Lecture 6

Engineering DNA Ends

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

Getting the right fragment

We know that restriction enzymes have specific short sequences in DNA that they recognize, and that these enzymes hydrolyze the phosphodiester backbone in specific and predictable places on each strand. If the cut sites in the backbones are not immediately opposite each other, overhanging (cohesive) ends will be created on each fragment. These "sticky ends" are held together weakly by base-pairing interactions (hydrogen bonds) and may be broken naturally by thermal motion in the solution.

One may use restriction enzymes as a type of "molecular scissors" to excise a desired DNA fragment and transfer it into a cloning vector. Restriction enzyme sites appear randomly in genomic DNA, so getting exactly the right fragment (without too much or too little DNA) is sometimes tricky.

We know that some enzymes cut frequently in DNA sequences, and some less frequently, depending on how likely it is that the enzyme's specific recognition sequence will be found. Longer recognition sequences are statistically less likely, so enzymes looking for those sites are likely to be "bored" most of the time!

Fortunately, you don't need to be bored while looking for restriction sites, because there are web-based programs that will scan sequences for you. An example is Webcutter.

Exercise

Here is an exercise that you can do on the web. In this exercise you will grab a DNA sequence (or part of it) from a DNA database and use a web-based program to prepare a restriction enzyme map and table for the sequence.

The coding sequence for Homo sapiens phenylalanine hydroxylase (PAH), Accession Number NM_000277 is available online through the National Center for Biotechnology Information, and here is the direct link to the sequence.

Go to the Webcutter program, and figure out how many Nco I sites (c^catgg) are in the PAH sequence.

	How many Nco I sites are in the PAH sequence?
☐	**No sites**
☐	**One site**
☐	**Two sites**
☐	**Five sites**

If you had trouble, then here's the trick - you need to copy the entire sequence and paste it into the Webcutter window, or you may specify the accession number in some cases. The information you are seeking is in the table of restriction enzymes, but you can also see the information in the form of a map.

Here's another example. I tried searching for the gene CXCR4 and GenBank gave me a list of different species types and patient examples. I chose the Gallus [chicken] CXCR4, and this was the first part of the output file from Webcutter:

```
            BstH2I
            Bsp143II                Alw21I                      MflI
         BsiI                       AspHI                       BstYI
tgcgctcgtggcgctcggacggcccggacctactcggtgctcggagtatggacggcagcatggacggtttggatc base pairs
acgcgagcaccgcgagcctgccgggcctggatgagccacgagcctcatacctgccgtcgtacctgccaaacctag 1 to 75
         BssSI                      Bbv12I                      BstX2I
            HaeII                   BsiHKAI                     XhoII
```

This is just the first segment of a long output file. The DNA is shown in red (in this figure) and the names of restriction enzymes that digest the DNA in that location are shown in black. For example, you can see the word Hae II on the bottom line, and it is underneath the sequence ggcgct on the top strand. Hae II recognizes the sequence rgcgc^y and ggcgct certainly fits that description.

I asked the program to generate a table of restriction enzyme sites by position, and this is what it produced (for the first line).

```
Table by Site Position
Cut     Enzyme        No.  Positions             Recognition
site    name          cuts of other sites        sequence
10      BsiI          1                          ctcgtg
10      BssSI         1                          ctcgtg
14      Bsp143II      1                          rgcgc/y
14      HaeII         1                          rgcgc/y
14      BstH2I        1                          rgcgc/y
41      AspHI         2    919                   gwgcw/c
41      Bbv12I        2    919                   gwgcw/c
41      Alw21I        2    919                   gwgcw/c
41      BsiHKAI       2    919                   gwgcw/c
71      BstYI         4    190 925 1186          r/gatcy
71      BstX2I        4    190 925 1186          r/gatcy
71      MflI          4    190 925 1186          r/gatcy
71      XhoII         4    190 925 1186          r/gatcy
```

As you can see from the table, the cut site is listed in the first column (in increasing order), followed by the enzyme name, number of total cuts (in the 1325 nt example I

chose), positions of the other cut sites, and enzyme recognition sequence. From this we can determine that the Hae II enzyme cuts only at position 14 -- we say then that Hae II is a unique enzyme.

I also asked for a table of enzyme cuts sites to be generated, listed alphabetically by enzyme name. This is useful for looking up individual enzymes that you may have in your lab freezer. I've reprinted part of the table below, and you can see for example that Afl III cuts the sequence only once (at position 862) and Apo I cuts the sequence twice (at 96 and 1238).

```
Table by Enzyme Name
Enzyme       No.  Positions                          Recognition
name         cuts of sites                           sequence
AccB1I        1   993                                g/gyrcc
AccB7I        1   379                                ccannnn/ntgg
AcsI          2   96 1238                            r/aatty
AflIII        1   862                                a/crygt
Alw21I        2   41 919                             gwgcw/c
Alw44I        1   915                                g/tgcac
AlwNI         1   585                                cagnnn/ctg
ApaLI         1   915                                g/tgcac
ApoI          2   96 1238                            r/aatty
AspHI         2   41 919                             gwgcw/c
```

In the example I've given, I restricted the Webcutter output so that it would only show enzymes with recognition sequences of length 6 or greater. What type of map would I get if I looked at all enzymes, including 4 and 5-cutters? See below.

```
    AspLEI HspAI Bsp143II CviJI HpaII Bme18I AspS9I BsiHKAI   Hsp92II  MboI
  Hin6I   BssSI CfoI Cfr13I PalI MspI SinI AvaII BmyI   Fsp4HI         BstX2I
  HinP1I  BsiI AspLEI Sau96I BsiSI MspR9I HgiEI AspHI   BsoFI NlaIII   BstYI
tgcgctcgtggcgctcggacggcccggacctactcggtgctcggagtatggacggcagcatggacggtttggatc base pairs
acgcgagcaccgcgagcctgccgggcctggatgagccacgagcctcatacctgccgtcgtacctgccaaacctag 1 to 75
  HspAI MwoI HhaI BstH2I HaeIII BcnI Cfr13I Eco47I Bbv12I  BbvI        DpnII
    HhaI   HinP1I  MwoI AspS9I NciI ScrFI AsuI SduI Alw21I Bst71I      NdeII
    CfoI   Hin6I HaeII AsuI BsuRI HapII Sau96I Bsp1286I ItaI           Sau3AI
```

What a mess! Most of these enzymes would be of little use in cloning because they cut the DNA in too many different places. For example, Sau96I shown in green, cuts at g^gncc, and that appears twice in this short segment.

Your turn...

Experiment on your own with these web-based programs: the NCBI sequence viewer and the Webcutter program.

Now you've had some practice using the database, and a web-based restriction enzyme analysis program. Let's talk about some of the properties of restriction enzymes, and in particular the types of ends they leave after cleavage.

Recognition sequence and DNA ends

Take a look at the following examples of DNA restriction enzyme sequences:

Kas I (G^GCGCC)
Nar I (GG^CGCC)
Ehe I (GGC^GCC)
Bbe I (GGCGC^C)

You see that the same sequence is recognized by four isoschizomers that break the phosphodiester backbone differently. The ends generated by these four would consequently be different:

```
Kas I   NNNNNG       GCGCCNNNNNN
        NNNNNCCGCG       GNNNNNN

Nar I   NNNNNGG       CGCCNNNNNN
        NNNNNCCGC       GGNNNNNN

Ehe I   NNNNNGGC       GCCNNNNNN
        NNNNNCCG       CGGNNNNNN

Bbe I   NNNNNGGCGC       CNNNNNN
        NNNNNC       CGCGGNNNNNN
```

The point here is that enzymes leave different types of DNA ends, and this is a matter that is independent of recognition sequence. In the example above, a digestion product using Kas I would not be compatible with a digestion product of Nar I, because they could not hydrogen bond.

```
Kas I   NNNNNG       CGCCNNNNNN  Nar I
        NNNNNCCGCG     GGNNNNNN
```

Bbe I has a GCGC-3' overhanging end, and similarly it cannot anneal to any of the other three examples. It does have ends that are compatible with ends generated by Hae II (RGCGC^Y) however. Here is an example:

```
Hae II NNNNNAGCGC       CNNNNNN  Nar I
       NNNNNT       CGCGGNNNNNN
```

In this situation, the ends would match perfectly and the phosphodiester bonds could be sealed with the enzyme T4 DNA ligase.

Question time:
Would the ligated sequence NNNNNAGCGCCNNNNN be a Nar I site anymore?

	Is it a Nar I site?
☐	**Yes**
☐	**No**

Would NNNNNAGCGCCNNNNN be a Hae II site anymore?

	Is it a Hae II site?
☐	**Yes**
☐	**No**

Are there any HaeII-NarI site fusions that would preserve both enzyme sites in the ligated product?

	Could it be done?
☐	**Yes**
☐	**No**

Blunt ends: the great equalizer

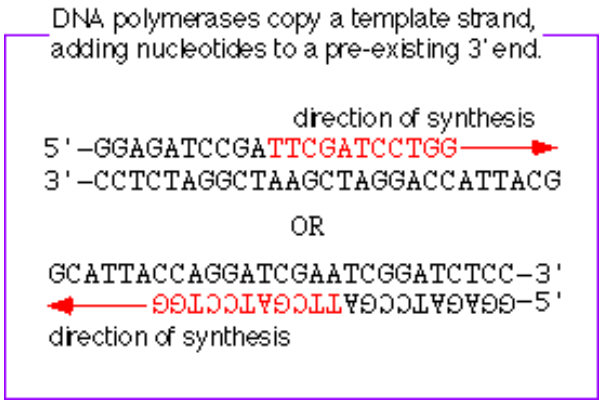
Blunt ends are always compatible with each other, because there are no H-bonds being formed that would define compatibility or incompatibility. So, a DNA end generated by Ehe I is compatible with a DNA end generated by EcoRV (GAT^ATC):

```
Ehe I  NNNNNGGC      ATCNNNNNN  EcoRV
       NNNNNCCG      TAGNNNNNN
```

This is a mixed blessing, for while the ends will always fit together there is a lack of specificity in assembly. Having cohesive ends gives better control of the assembly process because you can force the DNA fragment to be inserted in a single orientation. For example:

```
  BamHI        BamHI   EcoRI         EcoRI
NNNNNNG        GATCCNNNNNNNG         AATTCNNNNNNNNN
NNNNNNCCTAG        GNNNNNNNCTTAA         GNNNNNNNNN
```

In this example, the green DNA fragment (center) can only be inserted with the BamHI site on the left and EcoRI site on the right. This is called forced cloning, and it is not possible when the ends are blunt.
We can make a cohesive end into a blunt end using DNA polymerases such as Klenow (fragment of E. coli DNA polymerase I), T4 DNA polymerase, or Pfu polymerase. Let's review:



In this case, a 5' overhanging end is being filled in with newly-synthesized DNA. Restriction enzymes typically leave small overhanging ends, and they are usually of the 5' overhanging type.

```
Bam HI  NNNNNNG                 GATCCNNNNN
        NNNNNNCCTAG                 GNNNNN
```

Fill in one G nucleotide, and you have:

```
        NNNNNNGG                GATCCNNNNN
        NNNNNNCCTAG                 GGNNNNN
```

Fill in the next A nucleotide, and you have:

```
        NNNNNNGGA               GATCCNNNNN
        NNNNNNCCTAG                 AGGNNNNN
```

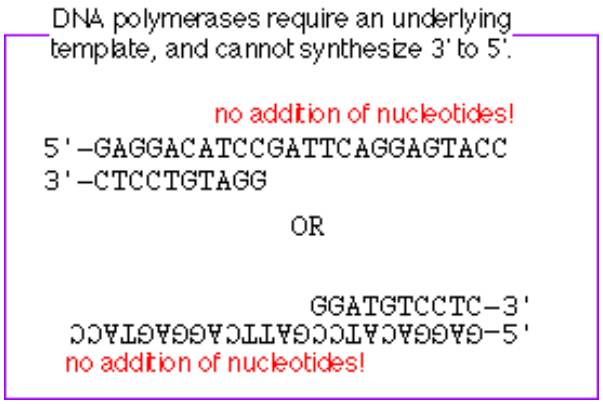
Then the next T nucleotide, and you have:

```
        NNNNNNGGAT              GATCCNNNNN
        NNNNNNCCTAG                 TAGGNNNNN
```

Finally the next C nucleotide, and you have a blunt end:

```
        NNNNNNGGATC             GATCCNNNNN
        NNNNNNCCTAG                 CTAGGNNNNN
```

Now the enzyme cannot add additional nucleotides to the 3' end because it requires a template:



If there is a 3' overhanging end, then the 3' to 5' exonuclease removes it, leaving a blunt end also.

Here are some examples of what the enzymes mentioned earlier (Klenow, T4 DNA polymerase, or Pfu) would do to ends left by the four restriction enzymes mentioned earlier:

```
Kas I   NNNNNG       GCGCCNNNNNN
        NNNNNCCGCG       GNNNNNN

...would be filled in to yield:

        NNNNNGGCGC   GCGCCNNNNNN
        NNNNNCCGCG   CGCGGNNNNNN

Nar I   NNNNNGG       CGCCNNNNNN
        NNNNNCCGC       GGNNNNNN

...would be filled in to yield:

        NNNNNGGCG     CGCCNNNNNN
        NNNNNCCGC     GCGGNNNNNN

Ehe I   NNNNNGGC       GCCNNNNNN
        NNNNNCCG       CGGNNNNNN

...would be unchanged

Bbe I   NNNNNGGCGC       CNNNNNN
        NNNNNC       CGCGGNNNNNN

...would be subject to the 3'-5' exo, leaving:

        NNNNNG           CNNNNNN
        NNNNNC           GNNNNNN
```

Having modified the DNA ends left by these four enzymes, all are now mutually compatible, and would be compatible with other blunt ends. Note that where modifications have taken place, the enzyme site is generally destroyed upon religation. Sometimes that's exactly what you want.

Protocol example:

Here is a protocol for Klenow treatment, downloaded from the Fermentas Inc. site

Protocol for Filling-in Recessed 3'-termini of Double-stranded DNA (with Klenow Fragment)

1.Dissolve 0.1-4µg of digested DNA in 10-15µl of water.
2.Add:

10X reaction buffer 2µl,
2mM 4dNTP mix 0.5µl (0.05mM - final concentration),
Klenow fragment 1-5u,
deionized water up to 20µl.

3.Incubate the mixture at 37°C for 10 minutes.
4.Stop the reaction by heating at 70°C for 10 minutes.

Reference

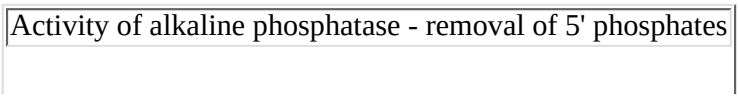
1.Current Protocols in Molecular Biology, vol. 1 (Ausubel, F.M., et al., ed.), John Wiley & Sons, Inc., Brooklyn, New York, 3.5.7-3.5.10, 1994-1997.

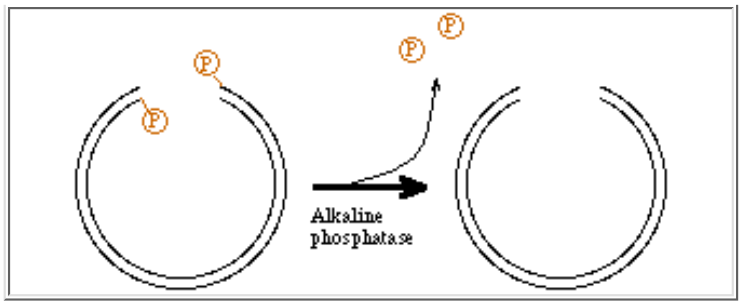
Source: http://fermentas.com/techinfo/modifyingenzymes/protocols/p_filrec3termdblstrdna_kf.htm

You may also be interested in reading through the kit instructions for the Fermentas DNA Blunting and Ligation kit. This site explains some of the content that has been discussed to this point.

Putting it together - the right way.

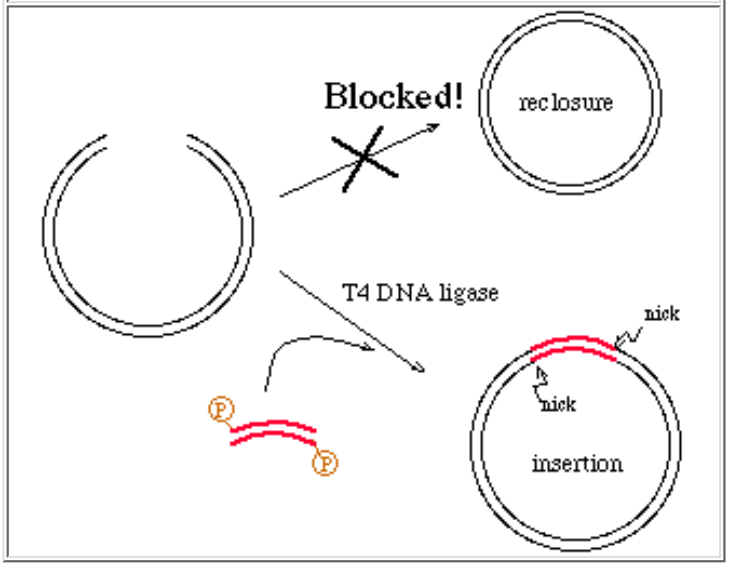
As we've discussed in class, we use the enzyme T4 DNA ligase to make covalent connections in the phosphodiester backbone. It was indicated in a previous lecture that 5' ends of DNA usually have a phosphate group, and we know that the phosphate group is required for ligase activity (as is ATP as a source of energy). We've also already discussed an enzyme (T4 polynucleotide kinase) that can be used to add a 5' phosphate where one is lacking, for example on a PCR oligonucleotide primer. When DNA is treated with the enzyme alkaline phosphatase, the 5' phosphate groups are removed.





Here's a nice application: If a linearized vector is dephosphorylated in this way, it cannot reclose upon itself because the enzyme T4 DNA ligase requires that a 5' phosphate group be present. A DNA fragment that has 5' phosphates still present can form a bridge between the dephosphorylated ends, so insertions are favored! When you are trying to combine two molecules, this removal of 5' phosphates from the vector (alone) keeps it from reclosing on itself and spoiling the construction.

Preventing reclosures by use of a dephosphorylated vector



What you get in the end: There are two widely separated nicks in the final product, because two of the four ligation events were prevented by the lack of 5' phosphates. Still, two out of four is good enough! The bacteria will fix the remaining nicks after the DNA is transformed.

Two sources of alkaline phosphatase are commonly used for this work:

- Calf intestinal alkaline phosphatase (CIAP, or CIP)
- Shrimp alkaline phosphatase (SAP)

The shrimp alkaline phosphatase is heat sensitive (it is derived from an Arctic shrimp that loves the cold!), so the enzyme can easily be inactivated at a moderately high temperature (65 degrees, 15 minutes). The calf intestinal alkaline phosphatase is relatively stable, so it must be inactivated at higher temperature, or via digestion with proteinase-K enzyme.

Why is it so important to inactivate the alkaline phosphatase enzyme? Because if it contaminates your ligation reaction, it will strip the 5' phosphates off of the DNA insert as well. That will block all ligation events, including the ones you want!

Protocol example: Here is a protocol for CIAP treatment, downloaded from the Fermentas Inc. site

Protocol for Dephosphorylation of DNA 5'-termini (with Calf Intestine Alkaline Phosphatase)

1.Dissolve DNA (1-20 picomoles of DNA termini) in 10-40µl deionized water.
2.Prepare reaction mixture by adding the following:

DNA solution 10-40µl,
10X reaction buffer 5µl,
deionized water to 49µl,
alkaline phosphatase 1u/µl.

3.Incubate at 37°C for 30 minutes.
4.Stop reaction by heating at 85°C for 15 minutes or extract DNA with phenol/chloroform and then precipitate with ethanol.

Note
Dephosphorylation can be performed by adding calf intestine alkaline phosphatase directly in mixture after DNA cleavage with a restriction endonuclease. We recommend using 0.05 units calf intestine alkaline phosphatase for dephosphorylation of 1 picomole DNA termini. The enzyme can be diluted with 1X reaction buffer.

Reference
Current Protocols in Molecular Biology, vol. 1 (Ausubel, F.M. et al., ed.), John Wiley & Sons, Inc., Brooklyn, NY, 3.10.1-3.10.2, 1994-1997.

Source: http://www.fermentas.com/techinfo/modifyingenzymes/protocols/p_dephosph53dna_cip.htm

The reclosure problem and the statistics of ligation

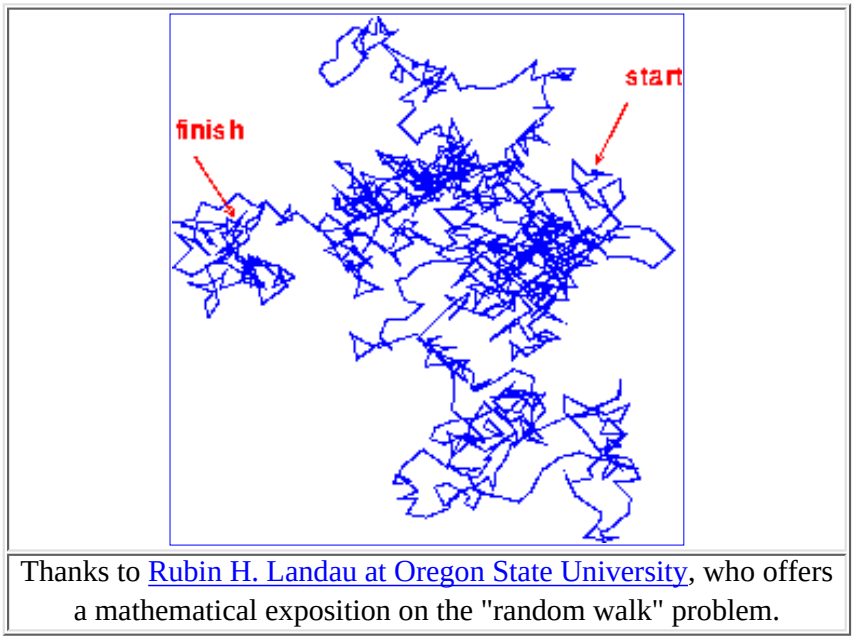
The method just described, of treating a vector with CIAP, is one of many that are used to prevent reclosure of a plasmid vector without inserted DNA. The process of combining two pieces of DNA (a bimolecular ligation) is tricky because it is often the case that the two ends of one piece of DNA are closer to each other in solution than DNA ends from two different molecules. Why? Because the two ends of a single DNA molecule are tethered to each other through the DNA molecule. On the other hand, in a dilute ligation reaction the two ends of different DNA molecules may rarely bump into each other.

If you attempt to reclose a plasmid by ligation, what are you actually doing? You are asking that the two ends of the DNA "find each other" in the solution, and that T4 DNA ligase covalently connect the phosphodiester backbones (using a bit of energy taken from an ATP molecule).

The question you may then ask, is:

"How easy is it for two DNA ends to find each other in solution?"

To answer that, we need to delve into the subjects of public drunkenness and physical chemistry. Come to think of it, those are pretty much the same thing. Picture in your mind a drunken person, hanging onto a lamppost for support. As he staggers away from the post, the direction of each step is random. How far away from the lamppost will he tend to be if he has taken N steps? The path of the person may look something like this, after 1000 steps.



Thanks to Rubin H. Landau at Oregon State University, who offers a mathematical exposition on the "random walk" problem.

Where R is the distance from the lamppost to the drunk, we have the greatest likelihood that:

$$R = \sqrt{N} r_{rms}$$

Where N is the total number of steps taken, and r_{rms} is the square root of the average squared step size or root mean squared step size. Of course the result has a statistical outcome, and this is just the distance with the highest probability in a distribution. If you move the drunk back to the lamppost and have him try again, you will most likely get a different result each time.

A linear piece of DNA tends to spread out in solution by a "random walk" of short segments in much the same way as the drunk staggers away from the lamppost. The distance from one end of the DNA to the other is most likely to be $\sqrt{N} r_{rms}$

where

> r_{rms} represents a characteristic (root mean square) step length (called the "persistence length") which is related to the flexibility of the DNA (its sequence characteristics) and the hydration of the nucleic acid (the solution characteristics). If the DNA is more flexible, the step length is smaller.
>
> N represents the number of these segments or "steps" in the entire DNA.

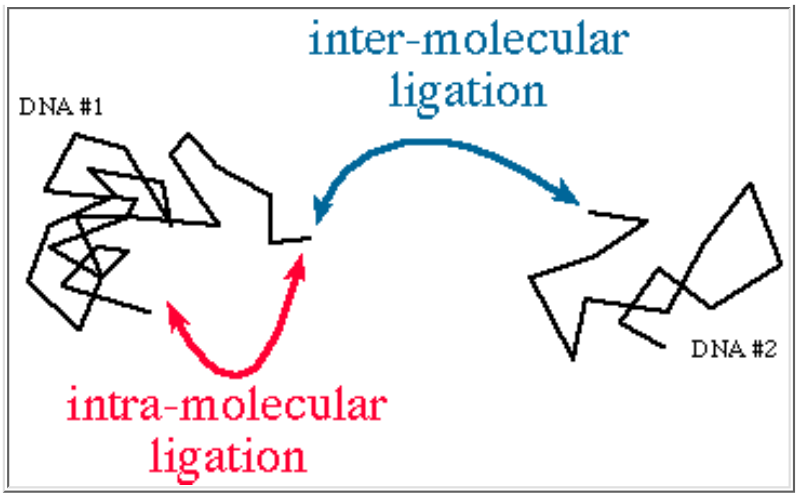
In the case of our attempting to ligate DNA ends in a single linear molecule, we want the two ends of a DNA molecule to "find each other" in solution. That is analogous to expecting the drunk to find their way back to the lamppost after taking N steps in random directions. The probability of this happening is proportional to time (because the DNA is always changing conformation and trying different "random walks") and the square root of the length of the DNA (by the above equations).

Let us now attempt to sober-up!

So far we have been discussing the behavior of a single linear DNA molecule, and I think you will agree that the chances that one end of a **single molecule** will find the other end of the same molecule do not depend on the concentration of DNA in the solution. On the other hand, if we want the ends of **two different DNA molecules** to find each other in solution, their independent concentrations will be extremely important! If the concentration of either DNA end in solution is too low, the other end cannot find it in a reasonable period of time.

Suppose you now want to insert a DNA fragment into a linearized vector. This isn't quite the same situation as we had considered before, because now the efficiency of the reaction will depend on DNA concentration. We want to have inter-molecular ligation; ligation between the two different molecules. At the same time, we want to avoid reclosure of the plasmid without an inserted DNA fragment (intra-molecular ligation)

What are the chances of intra- vs. inter-molecular ligation?



If the DNA is very concentrated, inter-molecular ligation is more likely. If the DNA is dilute in solution, then intra-molecular ligations are more likely.

Therefore, the following is sound advice:

- If you desire to circularize a single linear molecule, the ligation should be performed at low DNA concentration.
- If you desire to combine two different DNA molecules by ligation, the concentration of each should be high - on the order of 1 micromolar. The optimal **molar ratio** of "insert" to "vector" is generally taken to be about 2:1.

Remembrance of things past...

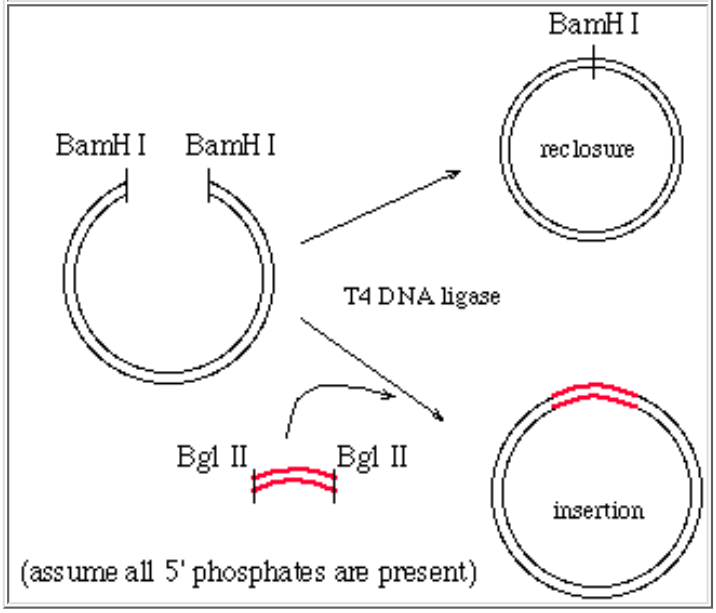
Do you remember in your old chemistry class, how a careful distinction was made between the concept of "concentration" and "activity?" You may recall that square brackets **[DNA]** were used to indicate concentration in an equation, whereas round brackets **(DNA)** were used to indicate activity. Here's where that little pearl of wisdom will finally become useful! It is possible for us to change the solution characteristics so that **(DNA)** > **[DNA]**. If we add **polyethylene glycol** (PEG) to a ligation reaction, it ends up taking over and monopolizing some of the aqueous solution volume. The DNA has less space, because it has to share the solution with the PEG molecules which are real hydrogen bond hogs! Since the DNA has less space, it can find other molecules of DNA more easily. Its concentration (i.e. moles per liter or grams per liter) hasn't changed, but its "activity" is higher.

A different trick

How do we solve the problem of reclosure, if we don't want to treat the vector with alkaline phosphatase? Here's another trick that works in some circumstances. If an unwanted ligation product (for example, a reclosure) has a restriction enzyme site, and that site is not present in the desired product, the unwanted product can be specifically linearized with the enzyme **after** ligation is completed.

For example, suppose we were attempting to clone a fragment that has Bgl II ends (Bgl II cuts at A^GATCT), into a vector that had been linearized with BamHI (G^GATCC). These ends are compatible but their ligation eliminates both the BglII and BamHI recognition sites.

The unwanted reclosure has a BamHI site!



The reclosed plasmid has a BamHI site (because the two halves just reformed into a complete site), but the desired product does not! If both are present in the ligated material, the reclosures can be specifically linearized by treatment with BamHI. That is, their ligation can be effectively reversed!

Yet another trick:

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.

Linkers and adapters can create a new restriction site

Suppose you have a collection of DNA fragments with blunt ends, and you want to introduce them efficiently into a vector that is linearized at an EcoRI site. From what you have learned so far, you might think that the best thing to do would be to make the vector ends blunt as well so that they will be compatible. You could...but that would not give you a high efficiency of ligation (meaning fewer candidate colonies per transformation). It may also be the case that you would like your product to retain EcoRI sites. If you make the vector ends blunt, you will destroy those EcoRI sites.

One solution is to ligate a small double-stranded DNA containing an EcoRI site to the end of the blunt DNA fragments. Since this small DNA can be added to the ligation reaction in great excess, it is an efficient reaction. Then the DNA can be digested with EcoRI to create the proper overhanging ends that will be compatible with the vector. We call this small piece of DNA a **linker**.

Ligation and digestion of linker	
NNNNN NNNNN	Blunt end - before addition of linker
GCCGGAATTCCGGNNNNNN CGGCCTTAAGGCCNNNNNN	After ligation of linker
AATTCCGGNNNNNN GGCCNNNNNN	After digestion of linker

Note that if the linker has 5' phosphates (i.e. if it is phosphorylated) then a great many linkers may be ligated to the fragment in a tandem repeat. These will all be digested away by the enzyme, leaving only the proximal sequence as shown.

One may also use a non-phosphorylated linker, in which case only one ligation event will occur - between the 5' end of the fragment and one of the 3' ends of the linker. In the example given:

AATTCCGGNNNNNN GGCCNNNNNN	After digestion of linker

If the linker had not been phosphorylated, the GGCC of the lower strand would not be ligated to the NNNNNN because the C-5' would have lacked a phosphate at the time of ligation.

A problem exists with linkers however. What happens if you add an EcoRI linker (as shown above) but your fragment has an EcoRI site right in the middle of it? If you were to treat your ligated linker plus fragment with the enzyme, you would digest your fragment right in the middle, which would probably make you sad!

Adapters solve the problem, without need for digestion

An adapter is (at least in some terminologies) a pair of non-phosphorylated single strand DNAs that hydrogen bond to create one overhanging and one blunt end.

Ligation of non-phosphorylated adapter	
AATTGGCCGC NNNNN CCGGCG NNNNN	Adapter andblunt end
AATTGGCCGCNNNNN CCGGCGNNNNN	After ligation

The cohesive end is added automatically - there's no need to treat the fragment with restriction enzyme, so internal sites are safe! These linkers and adapters are some of the tricks of the trade for engineering DNA ends, but don't forget that one of the most powerful ways of controlling DNA ends is a method we've already discussed: The polymerase chain reaction.

Don't touch that dial! In the next lecture, we'll look at a few more ways of controlling the ends of DNA, and controlling the ligation reaction without using restriction enzymes or DNA ligase!

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