Teaching Cre to follow directions

Gregory D. Van Duyne[1]

Department of Biochemistry and Biophysics, Howard Hughes Medical Institute, University of Pennsylvania School of Medicine, 242 Anatomy–Chemistry Building, Philadelphia, PA 19104

Tyrosine recombinases carry out a variety of important functions involved in moving and maintaining genomes (1). Several members of this family, including the bacteriophage P1 Cre recombinase and the bacteriophage λ-integrase, have also emerged as powerful tools with broad applications in genetic engineering and molecular biology (2, 3). Although structural, biochemical, and genetic experiments have revealed in some detail how the tyrosine recombinases carry out the recombination process, it is less clear how the more complex members of this family achieve a high level of regulation or how the tightly regulated recombinases have evolved such complexity. An article in a recent issue of PNAS (4) provides important new insights into both questions. By making a chimeric protein composed of Cre recombinase fused to a small DNA-binding domain of λ-integrase, Warren *et al.* (4) have turned the normally unregulated Cre into a regulated integrase with the same requirements and directionality found in the λ-integrase system.

Although Cre and λ-integrase carry out very similar reactions, they differ dramatically in their level of regulation. Cre is promiscuous and simple, efficiently recombining minimal 34-bp *loxP* sites with little regard to topological context and with no requirement for host or phage-encoded protein factors (5). Because the products of recombination are also *loxP* sites, the reaction is readily reversible. The integration and excision reactions catalyzed by λ-integrase, however, are strongly directional and highly regulated (6). Integration of the phage genome into the bacterial chromosome occurs between DNA sequences named *attP* and *attB* (the phage and host DNA sites, respectively) and requires the integration host factor (IHF) protein. Excision of the phage genome occurs between *attL* and *attR* sites (generated as a result of integration) and requires both IHF and the phage-encoded excisionase (Xis) protein.

The Cre and λ-integrase proteins themselves also have an important difference. Both share similar catalytic and core-binding (CB) domains that are responsible for carrying out the basic recombination biochemistry, but λ-integrase has an additional domain on its N terminus, termed the "N-domain" or "arm-binding domain" (Fig. 1). This N-domain binds to DNA sequences that are located outside of the core "*loxP*-like" site where strand exchange takes place. Thus, λ-integrase binds simultaneously to two different DNA sequences: the CB and catalytic domains engage the core site and the N-domains bind to sequences located in the "arms" of the *att* sites.

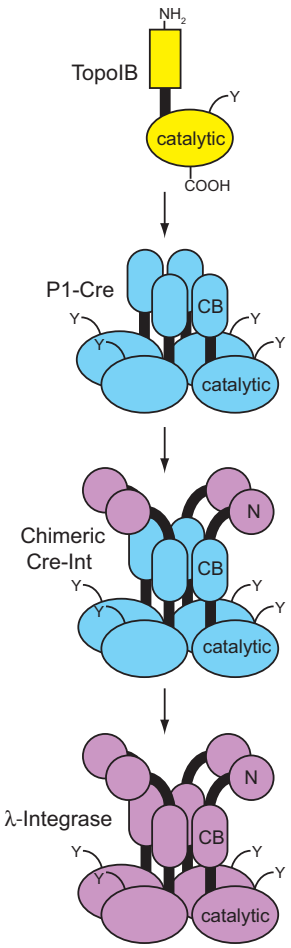

Fig. 1. Possible evolutionary pathway for tyrosine recombinases. The simplest recombinases, typified by Cre, may have evolved from an ancestral topoisomerase that is related to modern TopoIb enzymes (10). The simple recombinases may then have gained regulatory control by attenuation of their own DNA-binding activity, together with acquisition of auxiliary DNA-binding domains. The chimeric Cre protein is a simulation of that process, where Cre has gained the N-domain of λ-integrase. The TopoIb and tyrosine recombinase enzymes share similar catalytic domains, active sites, and mechanisms of phosphoryl transfer. The CB domains of Cre and λ-integrase are similar to one another, but quite different from the corresponding domain in the TopoIb enzymes.

Because the presence of an arm-binding N-domain is the most obvious difference between the Cre and λ-integrase proteins, Warren *et al.* (4) asked whether adding an N-domain to Cre would lead to a novel recombinase with the regulatory properties of λ-integrase. To test this idea, they generated modified *attP* and *attB* sites for integration (called *lotP* and *lotB*) and modified *attL* and *attR* sites for excision (called *lotL* and *lotR*) in which the core λ-integrase binding sites were replaced by weakened Cre binding sites. By weakening the interaction between Cre and its binding sites, they were able to mimic the situation in the λ-integrase system, where the core site interaction is weaker than the arm site interaction.

Remarkably, the resulting Cre chimera is not only functional, but displays the same regulated directionality as λ-integrase. Chimeric Cre requires IHF for efficient *lotP* × *lotB* recombination and requires both IHF and Xis for the *lotL* × *lotR* reaction. The chimeric Cre integration reaction is also inhibited by Xis and the excision reaction is stimulated by *Escherichia coli* Fis protein, both well-established features of the native λ-integrase reactions (6). Warren *et al.* (4) also asked whether more subtle features of the λ-integrase system were reproduced by the chimeric Cre enzyme. There are five N-domain binding sites in the arms of λ-*attP*, named P1, P2, P′1, P′2, and P′3. Previous experiments had confirmed that P1 but not P2 is required for the λ integration reaction (7). Similarly, P2 but not P1 is required for the excision reaction. With substrates in which the P1 or P2 sites had been inactivated by mutation, chimeric Cre displayed the same N-domain binding site requirements in the *att* site arms as did λ-integrase. Together, the results strongly support the idea that chimeric Cre is carrying out regulated recombination by using the same mechanistic pathway as λ-integrase.

One interesting implication of the chimeric Cre experiments relates to the extent of interaction between the

Author contributions: G.D.V.D. wrote the paper.

The author declares no conflict of interest.

See companion article on page 18278 in issue 47 of volume 105.

[1]E-mail: vanduyne@mail.med.upenn.edu.

www.pnas.org/cgi/doi/10.1073/pnas.0811624106

N-domains and the CB domains of λ-integrase when assembled into a protein–DNA complex competent for recombination. The only structural models available that include the intact integrase protein are low-resolution crystal structures where the four N-domains are bound to two DNA duplexes (8). Although the N-domain site arrangement in these structures is artificial, it allowed the complex to be crystallized and the architecture of the full-length tetrameric complex to be visualized. The N-domains form a tetramer in these structures that appears to be tightly coupled to the CB domains through an interdomain linker. This coupling was proposed to form the basis of an allosteric mode of regulation, whereby the N-domains modulate the activity of the CB and catalytic domains (8).

Because the sequence similarity between Cre and λ-integrase is low, it is unlikely that a functional interaction between the λ-integrase N-domain and the Cre CB-domain occurs in the context of the chimeric Cre system. Indeed, a rather long linker is present in the Cre-integrase fusion construct, suggesting that the N-domain is tethered, but not tightly coupled to the CB domain. Given that chimeric Cre has similar regulated directionality as λ-integrase and has the same N-domain binding site requirements, it is reasonable to conclude that tight coupling of the N-domain and CB domain is not required for regulated directionality. This idea is consistent with the presence of a flexible, protease-sensitive linker between domains in λ-integrase and helps to explain findings that the N-domain binding site requirements in the P and P′ arms might require more flexibility than is implied by the compact N-domain tetramer observed in the crystal structures (9).

Warren *et al.* have turned the normally unregulated Cre into a regulated integrase.

It is remarkable that Warren *et al.* (4) were able to transfer a complex regulatory framework to an unregulated recombinase by making only a small number of rationally designed changes. They have suggested that the relative ease of this conversion of function might indicate an evolutionary pathway by which simpler recombinases represented by Cre could have acquired higher-level regulation (Fig. 1). Two steps would have been required in this process: weakening of the recombinase–DNA interaction and fusion of an auxiliary DNA-binding domain to the N terminus. The first could occur by simple mutation of the recombinase and/or the binding sites. The second could occur as the outcome of homologous recombination events, but Warren *et al.* (4) point out that an alternative, more efficient mechanism could also play an important role. Many recombinase genes are located adjacent to their *att* sites, setting the stage for fusion with new sequences via site-specific recombination with weak *att* sites located within the recombinase gene and elsewhere in the bacterial genome.

The finding that a chimeric Cre-integrase has gained regulated directionality by adding only a 60-residue DNA-binding domain from λ-integrase is intriguing for a number of reasons. In addition to the mechanistic and evolutionary insights noted above, this work opens the door for some interesting new directions. For example, new hybrid recombinases could be prepared by using similar strategies and tested for properties that would be useful in experiments involving programmed gene rearrangements. As noted by Warren *et al.* (4), a process of self-promoted recombination that the chimeric Cre construct may represent could also be exploited as an experimental platform for exploring new sequences. Of course, we are also reminded once again that there is much to be learned from λ.

1. Nash HA (1996) Site-specific recombination: Integration, excision, resolution, and inversion of defined DNA segments. *Escherichia coli and Salmonella: Cellular and Molecular Biology*, eds Neidhardt FC, *et al.* (Am Soc Microbiol, Washington, DC), pp 2363–2376.
2. Branda CS, Dymecki SM (2004) Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. *Dev Cell* 6:7–28.
3. Hartley JL, Temple GF, Brasch MA (2000) DNA cloning using in vitro site-specific recombination. *Genome Res* 10:1788–1795.
4. Warren D, Laxmikanthan G, Landy A (2008) A chimeric cre recombinase with regulated directionality. *Proc Natl Acad Sci USA* 105:18278–18283.
5. Abremski K, Hoess R (1984) Bacteriophage P1 site-specific recombination: Purification and properties of the Cre recombinase protein. *J Biol Chem* 259:1509–1514.
6. Azaro MA, Landy A (2002) λ integrase and the λ Int family. *Mobile DNA II*, eds Craig NL, Craigie R, Gellert M, Lambowitz AM (Am Soc Microbiol, Washington, DC), pp 118–148.
7. Numrych TE, Gumport RI, Gardner JF (1990) A comparison of the effects of single-base and triple-base changes in the integrase arm-type binding sites on the site-specific recombination of bacteriophage lambda. *Nucleic Acids Res* 18:3953–3959.
8. Biswas T, *et al.* (2005) A structural basis for allosteric control of DNA recombination by lambda integrase. *Nature* 435:1059–1066.
9. Hazelbaker D, Azaro MA, Landy A (2008) A biotin interference assay highlights two different asymmetric interaction profiles for lambda integrase arm-type binding sites in integrative versus excisive recombination. *J Biol Chem* 283:12402–12414.
10. Cheng C, Kussie P, Pavletich N, Shuman S (1998) Conservation of structure and mechanism between eukaryotic topoisomerase I and site-specific recombinases. *Cell* 92:841–850.

Van Duyne