Rational Engineering of a site-specific anti-HIV recombinase

Engineered site-specific recombinases hold great promise for advanced biomedical applications. However, the generation of enzymes with desired properties remains challenging. Here we compare the Cre/loxP system and several newly described Cre-like site-specific recombinases in order to understand their specificity and to be able to rationally engineer a specific and efficient anti-HIV recombinase.

Site-specific recombinases (SSRs) are essential for a variety of diverse biological processes, including the integration, excision and inversion of genomic DNA segments [1]. Recently, several recombinases have become very popular tools for manipulating DNA *in vitro* and *in vivo* [2]. In the last decade, numerous studies aimed to mutate or redesign well-characterized DNA recombinases so that they specifically recognize new target sequences[3-5]. This represents an important step in broadening the use of these recombinases, so that their application can be extended to artificial/engineered DNA target sites. Applied site-specific recombination has become an important technology to precisely manipulate the genome in a broad range of organisms[6, 7]. However, DNA sequence constraints imposed by site-specific recombinases make the integration of new recognition sites into genomes difficult. Hence, different methodologies emerged, including substrate-linked protein evolution[5] or targeted mutagenesis with positive selection[3, 4] to generate recombinases recognizing homologous DNA sequences already present in the genome and efficiently recombining them. However, engineered site-specific recombinase variants often lead to a diversity of binding profiles with considerable plasticity and a low level of specificity[8, 9], an effect which is challenging when it comes to therapeutic application.

The bacteriophage P1 Cre recombinase is a member of the tyrosine recombinase family. Cre catalyzes site-specific recombination between 34 bp loxP (locus of X-over P1) sites consisting of two 13 bp inverted repeats surrounding an 8 bp spacer[10] (Figure 1A). The synaptic complex is formed by two loxP sites bound each by two Cre molecules. The recombination starts when the first Cre molecule attacks through a nucleophilic tyrosine each loxP strand creating a covalent DNA-protein phosphotyrosine linkage. This step will lead to the formation of a Holliday junction (HJ), which will be resolved by the nucleophilic attack of the second Cre molecule recreating a covalent complex. The phosphate will be released to obtain a product complex made out of recombined DNA (Figure 1B).

The Cre/loxP system is widely used in genomic engineering[11] because it can catalyze the complete recombination reaction between loxP substrates *in vitro* in the absence of accessory proteins and in the absence of auxiliary DNA-binding domains[10]. In particular, the development of genome engineering in living organisms has made the Cre/lox system an invaluable instrument for advanced genetic studies.

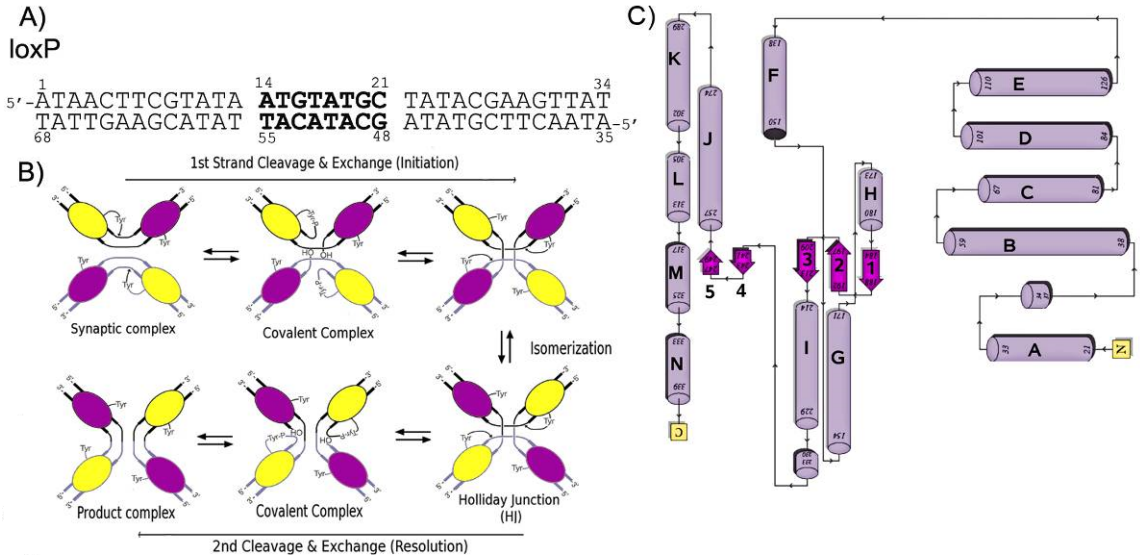

Figure 1: Cre/loxP recombination mechanism and Cre secondary structure.
A) The sequence of loxP consists of a 13 inverted repeated with an 8 bp spacer (in bold). B) Scheme of the loxP recombination process by Cre. The active and non-active Cre proteins are shown in purple and yellow, respectively. C) A topology diagram of the Cre crystal structure (taken from PDBsum [12]). The alfa-helices are named A to N and the β-sheets are numbered from 1 to 5.

One of the features of Cre recombinase is that it can specifically remove a DNA segment flanked by its loxP target sites. A recent breakthrough was the engineering of an evolved recombinase, named Tre, that would recombine a DNA sequence, named loxLTR, present in the 5′-LTR and 3′-LTR of an integrated provirus (HIV-1)[13]. Tre was evolved from Cre using a substrate-linked protein evolution methodology[5] and was demonstrated to efficiently excise the provirus from the genomic DNA of human HIV-1 infected cells[13]. The generation of new recombinases like Tre proved the capacity for Cre to adapt to divergent target sites. However, an immediate consequence of the Cre-to-Tre evolution process was that most of the candidates isolated from the evolved Tre library showed a relaxed specificity recombining both loxLTR and loxP. Typically, the evolved library is a diverse mix of clones with different levels of activity and specificity and needs to be extensively screened in order to find clones that are both efficiently and specifically recombining the new target sequence. Therefore, a rational approach for aiding directed evolution towards DNA specificity of site-specific recombinase was elaborated[14]. This combined methodology led to the generation of a new efficient recombinase, sTre, exhibiting the designed specificity. Despite this advance in specificity design, engineering a new recombinase for a new target with desired properties remains a challenge. Hence, it is important to study and understand how the specificity is evolving through the family members of SSRs. Recently a set of Cre-like proteins of the SSRs have been described, Dre[15], VCre[16], SCre[16] and vika[17], which recombine different DNA targets and do not cross-recombine. We have used available structural and sequence data to study the convergence and divergence of these SSRs in terms of specificity. The available crystal structures of Cre at high resolution reveal at atomic detail the recognition interface and the key regions for the specific recognition of loxP[18, 19]. Sequence analysis shows the low conservation rate of these regions in those SSRs. A detailed analysis of these regions in the different proteins and their corresponding targets gives insights on the evolution of SSRs specificity, which represents a first step in the establishment of a rationale for the engineering of new recombinases with desired specificity profiles.

Results

Alignment of Cre with the new site-specific recombination enzymes

The new site-specific recombinases share a percentage of identity ranging from 23 to 39%. However, they share a high percentage of similarity, which ranges from 43 to 62% (Table 1).

	Cre	Vika	Vcre	Dre	Scre
Cre	100/100				
Vika	25 / 47	100/100			
Vcre	23 / 43	37 / 62	100/100		
Dre	39 / 61	23 / 44	25 / 43	100/100	
Scre	25 / 44	29 / 49	33 / 51	23 / 44	100/100

Table 1: Percentage of identity and similarity of the site-specific recombinases.
For each pair of proteins percentage of identity / percentage of similarity is shown.

SSRs proteins consist of two domains connected by a linker where the C-terminus is more

conserved than the N-terminus (Figure 1C). The multiple sequence alignment shows the conservation of the catalytical triad in the C-terminus and the divergence in the N-terminus between these 5 proteins (Figure 2). An insertion or deletion in the beginning of the N-termini and around the helix J in Cre can be observed, which can originate a different disposition in 3D of the residues in these regions among SSRs.

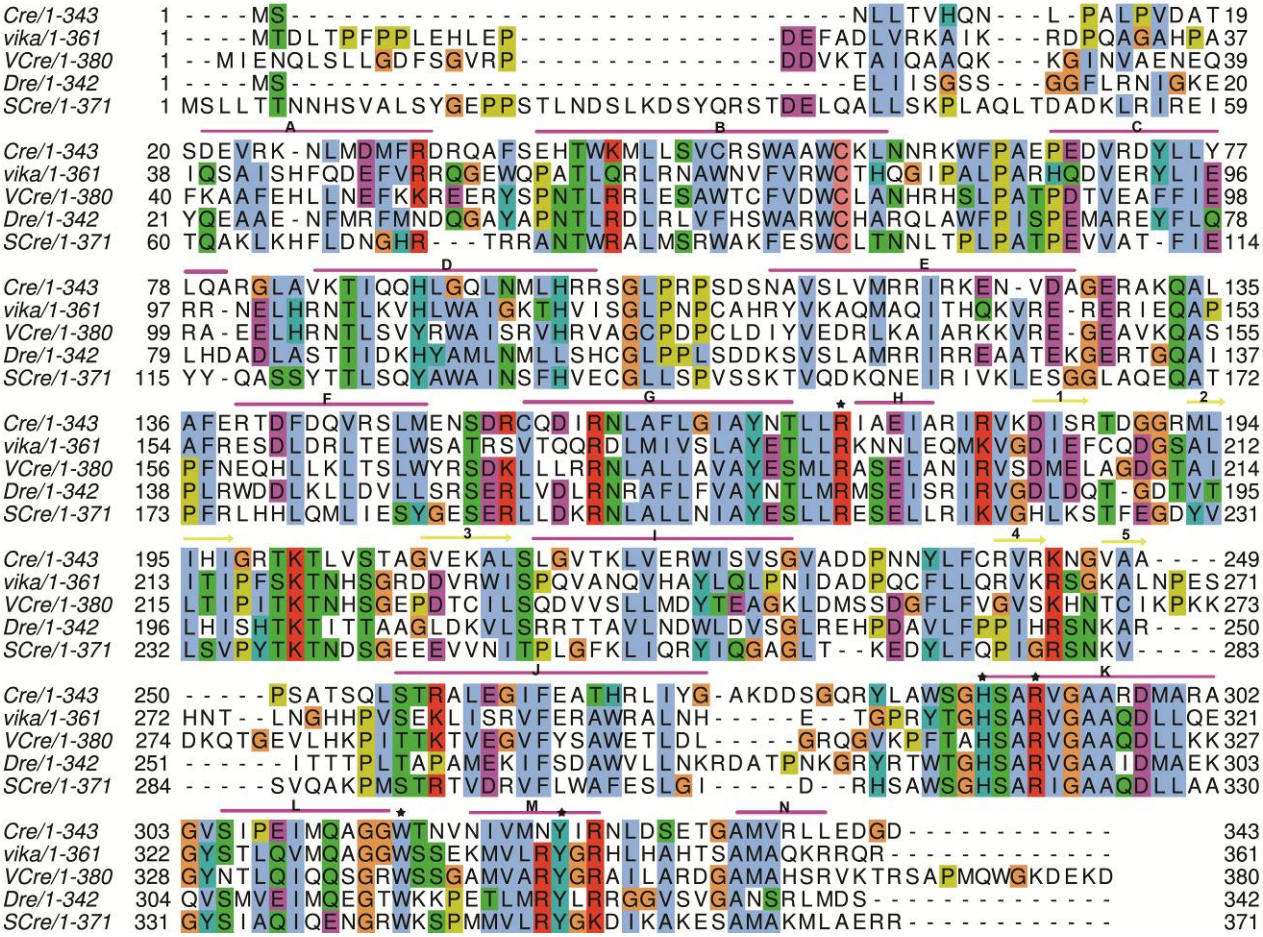

Figure 2: Sequence alignment of Cre with the newly described SSRs.
Cre, vika, Vcre, Dre and Scre are aligned using Clustal Omega and colored by the Clustalx method [20]. The secondary structure of Cre is represented above the sequences: alpha-helices in purple and β-sheets in yellow. The residues forming the catalytical triad are indicated with an asterisk (*).

The surface of interaction in Cre/loxP system is defined as residues within a radius of 6 Å to loxP. In order to account of this surface, we mapped onto the Cre structure the residues interacting with loxP. It can be observed that the residues in contact with the DNA are not conserved among the recombinases, neither in their N-termini nor in the C-termini, except for the region of the catalytical triad. A region of conserved residues can be observed in the C-terminus formed by the catalytical triad and its surrounding residues in sequence and in 3D (Figure 2).

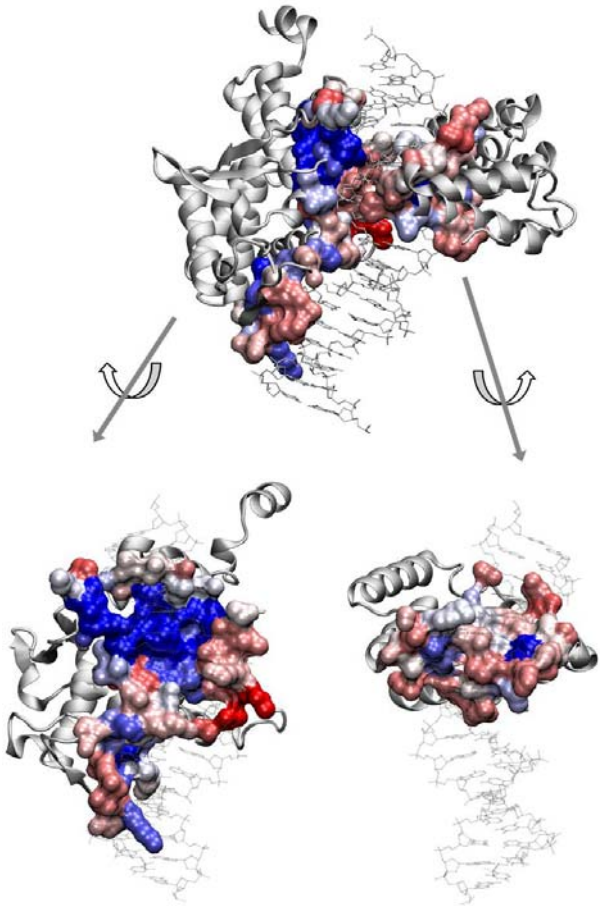

Figure 3: Mapping of the conservation rate of the SSRs on Cre crystal structure.
The top panel shows the surface of interaction of Cre/loxP. The bottom panel shows more in detailed the C-terminus (left) and the N-terminus (right). The scale goes from blue (conserved by type of amino acids) to red (not conserved).

Cre/loxP interfacial interactions

In the detailed analysis of the surface of interaction between Cre and loxP, we distinguish between a direct readout (*i.e.* residues interacting with the grooves of the DNA) and indirect readout (*i.e.* residues interacting with the backbone of the DNA). We observe on both arms of loxP the same pattern of interaction. In order to study and understand how Cre specifically interacts with its substrate, we have focused on the specific interactions made through the DNA grooves. We can describe 3 regions of interaction between one Cre molecule and one loxP arm. In fact, the loxP arm is contacted in its beginning, middle and end. The beginning of the loxP arm is heavily contacted in its major and minor groove by residues H40, K43, M44, S47, K86, Q90, Q94, K201 and R282. The middle part of loxP is contacted through the helix J in the major groove by residues R259 and E262. And finally, the end part of loxP is contacted in the minor groove by residues R241, R243 and K244.

Conservation of the specificity regions in the SSRs

In the alignment of the regions of Cre containing the residues described above with the Cre-like recombinases (Figure 3*Figure*), we observe that the conservation rate through the family is low in most of all residues involved in the direct interaction with the DNA. Furthermore, in our recent studies, we have been able to observe that the evolution from Cre to sTre induced mutations in these regions of interaction, which led to a switch in specificty[14]. This suggests the importance of these regions in the design of specificity.

Among the residues implicated in the specific interactions, residue K201 presents a different profile. This residue is conserved among these proteins and as well in the family of DNA breaking-rejoining enzymes, C-terminal catalytic domain (cd00397). K201 interacts in the minor groove

proximal to the crossover region, and it was shown to be important for the strand exchange order as well as for the formation of the HJ [21, 22]. The mutation of K201 to any other amino acid results in a defect in the recombination [23]. This analysis points out that when designing specificity of a new recombinase to a new DNA target, these regions are to be taken into account.

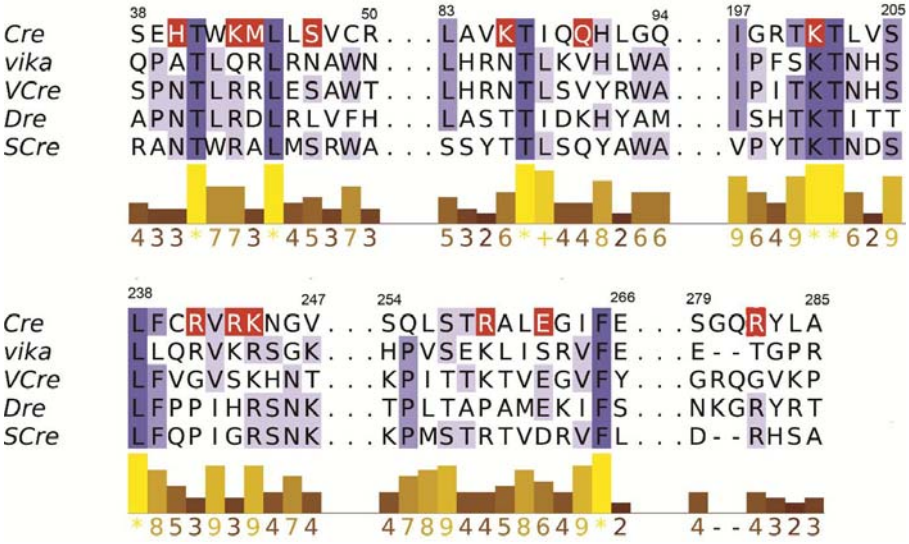

Figure 3: Alignment of the protein regions that include amino acids of Cre contacting specifically loxP.

Amino acids interacting specifically in the groove of loxP in the Cre/loxP system are indicated in red. The conservation of the amino acid between these recombinases is shown as a histogram. The alignment is colored by percentage of identity. The numbers above the alignments represent the residue numbering in Cre.

Discussion

The rational design of a protein with a desired specific DNA binding profile remains a remote goal. For achieving this goal, it is essential to study the binding interface between the protein and its substrate. Therefore, the access to high resolution three-dimensional structures of the complexes as well as the sequences other family members provide a rich ground to investigate how proteins recognize specifically the DNA by using molecular modeling techniques. Recently, by using a combined methodology of substrate-linked evolution and structure-aided computational design we have been able to engineer an anti-HIV recombinase called sTre [14]. However, many challenges remain to be able to design “a la carte” specificity in such systems. In our work, we show how by studying different members of the Cre-like proteins we can designate regions implicated in specificity. We show that in Cre/loxP the residues contained in helix B, D, J, β-sheet 4 and the loop between helix J and K ,which contact specifically the DNA grooves, are not conserved through the family. This indicates that when rationally engineering an anti-HIV recombinase these regions should be targeted for designing specificity.

We put in evidence the conservation of residue K201 among the SRRs members, which contacts specifically the DNA minor groove in Cre/loxP. K201 is a key player in the recombination mechanism and should be considered as part as the active site.

Engineered recombinases are important tools for biomedical application. However, to be able to use these recombinases they have to possess a high affinity and specificity profile towards their targets to avoid any undesired effect on the genome. Therefore, a detailed analysis of available information about family members is crucial to establish a rationale that will allow us to narrow down the regions where we can mutate or redesign to obtain a desired profile.

References

1. Grindley, N.D., K.L. Whiteson, and P.A. Rice, *Mechanisms of site-specific recombination.* Annu Rev Biochem, 2006. **75**: p. 567-605.
2. Akopian, A. and W. Marshall Stark, *Site-specific DNA recombinases as instruments for genomic surgery.* Adv Genet, 2005. **55**: p. 1-23.
3. Santoro, S.W. and P.G. Schultz, *Directed evolution of the site specificity of Cre recombinase.* Proc Natl Acad Sci U S A, 2002. **99**(7): p. 4185-90.
4. Baldwin, E.P., et al., *A specificity switch in selected cre recombinase variants is mediated by macromolecular plasticity and water.* Chem Biol, 2003. **10**(11): p. 1085-94.
5. Buchholz, F. and A.F. Stewart, *Alteration of Cre recombinase site specificity by substrate-linked protein evolution.* Nat Biotechnol, 2001. **19**(11): p. 1047-52.
6. Kilby, N.J., M.R. Snaith, and J.A. Murray, *Site-specific recombinases: tools for genome engineering.* Trends Genet, 1993. **9**(12): p. 413-21.
7. Glaser, S., K. Anastassiadis, and A.F. Stewart, *Current issues in mouse genome engineering.* Nat Genet, 2005. **37**(11): p. 1187-93.
8. Ernst, A., et al., *Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing.* Mol Biosyst. **6**(10): p. 1782-90.
9. Aharoni, A., et al., *The 'evolvability' of promiscuous protein functions.* Nat Genet, 2005. **37**(1): p. 73-6.
10. Abremski, K. and R. Hoess, *Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein.* J Biol Chem, 1984. **259**(3): p. 1509-14.
11. Garcia-Otin, A.L. and F. Guillou, *Mammalian genome targeting using site-specific recombinases.* Front Biosci, 2006. **11**: p. 1108-36.
12. Laskowski, R.A., *PDBsum new things.* Nucleic Acids Res, 2009. **37**(Database issue): p. D355-9.
13. Sarkar, I., et al., *HIV-1 proviral DNA excision using an evolved recombinase.* Science, 2007. **316**(5833): p. 1912-5.
14. Abi-Ghanem, J., et al., *Engineering of a target site-specific recombinase by a combined evolution- and structure-guided approach.* Nucleic Acids Res. **41**(4): p. 2394-403.
15. Anastassiadis, K., et al., *Dre recombinase, like Cre, is a highly efficient site-specific recombinase in E. coli, mammalian cells and mice.* Dis Model Mech, 2009. **2**(9-10): p. 508-15.
16. Suzuki, E. and M. Nakayama, *VCre/VloxP and SCre/SloxP: new site-specific recombination systems for genome engineering.* Nucleic Acids Res. **39**(8): p. e49.
17. Karimova, M., et al., *Vika/vox, a novel efficient and specific Cre/loxP-like site-specific recombination system.* Nucleic Acids Res. **41**(2): p. e37.
18. Guo, F., D.N. Gopaul, and G.D. van Duyne, *Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse.* Nature, 1997. **389**(6646): p. 40-6.
19. Van Duyne, G.D., *A structural view of cre-loxp site-specific recombination.* Annu Rev Biophys Biomol Struct, 2001. **30**: p. 87-104.
20. Goujon, M., et al., *A new bioinformatics analysis tools framework at EMBL-EBI.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W695-9.
21. Ghosh, K., et al., *Preferential synapsis of loxP sites drives ordered strand exchange in Cre-loxP site-specific recombination.* Nat Chem Biol, 2005. **1**(5): p. 275-82.
22. Ghosh, K., F. Guo, and G.D. Van Duyne, *Synapsis of loxP sites by Cre recombinase.* J Biol Chem, 2007. **282**(33): p. 24004-16.
23. Gibb, B., et al., *Requirements for catalysis in the Cre recombinase active site.* Nucleic Acids Res, 2010. **38**(17): p. 5817-32.