Site-specific recombination caught in the act

David MJ Lilley

The recently determined crystal structure of Cre recombinase from bacteriophage P1, bound to its target DNA, sheds considerable light on the mechanism underlying a major type of site-specific DNA recombination, and shows how strand exchange can be achieved without reorganisation of the protein complex.

Address: CRC Nucleic Acid Structure Research Group, Department of Biochemistry, The University of Dundee, Dundee DD1 4HN, UK.

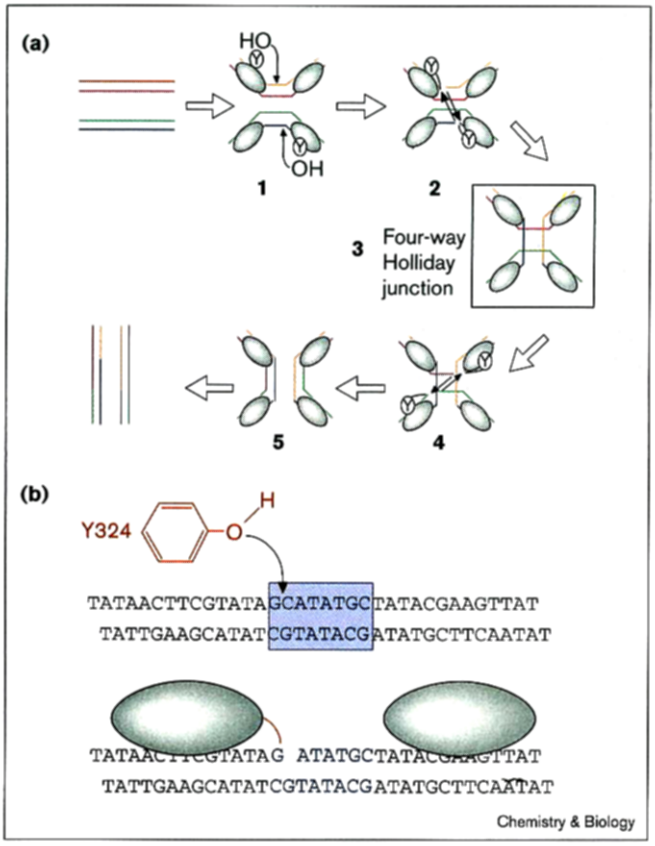
Correspondence: David MJ Lilley
E-mail: dmjlilley@bad.dundee.ac.uk

Chemistry & Biology October 1997, 4:717–720
http://biomednet.com/elecref/1074552100400717

Site-specific recombination is the breakage and rejoining of DNA in new combinations at a specific target site. It can be used to join two DNA circles together or to resolve them, or to invert a particular segment of DNA [1]. Recombination is essential for genetic variation and it also allows cells to repair and recover sequences that would otherwise be lost when DNA is damaged. Generally, site-specific recombination has a regulatory role — it regulates gene expression, particularly in prokaryotes, by moving DNA sections, resulting in the expression of different genes. More recently it has also been exploited experimentally, for example in the generation of gene 'knock-out' mice. Recently, our view of the mechanism of site-specific recombination events has been changing. This is particularly true for the family of systems, exemplified by that used for the integration and excision of bacteriophage λ, that includes the Cre–*lox* [2,3] and *xer* [4] systems. A new crystal structure of the phage P1 recombinase Cre bound to its target DNA [5] allows a new mechanistic insight into these processes and strongly supports the new thinking in this area.

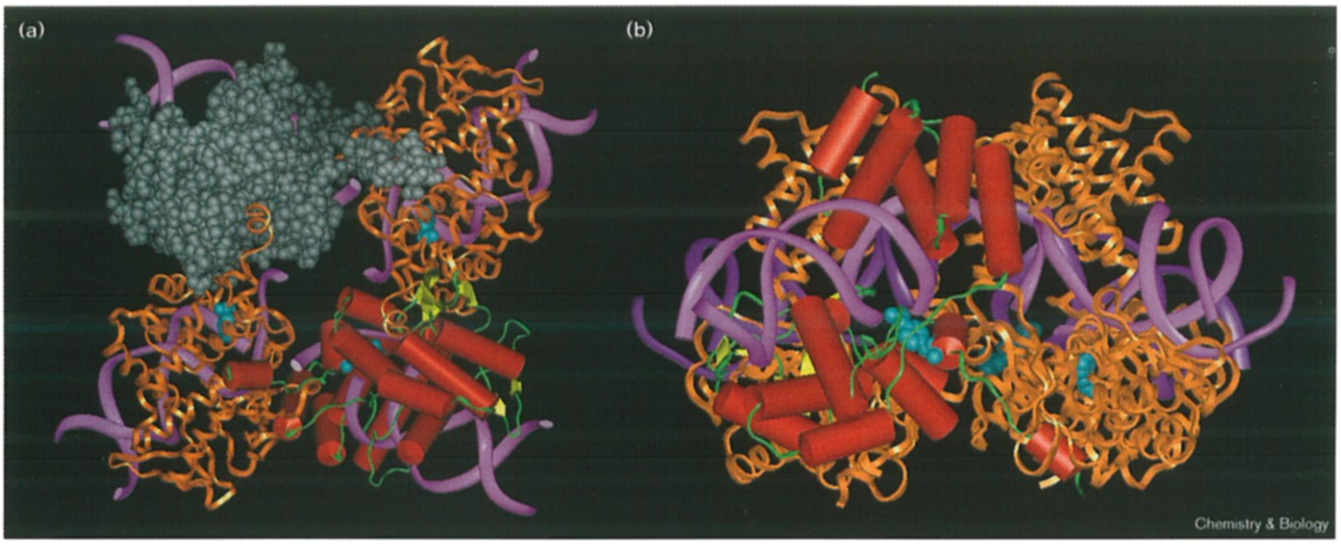
In outline, the mechanism of recombination of the λ-integrase type has been known for some time. Recombination is initiated by cleavage at a unique site via a covalent phosphotyrosine linkage with the recombinase protein [6]. The process involves two sets of strand exchanges; a four-way 'Holliday junction' is the intermediate between the two events (Figure 1). Each target site consists of an inverted repeat that has a central region over which the strand exchanges are made [7]. As the two sets of exchanges occur 6–8 base pairs away from each other, it was originally proposed that the four-way junction would form at the first site, and then it would have to branch migrate to the second position before the junction could be resolved to

Figure 1



Site-specific recombination in the Cre–*lox* system. **(a)** An outline of recombination in the λ-integrase family. The four strands of the two recombining DNA duplexes are different colours. The reaction proceeds in two distinct stages, forming a four-way (Holliday) junction as the central intermediate. Each duplex is bound by two Cre recombinase molecules (grey) at the target *loxP* site. The bound sites are shown bent in the light of the new crystallographic data. The first phase (**1–3**) involves cleavage by the active site tyrosine (Y) hydroxyl group of one subunit on each duplex, forming a covalent linkage to the DNA (**1**). The 3′-phosphotyrosine linkages are then attacked by the 5′ hydroxyl group of the other site **(2)** generating the four-way junction (**3**). In the second phase (**3–5**), the tyrosines of the two remaining Cre monomers cleave the opposite strands of each *loxP* site (**4**), and finally attack by the 5′ hydroxyl groups leads to formation of two recombinant duplex species (**5**). **(b)** The symmetrised *lox* site used by Van Duyne and coworkers [5]. The 34 base-pair site consists of a central strand-exchange region (blue) flanked by 13 base-pair inverted repeat sequences. Two molecules of Cre bind at the *lox* site. Each of the cleavage reactions occurs between GpC sequences, so the first and second stage cleavage sites are separated by six base pairs. In the suicide substrates used for crystallisation, this region contained a nick in the backbone at the sequence GC/ATATGC. The active site tyrosine in the first stage attacks the GpC phosphate to generate the covalent intermediate but the reaction can proceed no further because the cytosine diffuses away and thus the nucleophile required for the second stage is lost. Thus the species is trapped at stage **1** of (a).

Figure 2



The structure of the Cre–*lox* complex, drawn from coordinates kindly supplied by Dr G. Van Duyne. **(a)** A view from the major-groove side of the DNA, looking down on the carboxy-terminal domains of the Cre recombinase. The strands of the DNA are represented by the purple ribbons. One of the four Cre subunits is shown as a secondary structure cartoon (red, green and yellow, bottom right), one in space filling representation (grey, top left) and two as gold ribbons. The active site Tyr324 residues are blue. **(b)** A side view, edge-on to the DNA. The amino-terminal domains are above and the carboxy-terminal domains are below the DNA in this view. These images were prepared using INSIGHTII (MSI-Biosym) with the assistance of Dr D.G. Norman.

generate recombinant products. Simone Nunes-Düby and Art Landy suggested a new model [8] that involved 'peeling back' the strands after cleavage, thus locating the junction in the middle of the overlap region between cleavage sites. The junction was then proposed to undergo an exchange of helical stacking partners such that the second set of strand exchanges would be carried out in the same stereochemical environment as the first set.

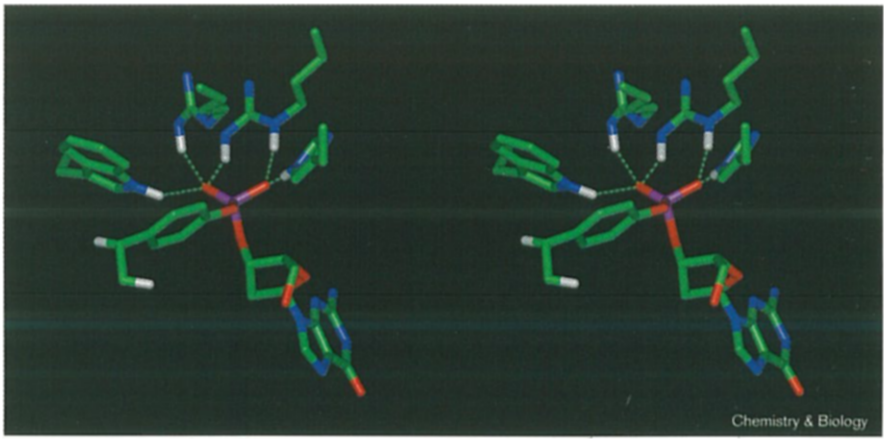
The new crystal structure by Van Duyne and coworkers [5] provides a deeper insight into the mechanism of recombination and support for the Nunes-Düby/Landy model. Guo *et al.* [5] crystallised the 38 kDa Cre protein bound to a 34 base-pair synthetically symmetrised version of its *loxP* target site (Figure 1b) and solved the structure to a resolution of 2.4 Å. Using a 'suicide substrate' [9] which cannot be successfully recombined, they trapped the reaction at the covalent intermediate stage of the first strand exchange. This was achieved by the simple expedient of placing a nick in the phosphodiester backbone so that when the site-specific transesterification reaction occurred, the 3′ nucleotide diffused away. The resulting strand cannot provide the 5′ hydroxyl group that normally serves as the nucleophile required to attack the other phosphotyrosine linkage to generate the recombinant strand, so the whole process is halted at the first step (**1**, Figure 1a).

The most striking feature of the new structure is the overall architecture of the complex (Figure 2). Each 34 base-pair *lox* site is bound by two Cre proteins, one of which is covalently linked to the DNA by Tyr324. The proteins are bound chiefly to the outer sections of the *lox* sites, and the DNA is bent by over 90° around the central strand-exchange region. Two *lox* sites are associated in the crystal, generating a pseudo-fourfold symmetrical arrangement that closely resembles a four-way junction in its open-square conformation (i.e. the form adopted by a junction in the absence of added metal ions [10]). Indeed, Van Duyne and coworkers [5] have modelled the formation of such a junction by relatively minor reconnections of the structure of the DNA that is found in the crystal. Thus, it can readily be seen from the structure how all the required chemical reactions of strand breakage and rejoining may occur within the central strand-exchange cavity of the synapsed DNA, and an open-square four-way junction is the central intermediate. No alteration in the quaternary structure is required as the reaction proceeds. Closer examination of the DNA structure reveals a local unpairing within the strand-exchange region of exactly the kind that would be required in the Nunes-Düby/Landy model, obviating any requirement for branch migration in the process.

The Cre protein comprises two distinct domains, both of which are highly α-helical. The amino-terminal 130 amino acids comprise five α helices, three of which form an antiparallel helical bundle with the remaining two lying sideways on to these three. Two of the helices sit in the major groove, forming direct contacts with the bases. The

Figure 3

Parallel-eye stereo view of the active site of the Cre recombinase, showing the 3′-phosphotyrosine linkage. Note the extensive hydrogen bonding (dotted lines) of the phosphate by tryptophan, two arginine and histidine sidechains.



remaining carboxy-terminal part of the protein is connected by an extended linker polypeptide. It is constructed primarily of nine α helices, with a small section of β sheet packed on one side. This domain makes contacts with almost the whole face of the DNA half-site, having extensive contacts in both grooves as well as the phosphodiester backbone. Only one direct base contact is seen in the major groove (a familiar bidentate arginine–guanine interaction) but a number of base contacts are made in the minor groove. The overall effect is that the protein totally envelopes the DNA, forming a tight clamp around it. Around 5000 Å^2 of the DNA–protein interface is buried for each Cre–half-site interaction, and extensive electrostatic contacts with the backbone help to lock the entire complex into place.

The amino- and carboxy-terminal domains of the Cre protein sit on opposite faces of the junction-like DNA structure. The DNA is not quite planar; the region that becomes the major-groove side of the eventual junction is slightly concave. The amino-terminal domain is located on the minor-groove side, whereas the carboxy-terminal domain is bound on the major-groove side. The active-site residues are all located on the carboxy-terminal domain, and thus all the chemical reactions are performed on the major-groove face. The key conserved residues [11] are three basic sidechains and a tryptophan, as well as the nucleophilic tyrosine [12]. Unlike the corresponding Flp recombinase of yeast [13], all the residues are provided by a single subunit. There are two distinct types of active site seen in the crystal structure, corresponding to the two Cre subunits bound to each duplex. One is poised, ready to undergo the reaction, whereas the other is trapped in the covalent 3′-phosphotyrosine complex, awaiting attack by a 5′-OH from the strand in the opposing duplex. The scissile phosphate sits in a positively charged pocket (termed a 'proton cradle' by Van Duyne and coworkers [5]), making a number of hydrogen bonds between the sidechains and the non-bridging oxygen atoms (Figure 3).

The overall pseudo fourfold symmetrical character of the synapse is set up by protein–protein interactions. There are extensive contacts between the subunits in which both amino- and carboxy-terminal domains play a part, giving a total buried surface area of 13,800 Å^2. The carboxy-terminal α helix of each subunit is donated to the adjacent subunit in a cyclic manner. The overall effect is strongly reminiscent of another protein, RuvA, the substrate of which is also a four-way DNA junction. The role of RuvA is to bind to a junction whereupon it (in conjunction with RuvB hexamers) promotes branch migration [14]. Like Cre, RuvA forms a tetramer that undergoes a similar cyclic exchange of domains [15]. The structure seems to be perfectly designed as a template for an open-square junction and it seems logical that it functions by opening the junction in this way in order to facilitate the exchange of base pairing. By contrast, in the stacked conformation of the junction that exists in free solution containing magnesium ions this process would expected to be slow. So, by opening the structure a considerable rate enhancement can be expected and this has been demonstrated experimentally for the free junction [16]. Such an opening of junctions appears to be quite general, and all the junction-resolving enzymes studied to date open the structure in one way or another [17]. Perhaps the stacked form of the junction is just too inaccessible to allow the kinds of manipulation that are required for recombination. The structure of the Cre complex shows how, by holding the DNA in the form of an open junction, all the required chemical reactions can be performed within the central exchange cavity while requiring no movements in the overall geometry of the synapse.

The structure of the Cre–*lox* complex tells us a great deal about this fascinating recombination system, but also more

generally about the way to manipulate DNA molecules in rearrangement reactions. Further structures of recombinase proteins, such as the XerD protein [18], are emerging and we can expect rapid progress in this area.

References

1. Stark, W.M., Boocock, M.R. and Sherratt, D.J. (1993) Catalysis by site-specific recombinases. *Trends Genet.* **8**, 432-439.
2. Hoess, R., Abremski, K. and Sternberg, N. (1984) The nature of the interaction of the P1 recombinase Cre with the recombining site *loxP*. *Cold Spring Harb. Symp. Quant. Biol.* **49**, 761-768.
3. Hoess, R. and Abremski, K. (1984) Interaction of the bacteriophage P1 recombinase Cre with the recombining site *loxP*. *Proc. Natl Acad. Sci. USA* **81**, 1026-1029.
4. Blakely, G., Sykora, P., Szatmari, G. and Sherratt, D.J. (1990) Recombination at ColE1 *cer* requires the *Escherichia coli xerC* gene product, a member of the Lambda integrase family of site-specific recombinases. *J. Bacteriol.* **172**, 6973-6980.
5. Guo, F., Gopaul, D.N. and Van Duyne, G.D. (1997) Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* **389**, 40-46.
6. Pargellis, C.A., Nunes-Düby, S.E., de-Vargas, L.M. and Landy, A. (1988) Suicide recombination substrates yield covalent lambda integrase-DNA complexes and lead to identification of the active site tyrosine. *J. Biol. Chem.* **263**, 7678-7685.
7. Abremski, K., Wierzbicki, A., Frommer, B. and Hoess, R. (1986) Bacteriophage P1 Cre-*loxP* site-specific recombination. Site-specific DNA topoisomerase activity of the Cre recombination protein. *J. Biol. Chem.* **261**, 391-396.
8. Nunes-Düby, S.E., Azaro, M.A. and Landy, A. (1995) Swapping DNA strands and sensing homology without branch migration in λ site-specific recombination. *Curr. Biol.* **5**, 139-148.
9. Nunes-Düby, S.E., Matsomoto, L. and Landy, A. (1987) Site-specific recombination intermediates trapped with suicide substrates. *Cell* **50**, 779-788.
10. Duckett, D.R., Murchie, A.I.H., Diekmann, S., von Kitzing, E., Kemper, B. and Lilley, D.M.J. (1988) The structure of the Holliday junction and its resolution. *Cell* **55**, 79-89.
11. Argos, P., Landy, A., Abremski, K., Egan, J.B., Haggard-Ljungquist, E., Hoess, R.H., Kahn, M.L., Kalionis, B., Narayana, S.V. and Pierson, L.S. (1986) The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO. J.* **5**, 433-440.
12. Abremski, K.E. and Hoess, R.H. (1992) Evidence for a second conserved arginine residue in the integrase family of recombination proteins. *Protein Eng.* **5**, 87-91.
13. Chen, J.W., Lee, J. and Jayaram, M. (1992) DNA cleavage in trans by the active site tyrosine during Flp recombination: switching protein partners before exchanging strands. *Cell* **69**, 647-658.
14. Muller, B., Tsaneva, I.R. and West, S.C. (1993) Branch migration of Holliday junctions promoted by the *Escherichia coli* RuvA and RuvB proteins.1. Comparison of RuvAB-mediated and RuvB-mediated reactions. *J. Biol. Chem.* **268**, 17179-17184.
15. Rafferty, J.B., Sedelnikova, S.E., Hargreaves, D., Artymiuk, P.J., Baker, P.J., Sharples, G.J., Mahdi, A.A., Lloyd, R.G. and Rice, D.W. (1996) Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction. *Science* **274**, 415-421.
16. Panyutin, I.G., Biswas, I. and Hsieh, P. (1995) A pivotal role for the structure of the Holliday junction in DNA branch migration. *EMBO J.* **14**, 1819-1826.
17. White, M.F., Giraud-Panis, M.-J.E., Pöhler, J.R.G. and Lilley, D.M.J. (1997) Recognition and manipulation of branched DNA structure by junction-resolving enzymes. *J. Molec. Biol.* **269**, 647-664.
18. Subramanya, H.S., Arciszewska, L.K., Baker, R.A., Bird, L.E., Sherratt, D.J. and Wigley, D.B. (1997) Crystal structure of the site-specific recombinase, XerD. *EMBO J.* **16**, 5178-5187.