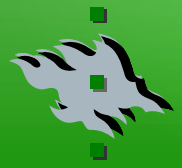


Mutagenesis

Types of mutations you may want to introduce:

- Point mutations
- Deletions
- Insertions

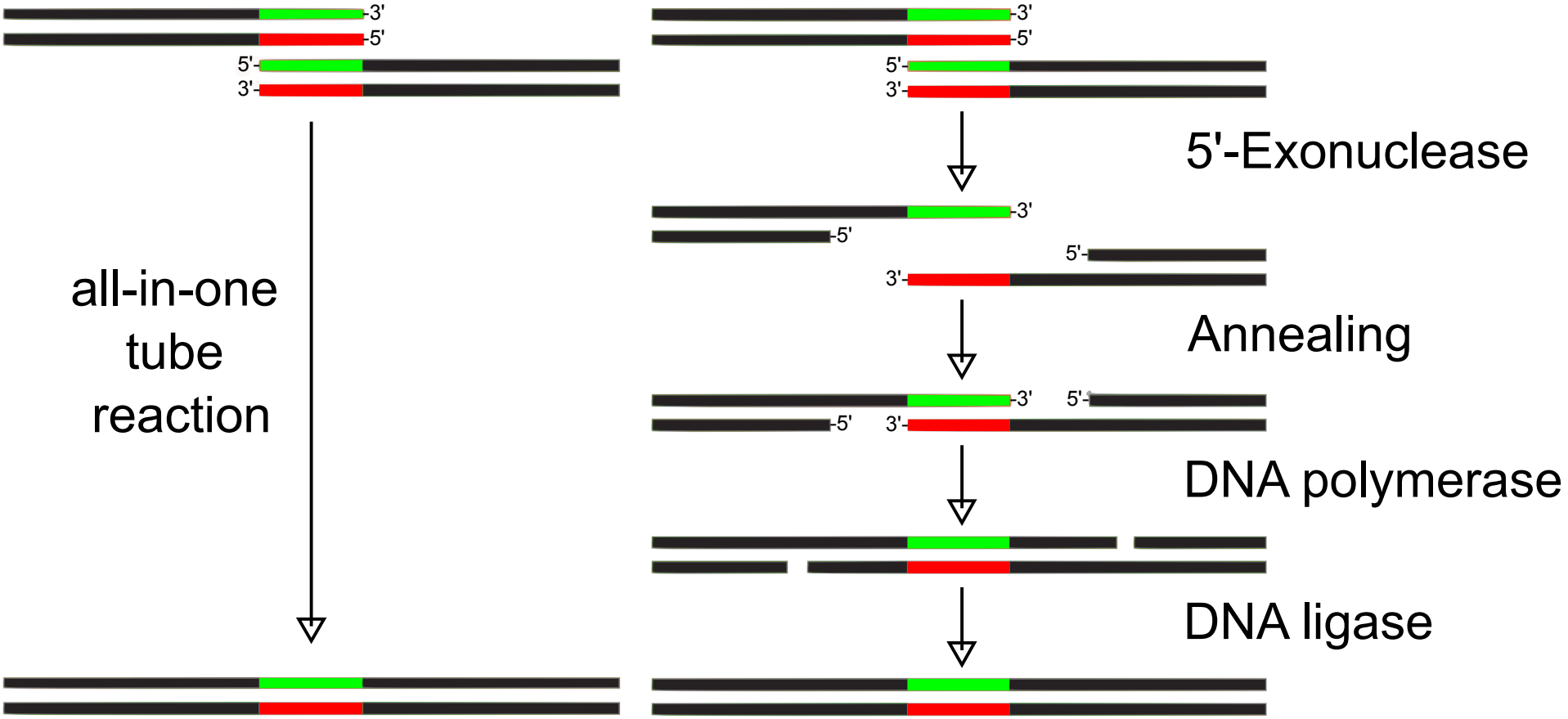
Technically, cloning an insert into a vector is equivalent to insertional mutagenesis. Mutagenesis is frequently done using “new cloning methods”, hence we have to introduce them before we talk about mutagenesis and in-vitro evolution.

“New” “cloning” techniques

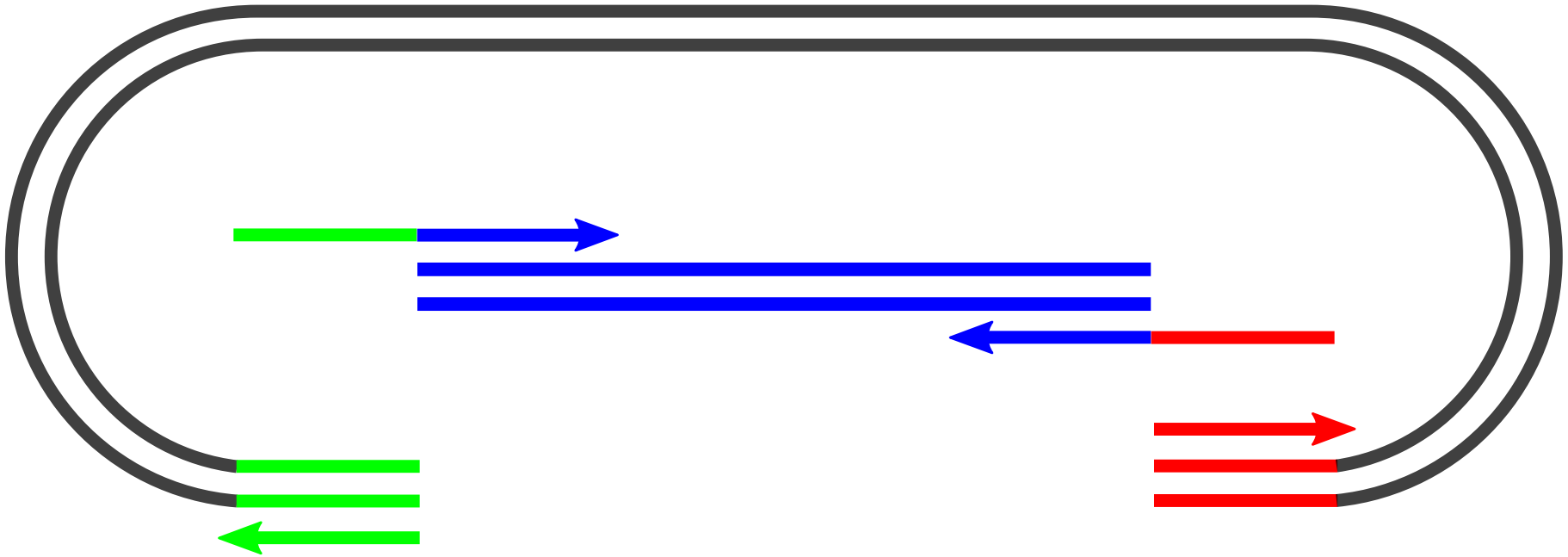
- Gibson Assembly
- In-Fusion
- SLICE (Seamless Ligation Cloning Extract)
- Ligation-Independent Cloning (LIC):
  PIPE (Polymerase Incomplete Primer Extension)
  SLIC (Sequence and Ligation-Independent Cloning)
  OEC (Overlap Extension Cloning)
- T/A cloning

- Recombineering
- Genome Editing:
  ZFN (Zinc Finger Nucleases)
  TALENs (Transcription Activator-Like Effector Nucleases)
  CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeat/CRISPR associated)
  Engineered/Hybrid Meganucleases
  rAAV (recombinant Adeno-Associated Virus)

Gibson Assembly

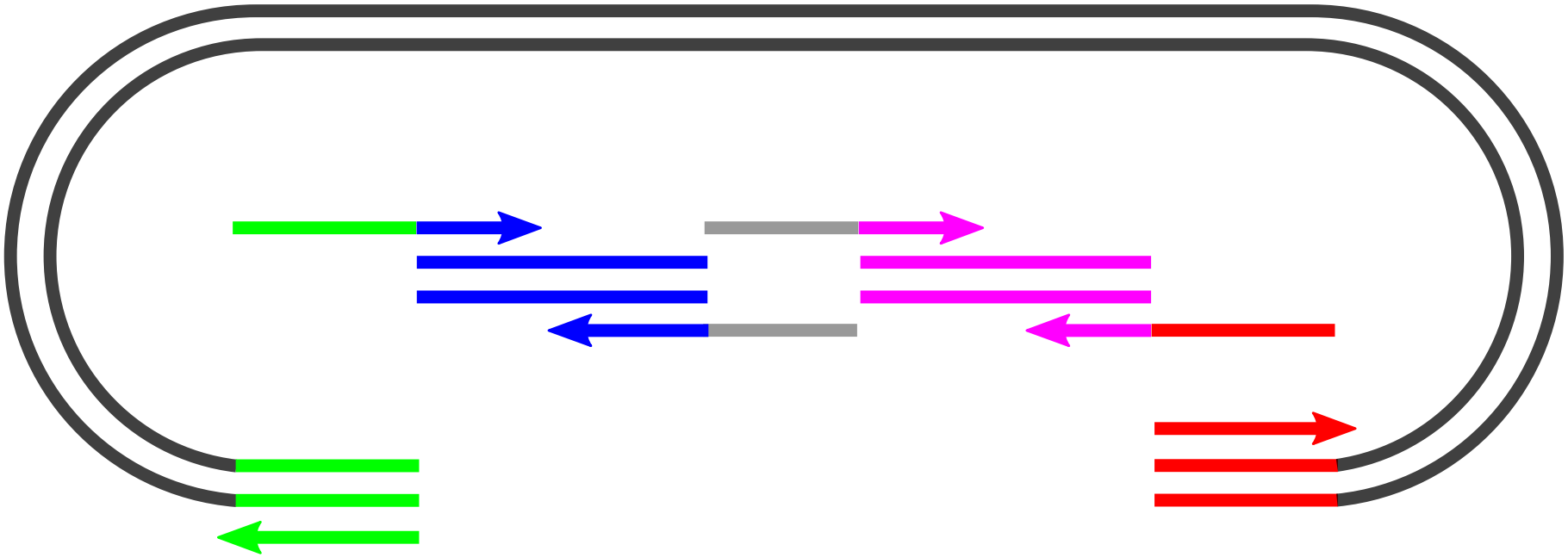


Gibson Assembly



Gibson Assembly

Gibson Assembly

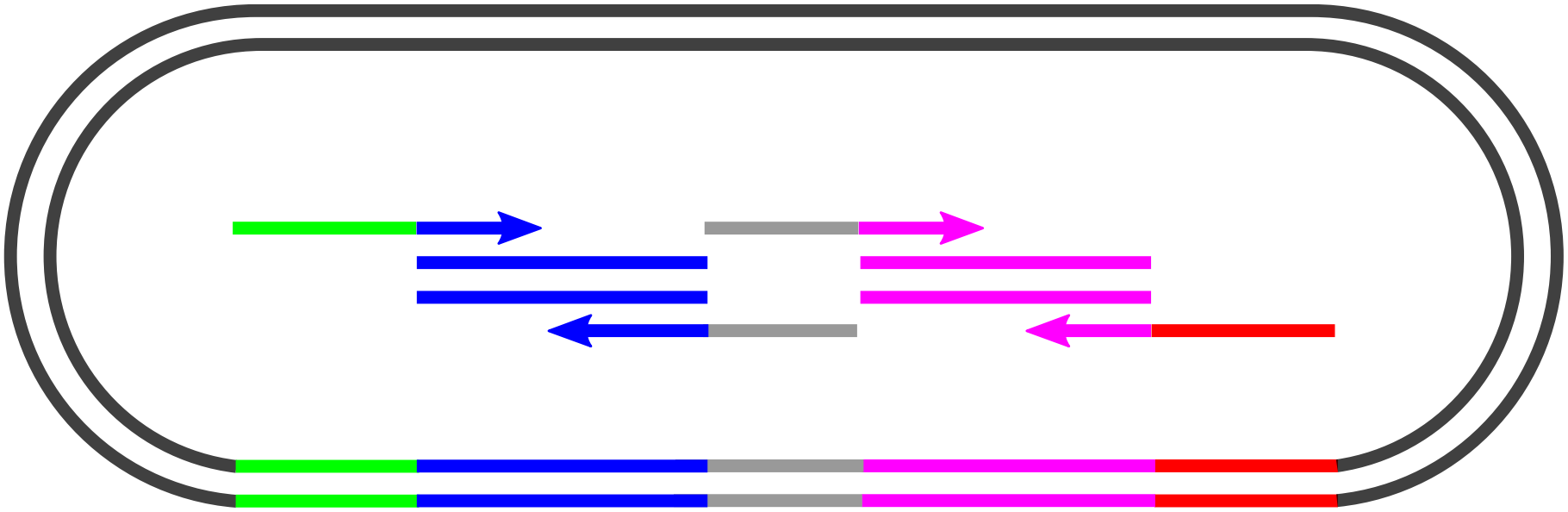


Simultaneous assembly of up to 8-10 fragments

Fragments can be short (= annealed oligos without PCR amplification)

Vector can be either PCR-amplified or linarized by restriction enzyme(s)

Gibson Assembly

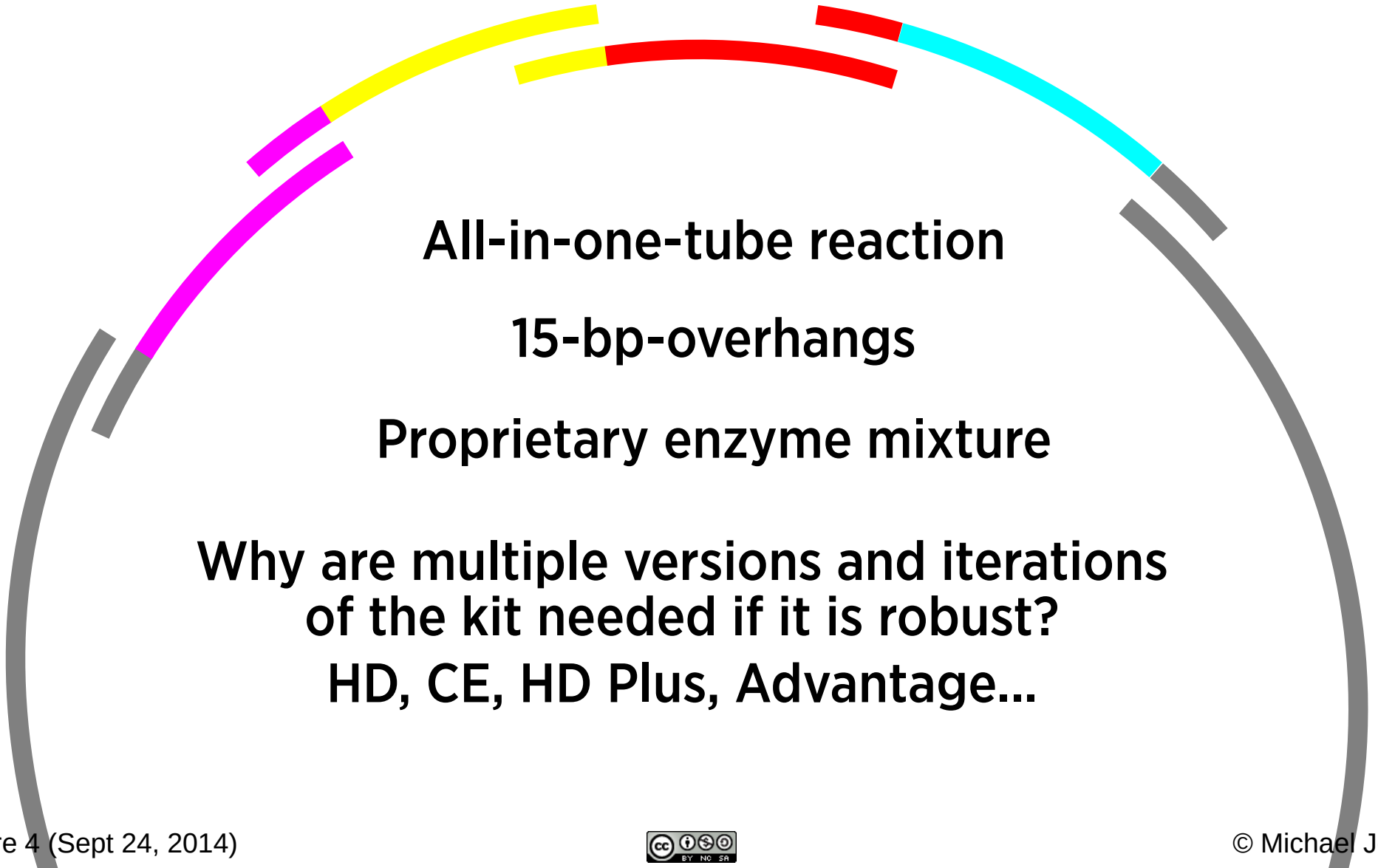


Simultaneous assembly of up to 8-10 fragments

Fragments can be short (= annealed oligos without PCR amplification)

Vector can be either PCR-amplified or linarized by restriction enzyme(s)

In-Fusion (Clontech)



All-in-one-tube reaction

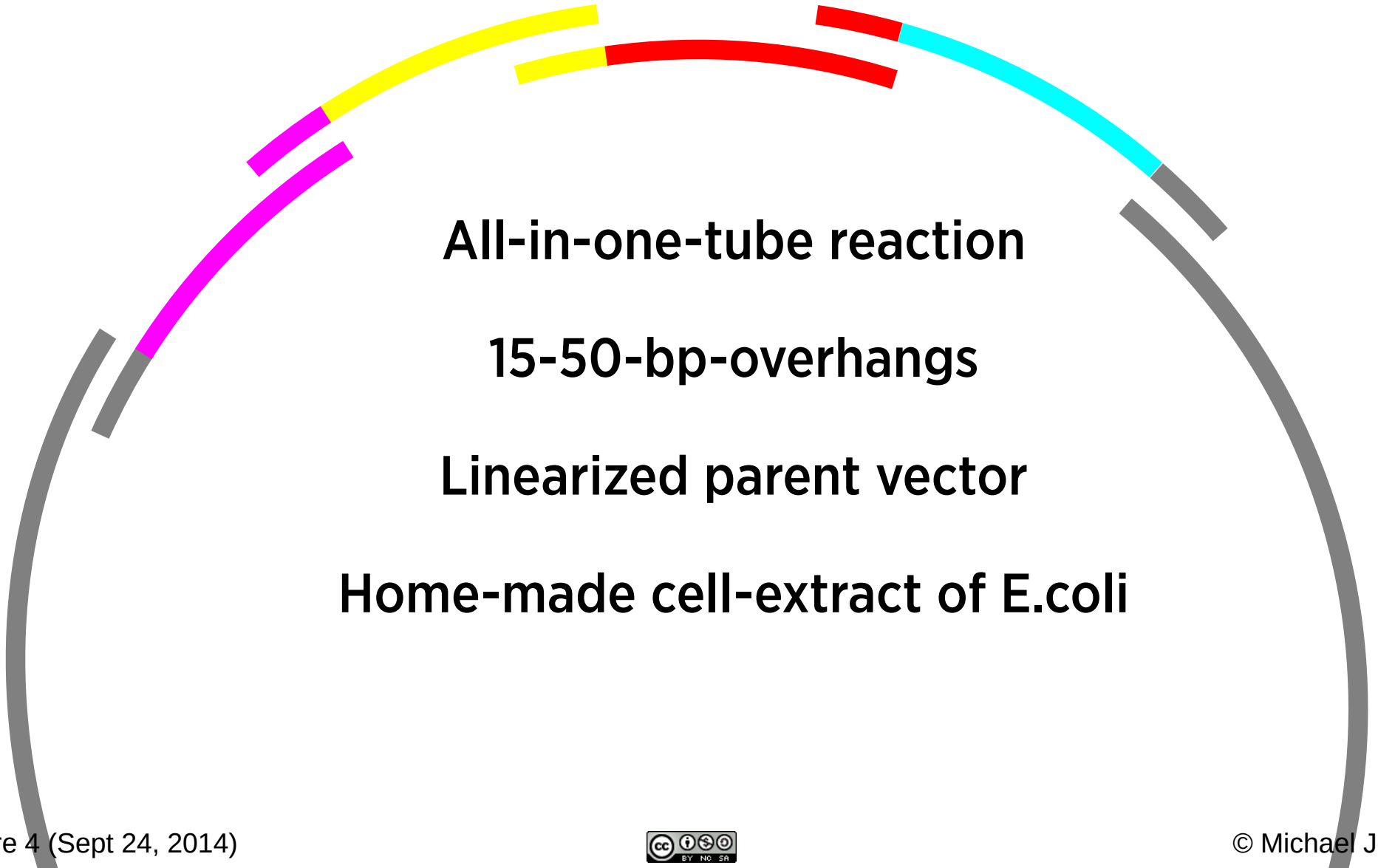
15-bp-overhangs

Proprietary enzyme mixture

Why are multiple versions and iterations of the kit needed if it is robust?

HD, CE, HD Plus, Advantage...

SLICE



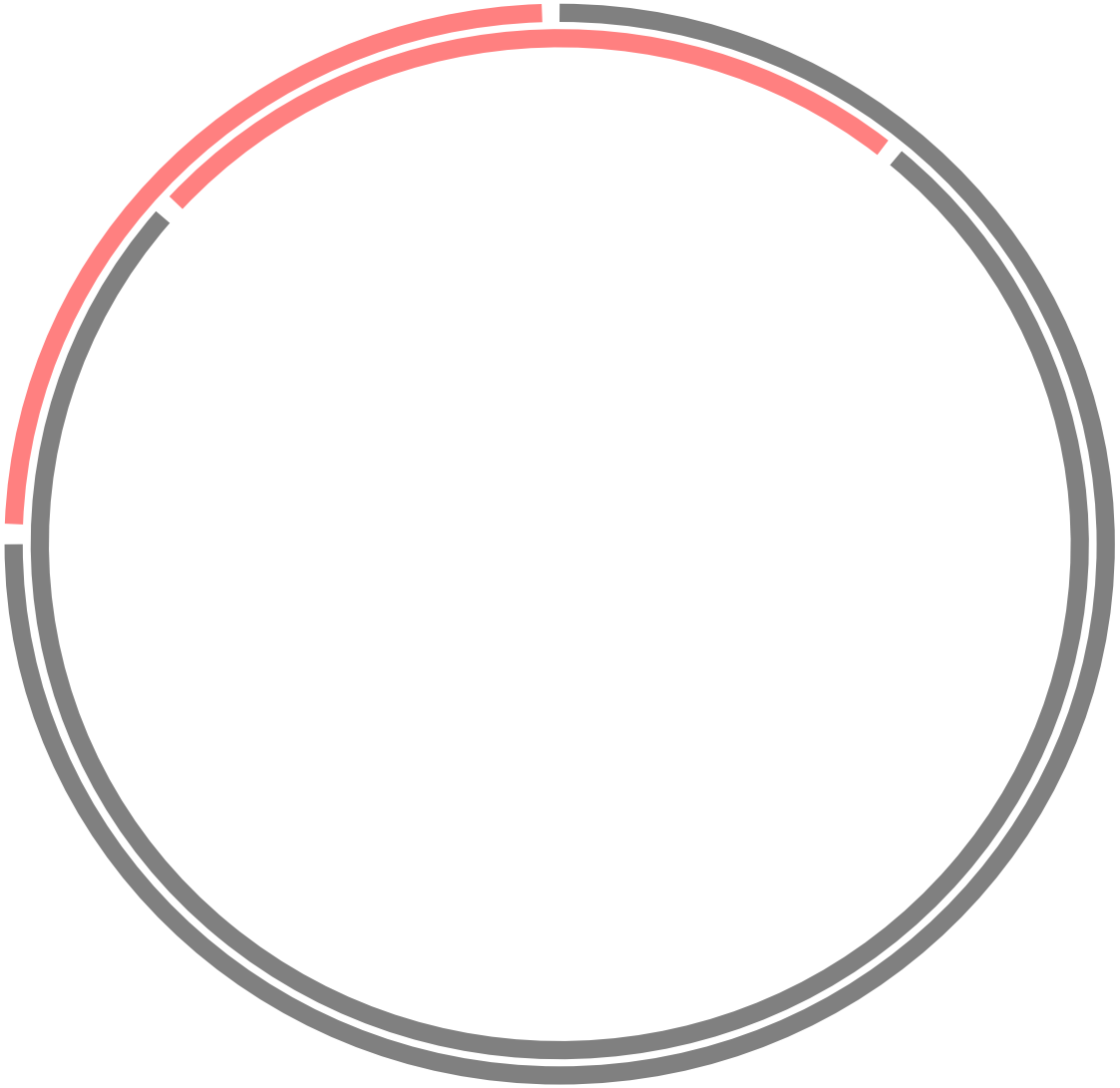
All-in-one-tube reaction

15-50-bp-overhangs

Linearized parent vector

Home-made cell-extract of E.coli

Ligation-Independent Cloning (LIC)



Ligation-Independent Cloning (LIC)

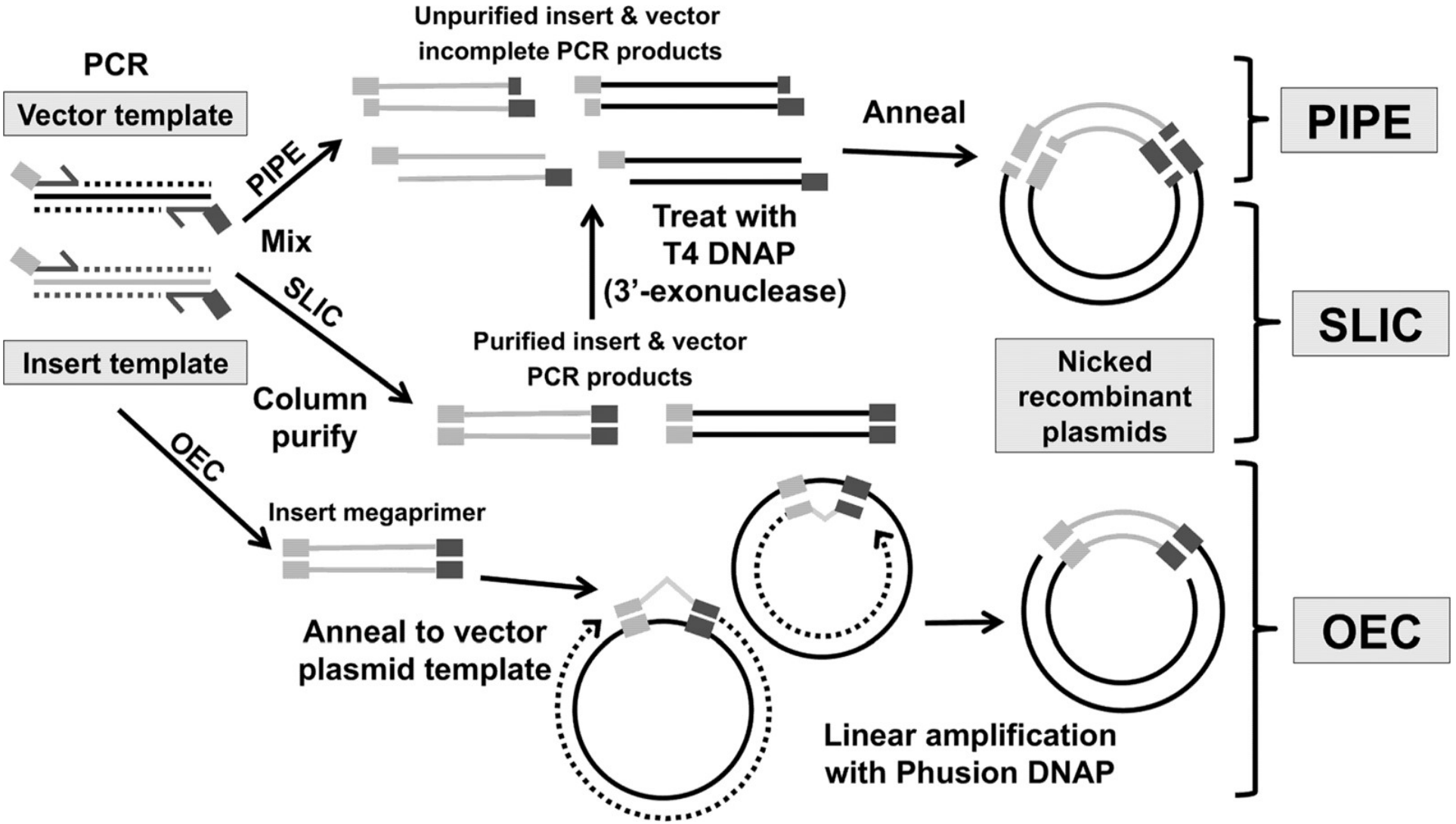
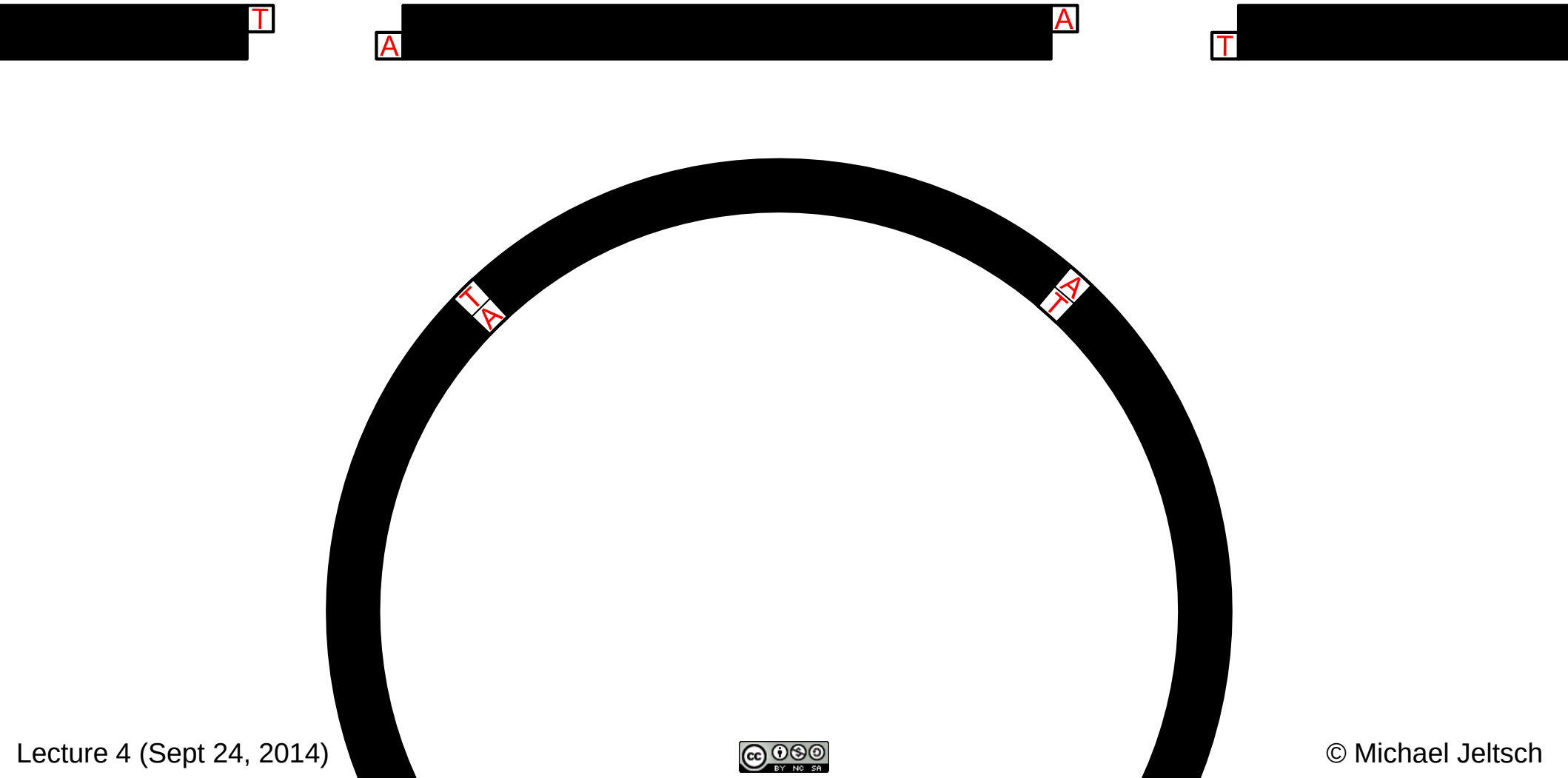


Figure from Stevenson et al.

T/A-Cloning



T/A-Cloning

PCR with a non-proofreading DNA polymerase

```
GGTTAGTTGT      TAATGAGGATACGGAGATACGGATGCGA      AGGATGTCC
GGTTTCAAC      AATTACTCCTATGCCTCTATGCCTACGC      TTCCTACAGG
```

- No directional cloning possible
- Mostly used as a kit (“TOPO-T/A” from Invitrogen)
- Ligation needed (in the TOPO kit, the ligation is performed by the covalently bound *Vaccinia* topoisomerase I)
- If a proofreading polymerase has been used, you can add A-overhangs by a final incubation with Taq or Dynazyme (immediate use or phenol/chloroform extraction) or use “Zero-Blunt TOPO”

Recombineering

- Recombineering = **recombin**ation-mediated genetic engin**eering**
- To modify large DNA constructs (**B**acterial **A**rtificial **C**hromosomes = BACs)
- Homologous recombination in (Escherichia coli) mediated by bacteriophage proteins (Red from λ)
- Linear DNA (ss or ds) is introduced into RecA-negative E.coli with 35-40 bp homologies at the ends (electroporation), Red is induced (by temperature shift)
- efficiency below 1%, selection!
- 2-step process to leave no trace behind: 1. positive selction, 2. negative selection (SacB ⟶ sucrose-sensitivity, GalK ⟶)

Recombineering

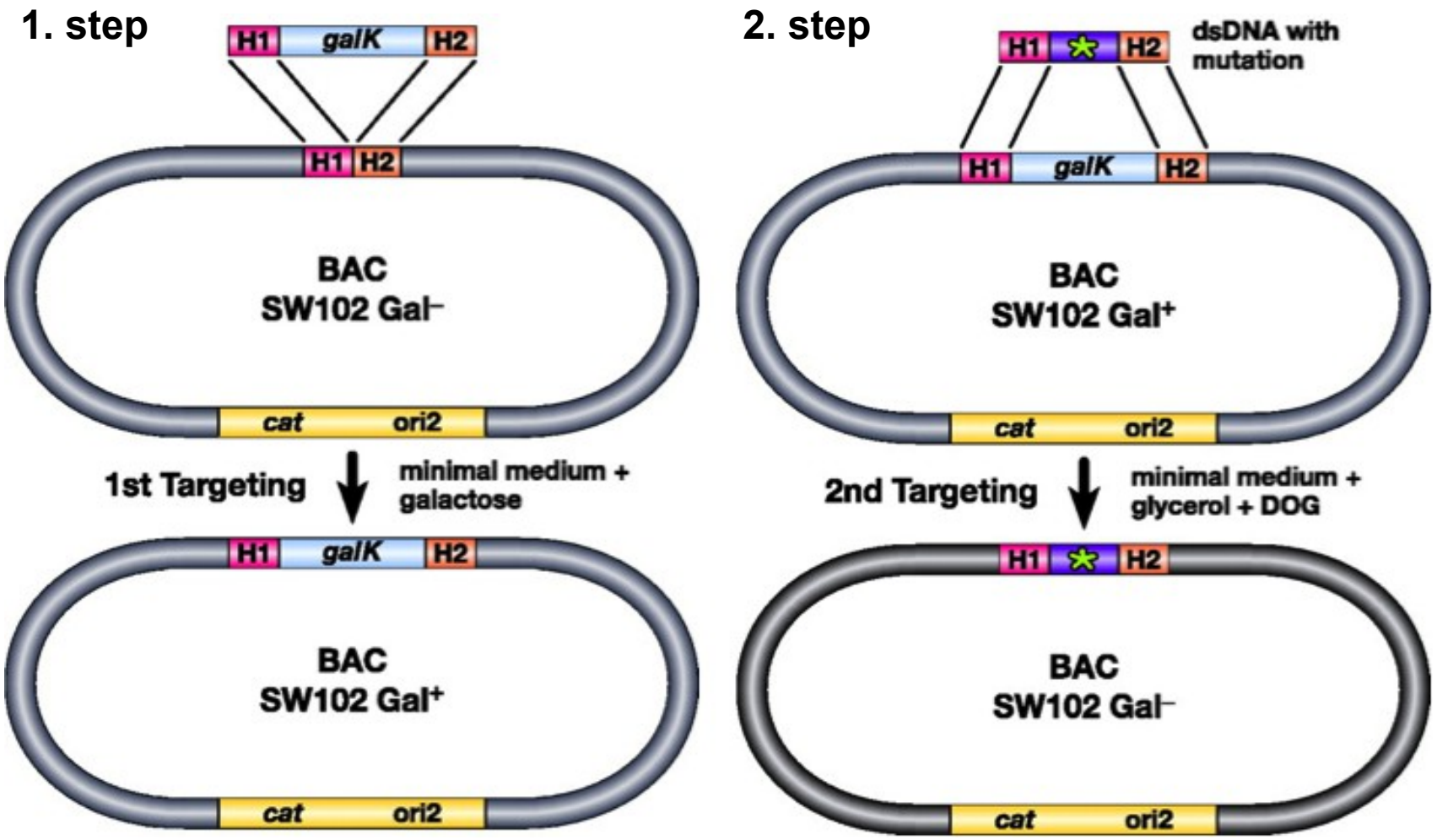


Figure by Warming S et al. Nucl. Acids Res. 2005;33:e36

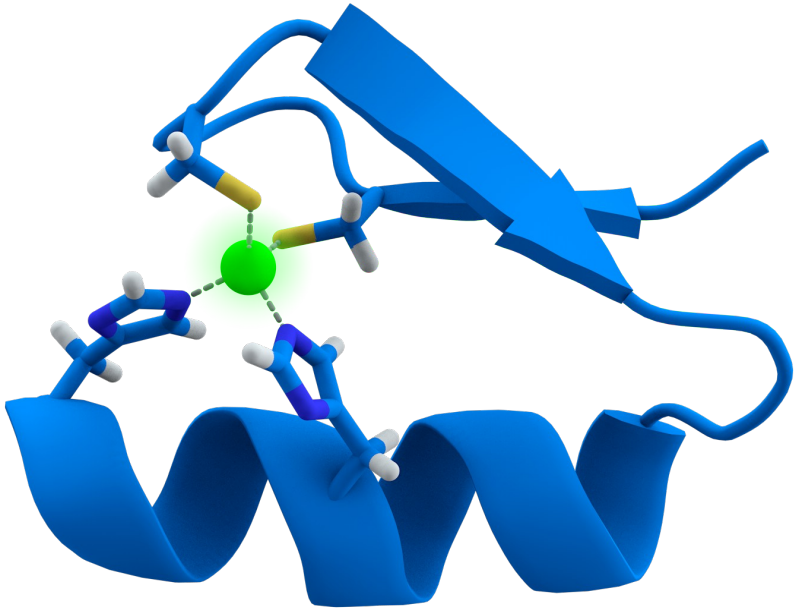
Genome Editing

- ZFN (Zinc Finger Nucleases), 2002/2003
- TALENs (Transcription Activator-Like Effector Nucleases), 2010/2011
- Engineered/Hybrid Meganucleases
- CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeat/CRISPR associated), 2013
- rAAV (recombinant Adeno-Associated Virus)

All (except rAAV) are based on Double Strand Break (DSB) repair by nonhomologous end joining (NHEJ) or homology-directed repair (HDR) pathways

ZFN

- Fusion protein consisting of a sequence-specific DNA-binding protein (Zinc Finger) with a cleavage domain of a restriction enzyme (Nuclease)



ZFN

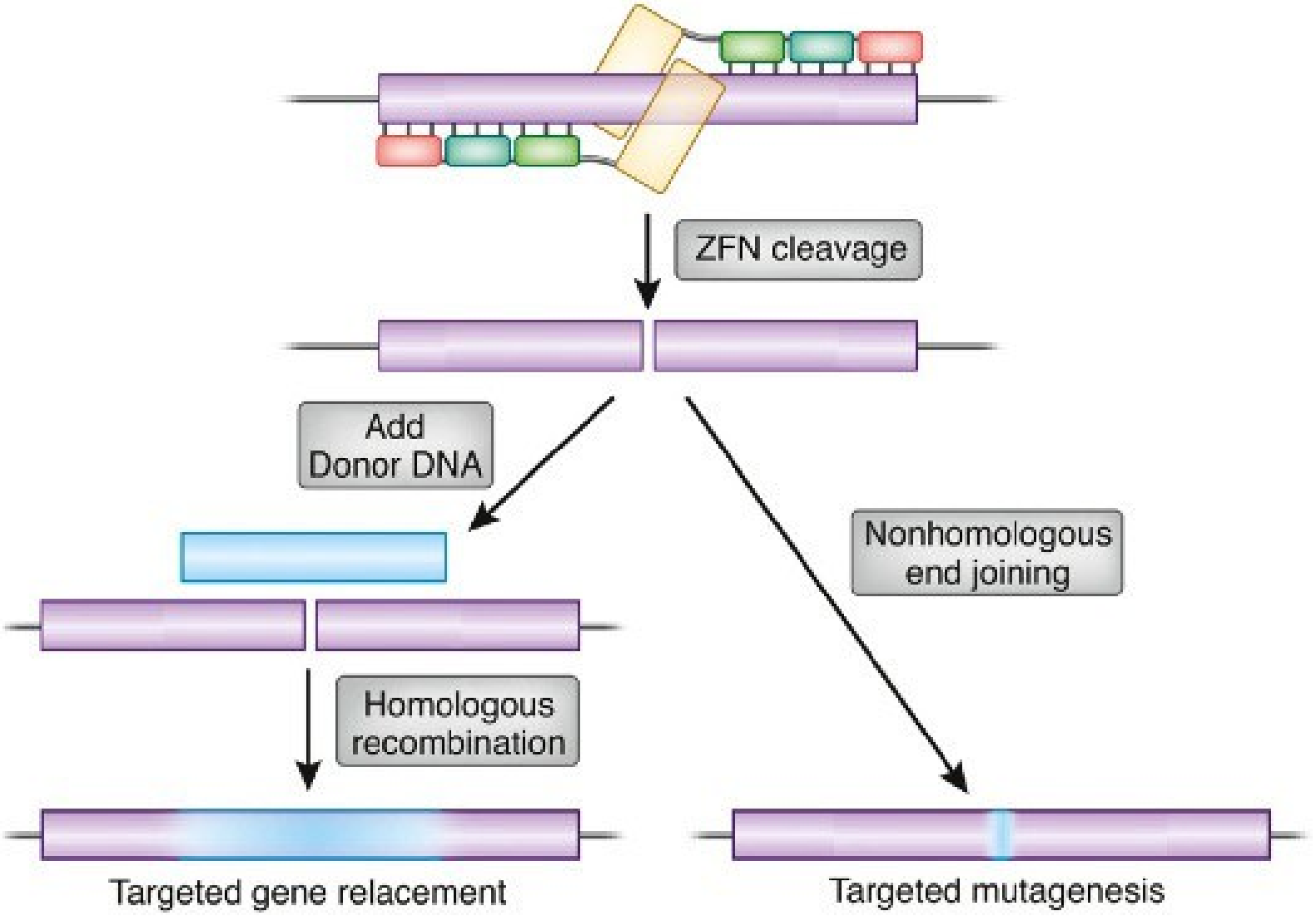
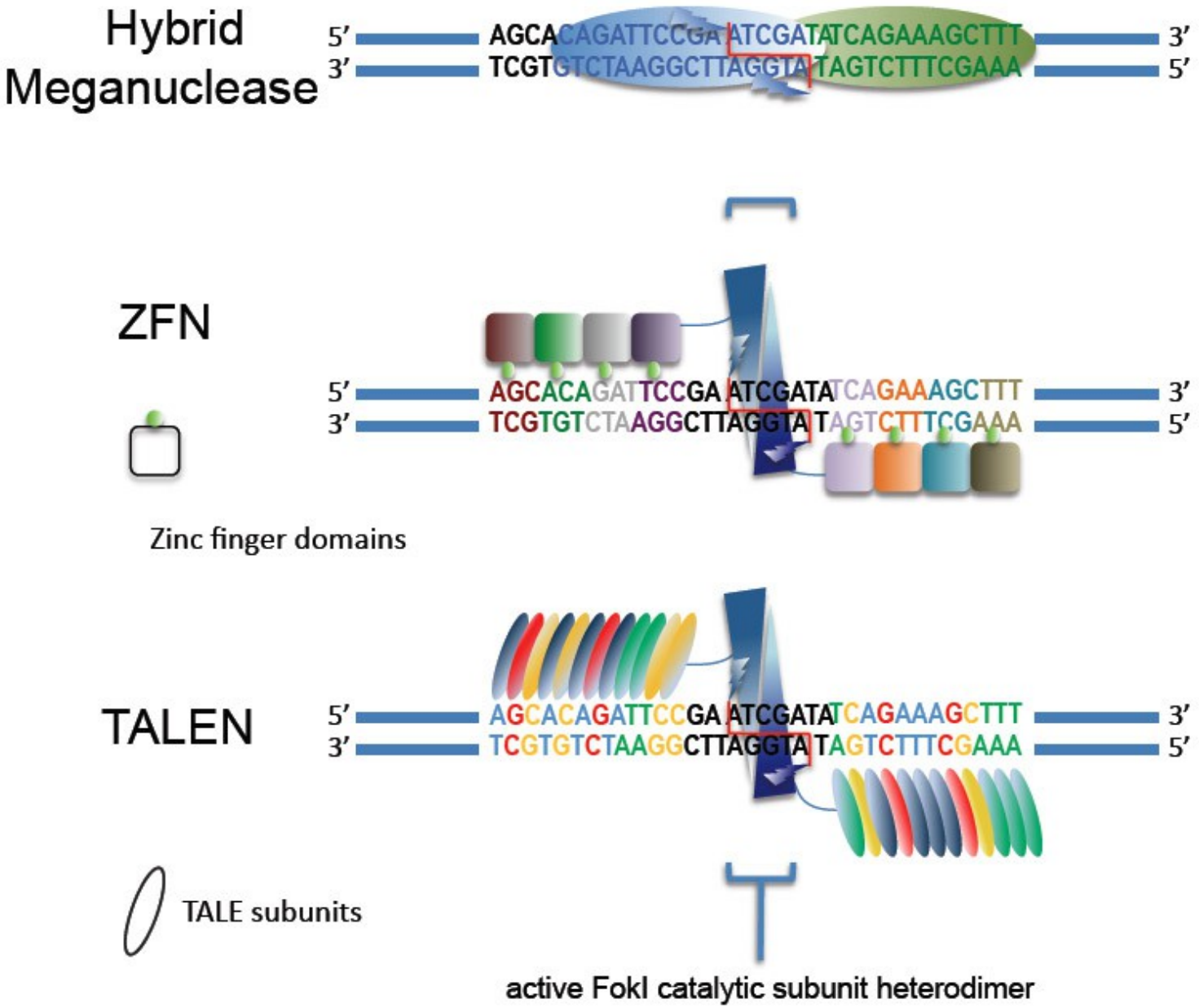
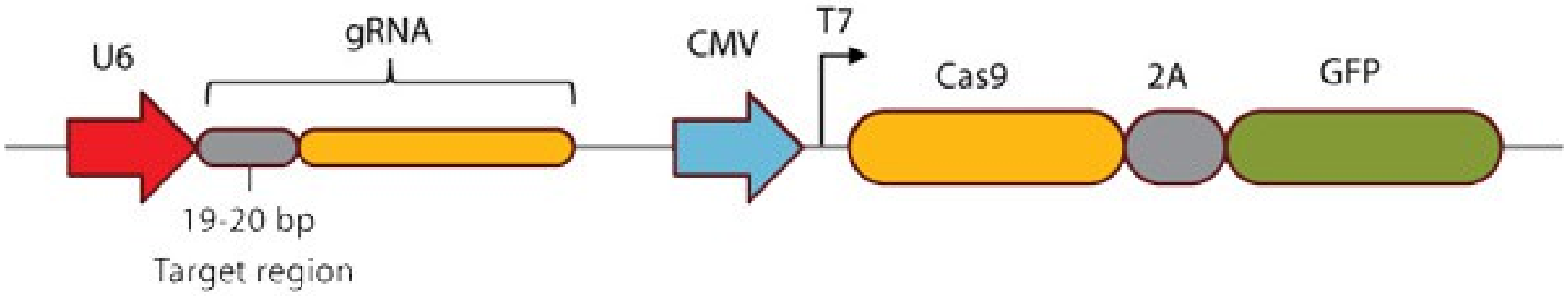


Figure by Caroll D: Genetics 2011 188: 773-782

TALENs & Meganucleases



CRISPR/Cas



CRISPR/Cas Single Plasmid Format

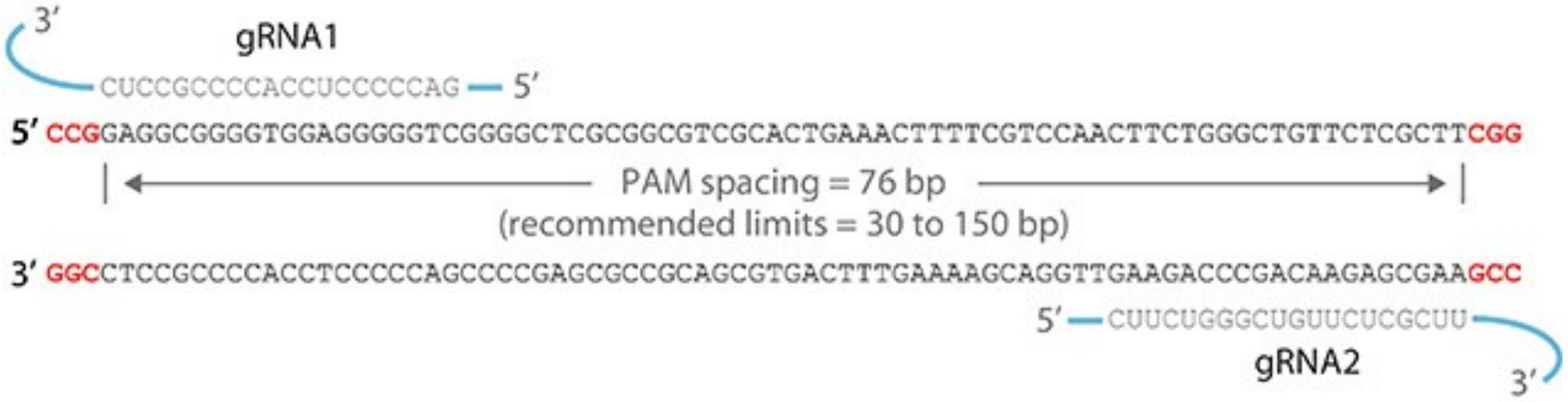
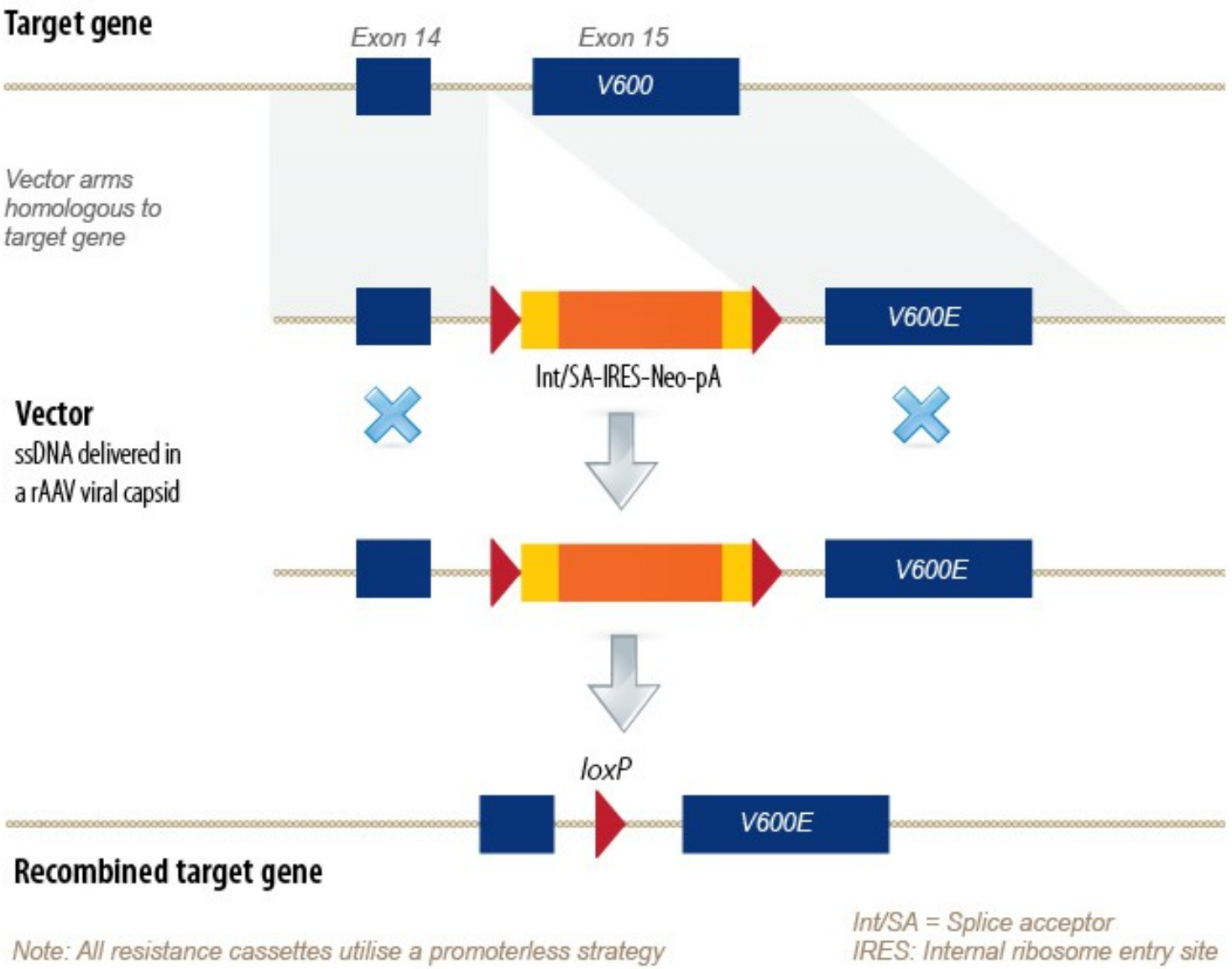


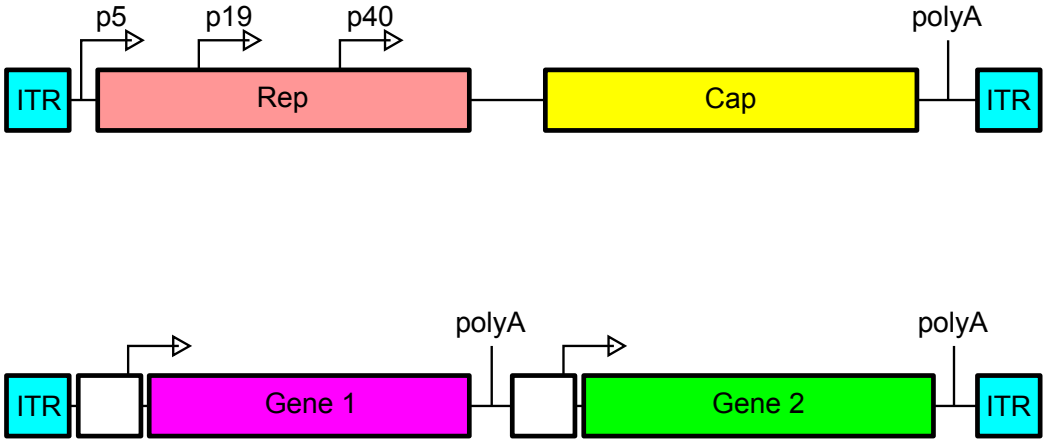
Figure by Sigma

rAAV



rAAV

- linear ssDNA virus
- Non-pathogenic, replication-defective, mild immune response
- Max. cargo of “emptied” genome:
- 4.7kb
- Infects most (quiescent and proliferating) human cells (serotype specific!)
- Integrates into the genome (randomly)
- Induces homologous recombination if suitable homology is present.



Comparison

- ZFN: expensive, difficult to make, Sigma Aldrich/Merck
- TALEN: cheaper, easier to make (can be home-made)
- CRISPR/Cas: cheap, easy, efficient, universal (shRNA and siRNA replacement), multiplexable
- Meganucleases: Very few target sequences, very difficult to create new target sequences, not modular
- rAAV: Mechanism unclear, cargo limit to ~4.7 kb

Comparison

- ZFN: expensive, difficult to make, Sigma Aldrich/Merck
- TALEN: cheaper, easier to make (can be home-made)

1st choice:

CRISPR/Cas: cheap, easy, efficient, universal (shRNA and siRNA replacement), multiplexable

- Meganucleases: Very few target sequences, very difficult to create new target sequences, not modular
- rAAV: Mechanism unclear, cargo limit to ~4.7 kb

Further reading

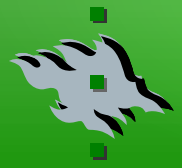
For in-depth information about Gibson Assembly, In-Fusion, SLICE, LIC, T/A cloning, ZFN, TALEN, CRISPR/Cas and rAAV, see the following pages or https://jeltsch.org/practical_molecular_biology

For recombineering, there are several good online resources, e.g.:

http://www.biotec.tu-dresden.de/research/stewart/group-page/recombineering-guide.html

http://redrecombineering.ncifcrf.gov/Home.html

http://ncifrederick.cancer.gov/research/brb/recombineeringinformation.aspx



Practical course

Contact Michael Jeltsch until end of October (michael@jeltsch.org) to agree on a project, create teams/groups and discuss and start necessary preparations such as getting cDNA, sequence verification, ordering oligonucleotides and other materials.

Gibson Assembly™ – Building a Synthetic Biology Toolset

In the quest to create the first bacterial cell controlled by a synthetic genome, the J. Craig Venter Institute (JCVI), with support from Synthetic Genomics, Inc. (SGI), developed a variety of powerful new DNA synthesis and assembly methodologies (1–5) to manipulate large, complex DNAs. These methods include a simple, one-step isothermal *in vitro* recombination technology capable of joining DNAs ranging from relatively short oligonucleotides to fragments hundreds of kilobases in length. This approach, commonly referred to as "Gibson Assembly," is now being used in laboratories around the world to construct DNA fragments. It has the potential to improve upon traditional cloning methods and opens up a range of innovative and ultimately very useful real-world applications.

Daniel G. Gibson, Ph.D., Synthetic Genomics, Inc. and Salvatore Russello, Ph.D., New England Biolabs, Inc.

Introduction:

The use of recombinant DNA technology began soon after the discovery of DNA ligase and restriction endonucleases. Soon after, advent of the polymerase chain reaction (PCR) opened up new possibilities for amplification of specific DNA sequences from a complex mixture of genomic DNA. These technologies have been a mainstay in the modern scientific laboratory for several decades and remain useful methods for cloning potentially valuable or interesting DNA today. However, as scientists seek to work with larger DNA fragments, conduct extensive re-engineering of genetic elements, synthesize whole genomes and move towards automated approaches, the technologies required to manipulate DNA also need to evolve.

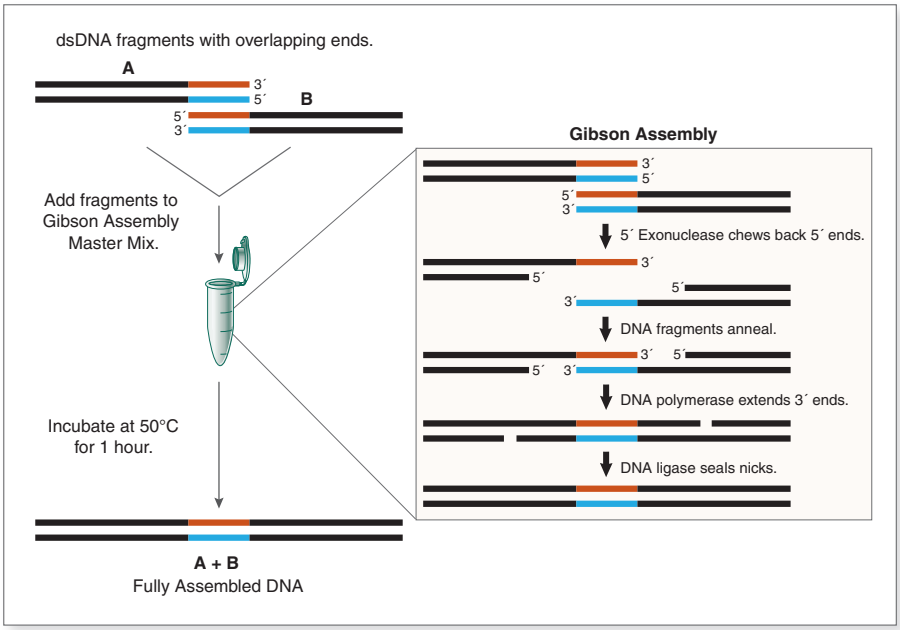
Investigators at the J. Craig Venter Institute (JCVI) have developed a number of *in vitro* enzymatic strategies to assemble short oligonucleotides into larger double-stranded DNA constructs (1-4). In 2003, JCVI made a significant advancement in the production of a synthetic genome by assembling the 5,386 bp genome of phiX174, a virus that infects bacteria, in just 14 days (5). This approach involved joining synthetic oligonucleotides by polymerase cycling assembly, and subsequently amplifying them by PCR (5-6). The unprecedented speed with which this was completed laid the foundation for constructing larger and more complex genomes.

In 2004, JCVI began synthesizing the *Mycoplasma genitalium* genome. It was found that overlapping DNA molecules could be efficiently joined using three enzyme specificities: (i) exonuclease activity, that chews back the ends of DNA fragments and exposes ssDNA overhangs that can anneal to their ssDNA complement; (ii) DNA polymerase activity, that fills gaps in the annealed products, and (iii) DNA ligase activity, that covalently seals the resulting nicks in the assembly. A two-step thermocycle-based *in vitro* recombination method utilizing these enzymes was used to join 101 overlapping DNA cassettes into four parts of the *M. genitalium* genome, each between 136 kb and 166 kb in size. This milestone marked the first assembly of a genome derived from a free-living organism. At 582,970 bp, this synthetic genome was the largest chemically defined DNA structure synthesized in a laboratory, and was 18 times larger than any DNA that had previously been synthesized (4).

Since then, two additional *in vitro* recombination methods have been developed by JCVI to join and clone DNA molecules larger than 300 kb in a single step (2-4). The simplest of these methods is Gibson Assembly, a one-step isothermal approach that utilizes the same three enzymatic activities described previously. This method can be used to join both ssDNA and dsDNAs.

Gibson Assembly has become the most commonly used of the *in vitro* assembly methods discussed above, as it is easy-to-use, flexible and needs little or no optimization, even for large, complex assemblies. All that is required is input DNA with appropriate overlaps, and an appropriate mix of the three enzymes – the Gibson Assembly Master Mix. DNA fragments are added to the master mix and incubated at 50°C for 1 hour; the resulting assembly product is a fully sealed dsDNA suitable for a range of downstream applications (Figure 1). JCVI has used Gibson Assembly to rapidly synthesize the entire 16,520 bp mouse mitochondrial genome from 600 overlapping 60-base oligonucleotides (3). It was also used in combination with yeast assembly to synthesize the 1.1 Mbp *Mycoplasma mycoides* genome, which was then activated in a recipient cell to produce the first synthetic cell (1).

Figure 1. Overview of Gibson Assembly.



Applications of Gibson Assembly:

Cloning. Gibson Assembly eliminates the need to engineer restriction enzyme cut sites within DNA when assembling fragments together. DNA molecules are designed such that neighboring fragments contain a 20-40 bp overlapping sequence. If the DNA fragments originate from PCR products, the overlapping sequence is introduced at the 5′ ends of the primers used in the amplification reaction (Figure 2). DNA fragments can also be assembled with restriction enzyme digested or PCR amplified vector to form circular products suitable for cloning, or for use in downstream applications, such as rolling circle amplification (RCA). To produce these vectors by PCR, each primer needs to include an overlap with one end of the vector, a restriction site (e.g., Not I) not present within the insert or inserts to enable it to be released from the vector, and an overlap with the ends of the DNA fragment assembly or insert. JCVI has been using this approach to combine DNA fragments with vectors, which are then transformed into *E. coli*. One or more fragments have been routinely assembled with general cloning vectors, such as pUC19, and assembled into NEB's pTYB1 expression vector (NEB #N6701). The latter approach was used to express several methylase genes, which aided the genome transplantation efforts at JCVI (8).

Assembly of large DNA constructs. Laboratories worldwide are beginning to explore the use of synthetic biology approaches in the production of pharmaceuticals, industrial compounds, antibiotics, cosmetics and alternative energy sources (7). This often requires the assembly of a genetic pathway consisting of multiple enzymes and their associated regulatory elements. Although template DNA is still required, Gibson Assembly simplifies construction of these types of molecules from component fragments. A long stretch of desirable DNA sequence (e.g., a 40 kb genetic pathway) can be broken down into several overlapping PCR products (e.g., eight, 5 kb pieces), which can then be amplified by conventional PCR and combined using Gibson Assembly. This approach has been used to move genetic pathways from one organism to another and to rapidly swap genes, promoters, terminators and ribosome binding sites. DNAs up to ~1 Mbp have been assembled *in vitro* using Gibson Assembly (2).

Assembly of chemically-synthesized oligonucleotides into dsDNA fragments. Gibson Assembly can also be used to directly assemble oligonucleotides into a cloning vector, such as pUC19 (3). A common problem observed when chemically synthesizing long stretches of oligonucleotides is the introduction of errors (9). To ensure that error-free molecules are obtained at a reasonable efficiency, a strategy employed by SGI and JCVI involves the assembly of only eight to twelve 60-base oligonucleotides (with 30 bp overlaps) at one time. The resulting dsDNA molecules are sequence-verified and assembled into larger DNA fragments using the same approach. Because assembly itself does not generally introduce new errors, the final assembled product can be retrieved at high efficiencies. Using this approach, many of the costly and time consuming steps currently used to synthesize DNA, including PCR and an error correction, are eliminated.

Site-directed mutagenesis. Gibson Assembly can also be used to make rapid changes to DNA fragments, including substitutions, deletions and insertions. To use Gibson Assembly for mutagenesis, the desired changes are introduced into the PCR primers, within the overlapping sequences at assembly points (Figure 3). To modify a DNA sequence in this way, two PCR primers are required: the first contains the desired nucleotide changes, and the second contains the reverse complement of the first primer at the overlapping region. Following amplification and assembly of the fragments, the designed changes are incorporated into the final product. The number of changes that can be made at once depends on the number of fragments simultaneously assembled. For example, an eight-piece assembly, which contains eight assembly points, provides eight opportunities to introduce changes in the DNA sequence. Because the method can be used to assemble large DNA fragments, mutations can rapidly be made to very large pieces of DNA. For example, eight modifications can be introduced into an 80 kb DNA molecule following the assembly of eight 10 kb PCR fragments. This site-directed mutagenesis strategy was used during synthesis of the *M. mycoides* genome. The cassettes comprising the synthetic genome were ordered based on an imperfect draft sequence, which resulted in small differences between the synthetic cassettes and the desired *M. mycoides* genome sequence. The sequences of 16 cassettes were successfully edited using this approach (1).

Combinatorial synthesis of DNA Fragments. In the near future, chromosomes will be designed and synthesized for processes ranging from biofuel production to pharmaceutical manufacture. Bacteria and plants often carry out syntheses that far exceed what can be readily achieved by the best organic chemists. The genes that control desirable pathways can be chemically synthesized, placed in artificial chromosomes, and "installed" in suitable host cells, including bacteria, yeast or plant cells. These multigene pathways can be constructed in a combinatorial fashion, such that each member of the library has a different combination of gene variants. Using screening and selection methods, cells bearing the pathway with the desirable trait (highest yield of a compound, for instance) can be obtained. The engineered host organism then becomes a biologic factory used to manufacture the product specified by the synthetic pathway. Gibson Assembly has the potential to be used to produce combinatorial libraries of synthetic or semisynthetic chromosomes carrying thousands of genes. Figure 4 demonstrates the combinatorial assembly of cassettes produced from 60-mer oligonucleotides. Here, 1,024 (2^{10}) variants of a 1 kb gene, containing 10 single nucleotide changes, are produced from 30 sequence-verified cassettes.

Moving Forward

Synthetic & Minimal Cells. For the past 17 years, the genomes of many organisms have been sequenced and deposited in databases. It has recently been shown that it is possible to reverse this process and synthesize bacterial cells from digitized information (1). In order to realize this vision, researchers at JCVI needed tools and technologies to sequence, synthesize and transplant genomes. Although many hurdles needed to be overcome, synthetic cells can now be produced in the laboratory. As proof of concept, the 1.08 Mbp *M. mycoides* JCVI-syn1.0 genome was designed, synthesized

Figure 2. PCR-Generated Vector and Insert Assembly.

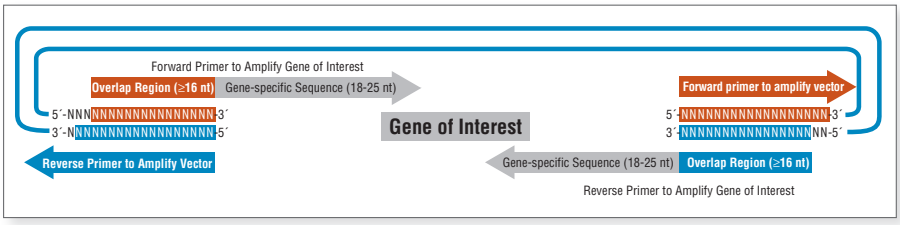
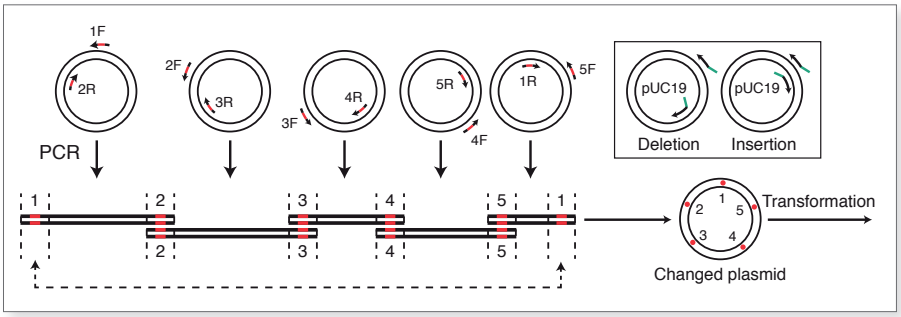


Figure 3. Introducing changes to a desired DNA using Gibson Assembly.



In the example above, five changes were introduced into a DNA plasmid using Gibson Assembly. Each primer contains a change introduced by PCR. The overlapping PCR fragments with each change at the assembly point are then combined using Gibson Assembly. (credit: Mikkel Algire, JCVI).

and assembled, starting from the digitized genome sequence, and transplanted into a *Mycoplasma capricolum* recipient cell to create new *M. mycoides* cells controlled only by the synthetic chromosome. The only DNA present in the cells is the designed synthetic DNA, including "watermark" sequences, and other designed gene deletions and polymorphisms and mutations acquired during the building process. The new cells have the expected phenotypic properties and are capable of continuous self-replication (1). The *M. mycoides* genome is currently the largest chemically defined DNA structure that has been synthesized in a laboratory. It is almost twice as large as the synthetic *M. genitalium* genome reported in 2008, and more than an order of magnitude larger than any reported DNA sequence synthesized outside JCVI. What has been learned in this "proof of concept" experiment can now be applied to designing and producing new organisms with useful properties.

Further, researchers at JCVI have already begun working on their ultimate objective: to synthesize a minimal cell with only the machinery necessary for independent life. Now that a living cell can be produced from a synthetic genome, components of a synthetic genome can be removed and transplanted in an iterative fashion until only the essential genes are present and the genome is as small as possible. This will help to better understand the function of every gene in a cell and what DNA is required to sustain life in its simplest form. Gibson Assembly is one of the core technologies that will be used to achieve these goals.

Conclusion

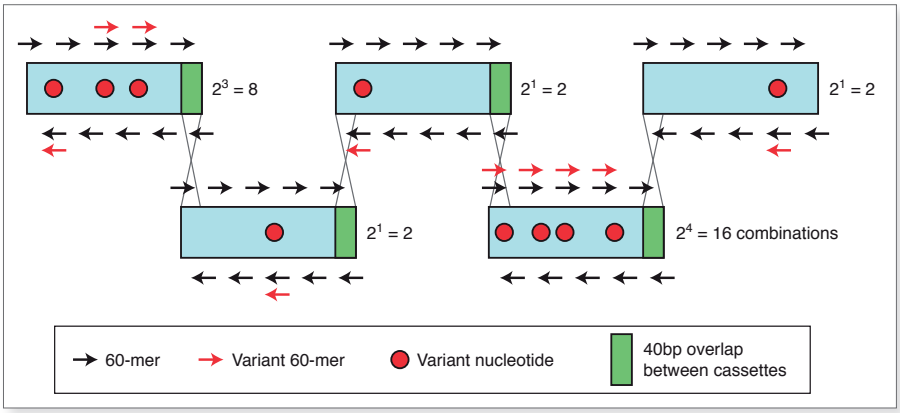
Gibson Assembly is a simple and robust method that enables the simultaneous production of many different combinations of genes and pathways, accelerating the progress of synthetic biology. Furthermore, this powerful technology has the potential to help turn DNA sequence into genes and pathways useful in the production of biofuels, industrial compounds, pharmaceuticals and vaccines. The synthesis of genes and pathways, and even small genomes, has been made easier with Gibson Assembly, helping to move the field of synthetic biology forward. As the power of DNA sequencing increases and sequencing costs decrease, DNA databases will continue to fill with novel genes and pathways waiting to be identified, optimized and expressed in a heterologous host organism. It is time to better understand how to turn these DNA sequences into useful applications.

The ability to quickly construct whole genes and genomes has the potential to accelerate research in a variety of other fields. This capability may also make it possible to quickly respond to emerging threats, and may allow researchers to understand how "life" works. The power of large scale DNA synthesis will dramatically impact the way research is done and vastly accelerate the pace of science. The Gibson Assembly Master Mix provides a new and powerful tool for biotechnology, whose most far-reaching benefits may not yet even be envisioned.

References

1. Gibson, D.G. et al. (2010) *Science*, 239, 52–56.
2. Gibson, D.G. et al. (2009) *Nature Methods*, 6, 343–345.
3. Gibson, D.G. et al. (2010) *Nature Methods*, 7, 901–903.
4. Gibson, D.G. et al. (2008) *Science*, 319, 1215–1220.
5. Smith et al. (2003) *PNAS*, 100, 15440–15445.
6. Stemmer, W.P. et al. (1995) *Gene*, 164, 49–53.
7. Endy, D. (2005) *Nature*, 438, 449–453.
8. Lartique, C. et al. (2009) *Science*, 325, 1693–1696.
9. Carr et al. (2004) *Nucleic Acids Res.* 32, e162.

Figure 4. Combinatorial gene synthesis.



Ten overlapping 60-mer oligonucleotides are recombined into five cassettes (light blue rectangles). Following sequence verification of these cassettes, they are recombined (indicated by the X) into the full-length gene in a second stage. Recombination of the cassettes is possible because 40 bp overlaps have been added to the cassettes. Within each cassette, there are between 1 and 4 positions in which 2 nucleotides are acceptable. At these locations, a variant 60-mer oligonucleotide can also be assembled. This equates to 2-16 different combinations for each cassette. In order to produce every different combination ($2^3 \times 2^1 \times 2^1 \times 2^4 \times 2^1 = 1{,}024$) of the full-length gene, every different cassette can be pooled and assembled into a cloning vector to form a circular DNA, cloned into E. coli, and then sequenced.

Gibson Assembly™ Master Mix

Gibson Assembly was developed by Dr. Daniel Gibson and his colleagues at the J. Craig Venter Institute and licensed to NEB by Synthetic Genomics, Inc. It allows for the successful assembly of multiple DNA fragments, regardless of fragment length or end compatibility. It has been rapidly adopted by the synthetic biology community due to its ease-of-use, flexibility and suitability for large DNA constructs.

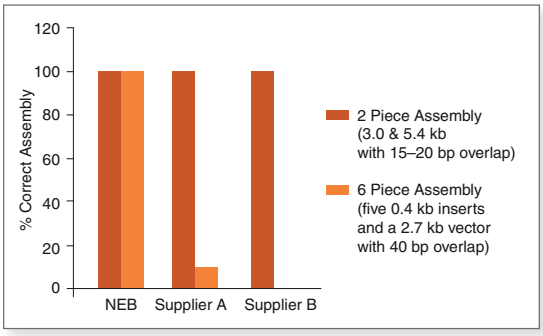
Gibson Assembly efficiently joins multiple overlapping DNA fragments in a single-tube isothermal reaction (1,2). The Gibson Assembly Master Mix includes three different enzymatic activities that perform in a single buffer (see Figure 1, page 3):

- The exonuclease creates a single-stranded 3´ overhang that facilitates the annealing of fragments that share complementarity at one end.
- The polymerase fills in gaps within each annealed fragment.
- The DNA ligase seals nicks in the assembled DNA.

The end result is a double-stranded, fully-sealed DNA molecule that can serve as template for PCR, RCA or a variety of other molecular biology applications, including direct transformation. The method has been successfully used by Gibson's group and others to assemble oligonucleotides, DNA with varied overlaps (15–80 bp) and fragments hundreds of kilobases long (1–3).

In contrast to other methods, Gibson Assembly is suitable for a wide range of different types of assemblies, including large numbers of DNA fragments. The figure below shows that Gibson Assembly outperforms other methods for assembly of six fragments.

Assembly of DNA fragments using methods from three different suppliers.



The assemblies were conducted in accordance with the manufacturer's recommended protocols. Following assembly and transformation, a portion of the resulting colonies were tested for the presence of a PCR fragment corresponding to a correct, fully assembled construct, by colony PCR, using primers that amplify the entire fragment. While all three vendors' methods perform a 2 fragment assembly well, the Gibson Assembly Master Mix outperforms other suppliers for larger numbers of fragments.

Ordering Information

PRODUCT	NEB #	SIZE	PRICE
Gibson Assembly Master Mix	E2611S/L	10/50 reactions	150 € /600 €

Gibson Assembly™ is a trademark of Synthetic Genomics, Inc.

Advantages

- Increased number of successful assembly products, particularly for longer or greater number of fragments
- Flexible sequence design with no need to engineer cloning sites (scarless cloning)
- Does not require clean up of PCR products prior to assembly
- Complex assembly achieved in 1 hour
- DNA can be used immediately for transformation, or as template for PCR or RCA
- Easily adapted for multiple DNA manipulations, including site-directed mutagenesis, insertions and deletions

1. Gibson, D.G. et al. (2009) *Nature Methods*, 343–345.
2. Gibson, D.G. et al. (2010) *Nature Methods*, 901–903.
3. Gibson, D.G. Personal communication.

Manufactured by New England Biolabs, Inc. under license from Synthetic Genomics, Inc.

SYNTHETIC GENOMICS®

Fun in the Lab: See what Gibson Assembly fans have posted on www.youtube.com!

Source: www.youtube.com, uploaded by the Cambridge iGEM Team

NEW ENGLAND BioLabs® GmbH
enabling technologies in the life sciences

New England Biolabs GmbH
Brüningstr. 50; Geb B852
D-65926 Frankfurt am Main

Telefon: +49 (0) 69/305-23 14-0
Telefax: +49 (0) 69/305-23 14-9
E-Mail: info@de.neb.com

www.neb-online.de

In-Fusion™ assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations

Baogong Zhu, Guifang Cai, Emily O. Hall, and Gordon J. Freeman
Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA

BioTechniques 43:354-359 (September 2007)
doi 10.2144/000112536

In-Fusion™ can join any two pieces of DNA that have a 15-bp overlap at their ends. The result is equivalent to a recombination event at the ends of the DNAs. The 15-bp overlap may be engineered by inclusion in primers used to PCR amplify a segment of DNA. Originally described for inserting one piece of DNA into a restriction enzyme-digested plasmid, we have found In-Fusion can join four or more pieces of DNA in a single reaction. We used this insight to construct seamless fusion proteins, modular vectors with readily interchangeable segments, and novel mutagenesis strategies. Replacement In-Fusion can be used to delete any desired DNA segment in a plasmid and replace it with any desired new DNA segment without limitations on position or size.

INTRODUCTION

DNA constructs are typically joined by ligation at restriction enzyme sites and construct options are limited by the available unique sites in the vector and gene. In contrast, an In-Fusion™ enzyme reaction can join any two pieces of DNA that have 15 bp of identity at their ends. The 15-bp overlap may be engineered by inclusion in primers used to PCR amplify a segment of DNA. The pieces of DNA may be generated by PCR and have blunt ends or by restriction digest of plasmid DNA and have sticky or blunt ends depending on the enzyme used. The In-Fusion mechanism is ligation-independent and while proprietary, likely uses the unique properties of the 3′–5′ exonuclease activity of poxvirus DNA polymerase (1–3). When incubated with linear duplex DNAs with homologous ends in the presence of Mg^{2+} and low concentrations of dNTP, the 3′–5′ proofreading activity of poxvirus DNA polymerase progressively removes nucleotides from the 3′ end. This exposes complementary regions on substrate DNAs that can then spontaneously anneal through base pairing, resulting in joined molecules containing a hybrid region flanked by nicks, 1–5 nucleotide gaps, or short overhangs (Figure 1A). The annealed structures are metastable because the poxvirus DNA polymerase has a lower affinity for nicked or gapped DNA ends than for duplex ends. Introduction into *Escherichia coli* repairs any single-stranded gaps. Thus, one copy of the overlap is present in the final DNA product, and the result is equivalent to a recombination event at the ends of the DNAs. Originally described for inserting one piece of DNA into a restriction enzyme-digested plasmid, we have found that In-Fusion can join four or more pieces of DNA. We used this insight to develop seamless fusion proteins, modular vectors, and novel mutagenesis strategies.

MATERIALS AND METHODS

Design of a construct. The desired pieces of a DNA construct are assembled in a DNA manipulation program such as Sequencher™ (Gene Codes, Ann Arbor, MI, USA). For example, as shown in Figure 1B, DNA segments encoding the interleukin-2 (IL-2) signal sequence (4,5), the extracellular domain of CD101 minus its endogenous signal sequence (6), and the fragment crystallizable (Fc) domain of murine immunoglobulin G3 (IgG3) (7) are assembled with a mammalian expression vector.

Design of overlap primers. Sense and antisense PCR primers are designed, which contain a 15-bp overlap with the adjacent segment of the construct and 20–30 bp of segment-specific sequence. The junction between two pieces of DNA can be made seamless by including no additional DNA sequence. Alternatively, short pieces of DNA such as restriction sites, translation initiation sites, linkers, or epitope tags can be added by inclusion in the primer sequences. Table 1 and Table 3 give the primers for the constructs in Figure 1B and Figure 3, A and B. Vector segments can be generated by restriction enzyme digest of a plasmid or by PCR. Where a primer is designed to overlap a restriction-digested DNA fragment, the 15-bp overlap is counted from the cleavage site on the antisense DNA strand as described in the Clontech In-Fusion user manual. Most primers are 35–55 bp, and we find quality control by mass spectroscopy to provide sufficient purity (Midland Certified, Midland, TX, USA).

Generation of DNA segments. The DNA segments were PCR amplified from appropriate templates with overlap primers (designed as described in the section entitled Design of overlap primers) and *PfuUltra*® II Fusion Hot Start polymerase (Stratagene, La Jolla, CA, USA), gel-purified, and quantitated. The use of a high-fidelity polymerase reduces errors, but *Taq* polymerase PCR products will also work.

In-Fusion reaction. Twenty-five to one hundred nanograms of restriction enzyme-digested, gel-purified vector were mixed at a molar ratio of 1 vector to 2 of each DNA segment in a total of 10 µL water in one tube of In-Fusion Dry-Down reaction mix (Clontech, Mountain View, CA, USA). The reaction was incubated at 42°C for 30 min, transferred to ice, and 40 µL Tris EDTA (TE) were added. Four microliters were transformed into One Shot® TOP10 competent *E. coli* (1×10^9 cfu/µg; Invitrogen, Carlsbad, CA, USA), miniprepped, and characterized by restriction enzyme digest and sequencing.

RESULTS AND DISCUSSION

DNA constructs often include undesired amino acids encoded by restriction sites engineered to provide a joining point, or seam at which two DNAs can be ligated. This is particularly undesirable for fusion proteins or recombinant antibodies, since the undesired amino acids may perturb structure, reduce expression, or be antigenic. We have found that In-Fusion will join multiple pieces of DNA together, thereby facilitating the design of seamless fusion protein constructs that contain only the desired protein sequence and are not constrained by the presence or absence of restriction enzyme sites. We designed a seamless CD101-IgG3 fusion protein by joining in silico four pieces of DNA encoding a secretory signal, CD101, and IgG3 Fc with an expression vector. No restriction sites are incorporated between the fusion protein domains, and *Nco*I and *Sal*I sites are present at the insertion site into the vector. PCR primers were designed to amplify each segment and contain a 15-bp overlap with the adjacent segment (Table 1). As shown in Figure 1B, DNAs encoding (*i*) three different secretory signal sequences [IL-2, erythropoietin (EPO), IgG1) to optimize protein expression; (*ii*) the extracellular domain of CD101 minus its endogenous signal sequence; and (*iii*) the Fc domain of murine IgG3 were generated by PCR using primers that contained a 15-bp overlap with the adjacent segment and 20–30 bp of segment-specific sequence.

The three DNA segments and gel-purified expression vector were joined in an In-Fusion reaction by mixing at a molar ratio of 1 vector to 2 of each signal-CD101-Fc DNA segment in one tube of In-Fusion Dry-Down reaction mix (Table 2). The reaction was transformed into TOP10 *E. coli*; 10 colonies per construct were miniprepped and characterized by *Nco*I + *Sal*I restriction enzyme digest and sequencing (Figure 2). All 10 minipreps of each construct contained all four pieces of DNA. Of note, the insert contains two internal *Nco*I sites, but since only the vector and not the inserts need to be digested with *Nco*I, internal restriction sites in PCR products do not limit the design of the construct. Sequencing of six minipreps was sufficient to identify error-free isolates of the 3.6-kb coding region containing IL-2 signal-CD101-Fc, EPO signal-CD101-Fc, and IgG1 signal-CD101-Fc. Of the three unsuccessful isolates, two contained 1- or 2-bp deletions in a junction region, and one contained two copies of the IgG1 signal sequence. When transiently transfected into COS cells, the IL-2 signal-CD101-Fc construct was expressed well [optical density (OD) 0.09], but the EPO signal-CD101-Fc (undetectable) and IgG1 signal-CD101-Fc (undetectable) were not thus identifying the IL-2 signal sequence as one that could successfully direct the secretion of this large, difficult to express, protein.

An *Nco*I site (CCATGG) in a vector is a universal cloning site for the start of a coding region, as it includes an ATG as well as the end of a Kozak consensus translational start site (GCCACCATG) (8). Since the 15-bp overlap for In-Fusion is counted from the restriction enzyme cut site on the antisense DNA strand, the 15-bp overlap for an *Nco*I

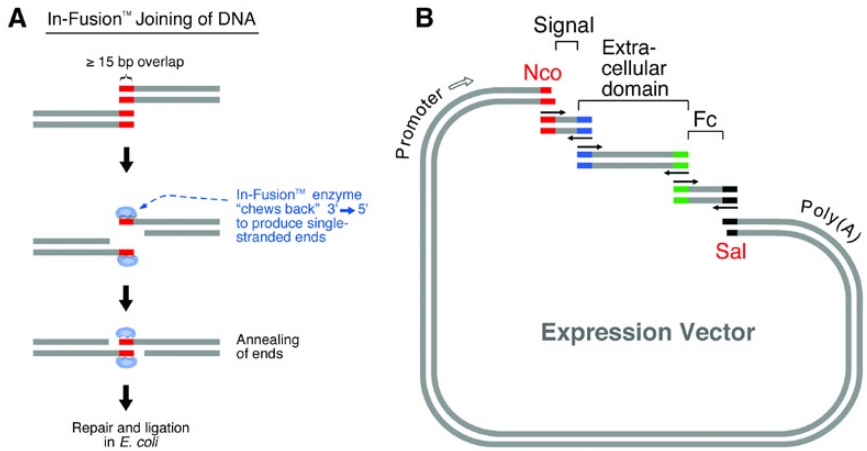


Figure 1. Mechanism of an In-Fusion reaction and its use in constructing a three-piece fusion protein. (A) In-Fusion reaction mechanism. (B) Seamless construction of an immunoglobulin fusion protein. Colored regions indicate overlap regions with 15 bp of identity. Arrows indicate PCR primers. Each segment is generated by PCR with primers that include the overlaps and joined to *Nco*I-*Sal*I-digested vector in an In-Fusion reaction.

Table 1. Primers for PCR Amplification of IL-2 Signal, CD101, and Murine IgG3 Segments

IL-2 Signal with Overlaps to *Nco*I-Digested Vector and CD101, 88-bp PCR Product
Sense (*Nco*I underlined)
5′-TTCAAATCCACCATGGATAGAATGCAATTGTTG-3′
Antisense
5′-CTGTTACTTCTCTCTGAGAATTCGTAACCAAAGCCAAAGACAAAGCAATCA-3′
CD101, 2799-bp PCR Product
Sense
5′-CAGAGAGAAGTAACAGTTCAGAAA-3′
Antisense
5′-GGCCGAGGAGCAGATCCTGGAA-3′
Murine IgG3 with Overlaps to CD101 and *Sal*I-Digested Vector, 771-bp PCR Product
Sense
5′-ATCTGCTCCTCGGCCCCTAGAATACCCAAGCCCAGTACC-3′
Antisense (*Sal*I underlined)
5′-**AGTAACGTTAGTCGAC**TCAGTGTCTTGTAAGACCCGAGGA-3′
Overlaps are colored to match Figure 1B. IL-2, interleukin 2; murine IgG3, murine immunoglobulin G3.

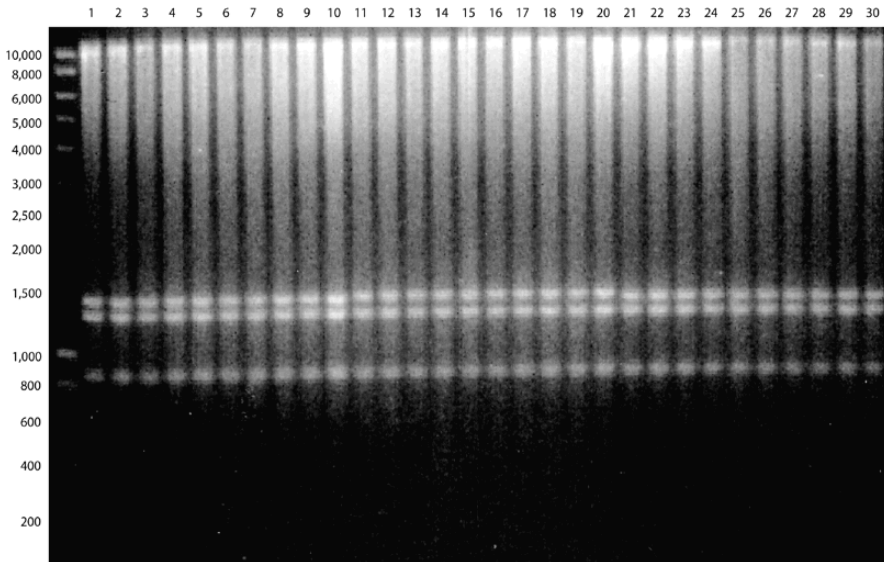


Figure 2. Plasmid minipreps of an In-Fusion reaction. Individual plasmids from the four-piece In-Fusion reaction illustrated in Figure 1B were digested with NcoI-*Sal*I and separated by electrophoresis on an agarose gel. Lanes 1–10, IL-2 signal-CD101-murine IgG3 Fc-vector; lanes 11–20, EPO signal-CD101-murine IgG3-vector; lanes 21–30, murine IgG1 signal-CD101-murine IgG3-vector. All plasmids had the 11.4-kb vector band and three other bands (there are two internal *Nco*I sites in CD101), indicating all pieces of DNA were incorporated into the construct. IL-2, interleukin 2; IgG, immunoglobulin G; EPO, erythropoietin; Fc, fragment crystallizable.

site is 5′-10 bp of vector-CCATG-3′, and the next nucleotide in the gene-specific region of the primer can match the natural nucleotide in the gene, thus requiring no change in the natural coding sequence. An *Nco*I site will be recreated if the next nucleotide is a G.

Technically, we see fewer background colonies if the insertion site in a plasmid vector is created by digestion with two restriction enzymes rather than one. In-Fusion may inefficiently facilitate the joining of the homologous ends of a singly digested vector. Phosphatasing of the vector is unnecessary and reduces efficiency. Accurate quantitation and ratios of vector and DNA segments are critical for success of multipiece constructs. Of 17 recent constructs, 28 of 38 minipreps sequenced were error-free and none required sequencing of more than 3 isolates to obtain the desired sequence. Of these 17 constructs, two-, three-, and four-piece In-Fusion reactions resulted in an average of 671, 539, and 56 colonies, respectively, on a plate spread with 0.1 mL of a total 0.3-mL transformation reaction. The use of highly competent *E. coli* (1×10^9 cfu/µg) is recommended for multipiece In-Fusion reactions as the number of transformants decreases as the number of pieces of DNA in the In-Fusion reaction increases. The size of the DNA segments appears to be unimportant and has varied from 83–12,000 bp. The length of the overlap between segments should be at least 15 bp, and slightly longer also works. The overlap sequence in the primers is not limited by G/C content or DNA sequence and may engineer in new restriction sites, but should avoid long hairpins at the end.

In-Fusion can be used to facilitate mutation of DNA via two strategies. As illustrated in Figure 3A, two unique restriction sites are identified that flank a desired mutation in a plasmid. Ideally, these are located 200–400 bp from the desired mutation so PCR products are >200 bp. The desired mutation is designed in silico and sense and antisense primers that incorporate the mutation (base pair changes, deletions, additions) and 20–30 bp of gene-specific priming sequence are synthesized. The mutagenic primers must overlap by 15 bp at their 5′ ends. Primers are also made that have 15 bp of identity flanking the unique restriction sites and 20–30 bp of gene-specific priming sequence. The two DNA segments are PCR amplified and joined with restriction enzyme-digested plasmid by In-Fusion, thereby incorporating the mutation.

As an example, primers for mutation W45A of human TIM-4 (hTIM-4) (9) are shown in Table 3. A TIM-4 template for PCR was made by *Xho*I digest of an hTIM-4 cDNA clone in pEF6 and gel purification of a 1566-bp fragment containing the cDNA. A 206-bp 5′ PCR product containing the W45A mutation was made using the antisense mutant and 5′ sense primers of Table 3. A 509-bp 3′ PCR product containing the W45A mutation was made using the sense mutant primer and 3′ antisense primers of Table 3. The hTIM-4 cDNA clone in pEF6 was digested with *Spe*I and *Bsp*EI, the 6375-bp fragment containing the vector and 3′ end of hTIM-4 was gel-purified, and 24 ng was combined with 1.6 ng 206-bp 5′ mutant PCR product and 3.8 ng of 509-bp 3′ mutant PCR product in a 10 µL In-Fusion reaction (1:2:2 molar ratio). Of five TIM-4 mutants made in this fashion, sequencing revealed 3 of 3, 3 of 3, 2 of 3, 2 of 3, and 1 of 8 to contain an error-free sequence.

An alternative strategy (Figure 3B) for mutagenesis is to use the mutagenic sense and antisense primers in a PCR with circular plasmid as template. A linear PCR product is generated that

Table 2. Four-Piece In-Fusion Reaction of Figure 1B and Figure 2

Secretory Signal	CD101 Domain	Murine IgG3 Fc Tail	Vector	Total Transformants	Error-Free/ Sequence (No.)
IL-2, 88 bp, 1.6 ng	2799 bp, 49.3 ng	771 bp, 13.6 ng	11,365 bp, 100 ng	690	1/2
EPO,108 bp, 1.9 ng	2799 bp, 49.3 ng	771 bp, 13.6 ng	11,365 bp, 100 ng	1150	1/2
Murine IgG1, 84 bp, 1.5 ng	2799 bp, 49.3 ng	771 bp, 13.6 ng	11,365 bp, 100 ng	2130	1/2

IgG1, immunoglobulin G1; Fc, fragment crystallization; IL-2, interleukin 2; EPO, erythropoietin.

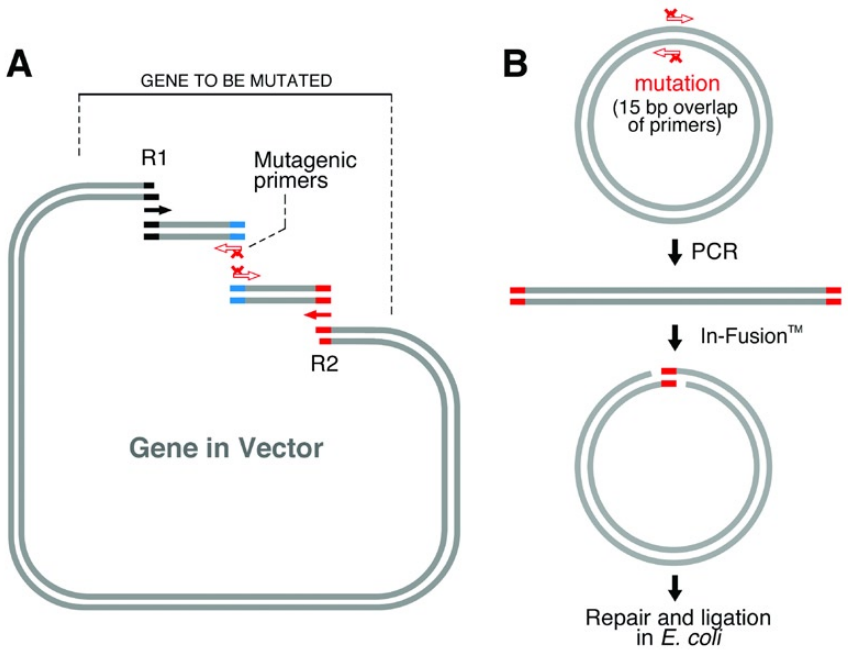


Figure 3. DNA mutagenesis strategies. (A) R1 and R2 indicate unique restriction sites flanking the desired mutation. Two DNA segments are amplified with the mutation incorporated at the DNA end via mutagenic primers and re-introduced into R1 + R2 digested plasmid via In-Fusion. (B) The mutagenic primers are used in PCR with circular plasmid as template to generate a linear molecule that is gel-purified and re-circularized via In-Fusion. Colored regions indicate overlap regions with 15 bp of identity.

Table 3. Primers for Mutagenesis

Primers for PCR Amplification to Make Human TIM-4 W45A Mutant (for Figure 3A)
5′ Sense Primer (Overlap with Vector Underlined, SpeI Site Bold)
5′-GCTCGGATCC**ACTAGT**CCAGTGTG-3′
Antisense Mutant Primer (Mutation Underlined)
206-bp PCR Product
5′-GAAGCGGATGAGTACAGACAGGGCAAAGTC-3′
Sense Mutant Primer (Mutation Underlined)
5′-TGTACTCATCCGCTTCTCACAACAGCAACAGC-3′
3′ Antisense Primer (BspEI Site Bold)
509-bp PCR product
5′-TGTGGCTTCC**TCCGGA**AGGGTGCTTGGGGTTA-3′
Primers for PCR Amplification to Make Human IgG1 Hinge C103S Mutant (for Figure 3B)
4609-bp PCR Product
Sense Mutant Primer (Mutation Underlined)
5′-AGTTGAGCCCAAATCTTCCGACAAAACTCACACA-3′
Antisense Primer
5′-GATTTGGGCTCAACTTTCTTGTCCACCTTGGTGT-3′
Overlaps are colored to match Figure 3. IgG1, immunoglobulin G1.

can then be recircularized in an In-Fusion reaction. As an example, 0.1 ng of a supercoiled plasmid containing human IgG1 cDNA in pCR-Blunt II was used as a template in a PCR with the human IgG1 sense mutant and antisense primers shown in Table 3. Cycling conditions were one cycle of 95°C for 2 min; 30 cycles of 95°C for 20 s, 59°C for 20 s, and 72°C for 2 min; and one cycle of 72°C for 3 min. The 4609-bp linear PCR product was gel-purified, and 25 ng were incubated in a 10-µL In-Fusion reaction, resulting in circularization of the plasmid via the 15-bp overlap at the ends. One of three plasmids sequenced had the desired sequence. The strategy of Figure 3A is more labor intensive but requires re-sequencing of a smaller region of DNA and is very reliable. The strategy of Figure 3B may be limited by the difficulty of PCR amplifying a large segment of DNA and likely requires re-sequencing of a larger region; however, it does not require the identification of unique restriction sites.

In-Fusion can also be used to replace any DNA segment in a plasmid with any desired new DNA segment. We term this strategy replacement In-Fusion, and two variations are illustrated in Figure 4. In the first strategy (Figure 4A), two unique restriction sites are identified that flank the desired replacement in a plasmid. Ideally, these are located 200–400 bp from the desired replacement, so PCR products are >200 bp. Primers are designed to amplify the two regions between the boundary of the replacement and each unique restriction site plus 15-bp overlaps. The replacement segment is amplified from an appropriate template using primers with a 15-bp overlap to the ends of the plasmid PCR product and 20–30 bp of segment specific sequence. The two PCR products, the replacement segment, and plasmid digested with the two unique restriction enzymes are joined in an In-Fusion reaction.

An alternative strategy has been described (10) and is shown in Figure 4B. An antisense primer is designed at the 5′ boundary of the segment of plasmid to be replaced, and a sense primer is designed at the 3′ boundary of the segment of plasmid to be replaced. Using the circular plasmid as a template, these primers are used to PCR amplify the desired region of the plasmid as a linear molecule lacking the segment to be replaced. The replacement segment is amplified from an appropriate template using primers with a 15-bp overlap to the ends of the plasmid PCR product and 20–30 bp of

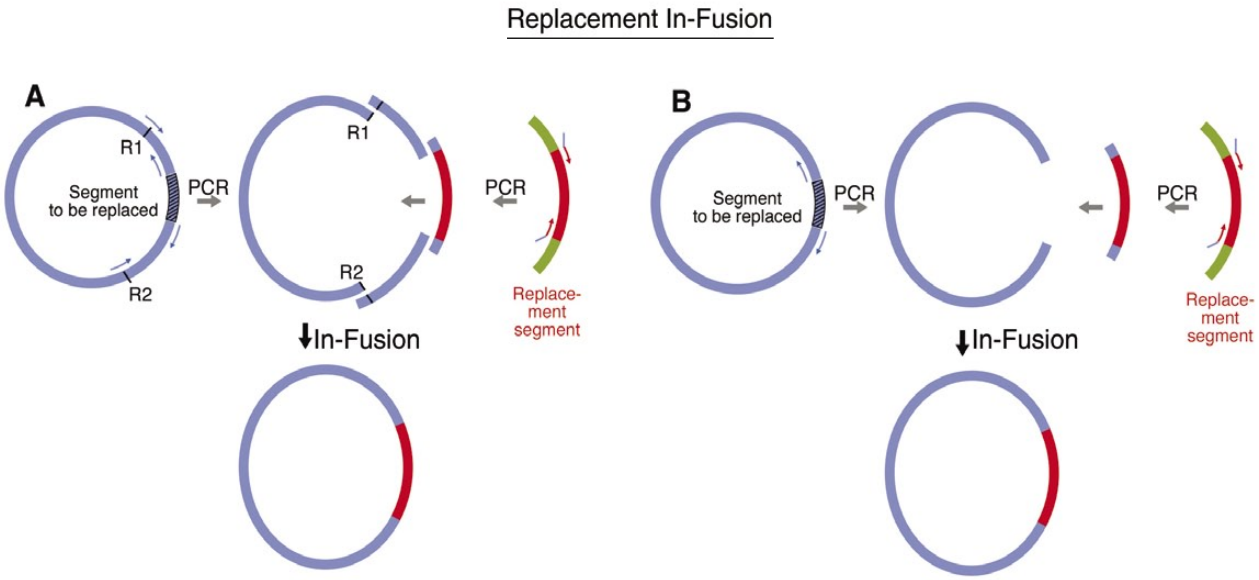


Figure 4. Replacement In-Fusion. The segment to be replaced is indicated by hatched lines. Sense and antisense primers are designed at the boundaries of the DNA segment to be replaced. The 3′ ends of the primers define the point of deletion. The replacement segment, indicated in red, is PCR amplified from an appropriate template using primers with a 15-bp overlap to the ends of the plasmid PCR product and 20–30 bp of replacement segment-specific sequence. Primers are indicated by arrows and are color-matched to their regions of identity. (A) Primers are designed to amplify the regions between the boundaries of the replacement and unique restriction enzymes sites R1 and R2 plus a 15-bp overlap. The two PCR products, the replacement segment, and plasmid digested with R1 and R2 are joined in a four piece In-Fusion reaction. (B) Sense and antisense primers at the boundaries of the DNA segment to be replaced are used to PCR amplify the desired region of the plasmid as a linear molecule lacking the segment to be replaced. The linear plasmid and replacement segment DNAs are joined by an In-Fusion reaction.

segment-specific sequence. The two DNAs are then joined by an In-Fusion reaction. No restriction sites are used in this strategy. Difficulty in PCR amplifying a very large segment of a plasmid suggests the use of the strategy shown in Figure 4A.

In-Fusion facilitates the construction of modular vectors and synthetic genes (Figure 5). One develops a toolbox of DNA segments that can be joined as desired, including fusion protein partners [Fc, green fluorescent protein (GFP), etc.], epitope tags, linkers, promoters, poly(A) sites, selectable markers, and origins of replication. A modular vector with the desired DNA segments can be designed in silico. Restriction enzymes that do not cut are identified, and a unique site is incorporated in silico between each segment. The toolbox DNAs can be used as templates for PCR with primers that include 15-bp overlaps at the desired junctions. The 15-bp overlaps may incorporate a different unique restriction enzyme site in each junction. Up to four pieces at a time may be joined by In-Fusion reaction, creating a modular vector whose individual components may be readily excised and replaced.

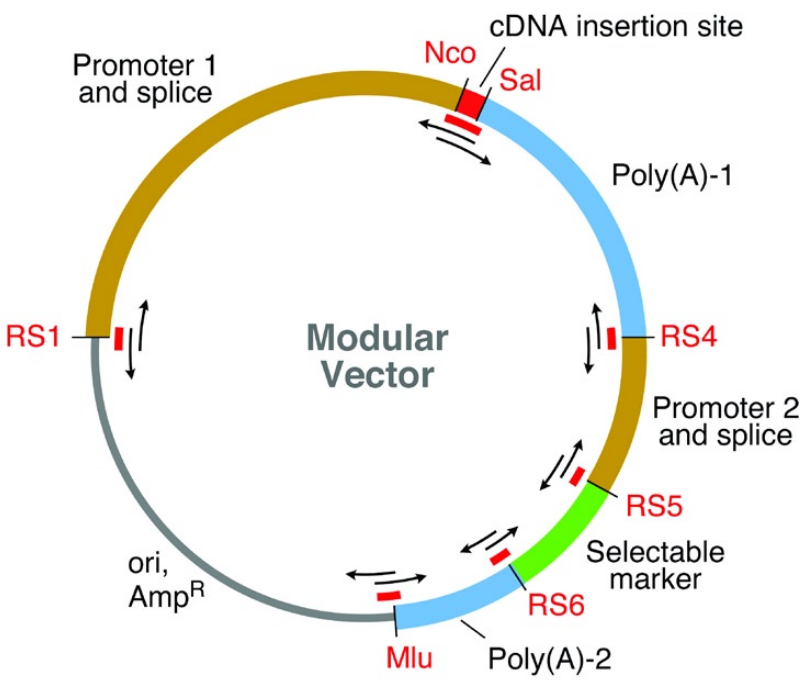


Figure 5. Modular vector. Red bars indicate junctions between segments. Different unique restriction sites (RS) may be engineered into the junctions via the primers used to amplify the segment. Each segment is amplified with primers that include 15 bp of overlap with the adjacent segment. DNAs are joined via In-Fusion reaction so no restriction digest is required. *Nco*I is a useful site as it is a universal acceptor for a coding region. *Mlu*I is a convenient site for linearization of vector for transfection.

Artificial genes can be constructed via PCR with overlapping sets of primers; however, this method is difficult for sequences longer than a few hundred base pairs (11,12). The construction of larger artificial genes can be facilitated by PCR amplifying the gene in 300-bp segments with 15-bp overlaps at the DNA ends and joining the pieces and vector via an In-Fusion reaction.

The use of two-piece In-Fusion reactions in high-throughput applications has been described (3,13) and additional work is needed to test whether multipiece In-Fusion reactions are adaptable to high-throughput. In-Fusion assembly provides a powerful way to join multiple pieces of DNA and facilitates the construction of seamless fusion proteins and modular vectors with readily interchangeable segments. Replacement In-Fusion can be used to replace any DNA segment in a plasmid with any desired new DNA segment without limitations to position or size.

ACKNOWLEDGMENTS

This work was supported by NIH grants AI39671, AI54456, and AI56299 (to G.J.F.).

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

1. **Evans, D.H., D.O. Willer, and X.-D. Yao.** 2003. DNA joining method, p. 1-52. In patent application 20030162265, USA.
2. **Hamilton, M.D., A.A. Nuara, D.B. Gammon, R.M. Buller, and D.H. Evans.** 2007. Duplex strand joining reactions catalyzed by vaccinia virus DNA polymerase. Nucleic Acids Res. *35*:143-151.
3. **Marsischky, G. and J. LaBaer.** 2004. Many paths to many clones: a comparative look at high-throughput cloning methods. Genome Res. *14*:2020-2028.
4. **Zhang, L., Q. Leng, and A.J. Mixson.** 2005. Alteration in the IL-2 signal peptide affects secretion of proteins in vitro and in vivo. J. Gene Med. *7*:354-365.
5. **Lim, L.H., H.Y. Li, N. Cheong, B.W. Lee, and K.Y. Chua.** 2004. High-level expression of a codon optimized recombinant dust mite allergen, Blo t 5, in Chinese hamster ovary cells. Biochem. Biophys. Res. Commun. *316*:991-996.
6. **Hall, K.T., L. Boumsell, J. Croopnick, A. Bensussan, L.M. Nadler, and G.J. Freeman.** 1997. T-cell genetic studies: molecular cloning of CD101 reveals it to contain seven Ig-V domains with identity to the V7 gene, p. 120-121. In T. Kishimoto, H. Kikutana, A.E.G. von dem Borne, S.M. Goyert, D.Y. Mason, M. Miyasaka, L. Moretta, K. Okumura, et.al. (Eds.), Leukocyte Typing VI: White Cell Differentiation Antigens. Garland Publishing, New York.
7. **Gajewski, T.F., F. Fallarino, C. Uyttenhove, and T. Boon.** 1996. Tumor rejection requires a CTLA4 ligand provided by the host or expressed on the tumor: superiority of B7-1 over B7-2 for active tumor immunization. J. Immunol. *156*:2909-2917.
8. **Kozak, M.** 2005. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene *361*:13-37.
9. **Kuchroo, V.K., D.T. Umetsu, R.H. DeKruyff, and G.J. Freeman.** 2003. The TIM gene family: emerging roles in immunity and disease. Nat. Rev. Immunol. *3*:454-462.
10. **Benoit, R.M., R.N. Wilhelm, D. Scherer-Becker, and C. Ostermeier.** 2006. An improved method for fast, robust, and seamless integration of DNA fragments into multiple plasmids. Protein Expr. Purif. *45*:66-71.
11. **Hoover, D.M. and J. Lubkowski.** 2002. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. Nucleic Acids Res. *30*:e43.
12. **Stemmer, W.P.C., A. Crameri, K.D. Ha, T.M. Brennan, and H.L. Heyneker.** 1995. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. Gene *164*:49-53.
13. **Berrow, N.S., D. Alderton, S. Sainsbury, J. Nettleship, R. Assenberg, N. Rahman, D.I. Stuart, and R.J. Owens.** 2007. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic Acids Res. *35*:e45.

Received 14 May 2007; accepted 27 June 2007.

Address correspondence to Dr. Gordon Freeman, Department of Medical Oncology, Dana-Farber Cancer Institute, Department of Medicine, Harvard Medical School, Boston, MA 02115, USA. e-mail: Gordon_Freeman@dfci.harvard.edu

Published online 12 January 2012

Nucleic Acids Research, 2012, Vol. 40, No. 8 ***e55***
doi:10.1093/nar/gkr1288

SLiCE: a novel bacterial cell extract-based DNA cloning method

Yongwei Zhang*, Uwe Werling and Winfried Edelmann*

Department of Cell Biology, Albert Einstein College of Medicine, Bronx, NY 10461, USA

Received July 28, 2011; Revised December 13, 2011; Accepted December 15, 2011

ABSTRACT

We describe a novel cloning method termed SLiCE (Seamless Ligation Cloning Extract) that utilizes easy to generate bacterial cell extracts to assemble multiple DNA fragments into recombinant DNA molecules in a single *in vitro* recombination reaction. SLiCE overcomes the sequence limitations of traditional cloning methods, facilitates seamless cloning by recombining short end homologies (≥15 bp) with or without flanking heterologous sequences and provides an effective strategy for directional subcloning of DNA fragments from Bacteria Artificial Chromosomes (BACs) or other sources. SLiCE is highly cost effective as a number of standard laboratory bacterial strains can serve as sources for SLiCE extract. In addition, the cloning efficiencies and capabilities of these strains can be greatly improved by simple genetic modifications. As an example, we modified the DH10B *Escherichia coli* strain to express an optimized λ prophage Red recombination system. This strain, termed PPY, facilitates SLiCE with very high efficiencies and demonstrates the versatility of the method.

INTRODUCTION

The generation of recombinant DNA molecules is an essential tool in modern molecular biology. The conventional DNA cloning strategies that have been used for several decades typically involve the use of type II restriction enzymes to generate appropriate DNA fragments, the modification of DNA ends to generate blunt or sticky ends and the ligation of the DNA fragments to generate plasmid or other type DNA vectors (1–3). However, these procedures depend on the presence of appropriate restriction sites to generate both vector and insert molecules and often leave unwanted sequences at the junction sites. In addition, the restriction enzymes and modifying enzymes required for these manipulations are often expensive making these procedures costly especially in high throughput settings. To circumvent these limitations, we developed a new restriction site independent cloning method that does not leave any unwanted sequences at the junction sites (seamless) and is based on *in vitro* recombination between short regions of homologies (15–52 bp) in bacterial cell extracts termed SLiCE (Seamless Ligation Cloning Extract). SLiCE allows for efficient restriction site independent cloning of DNA fragments generated by restriction digestion or PCR amplification into linearized vectors. In addition, SLiCE does not require the use of enzymes for the modification of vector and insert end sequences (such as Klenow or T4 DNA polymerase) or ligases. The SLiCE method can be used for virtually any type of cloning approach including the simple subcloning of PCR or restriction fragments, the generation of tagged expression vectors, the construction of more complex vectors such as gene targeting vectors or the directional subcloning of larger DNA fragments from more complex vectors such as bacterial artificial chromosomes (BACs). In addition, SLiCE allows the assembly of multiple DNA fragments in one cloning step, which may make it an ideal method for the assembly of multiple DNA fragments during gene synthesis applications.

The SLiCE method is based on bacterial extracts that can be derived from a variety of common RecA$^-$ *Escherichia coli* laboratory strains such as DH10B and JM109. These strains can also be further optimized by simple genetic modifications to improve SLiCE cloning efficiencies and capabilities making SLiCE highly versatile. For example, we established a DH10B-derived strain, termed PPY that was engineered to contain an optimized λ prophage Red recombination system (4–6). We found that extracts derived from this strain provide the highest cloning efficiencies thus far and that it can be used for all cloning approaches that are routinely used in the laboratory. The SLiCE method overcomes many problems related to conventional cloning procedures and provides a highly cost-effective approach for the

*To whom correspondence should be addressed. Tel: +1 718 678 1087; Fax: +1 718 678 1019; Email: yongwei.zhang@einstein.yu.edu
Correspondence may also be addressed to Winfried Edelmann. Tel: +1 718 678 1086; Fax: +1 718 678 1019;
Email: winfried.edelmann@einstein.yu.edu

generation of recombinant DNA molecules in a seamless and restriction site independent manner. In this report, we describe the SLiCE method, its principal features and applications.

MATERIALS AND METHODS

Bacteria strains

The following laboratory *E. coli* strains were used: DH10B (Invitrogen), JM109 (Promega), BL21(DE3) (Invitrogen), BLR(DE3) (Novagen) and ER2566(NEB).

The DH10B derived *E. coli* strain PPY was constructed by Suicide Plasmid Based Genome Modification (7) using plasmid pGT1 (PPY genotype: F^- *end*A1 *rec*A1 *gal*E15 *gal*K16 *nup*G *rps*LΔ*lac*X74 *Φ80lac*ZΔM15 *ara*D139 Δ(*ara,leu*)7697 *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) *cyn*X:: [*ara*C pBAD- *redα* EM7- *redβ* Tn5-*gam*] λ^-).

The competent cells used for transformation of SLiCE generated recombination products were ElectroMAX *DH10B*^{TM} cells and MAX Efficiency® *DH10B*^{TM} competent cells (Invitrogen).

Plasmids

The plasmid pBL was constructed by insertion of a 70-bp chemically synthesized multiple cloning site into the 2.5-kb PCR-generated plasmid backbone of pBluescript II KS(+) (Stratagene) and deletion of the *LacZα* ORF by conventional cloning. The plasmid pBL-DL was constructed by insertion of a 1-kb PCR fragment from pGEM®-luc (Promega) into the NotI/SalI sites of pBL by SLiCE. The suicide plasmid pGT1 was constructed by SLiCE-mediated insertion of a 830-bp PCR-amplified fragment spanning the 3′ region of the *E. coli* DH10B *cyn*X gene and an *ara*C-pBAD-*redα*/EM7-*redβ*/Tn5-*gam* expression cassette isolated from plasmid pBAD24 (8) and lambda phage DNA (NEB) into the *Sma*I site of plasmid pEL04 (9). pGT1 also contains a temperature-sensitive replicon and a chloramphenicol selection marker.

Preparation of SLiCE extract

Escherichia coli strains were grown at 37°C in 100 ml 2X YT medium until they reached $OD_{600} \approx 5.3$ (OD_{600} readings were calculated by diluting the sample to enable photometric measurement in the linear range between 0.1 and 0.5 OD_{600}). PPY was subsequently incubated for 2 h in 2X YT medium containing 0.2% L-arabinose to express λ prophage protein Redα. The cells were harvested by centrifugation at 5000g for 20 minutes at 4°C. The cells from 96 ml of original culture (wet weight ≈0.92 g) were washed 1 time with 200 ml ddH_2O and resuspended in 1.2 ml CelLytic^{TM} B Cell Lysis Reagent (Sigma). The resuspended cells were incubated at room temperature for 10 minutes to allow lysis to occur. The cell lysates were centrifuged at 20 000*g* for 2 min at room temperature to pellet the insoluble material. The resulting supernatants were carefully removed from the cell debris into a low binding 1.5 ml tube (Protein LoBind Tube 1.5 ml, Eppendorf). The cell extracts were mixed with equal volume of 100% glycerol, aliquoted into 40–60 ul portions in low binding 0.5 ml tubes (Protein LoBind Tube 0.5 ml, Eppendorf), and stored at −20°C for about 2 months without significant loss of activity. For long-term storage, the aliquoted cell extracts were stored at −80°C in 50% glycerol, which can be thawed on wet ice and refrozen up to 10 times without significant loss of activity.

SLiCE reaction and transformation

The vectors used for SLiCE were linearized by restriction digestion or PCR amplification. The cloning inserts were PCR amplified using primers containing 5′-end homologies to the vector or to other inserts for coassembly. Vector or insert DNAs that were generated by PCR amplification using plasmid DNA as templates were treated with DpnI prior to purification to remove residual plasmid template DNA. The linearized vectors and PCR inserts were subjected to gel electrophoresis and purified using the QIAEX II gel extraction kit. For SLiCE cloning of BAC DNA, the restriction digested BAC DNA was purified by phenol/chloroform extraction.

The standard SLiCE reaction mix contained the following ingredients: 50–200 ng linear vector, appropriate amount of insert DNA in a 1 : 1 to 10 : 1 molar ratio of insert to vector, 1 µl 10X SLiCE buffer (500 mM Tris–HCl (pH 7.5 at 25°C), 100 mM $MgCl_2$, 10 mM ATP, 10 mM DTT), 1µl SLiCE extract and ddH_2O to a total volume of 10 µl. The SLiCE reaction was incubated at 37°C for 1 h and subsequently 1 µl of the SLiCE reaction was electroporated into 20 µl ElectroMAX *DH10B*^{TM} cells (Invitrogen) or chemically transformed into 100 µl MAX Efficiency® *DH10B*^{TM} competent cells (Invitrogen) following the manufacturer's instructions. The transformation efficiencies of ElectroMAX *DH10B*^{TM} cells (Invitrogen) and MAX Efficiency® *DH10B*^{TM} competent cells (Invitrogen) were $\sim 1 \times 10^{10}$ and 1×10^9 transformants/µg of pUC19 DNA, respectively. The transformed cells were plated on ampicillin/Xgal agar plates or agar plates containing appropriate antibiotics.

RESULTS

Comparison of *E. coli* K12 strains for SLiCE

SLiCE is a cloning method that is based on *in vitro* recombination in bacterial extract. SLiCE is a simple and efficient procedure with the entire process consisting of three steps (Figure 1a): (i) The preparation of linear vector and insert fragments containing short end homologies introduced by PCR with primers having appropriate 5′ extension sequences; (ii) the SLiCE *in vitro* reaction and (iii) the standard transformation (electroporation or chemical transformation) of recombination products into suitable host bacteria. In this article, all bacterial transformations were performed by electroporation using ElectroMAX DH10B^{TM} cells (Invitrogen) unless otherwise noted.

To determine the most efficient bacterial strains for SLiCE, five standard laboratory *E. coli* K12 strains were tested including DH10B, JM109, ER2566, BL21 (DE3) and BLR (DE3). The main criteria for their selection

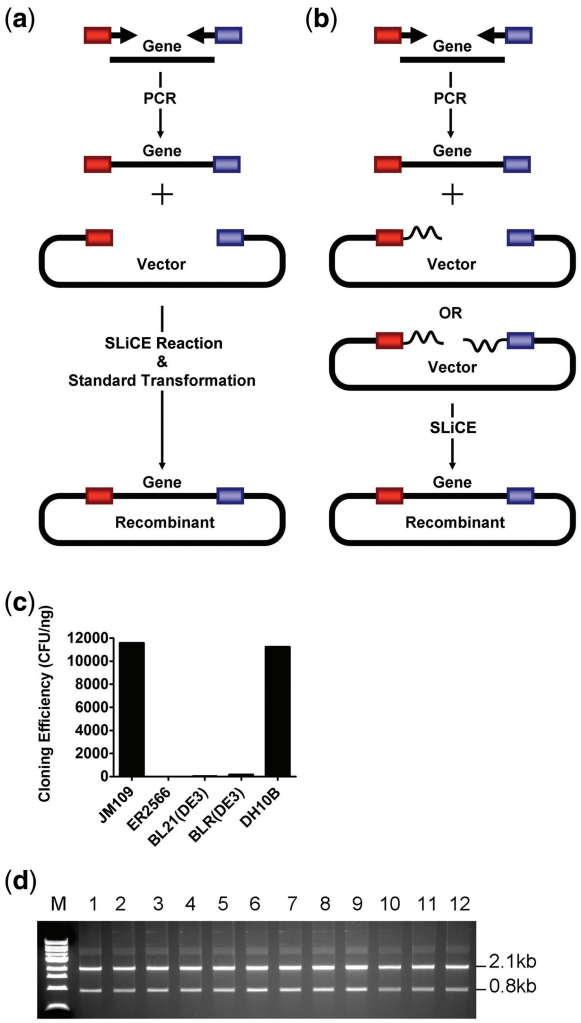


Figure 1. SliCE cloning. (**a**) Outline of SLiCE Cloning. (**b**) Schematic illustrating seamless SLiCE cloning with flanking heterologous sequences. (**c**) Comparison of SLiCE efficiency of *E. coli* K 12 strains. (**d**) BsaAI/SapI restriction analysis of the recombinants derived from SLiCE cloning. Plasmid DNAs from 12 independent ampicillin-resistant blue colonies (lanes 1–12) were digested with BsaAI/SapI. The digestion products were separated on a 1% agarose gel and visualized after ethidium bromide staining. Recombinant plasmids contain one BsaAI site and one SapI site yielding diagnostic 2.1- and 0.8-kb restriction fragments.

included their genetic status with regard to the RecA homologous recombination protein and the presence of restriction systems that could interfere with the stability of exogenous DNA (Supplementary Table S1).

To determine the suitability of extracts derived from these strains for SLiCE, a simple cloning strategy was devised. A plasmid vector termed pBL was linearized by restriction digestion and incubated together with a 500-bp PCR-amplified LacZα in the different bacterial extracts. To facilitate recombination, the LacZα fragment contained 42-bp end sequences that were homologous to the end sequences of pBL. According to the experimental design positive recombinant clones could be identified by blue/white selection after transformation and growth of bacteria on ampicillin/Xgal agar plates.

This analysis showed that extracts from the two RecA$^-$ strains DH10B and JM109 yielded the highest cloning efficiencies, indicating that SLiCE is facilitated by a RecA independent mechanism (Figure 1c).

Due to the high cloning efficiency, cell extract from the DH10B strain was further used for the analysis of SLiCE. Using the pBL–LacZα cloning strategy described above, the influence of several parameters on SLiCE cloning efficiency was tested including the lengths of end homologies, the vector/insert ratio, the overall DNA concentration and the transformation methods. To determine the effect of end homology length on SLiCE, homologies ranging from 0 to 100 bp (as counted from the 3′-ends of the vector) were tested (Table 1). This analysis showed that inserts without end homology or 10-bp end homology did not yield any recombinant colonies. In contrast, 15 bp of end homology already yielded an appreciable number of recombinant colonies at a cloning efficiency of 75 colony forming unit (CFU)/ng of vector, while 30 bp of end homology provided very robust cloning efficiencies (920 CFU/ng of vector). A further increase in end homology length resulted in even higher cloning efficiencies with an end homology length of 52 bp providing the highest efficiency (21 965 CFU/ng of vector). However, in contrast to *in vivo* homologous recombination cloning, the cloning efficiency dropped significantly when the end homology length was further increased (Table 1) indicating that SLiCE promotes *in vitro* recombination by a different pathway. Control reactions that contained the same vector/insert combinations with increasing end homologies but without SLiCE extract did not yield any recombinant colonies. Next, SLiCE was performed with varying vector/insert molar ratios at a vector concentration of 10 ng/μl and 42 bp of end homology. These studies showed that vector/insert ratios of 1 : 1, 1 : 2, 1 : 6 and 1 : 10 yielded 1335, 2330, 11 120 and 12 120 CFU/ng of vector, respectively, demonstrating that increased insert ratios could yield higher cloning efficiencies. Compared to the standard vector concentration of 10 ng/μl, SLiCE using low concentrations of vector (1 ng/μl) and a vector/insert at ratio of 1 : 1 led to a 200-fold decrease in cloning efficiency which is likely due to the instability of vector and insert DNA at these low concentration in the SLiCE reaction mix. All the data above were derived by electroporation of the SLiCE reaction products. We also performed chemical transformation of SLiCE reaction products with 42 bp of end homology using MAX Efficiency® *DH10B*^{TM} competent cells (Invitrogen), which yielded about a 10-fold lower cloning efficiency than electroporation (1063 CFU versus 10 480 CFU/ng of vector, chemical transformation versus electroporation).

Besides promoting recombination between homologous sequences at the ends of vector and inserts, SLiCE is also capable of facilitating recombination between DNA fragments that contain flanking heterologous sequences and of deleting the extra flanking sequences to generate precise junctions at the recombination sites. This feature of SLiCE provides a highly useful cloning tool, especially in those cases where the absence of suitable restriction sites in a vector prevent the seamless cloning of an insert

Table 1. Influence of End Homology Length on SLiCE Cloning

Homology length (bp)	Cloning efficiency		Cloning accuracy	
	DH10B SLiCE	PPY SLiCE	DH10B SLiCE	PPY SLiCE
Vector only	0	0	–	–
0	0	0	–	–
10	0	0	–	–
15	75	4 640	90%	99%
20	80	34 500	88%	99%
30	920	124 000	98%	99%
42	10 480	632 000	99%	99%
52	21 965	766 000	99%	99%
68	1795	162 000	99%	99%
78	3325	119 250	99%	99%
88	1890	68 000	99%	99%
100	3320	32 500	99%	99%

Cloning efficiencies using different lengths of end homologies are given as CFUs of blue colonies per nanogram of vector. Cloning accuracies are given as the percentage of blue colonies among the total number of all amp^r colonies (blue and white). The 2.5-kb vector pBL was linearized by NotI/SalI digestion and the 500–700-bp LacZα fragments were prepared by PCR. Experiments were performed using 10 ng/μl of vector and the corresponding amount of insert DNA at a 1 : 6 molar ratio of vector:insert. The blue colonies contain recombinant plasmid and the white colonies contain non-recombinant vector background.

fragment into a desired vector region (Figure 1b). To test this feature, DH10B SLiCE reactions were performed with vector pBL-DL that was designed to provide heterologous flanking sequences at the cloning sites. pBL-DL was linearized with appropriate restriction enzymes to generate flanking heterologous sequences on one side (319, 738 and 998 bp) or on both sides (45 bp plus 23 bp and 319 bp plus 738 bp) and was together with a LacZα fragment of 500 bp with 42 bp of end homologies (Figure 1b) subjected to SLiCE followed by blue/white selection. Our results showed that DH10B SLiCE can efficiently remove 45 bp plus 23 bp on both sides of flanking heterologous sequences but it cannot facilitate DNA cloning with longer flanking heterologous sequences of 319, 738, 998 bp on one side or 319 bp plus 738 bp on both sides (Supplementary Table S2).

Generation of a modified DH10B strain for the optimization of SLiCE

In vivo homologous recombination in *E. coli* can be facilitated by three different recombination pathways: the RecA dependent pathway, a RecA independent pathway of unknown nature and a RecA independent pathway that utilizes prophage Red/ET recombination systems (4–6,10–14). The studies above indicate that a RecA independent pathway catalyzes SLiCE. To optimize SLiCE and acquire even more efficient strains as a source for cell extracts, we modified the DH10B genome using a suicide plasmid based strategy to insert an optimized λ prophage Red recombination system into the bacterial genome. Specifically, the genome of DH10B bacteria were modified to constitutively express the λ phage *red*β and *gam* genes under the control of the EM7 and Tn5 promoters, respectively, and also the *red*α gene under the control of an arabinose-inducible pBAD promoter (*araC*-pBAD) (8). The modified DH10B strain was termed PPY and tested for SLiCE. Extracts derived from PPY bacteria yielded significantly higher cloning efficiencies and a more robust seamless cloning activity in the presence of heterologous flanking sequences than the DH10B extracts (see below) and were used in the following series of experiments for the analysis of optimized SLiCE capabilities.

Efficiency and fidelity of PPY SLiCE

In the first series of experiments we investigated the efficiency and fidelity of the improved PPY SLiCE extract. First, we examined the influence of end homology length on PPY SLiCE mediated cloning in more detail. Using the pBL–LacZα cloning strategy end homologies varying from 0 to 100 bp were examined (Table 1). Vector and insert fragments with no end homology or a short homology of 10 bp did not yield any recombinant colonies. Similar to DH10B SLiCE without the Red system, the minimum length of homology required for efficient cloning was 15 bp, however, the PPY SLiCE extract yielded a dramatic increase in the number of blue recombinant colonies (4 640 CFU versus 75 CFU per ng of vector, PPY SLiCE versus DH10B SLiCE) with a high cloning accuracy. Similar to DH10B SLiCE cloning, the cloning efficiency for PPY SLiCE increased with homology length in a range up to 52 bp but dropped significantly when the end homologies were further increased (Table 1).

PPY SLiCE-mediated cloning was also performed using another vector/insert combination (p3XFLAG-CMV-7.1 vector (Invitrogen) and a 800-bp PCR insert with end homologies ranging from 0 to 42 bp). These studies yielded similar results (data not shown).

The recombinant colonies derived from PPY SLiCE cloning were further analyzed by colony PCR, restriction digestion and DNA sequencing analyses. More than 300 blue colonies were screened using colony PCR (data not shown) and some of the colonies were verified by restriction digestion (Figure 1d). All of the analyzed clones contained the correct insert. The vector/insert junctions of

30 recombinant clones were sequenced and all of the clones contained the correct cloning junctions indicating that SLiCE fuses vector and insert in a precise manner. The small number of white background colonies that were observed in these test experiments could be traced back to spurious amounts of undigested pBL vector during linearization.

We next examined the fidelity of PPY SLiCE without prior selection of positive recombinant clones. For these studies we used a NotI/XbaI linearized 5.2-kb plasmid vector and a 1.4-kb PCR-amplified insert with 30 bp of end homologies. The entire insert and the junction regions of 20 positive recombinants were sequenced. Eighteen recombinants contained completely correct sequences and 2 recombinants presented one mutant nucleotide located in the PCR insert, which is consistent with the error rate of the DNA polymerase that was used for PCR amplification (Fastart Fidelity PCR system, Roche) at 2.4×10^{-6}/bp/cycle and 30 cycle amplification. The error rate and location of mutations indicate that these mutations were caused by PCR amplification of inserts and that SLiCE did not introduce further mutations.

We next examined the effect of various molar ratios of vector and insert and the overall DNA concentration on PPY SLiCE. The results were again similar to DH10B SLiCE mediated cloning. PPY SLiCE with pBL–LacZα vector/inserts at molar ratios of 1 : 1, 1 : 2, 1 : 6 and 1 : 10 with 20-bp end homologies yielded 19 240, 27 320, 34 500 and 65 350 CFU/ng of vector, respectively, showing that increasing the amount of insert yields slightly higher cloning efficiencies for PPY SLiCE. We also observed that PPY SLiCE at a low concentration of vector and insert (1 ng/μl) at a vector/insert ratio of 1 : 1 also resulted in a 10-fold reduced cloning efficiency.

We also determined the sequence dependence of PPY SLiCE in more detail. For this, we first generated four LacZα fragment and vector combinations containing 20-bp end homologies that each differed in GC content (ranging from 40% to 80%) and sequence. These vector/insert combinations were subjected to PPY SLiCE. We did not find a significant difference in the cloning efficiencies indicating that SLiCE is indeed sequence independent within the range of sequence diversity tested (Supplementary Table S3). Next, we generated four fragments with 20-bp end homologies containing a mismatch at different positions. We found that a single mismatched nucleotide at the end of the 20-bp homology produced one type of recombination product and the heterologous nucleotide was removed during SLiCE. A single mismatched nucleotide within the middle of the 20-bp homology produced two types of recombination products in which either one of the mismatched nucleotides was retained. In addition, we observed that the presence of a single mismatch only led to a slight reduction in cloning efficiency (Supplementary Table S4).

To determine the effect of insert length on PPY SLiCE cloning efficiencies we attempted to assemble vectors ranging in size from 2 to 15 kb and containing inserts ranging from 80 bp to 21 kb in size. We found that the cloning of larger fragments by SLiCE occurred at robust but somewhat reduced efficiencies. For example, the assembly of an 11-kb recombinant plasmid containing an 8-kb insert was achieved at a cloning efficiency of 140 CFU/ng of vector. The restriction analysis of 24 clones revealed that 22 contained the expected recombinant plasmid.

At present, we have successfully used SLiCE-mediated cloning to generate more than 100 recombinant plasmids employing various cloning strategies and many different vector/insert combinations. Our results indicate that SLiCE can be considered a universal cloning method for the generation of recombinant DNA at high fidelity. Furthermore, the nature of vector and insert ends such as blunt ends or 3′ or 5′ sequence overhangs did not influence SLiCE efficiency or accuracy. However, the use of vectors with complementary 5′ or 3′ overhanging ends for SLiCE increased the formation of empty vector background colonies, which is probably due to annealing of the single-stranded ends in the bacterial extracts or in the transformed host cells.

PPY SLiCE with flanking heterologous sequences

The seamless cloning activity of PPY SLiCE was examined using the same pBL-DL–LacZα cloning strategy as for DH10B SLiCE with flanking heterologous sequences at one side (2, 319, 738 and 998 bp) or on both sides (45 bp plus 23 bp and 319 bp plus 738 bp) and with various end homologies.

In comparison to DH10B extracts, PPY extracts presented a much stronger seamless activity, which can remarkably increase the efficiency of DNA cloning especially with shorter flanking heterologies (45 bp plus 23 bp on both sides; 7600 CFU versus 1265 CFU per ng of vector, PPY extract versus DH10B extract). In addition, PPY SLiCE can efficiently remove longer flanking heterologous sequences of up to 998 bp on one side or up to 319 bp plus 738 bp on both sides (Table 2). In general, vectors with shorter flanking sequences or flanking sequences on only one side yielded higher cloning efficiencies compared to vectors with longer and/or double-sided flanking heterologous sequences (Table 2). Similar to SLiCE cloning without flanking heterologous sequences, longer end homologies between vector and insert resulted in higher cloning efficiencies (Table 2).

SLiCE cloning with multiple fragments

The high cloning efficiency and fidelity of PPY SLiCE suggested it might be possible to generate more complex recombinant plasmids using multiple inserts in a single cloning reaction. To test this idea we designed two different SLiCE strategies for the cloning of multiple insert fragments. In the first strategy, we attempted to clone multiple inserts into one vector in one SLiCE reaction to generate a single recombinant DNA molecule derived from multiple fragments that was termed multiple-way SLiCE cloning (Figure 2a). The second strategy was designed to clone several different inserts carrying the same end homology into a vector in one SLiCE reaction in parallel. This strategy creates multiple different recombinant DNA molecules and was termed SLiCE batch cloning (Figure 2b).

Table 2. PPY SLICE with flanking heterologous sequences

Homology length (bp)	Flanking heterology length (bp)		Vector length (bp)	Cloning efficency	Cloning accuracy (%)
	Side 1	Side 2			
20	2	0	2500	10 000	99
42	319	0	2803	2270	99
30	319	0	2803	1250	98
42	738	0	3222	1232	87
30	738	0	3222	432	77
42	998	0	3482	570	94
42	45	23	2552	7600	81
30	45	23	2552	1288	63
20	45	23	2552	710	59
42	319	738	3541	5	98

Cloning efficiencies are given as CFUs of blue colonies per nanogram of vector. Cloning accuracies are given as the percentage of blue colonies among the total number of all amp^r colonies (blue and white). Vectors containing different end heterologies were derived from plasmid pBL-DL by digesting with various restriction enzymes. LacZ inserts of 500-bp size containing the indicated end homologies were generated by PCR. The experiments were performed using 10–40 ng/μl vector DNA and the corresponding amount of insert DNA at a 1 : 6 molar ratio of vector:insert in a 10 μl reaction volume. The blue colonies contain recombinant plasmid and the white colonies contain non-recombinant vector background.

To examine the potential of SLiCE for multiple-way cloning, three-, four- and seven-way SLiCE cloning was performed using the pBL vector and three different sets of PCR amplified inserts with 42 bp of end homology that can assemble into a single 1.9-kb DNA fragment expressing LacZα activity (Figure 2c and Table 3). Our studies showed that PPY SLiCE mediated three-, four- and seven-way cloning occurred at significant efficiencies and high accuracies (Table 3). DNA sequencing analysis showed that all of the multiple fragments were precisely joined by SLiCE-mediated multiple-way cloning.

We next examined multiple-way SLiCE cloning using other vectors and inserts. A three-way SLiCE using a 4.7-kb vector (p3XFLAG-CMV-7.1, Invitrogen) and two 250-bp inserts with 24 bp of end homologies produced about 500 CFU/ng of vector with a 10-fold stimulation over non-recombinant background colonies. In another three-way experiment we successfully assembled a 3-kb vector and two 2.6- and 2.5-kb inserts using 42 bp of end homology with 80% accuracy and a cloning efficiency of 60 CFU/ng of vector. To determine the ability of multiple-way SLiCE cloning to assemble highly complex vector/insert combinations, a four-way cloning strategy (42-bp end homology, 2.5 kb of vector, inserts of 500 bp, 1.4 and 2.5 kb) and a seven-way strategy with shorter end homology (24-bp homology, 2.5 kb of vector and six inserts totaling 2 kb) were performed. For both multiple-way cloning strategies the cloning efficiencies were reduced but still provided at least 20 CFU/10 ng of vector.

For SLiCE batch cloning, two sets of experiments were performed. Six PCR inserts varying from 300 bp to 1 kb with 30 bp of end homology were mixed together with a linearized 6.7-kb prokaryotic expression vector (PTXB1, NEB) and incubated in PPY SLiCE extract (Figure 2b). About 340 CFU/ng of vector were obtained. The analysis of 32 colonies showed that all six possible recombinant vector/insert combinations were obtained (Figure 2d and e). Another experiment using a 4.7-kb mammalian expression vector (p3XFLAG-CMV-7.1, Invitrogen) and three inserts of 1, 1.5 and 2.5 kb with 24-bp homology yielded similar results (data not shown).

SLiCE cloning of genomic fragments from BAC clones

It is often challenging to subclone a genomic DNA fragment from larger DNA vectors such as BACs into a plasmid vector. Due to the high cloning efficiency of PPY SLiCE, we tested whether SLiCE could also facilitate this type of cloning. A SLiCE cloning strategy was designed to subclone individual genomic DNA fragments from BAC vectors (Figure 3a). Specifically, BAC DNA isolated from clone *RP23-303G13* (CHORI), 165 kb in size and containing 66 *Bgl*II and 19 *Eco*RV restriction sites, was digested with either *Bgl*II or *Eco*RV to generate a complex pool of DNA fragments. The digested BAC DNA was phenol/chloroform purified and subjected to PPY SLiCE cloning with PCR-generated pBluescript II KS(+) (Stratagene) derived vectors that contained end homologies to different *Bgl*II or *Eco*RV BAC restriction fragments. We attempted SLiCE cloning of several *Bgl*II BAC fragments of different sizes (830 bp, 3.7, 6.7, 8.7 and 14 kb) with 42 or 52 bp of end homology. In addition, SLiCE cloning was also performed for several *Eco*RV BAC fragments of larger sizes (5.3, 6.3, 12.2 and 21 kb) (Table 4). In all cases we were able to obtain recombinant clones carrying the different BAC fragments with high or acceptable cloning efficiencies (Table 4 and Figure 3b and c), indicating that SLiCE cloning is an effective strategy for the directional subcloning of small or large BAC genomic fragments.

DISCUSSION

The observation that bacterial cell extracts can efficiently recombine DNA molecules using short-end homologies was a serendipitous discovery in our laboratory. After initial characterization and further optimization, we were able to establish a novel restriction site

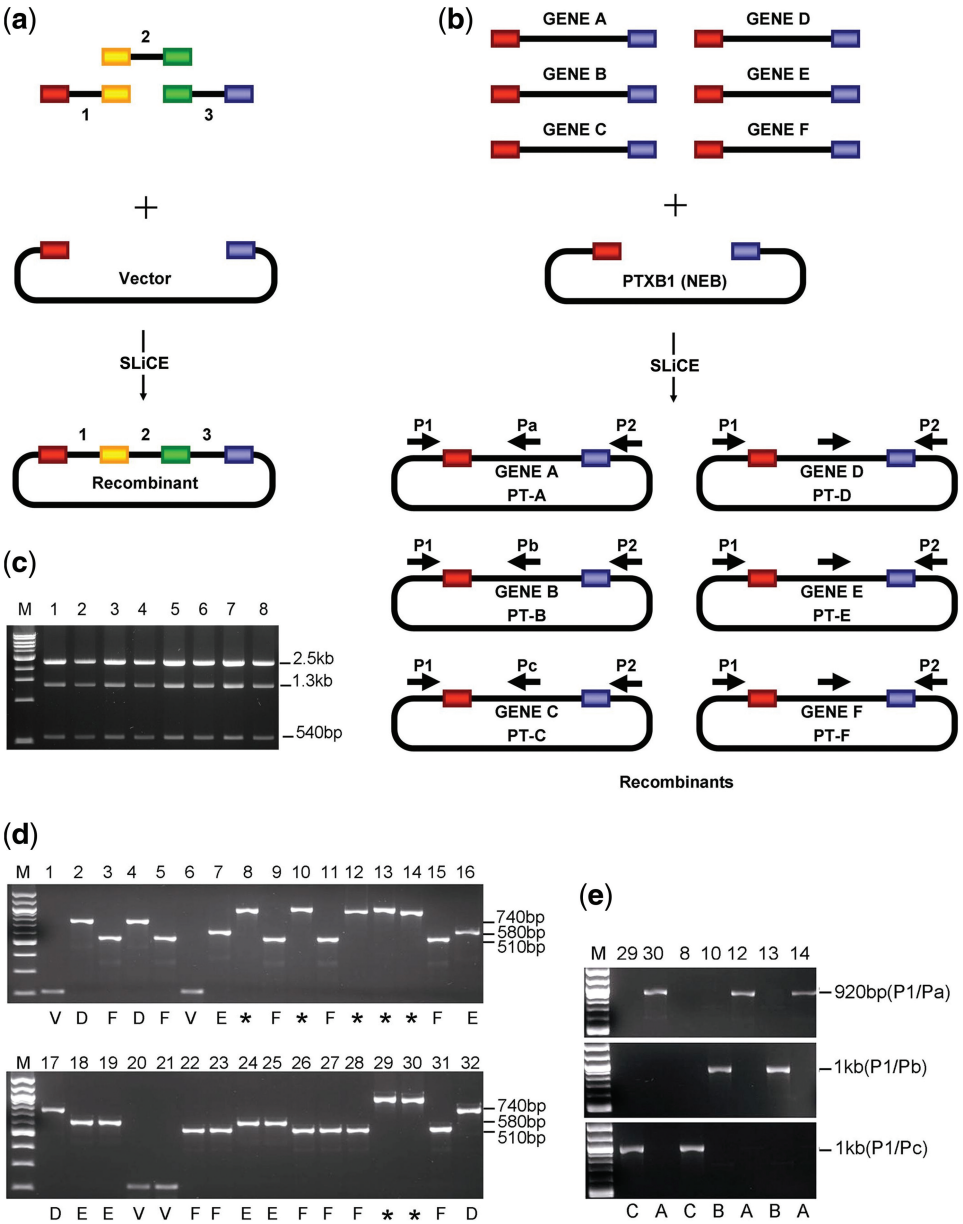


Figure 2. SLiCE cloning with multiple fragments. (**a**) Schematic illustrating multiple-way SLiCE cloning. A three-way cloning approach is shown. (**b**) Schematic illustrating SLiCE batch cloning. Six PCR inserts with 30-bp homology and plasmid vector PTXB1 (NdeI/SapI) were subjected to SLiCE batch cloning. P1, P2, Pa, Pb and Pc refer to the primers for colony PCR screening. (**c**) BsaAI/XmnI restriction analysis of the recombinants derived from seven-way SLiCE cloning. Plasmid DNAs from eight independent ampicillin-resistant colonies (lanes 1–8) were digested with BsaAI/XmnI. Recombinants contain one BsaAI site and one XmnI site located within the vector and one BsaAI site located in the insert yielding diagnostic 2.5-, 1.3-kb and 540-bp restriction fragments. (**d**) PCR screening of recombinants derived from SLiCE batch cloning of six different inserts. Thirty-two independent colonies (lanes 1–32) were subjected to PCR analysis with primer pair P1/P2. Recombinant plasmids PT-A, PT-B and PT-C all yielded PCR products of ~1 kb (lanes labeled with asteriscs). Recombinant plasmids PT-D, PT-E and PT-F yielded PCR products of 740, 586 and 510 bp, respectively (lanes labeled D, E and F). Non-recombinant vector pTXB1 yielded a PCR product of 210 bp (lanes labeled V). (**e**) To identify PT-A, PT-B and PT-C recombinants were further analyzed using primer pairs P1/Pa, P1/Pb and P1/Pc. The PCR products of PT-A using primer pair P1/Pa, PT-B using primer pair P1/Pb and PT-C using primer pair P1/Pc were 920 bp, 1 and 1 kb, respectively (lanes labeled A, B and C).

independent and seamless cloning method that we termed SLiCE. We found that simple cell extracts from two RecA deficient laboratory strains, JM109 and DH10B, were able to efficiently recombine vector and insert DNA containing short end homologies. In addition, we found that these strains can be further optimized for SLiCE by simple genetic modification. In this study we introduced the genes of the λ prophage Red recombination system and the *gam* gene into the DH10B genome to generate a new bacterial strain termed PPY. This strain currently provides the highest cloning efficiencies and facilitates SLiCE in a wide variety of cloning applications.

Bacterial extracts have been shown to efficiently catalyze RecA-dependent homologous recombination (15). However, SLiCE-mediated cloning is working most efficiently in RecA-deficient extracts indicating that it

Table 3. Multiple-way PPY SLICE cloning

Multiple-way SLiCE	Inserts length (bp)	Cloning efficiency	Cloning accuracy (%)
Three-way	520 and 1400	6610	88
Four-way	520, 762 and 760	4080	96
Seven-way	420, 400, 330, 220, 280 and 550	3260	90

Cloning efficiencies are given as CFUs of blue colonies per nanogram of vector. Cloning accuracies are given as the percentage of blue colonies among the total number of all amp^r colonies (blue and white). The 2.5-kb vector pBL was linearized by NotI/SalI digestion and the inserts were prepared by PCR. Experiments were performed using 10 ng/µl of vector and the corresponding amount of insert DNA at a 1 : 6 molar ratio of vector:insert. The blue colonies contain recombinant plasmid and the white colonies contain non-recombinant vector background.

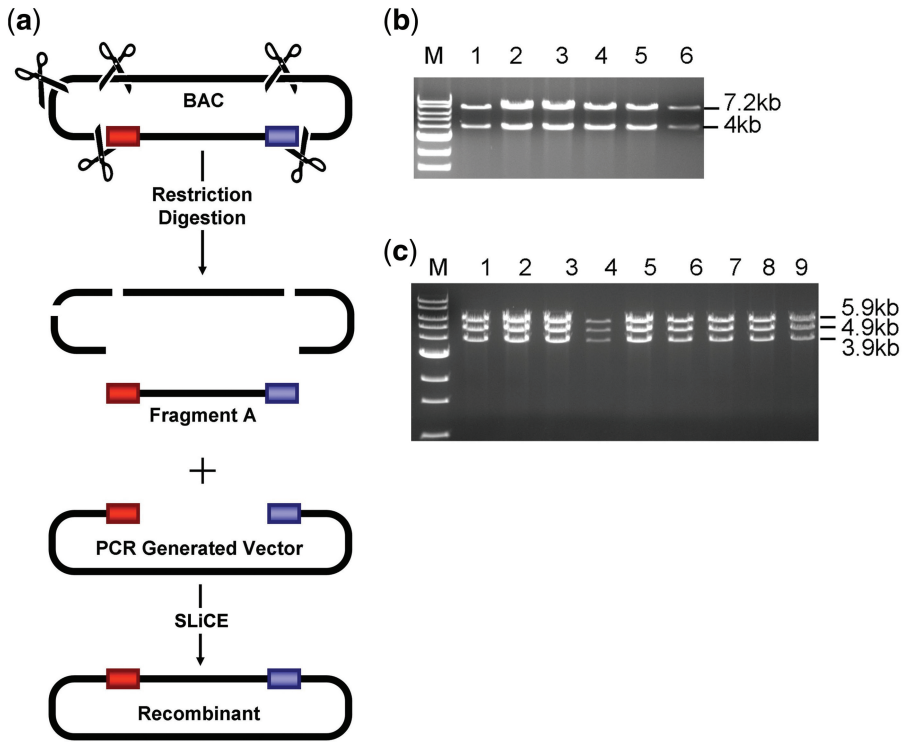


Figure 3. BAC SLiCE Cloning. (**a**) Schematic illustrating BAC SLiCE cloning. (**b**) XmnI/XhoI restriction analysis of recombinants derived from SLiCE cloning of an 8.7-kb BglII BAC fragment. Plasmid DNAs from six independent ampicillin-resistant colonies (lanes 1–6) were digested with XmnI and XhoI. Recombinants contain one XmnI site within the vector and one XhoI site within the insert yielding diagnostic 7.2- and 4-kb restriction fragments. (**c**) BamHI/BglII restriction analysis of recombinants derived from SLiCE cloning of a 12.2-kb EcoRV BAC fragment. Plasmid DNAs from nine independent ampicillin-resistant colonies (lanes 1–9) were digested with BamHI and BglII. Recombinants contain one BglII site within the vector and two BamHI sites in the insert yielding diagnostic 5.9-, 4.9- and 3.9-kb restriction fragments.

Table 4. BAC SLICE cloning

End homology length (bp)	Restriction enzyme	Insert length (kb)	Cloning efficiency	Cloning accuracy (%)
42	BglII	0.83	97	75
42	BglII	3.7	37	44
42	BglII	6.7	171	47
52	BglII	8.7	52	12
52	BglII	14	42	14
42	EcoRV	5.3	197	62
42	EcoRV	6.3	277	76
52	EcoRV	12.2	130	66
52	EcoRV	21	19	53

Cloning efficiencies are given as CFUs per nanogram of vector. Cloning accuracies are given as the percentage of correct clones among the total number of amp^r clones. pBluescript II KS+ (Stratagene) was used as template to PCR amplify linear vectors containing end homologies corresponding to various BglII or EcoRV restriction fragments in BAC clone RP23-303G13. Vector DNA (10–20 ng/µl) and total BglII or EcoRV digested BAC DNA (1 ug/µl) were subjected to PPY SLiCE cloning.

utilizes a different recombination pathway. The existence of a RecA-independent recombination pathway in *E. coli* that mediates exchange at short homologies has been proposed (10). It was suggested that this RecA-independent recombination mechanism involves the generation of short 5′ and 3′ tailed strands that anneal to homologous molecules (11) and that *E. coli* single-strand exonucleases such as RecJ, ExoVII, ExoI and ExoX could degrade these tails and abort the exchange reaction. Consistent with this notion, the RecA-independent recombination was stimulated in the absence of these single-strand exonucleases. However, the mechanism of RecA-independent recombination is unknown and it remains to be seen whether this recombination pathway is also responsible for SLiCE-mediated recombination. It is likely that SLiCE-mediated cloning involves the activities of exonucleases or helicases for the generation of single-strand tails at the ends of vectors and inserts, single-strand binding proteins for the stabilization of these single-strand overhangs and possibly other factors that protect 5′ and 3′ tails from degradation and facilitate their annealing. Consistent with this idea we found, that although SLiCE mediated recombination is efficient in the absence of the prophage Red/ET recombination systems, the introduction of the *red*α, *red*β and *gam* genes, which facilitate similar transactions at single stranded DNA ends, into DH10B bacteria greatly enhanced SLiCE mediated cloning.

Although SLiCE shares some features with other recently developed *in vivo* recombineering methods (16–19), it is different in several aspects. Recombineering methods provide useful tools for DNA modification and depend on homologous recombination that is mediated by the λ-prophage encoded Red recombination system *in vivo* in bacterial cells. In contrast, SLiCE is an *in vitro* recombination method facilitated by bacterial cell extracts. The Red recombination system is not required for SLiCE but can be used to further increase the cloning efficiencies of SLiCE. In addition, the main application of recombineering is the modification of large DNA molecules such as BACs or bacterial genomes, while SLiCE can be used as a general cloning method for the generation of recombinant plasmids.

Compared to conventional ligation dependent cloning methods including the cloning of DNA fragments with sticky or blunt ends generated by restriction digestion or the TA cloning of PCR fragments, SLiCE has several important advantages: (i) It is a time and labor saving method that consists of a one-hour/one-tube reaction followed by standard transformation of host bacteria. (ii) It does not require any prior treatment of end sequences. (iii) It can be used to directionally clone one or more fragments into any vector with high efficiency and fidelity. (iv) It promotes seamless cloning without leaving any unwanted sequences at the cloning junctions.

In addition to SLiCE, several other *in vitro* homologous recombination based cloning systems have recently been described, such as LIC-PCR and SLIC (20–23), Exonuclease III induced ligase-free directional subcloning of PCR products (24), Ribocloning (25), Enzyme-free cloning (26) and six commercially available cloning systems including In-Fusion™ PCR Cloning (27–29) (Clontech), Cold Fusion Cloning Kit (SBI), Fast Seamless Cloning Kit (Dogene), CloneEZ® Kit(Genescript), and GENEART® Seamless Cloning and Assembly Kit (invitrogen). SLIC uses the 3′–5′ exonuclease activity of T4 DNA polymerase to generate ssDNA overhangs in insert and vector which are required for the fusion of vector and insert fragments by single strand annealing with or without the addition of RecA. Exonuclease III induced ligase-free directional subcloning uses 3′–5′ exonuclease activity of Exonuclease III to generate ssDNA overhangs which facilitate cloning (24). Ribocloning uses Rnase A to cleave at single rC or rU bases that were introduced by PCR into vector and insert and subsequent heating to generate ssDNA overhangs for cloning. Enzyme-free cloning creates complementary ssDNA overhangs by PCR with tailed primer sets and post-PCR denaturation-hybridization reactions. In-Fusion™ PCR Cloning promotes PCR cloning by the In-Fusion enzyme, a poxvirus DNA polymerase with 3′–5′ exonuclease activity. The mechanisms or enzymatic activities involved in the other commercial cloning systems have not been disclosed by the suppliers, but it is likely that they utilize processes that are similar to that of SLIC and In-Fusion™ PCR Cloning.

Exonuclease III induced ligase-free directional subcloning is only useful for cloning fragments with blunt or 5′ protruding ssDNA ends but not compatible for cloning of fragments with 3′ protruding ssDNA ends. Furthermore, the method probably depends partially on the helical structure of the DNA fragments and displays sequence dependence (C > A = T > G) (24). Ribocloning and Enzyme-free cloning require PCR amplification to generate both vector and insert and cannot facilitate the cloning of DNA fragments generated by restriction digestion. In addition, Ribocloning requires special PCR primers containing ribonucleotides (rC or rU) at the 3′-ends and enzyme-free cloning requires four pairs of primers for one reaction and is not suitable for multiple-way cloning (25).

In comparison to SLiCE, SLIC requires optimization. Vector and insert fragments for SLIC need to be treated with T4 DNA polymerase and the treatment duration is not always constant but depends on the homology length (23).

The efficient seamless cloning activity is one of the most important features of SLiCE as it allows the recombining of vector and inserts *in vitro* even in the presence of flanking heterologous sequences of up to 998 bp on one side or 319 bp plus 738 bp on both sides. This property decreases the sequence dependence of end cloning by SLiCE and greatly extends its usefulness for many applications such as replacing unwanted sequences adjacent to the cloning sites without any prior treatment even in the absence of appropriate restriction sites. For example, in a single SLiCE reaction, we subcloned the open reading frame of the Pms2 gene into the mammalian expression vector p3XFLAG-CMV-7.1 (Invitrogen) and deleted 72 bp of an unwanted sequence tag flanking the cloning site within the vector, which could not be deleted by restriction digestion. SLIC also has seamless cloning

activity; however, it is limited to flanking heterologies of only up to 20 bp. There are no reports that In-Fusion™ PCR Cloning (Clontech) or any of the other *in vitro* cloning systems have such activity. Furthermore, SLiCE is the only known *in vitro* recombination based method for the directional subcloning of genomic BAC fragments into plasmid vectors. At present it is not clear if the other *in vitro* cloning systems can be used for this application.

In summary, SLiCE is an easy, efficient and inexpensive cloning method that allows the generation of recombinant plasmid vectors in a seamless and precise fashion. It requires the generation of simple bacterial cell extracts from readily available lab strains and does not require the use of restriction enzymes or DNA end modification enzymes such as Klenow or T4 DNA polymerase. In addition, the joining of vector and insert fragments by DNA ligase is not required. SLiCE is also a highly versatile method, and its capabilities can be expanded by the generation of additional optimized bacterial strains in the future.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–4.

ACKNOWLEDGEMENTS

The PPY strain can be obtained upon request (yongwei.zhang@einstein.yu.edu).

FUNDING

Funding for open access charge: National Institute of Health (1R01CA76329, 1R01CA93484, to W.E.).

Conflict of interest statement. None declared.

REFERENCES

1. Smith,H.O. and Wilcox,K.W. (1970) A restriction enzyme from Hemophilus influenzae. I. Purification and general properties. *J. Mol. Biol.*, **51**, 379–391.
2. Danna,K. and Nathans,D. (1971) Specific cleavage of simian virus 40 DNA by restriction endonuclease of Hemophilus influenzae. *Proc. Natl Acad. Sci. USA*, **68**, 2913–2917.
3. Cohen,S.N., Chang,A.C., Boyer,H.W. and Helling,R.B. (1973) Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl Acad. Sci. USA*, **70**, 3240–3244.
4. Little,J.W. (1967) An exonuclease induced by bacteriophage lambda. II. Nature of the enzymatic reaction. *J. Biol. Chem.*, **242**, 679–686.
5. Radding,C.M. and Carter,D.M. (1971) The role of exonuclease and beta protein of phage lambda in genetic recombination. 3. Binding to deoxyribonucleic acid. *J. Biol. Chem.*, **246**, 2513–2518.
6. Carter,D.M. and Radding,C.M. (1971) The role of exonuclease and beta protein of phage lambda in genetic recombination. II. Substrate specificity and the mode of action of lambda exonuclease. *J. Biol. Chem.*, **246**, 2502–2512.
7. Fehér,T., Karcagi,I., Gyorfy,Z., Umenhoffer,K., Csörgo,B. and Pósfai,G. (2008) Scarless. Engineering of the Escherichia coli genome. *Methods Mol. Biol.*, **416**, 251–259.
8. Guzman,L.M., Belin,D., Carson,M.J. and Beckwith,J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.*, **177**, 4121–4130.
9. Lee,E.C., Yu,D., Martinez de Velasco,J., Tessarollo,L., Swing,D.A., Court,D.L., Jenkins,N.A. and Copeland,N.G. (2001) A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics*, **73**, 56–65.
10. Lovett,S.T., Hurley,R.L., Sutera,V.A. Jr, Aubuchon,R.H. and Lebedeva,M.A. (2002) Crossing over between regions of limited homology in *Escherichia coli*. RecA-dependent and RecA-independent pathways. *Genetics*, **160**, 851–859.
11. Dutra,B.E., Sutera,V.A. and Lovett,S.T. (2007) RecA-independent recombination is efficient but limited by exonucleases. *Proc. Natl Acad. Sci. USA*, **104**, 216–221.
12. Muyrers,J.P., Zhang,Y., Buchholz,F. and Stewart,A.F. (2000) RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Dev.*, **14**, 1971–1982.
13. Kuzminov,A. (2002) Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. *Microbiol. Mol. Biol. Rev.*, **63**, 751–813.
14. Persky,N.S. and Lovett,S.T. (2008) Mechanisms of recombination: lessons from *E. coli*. *Crit. Rev. Biochem. Mol. Biol.*, **43**, 347–370.
15. Kolodner,R. (1980) Genetic recombination of bacterial plasmid DNA: electron microscopic analysis of in vitro intramolecular recombination. *Proc. Natl Acad. Sci. USA*, **77**, 4847–4851.
16. Murphy,K.C. (1998) Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*. *J. Bacteriol.*, **180**, 2063–2071.
17. Zhang,Y., Buchholz,F., Muyrers,J.P. and Stewart,A.F. (1998) A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat. Genet.*, **20**, 123–128.
18. Yu,D., Ellis,H.M., Lee,E.C., Jenkins,N.A., Copeland,N.G. and Court,D.L. (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **97**, 5978–5983.
19. Muyrers,J.P., Zhang,Y., Testa,G. and Stewart,A.F. (1999) Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.*, **27**, 1555–1557.
20. Aslanidis,C. and de Jone,P.J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.*, **18**, 6069–6074.
21. Yang,Y.S., Watson,W.J., Tucker,P.W. and Capra,J.D. (1993) Construction of recombinant DNA by exonuclease recession. *Nucleic Acids Res.*, **21**, 1889–1893.
22. Li,M.Z. and Elledge,S.J. (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods*, **4**, 251–256.
23. Gibson,D.G., Benders,G.A., Andrews-Pfannkoch,C., Denisova,E.A., Baden-Tillson,H., Zaveri,J., Stockwell,T.B., Brownley,A., Thomas,D.W., Algire,M.A. *et al.* (2008) Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma genitalium Genome. *Science*, **319**, 1215.
24. Hsiao,K. (1993) Exonuclease III induced ligase-free directional subcloning of PCR products. *Nucleic Acids Res.*, **21**, 5528–5529.
25. Barnes,W.M. (2003) Ribocloning: DNA cloning and gene construction using PCR primers terminated with a ribonucleotide. In: Carl,W.Dieffenbach (ed.), *PCR Primer*, Vol. 2. Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 441–449.
26. Tillett,D. and Neilan,B.A. (1999) Enzyme-free cloning: a rapid method to clone PCR products independent of vector restriction enzyme sites. *Nucleic Acids Res.*, **27**, e26.
27. Clontechniques. (2002) BD In-Fusion cloning kit. Precise, directional cloning of PCR products without restriction enzymes. *CLONTECHniques*, **XVII**, 10–11.
28. Hamilton,M.D., Nuara,A.A., Gammon,D.B., Buller,R.M. and Evans,D.H. (2007) Duplex strand joining reactions catalyzed by vaccinia virus DNA polymerase. *Nucleic Acids Res.*, **35**, 143–151.
29. Zhu,B., Cai,G., Hall,E.O. and Freeman,G.J. (2007) In-Fusion™ assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations. *Biotechniques*, **43**, 354–359.

OPEN ACCESS Freely available online

PLOS ONE

A Practical Comparison of Ligation-Independent Cloning Techniques

Julian Stevenson[1*☯], James R. Krycer[1,2*☯], Lisa Phan[1], Andrew J. Brown[1*]

1 School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, New South Wales, Australia, **2** Diabetes and Obesity Program, Garvan Institute of Medical Research, Sydney, New South Wales, Australia

Abstract

The precise assembly of specific DNA sequences is a critical technique in molecular biology. Traditional cloning techniques use restriction enzymes and ligation of DNA *in vitro*, which can be hampered by a lack of appropriate restriction-sites and inefficient enzymatic steps. A number of ligation-independent cloning techniques have been developed, including polymerase incomplete primer extension (PIPE) cloning, sequence and ligation-independent cloning (SLIC), and overlap extension cloning (OEC). These strategies rely on the generation of complementary overhangs by DNA polymerase, without requiring specific restriction sites or ligation, and achieve high efficiencies in a fraction of the time at low cost. Here, we outline and optimise these techniques and identify important factors to guide cloning project design, including avoiding PCR artefacts such as primer-dimers and vector plasmid background. Experiments made use of a common reporter vector and a set of modular primers to clone DNA fragments of increasing size. Overall, PIPE achieved cloning efficiencies of ~95% with few manipulations, whereas SLIC provided a much higher number of transformants, but required additional steps. Our data suggest that for small inserts (<1.5 kb), OEC is a good option, requiring only two new primers, but performs poorly for larger inserts. These ligation-independent cloning approaches constitute an essential part of the researcher's molecular-tool kit.

Citation: Stevenson J, Krycer JR, Phan L, Brown AJ (2013) A Practical Comparison of Ligation-Independent Cloning Techniques. PLoS ONE 8(12): e83888. doi:10.1371/journal.pone.0083888

Editor: Jean Peccoud, Virginia Tech, United States of America

Received October 1, 2013; **Accepted** November 8, 2013; **Published** December 23, 2013

Funding: This work was supported by the National Health and Medical Research Council (1008081, www.nhmrc.gov.au/); and the National Heart Foundation of Australia (G11S5757, http://www.heartfoundation.org.au). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: aj.brown@unsw.edu.au (AJB); j.stevenson@unsw.edu.au (JS); j.krycer@garvan.org.au (JRK)

☯ These authors contributed equally to this work.

Introduction

The precise assembly of specific DNA sequences is a critical technique in molecular biology. Traditional cloning makes use of restriction enzymes and ligation of DNA *in vitro*. Restriction endonuclease digestion and ligation increase the complexity of cloning projects, for example by requiring selection of appropriate restriction-sites and inefficient ligation steps. Consequently, several ligation-independent cloning (LIC) methods have since been developed that are simpler, faster, and highly efficient. These strategies rely on the generation of DNA fragments with single-stranded complementary ends to allow directional cloning of any insert, independent of restriction enzymes and *in vitro* ligation. The most effective and convenient methods include polymerase incomplete primer extension (PIPE) cloning [1], sequence and ligation-independent cloning (SLIC) [2], and overlap extension cloning (OEC) [3,4] (Figure 1). In this study, we will compare these cloning strategies.

All three techniques amplify the gene of interest by polymerase chain reaction (PCR). The 3′ ends of the PCR primers are template-specific, whilst the 5′ ends incorporate tails specific for the cloning junction.

PIPE relies on the observation that a significant portion of PCR products are incomplete, having 3′-recessed ends [1], particularly in the absence of a final extension step. SLIC uses a brief treatment of purified PCR products with the 3′→5′ exonuclease activity of T4 DNA polymerase to generate a higher proportion of recessed ends [2,5]. Cloning is possible due to the complementary 5′ ends of the insert and vector fragments. These are analogous to the much shorter 'sticky ends' generated by restriction enzymes. After mixing, these single-stranded overhangs anneal and can be efficiently ligated *in vivo* after transformation by the bacterial DNA recombination and repair machinery.

In contrast, OEC has two rounds of amplification. Firstly, the insert is PCR-amplified using primers with 5′ ends complementary to the target site in a circular destination vector. The 3′ ends of this product ('megaprimer') can consequently anneal and amplify the destination vector by overlap extension. This yields a nicked plasmid that is repaired after transformation. A variant of the OEC strategy is utilised by the commonly used QuikChange site-directed mutagenesis kit (Agilent Technologies).

In an age of high-throughput molecular biology, there is a need to move away from traditional cloning methods. We believe that LIC methods offer a more robust, efficient means for DNA cloning. Furthermore, variations between strategies suggest that certain techniques may be better suited for particular situations. Consequently, here we characterise and compare the efficiency, convenience, and utility of three major LIC techniques.

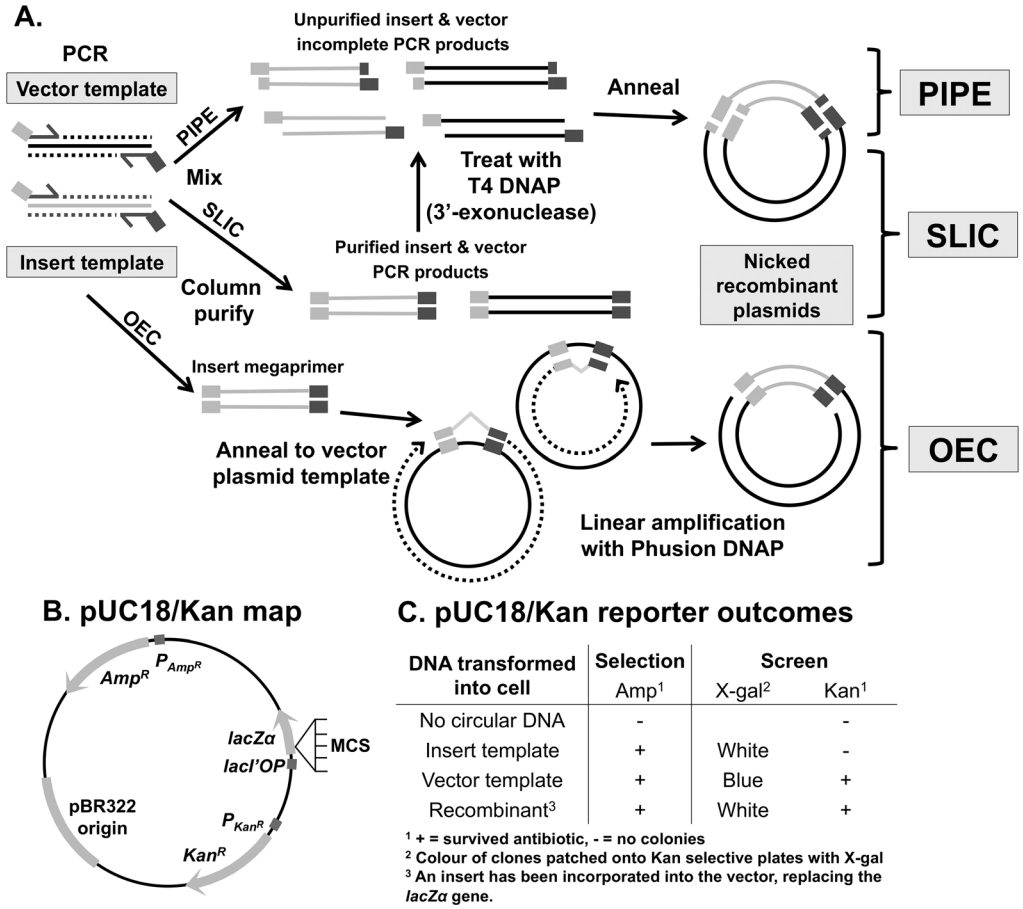


Figure 1. Principles of polymerase incomplete primer extension (PIPE) cloning, sequence and ligation-independent cloning (SLIC) and overlap extension cloning (OEC). (A) In PIPE, incomplete extension during PCR generates 3′-recessed ends. In SLIC, purified PCR products are treated with T4 DNA polymerase (DNAP) so that the exonuclease activity will increase the proportion of recessed ends. In both these techniques, by amplifying vector and insert with primers containing complementary 5′-tails and mixing the products, the overhangs can anneal and are joined *in vivo* after transformation into *E. coli*. In OEC, the insert PCR product acts a megaprimer to generate a nicked plasmid by overlap extension *in vitro* in a second round of amplification. The nicks are also repaired *in vivo*. For all techniques, a DpnI digestion step is included to remove plasmid template (not shown). (B) Design of the reporter vector, encoding resistance for ampicillin and kanamycin, and the alpha-fragment of beta-galactosidase. (C) Colonies from ampicillin plates were patched onto kanamycin/X-gal plates to distinguish recombinants from unwanted insert vector or empty pUC18/Kan background colonies.
doi:10.1371/journal.pone.0083888.g001

Materials and Methods

Primer combinations

Primer sequences are listed in Table S1 in File S1. Gene accession numbers and plasmid vector backbones are listed in Table S2 in File S1. The pUC18/Kan reporter vector (Figure 1B) was prepared using PIPE cloning. The vector backbone of pUC18 was obtained by PCR with primers pUC18–F and pUC18–R, and the kanamycin resistance cassette from pEGFP-C1 (Clontech) with primers KanR-in-pUC-F and KanR-in-pUC-R. The complete pUC18/Kan vector DNA sequence is included as a text file in Supporting Information (File S2).

A 24 bp FLAG epitope insertion (84 bp fragment) was achieved through PIPE or SLIC with primers FLAG-pUC18-F and FLAG-pUC18-R to amplify pUC18/Kan. The 85 bp FLAG megaprimer for OEC was prepared through 40 uL thermal cycling of the primers alone, with 4 μM FLAG-pUC18-overlap-F and FLAG-pUC18-overlap-R with 1× HF Buffer, 0.8 U of Phusion Hot Start II High-Fidelity DNA Polymerase (New England Biolabs) and cycling conditions: (98°C 30 s, 63°C 30 s, 72°C 1 min) ×5.

The pUC18/Kan vector backbone was amplified with primers pUC18-L30-F and pUC18-L30-R, which are the reverse complement of the common 5′-tails of the insert primers. A 350 bp fragment (Gluc) was amplified from FLAG-hLXRβ-hGluc(1) [6] with primers Gluc-pUC18-F and BGH-pUC18-R. LXRβ (1.4 kb) was similar amplified with primers LXRb-pUC18-F and LXRb-pUC18-R. A 2.2 kb sequence (AR) from pTK-AR-V5 [6] was obtained with primers AR-pUC18-F and AR-pUC18-R. SRC-1 (4.3 kb) was amplified with primers SRC1-pUC18-F and SRC1-pUC18-R from pCR3.1-SRC1 [7]. 1.4 kb Insig-1 and 4.3 kb SCAP fragments were obtained with primers T7-pUC18-F and BGH-pUC18-R from pCMV-Insig-1-Myc or pCMV-SCAP respectively [8].

PCR

PCR was performed in 50 μL reactions using 0.5 ng of template, 0.5 μM forward and reverse primers, 5% DMSO, HF Buffer and 1 U of the non-strand displacing enzyme Phusion Hot Start II High Fidelity DNA polymerase, and cycling conditions:

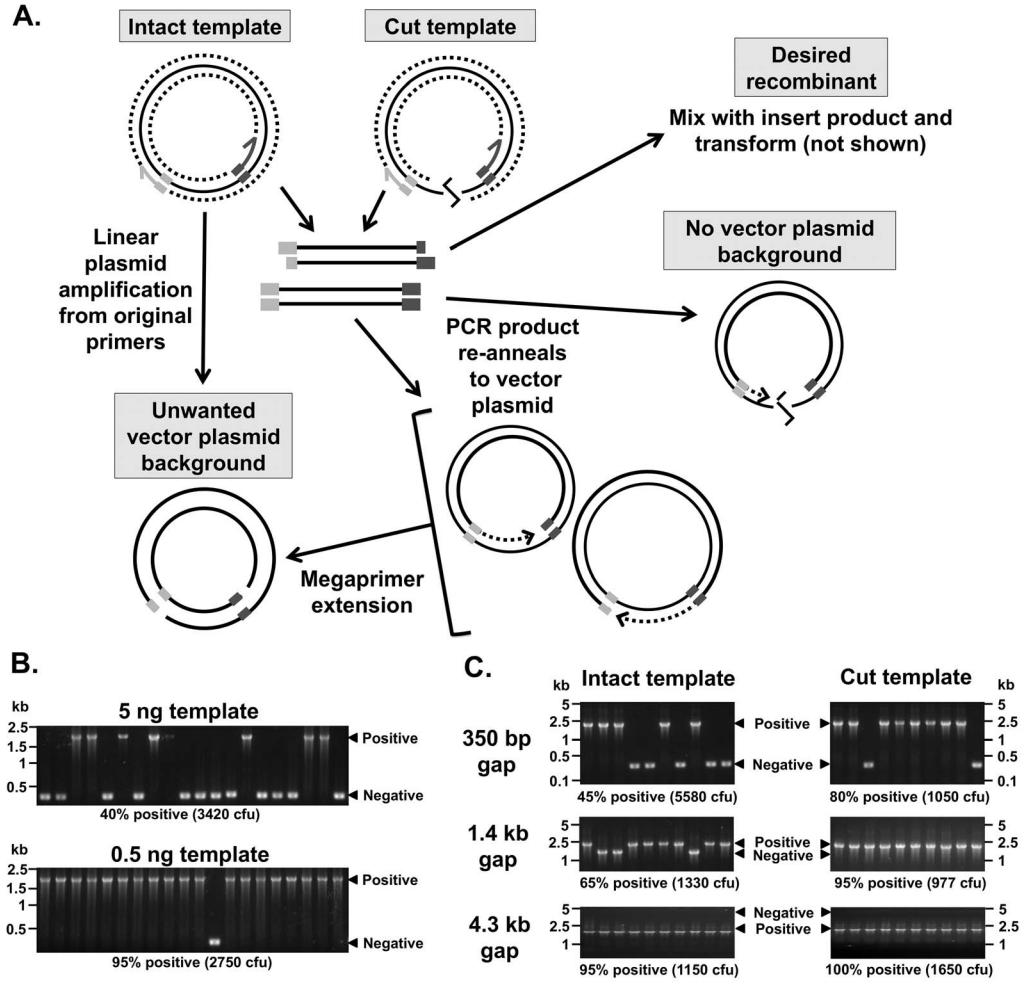


Figure 2. Generation of nicked vector plasmid can reduce PIPE cloning efficiency. Unwanted copies of the original vector plasmid template can be generated by overlap extension in the process of obtaining vector-backbone PCR product (A). This can be prevented by reducing the template concentration (B), cutting the plasmid template first or cloning into a vector with an existing large insert (C). See main text for details.
doi:10.1371/journal.pone.0083888.g002

98°C 3 min, (98°C 30 s, 63°C 30 s, 72°C 3 min for products >1.5 kb or 1 min for <1.5 kb) ×30.

Reaction products were column purified using a QIAquick PCR purification or gel extraction kit (Qiagen) and quantified by spectroscopy.

PIPE cloning

0.025 pmol of vector and 0.0625 pmol insert purified products (2.5:1 insert:vector ratio) were digested for 3 hr at 37°C with 10 U of DpnI in 10 µL CutSmart Buffer (NEBuffer 4.1/0.1 mg/mL BSA), diluted 1:1 with 1× HF buffer (to control for buffer composition between purified and unpurified products, as 1× HF buffer halved transformation efficiency), and 2 µL used to transform 18 µL of XL10 Gold ultracompetent cells (Agilent Technologies) according to manufacturer's instructions.

SLIC

DpnI digested purified PIPE products were incubated at 25°C for 5 min with 0.75 U of T4 DNA polymerase, immediately placed on ice for 10 min, diluted 1:1 with ice-cold 1× HF Buffer, and 2 µL used for transformation as described above.

OEC

20 µL OEC reactions were performed using 10–250 fmol of insert product −84 bp: 250 fmol (12.5 nM); 350 bp: 100 fmol (5 nM); 1.4 kb: 25 fmol (1.25 nM); 4.3 kb: 10 fmol (0.5 nM) - as megaprimer and 25 ng of pUC18/Kan vector template with 5% DMSO, 0.4 U of Phusion Hot Start II High Fidelity DNA polymerase, and cycling conditions: 72°C 5 min (to blunt the megaprimer), 98°C 3 min, (98°C 30 s 63°C 30 s 72°C 3 min) ×30. 5 µL fractions were diluted 1:1 with CutSmart Buffer (to control for buffer and assist DpnI activity), digested for 3 hr at 37°C with 20 U of DpnI, and 2 µL used for transformation as described above.

Colony counting and screening

Serial dilutions of transformation mixture were spread onto ampicillin selective plates to allow counting of the number of colony forming units (CFU), adjusted to account for total volume, and rounded to 3 significant figures. 40 colonies were picked and streaked onto kanamycin selective plates in the presence of X-gal/IPTG to identify recombinants. For one replicate experiment of each set, colonies were also tested by colony PCR across the cloning junctions or sequenced to validate the blue/white screening.

Step by step protocols for each technique are outlined in the supporting information (File S3).

Results

Design of the reporter system

To efficiently compare the different cloning techniques, we prepared a reporter vector plasmid, pUC18/Kan (Figure 1B). It contains resistance cassettes for ampicillin and kanamycin, as well as a polylinker encoding the *lacZ*α fragment. Each pair of insert-specific primers had the same 30 bp 5′-tail – this modular design allowed replacement of the multiple-cloning site of pUC18/Kan with a collection of inserts. This allows simple identification of the type of plasmid in each colony – insert plasmid, vector plasmid or desired recombinant plasmid – as the insert plasmids lack kanamycin resistance, and the vector plasmid template contains undisrupted *lacZ*α. Hence, bacteria were first plated onto ampicillin plates and then patched onto kanamycin selective plates in the presence of IPTG for blue/white colony screening. Thus, positive colonies are white and resistant to both ampicillin and kanamycin (Figure 2C), whilst blue colonies possess pUC18/Kan and Kan-sensitive colonies possess insert-containing plasmids.

Generation of nicked vector plasmid reduces PIPE cloning efficiency

We began our optimisation with PIPE. After first transforming the DpnI-digested vector PCR product alone we observed a significant number of blue (empty pUC18/Kan) colonies, but no colonies without the PCR (data not shown). Given that DpnI only digests methylated DNA, this indicates that vector plasmid is being generated during the PCR. This can occur because extension of the vector primers all the way around the plasmid template can regenerate the empty vector plasmid (Figure 2A). Note that this is linear amplification, generating less product in contrast to the desired PCR product between the primers, which is amplified exponentially by PCR. However, the vector PCR products can also re-anneal to the vector and extend analogously to the megaprimers in OEC, although this appeared to be less important, as addition of three 5′ non-template thymidines to the vector primers did not affect cloning efficiency (data not shown). Nevertheless, these situations can generate nicked copies of the vector *in vitro* that will escape DpnI digestion (Figure 2A). This leads to unwanted vector background colonies.

To confirm this mechanism, we hypothesised that nicked vector background should be reduced by 1) increasing the gap between the 5′ ends of the vector primers, or 2) cutting the template in between the vector primers (Figure 2). To test this, we replaced 350 bp, 1.4 kb, or 4.3 kb inserts in pUC18/Kan (i.e. different gap sizes) with the same 2.2 kb insert. The insert primer 5′ vector overhangs were complementary to the vector backbone, independent of the pre-existing insert sequences. This allows us to determine PIPE cloning efficiency – the proportion of successful recombinants - based on insert size differences using colony PCR. This was performed with or without linearising the vector using PstI. We found that both increasing the gap and linearisation reduced background (Fig 2C). However, because these strategies restrict the utility of PIPE-cloning, we sought a more flexible solution. We reasoned that reducing the template should dilute out the background – since the complete vector is linearly amplified, it is more dependent on template concentration and thus a lower template concentration would favour PCR amplification of the desired product. Accordingly, PIPE cloning efficiency increased from approximately 40% to greater than 95% when using ten-fold less template (Figure 2B), with similar results for insertion of the 350 bp, 1.4 kb and 4.3 kb inserts into pUC18/Kan (Table S3 in File S1). We adopted this latter strategy in subsequent experiments.

Optimisation of SLIC, PIPE and OEC

PIPE relies on incomplete extension in PCR. Although the proportion of incomplete products is sufficient at 25 cycles [1], we reasoned that additional cycles might deplete PCR components or generate DNA products that inhibit extension, leading to a greater proportion of recessed ends and thus more clones. However, PIPE cloning efficiency did not increase using equal amounts of DNA product taken from 25 to 40 PCR cycles (Table S4 & S5 in File S1). We maintained subsequent reactions at 35 cycles to ensure that good product yields are achieved, particularly for difficult templates or inefficient primers. PIPE cloning tolerated a wide range of insert:vector molar ratios for a range of insert sizes (Table S6 in File S1), with an optimum of 2.5:1, similar to previously reported values for SLIC [2,5].

We made use of a one-tube version of SLIC [5,9]. For a convenient volume of T4 polymerase (0.75 U/0.25 μL), the highest efficiency was observed after 5–10 min treatment at 25°C, although 5 min was most robust (Table S7 in File S1), followed by immediate incubation on ice to halt the reaction.

OEC uses linear amplification, generating less product than the exponential amplification in PCR. We consequently used 25 ng of template for OEC, 50-fold higher than for the PCR-based PIPE and SLIC. This required doubling the concentration of DpnI to account for the increased template. The effectiveness of OEC was dependent upon megaprimer concentration, with the optimum being inversely proportional to megaprimer size: high concentrations of small insert and vice versa (Table S8 in File S1). Thus concentrations of ~25–50 fmol (1.25–2.5 nM) were optimal for products >1.5 kb, ~50–100 fmol (2.5–5 nM) <1.5 kb, and ~100–300 fmol (5–15 nM) <350 bp. This is likely due to the increasing propensity of larger products to anneal to themselves rather than to the plasmid template. We observed more kanamycin-sensitive colonies when the megaprimer was originally prepared from insert PCRs with higher template (5 ng). This is likely to be insert-vector background from the insert PCR, and was removed by reducing the template amount to 0.5 ng (Table S9 in File S1). Increasing the number of cycles of overlap extension gave progressively more colonies, without clearly increasing cloning efficiency (Table S10 in File S1). Hence 30 thermal cycles were used for comparisons to the other techniques.

OEC does not tolerate the presence of primer-dimers

To test the ability of the techniques to tolerate the presence of primer-dimer, we attempted to clone a 1.4 kb gene product where the PCR product also included a marked amount of ~150 bp primer-dimer or mispriming product. Since these products contain 5′ ends complementary to the vector, they may be incorporated into recombinant clones, also yielding white colonies in our screening assay. Primer-dimer had little effect on PIPE or SLIC, as screened white colonies contained the correct insert, confirmed by colony PCR and sequencing (data not shown). In contrast, for OEC, most white colonies contained unwanted primer-dimer instead (Figure 3). Prior gel extraction of the insert megaprimer and using this in the overlap extension step ensured that nearly all white colonies contained the desired insert instead (Figure 3). An alternative method to remove primer-dimer before OEC was to pretreat the purified insert mixture with T4 DNA polymerase in the absence of dNTPs (Figure S1 in File S1). This allowed digestion of the primer-dimer, leaving the larger desired product to

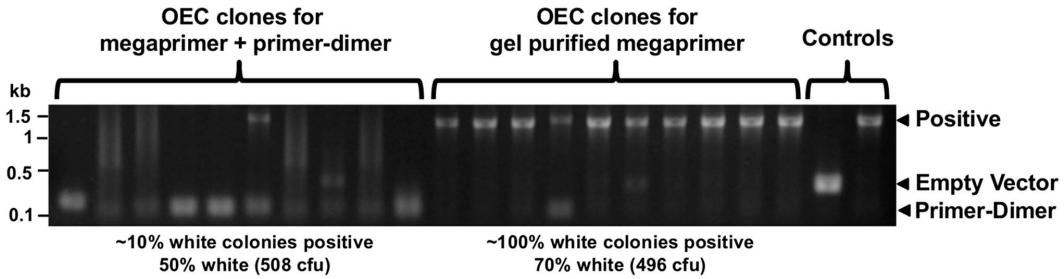


Figure 3. OEC does not tolerate primer-dimers. Megaprimer and primer-dimer contaminant or 1.4 kb LXR megaprimer alone were gel purified and used for OEC. Ten white colonies for each were screened by PCR across the cloning junctions.
doi:10.1371/journal.pone.0083888.g003

be end-filled - during the 5 min, 72°C step at the start of the overlap extension reaction - and cloned.

Direct comparison

To directly compare the effectiveness of the different cloning strategies, the reporter vector and different inserts were first amplified and purified, then fed into PIPE, SLIC and OEC with equivalent amounts of DNA using the optimised protocols (as described in the Materials and Methods). Reaction products were diluted with either HF buffer or NEBuffer to control for transformation buffer composition, as the Phusion 1× HF PCR buffer halved transformation efficiency (data not shown). The amount of DNA (25 fmol vector and 62.5 fmol insert) used for transformation was equivalent to that of unpurified PIPE cloning from a single pair of PCRs. 24 bp, 350 bp, 1.4 kb and 4.3 kb fragments were chosen to gauge the effect of increasing size on cloning efficiency for each strategy.

The 24 bp insertion is a special case due to its small size, in that PIPE and SLIC used a single primer pair of reverse design for PCR – with 3′ ends complementary to the vector and 5′-tails containing the entire insert – thus requiring only one PCR, such that the product anneals to itself to form a circular product. For very small inserts, using two primer sets to amplify vector and insert would be wasteful and very inefficient for PIPE and SLIC. The standard insert primer design was used for OEC, with annealing and 3′ end-filling of partially overlapping primers alone to first prepare the insert megaprimer.

PIPE consistently performed well, yielding hundreds of colonies with close to 100% cloning efficiency (Tables 1, S6 and S10). Addition of T4 DNA polymerase exonuclease treatment for SLIC increased the number of transformants by ~4–100 fold for all inserts, retaining the high efficiency.

OEC yielded very high numbers of transformants for the 24 bp (84 bp megaprimer) insertion, with close to 100% efficiency. Moderate to high efficiencies were also observed for the 350 bp insertion. Colony number fell proportionally as insert size increased, which was associated with a corresponding decline in cloning efficiency (Table 1). Cloning efficiencies and colony numbers were far more variable for larger fragments, reflecting lower robustness of OEC compared to the other LIC techniques (Table S8–S11 in File S1).

Table 1. Direct comparison of PIPE, SLIC and OEC for various insert sizes.

Insert Size[a]	Technique	Cloning Efficiency	Colonies	Fold Increase[b]
85 bp	PIPE	100%	1560	1
	SLIC	100%	15800	10
	OEC	95%	15100	10
350 bp	PIPE	98%	176	1
	SLIC	98%	18500	105
	OEC	90%	9200	52
1.4 kb	PIPE	95%	705	1
	SLIC	100%	2760	4
	OEC	73%	1450	2
4.3 kb	PIPE	100%	309	1
	SLIC	100%	6420	21
	OEC	45%	251	1

See main text for details.
[a]Single representative experiments are shown for FLAG (85 bp), Gluc (350 bp), Insig-1 (1.4 kb), and SCAP (4.3 kb).
[b]The increase in colonies relative to PIPE is shown.
doi:10.1371/journal.pone.0083888.t001

Discussion

Ligation-independent cloning techniques can be used to introduce DNA fragments into cloning vector plasmids quickly, cheaply, and with high efficiency. Anything that can be amplified by PCR can be introduced into any position of any vector of choice in a single cloning step without unwanted additional nucleotides, so called 'scarless cloning'. We observed that PIPE worked very well with limited manipulations, as long as the template concentration was kept to a minimum to avoid overlap extension vector background. SLIC consistently achieved the highest efficiencies and number of transformants, but required additional resources. OEC worked well for smaller fragments, but was less effective for larger fragments. It was also very vulnerable to the presence of primer-dimers.

The methods compared here work either through the generation of complementary single-stranded overhangs for *in vivo* homologous recombination (PIPE, SLIC) or by generating a nicked plasmid *in vitro* by overlap extension (OEC). This can require the design of new primers with longer complementary tail sequences than used for restriction cloning, but the time savings and increased robustness outweigh these nominal costs, even for routine applications. Modular primers can also be designed to allow parallel cloning of a gene into different vectors with identical linkers [10], as well as inserting different genes with the same 5′-tails into the same vector, as in this study.

The identical insert primer design also allowed the use of the same PCR product with the three different techniques. This can enable rescue of a failed cloning attempt using a different strategy

with a higher efficiency but requiring additional steps and resources (Figure 4 and Table 2). The first method of choice is dependent upon the nature of the insert, the availability of existing primers and the level of efficiency required. For genes of up to 1.5 kb, OEC is a good choice, as useful efficiencies (50%+) can be achieved with one primer pair. This will often be suitable for insertion of affinity tags or fluorescent proteins, such as the green fluorescent protein. However, cloning efficiency and robustness for OEC decrease with increasing insert size, introducing the risk of failure for larger genes. Overlap extension also requires relatively long (~30 bp) target-vector-tail sequences in the primers to allow stable annealing, whereas PIPE and SLIC can work well with overhangs of ~15 bp, allowing synthesis of shorter primers [1,2,5]. Purification of the insert PCR is recommended for OEC, but for smaller inserts, a wide range of megaprimer concentrations are tolerated, safely allowing use of a small fraction of unpurified PCR product as megaprimer (Table S12 in File S1 and data not shown). PIPE cloning is generally simpler and robustly performs well, at the cost of requiring a second primer pair.

For cloning projects where vector primers are already available or higher efficiencies are desired to clone inserts greater than 1.5 kb, PIPE is usually the best choice, achieving high efficiencies for a wide range of gene sizes [1] (Table 1 and Table S11 in File S1). Unpurified PIPE PCR products can be used for transformation directly after DpnI treatment, making it extremely fast, especially useful for high-throughput cloning projects [1]. This simple approach is the mainstay LIC technique in our laboratory and our default choice for any fragment size (See protocols in File S3). However, purification to remove inhibitory buffer components and concentrate the DNA provides better results (Table S12 in File S1 and data not shown). If PIPE cloning fails or a greater number of transformants are required, such as for library construction, then SLIC is superior, typically increasing the number of transformants by more than 5-fold in one step after purification, DpnI and T4 exonuclease treatment. SLIC is also effective after combining purified fragments without quantification (Table S12 in File S1). SLIC can also be performed using restriction enzyme-cleaved plasmid, minimising primer design at the cost of additional manipulations over PIPE [2,5,9].

SLIC has the advantage that it can be used to assemble multiple insert fragments in a single cloning step, particularly with addition of recombinant RecA to enhance annealing *in vitro*, albeit with reduced efficiency compared to single-fragment cloning [2]. This can also be achieved with high efficiency using the Gibson method

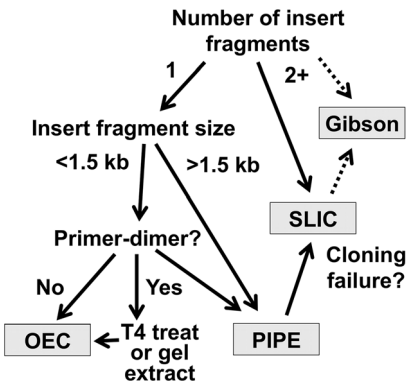


Figure 4. Technique selection flowchart for a new cloning project. The number of fragments, their size, primer availability and the presence of primer-dimers will determine the optimal cloning strategy. See main text for details.
doi:10.1371/journal.pone.0083888.g004

Table 2. Summary of effectiveness and resource use.

LIC Techniques	PIPE	SLIC	OEC
Cloning Efficiency <1.5 kb	Very High	Very High	High
Colony Number <1.5 kb	High	Very High	High
Cloning Efficiency >1.5 kb	Very High	Very High	Low
Colony Number >1.5 kb	High	Very High	Low
Primer Pairs	2	2[a]	1
Purification	Optional	Yes	Yes[b]
Thermal Cycling Rounds	1	1	2
T4 Treatment Step	No	Yes	No

See main text for details.
[a]Only 1 primer pair is required for SLIC if cleaved plasmid vector is used.
[b]Optional for small megaprimers (~500 bp).
doi:10.1371/journal.pone.0083888.t002

[11], which uses similar primer design to PIPE and SLIC, but involves use of a 5′ exonuclease, DNA ligase and DNA polymerase *in vitro* to join previously generated PCR fragments. This has proved to be particularly useful for synthetic biology projects requiring assembly of very large DNA fragments. However, for single fragment insertion, we do not believe that the increased resources required are justified when PIPE and SLIC robustly achieve high efficiencies. However, Gibson assembly [11] can prove a highly effective alternative when even SLIC fails (Figure 4).

A number of commercial cloning kits make use of site-specific recombination, requiring special vectors and costly proprietary enzymes, such as the Gateway® system (Life Technologies). Heterostagger or mixed PCR cloning only requires primers and polymerase to precisely engineer recessed ends similar to PIPE and SLIC [2,12]. Although very effective, it is also costly as it requires double the number of PCRs and primers used in SLIC, and long denaturing and annealing steps, without performing better [2].

We found that PIPE, SLIC and OEC were also very efficient for site-directed mutagenesis. PIPE or SLIC are optimal for large deletions [1,10], and extremely effective for substitutions or small insertions like epitope tags (Table 1). However, to create point mutations we routinely use an alternative two-stage cycling strategy similar to OEC, adapted from QuikChange [13,14]. It uses one mutagenic primer and a shorter pre-existing one, often a sequencing primer, to create a megaprimer by PCR, which then generates the mutant plasmid by overlap extension. This normally provides adequate efficiencies at minimal expense.

The LIC techniques tested in this study can be highly efficient and technically simple, but can be compromised by problems such as nicked vector plasmid background, cloning of primer-dimers and amplification of non-specific PCR products. However, these can be overcome through careful design, checking PCR products, and taking additional measures accordingly. Primer-dimers and non-specific products are only likely to be a problem if visible on the agarose gel or with a Bioanalyzer. However, nicked copies of the vector are often below the limit of detection, yet can still generate significant unwanted background (data not shown). Vector background colonies are generated when the vector plasmid templates are amplified rather than just the insert, or desired vector-backbone product (Figure 2). This potential problem is most relevant to PIPE, due to a relatively low number of transformants. The simplest effective method to reduce this background for PIPE is to use less template. If that fails or if greater robustness is desired, a restriction enzyme can be used to first cut the template outside of the primer binding sites. Increasing

the gap between the primers was also beneficial for our pUC18/Kan reporter, likely due to the difficulty of amplifying larger products in their entirety. We have previously observed higher efficiencies when replacing genes rather than PIPE cloning into an empty vector [6]. Using a vector containing an existing insert of distinguishable size can be particularly beneficial if it includes the lethal *ccdB* gene to select against vector background [1].

Non-specific PCR products and primer-dimers can be an important problem. Primer-dimers had little effect on PIPE or SLIC, likely due to their small size, such that they would lack the 5′-tails required for single-strand annealing and recombination. However, even low concentrations of primer-dimer are of concern for OEC, since cloning is extremely efficient for very small inserts, which clone preferentially. Primer-dimers could be removed by gel extraction or through pre-incubation with T4 polymerase to digest primer-dimer. Larger non-specific products will also compete with the desired insert in PIPE and SLIC, reducing cloning efficiency, which may require gel extraction. In our experience, these non-specific products can also be reduced through design of longer 3′ ends, increasing annealing temperatures, and reducing primer concentration. To avoid the requirement for gel extraction or PCR optimisation, alternative primer design can be used for Quick and Clean Cloning [9], a more specific variant of SLIC. This involves designing longer primers for one of the cloning junctions where the vector overhang is complementary to the region within the target insert PCR product, rather than insert primer. The resulting fragments can still be joined because short (~20 bp) ends of non-complementary sequence adjacent to the desired sequence are removed and successful recombination can occur *in vivo*, as long as the internal complementary sequence is single-stranded [2,9].

Ligation-independent cloning approaches constitute an essential part of the biomedical researcher's molecular-tool kit. With the extremely high fidelity of modern polymerases and availability of modular vector-specific primers, we find that ligation-independent methods are preferable even for simple subcloning projects. Due to their robustness, speed and low cost, they may largely supplant restriction enzyme and ligation-dependent cloning in many laboratories.

Supporting Information

File S1 Supporting Figure and Tables.
(PDF)

File S2 pUC18/Kan Reporter Vector Sequence.
(TXT)

File S3 Supporting Protocols.
(PDF)

Acknowledgments

We thank the members of the Brown lab for critically reviewing the manuscript.

Author Contributions

Conceived and designed the experiments: JS JRK. Performed the experiments: JS JRK LP. Analyzed the data: JS JRK LP AJB. Wrote the paper: JS JRK. Supervised the project: AJB.

References

1. Klock HE, Koesema EJ, Knuth MW, Lesley SA (2008) Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. Proteins 71: 982–994.
2. Li MZ, Elledge SJ (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 4: 251–256.
3. Bryksin AV, Matsumura I (2010) Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. Biotechniques 48: 463–465.
4. Unger T, Jacobovitch Y, Dantes A, Bernheim R, Peleg Y (2010) Applications of the Restriction Free (RF) cloning procedure for molecular manipulations and protein expression. J Struct Biol 172: 34–44.
5. Jeong JY, Yim HS, Ryu JY, Lee HS, Lee JH, et al. (2012) One-step sequence- and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies. Appl Environ Microbiol 78: 5440–5443.
6. Krycer JR, Brown AJ (2011) Cross-talk between the androgen receptor and the liver X receptor: implications for cholesterol homeostasis. J Biol Chem 286: 20637–20647.
7. Jenster G, Spencer TE, Burcin MM, Tsai SY, Tsai MJ, et al. (1997) Steroid receptor induction of gene transcription: a two-step model. Proc Natl Acad Sci U S A 94: 7879–7884.
8. Yabe D, Brown MS, Goldstein JL (2002) Insig-2, a second endoplasmic reticulum protein that binds SCAP and blocks export of sterol regulatory element-binding proteins. Proc Natl Acad Sci U S A 99: 12753–12758.
9. Thieme F, Engler C, Kandzia R, Marillonnet S (2011) Quick and clean cloning: a ligation-independent cloning strategy for selective cloning of specific PCR products from non-specific mixes. PLoS One 6: e20556.
10. Gill S, Stevenson J, Kristiana I, Brown AJ (2011) Cholesterol-dependent degradation of squalene monooxygenase, a control point in cholesterol synthesis beyond HMG-CoA reductase. Cell Metab 13: 260–273.
11. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6: 343–345.
12. Tillett D, Neilan BA (1999) Enzyme-free cloning: a rapid method to clone PCR products independent of vector restriction enzyme sites. Nucleic Acids Res 27: e26.
13. Sanchis J, Fernandez L, Carballeira JD, Drone J, Gumulya Y, et al. (2008) Improved PCR method for the creation of saturation mutagenesis libraries in directed evolution: application to difficult-to-amplify templates. Appl Microbiol Biotechnol 81: 387–397.
14. Tseng WC, Lin JW, Wei TY, Fang TY (2008) A novel megaprimed and ligase-free, PCR-based, site-directed mutagenesis method. Anal Biochem 375: 376–378.

Curr. Issues Mol. Biol. (2000) 2(1): 1-7.

Universal TA Cloning

Ming-Yi Zhou and Celso E. Gomez-Sanchez*

Endocrinology, The University of Mississippi
Medical Center, 2500 North State Street,
Jackson, MS 39216, USA
G.V. (Sonny) Montgomery VA Medical Center, Jackson,
MS and The University of Missouri-Columbia, MO USA

Abstract

TA cloning is one of the simplest and most efficient methods for the cloning of PCR products. The procedure exploits the terminal transferase activity of certain thermophilic DNA polymerases, including *Thermus aquaticus* (*Taq*) polymerase. *Taq* polymerase has non-template dependent activity which preferentially adds a single adenosine to the 3'-ends of a double stranded DNA molecule, and thus most of the molecules PCR amplified by *Taq* polymerase possess single 3'-A overhangs. The use of a linearized "T-vector" which has single 3'-T overhangs on both ends allows direct, high-efficiency cloning of PCR products, facilitated by complementarity between the PCR product 3'-A overhangs and vector 3'-T overhangs. The TA cloning method can be easily modified so that the same T-vector can be used to clone any double-stranded DNA fragment, including PCR products amplified by any DNA polymerase, as well as all blunt- and sticky-ended DNA species. This technique is especially useful when compatible restriction sites are not available for the subcloning of DNA fragments from one vector to another. Directional cloning is made possible by appropriate hemi-phosphorylation of both the T-vectors and the inserts. With a single T-vector at hand, any DNA fragment can be cloned without compromising the cloning efficiency. The universal TA cloning method is thus both convenient and labor-saving.

Introduction

The cloning of a DNA fragment into a plasmid vector is a routine procedure in recombinant DNA technology. Cloning methods can be divided into two classes, depending on whether or not ligase is required. The most common method for cloning and subcloning requires the use of DNA ligase to covalently link the compatible ends of the DNA fragment and the linearized plasmid, forming a single cyclic molecule that is capable of autonomous replication in host cells (1, 2). Recently, ligation-independent or ligase-free cloning methods have been developed for the cloning of polymerase chain reaction (PCR) amplified DNA fragments. These ligation-independent methods involve the generation of long (10-12 bases) protruding 3'-ends on PCR amplified DNA fragments which are annealed specifically to complementary DNA sequences of an appropriate plasmid vector, and the annealed products are subsequently transformed directly into competent cells. The application of the ligase-free procedure is limited by its requirement for extra-long primer synthesis and enzymatic modification steps (3-6).

There are two subclasses of ligase-dependent methods employed for the covalent joining of a DNA insert to a plasmid vector. One approach is cohesive-end ligation, in which cohesive ends on the insert and plasmid are generated by digestion with appropriate restriction enzymes, and the complementary ends of the plasmid and insert are then joined by DNA ligase. This approach is the most efficient of the two. However, for every restriction fragment having distinct cohesive ends, one linearized plasmid vector compatible with that (those) particular restriction site(s) has to be prepared. The other, less efficient approach is blunt-end ligation in which blunt-ended DNA fragments are ligated to a linearized blunt-ended plasmid. Both of the ligase-dependent methods require multiple purification and/or enzymatic modification steps.

Although the development of PCR made the rapid amplification of DNA fragments generally quite simple, the cloning of PCR products was often found to be quite difficult using early technology. Several methods have now been developed for the cloning of PCR-amplified DNA molecules, including cohesive-end cloning by the introduction of restriction sites at the 5'-ends of PCR primers (7), blunt-end cloning (8, 9), ligation-independent cloning (3-6), and TA cloning (10, 11, 12). It has been observed that many DNA polymerases, including *Thermus aquaticus* (*Taq*) DNA polymerase, are capable of adding an additional non-template directed nucleotide to the 3'-ends of a blunt-ended DNA fragment. *Taq* polymerase preferentially adds a single 3'-A to blunt-ended, double-stranded DNA via this terminal transferase-like activity (13, 14). Most of the PCR products amplified by *Taq* polymerase thus possess a single 3'-A overhang at both ends. To directly clone PCR products which have 3'-A overhangs, one can use a linearized "T-vector" which has a 3'-T overhang at each end (10, 11, 12). The complementarity between the vector 3'-T overhangs and PCR product 3'-A overhangs allows direct ligation of *Taq*-amplified PCR products into the T-vector, and this strategy is commonly referred to as "TA cloning." The TA cloning strategy is both simple and much more efficient than blunt-ended ligation for the cloning of PCR products.

With minor modifications, the TA cloning method, originally designed to facilitate the cloning of PCR products, is easily converted to a universal cloning method (15). All DNA fragments, with the exception of those with protruding 3'-ends, are directly converted to double stranded DNA molecules having 3'-A overhangs by incubation with *Taq* polymerase in the presence of the four dNTPs. These modified inserts are then ligated to T-vectors. In the case of DNA fragments having protruding 3'-ends, the DNA fragments are first made blunt by incubation with T4 DNA polymerase, and are then treated with *Taq* polymerase, as described above (Figure 1). A cloning efficiency of up to 90% is routinely achieved. The universal TA cloning technique is especially useful when compatible restriction sites are not available for subcloning DNA fragments from

*Corresponding author

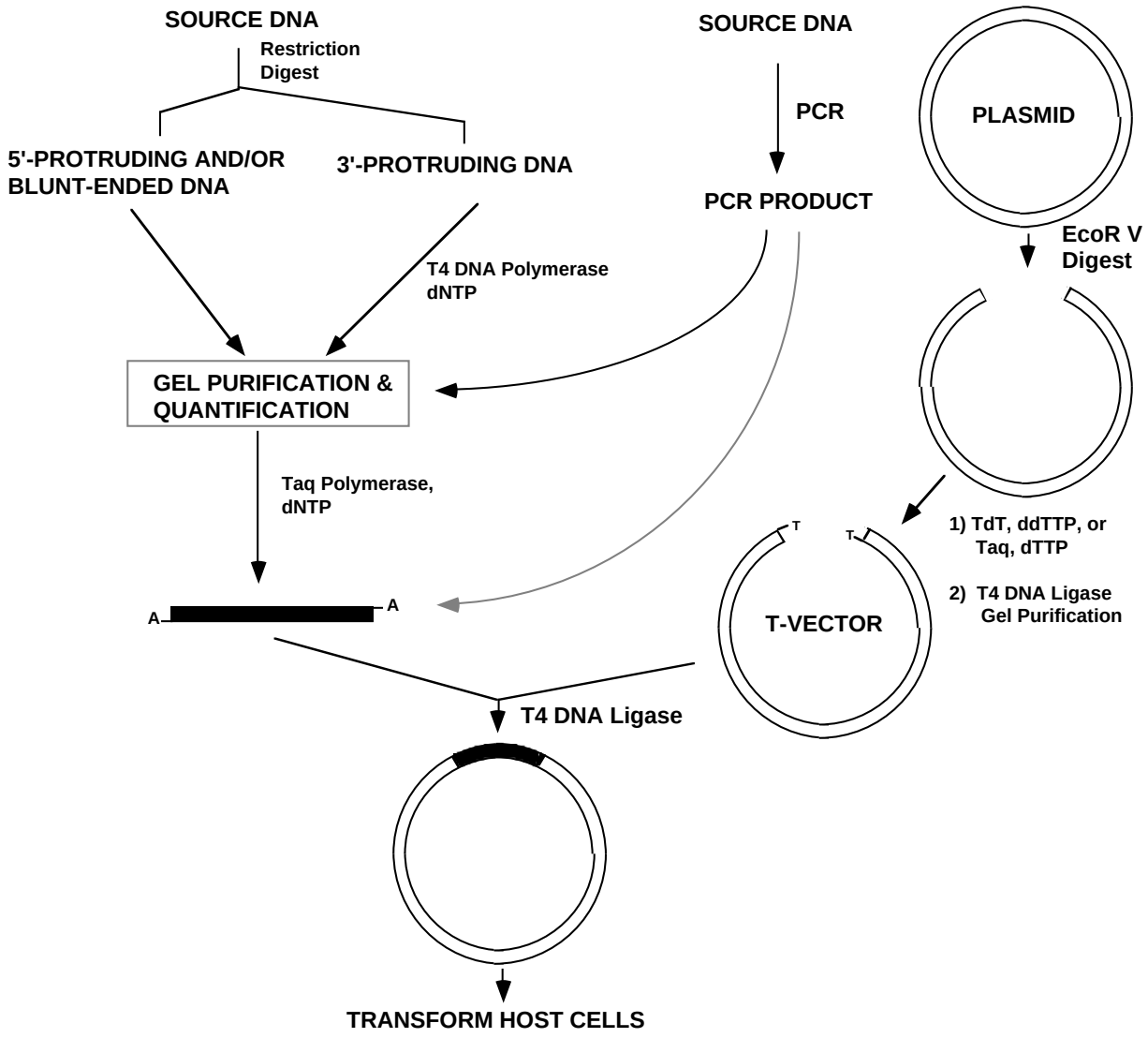


Figure 1. Schematic diagram outlining the cloning procedures using universal T-vectors (see text for details).

one vector into another. Directional cloning is made possible by the appropriate hemi-phosphorylation of both the T-vectors and the inserts. Thus, with a single T-vector at hand, any DNA fragment can be cloned without compromising the cloning efficiency. The universal TA cloning method is a convenient and labour-saving alternative to traditional, restriction endonuclease-mediated cloning strategies.

Protocols

Materials

dNTPs and ddTTP (United States Biochemical), high and low melting temperature agaroses (Fisher Scientific). *Taq* DNA polymerase (Promega), Terminal deoxynucleotidyl-transferase (Boehringer Mannheim), T4 DNA ligase (Promega), T4 DNA polymerase and Restriction enzymes (New England Biolabs). Appropriate buffers are provided or specified by the suppliers. Wizard DNA clean-up Kit, Wizard PCR Preps Kit and Wizard Minipreps DNA Purification Kit (Promega). One Shot or Ultracomp Top10F' (Invitrogen). pBluescript (Stratagene), pcDNA3 (Invitrogen).

Protocol 1: T-Vector Preparation

You may choose any plasmid which satisfies your needs as long as the plasmid contains a unique blunt-end restriction site in the multiple cloning sites polylinker (for example, the *Eco*R V site of the pBluescript plasmid, which has blue/white color selection, may be used for cloning). For expression in eukaryotic cells the pcDNA3 plasmid can be used. The T-vectors are prepared in batch and stored as aliquots at -20°C for future use.

Digest Plasmids with Blunt-End Restriction Enzyme

1. In a 1.5 ml microcentrifuge tube, add:

plasmid DNA	10 μg
10X *Eco*R V buffer	5 μl
*Eco*R V	50 units
water to final total volume of	50 μl

   Mix by gentle vortex and centrifuge briefly.
2. Incubate at 37°C for 1-2 h.
3. After 1 h of digestion, 1-2 μl of mix are taken and run on a 1% agarose gel to make sure the digestion is complete.
4. Add 2 μl of 0.5 M EDTA to the tube at the end of digestion.

Purification of Blunt-Ended Plasmid DNA

We routinely use the Wizard DNA Clean-up System to purify DNA, and the major steps are listed below. When using the Wizard DNA Clean-up kit, it is important to elute the DNA from the resin with water. EDTA in the TE buffer will chelate Mg^{++}, thus reducing the working concentration of Mg^{++} in the next enzyme incubation step. The phenol/chloroform extraction and ethanol precipitation method also works well.

1. Add 1 ml of well mixed Wizard DNA Clean-up resin to the tube containing the enzyme digestion products, and mix by inverting for 1 min.
2. Attach a 3 ml disposable syringe barrel (without the plunger) to the minicolumn provided, and insert the tip of the minicolumn/syringe barrel assembly into a vacuum manifold.
3. Pipet the DNA/resin mix into the syringe barrel. Apply the vacuum to draw the mix into the minicolumn.
4. Pipet 2 ml of 80% isopropanol into the syringe barrel, and apply the vacuum to draw the solution through the minicolumn.
5. Remove the syringe barrel and transfer the minicolumn to a 1.5 ml centrifuge tube. Centrifuge the minicolumn for 2 min at 12,000 x g.
6. Attach the minicolumn to the vacuum source, and apply the vacuum for another 2 min.
7. Transfer the minicolumn to a new 1.5 ml microcentrifuge tube, add 50 μl of water (pre-warmed to 65°C) to the minicolumn, and wait for 1 min. Centrifuge the minicolumn for 20 s at 12,000 x g. The DNA is collected in the microcentrifuge tube.

Addition of T-overhangs to the Blunt-ended Plasmid Vector

Two widely used methods are described. The first method employs terminal deoxynucleotidyl-transferase (TdT) and dideoxythymidine triphosphate (ddTTP) to add a ddT residue to the blunt-ended plasmid, which ensures the addition of only a single T residue (10). In the second method, *Taq* polymerase and dTTP are used to add a 3'-T to the blunt-ended plasmid (11). Since *Taq* polymerase adds the 3'-T residues inefficiently, only a portion of the blunt-ended plasmids are T-tailed at both ends, whereas the remainder are T-Tailed only at one end or do not have T-tails at all. Therefore, the vectors without T-overhangs are eliminated from the T-vectors by gel purification after self-ligation and/or concatemerization by incubation with T4 DNA ligase (16). Unlike *Taq* polymerase, terminal transferase adds ddT to 3'-ends very efficiently. Therefore consistent and high cloning efficiency is obtained using T-vectors prepared by TdT, even without ligation and gel purification steps. However, the efficiency can be further increased by incubation with T4 DNA ligase, followed by gel separation.

A. Terminal Transferase Method

This method originated from Holton and Graham (10), with minor modifications in our laboratory. Terminal transferase from Boehringer Mannheim is strongly recommended. TdT from other sources may not work as well. The final concentration of ddTTP in the reaction is 20 μM; higher concentrations are unnecessary.

1. Pipet 46 μl of purified blunt-ended plasmid DNA (from step 7 of "*Purification of Blunt-ended Plasmid DNA*") into a 1.5 ml centrifuge tube (if the volume is less, add sterile water to bring the volume to 46 μl), then mix the following reagents:

purified Blunted plasmid DNA	46 μl
5 X TdT buffer	15 μl
25 mM $CoCl_2$	7.5 μl
1 mM ddTTP	1.5 μl
Terminal Transferase (25 U/μl)	5 μl

   Mix by gentle vortexing and centrifuge briefly.
2. Incubate at 37°C for 1.5 h, then proceed to the *"Purification of the T-tailed Vectors"* (see below).

B. *Taq* Polymerase Method

1. Adjust the volume of the purified blunted plasmid DNA (from step 7 of "*Purification of Blunt-ended Plasmid DNA*") to 83 μl by adding sterile water, then mix the following reagents:

Purified blunted plasmid DNA	83 μl
10 x PCR buffer ($MgCl_2$ free)	10 μl
50 mM $MgCl_2$	5 μl
100 mM dTTP	1 μl
Taq polymerase (5 U/μl)	1 μl

   Mix well and centrifuge briefly.
2. Incubate at 72°C for 2 h, then proceed to the *"Purification of the T-tailed Vectors"* (see below).

Purification of the T-tailed Vectors

The T-tailed plasmid DNA solution obtained by any one of the methods described above is purified by employing the Wizard DNA Clean-up System as described above or phenol/chloroform extraction and ethanol precipitation.

Ligation of Vectors without T-tails

If you use the *Taq* polymerase method to make the T-vectors, removal of the vectors without T-tails is necessary and improves the cloning efficiency up to 80% (16). If the terminal transferase method and a plasmid without blue/white selection (for example pcDNA3) is used to make T-vectors, the cloning efficiency is about 70%, without further purification. Elimination of the vectors without T-overhangs by gel separation, following incubation with T4 DNA ligase, increases the cloning efficiency to 90%. If you use the terminal transferase method to make T-vectors, and the plasmid has blue/white selection capabilities, ligation and gel purification may not be necessary (Go directly to "*Quantification and Storage of the Purified T-vector"* below if time savings is desired).

1. To a 1.5 ml microcentrifuge tube, add

T-tailed plasmid DNA (as purified above)	44 μl
10x ligase buffer	5 μl
T4 DNA ligase (3 U/μl)	1 μl

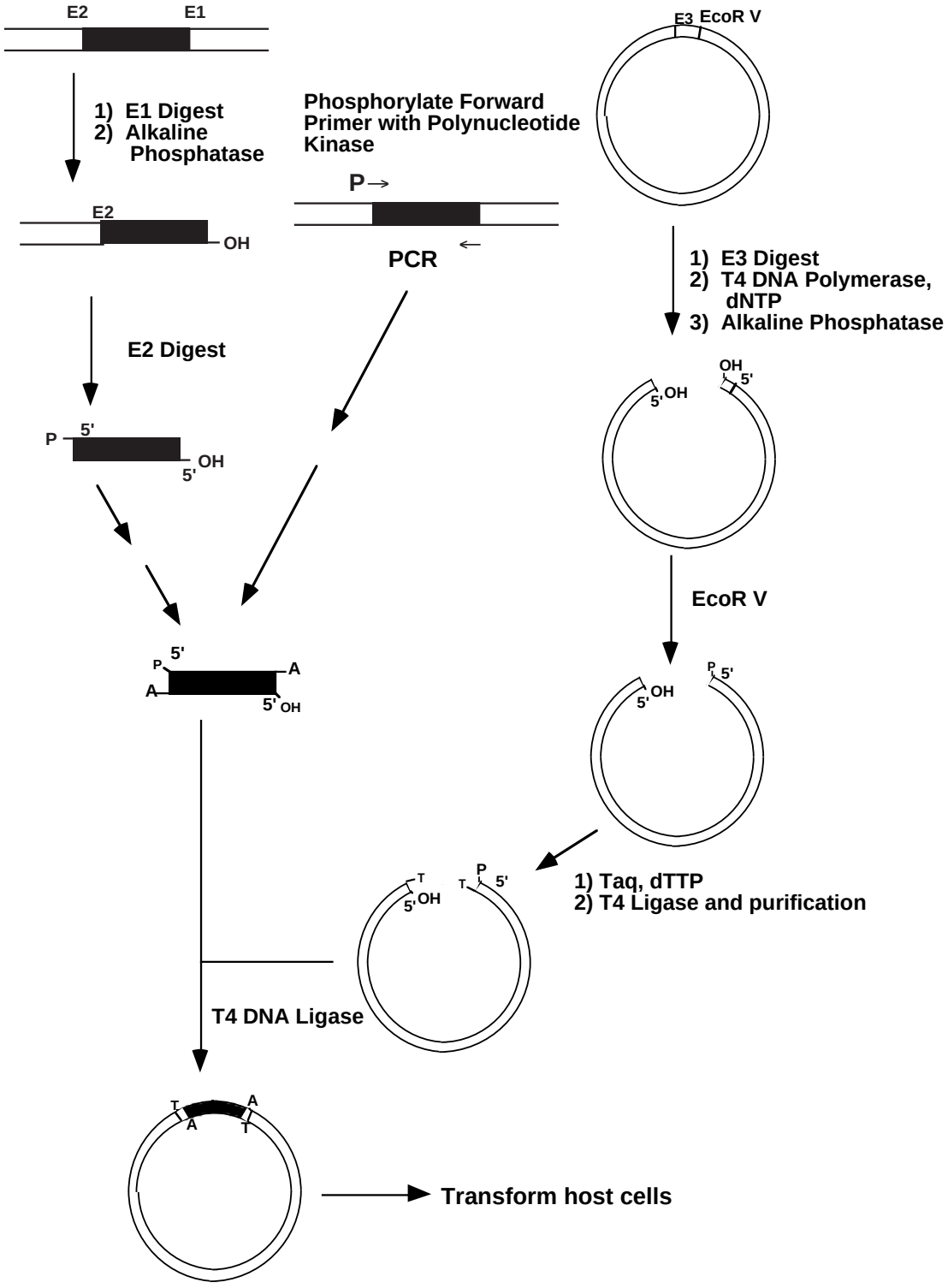


Figure 2. Schematic diagram demonstrating the strategies for directional TA cloning. E1, E2 and E3 are restriction enzymes. When E3 is an enzyme which generates blunt ends, the incubation step with T4 DNA polymerase and dNTP is not necessary.

Mix well and centrifuge briefly.
2. Incubate at 16°C overnight.

Separation and Recovery of T-vectors on an Agarose Gel
T-vectors are separated from the self-ligated and concatemerized plasmid DNA on a 1% agarose gel. The T-vector band is cut out and recovered using the Wizard PCR Preps Kit.
1. Separate the ligation mixture (from above) on a 1% low-melting agarose gel with 0.5 μg/ml of ethidium bromide. Run the gel in TAE buffer at 5 volts/cm.
2. Visualize the DNA bands with a hand-held long wave (365 nm) UV light. Excise the T-vector band with a sterile scalpel. Transfer the gel slice to a 2 ml microcentrifuge tube.
3. Incubate the gel slice in a 65°C water bath for 5 min, or until melted.
4. Add 1 ml of Wizard PCR preparation resin (which has been thoroughly mixed) to the melted agarose slice. Mix thoroughly by inverting for 1 min.
5. Follow steps 2 to 7 of the " *Purification of Blunt-Ended Plasmid DNA"* to complete the purification.

Quantification and Storage of the Purified T-vector
The concentration of the purified T-vector or other DNA can be conveniently estimated by running an aliquot of the DNA side by side with a known amount of a DNA standard in a 1% agarose gel, and then comparing the relative brightnesses of the bands. The electrophoresis is run at a higher than normal voltage for a short time (usually 10-15 min) to allow the DNA to more efficiently migrate into the gel. Standard loading buffer contains bromphenol blue and xylene cyanol which may interfere with the visualization of band brightness. Therefore, it is strongly suggested that a 6 X loading buffer without these dyes be used for quantification purposes. The T-vectors are aliquoted and stored at -20°C to reduce multiple freeze-thaw cycles.
1. Run 1 μl of the purified T-vector and a known amount of standard DNA side by side (use loading buffer without any dyes) in a 1% agarose gel at 10 volts/cm for 10-15 min.
2. Compare the band brightnesses to determine the concentration of the T-vector.
3. Aliquot the T-vector to several tubes and store them at -20°C.

Protocol 2: TA Cloning Procedure
Preparation of Source DNA
A. DNA Fragments from Restriction Enzyme Digest

Usually two restriction enzymes are needed to release the insert DNA. It is frequently possible to choose a buffer in which the two enzymes will remain active, and therefore the two enzyme digestions can be carried out simultaneously. The protocol described below gives an example of two digestions performed in the same buffer. In cases of either buffer or temperature incompatibility, separate digestions may be required. The reader is advised to refer to some basic manuals (1). If the insert has one or two 3'-overhangs, they must be removed with T4 DNA polymerase.

1. In a 1.5 ml microcentrifuge tube, add:

DNA	5 μg
10 X buffer	4 μl
Enzyme 1	20 Units
Enzyme 2	20 Units
sterile water to final total volume of 40 μl	

Mix well and centrifuge briefly.
2. Incubate at the recommended temperature (usually 37°C) for 1-2 h. If the enzymes produce inserts with 5'-overhangs or blunt-ends, proceed to the gel purification step. If inserts have 3'-overhang(s), continue with steps 3 to 5.
3. Steps 3 to 5 are only for the enzymes which produce insert DNAs with 3'-overhangs. Put the above digestion solution on ice, and add:

dNTP mix, 10 mM each	0.5 μl
T4 DNA polymerase	6 Units

Mix well and centrifuge briefly.
4. Incubate at 11°C for 20 min.
5. Add 2 μl of 0.5 M EDTA to stop the reaction. Then proceed to "*Separation and Recovery of DNA Fragments from an Agarose Gel*".

B. DNA Fragments Amplified by PCR

Large amounts of DNA can be obtained by PCR. Many variations and applications of this technique, such as RT-PCR (reverse transcription PCR), 5'-RACE (rapid amplification of 5'-cDNA ends), and XL-PCR (extra long PCR), have been developed to serve various needs. A successful PCR amplification starts with the selection of appropriate primer pairs and then optimization of the PCR conditions (such as annealing temperature, extension time, and Mg^{++} concentration). Several thermal stable enzymes such as *Taq*, *Vent*, *pfu*, and *Tth* polymerases, are available. General PCR references should be consulted to select the enzyme best suited to your needs (17, 18). After successful PCR amplification, proceed to the next stage to clone the amplified DNA product.

Separation and Recovery of DNA Fragments from an Agarose Gel
The desired DNA fragments from restriction digestion are separated from other DNA fragments by agarose gel electrophoresis and the desired bands are cut out and recovered. PCR products consisting of a single specific band amplified by *Taq* polymerase can be directly ligated to T-vectors. However, the removal of the impurities such as primers and trace amounts of nonspecific products by gel separation or the use of direct DNA clean-up systems increases the percentage of colonies containing the correct inserts. If you use a thermo-stable DNA polymerase which does not preferentially add a 3'-A at the ends of the PCR products, removal of that enzyme by either gel separation or use of a direct DNA clean-up system is essential. When two or more desired PCR products are amplified in the same PCR reaction, or when the nonspecific amplification products are abundant enough to be visualized on a gel, it is necessary to purify the desired DNA fragments by gel electrophoresis (to avoid screening the positive colonies later by more complicated hybridization procedures). Purify and recover the DNA from agarose gels as described in the *"Separation and Recovery of T-vectors on an Agarose Gel"* section of protocol 1. In addition, the concentration of the purified DNA is estimated as described in protocol 1.

A-tailing of the Purified DNA using Taq Polymerase

In the presence of all four dNTPs, *Taq* polymerase will first fill in the 3'-recessed bases of the DNA fragment with 5'-overhangs and then preferentially add A's to the 3'-ends. Alternatively, the *Taq* enzyme will add an A directly to the 3'-ends of a blunt-ended DNA fragment. In the case of PCR products or blunt-ended fragments, only dATP is necessary in the incubation mixture. However, a mixture of dNTPs works well in all cases, since *Taq* polymerase preferentially adds an A to the 3'-ends in the presence of all four dNTPs.

Before starting this step, calculate the concentration of the DNA fragment needed in a 10 μl ligation reaction to give a 1:2 molar ratio of the T-vector to the inserts. Since no more than 2 μl of A-tailed DNA is used in the ligation reaction, the final concentration of the DNA fragment in the A-tailing reaction mixture should be adjusted so that the amount of DNA fragment required for ligation will be contained in a volume of 2 μl or less. Should the DNA concentration be too low, you may concentrate the DNA before the *Taq* polymerase incubation step.

1. To a 0.5 ml sterile centrifuge tube, add:

Purified DNA fragment	7.7 μl
10 X PCR buffer, Mg^{++}- free	1 μl
25 mM $MgCl_2$	1 μl
dNTP mix, 10 mM each	0.2 μl
Taq polymerase (5 U/μl)	0.1 μl

   Mix well and centrifuge briefly.
2. Incubate at 72°C for 25 min, then transfer the tube to ice.

Ligation of the A-tailed DNA Fragment to the T-vector

Usually 50-60 ng of T-vector is enough for each ligation. A 1:2 molar ratio of T-vector to insert DNA is recommended, but do not add more than 2 μl of A-tailed DNA solution in a 10 μl ligation (see notes in previous section).

1. In a sterile 0.5 ml microcentrifuge tube, add

10 X ligation buffer	1 μl
A-tailed DNA (from previous step)	≤2 μl
T-vector	60 ng
T4 DNA ligase (2-3 U/μl)	1 μl
Sterile water	to final total volume of 10 μl

   Mix gently and centrifuge briefly.
2. Incubate at 14°C for 16 h or overnight. You may store the tube at -20°C if you do not plan to transform the cells immediately.

Transformation of the Ligated Vector

LB agar plates with appropriate antibiotics (most commonly 100 μg/ml of ampicillin) should be available and warmed to room temperature. If a plasmid with blue/white color selection is used, the LB plates should be spread with 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal) and isopropylthio-β-D-galactoside (IPTG) before transformation. A 42°C water bath should be available for the transformation.

When a control ligation reaction with the T-vector alone is transformed, a considerable number of colonies may be obtained. When ligating inserts to the T-vectors, up to 90 % colonies should be positive. A successful ligation and transformation should yield hundreds of colonies. When few colonies are obtained, the ligase activity or the competency of the cells should be checked.

1. Thaw on ice one 50 μl aliquot of appropriate frozen competent cells (for example, Top 10F') for each ligation reaction.
2. Add 2 μl of 0.5 M β-mercaptoethanol to the competent cells and mix by stirring gently with the pipette tip.
3. Add 2 μl of the ligation reaction to the competent cells, and mix gently by stirring with the pipette tip.
4. Incubate on ice for 30 min.
5. Heat shock the cells for 30 s in the 42°C water bath, then immediately place the cells on ice for 2 min.
6. Add 450 μl of SOC medium to the transformed cell vial.
7. Shake the cells at 37°C for 45-60 min at 225 rpm in a rotary shaking incubator.
8. Spread 100 μl of the transformed cells on each labeled LB-ampicillin plate. If plasmid with blue/white selection is used, the LB-plates should be spread with X-Gal and IPTG before spreading the cells.
9. Invert the plates and place them in a 37°C incubator overnight.

Identification of Positive Colonies

After transformation, the positive colonies may be identified either by restriction mapping of the miniprepared plasmids or screening by PCR. PCR screening can be carried out with primers that flank the vector cloning sites. In the case of the pBluescript plasmid, T3 and T7 primers may be used (19). The length of the amplified PCR product of a positive clone will be the size of the insert plus the bases flanking the two primers on the plasmid. This method is rapid, but may not work in some cases, especially when the insert is very large. Specific primers for the cloned insert may also be used as PCR screening primers. For screening by restriction mapping, plasmids are miniprepared, the insert is released by digestion with two unique restriction enzymes from the multiple cloning sites, and the insert size is confirmed by agarose gel electrophoresis. The orientation of the insert in the plasmid can also be determined by appropriate restriction digestion, in some cases. If needed, the orientation and identity of the clone may be further confirmed by sequencing. Because of the high cloning efficiency, in most cases only four to six colonies need to be screened.

Screening by Restriction Mapping

1. Randomly pick 4-6 colonies (in case of blue/white color selection, pick white colonies), and grow overnight at 37°C in 4 ml of LB medium with 50 μg/ml of ampicillin.
2. Pellet 2-3 ml of the cells by centrifuging for 2 min at top speed in a microcentrifuge. Isolate plasmids using Wizard Minipreps DNA purification system. The remaining cells may be stored at 4°C for 1-2 days. For long term storage, add glycerol to 15%, and store at -70°C.
3. Isolate plasmid DNA and digest 10-50% of the total preparation with the appropriate restriction enzymes.
4. Run an agarose gel to determine which colonies have the correct insert size and/or orientation.
5. Confirm the clone by sequencing when necessary.

Notes

The universal TA cloning procedure relies on the assumption that all DNA fragments can be easily converted to double stranded DNA with single 3'-A overhangs at each end, and thus the T-vector becomes a ready-for-ligation universal cloning vector. The T-vectors can be prepared in bulk and stored as aliquots at -20°C for future use. Long term storage may decrease the cloning efficiency and increase the background. New batches of T-vectors may be prepared using either the terminal transferase or *Taq* polymerase methods. T-vectors prepared employing terminal transferase produce more consistent results than those prepared by *Taq* polymerase. TA cloning kits which are convenient and have a high cloning efficiency are commercially available. However, cost may be a factor to consider.

The insert DNA may be ligated to the T-vectors in either direction, and TA cloning is thus ideal when both orientations of the insert DNA in the plasmid are required. The orientation of the insert in the plasmid can be determined by sequencing, and in some cases, it can be simply determined by restriction mapping. Unidirectional TA cloning can be achieved by hemi-phosphorylation of one end and dephosphorylation of the other end of both the T-vectors and the insert (Figure 2). Thus, only ligation in the designated direction can form a cyclic plasmid (with two nicks) which can be efficiently transformed into competent cells. The other ligation products will be linear heterodimers which can not be efficiently transformed into host cells. Hemi-phosphorylated T-vectors may be prepared by the *Taq* polymerase method as illustrated in Figure 2. The terminal transferase method is not suitable for the preparation of the hemiphosphorylated T-vectors, because the presence of 3'-ddT at the terminus of one strand and the lack of a 5'- phosphate on the other strand will prevent the formation of transformable cyclic plasmids. To prepare the PCR product for directional TA cloning, a single phosphorylated PCR primer, which can be prepared employing T4 polynucleotide kinase, is used to generate a hemi-phosphorylated PCR product. For DNA generated by restriction digestion, the substrate is first digested with one enzyme, dephosphorylated with alkaline phosphatase, and then digested with the second enzyme (Figure 2; Note: The alkaline phosphatase must be inactivated before the second enzyme digestion). The rest of the procedures will be the same as for the bi-directional TA cloning protocols.

References

1. Ausubel,F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. 1987. Current Protocols in Molecular Biology. John Wiley & Sons, New York, NY.
2. Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
3. Rashtchian, A. 1995. Novel methods for cloning and engineering genes using the polymerase chain reaction. Current Opinion in Biotech. 6: 30-36.
4. Hsiao, K. 1993. Exonuclease III induced ligase-free directional subcloning of PCR products. Nucleic Acids Res. 21: 5528-5529.
5. Aslanidis, C. and de Jong, P.J. 1991. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. 18: 6069-6074.
6. Kaluz, S. and Flint, A.P.F. 1994. Ligation-independent cloning of PCR products with primers containing nonbase residues. Nucleic Acids Res. 22: 4845.
7. Jung, V., Pestka, S.B. and Pestka, S. 1993. Cloning of polymerase chain reaction-generated DNA containing terminal restriction endonuclease recognition sites. Methods in Enzymology. 208: 357-362.
8. Crowe, J.S., Cooper, H.J., Smith, M.A., Sims, M.J., Parker, D. and Gewert, D. 1991. Improved cloning efficiency of polymerase chain reaction (PCR) products after proteinase K digestion. Nucleic Acids Res. 19: 184.
9. Costa, G.L. and Weiner, M.P. 1994. Polishing with T4 or *Pfu* polymerase increases the efficiency of cloning of PCR fragments. Nucleic Acids Res. 22: 2423.
10. Holton, T. A., and Graham, M. W. 1991. A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucleic Acids Res. 19: 1156.
11. Marchuk, D., Drumm, M., Saulino, A. and Collins, F. S. 1991. Construction of T- vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res. 19: 1154.
12. Mead, D.A., Pey, N.K., Herrnstadt, C., Marcil, R.A. and Smith, L.M. 1991. A universal method for the direct cloning of PCR amplified nucleic acid. Bio/Technology. 9: 657-663.
13. Clark, J. M. 1988. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. Nucleic Acids Res. 16: 9677- 9686.
14. Hu, G. 1993. DNA polymerase-catalyzed addition of nontemplated extra nucleotides to the 3' end of a DNA fragment. DNA Cell Biol. 12: 763-770.
15. Zhou, M.Y., Clark, S.E. and Gomez-Sanchez, C.E. 1995. Universal cloning method by TA strategy. BioTechniques. 19:34-35.
16. Hadjeb, N. and Berkowitz, G.A. 1996. Preparation of T-overhang vectors with high PCR product cloning efficiency. BioTechniques. 20: 20-22.
17. Erlich, H.A. 1989. PCR Technology. Stockton Press, New York, NY.
18. Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. 1990. PCR Protocols: A guide to Methods and Applications. Academic Press, Inc., San Diego, CA.
19. Trower, M.K. and Elgar, G.S. 1994. PCR cloning using T-vectors. In: Methods in Molecular Biology, Vol.31: Protocols for Gene Analysis. A.J. Harwood, ed. Humana Press Inc., Totowa, NJ.

Review

Cell PRESS

ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering

Thomas Gaj[1,2,3], Charles A. Gersbach[4,5], and Carlos F. Barbas III[1,2,3]

[1] The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA, USA
[2] Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA, USA
[3] Department of Chemistry, The Scripps Research Institute, La Jolla, CA, USA
[4] Department of Biomedical Engineering, Duke University, Durham, NC, USA
[5] Institute for Genome Sciences and Policy, Duke University, Durham, NC, USA

Zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) comprise a powerful class of tools that are redefining the boundaries of biological research. These chimeric nucleases are composed of programmable, sequence-specific DNA-binding modules linked to a nonspecific DNA cleavage domain. ZFNs and TALENs enable a broad range of genetic modifications by inducing DNA double-strand breaks that stimulate error-prone nonhomologous end joining or homology-directed repair at specific genomic locations. Here, we review achievements made possible by site-specific nuclease technologies and discuss applications of these reagents for genetic analysis and manipulation. In addition, we highlight the therapeutic potential of ZFNs and TALENs and discuss future prospects for the field, including the emergence of clustered regulatory interspaced short palindromic repeat (CRISPR)/Cas-based RNA-guided DNA endonucleases.

Classical and contemporary approaches for establishing gene function

With the development of new and affordable methods for whole-genome sequencing, and the design and implementation of large genome annotation projects, scientists are poised to deliver upon the promises of the genomic revolution to transform basic science and personalized medicine. The resulting wealth of information presents researchers with a new primary challenge of converting this enormous amount of data into functionally and clinically relevant knowledge. Central to this problem is the need for efficient and reliable methods that enable investigators to determine how genotype influences phenotype. Targeted gene inactivation via homologous recombination is a powerful method capable of providing conclusive information for evaluating gene function [1]. However, the use of this technique has been hampered by several factors, including the low efficiency at which engineered constructs are correctly inserted into the chromosomal target site, the need for time-consuming and labor-insensitive selection/screening strategies, and the potential for adverse mutagenic effects. Targeted gene knockdown by RNAi (see Glossary) has provided researchers with a rapid, inexpensive, and high-throughput alternative to homologous recombination [2]. However, knockdown by RNAi is incomplete, varies between experiments and laboratories, has unpredictable off-target effects, and provides only temporary inhibition of gene function. These restrictions impede researchers' ability to link directly phenotype to genotype and limit the practical application of RNAi technology.

Glossary

CRISPR/Cas (CRISPR associated) systems: clustered regulatory interspaced short palindromic repeats are loci that contain multiple short direct repeats, and provide acquired immunity to bacteria and archaea. CRISPR systems rely on crRNA and tracrRNA for sequence-specific silencing of invading foreign DNA. Three types of CRISPR/Cas systems exist: in type II systems, Cas9 serves as an RNA-guided DNA endonuclease that cleaves DNA upon crRNA–tracrRNA target recognition.
crRNA: CRISPR RNA base pairs with tracrRNA to form a two-RNA structure that guides the Cas9 endonuclease to complementary DNA sites for cleavage.
DSB: the product of ZFN, TALEN, and CRISPR/Cas9 action, double-strand breaks are a form of DNA damage that occurs when both DNA strands are cleaved.
HDR: homology-directed repair is a template-dependent pathway for DSB repair. By supplying a homology-containing donor template along with a site-specific nuclease, HDR faithfully inserts the donor molecule at the targeted locus. This approach enables the insertion of single or multiple transgenes, as well as single nucleotide substitutions.
NHEJ: nonhomologous end joining is a DSB repair pathway that ligates or joins two broken ends together. NHEJ does not use a homologous template for repair and thus typically leads to the introduction of small insertions and deletions at the site of the break, often inducing frame-shifts that knockout gene function.
PAM: protospacer adjacent motifs are short nucleotide motifs that occur on crRNA and are specifically recognized and required by Cas9 for DNA cleavage.
RNAi: RNAi is the process by which RNA molecules inhibit or knockdown gene expression. More broadly, RNAi is a natural mechanism that occurs in response to the introduction of many types of RNA molecules into cells.
TALENs: transcription activator-like effector nucleases are fusions of the *Fok*I cleavage domain and DNA-binding domains derived from TALE proteins. TALEs contain multiple 33–35-amino-acid repeat domains that each recognizes a single base pair. Like ZFNs, TALENs induce targeted DSBs that activate DNA damage response pathways and enable custom alterations.
tracrRNA: trans-activating chimeric RNA is noncoding RNA that promotes crRNA processing and is required for activating RNA-guided cleavage by Cas9.
ZFNs: zinc-finger nucleases are fusions of the nonspecific DNA cleavage domain from the *Fok*I restriction endonuclease with zinc-finger proteins. ZFN dimers induce targeted DNA DSBs that stimulate DNA damage response pathways. The binding specificity of the designed zinc-finger domain directs the ZFN to a specific genomic site.
ZFNickases: zinc-finger nickases are ZFNs that contain inactivating mutations in one of the two *Fok*I cleavage domains. ZFNickases make only single-strand DNA breaks and induce HDR without activating the mutagenic NHEJ pathway.

Corresponding author: Barbas, C.F. III (carlos@scripps.edu).
Keywords: zinc-finger nuclease; transcription activator-like effector nuclease; clustered regulatory interspaced short palindromic repeat; genome engineering..

0167-7799/$ – see front matter
 http://dx.doi.org/10.1016/j.tibtech.2013.04.004

In the past decade, a new approach has emerged that enables investigators to manipulate virtually any gene in a diverse range of cell types and organisms. This core technology – commonly referred to as 'genome editing' – is based on the use of engineered nucleases composed of sequence-specific DNA-binding domains fused to a nonspecific DNA cleavage module [3,4]. These chimeric nucleases enable efficient and precise genetic modifications by inducing targeted DNA double-strand breaks (DSBs) that stimulate the cellular DNA repair mechanisms, including error-prone nonhomologous end joining (NHEJ) and homology-directed repair (HDR) [5]. The versatility of this approach is facilitated by the programmability of the DNA-binding domains that are derived from zinc-finger and transcription activator-like effector (TALE) proteins. This combination of simplicity and flexibility has catapulted zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) to the forefront of genetic engineering. Here, we review recent advances in site-specific nuclease technologies and discuss applications of these reagents for targeted genome engineering and analysis in eukaryotic cells and model organisms. We also discuss the therapeutic potential of these technologies and examine future prospects, including the development and application of clustered regularly interspaced short palindromic repeats CRISPR/Cas (CRISPR-associated)-based RNA-guided DNA endonucleases.

Custom DNA-binding domains

The versatility of ZFNs and TALENs arises from the ability to customize the DNA-binding domain to recognize virtually any sequence. These DNA-binding modules can be combined with numerous effector domains to affect genomic structure and function (Box 1), including nucleases, transcriptional activators and repressors, recombinases, transposases, DNA and histone methyltransferases, and histone acetyltransferases. Thus, the ability to execute genetic alterations depends largely on the DNA-binding specificity and affinity of designed zinc-finger and TALE proteins. Below, we highlight several of the most successful approaches for assembling these modular DNA-binding domains.

Cys_2–His_2 zinc-finger proteins

The Cys_2–His_2 zinc-finger domain is among the most common types of DNA-binding motifs found in eukaryotes and represents the second most frequently encoded protein domain in the human genome. An individual zinc-finger consists of approximately 30 amino acids in a conserved $\beta\beta\alpha$ configuration [6] (Figure 1A). Several amino acids on the surface of the α-helix typically contact 3 bp in the major groove of DNA, with varying levels of selectivity. The modular structure of zinc-finger proteins has made them an attractive framework for the design of custom DNA-binding proteins. Key to the application of zinc-finger proteins for specific DNA recognition was the development of unnatural arrays that contain more than three zinc-finger domains. This advance was facilitated by the structure-based discovery of a highly conserved linker sequence that enabled construction of synthetic zinc-finger proteins that recognized DNA sequences 9–18 bp in length [7]. Because 18 bp of DNA sequence can confer specificity within 68 billion bp of DNA, this method allowed for specific sequences to be targeted in the human genome for the first time [8,9]. Although initially controversial [10], this design has proven to be the optimal strategy for constructing zinc-finger proteins that recognize contiguous DNA sequences that are specific in complex genomes [6–9,11–15].

Following this proof-of-principle work, several methods for constructing zinc-finger proteins with unique DNA-binding specificity were developed. The 'modular assembly' approach involves the use of a preselected library of zinc-finger modules generated by selection of large combinatorial libraries or by rational design [6,16]. Because zinc-finger domains have been developed that recognize nearly all of the 64 possible nucleotide triplets, preselected zinc-finger modules can be linked together in tandem to target DNA sequences that contain a series of these DNA triplets [6,8,13–15,9]. Alternatively, selection-based approaches, such as oligomerized pool engineering (OPEN) can be used to select for new zinc-finger arrays from randomized libraries that take into consideration context-dependent interactions between neighboring fingers [17]. Approaches have also been developed that combine the methods described above, utilizing zinc-finger modules preselected for con-

Box 1. Beyond nucleases: recombinases, transposases, and transcription factors

Site-specific nucleases are currently the most well-characterized, widely used and broadly applicable tool for inducing custom modifications in cells and model organisms. However, several limitations of targeted nucleases are driving the development of alternative types of programmable enzymes for genome engineering. For example, off-target effects created by site-specific nucleases can be toxic to cells, and difficult to predict and monitor comprehensively. Additionally, because targeted nucleases rely on NHEJ and HDR to induce genetic alterations, this technology may be limited by the availability of the desired DNA repair mechanism in particular cell types. To address these concerns, zinc-finger proteins and TALEs have been fused to enzymatic domains, including site-specific recombinases [28,109–111] and transposases [116], that catalyze DNA integration, excision, and inversion. Because these enzymes perform DNA cleavage and religation autonomously, potentially toxic DNA DSBs should not accumulate in the genome. Additionally, for applications that require targeted gene addition, recombinase and transposase activity is marked by the insertion of donor DNA into the genome, thereby enabling off-target effects to be monitored directly. Moreover, the mechanism of recombination and transposition is independent of cellular DNA repair pathways. As a result, these approaches should be functional in nearly any cell type and cell cycle stage. The efficiency of these processes can also be improved by directed evolution [117]. However, in order for recombinases and transposases to achieve the level of general utility afforded by site-specific nucleases, significant improvements in their performance and flexibility are needed. In particular, recombinase catalytic domains retain sequence specificity from the parental enzyme, and require significant re-engineering toward user-defined DNA targets [109,111]. Although transposase fusions demonstrate high activity at their intended genomic targets, these chimeric proteins also suffer from significant off-target activity [118]. Finally, synthetic zinc-finger and TALE transcription factors offer an alternative approach for inducing targeted modifications by providing stringent control over gene expression [6,8,9,26,27,114,115]. Collectively, these proteins and enzymes represent an exciting suite of tools that can be customized for diverse genome engineering applications.

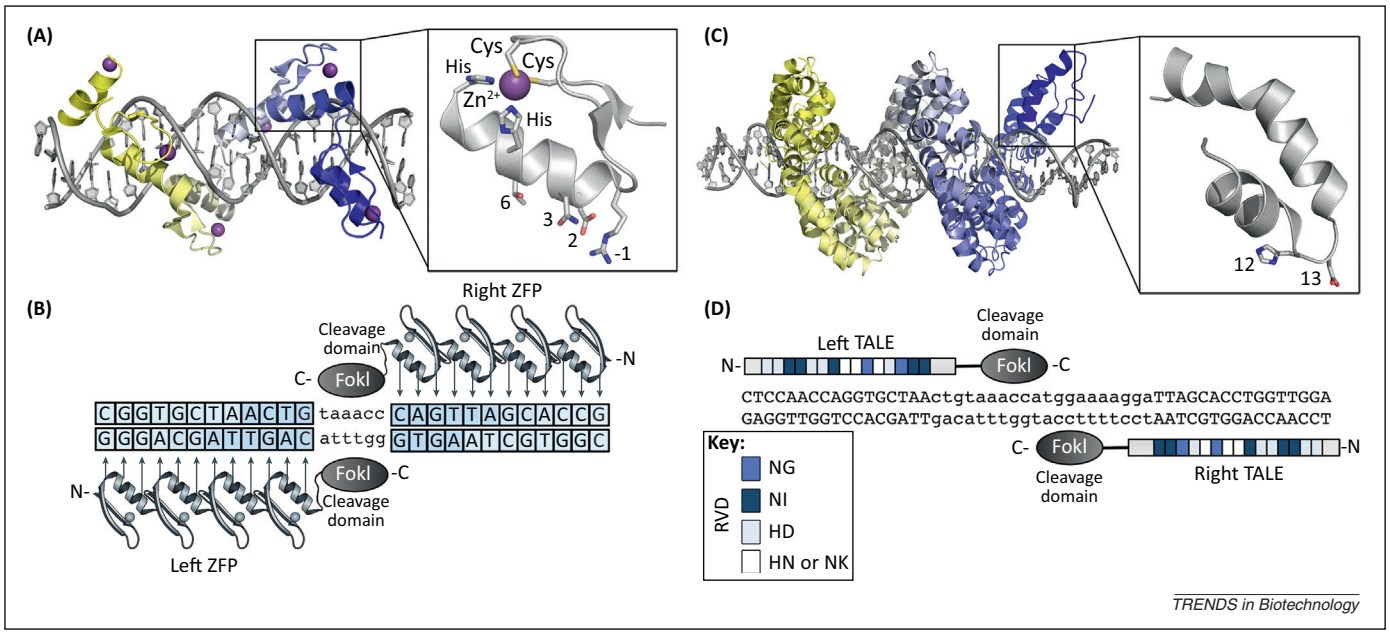


Figure 1. Structure of zinc-finger and transcription activator-like effectors (TALEs). **(A) (Top)** Designed zinc-finger protein in complex with target DNA (grey) (PDB ID: 2I13). Each zinc-finger consists of approximately 30 amino acids in an ββα arrangement (inset). Surface residues (-1, 2, 3 and 6) that contact DNA are shown as sticks. Each zinc-finger domain contacts 3 or 4 bp in the major groove of DNA. The side chains of the conserved Cys and His residues are depicted as sticks in complex with a Zn^{2+} ion (purple). **(B)** Cartoon of a zinc-finger nuclease (ZFN) dimer bound to DNA. ZFN target sites consist of two zinc-finger binding sites separated by a 5–7-bp spacer sequence recognized by the *FokI* cleavage domain. Zinc-finger proteins can be designed to recognize unique 'left' and 'right' half-sites. **(C) (Top)** TALE protein in complex with target DNA (grey) (PDB ID: 3UGM). Individual TALE repeats contain 33–35 amino acids that recognize a single base pair via two hypervariable residues (repeat-variable diresidues; RVDs) (shown as sticks) (inset). **(D)** Cartoon of a TALE nuclease (TALEN) dimer bound to DNA. TALEN target sites consist of two TALE binding sites separated by a spacer sequence of varying length (12–20 bp). TALEs can be designed to recognize unique left and right half-sites. RVD compositions are indicated.

text-dependency to assemble longer arrays by modular assembly [18,19]. For many years, zinc-finger protein technology was the only approach available to create custom site-specific DNA-binding proteins and enzymes. Engineered zinc fingers are also available commercially; Sangamo Biosciences (Richmond, CA, USA) has developed a propriety platform (CompoZr) for zinc-finger construction in partnership with Sigma–Aldrich (St. Louis, MO, USA), allowing investigators to bypass zinc-finger construction and validation altogether, and many thousands of proteins are already available. Broadly, zinc-finger protein technology enables targeting of virtually any sequence.

TALEs

The recent discovery of a simple modular DNA recognition code by TALE proteins [20,21] has led to the explosive expansion of an alternative platform for engineering programmable DNA-binding proteins. TALEs are naturally occurring proteins from the plant pathogenic bacteria genus *Xanthomonas*, and contain DNA-binding domains composed of a series of 33–35-amino-acid repeat domains that each recognizes a single base pair (Figure 1B). TALE specificity is determined by two hypervariable amino acids that are known as the repeat-variable di-residues (RVDs) [22,23]. Like zinc fingers, modular TALE repeats are linked together to recognize contiguous DNA sequences. However, in contrast to zinc-finger proteins, there was no re-engineering of the linkage between repeats necessary to construct long arrays of TALEs with the ability of targeting single sites in a genome. Following nearly two decades of pioneering work based on zinc-finger proteins, numerous effector domains have been made available to fuse to TALE repeats for targeted genetic modifications, including nucleases [24–26], transcriptional activators [26,27], and site-specific recombinases [28]. Although the single base recognition of TALE–DNA binding repeats affords greater design flexibility than triplet-confined zinc-finger proteins, the cloning of repeat TALE arrays presents an elevated technical challenge due to extensive identical repeat sequences. To overcome this issue, several methods have been developed that enable rapid assembly of custom TALE arrays. These strategies include 'Golden Gate' molecular cloning [29], high-throughput solid-phase assembly [30,31], and ligation-independent cloning techniques [32]. Several large, systematic studies utilizing various assembly methods have indicated that TALE repeats can be combined to recognize virtually any user-defined sequence [30,32]. The only targeting limitation for TALE arrays for which there is consensus in the literature is that TALE binding sites should start with a T base. Indeed, the ease with which TALE repeats can be assembled is evident in a recent study reporting the construction of a library of TALENs targeting 18 740 human protein-coding genes [33]; a technological feat that will not only facilitate numerous new studies, but will also encourage other, equally ambitious endeavors. Custom-designed TALE arrays are also commercially available through Cellectis Bioresearch (Paris, France), Transposagen Biopharmaceuticals (Lexington, KY, USA), and Life Technologies (Grand Island, NY, USA).

Genome editing with site-specific nucleases

Historically, targeted gene inactivation, replacement, or addition has been achieved by homologous recombination; however, the low efficiency of homologous recombination in mammalian cells and model organisms dramatically limits

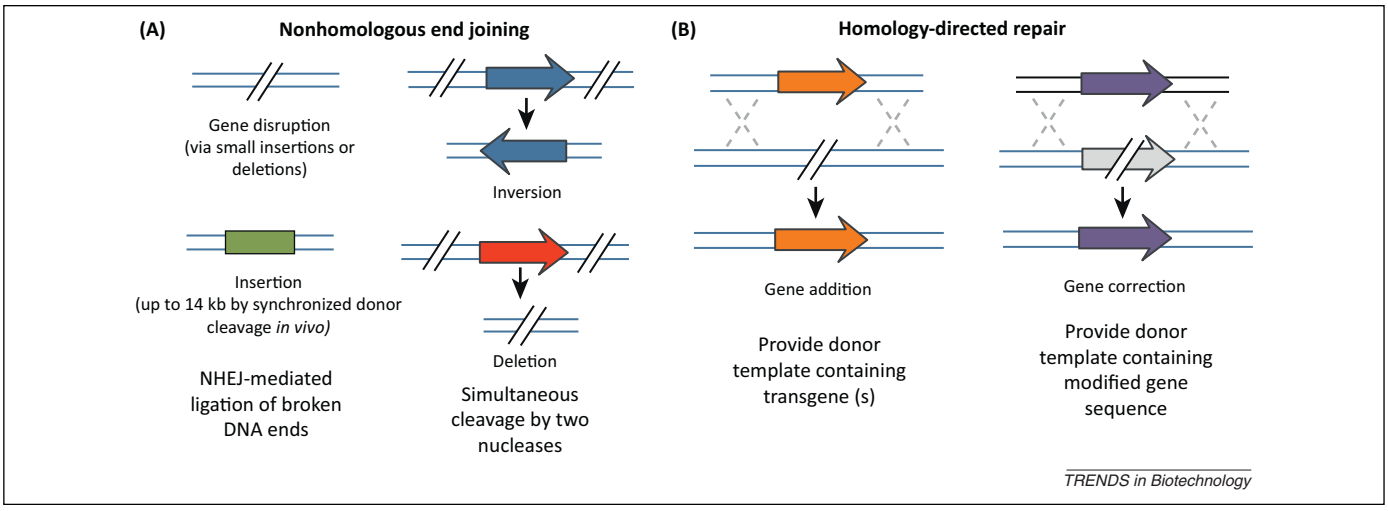


Figure 2. Overview of possible genome editing outcomes using site-specific nucleases. Nuclease-induced DNA double-strand breaks (DSBs) can be repaired by homology-directed repair (HDR) or error-prone nonhomologous end joining (NHEJ). **(A)** In the presence of donor plasmid with extended homology arms, HDR can lead to the introduction of single or multiple transgenes to correct or replace existing genes. **(B)** In the absence of donor plasmid, NHEJ-mediated repair yields small insertion or deletion mutations at the target that cause gene disruption. In the presence of double-stranded oligonucleotides or *in vivo* linearized donor plasmid, DNA fragments up to 14 kb have been inserted via NHEJ-mediated ligation. Simultaneous induction of two DSBs can lead to deletions, inversions and translocations of the intervening segment.

the utility of this approach. Following the discovery that induction of a DSB increases the frequency of HDR by several orders of magnitude, targeted nucleases have emerged as the method of choice for improving the efficiency of HDR-mediated genetic alterations. By co-delivering a site-specific nuclease with a donor plasmid bearing locus-specific homology arms [34], single or multiple transgenes can be efficiently integrated into an endogenous locus (Figure 2A). Linear donor sequences with <50 bp of homology [35], as well as single-stranded DNA oligonucleotides [36], can also be used to induce mutations, deletions, or insertions at the target site. Significantly, nuclease-mediated targeted integration normalizes for positional effects that typically confound many types of genetic analysis and enables study of structure–function relations in the complex and native chromosomal environment. In addition to their role in facilitating HDR, site-specific nucleases also allow rapid generation of cell lines and organisms with null phenotypes; NHEJ-mediated repair of a nuclease-induced DSB leads to the introduction of small insertions or deletions at the targeted site, resulting in knockout of gene function via frameshift mutations [37] (Figure 2B). Site-specific nucleases can also induce deletions of large chromosomal segments [38,39]. This method has been shown to support large chromosomal inversions [40] and translocations [41]. Finally, by synchronizing nuclease-mediated cleavage of donor DNA with the chromosomal target, large transgenes (up to 14 kb) have been introduced into various endogenous loci via NHEJ-mediated ligation [42,43]. Together, these approaches support the study of gene function and the modeling of disease states by altering genes to mimic both known and as yet uncharacterized genotypes. Many of these approaches have been extended to progenitor cell types, including embryonic stem (ES) cells [44] and induced pluripotent stem (iPS) cells [45,46], encouraging their further development for modeling a broad range of genetic conditions [47,48] (Table 1). Extension of this technology to study the role of noncoding DNA in the regulation and expression of coding genes can also be envisioned [49,50], including the use of multiplexed approaches as a means to identify unknown regulatory sites for genes of interest [51].

Improving the performance of site-specific nucleases

In order for customizable nucleases to carry relevance for genetic analysis and clinical application, they must demonstrate strict specificity toward their intended DNA targets. Complex genomes, however, often contain multiple copies of sequences that are identical or highly homologous to the intended DNA target, leading to off-target activity and cellular toxicity. To address this problem, structure [52,53] and selection-based [54,55] approaches have been used to generate improved ZFN and TALEN heterodimers with optimized cleavage specificity and reduced toxicity. Our laboratory has utilized directed evolution to generate a hyperactivated variant of the *Fok*I cleavage domain, *Sharkey*, that exhibits a >15-fold increase in cleavage activity in comparison to traditional ZFNs [55], and is directly compatible with various ZFN architectures [54]. Furthermore, there is mounting evidence to suggest that 4–6 zinc-finger domains for each ZFN half enzyme significantly enhances activity and specificity [13,55–57]. Additional methods for improving ZFN activity include the use of transient hypothermic culture conditions to increase nuclease expression levels [58], co-delivery of site-specific nucleases with DNA end-processing enzymes [59], and the use of fluorescent surrogate reporter vectors that allow for the enrichment of ZFN- and TALEN-modified cells [60]. The specificity of ZFN-mediated genome editing has been further refined by the development of zinc-finger nickases (ZFNickases) [61–63], which take advantage of the finding that induction of nicked DNA stimulates HDR [64] without activating the error-prone NHEJ repair pathway. Consequently, this approach leads to fewer off-target mutagenesis events than conventional DSB-induced methods for genome editing; however, the frequency of HDR by ZFNickases remains lower than those achieved with traditional ZFNs. Finally, conventional DNA- and

Table 1. Abbreviated list of examples of ZFN, TALEN, and CRISPR/Cas-mediated genome editing in human cells and model organisms

Type of modification	Organism	Genes	Nucleases	Refs
Gene disruption	Human	*CCR5*	ZFN	[65,91,92]
			TALEN	[25,52]
			CRISPR/Cas	[101]
	Human	TCR (T cell receptor)	ZFN	[94,95]
	Zebrafish	*gol* (Golden), *ntl* (No tail), *kra*	ZFN	[66,68]
	Pig	*GGTA1* (α1, 3-galactosyltransferase)	ZFN	[77]
		LDLR (LDL receptor)	TALEN	[76]
	Bovine	*ACAN12, p65*	TALEN	[76]
	Human	*EMX1, PVALB*	CRISPR/Cas	[102]
	Rat	*IgM, Rab38*	ZFN	[70]
	Arabidopsis	*ADH1, TT4*	ZFN	[81]
	C. elegans	*ben-1, rex-1, sdc-2*	ZFN/TALEN	[78]
	Hamster	*DHFR*	ZFN	[37]
	Drosophila	*yellow*	ZFN	[72]
	Rice	*OsSWEET14*	TALEN	[84]
Gene addition	Human	*OCT4, PITX3*	ZFN/TALEN	[45,46]
	Human	*CCR5*	ZFN	[97]
	Human	*F9* (Coagulation Factor IX)	ZFN	[86]
	Mouse	*Rosa26*	ZFN	[57]
	Human	AAVS1	ZFN	[45,96,97]
			TALEN	[46]
			CRISPR/Cas	[103]
	Human	*VEGF-A*	ZFN	[17]
	Zebrafish	*th* (tyrosine hydroxylase), *fam46c, smad5*	TALEN	[80]
	Maize	*IPK1*	ZFN	[82]
Gene correction	Human	*IL2RG*	ZFN	[44,85]
		A1AT (α_1-antitrypsin)	ZFN	[89]
		HBB (β-globin)	ZFN	[87,88]
		SNCA (α-synuclein)	ZFN	[90]
	Tobacco	*SuRA, SurRB* (acetolactate synthase)	ZFN	[83]
	Drosophila	*yellow*	ZFN	[71]

mRNA-based methods for delivering ZFNs into cells are restricted to certain cell types and are associated with undesirable side effects, including insertional mutagenesis, toxicity, and low efficiency (Box 2). To address these limitations, we recently developed a simple alternative based on the direct delivery of purified ZFN proteins into cells. This approach does not carry the risk of insertional mutagenesis and leads to comparatively fewer off-target effects than ZFN gene-delivery systems that rely on expression from nucleic acids [65]. This type of delivery platform thus may represent an optimal strategy for studies that require precise genome engineering in cells.

Site-specific nucleases in model organisms

Site-specific nucleases have enabled the introduction of targeted modifications in several model organisms common to biological research, including zebrafish [66–68], rats and mice [69,70], *Drosophila* [71,72], *Caenorhabditis elegans* [73], and many other species for various applications, including the monarch butterfly [74], frogs [75], and livestock [76,77]. ZFNs and TALENs have also allowed investigators to compare gene function across related species, such as *C. elegans* and *Caenorhabditis briggsae* [78], shedding light on the similarities and differences between closely related organisms and making analyses between orthologous gene pairs possible. By microinjecting single-cell embryos with TALEN mRNA and single-stranded DNA oligonucleotides [79] or donor plasmid with extended (>800 bp) homology arms [80], TALENs have achieved targeted integration in zebrafish, enabling the generation of loxP engineered chromosomes and the possibility for conditional gene activation in this model organism. In addition to valuable animal models, both ZFNs and TALENs have been used to introduce targeted alterations in plants, including *Arabidopsis* [81] and several crop species [82,83], allowing the incorporation of valuable traits, such as disease [84] and herbicide resistance [82,83]. The diversity of organisms modified by these site-specific nucleases will undoubtedly continue to grow, expanding the repertoire of model systems for basic research and knowledge of the intricacies and opportunities of genome biology.

Therapeutic applications of site-specific nucleases

The use of site-specific nucleases for therapeutic purposes represents a paradigm shift in gene therapy. Unlike conventional methods, which either temporarily address disease symptoms or randomly integrate therapeutic factors in the genome, ZFNs and TALENs are capable of correcting the underlying cause of the disease, therefore permanently eliminating the symptoms with precise genome modifications. To date, ZFN-induced HDR has been used to directly correct the disease-causing mutations associated with X-linked severe combined immune deficiency (SCID) [85],

Box 2. Methods for delivering site-specific nucleases into cells

Although site-specific nucleases provide a means for introducing diverse custom alterations at specific genomic locations, this technology is still limited by methods for delivering these enzymes into relevant cell types. Typically, nuclease-encoded genes are delivered into cells by plasmid DNA, viral vectors, or *in vitro* transcribed mRNA. The delivery method can be tailored to some degree toward the application or cell type of interest; however, the deficiencies of contemporary viral and nonviral gene delivery systems restrict the possible applications of site-specific nucleases. In particular, transfection of plasmid DNA or mRNA by electroporation or cationic lipid-based reagents can be toxic and restricted to certain cell types. Viral vectors also present limitations, because they are complex, difficult-to-produce, potentially immunogenic, and involve additional regulatory hurdles. Despite these difficulties, clinical trials based on adenovirus-mediated ZFN gene delivery into T lymphocytes are ongoing [91], however, future endeavors would benefit greatly from improved delivery methods.

Integrase-deficient lentiviral vectors (IDLVs) are an attractive alternative for delivering ZFNs into transfection-resistant cell types [44]; however, this method does not appear to be compatible with highly repetitive TALEN sequences [107]. Despite the apparent ease with which TALENs can be engineered, these enzymes may prove more difficult to deliver into cells than ZFNs. AAV is a promising vector for ZFN delivery that has been used to enhance the efficiency of ZFN-mediated HDR [108,119] and drive ZFN-mediated gene correction *in vivo* [86]. Efficient packaging of AAV occurs only for expression cassettes less than 4.2 kb in length. Although this is sufficient to accommodate both ZFN monomers and an engineered donor construct, only a single TALEN monomer with a minimal promoter sequence can be inserted into this vector.

As an alternative to ZFN gene-delivery systems, our group recently reported that purified ZFN proteins are capable of crossing cell membranes and inducing endogenous gene disruption [65]. This approach has several advantages over gene-based delivery methods. First, this approach reduces off-target activity by limiting the time that cells are exposed to ZFNs and thus minimizing opportunities for off-target activity. Second, this method circumvents the cell-type dependency and toxicity of viral and nonviral gene delivery systems. Third, this approach overcomes several safety and regulatory hurdles for developing ZFN-based therapies by allowing the knockout of human genes without exposing cells to any genetic material. It remains unknown whether purified TALEN proteins can also introduced into cells in the same manner.

hemophilia B [86], sickle-cell disease [87,88], and α_1-antitrypsin deficiency [89]. Moreover, ZFNs have been used to genetically repair Parkinson's disease-associated mutations within the SNCA gene in patient-derived human iPS cells [90]. Targeted gene knockout via ZFN-induced NHEJ-mediated repair has also proven a potentially powerful strategy for combating HIV/AIDs. ZFNs have been used to confer HIV-1 resistance by disabling the HIV coreceptor C-C chemokine receptor type 5 (CCR5) in primary T cells [91] and hematopoietic stem/progenitor cells [92]. This approach is currently in clinical trials (NCT01252641, NCT00842634 and NCT01044654). More recently, ZFN-mediated targeted integration of anti-HIV restriction factors into the *CCR5* locus has led to the establishment of T cells that show near-complete protection from both R5- and X4-tropic strains of HIV [93]. Additionally, ZFNs have been used to improve the performance of T-cell-based immunotherapies by inactivating the expression of endogenous T cell receptor genes [94,95], thereby enabling the generation of tumor-specific T cells with improved efficacy profiles. Finally, site-specific nucleases afford the unique possibility of safely inserting therapeutic transgenes into specific 'safe harbor' locations in the human genome [96,97]. Although the overall utility of site-specific nucleases is currently limited in somatic cells, continued progress in stem cell research, including the production and manipulation of iPS cells, will ultimately open countless new directions for gene therapy, including treatments based on autologous stem cell transplantation.

Genome editing using programmable RNA-guided DNA endonucleases

Distinct from the site-specific nucleases described above, the CRISPR/Cas system has recently emerged as a potentially facile and efficient alternative to ZFNs and TALENs for inducing targeted genetic alterations. In bacteria, the CRISPR system provides acquired immunity against invading foreign DNA via RNA-guided DNA cleavage [98]. In the type II CRISPR/Cas system, short segments of foreign DNA, termed 'spacers' are integrated within the CRISPR genomic loci and transcribed and processed into short CRISPR RNA (crRNA). These crRNAs anneal to trans-activating crRNAs (tracrRNAs) and direct sequence-specific cleavage and silencing of pathogenic DNA by Cas proteins. Recent work has shown that target recognition by the Cas9 protein requires a 'seed' sequence within the crRNA and a conserved dinucleotide-containing protospacer adjacent motif (PAM) sequence upstream of the crRNA-binding region [99]. The CRISPR/Cas system can thereby be retargeted to cleave virtually any DNA sequence by redesigning the crRNA. Significantly, the CRISPR/Cas system has been shown to be directly portable to human cells by co-delivery of plasmids expressing the Cas9 endonuclease and the necessary crRNA components [100–103]. These programmable RNA-guided DNA endonucleases have demonstrated multiplexed gene disruption capabilities [102] and targeted integration in iPS cells [103]. Cas9 endonucleases have also been converted into nickases [102], enabling an additional level of control over the mechanism of DNA repair. In addition to human cells, CRISPR/Cas-mediated genome editing has been successfully demonstrated in zebrafish [104] and bacterial cells [105]; however, more exhaustive studies are required in order to thoroughly evaluate the utility of this system, including the potential for off-target effects. In particular, it remains unclear whether CRISPR/Cas system affords the requisite recognition selectivity necessary to ensure single-site specificity in complex genomes.

Concluding remarks and future directions

ZFNs, TALENs, and RNA-guided DNA endonucleases are transformative tools that have the potential to revolutionize biological research and affect personalized medicine. Indeed, these emerging technologies have dramatically expanded the ability to manipulate and study model organisms, and support the promise of correcting the genetic causes behind many diseases. However, in order to achieve the full potential of this technology, many important questions and challenges must be addressed (Box 3). Chief among these is the relative specificity of each nuclease platform. In the future, the use of high-throughput methods that enable comprehensive profiling of off-target cleavage sites [106] should provide insight into the stringency of

Box 3. Outstanding questions

- How effective are ZFNs and TALENs as therapeutic agents?
- What are the best methods for delivering site-specific nucleases into cells, and how can TALENs be delivered into cells by lentivirus?
- Can the Cas9 endonuclease be co-opted as a DNA-binding domain and be fused to enzymatic domains?
- How specific and safe are CRISPR/Cas9 systems, and how does the efficiency of Cas9-mediated genome editing compare to ZFN and TALEN-based approaches?
- What is the optimal RNA scaffold for application of CRISPR/Cas9 in mammalian cells?

target recognition inherent in each system. Questions also remain regarding the optimal methods for delivering these nucleases into cells and organisms. In particular, although adenoviral vectors can accommodate and deliver full-length TALEN genes into human cells, lentiviral plasmid vectors harboring TALEN sequences are prone to rearrangements after transduction [107]. Furthermore, the large size of TALENs may limit their delivery by size-restricted vectors such as recombinant adeno-associated virus (AAV), which has been shown to accommodate ZFN genes [108]. These findings suggest that the development of new TALEN delivery systems will be a critical area of future research. Although CRISPR/Cas systems show great promise and flexibility for genetic engineering, sequence requirements within the PAM sequence may constrain some applications. Directed evolution of the Cas9 protein should offer a path toward PAM independence, and may also provide a means to generate an even more efficient Cas9 endonuclease. Additional studies will also be required to evaluate the specificity and toxicity of RNA-guided DNA endonucleases *in vitro* and *in vivo*. Finally, the continued development of conditional methods that rely on customizable recombinases [109–111] and transcription factors [6,9,112–115] for affecting genomic structure and function will complement existing and future nuclease technologies. Together, these technologies promise to expand our ability to explore and alter any genome and constitute a new and promising paradigm to understand and treat disease.

Acknowledgments

The authors are supported by the National Institutes of Health (Pioneer Award DP1CA174426 (CB) and DP2OD008586 (CG) and National Science Foundation (CBET-1151035). T.G. was supported by National Institute of General Medicine Sciences fellowship (T32GM080209). We apologize to those investigators whose important contributions may have been omitted due to space constraints.

References

1 Capecchi, M.R. (2005) Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat. Rev. Genet.* 6, 507–512
2 McManus, M.T. and Sharp, P.A. (2002) Gene silencing in mammals by small interfering RNAs. *Nat. Rev. Genet.* 3, 737–747
3 Urnov, F.D. *et al.* (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* 11, 636–646
4 Carroll, D. (2011) Genome engineering with zinc-finger nucleases. *Genetics* 188, 773–782
5 Wyman, C. and Kanaar, R. (2006) DNA double-strand break repair: all's well that ends well. *Annu. Rev. Genet.* 40, 363–383
6 Beerli, R.R. and Barbas, C.F., III (2002) Engineering polydactyl zinc-finger transcription factors. *Nat. Biotechnol.* 20, 135–141
7 Liu, Q. *et al.* (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc. Natl. Acad. Sci. U.S.A.* 94, 5525–5530
8 Beerli, R.R. *et al.* (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14628–14633
9 Beerli, R.R. *et al.* (2000) Positive and negative regulation of endogenous genes by designed transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* 97, 1495–1500
10 Kim, J.S. and Pabo, C.O. (1998) Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc. Natl. Acad. Sci. U.S.A.* 95, 2812–2817
11 Segal, D.J. *et al.* (2006) Structure of Aart, a designed six-finger zinc finger peptide, bound to DNA. *J. Mol. Biol.* 363, 405–421
12 Neuteboom, L.W. *et al.* (2006) Effects of different zinc finger transcription factors on genomic targets. *Biochem. Biophys. Res. Commun.* 339, 263–270
13 Bhakta, M.S. *et al.* (2013) Highly active zinc-finger nucleases by extended modular assembly. *Genome Res.* 23, 530–538
14 Kim, S. *et al.* (2011) Preassembled zinc-finger arrays for rapid construction of ZFNs. *Nat. Methods* 8, 7
15 Gonzalez, B. *et al.* (2010) Modular system for the construction of zinc-finger libraries and proteins. *Nat. Protoc.* 5, 791–810
16 Segal, D.J. *et al.* (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences. *Proc. Natl. Acad. Sci. U.S.A.* 96, 2758–2763
17 Maeder, M.L. *et al.* (2008) Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell* 31, 294–301
18 Sander, J.D. *et al.* (2011) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat. Methods* 8, 67–69
19 Gupta, A. *et al.* (2012) An optimized two-finger archive for ZFN-mediated gene targeting. *Nat. Methods* 9, 588–590
20 Boch, J. *et al.* (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509–1512
21 Moscou, M.J. and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science* 326, 1501
22 Mak, A.N. *et al.* (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 335, 716–719
23 Deng, D. *et al.* (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 335, 720–723
24 Christian, M. *et al.* (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186, 757–761
25 Mussolino, C. *et al.* (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res.* 39, 9283–9293
26 Miller, J.C. *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* 29, 143–148
27 Zhang, F. *et al.* (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.* 29, 149–153
28 Mercer, A.C. *et al.* (2012) Chimeric TALE recombinases with programmable DNA sequence specificity. *Nucleic Acids Res.* 40, 11163–11172
29 Cermak, T. *et al.* (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* 39, e82
30 Reyon, D. *et al.* (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.* 30, 460–465
31 Briggs, A.W. *et al.* (2012) Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res.* 40, e117
32 Schmid-Burgk, J.L. *et al.* (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat. Biotechnol.* 31, 76–81

33 Kim, Y. *et al.* (2013) A library of TAL effector nucleases spanning the human genome. *Nat. Biotechnol.* 31, 251–258
34 Moehle, E.A. *et al.* (2007) Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc. Natl. Acad. Sci. U.S.A.* 104, 3055–3060
35 Orlando, S.J. *et al.* (2010) Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic Acids Res.* 38, e152
36 Chen, F. *et al.* (2011) High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat. Methods* 8, 753–755
37 Santiago, Y. *et al.* (2008) Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5809–5814
38 Lee, H.J. *et al.* (2010) Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res.* 20, 81–89
39 Sollu, C. *et al.* (2010) Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic Acids Res.* 38, 8269–8276
40 Lee, H.J. *et al.* (2012) Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Res.* 22, 539–548
41 Brunet, E. *et al.* (2009) Chromosomal translocations induced at specified loci in human stem cells. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10620–10625
42 Cristea, S. *et al.* (2013) In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. *Biotechnol. Bioeng.* 110, 871–880
43 Maresca, M. *et al.* (2013) Obligate Ligation-Gated Recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Res.* 23, 539–546
44 Lombardo, A. *et al.* (2007) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat. Biotechnol.* 25, 1298–1306
45 Hockemeyer, D. *et al.* (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.* 27, 851–857
46 Hockemeyer, D. *et al.* (2011) Genetic engineering of human pluripotent cells using TALE nucleases. *Nat. Biotechnol.* 29, 731–734
47 Ding, Q. *et al.* (2013) A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem Cell* 12, 238–251
48 Zou, J. *et al.* (2009) Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* 5, 97–110
49 Sanyal, A. *et al.* (2012) The long-range interaction landscape of gene promoters. *Nature* 489, 109–113
50 Gutschner, T. *et al.* (2011) Noncoding RNA gene silencing through genomic integration of RNA destabilizing elements using zinc finger nucleases. *Genome Res.* 21, 1944–1954
51 Dunham, I. *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74
52 Miller, J.C. *et al.* (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* 25, 778–785
53 Szczepek, M. *et al.* (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat. Biotechnol.* 25, 786–793
54 Doyon, Y. *et al.* (2011) Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat. Methods* 8, 74–79
55 Guo, J. *et al.* (2010) Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *J. Mol. Biol.* 400, 96–107
56 Sood, R. *et al.* (2013) Efficient methods for targeted mutagenesis in zebrafish using zinc-finger nucleases: data from targeting of nine genes using CompoZr or CoDA ZFNs. *PLoS ONE* 8, e57239
57 Perez-Pinera, P. *et al.* (2012) Gene targeting to the ROSA26 locus directed by engineered zinc finger nucleases. *Nucleic Acids Res.* 40, 3741–3752
58 Doyon, Y. *et al.* (2010) Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. *Nat. Methods* 7, 459–460
59 Certo, M.T. *et al.* (2012) Coupling endonucleases with DNA end-processing enzymes to drive gene disruption. *Nat. Methods* 9, 973–975
60 Kim, H. *et al.* (2011) Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat. Methods* 8, 941–943
61 Kim, E. *et al.* (2012) Precision genome engineering with programmable DNA-nicking enzymes. *Genome Res.* 22, 1327–1333
62 Wang, J. *et al.* (2012) Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. *Genome Res.* 22, 1316–1326
63 Ramirez, C.L. *et al.* (2012) Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res.* 40, 5560–5568
64 McConnell Smith, A. *et al.* (2009) Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5099–5104
65 Gaj, T. *et al.* (2012) Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nat. Methods* 9, 805–807
66 Doyon, Y. *et al.* (2008) Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat. Biotechnol.* 26, 702–708
67 Sander, J.D. *et al.* (2011) Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat. Biotechnol.* 29, 697–698
68 Meng, X. *et al.* (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat. Biotechnol.* 26, 695–701
69 Tesson, L. *et al.* (2011) Knockout rats generated by embryo microinjection of TALENs. *Nat. Biotechnol.* 29, 695–696
70 Geurts, A.M. *et al.* (2009) Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* 325, 433
71 Bibikova, M. *et al.* (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science* 300, 764
72 Bibikova, M. *et al.* (2002) Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* 161, 1169–1175
73 Morton, J. *et al.* (2006) Induction and repair of zinc-finger nuclease-targeted double-strand breaks in *Caenorhabditis elegans* somatic cells. *Proc. Natl. Acad. Sci. U.S.A.* 103, 16370–16375
74 Merlin, C. *et al.* (2013) Efficient targeted mutagenesis in the monarch butterfly using zinc-finger nucleases. *Genome Res.* 23, 159–168
75 Young, J.J. *et al.* (2011) Efficient targeted gene disruption in the soma and germ line of the frog *Xenopus tropicalis* using engineered zinc-finger nucleases. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7052–7057
76 Carlson, D.F. *et al.* (2012) Efficient TALEN-mediated gene knockout in livestock. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17382–17387
77 Hauschild, J. *et al.* (2011) Efficient generation of a biallelic knockout in pigs using zinc-finger nucleases. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12013–12017
78 Wood, A.J. *et al.* (2011) Targeted genome editing across species using ZFNs and TALENs. *Science* 333, 307
79 Bedell, V.M. *et al.* (2012) In vivo genome editing using a high-efficiency TALEN system. *Nature* 491, 114–118
80 Zu, Y. *et al.* (2013) TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nat. Methods* 10, 329–331
81 Zhang, F. *et al.* (2010) High frequency targeted mutagenesis in *Arabidopsis thaliana* using zinc finger nucleases. *Proc. Natl. Acad. Sci. U.S.A.* 107, 12028–12033
82 Shukla, V.K. *et al.* (2009) Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* 459, 437–441
83 Townsend, J.A. *et al.* (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* 459, 442–445
84 Li, T. *et al.* (2012) High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat. Biotechnol.* 30, 390–392
85 Urnov, F.D. *et al.* (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 435, 646–651
86 Li, H. *et al.* (2011) In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature* 475, 217–221
87 Zou, J. *et al.* (2011) Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. *Blood* 118, 4599–4608
88 Sebastiano, V. *et al.* (2011) In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. *Stem Cells* 29, 1717–1726
89 Yusa, K. *et al.* (2011) Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* 478, 391–394
90 Soldner, F. *et al.* (2011) Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell* 146, 318–331

91 Perez, E.E. *et al.* (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat. Biotechnol.* 26, 808–816
92 Holt, N. *et al.* (2010) Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nat. Biotechnol.* 28, 839–847
93 Voit, R.A. *et al.* (2013) Generation of an HIV resistant T-cell line by targeted "stacking" of restriction factors. *Mol. Ther.* 21, 786–795
94 Torikai, H. *et al.* (2012) A foundation for universal T-cell based immunotherapy: T cells engineered to express a CD19-specific chimeric-antigen-receptor and eliminate expression of endogenous TCR. *Blood* 119, 5697–5705
95 Provasi, E. *et al.* (2012) Editing T cell specificity towards leukemia by zinc finger nucleases and lentiviral gene transfer. *Nat. Med.* 18, 807–815
96 DeKelver, R.C. *et al.* (2010) Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Res.* 20, 1133–1142
97 Lombardo, A. *et al.* (2011) Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat. Methods* 8, 861–869
98 Wiedenheft, B. *et al.* (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482, 331–338
99 Jinek, M. *et al.* (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821
100 Jinek, M. *et al.* (2013) RNA-programmed genome editing in human cells. *eLife* 2, e00471
101 Cho, S.W. *et al.* (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* 31, 230–232
102 Cong, L. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823
103 Mali, P. *et al.* (2013) RNA-guided human genome engineering via Cas9. *Science* 339, 823–826
104 Hwang, W.Y. *et al.* (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 227–229
105 Jiang, W. *et al.* (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* 31, 233–239
106 Gabriel, R. *et al.* (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.* 29, 816–823
107 Holkers, M. *et al.* (2013) Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res.* 41, e63
108 Ellis, B.L. *et al.* (2013) Zinc-finger nuclease-mediated gene correction using single AAV vector transduction and enhancement by Food and Drug Administration-approved drugs. *Gene Ther.* 20, 35–42
109 Gaj, T. *et al.* (2013) A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. *Nucleic Acids Res.* 41, 3937–3946
110 Gersbach, C.A. *et al.* (2011) Targeted plasmid integration into the human genome by an engineered zinc-finger recombinase. *Nucleic Acids Res.* 39, 7868–7878
111 Gaj, T. *et al.* (2011) Structure-guided reprogramming of serine recombinase DNA sequence specificity. *Proc. Natl. Acad. Sci. U.S.A.* 108, 498–503
112 Beerli, R.R. *et al.* (2000) Chemically regulated zinc finger transcription factors. *J. Biol. Chem.* 275, 32617–32627
113 Polstein, L.R. and Gersbach, C.A. (2012) Light-inducible spatiotemporal control of gene activation by customizable zinc finger transcription factors. *J. Am. Chem. Soc.* 134, 16480–16483
114 Perez-Pinera, P. *et al.* (2013) Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat. Methods* 10, 239–242
115 Maeder, M.L. *et al.* (2013) Robust, synergistic regulation of human gene expression using TALE activators. *Nat. Methods* 10, 243–245
116 Yant, S.R. *et al.* (2007) Site-directed transposon integration in human cells. *Nucleic Acids Res.* 35, e50
117 Gersbach, C.A. *et al.* (2010) Directed evolution of recombinase specificity by split gene reassembly. *Nucleic Acids Res.* 38, 4198–4206
118 Owens, J.B. *et al.* (2012) Chimeric piggyBac transposases for genomic targeting in human cells. *Nucleic Acids Res.* 40, 6978–6991
119 Handel, E.M. *et al.* (2012) Versatile and efficient genome editing in human cells by combining zinc-finger nucleases with adeno-associated viral vectors. *Hum. Gene Ther.* 23, 321–329

nature biotechnology

REVIEW

CRISPR-Cas systems for editing, regulating and targeting genomes

Jeffry D Sander[1,2] & J Keith Joung[1,2]

Targeted genome editing using engineered nucleases has rapidly gone from being a niche technology to a mainstream method used by many biological researchers. This widespread adoption has been largely fueled by the emergence of the clustered, regularly interspaced, short palindromic repeat (CRISPR) technology, an important new approach for generating RNA-guided nucleases, such as Cas9, with customizable specificities. Genome editing mediated by these nucleases has been used to rapidly, easily and efficiently modify endogenous genes in a wide variety of biomedically important cell types and in organisms that have traditionally been challenging to manipulate genetically. Furthermore, a modified version of the CRISPR-Cas9 system has been developed to recruit heterologous domains that can regulate endogenous gene expression or label specific genomic loci in living cells. Although the genome-wide specificities of CRISPR-Cas9 systems remain to be fully defined, the power of these systems to perform targeted, highly efficient alterations of genome sequence and gene expression will undoubtedly transform biological research and spur the development of novel molecular therapeutics for human disease.

The introduction of targeted genomic sequence changes into living cells and organisms has become a powerful tool for biological research and is a potential avenue for therapy of genetic diseases. Frameshift knockout mutations enable reverse genetics and identification of gene functions; sequence insertions can fuse epitope tags or other functional domains, such as fluorescent proteins, to endogenous gene products; and specific sequence alterations can induce amino acid substitutions for disease modeling, transfer traits in agricultural crops and livestock, and correct defective genes for therapeutic applications. For many years, strategies for efficiently inducing precise, targeted genome alterations were limited to certain organisms (e.g., homologous recombination in yeast or recombineering in mice) and often required drug-selectable markers or left behind 'scar' sequences associated with the modification method (e.g., residual *loxP* sites from Cre recombinase-mediated excision). Targeted genome editing using customized nucleases provides a general method for inducing targeted deletions, insertions and precise sequence changes in a broad range of organisms and cell types. The high efficiency of genome editing obviates the need for additional sequences, such as drug-resistance marker genes, and therefore the need for additional manipulations to remove them.

A crucial first step for performing targeted genome editing is the creation of a DNA double-stranded break (DSB) at the genomic locus to be modified[1]. Nuclease-induced DSBs can be repaired by one of at least two different pathways that operate in nearly all cell types and organisms: nonhomologous end-joining (NHEJ) and homology-directed repair (HDR) (**Fig. 1**). NHEJ can lead to the efficient introduction of insertion/deletion mutations (indels) of various lengths, which can disrupt the translational reading frame of a coding sequence or the binding sites of *trans*-acting factors in promoters or enhancers. HDR-mediated repair can be used to introduce specific point mutations or to insert desired sequences through recombination of the target locus with exogenously supplied DNA 'donor templates'. With targeted nuclease-induced DSBs, the frequencies of these alterations are typically greater than 1% and, in some cases, over 50%; at these rates, desired mutations can be identified by simple screening, without drug-resistance marker selection.

Early methods for targeting DSB-inducing nucleases to specific genomic sites relied on protein-based systems with customizable DNA-binding specificities, such as meganucleases, zinc finger nucleases (ZFNs) and transcription activator–like effector nucleases (TALENs). These platforms have made possible important advances, but each has its own set of associated advantages and disadvantages (**Box 1**). More recently, a platform based on a bacterial CRISPR-associated protein 9 nuclease from *Streptococcus pyogenes* (hereafter referred to as Cas9) has been developed; it is unique and flexible owing to its dependence on RNA as the moiety that targets the nuclease to a desired DNA sequence. In contrast to ZFN and TALEN methods, which use protein-DNA interactions for targeting, RNA-guided nucleases (RGNs) use simple, base-pairing rules between an engineered RNA and the target DNA site.

In this Review, we describe how this RNA-guided system works and how it has been applied to perform genome editing across a wide variety of cell types and whole organisms. We also discuss the advantages and limitations of this system, and assess off-target effects and recent strategies for improving specificity and how the system can be repurposed for other applications, such as regulation of gene expression and selective labeling of the genome (**Fig. 2** summarizes different applications). Finally, we consider the challenges that will need to be addressed for this emerging genome editing platform.

[1]Molecular Pathology Unit, Center for Computational and Integrative Biology, Center for Cancer Research, Massachusetts General Hospital, Charlestown, Massachusetts, USA. [2]Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA. Correspondence should be addressed to J.K.J. (jjoung@mgh.harvard.edu).

Received 2 September 2013; accepted 31 January 2014; published online 2 March 2014; doi:10.1038/nbt.2842

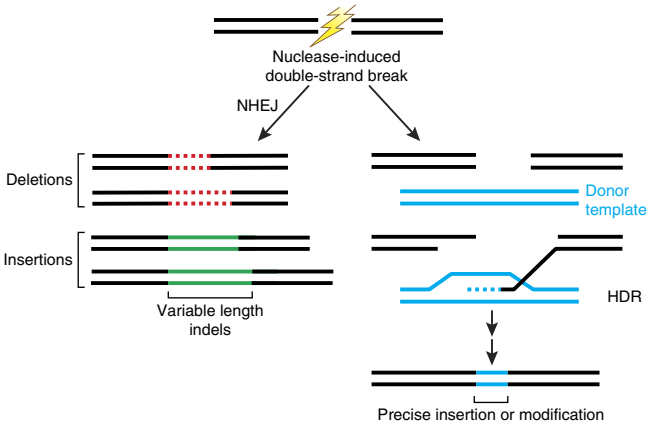


Figure 1 Nuclease-induced genome editing. Nuclease-induced double-strand breaks (DSBs) can be repaired by nonhomologous end joining (NHEJ) or homology-directed repair (HDR) pathways. Imprecise NHEJ-mediated repair can produce insertion and/or deletion mutations of variable length at the site of the DSB. HDR-mediated repair can introduce precise point mutations or insertions from a single-stranded or double-stranded DNA donor template.

From a bacterial CRISPR immune system to engineered RGNs

CRISPR systems are adaptable immune mechanisms used by many bacteria to protect themselves from foreign nucleic acids, such as viruses or plasmids[2–5]. Type II CRISPR systems incorporate sequences from invading DNA between CRISPR repeat sequences encoded as arrays within the bacterial host genome (**Fig. 3a**). Transcripts from the CRISPR repeat arrays are processed into CRISPR RNAs (crRNAs), each harboring a variable sequence transcribed from the invading DNA, known as the "protospacer" sequence, and part of the CRISPR repeat. Each crRNA hybridizes with a second RNA, known as the transactivating CRISPR RNA (tracrRNA)[6], and these two RNAs complex with the Cas9 nuclease[7]. The protospacer-encoded portion of the crRNA directs Cas9 to cleave complementary target-DNA sequences, if they are adjacent to short sequences known as protospacer adjacent motifs (PAMs). Protospacer sequences incorporated into the CRISPR locus are not cleaved presumably because they are not next to a PAM sequence.

The type II CRISPR system from *S. pyogenes* has been adapted for inducing sequence-specific DSBs and targeted genome editing[7]. In the simplest and most widely used form of this system, two components must be introduced into and/or expressed in cells or an organism to perform genome editing: the Cas9 nuclease and a guide RNA (gRNA), consisting of a fusion of a crRNA and a fixed tracrRNA (**Fig. 3b**). Twenty nucleotides at the 5′ end of the gRNA (corresponding to the protospacer portion of the crRNA; **Fig. 3c**) direct Cas9 to a specific target DNA site using standard RNA-DNA complementarity base-pairing rules. These target sites must lie immediately 5′ of a PAM sequence that matches the canonical form 5′-NGG (although recognition at sites with alternate PAM sequences (e.g., 5′-NAG) has also been reported, albeit at less efficient rates[7–9]). Thus, with this system, Cas9 nuclease activity can be directed to any DNA sequence of the form N_{20}-NGG simply by altering the first 20 nt of the gRNA to correspond to the target DNA sequence. Type II CRISPR systems from other species of bacteria that recognize alternative PAM sequences and that utilize different crRNA and tracrRNA sequences have also been used for targeted genome editing[10–12]. However, because the most commonly used and extensively characterized system is based on the *S. pyogenes* system, the remainder of this Review focuses on this particular platform and its components, unless otherwise noted.

Following the initial demonstrations in 2012 that Cas9 could be programmed to cut various DNA sites *in vitro*[7], a flurry of papers published in 2013 showed that this platform also functions efficiently in a variety of cells and organisms. Initial proof-of-principle studies showed that Cas9 could be targeted to endogenous genes in bacteria[8], cultured transformed human cancer cell lines and human pluripotent stem cells in culture[13–16], as well as in a whole organism, the zebrafish (J.K.J. and colleagues[17]). Subsequently, Cas9 has been used to alter

Box 1 Meganucleases, ZFNs and TALENs

Meganucleases, ZFNs and TALENs have been used extensively for genome editing in a variety of different cell types and organisms. Meganucleases are engineered versions of naturally occurring restriction enzymes that typically have extended DNA recognition sequences (e.g., 14–40 bp). ZFNs and TALENs are artificial fusion proteins composed of an engineered DNA binding domain fused to a nonspecific nuclease domain from the FokI restriction enzyme. Zinc finger and TALE repeat domains with customized specificities can be joined together into arrays that bind to extended DNA sequences.

The engineering of meganucleases has been challenging for most academic researchers because the DNA recognition and cleavage functions of these enzymes are intertwined in a single domain[86,87]. By contrast, the DNA binding domains of ZFNs and TALENs are distinct from the FokI cleavage domain[88], thereby making it more straightforward to modify the DNA-binding specificities of these nucleases. However, robust construction of engineered zinc finger arrays has also proven to be difficult for many laboratories because of the need to account for context-dependent effects between individual finger domains in an array[89]. Despite the availability of various publicly available methods designed to simplify the challenge of creating ZFNs[90–96], (J.K.J. and colleagues[92,93,95]), these nucleases have not been engineered by a wide range of laboratories.

In contrast to zinc fingers, TALE repeat domains seem to have fewer context-dependent effects and can be assembled robustly in a modular fashion to recognize virtually any DNA sequence (J.K.J. and colleagues[97]), using a simple one-to-one code between individual repeats and the four possible DNA nucleotides[98,99]. Although TALE repeat domains are much simpler to design than meganucleases or ZFNs, the assembly of DNA molecules encoding large numbers of highly conserved TALE repeats can require the use of nonstandard molecular biology cloning methods. Many user-friendly methods for making such assemblies have been described in the literature (J.K.J. and colleagues[100]) but the highly repetitive nature of TALEN-coding sequences also creates barriers to their delivery using certain viral vectors, such as lentiviruses[101]. Nonetheless, the greater simplicity of TALENs relative to meganucleases and ZFNs has led to their adoption over the past several years by a broad range of scientists. The question of whether to utilize these platforms for a given application must be considered on a case-by-case basis, and we refer the reader to recent reviews on these different technologies for additional information[86,88,100].

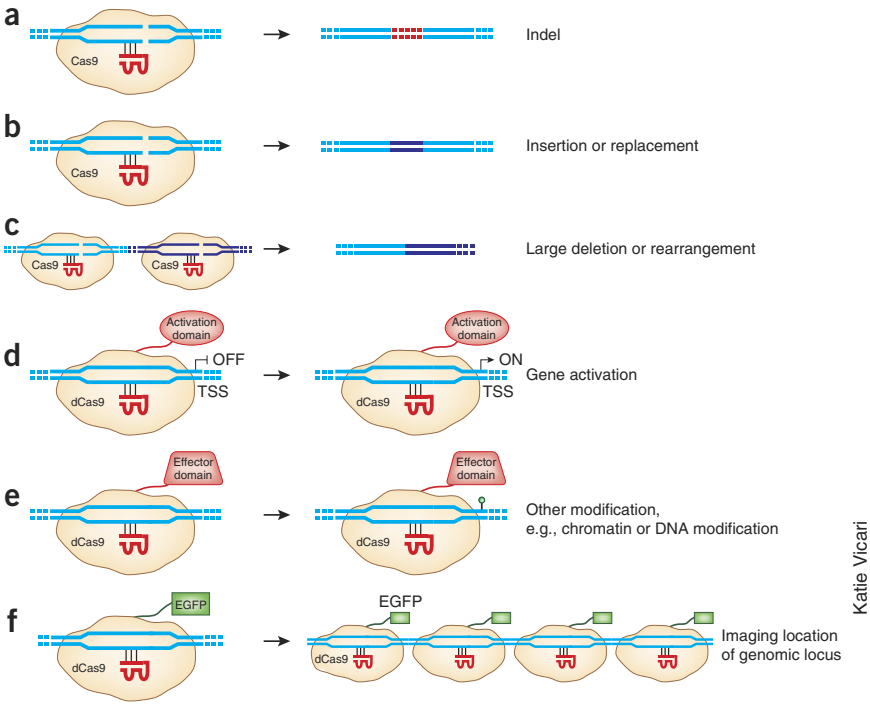


Figure 2 Overview of various Cas9-based applications. (**a**,**b**) gRNA-directed Cas9 nuclease can induce indel mutations (**a**) or specific sequence replacement or insertion (**b**). (**c**) Pairs of gRNA-directed Cas9 nucleases can stimulate large deletions or genomic rearrangements (e.g., inversions or translocations). (**d**–**f**) gRNA-directed dCas9 can be fused to activation domains (**d**) to mediate upregulation of specific endogenous genes, heterologous effector domains (**e**) to alter histone modifications or DNA methylation, or fluorescent proteins (**f**) to enable imaging of specific genomic loci. TSS, transcription start site.

genes in yeast[18], tobacco[19,20], thale cress[19], rice[21,22], wheat[21], sorghum[23], mice[24,25], rats[26], rabbits[27], frogs[28], fruit flies[29,30], silkworms[31] and roundworms[32] (see **Table 1** for a list of these published reports). Cas9-induced DSBs have been used to introduce NHEJ-mediated indel mutations as well as to stimulate HDR with both double-stranded plasmid DNA and single-stranded oligonucleotide donor templates. Being able to introduce DSBs at multiple sites in parallel with Cas9 is a unique advantage of this RNA-guided genome editing platform relative to meganucleases, ZFNs or TALENs. For example, expression of Cas9 and multiple gRNAs has been used to induce small and large deletions or inversions between the DSBs[14,33–35], to simultaneously introduce mutations in three genes in rat cells[36], five genes in mouse ES cells[24] and five genes in the somatic cells of a single zebrafish[37].

The simplicity of Cas9 targeting has also inspired the generation of large gRNA libraries using array-based oligonucleotide synthesis. These libraries can be engineered to encompass multiple gRNAs for almost every gene in a host organism, thereby facilitating forward genetic screens and selection. In contrast to short hairpin RNA libraries, which mediate only gene knockdown, these gRNA libraries have been used with Cas9 nuclease to generate libraries of cells with knockout mutations. Libraries consisting of between ~64,000 and ~87,000 distinct gRNAs have demonstrated this strategy for positive and negative forward genetic phenotype screens in human and mouse cells[38–40].

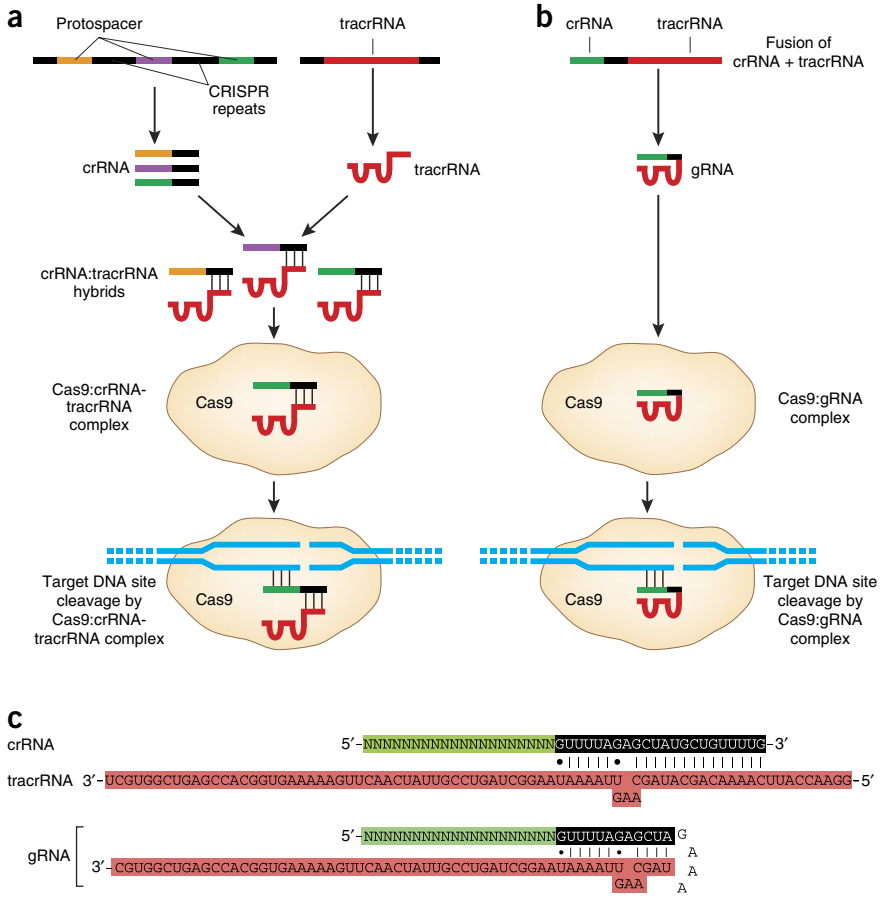


Figure 3 Naturally occurring and engineered CRISPR-Cas systems. (**a**) Naturally occurring CRISPR systems incorporate foreign DNA sequences into CRISPR arrays, which then produce crRNAs bearing "protospacer" regions that are complementary to the foreign DNA site. crRNAs hybridize to tracrRNAs (also encoded by the CRISPR system) and this pair of RNAs can associate with the Cas9 nuclease. crRNA-tracrRNA:Cas9 complexes recognize and cleave foreign DNAs bearing the protospacer sequences. (**b**) The most widely used engineered CRISPR-Cas system utilizes a fusion between a crRNA and part of the tracrRNA sequence. This single gRNA complexes with Cas9 to mediate cleavage of target DNA sites that are complementary to the 5′ 20 nt of the gRNA and that lie next to a PAM sequence. (**c**) Example sequences of a crRNA-tracrRNA hybrid and a gRNA.

Table 1 Published examples of cell types and organisms modified by Cas9

Cell type or organism	Cas9 form	Cell type	Reference numbers
Human cells	Cas9 nuclease	HEK293FT, HEK293T, HEK293, K562, iPSC, HUES9, HUES1, BJ-RiPS, HeLa, Jurkat, U2OS	9,13–16,47, 49–51,54,59, 84,85
	Cas9 nickase	HEK293FT, HEK293T	13,14,47,49
	dCas9 (gene regulation)	HEK293FT, HEK293T	70–72,74,82
	dCas9 (imaging)	HEK293T, UMUC3, HeLa	81
Mouse or mouse cells	Cas9 nuclease	Embryos	14,24–26
	Cas9 nickase	Embryos	47
	dCas9 (gene regulation)	NIH3T3	74
Rat	Cas9 nuclease	Embryos	26,36
Rabbit	Cas9 nuclease	Embryos	27
Frog	Cas9 nuclease	Embryos	28
Zebrafish	Cas9 nuclease	Embryos	17,33,37,60,85
Fruit fly	Cas9 nuclease	Embryos	29,30,61
Silkworm	Cas9 nuclease	Embryos	31
Roundworm	Cas9 nuclease	Adult gonads	32,62–67
Rice	Cas9 nuclease	Protoplasts, callus cells	21,23
Wheat	Cas9 nuclease	Protoplasts	21
Sorghum	Cas9 nuclease	Embryos	23
Tobacco	Cas9 nuclease	Protoplasts, leaf tissue	19,20,23
Thale cress	Cas9 nuclease	Protoplasts, seedlings	19,23
Yeast	Cas9 nuclease	*Saccharomyces cerevisiae*	18
Bacteria	Cas9 nuclease	*Streptococcus pneumoniae, E. coli*	8
	dCas9 (gene regulation)	*E. coli*	69,70

HEK, human embryonic kidney; iPSCs, induced pluripotent stem cells; UMUC3, human bladder cancer.

Cas9 variants that cut one strand rather than both strands of the target DNA site (known as 'nickases') have also been shown to be useful for genome editing. Introduction of a D10A or H840A mutation into the RuvC1- or HNH-like nuclease domains in Cas9 (**Fig. 4a**)[41,42] results in the generation of nickases that cut either the complementary or noncomplementary DNA target strands, respectively, *in vitro*[7,12,43] (**Fig. 4b**,**c**). Consistent with previous studies with ZFN-derived nickases[44–46] (J.K.J. and colleagues[44]), Cas9 nickases can, at some sites, induce HDR with reduced levels of concomitant NHEJ-mediated indels[13,14]. However, although at some sites Cas9 nickases can induce HDR with efficiencies similar to those of the original Cas9 nuclease[13,14], these rates can be much lower at other sites[47]. Notably, the frequencies of indel mutations introduced by nickases have also been high at certain sites[13,47–49] (J.K.J. and colleagues[48]). Although the precise DNA repair pathways by which these various alterations are induced remain undefined, one potential mechanism that has been postulated is that passage of a replication fork through a nuclease-induced nick site might result in a DNA DSB. Additional studies with Cas9 nickases are needed to better understand locus-dependent differences in the relative efficiencies of HDR and indel mutation induced by these enzymes.

Determining the specificities of RNA-guided Cas9 nucleases

Although RGNs generally cleave their intended target sites reliably, an important question is, to what extent do these nucleases induce off-target cleavage events (and therefore unwanted NHEJ-induced indel mutations)? To assess RGN specificity, several groups have created gRNA variants containing one to four nucleotide mismatches in the complementarity region and have then examined the abilities of these molecules to direct Cas9 nuclease activity in human cells at reporter gene (J.K.J. and colleagues[50]) or endogenous gene[14,51] target sites. These studies showed that mismatches are generally better tolerated at the 5′ end of the 20-nt targeting region of the gRNA than at the 3′ end; this result is consistent with previous experiments performed *in vitro* and in bacterial cells, which suggested that the 8–12 bp at the 3′ end of the targeting sequence (also known as the 'seed' sequence) are crucial for target recognition[7,8,14,52,53]. However, the effects of single and double mismatches are not always predictable based on their location within the gRNA targeting region; some mismatches in the 5′ end can have dramatic effects, whereas some in the 3′ end do not greatly affect Cas9 activity[50]. In addition, not all nucleotide substitutions at a given position necessarily have equivalent effects on activity[51].

A reciprocal, and perhaps more relevant, approach for studying specificity is to assess the activities of Cas9 at potential off-target genomic DNA target sites, (i.e., sites that have a few nucleotide differences compared to the intended target). A number of studies have examined potential off-target sites that differ at one to six positions from the on-target site in human cells[9,47,48,50,51,54]. Collectively, these reports have found cases of off-target mutations at sites that differ by as many as five positions within the protospacer region[50] and/or that have an alternative PAM sequence of the form NAG[51]. Interestingly, indel mutation frequencies at these off-target sites can be high enough (>2–5%) to detect using the relatively insensitive T7 endonuclease I (T7E1) mutation mismatch assay and sometimes are comparable to the on-target site mutation frequency[48,50,54]. In addition, more sensitive deep sequencing assays have identified lower frequency off-target mutations[48,51,55]. It is important to note that all of these directed studies examined only a subset of the much larger number of potential off-target sites in the genome. For example, any given 20-nt protospacer will typically have hundreds to thousands of potential off-target sites that differ at four or five positions, respectively, in 6×10^9 bases of random DNA. In addition, although it has been suggested that higher GC content at the RNA:DNA hybridization interface might potentially help to stabilize binding of the RGN to DNA, high rates of mutagenesis have been observed for off-target sites with as little as 30% matched GC content[9,50].

A somewhat more comprehensive strategy for examining Cas9 specificities is to identify off-target sites from a partially degenerate library of variants that is based on the intended on-target sequence. One recent report identified sites from such libraries based on their abilities to be bound by a catalytically inactive form of Cas9 fused to a transcriptional activation domain (see **Fig. 4** and further discussion below)[49]. This study found sites that were mismatched by as many as three (and possibly more) positions relative to the on-target site[49]. These results are similar to those of another study, which used *in vitro* selection for Cas9 nuclease cleavage activity to identify potential off-target sites from a partially degenerate library of target site variants. Some of the off-target sites identified by these *in vitro* selections (with up to four mismatches) were also shown to be mutated in human cells[9].

A recent study using whole-exome sequencing did not find evidence of Cas9-induced, off-target mutations in three modified human K562 cell line clones[56]. Although the authors acknowledge that the high false-negative rate associated with exome sequencing analysis limits interpretation of these data, these results do suggest that with careful target selection, it may be possible to isolate Cas9-edited cells with otherwise intact exomes. Additional examples with deeper sequencing coverage and whole genome (rather than whole exome) sequencing will be needed to determine how readily cells that do not have off-target mutations can be isolated. The ability to do so would encourage broader research application of Cas9 technology. However, it is worth noting that deep sequencing the genomes of individual cell

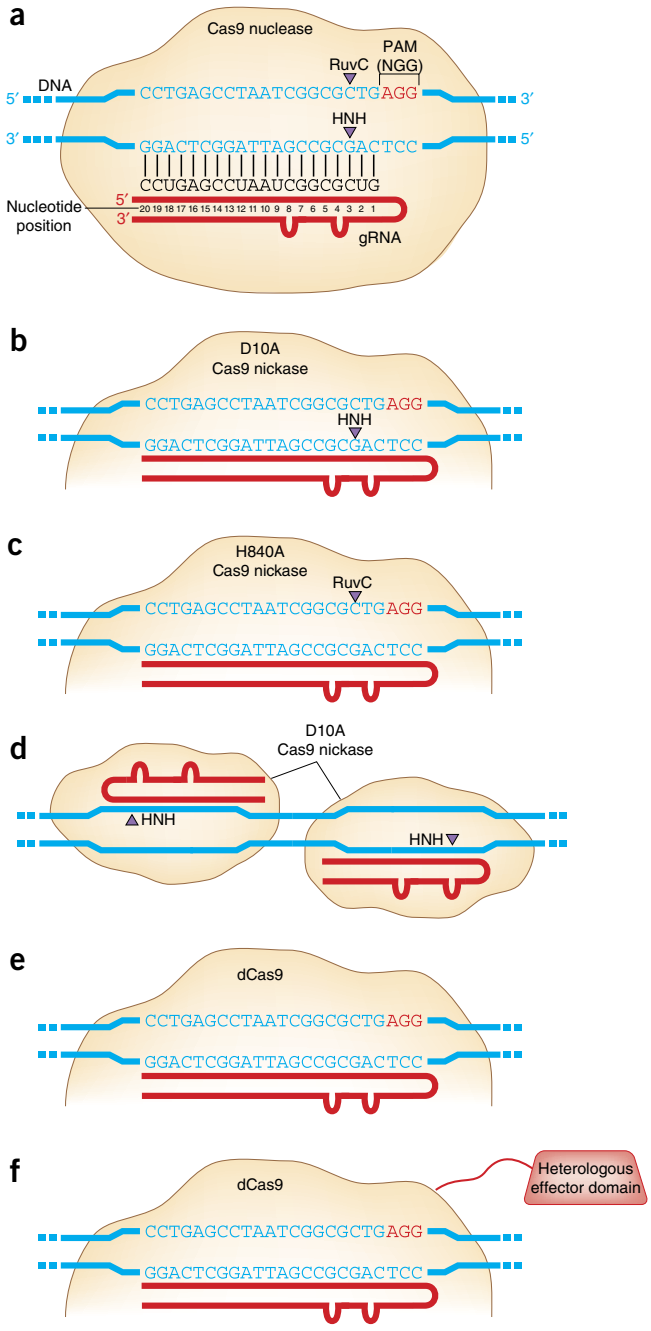


Figure 4 Cas9-based systems for altering gene sequence or expression. (**a**) Cas9 nuclease creates double-strand breaks at DNA target sites with complementarity to the 5′ end of a gRNA. Cas9 contains RuvC and HNH nuclease domains (arrowheads). (**b**) Cas9 nickase created by mutation of the RuvC nuclease domain with a D10A mutation. This nickase cleaves only the DNA strand that is complementary to and recognized by the gRNA. (**c**) Cas9 nickase created by mutation of the HNH nuclease domain with a H840A mutation. This nickase cleaves only the DNA strand that does not interact with the gRNA. (**d**) Paired nickase strategy for improving Cas9 specificity. Two D10A Cas9 nickases are directed by a pair of appropriately oriented gRNAs. This leads to induction of two nicks that, if introduced simultaneously, would be expected to generate a 5′ overhang. (**e**) Catalytically inactive or 'dead' Cas9 (dCas9) (e.g., with mutations in both the RuvC and HNH domains). This can be recruited by a gRNA without cleaving the target DNA site. (**f**) Catalytically inactive dCas9 can be fused to a heterologous effector domain.

clones is expected to be neither sensitive nor effective for defining the full genome-wide spectrum of Cas9 off-target sites because each clone would likely only carry mutations at a small proportion of, if any, possible off-target sites.

Overall, the various published studies strongly suggest that off-target sites of RNA-guided Cas9 nucleases can be variable in frequency and challenging to predict. For any given target site, it is not currently possible to predict how many mismatches can be tolerated, nor do we fully understand why some sites are cleaved whereas other are not. We also do not know how genomic and/or epigenomic context might affect the frequency of cleavage. Although some initial evidence suggests that DNA methylation does not inhibit Cas9-based genome editing[51], it seems plausible and likely that chromatin structure could play a role in off-target site accessibility. A more comprehensive understanding of Cas9 off-target effects will have to await the development of unbiased, global measures of Cas9 specificity in cells.

Methods for reducing off-target effects of Cas9 nucleases

Even with an incomplete understanding of RNA-guided Cas9 nuclease specificity, researchers have begun to explore various approaches to reduce off-target mutagenic effects. One potential strategy is to test the effects of reducing the concentrations of gRNA and Cas9 expressed in human cells. Results with this approach have been mixed; one group observed proportionately larger decreases in rates of off-target relative to on-target mutagenesis for two gRNAs[51], whereas our group observed nearly proportionate decreases at both off-target and on-target sites for two other gRNAs[50]. The use of modified gRNA architectures with truncated 3′ ends (within the tracrRNA-derived sequence) or with two extra guanine nucleotides appended to the 5′ end (just before the complementarity region) also yielded better on-target to off-target ratios but generally with considerably lower absolute efficiencies of on-target genome editing[9,56].

Another proposed approach for improving specificity involves the use of 'paired nickases' in which adjacent off-set nicks are generated at the target site using two gRNAs and Cas9 nickases[47,49,56] (**Fig. 4d**), a strategy analogous to one originally performed with pairs of engineered zinc finger nickases[46]. In contrast to single Cas9 nickases (which can at some sites more favorably induce HDR events relative to NHEJ indels), paired Cas9 nickases targeted to sites on opposite DNA strands separated by 4 to 100 bp can efficiently introduce both indel mutations and HDR events with a single-stranded DNA oligonucleotide donor template in mammalian cells[47,49,56]. It has been proposed by some that the concerted action of paired nickases create a DSB that is then repaired by NHEJ or HDR[47,55]. Importantly, paired nickases can reduce Cas9-induced off-target effects of gRNAs in human cells; the addition of a second gRNA and substitution of Cas9 nickase for Cas9 nuclease can lead to lower levels of unwanted mutations at previously known off-target sites of the original gRNA[47]. However, an as-yet unanswered question is whether the second gRNA can itself induce its own range of Cas9 nickase–mediated off-target mutations in the genome. Multiple studies have shown that single monomeric Cas9 nickases can function on their own to induce indel mutations at certain genomic loci[13,47–49], perhaps because an individual nick might be converted to a DSB when a replication fork passes through the locus[57,58]. Thus, an important improvement needed for the paired nickase system will be to make the activities of the two nickase monomers strictly co-dependent on each other for genome editing activity—that is, so that these nickase monomers are only active for genome editing when bound to DNA in close proximity to the other.

Our group has recently shown that off-target effects can be substantially reduced simply by using gRNAs that have been shortened

at the 5′ end of their complementarity regions[48]. These truncated gRNAs (which we refer to as 'tru-gRNAs') have 17 or 18 nucleotides of complementarity; they generally function as efficiently as full-length gRNAs in directing on-target Cas9 activity but show decreased mutagenic effects at off-target sites and enhanced sensitivity to single or double mismatches at the gRNA:DNA interface[48]. This strategy avoids the technical challenges associated with expressing multiple gRNAs in a single cell for the paired nickase approach and should be straightforward to implement. tru-gRNAs could also be used in conjunction with other strategies for improving Cas9 specificity (e.g., we showed that tru-gRNAs further improve the specificity of paired nickases[48]) as well as the specificities of dCas9 fusion proteins for non-nuclease applications (described below).

Practical considerations for implementing CRISPR-Cas technology

Owing to rapid progress in the field, potential users face a variety of choices about how to implement CRISPR-Cas technology. Here we discuss some of the parameters to consider when applying the methodology.

Choice of gRNA platform. It is important to note that the efficiency of Cas9 activity for any given locus can be influenced by the architecture of gRNA(s) used. As described above, most recent studies have used a single gRNA that is a fusion of a programmable crRNA and part of the tracrRNA, but earlier studies also used a 'dual gRNA' configuration in which the crRNA and tracrRNA are expressed separately. In general, studies using single gRNAs have consistently reported substantially higher editing rates than those using dual gRNAs[13,14,17,59]. These findings suggest that the single gRNA system may be more active than the double gRNA system, presumably because two components can assemble more efficiently than three components.

In addition, single gRNAs harboring variable lengths of tracrRNA sequence at their 3′ ends have been used by different groups (**Supplementary Table 1**). Systematic comparisons have generally demonstrated that longer single gRNAs (containing more of the 3′ portion of the tracrRNA sequence) yield higher editing rates than shorter ones[51]. The most commonly used single gRNA design to date is ~100 nt in length (**Supplementary Table 1**). The tru-gRNAs described above are shortened versions of this ~100-nt single gRNA.

Targeting range and choice of gRNA target sites. The choice of promoter used to express gRNAs can limit the options for potential target DNA sites. For example, the RNA polymerase III–dependent U6 promoter or the T7 promoter require a G or GG, respectively, at the 5′ end of the sequence of the RNA that is to be transcribed (top panels of **Fig. 5a**,**b**). As a result, standard full-length or tru-gRNAs expressed from these promoters are limited to targeting sites that match the forms $GN_{16\text{-}19}NGG$ or $GGN_{15\text{-}18}NGG$; such sites are expected to occur every 1 in 32 bp or 1 in 128 bp, respectively, in random DNA sequence. Paired nickase strategies require the identification of two sites on opposite strands of DNA with appropriate spacing in between (as described above). One option to reduce these targeting range restrictions is to choose sites without regard to the identities of the first or first two bases at the 5′ end (that is, making gRNAs that are mismatched at these positions). Another potential strategy to bypass these restrictions is to append the required G or GG to the 5′ end of the gRNA, thereby encoding gRNA transcripts that are 1 or 2 bp longer (bottom panels of **Fig. 5a**,**b**). Both of these strategies have been used successfully to produce active gRNAs but with variable efficiencies in genome-editing activities with Cas9 nucleases[47,54,56,60]. Larger-scale studies are needed to clarify the effects of using either

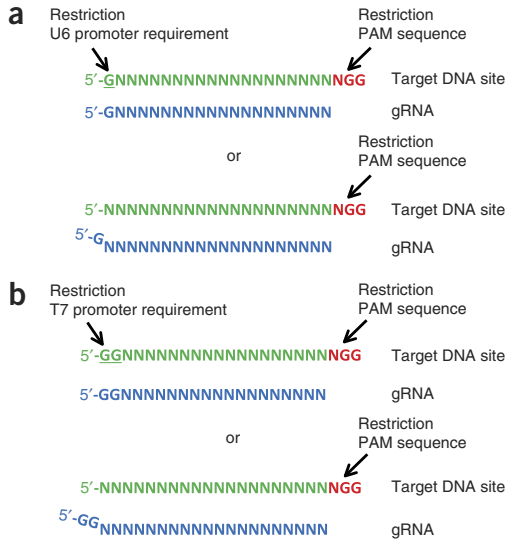


Figure 5 Sequence limitations on the targeting range of guide RNAs. (**a**) Ranges of potential target sites for gRNAs expressed from a U6 promoter. Target DNA sequence restrictions are imposed by the requirement for a G at the first 5′ nucleotide of the gRNA (blue letters) targeted to the DNA site (required for efficient expression from a U6 promoter) and by the need for an NGG (the PAM sequence; red letters) adjacent to the complementarity region of target site (green letters) (top panel). One strategy to avoid the requirement for a 5′ G in the target site is to append an extra G to the 5′ end of the gRNA (bottom panel). (**b**) Ranges of potential target sites for gRNAs expressed from a T7 promoter. Target DNA sequence restrictions are imposed by the requirement for a GG at the first two nucleotides of the gRNA targeted to the DNA site (required for efficient expression from the T7 promoter) and by the need for an NGG adjacent to the complementarity region target site (top panel). One strategy to avoid the requirement for a GG dinucleotide in the target sequence is to append an extra GG dinucleotide to the 5′ end of the gRNA (bottom panel).

mismatched or extended gRNAs on the efficiencies and specificities of RGN-mediated cleavage. Several groups have provided web-based software that facilitates the identification of potential CRISPR RGN target sites in user-defined sequences (e.g., the ZiFiT Targeter software[17,48] (http://zifit.partners.org/) and the CRISPR Design Tool[51] (http://crispr.mit.edu/)).

Delivery of CRISPR-Cas components. RGNs have been delivered to a broad range of cell types and organisms using a variety of delivery methods. In cultured mammalian cells, researchers have used electroporation[59], nucleofection[13,50] and Lipofectamine-mediated transfection[13,14,50] of nonreplicating plasmid DNA to transiently express Cas9 and gRNAs. Lentiviral vectors have also been used to constitutively express Cas9 and/or gRNAs in cultured human[38,39] and mouse[40] cells. RNAs and/or plasmid DNA transcribed *in vitro* have been injected directly into the embryos of zebrafish[17], fruit flies[29,30,61], mice[24,26] and rats[26]. Plasmid DNA and RNA have also been injected into the gonads of adult roundworms[32,62–66], and in one study purified Cas9 protein complexed with gRNA was injected into roundworm gonads[67]. In addition to animal models and cell lines, Cas9 has been used successfully in multiple plant species including wheat, rice, sorghum, tobacco and thale cress using a range of standard delivery methods including PEG-mediated transformation of protoplasts, *Agrobacterium*-mediated transfer in embryos and leaf tissue, and/or bombardment of callus cells with plasmid DNA[19–21,23]. For most RGN applications, transient expression of gRNAs and Cas9 is typically sufficient to induce

efficient genome editing. Although constitutive expression of RGN components might potentially lead to higher on-target editing efficiencies, extended persistence of these components in the cell might also lead to increased frequencies of off-target mutations, a phenomenon that has been previously reported with ZFNs[68].

Experimental strategies to control for RGN off-target effects. The existence of CRISPR RGN-induced off-target effects and our current inability to comprehensively identify these alterations on a genome-wide scale mean that investigators need to account for the potentially confounding effects of these undesired mutations. Several strategies can be used to rule out off-target mutations as a potential alternative explanation for any phenotypes observed. For example, complementation with reintroduction of a wild-type gene can be used to confirm the effects of knockout mutations. In addition, similar to the strategy of targeting a gene with multiple RNA interference hairpins, one could easily create mutations in the same gene using gRNAs targeted to different sites. Presumably, each gRNA will be expected to have a different range of off-target effects and therefore if the same phenotype is observed with each of these different gRNAs it would seem unlikely that undesired mutations are the cause. The ease with which multiple gRNAs can be rapidly designed and constructed makes it simple and feasible to implement this type of strategy with the Cas9 system. The high efficacy of the Cas9 nuclease for inducing mutations makes it an attractive choice for creating mutant cell lines and whole organisms in spite of the need to account for off-target effects.

Applications of CRISPR-Cas beyond genome editing

Beyond enabling facile and efficient targeted genome editing, the CRISPR-Cas system has the potential to be used to regulate endogenous gene expression or to label specific chromosomal loci in living cells or organisms. Catalytically inactive or "dead" Cas9 (dCas9)—a variant bearing both the D10A and H840A mutations that does not cleave DNA—can be recruited by gRNAs to specific target DNA sites[7,12] (**Fig. 4e**). Targeting of dCas9 to promoters was initially shown to repress gene expression in both *Escherichia coli* and human cells[69,70]. Interestingly, dCas9 repressed a bacterial promoter efficiently when recruited with gRNAs that interacted with either strand of sequences upstream of the promoter; however, when targeting sites downstream of the transcription start point, only gRNAs that interacted with the nontemplate strand induced dCas9-mediated repression[69]. dCas9 also provides a general platform for recruitment of heterologous effector domains to specific genomic loci (**Figs. 2d–f** and **4f**). For example, dCas9 fusions to a transcriptional activation domain (VP64 or the p65 subunit of nuclear factor kappa B; NF-κB) or a transcriptional repression domain (the Krüppel-associated box (KRAB) domain) have been shown to regulate the expression of endogenous genes in human[71–73,82] and mouse cells[74]. Changes in gene expression induced by these dCas9 fusions in human cells thus far seem to be generally lower than those induced by similar TALE-based transcription factors[49,75–78]. However, multiplex recruitment of dCas9-based activators using between 2 and 10 sgRNAs targeted to the same promoter can result in substantially higher levels of human gene activation, presumably due to the phenomenon of activator synergy[49,71,72,74] (J.K.J. and colleagues[71]). This capability of dCas9-based activators to function synergistically is consistent with previous observations for TALE-based activators[75,76] (J.K.J. and colleagues[75]) in human cells. In future experiments it will be interesting to see whether dCas9 fusions to histone modifiers and proteins involved in altering DNA methylation, such as the ten-eleven translocation (TET) proteins, can also be used to perform targeted 'epigenome editing' (**Fig. 2e**), including the alteration of specific histone modifications and demethylation of particular cytosine bases in human cells as has been recently described with TALE DNA-binding domains[73,79] (J.K.J. and colleagues[80]).

An alternative strategy for tethering heterologous effector domains to DNA-bound gRNA:dCas9 complexes is to exploit well-defined, RNA-protein interaction pairs. This approach uses two engineered components: a gRNA that has one or two RNA binding sites for the phage MS2 coat protein fused to its 3′ end; and a fusion of MS2 coat protein to an effector domain[49]. Addition of the MS2 RNA binding sequences to the gRNA does not abolish its ability to target dCas9 to specific DNA sites. Furthermore, co-expression of the MS2 coat protein fusion with the hybrid gRNA and dCas9 has been used to recruit activation domains to a gene promoter in human cells[49]. Although the activation observed seems to be somewhat less robust than direct fusions to dCas9, this type of configuration might provide additional options and flexibility for recruitment of multiple effector domains to a promoter by, for example, using multiple gRNAs and MS2 coat protein binding sites on each gRNA to recruit many copies of different domains to the same promoter.

It has also been demonstrated that an EGFP-dCas9 fusion can be used to visualize DNA loci harboring repetitive sequences, such as telomeres, with a single gRNA or nonrepetitive loci using 26 to 36 gRNAs tiled across a 5-kb region of DNA[81] (**Fig. 2f**). This imaging strategy provides a powerful tool for studying chromosome dynamics and structure and extends the dCas9 system beyond gene expression–based applications.

Thus far, evidence suggests that the effects of the small number of dCas9-activation or repression domain fusions tested to date can be highly specific in mammalian cells, as judged by RNA-seq or expression microarray experiments[74,82]; however, this may be because not all binding events lead to changes in gene transcription. It remains to be determined whether dCas9 fusions are truly specific for single sites in their cellular activities or if, like their nuclease counterparts, strategies (such as the use of tru-gRNAs) will be needed to improve specificity.

Future directions

Progress in the development of Cas9-based technologies over the past 18 months has been stunning, but many interesting questions and applications remain to be addressed and explored.

First, methods for expanding the targeting range of RNA-guided Cas9 will be important for inducing precise HDR or NHEJ events as well as for implementing multiplex strategies, including paired nickases. As noted above, the targeting range for Cas9, paired Cas9 nickases and dCas9 fusions is restricted mainly by the need for a PAM sequence matching the form NGG. Alternative PAM sequences of the form NAG or NNGG can be exploited, as has been noted[7,8,51], but more experiments are needed to ascertain how robustly these sequences are recognized and cleaved. Other gRNA:Cas9 platforms with different PAM sequences isolated from *Streptococcus thermophilus*, *Neisseria meningitidis* and *Treponema denticola* have also been characterized[10,11,14] and identification of more of these systems from other species[83] could further enhance the targeting range of the platform.

Second, the field urgently needs to develop unbiased strategies to globally assess the off-target effects of Cas9 nucleases or paired nickases in any genome of interest. Such methods will be crucial for evaluating how effectively improvements described to date enhance the specificity of the platform. In addition, although tru-gRNAs and paired nickases can reduce off-target effects, it is likely that further improvements will be needed, especially for therapeutic applications. Ideally, new strategies could be combined with existing approaches.

Examples of such improvements might involve using protein engineering to modify Cas9 and/or modifying the nucleotides used by the gRNA to mediate recognition of the target DNA site. Alternatively, the construction of inducible forms of Cas9 and/or gRNAs might provide a means to regulate the active concentration of these reagents in the cell and thereby improve the ratio of on- and off-target effects.

Third, methods for efficient delivery and expression of CRISPR-Cas system components will undoubtedly need to be optimized for each particular cell-type or organism to be modified. For example, some cell types or tissues might be refractory to transfection and/or infection by standard viral vectors. A related challenge will be to develop methods that enable expression of either the gRNAs or the Cas9 nuclease that is specific to a tissue, cell type or developmental stage. Strategies that ensure efficient expression of large numbers of different gRNAs simultaneously from one vector would also allow more extensive use of the multiplex capability of CRISPR-Cas systems. Collectively, these advances will be important for research use and therapeutic applications.

Lastly, strategies for shifting the balance away from NHEJ-mediated indel mutations and toward HDR-driven alterations remain a priority for development. Although high rates of HDR can be achieved with the CRISPR RGNs and single-stranded DNA oligonucleotides, competing mutagenic NHEJ also occurs simultaneously. This limitation is particularly problematic when using HDR to induce point mutation changes (as opposed to insertions) in the protospacer part of the target site; alleles that have been successfully altered in this way can still be efficiently re-cut and then mutagenized by NHEJ, thereby reducing the yield of correctly edited sequences. One of the drawbacks to developing an approach to improve the HDR:NHEJ ratio is that inhibition of NHEJ is likely to be poorly tolerated by most cells, given its central role in normal DNA repair. For therapeutic applications seeking to exploit HDR, reduction or elimination of competing NHEJ will be crucially important.

The simplicity, high efficiency and broad applicability of the RNA-guided Cas9 system have positioned this technology to transform biological and biomedical research. The ease with which researchers can now make changes in the sequence or expression of any gene means reverse genetics can be performed in virtually any organism or cell type of interest. In addition, the construction of large libraries of gRNAs for altering or regulating genes of interest will enable facile, comprehensive forward genetic screens. All of these systems can also be multiplexed by expressing multiple gRNAs in a single cell, thereby further extending the complexity of forward and reverse genetic experiments that can be done. Although the off-target effects of Cas9 remain to be defined on a genome-wide scale, much progress has already been made toward improving specificity, and further advances will undoubtedly come rapidly, given the intensity of research efforts in this area. All of these recent advances—and those to come—in developing and optimizing Cas9-based systems for genome and epigenome editing should propel the technology toward therapeutic applications, opening the door to treating a wide variety of human diseases.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

J.K.J. is grateful for support from the US National Institutes of Health (NIH) (grants DP1 GM105378 and R01 GM088040), the Defense Advanced Research Projects Agency (grant W911NF-11-2-0056) and The Jim and Ann Orr Massachusetts General Hospital Research Scholar Award. This material is based upon work supported fully or in part by the US Army Research Laboratory and the US Army Research Office under grant number W911NF-11-2-0056.

The authors apologize to colleagues whose studies were not cited due to length and reference constraints.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Carroll, D. Genome engineering with zinc-finger nucleases. *Genetics* **188**, 773–782 (2011).
2. Wiedenheft, B., Sternberg, S.H. & Doudna, J.A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331–338 (2012).
3. Fineran, P.C. & Charpentier, E. Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* **434**, 202–209 (2012).
4. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167–170 (2010).
5. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
6. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
7. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
8. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).
9. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843 (2013).
10. Esvelt, K.M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* **10**, 1116–1121 (2013).
11. Hou, Z. *et al.* Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. *Proc. Natl. Acad. Sci. USA* **110**, 15644–15649 (2013).
12. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. USA* **109**, E2579–E2586 (2012).
13. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
14. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
15. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, e00471 (2013).
16. Cho, S.W., Kim, S., Kim, J.M. & Kim, J.S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**, 230–232 (2013).
17. Hwang, W.Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 227–229 (2013).
18. DiCarlo, J.E. *et al.* Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* **41**, 4336–4343 (2013).
19. Li, J.F. *et al.* Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat. Biotechnol.* **31**, 688–691 (2013).
20. Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J.D. & Kamoun, S. Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**, 691–693 (2013).
21. Shan, Q. *et al.* Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 686–688 (2013).
22. Xie, K. & Yang, Y. RNA-guided genome editing in plants using a CRISPR-Cas system. *Mol. Plant* **6**, 1975–1983 (2013).
23. Jiang, W. *et al.* Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice. *Nucleic Acids Res.* **41**, e188 (2013).
24. Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).
25. Shen, B. *et al.* Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res.* **23**, 720–723 (2013).
26. Li, D. *et al.* Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 681–683 (2013).
27. Yang, D. *et al.* Effective gene targeting in rabbits using RNA-guided Cas9 nucleases. *J. Mol. Cell Biol.* doi:10.1093/jmcb/mjt047 (8 January 2014).
28. Nakayama, T., *et al.* Simple and efficient CRISPR/Cas9-mediated targeted mutagenesis in *Xenopus* tropicalis. *Genesis* **51**, 835–843 (2013).
29. Yu, Z. *et al.* Highly efficient genome modifications mediated by CRISPR/Cas9 in *Drosophila*. *Genetics* **195**, 289–291 (2013).
30. Bassett, A.R., Tibbit, C., Ponting, C.P. & Liu, J.L. Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. *Cell Reports* **4**, 220–228 (2013).
31. Wang, Y. *et al.* The CRISPR/Cas system mediates efficient genome engineering in Bombyx mori. *Cell Res.* **23**, 1414–1416 (2013).

32. Friedland, A.E. *et al.* Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat. Methods* **10**, 741–743 (2013).
33. Xiao, A. *et al.* Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic Acids Res.* **41**, e141 (2013).
34. Upadhyay, S.K., Kumar, J., Alok, A. & Tuli, R. RNA-guided genome editing for target gene mutations in wheat. *G3 (Bethesda)* **3**, 2233–2238 (2013).
35. Horii, T. *et al.* Genome engineering of mammalian haploid embryonic stem cells using the Cas9/RNA system. *PeerJ* **1**, e230 (2013).
36. Li, W., Teng, F., Li, T. & Zhou, Q. Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 684–686 (2013).
37. Jao, L.E., Wente, S.R. & Chen, W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc. Natl. Acad. Sci. USA* **110**, 13904–13909 (2013).
38. Wang, T., Wei, J.J., Sabatini, D.M. & Lander, E.S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
39. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
40. Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera, M.D.C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).
41. Haft, D.H., Selengut, J., Mongodin, E.F. & Nelson, K.E. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* **1**, e60 (2005).
42. Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I. & Koonin, E.V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **1**, 7 (2006).
43. Sapranauskas, R. *et al.* The Streptococcus thermophilus CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* **39**, 9275–9282 (2011).
44. Ramirez, C.L. *et al.* Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res.* **40**, 5560–5568 (2012).
45. Wang, J. *et al.* Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. *Genome Res.* **22**, 1316–1326 (2012).
46. Kim, E. *et al.* Precision genome engineering with programmable DNA-nicking enzymes. *Genome Res.* **22**, 1327–1333 (2012).
47. Ran, F.A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
48. Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M. & Joung, J.K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* doi:10.1038/nbt.2808 (26 January 2014).
49. Mali, P., *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013).
50. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013).
51. Hsu, P.D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
52. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. USA* **108**, 10098–10103 (2011).
53. Wiedenheft, B. *et al.* RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. USA* **108**, 10092–10097 (2011).
54. Cradick, T.J., Fine, E.J., Antico, C.J. & Bao, G. CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.* **41**, 9584–9592 (2013).
55. Ran, F.A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
56. Cho, S.W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24**, 132–141 (2014).
57. Kuzminov, A. Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proc. Natl. Acad. Sci. USA* **98**, 8241–8246 (2001).
58. Cortes-Ledesma, F. & Aguilera, A. Double-strand breaks arising by replication through a nick are repaired by cohesin-dependent sister-chromatid exchange. *EMBO Rep.* **7**, 919–926 (2006).
59. Ding, Q. *et al.* Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell* **12**, 393–394 (2013).
60. Hwang, W.Y. *et al.* Heritable and precise zebrafish genome editing using a CRISPR-Cas system. *PLoS ONE* **8**, e68708 (2013).
61. Gratz, S.J. *et al.* Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. *Genetics* **194**, 1029–1035 (2013).
62. Lo, T.W. *et al.* Precise and heritable genome editing in evolutionarily diverse nematodes using TALENs and CRISPR/Cas9 to engineer insertions and deletions. *Genetics* **195**, 331–348 (2013).
63. Waaijers, S. *et al.* CRISPR/Cas9-targeted mutagenesis in *Caenorhabditis elegans*. *Genetics* **195**, 1187–1191 (2013).
64. Tzur, Y.B. *et al.* Heritable custom genomic modifications in *Caenorhabditis elegans* via a CRISPR-Cas9 system. *Genetics* **195**, 1181–1185 (2013).
65. Katic, I. & Grosshans, H. Targeted heritable mutation and gene conversion by Cas9-CRISPR in *Caenorhabditis elegans*. *Genetics* **195**, 1173–1176 (2013).
66. Chiu, H., Schwartz, H.T., Antoshechkin, I. & Sternberg, P.W. Transgene-free genome editing in *Caenorhabditis elegans* using CRISPR-Cas. *Genetics* **195**, 1167–1171 (2013).
67. Cho, S.W., Lee, J., Carroll, D. & Kim, J.S. Heritable gene knockout in *Caenorhabditis elegans* by direct injection of Cas9-sgRNA ribonucleoproteins. *Genetics* **195**, 1177–1180 (2013).
68. Gaj, T., Guo, J., Kato, Y., Sirk, S.J. & Barbas, C.F. III. Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nat. Methods* **9**, 805–807 (2012).
69. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41**, 7429–7437 (2013).
70. Qi, L.S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
71. Maeder, M.L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10**, 977–979 (2013).
72. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **10**, 973–976 (2013).
73. Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472–476 (2013).
74. Cheng, A.W. *et al.* Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* **23**, 1163–1171 (2013).
75. Maeder, M.L. *et al.* Robust, synergistic regulation of human gene expression using TALE activators. *Nat. Methods* **10**, 243–245 (2013).
76. Perez-Pinera, P. *et al.* Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat. Methods* **10**, 239–242 (2013).
77. Miller, J.C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**, 143–148 (2011).
78. Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.* **29**, 149–153 (2011).
79. Mendenhall, E.M. *et al.* Locus-specific editing of histone modifications at endogenous enhancers. *Nat. Biotechnol.* **31**, 1133–1136 (2013).
80. Maeder, M.L. *et al.* Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat. Biotechnol.* **31**, 1137–1142 (2013).
81. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
82. Gilbert, L.A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
83. Chylinski, K., Le Rhun, A. & Charpentier, E. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol.* **10**, 726–737 (2013).
84. Ebina, H., Misawa, N., Kanemura, Y. & Koyanagi, Y. Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. *Sci Rep* **3**, 2510 (2013).
85. Chang, N. *et al.* Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Res.* **23**, 465–472 (2013).
86. Silva, G. *et al.* Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr. Gene Ther.* **11**, 11–27 (2011).
87. Smith, J. *et al.* A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res.* **34**, e149 (2006).
88. Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. & Gregory, P.D. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11**, 636–646 (2010).
89. Wolfe, S.A., Nekludova, L. & Pabo, C.O. DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 183–212 (2000).
90. Gupta, A. *et al.* An optimized two-finger archive for ZFN-mediated gene targeting. *Nat. Methods* **9**, 588–590 (2012).
91. Kim, S., Lee, M.J., Kim, H., Kang, M. & Kim, J.S. Preassembled zinc-finger arrays for rapid construction of ZFNs. *Nat. Methods* **8**, 7 (2011).
92. Maeder, M.L. *et al.* Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell* **31**, 294–301 (2008).
93. Sander, J.D. *et al.* Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat. Methods* **8**, 67–69 (2011).
94. Gonzalez, B. *et al.* Modular system for the construction of zinc-finger libraries and proteins. *Nat. Protoc.* **5**, 791–810 (2010).
95. Wright, D.A. *et al.* Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat. Protoc.* **1**, 1637–1652 (2006).
96. Carroll, D., Morton, J.J., Beumer, K.J. & Segal, D.J. Design, construction and in vitro testing of zinc finger nucleases. *Nat. Protoc.* **1**, 1329–1341 (2006).
97. Reyon, D. *et al.* FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.* **30**, 460–465 (2012).
98. Boch, J. & Bonas, U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.* **48**, 419–436 (2010).
99. Moscou, M.J. & Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009).
100. Joung, J.K. & Sander, J.D. TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* **14**, 49–55 (2013).
101. Holkers, M. *et al.* Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res.* **41**, e63 (2013).

Recombineering: a homologous recombination-based method of genetic engineering

Shyam K Sharan[1], Lynn C Thomason[2], Sergey G Kuznetsov[1] & Donald L Court[3]

[1]Mouse Cancer Genetics Program, Center for Cancer Research, National Cancer Institute at Frederick, Frederick, Maryland 21702, USA. [2]SAIC-Frederick Inc., Gene Regulation and Chromosome Biology Laboratory, Basic Research Program, Frederick, Maryland, 21702, USA. [3]Gene Regulation and Chromosome Biology Laboratory, Center for Cancer Research, National Cancer Institute at Frederick, Frederick, Maryland 21702, USA. Correspondence should be addressed to S.K.S. (sharans@mail.nih.gov) or D.L.C. (court@mail.ncifcrf.gov).

Published online 29 January 2009; doi:10.1038/nprot.2008.227

Recombineering is an efficient method of *in vivo* genetic engineering applicable to chromosomal as well as episomal replicons in *Escherichia coli*. This method circumvents the need for most standard *in vitro* cloning techniques. Recombineering allows construction of DNA molecules with precise junctions without constraints being imposed by restriction enzyme site location. Bacteriophage homologous recombination proteins catalyze these recombineering reactions using double- and single-stranded linear DNA substrates, so-called targeting constructs, introduced by electroporation. Gene knockouts, deletions and point mutations are readily made, gene tags can be inserted and regions of bacterial artificial chromosomes or the *E. coli* genome can be subcloned by gene retrieval using recombineering. Most of these constructs can be made within about 1 week's time.

INTRODUCTION

Recombineering is an *in vivo* method of genetic engineering used primarily in *Escherichia coli* that uses short 50 base homologies[1–5]. As recombineering is based on homologous recombination, it allows insertion, deletion or alteration of any sequence precisely and is not dependent on the location of restriction sites (**Fig. 1**). Linear DNAs, either double-stranded (ds), usually in the form of PCR products[3,6–8], or single-stranded (ss) synthetic oligonucleotides[2,9] are introduced by electroporation and provide the homologous substrates (i.e., targeting constructs) to create genetic changes. Recombineering is catalyzed by bacteriophage-encoded homologous recombination functions, such as the coliphage λ Red system[3] and the RecET system from the Rac prophage[4,10].

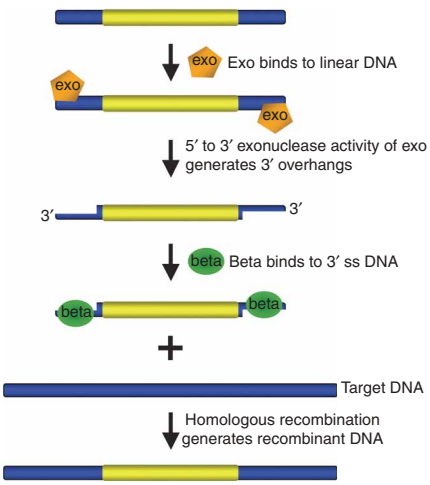


Figure 1 | Overview of bacteriophage λ recombination system used for recombineering. Exo has a 5′- to 3′-dsDNA exonuclease activity, which can generate 3′-overhangs on linear DNA. Beta binds the single-stranded DNA (3′-overhangs), promotes ss-annealing and generates recombinant DNA. An additional protein, Gam (not shown here), which prevents RecBCD nuclease from degrading double-stranded linear DNA fragments, is also required for dsDNA recombineering.

This protocol will emphasize modification of bacterial artificial chromosomes (BACs) and multicopy plasmids but the procedures described are generally applicable to other replicons. A basic knowledge of molecular and microbiological techniques is required to execute the recombineering protocols described here. These basic techniques are described in detail by Ausubel *et al.*[11]. Recombineering protocols for manipulation of the bacterial and phage chromosomes are described elsewhere[12,13]. Recombineering can also be used to modify episomal DNAs such as the low-copy plasmid derivatives of P1 and F that carry artificial chromosomes and are called PAC (P1 artificial chromosome)[14] and BAC[15,16], respectively. Although multicopy plasmids may be the ideal choice of vector when the insert size is relatively small (up to 50 kb), PACs, which accommodate inserts of 50–100 kb, and BACs, which allow inserts of >100 kb, are used for cloning large genomic fragments. A BAC is the vector of choice for cloning and manipulating large DNA fragments. BACs may contain genomic segments that include all of the extragenic *cis*-regulatory elements (promoter, terminator and enhancers) of a gene of interest. BACs are therefore ideal for generating transgenic mice because the insert size may allow expression of the cloned gene under the control of its own regulatory elements, mimicking the endogenous expression pattern. In addition, because of the ease of obtaining desirable BAC clones coupled with simple DNA purification protocols, BACs are widely used in gene mapping and functional studies, analysis of regulatory elements and expression of a transgene under the control of a heterologous promoter[17]. They are also used for genetic analysis of mutations that have been identified in human diseases[18,19]. BACs are modified by recombineering in *E. coli*, and these modified constructs can then be used to generate knockout and knock-in mouse models using embryonic stem cell technology[20,21].

The recombineering protocols described here use the bacteriophage λ Red system that includes the phage recombination genes *gam*, *bet* and *exo*. The *gam* gene function, Gam, prevents an *E. coli* nuclease, RecBCD, from degrading linear DNA fragments[22,23], thus

allowing preservation of transformed linear DNA *in vivo*. The *bet* gene product, Beta, is an ssDNA-binding protein that promotes annealing of two complementary DNA molecules[24–26], and the *exo* gene product, Exo, has a 5′ to 3′ dsDNA exonuclease activity[27,28]. Working together, these latter two proteins insert linear DNA at the desired target, creating genetic recombinants (**Fig. 1**)[29,30]. For dsDNA, Red Exo is thought to degrade from both 5′-ends, exposing ssDNA that is bound by Red Beta. Use of the phage λ Red system for *in vivo* genetic engineering was pioneered by Murphy *et al.*[6,7], who demonstrated dsDNA recombination with the Red functions expressed from the *lac* promoter, both on a multicopy plasmid and as an insertion on the *E. coli* chromosome. Murphy *et al.*[6,7] used linear substrate DNA targeting homologies that were greater than 1 kb. Zhang *et al.*[4] demonstrated that the phage RecET system catalyzes recombination using targeting homologies of only 40–60 bp. This short-length requirement allowed the homologies to be incorporated into PCR primers, substantially advancing the technology. The λ Red system was also shown to act on short homologies[3]. Ellis *et al.*[2] demonstrated that the λ Beta protein promotes efficient *in vivo* recombination with ssDNA, provided as 70-mer ss-oligos. Models for how recombineering occurs[31] propose that the single-strand regions of the incoming linear DNA bound by the Beta protein are annealed to complementary single-strand gaps arising at the replication fork during DNA replication. Consistent with this model, an oligo able to anneal to the discontinuously replicated lagging strand gives a higher recombination frequency than its complementary 'leading strand' oligo[2].

When engineering DNA *in vivo*, it is important to provide a brief high-level pulse of the phage recombination proteins. Limited expression minimizes the toxic effects of the Gam protein[32] and minimizes undesirable genome rearrangements between repetitive DNA sequences. Although the phage recombination functions have been produced from multicopy plasmids under control of the IPTG-inducible *lac* promoter[6,7], the *lac* promoter has a high basal level and requires the *lacIQ* repressor gene *in cis* for tight regulation. A better choice is the arabinose-inducible *pBAD* promoter[33,34]. In *Ara*$^+$ strains, the *pBAD* promoter can be tightly repressed by addition of glucose to the growth medium; however, in *ara* mutants, glucose-mediated repression is less effective, even with the *araC* gene on the plasmid[35]. When the pBAD plasmids are used for recombineering[34], glucose is not added during cell growth and arabinose is usually added at least a generation before the cells are made competent for electroporation. This procedure, although convenient for induction, does not yield the tightest possible repression[35]. In contrast, expression from the λ prophage system is based on the endogenous λ regulatory system, which is the natural method for expression of these recombination functions. The most recent studies on λ (see refs. 36,37) show that its repression system is uniquely strong: the λ repressor binds cooperatively at the three operator sites present at both the pL and pR promoters, and these two sets of repressor-bound operators interact with each other by protein–protein-mediated looping between pL and pR to generate a handcuff of 12 repressor proteins. To ensure this same tight repression during recombineering, both sets of operators are present on all of the prophage constructs, whether they are in the bacterial chromosome or on low-copy plasmids, and a temperature-sensitive repressor is expressed from the *cI857* gene. At low temperatures (30–34 °C), the repressor is active and the recombination genes are not expressed. When the temperature of the bacterial culture is shifted to 42 °C for 15 min, the repressor is rapidly inactivated, and the recombination genes are expressed at high levels from the powerful λ*p*L promoter. Repressor is renatured and tight repression is restored by lowering the temperature after 15 min. The short induction time minimizes adventitious recombination and cellular stress[32,38]. An advantage of the natural phage system is that the λ repressor is autoregulated and, thus, better controlled than when recombination genes on multicopy plasmids are expressed from heterologous promoters, which are often leaky, causing unwanted expression and side effects. Red recombination does not require the *E. coli* RecA function[3], and when recombineering is performed in a strain with a *recA* mutation, all extraneous homologous recombination is prevented.

Owing to its high efficiency and short homology requirements, recombineering can be used for a wide range of applications (**Fig. 2**). Recombineering can be used to insert selectable or nonselectable markers in plasmids, bacterial chromosomal DNA or BACs. It can be used to generate gene-targeting constructs to be used for making knockout or knock-in alleles in embryonic stem cells. This technology is also suitable for generating transgenic reporter constructs using BACs to express *lacZ*, fusion tags (e.g., *GFP*, tandem affinity, FLAG, HIS and so on), site-specific recombinases (e.g., *Cre*, *Flp*), selectable markers (e.g., neomycin resistance gene) or any cDNA under the control of a tissue-specific promoter or to make fusion proteins. It can also be used to generate subtle alterations in BACs or bacterial chromosomal DNA without the use of any selectable marker or site-specific recombination system.

In spite of being a very tractable and versatile technology, recombineering has limitations. For example, as very short regions of homology are sufficient for the recombination, manipulating regions containing repetitive sequences can be problematic. As most recombineering-based methods use PCR-amplified products as substrate DNA, the recombinant products may occasionally acquire mutations. However, confirming the integrity of the recombinant target DNA by sequencing helps to discard such products. Another limitation of recombineering is that the sequence of the target region must be known. However, because the entire genome of many organisms has been sequenced, this is not a limitation for most commonly used organisms.

Experimental design

Each recombineering experiment involves the following six steps, which are illustrated in the general flowchart (**Fig. 2**).

(1) Generation of the appropriate linear targeting substrate DNA
(2) Provision of the λ Red recombination genes
(3) Induction of the λ recombination genes
(4) Preparation of electrocompetent cells and electroporation of the linear targeting substrate DNA
(5) Outgrowth following electroporation
(6) Identification and confirmation of the recombinant clones.

Substrate DNA design and generation. Depending upon the application, different DNA substrates are required (**Fig. 2**). The creation of appropriate double-stranded and single-strand linear targeting substrates for specific applications is described below and their construction is detailed in Step 3 of PROCEDURE.

Double-stranded DNA recombination. Linear dsDNA substrates for recombineering consist of a region to be inserted flanked by two homology arms. The inserted region may be either a selectable

marker or a nonselectable DNA that is used to replace a counter-selectable marker.

Recombineering can be used to insert a selectable marker (drug resistance gene or a prototrophic marker) into the bacterial chromosome or any episomal DNA. This is done by using linear DNA containing the desired selectable marker flanked by 50 bases of homology to the target site (**Figs. 1** and **3**). Such substrates can be made with the PCR by amplifying the selectable markers using a pair of chimeric primers, each about 70 bases in length. Each primer will have 50 bases at the 5′-end corresponding to the region to be targeted to provide homology and ~20 bases at the 3′-end to prime amplification of the selectable marker as shown in **Figure 3**. Although flanking 50 base homologies are more than sufficient for recombination, longer homologies of 150–200 bases have been used to further improve the targeting efficiency; however, this substantially increases the time and effort required[6,21]. To use longer homologies, each homology arm is amplified independently by PCR and directionally cloned into a plasmid so as to flank the selectable marker already present in that plasmid. High-fidelity *Taq* DNA polymerase with proofreading ability (such as Invitrogen High Fidelity Platinum *Taq* or Roche Expand High Fidelity) is used for generating dsDNA PCR products for recombineering substrates; for confirming constructs, standard *Taq* polymerase can be used. Try not to use supercoiled DNA as template for PCR amplification, as residual intact plasmid will give a high background of transformants, making it difficult to identify actual recombinants. The TKC strain (**Table 1**) contains the three genes for antibiotic resistance to tetracycline, kanamycin and chloramphenicol, as well as their regulatory elements. This strain can be used to amplify any of these markers with the colony PCR technique using the primer pairs specified in **Table 2** (see option A in Step 3 of PROCEDURE, **Fig. 3a**). Plasmid templates with selectable markers flanked by *loxP* and *FRT* sites are available through the NCI recombineering website (http://recombineering.ncifcrf.gov/; see option B in Step 3 of PROCEDURE, **Fig. 3a**). Drug cassettes with dual promoters that allow the use of the same selectable marker (e.g., neomycin, hygromycin and blasticidin) in both bacterial and mammalian cells are also available.

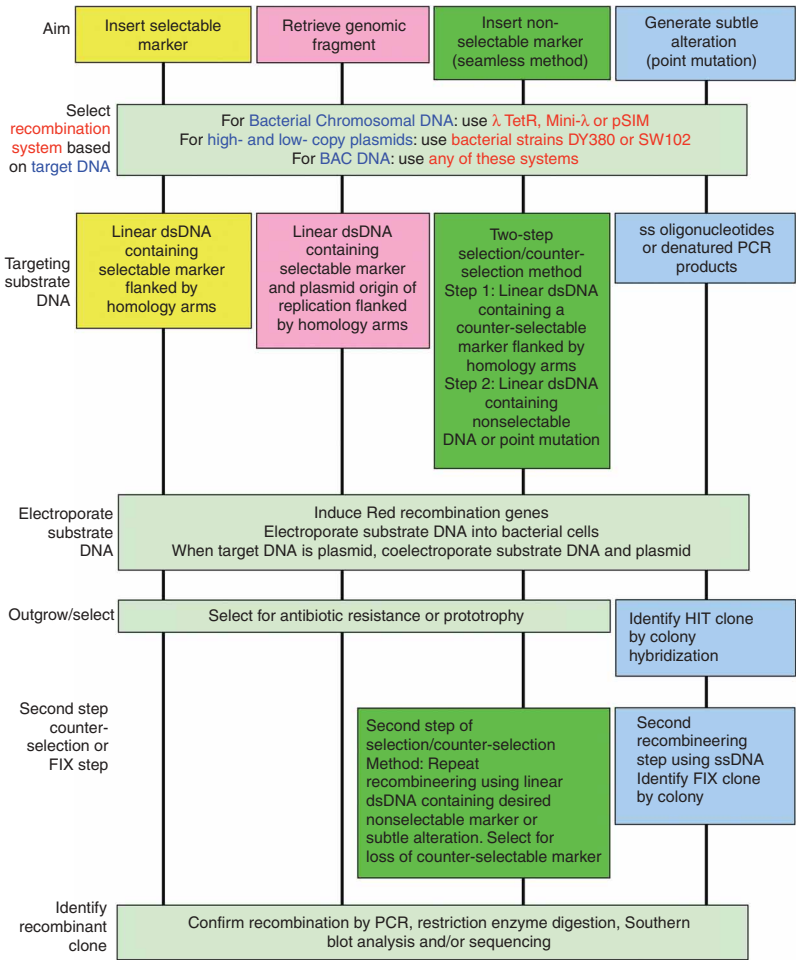


Figure 2 | A flowchart of recombineering procedures. Schematic representation of various steps involved in recombineering. An appropriate system should be selected on the basis of the choice of target DNA. The type of substrate DNA depends upon the choice of method used. The outgrowth procedures and methods to identify the recombinant clone are based on the use a selectable marker, selection/counter-selection method or lack of any selectable marker in the substrate DNA.

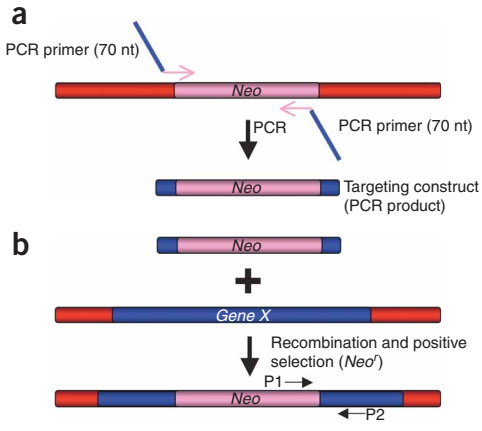


Figure 3 | Insertion of a selectable marker by recombineering. (**a**) Targeting construct can be generated by PCR to introduce the region of homology (in blue) and a selectable marker (e.g., *Neo*, kanamycin/neomycin-resistance gene). The PCR primers used to generate the targeting construct are 70-mer oligonucleotides with 50 nucleotides corresponding to the target site (e.g., *Gene X*, in blue) sequence to introduce the homology arm and 20 bases from the ends of the selectable marker (*Neo*, in pink). (**b**) The targeting construct is electroporated into the bacterial cells that are induced to express the phage recombination genes. Recombinant clones are selected as kanamycin-resistant colonies in this case and confirmed with PCR using test primers P1 and P2.

TABLE 1 | Bacterial strains used in recombineering.

Strain	Primary use	Relevant genotype	Reference	Available from
For introduction of BACs				
DH10B	Parent strain	*mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU pglΔ8 rpsL nupG*	55	Invitrogen
DY380	BAC recombineering with dsDNA	DH10B *λ(cI857ind1) Δ{(cro-bioA)<>tetRA} (TetR) gal490*	44	Court lab (court@ncifcrf.gov)
EL250	BAC recombineering: for removing selectable markers flanked by frt sites	DY380 *(cro-bioA) <> araC-PBAD Flpe*	44	The NCI recombineering website
EL350	BAC recombineering: for removing selectable markers flanked by loxP sites	DY380 *(cro-bioA) <> araC-PBAD Cre*	44	The NCI recombineering website
SW102	BAC recombineering: galK counter-selection	DY380 *ΔgalK*	40	The NCI recombineering website
SW105	BAC recombineering: galK counter-selection	EL250 *ΔgalK*	40	The NCI recombineering website
SW106	BAC recombineering: galK counter-selection	EL350 *ΔgalK*	40	The NCI recombineering website
For plasmid propagation and recombineering				
DH5α	Plasmid isolation	*φ80dlacZΔM15 Δ(lacZYA-argF)U169 deoR recA1 endA1 hsdR17(rk-, mk+) phoA supE44 thi-1 gyrA96 relA1*	56	Invitrogen
HME70	Plasmid recombineering: MMR-deficient strain for high-efficiency oligo recombination	W3110 *Δ(argF-lac)U169 galKtyr145UAG mutS<>cat Δ(srlA-recA)::Tn10 {λcI857 Δ(cro-bioA)}*	50	Court lab (court@ncifcrf.gov)
HME71	Plasmid recombineering	W3110 *Δ(argF-lac)U169 galKtyr145UAG Δ(srlA-recA)::Tn10 {λcI857 Δ(cro-bioA)}*	50	Court lab (court@ncifcrf.gov)
Drug cassette amplification				
TKC	Template for tetracycline, kanamycin and chloramphenicol	*tetA, cat, kan*		Court lab (court@ncifcrf.gov)
LE392	For propagation of λTet phage	*e14- glnV44 supF58 (lacY1 or ΔlacZY) galK2 galT22 metB1 trpR55 hsdR514(rK-mK+)*	NIH Strain Collection	Court lab (court@ncifcrf.gov)
DH10B-containing BACs		DH10B-containing BAC libraries		BACPAC Resources Center (BPRC)

Other strains for *E. coli* recombineering are listed elsewhere[12,13].

A nonselectable DNA fragment (e.g., a reporter gene, such as *lacZ*, a transgene, such as *cre*, *flp* or *GFP*, or a HIS-tag) can also be efficiently inserted into a genomic region using recombineering. There are two approaches to achieve this. The first approach is a two-step selection/counter-selection method that allows insertion of the desired DNA without leaving other undesirable modifications, in a 'seamless' event (**Fig. 4a**). Generally, a counter-selectable cassette is first inserted at the site to be changed; this cassette is then

TABLE 2 | PCR primers and possible source of template for drug cassette amplification[a].

Gene cassette	Source	Primer sequence	Annealing temperature (°C)	Size (kb)
Ampicillin (*amp*)	pBluescript SK(+) (Stratagene)	5′-CATTCAAATATGTATCCGCTC-3′ 5′-AGAGTTGGTAGCTCTTGATC-3′	53	1.2
Chloramphenicol (*cat*)	pPCR-Script Cam (Stratagene)	5′-TGTGACGGAAGATCACTTCG-3′ 5′-ACCAGCAATAGACATAAGCG-3′	53	0.86
Kanamycin (*kan*)	*Tn5*	5′-TATGGACAGCAAGCGAACCG-3′ 5′-TCAGAAGAACTCGTCAAGAAG-3′	55	0.95
Tetracycline (*tetA*)	*Tn10*	5′-TCCTAATTTTTGTTGACACTCTA-3′ 5′-CTCTTGGGTTATCAAGAGGG-3′	55	1.34

[a]When amplified with these primer pairs, all drug cassettes will contain a promoter and all but *kan* will have a transcriptional terminator.

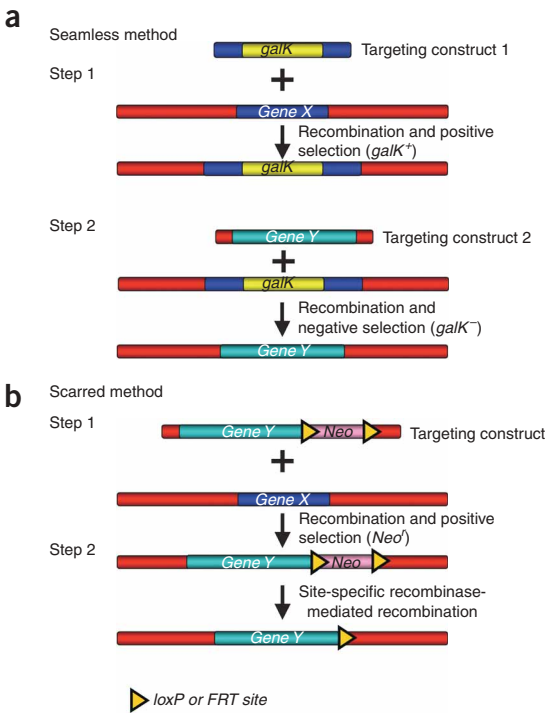


Figure 4 | Insertion of a nonselectable DNA fragment by recombineering. (**a**) 'Seamless' method to insert nonselectable DNA fragment makes use of selectable markers that can be used for positive as well as negative selection (e.g., *galK*). In this two-step method, first the selectable *galK* marker is targeted to the site (*Gene X*) where the nonselectable DNA fragment (*Gene Y*) is to be inserted. In the second step, a targeting construct containing the nonselectable DNA fragment flanked by the same 50 bp of homology to the target site is electroporated into Gal$^+$ bacterial cells containing the recombinant DNA from Step 1. Clones in which the *Gene Y* DNA fragment is correctly targeted are counter-selected for loss of the *galK* gene. (**b**) The scarred method: this method targets both the selected (*Neo*) and nonselected (*Gene Y*) DNAs jointly. In Step 1, the nonselectable DNA fragment (*Gene Y*) is introduced along with a selectable marker, Neo, which is flanked by *loxP* or *FRT* sites. Recombinants are selected for the presence of *Neo*. In Step 2, *Neo* is deleted by site-specific recombinase-mediated recombination (Cre for *loxP* sites and Flp for *FRT* sites). Unlike the 'seamless' method, a single *loxP* or *FRT* site is retained after recombination.

replaced with the desired alteration in a second recombineering reaction. In all such procedures, each step can be selected and the final construct will not have a drug marker or other genetic scar. A number of counter-selections are available: these include the *sacB*[39] gene linked with a drug marker, such as *cat* or *kan*, the *galK* gene[40], *rpsL*[41], *thyA*[42] and *tolC*[43]. In many situations, such as when using *galK*, the same gene can be selected both for and against (**Fig. 4a**). The *galK* counter-selection has been optimized for use with BACs and uses strains that have been deleted for the *galK* gene at its normal chromosome location; such strains include SW102, SW105 and SW106. A detailed protocol for the *galK* counter-selection has been published[40]. For convenience, the strains and plasmids used in the *galK* selection are listed in **Tables 1** and **3**. In the second approach, a selectable marker is linked with a nonselectable DNA fragment and the two are inserted together (**Fig. 4b**). If subsequent removal of the selectable marker is desired, it can be flanked with *lox*P or *FRT* sites, thus allowing excision by expression of the Cre or Flp proteins[44]. Template drug cassettes with flanking *loxP* and *FRT* sites that can be used for PCR amplification are also available from the National Cancer Institute recombineering website. As a single *loxP* or *FRT* site remains after the genetic manipulation, we refer to this as the 'scarred' method (**Fig. 4b**).

Recombineering is now routinely used as a subcloning technique. Whereas classical methods of subcloning using restriction enzymes and DNA ligase allow manipulation of small DNA fragments, recombineering allows precise retrieval of even a large DNA segment from a genomic fragment in a BAC or from the bacterial chromosome[21,44]. Recombineering allows the junctions of the subcloned fragment to be exactly determined on the basis of need rather than on chance location of restriction sites. For rapid subcloning by recombineering, the small retrieval vectors are generated with PCR using a pair of primers, each with 50 base homologies at the 5′-ends corresponding to the region to be retrieved and ~20 bases of plasmid sequence for PCR amplification (**Fig. 5**). As mentioned above, 150–200 bp homology arms can

TABLE 3 | Mini-λs and plasmids.

Mini-λ or plasmid	Relevant genotype	Antibiotics concentration (μg ml^{-1} of media)	Source or reference
Primary use: express red functions for recombineering			
Mini-λ Kan (kanamycin)	Contained in DH10B	30	31
Mini-λ Tet (tetracycline)	Contained in DH10B	12.5	31
Mini-λ Cat (chloramphenicol)	Contained in DH10B	12.5	31
Mini-λ Amp (ampilcillin)	Contained in DH10B	30	31
pSIM5	pSC101ts gam exo bet chloramphenicol resistance (CmR)	12.5	45
pSIM6	pSC101ts gam exo bet ampicillin resistance (AmpR)	100	45
pSIM7	pBBR1 gam exo bet chloramphenicol resistance (CmR)	12.5	45
pSIM8	pBBR1 gam exo bet ampicillin resistance (AmpR)	100	45
pSIM9	pRK2 gam exo bet chloramphenicol resistance (CmR)	12.5	45
pSIM17	pSC101ts gam exo bet blasticidin resistance (BsdR)	50	45
pSIM18	pSC101ts gam exo bet hygromycin resistance (HygR)	50	45
pSIM19	pSC101ts gam exo bet spectinomycin resistance (SpecR)	100	45
Primary use: amplify galK for counter-selection			
pgalK	pBluescript SK- pEm7-galK ampicillin resistance (AmpR)	100	40

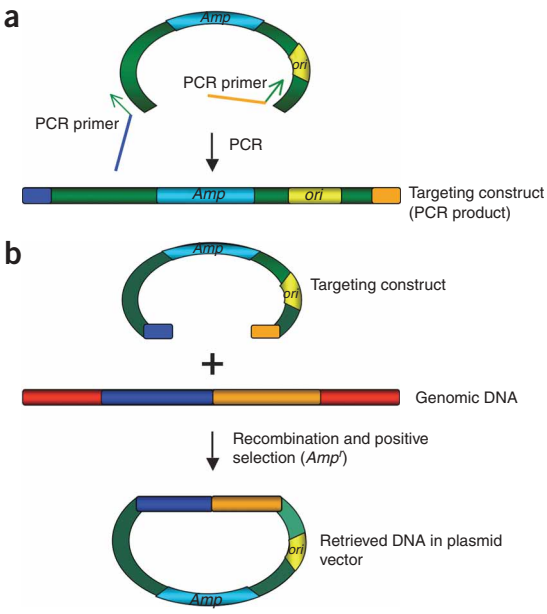


Figure 5 | Subcloning DNA fragments from genomic DNA. (**a**) To subclone or retrieve a genomic DNA fragment, generate a targeting construct by PCR. Each primer consists of 50 bases from the end of the genomic DNA that needs to be subcloned (blue and orange) and 20 bases from the plasmid sequence (in green) flanking the region of the origin of replication (*ori*) and an antibiotic resistance marker (*Amp*, ampicillin resistance gene). A linear plasmid DNA is used as template for PCR. (**b**) The linear targeting construct (PCR product) is electroporated into the bacterial cells that are induced to express the phage recombination genes. Recombination between the targeting construct and the genomic DNA results in the formation of a circular plasmid by gap repair. The circular plasmid contains the desired DNA fragment.

be cloned to improve retrieval efficiency if necessary[21]. A plasmid template must be used when amplifying the linear retrieval vector backbone, usually consisting of an origin of DNA replication and a selectable marker (option C in Step 3 of PROCEDURE, **Fig. 5a**). When the linear retrieval vector contains a selectable marker, however, a potential side reaction of end-joining can occur, generating a circular plasmid carrying only the selectable marker[5]. This is avoided by using a retrieval vector that contains only the origin of replication; in this approach, the selectable marker is first inserted adjacent to the target DNA to be retrieved. See Datta *et al.*[45] for details. The linear targeting construct recombines with the target DNA and retrieves the desired fragment by gap repair. This results in the generation of a circular plasmid (**Fig. 5b**). When gap repair is used for retrieving DNA, it is not necessary to know the DNA sequence of the entire piece to be retrieved, only that of the flanking region used as targeting homology. When the length of the fragment to be retrieved is less than 15 kb, a high-copy plasmid (e.g., pBluescript) can be used, but a lower-copy plasmid (e.g., pBR322) should be used for larger fragments[44]. In a variation of this method known as '*in vivo* cloning'[5], a linear DNA to be incorporated onto the plasmid, often generated by PCR, is coelectroporated with the linear plasmid. DNA can also be transferred by retrieval from one BAC to another as has been described[46,47].

Single-strand DNA recombination. Recombineering can also be performed using synthetic oligonucleotides (see Step 2 in PROCEDURE) or short denatured PCR products[2,8,9,48]. These oligos (ss-oligos) or ssDNAs can be used to create single base changes, insert or substitute short DNA sequences and generate deletions. Under optimal conditions, creation of small changes with recombineering using ssDNA is highly efficient and often completely eliminates the requirement for any selection. Two different variables contribute to creation of these optimal conditions. One variable is use of a 'lagging-strand' oligo, i.e., a single strand corresponding in sequence to the DNA chain that is replicated discontinuously[2,9]. Which of the two complementary oligos corresponds to the lagging-strand oligo will depend on the direction of DNA replication through the region of the chromosome or episome to be modified, and often it is easier to just try both strands; one should recombine with a 20- to 30-fold higher efficiency than the other. The other important variable includes avoidance of the *E. coli* methyl-directed mismatch repair (MMR) system: preventing MMR increases the effective recombination frequency about 100-fold[9,49]. As strains mutant for the MMR system acquire adventitious mutations, procedures that avoid the use of MMR mutants have been devised. The simplest of these tricks is to use an oligo that creates a C–C mispair when annealed to the target DNA, as the MMR system does not recognize and remove this particular mismatch. This method has limited utility, however, and other more generally useful two-step procedures are illustrated below.

A 'hit and fix' two-step recombineering approach[48], in combination with screening by colony hybridization, can be used to generate and find subtle changes in BAC DNA (**Fig. 6**). This approach is particularly useful for modifying DNA regions with repetitive motifs, which can be difficult to screen by PCR. This method uses short (~180 bp) dsDNA fragments generated by PCR that can be denatured to provide ss-oligonucleotide substrates for recombineering. In the first 'hit' step, a stretch of 20 nucleotides is changed, including the nucleotide(s) to be mutated in the final construct. The altered sequence is designed to contain recognition sites for a restriction enzyme. In the second 'fix' step, the bases that were modified in the first step are restored to the original sequence, with the exception of the desired mutation(s). As a run of nucleotides is changed at each step, the recombinant BACs can be detected by colony hybridization using a primer specific for the altered bases. Colony hybridization allows thousands of individual colonies to be screened at once. The presence or absence of the restriction site provides an additional tool to rapidly confirm both the 'hit' and 'fix' steps. The central 20-bp region can be the same for all 'hit' PCR products, allowing one 20 base universal probe to be used to identify the inserts at the 'hit' step. See **Box 1** for generation of 'hit and fix' substrates and Step 19D of PROCEDURE for identifying recombinants by the 'hit and fix' method.

Oligonucleotides can also be used to precisely delete unwanted regions of DNA by designing a 70-mer lagging strand oligo that spans the portion to be deleted, with 35 bases of homology at either side. The efficiency of deleting large segments is lower than that of engineering small single-base changes, and a selection or screen will be needed to identify recombinants[2,8,9,50]. As an example, a counter-selection can be used in combination with an oligo to generate deletions when needed. Alternatively, colony hybridization, using a 20-base probe spanning the deletion with ten bases on each side, can be used to identify recombinants.

Modifying multicopy plasmids. Recombineering is useful for making point mutations and small changes on plasmids[3,5,50]. For inserting or deleting larger pieces of DNA without a selection, classical cloning methods may sometimes be easier, but the precision of making junctions is usually lost. The plasmid to be engineered can be introduced into the Red-expressing cells by co-electroporation with the linear DNA after the recombination functions are induced; alternatively, it may be resident in the recombineering host. Coelectroporation of the plasmid with the modifying DNA, as described in Step 16 of PROCEDURE, is often preferable, as it helps control the number of plasmids present at the time of recombineering and may minimize opportunities for plasmid multimer formation. Recombineering will occur only in cells receiving the plasmid, so when introducing the plasmid by coelectroporation with the linear substrate, enough plasmid molecules must be added to obtain a high transformation frequency, but not so high as to introduce large numbers of plasmid molecules per cell. This is generally about 10 ng of plasmid DNA per electroporation, but the appropriate amount of plasmid DNA to add may have to be determined empirically by generation of a transformation efficiency curve, as the transformation efficiency of plasmids may differ (see **Box 2**). Ideally, each cell should receive about one plasmid molecule. If the plasmid is very large and has a low copy number, coelectroporation may not be practical; in this case, a resident plasmid should be used for modification.

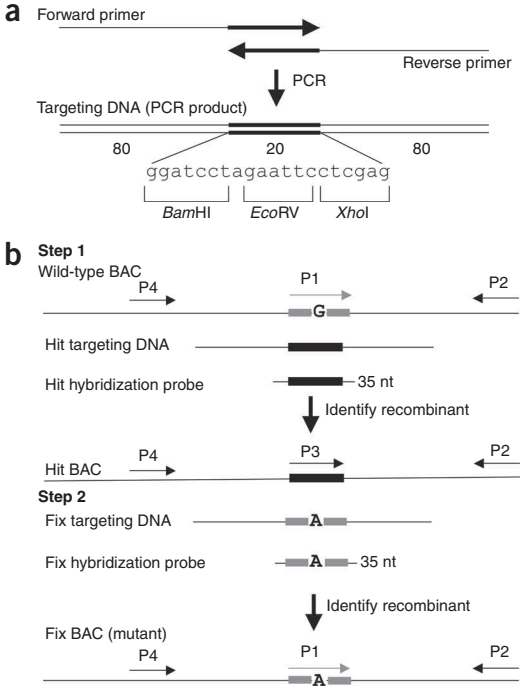


Figure 6 | Two-step 'hit and fix' method to generate subtle mutations using ss short PCR product or oligonucleotides as targeting vector. (**a**) The ss oligonucleotides containing 160 bases of homology and 20 unique bases are generated by using two 100-mer oligonucleotides in a PCR. The two 100-mer oligonucleotides have 20 complementary bases (in this case, the 20 bp contains restriction sites *Bam*HI, *Eco*RV and *Xho*I) at the 3′-end. The 180 bp PCR product can be denatured to obtain ss oligonucleotides that can be used as targeting construct. (**b**) Schematic representation of the two steps involved in 'hit and fix' method to generate subtle alterations (e.g., G to A) without the use of a selectable marker. In Step 1, a 180-mer ss oligonucleotide is used to replace 20 nucleotides (gray box) around the target site with 20 heterologous nucleotides (black box). Recombinants can be identified by colony hybridization using an end-labeled 35-mer oligonuclotide that can specifically anneal only to the recombinant DNA. A primer set specific for the heterologous 'hit' sequence (P3 and P2) can be used to confirm the presence of recombinant clones by PCR. A second primer set (P1 and P2) can be used as a control to amplify only the nonrecombinant DNA. Generation of a correct recombinant clone can be confirmed by digesting with *Bam*HI, *Eco*RV or *Xho*I the PCR product (approximately 300–500 bp) of primers P2 and P4. In Step 2, the 20 nucleotides are restored to the original sequence, except for the desired mutation. Such clones can be identified by colony hybridization using a 35-mer oligonucleotide as probe and further confirmed by PCR amplification using primers P1 and P2, by testing for loss of the restriction sites inserted in Step 1, by digesting the PCR product of primers P2 and P4 and by sequencing.

The intracellular copy number of commonly used plasmids such as pUC derivatives may exceed 500 copies per cell[51]. Such a high copy number means that most copies will not be altered during a recombineering reaction, and cells that contain recombinant plasmid molecules will also contain many unmodified parental molecules. To create pure clonal populations of the recombinant, it is necessary to separate the recombinant class from the unmodified class; this is done by isolating the plasmid DNA after recombineering and retransforming it at a low concentration of less than one molecule per cell. If plasmid DNA has been deleted during the recombineering, often the plasmid mixture can be digested with a restriction enzyme that cuts only the parental plasmid species but not the desired recombinant, thus enriching for the recombinant class.

When making changes on plasmids, it is important to begin the procedure with a pure monomer plasmid species. By the same token, it is also important to use a host defective in RecA-mediated homologous recombination to prevent plasmid multimer formation by host recombination pathways. However, it has been observed[3,50] that when plasmids are modified with recombineering,

BOX 1 | GENERATION OF 'HIT' AND 'FIX' PCR PRODUCTS

1. Synthesize a linear dsDNA cassette with PCR using two 100-mer oligonucleotide primers that overlap by 20 nucleotides (nt) at their 3′-ends, designing the primers so that the resulting DNA contains 80-nt homology arms flanking a 20-nt heterologous region. For each change to be created, the 'fix' DNA should contain the same homologies as the corresponding 'hit' DNA. Only the central region of the two PCR products will vary. The exact sequence of the 'hit' 20-nt central region is not critical as long as it is different from the wild-type sequence(s) and contains a restriction site not present in the flanking homology; however, we recommend using our universal 20 base sequence: 5′-GGATCCTAGAATTCCTCGAG-3′.
2. For the 'fix' DNA, replace the central 20-bp sequence with the endogenous sequence but include the desired mutational change(s).
3. Verify the size of the PCR products (180 bp) by agarose gel electrophoresis and purify them with a PCR clean-up kit.

BOX 2 | DETERMINATION OF PLASMID DNA TRANSFORMATION EFFICIENCY ● TIMING 1 D

1. Make electrocompetent cells as described in the protocol (PROCEDURE Steps 9–15), adding a different amount of DNA to each of several electroporations (suggested range is 0.1, 1, 10 and 100 ng).
2. Outgrow the electroporation mixes for 2 h and plate tenfold serial dilutions on Petri plates, selecting for the plasmid drug resistance. Use a range of dilutions to obtain countable number of colonies for each transformation.
3. The next day, count colonies and generate a semilog graph of the number of transformants (ordinate) versus the amount of DNA added (abscissa).
4. From the graph, determine the minimum amount of DNA necessary to achieve maximal transformation efficiency (i.e., without adding excess DNA), and use this amount of DNA for each recombineering reaction.

? TROUBLESHOOTING

some larger multimeric circles of these plasmids arise. As these multimers form during and are dependent upon the recombineering reaction, they cannot be totally eliminated. There is anecdotal evidence that other replicons, including BACs, may also form larger-order multimers during recombineering. Ways to deal with these multimers are suggested in the Troubleshooting section.

Provision of the λ Red recombination genes. There are several ways to provide the prophage system that will express the phage recombination proteins in the bacterial cells where recombineering is performed (**Fig. 6**). All variations allow highly efficient recombination, and the choice of which to use will depend on the target DNA (see **Fig. 2**). When high- or low-copy plasmids have to be manipulated, bacterial strains containing the prophage recombination system on the chromosome (e.g., DY380 and its derivatives, such as SW102) should be used. When the bacterial chromosome or BAC DNA is modified, mobile recombineering systems, such as the pSIM vectors, mini-λ or the replication-defective λ phage (λTetR), can be introduced into the bacterial cells to be engineered (**Fig. 7a–d**). BAC DNA can also be transformed into bacterial strains containing the repressed recombination system. In all cases, it is advisable to check the integrity of the BAC DNA by restriction analysis before proceeding with the recombineering procedure. The individual options are discussed in more detail below.

Bacterial strains with defective prophage; DY380 and derivatives. A number of bacterial strains have been generated that harbor a defective λ prophage that is stably situated in the bacterial chromosome (see Step 4A of PROCEDURE, **Fig. 7b** and **Table 1**). Such host strains are ideal for manipulation of plasmid DNA, which can be introduced into these bacterial strains. Another major advantage of using the strains containing the integrated λ prophage is that a drug selection need not be applied to maintain the recombineering system. Strains routinely used for BACs include DY380[44], SW102 and their derivatives[40]. To use these strains, the BAC or plasmid to be modified must be introduced into the recombineering strain by transformation (see **Box 3**). A mini-prep method described below

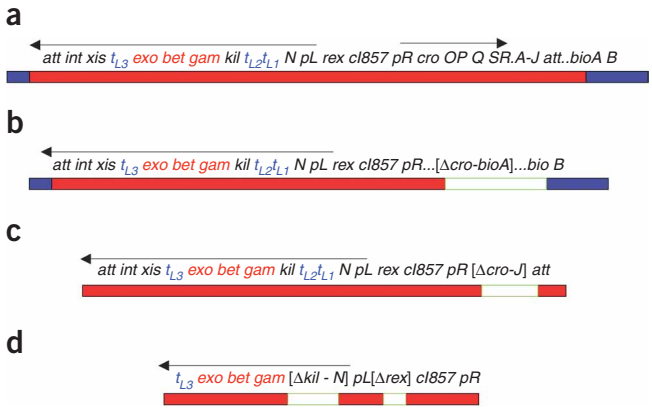


Figure 7 | Schematic representation of λ phage constructs used for recombineering. (**a**) The λ prophage with some of its genes is shown integrated in the bacterial chromosome. It is adjacent to the biotin gene *bioAB*. Bacterial DNA is shown as a blue bar; phage DNA is shown as a red bar; and a region containing a deletion and/or a substitution is shown as a white bar in **b**–**d**. The complete λ prophage is flanked by its attachment sites *att*, where integration occurred. The *int* and *xis* genes located in the *pL* operon encode functions to integrate and excise the phage DNA into and out of the bacterial chromosome. Two operons with their promoters *pL* and *pR* are indicated. Transcription of the promoters is controlled by the λ CI857 repressor. This repressor is temperature sensitive in that it is active and represses the promoter at 30–32 °C but is inactive at 42 °C allowing for transcription. The N protein functions as a transcription anti-terminator and prevents RNA polymerase termination of *pL* transcripts at terminators *tL1*, *tL2* and *tL3*. The red genes *exo*, *bet* and *gam* encoding the homologous recombination functions are also in the *pL* operon. The *kil* gene is adjacent to *gam* and, when expressed for over 1 h, kills the bacterial cell[32]. The replication genes *O* and *P* are in the *pR* operon. Cro functions as a partial repressor of the *pL* and *pR* operons when CI is inactive at 42 °C. The lysis and structural genes, *SRA–J*, are shown located beyond their regulator *Q*. The λ TetR phage used for recombineering is identical to the above complete prophage with the exception of four changes. The replication gene *P* has an amber mutation, the *cro* gene also has an amber mutation, and the *tetA* gene encoding tetracycline resistance replaces *rex*. (**b**) The defective prophages in DY380, SW102 and the HME strains have the lytic *pR* operon deleted from *cro* through the *bioA* genes as shown in brackets. The right *att* site is also deleted, preventing any excision of this prophage. In DY380 and SW102, *cro* through *bioA* is replaced by the tetracycline resistance cassette, *tetRA*. (**c**) The mini-λ phage DNA is shown with the lytic genes *cro* through *J* deleted. In different constructs of the mini λ, the *cro-J* region is replaced with various drug-resistant cassettes. Mini-λ has both *att* sites plus *int* and *xis*, which allows for its integration and excision. (**d**) The λ segment on the pSIM plasmids is shown. The plasmid backbone is not shown. The genes from *cro* to *att* are removed as well as genes beyond *tL3*, including *int* and *xis*. The *red* genes are connected directly to *pL* by a deletion that removes *kil* through *N*. The drug-resistant marker characteristic of each SIM plasmid replaces the *rex* gene adjacent to *cI857*. The basic features are conserved in all of these Red expression constructs. Red is left under the native phage-controlling elements for optimal expression and regulation. The *pL* promoter drives gene expression and is controlled by the temperature-sensitive but renaturable λ CI857 repressor. In all, the *cro* gene is inactive to maximize *pL* expression, and the replication genes are absent or inactive to prevent lethal effects on the cell.

BOX 3 | ELECTROPORATION OF BAC DNA INTO BACTERIAL STRAIN OF CHOICE ● TIMING 1 D

1. Pick an isolated colony from an LB plate and grow overnight in 3–5 ml of LB at 32 °C.
2. Next morning, add 0.5 ml of the culture to 25 ml of LB in a 250-ml flask and grow at 32 °C to an OD_{600} of 0.50–0.60. Transfer the culture to a 50-ml Oak Ridge tube and spin at 6,000*g* in prechilled rotor for 10 min at 4 °C.
3. Wash the cell pellet with 20 ml of ice-cold H_2O once and then resuspend in 1 ml of H_2O and transfer to a chilled 1.5-ml tube. Spin at 10,000*g* for 20–30 s at 4 °C.
4. Wash the cells two more times with 1 ml of ice cold H_2O. Resuspend the cell pellet in H_2O in a final volume of 100 µl and keep on ice.
5. Mix 100 ng of BAC DNA with 50 µl of electrocompetent cells and chill on ice for 5 min then transfer into a 0.1-cm cuvette. Introduce the BAC DNA into the cells by electroporation (1.8 kV, 25 µF capacitance and 200 Ω resistance). Also do a control electroporation where no DNA is added. After electroporation, immediately add 1 ml of LB and transfer the cells to a standard sterile culture tube. Incubate cells at 32 °C for 1 h. Spin down cells for 20–30 s in a microcentrifuge. Resuspend the pellets in 200 µl of LB. Plate each aliquot of cells on a single LB plate containing chloramphenicol (12.5 µg ml^{-1}) and incubate for 20–22 h at 32 °C. Only the BAC DNA electroporation should have colonies.
6. Pick 5–10 individual colonies and grow in 3 ml of LB containing 12.5 µg ml^{-1} chloramphenicol. Extract BAC DNA and perform restriction digestion with common enzymes like *Eco*RI, *Bam*HI, *Pst*I, *Hin*dIII, *Eco*RV. Run the sample on 0.8% agarose gel. Compare the restriction pattern with that of the original BAC DNA.

(see **Box 4**) works well for rapid extraction of BAC DNA and produces DNA that is suitable for restriction digestion, PCR analysis and sequencing[52]. For targeting both BACs and any multicopy plasmids, it is important to use *recA* mutant strains such as these.

Mobile recombineering systems; mini-λ and pSIM plasmids. There are mobile recombineering systems available that can be introduced into any bacterial strain to allow manipulation of the BAC or chromosomal DNA. Such mobile elements include 'mini-λ'[53] and the plasmids collectively called the pSIM vectors[45] (see Steps 4B and 4C of PROCEDURE, respectively). Both mini-λ and the pSIM plasmids provide the same endogenous phage-controlling elements and Red recombination functions present in the above-mentioned defective prophage strains (see **Fig. 7c,d**); however, these must be introduced into the BAC-containing strain. Mini-λ consists of a defective nonreplicating, circular phage DNA, which, once introduced into the strain, integrates into the bacterial chromosome using site-specific recombination[53]. The integrants are selected by the presence of the antibiotic resistance gene carried by the mini-λ. Once integrated, mini-λ is stable, replicates as part of the host chromosome and does not require drug selection for maintenance. The mini-λ includes the *pL* operon from which are expressed the *int* and the *xis* genes that catalyze the site-specific integration/excision events, as well as the *exo*, *bet* and *gam* recombination genes for recombineering (**Fig. 7c**). In addition to being easily introduced into any bacterial cell, the mini-λ can be readily excised to cure the cells of the phage DNA. The excised mini-λ DNA circle can also be purified from the induced bacterial cells using a standard plasmid purification protocol. The pSIM vectors[45] consist of the elements of the prophage necessary for recombineering (see **Fig. 7d**) carried on a number of different plasmid origins. Especially useful is a pSC101 plasmid derivative with temperature-sensitive DNA replication; this plasmid has a low copy number and the *ts* replicon allows loss of the plasmid after recombineering is completed. These plasmids are available with a variety of selectable antibiotic resistance markers (**Table 3**). Unlike the prophage and the mini-λ, they require drug selection for stable maintenance.

λ phage carrying tetracycline resistance gene; λTetR. A further means of expressing the Red system is from a λ phage carrying a tetracycline resistance gene[20] (see Step 4D of PROCEDURE). This phage system combines the advantages of stability and mobility. The genotype of this phage is λ *cI857 ind1 rexAB*<> *tetAR* $cro_{TYR26am}$ $P_{GLN59am}$, and it is referred to here as λTetR (see **Fig. 7a**). The λTetR is easy to prepare as a high titer phage lysate; using this lysate, lysogens are readily created by infection of the desired host and TetR selection. The efficiency of lysogeny is greater than that of plasmid transformation, and once introduced into the BAC-containing strain, the prophage is very stable and drug selection is no longer required.

BOX 4 | SMALL-SCALE MINIPREP FOR BAC DNA PREPARATION ● TIMING 1.5 H

1. Inoculate a single colony of DH10B cells containing the BAC into 3 ml of LB containing chloramphenicol (12.5 µg ml^{-1}) and grow overnight at 32 °C (because of the presence of the Red system) in a shaking incubator at 200–250 r.p.m.
2. Next day, pellet 1.5 ml of culture by centrifugation in a 1.5-ml tube at 10,000*g* at room temperature (20–25 °C) for 1 min.
3. Add 100 µl of chilled alkaline lysis solution I. Resuspend by pipetting the cells up and down using the pipette tip. Place on ice for 5 min.
4. Add 200 µl of freshly prepared alkaline lysis solution II. Gently invert the tubes until the lysate is clear. Place on ice for 5 min.
5. Add 150 µl of alkaline lysis solution III and mix gently until a white precipitate appears. Place on ice for 5 min. Centrifuge the tubes at 10,000*g* for 5 min at 4 °C and then transfer the supernatant to a clean 1.5-ml tube.
6. Add two volumes of 95% ethanol to the supernatant. Mix and place on ice for 30 min. Centrifuge at 10,000*g* for 10 min at 4 °C. Discard supernatant and wash the pellet with 70% ethanol. Air-dry the pellet and resuspend in 30–50 µl of 1× TE (10 mM Tris-HCl, pH 8.0; 1.0 mM EDTA).

Induction of the λ recombination genes. The Red system is induced by incubating the mid-log bacterial culture in a 42 °C water bath shaking at 200 r.p.m. for 15 min as detailed in Steps 6–8 of PROCEDURE. It is important to note that air shakers are not satisfactory for this purpose because the heat transfer is less efficient with an air shaker. Immediately after the heat pulse, the cells should be placed in an ice-water slurry for quick chilling. Failure to rapidly chill the cells may result in premature decay of the recombination activity. Once the cells are induced for the Red functions, it is best to complete the procedure promptly.

Preparation of electrocomptent cells and electroporation of the linear targeting substrate DNA (see Steps 9–18 of PROCEDURE). The chilled cells are gently washed twice with cold sterile water to remove salts that will cause problems during electroporation. The washed cells are suspended in a small volume of water and mixed with the linear DNA substrate, then aliquoted into chilled electroporation cuvettes. Luria broth (LB) is added immediately after electroporation to maximize recovery of viable cells.

Outgrowth following electroporation. The cells are fragile after electroporation and require a minimal 30-min recovery period in LB before plating. An outgrowth period may also be needed to allow expression of cloned genes, such as those encoding antibiotic resistance. The details of the outgrowth procedure will vary according to the type of recombineering reaction performed and how the recombinants will be identified. The various types of outgrowth are described in Step 19A–D of PROCEDURE.

Identification and confirmation of the recombinant clones. Drug-resistant recombinants are first selected using the appropriate antibiotic resistance and then analyzed with PCR to verify that the insertion has gone to the proper location. In certain cases, such as for multicopy plasmids, restriction analysis may be used to confirm the recombinant clones. Southern blots and sequencing may be used to analyze BAC or genomic recombinants. When a point mutation or other subtle change is made, DNA sequencing must be used to confirm correctness of recombinants. Recombinant identification is described in Steps 19A–D of PROCEDURE.

MATERIALS

REAGENTS

- One of the following is required to express the Red functions (see Experimental design for further details):
  - Bacterial strain containing defective λ prophage (see **Table 1**)
  - Mini-λ DNA (see **Table 3**)
  - pSIM plasmid (see **Table 3**)
  - λTetR phage with titer of at least 10^9 plaque-forming units ml^{-1}
- Ampicillin (A0166, Sigma)
- Kanamycin (K0254, Sigma)
- Chloramphenicol (C7795, Sigma)
- Tetracycline (T7660, Sigma)
- Hygromycin (H3274, Sigma)
- Sterile distilled H_2O, chilled on ice
- Minimal salts solution, such as M9
- Plasmid DNA isolation reagents (12163, Qiagen maxiprep kit; 27104, Qiagen miniprep kit)
- PCR purification kit (28104, Qiagen)
- Gel extraction kit to purify DNA from agarose gels (28604, Qiagen)
- High-fidelity Taq polymerase with proofreading ability (11304-029, Invitrogen)
- Standard *Taq* polymerase (11342-020, Invitrogen)
- dNTP mixture, 10 mM each, PCR grade, (10297-018, Invitrogen)
- Agarose (SeaKem LE, ISC Bioexpress)
- Restriction enzymes (New England Biolabs)
- Glycerol (BP229-1, Fisher Scientific)
- Chimeric primers for PCR amplification of recombineering substrates, 25 pmol μl^{-1} in H_2O (Integrated DNA Technologies)
- Bacto-tryptone (Difco)
- Yeast extract (Difco)
- One of the following sources of pure linear DNA for recombineering (see Experimental design for further details), suspended in sterile H_2O:
  - Purified PCR product with 50 bp of homology to the target, ~100 ng μl^{-1}
  - 70-mer ss-oligo (salt-free but otherwise unpurified) with desired changes(s) at center and with ~35 bases flanking homology, 0.5 pmol μl^{-1} (Integrated DNA Technologies)
  - Plasmid backbone with flanking homology to DNA to be retrieved, ~100 ng μl^{-1}
  - Lysozyme (Roche Molecular Biochemicals)

EQUIPMENT

- Constant temperature bacterial incubator set at 30–34 °C (2005 Low temp Incubator, VWR)
- Two shaking H_2O baths (200 r.p.m.) set at 30–32 °C and at 42 °C (G-76, New Brunswick Scientific)
- Spectrophotometer and cuvettes (DU 530, Beckman-Coulter)
- Electroporator (Genepulser II with Pulse Controller II, Bio-Rad)
- Electroporation cuvettes with 0.1-cm gap (870582, Bio-Rad), labeled and prechilled
- Floor model low speed centrifuge at 4 °C (Avanti, J-25I, Beckman)
- Refrigerated microcentrifuge at 4 °C (Micromax, RF, IEC)
- Thermal cycler and accessories for PCR (MyCycler, Bio-Rad)
- Agarose gel electrophoresis apparatus (Bio-Rad)
- Sterile 125- and 250-ml Erlenmeyer flasks, preferably baffled (2540-00125 and 2540-00250 respectively, Bellco)
- Sterile 35- to 50-ml centrifuge tubes (03530, Thermo Scientific)
- 1.5-ml microfuge tubes (022363204, Eppendorf)
- 0.2-ml flat cap PCR tubes (TFI0201, Bio-Rad)
- Insulated ice buckets (35754-100, VWR)
- Sterile glass culture tubes (16 mm × 150 mm) for overnight growth of bacterial cultures (89000-506, VWR)
- Stainless steel caps for culture tubes (60825-882, VWR)
- Pipetters of various volumes (Gilson) with aerosol-resistant sterile tips
- Petri plates, 100 mm × 15 mm (25834-302, VWR)
- Computerized DNA analysis program (Gene Construction Kit, Textco Biosoftware; Vector NTI, Invitrogen)

REAGENT SETUP

M9 medium Mix 3.0 g of KH_2PO_4, 12.8 g of $Na_2HPO_4 \cdot 7H_2O$, 1.0 g of NH_4Cl, 0.5 g of NaCl and 1 liter of distilled H_2O

LB Mix 10 g of Bacto-tryptone, 5 g of yeast extract, 5 g NaCl and 1 liter H_2O, pH 7.2. ▲ CRITICAL Both M9 and LB should be autoclaved for sterility and stored at room temperature (20–25 °C).

Alkaline lysis solution I 25 mM Tris-HCl, pH 8.0, 50 mM glucose, 10 mM EDTA, pH 8.0, 2.5 mg/ml of Lysozyme and 100 μg/ml RNase A

Alkaline lysis solution II 0.2 N NaOH and 1% Sarcosyl

Alkaline lysis solution III 5 M Potassium acetate, pH 4.8

EQUIPMENT SETUP

LB and Petri plates LB and Petri plates (100 mm × 15 mm) containing 1.5% (wt/vol) Difco agar, with antibiotics as needed (see **Table 4**)

TABLE 4 | Antibiotics and concentrations (μg ml^{-1}) used in media for selection.

Antibiotic	Single copy[a]	Multicopy plasmids
Ampicillin	30	100
Kanamycin	30	50
Chloramphenicol	10	20
Tetracycline	12.5	25
Hygromycin	Not used in single copy	50
Spectinomycin	100	100
Blastocidin	50	50

[a]Use these concentrations for drug markers on BACs or the *E. coli* chromosome and for single-copy plasmids.

PROCEDURE

Design and generation of linear DNA substrate for recombineering

1| Before beginning recombineering, precisely design your desired DNA sequence *in silico*. This is greatly facilitated by use of computerized DNA analysis programs, such as Gene Construction Kit (Textco Software) or Vector NTI (Invitrogen).
▲ CRITICAL STEP Design the desired molecule on the computer before ordering primers!

2| From the *in silico* design of the new DNA molecule, generate the sequence(s) of required primers or 70-mer ss-oligos (see Experimental design) and order or synthesize them as salt-free and otherwise unpurified.
▲ CRITICAL STEP Be sure to include the 50 bases of flanking homology to the 5′-ends of your primers (see Experimental design).
▲ CRITICAL STEP ss-oligos can be directly used as substrates for recombineering after appropriate dilution.

3| To generate dsDNA substrates for recombineering by PCR, different templates can be used as follows: option A for bacterial DNA (**Fig. 3a**), option B for plasmid DNA (**Fig. 3a**) and option C for plasmid origin amplification (**Fig. 5a**).

(A) PCR-amplification of dsDNA substrate for recombineering from bacterial DNA ● TIMING **3–4 h**

(i) Antibiotic resistance genes containing tetracycline, kanamycin and chloramphenicol cassettes can be amplified from the TKC strain using 'colony PCR' and primer pairs listed in **Table 2**. To perform colony PCR, prepare a PCR without allowing any volume for the template DNA; for a typical 50-µl PCR, using the high-fidelity Platinum *Taq* kit from Invitrogen, mix the following in a 0.2-ml PCR tube using sterile technique (for multiple PCRs, prepare a master mix and dispense aliquots as appropriate):
- 5 µl of 10× PCR buffer
- 2 µl of 50 mM $MgSO_4$
- 1 µl of 10 mM dNTP mixture
- 1 µl of each primer (25 pmol μl^{-1} in H_2O)
- 0.5 µl of platinum *Taq* high fidelity
- 39.5 µl of sterile distilled H_2O

(ii) Lightly touch a fresh colony of the *E. coli* strain with a sterile inoculating loop and mix cells into the PCR.

(iii) Use the following PCR program suitable for amplifying any of the drug cassettes referred to in **Table 2** (using optimal conditions for Invitrogen Platinum *Taq* high fidelity):
- Initial denaturation: 94 °C, 5 min
- Subsequent denaturation: 94 °C, 30 s
- Anneal primers: 53 °C, 30 s
- Extension: 68 °C, 1.5 min
- Repeat the last three steps 29 times
- Final extension: 68 °C, 10 min
- Hold at 4 °C

(iv) Visualize the PCR on an agarose gel as described elsewhere[11].

(v) Remove salt and clean up the PCR product by ethanol precipitation as described elsewhere[11] or by using a commercially available kit and following the manufacturer's instructions.

(vi) Suspend the product in small volume of H_2O so as to have a concentration of ~100 ng μl^{-1}.

(B) PCR-amplification of dsDNA substrate for recombineering from plasmid DNA ● TIMING **3–4 h**

(i) Plasmid templates must be linearized before PCR amplification. Digest plasmid DNA with a restriction enzyme at sites present in the vector sequence but absent in the sequence to be amplified[11].

(ii) Run an aliquot of the DNA on an agarose gel to verify complete digestion[11].
▲ CRITICAL STEP Be aware that super-coiled plasmid migrates at ~70% of the size of a full-length linear species.

(iii) If necessary, cut the band containing the linearized DNA from the gel and extract the DNA[11].

(iv) Resuspend the linear DNA in 20–30 µl TE (pH 8.0), check the concentration using a spectrophotometer (read absorbance at 260 nm) and dilute to 0.5–1.0 ng μl^{-1}. A purified fragment containing just the drug cassette can be stored and reused as template.

(v) Amplify the drug cassette with a pair of chimeric primers, using PCR conditions based on information in **Table 2** and a *Taq* DNA Polymerase with proofreading ability.

(vi) Examine the PCR product by agarose gel electrophoresis[11].

(vii) Gel-purify the desired band[11] to eliminate contamination of supercoiled plasmid.

(viii) If linear plasmid DNA was used as template and the product was not gel-purified, digest the PCR with 3–5 U of the modification-dependent restriction enzyme *Dpn*I per µg DNA (at 37 °C for 2 h), which will cut the plasmid DNA but not the unmodified PCR product. *Dpn*I digestion may not totally eliminate plasmid background, however.
▲ CRITICAL STEP Do not expose the PCR product to direct UV light, which may damage it and result in abnormal recombination frequencies.

(ix) Clean up the PCR product as described in Step 3A(v).

▲ CRITICAL STEP Do not use supercoiled DNA as template for PCR amplification. To avoid having to gel-purify the DNA or treat with *Dpn*I, use the TKC strain as a template when possible.

(C) Generation of dsDNA substrate for recombineering by PCR amplification of plasmid origin for gene retrieval

● TIMING **3–4 h**

(i) Digest the plasmid with one or more restriction enzyme(s) that do not cut within the region to be amplified.

(ii) Amplify the origin and selectable marker from the linear plasmid by PCR using the primers generated in Step 2 (reaction conditions will need to be established empirically). Use the least amount of linear plasmid possible for the PCR template, to minimize residual uncut plasmid.

(iii) Digest the completed PCR with *Dpn*I to further remove the template plasmid as in Step 3B(viii).

(iv) Purify the PCR product to remove salt before proceeding (as in Step 3A(v)). The amplified product will be a linear plasmid with flanking homology to either side of the region to be rescued from the chromosome. It is possible to avoid the potential side reaction of plasmid end-joining by using an alternative approach: amplify a retrieval vector containing only the origin of replication, after first inserting a selectable marker different than the one on the starting plasmid next to the target DNA to be retrieved[45]. See **Box 1** for generation of 'hit' and 'fix' PCR products.

? TROUBLESHOOTING

Provision of the λ Red recombination genes

4| Provide the λ recombination genes from the defective prophage present in DY380 and its derivatives (option A), the mini-λ (option B), a pSIM plasmid (option C) or the replication-defective λTet phage (option D).

▲ CRITICAL STEP It is critical to realize that in all cases, the Red genes are under control of a temperature-sensitive repressor, and strains containing them should always be propagated at a low temperature (30–32 °C) except during induction of the Red system.

(A) Provision of the λ Red recombination genes from DY380 and derivatives

(i) DY380 and derivatives already contain the Red functions and so require no further preparation before using for recombineering. See **Box 3** for electroporation of BAC DNA into bacterial strain of choice and **Box 4** for small-scale miniprep for BAC DNA preparation.

(B) Provision of the λ Red recombination genes from Mini-λ ● TIMING **2 d**

(i) To prepare the mini-λ DNA, inoculate a single colony of *E. coli* with an integrated mini-λ prophage into 125 ml of LB containing appropriate antibiotics in a one-liter flask and culture overnight at 30–32 °C.

(ii) The next day, incubate the culture at 42 °C for 15 min to induce excision of the mini-λ DNA circles.

(iii) Chill the culture on ice for 15 min and extract the mini-λ DNA using a standard plasmid midi-prep purification kit (e.g., Qiagen) as described previously[53].

(iv) Introduce the mini-λ into the desired *E. coli* strain that will be used for recombineering by mixing 1 μl of mini-λ DNA (25–35 ng) with 25 μl of electrocompetent bacterial cells and introduce the DNA by electroporation. Also do a control electroporation of cells alone without added DNA.

(v) Grow the transformed cells at 32 °C for 1 hour in LB medium and plate dilutions on an LB agar plate containing the appropriate antibiotic to select the strain with mini-λ (see **Table 4**, use antibiotic concentration indicated for single-copy plasmids). Incubate overnight at 32 °C.

(vi) Pick isolated colonies for recombineering and test for the presence of the prophage by confirming that the strain is temperature sensitive for colony formation on antibiotic plates at 42 °C.

(C) Provision of the λ Red recombination genes from pSIM plasmids ● TIMING **1 d**

(i) Introduce the pSIM plasmid DNA into the bacterial strain where recombineering will be performed by selecting for the appropriate antibiotic resistance of the plasmid and plating the cells at 30–32 °C (see **Table 4**, use antibiotic concentration indicated for multicopy plasmids).

(ii) Purify a single colony 30–32 °C and grow an overnight culture from it with the appropriate antibiotic using methods described elsewhere[11].

(D) Provision of the λ Red recombination genes from replication-defective λTet phage ● TIMING **1 d**

(i) Grow a 1-ml culture of the appropriate bacterial strain that will be used for recombineering overnight in LB, adding chloramphenicol (10 μg ml^{-1}) if a BAC is present.

(ii) The next day, pellet the cells by centrifuging at 10,000*g* at 20–25 °C for 1 min in a microfuge and suspend them in 0.1 ml of 10 mM $MgSO_4$.

(iii) Add 1 μl of high titer phage stock (at least 10^9 ml^{-1}) to the cells and incubate at 32 °C for 20 min.

(iv) Add 1 ml of LB and incubate for 1 h at 30–32 °C, then plate a series of tenfold dilutions on LB-Tet plates (tetracycline concentration: 12.5 μg ml^{-1}) and incubate at 30–32 °C overnight.

(v) Purify tetracycline-resistant colonies by streaking the bacteria and picking isolated colonies on LB-Tet plates.

? TROUBLESHOOTING

5| Inoculate the bacterial strain containing the Red functions from a single colony into 3–5 ml of LB medium. Incubate the culture at 32 °C overnight with aeration.
▲ CRITICAL STEP Maintain antibiotic selection for the BAC (chloramphenicol 10 μg ml^{-1}), and if a pSIM plasmid is used, also include the antibiotic for the pSIM (see **Table 4**, use antibiotic concentration indicated for multicopy plasmids).
■ PAUSE POINT Once the recombineering strains are created, they can be frozen and stored indefinitely for future use at −70 °C as a glycerol stock prepared by mixing 1 ml of a freshly grown overnight culture with 1 ml of 50% glycerol (dilute glycerol in water (vol/vol) and autoclave).

Induction of the λ recombination genes ● TIMING **3 h**
6| Add 0.5 ml of the overnight culture to 35 ml of LB medium in a 250-ml baffled Erlenmeyer flask. Supplement with appropriate antibiotic to maintain plasmids, if needed. Use antibiotic concentration indicated in **Table 4**.
▲ CRITICAL STEP Dilute the overnight culture by at least 70-fold.

7| Place the flask in the 32 °C H_2O bath and grow cells with shaking for about 2 h or until the cells reach an OD_{600} of 0.4–0.6.
▲ CRITICAL STEP The growth time may vary with different strains. The cells are ready when the OD_{600} is 0.4–0.6. It is important not to overgrow the cells, as stationary-phase cells are not optimal for recombineering.

8| Transfer half of the culture to a 125-ml Erlenmeyer flask and place the flask to shake in a 42 °C H_2O bath; keep the other flask at 32 °C. Shake for 15 min at 200–220 r.p.m. The culture at 42 °C is induced for recombination functions and the 32 °C culture is the uninduced control.
? TROUBLESHOOTING

Preparation of electrocompetent cells and electroporation of the linear targeting substrate DNA ● TIMING **2 h**
9| Immediately after the 15-min induction, rapidly chill both cultures in an ice-water slurry; swirl the flasks gently. Leave on ice for 5–10 min. Label and chill two 35- to 50-ml centrifuge tubes for each set of induced and uninduced cells.

10| Transfer both the induced and uninduced cultures to the centrifuge tubes and centrifuge at 4,600*g* (6,700 r.p.m. in a Sorvall SA-600 rotor) for 7 min at 4 °C. Using sterile technique, aspirate or pour off supernatant.

11| Add 1 ml of ice-cold sterile distilled H_2O to the cell pellet and gently suspend cells with a large pipette tip (do not vortex). Add another 30 ml of ice-cold distilled H_2O to each tube, seal and gently invert to mix, again without vortexing. Centrifuge tubes again as mentioned in Step 10.

12| Promptly decant the 30-ml supernatant very carefully from the soft pellet in each tube and gently suspend each cell pellet in 1 ml of ice-cold distilled H_2O.
▲ CRITICAL STEP Remove tubes from the centrifuge promptly at the end of spin. The pellet is very soft and care should be taken not to dislodge it or lose the cells, especially when processing multiple tubes. If necessary, leave a little supernatant in the tube.
? TROUBLESHOOTING

13| Transfer the suspended cells to pre-chilled microcentrifuge tubes. Centrifuge at 10,000*g* in refrigerated microcentrifuge for 30 s at 4 °C. Carefully aspirate the supernatant and suspend cells in 1 ml of ice-cold H_2O, centrifuge again and aspirate supernatant, being extremely careful with the pellet.

14| Suspend the cell pellet in 200 μl of sterile cold distilled H_2O and keep on ice until used.
▲ CRITICAL STEP For highest efficiency, use freshly processed cells.

15| To introduce the linear DNA recombineering substrate (from Step 3) into the electrocompetent cells (from Step 14) by electroporation, chill the desired number of labeled 0.1-cm electroporation cuvettes on ice. Turn on the electroporator and set to 1.80 kV, 25 μF capacitance and 200 Ω resistance.
? TROUBLESHOOTING

16| In labeled microcentrifuge tubes on ice, mix 1 μl of DNA from Step 3 (~100 ng of salt-free PCR fragment or 0.5 pmol (~100 ng) ss-oligo) with 50 μl of electrocompetent induced or uninduced cells from Step 14. When modifying a multicopy plasmid introduced by coelectroporation, also add 1 μl of the multicopy plasmid DNA (~20 ng μl^{-1}) at this point. Include the following controls:
- Induced cells without DNA. If colonies are present on this control plate, either the selection is not working properly or the cells have a high reversion frequency for the property selected.
- Uninduced cells plus DNA. This is a control to estimate the number of background colonies due to some contaminating factor in the DNA, such as intact plasmid template from a PCR.

▲ CRITICAL STEP Never leave the DNA–cell mixes on ice for more than ~10 min.
? TROUBLESHOOTING

17| Transfer the DNA–cell mixes to chilled electroporation cuvettes as quickly as possible and introduce the DNA into the cells by electroporation (according to settings outlined in Step 15).
▲ CRITICAL STEP Make sure that the time constant is greater than 5 ms for optimal results. Lower time constants generally indicate impurities or salts in the cells or the DNA.
? TROUBLESHOOTING

18| Immediately after electroporation, add 1 ml of LB medium to the cuvette. Do this before proceeding to the next electroporation. After the electroporations are completed, transfer the electroporation mixes in LB to sterile culture tubes and incubate with shaking at 32 °C for a minimum of 30 min.
▲ CRITICAL STEP Failure to allow 30 min outgrowth gives poor cell viability.

Outgrowth following electroporation, identification and confirmation of the recombinant clones ● TIMING 1 d to 2 weeks
19| The outgrowth and screening procedures will vary according to the type of recombineering reaction: use option A to select recombinant clones by antibiotic resistance or prototrophy, option B to select recombinant clones containing a genomic fragment subcloned by recombineering, option C to select modified multicopy plasmids and option D to screen 'hit and fix' recombinants by colony hybridization.

(A) Outgrowth and selection of recombinant clones by antibiotic resistance selection or prototrophy ● TIMING 2 d

(i) Continue outgrowth in 1 ml of LB without any antibiotics for at least 2 h at 30–32 °C to allow for expression of the antibiotic resistance or metabolic gene to be selected.

(ii) Following the outgrowth, make six 1:10 serial dilutions of the experimental cultures in a buffered medium lacking a carbon source, such as M9 salts.

(iii) To select recombinants, spread 0.1 ml of the undiluted culture and the 10^{-1} and 10^{-2} dilutions on plates selective for the recombinant at 30–32 °C. Also assay total viable cells by plating 0.1 ml of the 10^{-4}, 10^{-5} and 10^{-6} dilutions on LB; determining this viable cell count will allow calculation of a recombinant frequency. If the number of viable cells is too low (less than 10^{7} ml^{-1}), recombinants may not be found. For the control cultures, both the uninduced (32 °C) and the induced (42 °C) to which no DNA was added, plate 0.1 ml of the undiluted culture on a single selective plate.
▲ CRITICAL STEP Do not add chloramphenicol for BAC maintenance while outgrowing or selecting the recombinant on plates.

(iv) Incubate plates until colonies appear (generally 22–24 h for rich plates at 32 °C). Generally no or only a few colonies are obtained on the control plates, indicating that most of the colonies on the experimentally induced plus DNA plates represent true recombinant clones.

(v) Once potential recombinant clones are identified, confirm the construct by PCR analysis, restriction digestion analysis and sequencing where necessary. For confirming insertion of a dsDNA, such as a drug cassette, use two pairs of primers, each pair having one primer outside the targeted flanking region and one primer in the inserted DNA, and amplify the two junctions. Another PCR, using the two outside flanking primers, should also be performed to confirm the absence of the gene to be removed, thus ruling out the possibility of a duplication event[54].
? TROUBLESHOOTING

(B) Outgrowth and selection of recombinant colonies containing genomic fragments retrieved onto a plasmid by subcloning using recombineering: ● TIMING 2 d

(i) Following electroporation, add the 1 ml cell mixture to 9 ml of LB medium and grow the culture overnight nonselectively at 32 °C.

(ii) The next day, isolate plasmid DNA using a standard miniprep protocol. It is not necessary to visualize this DNA by agarose gel electrophoresis.

(iii) Introduce the plasmid DNA into a high-efficiency cloning strain by transformation, selecting the drug resistance of the plasmid or the retrieved selectable marker. Use a low concentration of DNA to minimize uptake of multiple plasmids into the same cell.

(iv) Purify drug-resistant clones.

(v) Grow 5 ml of overnight cultures from ~20 isolates with antibiotic selection.

(vi) Isolate plasmid DNA and examine by restriction analysis.
? TROUBLESHOOTING

(C) Modifying multicopy plasmids ● TIMING 3 d

(i) Outgrow the electroporated cultures in at least 1 ml of LB with aeration for at least 2 h at 30–32 °C before applying selection.

(ii) After outgrowth, add 9 ml of LB and drug for plasmid selection and grow overnight at 30–32 °C with aeration.

(iii) Isolate plasmid DNA using a standard mini-prep procedure.

(iv) When possible, eliminate parental plasmid by cutting with a restriction enzyme that does not cut the recombinant molecule.

(v) Introduce plasmid DNA by transformation into a *recA* strain of *E. coli* (e.g., DH10B) at a low DNA concentration (less than one plasmid per cell). Generally, the amount of DNA should be around 0.1 ng. See **Box 2** for determination of plasmid DNA transformation efficiency.

(vi) Select or screen for desired phenotype. Possible methods of screening include restriction digestion, identification of plasmids with altered size as assayed by migration on agarose gels, sequencing and PCR analysis.

BOX 5 | TESTING COLONY PURITY BY RESTRICTION DIGESTION

For this approach, the region including the modification site is amplified using PCR primers hybridizing outside the region of homology. The purified PCR product should be digested with a restriction enzyme recognizing the 'hit' cassette (**Fig. 6**). Complete digestion of the PCR product indicates that the recombinant BAC is correct and is a pure culture.

▲ CRITICAL STEP Some of the recombinant plasmid molecules will have undergone multimer formation. Avoid these. If monomer recombinants are not apparent, a recombinant multimer can be digested with a restriction enzyme having a unique site in the monomer plasmid and the resulting monomer DNA can be ligated at a low DNA concentration.

(vii) Confirm the desired change by sequencing where necessary.

? TROUBLESHOOTING

(D) Screening 'hit and fix' recombinants with colony hybridization ● TIMING 2 weeks

(i) After 30 min of outgrowth at 32 °C, dilute the cell suspension 10^{-2} and 10^{-3} in LB. Spread 200 µl of each diluted suspension onto a 150 mm × 10 mm (large) agar plate containing appropriate antibiotic for BACs selection. Try to achieve uniform distribution of colonies throughout the plate. Incubate plates for 18–22 h at 32 °C. Approximately 1–5 × 10^3 colonies are expected.

(ii) Identify recombinant clones by colony hybridization using a γ-^{32}P-end labeled (or a nonradioactively labeled) oligonucleotide (35-mer) containing the 20 nt heterologous sequence as described in detail by Ausubel *et al.*[11]

(iii) Using a bacteriological loop, pick bacterial colonies giving a positive signal and suspend them in 1 ml of LB media. Screen at least 3–5 positive areas for each BAC modification. As the recombinant colonies also contain parental cells and because many colonies are close or touching each other on the original plate, a second screening step is required.

(iv) Make 10^{-2}, 10^{-3} and 10^{-4} dilutions of a bacterial suspension for each colony picked.

(v) Plate 50 µl of each dilution on a 15 mm × 100 mm agar plate containing the appropriate antibiotic for BAC selection. Incubate the plates at 32 °C for 18–22 h.

(vi) Select plates with 20–50 clearly isolated colonies and repeat colony hybridization.

(vii) Pick two well-isolated positive colonies from each plate for further analysis. Confirm the purity of the clones by colony PCR. Set up two PCRs for each clone using primers (20 nt) that will specifically amplify either the recombinant ('hit' clone) or the nonrecombinant DNA (**Fig. 6**). Test a total of ten individual colonies corresponding to the 3–5 positive areas on the original plate for each BAC modification.

▲ CRITICAL STEP Only one of the two reactions should be positive if the colony is pure. Occasionally, some colonies will amplify PCR products with both sets of primers, implying that at least one modified and one unmodified copy of the BAC is present. These colonies may be due to BAC multimer formation and should be discarded.

(viii) Once correctly targeted pure 'hit' colonies have been identified, repeat the preceding Steps 19D (ii–vii) used for the 'fix' step, beginning with the newly created 'hit' strain.

▲ CRITICAL STEP When using either the mini-λ or the pSIM plasmids for recombineering, streak the correct 'hit' clones on the appropriate antibiotic plate to determine whether they have retained the λ Red system. If necessary, reintroduce the Red system before proceeding.

(ix) Confirm the presence of correct mutation after the 'fix' step by sequencing.

(x) Digest both the original and modified BACs with a few commonly used restriction enzymes (e.g., *Bam*HI, *Eco*RI, *Eco*RV, *Hind*III, *Pst*I and so on). Confirm the integrity of the new BAC by comparing restriction digestion pattern with the original BAC. See **Box 5** for testing colony purity by restriction digestion and **Box 6** for detection of selectable marker flanked by *loxP* or *FRT* sites.

? TROUBLESHOOTING

BOX 6 | DELETION OF SELECTABLE MARKER FLANKED BY *loxP* OR *FRT* SITES

1. Grow an isolated colony of the recombinant strain containing the correct insertion of the transgene with its selectable marker flanked by *lox*P or *FRT* sites at 32 °C overnight in 3 ml of LB containing appropriate antibiotics.
2. Dilute 1 ml of the culture into 50 ml of LB containing antibiotics in a 250-ml flask and grow at 32 °C.
3. Once the culture reaches an OD_{600} of 0.3, add 500 µl of 10% filter-sterilized Arabinose to a final concentration of 0.1%. This induces the *flp* gene in SW105 or the *cre* gene in SW106.
4. Grow the culture at 32 °C until it reaches an OD_{600} of 0.5. Plate 100 µl of 10^{-2}, 10^{-3}, 10^{-4} dilutions on LB plates containing chloramphenicol (12.5 µg ml^{-1}) to maintain the BAC and grow overnight at 32 °C.
5. Screen for clones that have lost the selectable marker by site-specific recombination, these can be identified by antibiotic sensitivity. Streak approximately 50–100 colonies in duplicate on LB plates and LB containing the drug. Colonies that grow on LB plates but not in the presence of the drug are the desired clones that have recombined and retain only a single *lox*P or *FRT* site.
6. Confirm the loss of the selectable marker by PCR and sequencing.

● TIMING

Steps 1–3, design and generation of linear dsDNA substrates for recombineering by PCR: 3–4 h
Steps 4 and 5, introduction of the λ Red recombination system: 1–2 d
Steps 6–8, cell growth and induction of the λ Red genes: 3 h
Steps 9–14, preparation of electrocompetent cells: 1 h
Steps 15–18, electroporation of the linear DNA: <1 h
Step 19: outgrowth, identification and confirmation of the recombinant clones: 1 d to 2 weeks

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 5**.

TABLE 5 | Troubleshooting table.

Step	Problem	Possible cause(s)	Possible solution(s)
19A	No recombinants obtained with drug selection	Inadequate outgrowth time	Allow greater than 2 h outgrowth if needed. Increase LB volume so that more cell doublings occur
	No recombinants or transformants with double drug selection	Double drug selection issues	Try to use only one drug selection at a time rather than combining drugs. Do not maintain selection for BAC or pSIM plasmid when selecting drug-resistant recombinants
19B	No DNA retrieved, only empty vector	Vector is recircularizing by microhomologies	Either redesign the primers or place a selectable marker next to the DNA to be retrieved and retrieve with plasmid *ori* only
19C	Difficulty obtaining pure clones: multicopy plasmids	Contamination with parental plasmid	Transform DNA prep at low DNA concentration (less than one molecule per cell), select or screen for recombinant
	No recombinants	Poor plasmid transformation efficiency	Determine plasmid transformation efficiency as described in optional procedure
		Heteroallelic multimer	Either screen other candidates or digest heteroallelic multimer with a unique cutter and religate at a low DNA concentration
	Multimers with plasmid recombineering	Inherent nature of recombineering	Screen other isolates
			Use restriction digestion to convert to monomer, retransform and screen for recombinant plasmids
		Use of RecA$^+$ host	Use only a *recA* mutant host for plasmid modification
		Plasmid was a multimer initially	Start with a pure monomer plasmid species
19A–C	Colonies in uninduced control reactions	Incomplete digestion of plasmid template used for PCR, so intact plasmid is transformed (Step 3)	Digest plasmid before using as template, digest PCR product with *Dpn*I following amplification
		Leaky selection	Alter selection, e.g., increase drug concentration
		Contamination	Use sterile technique
			Test solutions for source of contamination and discard if necessary
19A–D	No recombinants obtained	Problem with bacterial strain (Step 4)	Never place recombineering strains at temperatures above 35 °C except during induction
			For strains containing defective prophage and λTet, verify that strain is temperature sensitive by plating dilutions at 32 and 42 °C
			For strains containing mini-λ and pSIMs, verify drug resistance

(continued)

TABLE 5 | Troubleshooting table (continued).

Step	Problem	Possible cause(s)	Possible solution(s)
			Do control recombineering reaction (i.e., drug cassette insertion) to check strain
		Cell loss during washing for electroporation (Step 12)	Titer viable cells as well as plating for recombinants—if viables are less than 10^7 ml^{-1}, recombinants may not be obtained. Be careful not to lose cells during H_2O wash steps
		Problem with induction procedure (Step 8)	Using shaking H_2O baths for both 32 °C growth and 42 °C induction, do not use air shaker. Confirm temperatures of H_2O baths. Do not exceed 15 min, 42 °C induction. Be sure to quick-chill flasks in ice bath after induction. If scaling up procedure, maintain 1/10 media/flask ratio
		Problem with linear DNA targeting construct (Step 3)	Remake PCR product or obtain new oligos
		Problem with electrocompetent cells (Steps 15–17)	Test cells by transforming supercoiled plasmid
			Use correct electroporation conditions. Be sure that the time constant for electroporation is ~5 ms. Add LB immediately after electroporation
		Drug concentration too high (see **Table 4**)	Be sure to use antibiotic concentrations for single-copy molecules when appropriate
19A–D	Difficulty in obtaining pure clones: BAC, PAC or *E. coli* chromosome	Incomplete colony purification	Restreak for single colonies and retest
	Difficulty in obtaining pure clones: BAC or PAC	Possible multimer formation	Screen more candidates

Troubleshooting guide for Recombineering (also see http://recombineering.ncifcrf.gov/faq.asp).

ANTICIPATED RESULTS

Recombineering is an efficient and relatively straightforward method of *in vivo* genetic engineering that is rapidly becoming routine for modification of the bacterial genome, plasmids, PACs and BACs. As recombineering allows DNA to be precisely inserted, altered or deleted, it is very useful for generation of various genomic constructs for functional studies. Linear DNA, both PCR products and synthetic oligonucleotides (oligos) can be used as substrates to create null mutations in genes as well as more subtle genetic changes. Generally, recombinants are obtained at a frequency of 10^{-4}–10^{-5} when dsDNA is used to make insertions. For optimized recombineering with ss-oligos or largely homologous denatured PCR fragments, the frequency of recombinants is in the range of 0.1–10%. The frequency of cloning by gene retrieval is lower, in the range of 10^{-5}–10^{-6}. Insertion of reporter genes at the 5′- or 3′-ends of genes cloned in BACs can be easily achieved. There are a number of critical steps in the recombineering procedure; each step should be optimized. Deviations from the protocol may result in suboptimal results; the overall effect of not optimizing each step can be a reduction in recombination efficiency and even failure to generate recombinants.

ACKNOWLEDGMENTS We thank Nina Costantino and James Sawitzke for helpful comments. This research was supported by the Intramural Research Program of NIH, National Cancer Institute, Center for Cancer Research. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract N01-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Published online at http://www.natureprotocols.com/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Copeland, N.G., Jenkins, N.A. & Court, D.L. Recombineering: a powerful new tool for mouse functional genomics. *Nat. Rev. Genet.* **2**, 769–779 (2001).
2. Ellis, H.M., Yu, D., DiTizio, T. & Court, D.L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci. USA* **98**, 6742–6746 (2001).
3. Yu, D. *et al.* An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **97**, 5978–5983 (2000).
4. Zhang, Y., Buchholz, F., Muyrers, J.P. & Stewart, A.F. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat. Genet.* **20**, 123–128 (1998).
5. Zhang, Y., Muyrers, J.P., Testa, G. & Stewart, A.F. DNA cloning by homologous recombination in *Escherichia coli*. *Nat. Biotechnol.* **18**, 1314–1317 (2000).

6. Murphy, K.C. Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*. *J. Bacteriol.* **180**, 2063–2071 (1998).
7. Murphy, K.C., Campellone, K.G. & Poteete, A.R. PCR-mediated gene replacement in *Escherichia coli*. *Gene* **246**, 321–330 (2000).
8. Swaminathan, S. *et al.* Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis* **29**, 14–21 (2001).
9. Costantino, N. & Court, D.L. Enhanced levels of λ Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. USA* **100**, 15748–15753 (2003).
10. Datta, S., Costantino, N., Zhou, X. & Court, D.L. Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc. Natl. Acad. Sci. USA* **105**, 1626–1631 (2008).
11. Ausubel, F.M. *et al.* (eds.) *Current Protocols in Molecular Biology* Vol. 1, Chapters 1 and 6 (John Wiley & Sons Inc., Hoboken, NJ, 2008).
12. Sawitzke, J.A. *et al.* Recombineering: *in vivo* genetic engineering in *E. coli*, *S. enterica*, and beyond. *Methods Enzymol.* **421**, 171–199 (2007).
13. Thomason, L.C. *et al.* In *Current Protocols in Molecular Biology* Vol. 1, Chapter 1, Unit 16 (eds. Ausubel, F.M. *et al.*) 1–24 (John Wiley & Sons Inc., Hoboken, NJ, 2007).
14. Sternberg, N.L. Cloning high molecular weight DNA fragments by the bacteriophage P1 system. *Trends Genet.* **8**, 11–16 (1992).
15. Shizuya, H. *et al.* Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**, 8794–8797 (1992).
16. Muyrers, J.P. *et al.* In *Methods in Molecular Biology: Bacterial Artificial Chromosomes Volume. 2: Functional Studies* Vol. 256 (eds. Zhao, S. & Stodolsky, M.) 107–121 (Humana Press, Totowa, NJ, 2004).
17. Heaney, J.D., Rettew, A.N. & Bronson, S.K. Tissue-specific expression of a BAC transgene targeted to the Hprt locus in mouse embryonic stem cells. *Genomics* **83**, 1072–1082 (2004).
18. Kuznetsov, S. *et al.* RAD51C deficiency in mice results in early prophase I arrest in males and sister chromatid separation at metaphase II in females. *J. Cell Biol.* **176**, 581–592 (2007).
19. Yang, Y., Swaminathan, S., Martin, B.K. & Sharan, S.K. Aberrant splicing induced by missense mutations in BRCA1: clues from a humanized mouse model. *Hum. Mol. Genet.* **12**, 2121–2131 (2003).
20. Chan, W. *et al.* A recombineering based approach for high-throughput conditional knockout targeting vector construction. *Nucleic. Acids Res.* **35**, e64 (2007).
21. Liu, P., Jenkins, N.A. & Copeland, N.G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13**, 476–484 (2003).
22. Karu, A.E., Sakaki, Y., Echols, H. & Linn, S. The gamma protein specified by bacteriophage λ. Structure and inhibitory activity for the RecBC enzyme of *Escherichia coli*. *J. Biol. Chem.* **250**, 7377–7387 (1975).
23. Murphy, K.C. λ Gam protein inhibits the helicase and chi-stimulated recombination activities of *Escherichia coli* RecBCD enzyme. *J. Bacteriol.* **173**, 5808–5821 (1991).
24. Karakousis, G. *et al.* The β protein of phage λ binds preferentially to an intermediate in DNA renaturation. *J. Mol. Biol.* **276**, 721–731 (1998).
25. Kmiec, E. & Holloman, W.K. β protein of bacteriophage λ promotes renaturation of DNA. *J. Biol. Chem.* **256**, 12636–12639 (1981).
26. Muniyappa, K. & Mythili, E. Phage λ β protein, a component of general recombination, is associated with host ribosomal S1 protein. *Biochem. Mol. Biol. Int.* **31**, 1–11 (1993).
27. Cassuto, E. & Radding, C.M. Mechanism for the action of λ exonuclease in genetic recombination. *Nature New Biol.* **229**, 13–16 (1971).
28. Little, J.W. An exonuclease induced by bacteriophage λ. II. Nature of the enzymatic reaction. *J. Biol. Chem.* **242**, 679–686 (1967).
29. Carter, D.M. & Radding, C.M. The role of exonuclease and β protein of phage λ in genetic recombination. II. Substrate specificity and the mode of action of lambda exonuclease. *J. Biol. Chem.* **246**, 2502–2512 (1971).
30. Cassuto, E., Lash, T., Sriprakash, K.S. & Radding, C.M. Role of exonuclease and β protein of phage λ in genetic recombination. V. Recombination of λ DNA *in vitro*. *Proc. Natl. Acad. Sci. USA* **68**, 1639–1643 (1971).
31. Court, D.L., Sawitzke, J.A. & Thomason, L.C. Genetic engineering using homologous recombination. *Annu. Rev. Genet.* **36**, 361–388 (2002).
32. Sergueev, K., Yu, D., Austin, S. & Court, D. Cell toxicity caused by products of the p_L operon of bacteriophage lambda. *Gene* **272**, 227–235 (2001).
33. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97**, 6640–6645 (2000).
34. Wang, J. *et al.* An improved recombineering approach by adding RecA to λ Red recombination. *Mol. Biotechnol.* **32**, 43–53 (2006).
35. Guzman, L.M., Belin, D., Carson, M.J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose pBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
36. Court, D.L., Oppenheim, A.B. & Adhya, S.L. A new look at bacteriophage λ genetic networks. *J. Bacteriol.* **189**, 298–304 (2007).
37. Dodd, I.B., Shearwin, K.E. & Egan, J.B. Revisited gene regulation in bacteriophage lambda. *Curr. Opin. Genet. Dev.* **15**, 145–152 (2005).
38. Murphy, K.C. & Campellone, K.G. Lambda Red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic *E. coli*. *BMC Mol. Biol.* **4**, 11 (2003).
39. Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T. & Kado, C.I. Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J. Bacteriol.* **164**, 918–921 (1985).
40. Warming, S., Costantino, N., Court, D.L., Jenkins, N.A. & Copeland, N.G. Simple and highly efficient BAC recombineering using *galK* selection. *Nucleic. Acids Res.* **33**, e36 (2005).
41. Rivero-Muller, A., Lajic, S. & Huhtaniemi, I. Assisted large fragment insertion by Red/ET-recombination (ALFIRE)—an alternative and enhanced method for large fragment recombineering. *Nucleic. Acids Res.* **35**, e78 (2007).
42. Wong, Q.N. *et al.* Efficient and seamless DNA recombineering using a thymidylate synthase A selection system in *Escherichia coli*. *Nucleic. Acids Res.* **33**, e59 (2005).
43. DeVito, J.A. Recombineering with *tolC* as a selectable/counter-selectable marker: remodeling the rRNA operons of *Escherichia coli*. *Nucleic. Acids Res.* **36**, e4 (2008).
44. Lee, E.C. *et al.* A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**, 56–65 (2001).
45. Datta, S., Costantino, N. & Court, D.L. A set of recombineering plasmids for gram-negative bacteria. *Gene* **379**, 109–115 (2006).
46. Kotzamanis, G. & Huxley, C. Recombining overlapping BACs into a single larger BAC. *BMC Biotechnol.* **4**, 1 (2004).
47. Venken, K.J., He, Y., Hoskins, R.A. & Bellen, H.J. P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science* **314**, 1747–1751 (2006).
48. Yang, Y. & Sharan, S.K. A simple two-step, 'hit and fix' method to generate subtle mutations in BACs using short denatured PCR fragments. *Nucleic Acids Res.* **31**, e80 (2003).
49. Li, X.T. *et al.* Identification of factors influencing strand bias in oligonucleotide-mediated recombination in *Escherichia coli*. *Nucleic Acids Res.* **31**, 6674–6687 (2003).
50. Thomason, L.C., Costantino, N., Shaw, D.V. & Court, D.L. Multicopy plasmid modification with phage lambda Red recombineering. *Plasmid* **58**, 148–158 (2007).
51. Sambrook, J. & Russell, D.W. *Molecular Cloning: A Laboratory Manual/Third Edition* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2001).
52. Sinnett, D., Richer, C. & Baccichet, A. Isolation of stable bacterial artificial chromosome DNA, using a modified alkaline lysis method. *Biotechniques* **24**, 752–754 (1998).
53. Court, D.L. *et al.* Mini-λ: a tractable system for chromosome and BAC engineering. *Gene* **315**, 63–69 (2003).
54. Bubunenko, M., Baker, T. & Court, D.L. Essentiality of ribosomal and transcription antitermination proteins analyzed by systematic gene replacement in *Escherichia coli*. *J. Bacteriol.* **189**, 2844–2853 (2007).
55. Durfee, T. *et al.* The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. *J. Bacteriol.* **190**, 2597–2606 (2008).
56. Hanahan, D. Techniques for transformation of *E.coli*. In *DNA cloning techniques: a practical approach* **1** (ed. Glover, D.M.) 109–135 (IRL Press, Oxford, 1985).

OPEN ACCESS Freely available online

PLoS one

AAV Recombineering with Single Strand Oligonucleotides

Matthew L. Hirsch[1], Francesca Storici[3], Chengwen Li[1], Vivian W. Choi[4], R. Jude Samulski[1,2]*

1 UNC Gene Therapy Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **2** Department of Pharmacology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **3** School of Biology, Georgia Institute of Technology, Atlanta, Georgia, United States of America, **4** Harvard Medical School, Cambridge, Massachusetts, United States of America

Abstract

Adeno-associated virus (AAV) transduction initiates a signaling cascade that culminates in a transient DNA damage response. During this time, host DNA repair proteins convert the linear single-strand AAV genomes to double-strand circular monomers and concatemers in processes stimulated by the AAV inverted terminal repeats (ITRs). As the orientation of AAV genome concatemerization appears unbiased, the likelihood of concatemerization in a desired orientation is low (less than 1 in 6). Using a novel recombineering method, Oligo-Assisted AAV Genome Recombination (OAGR), this work demonstrates the ability to direct concatemerization specifically to a desired orientation in human cells. This was achieved by a single-strand DNA oligonucleotide (oligo) displaying homology to distinct AAV genomes capable of forming an intermolecular bridge for recombination. This DNA repair process results in concatemers with genomic junctions corresponding to the sequence of oligo homology. Furthermore, OAGR was restricted to single-strand, not duplexed, AAV genomes suggestive of replication-dependent recombination. Consistent with this process, OAGR demonstrated oligo polarity biases in all tested configurations except when a portion of the oligo targeted the ITR. This approach, in addition to being useful for the elucidation of intermolecular homologous recombination, may find eventual relevance for AAV mediated large gene therapy.

Citation: Hirsch ML, Storici F, Li C, Choi VW, Samulski RJ (2009) AAV Recombineering with Single Strand Oligonucleotides. PLoS ONE 4(11): e7705. doi:10.1371/journal.pone.0007705

Editor: Jean Peccoud, Virginia Tech, United States of America

Received June 29, 2009; **Accepted** September 28, 2009; **Published** November 2, 2009

Funding: This work was supported by NIH grants GM0529299, HL066973, HL051818, and AI072176 awarded to R.J.S and a UNC Lineberger Fellowship to M.L.H. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: RJS@med.unc.edu

Introduction

Adeno-associated virus (AAV) is a single-strand (ss) DNA dependovirus of the family, Parvoviridae. The 4.7 kb genome consists of two genes, *rep* (replication) and *cap* (capsid), flanked by 145 nucleotide (nt) inverted terminal repeats [ITRs, 1]. AAV transduction results in both polarity genomes within the host nucleus where they form duplexed DNA, either by annealing or by extension of the free 3′ end of the ITR [2]. The generation of the duplexed genomes is perhaps coincident with their circularization, preferably as monomers but also as concatemers [3–5]. The circularization events are stimulated by the ITRs, and the created double-strand episomes are stable and capable of long-term gene expression.

The past 20 years has demonstrated the modulation and effective application of AAV for gene therapy in humans. This approach relies on replacing the viral genes with a therapeutic cassette such that only the ITRs remain [referred to as recombinant or rAAVs, 6]. Regarding *in vivo* DNA delivery for therapeutic applications, rAAV is perhaps currently the most desirable method as the virus is considered non-pathogenic, has a minimal tendency for chromosome integration, results in sustained gene expression and the capsid can be designed for targeted transduction [7]. However, a major limitation of rAAV is its small packaging capacity (≈4.7 kb), rendering it deficient for therapies requiring large promoters and or transgenes. To overcome this limitation, research has exploited host cellular machinery to concatemerize distinct genomes delivered by separate capsids, resulting in a larger DNA molecule. Two slightly different approaches have been used *in vitro* and *in vivo* termed overlapping vectors, and more successfully, *trans*-splicing AAV vectors [referred to herein as split vectors as splicing technically occurs in *cis*; 4, 5, 8]. In fact, the later approach has demonstrated therapeutic relevance in a mouse disease model [9]. However, as AAV split vector genome concatemerization is unbiased, the efficiency of functional transgene reconstruction is low. An additional rate-limiting step is the resultant ITR junction, which impedes transcription and splicing of the reconstructed transgene up to 50% [10]. Therefore, improvement on the split vector approach should address the ability to exclusively direct productive dimerization while eliminating the inhibitory ITR junction.

Several reports demonstrate that AAV genomes induce a DNA damage response in the host cell including; I) AAV infection induces G2/M arrest [11], II) the DNA repair proteins involved in non-homologous end joining (NHEJ) and homologous recombination, Ku86 and Rad52 respectively, directly bind AAV genomes [12], III) microinjection of ITRs induces ATM checkpoint activation similar to DNA double-strand breaks [DSBs; 13], IV) AAV2 DNA triggers a damage response similar to a stalled replication fork which depends on ATR and Chk1 signaling [14,15], and V) DNA-PKcs and Artemis process the AAV ITRs [16]. Collectively, these results provide evidence for the two canonical DNA repair pathways, NHEJ and homologous

recombination, in the circularization/concatemerization of AAV genomes.

The efficiency of homologous recombination is stimulated several orders of magnitude by a DNA DSB within the repair target [17]. Consistently, DNA inverted repeat sequences are processed to DSBs [18] and stimulate gene correction with a homologous substrate, such as a single-strand oligonucleotide [oligo; 18, 19]. In the context of AAV concatemerization, it is possible that homologous recombination is also stimulated following processing of the ITRs [16]. Therefore, we hypothesized that the orientation of AAV concatemerization could be directed using a single-strand oligo complementary to distinct genomes functioning in ITR repair.

This report describes a novel recombineering method: Oligo-Assisted AAV Genome Recombination (OAGR). In this approach, a DNA oligo with homology to distinct transduced AAV genomes directs concatemerization via homologous recombination. Oligos targeting different regions of the transduced genomes varied in their ability to stimulate intermolecular recombination, which in some cases approached the efficiency of single vector transduction. Interestingly, oligo polarity biases for OAGR were observed similar to other oligo-mediated DNA repair processes in bacteria, yeast and other human contexts [20–23]. Characterization of the recombination junctions suggests a process in which the oligo homology tethers distinct single-strand genomes and host enzymes generate a single larger molecule precisely at that position. OAGR, in addition to being useful for the elucidation of intermolecular homologous recombination in a human context, may find eventual relevance for AAV mediated large gene therapy.

Results

To test the hypothesis that DNA oligos homologous to separate AAV genomes enhance functional concatemerization we initially employed ssAAV2 split *gfp* vectors (Fig. 1*A*). These consist of two distinct genomes, each having ITRs, with the respective features (termed vectors A and B throughout): A. a CMV promoter followed by the first 381 nt of *gfp* coding sequence at which point a canonical splice donor site was added and B. the human chorionic gonadotropin (*hcg*) intron and the remaining *gfp* coding sequence followed by a poly-adenylation sequence (Fig. 1*A*). Transduction by either vector alone does not result in functional GFP (Fig. 2*A*). We chose to evaluate this effect in human embryonic kidney cells (293s) due to the high efficiency of AAV2 transduction and DNA oligo transfection using PEI. Results from other gene targeting experiments demonstrated that DNA 80-mers served efficiently for chromosomal and episomal repair (Fig. S1; data not shown, DNS). Therefore, we exploited the function of 80-mers for the majority of experiments. As negative controls, several oligos without split vector genome homology were evaluated and similar results were obtained, the value of which is defined as natural concatemerization (Fig. 2*A*, no homology; DNS).

The initial oligo tested contained 40 nt of 3′ terminal *gfp* coding sequence homology to vector A and 40 nt of 5′ *gfp* coding sequence homology on vector B (Fig. 1*C*, GFP). Co-transduction of the ssAAV2 split *gfp* system, at a multiplicity of infection (MOI) of 200 for each vector, followed by transfection of the forward orientation (sense = non-template strand = Fwd) of the GFP oligo resulted in a significant increase (3 to 4 fold) of GFP+ cells compared to the non-homologous control after 72 h (Fig. 2*A*). The anti-sense version (Rev) of the same oligo did not alter the amount of GFP+ cells significantly from the non-homologous control oligo (Fig. 2*A*).

Oligo-Assisted Intermolecular AAV Genome Recombination Depends on Oligo Target Homology

Next, oligos with different AAV genome targets were evaluated; i) 40 nt of homology to unique single-strand intron sequences of each genome (Fig. 1*C*, intron A-intron B = I_a-I_b), ii) an oligo with 40 nt of homology to the duplexed AAV Rep binding element (RBE) present in all AAV ITRs and 40 nt of homology to single-strand intron sequence exclusively on vector B (Fig. 1*C*, hairpin-intron B = H-I_b) and iii) an oligo with the described H homology as an inverted repeat capable of binding all ITRs (H-H, Fig. 1*C*; Table S1). Co-transduction by the split single-strand *gfp* vectors followed by the transfection of the Fwd versions of I_a-I_b or H-I_b resulted in an 8-fold or 10-fold increase in the amount of GFP+ cells, respectively, over the non-homologous control after 72 hours (Fig. 2). A decrease in GFP+ cells, below the non-homologous control, was observed after transfection of the H-H oligo in either orientation (Fig. 2*A*). Similar to the Gfp Rev oligo result, the I_a-I_b Rev oligo did not perform as well as the Fwd version with only a modest enhancement of GFP+ cells (Fig. 2*A*). Interestingly, this was not the case for the H-I_b Rev oligo which increased the amount of GFP+ cells to the same extent as the equivalent Fwd version (10-fold).

It should be noted that the OAGR effect was found independent of the oligo manufacturer and comparable with standard desalting and PAGE purification methods (DNS). Additionally, split *gfp* AAV2 particles purified from cesium chloride gradients and different chromatographies were all equally subject to OAGR (DNS). Furthermore, OAGR effects were independent of the AAV capsid serotype (Fig. S2).

The effect of the H-I_b oligo on *gfp* reconstruction after ssAAV2 split *gfp* co-transduction was compared to the transduction efficiency of a single AAV2 vector containing the exact genetic elements of the split system, CMV-*gf-hcg* intron-*p*-polyA, on a single-strand genome. It was first noted that the H-I_b Fwd or Rev oligos did not alter the GFP+ transduction profile, or GFP abundance, of the single intact particle (Fig. 2*C*; Fig. S3). Then, the enhancement of the H-I_b Fwd oligo on split *gfp* vector co-transduction was compared to the efficiency of the intact *gfp* single vector in the presence of the same oligo. Following transduction (or co-transduction) with an equivalent number of viral genomes, there was an approximate 2-fold increase in GFP+ cells for the single intact *gfp* vector compared to cells treated with single-strand split *gfp* vectors and the H-I_b as measured by flow cytometry (Fig. S3). Quantitation of the GFP directly by Western analysis demonstrates that OAGR increases the efficiency of split vector gene reconstruction greater than 60-fold (Fig. 2C). This OAGR effect resulted in 28% of the GFP produced by the single intact vector while natural concatemerization produced only 0.4% (Fig. 2C).

The time course of the H-I_b Fwd oligo effect on split *gfp* vector co-transduction was compared to a non-homologous control oligo. In this experiment, the cells were maintained for continuous replication. An increase in GFP+ cells was noted after 42 h for cells treated with the H-I_b oligo. Thereafter, the number of GFP+ cells increased with time in a linear fashion to over 20-fold greater than the control at day 5 (Fig. 3*A*). Southern blot analysis of intracellular oligo persistence demonstrated the greatest abundance at 24 h and stability out to day 5 under the PEI transfection conditions used herein (Fig. S4).

Oligo-Assisted Intermolecular AAV Genome Recombination is Dependent on MOI

Next, the effect of H-I_b Fwd oligo was characterized at increasing MOIs of the AAV2 split *gfp* vectors. Three days after

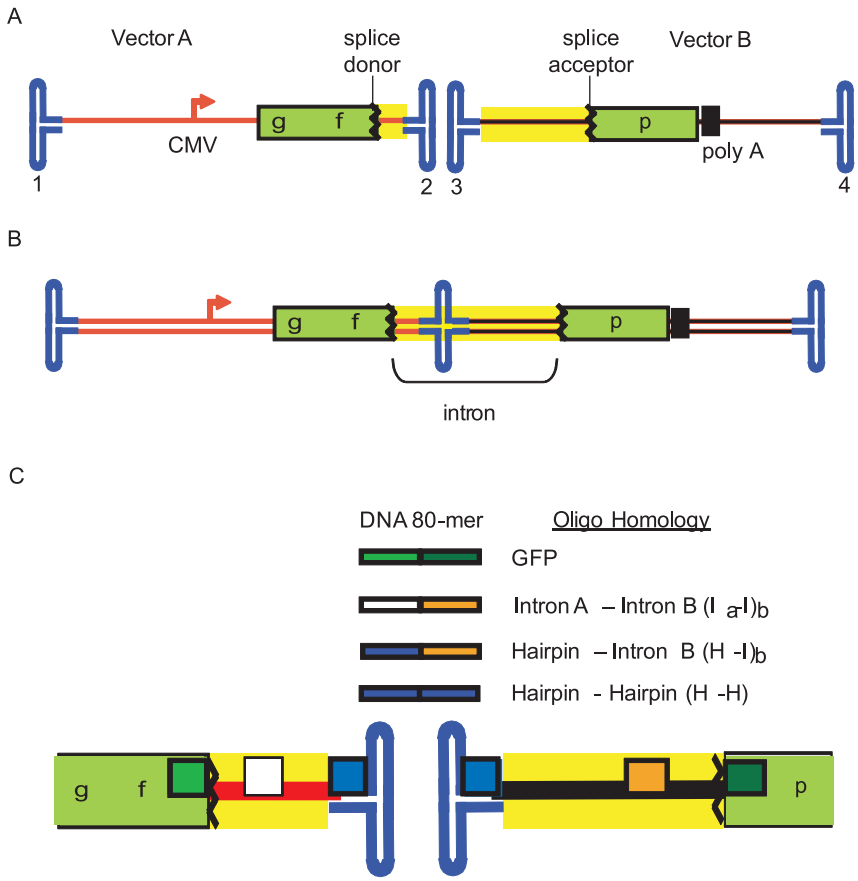


Figure 1. AAV Single-Strand Split *gfp* Vector System. A) Two AAV genomes are depicted, vectors A and B, that have the following genetic arrangement: i) vector A contains the CMV promoter upstream of a partial *gfp* coding sequence (cds) followed by a splicedonor site, and ii) vector B contains intron elements, the remainder of the *gfp* cds followed by a poly-adenylation signal. The inverted terminal repeats (ITRs) of the distinct genomes are labeled 1–4 and yellow highlights the intron region. B) The functional concatemer: a single DNA molecule for the ITR 2–3 linkage which results in functional GFP after splice removal of the intron. C) DNA 80-mers are depicted individually having 40 nt of homology to the color matched positions on vectors A and B. The crooked lines represent the splice donor and acceptor sequences and the intron is highlighted in yellow.
doi:10.1371/journal.pone.0007705.g001

co-transduction/transfection, GFP+ cells were quantitated. At an MOI of 20, the H-I_b oligo increased GFP+ cells over 20-fold (Fig. 3*B*). This value was not statistically different from the use of 10-fold more particles and a non-homologous oligo (Fig. 3*B*). When a capsid less efficient for 293 cell transduction (AAV8) was used at an MOI of 20, functional reconstruction was only observed in the presence of the H-I_b oligo, while no GFP+ cells were noted for natural concatemerization (Fig. S3). These results highlight the inefficiency of AAV split vectors as well as emphasize the ability of OAGR to allow a 10-fold decrease in particle administration while maintaining the same level of gene reconstruction.

Oligo-Assisted Intermolecular AAV Genome Recombination is Dependent on Oligo Abundance and Size

Initially, a high oligo concentration of 80 nM was chosen to emphasize any stimulatory effect (Fig. 2, Fig. 3*A*, *B*). We also transfected decreasing concentrations of the H-I_b Fwd oligo (or a non-homologous control) and measured GFP+ cells 72 h post-co-transduction (MOI 200). In this context, the number of GFP+ cells at increasing concentrations of the non-homologous oligo did not change and is statistically equivalent to the 0 nM concentration (Fig. 3*C*, DNS). A dose-dependent response to the H-I_b Fwd oligo was apparent with 26 nM being the first tested concentration at which an effect was significant (Fig. 3*C*). No significant difference in GFP+ cells was noted for concentrations 53 nM and 80 nM (Fig. 3*C*).

To determine whether smaller oligos could direct recombination of split vector genomes, DNA 60-mers and 40-mers based on the H-I_b Fwd oligo sequence, were also investigated. Decreasing the size of the oligo by 20 nt maintained a 6-fold OAGR effect while the 40-mer (20 nt arms of homology) was not significantly different from natural concatemerization (Fig. 3*D*). This observation is consistent with results from oligo-mediated DSB repair in yeast cells [19].

Sequence Confirmation of Oligo-Directed Concatemerization

If oligo directed concatemerization resulted from homologous recombination, then repair of distinct AAV genomes would create a junction corresponding to the sequence of the oligo. To investigate this, we performed PCR analysis of the functional concatemer junctions using DNA from co-transduced cells treated with the H-I_b Fwd oligo or the non-homologous control at different concentrations. PCR primers were designed and used to amplify recovered DNA external of any *gfp* functional junction (Fig. 4*A*; Materials and Methods). DNA harvested on day 3 from cells treated with the single-strand split *gfp* vectors and the H-I_b

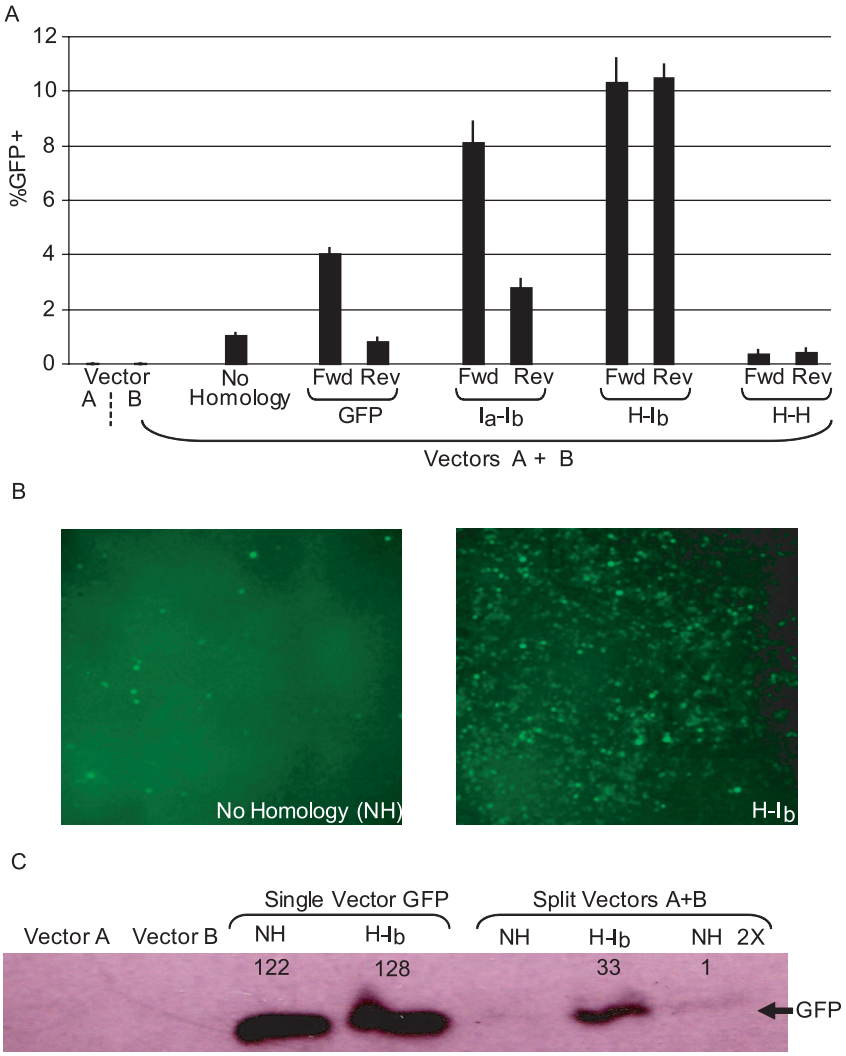


Figure 2. Oligonucleotides Direct AAV2 Concatemerization. A) Human embryonic kidney cells were co-infected with the split *gfp* vectors (A and B; Fig. 1) at a multiplicity of infection of 200 for each, vector A and B. Then, DNA oligonucleotides (80 nM) displaying 40 nucleotides of different sequence homology to vector A and B genomes were transfected in (see Fig. 1, non-homologous = NH). Three days post-treatment, GFP+ cells were quantitated by flow cytometry. An 80-mer without vector A and B homology was used as the non-homologous control (NH). The tested oligos are described (Fig. 1 and in the text) and opposite polarities were also investigated (Fwd = sense = non-template strand, Rev = anti-sense). Results are presented as %GFP + cells. B) Fluorescent microscopy of cells treated as in A) with a NH oligo or the H-I_b Fwd oligo used for transfection. C) Cells were treated as described in A). Additionally a single vector with the identical split gfp reporter as the split vectors (a CMV promoter upstream of the *gfp* coding sequence interrupted by the *hcg* intron) was investigated (single vector Gfp). Cells treated as in A) with the split vectors, or the single vector, followed by oligo transfection were harvested 3 days post-treatment. Western blotting was then performed to quantitate Gfp protein (NH 2X = twice the total protein loaded compared to all other lanes). Using densitometry, the relative levels of Gfp abundance was determined and the acquired values are also depicted.
doi:10.1371/journal.pone.0007705.g002

Fwd oligo (80 nM) resulted in a unique PCR product corresponding to the predicted size of the concatemer junction if recombination were to occur at H-I_b homology, (an approximate 400 bp deletion compared to the natural concatemer; Fig. 4). This product was not observed in cells receiving the non-homologous oligo and was subsequently cloned and sequenced. Sequencing of 20 clones demonstrated that the OAGR concatemer junction corresponded to the sequence of the oligo, effectively making a deletion that includes the ITR junction.

Oligo-Assisted Self-Complementary AAV Recombination

To determine if OAGR prefers a single-strand or a duplexed template, a self-complementary (sc) AAV2 split *gfp* system [5] was employed in our oligo assay (Fig. 5). In the sc context, a small deletion on only one of the ITRs results in transduced genomes that likely have a hinge (closed end = C) at one end of the complementarity and free, or open (O), ends on the other side (Fig. 5*A*). Thus, AAV genome dimerization resulting in functional GFP was investigated in all contexts (vector A-vector B; C-C, C-O, O-C, O-O; Fig. 4*A*). It is important to note that oligos tested on the single-strand system are also homologous to the scAAV2 genomes and that individual vector transduction does not result in GFP synthesis [5]. As performed above, 293 cells were co-transduced by these scAAV split vector combinations at an MOI of 200 for each vector, followed by transfection of an oligo and determination of GFP+ cells on day 3. Consistent with a previous report, concatemerization appeared most efficient with the C-C scAAV2 split *gfp* vector orientation and least efficient arranged in the O-O configuration (Fig. S5, 5). However, in contrast to the H-I_b Fwd OAGR effect on the single-strand vectors (Fig. 2*A*) only a

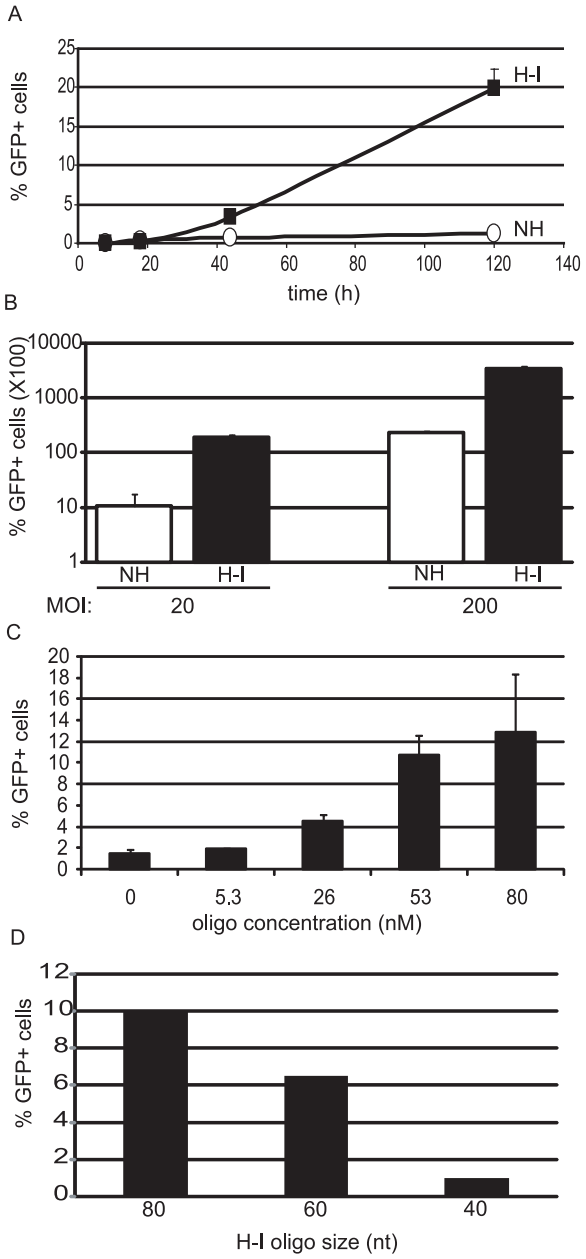


Figure 3. Characterization of Oligo-Assisted AAV Genome Concatemerization. A) Human embryonic kidney (HEK) cells were co-infected by AAV2 single strand split *gfp* vectors at a multiplicity of infection (MOI) of 200 for each vector. Then, either H-I_b Fwd or the non-homologous (NH) oligos were used for transfection (80 nM). GFP positive cells were quantitated by flow cytometry at the indicated times. B) HEK cells were co-transduced with the indicated MOI of each split *gfp* vector and subsequently transfected with the H-I_b Fwd oligo or a NH control (80 nM). Three days later the number of GFP+ cells was determined by flow cytometry. C) HEK cells were co-transduced with the AAV2 split *gfp* vectors at a MOI of 200 and the indicated concentrations of H-I_b Fwd was used for transfection. GFP+ cells were determined on day 3. The presented results were not normalized to the efficiencies of co-infection and oligo transfection. D) Oligo size analysis for stimulation of OAGR. Cells were co-infected with the split vector system as above and then either the H-I_b Fwd oligo or derivatives of that sequence were then used for transfection. %GFP positive cells were determined on day 3 by flow cytomtery. nt = nucleotide.
doi:10.1371/journal.pone.0007705.g003

minor to no effect was observed on the different self-complementary orientations (Fig. 5*B*). As the effect of the H-I_b Fwd oligo was more distinguishable at low MOIs, we also investigated the C-C vector combination under these conditions, however, no effect greater than 2-fold was observed (DNS). Additionally, no significant effect was observed with the GFP oligo (Fig. 1*C*) which targets a portion of the *gfp* coding sequence on the distinct genomes (DNS).

Oligo-Assisted Intramolecular Genome Recombination

GFP synthesis after ssAAV split *gfp* co-transduction results from a functional dimerization of vector A and B genomes. In comparison, single-strand and self-complementary vectors were generated, (using the same genetic elements as both the single-strand and self-complementary split *gfp* genomes described herein) that result in GFP synthesis after intramolecular linkage [circularization dependent; 5]. The genetic organization of these molecules follows: ITR-*hcg* intron-downstream *gfp* coding sequence-polyA-CMV promoter-upstream *gfp* coding sequence-intron-ITR (Fig. 6*A*). At low MOIs, these vectors depend on genome circularization (monomerization) for GFP synthesis, which has been reported by several groups to be preferred, in comparison to concatemer formation, at least initially [3,5,24]. Consistently, it was also reported that at low MOIs the self-complementary circularization dependent vectors were similar to genomes containing an intact transcriptional *gfp* cassette [5]. To investigate whether the H-I_b Fwd oligo promotes intramolecular circularization, 293 cells were transduced with either a single-strand or self-complementary version of the circularization dependent split *gfp* vector at increasing MOIs. At all vector to cell ratios tested, transfection of the H-I_b Fwd oligo slightly (less than 2-fold), yet significantly ($p<0.03$), increased the number of GFP+ cells compared to values determined with the control oligo for the single-strand vector (Fig. 6*B*). In contrast, no significant effect of the H-I_b oligo was noted for the self-complementary version (Fig. 6*B*).

Discussion

In most dividing cells, AAV DNA initiates a signaling cascade that culminates in a transient G2/M cell cycle arrest. Following infection, there is pan-nuclear activation of γ-H2AX and host proteins involved in DNA repair process the linear single-strand viral DNA to double-strand circular monomers and concatemers. The results described herein exploit this aspect of the AAV life cycle to demonstrate the first example of directed concatemerization, specifically and exclusively, of distinctly transduced AAV genomes. This intracellular genome engineering technique relies on the ability of an oligo to tether DNA molecules and induce recombination at the site of homology. A dependence on a single-strand DNA substrate is suggestive of replication-dependent recombination and consistent with the observed oligo polarity biases in bacteria [25] and yeast [22]. In addition to serving as a unique example of facilitated recombination, the successful translation of OAGR to *in vivo* settings may allow efficient large gene delivery by distinct rAAV split vectors.

Using ssAAV split *gfp* vector co-infection followed by the transfection of oligos having complementarity to both split genomes, dramatic effects on GFP abundance/function were observed. These effects exhibited a broad range of variation dependent upon the region of the AAV genome targeted by the oligo (Fig. 2). The choice of the most efficient oligo target homology is not completely understood. However, analysis of additional oligos were not consistent with the obvious theories that

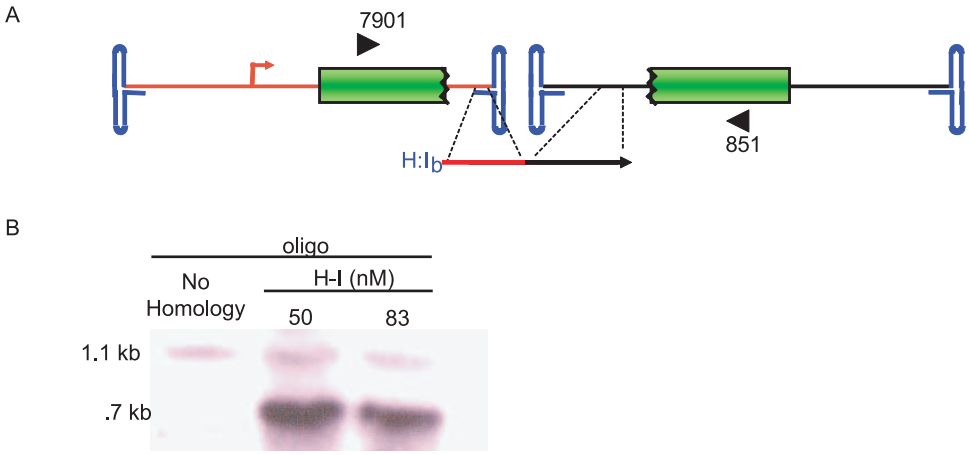


Figure 4. Junction Characterization of Functional Concatemers. A) This cartoon depicts our AAV split *gfp* vector system and the position of primers (block arrows) used for functional concatemer amplification. B. Evaluation of amplification products by southern blotting (*gfp* probe) showed a unique band formed only from DNA isolated for H-I treated cells. This band was directly cloned and sequenced.
doi:10.1371/journal.pone.0007705.g004

decreasing the distance from the ITR (Fig. S6) or increasing the oligo-target annealing temperature result in an increased OAGR effect (Fig. S7). Surprisingly, the most effective oligo, H-I_b, contains a 5′ 40 nt region capable of annealing to all ITRs and the 3′ 40 nt provide direction specificity by targeting an unique intron region found on vector B. If the 5′ region is shifted upstream of "H" homology to target a unique region of vector A, while maintaining the exact vector B homology ("I_b") there is slight reduction in the OAGR effect (Fig. 2*A*). Perhaps the observed decrease represents a separate additional effect specific to targeting the ITR (i.e. an effect on second-strand synthesis). Alternatively, the tetrad repeat of the ITRs (GACT) may represent a recombination hot-spot sequence as it is the origin of replication and involved in host chromosome integration as well [6].

A polarity bias for OAGR was noted for the GFP and I_a-I_b oligo sequences but not for the H- I_b oligo that proved most effective (Fig. 2*A*). Since the vector B homology was constant between I_a-I_b and H-I_b, this difference is attributed to the 5′ portion of the oligo. At this position, the I_a homology targets a single-strand region unique to the vector A intron while the "H" sequence targets a region common to all ITRs. Since the H target is paired, it is capable of binding both genome polarities while the vector A "I_a" homology would be restricted to its complement. Common to all oligos that demonstrated a polarity bias is the fact they target a region that undergoes second-strand synthesis, whereas the oligo that functions equally in both polarities (H-I_b) targets the duplexed region which is not synthesized in the generation of no end AAV genomes. This brings to mind an explanation for the polarity bias initially derived from single-strand oligo chromosome repair in *Escherichia coli* [20]. In that case, it was observed that the oligo strand bias correlated with direction of replication through the recombination site [20]. This result suggested that lagging strand synthesis would allow greater access to its template strand and consistently, oligos complementary to that strand out performed the opposite orientations. In the case of the AAV bi-polarity single-strand genomes, both oligo polarities have equal opportunity for genome binding and continuous replication is initiated from the free 3′ end and therefore, imposes no obvious bias. As the OAGR event is postulated similar to replication-dependent recombination, an alternative model in which the ITRs are biased for their ability to initiate second-strand synthesis is proposed. In fact, polarity biases for AAV replication and genome packaging have been suggested previously [26,27]. If replication-dependent recombination is responsible for the bias observed in this work, it argues that the oligo acts before second-strand synthesis which is supported by no significant OAGR effect on scAAV, or for an ITR containing split *gfp* plasmid system (Fig. 5, Fig. S5 and DNS).

In contrast to the results using the single-strand split *gfp* vectors, the scAAV genomes, and double-strand plasmids (DNS), were not responsive to homologous oligos, nor was there an effect for the intramolecular recombination of the scAAV circularization dependent vector (Fig. 6). This OAGR requirement for ssAAV genomes likely reflects the ability of the oligo to access its target. For example, the self-complementary molecule is essentially a 2 kb hairpin precluding oligo annealing. Another theory for the lack of the OAGR effect on the scAAV genomes is that OAGR is coupled to second-strand synthesis. As replication forks are "inherently recombinogenic" one can envision the annealed oligo blocking second-strand synthesis in a manner that would stimulate recombination of tethered, yet distinct, genomes [25]. Interestingly, only a minor OAGR effect of the H- I_b oligo was observed for the single-strand circularization dependent vectors which, at higher MOIs, could also undergo dimerization to generate GFP. This result supports that intramolecular genome linkage is very efficient, as has been reported [5,24]. In this scenario, monomerization readily occurs and the modest increase observed with the homologous oligo is perhaps a result of intermolecular dimerization by OAGR, which is evident when monomer formation does not result in reporter activity (split vectors; Fig. 2*A*). This result supports a role for oligo tethering of distinct DNA molecules as opposed to a stimulation of intramolecular ITR processing or a simple model in which the oligo opens the ITRs. The absence of any oligo effect with the scAAV circularization dependent vector is consistent to the results obtained with the scAAV split *gfp* co-infection, and could be explained in the same manner (see above). Given that the activity of the single-strand and self-complementary circularization dependent vectors are somewhat similar without a homologous oligo (Fig. 6), the explanation for genome tethering by the oligo is further supported, especially when the 145 nt ITRs are likely the major local attraction. This also suggests that processing of single-strand and self-complementary genomes for concatemerization are similar. Additionally, we speculate that natural ITR-mediated circularization/concatemerization is primarily reliant on NHEJ which is more efficient than

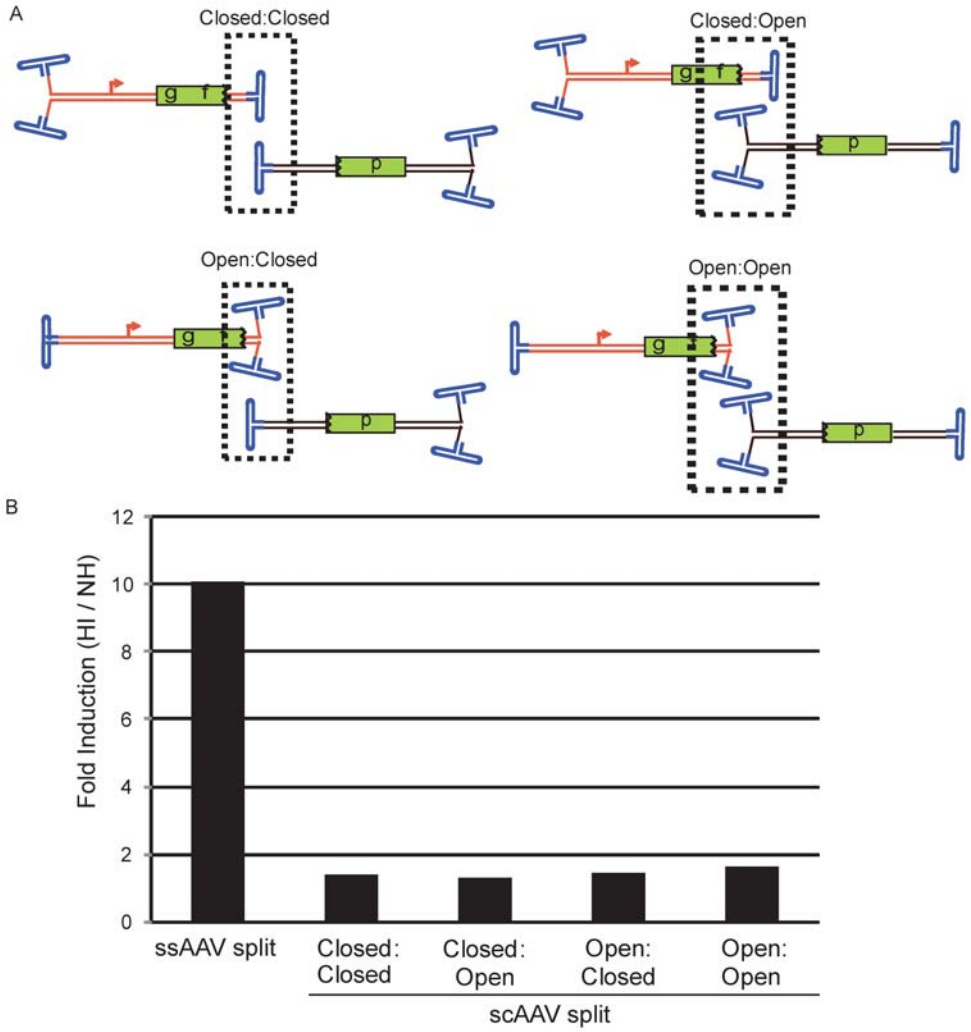


Figure 5. Oligo-Assisted AAV Genome Concatemerization of Self-Complementary Split *gfp* Vectors. A) Four split *gfp* systems are depicted for self-complimentary (sc) AAV2 genomes. The genomes differ in the orientation of their respective cassettes (see text for details) in relation to the open or closed ends of the molecule. A broken box encases the type of junction necessary for functional GFP. B) The pairs of split *gfp* vectors depicted in A were used for co-infection of human embryonic kidney cells at a multiplicity of infection (MOI) of 200 for each particle type. Then, either H-I_b Fwd or a non-homologous (NH) oligo was used for transfection. Three days later, GFP+ cells were determined by flow cytometry. The results are presented as fold-induction which is the number of GFP+ cells given the H-I_b oligo divided by the number of GFP+ cells given the NH oligo.
doi:10.1371/journal.pone.0007705.g005

OAGR's reliance on homologous recombination, especially for an intramolecular linkage event.

The collective results support the "annealing–integration" model as the mechanism of OAGR (Fig. 7), similar to the "rounds of strand annealing" mechanism of DSB repair driven by single-strand DNA oligos in yeast cells [18] and in line with the model proposed for single-strand oligo recombineering in *E. coli* [20]. Following the capsid release of AAV genomes, oligos with homology to each single-strand vector genome anneal to their target, displacing the remaining terminal ends of the genome. This annealing event promotes the desired genome dimer orientation and exploits host replication machinery and DNA repair enzymes for intermolecular recombination with the resultant junction sequence corresponding to the sequence of the oligo. Evidence from additional modified oligos suggests that the annealed oligo physically incorporates into the newly formed dimer molecule, as demonstrated in both prokaryotes and human cell lines, acting as a split with or without oligo-primed synthesis [28,29]. However, the exact mechanism is still under investigation.

Materials and Methods

Maintenance of cells

Human embryonic kidney (HEK) 293 cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% heat-inactivated fetal bovine serum and 1X Penn/Strep (DMEM+, Sigma). Cells were grown at 37°C in a 5% CO_2 humidified incubator.

rAAV vector construction, production and purification

The self-complementary (sc) split *gfp* vectors and the sc *gfp* circularization dependent (cd) vectors have been previously reported [5]. The construction of the single-strand (ss) AAV-cd and ss split gene vectors was the same as the scAAV-cd and sc split-gene vectors, with the exceptions that different pieces of stuffer fragments were used, and that the expression cassettes were cloned into the pSSV9 [6] vector to generate conventional ssAAV vectors. Lambda Phage *Dra* I 3332 bp fragment was used as a stuffer for the upstream (vector A; Fig. 1*A*) ss split vector ssAAV-Gf using *Dra* I-5′-

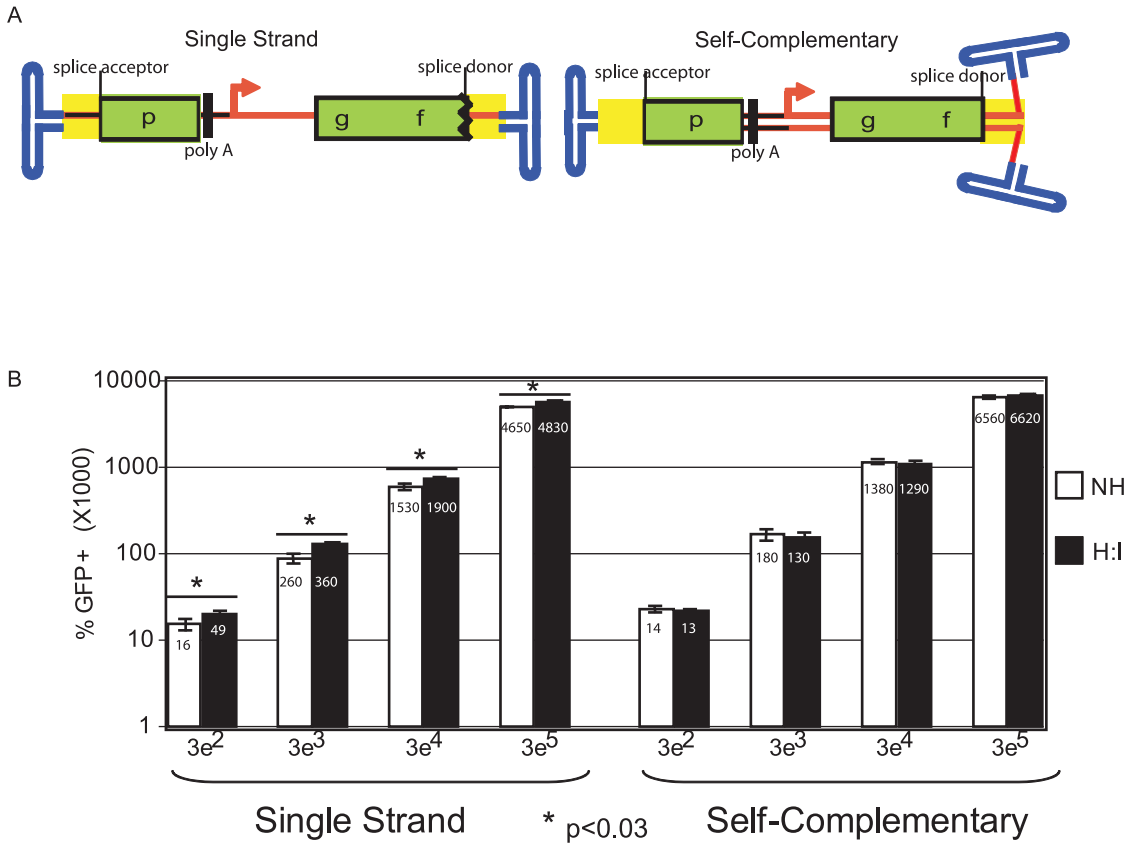


Figure 6. Oligo-Assisted AAV Genome Recombination of Circularization Dependent AAV. A) The cartoon depicts single vector split *gfp* systems that are capable of GFP production after either intra- or inter-molecular dimerization. The relevant system features are depicted; the arrow represents the CMV promoter followed by the 5′ portion of the *gfp* coding sequence (cds) and an splice donor sequence. On the far left side is the *hcg* intron and the remaining *gfp* cds followed by a poly-adenylation sequence. The entire intron is highlighted in yellow. Both the single-strand (ss) and self-complementary (sc) versions are shown. B) AAV2 capsids packaged with either genome depicted in A were used to separately infect human embryonic kidney cells at the indicated viral genomes per cell. Then, either the H-I_b Fwd or a non-homologous (NH) oligo was used for transfection. Three days later, GFP+ cells were determined by flow cytometry. In all cases the single-strand construct demonstrated a significant (unpaired student t test) increase, yet less than 2-fold, with the H-I_b Fwd oligo compared to the NH control. No significant changes were note for the sc construct.
doi:10.1371/journal.pone.0007705.g006

2 5′CTA CGC GTC ATC GCC AAT AAA AGT GGC G as a forward primer and *Dra* I-3′ 5′ GCA CGC GTA AAA GGC AGG TGG GC as a reverse primer. Lambda *Msc* I 3605 bp fragment was the stuffer for the downstream (vector B; Fig. 1*A*) ss split vector ssAAV-Gfp using *Msc* I-5′ 5′CTA CGC GTC CAV GCA GCT TGC AG as a forward primer and *Msc* I-3′ 5′GTA CGC GTC CAG CAT GAT ACG TCC CG as a reverse primer. The lambda *Mlu* I 956 bp fragment described above was used as a stuffer for ssAAV-*gfp* for an intact *gfp* cassette clone [5].

rAAV was generated using the triple transfection method with AAV serotype 2 capsid and purified by discontinuous iodixanol gradient separation and heparin chromatography [5,30,31]. Additional viral preparations, including rAAV8, were purified on a CsCl gradient, followed by peak fraction collection and dialysis against phosphate-buffered saline (PBS)[31]. The number of genome containing particles was determined by Southern dot blot hybridization while rAAV2 infectious units (MOI) were determined by infectious center assays in C12 cells [32].

rAAV transduction

50,000 293 cells were seeded in a 24 well plate 24 h prior to transduction. The next day, the total medium volume was adjusted to 700 ul and virus particles were diluted in 50 ul of DMEM+ prior to well addition. A MOI of 200 I.U. of each split vector was used for transduction, unless otherwise noted. At the indicated time point (usually 72 h), cells were harvested by gentle aspiration and re-suspended in PBS for flow cytometric analysis. In the experiments described herein, the co-transduction efficiency was not explicitly determined, thus the experimental results were not corrected for co-transduction and likely under represent the actual values. All experiments were performed in triplicate on at least 3 different days and the results are presented as averages.

Oligo transfection

The sequence of the oligos used in these studies listed in Table S1. The indicated concentration of oligo (80 nM if not indicated) was transfected 10 min post co-transduction using linear polyethylenimine (PEI) transfection reagent prepared in house. In these experiments, the oligo was diluted in DMEM without supplements, vortexed in the presence of PEI, and the mixture was added to the well 10 min later where it remained unless otherwise noted. Under these conditions, transfection of an equivalent amount of GFP expressing plasmid results in 70% GFP+ cells. However, the % GFP+ values presented in this work are not corrected for the efficiency of oligo transfection and thus, under represent the actual efficiency of the approach.

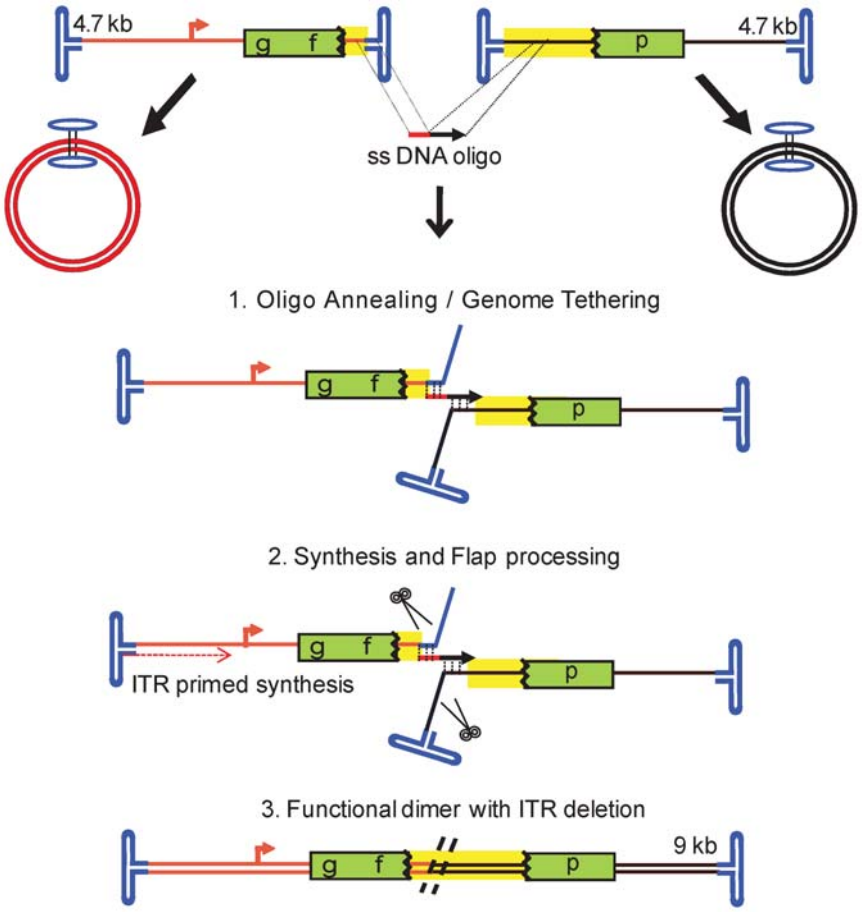


Figure 7. A Model for Oligo-Assisted AAV Genome Recombination. Following the capsid release of the 4.7 kb AAV genomes, presumably in the nucleus, oligos with homology to each vector anneal to their single-strand (ss) target, displacing the remaining terminal ends of the genome. This annealing event promotes the desired genome dimer orientation and viral genome replication using the inverted terminal repeats for initiation. Extension of the 3′ oligo end is also a possibility. Exploitation of host repair enzymes for intermolecular recombination follows resulting in a double strand product with the vector junction sequence corresponding to that of the oligo. In the case of the H-I_b oligo this recombination event results in an approximate 400 bp deletion of the ITR junction. The recombined dimer is near double the length of the independent genomes. Although it was not conclusively demonstrated, this model favors physical incorporation of the transfected oligo into the oligo-directed concatemer (described in discussion). Also depicted are perhaps the naturally favored products initially following transduction, circular monomers, which do not produce GFP.
doi:10.1371/journal.pone.0007705.g007

Analysis of GFP+ cells

GFP+ cells were visualized by fluorescent microscopy and quantitated by flow cytometry. A minimum of 10,000 similar sized/shaped single cells were counted, for each replicate, at the indicated times at an event per second rate of 1500–2000. The GFP+ gate was designed such that untreated cells did not give a single GFP+ event beyond a million counted cells. It was also positioned away from false, or transitioning, GFP+ cells to obtain no false positives. Visualization of pooled GFP+ cells after FACs sorting confirmed the GFP+ phenotype (DNS).

Concatemer junction sequence verification

GFP cells were pooled by FACs analysis and total DNA was recovered using a DNA-EZ kit (Qiagen). PCR reaction conditions were performed by standard procedures using pfu turbo polymerase (stratagene). The sequence of primers used to initially amplify a functional *gfp* concatemer follow (5′ to 3′ orientation): i) early *gfp* sequence on the upstreamsplit genome, CGGCGAGGGC GAGGGCGATG CCACC and ii) late *gfp* coding sequence (cds) on the downstream split genome, GACCGCCGCC GGGATCACTC TCGGC. Nested primers amplified the gel isolated product for cloning and subsequent sequencing: EcoRI-fwd CGCGAATTCGCTTCAGCCGC TACCCCGACC ACATG, Xba-rev gcga tctaga AGGTAGTGGTT GTCGGGCAGC AGC. In validation of this primer set, capsid purified A and B vector genomes (at different concentrations) served as templates in a PCR reaction where different amounts of the H-I_b oligo were spiked in the reaction (DNS). In these instances, a product corresponding to the oligo-directed concatemer junction could not be generated (DNS).

Supporting Information

Table S1 Table of oligonucleotide sequences used in this study Found at: doi:10.1371/journal.pone.0007705.s001 (0.03 MB DOC)

Figure S1 Repair Oligonucleotide Size Influences Gene Correction. The chromosomal homologous recombination GFP reporter system of M. Porteus (Porteus et al. Mol Cell Biol. 2003 May; 23 (10):3558-65) was used to evaluate gene correction by different sized homologous oligonucleotides in the presence and absence of a Sce-induced double-strand break (DSB). In this system the stop codons in all reading frames and the Sce binding

site is inserted into the gfp coding sequence. This defective reporter does not produce GFP until the insertion is releived. The numbers refer to the size of the oligonucleotides used for repair having equal sized gfp homologous arms to each side of the disrupting insertion. GFP+ cells indicates chromosomal reporter repair and fluorescence was determined 48 h post-transfection (oligonucleotide with or without a Sce expression plasmid) by flow cytometry. NH = non-homologous control oligonucleotide.
Found at: doi:10.1371/journal.pone.0007705.s002 (0.97 MB EPS)

Figure S2 Oligo-Assisted Genome Recombination using AAV serotype 8 capsids. Human embryonic kidney cells were co-infected with the split gfp vectors (A and B; Fig. 1) packaged in serotype capsid 8 at a multiplicity of infection of 50,000 for each, vectors A and B. Then, the H-I DNA oligonucleotide (80 nM) displaying 40 nucleotides of sequence homology to vector A and B genomes was used for transfection (see Fig. 1, non-homologous = NH). Three days post-treatment, GFP+ cells were quantitated by flow cytometry. An 80-mer without vector A and B homology was used as the negative control (NH).
Found at: doi:10.1371/journal.pone.0007705.s003 (1.45 MB EPS)

Figure S3 Comparison of Oligo-Assisted Genome Recombination on Intact and Split AAV vectors. Human embryonic kidney cells were infected with either a single vector having a split gfp gene system (see text) or the identical split gfp gene system encoded on separate vectors (A and B; Fig. 1) at a multiplicity of infection of 200. Then, the H-I DNA oligonucleotide (80 nM) displaying 40 nucleotides of different sequence homology to vector A and B genomes, or non-contiguous homology to the single intact gfp vector (see text), was used for transfection (see Fig. 1, non-homologous = NH). Three days post-treatment, GFP+ cells were quantitated by flow cytometry. An 80-mer without vector A and B homology was used as the negative control (NH).
Found at: doi:10.1371/journal.pone.0007705.s004 (0.95 MB EPS)

Figure S4 Oligonucleotide persistence in 293 cells. The NH or H-I oligonucleotides (described in text) were transfected to 293 cells using PEI. At the indicated timepoints total DNA was recovered and separated on an alkaline gel. Then, the DNA was transferred to a membrane and probed with the H-I oligonucleotide labeled with P-32. The positive control is 10 ng of the H-I oligonucleotide loaded on the gel directly.
Found at: doi:10.1371/journal.pone.0007705.s005 (2.34 MB EPS)

Figure S5 Oligo-Assisted AAV Recombination of Self-Complementary AAV Split Vectors. Gene reconstruction from distinct transduced self-complementary AAV genomes was monitored in the presence of an oligonucleotide (H:Ib) having homology to both split vectors (see text for split vector description and Figure 5 for a cartoon depicting the vectors). Given the nature of self-complementary genomes 4 orientations exist in which gene reconstruction was evaluated (x-axis; see text and Figure 5). Gene reconstruction results in GFP+ cells which were evaluated on day 3 post-transduction and transfection by flow cytometry.
Found at: doi:10.1371/journal.pone.0007705.s006 (1.10 MB EPS)

Figure S6 Evaluation of Oligonucleotide Target Homology for Oligo-Assisted AAV Genome Recombination. The sequence evaluated for oligo-assisted AAV genome recombination of the single-strand AAV split vectors (see text) is shown (the dotted lines project the evaluated region). A single oligonucleotide consists of 40 nucleotides (nt) of homology to Vector A (numbers) and 40 nt of homology to Vector B (letters). The single-strand gfp split vector system (see text and Fig. 1) was used for 293 cell transduction followed by transfection of the depicted oligos (the gfp oligo is described in the text). Three days later gene reconstruction was evaluated by flow cytometry. RBE = the AAV Rep binding element, TR = terminal repeat, cds = coding sequence.
Found at: doi:10.1371/journal.pone.0007705.s007 (1.25 MB EPS)

Figure S7 Analysis of melting temperature in relation to Oligo-Assisted AAV Genome Recombination. Human embryonic kidney cells were co-infected with the single strand split gfp vectors packaged in serotype capsid 2 at a multiplicity of infection of 200 for vector A and B. Then, the oligonucleotides depicted displaying 40 nucleotides of sequence homology to vector A (H, H2) and B (I, I2) genomes were used for transfection (see Fig. 1). In these cases, the homology of the oligo to the AAV vector system was shifted slightly to significantly increase oligonucleotide/vectors annealing. Three days post-treatment, GFP+ cells were quantitated by flow cytometry. An 80-mer without vector A and B homology was used as the negative control (NH). Tm = melting temperature.
Found at: doi:10.1371/journal.pone.0007705.s008 (1.28 MB EPS)

Acknowledgments

We thank Dr. Dale Ramsden for critiquing this manuscript and for speculation as to the mechanism of recombination.

Author Contributions

Conceived and designed the experiments: MLH FS RJS. Performed the experiments: MLH FS CL. Analyzed the data: MLH FS CL RJS. Contributed reagents/materials/analysis tools: MLH CL VWC. Wrote the paper: MLH FS. Final approval: RJS.

References

1. Srivastava A, Lusby EW, Berns KI (1983) Nucleotide sequence and organization of the adeno-associated virus 2 genome. J Virol 45: 555–64.
2. Ferrari FK, Samulski T, Shenk T, Samulski RJ (1996) Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. J Virol 70: 3227–34.
3. Nakai H, Storm TA, Kay MA (2000) Recruitment of single-stranded recombinant adeno-associated virus vector genomes and intermolecular recombination are responsible for stable transduction of liver in vivo. J Virol 74: 9451–63.
4. Duan D, Yue Y, Engelhardt JF (2001) Expanding AAV packaging capacity with trans-splicing or overlapping vectors: a quantitative comparison. Mol Ther 4: 383–91.
5. Choi VW, Samulski RJ, McCarty DM (2005) Effects of adeno-associated virus DNA hairpin structure on recombination. J Virol 79: 6801–7.
6. Samulski RJ, Chang LS, Shenk T (1989) Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression. J Virol 63: 3822–8.
7. Wu Z, Asokan A, Samulski RJ (2006) Adeno-associated virus serotypes: vector toolkit for human gene therapy. Mol Ther 14: 316–27.
8. Halbert CL, Allen JM, Miller AD (2002) Efficient mouse airway transduction following recombination between AAV vectors carrying parts of a larger gene. Nat Biotechnol 20: 697–701.
9. Ghosh A, Yue Y, Lai Y, Duan D (2008) A hybrid vector system expands adeno-associated viral packaging capacity in a transgene-independent manner. Mol Ther 16: 124–30.
10. Xu Z, Yue Y, Lai Y, Ye C, Qiu J, et al. (2004) Trans-splicing adeno-associated viral vector-mediated gene therapy is limited by the accumulation of of spliced mRNA but not dual vector co-infection efficiency. Hum Gene Ther 15: 896–905.
11. Hermanns J, Schulze A, Jansen-Db1urr P, Kleinschmidt JA, Schmidt R, et al. (1997) Infection of primary cells by adeno-associated virus type 2 results in a modulation of cell cycle-regulating proteins. J Virol 71: 6020–7.
12. Schwartz RA, Palacios JA, Cassell GD, Adam S, Giacca M, et al. (2007) The Mre11/Rad50/Nbs1 complex limits adeno-associated virus transduction and replication. J Virol 81: 12936–45.
13. Raj K, Ogston P, Beard P (2001) Virus-mediated killing of cells that lack p53 activity. Nature 412: 914–7.

14. Jurvansuu J, Fragkos M, Ingemarsdotter C, Beard P (2007) Chk1 instability is coupled to mitotic cell death of p53-deficient cells in response to virus-induced DNA damage signaling. J Mol Biol 372: 397–406.
15. Jurvansuu J, Raj K, Stasiak A, Beard P (2005) Viral transport of DNA damage that mimics a stalled replication fork. J Virol 79: 569–80.
16. Inagaki K, Ma C, Storm TA, Kay MA, Nakai H (2007) The role of DNA-PKcs and artemis in opening viral DNA hairpin termini in various tissues in mice. J Virol 81: 11304–21.
17. Jasin M (1996) Trends Genet. 12: 224–8.
18. Storici F, Snipe JR, Chan GK, Gordenin DA, Resnick MA (2006) Conservative repair of a chromosomal double-strand break by single-strand DNA through two steps of annealing. Mol Cell Biol 26: 7645–57.
19. Storici F, Durham CL, Gordenin DA, Resnick MA (2003) Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. Proc Natl Acad Sci U S A 100: 14994–9.
20. Ellis HM, Yu D, DiTizio T, Court DL (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proc Natl Acad Sci U S A 98: 6742–6.
21. Radecke F, Peter I, Radecke S, Gellhaus K, Schwarz K, et al. (2006) Targeted chromosomal gene modification in human cells by single-stranded oligodeoxynucleotides in the presence of a DNA double-strand break. Mol Ther 14: 798–808.
22. Yamamoto T, Moerschell RP, Wakem LP, Komar-Panicucci S, Sherman F (1992) Strand-specificity in the transformation of yeast with synthetic oligonucleotides. Genetics 131: 811–9.
23. Liu L, Rice MC, Drury M, Cheng S, Gamper H, et al. (2002) Strand bias in targeted gene repair is influenced by transcriptional activity. Mol Cell Biol 22: 3852–63.
24. Duan D, Sharma P, Yang J, Yue Y, Dudus L, Zhang Y, et al. (1998) Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term episomal persistence in muscle tissue. J Virol 72: 8568–77.
25. Kreuzer KN (2005) Interplay between DNA replication and recombination in prokaryotes. Annu Rev Microbiol 59: 43–67.
26. Samulski RJ, Srivastava A, Berns KI, Muzyczka N (1983) Rescue of adeno-associated virus from recombinant plasmids: gene correction within the terminal repeats of AAV. Cell 33: 135–43.
27. King JA, Dubielzig R, Grimm D, Kleinschmidt JA (2001) DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. Embo J 20: 3282–91.
28. Huen MS, Li XT, Lu LY, Watt RM, Liu DP, et al. (2006) The involvement of replication in single stranded oligonucleotide-mediated gene repair. Nucleic Acids Res 34: 6183–94.
29. Radecke S, Radecke F, Peter I, Schwarz K (2006) Physical incorporation of a single-stranded oligodeoxynucleotide during targeted repair of a human chromosomal locus. J Gene Med 8: 217–28.
30. Choi VW, McCarty DM, Samulski RJ (2006) Host cell DNA repair pathways in adeno-associated viral genome processing. J Virol 80: 10346–56.
31. Grieger JC, Choi VW, Samulski RJ (2006) Production and characterization of adeno-associated viral vectors. Nat Protoc 1: 1412–28.
32. Zolotukhin S, Byrne BJ, Mason E, Zolotukhin I, Potter M, et al. (1999) Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. Gene Ther 6: 973–85.