

Engineering of a target site-specific recombinase by a combined evolution- and structure-guided approach

Josephine Abi-Ghanem¹, Janet Chusainow^{2,3}, Madina Karimova^{2,3},
Christopher Spiegel^{2,3}, Helga Hofmann-Sieber⁴, Joachim Hauber⁴,
Frank Buchholz^{2,3,*} and M. Teresa Pisabarro^{1,*}

¹Structural Bioinformatics, BIOTEC TU Dresden, Tatzberg 47-51, 01037 Dresden, ²University Carl Gustav Carus and Medical Faculty, TU Dresden, Fetscherstrasse 74, ³Max-Planck Institute for Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden and ⁴Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Martinistrasse 52, 20251 Hamburg, Germany

Received August 24, 2012; Revised and Accepted November 15, 2012

ABSTRACT

Site-specific recombinases (SSRs) can perform DNA rearrangements, including deletions, inversions and translocations when their naive target sequences are placed strategically into the genome of an organism. Hence, in order to employ SSRs in heterologous hosts, their target sites have to be introduced into the genome of an organism before the enzyme can be practically employed. Engineered SSRs hold great promise for biotechnology and advanced biomedical applications, as they promise to extend the usefulness of SSRs to allow efficient and specific recombination of pre-existing, natural genomic sequences. However, the generation of enzymes with desired properties remains challenging. Here, we use substrate-linked directed evolution in combination with molecular modeling to rationally engineer an efficient and specific recombinase (sTre) that readily and specifically recombines a sequence present in the HIV-1 genome. We elucidate the role of key residues implicated in the molecular recognition mechanism and we present a rationale for sTre's enhanced specificity. Combining evolutionary and rational approaches should help in accelerating the generation of enzymes with desired properties for use in biotechnology and biomedicine.

INTRODUCTION

Applied site-specific recombination has become an important technology to precisely manipulate the genome in a broad range of organisms (1,2). Especially the Cre/loxP system has found widespread utility because of its efficacy and ease of use (3). Cre originates from the bacteriophage P1 and efficiently and specifically recombines its DNA target site loxP, which is composed of two 13-bp inverted repeats separated by an 8-bp spacer (Figure 1A). Importantly, recombination occurs in heterologous hosts without the aid of accessory proteins (3). The Cre/loxP system and various recombination intermediates have been intensively studied structurally (4).

Because of the simplicity of its enzymatic reaction, the Cre/loxP system has become a very popular tool for manipulating DNA *in vitro* and *in vivo* (2,5). In particular, the development of genome engineering in living organisms has made the Cre/lox system an invaluable instrument for advanced genetic studies. By introducing lox sites strategically into the genome of model organisms, experiments such as conditional gene knockouts (6,7), site-specific integration (8,9), induced chromosomal translocations (10,11), lineage tracing (12,13) and recombinase-mediated cassette exchange have become possible (14).

A limitation of the technology is that the recombinase target site has to be introduced into the genome of the organism of interest. To overcome this limitation, numerous studies have aimed at redesigning well-characterized recombinases so that they recognize new

*To whom correspondence should be addressed. Tel: +49 351 463 40288; Fax: +49 351 463 40289; Email: frank.buchholz@tu-dresden.de
Correspondence may also be addressed to M. Teresa Pisabarro. Tel. +49 351 463 40071; Fax: +49 351 463 40087; Email: mayte@biotec.tu-dresden.de

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

target sequences (15). This comes as an important step in broadening the use of these recombinases, so that their application can be extended to predefined DNA target sites (16,17). Different methodologies have emerged, including substrate-linked directed evolution (18) or targeted mutagenesis with positive selection (19), to generate recombinases recognizing non-native genomic DNA sequences. Recently, an extensive substrate-linked directed evolution strategy has been used to evolve the recombinase Tre from Cre. Tre recognizes a sequence, loxLTR, which is present in the long terminal repeat region of a primary HIV-1 isolate and shares 50% sequence identity with loxP (Figure 1A). Expression of Tre in HIV-1-infected human cells leads to the eradication of the integrated provirus (20), thereby offering a new strategy for a possible cure for AIDS (21).

Tre was selected by experimentally testing hundreds of individual clones randomly picked after more than 120 cycles of evolution (20). This was necessary because the evolved library is a diverse mix of clones with different levels of activity and specificity. Hence, a large number of enzymes needed to be examined for their recombination properties in order to identify an appropriate clone. Thus, the *in vitro* evolution process is tedious and time-consuming because it requires many cycles (>100) of evolution and screening of a large number of individual clones to identify an enzyme with acceptable activity and specificity for a new DNA target site.

Here, we investigate whether sequence information collected during the course of evolution together with available high-resolution structures of the Cre/loxP system can be exploited to rationally target specificity. We first continued the evolution cycles to enrich for conserved mutations. We then used a rational approach to establish the determinants of specificity found in the evolved Tre/loxLTR system. We present the rationally engineered active and specific recombinase sTre, which was rapidly generated through the identification of conserved residues conferring robust enzymatic activity, followed by structure-based modeling and simulation to identify key residues providing target-site specificity.

MATERIALS AND METHODS

Directed evolution and library evaluation

The Tre library was generated as described previously (20) with minor modifications (see Supplementary Data). Briefly, for each evolution cycle, recombinase coding sequences encoding active enzymes were PCR amplified and cloned back into the original evolution vector. After every third evolution cycle, DNA shuffling was performed to combine enriched mutations (18). Tre clones (163) with differing recombination activities were selected and sequenced. The sequences were aligned to Cre in order to map all mutations and to identify evolutionary conserved mutations in the evolved library. To generate the described mutants, site-directed mutageneses were performed using the QuikChange[®] Site-Directed Mutagenesis Kit (Stratagene) following the manufacturer's instructions.

Recombination assays and specificity tests

To assay the recombination efficiency of Tre recombinase clones, plasmid DNA was isolated from L-arabinose-induced overnight cultures in *Escherichia coli* or by recombination assays in human cells and investigated as described in Supplementary Data.

Structure-based modeling and molecular dynamics

Molecular dynamics (MD) of the Cre/loxP complex [1Q3U; resolution 2.9 Å (22)] was performed using AMBER10 (23) and the Parm99SB force field (24) (see Supplementary Materials and Methods for details). Tre14 and sTre were modeled with Modeller in Discovery Studio (Accelrys) using 1Q3U as template. The models were refined as described earlier and in Supplementary Data for the wild-type (WT) complex. Tre14 was simulated for 100 ps, and the complexes Tre14/loxP and Tre14/loxLTR for 400 ps applying distance restraints between the catalytic tyrosine and the phosphate being cleaved. The backbone of the N-terminus and C-terminus helices was constrained (5.0 kcal/mol/Å² force constant). loxLTR (13 bp) was constructed as canonical DNA in nucgen. We applied restraints based on the TRX scale (25) and followed a previously established protocol (26) with 15 ns MD. Analyses of hydrogen bonds were done using HBPLUS (27) and plotted using nuclplot (28). We established a criterion to consider dynamic formation of hydrogen bonds along the simulation time in which a hydrogen bond was considered formed when found >10% of the total simulation time. Simulation analysis and image rendering were performed with tachyon in VMD (29).

RESULTS

Generation of a consensus recombinase

During substrate-linked directed molecular evolution, complex clone libraries are generated through continuous cycles of PCR mutagenesis and DNA shuffling. In order to obtain data on the complexity and variability of a library after an extended evolution process, we performed 13 additional cycles (20) to enrich for conserved mutations (Supplementary Table S1) and isolated 288 random clones from the Tre library evolution cycle 139 to test their individual recombination activities on loxLTR in *E. coli* by a PCR-based assay. The recombination rates of these clones varied largely, ranging from no detectable to competent recombination (Supplementary Figure S1). When testing a selection of active clones in restriction enzyme-based recombination and specificity assays, the evolved Tre library was found to be a diverse mix of clones with differing levels of activity and specificity. To investigate the complexity of the recombinase pool in more detail, we sequenced 163 randomly selected clones of evolution cycle 139 and mapped the obtained mutations onto the original Cre sequence (Figure 1C). On average, 23 ± 3 mutations per Tre clone were observed, with a minimum of 16 and a maximum of 32 mutations. Strikingly, while some mutations were repeatedly found in many clones,

none of the sequenced recombinases was identical (Supplementary Table S2). Hence, isolation of the best enzyme from a complex pool of extensively evolved recombinases is not straightforward. As a consequence, many individual clones have to be tested to identify an enzyme with desired features. In order to shorten this laborious process of finding the best clones, we decided to investigate the applicability of a more rational approach.

Mutations that significantly improve fitness in a population of clones become fixed with time during the evolution process (30). We reasoned that the generation of a clone that only carries the repeatedly identified mutations might result in a highly active enzyme. We identified 14 mutations by using an 85% conservation threshold (Figure 1C) and we generated a consensus recombinase, designated Tre14, which exclusively contained these mutations. Tre14 efficiently recombined the target site loxLTR with a recombination rate that matched or exceeded the individually tested clones from the original library (Figure 1B; Supplementary Figures S1 and S2). We also tested the contribution to loxLTR recombination of each of these 14 mutations by back-mutating them to the original amino acid found in Cre. Strikingly, most of these mutants showed reduced activity on loxLTR, while no activity change was observed on loxP (Supplementary Figure S3). These data suggest that the generation of a consensus recombinase is indeed an efficient way to generate a highly active enzyme from a pool of sequenced clones.

During the evolution cycles, little selection pressure is enforced on the specificity of the enzyme. Indeed, target promiscuity has been widely seen in directed evolution studies (31). To investigate the recombination specificity of Tre14, we tested the enzyme on the target site of the Cre-related recombinase Dre (32). The recognition site of Dre, rox, shares similarity to loxP and loxLTR (Figure 1A) and it is not present in the *E. coli* genome. Hence, it can serve to probe the overall specificity of Tre14. Expression of Tre14 in a Dre reporter plasmid did not show any recombination (Figure 1B), indicating that Tre14 does not recombine rox sites. Therefore, extended substrate-linked directed evolution of site-specific recombinases does not lead to the generation of enzymes with widely relaxed target specificity. We also tested Tre14 on the original Cre DNA target site, loxP. Tre14 recombined the loxP sites with recombination rates comparable to the rates measured on loxLTR (Figure 1B) and is therefore more promiscuous than the previously reported Tre recombinase (20). Thus, Tre14 maintained activity on the target site that served as the starting point before carrying out the substrate-linked directed evolution process. We conclude that the generation of a consensus recombinase containing all highly abundant mutations emerging during the evolution process yields an efficient recombinase with activity on both, the original target site (loxP) and the new target site (loxLTR).

Building atomic 3D models for a consensus recombinase and its substrates

Before addressing the improvement of target site specificity, we first investigated the role of the 14 conserved

mutations in Tre14 based on an existing Cre/loxP co-crystal structure (22). We performed a MD simulation in order to account for all possible protein–DNA interactions that could potentially be established in addition to the ones observed in this template and other available crystal structures in the Brookhaven Protein Databank (see ‘Materials and Methods’ section and Supplementary Figure S4).

After investigating the mutations on the crystal structure of Cre/loxP, we generated a model of Tre14 by using the crystal structure as template. After energy refinement, the Tre14 model presented an $\text{RMSD}_{\text{C}\alpha}$ of $2.2 \pm 0.1 \text{ \AA}$ with respect to the template. In order to model the complex structure of Tre14 with loxP, we manually docked the loxP structure of the crystallographic complex (1Q3U) onto the Tre14 model. The obtained complex was refined without restraints, and the resulting structure presented an $\text{RMSD}_{\text{C}\alpha}$ of $1.7 \pm 0.1 \text{ \AA}$ for Tre14 and RMSD of $2.0 \pm 0.3 \text{ \AA}$ for all atoms of loxP, indicating only minor structural differences with respect to the template crystal structure. Subsequently, we investigated possible structural features of the LoxLTR target site. Then, the Tre14/loxLTR complex was energy minimized applying distance restraints between the catalytic tyrosine and the phosphate being cleaved ($\text{RMSD}_{\text{C}\alpha} = 1.5 \pm 0.3 \text{ \AA}$).

The detailed analysis of the Cre/loxP, Tre14/loxP and Tre14/loxLTR structures allowed us now to investigate the mutations in Tre14 in an attempt to obtain mechanistic insights to rationally improve its DNA-binding specificity.

Analysis of Tre14/loxP and Tre14/loxLTR 3D models

In order to understand the impact of the 14 conserved mutations in Tre, we analyzed in detail each of the residues in the Tre14/loxP and Tre14/loxLTR models and the corresponding WT residues in the Cre/loxP complex (Figure 1D). Based on these 3D models and their behavior in our MD simulations, we were able to distinguish three distinct groups among the 14 mutations.

Group I (non-DNA contacting): V7I, P12S, P15L, Y77H

The first 20 amino acids of the protein sequence were found mutated with a high frequency (Figure 1C). This part of the protein is not resolved in any of the available crystal structures, most likely because it is flexible and not structured (4). As recently reported, an amino-terminal deletion mutant (first 12 residues) of Cre retains its recombination activity with similar kinetics to the WT protein (33). We therefore assume that the conserved mutations at the N-terminus (V7I, P12S and P15L) do not have a direct effect on the substrate recognition of the recombinase, although they might, for instance, play an important role in the stability of the evolved protein. The conserved mutation Y77H is located in helix C and is not directly involved in the protein/DNA interface or protein/protein interactions. This residue does not seem to have an obvious effect on specificity or activity, but we cannot rule out that it may be important for the general structure of the evolved protein or it perhaps has a role on

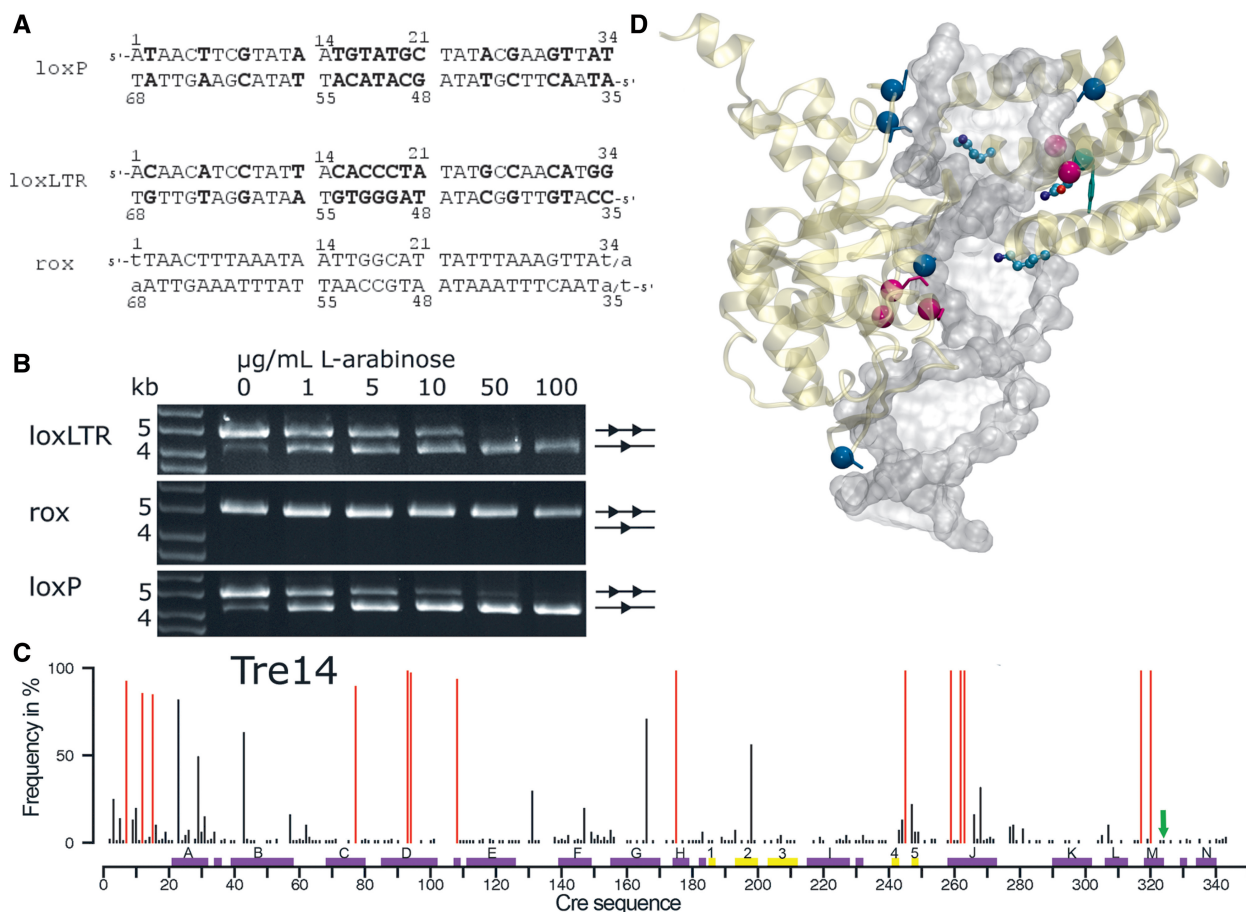


Figure 1. Tre14 recombination sites, activity and evolutionary mutations. (A) loxP, loxLTR and rox recombination sites alignment. Bases differing in loxP and loxLTR are shown in bold. (B) Recombination efficiency and specificity of Tre14 on loxLTR, rox and loxP. Recombinase expression was induced with L-arabinose at indicated concentrations. Recombination was assayed by restriction enzyme digest, resulting in a smaller fragment for recombined (one triangle) and a larger fragment for non-recombined substrate (two triangles). 1 kb Ladder marker is shown on the left. (C) *In vitro* evolution mutations obtained with Tre and their respective frequencies are mapped onto the Cre sequence. Cre α -helices and β -sheets are shown (purple and yellow blocks, respectively). The catalytic tyrosine is shown as a green arrow, and in red the mutations present in Tre14. (D) Tre14 mutations shown as spheres on C α atoms and licorice in side chains are mapped onto the Cre structure 1Q3U (22). Cre protein (yellow ribbon) is bound to one loxP arm (gray surface). Group I is shown in green, Group II in blue and Group III in purple. Residues K43, K86 and Q94 are shown in CPK.

specificity and/or activity through long-range electrostatic interactions.

Group II (DNA backbone interacting): S108G, A175S, N245Y, E262Q, N317T, I320S

The structure of the DNA backbone is sequence-dependent, i.e. the phosphate conformation depends on the nature of each dinucleotide step (25,34,35). Therefore, the sequence variation between the loxP and the loxLTR target sites confers a slightly different structure to its backbone (Supplementary Figure S5). We predict that the altered residues in Tre14 that contact the DNA backbone give the protein, on one hand, the flexibility to adapt slightly different structures. For instance A175S, which leads to the appearance of a hydrogen donor, could strengthen the interaction towards loxP and loxLTR (Supplementary Figure S6). On the other hand, mutations that decrease interaction with the DNA might offer fewer restraints to the protein–DNA interaction. For example, S108G leads to the absence of side chain and therefore no hydrogen donor. This could preclude hydrogen bond interaction between Tre14 and loxLTR,

which could give loxLTR a less restrictive environment for binding in this region and, as a consequence, could increase the conformational space available for the positioning of the substrate. In all available Cre/loxP structures, E262 is in close contact with the phosphates of the DNA and it has been shown previously to be essential for specificity in the Cre/loxP complex. This residue was extensively studied and its mutation to glutamine allows Cre to recombine divergent targets with increased activity, including the WT loxP sequence (36,37). Thus, we hypothesize that this group of mutations confers plasticity to Tre14 to recombine divergent DNA targets consisting of different structure profiles.

Group III (DNA major groove contacting): G93C, Q94R, R259Y, G263R

The mutations of Group III are located in helices D and J, which are the only secondary structure elements of the protein in contact with the major groove in both loxP and loxLTR.

Arginine is one of the most probable amino acids found establishing interactions with the major groove of

DNA (28). Residue R259 in Cre specifically interacts with the major groove of loxP (38), namely with bases G₂₇ and G₆₁. In loxLTR, a C₂₇ or G₆₁ (Figure 1A) will be in contact with the protein residue in position 259. A cytosine possesses only a hydrogen bond donor, while a guanine has two acceptors. In our model, we observe that the mutation of R259 to a tyrosine at this position (R259Y) allows interactions with both bases, because tyrosine can act as a hydrogen donor and also as an acceptor. In the case of the base G₆₁, Y259 will assume a hydrogen bond donor role, whereas in the case of base C₂₇ a hydrogen bond acceptor. In the Tre14/loxLTR model, we observed that the elongation of the amino acid side chains in mutations G263R and Q94R implies the appearance of five hydrogen bond donors in each arginine side chain, which promote the formation of interactions with the backbone and also with the bases in the major groove of loxLTR and, by doing so, they may strengthen the binding to the new DNA substrate by a direct or indirect readout. In contrast to Tre14/loxLTR, in our Tre14/loxP model arginine at position 94 does not contact bases in the major groove and only interacts with the DNA backbone (Figure 2A).

Residue G93 and its mutation G93C are disposed at a van der Waals (vdw) distance to the DNA. However, we did not observe any hydrogen bond contacts to the DNA in the Cre/loxP crystallographic complex, the Tre14/loxP and Tre14/loxLTR models. Furthermore, cysteine is one of the less probable residues to interact with DNA (28). Hence, the mutation to cysteine (G93C) is most probably not directly involved in contacting the DNA. However, the G93C mutation always appeared simultaneously with Q94R (Figure 1C). Both residues are located in helix D and, since glycine is noted as an amino acid with intrinsic propensities to break α helices (39), we predict that the mutation G93C might have a structural role by circumventing possible breaking of the helix.

Thus, this group of mutations highlights key residues that are selected during the evolution process to aid Tre14 to interact with loxLTR. Nevertheless, these mutations still retain some physico-chemical features that allow them to achieve binding to loxP.

In conclusion, during the evolution process key residues emerged that adapt to the new conformation of the DNA backbone and the different DNA bases. However, all mutated residues adopt a versatile interaction being able to accommodate both loxP and loxLTR. Hence, these observations explain the relaxed specificity of Tre14 and establish the bases for its rational improvement.

Rational engineering of target site specificity (sTre)

Specific interactions within a protein/DNA complex are established in most cases through direct contact of protein residues with the bases in the major and minor grooves of the DNA or through the indirect readout (structure of the backbone of the DNA) (40,41). In our case the major groove, which confers a unique profile of hydrogen bond donor/acceptor groups for each base pair, was chosen as the most appropriate to direct specificity through it.

We first investigated possible residues only essential for loxP recombination in order to design a Tre-based enzyme that specifically recombines loxLTR without activity on loxP. For this we established a specificity profile for Cre/loxP by screening the hydrogen bond interactions occurring in the major and minor grooves of the DNA in a 50-ns MD simulation. As depicted in Figure 3, there are six residues in Cre (H40, K43, K86, Q90, R259 and N317) that specifically contact the major groove of loxP. Two of these residues, K86 and K43, are the only ones interacting in a dissimilar region in our Tre14/loxLTR model (Figure 2B and C). In loxP, K43 binds to both half sites (T₁₀, T₅₈ and T₂₄, T₄₄) (see Figure 1A for base numbering). A negatively charged amino acid in this position might affect the recombination activity. At the same time, this residue change might allow the interaction with the C tracts present in loxLTR (C_{8p}C₉; C_{26p}C₂₇ and C₄₄) (Figure 1A). A cytosine in the major groove contains only a hydrogen bond donor, which would be a suitable atom to bind a negatively charged amino acid. In fact, the mutation K43E already appeared during the evolution process (Figure 1C), indicating its conformance with loxLTR recognition. Residue K86 has been shown to play an important role in the Cre/loxP complex by being implicated in establishing the site of initial strand exchange (42). Based on the X-ray structure of Cre/loxP and our models, K86 is shown to be interacting in the major groove of the DNA with A₁₃ or A₃₇ and G₄₈ (Figures 2C, 3 and 4). We hypothesized that a mutation of K86 into a negatively charged amino acid could result in a loss of activity of the enzyme on both loxP and loxLTR target sites due to the weakness of interaction in this region. In this scenario, if we intent to impair activity on loxP but not on loxLTR, i.e. gain specificity to loxLTR, another residue would be needed to interact with loxLTR in order to take over the role of K86. We therefore analyzed all base-specific interactions of Tre14 to loxLTR in our models. Interestingly, the mutation Q94R allows the protein to establish direct contacts to the base A₅₆ in the major groove of loxLTR (Figure 4C). In contrast, our Tre14/loxP model indicates that R94 is not implicated in binding to bases in the major groove but only interacts with the DNA backbone (Figure 4A and B). The methyl group of the thymine T₂₂ and the lack of hydrogen bond acceptors on the C₂₁ in the major groove of loxP sterically and electrostatically hinder R94 to bind to the major groove. On the other hand, in the Tre14/loxLTR model, we observed that R94 is positioned in the vicinity of adenosine A₅₆, where it is involved in a base-specific interaction (Figure 4C). More importantly, R94 binds the major groove in the vicinity of K86 in loxLTR. Hence, R94 might be able to take over the role of K86 when this residue is mutated to glutamate and thus provides specificity toward loxLTR.

Based on these observations, we hypothesized that a newly engineered recombinase containing the mutations K43E and K86E (Tre14_{K43E_K86E}; referred from here on as sTre) could efficiently recombine loxLTR, while loxP should not be a preferred substrate.

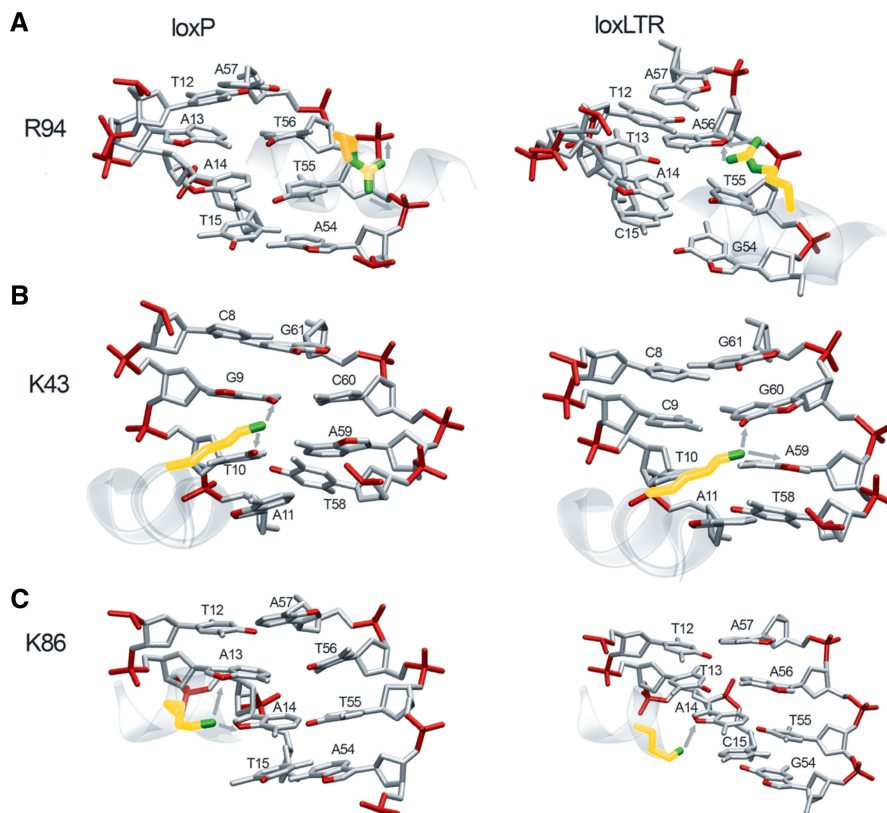


Figure 2. Interactions of Tre14 residues R94, K43 and K86 with loxP and loxLTR. Predicted interactions are indicated with arrows. (A) R94: arginine can interact only with the backbone in loxP, and also with the bases in loxLTR. (B) K43: lysine can recognize both substrates. In loxP, a negatively charged amino acid could impair binding, but the presence of the cytosine in loxLTR can mediate interactions. (C) K86: lysine can recognize both substrates. A negatively charged residue would impair binding in both cases.

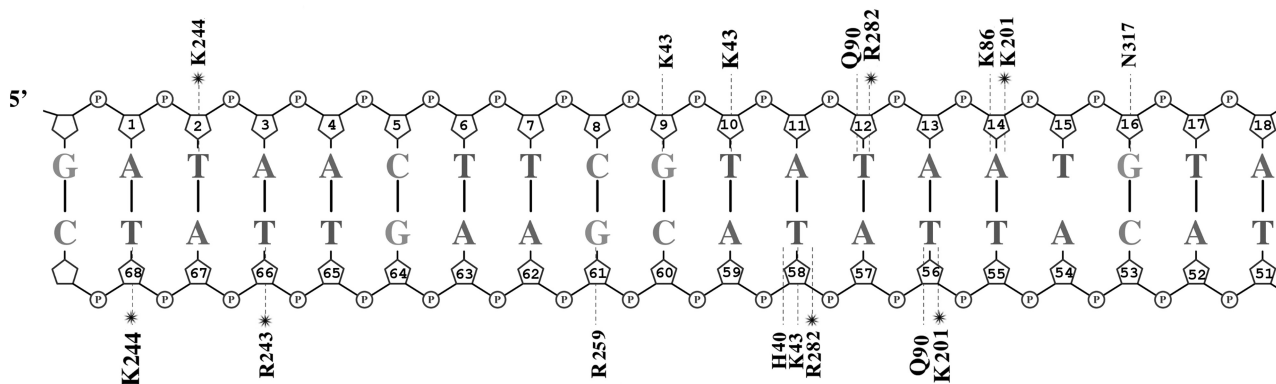


Figure 3. Projection of hydrogen bonds between each of the 13-bp half-sides of loxP and Cre based on MD simulation. Residue numbers are shown. The interactions between the two inverted repeats are symmetric. Asterisks indicate residues interacting in the DNA minor groove, while the rest interact within the major groove.

Experimental validation of the enhanced properties of sTre

To test whether our structure-based prediction could indeed improve specificity, we generated recombination reporter constructs for sTre and tested its recombination activity on loxLTR and loxP in *E. coli* employing two different assays. In the first assay, the recombinase was expressed from a plasmid that also contained the

respective recombinase recognition sites, loxLTR or loxP. Like Tre14, sTre recombined loxLTR with high efficiency, approaching a recombination rate of 100% when sTre expression was induced with L-arabinose at 50 µg/ml or higher. Recombination was already detected at L-arabinose concentrations of as little as 1 µg/ml (Figure 5A), indicating activity of the enzyme even at very low expression levels. These data also demonstrate the improved

performance of sTre in comparison to the original Tre recombinase (20) (Figure 5A). In contrast to Tre14 however, sTre did not recombine loxP, demonstrating that it has significantly gained specificity to the preferred target site loxLTR. In the second assay in *E. coli*, a lacZ-based reporter assay was employed. Recombination by Tre in cells containing both the Tre expression plasmid and the lacZ reporter plasmid removes the promoter driving lacZ expression. Hence, colonies harboring the recombined reporter plasmid appear in white on Xgal plates, whereas colonies that contain the reporter plasmid in its original form display blue staining on the same plates (Figure 5B). Consistent with the previous

results, we noted an obvious switch in specificity of sTre in comparison to Tre14 (Figure 5B). Importantly, sTre activity could also be demonstrated in human cells in a transient co-transfection experiment (Supplementary Figure S7A) and in an assay to remove an integrated HIV-1 provirus (Supplementary Figure S7B). In summary, we conclude that sTre containing the mutations K43E and K86E recombined loxLTR efficiently and with largely improved specificity.

After obtaining sTre and establishing the role of K43 and K86 in both loxP and loxLTR recombination, we next experimentally investigated the role of the mutation Q94R. We mutated R94 to alanine, which does not

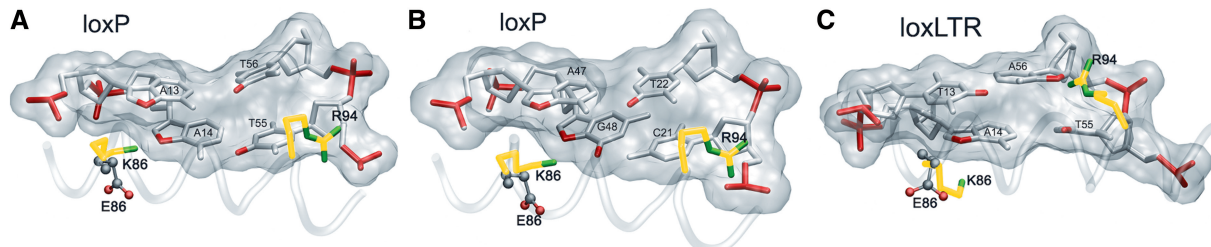


Figure 4. Location of mutations K86E and Q94R on Tre14/loxP and Tre14/loxLTR models. Helix D of Tre14 (gray ribbon) is shown bound to: (A) loxP right half-site, (B) loxP left half-site and (C) loxLTR right half-site. Hydrogen bond acceptors are shown in red and donors in green. Nucleotides are labeled and residues K86 and R94 are shown in licorice (yellow) and E86 in CPK.

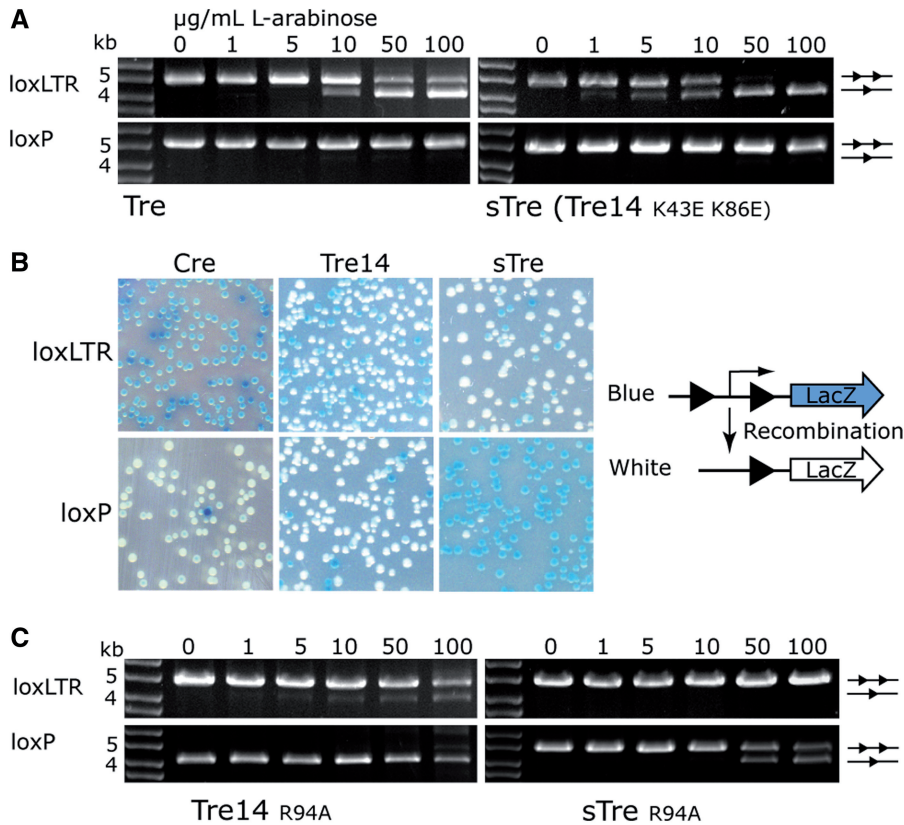


Figure 5. Recombination efficiency and specificity of sTre and effect of mutations K43E, K86E and R94A. Recombination efficiency at indicated induction conditions with L-arabinose at different concentrations is shown for sTre and Tre (20) in (A) and for Tre14_{R94A} and sTre_{R94A} in (C). In (B), the LacZ-based recombination reporter assay in *E. coli* is shown together with a scheme of the assay. Colonies grown on LB plates stained with X-gal for Cre, Tre14 and sTre are shown.

contain hydrogen bond donor groups. We generated R94A mutants of both Tre14 and sTre (Tre14_{R94A} and sTre_{R94A}) and tested these clones for their recombination activities on loxLTR and loxP. In contrast to Tre14, which recombined both loxLTR and loxP (Figure 1B), Tre14_{R94A} showed a marked decrease in the rate of recombination on loxLTR, whereas recombination activity on loxP was high (Figure 5C). Unlike sTre, which efficiently recombined loxLTR (Figure 1B), sTre_{R94A} showed a total loss of recombination on loxLTR (Figure 5C), demonstrating the importance of R94 for the recombination of this target site. We also tested whether the mutations K43E, K86E and Q94R alone are sufficient to switch the specificity from loxP to loxLTR. However, no recombination activity of this enzyme was seen on loxLTR, indicating that the other amino acid changes are also required to allow loxLTR recombination (data not shown). Hence, these data confirm our predictions and establish a crucial role of Q94R for loxLTR recombination. These experimental results also validate our assumption that K43 and K86 could be key players in the recombination in the Cre/loxP system, whereas R94 is selectively essential for the recombination activity of sTre on loxLTR (Figure 6 and Supplementary Figure S8).

In conclusion, our results demonstrate that through a combined approach linking *in vitro* evolution with modeling and simulation techniques, we have been able to elaborate a rationale on the role of mutations selected by *in vitro* evolution on activity and specificity. Furthermore, combining evolutionary and rational engineering approaches has been key to engineer a new recombinase, sTre, which efficiently and specifically recombines a sequence that is present in an HIV-1 LTR.

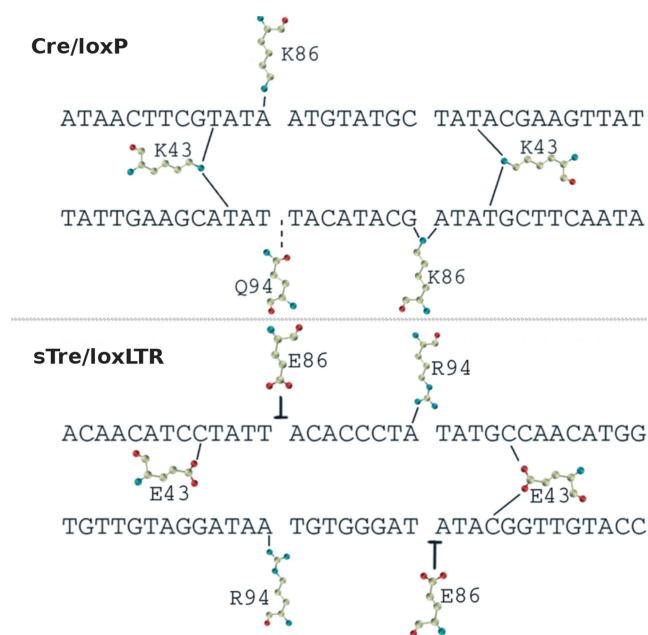


Figure 6. Rationale for specificity and role of K43E, K86 and Q94R on targeted recombination on loxP and loxLTR. Solid lines show hydrogen bonds and dashed lines van der Waals interactions. The loss of hydrogen bond with K86E mutation is represented by an inhibition symbol (\perp).

DISCUSSION

Directed molecular evolution is an elegant strategy to generate proteins with altered properties, including changed DNA-binding specificities. This approach does not rely on pre-existing knowledge and uses the forces of iterative rounds of mutation and screening/selection to evolve an enzyme with desired function. While the success of this approach is remarkable, it represents an arduous path to achieve the desired enzyme properties in an acceptable number of evolution cycles.

Some DNA-recognizing proteins are now understood in much detail, including the zinc finger domain (43) and the TAL effector central repeat domain (44). However, these domains remain the exception from the many proteins that specifically bind DNA sequences. Hence, the rational design of proteins with predefined DNA binding properties remains a distant goal. Key to achieving this goal is the understanding at the atomic level of all properties within the protein and substrate that contribute to an interaction. Access to 3D structures of enzymes and their substrates provides a rich starting point to investigate how proteins recognize DNA (45). However, engineering proteins to specifically bind a DNA sequence remains difficult because very small changes in structure may have big effects on DNA recognition (46). Recently, directed evolution techniques have been used as a tool to enhance computationally derived designs (47–49). These examples show that directed evolution following rational engineering can accelerate the generation of proteins with desired properties. However, this approach is dependent on the modeling-based pre-selection of positions that might alter specificity. This early restriction of residues to mutate may result in amplification of wrong choices as it may also affect activity. In our work we reveal that the strength of both methodologies, directed evolution and structure-aided computational design, can be boosted in a reverse approach where an extended molecular evolution is preceding and ensuring activity and the fine tuning on specificity is then provided by a structure-based rationale. Thus, in our approach, molecular evolution defines a new set of mutations from which an active consensus enzyme is derived and, then, molecular modeling and simulation is performed to establish a rationale for the evolution process, which allows rational design of new mutations that achieve the desired specificity profile. The development of sTre, which specifically recognizes loxLTR and does not cross-recombine loxP, by coupling substrate-linked directed evolution to structure-based modeling and simulation, demonstrates that proteins with complex DNA-binding behavior can be generated with relative ease. In this way, modeling complements the directed evolution process, standing as a suitable shortcut to gain insights into the mechanism driving specificity. A noteworthy aspect to be considered when attempting design of recombinase specificity, and likely other enzymes, is the fact that the investigations should not only focus on concrete residues that may be considered indispensable for specificity. The vicinity of these particular residues should also be explored to discover others with a potential to overtake this role.

Considering these ‘compensative substitutions’ may provide insights into the functionality of specific residues and how to rationally make them dispensable. In general, with the development of sTre, our results underline the great potential that structure-based computational methods represent in order to understand molecular recognition through evolution processes and to rationally design specificity for large and complex biological systems in an efficient manner.

Engineered recombinases have been proposed for future medical applications. These include novel antiviral strategies (20) and site-specific delivery of DNA for gene therapy approaches (50). For such applications these enzymes have to be efficient and highly specific to avoid unintended genetic alterations to the host genome. The combined evolutionary and rational approach presented here should help to develop such safe enzymes useful for therapeutic applications. Obviously, the engineered enzymes will have to be thoroughly tested before their clinical application, but without side effects these enzymes could be employed for *in vivo* DNA surgery.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2, Supplementary Figures 1–8, Supplementary Materials and Methods and Supplementary References [51–58].

ACKNOWLEDGEMENTS

We are grateful to Ralf Gey for technical support, to the ZIH at the Technische Universität Dresden for computational resources and assistance and to the members of the Buchholz and Pisabarro groups for helpful discussions.

FUNDING

Klaus Tschira Stiftung gGmbH (to J.A.-G.). Funding for open access charge: Institutional (University).

Conflict of interest statement. None declared.

REFERENCES

1. Glaser, S., Anastassiadis, K. and Stewart, A.F. (2005) Current issues in mouse genome engineering. *Nat. Genet.*, **37**, 1187–1193.
2. Kilby, N.J., Snaith, M.R. and Murray, J.A. (1993) Site-specific recombinases: tools for genome engineering. *Trends Genet.*, **9**, 413–421.
3. Nagy, A. (2000) Cre recombinase: the universal reagent for genome tailoring. *Genesis*, **26**, 99–109.
4. Van Duyne, G.D. (2001) A structural view of cre-loxP site-specific recombination. *Annu. Rev. Biophys. Biomol. Struct.*, **30**, 87–104.
5. Buchholz, F. and Bishop, M. (2001) LoxP-directed cloning: use of Cre recombinase as a universal restriction enzyme. *Biotechniques*, **31**, 906–908, 910, 912, 914, 916, 918.
6. Le, Y. and Sauer, B. (2000) Conditional gene knockout using cre recombinase. *Methods Mol. Biol.*, **136**, 477–485.
7. Wilson, T.J. and Kola, I. (2001) The LoxP/CRE system and genome modification. *Methods Mol. Biol.*, **158**, 83–94.
8. Hirano, N., Muroi, T., Takahashi, H. and Haruki, M. (2011) Site-specific recombinases as tools for heterologous gene integration. *Appl. Microbiol. Biotechnol.*, **92**, 227–239.
9. Wirth, D., Gama-Norton, L., Riemer, P., Sandhu, U., Schucht, R. and Hauser, H. (2007) Road to precision: recombinase-based targeting technologies for genome engineering. *Curr. Opin. Biotechnol.*, **18**, 411–419.
10. Buchholz, F., Refaeli, Y., Trumpp, A. and Bishop, J.M. (2000) Inducible chromosomal translocation of AML1 and ETO genes through Cre/loxP-mediated recombination in the mouse. *EMBO Rep.*, **1**, 133–139.
11. Collins, E.C., Pannell, R., Simpson, E.M., Forster, A. and Rabbitts, T.H. (2000) Inter-chromosomal recombination of Mll and Af9 genes mediated by cre-loxP in mouse development. *EMBO Rep.*, **1**, 127–132.
12. Livet, J., Weissman, T.A., Kang, H., Draft, R.W., Lu, J., Bennis, R.A., Sanes, J.R. and Lichtman, J.W. (2007) Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature*, **450**, 56–62.
13. Zovein, A.C., Hofmann, J.J., Lynch, M., French, W.J., Turlo, K.A., Yang, Y., Becker, M.S., Zanetta, L., Dejana, E., Gasson, J.C. *et al.* (2008) Fate tracing reveals the endothelial origin of hematopoietic stem cells. *Cell Stem Cell*, **3**, 625–636.
14. Turan, S., Galla, M., Ernst, E., Qiao, J., Voelkel, C., Schiedlmeier, B., Zehe, C. and Bode, J. (2011) Recombinase-mediated cassette exchange (RMCE): traditional concepts and current challenges. *J. Mol. Biol.*, **407**, 193–221.
15. Buchholz, F. (2009) Engineering DNA processing enzymes for the postgenomic era. *Curr. Opin. Biotechnol.*, **20**, 383–389.
16. Akopian, A. and Marshall Stark, W. (2005) Site-specific DNA recombinases as instruments for genomic surgery. *Adv. Genet.*, **55**, 1–23.
17. Bolusani, S., Ma, C.H., Paek, A., Konieczka, J.H., Jayaram, M. and Voziyanov, Y. (2006) Evolution of variants of yeast site-specific recombinase F1p that utilize native genomic sequences as recombination target sites. *Nucleic Acids Res.*, **34**, 5259–5269.
18. Buchholz, F. and Stewart, A.F. (2001) Alteration of Cre recombinase site specificity by substrate-linked protein evolution. *Nat. Biotechnol.*, **19**, 1047–1052.
19. Santoro, S.W. and Schultz, P.G. (2002) Directed evolution of the site specificity of Cre recombinase. *Proc. Natl Acad. Sci. USA*, **99**, 4185–4190.
20. Sarkar, I., Hauber, I., Hauber, J. and Buchholz, F. (2007) HIV-1 proviral DNA excision using an evolved recombinase. *Science*, **316**, 1912–1915.
21. van Lunzen, J., Fehse, B. and Hauber, J. (2011) Gene therapy strategies: can we eradicate HIV? *Curr. HIV/AIDS Rep.*, **8**, 78–84.
22. Ennifar, E., Meyer, J.E., Buchholz, F., Stewart, A.F. and Suck, D. (2003) Crystal structure of a wild-type Cre recombinase-loxP synapse reveals a novel spacer conformation suggesting an alternative mechanism for DNA cleavage activation. *Nucleic Acids Res.*, **31**, 5449–5460.
23. Case, D.A., Darden, T.A., Cheatham, T.E. III, Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Crowley, M., Walker, R.C., Zhang, W. *et al.* (2008) AMBER 10. University of California, San Francisco.
24. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, **65**, 712–725.
25. Heddi, B., Oguey, C., Lavelle, C., Foloppe, N. and Hartmann, B. (2010) Intrinsic flexibility of B-DNA: the experimental TRX scale. *Nucleic Acids Res.*, **38**, 1034–1047.
26. Abi-Ghanem, J., Heddi, B., Foloppe, N. and Hartmann, B. (2010) DNA structures from phosphate chemical shifts. *Nucleic Acids Res.*, **38**, e18.
27. McDonald, I.K. and Thornton, J.M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.*, **238**, 777–793.
28. Luscombe, N.M., Laskowski, R.A. and Thornton, J.M. (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.*, **29**, 2860–2874.
29. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38, 27–38.
30. Rozen, D.E., de Visser, J.A. and Gerrish, P.J. (2002) Fitness effects of fixed beneficial mutations in microbial populations. *Curr. Biol.*, **12**, 1040–1045.

31. Aharoni, A., Gaidukov, L., Khersonsky, O., Mc, Q.G.S., Roodveldt, C. and Tawfik, D.S. (2005) The 'evolvability' of promiscuous protein functions. *Nat. Genet.*, **37**, 73–76.
32. Sauer, B. and McDermott, J. (2004) DNA recombination with a heterospecific Cre homolog identified from comparison of the pac-c1 regions of P1-related phages. *Nucleic Acids Res.*, **32**, 6086–6095.
33. Rongrong, L., Lixia, W. and Zhongping, L. (2005) Effect of deletion mutation on the recombination activity of Cre recombinase. *Acta Biochim. Pol.*, **52**, 541–544.
34. Gorenstein, D.G. (1994) Conformation and Dynamics of DNA and Protein-DNA Complexes by 31P NMR. *Chem. Rev.*, **94**, 1315–1338.
35. Heddi, B., Abi-Ghanem, J., Lavigne, M. and Hartmann, B. (2010) Sequence-dependent DNA flexibility mediates DNase I cleavage. *J. Mol. Biol.*, **395**, 123–133.
36. Gelato, K.A., Martin, S.S., Wong, S. and Baldwin, E.P. (2006) Multiple levels of affinity-dependent DNA discrimination in Cre-LoxP recombination. *Biochemistry*, **45**, 12216–12226.
37. Rufer, A.W. and Sauer, B. (2002) Non-contact positions impose site selectivity on Cre recombinase. *Nucleic Acids Res.*, **30**, 2764–2771.
38. 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.
39. Chakraborty, A., Schellman, J.A. and Baldwin, R.L. (1991) Large differences in the helix propensities of alanine and glycine. *Nature*, **351**, 586–588.
40. Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B. and Mann, R.S. (2010) Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.*, **79**, 233–269.
41. Seeman, N.C., Rosenberg, J.M. and Rich, A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA*, **73**, 804–808.
42. Lee, L., Chu, L.C. and Sadowski, P.D. (2003) Cre induces an asymmetric DNA bend in its target loxP site. *J. Biol. Chem.*, **278**, 23118–23129.
43. Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. and Gregory, P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.*, **11**, 636–646.
44. Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509–1512.
45. Gaj, T., Mercer, A.C., Gersbach, C.A., Gordley, R.M. and Barbas, C.F.III (2011) Structure-guided reprogramming of serine recombinase DNA sequence specificity. *Proc. Natl Acad. Sci. USA*, **108**, 498–503.
46. Romero, P.A. and Arnold, F.H. (2009) Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.*, **10**, 866–876.
47. Karanicolas, J., Corn, J.E., Chen, I., Joachimiak, L.A., Dym, O., Peck, S.H., Albeck, S., Unger, T., Hu, W., Liu, G. *et al.* (2011) A de novo protein binding pair by computational design and directed evolution. *Mol. Cell*, **42**, 250–260.
48. Rothlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J.L., Althoff, E.A., Zanghellini, A., Dym, O. *et al.* (2008) Kemp elimination catalysts by computational enzyme design. *Nature*, **453**, 190–195.
49. Savile, C.K., Janey, J.M., Mundorff, E.C., Moore, J.C., Tam, S., Jarvis, W.R., Colbeck, J.C., Krebber, A., Fleitz, F.J., Brands, J. *et al.* (2010) Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science*, **329**, 305–309.
50. Chavez, C.L. and Calos, M.P. (2011) Therapeutic applications of the PhiC31 integrase system. *Curr. Gene Ther.*, **11**, 375–381.
51. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
52. Mahoney, M. and Jorgensen, W. (2000) A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.*, **112**, 8910–8922.
53. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Dinola, A. and Haak, J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.
54. van Gunsteren, W.F. and Berendsen, H.J.C. (1977) Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.*, **34**, 1311–1327.
55. Darden, T., York, D. and Pedersen, L. (1993) Particle mesh Ewald: an N [center-dot] log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
56. Gibb, B., Gupta, K., Ghosh, K., Sharp, R., Chen, J. and Van Duyne, G.D. (2010) Requirements for catalysis in the Cre recombinase active site. *Nucleic Acids Res.*, **38**, 5817–5832.
57. Kim, S.T., Kim, G.W., Lee, Y.S. and Park, J.S. (2001) Characterization of Cre-loxP interaction in the major groove: hint for structural distortion of mutant Cre and possible strategy for HIV-1 therapy. *J. Cell. Biochem.*, **80**, 321–327.
58. Lee, L. and Sadowski, P.D. (2001) Directional resolution of synthetic holliday structures by the Cre recombinase. *J. Biol. Chem.*, **276**, 31092–31098.