain of the site-specific recombinase Tn916 is functional and recognizes a target sequence that resembles the junction of the circular form of the Clostridium perfringens transposon Tn451 J. Bacteriol. 179, 5148-5156


Received 24 July 2009
doi:10.1042/BST0380388