Biomolecular chemistry

7. Molecular Biology

Required reading: Sections 8.1 to 8.5 of Mikkelsen and Cortón, Bioanalytical Chemistry

Source Material
- Molecular Cell Biology 4th ed.: Lodish, Harvey; Berk, Arnold; Zipursky, S. Lawrence; Matsudaira, Paul; Baltimore, David; Darnell, James E. (NCBI bookshelf).
- Many figures and the descriptions for the figures are from the educational resources provided at the Protein Data Bank (http://www.pdb.org/)
- Most of these figures and accompanying legends have been written by David S. Goodsell of the Scripps Research Institute and are being used with permission. I highly recommend browsing the Molecule of the Month series at the PDB (http://www.pdb.org/pdb/101/motm_archive.do)

Some objectives for this section
- You will know what the 3 primary tools of molecular biology are
- You will have a thorough understanding of what PCR is
- You will have some appreciation of factors to consider when running a PCR reaction
- You will have an overview of how PCR is used in the manipulation and modification of DNA
- You will be familiar with several methods for site-directed mutagenesis of DNA
- You will know how the Quikchange protocol works
- You will understand how these techniques can be combined to achieve such things as the cloning of a ScFv or the generation of an RNA aptamer
The term ‘molecular biology’ is generally used to refer specifically to the practice of manipulating and modifying DNA.

You might think that the focus on DNA is a bit too specific for a term as all-encompassing as ‘molecular biology’. After all, shouldn’t the modification of RNA and proteins (and lipids, carbohydrates, and metabolites for that matter) be equally deserving of being called molecular biology?

I suppose that the justification for this specific use of term is that manipulation of all of these other groups of biomolecules, in the context of live cells, can only (or at least most easily) be achieved through the modification of DNA.

Fortunately, the techniques of molecular biology are very few in number. It is a classic case of ‘A minute to learn, a lifetime to master’.
Restriction endonucleases (also just called restriction enzymes), recognize specific base sequences in double-helical DNA and cleave, at specific places, both strands of a duplex containing the recognized sequences.

The booming field of biotechnology was made possible with the discovery of restriction enzymes by Werner Arber and Hamilton Smith in the early 1950's. Daniel Nathans pioneered their use for cutting DNA at specific sequences during the late 1960s.

To biochemists, these exquisitely precise scalpels are marvellous gifts of nature. They are indispensable for analyzing chromosome structure, sequencing very long DNA molecules, isolating genes, and creating new DNA molecules that can be cloned.

Do you recall any instance in our introduction to DNA replication, transcription, and translation where DNA was cut? No? Good. There isn’t any.

Restriction enzymes are found in a wide variety of prokaryotes. Their biological role is to cleave foreign DNA molecules. The cell's own DNA is not degraded, because the sites recognized by its own restriction enzymes are methylated. It follows that a bacteria must have a pair of enzymes, a restriction enzyme to cleave foreign DNA and a methyltransferase to protect its own DNA.
Restriction enzymes generally cleave palindromic sequences

Many restriction enzymes recognize specific sequences of four to eight base pairs and hydrolyze a phosphodiester bond in each strand in this region. A striking characteristic of these cleavage sites is that they almost always possess twofold rotational symmetry. In other words, the recognized sequence is palindromic (def. A word, sentence, or verse that reads the same from right to left as it does from left to right).

This symmetry reflects that of structures of the restriction enzymes themselves. The enzymes are generally homodimers. Some restriction enzymes recognize asymmetric sequences because they are heterodimers. But this still does not answer the more fundamental question of why these enzymes are homodimers that cut palindromic sequences? Can you rationalize why evolution might have selected for this situation, as opposed to monomeric or heterodimeric enzymes, or enzymes that cut non-palindromic sequences?

Hundreds of restriction enzymes have been purified and characterized. Their names consist of a three-letter abbreviation for the host organism (e.g., Eco for Escherichia coli, Hin for Haemophilus influenzae, Hae for Haemophilus aegyptius) followed by a strain designation (if needed) and a roman numeral. The specificities of several of these enzymes are shown above. Note that the cuts may be staggered or even.

Restriction enzymes are used to cleave DNA molecules into specific fragments that are more readily analyzed and manipulated than the entire parent molecule. Indeed, complex chromosomes containing hundreds of millions of base pairs can be mapped by using a series of restriction enzymes.

Question: I do not understand how AgeI and Xma I A/CCGGT and C/CCGGG, can form complimentary sticky ends.

Answer: Try drawing out the ends that you would get after digestion and it will make more sense. The A (of the Age1 substrate) and the C (of the Xma1 substrate) are part of the substrate recognition sequence of, but they do not form part of the stick ends. The sticky ends formed by these two enzymes are actually identical. They do differ in the first residue of the double helix, but this is not part of the sticky end. Two pieces of DNA, one cut with Age1 and one cut with Xma1, could be ligated together. BUT, the product of the ligation could not be cut by either enzyme. In contrast, the product of a ligation of two pieces of DNA cut by the same enzyme could always be cut again by the same enzyme (since the exact substrate has been remade).
Restriction Enzymes catalyze a double hydrolysis of the backbone of DNA

When the restriction enzyme cuts the DNA, it leaves overhanging chains: these are termed "sticky ends" because the base pairs formed between the two overhanging portions will glue the two pieces together, even though the backbone is cut. Sticky ends are an essential part of genetic engineering, allowing researchers to cut out little pieces of DNA and place them in specific places, where the sticky ends match.

The PDB contains structures for many restriction enzymes. Another example from Escherichia coli (specifically EcoRV) is shown here. The structure at the top, taken from PDB entry 1rva, shows the enzyme bound to a short piece of DNA. The arrow shows the phosphate group that will be cut. The lower illustration, taken from PDB entry 1rvc, shows the structure after the DNA has been cut. A water molecule has been inserted, so there are now two oxygen atoms, close to one another but not bonded together, where there was a single bonded oxygen atom in the intact DNA. In both illustrations, the protein is shown with a simple backbone representation and one DNA strand is colored green.

Question: When talking about the restriction enzyme, Why the evolution select it to be homodimers rather than heterodimers?

Answer: To cut through both strands of DNA, a restriction enzyme must have two active sites. The way to most efficiently accomplish this is to have a homodimeric restriction enzyme. Other, less efficient ways would be to have one larger enzyme with two active sites or two different genes that encode for the two proteins of a heterodimer. These way are less efficient since they require the maintenance of larger sections of DNA, the formation of longer (or more) mRNA molecules, the formation of longer (or more) proteins. All of this requires the cell to use more energy that it could be using for other purposes.
The NEB catalog is the definitive resource for all things to do with restriction enzymes. A catalog can be requested through their website.

- This catalog provides details of all the enzymes they sell; which is a very large fraction of all known activities.
- For each enzyme, a variety of details are provided:
  - Reaction conditions: the buffer supplied with the enzyme. There are 4 standard buffers plus a handful of special buffers for very popular or very fussy enzymes.
  - Ligation and recutting: a standard test for the fidelity of the enzyme
  - Concentration: Standard concentration of 5-20 units/μl. Unit definition: 1 unit of a restriction enzyme will completely digest 1 μg of substrate DNA in a 50 μl volume reaction in 60 minutes. Typically add 10x extra enzyme to ensure complete reaction.
  - Storage: stored at -20 ºC in cryoprotectant so solution does not freeze.

- Heat inactivation: some enzyme are from thermophiles and thus may not denature at high temperature. This would generally be used to stop the reaction.

- Question: I have found that you need some additional DNA sequence between the restriction site and the end of the PCR product, but I am not sure how to come up with these sequences
  - Answer: I didn't mention it in class, but restriction enzymes often have less activity when they are cutting substrates at the very end of a double stranded DNA. These are enzymes that evolved to cut the middle of very long pieces of DNA, not at the end of DNA sequences. That is, they are endonucleases not exonucleases. So it is often necessary to add some extra bases to the primers to trick the enzyme into thinking that is is cutting in the middle of a long piece of DNA and not near an end. Here is a table with further information: http://www.neb.com/nebecomm/tech_reference/restriction_enzymes/cleavage_linearized_vector.asp. The exact sequence of the extra bases that are added doesn't matter too much, but as a general rule people like to include more Gs and Cs than A's and T's in order to make them more stable. The one thing to watch out for is that you are not extending the palindrome region of the restriction site. One thing that might make a primer 'not work' is that it forms a strong self dimer (another bad thing is stable hairpin). To be a self dimer it must have a substantial palindromic region. Now, every primer that contains a restriction site must have at least a 6 base palindrome that is, of course, a potential site of self-dimerization. However, it won't have a high enough melting temperature to cause problems during PCR. That is, as long as you don't accidentally make it into a longer palindrome that will have a higher melting temperature and actually cause some problems.
As we have seen earlier in the course, polynucleotide ligases are ubiquitous proteins that are required for a number of important cellular processes, including the replication, repair and recombination of DNA.

DNA ligases catalyse the formation of 3'-phosphodiester bonds at single-strand breaks between adjacent 3'-hydroxyl and 5'-phosphate termini in double-stranded DNA.

DNA ligases have also found widespread use as a tool for in vitro DNA manipulation and cloning techniques.

DNA ligases can be divided into two broad classes, those requiring nicotinamide adenine dinucleotide (NAD+) as a substrate (structure shown on next slide) and those requiring ATP. The eukaryotic, viral and archaeabacteria encoded enzymes all require ATP. NAD+-requiring DNA ligases are found exclusively in eubacteria.

The ATP-dependent ligases range in size from 30 to >100 kDa. The NAD+-dependent enzymes are highly homologous and are monomeric proteins of 70–80 kDa. The results of structural studies revealed that the ATP- and NAD+-utilizing enzymes are homologous even though this homology could not be deduced from their amino acid sequences alone.

Note that, for molecular biology purposes, we use ligases to join double stranded DNA fragments together. Most ligases in nature are used to fix single strand ‘nicks’ in the context of double stranded DNA (think back to Okazaki fragments, for example). There are some naturally occurring ligases that are used to repair double stranded breaks in DNA, however to the best of my knowledge, these are not used in molecular biology applications.

**Question:** You mentioned that we must first treat the plasmid with alkaline phosphatase to remove the phosphate group. I don't understand why should we do this? Don't we need the 5'-P for the ligation?

**Answer:** This is not something that must be done, but it is something that can be done be done. By treating a cut plasmid with alkaline phosphatase you can prevent it from ligating to itself. This is useful for preventing plasmid from re-circularizing without including the fragment that you want to insert. This is particularly a problem when you are using a single restriction site. It would be less of a problem when two restriction sites are used, since the sticky ends are not complementary. If the fragment to be inserted still has it's 5' phosphate, you don't need a 5' phosphate on the plasmid.
Restriction enzymes plus ligases allow us to manipulate genes and introduce them into bacteria in the form of plasmids.

- Plasmids are circular, double-stranded DNA (dsDNA) molecules that are separate from a cell's chromosomal DNA. These extra-chromosomal DNAs, which occur naturally in bacteria, yeast, and some higher eukaryotic cells, exist in a parasitic or symbiotic relationship with their host cell. Plasmids range in size from a few thousand base pairs to more than 100 kilobases (kb). Like the host-cell chromosomal DNA, plasmid DNA is duplicated before every cell division. During cell division, at least one copy of the plasmid DNA is segregated to each daughter cell, assuring continued propagation of the plasmid through successive generations of the host cell.
- The plasmids most commonly used in recombinant DNA technology replicate in E. coli. Generally, these plasmids have been engineered to optimize their use as vectors in DNA cloning. For instance, to simplify working with plasmids, their length is reduced; many plasmid vectors are only ~3-4 kb in length, which is much shorter than in naturally occurring E. coli plasmids. (The circumference of plasmids usually is referred to as their “length,” even though plasmids are almost always circular DNA molecules.)
- Most plasmid vectors contain little more than the essential nucleotide sequences required for their use in DNA cloning: a replication origin, a drug-resistance gene, and a region (poly linker) in which exogenous DNA fragments can be inserted.
- If there is a promoter and a ribosome binding site at the 5' end of the poly linker, the plasmid can be used to express the gene and make the corresponding protein. Researchers typically use plasmids that are inducible. That is, they can be ‘turned on’ by addition of a small molecule. The classic example is the IPTG-inducible lac promoter (part of the lac operon).
- Note that inserting a DNA fragment encoding a gene could be done using either one or two restriction sites. If one is used, the DNA fragment could insert in two possible directions. If two are used, the orientation of the inserted DNA is fixed by the experimental design. If the plasmid will be used for gene expression, two restriction sites are obviously preferred.
- When restriction enzymes are used in pairs to cut out a certain chunk of DNA, the reaction conditions must be compatible with both enzymes.
In 1984, Kary Mullis devised an ingenious method called the polymerase chain reaction (PCR) for amplifying specific DNA sequences. Consider a DNA duplex consisting of a target sequence surrounded by nontarget DNA. Millions of the target sequences can be readily obtained by PCR if the flanking sequences of the target are known. PCR is carried out by adding the following components to a solution containing the target sequence:
- a pair of primers that hybridize with the flanking sequences of the target,
- all four deoxyribonucleoside triphosphates (dNTPs), and
- a heat-stable DNA polymerase.

The solution is then ‘thermocycled’ between the temperatures of ~50 °C, ~70 °C, and ~95 °C approximately 20 to 30 times.

This process leads to an exponential amplification of the target DNA sequence of interest.

The Nobel Prize in Chemistry 1993 was awarded "for contributions to the developments of methods within DNA-based chemistry". The two recipients were Kary Mullis "for his invention of the polymerase chain reaction (PCR) method" and Michael Smith of UBC "for his fundamental contributions to the establishment of oligonucleotide-based, site-directed mutagenesis and its development for protein studies". http://nobelprize.org/nobel_prizes/chemistry/laureates/1993/
Polymerase Chain Reaction (PCR)

- If the picture on the left is confusing, it's because it is a frame from a movie. I highly recommend watching the whole movie in order to better understand PCR: [http://www.chem.ualberta.ca/~campbell/teaching_private/Chem-419-511-519/PCR.mov](http://www.chem.ualberta.ca/~campbell/teaching_private/Chem-419-511-519/PCR.mov)

- A PCR cycle consists of three steps:
  - Strand separation. The two strands of the parent DNA molecule are separated by heating the solution to 95 °C for 15 s.
  - Hybridization of primers. The solution is then abruptly cooled to 54 °C to allow each primer to hybridize to a DNA strand. One primer hybridizes to the 3'-end of the target on one strand, and the other primer hybridizes to the 3'-end on the complementary target strand. Parent DNA duplexes do not form, because the primers are present in large excess. Primers are typically from 20 to 30 nucleotides long.
  - DNA synthesis. The solution is then heated to 72 °C, the optimal temperature for Taq DNA polymerase. This heat-stable polymerase comes from Thermus aquaticus, a thermophilic bacterium that lives in hot springs. The polymerase elongates both primers in the direction of the target sequence because DNA synthesis is in the 5'-to-3'-direction. DNA synthesis takes place on both strands. In the third cycle of PCR, the first copy of the target sequence with no flanking sequences is generated.

- Repetition of these 3 steps leads for ~30 times leads to the production of ~ 1 billion copies of the target sequence without flanking sequence.

- There has been a long and bitter battle over the patent for Taq polymerase. In a nutshell: PCR was invented by Kary Mullis at Cetus Pharmaceuticals, and a patent was granted in 1989. In 1992 Hoffman-La Roche purchased the rights to the use of Taq polymerase for PCR from Cetus, including the licenses. Cetus had granted a license to the Promega Corporation in 1990 to sell Taq polymerase, the critical component in PCR, for purposes other than PCR. In 1992 Hoffman-La Roche sued Promega for breach of license agreement. Maintaining that there are other applications of Taq polymerase and that end users are responsible for ensuring that they are licensed to use Taq polymerase for PCR, Promega denied the breach of contract claim and countered by challenging the validity of the Taq patent (from [http://biotech.about.com/library/weekly/aa083099.htm](http://biotech.about.com/library/weekly/aa083099.htm)). This dispute was finally settled in 2005 which, not coincidently, is the year the patent expired.
PCR frees us from the tyranny of restriction sites

Three approaches to ligating PCR products into plasmids:
- Complementary restriction sites
- Blunt ends
- TA overhangs

• PCR is a very handy method of adding restriction sites to the ends of a gene for cloning into a plasmid vector. The particular restriction site can be designed into the primers used to amplify the gene sequence of interest. Recall that every copy of the PCR product contains the primer cDNA.
• The sections of DNA added to the ends during PCR amplification can have arbitrary sequence and length. These added bits might be up to 50 or so bases and may or may not contain a restriction site. Note that in the final PCR product, there will be a complementary strand to these overhangs.
• Using two different restriction sites allows for directional cloning into a plasmid vector.
• If a blunt ended PCR fragment is generated, it can be inserted into a blunt-end cut vector but the efficiency is much lower and there is no directionality.
• If Taq is used as the polymerase, the PCR product will have 3’ A overhangs. This PCR product can be directly ligated into an appropriate vector with 3’ T overhangs. The approach is also not directional.
• From a very practical standpoint, there is always a chance that a PCR product will not be exactly what you think it is due to random errors. To identify an error-free PCR product, it is often necessary to sequence several distinct clones. The higher the error rate, the more clones that will need to be sequenced to find one that is error free. With a 1 kb or less PCR product, you can be fairly certain that the product is error free if you used anything but Taq. If Taq is used, you should submit two separate clones for sequencing to give a high probability of one being error free. With larger PCR products Taq should be avoided because the cost of sequencing the whole thing is increasing along with the probability of there being an error. A typical DNA sequencing reaction might give 500 to 1000 bases.
• Enzymes with much lower error rates than Taq (e.g. Pfu from Stratagene) are also available.

• Question: Why can a linear DNA molecule not be used to transform bacteria?
  • Answer: I am not sure about the exact mechanism involved, but linear DNA is generally not effective for transforming E. coli. I suspect it is recognized as foreign (identified as potentially being phage) and so the cell does not methylate it in order to protect it from the endonucleases. Another reason it wouldn’t be effective is because it would not be able to replicate itself since the origin of replication would now be in the middle of a linear piece of DNA and so the full circle would not be synthesized by DNA polymerase. I suspect that DNA digested with a single restriction enzyme that leaves sticky ends might be used for transformation in some cases since the ends have at least a slim chance of being ligated together in the cell before the DNA is destroyed by the bacteria.
We use the techniques of Molecular Biology to manipulate DNA - however we are typically doing this in order to change the structure of a protein. For this and other examples that follow, it is important to think about the genetic changes in terms of how they would change the protein structure.

Question: I don't quite understand what the forward and backward primers are. Would you please give me a brief description?
Answer: A primer is relatively short piece of RNA or DNA that is complementary to a longer piece of DNA known as the template. It serves as the starting point for DNA polymerase to continue synthesis of a new strand in a 5’ to 3’ direction. We saw RNA templates at the start of the course when we were discussing how DNA replication happens in cells. In the lab we can copy DNA is a similar way but instead of using RNA primers we use synthetic DNA primers. For PCR to work we need to copy both strands of the template. This means that we need one 'forward' primer for making the sense strand and one 'reverse' primer for making the antisense strand. The forward primer uses the antisense strand as it's template while the reverse primer uses the sense strand as its template. Primers are designed such that the part that actually binds to the template in the first round has a melting temperature in the range of ~50 degrees C. You can add extra sequence to a primer (such as restriction sites), but since this isn't present in the template it doesn't count as being 'complementary sequence'. That is, you shouldn't count it when calculating the melting temperature. Though after the first few rounds of PCR effectively all of the template is the product of earlier cycles and the full length of the primer plus any added sequence will be complementary to these newly synthesized template molecules.

Question: I don't completely understand what the overlap region of the primers is exactly. Is it not just the segment between the two primers that is getting amplified? Also, when you say that the primer should be in the frame of the N terminal His tag, you mean of the plasmid yes?
Answer: The overlap region is the region of complementarity between the primer and the template. That is, it is the sequence of 16 or more nucleotides where the primer actually anneals onto the template. For the homework problem, you are ligating the digested PCR product into a plasmid that contains an N-terminal His-tag sequence. The reading frame of the inserted gene must be consistent with the reading frame of the His-tag. Another way of thinking of looking at this is that the start codon is already in the plasmid. The N-terminal methionine of the gene to be inserted is no longer a start codon, it is just a plain methionine (but it still needs to be in the correct reading frame of course).
Molecular Biology applications: deletion of a section of DNA

1. Do two separate PCRs:

2. Digest and ligate

- Can also use blunt ends if primers are 5'-phosphorylated.
Molecular Biology applications:

site-directed mutagenesis

- This one is a no-brainer for why you would want to do it to proteins. Site directed mutagenesis allows us to change one amino acid into any other amino acid. This is commonly used to investigate the role of specific residues in the active site of an enzyme. It can also be used, for example, to introduce a cysteine into a protein which could serve as a reactive site for attaching other molecules such as fluorescent dyes or perhaps biotin (more on this later).
Molecular Biology applications: random mutagenesis

DNA to be randomly mutated

Do one PCR:

Use Taq with some Mn$^{2+}$ instead of Mg$^{2+}$ and low concentrations of one or more dNTPs

• This one might be a bit more mysterious at first glance. Why would you want to randomly mutate a protein? The answer is that this is how you can generate diversity for molecular evolution in the laboratory. If you are interested in changing or ‘improving’ the function of a protein, you can generate libraries of randomly mutated variants. The tricky part is coming up with a way of screening or selecting the improved variants from the library. But that is another story...
Quikchange: the most popular method of site-directed mutagenesis

- Why not use Taq?
- Is it exponential or linear amplification during thermocycling?
- Why do you need a kit?

http://www.stratagene.com/

- The next slide shows this process is somewhat more detail.
Quikchange is not PCR

- Strands a and b are the original complementary strands of the double stranded plasmid DNA
- Heating them at 95 °C melts the DNA and gives the individual single strands
- Cooling to 55 °C allows the mutagenic primers that are mostly complementary to template plasmid, with the exception of the designed mutationed
- Heating to 65 °C allows DNA polymerase to extend the DNA to complete the full circle
- This cycle can be repeated many times. With each cycle a new copy of the template is created. The newly synthesized strands (in red) can not serve as templates for subsequent cycles. This is why this is not a PCR reaction. The amplification is linear, not exponential.
- Finally Dpn1 is added to the mixture and all of the DNA that contains at least one strand derived from the original template will be digested and effectively destroyed.

E. Coli will fix the 'nicks' in this double stranded DNA

\[ \text{X} = \text{digested by Dpn1} \]
How does Quikchange work? *E. coli* has enzymes that methylate DNA

*dam* = gene for adenine methyltransferase

In prokaryotes, MTases have most often been identified as elements of restriction/modification systems that act to protect host DNA from cleavage by the corresponding restriction endonuclease. Most laboratory strains of *E. coli* contain three site-specific DNA methylases.

- Dam methylase—methylation at the N6 position of the adenine in the sequence GATC.
- Dcm methylase—methylation at the C5 position of cytosine in the sequences CCAGG and CCTGG.
- EcoKI methylase—methylation of adenine in the sequences AAC(N6A)GTGC and GCAC(N6A)GTT.

Dpn1 is a restriction endonuclease that only only cuts at GATC when the A is methylated. This is different than most of the common restriction endonucleases that have their activity blocked by methylation.

Question: After purifying a plasmid from bacteria (which was transformed with this plasmid), it's still possible to cut this plasmid using REs. Would this plasmid DNA be methylated by the bacteria? Are special bacterial strains required for this molecular biology work, which don't have the enzymes required to methylate DNA?

Answer: All DNA isolated from bacteria is methylated to some extent. If the methylation site is within the substrate recognition site for a particular enzyme, it may or may not be blocked (http://www.neb.com/nebecomm/tech_reference/restriction_enzymes/dam_dcm_cpg_methylation.asp). This really depends on the specifics of how the enzyme interacts with the substrate. That is, if the enzyme makes an important interaction with the position where the methyl group is attached, the binding may be disrupted by the presence of the methyl group. If the enzyme does not interact with that particular position, the enzyme may be indifferent to the presence of the methyl group. The majority of enzymes that are used routinely are indifferent to the presence of the methylation. These means that they can cut DNA whether it comes from bacteria or PCR. You can get special strains of *E. coli* without the methylation enzymes (knockouts), but this is just for highly specialized applications. Dpn1 is interesting in that it actually requires the methylation to be present in order to cut (http://www.neb.com/nebecomm/products/productR0176.asp).

from the NEB technical resource
Variations on the Quikchange concept:

- Point mutations
- Small insertions
- Small deletions

The key concept that connects these strategies is that the whole plasmid is amplified.

Exponential amplification only occurs when the product of the extension reaction can itself serve as template. It turns out that any case (to the best of my knowledge) in which you produce overlapping stick ends (and thus a nick), the product can not serve as a template and so the amplification is linear.

Question: Would you please tell how we can find out if the product of amplification of a plasmid has "nick" or not? Would be linear or exponential?

Answer: A plasmid can not be PCR amplified to make intact and covalently linked circular double stranded DNA. It will either be linear or non-covalently circularized due to hydrogen bonds between overlapping sticky ends (as in the Quikchange rxn). The product of the Quickchange reaction is missing one phosphodiester bond missing in each strand of the plasmid. The location of this missing bond is the nick.
Applications of molecular biology: addition of affinity tags for protein purification

- In recent years it has become increasingly unnecessary to ever have to purify a wild-type protein from a natural source. If you know the DNA sequence of the protein that you are interested in, it is much more efficient to amplify the gene from a cDNA library and attempt to express a recombinant hybrid gene in E. coli bacteria.

- There are many plasmids available for protein expression where the start codon is already part of the plasmid. It is then followed by a peptide or protein sequence that is particularly useful for affinity purification (the affinity tag). The affinity tag is then followed by a sequence that encodes the substrate for a specific protease and then, finally, comes the polylinker (with the restriction sites) which is where the gene encoding your protein of interest would be inserted.

- The advantage is that the hybrid gene can be expressed with this polypeptide fusion partner, termed an affinity tag, to facilitate the purification of the target polypeptide. Many different proteins, domains, or peptides can be fused with the target protein. The advantages of using fusion proteins to facilitate purification and detection of recombinant proteins are numerous.

- The only time that you might need to purify a wild-type protein from a natural source is when you are going on a fishing expedition. That is, you have an assay for a function but you don’t know which protein is responsible for the activity. In this case the goal is to purify the protein of interest to homogeneity so you can then figure out what it is (probably by mass spectrometry).
It can sometimes be difficult to choose the right purification system for a specific protein of interest. The most frequently used and interesting systems include: Arg-tag, calmodulin-binding peptide, cellulose-binding domain, DsbA, c-myc-tag, glutathione S-transferase, FLAGtag, HAT-tag, His-tag, maltose-binding protein, NusA, Stag, SBP-tag, Strep-tag, and thioredoxin.

- His-tag is the single most popular. This is followed by glutathione S-transferase and maltose binding protein.

- Some typical protease sites might include TEV protease (ENLYFQG), Enterokinase (DDDDK), or others.

- For purification of antibodies, researchers generally use a type of affinity chromatography where the matrix is either protein A or protein G. These are bacterial proteins that are bind with relatively high affinity and specificity to the Fc fragment of most antibodies.
Applications of molecular biology: creating an ScFv

- Three scenarios:
  - starting with hybridoma expressing monoclonal antibody
  - starting with a population of B-cells
  - starting with synthetic gene diversity

- Note that Sfi1 is an interesting restriction endonuclease that has the following recognition site: 5' GGCCNNNN/NGGCC 3' where N = any nucleotide. What might be an advantage of using such an enzyme?


- Question: What is the advantage of using restriction enzyme that can recognize site containing unknown nucleotides such as 5' GGCCNNNN/NGGCC 3'. Thank you.
- Answer: The advantage is that you can design the sticky end to be whatever you want. In the phage display example, there is an Sfi1 site at each end of the gene, but the exact sequence of nucleotides at the 'N' positions will be different. After it is cut, you will have different sticky ends at each end of the gene.
Applications of molecular biology: systematic evolution of ligands by exponential enrichment (SELEX)

DNA Aptamer

RNA Aptamer

- RNA or DNA molecules engineered to have specific binding functions to molecules other than DNA and RNA are known as aptamers.

- Figure legend from paper cited above: “In vitro selection of high-affinity aptamers for in vivo applications. DNA and RNA aptamers are selected from a randomized oligonucleotide library by reiterative SELEX rounds. In case of the DNA aptamer, the double-stranded DNA pool needs to be denatured for purification of the single-stranded sense strand, which then is presented to its selection target. Following removal of unbound and low-affinity bound DNA molecules, target-bound DNA molecules are eluted and amplified by PCR in the presence of a biotinylated primer. Following denaturation, the unbiotinylated sense strand can be purified by polyacrylamide gel electrophoresis and used for the next SELEX cycle. In the case of development of an RNA aptamer, the double-stranded DNA pool is in vitro transcribed to the RNA pool which is then used for SELEX. Eluted target binders are reversed transcribed to cDNA, amplified by PCR, and again in vitro transcribed to give the RNA pool used in the next SELEX cycle. RNA and DNA molecules can be protected against nuclease attacks during the selection process by incorporation of modified nucleotides using enzymatic reactions or can be chemically modified after the SELEX process; ss = single-stranded; ds = double-stranded; NTP = nucleoside-triphosphate.”

- See also:
RNA aptamers are highly folded structures with distinct binding pockets for their ligands.

- For some applications, aptamers are now considered as a viable alternative to antibodies.
- Advantages of aptamers relative to antibodies
  - relatively inexpensive to generate
  - can be rapidly discovered in vitro
  - larger libraries can be screened
  - smaller molecular size
- Disadvantages of aptamers relative to antibodies
  - generally can not achieve as high of binding affinity, though there are many examples of high affinity aptamers as well
  - susceptible to hydrolysis (RNA aptamers)
  - smaller repertoire of functional groups

- Note that Pegaptanib is an FDA-approved pegylated anti-VEGF aptamer for wet macular degeneration. Vascular endothelial growth factor (VEGF) is the hormone that cells release in order to stimulate the growth of new blood vessels towards them. Bevacizumab (the FDA-approved antibody also known as Avastin) fights cancer (and macular degeneration) by the same mechanism.
Summary of Molecular Biology

• The basic tools of modern molecular biology are DNA polymerases, restriction enzymes, and ligases. With these 3 tools, a vast array of manipulations are possible.
• The polymerase chain reaction allows researchers to copy DNA with an exponential increase in the number of copies.
• Starting from one molecule of DNA, 30 cycles of PCR (a couple of hours) will result in millions of copies.
• Molecular biology techniques can be used to insert, delete, cut and paste, randomly mutate, and site-specifically mutate DNA.
• The standard protocol for site-directed mutagenesis is commercially available as the Quikchange kit from Stratagene.
• Variations on the Quikchange protocol are very useful for rapidly modifying DNA.
• These same tools allow for genes to be ‘cloned’ and expressed in other cell types.
• PCR can be used to discover aptamers using the SELEX procedure.