HHS Public Access

Author manuscript

Nat Commun. Author manuscript; available in PMC 2014 April 01.

Published in final edited form as:

Nat Commun. 2013 ; 4: 2509. doi:10.1038/ncomms3509.

Mutants of Cre recombinase with improved accuracy

Nikolai Eroshenko[1] and **George M. Church**[2,3]

[1]Harvard School of Engineering and Applied Science, New Research Building 238, 77 Ave. Louis Pasteur, Boston, MA 02135

[2]Harvard Medical School, Department of Genetics, New Research Building 238, 77 Ave. Louis Pasteur, Boston, MA 02135

[3]Wyss Institute for Biologically Inspired Engineering, New Research Building 238, 77 Ave. Louis Pasteur, Boston, MA 02135

Abstract

Despite rapid advances in genome engineering technologies, inserting genes into precise locations in the human genome remains an outstanding problem. It has been suggested that site-specific recombinases can be adapted towards use as transgene delivery vectors. The specificity of recombinases can be altered either with directed evolution or via fusions to modular DNA-binding domains. Unfortunately, both wildtype and altered variants often have detectable activities at off-target sites. Here we use bacterial selections to identify mutations in the dimerization surface of Cre recombinase (R32V, R32M, and 303GVSdup) that improve the accuracy of recombination. The mutants are functional in bacteria, in human cells, and *in vitro* (except for 303GVSdup, which we did not purify), and have improved selectivity against both model off-target sites and the entire *E. coli* genome. We propose that destabilizing binding cooperativity may be a general strategy for improving the accuracy of dimeric DNA-binding proteins.

Introduction

Safe delivery of transgenes into the human genome remains an open problem of critical importance to clinical genetics. Many existing technologies have major limitations. For instance, retroviruses, lentiviruses, and transposons integrate non-specifically and can therefore cause cancer by mutagenesis[1,2]. Transgenes can also be integrated using the endogenous homologous repair pathways, although this process must be stimulated by generating double-stranded breaks at the target site using programmable nuclease technologies such as meganucleases[3], zinc finger nucleases[4], TALE nucleases[5,6], or the RNA-guided Cas9 protein[7,8]. This technique is limited by the fact that homologous

Correspondence should be addressed to N.E., eroshenk@seas.harvard.edu.

Author Contributions

N.E. conceived the study, designed and performed all experiments, and wrote the manuscript. G.M.C. supervised the project and provided critical advice.

Competing Financial Interests

N.E. and G.M.C. are named inventors on a patent application on technologies described in this article.

recombination in humans is less efficient than the competing mutagenic non-homologous end joining pathway[9,10].

Site-specific recombinases, which catalyze recombination at precise sites, have properties that make them promising candidates for use as safe gene delivery vectors. For instance, many do not require host-encoded factors[11]. Additionally, the size of the integrated cassette is less restricted than for other methods. Recombinases' specificities can be altered either by direction evolution or by fusion to modular DNA-binding domains[12–20]. Unfortunately, many reprogrammed variants are promiscuous in their activity. This problem isn't restricted to artificial variants, as activity at off-target human genomic loci has been reported for some wild-type (WT) recombinases[21–24]. If recombinases are to be used as gene delivery vectors, it is imperative to identify ways to enhance their accuracy.

The accuracy of DNA-binding proteins can be altered by varying the ratio of specific to non-specific DNA-protein interactions[25]. While powerful, this approach can be inconvenient if the goal is to generate variants of a protein with different specificities: a specificity change alters the DNA-protein interaction, requiring re-optimization of accuracy. Therefore, there is a need for methods to systematically enhance accuracy without changing the DNA-protein interface.

In this work we attempt to discern such principles using Cre recombinase of the phage P1 as a model system. Cre catalyzes a reversible, directional recombination between two 34 base-pair (bp) *loxP* sites that consist of a pair of 13 bp inverted repeats flanking an 8 bp asymmetrical spacer[26–28]. Mutagenic studies of *loxP* have shown that many mutations have non-catastrophic effects on recombination efficiency[29–31]. Using a theoretical model we predict that reducing the cooperativity of binding should increase accuracy. We mutagenize a region involved in the formation of Cre dimers and perform bacterial selections for functional and accurate mutants. We isolate three mutants, all of which were able to recombine *loxP* sites with high efficiency and exhibited improved accuracy with respect to both model off-target sites and the entire *E. coli* genome.

Results

A theoretical model of DNA binding accuracy

Under the currently accepted mechanism of Cre recombination, the binding of a Cre monomer to one half of a *loxP* site is followed by the formation of an asymmetrical homodimer when a second Cre molecule binds to the other half of *loxP*. Next, the two *loxP*-bound dimers associate to form a tetramer, and recombination proceeds via a Holiday Junction intermediate[28]. We reasoned that the formation of the dimer of dimers is not site specific in the sense that it involves no new DNA binding events, leading us to conclude that the precision of dimer formation determines the accuracy of recombination. Dimer formation on target sites is described with:

$$P + D \xleftrightarrow{2K} P \cdot D \quad (1)$$

$$P+P\cdot D \xleftrightarrow{K_{dim}} 2P\cdot D \quad (2)$$

where P is the unbound protein monomer, D is the full DNA binding site, K is the affinity of each monomer for half of the binding site, and K_{dim} is the affinity of the protein dimer for the full binding site. If it is assumed that the cooperative energy is sequence-independent then $K_{\mathrm{dim}} = KK_{\mathrm{coop}}$, where K_{coop} is the protein-protein affinity. A competing set of binding events occurs between off-target DNA and the protein:

$$P+\mathrm{D}_{\mathrm{OT}} \xleftrightarrow{2\mathrm{K}_{\mathrm{OT}}} \mathrm{P}\cdot\mathrm{D}_{\mathrm{OT}} \quad (3)$$

$$P+P\cdot D_{OT} \xleftrightarrow{K_{OT}K_{coop}} 2P\cdot D_{OT} \quad (4)$$

where $[D_{\mathrm{OT}}]$ is the off-target DNA concentration and K_{OT} is the affinity of the protein for off-target DNA. Accuracy can be defined as the ratio of on-target protein dimers to off-target protein dimers:

$$A \triangleq \frac{[2P\cdot D]}{[2P\cdot D_{OT}]}=\left(\frac{K}{K_{OT}}\right)^2\frac{[D]}{[D_{OT}]} \quad (5)$$

The free DNA concentration $[D]$ is related to the total DNA concentrations $[D_{\mathrm{total}}]$ via:

$$[D_{total}]=[D]+[P\cdot D]+[2P\cdot D]=[D]\left\{2K_{coop}K^2[P]^2+2K[P]+1\right\} \quad (6)$$

An analogous expression relates $[D_{\mathrm{OT}}]$ and $[D_{\mathrm{OT,total}}]$. Equation (5) can be expressed in terms of total DNA concentrations:

$$A=\left(\frac{K}{K_{OT}}\right)^2\frac{[D_{total}]}{[D_{OT,total}]}\frac{2K_{coop}K_{OT}^2[P]^2+2K_{OT}[P]+1}{2K_{coop}K^2[P]^2+2K[P]+1} \quad (7)$$

To model accuracy of Cre in E. coli we used the in vitro affinity coefficients K of 1.5×10^8 M^{-1} and a K_{coop} of 1.7×10^3 M^{-1}, with both values obtained from previous *in vitro* measurements[32]. We assumed that $K/K_{OT} = 10^4$, which is in the same order of magnitude as the experimentally determined K/K_{OT} of EcoRV and BamHI[33]. Assuming a single target site in an E. coli cell of a 0.5 μm radius gives a $[D_{\mathrm{total}}]$ of 3.2×10^{-9} M. If off-target sites exist at 1bp windows along both strands of the 4.6 Mbp E. coli genome then $[D_{\mathrm{OT,total}}]$ is approximately 3.2×10^{-9} M bp^{-1} * 9.2×10^6 bp = 2.9×10^{-2} M. This value of $[D_{\mathrm{OT,total}}]$ may be an overestimate due to competitive binding of other DNA-associated proteins to the genome.

We plotted the predicted accuracy of dimer formation as a function of total protein concentration for both WT Cre and for mutants with reduced cooperativity (Fig. 1). The model predicts that accuracy increases with both a reduction in cooperativity and a decrease

in protein expression levels. This is an intuitive result because accuracy should increase as the two monomer binding events become more independent from each other. We conclude that while a reduction in cooperatively will affect both target and off-target binding, off-target binding will be destabilized to a greater degree.

Identifying candidate mutations using bacterial selection

Our theoretical model predicted that accuracy could be improved by decreasing the cooperative binding moment. We therefore targeted our mutagenesis towards a domain directly involved in the dimer interaction but distant from the Cre-DNA interaction: the alpha helix closest to the amino-terminus[34]. To find mutations that improve accuracy while maintaining proper function with respect to *loxP* we performed two rounds of bacterial selection. The first round was designed to identify functional mutants while the second round selected accurate mutants. To select functional mutants we used a resistance marker flanked by *loxP* sites in an inverted orientation relative to each other. The reading frame was inverted with respect to the promoter, such that Cre-mediated inversion would result in gain of antibiotic resistance. An out-of-frame toxic *ccdB* was used to apply a selective pressure against inaccurate recombination reactions that produced frame-shifts (Fig. 2a). To minimize false negatives due to reversal of the inversion we placed the selection cassette on a high copy plasmid. The selection resulted in the recovery of 1,690 library-transformed colonies, which corresponds to 38% of the total transformation efficiency. The library produced fewer clones than the positive WT control, suggesting that the selection is functional (Fig. 2b).

To identify accurate constructs we first found an off-target site to serve as bait in the negative selection. To achieve a high selective pressure we wanted a bait sequence that would be recombined with a high efficiency. At the same time, we wanted to select for improved accuracy across the entire protein-DNA interaction, and therefore we wanted a bait sequence with little similarity to *loxP*. We found such a site by performing a selection for pseudo-*loxP* sites and characterizing their *in vitro* recombination efficiency. The site, which we named *loxBait*, is recombined with 37% the efficiency of *loxP* despite differences in 9 out of 13 bases within a single inverted repeat (Figs. 3a and 3b).

We performed the counter-selection by flanking an in-frame antibiotic resistance marker with *loxBait* and *loxP* oriented in the same direction. The toxic *ccdB* gene was placed 3′ of *loxP*. Cre-mediated excision would result in both loss of the resistance marker and expression of the toxic gene (Fig. 2c). We subjected the expression plasmids recovered from the positive selection to one round of counter-selection. In this case the catalytically inactive mutant, Y324F, served as the control for growth without selection (Fig. 2d).

We randomly isolated two of the recovered mutants, R32V and R32M, for further characterization. R32 is involved in an inter-monomer salt bridge with E69, so its disruption in the two mutants can be expected to reduce the protein-protein affinity (Fig. 4a). Two WT colonies survived the counter-selection, one of which contained a clone with a *de novo* duplication of residues 303–305 (303GVSdup). This region is a loop that makes close contact to the other monomer in the dimer structure (Fig. 4b).

Mutants better discriminate a model off-target site

We measured the activity of Cre and the isolated mutants on *loxP* and *ψLox h7q21*, a known human off-target site[21], using a plasmid-based inversion assay. As in the selections the proteins were expressed from the P_{bad} promoter. Since our theoretical model suggests that greater accuracy is achieved at low protein expression levels, we grew the cells on repressive LB/glucose medium. Indistinguishable results were achieved with growth on LB in the absence of glucose. To test the mutants for improved accuracy we analyzed their ability to recombine a known human pseudo-*loxP* site *ψLox h7q21* and *ψCore h7q21* (which consists of inverted repeats from *loxP* but a spacer that matches *ψLox h7q21F*) (Fig. 3b).

Following recombination, substrate plasmids were isolated and digested in a way that resulted in the substrate producing a 631 bp band and its inverted product producing a 386 bp band (Fig. 5a). Y324F produced only the 631 bp parent product while R32V, R32M, and 303GVSdup resulted in both inverted and parental plasmids (Fig. 5b). WT unexpectedly produced the 898 bp band that corresponds to a deletion; there were no detectable bands for the inverted and parental forms. We also observed a 1,795 bp band formed when ScaI failed to cut. This product could be ignored for Y324F, R32V, R32M, and 303GVSdup because ScaI cutting efficiency should be independent of cassette orientation. However the 1,795 bp band did need to be taken into account when analyzing recombination by WT: we normalized the 898 bp deletion product band to the total amount of DNA in the 898 bp and 1,795 bp bands (Figs. 5c & 5d). Incomplete ScaI digests also produced a 2,573 bp product that is formed via an intermolecular insertion followed by inversion between recombination sites that originated from different molecules (Supplementary Fig. S1). We excluded this product from our analysis.

Achieving equilibrium in the inversion assay should result in 50% inversion. Recombination of two *loxP* sites by R32V, R32M, and 303GVSdup resulted in approximately 50% inversion and no detectable deletion products. In contrast, WT resulted in deletion in 98.7% (95% CI: 97.5 – 100%) of plasmids (Fig. 5c). Unlike inversions, which can be reversed, excisions are selected for in a dividing cell population because the excised products cannot replicate. Therefore, the data should not be taken to mean that WT incorrectly excises nearly 100% of the time, but rather that the vast majority of substrate plasmids experienced at least one excision event during the 12 hours of growth. Recombination by WT of pseudo-loxP sites *ψLox h7q21* and *ψCore h7q21* resulted in improper excision with a frequency of 97.5% (95% CI: 96.5–98.4%). R32V, R32M, and 303GVSdup produced no detectable recombination products (Fig. 5d). In aggregate, the data provided evidence that the isolated mutants are better able to distinguish on-target and off-target sites than WT.

Mutants are functional in vitro and in human cells

Given that *in vivo* recombination between two *loxP* sites reached equilibrium (~100% deletion for WT and ~50% inversion for R32V, R32M, and 303GVSdup), there are two explanations for the improved selectivity against pseudo-loxP sites observed for the mutant recombinases; either the mutants were better able than WT to discriminate against pseudo-*loxP* sites or the mutants were equally less efficient at recombining *loxP* and pseudo-*loxP* sites than WT, but the high enzyme concentration and long reaction times were sufficient to

drive *loxP* recombination to completion. These two hypotheses can be distinguished by comparing recombination of pseudo-*loxP* sites under conditions at which WT and mutants have comparable, sub-equilibrium efficiencies of *loxP* recombination. Since such conditions can be easily found *in vitro* we attempted to purify the WT and mutant Cre recombinases. Affinity purification of using a maltose binding protein (MBP) domain fusion followed by scarless protease cleavage of at the MBP domain allowed isolation of WT, R32V and R32M (Supplementary Fig. S2). We attribute our failure to purify 303GVSdup to its lack of solubility following MBP removal.

We measured recombination kinetics using an intramolecular excision assay on linearized plasmid substrates. To test off-target activity we used a pseudo-*loxP* site *lox80* (Fig. 3b), which preliminary tests revealed to be efficiently recombined by WT *in vitro*. WT, R32V and R32M were all able to recombine *loxP* within the 8 h assay. In contrast to WT, R32V and R32M did not produce detectable *loxP*x *lox80* reaction products (Figs. 6a–c). WT recombination of the *loxP*x *lox80* substrate produced a small amount of an unexpected product that migrated at approximately twice the size of the linear recombination product. This product, which we excluded from our quantitative analysis, may represent recombination with a cryptic pseudo-*loxP* site in the vector backbone.

Cre recombined *loxP* x *loxP* and *lox80* x *loxP* reactions with comparable kinetics, although it was significantly less effective recombining *lox80* at the 5 min time point (Figs. 5d and 5e). R32V and R32M recombined *loxP* sites significantly more slowly than WT, with both mutants failing to achieve a steady state within the 8 h experiment. The kinetic data enabled pseudo-*loxP* recombination efficiencies of the different mutants to be compared at similar *loxP* x *loxP* recombination efficiencies. For instance, after 5 min WT had recombined *loxP* sites with 21% efficiency, after 8 h R32V had recombined *loxP* sites with 29% efficiency, and after 4 h R32M recombined the *loxP* substrate at an efficiency of 27%. Despite the slightly higher *loxP* recombination efficiencies, R43V and R32M had no detectable off-target activity while WT recombined *lox80* and *loxP* with 4% efficiency. A similar conclusion can be reached by comparing the 46% *loxP* x *loxP* efficiency and 35% *lox80* × *loxP* achieved by WT after 30 min to the 44% *loxP* x *loxP* efficiency and absence of off-target recombination seen with R32M after 8 h. In aggregate, these data support the model under which the mutations we identified destabilize both *loxP* and pseudo-*loxP* recombination, but with a much greater reduction in pseudo-*loxP* efficiency than in *loxP* efficiency.

Our observation that R32V and R32M are functional *in vitro* suggested that the mutants, like Cre, could catalyze *loxP* recombination in mammals. To test the functionality of Cre mutants in human HEK293 cells we expressed R32V, R32M, and 303GVSdup under the control of the CMV promoter. The substrate plasmid contained a *loxP*-flanked GFP with downstream lacZ gene, such that recombination excised GFP and made the cells turn blue when stained with X-gal (Supplementary Fig. 3a). 2.5 days following transfection with R32V, R32M, and 303GVSdup the cells had lost the GFP signal and stained positive for β-galactosidase activity (Supplementary Fig. 3b). As in *E. coli*, the *in vivo* reaction approached completion; the transfection efficiency-normalized recombination efficiencies were 95.2%

for R32V (95% C.I.: 85.9 – 100%), 99% for R32M (95% C.I: 97.1 – 100%), and 100% for 303GVSdup (with no variation across replicates) (Supplementary Fig. 3c).

Mutants have less off-target activity genome-wide

One explanation for the improved accuracy we observed with R32V, R32M, and 303GVSdup is that the mutations altered the preferred off-target sites without changing the overall accuracy. To test this possibility we measured the efficiency of off-target insertions across the entire E. coli genome. Strains carrying the arabinose-inducible recombinase expression plasmids were transformed with a plasmid containing a *loxP* site, an R6k γ origin of replication, and a kanamycin resistance gene. Since the R6k γ origin cannot replicate in the *pir*-strain we used, only insertion of the plasmid into the genome would result in the replication of the kanamycin resistance gene. To control for variable transformation efficiencies we normalized the number of R6k γ colonies by the number of colonies arising from transformation with a plasmid lacking *loxP* but containing a functional origin of replication.

We were unable to obtain Cre-mediated integration on LB/glucose. However, when we briefly pulsed the cells with arabinose prior to the growth on LB/glucose, we obtained a WT-mediated integration frequency of 1.3×10^{-4} (Table 1). 303GVSdup had an integration frequency ~100-fold lower than WT. The integration frequencies of R32V and R32M were lower than that of 303GVSdup and could not be distinguished from Y324F background. These data strongly suggest that the higher accuracy of the mutants was not restricted to *loxBait*, *ψLox h7q21*, and *lox80*.

Discussion

Our data suggest that R32V, R32M, and 303GVSdup can efficiently recombine *loxP* in bacteria and in human cells. We also observed that the mutants exhibit better directionality and better accuracy with respect to the pseudo-*loxP* site *ψLox h7q21* than WT. R32V and R32M were functional *in vitro* and were better able to discriminate against the *lox80* site than WT Cre. All three mutants more rarely integrated a *loxP*-carrying plasmid into the *E. coli* genome than WT Cre. Crystal structures of Cre strongly suggest that R32V and R32M disrupt a strong salt-bridge in the dimer interface (Fig. 4a). The duplicated residues in the *de novo* 303GVSdup mutant are also located at the dimer interface, although the exact biochemical consequences of the mutation are unclear.

Despite our model's prediction that higher protein concentrations should reduces the accuracy of recombination (Fig. 1), we saw identical results in our bacterial plasmid recombination assay when the P_{bad} promoter controlling protein expression was repressed with glucose (Figs. 5c, d) and when it was weakly induced by LB without glucose. Despite the apparent contradiction, the data do not disprove the model. Even under glucose repression WT Cre was able to drive both the *loxP* and pseudo-*loxP* reactions to equilibrium within the time frame of the experiment, therefore any increases in recombination efficiency caused by a higher protein concentrations could not be measured. The same logic applies to R32V, R32M, and 303GVSdup recombining *loxP* sites. The model predicts that higher protein concentrations should cause Cre mutants to recombine pseudo-*loxP* sites more

efficiently. The data neither confirm nor deny this prediction since both expression conditions resulted in a recombination frequency that was below the detection level.

The main limitation of our accuracy model is that it considers only binding, ignoring catalysis. The binding site sequence likely contributes to both establishing the proper alignment of the catalytic site and creating a tertiary DNA structure within the recombinase complex that is energetically favorable for recombination. The DNA site may also contribute to protein folding[35]. It is therefore possible that R32V, R32M, and 303GVSdup mutations contribute to improved accuracy in steps of the catalytic pathway other than DNA binding. If that is the case, there may exist pseudo-*loxP* sites that are recombined by WT Cre but not at all by the mutants.

Although R32V and R32M had high *in vivo* recombination efficiencies, the mutants catalyzed recombination slower than WT *in vitro* (Fig. 6d). The slower kinetics may explain some of the decrease in apparent off-target integration in the genome-wide assay (Table 1). However, during the genome integration assay the cells were induced for 30 min and then recovered for 1 h in glucose media prior to antibiotic selection. If it is assumed that recombination occurred for 1 h (an underestimate since recombination can continue to happen after the addition of antibiotics) then the *in vitro* data predict that R32V and R32M will produce 10–20 fold fewer transformants than WT. In fact, R32V and R32M were at least 200 times more accurate than WT. We therefore believe that the majority of the decrease in the frequency of genomic integration was caused by an improvement in selectivity against pseudo-*loxP* sites and not by the mutants' slowed reaction kinetics.

Our observation that Cre make frequent deletions on a substrate containing inverted *loxP* has been made previously[36] and is consistent with the fact that Cre can occasionally recombine sites with non-matching spacers[30,31]. This looseness in the directionality of recombination may interfere with synthetic circuit designs that rely on Cre-mediated inversion as a form of genetic memory[37,38]. The three Cre mutants we isolated seem to have an improved directionality over WT and may therefore be of use in synthetic biology applications.

Since the binding specificity of nucleases is the major determinant of their toxicity, it may be possible to minimize nucleases' toxicity by reducing their DNA-binding cooperativity[39]. The main limitation of this approach is the fact that it inevitably decreases affinity for the target site. Strong monomer affinities, such as the 1–10 nM K_d of Cre for half of *loxP*[32,40], are a likely prerequisite. For comparison, DNA-binding domains composed of three Cys_2-His_2 zinc fingers have been reported to bind to their 9 bp recognition sequence with a K_d as low as 400 pM, while one 17.5 TALE repeat domain interacts with DNA with a K_d of 160 pM[41,42]. Although only the better zinc finger and TALE designs achieve such high binding energies, the reported values suggest that destabilizing the cooperativity of DNA-binding may be a viable strategy for increasing the accuracy of designer nucleases.

Changing the dimer interface has been previously used to improve the specificity of nucleases engineered to cut asymmetrical DNA sites. Because targeting asymmetric sites necessitates co-expression of two different DNA binding domains fused to a constant nuclease domain, catalytically active dimers can format four different sites: the desired

asymmetric site, an off-target asymmetric site, and the two symmetric sites targeted by the two possible homodimers. Converting the nuclease dimerization interface into a heterodimeric interface reduces activity at the two symmetric off-target sites[43–45]. The approach described in this paper is different from the obligate heterodimer strategy because destabilizing the cooperativity of a dimeric protein improves accuracy by increasing the energy difference between binding to the target and the off-target sites not by reducing the number of possible target sites. The obligate heterodimer and the destabilized cooperativity strategies should be compatible with each other as it should be possible to modulate the binding cooperativity of heterodimers.

Due to its high efficiency and the lack of necessary co-factors, Cre has found widespread use in genetics research[46]. However, Cre toxicity in the absence of *loxP* sites has been observed in a variety of animal and cell culture systems[47–51]. The source of this toxicity is not known. However, a number of observations, including absence of toxicity from catalytically inactive Cre mutants, an increase in the frequency of chromosomal rearrangements, and evidence of activation of DNA damage response pathways all point at recombination at off-target pseudo-*loxP* sites as the cause[47,49–51]. We envision that the Cre mutants isolated in this study may be useful for alleviating the toxic phenotypes associated with WT Cre. This approach should be compatible with existing strategies for reducing Cre toxicity, which include placing the Cre gene in a self-excisable cassette, regulating Cre activity with a hormone binding domain of a steroid receptor, or using a drug-regulated fragment complementation strategy[50,52–54].

Methods

Plasmid construction

The positive selection substrate pCR-(loxP-ampR-loxP^{inv})^{inv} was built by amplifying ampR from pQL123[55] with primers ampR_f and ampR_r (primer sequences are provided in Supplementary Table S1), performing an extension PCR on the amplicon with ampR_loxP_f and ampR_loxP_inv_r, and TOPO cloning the product into pCR-Blunt II-TOPO (Life Technologies). Sequencing was used to screen for colonies in which the ampR gene was in reverse orientation with respect to the promoter. A similar workflow was used to construct pCR-(h7q21-ampR-ψlox h7q21^{inv})^{inv}, pCR-(lox80-ampR-loxP)^{inv}, and pCR-loxBait-ampR-loxP. The pZE2-loxP/loxP and and pZE2-ψCore h7q21/ψlox h7q21 *in vivo* recombination substrates were obtained by cloning the XhoI/BamHI fragment from pCR-(loxP-ampR-loxP^{inv})^{inv} or pCR-(h7q21-ampR-ψlox h7q21^{inv})^{inv} into XhoI/BamHI-digested pZE21G.

The Cre gene was obtained from pQL123, though we reverted the alanine at the second position to the serine found in WT Cre (GenBank sequence YP_006472). For the bacterial assays the Cre mutants or libraries were introduced in place of HpaII[51-358] in pARC8-HpaII[51-358](a derivative of pAR-MHhaI[29–327][56] containing residues 51-358 of HpaII in the place of HhaI) using Gibson assembly[57]. Cloning was performed with a backbone amplified using primers pARC8_f and pARC8_r. For protein purification WT, R32V, R32M, and 303GVSdup were amplified from the respective pARC8-based plasmids using primers Cre_not ATG_f and Cre_SalI_r. The genes were digested with SalI and ligated into SalI/XmnI-digested pMAL-c5x (New England Biolabs). For mammalian expression R32V,

R32M, and 303GVSdup were amplified from the corresponding pARC8-based expression plasmids using primers Cre_TA_f and Cre_TA_r and TOPO TA cloned into pcDNA3.3-TOPO (Life Technologies). Clones were screened for the correct orientation by PCR.

Identifying and characterizing functional loxP variants

Libraries of half-site variants were constructed by amplifying the pZE21G plasmid[58] first with primers pZE21G_f and pZE21G_r, and then pZE21G_2_lib and pZE21G_2_loxP, producing an amplification product that was 2,437 bp long and contained a *loxP* site and a random library site near each end. The libraries were purified using the QIAquick PCR Purification Kit (Qiagen), 10–20 ng of the DNA was treated with 1 U Cre (New England Biolabs) in Cre reaction buffer (10 mM $MgCl_2$, 33 mM NaCl, 50 mM Tris-HCl pH 7.5) in 20 μL total reaction volume for 1 hr at 37 °C, heat inactivated at 75–80 °C, then digested with DpnI. The DNA was then purified, digested with PlasmidSafe (Epicentre), and transformed into One Shot Top10 chemically competent cells (Life Technologies, *rec*A1 *ara*D139Δ(*ara-leu*)7697 Δ(*lac*)X74). Colonies were randomly selected for sequencing. Substrates for validating the selection hits were generated by performing extension PCR on pZE21G as for the selections except that sequences obtained from the selection were in place of the random library. 30 ng of purified products were treated with 1 U Cre in Cre reaction buffer in 20 μL total reaction volume for 1.5 hr at 37 °C followed by heat inactivation of the enzyme. The entire reaction was resolved on a 0.7% agarose gel stained with SYBR Green I (Life Technologies). Each recombination was performed in parallel with a no-enzyme negative control.

Negative and positive selections

The pCR-(loxP-ampR-$loxP^{inv}$)^{inv} and pCR-loxBait-ampR-loxP selection plasmids were maintained in NEB 10-beta cells (New England Biolabs, *rec*A1 *ara*D139 Δ(*ara-leu*)7697 Δ(*lac*)X74); cells were made electrocompetent using standard techniques[59]. The library of Cre variants was generated by mutagenic PCR using a pool of 19 oligonucleotides that substituted each of the 19 codons encoding S20–S38 for NNN (1st round of PCR: primers Cre_mut_1 – Cre_mut_19 and Cre_pARC8_r; 2nd round of PCR: primers Cre_pARC8_f and Cre_pARC8_r). The library was introduced into pARC8 using Gibson assembly, desalted by drop dialysis, and electroporated into competent cells carrying pCR-(loxP-ampR-$loxP^{inv}$)^{inv}. Control transformations were performed with 100 pg pARC8-Cre or pARC8-Y324F. Transformed cells were recovered in low-salt 2X LB (2% bacto-tryptone, 1% yeast extract, 0.5% NaCl, pH 7.5) at 37 °C for 30 min, induced with 0.2% arabinose at 37 °C for 30 min, and recovered in SOC with 200 μM IPTG at 37 °C for 1 hr. The cells were then grown overnight at 37 °C on 0.2% glucose, 100 μM IPTG, 12.5 μg/mL chloramphenicol, 50μg/mL kanamycin LB plates either with or without 100 μg/mL carbenicillin.

Colonies obtained from the positive selection of the library and of the controls were collected by scraping. DNA was isolated using the QIAprep Spin Miniprep kit (Qiagen) and was digested with XmaI and SpeI (which cut only the substrate plasmids.) The concentration of expression plasmid was quantified via agarose gels. Electrocompetent cells carrying the pCR-loxBait-ampR-loxP negative selection substrate were transformed with 100 pg of the

expression plasmid, recovered the cells in low-salt 2X LB for at 28 °C for 30 min, induced with 0.2% arabinose at 28 °C for 30 min, washed with SOC, and recovered in SOC with 200 μM IPTG at 37 °C for 1 hr. The cells were then grown overnight at 37 °C on 0.2% glucose, 100 μM IPTG, 12.5 μg/mL chloramphenicol, 50 μg/mL kanamycin, 100 μg/mL carbenicillin LB plates. To ensure clonality the isolated variants were amplified and re-cloned into pARC8 via Gibson assembly.

Bacterial recombination assay

Efficiencies were measured by co-transforming 75 ng of a pARC8-based expression plasmid (WT, Y324F, R32V, R32M, or 303GVSdup) and an equimolar amount of either pZE2-loxP/loxP or pZE2-ψCore h7q21/ψlox h7q21 into 50 μL vial of OneShot Top10 chemically competent cells. The cells were recovered in LB with 0.2% glucose at 37 °C for 1 hr, then grown at 37 °C for 12 hrs in 0.2% glucose, 12.5 μg/mL chloramphenicol, 50 μg/mL kanamycin LB. The plasmids were isolated using the QIAprep Spin Miniprep kit, eluting into 50 μL10 mM Tris-HCl pH 8. The purified plasmids were digested for 20 min at 37 °C in 60 μL reactions containing all of the collected DNA and 20 U of both ScaI-HF and NcoI-HF (both from New England Biolabs) in NEBuffer 4 (1 mM DTT, 50 mM potassium acetate, 20 mM tris-acetate, 10 mM magnesium acetate, pH 7.9). The digests quantified on 1% agarose gels following purification using the QIAprep Spin Miniprep kit.

In vitro recombination assay

The pMAL-based recombinase expression plasmids were maintained in One Shot Top10 chemically competent cells. An overnight culture of each expression clone was diluted to OD600 of 0.02 in 55 mL 0.2% glucose, 50 μg/mL kanamycin LB and grown for 2.5 hrs at 37 °C. IPTG was added to a final concentration of 300 μM and the cultures were grown at 37 °C for an additional 2 hrs. 50 mL of the culture was pelleted, resuspended in 5 mL lysis buffer (1 mM EDTA, 1 mM DTT, 200 mM NaCl, 20 mM Tris-HCl pH 8) supplemented with Halt Protease Inhibitor Cocktail (Thermo Scientific). The cell suspension lysed using eight 15 sec sonication cycles using a power setting of 4 on a Misonix Sonicator 3000. The lysate was centrifuged for 20 min at 5,000 g and the supernate was bound to 500 uL amylose resin (New England Biolabs). The resin was washed with 4 mL wash buffer (200 mM NaCl, 20 Tris-HCl pH 7.5), 12 mL of DNA-removal buffer (1 M NaCl, 20 mM Tris-HCl pH 7.5), and again with 4 mL wash buffer. The protein was eluted into 500 μL elution buffer (0.5% maltose, 2 mM $CaCl_2$, 100 mM NaCl, 20 mM Tris-HCl pH 7.5) and digested with 200 ng Factor Xa (New England Biolabs) for 13–14 hrs at room temperature. The protease activity was stopped by adding Halt Inhibitors to 1× and DTT to 1 mM. Precipitated protein was removed by centrifuging at 16,000 g for 10 min and passing the supernate through an Amicon Ultracel 100K filter. The flow-through was concentrated using an Amicon Ultracel 10K filter. Glycerol was added to 27% final volume, NaCl was added to adjust the concentration to 200 mM, and the proteins were stored at −20 °C. Catalytic assays were performed within 2 days of purification. Protein concentration was determined using a Qubit 2.0 fluorometer (Life Technologies). The concentration of protease-cleaved recombinase was determined by quantifying band intensity on a SimplyBlue (Life Technologies) stained PAGE gel (typical values were 15–30%), and by then normalizing the total protein concentration by the obtained value.

Linearized pLox2+ (New England Biolabs) and ScaI-digested pCR-(lox80-ampR-loxP)^{inv} were used as the recombination substrates. Recombination kinetics were measured by reacting 5 nM of the DNA substrate and 750 nM of each protease-cleaved recombinase in Cre reaction buffer with 100 ng/μL BSA. The reactions were performed at 37 °C and were stopped by heating them to 70 °C for 15–25 min. Excision was measured by quantifying bands on 1% agarose gels.

Human cell culture recombination assay

pCMV:GFP(loxP)lacZ[60] (Addgene: 31125) was used as the recombination substrate plasmid and pcDNA3.3-TOPO/lacZ (Life Technologies) was used as a lacZ staining control. Except for pcDNA3.3-TOPO/lacZ, which was prepared by the manufacturer, all substrate and expression plasmids were prepared for transfection using the HiSpeed Plasmid Maxi Kit (Qiagen).

Lenti-X 293T cells (Clontech) were grown on poly-D-lysine coated tissue-culture treated polystyrine in DMEM High Glucose with GlutaMAX (Life Technologies) supplemented with 10% FBS, 50 U/mL penicillin, and 50 ug/mL streptomycin at 37 °C and 5% CO_2 in a humidified incubator. Cells were subcultured using TrypLE Express (Life Technologies). Cells in 12-well plates were transformed with 500 ng pCMV:GFP(loxP)lacZ and 500 ng pcDNA3.3-TOPO-based expression plasmid using Fugene HD (Promega). Substrate only-controls were transformed with only 500 ng recombination substrate and the lacZ controls were trasfromed with 500 ng pcDNA3.3-TOPO/lacZ. Media was changed 1 d following transfection. Cells were subcultured 2 d following transfection. Staining and imaging was performed 2.5 d following transfection. Cells were fixed using 2% formaldehyde and 0.1% glutaraldehyde in PBS for 10 min at room temperature. After being washed with PBS the cells were stained with 1 mg/mL X-gal, 4 mM potassium ferricyanide, 4 mM potassium ferrocyanide, and 2 mM $MgCl_2$ in PBS for 30 min at 37 °C. Following staining the cells were washed with PBS and imaged.

Genome-wide off-target integration assay

Electrocompetent cells were prepared from each of the pARC8-based expression plasmids (WT, Y324F, R32V, R32M, or 303GVSdup) cloned in NEB 10-beta cells using 40 mL of culture per transformation. Each expression strain was transformed either with 200 ng pUNI10 (loxP^{+}, ori$^{R6K\gamma}$, kanR)[55] or an equimolar amount of pZE21G (loxP^{-}, ori^{ColE1}, kanR). The two transformations were done in parallel using competent cells made from aliquots of the same culture. The cells were recovered in LB for at 37 °C for 30 min, induced with 0.2% arabinose at 37°C for 30 min, and recovered in SOC at 37 °C for 1 hr. The cells were then grown overnight at 37 °C on 0.2% glucose, 12.5 μg/mL chloramphenicol, 50 μg/mL kanamycin LB plates.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the US Office of Naval Research Multidisciplinary University Initiative and the National Human Genome Research Institute Center for Excellence in Genomics Science. We thank G. Meister for pARC8-HpaII[51-358] and S. Elledge for pQL123 and pUNI10. We also thank P. Mali, J. DiCarlo and A. McElhinney for a critical reading of the manuscript and S. Kosuri, S. Raman, M. Lajoie and X. Rios for helpful advice and inspiring discussions.

References

1. Stocking C, et al. Distinct classes of factor-independent mutant can be isolated after retroviral mutagenesis of a human myeloid stem cell line. Growth Factors. 1993; 8:197–209. [PubMed: 8391284]
2. Hacein-Bey-Abina S, et al. A aerious adverse event after successful gene therapy for X-Linked severe combined immunodeficiency. New England Journal of Medicine. 2003; 348:255–256. [PubMed: 12529469]
3. Rouet P, Smih F, Jasin M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Molecular and Cellular Biology. 1994; 14:8096–8106. [PubMed: 7969147]
4. Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proceedings of the National Academy of Sciences. 1996; 93:1156–1160.
5. Li T, et al. TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Research. 2010; 39:359–372. [PubMed: 20699274]
6. Christian M, et al. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics. 2010; 186:757–761. [PubMed: 20660643]
7. Cong L, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013; 339:819–823. [PubMed: 23287718]
8. Mali P, et al. RNA-guided human genome engineering via Cas9. Science. 2013; 339:823–826. [PubMed: 23287722]
9. Mao Z, Bozzella M, Seluanov A, Gorbunova V. Comparison of nonhomologous end joining and homologous recombination in human cells. Journal of Molecular Biology. 2008; 7:1765–1771.
10. Hartlerode AJ, Scully R. Mechanisms of double-strand break repair in somatic mammalian cells. Biochemical Journal. 2009; 423:157–168. [PubMed: 19772495]
11. Grindley NDF, Whiteson KL, Rice PA. Mechanisms of site-specific recombination. Annual Review of Biochemistry. 2006; 75:567–605.
12. Buchholz F, Stewart AF. Alteration of Cre recombinase site specificity by substrate-linked protein evolution. Nature Biotechnology. 2001; 19:1047–1052.
13. Sclimenti CR, Thyagarajan B, Calos MP. Directed evolution of a recombinase for improved genomic integration at a native human sequence. Nucleic Acids Research. 2001; 29:5044–5051. [PubMed: 11812835]
14. Santoro SW, Schultz PG. Directed evolution of the site specificity of Cre recombinase. Proceedings of the National Academy of Sciences. 2002; 99:4185–4190.
15. Akopian A, He J, Boocock MR, Stark WM. Chimeric recombinases with designed DNA sequence recognition. Proceedings of the National Academy of Sciences. 2003; 100:8688–8691.
16. Gordley RM, Smith JD, Graslund T, Barbas CF III. Evolution of programmable zinc finger-recombinases with activity in human cells. Journal of Molecular Biology. 2007; 367:802–813. [PubMed: 17289078]
17. Sarkar I, Hauber I, Hauber J, Buchholz F. HIV-1 proviral DNA excision using an evolved recombinase. Science. 2007; 316:1912–1915. [PubMed: 17600219]
18. Keravala A, et al. Mutational derivatives of PhiC31 integrase with increased efficiency and specificity. Molecular Therapy. 2009; 17:112–120. [PubMed: 19002165]
19. Gersbach CA, Gaj T, Gordley RM, Barbas CF. Directed evolution of recombinase specificity by split gene reassembly. Nucleic Acids Research. 2010; 38:4198–4206. [PubMed: 20194120]

20. Abi-Ghanem J, et al. Engineering of a target site-specific recombinase by a combined evolution- and structure-guided approach. Nucleic Acids Research. 2012; 41:2394–2403. [PubMed: 23275541]
21. Thyagarajan B, Guimaraes MJ, Groth AC, Calos MP. Mammalian genomes contain active recombinase recognition sites. Gene. 2000; 244:47–54. [PubMed: 10689186]
22. Thyagarajan B, Olivares EC, Hollis RP, Ginsburg DS, Calos MP. Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. Molecular and Cellular Biology. 2001; 21:3926–3934. [PubMed: 11359900]
23. Keravala A, et al. A diversity of serine phage integrases mediate site-specific recombination in mammalian cells. Molecular Genetics and Genomics. 2006; 276:135–146. [PubMed: 16699779]
24. Semprini S, et al. Cryptic loxP sites in mammalian genomes: genome-wide distribution and relevance for the efficiency of BAC/PAC recombineering techniques. Nucleic Acids Research. 2007; 35:1402–1410. [PubMed: 17284462]
25. Nelson HCM, Sauer RT. Interaction of mutant λ repressors with operator and non-operator DNA. Journal of Molecular Biology. 1986; 192:27–38. [PubMed: 2950238]
26. Sternberg N, Hamilton D. Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. Journal of Molecular Biology. 1981; 150:467–486. [PubMed: 6276557]
27. Hoess RH, Ziese M, Sternberg N. P1 site-specific recombination: nucleotide sequence of the recombining sites. Proceedings of the National Academy of Sciences. 1982; 79:3398–3402.
28. Van Duyne GD. A structural view of cre-loxp site-specific recombination. Annual Reviews of Biophysics and Biomolecular Structure. 2001; 30:87–104.
29. Hartung M, Kisters-Woike B. Cre mutants with altered DNA binding properties. The Journal of Biological Chemistry. 1998; 273:22884–22891. [PubMed: 9722507]
30. Missirlis PI, Smailus DE, Holt RA. A high-throughput screen identifying sequence and promiscuity characteristics of the loxP spacer region in Cre-mediated recombination. BMC Genomics. 2006; 7:73. [PubMed: 16595017]
31. Sheren J, Langer SJ, Leinwand LA. A randomized library approach to identifying functional lox site domains for the Cre recombinase. Nucleic Acids Research. 2007; 35:5464–5473. [PubMed: 17702764]
32. Rufer A, Neuenschwander PF, Sauer B. Analysis of Cre loxP interaction by surface plasmon resonance: Influence of spermidine on cooperativity. Analytical Biochemistry. 2002; 308:90–99. [PubMed: 12234468]
33. Jen-Jacobson L. Protein-DNA recognition complexes: conservation of structure and binding energy in the transition state. Biopolymers. 1997; 44:153–180. [PubMed: 9354759]
34. Guo F, Gopaul D, Van Duyne G. Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. Nature. 1997; 389:40–46. [PubMed: 9288963]
35. Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci USA. 2000; 97:8868–8873. [PubMed: 10908673]
36. Aranda M, et al. Altered directionality in the cre-loxP site-specific recombination pathway. Journal of Molecular Biology. 2001; 311:453–459. [PubMed: 11492999]
37. Livet J, et al. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. Nature. 2007; 450:56–62. [PubMed: 17972876]
38. Friedland AE, et al. Synthetic gene networks that count. Science. 2009; 324:1199–1202. [PubMed: 19478183]
39. Cornu TI, et al. DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. Molecular Therapy. 2007; 16:352–358. [PubMed: 18026168]
40. Ringrose L, et al. Comparative kinetic analysis of FLP and cre recombinases: mathematical models for DNA binding and recombination. Journal of Molecular Biology. 1998; 284:363–384. [PubMed: 9813124]
41. Segal DJ, Dreier B, Beerli RR, Barbas CF III. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences. Proceedings of the National Academy of Sciences. 1999; 96:2758–2763.

42. Meckler JF, et al. Quantitative analysis of TALE-DNA interactions suggests polarity effects. Nucleic Acids Research. 2013; 41:4118–4128. [PubMed: 23408851]
43. Miller JC, et al. An improved zinc-finger nuclease architecture for highly specific genome editing. Nature Biotechnology. 2007; 25:778–785.
44. Szczepek M, et al. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nature Biotechnology. 2007; 25:786–793.
45. Doyon Y, et al. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nature Methods. 2011; 8:74–79. [PubMed: 21131970]
46. Branda CS, Dymecki SM. Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. Developmental Cell. 2004; 6:7–28. [PubMed: 14723844]
47. Schmidt EE, Taylor DS, Prigge JR, Barnett S, Capecchi MR. Illegitimate Cre-dependent chromosome rearrangements in transgenic mouse spermatids. Proceeedings of the National Academy of Sciences. 2000; 97:13702–13707.
48. Heidmann D, Lehner CF. Reduction of Cre recombinase toxicity in proliferating Drosophila cells by estrogen-dependent activity regulation. Development Genes and Evolution. 2001; 211:458–465. [PubMed: 11685583]
49. Loonstra A, et al. Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. Proceedings of the National Academy of Sciences. 2001; 98:9209–9214.
50. Silver DP, Livingston DM. Self-excising retroviral vectors encoding the Cre recombinase overcome Cre-mediated cellular toxicity. Molecular Cell. 2001; 8:233–243. [PubMed: 11511376]
51. Huh WJ, Mysorekar IU, Mills JC. Inducible activation of Cre recombinase in adult mice causes gastric epithelial atrophy, metaplasia, and regenerative changes in the absence of ‘floxed’ alleles. American Journal of Physiology - Gastrointestinal and Liver Physiology. 2010; 299:G368–80. [PubMed: 20413717]
52. Feil R, et al. Ligand-activated site-specific recombination in mice. Proceedings of the National Academy of Sciences. 1996; 93:10887–10890.
53. Kellendonk C, et al. Regulation of Cre recombinase activity by the synthetic steroid RU 486. Nucleic Acids Research. 1996; 24:1404–1411. [PubMed: 8628671]
54. Jullien N, Sampieri F, Enjalbert A, Herman JP. Regulation of Cre recombinase by ligand-induced complementation of inactive fragments. Nucleic Acids Research. 2003; 31:e131. [PubMed: 14576331]
55. Liu Q, Li MZ, Leibham D, Cortez D, Elledge SJ. The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. Current Biology. 1998; 8:1300–1309. [PubMed: 9843682]
56. Choe W, Chandrasegaran S, Ostermeier M. Protein fragment complementation in M.HhaI DNA methyltransferase. Biochem Biophys Res Commun. 2005; 334:1233–1240. [PubMed: 16040000]
57. Gibson DG, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods. 2009; 6:343–345. [PubMed: 19363495]
58. Isaacs FJ, et al. Engineered riboregulators enable post-transcriptional control of gene expression. Nature Biotechnology. 2004; 22:841–847.
59. Sambrook, J.; Russell, DW. Molecular Cloning: a Laboratory Manual. 2001.
60. Werdien D, Peiler G, Ryffel GU. FLP and Cre recombinase function in Xenopus embryos. Nucleic Acids Research. 2001; 29:E53–3. [PubMed: 11376165]

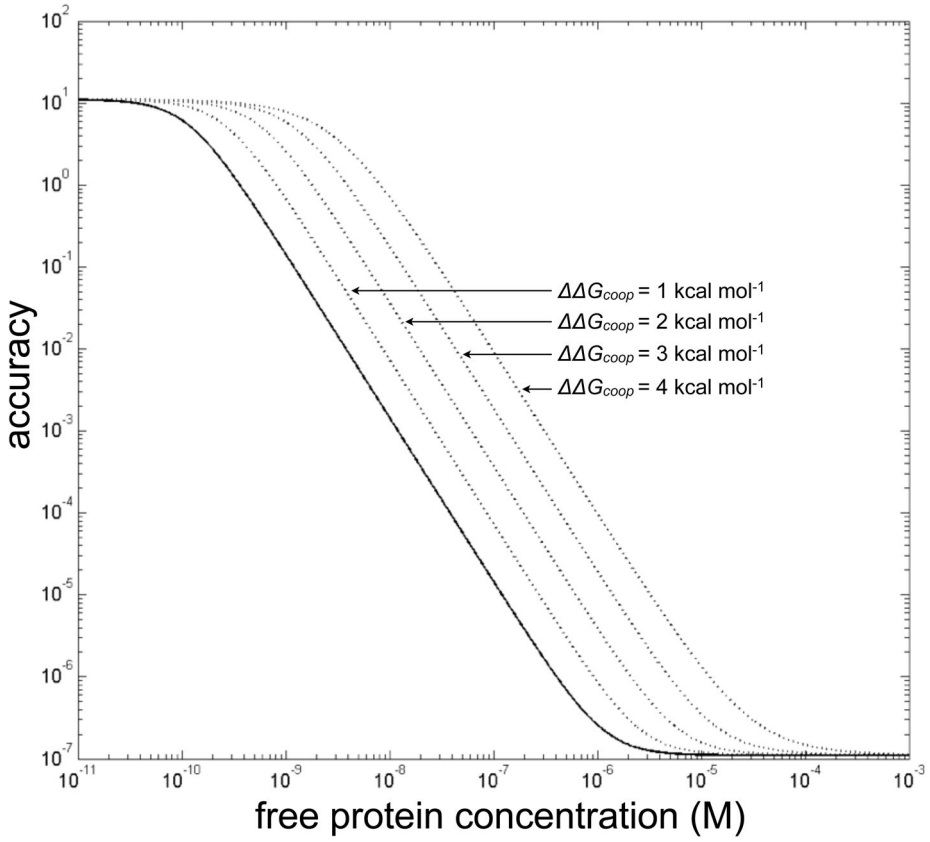


Figure 1. Model predicting an increase in Cre recombinase dimer binding accuracy as cooperativity decreases

The solid line indicates the accuracy predicted for WT Cre, while the dashed lines correspond to the expected accuracy of mutants in which the energy in the protein-protein interface has been reduced by the indicated amount.

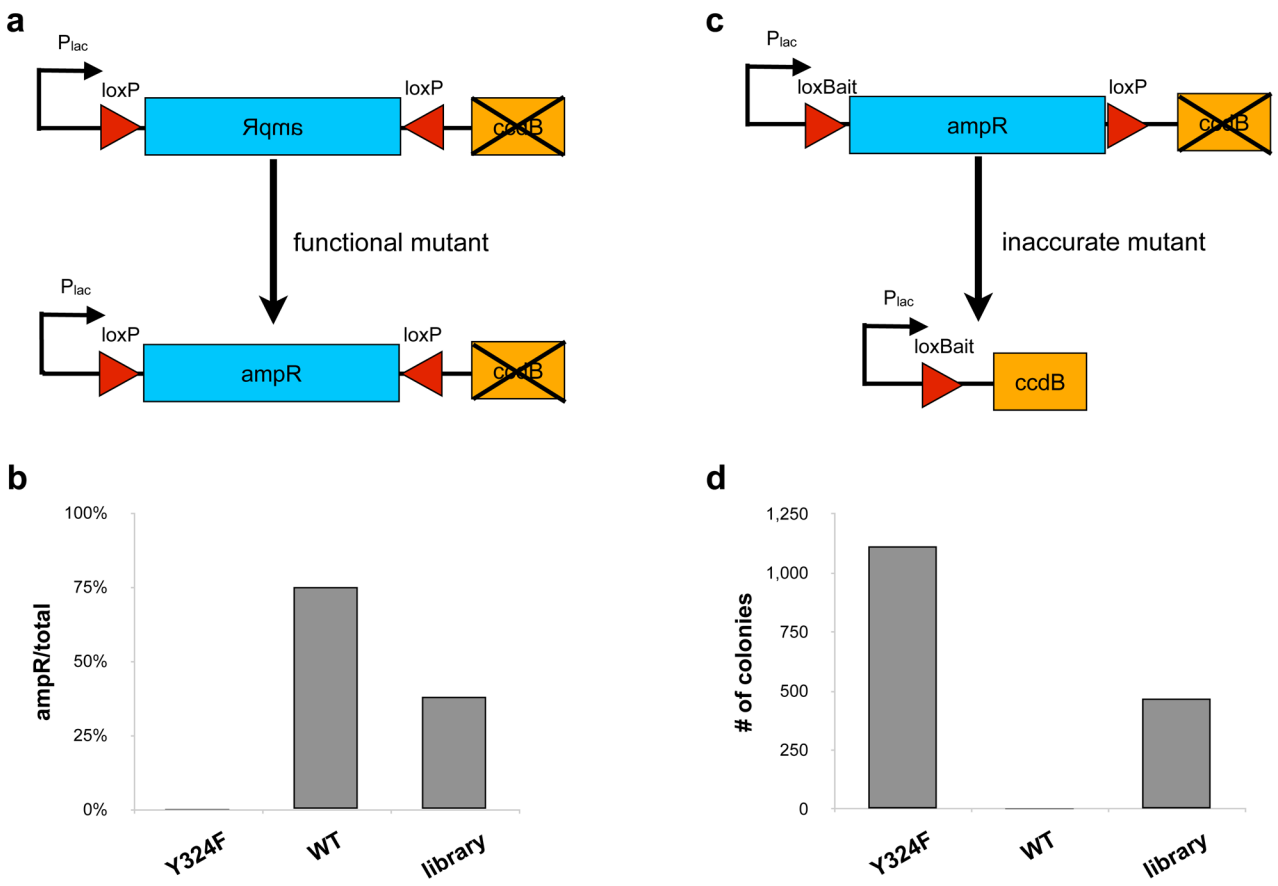
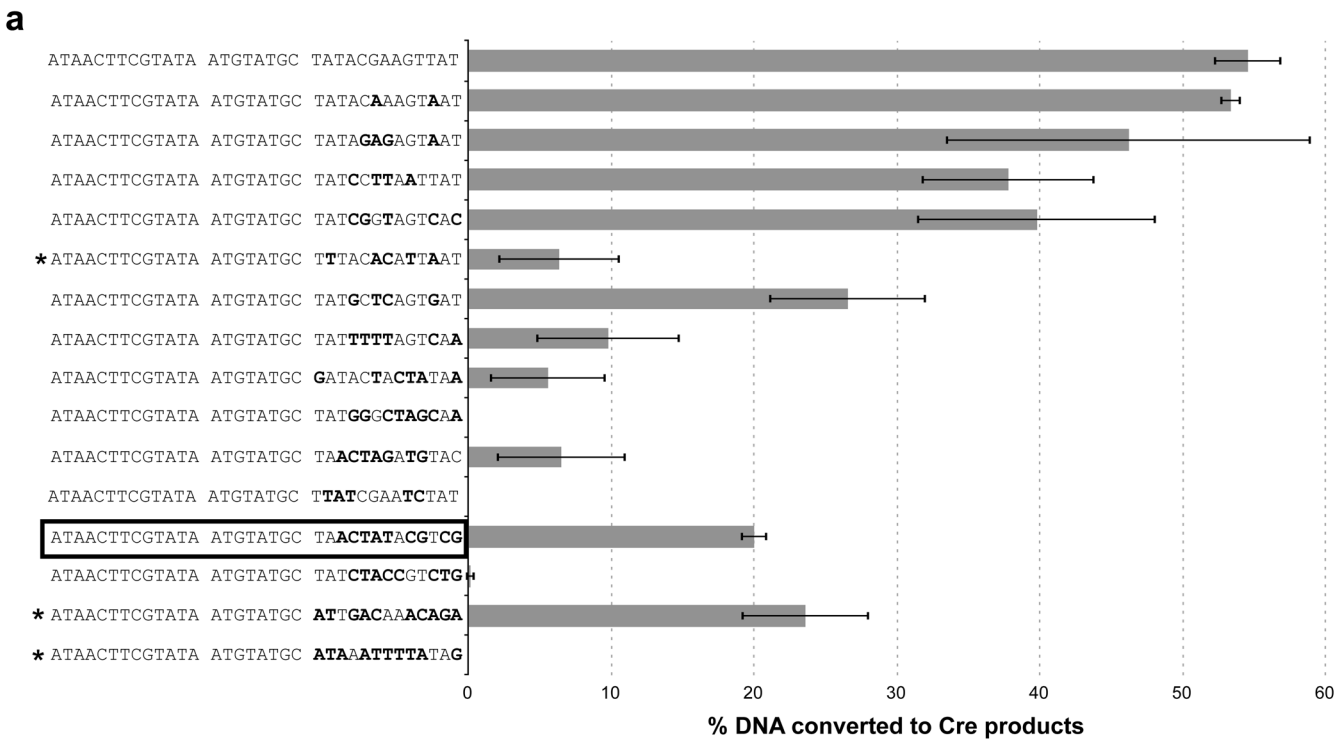


Figure 2. Selecting functional and accurate Cre variants

a) The substrate used to select for functional variants; proper recombination would place the ampicillin resistance gene (ampR) under the lac promoter (P_{lac})conferring resistance. The ccdB gene is crossed out to indicate that it's out of frame with respect to the ampR start codon. b) Ratio of ampicillin resistant to the total number of colonies isolated following the positive selection. c) The substrate used to select for accurate variants - recombination of *loxP* and *loxBait* sites would result in loss of ampicillin resistance and would place the toxic ccdB gene in frame with the promoter. d) Number of ampicillin resistant colonies recovered from the negative selection.



b

loxP	ATAACTTCGTATA ATGTATGC TATACGAAGTTAT
loxBait	ATAACTTCGTATA ATGTATGC TA**ACTAT**A**CG**T**CG**
ΨLox h7q21	ATA**CA**T**A**CGTATA **TATGTATA** TATAC**AT**A**TA**TAT
ΨCore h7q21	ATAACTTCGTATA **TATGTATA** TATACGAAGTTAT
lox80	ATAACTTCGTATA ATGTATGC TATAC**A**A**G**G**GCGA**

Figure 3. Activity and nomenclature of pseudo-*loxP* sites

a) Linear fragments with *loxP* onone end and the indicated sites on the other were treated with Cre and the products were quantified on an agarose gel. All new bands were counted towards the recombination efficiency. No recombination was observed for any of the sites in the absence of Cre. Bolded positions correspond to differences from *loxP*. Sites are listed in order of similarity with *loxP*. Asterisks mark sequences generated randomly; all others were obtained from a selection for functional pseudo-*loxP* sites. The box indicates *loxBait*. b) The names and sequences of recombination sites used in this study. Positions that are different from *loxP* are shown in bold. Error bars correspond to 95% C.I. (n = 2–3 experimental replicates).

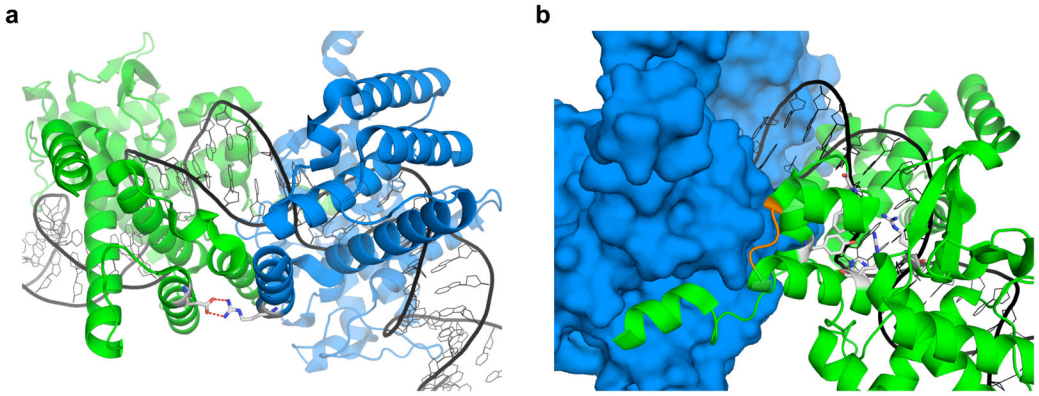


Figure 4. Structural context of the isolated mutants

a) R32V and R32M disrupt a putative salt bridge between two monomers (shown in blue and green) at R32 and E69. The two residues are shown as stick structures colored by atom identity (blue - N; red - O; gray - C). b) 303GVSdup duplicated the loop shown in orange. One of the monomers is shown as a space-filling model. The catalytic site residues (R173, H289, R292, W315, Y324) are shown as stick figures. The crystallographic data was obtained from PDB 3C29.

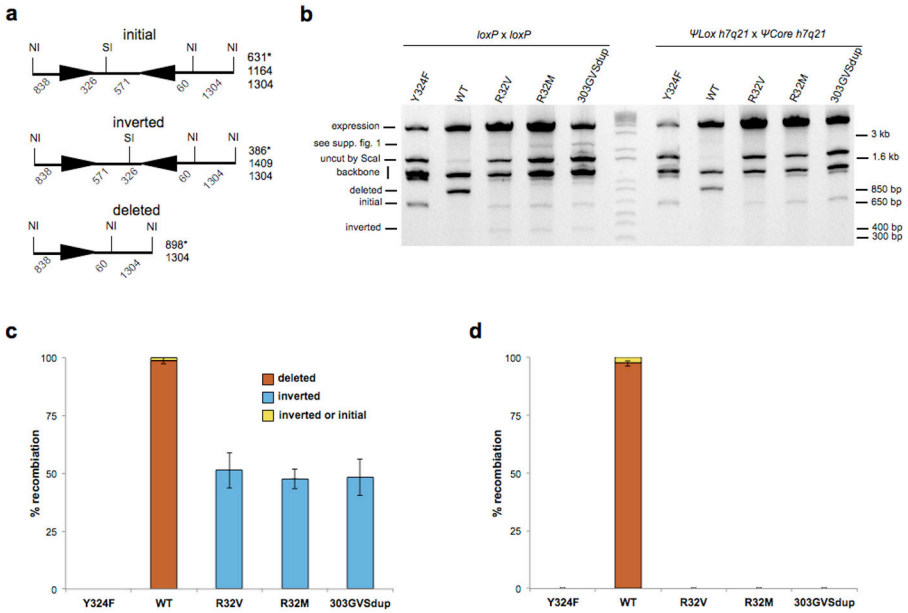


Figure 5. In vivo recombination of plasmids by mutants of Cre

a) Plasmid architecture of the three expected recombination products. Recombination sites are shown as triangles. For simplicity, the map is in linear form, where the NcoI sites at the ends represent a single NcoI site on the circular plasmids. The numbers under each segment indicate distance in bp (map not drawn to scale). The numbers adjacent to each map are the sizes of the expected digestion products, with the asterisks indicating the product size that is unique to the particular configuration. b) Digest analysis of *loxP* x *loxP* (left five lanes) and *ψlox h7q21 x ψCore h7q21* (right five lanes) recombination. (c& d) Inversion and recombination frequency of (c) *loxP* x *loxP* and (d) *ψCore h7q21* x *ψlox h7q21* recombination obtained by quantifying band intensities. Error bars correspond to 95% C.I. (n= 2 independent experiment). NI: NcoI site; SI: ScaI site.

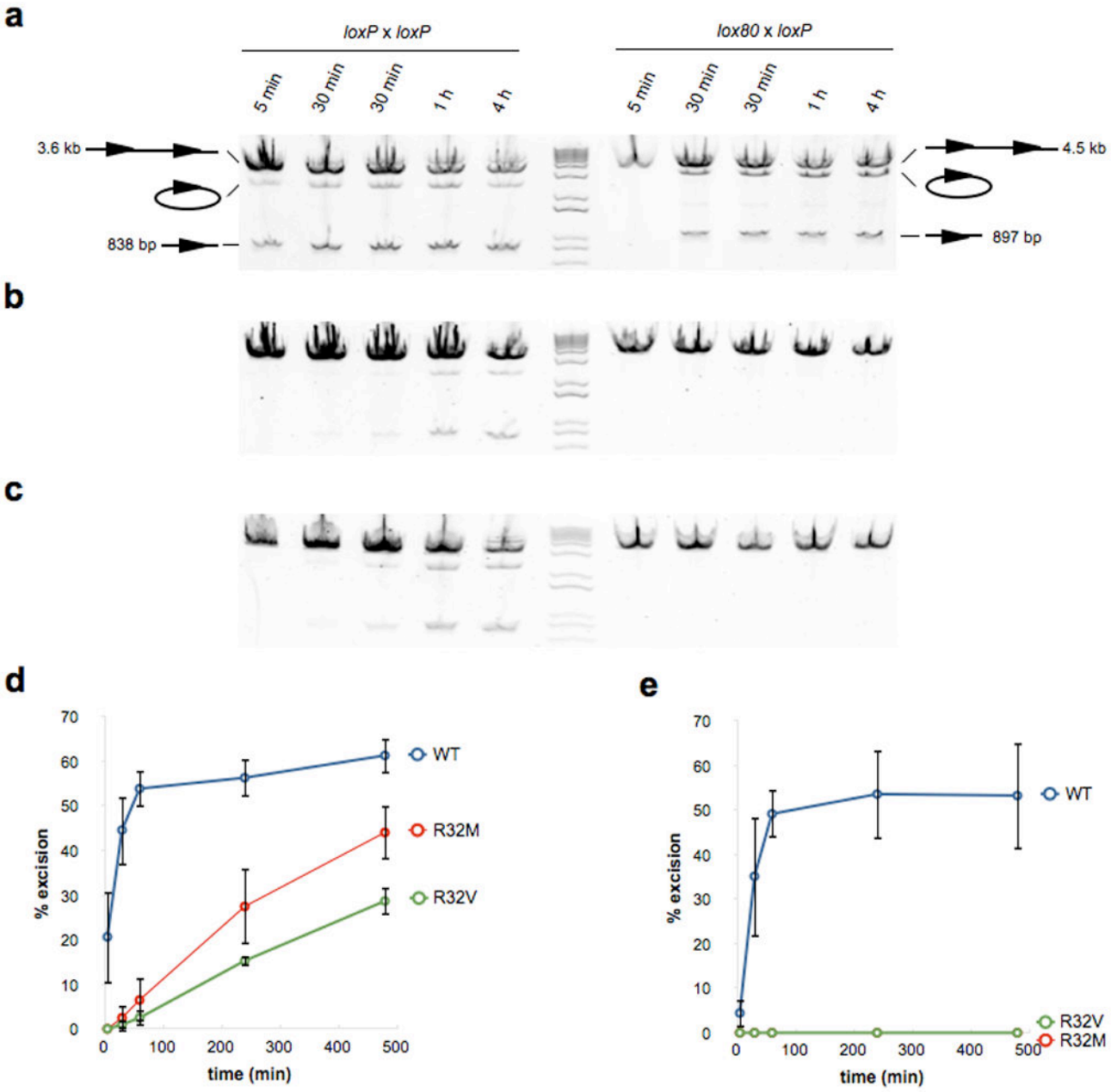


Figure 6. In vitro recombination of plasmids by mutants of Cre

a) WT Cre purification of *loxP* x *loxP* (left) and *lox80* x *loxP* substrates was performed for the indicated amount of time and resolved on an agarose gel. The linearized substrate plasmid contained two recombination sites such that catalysis resulted produced a circular and a short linear fragment. The same conditions were used to test b) R32V and c) R32M. The band intensities were quantified and plotted for d) *loxP* x *loxP* and e) *lox80* x *loxP* recombination. Error bars correspond to 95% C.I. (n = 2–4 experiments using different protein preparations).

Table 1
Genome-wide off-target integration frequency

Insertion frequency corresponds to the ratio of $loxP^+$ to $loxP^-$ colonies. 95% C.I. were computed using the Poisson variance. The data represent pooled observations from 3–4 independent experiments.

	***loxP*+ ori- kanR colonies**	***loxP*- ori+ kanR colonies**	***loxP* insertion frequency**	**95% C.I.**
Y324F	1	6.2×10^6	1.6×10^{-7}	$0 – 4.8 \times 10^{-7}$
WT	1,025	8.1×10^6	1.3×10^{-4}	$1.2 \times 10^{-4} – 1.4 \times 10^{-4}$
R32V	2	9.4×10^6	2.1×10^{-7}	$0 – 5.1 \times 10^{-7}$
R32M	3	6.3×10^7	4.8×10^{-8}	$0 – 1 \times 10^{-7}$
303GVSdup	46	2.6×10^7	1.7×10^{-6}	$1.2 \times 10^{-6} – 2.2 \times 10^{-6}$