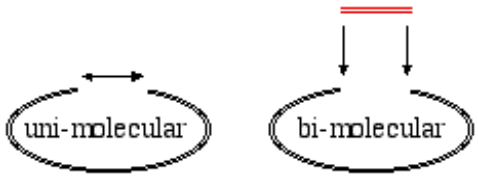
Lecture 7

Additional cloning methods

Today we're going to continue our discussion about some of the "fine finishing work" that goes into making DNA ends fit together.

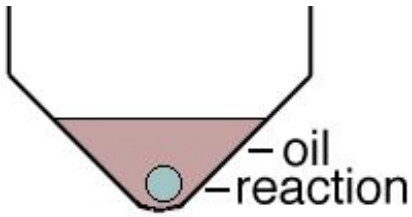
A fine distinction

As you recall from the last lecture, much was made of the difference between "uni-molecular" and "bi-molecular" ligation reactions:



Why is this so important? Because as genetic engineers you will often want to conduct a bi-molecular ligation, inserting a piece of DNA into a vector for example, and you will want to prevent the uni-molecular side reaction from taking place or interfering with our work.

Uni-molecular ligation reactions (reclosures) are easy to perform because they do not depend on the DNA concentration in the ligation. Two ends of a single DNA molecule may take a random walk away from each other in solution, but they can only go so far since they are tethered to each other. On the other hand, there is no limit to the separation between ends of different DNA molecules, and so bimolecular reactions are problematic in dilute solutions. We sometimes add substances such as polyethylene glycol (PEG) to try to absorb some of the solution volume in a ligation, increasing the activity of the ends. It is also helpful in this difficult case, to conduct the reaction in a minimal volume. You may, for example, run a ligation reaction in 2 ul, in a little droplet that is trapped under mineral oil. That keeps the reaction from adhering to the sides of the tube, and makes it easier to inject reaction components during assembly of the reaction. You do need a good pipet that can measure small volumes accurately, however.



Why make the reaction such a small volume? To increase the chances that the two molecules will find each other. It is a challenge of kinetics. The molecules with two compatible ends will have less trouble finding each other (they have better kinetics) because they are tethered by the length of the DNA.

Of course, besides the issue of the overall reaction volume, there is the issue of what goes into the reaction. For example, suppose we have 5 ng of a digested plasmid that is 5 kbp in size. The small piece that we want to insert is 250 bp in size, and we want to employ a 2:1 molar ratio of the insert to the vector during ligation. How much insert should we add to the 5 ng of plasmid to give us this ratio?

	How many nanograms of this insert should we use to give a 2:1 molar ratio with the vector?
☐	**10 ng**
☐	**2 ng**
☐	**0.5 ng**
☐	**0.01 ng**

How did we figure that out? Note that the insert is 20 times smaller than the vector, so if we wanted a 1:1 molar ratio we would need to use one twentieth of 5 ng. In fact, we want twice that so that we will have a 2:1 molar ratio, so we want one tenth of 5 ng, or 0.5 ng that is. We might also call it 500 pg. The lesson is, when you are setting up reactions with fragments of vastly different sizes, you need to think about the reaction on a molar basis rather than on a mass basis.

Some methods that were discussed in the last lecture are designed to limit the ability of a vector to reclose upon itself without a DNA insert. For example, the alkaline phosphatase enzyme from calf intestines is often used to treat the vector DNA ends, removing the 5' phosphates and the potential of the vector to conduct a unimolecular reaction. The DNA insert is not dephosphorylated, and can bridge together the ends of the vector (leaving two widely-spaced nicks on different strands).

Digestion of DNA products after ligation, with an enzyme that linearizes only unwanted products (as described in the previous lecture for Bgl II fragments cloned into a BamHI site), is a method that can also be employed in some cases.

Base-pairing incompatibility

A vector may also be designed so that reclosures are prevented by incompatibility of hydrogen bonding. We discussed in the second lecture a plasmid called pCRII-topo that has a single 3' overhanging nucleotide (a thymidine)

```
...GAATTCGCCCTT          AGGGCGAATTC...
...CTTAAGCGGGA          TTCCCGGTTAAG...
```

This vector cannot easily reclose upon itself because the 3' overhanging ends are incompatible (both being T bases). On the other hand, it is perfect at accepting PCR products made with Taq polymerase, an enzyme that has a

limited "terminal transferase" activity and tends to add a single 3' overhanging adenosine to its products of synthesis:

```
     NNNNNN...NNNNNNA
ANNNNNN...NNNNNN
```

When we prepare PCR fragments for cloning in a bi-molecular reaction using Pfu polymerase (which does not leave a 3' untemplated A nucleotide), we cannot use the pCRII-topo plasmid. Why? Because the blunt end of the PCR fragment would be incompatible with the overhanging end of the vector:

```
...GAATTCGCCCTTNNNNNNNN AGGGCGAATTC...
...CTTAAGCGGGA NNNNNNNNTTCCCGGTTAAG...
```

T4 DNA ligase cannot connect the 3' T of the vector to the 5' N of the blunt fragment, because there is an underlying gap of one nucleotide. There are enzymes (one in particular, named T4 RNA ligase) that will ligate single-stranded ends, but T4 DNA ligase will not do this job.

An aside...on filling gaps with patches

Consider the case of BamHI (G^GATCC, leaving a 5' overhang) and SacI (GAGCT^C, leaving a 3' overhang).

```
AGTCAG              CTTATC
TCAGTCCTAG      TCGAGAATAG
BamHI       ----        SacI
```

These could not be ligated together with T4 DNA ligase (our usual "ligase of choice" for genetic engineering work), but if you ask a company to synthesize an oligonucleotide with the sequence GATCAGCT this synthetic sequence would fit perfectly into the gap left by BamHI and SacI ends. T4 DNA ligase would be able to seal the nicks (two on the top strand, one on the bottom), provided you had made sure to phosphorylate the 5' ends of the oligonucleotides with T4 polynucleotide kinase.

```
AGTCAGGATCAGCTCTTATC
TCAGTCCTAGTCGAGAATAG
BamHI    Patch    SacI
```

In fact, now that oligonucleotide synthesis is so inexpensive ($0.70/nt), it is common for labs to assemble clones entirely from synthetic pieces. The oligonucleotides overlap after annealing, and are held together by hydrogen bonding:

For example:

```
GATCCAGAGTACAGGACGATAGTACAGAGTACAGTAGCACGATAAAAGACTAGACGATACCGAGTAGAG
    GTCTCATGTCCTGCTATCATGTCTCATGTCATCGTGCTATTTTCTGATCTGCTATGGCTCATCACCTAG
```

The cloning systems in which the vector has a 3' overhanging T and the fragment has a 3' overhanging A are sometimes called "TA cloning" methods, and they are widely used with PCR products generated with Taq polymerase. In this example, the insert prepared with Taq polymerase is shown in red, and has 3' overhanging "A" nucleotides. Taq has a natural (but limited) terminal transferase reaction, so it makes this type of "single A overhang" end at some high frequency. Taq lacks a 3' to 5' exonuclease activity. The vector arms shown have a "single T overhang" and so are compatible with the PCR product.

```
GGATACTAAGACTGT     AACTGAGA....AAGATTACGA     GAGGTACG
CCTATGATTCTGAC     ATTGACTCT....TTCTAATGC     TCTCCATGC
```

For PCR products made with Pfu polymerase, which has a 3' to 5' exonuclease activity and no terminal transferase activity, some methods advocate adding Taq polymerase at the end of the reaction so that a 3' terminal A is applied to the product (making it eligible for cloning in a TA cloning kit). Others suggest adding a single A nucleotide by use of the enzyme terminal transferase and ddATP as a substrate. You could always just clone products made with Pfu polymerase using vector arms having a blunt end.

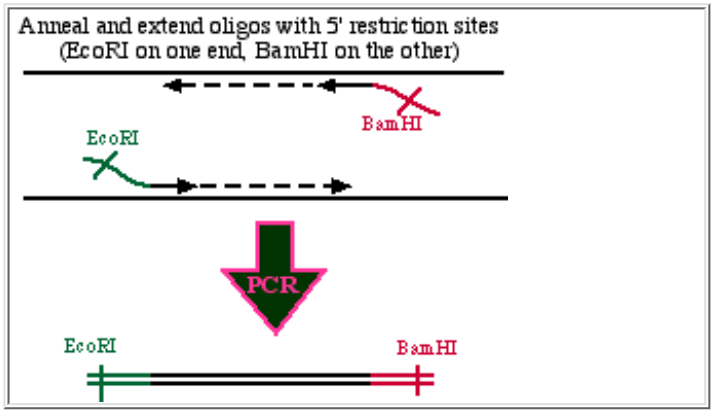
(A question or two for you: Why would that protocol lead to the addition of only one A nucleotide instead of many, and would such an added ddA participate in ligation?)

Prior to TA cloning methods, which is to say prior to about 1990, it was necessary to treat products of Taq polymerase with T4 DNA polymerase and the four dNTP substrates, so that the fragment could be cloned into a blunt-ended cloning site (T4 DNA polymerase has a powerful 3'-5' exonuclease that is several thousand times more active than the 3'-5' exonucleolytic activity of Klenow enzyme). It was a real joy to be able to clone PCR fragments directly into a TA cloning vector without this cleanup step.

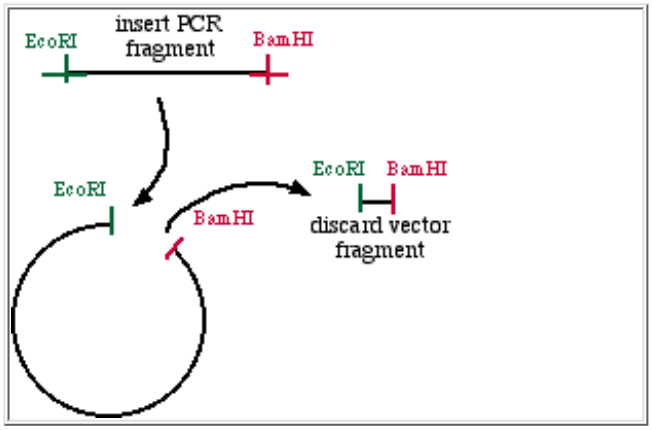
A question for you to ponder: How do you suppose the company supplying the TA vector makes the 3' overhanging dT ends? Do you think the ends were left by a restriction enzyme?

Using PCR to add a restriction site

TA cloning and blunt end cloning is not the only game in town, however. This is a point at which the polymerase chain reaction, and a type of site directed mutagenesis, can be put to good use. If the DNA fragment you wish to clone has not got the restriction enzymes you want, you may be able to simply add the sites to the 5' ends of specific oligonucleotides and incorporate them by PCR.



Your PCR fragment is now ready to digest with EcoRI and BamHI, and can be inserted into a similarly digested vector (usually exchanged for a small EcoRI to BamHI fragment in the vector, which is lost)



As we have discussed, this is called forced cloning because the fragment can only be inserted in one orientation, with the EcoRI sites matching on one end and the BamHI sites matching on the other end.
Forced cloning may also be used if you have one "blunt" site and one cohesive end site. For example, consider the following problem:

You have a DNA fragment that you have digested with EcoRV (GAT^ATC) and Sau3AI (^GATC). You could force it into a vector that had been digested with SmaI (CCC^GGG) and BamHI (G^GATCC). The EcoRV end would be compatible with the SmaI end (both are blunt) and the Sau3AI end would be compatible with the BamHI end (both have a 5' GATC overhanging end).

Digestion near the ends of DNA fragments

We talked about this in a previous lecture. Designing PCR products so that they carry a restriction enzyme sequence is no trouble at all! If you want to have one PCR oligonucleotide bind to the red sequence in the middle middle of this:
3'-TTACTAGACTAAGCTAGAGACTTACGAGACCATGAC-5'
for example, you might write it as a complementary base sequence:

5' TTCGATCTCTGAATGCTC 3'

Where can you add a restriction site such as EcoRI (G^AATTC)? You can't add it at the 3' end of the oligonucleotide, because then that end would no longer match the template and be extended:

```
3'-TTACTAGACTAAGCTAGAGACTTACGAGACCATGAC-5'
          5' TTCGATCTCTGAATGCTC          ...DOESN'T EXTEND
                                GAATTC 3'
                                EcoRI
```

You could however, add anything you want to the 5' end, and it will not interfere with polymerase function at the 3' end:

```
3'-TTACTAGACTAAGCTAGAGACTTACGAGACCATGAC-5'
             TTCGATCTCTGAATGCTC---->    ...IT WORKS!
    5' GAATTC
       EcoRI
```

So if you extend your oligo at the 5' end, adding an EcoRI site to make:

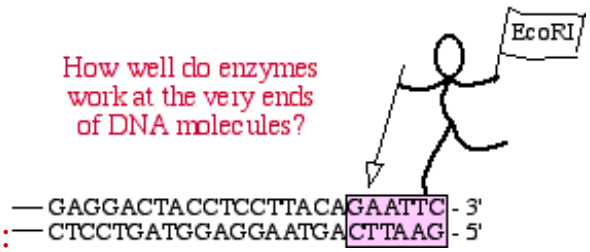
5' GAATTCTTCGATCTCTGAATGCTC 3'

Then you will create an enzyme recognition sequence right at the end of your DNA product. That is, the first nucleotide of the PCR product will also be the first nucleotide of the EcoRI recognition sequence.

Do you think the enzyme will be able to make use of that site, or is it too close to the end?

http://www.bfree.on.ca/bfinfo/sig/extreme/hanging.jpg

Hanging on...



A general question:

Answer: it depends on the enzyme

Some catalogs of enzymes provide anecdotal data on the efficiency of enzymes trying to work at the ends of DNA molecules. Generally, enzymes work better if they have a couple of extra nucleotides at the end - they don't do very well if they are perched on the end of a molecule.

Here are some examples:

			% Cleavage	
Enzyme	Oligo Sequence	Chain length	2 hr	20 hr
BamHI	CGGATCCG CGGGATCCCG CGCGGATCCGCG	8 10 12	10 >90 >90	25 >90 >90
EcoRI	GGAATTCC CGGAATTCCG CCGGAATTCCGG	8 10 12	>90 >90 >90	>90 >90 >90
HindIII	CAAGCTTG CCAAGCTTGG CCCAAGCTTGGG	8 10 12	0 0 10	0 0 75
Sma I	CCCGGG CCCCGGGG CCCCCGGGGG TCCCCCGGGGGA	6 8 10 12	0 0 10 >90	10 10 50 >90
Xba I	CTCTAGAG GCTCTAGAGC TGCTCTAGAGCA CTAGTCTAGACTAG	8 10 12 14	0 >90 75 75	0 >90 >90 >90

(Source: The New England Biolabs catalog)

Controlling the ends with a polymerase such as Klenow

The partial fill-in

If you want to use compatible ends during a ligation (to take advantage of the hydrogen bonds) but don't want to have vector reclosure problems, here's a trick that works in some circumstances. Suppose you've digested a vector with the enzyme XhoI (C^TCGAG), and you partially fill in the overhanging 5' ends with the enzyme Klenow and the substrates dCTP and dTTP. Note that the other two nucleoside triphosphates are excluded from the reaction. Here is what would happen:

The Xho I "partial fill-in" reaction	
Before digestion with Xho I	GAGGCTCGAGAATAC CTCCGAGCTCTTATG
After digestion with Xho I	GAGGC TCGAGAATAC CTCCGAGCT CTTATG

After partial fill-in with dCTP and dTTP	GAGGCTC TCGAGAATAC CTCCGAGCT CTCTTATG

Now you've created a two base 5' overhang that is incompatible with itself, so it cannot reclose naturally. On the other hand, the 5'-TC overhang is compatible with 5'-GA overhangs:

5' GA overhang	Source
GATCCNNNNN AGGNNNNN	BamHI end, partially filled in with dGTP and dATP
GATCTNNNNN AGANNNNN	Bgl II end, partially filled in with dGTP and dATP
GATCNNNNNN AGNNNNNN	Sau3AI end, partially filled in with dGTP and dATP
GATCANNNNN AGTNNNNN	Bcl I end, partially filled in with dGTP and dATP
GATCYNNNNN AGRNNNNN	Xho II end, partially filled in with dGTP and dATP

And so, if you prepare a DNA insert with one of these enzymes and partially fill in the ends (as shown), the problem of reclosures should be eliminated. Only the vector and insert ends can be ligated. For example, an Xho I site partially filled in, is compatible with a partially filled BamHI site (left) and BglII site (right):

```
GAGGCTC      GATCCNNN....NNNAAG      TCGAGAATAC
CTCCGAGCT      AGGNNN....NNNTCTAG      CTCTTATG
```

The partial chew back

Here's another way to control the ends - by the 3' to 5' exonuclease activity in the Klenow fragment or T4 DNA polymerase, we can make 5' overhanging ends of controlled composition.

I'll give you an example. Here's the sequence in the region of the multiple cloning site for the vector pUC18.

```
ATGCCTGCAGGTCGACTCTAGAGGAT
TACGGACGTCCAGCTGAGATCTCCTA
```

If we cut the vector with the enzyme Sma I (CCC^GGG), which cuts the sequence in only one place, then the two arms will look like this:

```
...ACTCTAGAGGATCCCC      GGGTACCGAGCTCG...
...TGAGATCTCCTAGGGG      CCCATGGCTCGAGC...
```

Now let's take the products of digestion and add Klenow fragment and only the substrate dATP. What do we get? It will look like this when the reaction reaches steady state.

```
...ACTCTAGAGGA              GGGTACCGAGCTCG...
...TGAGATCTCCTAGGGG            ATGGCTCGAGC...
```

I've marked the 3' terminal A nucleotides in red, because these will be removed and added over and over again, as long as there is a large excess of dATP in the reaction.

Now let us set that aside and design polymerase chain reaction primers to isolate this sequence:

```
CTAATACAGGACTACGGACT....CAGACTACGAGGACTACTACC
GATTATGTCCTGATGCCTGA....GTCTGATGCTCCTGATGATGG
```

For the oligonucleotide on the left, we design the following (and note the extension in blue):

```
5' TCCCCACTAATACAGGACTACGGA
```

For the oligonucleotide on the right, we design the following (and note the extension in blue):

```
5' CCCAGGTAGTAGTCCTCGTAGT
```

Our PCR product using these two oligonucleotides looks like this, initially:

```
TCCCCACTAATACAGGACTACGGACT....CAGACTACGAGGACTACTACCTGGG
AGGGGTGATTATGTCCTGATGCCTGA....GTCTGATGCTCCTGATGATGGACCC
```

But now suppose that we add Klenow fragment and only dTTP as a substrate. We would get this:

```
TCCCCACTAATACAGGACTACGGACT....CAGACTACGAGGACTACTACCT
      TGATTATGTCCTGATGCCTGA....GTCTGATGCTCCTGATGATGGACCC
```

Please note that this fits perfectly into our prepared vector, with a forced orientation:

```
...ACTCTAGAGGA                                GGGTACCGAGCTCG...
...TGAGATCTCCTAGGGG                              ATGGCTCGAGC...
```

It is important to note that we didn't have to touch our PCR product with a restriction enzyme to get these compatible ends. This is a great relief, because there are often times when our PCR product has an internal restriction site. If we digest the PCR product with a restriction enzyme, we run the risk of cutting it into smaller pieces than we had intended.

With this partial chew back method, we might need to treat the PCR product with T4 polynucleotide kinase to apply phosphates to the 5' ends (for efficient cloning with T4 DNA ligase) but the method is otherwise very straightforward. You have a considerable amount of control over the ends of DNAs, and can use the design process of PCR to make compatible ends.

Note: Precise chewing of the ends of the vector requires a large excess of substrate, perhaps 0.5 mM to 1 mM, and a short reaction time, because you don't want the chewing reaction to slip beyond the boundary you have set. For example, in the case of preparation of the vector, what you planned to make is this:

```
...ACTCTAGAGGA              GGGTACCGAGCTCG...
...TGAGATCTCCTAGGGG            ATGGCTCGAGC...
```

If the dATP concentration is too low, however, the exonuclease may occasionally remove an A nucleotide and not replace it. You may start to get this as a product (or worse!):

```
...ACTCTAGA                 GGGTACCGAGCTCG...
...TGAGATCTCCTAGGGG                    AGC...
```

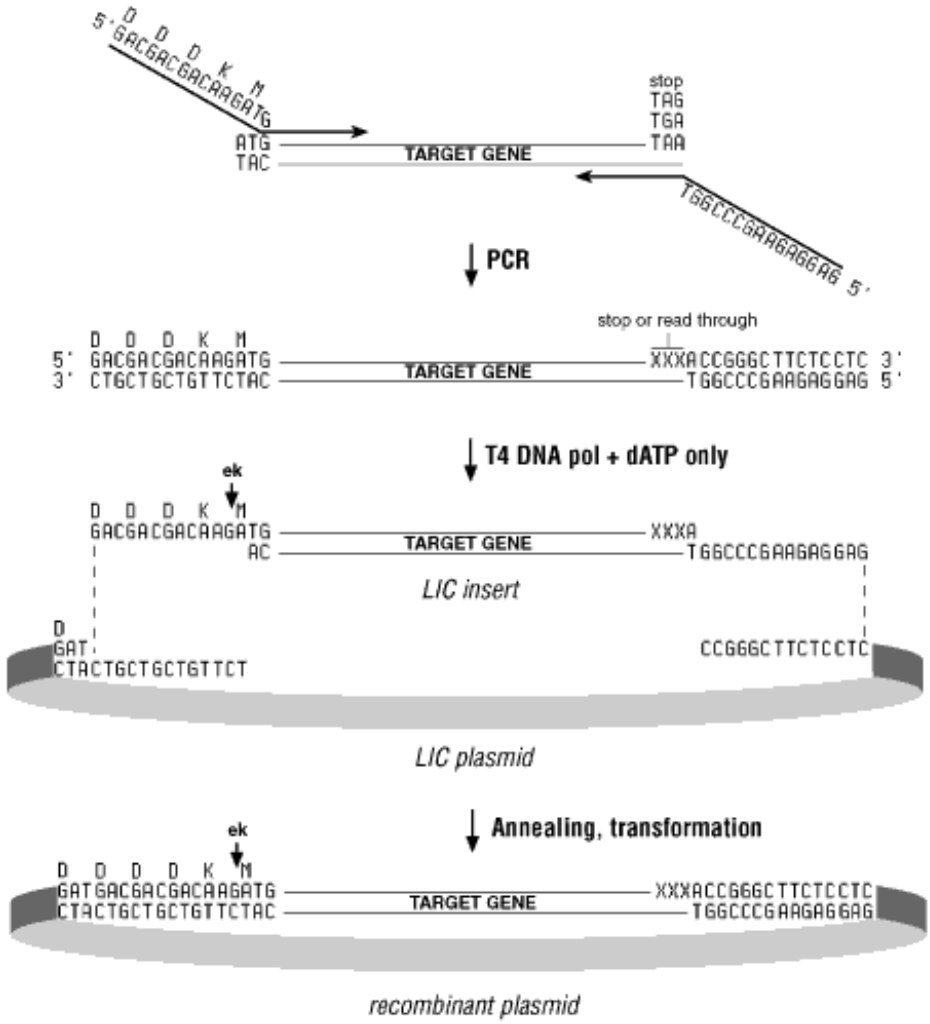
A better design would have a bit of redundancy so that a single "slip" of the exonuclease does not ruin the experiment.

Long chew-back methods can be taken to extremes, as in the case of "ligation independent cloning" described below.

The mighty H bond

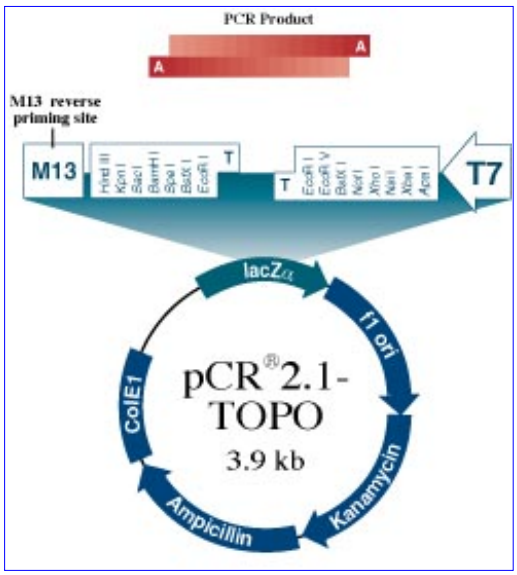
Here's an interesting idea - is it actually necessary to make the plasmid a covalently closed circle before transformation? Answer: No. If there are enough hydrogen bonds to hold an inserted DNA in place in a vector, as in this example called "Ligation Independent Cloning (LIC)". In this method, a polymerase chain reaction is conducted using oligonucleotides with "T-less" extensions. The complementary strand will have a long "A-less" stretch at the 3' end, and upon treating the product with dATP and a 3'-5' exonuclease (Pfu polymerase, T4 DNA polymerase, Klenow, etc), the "A-less" stretch will be removed. Degradation of the strand will stop at the first A nucleotide. In this case, a long (13-14 nt) 5' overhanging end will be generated.

The pLIC plasmid has a complementary sequence, and is similarly treated with a 3'-5' exonuclease and dTTP. The long 5' overhanging ends of the vector and PCR product can hydrogen bond stably, and the annealed product can be directly transformed into E. coli (which will repair the nicks).



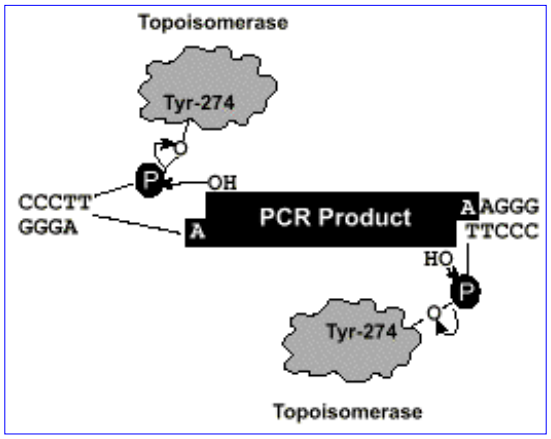
http://www.novagen.com/

Topoisomerase methods DNA ligase is not the only enzyme that can repair a phosphodiester bond. Topoisomerase I is another, and it is the principle behind the TOPO cloning kits from Invitrogen. For example, the vector pCR2.1-TOPO is designed to accept PCR products made with Taq DNA polymerase (and therefore having 3' overhanging A nucleotides).



Source: http://www.invitrogen.com/catalog_project/cat_topota.html

As they sell it, the vector is already linearized and has 3' overhanging T residues, and it is "activated" by having a Vaccinia virus topoisomerase I tethered to each end:



Source: http://www.invitrogen.com/pdf_manuals/topota_man.pdf

Because of the kinetic advantages of having the topoisomerase enzyme on a leash, the reaction only takes 30 seconds to 5 minutes to run! Compare that to a typical ligation reaction that might be run overnight.

The topoisomerase requires that the 5' end have a free hydroxyl, rather than a phosphate group. That's different, as you will recall, from DNA ligase which requires the phosphate. It is fortuitous, however, for PCR products that tend to be made with oligonucleotides having 5'-OH ends.

You just need to keep the specific requirements of the reactions in mind, and you won't have a problem.

Stan Metzenberg
Department of Biology
California State University Northridge
Northridge CA 91330-8303
stan.metzenberg@csun.edu