A randomized library approach to identifying functional lox site domains for the Cre recombinase

Jamie Sheren*, Stephen J. Langer and Leslie A. Leinwand

University of Colorado, Department of Molecular, Cellular and Developmental Biology, 347 UCB, Boulder, CO 80309-0347

Received May 31, 2007; Revised July 9, 2007; Accepted July 25, 2007

ABSTRACT

The bacteriophage P1 Cre/loxP site-specific recombination system is a useful tool in a number of genetic engineering processes. The Cre recombinase has been shown to act on DNA sequences that vary considerably from that of its bacteriophage recognition sequence, loxP. However, little is known about the sequence requirements for functional lox-like sequences. In this study, we have implemented a randomized library approach to identify the sequence characteristics of functional lox site domains. We created a randomized spacer library and a randomized arm library, and then tested them for recombination in vivo and in vitro. Results from the spacer library show that, while there is great plasticity, identity between spacer pairs is the most important factor influencing function, especially in in vitro reactions. The presence of one completely randomized arm in a functional loxP recombination reaction revealed that only three wild-type loxP arms are necessary for successful recombination in Cre-expressing bacteria, and that there are nucleotide preferences at the first three and last three positions of the randomized arm for the most efficiently recombined sequences. Finally, we found that in vitro Cre recombination reactions are much more stringent for evaluating which sequences can support efficient recombination compared to the 294-CRE system.

INTRODUCTION

The bacteriophage P1 Cre/loxP site-specific recombination system is widely used as a genetic engineering tool (1,2) due to its well-defined recognition sequence, lack of any necessary co-factors and efficacy in both bacterial and eukaryotic systems. Since its discovery, Cre/loxP has been applied to temporal and spatial gene activation/deactivation (3–5), site-specific genomic integration and deletion (6–10) as well as the construction of libraries (11) and cloning strategies (12). Interest in expanding and improving the types of reactions that the Cre recombinase can facilitate has led to the discovery of numerous functional recognition sequences for Cre; however, no sequence characteristics have been defined to evaluate what constitutes a functional site other than its need to bind the recombinase and the role spacer compatibility plays in efficient recombination events (16,19,22).

The Cre recombinase acts on a 34-bp sequence known as loxP (13,14). This recognition sequence consists of two 13-bp palindromic arms separated by an 8bp spacer region (14,15). In a site-specific recombination reaction two loxP sites are brought together, each arm binding one recombinase monomer (16), while strand exchange takes place within the spacer regions (17). LoxP sequences are maintained following recombination events. This characteristic is responsible for the reversible nature Cre/loxP reactions.

Successful site-specific recombination has been shown to occur between sites with sequences having varying degrees of similarity to loxP (8,9,15,18–21). Beginning with the description of the Escherichia coli loxB site (15,18), several endogenous genomic sequences from different organisms have been discovered that can serve as substrates for Cre-mediated site-specific recombination (9,20,21). Sauer (20) first identified endogenous cryptic lox sites in Saccharomyces cerevisiae. These cryptic sites contained as few as 14 out of 34 bases in common with loxP, and illustrated the apparent importance of the TATA sequence adjacent to the spacer region in functional sites. Later, other functional genomic lox sites were reported for yeast, human and mouse (9,21). Together, the description of these functional sites demonstrates flexibility in sequence recognition by the Cre recombinase.

The search for endogenous lox sequences within genomes for use in genomic engineering and gene therapy has been a driving force in the discovery of alternative functional lox sites. However, diversification and manipulation of lox site behavior has also identified many alternative substrates for Cre. Albert et al. (8) were the...
first to screen for arm mutations that could facilitate stable Cre/loxP integration reactions; while mutation studies of the spacer (19,22) documented the effects of single and double base mutations on recombination efficiency. Recently, two different surveys of partially randomized spacer libraries (NNNTANNNN) (23,24) added to the growing list of spacer sequences that are proficient in recombination.

Previous studies have expanded the list of lox sites with which the Cre recombinase is able to recombine without pursuing a full-scale randomized approach to the 34bp sequence. Here, we report on the first randomized sequence studies aimed at defining functional lox sequences. We have created a randomized spacer library and a randomized arm library, both of which have been tested in vivo in Cre-expressing bacteria and in vitro reactions. The results of these studies indicate that Cre is very flexible in the sequences it can recombine. We show that lox sites with matching spacers recombine more effectively than sites with non-matching spacers and that the central ‘TA’ is not required for efficient recombination. Our results also indicate that there is a sequence bias in functional arms and that the mode of recombination (in vivo or in vitro) is important when evaluating site function.

MATERIALS AND METHODS
Constructing a randomized spacer library

The spacer library plasmid backbone (pLK) was engineered from the previously described pPG3-loxR/loxP plasmid (23) modified to replace its loxP site with a HindIII–NotI–BglII linker at its XbaI site.

Two separate syntheses of the library template (5′ GCCGAATTCTGCCGCATAACTCGTGATANNNNNNNNTATACGAAGTTATAGATAGGACGGATCCAGTTGG 3′) were done in order to compensate for non-random base frequencies in the oligonucleotide manufacturing process. PCR was used to create double-stranded DNA from the templates using primers EcoRI and BamHI (5′-CCGGCGAATTCTGCAGCATAACTTCGTATANNNNNNTATACGAAGTTATAGATAGGACGGATCCAGTTGG 3′) followed by PCR with primers Hind3KAN and BamHIKAN (arrows) to form kan/lox fragment. (B.2) The final spacer library with reiterated lox sites (segmented triangles) and complete kanamycin resistance marker after subcloning of the PCR product at BglII/HindIII is depicted. Arrows represent the positions of the PLK sequencing primer and SYBRL, R primers for PCR detection of recombination. (B.3) Cre-mediated site-specific recombination results in the deletion of the kanamycin resistance marker. amp′ = gene for β-lactamase, kan′ = kanamycin resistance marker, lacZ = gene encoding β-galactosidase.

Figure 1. Constructs for spacer library. (A) Sequence of the template oligonucleotides used to create either the loxP control library site or the spacer library site are shown. The position of each base in the spacer is numbered 1–8. N is any nucleotide (A, T, G or C). loxP sequences are underlined.

(B.1) The pLK plasmid with the spacer library cloned as an EcoRI/BamHI fragment (segmented triangle, degenerate segment filled-in) followed by PCR with primers Hind3KAN and BamHIKAN (arrows) to form kan/lox fragment. (B.2) The final spacer library with reiterated lox sites (segmented triangles) and complete kanamycin resistance marker after subcloning of the PCR product at BglII/HindIII is depicted. Arrows represent the positions of the PLK sequencing primer and SYBRL, R primers for PCR detection of recombination. (B.3) Cre-mediated site-specific recombination results in the deletion of the kanamycin resistance marker. amp′ = gene for β-lactamase, kan′ = kanamycin resistance marker, lacZ = gene encoding β-galactosidase.
In vitro recombination. pLK + library + Kan/Lox clone DNA was isolated by QIAprep spin miniprep column (Qiagen) and eluted in EB buffer (10 mM Tris–HCl, pH 8.5). Cre reactions were performed with 75 ng of plasmid DNA, 1U of MBP-Cre extract and 3 μl of 10× Cre buffer (500 mM Tris–HCl, pH 7.5; 330 mM NaCl and 110 mM MgCl2) in a 30 μl total volume. Reactions were incubated at 37°C for 15 min and then transferred for storage at −20°C.

Recombination was detected through high-speed PCR utilizing iProof High-Fidelity DNA Polymerase (BioRad). A standard 25 μl PCR mix was used containing 2 ng of template DNA taken directly from in vitro Cre reactions. The primers used were pLK SYBR R 5’ GAGATAGGG TTGAGTTTGTCC 3’ and pLK SYBR L 5’ GACCT ACACCGAAGTACACC 3’. Cycling conditions included an initial denaturation at 98°C for 30 s, 17 cycles of 98°C for 10 s, 67°C for 10 s and 72°C for 2 min followed by a final extension of 72°C for 4 min.

PCR samples were pre-incubated with SYBR Green I nucleic acid stain and visualized by agarose gel electrophoresis. Gels were scanned via a Storm PhosphorImager system (Molecular Dynamics) and analyzed with ImageQuant 5.1 software.

Constructing an arm library
In vitro recombination. To assay recombination for arm library clones in vitro, the lacZ/loxP fragment lost as a result of recombination was added back to randomly chosen clones (Figure 4A,3). The loxP/lacZ 3.7 kb fragment was isolated from pSV-β-Gal + loxP by HindIII/XbaI restriction digest. After treatment with Klenow, the lacZ/loxP fragment was blunt-end ligated into arm library clones filled-in following NdeI digest. Reconstructed clones were propagated in DH5α. Plasmid DNA from reconstructed clones was isolated via QIAprep spin miniprep column (Qiagen) and eluted in EB buffer (10 mM Tris–HCl, pH 8.5). In vitro reactions consisted of 0.1 μg of DNA, 1 U of MBP-Cre extract and 3 μl of 10× Cre buffer (500 mM Tris–HCl, pH 7.5; 330 mM NaCl and 110 mM MgCl2) in a 30 μl total volume. Reactions were incubated at 37°C for 15 min and then transferred for storage at −20°C. Recombination was detected by blue/white color screen in DH5α following transformation of 2 μl of the in vitro reaction.

Sequencing
All spacer library clones were sequenced with primers PLK 5’ TAAATGAGCATCATGTTGG 3’ and M13F-pUC (–40) 5’ GTTTTCCCCAGTCACGAC 3’. All arm library clones and reconstructed arm library clones were sequenced with primers KpnI3X 5’ CAGGATCCATGTTGG 3’ and BstBI 5’ CCCTGCTTCTCGAACTCG 3’ to create double-stranded DNA from the arm library template oligonucleotide 5’ CAGGATCCATGTTGG 3’.
sequenced with the pSVβ 5′ CGACTGGAAAGCGGT CAGTG 3′ primer. The arm library loxP site was sequenced with the pQEPromotor 5′ CCCGAAAG TGCCACCTG 3′ primer. The pMAL-Cre expression plasmid was sequenced with primers pTYB11 #5 5′ TG TCGAAATCAGTGCGTTCCG 3′, pTYB11 #4 5′ CGAG TGTAGAGCTGCTGTTGT 3′ and pTYB11 #3 5′ CGAAGCMTGTTTGCCCG 3′.

**Library sequence analysis**

Sequence pools from both the spacer and arm libraries were subject to $\chi^2$ analysis: $\chi^2 = \sum(O-E)^2$, where $O$ is the observed nucleotide value and $E$ is the expected nucleotide value.

**Purification of Cre recombinase**

Active Cre recombinase was purified through a maltose-binding protein (MBP) tag previously described by Kolb and Siedel (26). The cre gene was PCR amplified from the pMC-Cre plasmid (27) with primers Cre ATG 5′-GGGATCCATGCTGAGCTAATCGCCATCTTCCAGCAGGCG-3′, digested with XhoI and cloned in frame into the pMAL-C2x vector (New England Biolabs) digested with SalI. pMAL-Cre was then transformed into E. coli K12 ER2508 (New England Biolabs) for expression.

Three hours after induction of a 11 culture, bacteria were harvested by centrifugation and resuspended in 80 ml of column buffer (20 mM Tris–HCl, pH 7.4; 400 mM NaCl and 1 mM EDTA) and then frozen at −20°C. Cells were lysed by 3 cycles of freeze/thaw using an ice water bath. MBP-Cre was purified by affinity chromatography on an amylose resin column following New England Biolabs' product manual (pMAL Protein Fusion and Purification System Instruction Manual V. 5.1). Storage buffer (100% glycerol; 1 mM EDTA and 30 mM Tris–HCl, pH 7.5) was added on a 1:1 ratio to aliquots followed by storage at −20°C.

For our purposes, 1 U of purified recombinase is defined as the amount of enzyme necessary for maximum recombination after a 15 min incubation at 37°C of either 75 ng of spacer library control (pLK + LoxP+2) or 0.1 µg of the arm library control plasmid (pSV-β-Gal + LoxP+) in a 30 µl total reaction volume.

**RESULTS**

**In vivo recombination of a spacer library**

What are the sequence requirements for functional lox spacers? In order to investigate what constitutes a functional spacer sequence, we created two randomized spacer libraries from template oligonucleotides consisting of two loxP arms separated by an 8 (N) randomized spacer region (Figure 1A). Since it was already known that the most efficient recombination events usually take place between lox sites with matching spacer sequences (22), we reiterated the library sites in order to test the function between matching spacer pairs (Figure 1B).

Fifty-five clones (40 from group one and 15 from group two) were randomly chosen for reiteration. Following reiteration, clones were sequenced at each lox site and then passaged through 294-CRE E. coli cells, which constitutively express Cre recombinase. Successful site-specific recombination events resulted in the deletion of the kanamycin resistance gene (Figure 1B). Restriction digests with XmnI linearized both parental (5026 bp) and recombined plasmids (3707 bp). Clones exhibited either complete recombination, a mixture of parental and recombined plasmids, or no recombination product at all (Figures 2 and 3). In addition, two of the clones (#41 & #205) had an unknown product along with their recombination product (data not shown).

Sequencing revealed that not all clones contained lox sites with matching spacers (Figure 3). A subset of unmatched spacers was cloned for comparison to the matched spacer pairs. The remaining unmatched pairs were the result of the cloning process and were also included for comparison.

All clones with full-length matching spacers (33/33) recombined to completion with no parental plasmid detected. Many clones with unmatched spacers also recombined to completion (12/21) or had some visible recombination product (6/21). Clone #23 contained shortened 7 bp matching spacers and no observable recombination product. Overall, this in vivo assay for recombination appears to differentiate, in some cases, between matched and unmatched spacer pairs. It does not provide information on the efficiency of recombination between the various spacer pairs, which recombined to completion in vivo.

**Sequence analysis of spacer library clones**

$\chi^2$ analysis of all unique library sequences indicated that there were positions in both template libraries that were not randomly represented in terms of nucleotide frequency. Instead of each nucleotide having a frequency of 25% at each of the 8 positions of the spacer, positions 2, 3, 5, 7 and 8 in the first library all had significantly unequal base distribution (Supplementary Figure 1). The second library contained only one position (6), which had statistically unequal nucleotide distribution. This means that these libraries do not represent statistically random sequence populations. The non-random base distributions could be a result of the oligonucleotide manufacturing process and/or the manipulations (PCR, cloning and propagation in E. coli) necessary to clone the library oligonucleotide (28).

**In vitro recombination of a randomized spacer library**

The in vivo analysis of the spacer library did not demonstrate any differences in recombination based on nucleotide composition at specific positions within the spacer. Therefore, we assayed for recombination in vitro in an attempt to better distinguish functional diversity due to variation in spacer sequences. All spacer library clones were tested for function in vitro using purified Cre recombinase. Standard reactions were allowed to proceed for 15 min at 37°C. Recombination was detected by PCR followed by SYBR Green I staining of the PCR product and quantitation of band intensities on an agarose gel.
Figure 3 contains the averaged results of three separate in vitro recombination experiments.

The amount of Cre used in these reactions was titrated to give maximum recombination for the loxP control (pLK + loxP). Due to the reversible nature of the Cre/loxP system, there is less than 100% recombination for an in vitro deletion reaction irrespective of reaction time (29,30). The 70.4% maximal recombination achieved here mirrors previously published data for in vitro Cre reactions (29,30).

Compared to the recombination results in the 294-CRE bacterial system, the in vitro system appears to be more stringent in terms of the types of recombination events allowed and the amount of recombination product produced. For example, whereas the range of average recombination in vitro for the set of matching spacer clones was 56.7–1.7%, these clones were indistinguishable from one another in vivo. Non-matching spacer clones that had functioned in vivo did not necessarily function well in vitro; 7 of 12 clones that showed full recombination in vivo had <3.0% recombination under our in vitro conditions.

Therefore, not all clones that appeared to be as functional as the wild-type loxP control in vivo were equivalent when compared in vitro. The LoxP control averaged 70.4% in vitro recombination while the best library clone averaged 56.7%. Of the clones with full-length matching spacers (33), the group averaged 30.5% in vitro recombination, while clones with non-matching spacers averaged 4.6%.

These results suggest that in vitro Cre reactions are much less permissive in the types of spacer pairs that can undergo successful recombination compared to in vivo 294-CRE reactions. Matching spacer pairs are favored in recombination events, and while in vivo it is difficult to see any difference in recombination across sequences, the in vitro environment clearly discriminates between spacer sequences. Further analysis of these sequences is hampered by the small number of clones sampled.

In vivo recombination of an arm library

Our second basic question addressed what arm sequences the Cre recombinase can use to facilitate site-specific recombination events. Cre is known to bind cooperatively to its recognition sequence (31,32), and its contact positions within the arms have been mapped (33,34). By randomizing only one arm out of the four that make-up the Cre/lox recombinase, we could study what minimal functional sequences Cre can cooperatively bind and recombine.

In order to easily screen large numbers of randomized arm library sites, we constructed a plasmid in which color selection could be used to visualize recombination events in a bacterial system. A template oligonucleotide containing one loxP arm plus the loxP spacer followed by a 13 N randomized region was used to create the arm library (Figure 4B). Once the template was made double stranded through PCR, it was cloned into the β-galactosidase reporter construct pSV-β-Gal + loxP. This reporter construct had been modified with a loxP site at the 5' end and a loxP site flanking the kanamycin resistance marker. The sequences of spacers for clones from the first and second library template oligonucleotides are shown in descending order of percent in vitro recombination. Clones from the second library template oligonucleotide are given below the dotted line. Spacer sequences for the lox sites 5' to the kanamycin resistance marker and 3' of the kanamycin marker are given. For clones with matching lox spacers only the 5' site sequence is shown. For clones with unmatched spacers the sequence of both sites is shown. Spacers containing only 7 bases are marked with asterisks (*). Spacers sequences that appear more than once are underlined. The type of recombination products observed for in vivo reactions is noted in the middle column. All clones displayed either full recombination (F), a mixture of parental and recombination products (M) or no recombination products (N). The average percent recombination for in vitro Cre reactions is shown for each clone. The average and SD were calculated from three separate in vitro experiments.

<table>
<thead>
<tr>
<th>clone</th>
<th>5' lox site spacer</th>
<th>3' lox site spacer</th>
<th>in vivo products</th>
<th>% recombination</th>
</tr>
</thead>
<tbody>
<tr>
<td>LoxP</td>
<td>GCATACT</td>
<td>same</td>
<td>F</td>
<td>70.4 ± 2.0</td>
</tr>
<tr>
<td>39</td>
<td>GGAGGGGT</td>
<td>same</td>
<td>F</td>
<td>54.7 ± 10.0</td>
</tr>
<tr>
<td>33</td>
<td>TCTGGGCC</td>
<td>same</td>
<td>F</td>
<td>48.1 ± 5.1</td>
</tr>
<tr>
<td>9</td>
<td>TCTGGCCG</td>
<td>same</td>
<td>F</td>
<td>47.1 ± 10.3</td>
</tr>
<tr>
<td>28</td>
<td>CCGTGGCG</td>
<td>same</td>
<td>F</td>
<td>45.2 ± 2.6</td>
</tr>
<tr>
<td>14</td>
<td>ACTAAAGC</td>
<td>same</td>
<td>F</td>
<td>37.9 ± 2.6</td>
</tr>
<tr>
<td>29</td>
<td>AGAATAC</td>
<td>same</td>
<td>F</td>
<td>34.0 ± 10.3</td>
</tr>
<tr>
<td>592</td>
<td>GCCGGGGGC</td>
<td>same</td>
<td>F</td>
<td>33.1 ± 3.5</td>
</tr>
<tr>
<td>64</td>
<td>GCGGTGTA</td>
<td>same</td>
<td>F</td>
<td>27.6 ± 6.0</td>
</tr>
<tr>
<td>63</td>
<td>AGGAGGTA</td>
<td>same</td>
<td>F</td>
<td>24.5 ± 7.3</td>
</tr>
<tr>
<td>37</td>
<td>CAGCGC</td>
<td>same</td>
<td>F</td>
<td>22.7 ± 2.6</td>
</tr>
<tr>
<td>25</td>
<td>CGATTGGA</td>
<td>same</td>
<td>F</td>
<td>19.5 ± 10.6</td>
</tr>
<tr>
<td>4</td>
<td>TTCTGTA</td>
<td>same</td>
<td>F</td>
<td>18.1 ± 16.9</td>
</tr>
<tr>
<td>10</td>
<td>GCTATGCT</td>
<td>same</td>
<td>F</td>
<td>17.8 ± 5.3</td>
</tr>
<tr>
<td>412</td>
<td>ACCGATCG</td>
<td>same</td>
<td>F</td>
<td>17.7 ± 7.6</td>
</tr>
<tr>
<td>7</td>
<td>AGGGAGGC</td>
<td>same</td>
<td>F</td>
<td>16.9 ± 13.4</td>
</tr>
<tr>
<td>512</td>
<td>GGGCGCGC</td>
<td>same</td>
<td>F</td>
<td>16.3 ± 1.6</td>
</tr>
<tr>
<td>17</td>
<td>TGGTGCGC</td>
<td>same</td>
<td>F</td>
<td>7.5 ± 4.2</td>
</tr>
<tr>
<td>6</td>
<td>GGCGCGTG</td>
<td>same</td>
<td>F</td>
<td>6.2 ± 1.1</td>
</tr>
<tr>
<td>21</td>
<td>GCCGATAG</td>
<td>same</td>
<td>F</td>
<td>6.9 ± 8.3</td>
</tr>
<tr>
<td>3</td>
<td>TTCGGTCG</td>
<td>same</td>
<td>F</td>
<td>1.7 ± 1.6</td>
</tr>
<tr>
<td>23</td>
<td>TAGCTGG</td>
<td>same N</td>
<td>0.0 ± 0.0</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>TTGGTGCG</td>
<td>same N</td>
<td>22.0 ± 3.7</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>CTACCGGGCGGAGG</td>
<td>M</td>
<td>18.9 ± 2.7</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>GTGCTGGC TTGAAAT</td>
<td>F4</td>
<td>14.1 ± 5.0</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>AGTGCGG GCCGGGGC</td>
<td>M</td>
<td>12.6 ± 6.9</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>GTCCGGTC CGGTCG</td>
<td>F</td>
<td>5.2 ± 1.1</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>AGCGTGC ATACGGC</td>
<td>F</td>
<td>2.8 ± 0.7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>TATCCTC CGGTGTT</td>
<td>F</td>
<td>2.6 ± 0.1</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>TTTCTGGC GGGGAC</td>
<td>M</td>
<td>2.4 ± 1.1</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>GCGGCGGGC GTGGTCTG</td>
<td>F</td>
<td>2.4 ± 1.8</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>GGTGGGC AGGCGC</td>
<td>N</td>
<td>1.6 ± 1.1</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>GTGACGCA ACCGAGTC</td>
<td>M</td>
<td>1.2 ± 1.1</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>TACGGCGG CGGCGGGG</td>
<td>M</td>
<td>0.9 ± 0.2</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>TGAGATG CCCGGGCG</td>
<td>N</td>
<td>0.8 ± 0.6</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>TGGAGG GCCGGTAT</td>
<td>F</td>
<td>0.6 ± 0.8</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>GGGCGGGT TTTTGGGC</td>
<td>F</td>
<td>0.6 ± 0.4</td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>GGGTGGCG GTGGTGTG</td>
<td>M</td>
<td>0.4 ± 0.6</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>GCGGGGT CTATCCAG</td>
<td>F</td>
<td>0.3 ± 0.5</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>GGGCTACT CACGTGAT</td>
<td>N</td>
<td>0.0 ± 0.0</td>
<td></td>
</tr>
<tr>
<td>492</td>
<td>CGGCTAGG CGGGCGG</td>
<td>F</td>
<td>0.0 ± 0.0</td>
<td></td>
</tr>
</tbody>
</table>

206 GTCTGTA | same | F | 50.7 ± 2.9 |
270 CCAGATG | same | F | 53.7 ± 1.9 |
271 CAAATCTC | same | F | 52.3 ± 3.7 |
268 CGAGATT | same | F | 51.9 ± 4.2 |
267 GTATCTC | same | F | 50.7 ± 3.1 |
207 TCTTGACT | same | F | 45.7 ± 6.2 |
209 TCCCAGT | same | F | 42.9 ± 10.1 |
205 TAAGTACT | same | F | 39.6 ± 8.9 |
202 GTCTGCAA | same | F | 32.6 ± 9.0 |
203 CGTGACG | same | F | 32.2 ± 2.1 |
208 TAAAGGCC | same | F | 23.4 ± 4.7 |
266 CATAAAGT | same | F | 13.1 ± 4.0 |
204 GAAGCCGT | same | F | 9.1 ± 4.2 |
201 CAGAAGA TACGAC | F | 4.6 ± 2.3 |
205 CTAATCTA GAGTTAAA | M | 2.6 ± 0.5 |

In vivo recombination of an arm library

Our second basic question addressed what arm sequences the Cre recombinase can use to facilitate site-specific recombination events. Cre is known to bind cooperatively to its recognition sequence (31,32), and its contact positions within the arms have been mapped (33,34). By randomizing only one arm out of the four that make-up the Cre/lox recombinase complex, we could study what minimal functional sequences Cre can co-operatively bind and recombine.
end of the lacZ coding region. With the arm library site in place at the 5' end, the lacZ gene was flanked by lox sites (Figure 4A.1). Following transformation into 294-CRE cells, any site-specific recombination would result in removal of lacZ and loss of β-galactosidase activity (Figure 4A.2).

Hundreds of colonies were screened by blue/white color selection via seeding of transformations on LB/X-gal agar plates. After overnight growth at 37°C, ~87.9% of colonies were stained dark blue, 3.5% were white with blue centers (white w/blue) while 8.6% were white only (data not shown). The pSV-β-Gal + LoxP2 control in 294-CRE resulted in white colonies only, while blue-stained colonies were observed for both control and arm library constructs in Cre/C0 DH5α E. coli.

It was assumed that blue colonies would contain only unrecombined, non-functional arm library sites, while white colonies would be the result of successful site-specific recombination between functional lox sites. White colonies with blue centers were unexpected, but could be explained by a slower rate of recombination, relative to the phenotypically white colonies, for the cells that seeded these colonies.

Upon examination of plasmid DNA from each type of colony (Figure 5), it appeared that all colonies contained DNA that had undergone successful recombination events. Preliminary analysis showed that all plasmid DNAs linearized with XhoI were the size of the deletion product. However, full-length parental product could be visualized for blue colonies after overloading the restriction samples on an agarose gel (data not shown). Retransforming blue colony DNA into DH5α (−Cre) cells resulted in a 1:75 ratio of blue to white colored colonies, pointing to a small population of unrecombined plasmid DNA present after ~48 h of growth in 294-CRE.

It was unclear whether blue clones had single or multiple library constructs present within a single clone. If so, this could account for the continued presence of a full length, non-functional plasmid. To further investigate, two blue clones (#3 and #5) were chosen randomly for retransformation of their plasmid DNA into DH5α followed by isolation and sequencing of resulting white and blue colony DNAs. Plasmids from these DH5α colonies were the expected sizes (blue colonies 6.5 kb, white 2.7 kb, data not shown). Sequencing revealed that both blue and white colonies were derived from the same arm library construct. It was concluded that each original 294-CRE blue clone contained a single arm library construct.

Figure 4. Arm library construction. (A.1) Plasmid pSV-β-Galactosidase with a loxP site at NotI/BamHI (segmented triangle) and the arm library (segmented triangle, degenerate arm shaded) PCR product cloned within the added Neol/HindIII linker as a BstBI/KpnI fragment. Arrows depict the positions of primers for inverse PCR (Kpn inver 1 and 2) and sequencing (PSVβ). (A.2) The result of Cre-mediated site-specific recombination is a 2727 bp plasmid containing the library arm within its lox site. Selected recombination products were chosen to be rebuilt by adding back the lacZ/loxP segment lost to recombination. (A.3) The resulting construct was then used in in vitro recombination reactions. (B) Sequences of the control loxP library site and the arm library template are shown. LoxP sequences are underlined. The degenerate arm sequence is shaded and a number is given to each base. Position 1 starts with the 3' nucleotide of the arm and continues towards the 5' end in order to position 13. amp r = gene for β-lactamase, lacZ = gene encoding β-galactosidase.

Figure 5. In vivo recombination. Arm library clones were chosen at random from 294-CRE colonies that resulted from the transformation of the pSV-β-Gal + LoxP + arm library ligation. Examples of XhoI restriction digests of plasmid DNAs visualized by agarose gel electrophoresis are shown. Full-length constructs were 6509 bp, while site-specific recombination resulted in a 2727 bp product. M = marker, C = control/−Cre, W = white colony, W/B = white colony with blue center, B = blue colony.
Sequencing was also used to identify clones, which might have multiple library constructs present. Background sequence seen on the sequence chromatograms only at the 13 base randomized arm region could be a result of more than one sequence present within a clone. Only 2 blue clones (#16 and #18) out of 37, along with 2 white with blue center clones (#5 and #12) out of 15, had strong evidence from their sequence chromatograms to suggest there might be multiple arm library constructs present in each clone. Evidence of unrecombined products could also be seen in sequence chromatograms, as slightly more than half (20 of 37) of the blue clones had strong background sequence at the recombination junction.

Finally, growth time could play a role in accumulation of recombination products and help explain the observed β-galactosidase activity in both blue and white with blue center colonies. If so, serial culturing of blue 294-CRE clones over several days would result in total loss of β-galactosidase activity. To test this, 10 blue colonies were chosen at random for extended serial culture. Following 5 days of continuous growth culture, colonies no longer stained positive for β-galactosidase when plated on LB agar containing X-gal (data not shown). The mechanism by which small numbers of unrecombined plasmids persist over time in the presence of the recombinase is unknown.

Sequence analysis of arm library clones

In total, 93 clones were sequenced (Figure 6). Forty-one sequences were from white colonies, 37 were from blue and 15 were from white colonies with blue centers. A random sequence population was derived from the ratios of blue:white:blue w/blue colonies (~37:4:1) observed upon plating of the library (Supplementary Figure 2). χ² analysis of the random library sequence population revealed again that the base distributions at several positions of the library template oligonucleotide were not statistically 25% for each nucleotide (Supplementary Figure 3). Positions 1, 2, 4, 5, 8, and 9 had significantly skewed base frequencies. Again, these nucleotide frequencies could be due to the manufacture of the oligonucleotide and/or the cloning process.

Despite the non-random nature of the arm library template, it is still possible to analyze the sequences from the efficiently recombined white clones to the overall library population from a statistical standpoint. In order to compare the arm sequences of white clones to that of the random library population, the frequencies for the expected (E) base distribution in the χ² analysis [χ² = Σ((O - E)²/E)] were taken directly from the random sequence population calculated above.

χ² analysis on all 41 white clone arm sequences (Supplementary Figure 3) showed statistically significant nucleotide changes at 6 arm positions. Positions 1, 2, 3, 11, 12 and 13, translating into the first 3 and last 3 arm positions, were statistically different from the random set population. Comparison of the nucleotide frequencies at these 6 positions shows a trend towards the sequence ‘ATA’ at both ends of the library arm for white colony clones (Supplementary Figure 4). This ‘ATA’ sequence is also present in the wild-type loxP site. While 5 of the 6 biased positions are known contacts for the recombinase, position 3 has not been previously identified as an important position for functional arm sequences. These data indicate that there are preferences at the ends of the arm for efficiently recombined lox sites.

Though the total number of sequences in the white colonies with blue centers group is low for statistical analysis, it too shows significant nucleotide distribution changes at positions 11 and 12 (Supplementary Figure 3). Comparison of the nucleotide frequencies at these two positions shows a trend towards the sequence ‘AT’ (Supplementary Figure 5). This, along with the simple calculation of the average number of conserved bases (Figure 6) for each colony type group, points to a continuum for the number of conserved bases in an arm relative to its probable function. The more bases conserved overall, and the more bases conserved at the ends of the arm, the more likely an arm is to recombine efficiently in the 294-CRE system.

Arm library in vitro recombination

Several blue and white clones were chosen at random to be tested in in vitro Cre reactions. The process of adding back the lacZ gene plus the loxP site resulted in a construct (Figure 4A.3) that was almost identical to the pSV-β-Galactosidase + loxP + arm library construct before it underwent recombination (Figure 4A.1). In vitro reactions were carried out as described for the spacer library except the amount of Cre input was titrated for maximum recombination of 0.1 µg of the arm library control, pSV-β-Gal + loxP². Recombination was detected via color selection following transformation of 2 µl of the in vitro reaction into DH5α cells and plating on LB/X-gal agar.

On average, a total of 878 colonies were counted for each reconstructed clone encompassing three replicates of their in vitro reactions. The loxP control reaction reached a recombination level of 66.0 ± 3.1%. Overall, there was a striking difference in recombination between constructs from white versus blue colonies (Figure 6). No reconstructed blue clone achieved >0.6% recombination in vitro on average, while a full range of total recombination (1.4–54.6%) was observed for white reconstructed clones.

These in vitro results are more in line with what was predicted for white and blue clones. Clearly, blue clone arm sequences do not support effective recombination in vitro, and as might be expected, average fewer conserved loxP bases (2.9, Figure 6). Supporting the idea that clones from white colonies with blue centers have characteristics that are part of a continuum between blue and white clones, their average number of conserved bases (4.9) falls between that of the blue and white clones. On average, white clone arms have almost twice as many conserved (both contact, 4.2 and non-contact, 5.4) bases compared to the blue clones.

It has already been noted that white clone arms tend to have conserved ‘ATA’ motifs at each end. Drawing precise conclusions about base conservation or preference in relation to percent recombination in vitro is precluded by...
the small overall sample size. However, it is interesting to note that the two best performing arms tested (White #2 and #8) maintained a conserved ‘TATA’ sequence next to the spacer.

**DISCUSSION**

We have demonstrated the utility of a randomized library approach to the understanding of the nature of the functional domains of the Cre recombinase recognition sequence, as well as the distinct differences in function of lox sequences between in vivo and in vitro environments. Although this study did not test a large sampling of all sequences possible in each library (spacer = 4 and arm = 4), its results are significant. By electing to randomize whole sections of the lox site, we have observed a great range of functional sequences, function tied to recombination environment...
and the first unbiased, systematic evaluation of lox arm sequences.

Two previous studies have concentrated on the sequence of the spacer region and its role in recombination efficiency. Lee and Saito (22) undertook the task of creating single and double-base mutations of the spacer and noted variability in the recombination efficiency across mutants. Several of their observations are supported by our work such as low or undetectable recombination in vitro between unmatched spacers, variable recombination between like mutant spacers, and no mutant spacer with equal or greater recombination efficiency than the wild-type loxP spacer.

Our study is, however, in conflict with a published report from Missirlis et al. (24). While we agree that mutant spacers and sites with mismatched spacers do recombine, our observations do not substantiate Missirlis et al.’s conclusion that a G-rich spacer is the most favored for recombination or that self-recombination is not significantly greater than recombination with other spacers. Though we cannot rule out the possibility of base preferences within the spacer for efficiently recombined sites, G-rich spacers were not the most effective in our data set.

We observed frequency imbalances in our libraries that were the result of the oligonucleotide manufacturing process and/or the cloning of the library. These biases were not from clonal selection since all clones in our study were chosen at random for further manipulation. By sequencing only successful recombination products, Missirlis et al. leave their study open to the possibility that their analyzed pool of products did not originate from a truly random population. Without published analysis of the pool of clones created from the PCR and ligation of the input oligonucleotides, it is unclear whether their observations are the result of nucleotide frequency biases present in the library clones or true base preferences of the Cre recombinase.

Second, we believe our observations support the general conclusion that recombination between matching spacers is more efficient in vitro and more likely to proceed to completion in vivo than recombination between mismatched spacer pairs. Based on our observations, we believe that Missirlis et al.’s conclusions based on equating efficiency of recombination with the number of times a recombination event appears in a shotgun library in vitro experiment could be misleading. This could easily be determined by testing Missirlis et al.’s spacer sequences in traditional in vitro and in vivo assays.

In our data set, the non-matching spacer clone (#42) with the highest percent recombination in vitro (22.0%) contained sites with homologous spacers with the exception of one position. The remaining unmatched spacer clones all had <20% recombination in vitro, compared to 21 of 33 matching spacer clones with >20% recombination. All matched spacer pairs recombined to completion in vivo compared to 12 of 21 unmatched.

This is the first study to report on a fully randomized (8N) spacer library. Results from this study show there to be great sequence plasticity for successfully recombined spacers. In past work, importance was given to the role of the central ‘TA’ dinucleotide as critical for recombination (19,22). Our work demonstrates that a central ‘TA’ is not necessary to achieve >45% recombination in vitro or complete recombination in 294-CRE. Also, complete identity through the central six bases is not required for recombination as previously thought (22). Unfortunately, our sample set was too small to statistically evaluate for base preferences in efficiently recombined spacers.

Results from both spacer and arm libraries suggest that recombination in the 294-CRE strain of E. coli is robust compared to in vitro recombinase reactions. Whether recombination is aided by replication and/or endogenous E. coli factors is unclear. This is an important issue to consider when testing for the function of new lox sites or libraries, as recombination in 294-CRE may not transfer well to the in vitro environment. It should also be noted that recombination efficiency in other in vivo systems might differ significantly from both the E. coli and in vitro systems tested here.

Though our results from the spacer library study were not unexpected, the fact that our arm library demonstrated effective recombination with the presence of one randomized arm was surprising. It appears that an arm sequence with as few as one conserved loxP base can comprise 1 of the 4 lox arm sequences necessary for recombination in 294-CRE, and as few as 4 of 13 conserved bases were sufficient for recombination in vitro. This could be explained by Cre’s cooperative binding properties, where three bound recombinase monomers might be able to recruit the final subunit to the complex leading to recombination in the absence of a fourth arm sequence that has any resemblance to loxP.

Our arm library testing system proved useful in the screening of hundreds of clones at once and in identifying the minimal arm sequences required for recombination. Despite not working with a truly random library, statistics reveal nucleotide preferences at 6 out of 13 arm positions for efficiently recombined clones. The ‘ATA’ motif at both ends of the arm was overrepresented in sequences from efficiently recombined clones. Though other studies have looked at the effects of base mutations on the arms (35,36), Sauer (20) was the first to speculate on the importance of the ‘TATA’ sequence flanking the spacer region. Ours is the first study to find evidence of base preferences for functional arm sequences and disproves the notion that bases at the outer end of the arm are not important for efficient recombination.

Clearly, the results from both libraries must be interpreted with the fact that they were tested in the context of the wild-type loxP sequence. The arm library in particular identifies the weakest arm sequences that can be tolerated in an otherwise wild-type reaction. Mutations in the loxP arm or spacer coupled to the fully randomized arm might exert greater selective pressure to conserve loxP bases in the library arm.

In the future, we plan to continue to test the function of various lox site libraries. How mutations in the spacer and in both arms combine to affect function has not been explored. Information from this study and future work would be of use in the expansion of the library of
functional lox sequences, the search for functional endogenous lox sequences in genomes, finding non-compatible lox site pairs, and the creation of sites capable of stable recombination events.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS
The authors would like to thank Elizabeth Ketner for construction of the pLK plasmid. Funding to pay the Open Access publication charges for this article was provided by NIH RO1 HL56510.

Conflict of interest statement. None declared.

REFERENCES