The Serine Recombinases

W. MARSHALL STARK
Institute of Molecular, Cell and Systems Biology, University of Glasgow, Bower Building, Glasgow G12 8QQ, Scotland, United Kingdom

ABSTRACT In site-specific recombination, two short DNA sequences ('sites') are each cut at specific points in both strands, and the cut ends are rejoined to new partners. The enzymes that mediate recognition of the sites and the subsequent cutting and rejoining steps are called recombinases. Most recombinases fall into one of two families according to similarities of their protein sequences and mechanisms; these families are known as the tyrosine recombinases and the serine recombinases, the names referring to the conserved amino acid residue that attacks the DNA phosphodiester and becomes covalently linked to a DNA strand end during catalysis. This chapter gives an overview of our current understanding of the serine recombinases, their types, biological roles, structures, catalytic mechanisms, mechanisms of regulation, and applications.

SITE-SPECIFIC RECOMBINATION: A BRIEF PRIMER

Introduction
The term site-specific recombination encompasses a group of biological processes that, unlike homologous recombination, promote rearrangements of DNA by breaking and rejoining strands at precisely defined sequence positions. In a canonical site-specific recombination event, two discrete sites (sequences of DNA, typically a few tens of base pairs long) are broken, and the ends are reciprocally exchanged and rejoined, resulting in recombinant products (Fig. 1). Site-specific recombination does not require extensive sequence homology; the sites are identified and brought together by protein–DNA and protein–protein interactions involving specialized recombinase proteins, unlike homologous recombination where DNA–DNA interactions define the loci of strand exchange. “Conservative” site-specific recombination systems form recombinants without any requirement for DNA synthesis or high-energy cofactors. Some other biological processes such as transposition are sometimes categorized with site-specific recombination because of common features including cleavage and rejoining of DNA strands at precise positions defined by protein–DNA interactions, but these processes may require DNA synthesis and/or ligase-mediated rejoining of DNA strands. The systems discussed in this chapter conform to the strict “conservative” definition. General aspects of site-specific recombination have been reviewed elsewhere (1, 2, 3).

Site-specific recombination can have different outcomes depending on the nature of the DNA substrate(s) (Fig. 2). Recombination between two sites, each on a separate linear DNA molecule, results in linear recombinants. Two outcomes are possible, depending on which “half-site” is joined to which, as shown in Fig. 2a. However, typical sites have a polarity, such that the “left half” of one site is joined to the “right half” of the other, and vice versa; thus, only one of these possibilities normally occurs. The origin of the site polarity is discussed below. If the two sites are on separate molecules but one or both molecules are circular (Fig. 2b), recombination will join the two molecules together (this is called integration or fusion). The product molecule contains two sites, oriented in a direct repeat (head-to-tail) relationship. Conversely, recombination of this two-site molecule splits it into two products (this is called excision or resolution). If two sites within a single DNA molecule are in an inverted relationship (Fig. 2c), recombination...
inverts the orientation of one DNA segment bounded by the sites, relative to the other. In most real site-specific recombination systems, restrictions imposed by the mechanism of recombinase-mediated catalysis allow only some of these possibilities (see below). Site-specific recombination is seemingly isoenergetic; the products, like the substrates, are normal double-stranded DNA molecules. Reactions might therefore be expected to reach a 1:1 equilibrium of substrates and recombinants. However, natural systems have evolved strategies to bias the reaction toward the desired products; some examples are described in the sections that follow.

Conservative site-specific recombination has been adopted widely for diverse programmed DNA rearrangements essential to the biology of bacteria, archaea and the mobile DNA elements that infest them (bacteriophages, plasmids and transposons) (2, 4, 5, 6). Curiously, however, there are only a few known conservative site-specific recombination systems in eukaryotes, and some of these may be associated with bacterial symbionts or bacterial-derived organelles, or may be recent acquisitions from horizontal transfer of mobile DNA (1, 5, 6, 7, 8). Roles of site-specific recombination systems include temperate bacteriophage DNA integration and excision from the host bacterial genomic DNA, transposon cointegrate resolution, monomerization of plasmid multimers, switching of gene expression by inversion of regulatory sequences relative to coding sequences and developmentally programmed excision of intervening genomic sequences. There is no clear distinction of the biological functions of conservative site-specific recombination systems in bacteria.

**FIGURE 1** Site-specific recombination. Two sites (pointed boxes) in double-helical DNA (shown as double lines) are recognized by a recombinase protein (not shown), and then cut and rejoined to form recombinants. doi:10.1128/microbiolspec.MDNA3-0046-2014.f1

**FIGURE 2** Site-specific recombination outcomes. (a) Recombination between two sites in separate linear DNA molecules results in two linear recombinant products. Usually, the sites have a polarity (indicated by the pointed boxes) such that the lower pathway (red arrow) is forbidden. (b) Recombination between two sites in separate DNA molecules, when one or both of the molecules is circular, results in a single product molecule containing two sites in direct repeat. This is called integration or fusion. The "reverse" reaction splits a molecule containing two sites into two product molecules, one or both of which are circular. This is called resolution, excision, or deletion (depending on the biological context). (c) Recombination between two sites in inverted repeat in a DNA molecule inverts the orientation of one segment of DNA relative to the other. This is called inversion. doi:10.1128/microbiolspec.MDNA3-0046-2014.f2
Mechanistic nuts and bolts

In this section, I will give a brief overview of the molecular mechanisms of conservative site-specific recombination.

In many systems, recombination takes place between two identical sites, and two identical sites are re-formed in the recombinants. However, there are examples (notably bacteriophage integrase systems; see “Regulation of recombination by large serine recombinases” below) where recombination is between two different sites. Sites range in length from about 25 up to several hundreds of base pairs. The shortest sites typically have imperfect 2-fold (dyad/palindrome) DNA symmetry, consistent with their observed or inferred property of binding a symmetric dimer of recombinase. The specific phosphodiester linkages that are cut and rejoined during recombination are located close to the center of the site (Fig. 3).

Longer sites comprise a “crossover site” conforming to the above description, which binds a recombinase dimer and within which are the points of strand exchange, as well as adjacent “accessory sequences” on one or both sides of the crossover site, which may include binding sites for additional recombinase subunits or other “accessory proteins,” or for looping interactions with recombinase subunits bound at the crossover site (Fig. 3a). The roles of the accessory sequences are in regulation of recombinase activity; initiation of catalysis typically depends on their presence and their correct interactions with other components of the system (1, 2, 3).

Recombinases do not cut the two DNA strands at the precise center of the site. Instead, the break points are symmetrically positioned off-center, so that there are a few base pairs between the top strand and bottom strand break points. These base pairs are often referred to as the “overlap sequence” because the top and bottom strands of this sequence in the recombinant sites originate from different parent sites. All serine recombinase systems examined in this respect have 2 bp overlap sequences with the strand breaks staggered as shown in Fig. 3b; in contrast, the overlap sequences for tyrosine recombinases vary in length (typically 6 to 8 bp), and the stagger is in the opposite direction. If the “half-sites” that are to be joined to form recombinants do not have complementary overlap sequences, the products would have mismatched base pairs. This scenario can arise if two identical crossover sites are misaligned in the catalytic intermediate such that strand exchange pairs two identical, noncomplementary ends. Serine recombinases do not normally form mismatched recombinants; this is one origin of the site polarity discussed above. However, reactions of “mismatched” sites can have other consequences (see “Subunit rotation” below).

Each crossover site binds a recombinase dimer. A critical subsequent step is when two crossover sites come together; this is called synapsis. The “synapse” or “synaptic complex” that is thus formed comprises the two crossover sites bridged by a recombinase tetramer, and it is in this intermediate that the chemical steps of strand cleavage, exchange and ligation will take place. In regulated systems, crossover-site synapsis is typically a control point that depends on interactions with accessory factors.

In any conservative site-specific recombination event, there are eight chemical steps: four strand cleavages and four ligations. Cleavage occurs when a nucleophile amino acid functional group at the recombinase active site attacks the scissile phosphodiester bond of a DNA strand; for the serine recombinases, this is the hydroxyl group of a serine residue. The immediate product of cleavage has a broken DNA strand, with a covalent phosphodiester linkage between one DNA end and the
recombinase at the break point. Serine recombinases become linked to the 5’ end of the DNA, leaving a 3’-hydroxyl group on the other end at the break. Serine recombinases cleave all four DNA strands in the synaptic complex, creating double-strand breaks at the center of each crossover site. Each half-site thus formed has a recombinase subunit covalently attached to its 5’ end, and 2-nt single-stranded protrusions terminated by a 3’-OH group (Fig. 4). The half-sites are then exchanged and re-ligated, creating recombinants. This mechanism contrasts with that of the tyrosine recombinases, which become linked to the 3’ end of the DNA at the strand break and do not make double-strand break intermediates. Instead, they cleave, exchange and re-ligate pairs of single DNA strands; thus, strand exchange proceeds via a “four-way junction” intermediate with two recombinant and two non-recombinant strands (see Chapters [X] and [Y], this volume).

SERINE RECOMBINASES
Some history
Following the discovery of the first site-specific recombinase, λ Int, in the 1970s, it was realized that the product of the tnpR gene encoded by the bacterial penicillin resistance transposon Tn3 has a similar function (9, 10). Detailed characterization of the tnpR gene product (resolvase) and the recombination site (res) soon followed (11, 12, 13). It quickly became apparent that there was a group of enzymes related to Tn3 resolvase, encoded by other bacterial transposons and DNA inversion systems. The group came to be known as the resolvase or resolvase/invertase family (14, 15). Pioneering in vitro studies of the ϕ6 transposon resolvase (closely related to Tn3 resolvase) by Reed and Grindley revealed basic mechanistic differences from λ Int and its relatives (16, 17, 18). It was later shown that the resolvase–DNA linkage is via a serine residue, unlike the tyrosine that is used by λ Int and its brethren (19, 20, 21, 22). In the 1990s, the two families came to be referred to as the “serine” and “tyrosine” recombinases (23, 24).

Serine recombinase proteins
All serine recombinases possess a characteristic catalytic domain, which implements the chemical steps of strand exchange. I will call it the “SR” (serine recombinase) domain throughout this review. The size of the SR domain is remarkably constant (usually about 150 amino acid residues). Several of its amino acid residues are highly conserved and are now known to contribute to the structure of the active site (3). All known serine recombinases have “attachments” to the SR domain, usually at the C terminus; these vary substantially and their specific properties have roles in definition of the recombinase function (25, 26) (Fig. 5). The recombinases studied in the early days (transposon resolvases and invertases) have a simple configuration with the SR domain at the N-terminus linked to a small C-terminal helix–turn–helix (HTH) DNA-binding domain, giving a total length of ~180–200 residues. These have come to be known as the “small serine recombinases”. However, as identification of putative serine recombinases by sequence analysis gathered pace, the diversity of the family became apparent (25, 26) (Fig. 5). Many sequences could be aligned with the entire length of the small serine recombinases but have extensions at the C-terminal end, such as the ISXc5 resolvase. Others have large C-terminal extensions immediately after the SR domain, in place of the HTH domain. An important subgroup of these “large serine recombinases” includes the bacteriophage serine integrases, the first to be identified being that of the Streptomyces phage ϕC31 (28). These proteins (~400 to 700 amino acids) have an N-terminal SR domain followed by a complex, variable multidomain region with DNA-binding and regulatory functions, which are still only partially characterized (28). At first, it seemed that a “rule” was that the SR domain should be at the extreme N terminus of the

FIGURE 4 The serine recombinase strand-exchange mechanism. A synaptic complex of two crossover sites bridged by a recombinase tetramer (yellow ovals) is shown. The four subunits are spaced out, so that the catalytic steps can be seen clearly. The catalytic serine residues are indicated by S-OH. The scissile phosphodiester bonds are represented as circled Ps, and in the first and last panels the 2-bp overlap sequence is indicated by vertical lines. For further details, see text. doi:10.1128/microbiolspec.MDNA3-0046-2014.f4
protein, but proteins with a HTH domain preceding the SR domain were then identified and are now known to be transposases (29, 30).

**Biological roles of serine recombinases**

The role of the patriarchs of the serine recombinase family, the transposon resolvases, is to divide (“resolve”) a large circular DNA molecule into two smaller circles. The natural substrate is a “cointegrate” molecule formed by replicative transposition, which contains two transposons, each with a res recombination site (31, 32). Closely related resolution systems are encoded by some bacterial plasmids and act to reduce plasmid multimers to monomers. DNA invertases promote flipping of the DNA sequence between two sites, thereby switching between different modes of gene expression (often to evade host defenses against infection). The activity of one invertase, Hin, from *Salmonella typhimurium*, is responsible for the phenomenon of flagellar phase variation studied since the 1920s. Other invertases are encoded in bacterial, bacteriophage and plasmid genomes (33). Bacteriophage serine integrases integrate and excise the DNA genomes of “temperate” or “lysogenic” phages to/from the bacterial host chromosomal DNA, like the famous tyrosine recombinase-based phage λ system (34, 35). Recently, some small serine recombinases with similarity to the DNA invertase group have also been shown to be phage integrases (36). As noted above, a group of transposases have an SR domain with a HTH domain at the N terminus (30). In addition, substantial numbers of proteins in the databases have homology to the SR domain but have unknown functions (M. R. Boocock, personal communication); there may still be surprises in store. More details on the functions of particular groups of serine recombinases are given in Chapters [X], [Y] and [Z], this volume.

**Serine recombinase structures**

There is now a substantial bank of structural data at atomic resolution on members of the serine recombinase family. In particular, a series of groundbreaking crystal structures of γδ resolvase obtained by the Steitz laboratory in Yale has gone hand in hand with our developing understanding of serine recombinase mechanisms (37, 38, 39, 40, 41, 42). Recently, crystallographic studies of a distantly related serine resolvase, Sin, have transformed our understanding of the regulatory mechanisms and catalytic active site (43, 44). Chapter [X] (this volume) gives an in-depth review of these data. Further insights have come from the structures of the “attachments” to the SR domain. Table 1 summarizes the crystallographic and nuclear magnetic resonance (NMR) structural data on serine recombinases available at the time of submission of this review (45, 46, 47, 48, 49, 50, 51, 52). Small-angle scattering-based methods have also provided important structural information (53).

As an example, Fig. 6a shows the structure of a γδ resolvase dimer bound to the crossover site of the γδ res recombination site (40). Each subunit (183 amino acids)
comprises an SR domain connected via a short linker peptide to a C-terminal HTH domain. The HTH domains recognize sequence motifs at the ends of the crossover site. Each SR domain comprises a core β-sheet decorated with α-helical and irregular regions, and ends in a long α-helix whose C-terminal region contacts the DNA minor groove near the center of the crossover site. The linker between this helix and the HTH domain lies in the minor groove and bears a structural resemblance to the “AT-hook” motif found in other DNA-binding proteins (54). The two SR domains make a complex network of interactions to form a dimer with imperfect 2-fold symmetry. The crossover site DNA is significantly bent but essentially “B-form.” The positions of the catalytic serine residues are shown in Fig. 6a; like other putative active-site residues, they are not in contact with the DNA in this structure. The resolvase is therefore considered to be bound to the DNA in a precatalytic conformation. However, subsequent structures, also solved by the Steitz group using “activated” γδ resolvase variants (see below), revealed a catalytic intermediate containing a resolvase tetramer bridging two crossover sites that have each been cleaved in both strands (Fig. 6b) (41, 42).

Taken together, the structural data (Table 1) show that the SR-domain fold is very well conserved throughout the serine recombinase family, despite substantial amino acid sequence divergence.

### Serine recombinase mechanism

The early studies of Reed and Grindley demonstrated that the resolvase catalytic mechanism was significantly different from that of λ Int, Cre and other members of the “integrase family” (now tyrosine recombinases) (17, 18). DNA cleavage and rejoining by γδ resolvase occur at precise positions within the 28-bp crossover site of res.

<table>
<thead>
<tr>
<th>PDB accession no.</th>
<th>Description</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2RSL</td>
<td>γδ resolvase; supersedes 1RSL</td>
<td>37, 39</td>
</tr>
<tr>
<td>1GDR</td>
<td>γδ resolvase</td>
<td>38</td>
</tr>
<tr>
<td>1GDT</td>
<td>γδ resolvase dimer bound to res site I DNA</td>
<td>40</td>
</tr>
<tr>
<td>1GHT, 1HX7</td>
<td>γδ resolvase catalytic domain (NMR structures)</td>
<td>45</td>
</tr>
<tr>
<td>1RES, 1RET</td>
<td>γδ resolvase DNA-binding domain (NMR structures)</td>
<td>46</td>
</tr>
<tr>
<td>1ZR2, 1ZR4</td>
<td>γδ resolvase activated mutant tetramer in cleaved-DNA synaptic intermediate</td>
<td>41</td>
</tr>
<tr>
<td>2GM4</td>
<td>γδ resolvase activated mutant tetramer in cleaved-DNA synaptic intermediate</td>
<td>42</td>
</tr>
<tr>
<td>2GM5</td>
<td>γδ resolvase mutant tetramer</td>
<td>42</td>
</tr>
<tr>
<td>1HCR</td>
<td>Hin invertase C-terminal domain bound to DNA motif</td>
<td>47</td>
</tr>
<tr>
<td>1J6+</td>
<td>Hin C-terminal domain bound to wild-type and mutant DNA motifs (also 1JW, 1J8B, 1J7O, 1J7P, and 1J7Q)</td>
<td>48</td>
</tr>
<tr>
<td>3UJ3</td>
<td>Gin activated mutant tetramer. Supersedes 3PLO</td>
<td>49</td>
</tr>
<tr>
<td>4M6F</td>
<td>Gin dimer bound to gix site DNA</td>
<td>50</td>
</tr>
<tr>
<td>2ROQ</td>
<td>Gin tetramer in synaptic complex with res site II DNA</td>
<td>43</td>
</tr>
<tr>
<td>3PKZ</td>
<td>Sin activated mutant tetramer</td>
<td>44</td>
</tr>
<tr>
<td>4KIS</td>
<td>Bacteriophage A118 integrase (C-terminal part bound to att site DNA)</td>
<td>51</td>
</tr>
<tr>
<td>4BQQ</td>
<td>Bacteriophage ΦC31 integrase (N-terminal part)</td>
<td>McMahon SA, McEwan AR, Smith MCM, and Naismith, JH, unpublished data</td>
</tr>
<tr>
<td>3BVP</td>
<td>Bacteriophage TP901-1 integrase catalytic domain</td>
<td>52</td>
</tr>
<tr>
<td>3G13</td>
<td>Transposase from Ctn7 of Clostridium difficile</td>
<td>Bagaria A, Burley SK, and Swaminathan S, unpublished data</td>
</tr>
<tr>
<td>3ILX</td>
<td>TnpA transposase from Sulfolobus solfataricus ISC1904</td>
<td>Chang C, Bigelow L, Bearden J, and Joachimiak A, unpublished data</td>
</tr>
</tbody>
</table>
Alteration of the reaction conditions allowed isolation of products with double-strand breaks at the center of the crossover site. A resolvase subunit is covalently linked to each 5′ end of the linearized DNA (18) (Fig. 4). The protein–DNA linkage was shown later to be a phosphodiester bond with the resolvase Ser10 (19, 21). The in vitro reaction is very efficient; under standard conditions, nearly all the substrate is converted into recombinant products within a few minutes. No cofactors or metal ions such as Mg2+ are required for activity. The analysis of Reed and Grindley also revealed that the product of resolution of a supercoiled plasmid substrate in vitro was a specific simple catenane in which the two product circles are linked as in a chain, an intriguing observation that led to many further studies and insights (see “Topological studies” below) (16).

Studies on related systems, including the resolution systems of Tn3 and Tn21, and the DNA invertases Gin, Hin and Cin, confirmed the generality of the mechanistic insights from the γδ resolvase system (32, 33). However, the products of the inversion systems, their site structures and their regulation are substantially different, as will be discussed below.

The products with DNA double-strand breaks were presumed to be derived from a recombination intermediate, and suggested a simple “cut-and-paste” mechanism of strand exchange (Fig. 4). Together with the specific, simple catenane or unknotted circle product topologies of resolvases and invertases, respectively, the data suggested that exchange of DNA ends by serine recombinases is a well-ordered process, taking place within a synaptic complex of two crossover sites and a resolvase tetramer, after double-strand cleavage of both sites (55, 56).

**Topological studies**

In the absence of protein structural information, most early analysis of the mechanism focused on the DNA reaction products. Analysis of the product topologies from supercoiled circular (plasmid) substrates was especially significant (57). Studies with the tyrosine recombinase λ Int (and later FLP and Cre) had revealed that a supercoiled two-site substrate could give products with a wide range of knot/catenane topologies. These results were interpreted by a “random collision” mechanism of synapsis; that is, the sites collide due to natural random motions of the supercoiled substrate molecule. Various numbers of coils/tangles are trapped as the two sites synapse. A subsequent simple strand exchange mechanism results in products with a range of topology (Fig. 7). Consistent with a random collision synapsis mechanism, these tyrosine recombinase systems did not distinguish between substrates with sites in different relative orientations: both “head-to-tail” (direct repeat) and “head-to-head” (inverted repeat) arrangements of sites were recombined equally well (57). The serine resolvases and invertases were clearly different. Resolvases yield almost exclusively simple catenane recombination products (Fig. 7a), and invertase recombination products are almost exclusively unknotted circles. Furthermore,
resolvases only recombine substrates with sites in direct repeat, and invertases only recombine sites in inverted repeat. These selectivities are very strong (for example, a >10⁴-fold rate difference for Tn3 resolvase), and persist even when the sites are separated by several kilobase pairs of DNA. Neither resolvases nor invertases recombine sites on separate supercoiled plasmids (57). The question therefore arose: how and why do these systems avoid the formation of random collision products? The phenomenon, which became known as “topological selectivity,” is discussed fully in Chapter [X], this volume.

To summarize very briefly, the catalytic activity of these serine recombinases is strictly regulated so as to take place only when an elaborate synaptic complex is properly formed. This structure includes the accessory DNA sequences and protein subunits, and involves intertwining of the sites (as shown for resolvase in Fig. 7). The twisting/writhing of the DNA involved in synaptic complex formation is energetically favorable only when the sites come together in a specific way, in a substrate with the correct relationship between the two sites. The regulatory properties of synaptic complexes are discussed further below.
Subunit rotation
Pioneering electron microscopy studies by Cozzarelli’s group revealed the precise topologies of a series of minor resolvase reaction products (Fig. 8a). These were proposed to be made by repeated rounds of strand exchange equivalent to half-turns of one pair of DNA ends relative to the other, in an intermediate with double-strand breaks in both crossover sites (58, 59, 60). The changes in DNA linkage that accompany the first round of the series (the standard resolution reaction) and its reverse reaction (catenane fusion) were determined and are consistent with this “simple rotation” mechanism (61). However, the simple hypothesis that the recombinase subunits attached to the half-sites rotate along with the DNA ends (“subunit rotation”; Fig. 8b) was difficult for many to accept, because of its radical biochemical implication; one half of the recombinase tetramer must rotate through 180° relative to the other half, but somehow disastrous dissociation of the two halves must be avoided. There is no biochemical precedent for this model; it was a “unicorn in the garden,” which would require extraordinarily rigorous testing.

A synapse with the recombining crossover sites on the outside of a recombinase tetramer core (“DNA-out”) was argued to be most consistent with the subunit rotation model (61). Later, alternative models that retained a fixed tetramer structure (and thus avoid the dissociation issue) were proposed. Some models placed the two recombining DNA double helices close to each other near the center of the tetramer (“DNA-in”). However, it was very difficult then to account for the observed topological changes after DNA strand exchange. Another model proposed that part of the DNA-out tetramer remains fixed, while the N-terminal parts of two subunits rotate with their attached DNA ends (38, 31).

A strange property of serine recombinase-mediated recombination, first discovered in the Gin DNA invertase system, led to strong experimental support for subunit rotation. If two recombining sites have different 2-bp sequences at their central “overlap,” the recombinants

![Figure 8 Subunit rotation mechanism of resolvase. (a) Topologies of first round and "iteration products" observed by Cozzarelli’s group (58, 59, 60). The upper part shows the products predicted by a rotation mechanism in a resolvase synapse with topology as shown in Fig. 7. The lower panels show the simplified topologies of these products. "Mismatched" substrates (see text) form only the nonrecombinant knot products, starting with the 4-noded knot. (b) Cartoon illustrating the proposed subunit rotation mechanism. DNA is represented as ribbons and recombinase subunits as ovals. The crystal structure of a proposed intermediate in subunit rotation is shown in Fig. 6b. doi:10.1128/microbiolspec.MDNA3-0046-2014.f8](ASMscience.org/MicrobiolSpectrum)
that would have mismatched base pairs are not formed (see above). Instead, a second round of strand exchange ensues, restoring the ends to a nonrecombinant configuration but leaving a record of the transaction as a change in the DNA topology (knotting) (62, 63, 64, 65). This behavior has since been shown to be general to many if not all serine recombinases, and per se suggests subunit rotation. In the case of resolvase, reaction of a “mismatched” substrate leads to a 4-noded knot product, consistent with a 360° rotation, and further double rounds of strand exchange give more complex knotted products (65; Fig. 8a). Further analysis of the products supported subunit rotation, but not alternative mechanisms (66, 67, 68, 69, 70). It was shown that knotting of a mismatched substrate proceeds with the DNA linkage changes predicted for subunit rotation, that the recombinase subunits move in concert with the DNA ends to which they are bound and that the knotting reaction of a mismatched substrate proceeds without any intermediate rejoining of the DNA ends that would allow a “reset” of the protein subunits, as would be necessary for all nonrotary mechanisms. The dimer interface seen in the early γδ resolvase crystal structures (37, 40) is quite rugged and apparently incompatible with subunit rotation. The structural breakthrough came from further crystallographic studies using “activated” γδ resolvase variants. Activated serine recombinase mutants (first identified in the Gin and Cin invertase systems) have lost their dependence on regulation by accessory factors (71, 72, 73, 74). Activated resolvase variants were shown to form synapses in vitro, comprising two crossover sites bridged by a resolvase tetramer (75, 76), and low-resolution structural studies confirmed that the sites were bound on the outside of the tetramer (53). The resolvase–DNA cocrytal structures of Li et al. and Kamtekar et al. (41, 42) revealed a synaptic intermediate with both crossover sites cleaved at their centers and resolvase subunits covalently attached to each 5′ end via the active-site serines, in line with the earlier biochemical experiments (Figs 4 and 6b). The conformation of the resolvase SR domains is dramatically different from that in the dimer structures; the tetramer has a remarkably flat, hydrophobic surface between the two “halves” that are predicted to rotate with respect to each other (Fig. 6b). It is proposed that a flat, greasy interface is maintained throughout rotation, and structure-based modeling has demonstrated that there would be no major energy barriers to this process (41).

Biochemical studies on invertases (69, 70, 77) and serine integrases (78, 79) have provided further support for a subunit rotation mechanism, and recent crystal structures of activated Sin resolvase and Gin invertase variants show recombinase tetramers with flat hydrophobic interfaces but different rotational relationships of the “rotating dimers” compared with the γδ resolvase structures (44, 49). It looks like all serine recombinases work this way.

The active site
Each of the four active sites in a serine recombinase tetramer has to perform two chemical steps during a round of recombination (a strand cleavage and then a ligation). We would like to understand the mechanism of catalysis, the contributions of individual subunits, and the choreography and reversibility of the reaction steps.

Alignments of serine recombinases reveal about a dozen well-conserved polar or charged residues that might contribute to catalysis at the active site, and studies involving mutagenesis and in vitro biochemical analysis have identified six key residues including the nucleophilic serine (3, 80, 81). Proposed roles for these residues include those typical of phosphoryl transfer enzymes: generation of a strong base to increase the nucleophilicity of the very weakly acidic serine hydroxyl (cleavage reaction) and deoxyribose 3′-hydroxyl (re-ligation reaction) groups, stabilization of the transition state geometry and/or charge, and provision of an acid for protonation of the leaving group during cleavage and re-ligation (82). Other active-site features that must be present include interactions to guide the incoming nucleotide bearing the 3′-OH to the active site for ligation and contacts that detect base pairing (or the lack of it) in the product overlap sequence.

REGULATION OF RECOMBINATION ACTIVITY

Introduction
The “programmed” DNA rearrangements promoted by natural site-specific recombinases typically involve sophisticated regulation to ensure that strand cleavages and subsequent events happen only at the right times and places. Serine recombinase-based systems adapted for resolution, inversion and integration have evolved distinct regulatory strategies, as will be discussed in the following sections.

All site-specific recombinases must have high fidelity for their target sites; off-target reactions are very likely to be deleterious. The C-terminal HTH domains of small serine recombinases recognize sequence motifs at the ends of the crossover site, but their sequence specificity
is limited (83) and DNA contacts by the SR (catalytic) domains make a substantial additional contribution to specificity (84). Even so, some variation of the crossover site sequence is tolerated (85, 86). The observed high site specificity of the complete systems is presumably due to tight dependence of catalytic activity on cooperative assembly of all the components including accessory factors.

Formation of the “catalytic module” by bringing together two recombinase dimer-bound crossover sites may be a key regulatory step for most systems. For example, the crossover-site DNA-bound γδ resolvase dimers in the crystals studied by Yang and Steitz (40) do not make a synaptic interaction despite their extremely high concentration, whereas “activated” mutants that are defective in regulation readily form tetramer-containing synaptic complexes (75, 87). This checkpoint apparently prevents wild-type resolvase catalysis until the full synaptic complex including the accessory sites and their bound subunits is correctly assembled (88) (see below).

**Regulation of resolvase recombination**

The serine resolvase systems are described in detail in Chapter [X], this volume. Recombination by resolvases takes place following formation of a specific synaptic complex involving intertwining of the res recombination site accessory sequences. As noted above, this complex forms only when the two res sites are in direct repeat in a negatively supercoiled DNA molecule. The synaptic complex also plays an important but as yet mysterious role in restricting strand exchange to a single half-turn, so that the first-round simple catenane resolution product is released and inert to further reaction (see Figs 7 and 8). The resolvase-bound accessory sites of Tn3/γδ res can pair and intertwine to form a “regulatory module” even in the absence of the crossover sites (Fig. 7b, “site I”); this property has been used to impose topological selectivity on normally nonselective recombinases (such as Cre) by putting their crossover sites in place of res site I (89, 90). The detailed molecular architecture of the Tn3/γδ regulatory module is still unclear. However, it has been shown that a specific protein interface between resolvase subunits plays a key role in coupling accessory site synopsis to synopsis of the crossover sites and activation of catalysis (43, 88, 91, 92). One hypothesis is that crossover-site synopsis and the dramatic concomitant protein conformational changes (see “Subunit rotation” above; Fig. 6b) are brought about simply by forcing the recombinase dimers into close proximity by their interactions with accessory subunits (“extreme mass action”). Alternatively, the interactions with accessory protein subunits might play an essential role in promoting the required conformational changes.

The arrangements of protein-binding sites within res-type recombination sites are quite diverse (93), as exemplified by Sin res, which contains just a single accessory binding site for Sin resolvase and a site for an “architectural” DNA-bending protein (HU/IHF). A crystal structure of Sin in a synaptic interaction with accessory DNA has led to a model for the complete synaptic complex formed by that system (43). The intertwining of the DNA and the contacts between Sin subunits bound at the crossover and accessory sites are strikingly similar to the corresponding features in current models of the Tn3/γδ resolvase complex. It seems plausible that many resolution systems adopt a similar strategy for activation of catalysis, despite significantly different regulatory module architectures.

**Regulation of invertase recombination**

The serine invertase systems are described in detail in Chapter [X], this volume. Most research on invertase mechanism has been on the Hin, Gin and Cin systems (33). Like resolvases, the invertases recombine at precise positions within dimer-binding crossover sites, but unlike the resolvases there are no adjacent accessory sequences. However, it was discovered that a sequence quite far from the crossover sites (called the enhancer, or sis) which binds an *Escherichia coli* protein FIS (factor for inversion stimulation) was essential for efficient recombination in each system (94, 95, 96, 97, 98). The lengths of DNA between the crossover sites and the enhancer could be varied without loss of activity (99). It is proposed that the invertase-bound crossover sites and FIS-bound enhancer come together to form a synaptic complex, as shown diagrammatically in Fig. 9. The molecular architecture of this complex is still incompletely understood, but structure-based models have been built following characterization of specific invertase–FIS interfaces, and it has been shown recently that Hin subunits make direct contacts with the enhancer DNA (100).

Like resolvases, invertases selectively recombine sites within the same supercoiled molecule, in a specific relative orientation, in this case inverted repeat. However, there is a notable difference. Resolvase has no activity at all on substrates with two res sites in an inverted repeat, whereas Hin (or Gin) invertase substrates with hix (or gix) sites in a direct repeat do not give recombinants but do undergo efficient double rounds of rotational strand exchange, giving knotted nonrecombinant products (62, 63, 64). It was concluded that the synaptic
complex (Fig. 9) is formed regardless of relative site orientation, but sites in a direct repeat are thereby misaligned in “antiparallel” such that strand exchange would give recombinants with mismatched base pairs; instead, double rounds of strand exchange give knotted nonrecombinant products (62) (see “Subunit rotation” above).

Regulation of recombination by large serine recombinases
Serine integrases and other large serine recombinases are the subject of Chapter [Y] of this book. The integrases do not display topological selectivity; they adopt a quite different strategy to ensure correct product formation. The phage (attP) and bacterial genome (attB) crossover sites are not identical, and recombination results in two further nonidentical sites attL and attR, flanking the integrated prophage DNA (Fig. 10). Unlike phage tyrosine integrases related to λ Int, the large serine integrases (and other large serine recombinases such as the Clostridium transposase TnpX) apparently do not have accessory DNA sequences (35, 101, 102). The details that follow derive from in vitro studies on the two best-characterized serine integrases, ϕC31 Int and Bxb1 Int (33). The att sites (∼40 bp) each bind an integrase dimer and have an asymmetric central 2-bp overlap sequence, which has been shown to be the sole determinant of site polarity in attP×attB recombination (103, 104). An attP site only recombines with an attB site, not with another attP, attL, or attR site; likewise, attB only recombines with attP. Synapsis is a key selective step; only the “correct” pair of sites (attP and attB) forms a stable complex (105, 106). The integrase alone does not recombine the “lysogen” sites attL and attR at all. However, a phage-encoded recombination directionality factor (RDF) protein binds to and transforms integrase so that it efficiently and specifically recombines attL×attR, whereas attP×attB recombination is inhibited (107, 108, 109).

FIGURE 9 Cartoon of proposed inversion synaptic complex. An invertase tetramer bridging the two crossover sites is shown as an orange oval, and contacts the enhancer DNA (brown) and a FIS dimer bound there (pink). The DNA in the complex is intertwined as shown. Supercoiled loops of DNA outside the complex are shown as dashed lines. doi:10.1128/microbiolspec.MDNA3-0046-2014.f9

FIGURE 10 Recombination by serine integrases. In nature, integrase promotes recombination between a crossover site, attP, in the circular bacteriophage DNA and a different crossover site, attB, in the host bacterial genome (indicated by a squiggly line). Integrase alone does not promote any reaction between the product sites attL and attR. However, the presence of a bacteriophage-encoded recombination directionality factor (RDF) protein alters the properties of integrase so that it preferentially promotes attL×attR recombination (red arrow). See text for further details. doi:10.1128/microbiolspec.MDNA3-0046-2014.f10
PROTEINS RELATED TO SERINE RECOMBINASES

There are no other families of proteins with known functions that can be unambiguously shown to be related to the serine recombinases. The SR fold has similarity to the structures of a group of 5′-to-3′ exonucleases and to the TOPRIM domain of type IA and type II topoisomerases (110, 111), but the active site residues (and thus presumably the catalytic mechanisms) are quite different, so it is not clear that there is any relationship by descent.

SERINE RECOMBINASES IN BIOTECHNOLOGY AND SYNTHETIC BIOLOGY

Many recombinase systems have been investigated as possible tools for mediation of precise, inducible DNA rearrangements in the fields of experimental genetics, biotechnology and gene therapy (112). However, with one notable exception (ϕC31 integrase), the utilization of serine recombinases has been relatively limited. ϕC31 integrase has been adopted for targeted transgene integration in a number of organisms including humans, and it is now in widespread use in experimental research, notably in the Drosophila field (112, 113).

Serine recombinases are currently being exploited in the field of synthetic biology, for construction of artificial genetic switches and circuits. Recent studies have shown how all the standard Boolean logic operations can be implemented on gene expression in E. coli by the combined action of two orthogonal serine integrases (ϕC31 Int and Bxb1 Int) (114). Other applications are in the assembly and manipulation of metabolic pathway genetic components (113, 115, 116).

ENGINEERING SERINE RECOMBINASES

Site-specific recombination has obvious potential as a tool for “genomic surgery” in organisms of interest to humans, but to realize this potential it will be necessary to engineer recombinases so that they recognize and act on sequences occurring in these organisms. Natural recombinases often require long complex sites, accessory factors, and DNA supercoiling, making this task seem quite daunting. However, the characterization of activated variants of small serine recombinases, which have simplified substrate requirements, has opened up engineering possibilities (71, 74, 117). The small serine recombinases are modular proteins with spatially distinct SR and HTH (DNA-binding) domains (see Fig. 6). Some sequence specificity changes were made by mutating the HTH domain or replacing it with a domain from a related serine recombinase (118, 119, 120). However, much more dramatic retargeting was achieved by linking the SR domain to a zinc-finger DNA-binding domain (121). These “zinc finger recombinases” can be adapted to use a wide range of new “crossover sites” including natural genomic sequences, by engineering zinc-finger-domain specificity, reducing or altering the residual sequence specificity of the SR domain, or using SR domains from different recombinases (86, 122, 123, 124, 125). Recently, transcription activator-like effector (TALE) DNA-binding domains have been used instead of zinc-finger domains to retarget SR domain activity; the modularity of TALE domains and thus the ease of creating new specificities may greatly enhance the applicability of these “designer recombinases” (126).

ACKNOWLEDGMENTS

I apologize to readers that, in this overview chapter on the serine recombinases, it has been painfully necessary for me to cover important aspects of the subject only sketchily or even not at all, and to curtail the reference list. However, much more detail on serine recombinases can be found in the three chapters in this book by Phoebe Rice, Maggie Smith and Reid Johnson. I also take this opportunity to acknowledge the many contributions of colleagues past and present to the advancement of this field, and to look forward to many more exciting developments and insights in the future.

REFERENCES


Stark


