Lecture 4

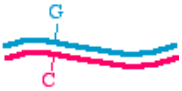
Site-directed mutagenesis

In which we learn how to obtain precise control over the coding content of DNA

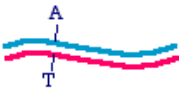
The principle of site-directed mutagenesis is that a mismatched oligonucleotide is extended, incorporating the "mutation" into a strand of DNA that can be cloned. In this lecture, I will present a number of current methods in use.

First, let's talk about the approaches in very general terms, because that will allow us to organize the specific methods in our minds. When we talk about making a specific mutation, let's call the molecule that we are starting with, the one without the mutation, the "parent" molecule. It might look like this:

The top strand is blue and the bottom strand is magenta. Let's be a bit more specific and point to a specific nucleotide pair that we intend to change, as follows:



So let us say that in the parent molecule there is a GC base pair that we want to change into an AT base pair. We want the mutated version to look like this:



The first general approach is to take the parent molecule and convert it to the mutated version by polymerase chain reaction. The mutation is made by having a mismatch between the parental template and one (or more) oligonucleotide primers. You can start with a very small amount of the parent molecule, and by PCR make a tremendous amount of the mutated version, so much in fact that the chances of cloning the mutated version from the product is essentially 99.999%. There is always a bit of room for error, of course, so you must carefully confirm your work by DNA sequencing.

To master this approach, we will need to learn a bit more about the design of primers. When you start thinking about making primers that are slightly mismatched with the template, you have to know what you can get away with and what you cannot.

Changing the end of a PCR fragment

Let's consider a very simple case, in which you want to make a change in the end of a DNA molecule. Suppose you have a PCR fragment that looks like this, where the dots indicate an extended sequence that is not shown:

```
TCTATGGACCAGTACGATACCAGTA.....CGACCTACGTAGACTAGACGGATAGAG
AGATACCTGGTCATGCTATGGTCAT.....GCTGGATGCATCTGATCTGCCTATCTC
```

The two oligonucleotides you used to make the fragment look like this:

Oligo on the left	TCTATGGACCAGTACGAT
Oligo on the right	CTCTATCCGTCTAGTCTA

These oligonucleotides are written 5' to 3'. If you don't understand the oligo on the right, then I suggest that you go back to review the lecture on PCR.

Suppose that you want to add an EcoRI site (GAATTC) to the end on the left, and a BamHI site (GGATCC) to the end on the right.

	Which of these might be correct for the new oligo on the right?
☐	5' GAATTCCTCTATCCGTCTAGTCTA
☐	5' CTCTATCCGTCTAGTCTAGGATCC
☐	5' GGATCCCTCTATCCGTCTAGTCTA
☐	5' ATCTGATCTGCCTATCTCCCATGG

No problem! We just make our two oligonucleotides a bit longer, and embed the extra sequence within.

Oligo on the left GCGAATTCTCTATGGACCAGTACGAT
Oligo on the right GCGGATCCCTCTATCCGTCTAGTCTA

The new PCR product would look like this:

```
GCGAATTCTCTATGGACCAGTACGATACCAGTA.....CGACCTACGTAGACTAGACGGATAGAGGGAT
CGCTTAAGAGATACCTGGTCATGCTATGGTCAT.....GCTGGATGCATCTGATCTGCCTATCTCCCTA
```

After we had digested it with the two enzymes, the bits on the ends would be lost and the product would be ready to clone:

```
    AATTCTCTATGGACCAGTACGATACCAGTA.....CGACCTACGTAGACTAGACGGATAGAGG

GAGATACCTGGTCATGCTATGGTCAT.....GCTGGATGCATCTGATCTGCCTATCTCCCTAG
```

Please note that we added a "GC" base pair to each end to make the enzymes work better - that is a subject for a future lecture, so don't worry about it just now. The important thing is that we managed to change the ends of the DNA, just by adding a bit of sequence to the 5' ends of each oligonucleotide.

We can do more than append a sequence - we could also change the parental sequence at the end without making the product any longer. Remember that our original oligonucleotide pair was:

Oligo on the left TCTATGGACCAGTACGAT
Oligo on the right CTCTATCCGTCTAGTCTA

First, I would like you to note that each is 18 nt in length, and that we would get exactly the same product if we extended one of them at the 3' end, so that it was 21 nt in length.

Oligo on the left TCTATGGACCAGTACGATACC

Why would we get the same product? Because it would look like this, which is the same as our original version (aside from the coloration)

```
TCTATGGACCAGTACGATACCAGTA.....CGACCTACGTAGACTAGACGGATAGAG
AGATACCTGGTCATGCTATGGTCAT.....GCTGGATGCATCTGATCTGCCTATCTC
```

Now let's have some fun, and make a change in the oligo on the left:

Oligo on the left TCGATGGACCAGTACGATACC

You see that we made a "G" at the third nucleotide instead of a "T". This will create a transversion mutation in the product:

```
TCGATGGACCAGTACGATACCAGTA.....CGACCTACGTAGACTAGACGGATAGAG
AGCTACCTGGTCATGCTATGGTCAT.....GCTGGATGCATCTGATCTGCCTATCTC
```

Both strands are affected, because the new version is simply copied into its complementary nucleotides on the bottom strand. So you see, we can make changes in the sequence that are internal.

Why did we extend the oligo on the left so that it was 21 nt in length? Well, we wanted to be sure that it would anneal correctly to the template in the very first cycle. If we introduce a mismatch, we want to be sure that there are an adequate number of matching nucleotides at the 3' end of the primer (18 is a safe number, in most cases). So you see that in this mutagenesis approach, the first annealing would be imperfect, and the 5' end of the oligo on the left would be single stranded for three nucleotides.

```
 5'T
   C
   G
   ATGGACCAGTACGATACC----->extension
AGATACCTGGTCATGCTATGGTCAT.....GCTGGATGCATCTGATCTGCCTATCTC
```

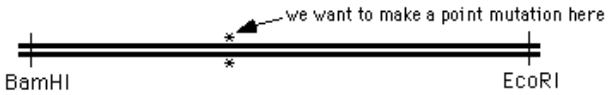
Of course, the mutation is copied into the product by extension from the oligo on the right, so once the PCR reaction is underway, the annealing will be perfect over the entire 21 nt of the primer:

```
5'TCGATGGACCAGTACGATACC----->extension
  AGCTACCTGGTCATGCTATGGTCAT.....GCTGGATGCATCTGATCTGCCTATCTC
```

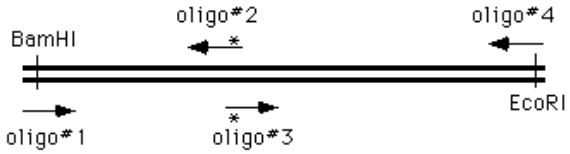
So you see, it is fairly straightforward to change a DNA sequence if it can be covered by an oligonucleotide during polymerase chain reaction.

Changing the middle of a sequence by two consecutive reactions; the 4 oligo method

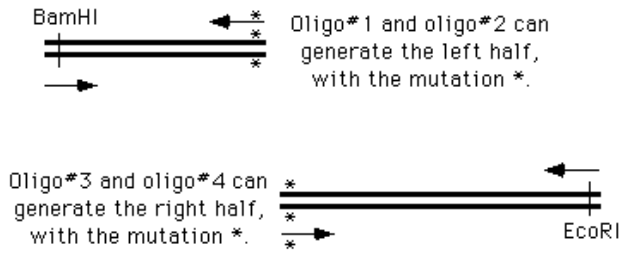
Suppose you want to do something a bit more challenging - creating a point mutation in the middle of a DNA sequence, at the position marked with an "*" in the figure:



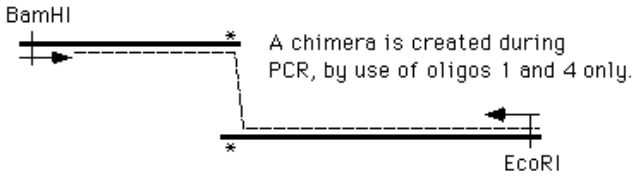
The ways of doing this in the old days were unspeakable, but now we can simply get on the phone and order four oligonucleotides; two of which are flanking and two of which cover and introduce the mutation into the amplified material:



We perform two PCR reactions to obtain the two halves of our final product, and combine them in a third reaction, using the two "outside" oligonucleotides to generate a chimeric product.



How does this happen? During the PCR process, the right side of the first molecule can prime the synthesis from the left side of the second.



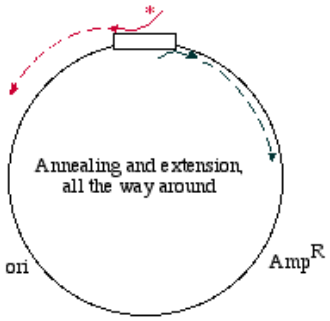
Now we can simply cut the PCR product with EcoRI and BamHI, and drop it into the vector, in place of the original version. Or, we can continue to manipulate the DNA by PCR.

To some extent, this is just using the end-based method we described in the first place, but doing it twice, and then combining the results into a single product.

Inverse PCR

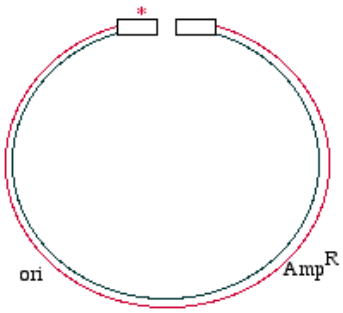
Here's a different approach, which would be appropriate if the DNA template is circular, for example in a plasmid:

We may start with a circular plasmid, and use two oligonucleotides to change a small region by PCR (see asterisk). The 5' ends of the oligonucleotides are shown not annealed - they do not base pair because they are mutagenized. The two oligos are situated in such a way that they re-copy the entire plasmid.



They point towards each other, but only going the "long way around." That is like deciding to go to Los Angeles for the day, but instead of heading down the 405 you go up the Pacific Coast Highway to Alaska, snowshoe over to Denmark, hop a train to Capetown, boat over to Tierra del Fuego, and bicycle up through South and Central America to Los Angeles! It's the long way to communte!

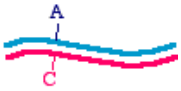
In this PCR example however, it makes sense because it means you don't need to combine two pieces for cloning. What you obtain in the end is a linear fragment, suitable for reclosure and cloning:



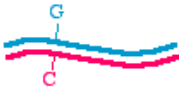
One comment however, is that the 5' ends of a PCR fragment are exactly the 5' ends of the oligonucleotides. You will need to have a 5' phosphate if you intend to use DNA ligase to reclose the circle, and so if your oligonucleotide does not have a 5' phosphate (which would be typical) then you need to apply a phosphate to each end using the enzyme T4 polynucleotide kinase.

One more thing - if you want this to work, you need to use very few template molecules in the reaction, perhaps 1000. If you start with, let's say just to be really gross, a nanogram of template, then you will have too much parental DNA lingering by during your transformation. That is, you will find that very few of your transformed bacteria actually have the mutation in the plasmid.

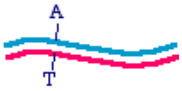
The second general approach to mutagenesis does not use polymerase chain reaction, but does use a polymerase. In effect, you change only one strand in the parent, and transform the bacteria with the heteroduplex, that might look like this:



You see that the blue strand has been mutated and is now mismatched with the magenta strand. At this point in the DNA, the mismatch would make a small bubble of single-stranded DNA. What will the bacteria do with this? When the DNA is replicated (typically as part of a plasmid), semiconservative replication will cause two different daughter molecules. One looks just like the parent, and the other has the mutation fixed in both strands:



Parent

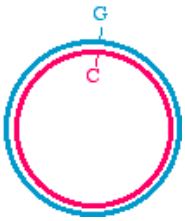


Mutated

That's great! Now half of the products will be mutated! Well, that would be true if we could efficiently make the heteroduplex in the first place (which is a bit of a dubious assumption), and if the heteroduplex could be transformed into cells with the same high efficiency as the parental version (which may not be true, since it has a bit of single-stranded DNA in it), and many other worries of a similar nature. We need to have some way of disfavoring the parental version in this contest. As we are making the mutation in one strand, we need to link that strand to some persuasive form of selection, like this:

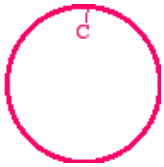
That is, we make the mutated (blue) strand a "happy strand" in some way or another, and the parental (magenta) strand into an "unhappy strand". By the strength of this linkage, we select for the mutated version by disfavoring the parental version.

When using this approach, it is common to employ single stranded DNA as a template (the magenta strand), because then you can simply apply a mutagenic oligonucleotide and make a second strand. The second strand you make will not have a parental strand with which to compete. Think of it as follows:



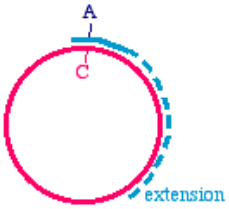
Start with a double-stranded plasmid with the parental sequence

Then, you make a single-stranded DNA containing just the inner strand of the plasmid, which would look like this:



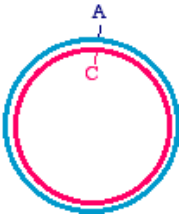
Just one of the parental strands

Then, apply an oligonucleotide that anneals to the single strand, and carries a mutation. Extend the primer with a DNA polymerase such as Klenow fragment:



Synthesize a second strand, incorporating the mutagenic primer

Note that there is an A/C mismatch at the top of the figure. Once extension is complete, it is double-stranded and might look like this. Of course, the Klenow enzyme does not make the blue strand covalently closed. There will be a "nick" where the synthesis ends, but don't worry about that - once this is transformed into bacteria, the bacterial host will repair the nick.



Heteroduplex intermediate, ready to transform into bacteria

Remember once again, that the mutagenized strand will be copied into half of the daughter molecules during replication.

So, that's just terrific, and I hear you wondering how we managed to get the single stranded template with which we started the method. It is a minor digression...

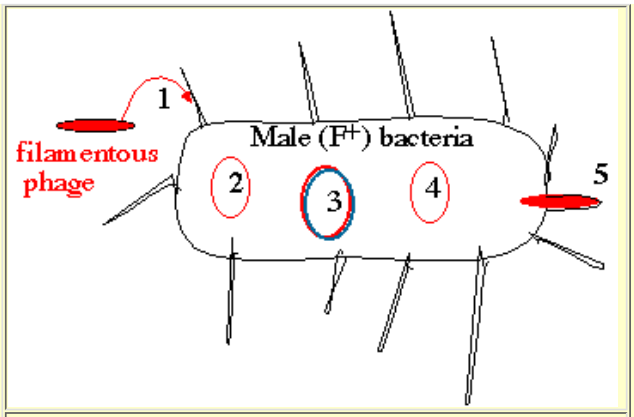
Origins of ssDNA replication - how to get single stranded DNA in the laboratory.

As we learned previously, an origin of DNA replication is a required element for ensuring plasmid maintenance. Origins of replication do come however, in several different "colors and styles." Most commercial plasmids are based on the ColE1 origin, a natural "high copy number origin" which fosters the accumulation of several hundred copies of a plasmid per bacterium. It is also not uncommon to find a second conditional origin of replication in some plasmids, derived from a filamentous bacteriophage such as M13, fd, or f1. These origins of replication have two important features:

1. They generate a single-stranded DNA product
2. They are only activated during co-infection with a helper phage

Why would we want to make single-stranded DNA? One reason would be to make a single stranded template for a sequencing reaction (a matter we will discuss later in the course), or a single stranded DNA probe. Site directed mutagenesis is sometimes facilitated by having a single-stranded plasmid to work with.

In any case, a slight digression to discuss the life-cycle of the filamentous bacteriophage might be in order:



Lifestyles of the filamentous bacteriophage

1. Phage enters through pili of male (F+/Hfr) bacteria
2. Introduced phage genome is single-stranded
3. A double-stranded "replicative form (RF)" is generated
4. New single-stranded genomes are copied from the RF
5. The new phage carrying the single-stranded genome buds out

Those poor male bacteria! They have to contend with invading filamentous phage - something that Rogaine just can't cure! We'll be talking more about male and female bacteria in a later lecture ("sex" in bacteria isn't quite the same concept as in eukaryotes).

What is significant here is that the **virion** of the filamentous phage (i.e. the viral particle) carries a single-strand of DNA - not a double helix. In the cell, this single-stranded genome (2.) is used as a template to synthesize a double-stranded replicative form (RF), which is essentially a plasmid (3.). The replicative form is used as a template to generate new single-stranded genomes (4.) that are packaged into virions (5.) to generate new phage. The cell doesn't die - it just grows more slowly and continues to secrete phage indefinitely.

The practical side of this story - if you use a cloning vector that is based on a filamentous bacteriophage (such as M13mp18 which is an engineered version of the phage M13) or merely contains an origin of replication from a filamentous bacteriophage (such as f1), then you can induce single-stranded DNA replication and collect the products in the form of secreted phage particles (which may be precipitated from the growth medium with polyethylene glycol). In the case of a plasmid that only contains an f1 origin of replication, and not the remaining genes from the phage, it is necessary to infect the plasmid containing cell with a filamentous "helper phage" that will activate the f1 origin of replication in the plasmid and foster viral secretion.

Now we see how to get single stranded DNA, but what about those persuasive selection methods? The smiley faces and frowning faces? What's that about?

Fundamentally, we're talking about approaches that allow us to distinguish one strand from another.

Restriction enzymes used to distinguish strands

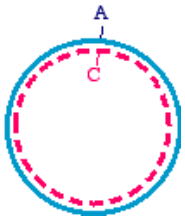
Suppose you had a parental DNA that had a unique restriction site in it. If you mutagenized the restriction site at the same time that you made a mutation in your gene of interest, then the parental strand would be sensitive to the enzyme and the other strand containing the mutation would not.

The company Clontech has such a method, called the Transformer Site-Directed Mutagenesis Kit. Digestion of the heteroduplex with the restriction enzyme debilitates the parental strand, because it introduces a "nick". The DNA

can then be transformed into a bacterial strain. The efficiency can be increased by extracting the pooled DNA from these cells and digesting a second time. This will eliminate the products of replication in the bacteria that are purely parental (homoduplex), and will spare the ones that are purely mutagenized (homoduplex). These plasmids can then be reintroduced into bacteria, and most of the surviving plasmids should be the mutagenized form.

Let me give you another example. There is a restriction enzyme named Dpn I that will cleave the sequence G^{Me}ATC where ^{Me}A means that the adenylate nucleotide is methylated. Dpn I will not cleave the unmethylated sequence GATC. We can methylate such sequences in a plasmid by growing the plasmid in a "dam+" strain of bacteria. Suppose then that we prepare a single stranded DNA template in such a "dam+" strain. The parental strand would be methylated at every GATC sequence (that is, approximately every 200 to 300 nt). When we apply an oligonucleotide primer to this template and extend it using Klenow fragment, however, the new DNA that is synthesized in vitro will be unmethylated. We therefore create a marked difference between the parental strand (methylated) and the mutagenized strand (unmethylated).

Once we have completed synthesis of the mutagenized strand, what would happen if we tried to digest the heteroduplex with Dpn I?



The answer is that the parental strand would be nicked (cleaved) in many places, but the mutagenized strand would not. By putting this extra damage into the parental strand, it is less favored during replication in the bacteria.

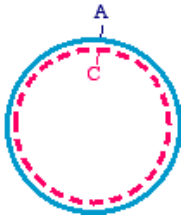
Uracil N-glycosylase used to distinguish strands

Here's another method, and this one involves taking advantage of the enzyme that we discussed in the first lecture that removes uridylate nucleotides from DNA (where they don't belong!)

How do we get a parental DNA that contains numerous uracil bases incorporated in place of thymidine bases? The answer is that we grow the plasmid in a strain that makes deoxyuridine triphosphate (a strain that is "dut-", meaning dUTPase deficient) and does not surveil the DNA for uracil to excise (a strain that is "ung-", meaning uracil N-glycosylase deficient). The bacteriologists usually don't say "minus" and "plus" by the way - they would just call such a strain "*dut*, *ung*", meaning that those two loci had mutated alleles.

So we make the single stranded parental DNA in a *dut*, *ung* strain of bacteria, apply the mutagenic oligonucleotide in vitro, and extend it using the usual DNA substrates and Klenow fragment. The newly synthesized DNA will not contain uracil bases, because we did not use dUTP as one of the substrates - only dATP, dGTP, dCTP, and dTTP.

What would happen if we treated this DNA with uracil N-glycosylase? Well, the parental strand would be shredded and the mutagenized strand would be untouched.



We don't actually have to add the enzyme ourselves - we could just take the heteroduplex and transform a wild type bacterium with it - one that was not *dut*, *ung* that is. The wild type bacteria would shred the parental strand specifically, because its uracil N-glycosylase would find the U-rich DNA to be offensive.

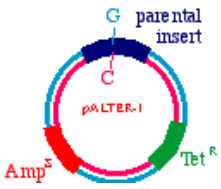
The Altered States® method.

"Altered States" - sounds like something that ought to have been invented by Kary Mullis, but you'll find it at Promega Inc.

1. Start with a plasmid carrying a defective selectable marker (e.g. Amp)
2. Link the mutation you are making elsewhere in the plasmid to a correction of the defective selectable marker.
3. Select for correction of the selectable marker, and you are likely to also find plasmids with your specific mutation introduced as well

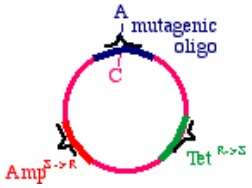
Sound's easy? Here's a diagram that help to explain it.

Start with a double stranded plasmid containing your DNA insert, the sequence you wish to mutagenize:

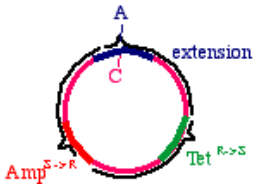


Note that there is a G/C base pair that we want to mutagenize to an A/T base pair, in our dark blue sequence (the parental insert). Also, there is a green sequence representing tetracycline resistance, and a red sequence that is a defective version of the ampicillin resistance gene. Since the Amp gene is defective, we will say that it is Amp^S meaning "sensitive".

Now we prepare a single stranded version of the plasmid, perhaps by simply denaturing them in alkali. We anneal THREE oligonucleotides to the circle: One to the DNA parental insert, that causes the mutation in our gene of interest (from a G to an A in this example), one to the tetracycline resistance gene that will debilitate it by the introduction of a mutation, and one to the ampicillin "sensitive" gene that will repair it by the introduction of a mutation.

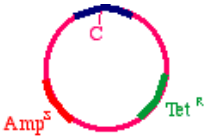


These oligonucleotides are extended clockwise around the plasmid using DNA polymerase Klenow fragment, so it looks like this:

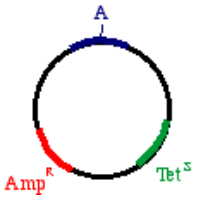


Note that this heteroduplex has three points of mismatch, in three entirely different places in the plasmid. The "inner strand" that contains the parental sequence is unchanged, but the outer strand will contain the three alterations. What happens when the bacteria replicates this?

The answer is that two types of products will appear. First, the replicative products of the inner strand:



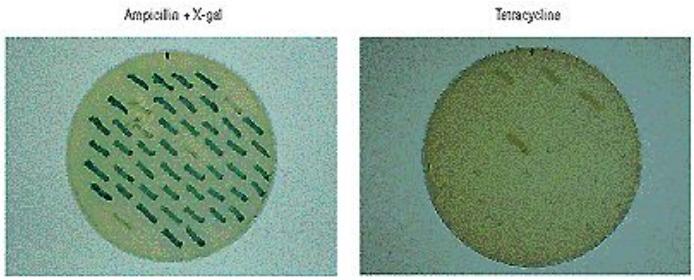
These will carry an intact tetracycline resistance gene and a nonfunctional ampicillin resistance gene. The cells that inherit these plasmids will die in ampicillin. On the other hand, the products of the outer strand:



These will carry a functional ampicillin resistance gene, a nonfunctional tetracycline resistance gene, and more importantly, the G->A mutation in the DNA insert.

So, by transforming the synthetic product into E. coli and

growing on ampicillin, we favor the mutated strand. Here's an example from the Promega site, showing the repair of the beta galactosidase gene



http://www.promega.com/pnotes/46/2259a/2259a.html

As you see, when the bacterial streaks are plated in ampicillin and the colorimetric substrate X-gal (panel on left), you get most showing blue color indicating repair of the gene. This is an indication of good concordance between ampicillin resistance and beta galactosidase gene mutation (repair). You also see that very few are tetracycline resistant (panel on right).

Now why should we want to debilitate the tetracycline gene? So that we can use the method to make additional changes, and while we're doing that we will repair the tetracycline gene and debilitate the ampicillin gene. That is, we can make a sequence of changes in our insertion, toggling between ampicillin resistance and tetracycline resistance.

A piece of cake!

A couple of interesting things to read:

Site-directed Mutagenesis using PCR, Michael P. Weiner, Tim Gackstetter, Gina L. Costa, John C. Bauer, and Keith A. Kretz (From: *Molecular Biology: Current Innovations and Future Trends.* Eds. A.M. Griffin and H.G.Griffin. ISBN 1-898486-01-8)

The QuikChange method: Student report, technical report, and application

Having problems getting whiter clothes? DIRECTED MUTAGENESIS AND PROTEIN ENGINEERING. Protein engineering of laundry detergent

Stanley Maloy's site: dut ung, and Altered States

Have no social life? Exciting researchers want to meet you and talk about their mutagenesis problems.

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