A possible way that φC31 integrase regulates the recombination direction

T.A. Liu[1], Y.X. Fang[1], H. Yin[1,2] and Z.Z. Jing[1,2]

[1]State Key Laboratory of Veterinary Etiological Biology,
Key Laboratory of Veterinary Public Health of Agricultural Ministry,
Lanzhou Veterinary Research Institute,
Chinese Academy of Agricultural Sciences, Lanzhou, China
[2]Jiangsu Co-Innovation Center for Prevention and Control of Important
Animal Infectious Dieases and Zoonoses, Yangzhou, China

Corresponding author: Z.Z. Jing
E-mail: zhizhongj@163.com

ABSTRACT. φC31 integrase encoded by *Streptomyces* phage can mediate site-specific recombination between phage and host genomes. The recombination direction is generally considered to be accurately regulated, but the regulatory mechanisms involved are still unclear. Recently, some hyperactive mutants of φC31 integrase that can bypass the regulatory steps have been isolated and extensively studied. A putative coiled-coil region is found to play a critical role in controlling recombination direction. Further analysis led us to the speculation that at least two regions in the N-terminal domain of φC31 integrase are involved in the tetrameric interfaces and that the putative coiled coil interacts with one of the regions to regulate the recombination direction.

Key words: φC31 integrases; Hyperactive mutants; Synapsis; Directional control

INTRODUCTION

To produce offspring, many temperate bacteriophages integrate their genomes into the host chromosomes by encoding an integrase (Campbell, 2006). φC31 integrase, which is encoded by *Streptomyces* phage, is a member of the serine recombinase family, which contains serine resolvases, invertases, transposases, and integrases (Smith and Thorpe, 1998). A relatively conserved catalytic domain is shared among the serine recombinases (Gorth and Calos, 2004). The full-length φC31 integrase consists of 605 residues and can be divided into two parts: the C terminal domain (CTD), which is responsible for DNA binding and controls the recombination direction, and the N terminal domain (NTD), which cleaves and rejoins the DNA substrate (Grindley et al., 2006; McEwan et al., 2009).

φC31 integrase has now been applied to all kinds of genomic manipulations due to its unidirectivity and high efficiency in catalyzing site-specific recombination (Brown et al., 2011; Liu et al., 2013). A number of mutations have been introduced to further define the properties of this enzyme and to unravel its reaction mechanism (Liesner et al., 2006; Rowley et al., 2008; Keravala et al., 2009; McEwan et al., 2009; Liu et al., 2010). As a result, several φC31 integrase mutants made significant improvements in recombination efficiency and specificity (Liesner et al., 2006; Keravala et al., 2009). Moreover, some mutants gain the extra ability of excision (the reverse reaction of integration) (Rowley et al., 2008; McEwan et al., 2009), which normally takes place in the presence of a recombination directionality factor, gp3 protein (Khaleel et al., 2011), which provides us with a clue to explore the possible mechanism of φC31 integrase controlling the reaction direction. Here, we focus on the hyperactive mutants that are both active in integration and excision and then present a possible way in which φC31 integrase regulates the recombination direction.

Hyperactive φC31 integrase mutants

Previous studies have proven that the recombination mediated by φC31 integrase is mainly regulated in the pre-synapsis step (Rowley et al., 2008; Smith et al., 2004, 2010). During recombination, φC31 integrases is first bound to attP or attB as a dimer, with each monomer occupying half of the att site in a head to head manner. The two dimers bound to attB and attP then pair to bring the substrates together to form a synapsis (Grindley et al., 2006). This process is commonly prevented when φC31 integrase is bound to other pairs of att sites (Smith et al., 2004), which, however, can be changed by some mutations (Rowley et al., 2008).

Rowley et al. (2008) isolated several mutated integrases that gain the ability to recombine attL and attR without affecting the attB×attP recombination. Interestingly, these mutants are all located in a putative coiled-coil region of CTD (Rowley et al., 2008), rather than NTD like in serine resolvases (Burke et al., 2004; Li et al., 2005; Keenholtz et al., 2011). The putative coiled-coil structure contains two regions, E449-D477 and R489-A519 (Rowley et al., 2008), which are termed C1 and C2, respectively, in this article. Several mutations that cause hyperactivity are all located in C1. The most active position is E449, which is at the start of C1 and occupies the "g" position of the repeating heptad. Substitutions at this position with different amino acids cause activating effects to various extents (Rowley et al., 2008). The putative coiled coil often mediates cooperativity between two integrases in the att site-binding process except attP, possibly due to a longer span in the attP site (McEwan et al., 2009). This cooperation may directly contribute to the conformational changes of both the

DNA substrates and the next binding integrase, which can facilitate DNA binding (Rowley et al., 2008; McEwan et al., 2009).

The putative coiled coil not only has a role in DNA binding but also acts in synapsis. It has been proven that the CTD synapsis is mediated by the putative coiled coil (McEwan et al., 2009). When bound to attB and attP but not to other pairs of att sites, the CTD of φC31 integrase can form an unstable synapsis, which implies that the CTD has a role in selecting which pair of att sites to be recombined. Given that E449K does not affect the oligomerization of the putative coiled coil and that integrase[E449K] can recombine both attB×attP and attL×attR (McEwan et al., 2009; Smith et al., 2010), we may expect that CTD[E449K] can synapse both of them. Surprisingly, the result is just the contrary: neither attB×attP nor attL×attR can be synapsed by CTD[E449K], which, on the other hand, indicates that the active synapsis is mainly composed of the NTDs (McEwan et al., 2009). Therefore, McEwan et al. (2009) proposed that the formation of a productive synapse by integrase, mainly the NTD of integrase, is inhibited by the putative coiled coil and that E449K mutation can unblock this inhibitory effect. But how the inhibition is achieved is still unknown. Moreover, some higher order complexes resembling the trimer or tetramer of CTD[E449K] are found in the attP binding process, indicating that E449K may reset the putative coiled coil so that it can interact with some unknown domains or motifs in CTD[E449K] (McEwan et al., 2009). This further supports the notion that E449K may change the position of the putative coiled coil and, therefore, interrupt the active synapsis of CTD[E449K] (McEwan et al., 2009; Smith et al., 2010).

Aside from φC31 integrase, some serine resolvases can also be activated by mutations (Burke et al., 2004; Li et al., 2005; Keenholtz et al., 2011). Their crystal structures have been well determined (Li et al., 2005; Keenholtz et al., 2011), pointing to the same activating principle that most of the mutations tend to stabilize the resolvase tetramer or destabilize its dimeric form (Li et al., 2005; Keenholtz et al., 2011). This is consistent with the observation that some φC31 integrase variants have more stable dimeric forms and display a decreased recombination activity (McEwan et al., 2011). Since the dimeric and tetrameric interfaces of serine recombinases are mainly composed of NTDs (Burke et al., 2004; Li et al., 2005; McEwan et al., 2009; Keenholtz et al., 2011), mutations in this domain can easily change the equilibrium between recombinase dimer and tetramer as in serine resolvases. Thus, it appears that the mutations in φC31 integrase CTD affect integrase synapsis in a rather intricate manner. One reasonable explanation is that CTD influences NTD synapsis (McEwan et al., 2009).

C1 possibly interacts with NTD to inhibit synapsis

Based on above analysis, it can be concluded that regulation of synapsis is mainly accomplished through the putative coiled coil. To further reveal the regulatory process, we compared wild-type φC31 integrase with φC31 integrase mutant E449K in the attB×attP synapse. Naturally, it is concluded that CTD synapsis mediated by the putative coiled coil and E449K mutation are equal in removing the inhibitory effect on NTD synapsis. Since both CTD synapsis and E449K mutation (Rowley et al., 2008) are able to change the position of the putative coiled coil, it can be inferred that the inhibition is mainly achieved through the interaction between the putative coiled coil and NTD (McEwan et al., 2009) and that blocking (positional change) this interaction can disrupt the inhibition (Figure 1). The interaction may be direct binding, indirect allosteric regulation or other forms. But there is one question left: why does CTD synapsis only occur with attB×attP? This question can be partly answered

by the complementary interactions where integrases bound to a P-type arm can only recognize integrases bound to a B-type arm in synapsis (Rowley et al., 2008). However, attL×attR, attL×attL and attR×attR, which comply with the complementary interaction principle, cannot form any synapsis by CTD either. For the putative coiled coil mediates CTD synapsis (McEwan et al., 2009), the difference of C1 and C2 in dimers bound to these four kinds of att sites may be the remaining answer to that question. C1 and C2 in a dimer bound to attP are uniquely unpaired (McEwan et al., 2009). The CTD synapsis may only happen in the presence of these free C1 and C2 (Figure 1).

Since E449K may alter the position of the putative coiled coil (McEwan et al., 2009), the putative coiled coil cannot interact with each other in synapsis (Figure 1). Integrase mutants L460P and Y475H gain the ability to recombine attL×attR at a very low level (McEwan et al., 2009), probably because of the impaired interaction between the putative coiled coil and NTD, coupled with the reduced affinity or cooperativity in binding to DNA substrate (McEwan et al., 2009). Thus, L460P and Y475H turn out to be defective in both integration and excision. Remarkably, there is still a D447K mutation that confers inactivity to integrase (Rowley et al., 2008). According to the aforementioned analyses, this mutation possibly enhances the interaction between the putative coiled coil and NTD, which may further prevent the putative coiled coil from being taken away by the CTD synapsis. Therefore, mutant D447K is inactive in integration. Since D447 is located in C1 (Rowley et al., 2008), it is likely that C1 interacts with NTD (Figure 1).
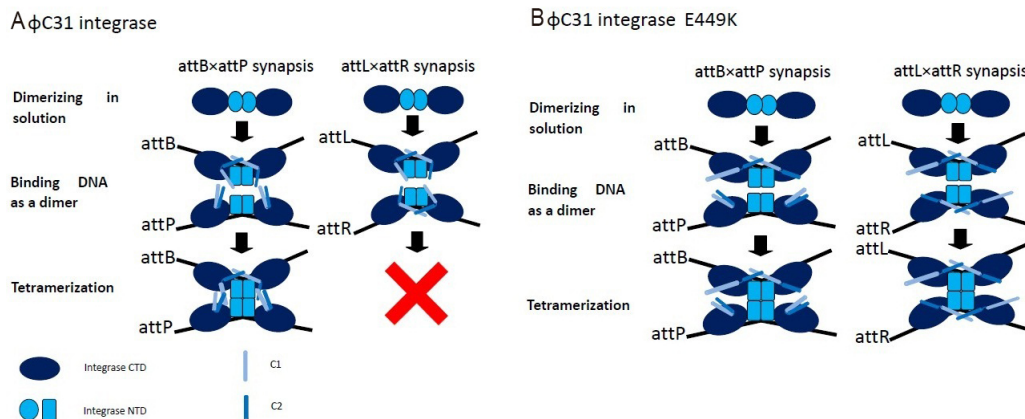

Figure 1. Model for synapsis by φC31 integrase. First two φC31 integrase monomers dimerize in solution, during which C1 and C2 are buried in the CTD (McEwan et al., 2009) and the conformation of the NTD is disordered. When a dimer binds DNA substrate, C1 and C2 become exposed, beginning to mediate the cooperativity between the dimer in attB, attL and attR binding process. Simultaneously the NTD gains the potential to form tetrameric interface, which is indicated by the transition from ellipse to rectangle in the figure. Finally the condition of C1 and C2 determines the tetramerization. **A.** Modeled synapsis by wild-type φC31 integrase. When dimers bind attB, attL and attR, the unbound C1 may make a contact with αA' of the NTD, which can inhibit the tetramerization of the NTD. The inhibition can be removed when the free C1 and C2 from a dimer bound to attP interact with the inhibitory C1. **B.** Modeled synapsis by φC31 integrase E449K. The mutation E449K may change the position of C1 and C2 (Rowley et al., 2008), leaving C1 beyond the reach of αA'. Thus, the NTD synapse is not blocked. And the misplaced C1 and C2 cannot interact with each other in synapsis either. The positional change of C1 and C2 is indicated by their different locations between Figure A and B. The C1 or C2 from integrase E449K bound to attP interacting with another integrase CTD is not shown in this figure.

There may be more than one domain involved in the tetrameric interface of NTD

Enlightened by the actively mutated resolvase, we all expected that the mutations in the region corresponding to the active loop preceding E helix of γδ resolvase would result in a similar activating effect in φC31 integrase. But several mutations at V129 that align with M103 of γδ resolvase all give no detectable attL×attR recombination. Moreover, the level of attB×attP recombination is also greatly lowered by most of the substitutions at this site (Rowley and Smith, 2008). Two reasons may account for this phenomenon: the conformational changes caused by all the mutations (V129L, V129M, V129F, V129G, V129A, and V129E) are not competent in forming a new tetrameric interface; unlike resolvases, the structure corresponding to the E helix of resolvases is not the only one responsible for forming a tetrameric interface in φC31 integrase. The mutations in V129 have no effect on other responsible domains. As V129G and V129M mutants act efficiently as normal integrase in synapsis but appear to be defective in attB×attP recombination, V129 may also play a critical role in activating the post-synapsis cleavage besides synapsis (Rowley and Smith, 2008), which is distinct from resolvases (Burke et al., 2004; Li et al., 2005; Keenholtz et al., 2011). Therefore, the latter explanation appears to be more reasonable.

Since the tertiary structure of φC31 integrase is still not available, it appears to be rather worthwhile to select a related recombinase whose structure has been illustrated for reference. TP901 integrase also belonging to the large serine integrase subfamily has similar components as φC31 integrase: a relatively conserved NTD for catalysis and a large unknown CTD (Grindley et al., 2006; Yuan et al., 2008). Moreover, the tetrameric structure of the NTD of TP901 integrase has been demonstrated (Figure 2) (Yuan et al., 2008). A sequence alignment is made between the NTDs of φC31 integrase and TP901 integrase (Figure 3). Even though they only share 33.5% similarity, the predicted secondary structures of φC31 integrase NTD resemble the actual ones of TP901 integrase NTD (Figure 3) (Cole et al., 2008; Yuan et al., 2008), indicating that their key structure elements are similar. From the tetrameric structure of TP901 integrase NTD, we can observe that its tetrameric interface contains three regions: αA, β5 and αE (Figure 2) (Yuan et al., 2008). This is consistent with our hypothesis that in φC31 integrase, there are likewise at least two domains involved in the tetrameric interface. Therefore, the putative coiled coil is very likely to exert an influence, directly or indirectly, on the regions responsible for tetramerization, thereby preventing the formation of the tetrameric interface (Figure 1).

In addition, there is another phenomenon that is worth discussing. The NTD of TP901 integrase crystallizes as a tetramer, but it exists as a dimer in solution. The dimeric interface of TP901 integrases in solution is composed of two adjacent αE helices (Figure 2). Possibly the tetrameric structure is favored during crystallization due to the high protein concentration (Yuan et al., 2008). Therefore, we speculate that the potential tetrameric interface is normally formed after the dimer is bound to DNA substrate (Figure 1). This can further explain why φC31 integrases behave as a dimer in solution since the putative coiled coils are buried and cannot inhibit NTD tetramerization (McEwan et al., 2009).

C1 may bind αA' directly to modulate the tetramerization of NTD

Since there is often interaction between coiled coils (Burkhard et al., 2001), we analyzed the NTD of φC31 integrase using the COILS program (Lupas et al., 1991), trying to find another coiled-coil structure. The output from the COILS program displayed a region with a >90% possibility to assume a coiled-coil structure in 14 scanning window, and this region just

overlapped αA', which aligns to αA in TP901 integrase (Figure 3).

Although the scanning window is so short that the prediction may be a false positive, it is consistent with the transient interaction between these two putative coiled coils. The contacts between these coiled coils are not stable and they can separate momentarily according to the stage of recombination, so the coiled coil does not seem to be too long. Given this consistency, αA' is the most possible structure in NTD to interact with C1. Therefore, we moved one step further to speculate that C1 binds αA' directly to enable NTD to tetramerize (Figure 1).
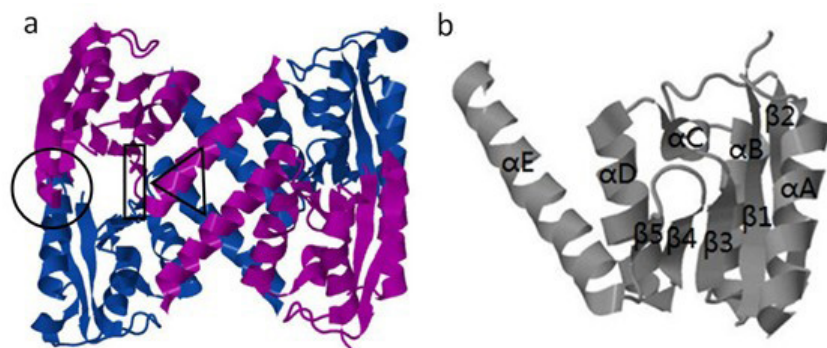

Figure 2. Structure of the catalytic domain of TP901 integrase. **a.** Tetrameric structure of the tetramer. Circle, rectangle and triangle represent tetrameric interface αA, β5 and αE, respectively. **b.** Tertiary structure of TP901 integrase NTD. The names of the secondary structures are labeled on the corresponding regions in Figure b. The figure is made from PDB ID: 3BVP.

Figure 3. Sequence alignment of catalytic domains between φC31 integrase and TP901 integrase. The alignment is made by EMBOSS Needle. The vertical line means identical amino acids. Dot indicates that the two amino acids have similar side chains with two dots meaning similar polarity. The diagonal arrows and columns above the sequence of φC31 integrase indicate the predicted β-pleated sheet and α-helix by Jpred3 (Cole et al., 2008). The dark black arrows and columns below TP901 sequence are indicative of β-pleated sheet and α-helix, respectively, with their names labeled below (Yuan et al., 2008). The predicted secondary structure elements of φC31 integrase NTD is named according to the actual ones of TP901 integrase NTD.

CONCLUSIONS

Since φC31 integrase is difficult to crystallize, other methods such as mutagenesis have been introduced to elucidate the precise mechanism of the catalysis of DNA recombination by φC31 integrase. A number of hyperactive φC31 integrase mutants have been isolated and extensively studied (Rowley et al., 2008). Here, we briefly review these studies and then speculate that φC31 integrase controls recombination direction through the direct interaction between the predicted C1 and αA'. This model provides some new insights into the directional control mediated by φC31 integrase, which may contribute to further investigation.

ACKNOWLEDGMENTS

REFERENCES

Brown WRA, Lee NCO, Xu Z, Smith M, et al. (2011). Serine recombinases as tools for genome engineering. *Methods* 53: 372-379.

Burke ME, Arnold PH, He J, Wenwieser SVCT, et al. (2004). Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation. *Mol. Microbiol*. 51: 937-948.

Burkhard P, Stetefeld J and Strelkov SV (2001). Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol.* 11: 82-88.

Campbell A (2006) General aspects of lysogeny. In: The bacteriophages (Calendar R, ed.). 2nd edn. Oxford University Press, Oxford, 66-73.

Cole C, Barber JD and Barton GJ (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Res*. 36: 197-201.

Gorth AC and Calos MP (2004). Phage integrase: biology and applications. *J. Mol. Biol*. 335: 667-678.

Grindley NDF, Whiteson KL and Rice PA (2006). Mechanisms of site-specific recombination. *Ann. Rev. Biochem*. 75: 567-605.

Keenholtz RA, Rowland SJ, Boocock MR, Stark WM, et al. (2011). Structural basis for catalytic activation of a serine recombinase. *Structure* 19: 799-809.

Keravala A, Lee S, Thyagarajan B, Olivares EC, et al. (2009). Mutational derivatives of phiC31 integrase with increased efficiency and specificity. *Am. Soc. Gene Ther*. 17: 112-120.

Khaleel T, Younger E, McEwan AR, Varghese AS, et al. (2011). A phage protein that binds φC31 integrase to switch its directionality. *Mol. Microbiol*. 80: 1450-1463.

Li W, Kamtekar S, Xiong Y, Sarkis GJ, et al. (2005). Structure of a synaptic γδ resolvase tetramer covalently linked to two cleaved DNAs. *Science* 309: 1210-1215.

Liesner R, Zhang W, Noske N and Ehrhardt A (2006). Critical amino acid residues within the φC31 integrase DNA-binding domain affect recombination activities in mammalian cells. *Hum. Gene Ther*. 21: 1104-1118.

Liu S, Ma J, Wang W, Zhang M, et al. (2010). Mutational analysis of highly conserved residues in the phage phiC31 Integrase reveals key amino acids necessary for the DNA recombination. *PLoS One* 5: e8863.

Liu T, Fang Y, Yao W, Jia H, et al. (2013). Fluorescent protein-based detection of φC31 integrase activity in mammalian cells. *Anal. Biochem*. 441: 104-108

Lupas A, Van Dyke M and Stock J (1991). Predicting coiled coils from protein sequences. *Science* 252: 1162-1164.

McEwan AR, Rowley RA and Smith MCM (2009). DNA binding and synapsis by the large C-terminal domain of φC31 integrase. *Nucleic Acids Res*. 37: 4764-4773.

McEwan AR, Raab A, Kelly SM, Feldmann J, et al. (2011). Zinc is essential for high-affinity DNA binding and recombinase

activity of φC31 integrase. *Nucleic Acids Res.* 39: 6317-6147.

Rowley PA and Smith MCM (2008). Role of the N-terminal domain of φC31 integrase in attB-attP synapsis. *J. Bacteriol.* 190: 6918-6921.

Rowley PA, Smith MCA, Younger E, Smith MCM, et al. (2008). A motif in the C-terminal domain of φC31 integrase controls the directionality of recombination. *Nucleic Acids Res.* 36: 3879-3891.

Smith MCA, Till R, Brady K, Soultanas P, et al. (2004). Synapsis and DNA cleavage in φC31 integrase-mediated site-specific recombination. *Nucleic Acids Res*. 32: 2607-2617.

Smith MCM and Thorpe HM (2002). Diversity in the serine recombinases. *Mol. Microbiol.* 44: 299-307.

Smith MCM, Brown WR, McEwan AR, Rowley PA, et al. (2010). Site-specific recombination by phiC31 integrase and other large serine recombinases. *Biochem. Soc. Trans*. 38: 388-394.

Yuan P, Gupta K and Duyne GDV (2008). Tetrameric structure of a serine integrase catalytic domain. *Structure* 16: 1275-1286.