Supplementary Data

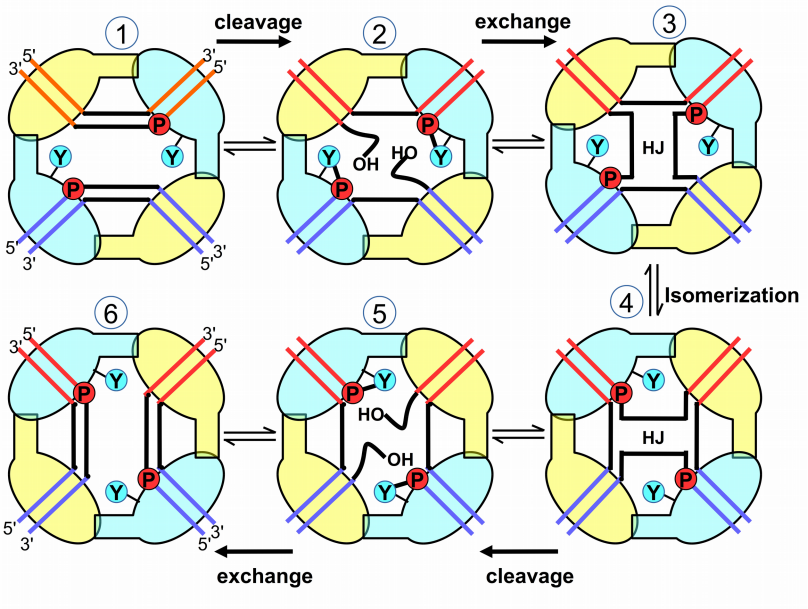
Crystal structure of an engineered, HIV-specific recombinase for removal of integrated proviral DNA.

Gretchen Meinke[1], Janet Karpinski[2], Frank Buchholz[2,3,4,5] and Andrew Bohm[1,*]

[1] Department of Developmental, Molecular and Chemical Biology, Tufts University School of Medicine, Boston, MA, 02111, USA
[2] Medical Systems Biology, UCC, Medical Faculty Carl Gustav Carus, TU Dresden, Germany
[3] Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany
[4] German Cancer Research Center (DKFZ), Heidelberg and German Cancer Consortium (DKTK) partner site Dresden, Germany
[5] National Center for Tumor Diseases (NCT), University Hospital Carl Gustav Carus, TU Dresden, Dresden, Germany

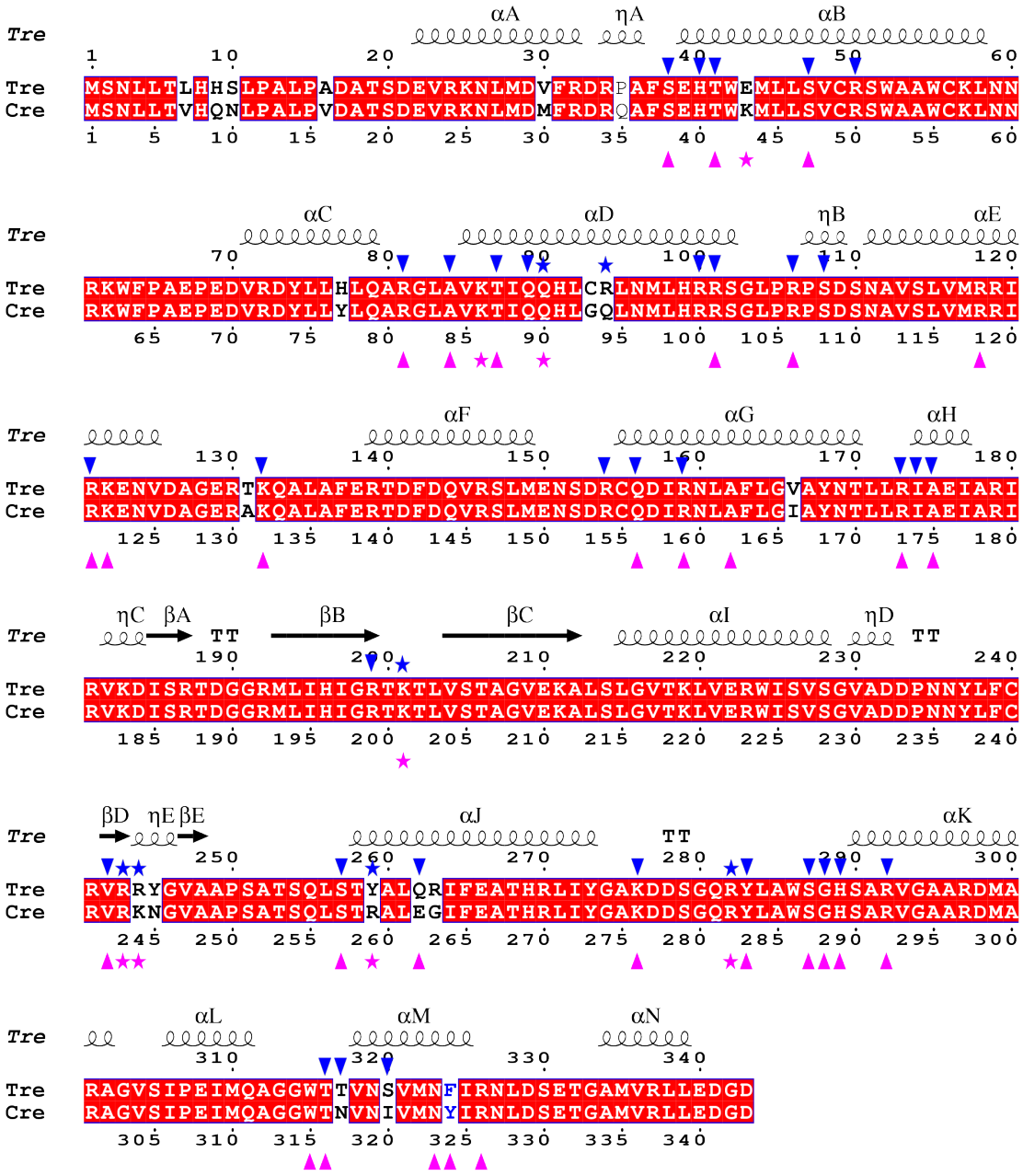
* To whom correspondence should be addressed. Email: Andrew.Bohm@Tufts.edu

Supplementary Figure 1



Supplementary Figure 1. Schematic Overview of the step-wise Cre recombination mechanism. Step 1. (left) Synapsis. Four molecules of Cre interact with two *loxP* substrates to form a tetrameric complex. This complex results in a large bend in the DNA shown schematically. The duplex DNA is represented by two lines, labeled 5' or 3' to indicate the DNA directionality. The recombinase is shown as an asymmetric shape bound to each half-site in either the "cleaving competent" conformation (cyan) or "non-cleaving" conformation (yellow). The nucleophilic tyrosine is shown only for the active conformation and indicated by a Y in a blue circle. The location of the activated scissile phosphate is indicated by a red circle around the letter P In addition, in the synaptic tetramer, the protomers are arranged head-to-tail and have a pseudo four-fold symmetry. Each protomer communicates its state in trans to its neighbor via C-terminal conformational changes shown schematic as a rectangle extension interacting with a neighboring protomer. Note that opposing protomers are active (and poised to interact with the scissile phosphate on one strand of DNA) or inactive. Step 2. Cleavage/Ligation. The nucleophilic tyrosine attacks the scissile phosphate forming a 3'-phosphotyrosine linkage, and resulting in a free 5'OH. Step 3. Strand exchange. Strand exchange occurs when the free 5'OH attacks the neighboring phosphotyrosine and forms a Holliday Junction (HJ) intermediate. Step 4. Isomerization. The inactive conformers adopt active conformations and the active conformers become inactive conformers. The complex is ready for the second round of strand exchange. Again, the catalytic tyrosine attacks the scissile phosphate to form the phosphotyrosine intemediate. Step 5. The free 5'OH group attacks the phosphotyrosine. Step 6. Resolution of the recombined target.

Supplementary Figure 2. Sequence Alignment of Tre_{Y324F} and Cre



Supplementary Figure 2 Legend. The secondary structure assignments for Tre are shown. Identical residues are in red boxes with white letters. Non-identical residues are shown as black text. The catalytic nucleophile (Tyr 324) is shown as blue text. Protein-DNA interactions for the cleaving-competent Tre (blue) and Cre (pink) are shown above and below the sequence, respectively. The triangles indicate polar interaction with the phosphate backbone, and a star indicates a sequence-specific polar interaction with a base. For Tre, these were calculated using the program PISA, and verified by examination of the structure. For Cre, these interactions were obtained from the Nucleic acid-Protein Interaction DataBase, http://npidb.belozersky.msu.ru/. This figure was made using the program ESPript3, http://espript.ibcp.fr/ESPript/ESPript/.

Supplementary Table 1. Structural Alignment Results of Tre/*loxLTR* and Cre/*loxP*

Reaction State	PDB ID	Res (Å)	AU:	Cre to Cre (PDB 1q3v) RMSD of Tet	Tre to Cre RMSD of Tet (outlier rejection)	Tre to Cre RMSD of Tet (no outlier rejection)	Tre monomer closest "cleaving" (#Ca)	Citation (PMID)
Cre-*loxP* Synaptic complex	1NZB	3.10	Tet	0.421 (1197)	0.739 (1185)	1.123 (1293)	0.89 (320) B	PMID: 12954782
	4CRX	2.20	Dimer	1.915 (1258)	2.047 (1256)	2.263 (1286)	0.83 (322) B	PMID: 10377382
	5CRX	2.70	Dimer	2.318 (1146)	2.415 (1139)	3.467 (1184)	0.65 (295) B	PMID: 10377382
	3C28	2.60	Dimer	1.515 (1183)	1.639 (1194)	1.878 (1226)	0.91 (311) B	(no reference)
	3C29	2.20	Tet	1.173 (1259)	1.361 (1257)	1.520 (1286)	0.76 (322) B	(no reference)
Cre-*loxP* Pre-cleavage intermediate	1Q3U	2.90	Tet	0.429 (1185)	0.787 (1198)	1.132 (1293)	0.81 (318) B	PMID: 12954782
	2HOF	2.40	Dimer	1.760 (1209)	1.831 (1192)	2.068 (1248)	0.88 (311) B	PMID: 17573343
	2HOI	2.60	Tet	1.174 (1251)	1.360 (1253)	1.538 (1286)	0.79 (322) B	PMID: 17573343
Cre-*loxP* Covalent intermediate	1Q3V	2.91	Tet	aligned	0.685 (1153)	1.025 (1293)	0.64 (319) B	PMID: 12954782
	1OUQ	3.20	Tet	0.436 (1291)	0.872 (1195)	1.152 (1293)	0.66 (319) B	PMID: 12954782
	1CRX	2.40	Dimer	1.926 (1210)	2.151 (1239)	2.390 (1286)	0.95 (317) B	PMID: 9288963
Cre-loxP Transition state intermediate	3MGV	2.29	Tet	1.223 (1255)	1.404 (1246)	1.603 (1286)	0.79 (322) B	PMID: 9288963
Cre-*loxP* Holliday junction	1XNS	2.80	Dimer	1.868 (1237)	2.053 (1249)	2.293 (1286)	0.88 (316) B	PMID: 15591069
	1XO0	2.00	Dimer	2.174 (1246)	2.316 (1251)	2.522 (1286)	0.94 (315) B	PMID: 15591069
	2CRX	2.50	Dimer	2.427 (1209)	2.555 (1209)	2.732 (1225)	1.19 (302) B	PMID: 9670032
	3CRX	2.50	Dimer	2.587 (1244)	2.761 (1255)	3.015 (1289)	1.02 (307) B	PMID: 9670032
	1KBU	2.20	Dimer	1.970 (1239)	2.135 (1236)	2.512 (1283)	0.94 (312) B	PMID: 12051940
Cre-*loxP* Cre mutant/mutant-*loxP*	1PVP	2.35	Dimer	1.992 (1200)	2.111 (1195)	2.553 (1279)	0.96 (313) B	PMID: 14652076
	1PVQ	2.75	Dimer	2.070 (1205)	2.297 (1226)	2.632 (1281)	0.85 (312) B	PMID: 14652076
	1PVR	2.65	Dimer	1.985 (1189)	2.187 (1202)	2.680 (1285)	0.87 (312) B	PMID: 14652076
	1MA7	2.30	Dimer	2.069 (1207)	2.243 (1215)	2.583 (1278)	0.98 (312) B	PMID: 12779336
Cre-three way Y junction (trimers)	1DRG	2.55		21.844 (750)	23.74 (777)	34.777 (937)	1.23 (295) A	PMID: 11601846
	1F44	2.05		22.338 (763)	23.60 (782)	34.797 (946)	1.17 (296) A	PMID: 11601846

Supplementary Table 1. The above table was made with the program ALIGN within the PyMOL using a rejection RMS cutoff of 2 Å, and 5 cycles of outlier rejection. Only the protein C-α atoms are included in this table. AU =asymmetric unit; D = dimer, Tet= tetramer. The tetramer was generated if necessary from symmetry mates. The RMSD is in Å, and the number of C-α atoms aligned shown in parentheses. The Cre-to-Cre column using PDB ID 1Q3V as the reference shows how the other Cre structures compare to 1Q3V. The Tre-to-Cre column shows how closely Tre compares to the available Cre structures. The chain ID of the closest monomer is also given. The unique PubMed ID (PMID) for each PDB code is given.

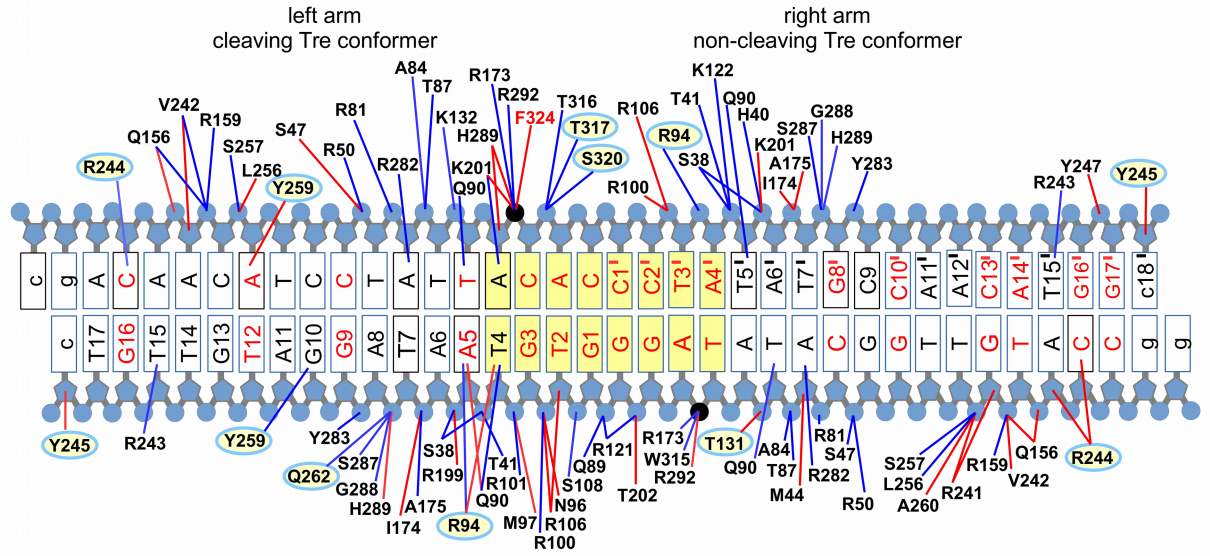
Supplementary Table 2. Structural and Functional Analysis of Tre mutations

Cre-Tre AA	Activity of revertant mutation on *loxLTR*	Contacts DNA?	Contacts adjacent monomer?	Role of Tre mutations in *loxLTR* recognition
V7L	↓	N/A	N/A	Not observed in electron density. Unclear role.
Q9H	↓	N/A	N/A	Not observed in electron density. Unclear role.
N10S	↓	N/A	N/A	Weak electron density but not traced. Unclear role.
V16A	↓	No	No	Weak electron density. Unclear role.
M30V (Fig 10B)	↑	No	Yes	**Weakened protein-protein interface**. M30V creates a void within a hydrophobic core, potentially destabilizing. Has perturbed the protein-protein interface, resulting in the loss of a salt bridge between R 32 (non-cleaving monomer) and E69 (cleaving monomer). Same for non-cleaving conformer.
Q35P (Fig 10A)	↑	No	Yes	**Weakened protein-protein interface**. Has perturbed the protein-protein interface, and results in the loss of a polar interaction present in Cre (Gln35-Glu123 of neighboring monomer). This decrease at the protein-protein interface may increase the accuracy of selectivity of monomers for the DNA. Pro also adds rigidity to the structure at a turn of a helix.
K43E (Fig 7)	↓	No	No	**Important for altered specificity**. Loss of base specific hydrogen bond present in Cre: Lys 43 makes base specific contacts with N7 of Gua 9. Gua 9 is altered to Cyt 9 in *LoxLTR*. K34E carboxyl group forms a water mediated contact with the O4 of Thy 7. Both K43 in Cre and K43E in Tre too far to contact DNA in non-cleaving protomer.
Y77K	↓	No	No	Both Tyr 77 and the Y77H rings occupy a similar orientation near the surface and pack against Trp 155 on one side and Arg 81 on the other. A water mediated H bond is observed between His N and Arg 81 Nζ. Arg 81 makes a phosphate interaction with the other Nζ. In Cre, Arg 81 forms a stronger phosphate interaction. Similar in non-cleaving conformer.
G93C (Fig 6B)	↓	No	No	Cys 93 side chain points toward major groove of DNA, and although too far for sequence specific contact, G93C may make water mediated interaction with DNA. There is a loss of flexibility in going from Gly to Cys. The non-cleaving conformer is similar. Functionally important for Tre.
Q94R (Fig 6B)	↓	Yes	No	**Important for altered specificity.** This mutation adds a new sequence-specific interaction to an altered base: Q94R to the N7 of Ade 5, and Q94R is buttressed by Gln 90 interaction. In contrast, Cre, Gln94 does not interact with DNA but stabilizes the orientation of Gln 90 which makes base specific contacts to the DNA. In the non-cleaving Q94R interacts with P (Ade 4'). In Cre, Gln 94 does not interact with DNA.
A131T (Fig 10C)	↓	Yes	Yes	In the cleaving conformer, T131 hydroxyl forms a hydrogen bond with the backbone of Leu 203. In the non-cleaving conformer, the side chain hydroxyl of A131T forms a hydrogen bond with Thr 206 on a neighboring monomer and adds phosphate backbone interaction on the inactive monomer (Ade 5'). These interactions cannot form in Cre/*loxP* and could help orient Tre on *loxLTR*.
I166V	↓	No	No	I166V mutation creates a tiny void in a hydrophobic pocket and is potentially destabilizing. Similar in non-cleaving conformer.

Cre-Tre AA	Activity of revertant mutation on *loxLTR*	Contacts DNA?	Contacts adjacent monomer?	Role of Tre mutations in *loxLTR* recognition
K244R (Fig 9)	↓	Yes	No	**Important for altered specificity.** This mutation adds sequence specific contacts to altered bases in the *loxLTR*. K244R contacts altered base Cyt 16'. In Cre, Lys 244 contacts Thy 16 and Thy 17. In the non-cleaving conformer, K244R contacts altered bases in both "arms" of *loxLTR*: Cyt 16' and Cyt 17'. In Cre, Lys 244 contacts Thy 16'.
N245Y (Fig 9)	↓	Yes	No	N245Y mutation stabilizes and packs along K244R side chain and widens the minor groove and prevents a steric collision with the DNA backbone (Cyt 17). This results in the loss of a phosphate backbone interaction that occurs in Cre between Asn 245 and P. Similar in non-cleaving conformer.
R259Y (Fig 8)	↓	Yes	No	**Important for altered specificity**. Cre Arg 259 forms a bidentate interaction with the O6 and N7 of Gua 10 in both the left and right arms. In the left arm, this base is the same in *loxLTR*. R259Y loses these two interactions. The Tyr side-chain superimposes well on the Arg side chain, and makes van der Waals contacts with the DNA, and an edge to face interaction. In one copy, the cleaving conformer a polar interaction is observed between Tyr 259 OH and N7 of Gua 10. In the right arm, this base is altered to Cyt 10' in *loxLTR*. The R259Y side chain superimposes well on the Cre Arg 259 side chain, but is too far for any base specific interaction.
E262Q (Fig 8)	↓	Yes	No	**Important for altered specificity**. This E262Q mutation makes direct phosphate interactions with the side chain in both conformers of Tre (with the phosphate on the bottom strand between Thy 7 and Ade 8 or between 7' and 8' on the top strand). In Cre, Glu 262 has been shown to be essential for *loxP* specificity and makes an unusual direct interaction with a phosphate.
G263R (Fig 8)	↓	Yes	No	The G263R mutation inserts in the major groove and its side chain packs along the Tyr 259 side chain and appears to stabilize Tyr 259 conformation. Both copies of the cleaving form have different conformations,and are too far for a direct phosphate or base interaction. In the non-cleaving conformer G263R has poor side chain density, which suggests multiple orientations.
N317T (Fig 10D)	↓	Yes	Yes	**Important to orient catalytic nucleophile Tyr 324.** The side chain of the T 317 forms an additional hydrogen bond to the phosphate backbone in the central DNA region (Ade 2) that is not present in wt Cre. This may help correctly orient the active site residues in the cleaving conformer. The N317T also prevents the hydrogen bond between N 317 and N 319 that occurs in wt Cre. Consequently, in Tre, Asn 319 forms a weak polar interaction with T hr 316 of the adjacent monomer. in contrast, in wt Cre, the monomers are 1.5 A farther apart from each other at this location and cannot interact.
I320S (Fig 10D)	↓	Yes	Yes	**Important to orient catalytic nucleophile Y324.** S 320 is adjacent to important catalytic residues Y324, W315. S320 is more polar and smaller than I. In the cleaving monomer, S 320 hydroxyl interacts with a phosphate backbone (Ade 2) and the hydroxyl of the N317T of the neighboring protomer. These may help correctly orient active site of protein in the cleaving conformer.

Supplementary Table 2. Structural and Functional analysis of Tre mutations. The first column indicates the Cre-to-Tre mutation. A relevant figure in the manuscript is indicated in parentheses. The second column summarizes the mutagenesis data shown in Figure 2. Arrows indicate whether mutation back to wildtype greatly decreased (fat downward arrow), decreased (thinner downward arrow), or greatly increased (fat upward arrow) its ability to recombine *loxLTR* relative to Tre. The remaining columns indicate whether the residue contacts DNA or an adjacent protein monomer, and a brief written summary.

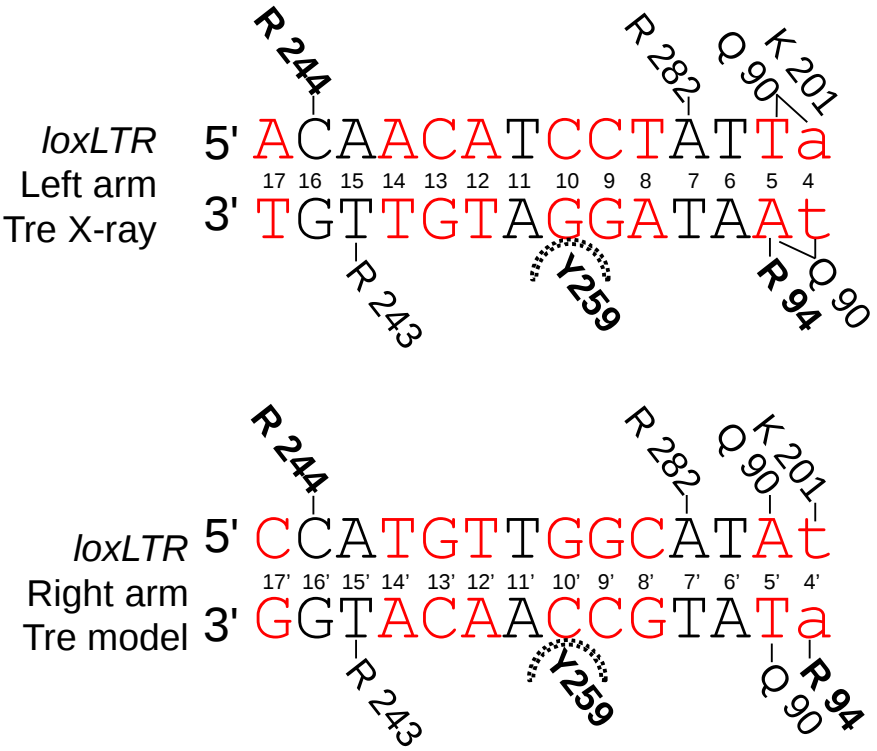
Supplementary Figure 3



Supplementary Figure 3. Protein/DNA interaction of Tre/*loxLTR*. Tre/*loxLTR* interaction diagram. The *loxLTR* sequence is numbered according to standard *loxP* numbering scheme. The bases in red differ from *loxP*. The yellow region indicates the location of the 8 bp central spacer region. The scissile phosphates are shown as solid black spheres. The residues that are mutated relative to Cre are shown in yellow circles. The blue lines indicate polar interactions, the red lines hydrophobic interactions. Lines extending into the boxes of the bases indicate a base specific interaction, lines touching the circles indicate a phosphate interaction. As there are two copies each of the cleaving and non-cleaving conformers of Tre in the asymmetric unit, if an interaction was observed in either copy, it was included here. Residues in black text were observed making the specified interaction in both copies of Tre whereas residues in grey text were observed in only one copy of Tre. The program PDBSUM was used as the basis for this schematic.

Supplementary Figure 4

Interaction diagram of cleaving competent Tre conformers with *loxLTR* arms



Supplementary Figure 4. Schematic Diagram comparing the "left" and "right" protein-DNA interactions in Tre cleaving-competent conformers. The duplex sequences of the each "arm" of the *loxLTR* are shown in the 5'->3' direction, and labeled "left" and "right". Nucleotides that are identical between the arms are colored black. Nucleotides that are not conserved are colored red. Interactions depicted one the left arm are observed in the cleaving-conformer in the Tre/*loxLTR* X-ray structure. To visualize the putative interactions of a cleaving-competent Tre conformer interacting with the "right" arm, a 3D model was generated. The right arm 3D model was made by using the active Tre conformer and the activated DNA backbone, but altering bases to match the right arm sequence, then analyzing the resultant Tre-DNA interactions. This model was used to analyze the interactions depicted in the right arm model scheme. Amino acids that make important base specific contacts are indicated, residues that are mutated relative to Cre are shown in bold. Dashed lines indicate van der Waals interactions.