Biomolecular chemistry

6. Molecular Biology

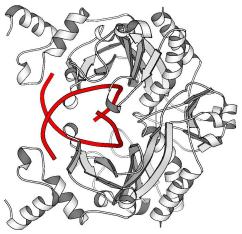
Suggested reading: **Sections 8.1 to 8.5** of
Mikkelsen and Cortón, *Bioanalytical Chemistry*

Source Material

- Biochemistry: Berg, Jeremy M.; Tymoczko, John L.; and Stryer, Lubert (NCBI bookshelf).
- Molecular Cell Biology 4th ed.: Lodish, Harvey; Berk, Arnold; Zipursky, S. Lawrence; Matsudaira, Paul; Baltimore, David; Darnell, James E. (NCBI bookshelf).
- Introduction to Genetic Analysis Anthony: J.F. Griffiths, Jeffrey H. Miller, David T. Suzuki, Richard C. Lewontin, William M. Gelbart (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)

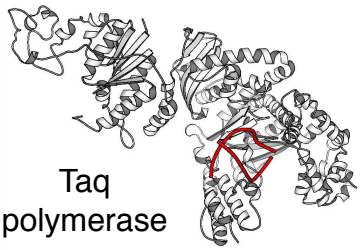
The 3 primary tools of molecular biology

Real-world tool metaphor?



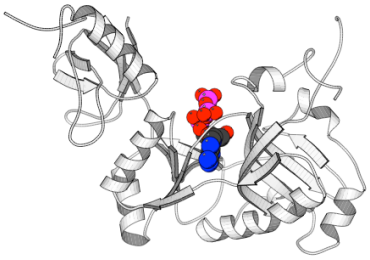
Restriction endonuclease

=



Taq polymerase

=

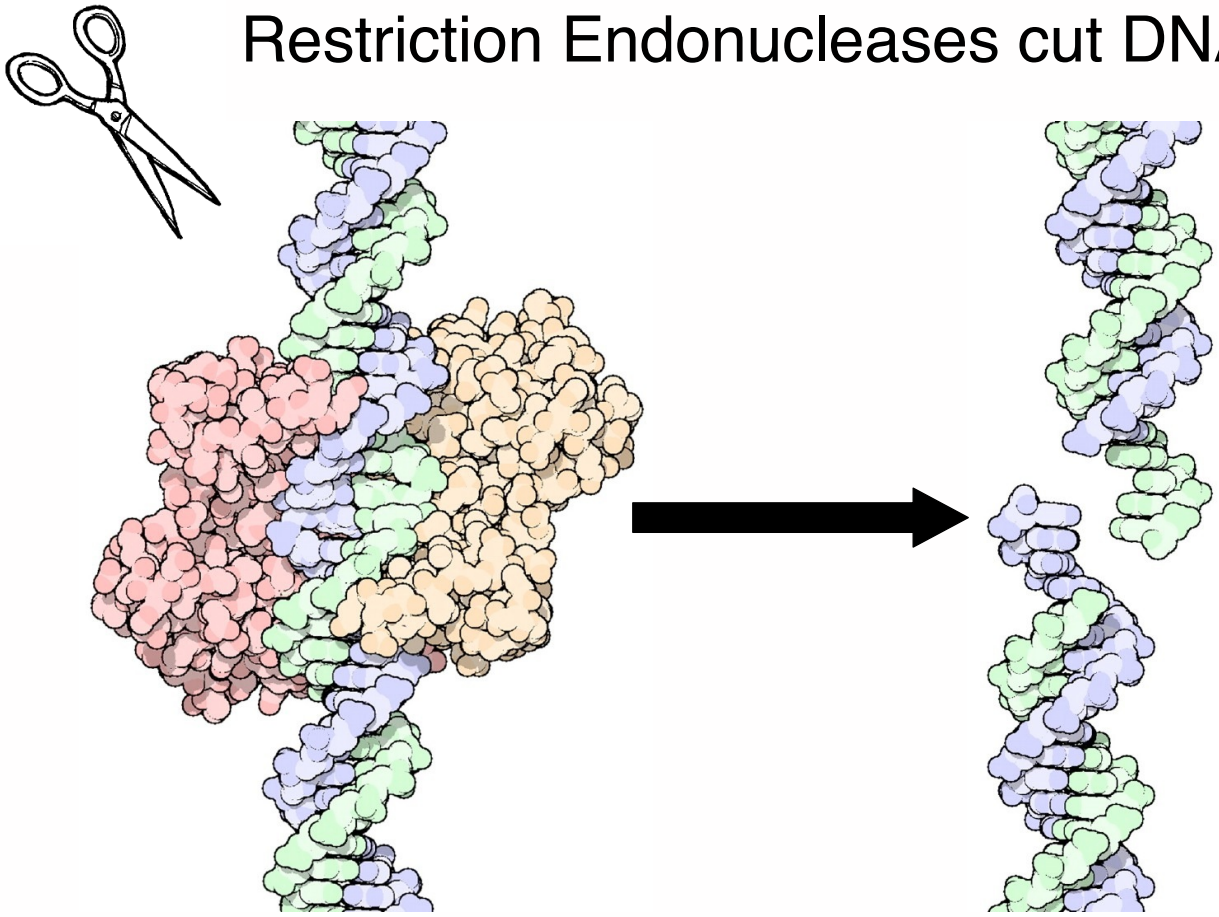


DNA ligase

=

- 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’.

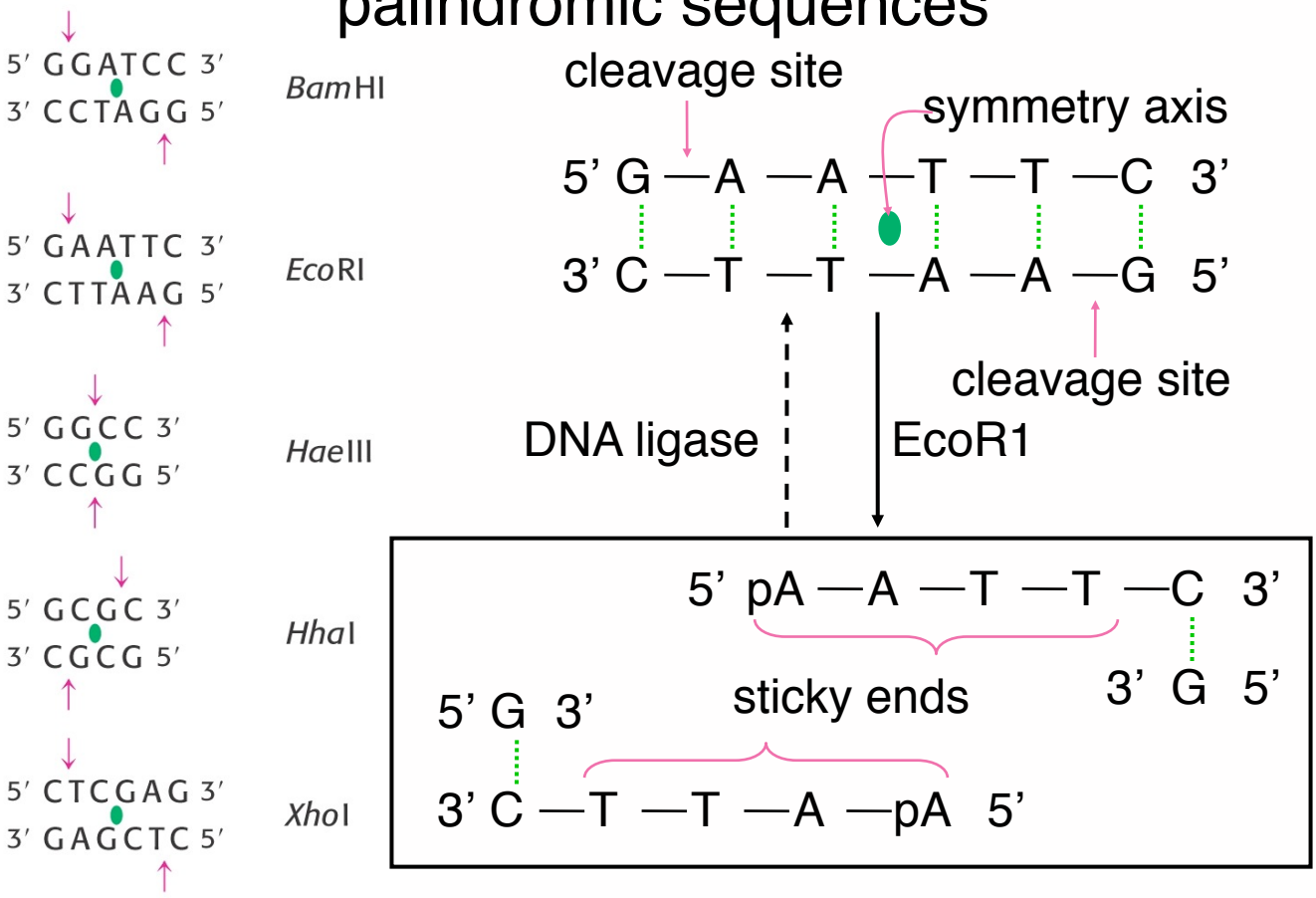
Restriction Endonucleases cut DNA



David S. Goodsell: The Molecule of the Month appearing at the PDB

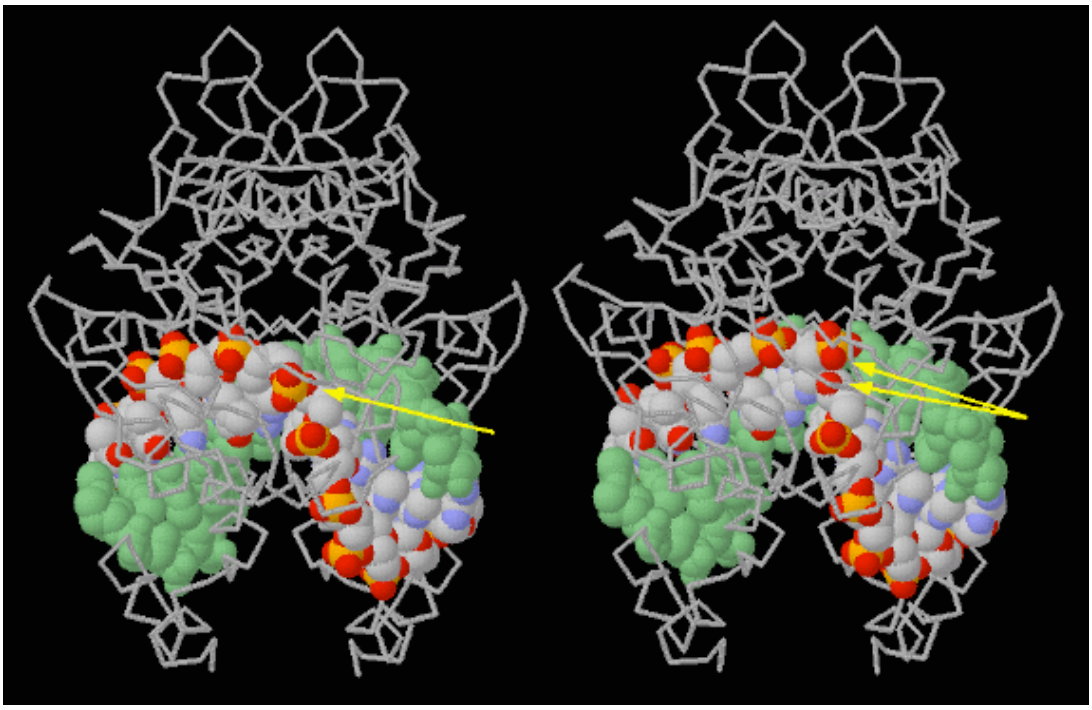
- Restriction endonucleases (also just called restriction enzymes), recognize specific base sequences in double-helical DNA and cleave, at specific places, both strands of a duplex containing the recognized sequences.
- The booming field of biotechnology was made possible with the discovery of restriction 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 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.
- As we will see soon enough, PCR allows us to add restriction sites to the end of the fragments, so we can add the restriction sites that are most convenient for inserting into the plasmid that we want to use. This means that we don't need to worry about finding specific restriction enzyme sites (the palindromes) in the DNA that we want to insert into he plasmid. The DNA produced by PCR amplification, with its designed restriction sites, still must be treated with restriction enzymes to produce the sticky ends for ligation. PCR gives researchers a lot of control and flexibility to manipulate 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.
- *Q. I do not understand how AgeI and Xma I A/CCGGT and C/CCGGG, can form complimentary sticky ends.*
- *A. 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).*
- *Q. Do we have to memorize the restriction sites of the popular restriction enzymes or will those be given in the exam?*
- *A. These will be provided.*

Restriction Enzymes catalyze a double hydrolysis of the backbone of DNA



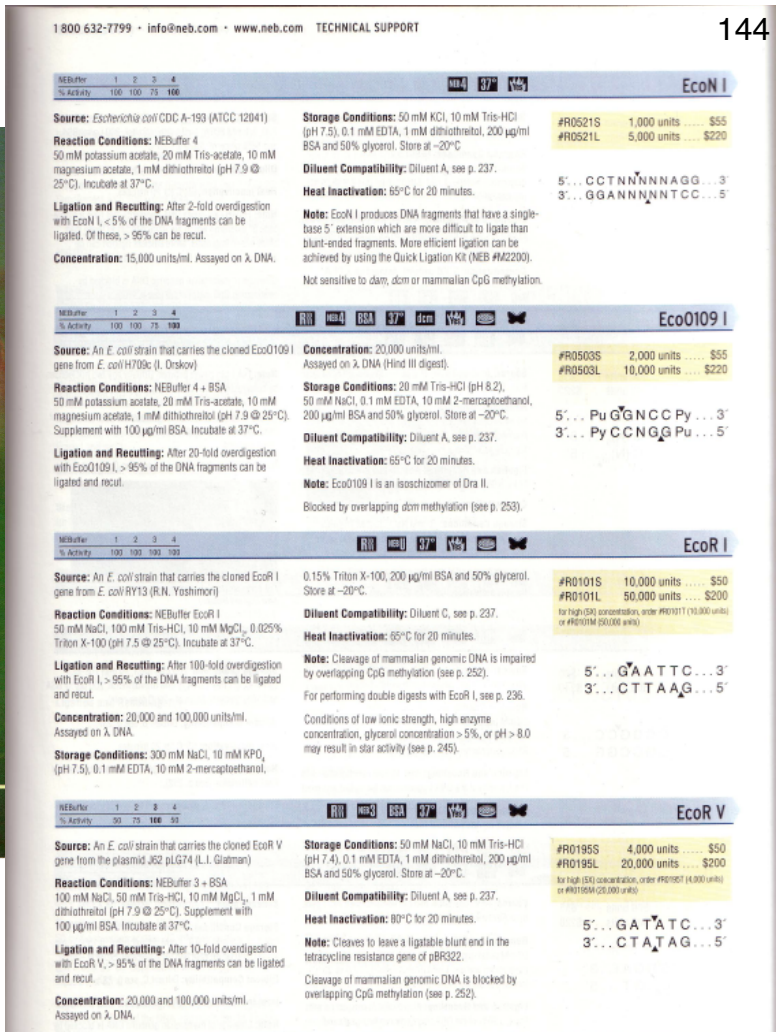
David S. Goodsell: The Molecule of the Month appearing at the PDB

- 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.
- *Q: When talking about the restriction enzyme, Why the evolution select it to be homodimers rather than heterodimers?*
- *A: 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 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.*
- *Q. Another question for the restriction enzymes to cut the DNA sequence, does it cut just the same specific sequence of it or it can also cut the complementary sequence to it?*
- *A. Recall that a restriction enzyme only cuts dsDNA and that the substrates are palindromes. Say an enzyme cuts this sequence:*

  GAATTC
  CTTAAG
  If I rotate it by 180 degrees I get:
  GAATTC
  CTTAAG
  It is of course exactly the same. So yes, an enzyme can always cut the complementary sequence, because it is exactly the same thing.

The NEB catalog

www.NEB.com



1 800 632-7799 · info@neb.com · www.neb.com TECHNICAL SUPPORT

EcoN I

Source: *Escherichia coli* CDC A-193 (ATCC 12041)

Reaction Conditions: NEBuffer 4
50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM dithiothreitol (pH 7.9 @ 25°C). Incubate at 37°C.

Ligation and Recutting: After 2-fold overdigestion with EcoN I, < 5% of the DNA fragments can be ligated. Of these, > 95% can be recut.

Concentration: 15,000 units/ml. Assayed on λ DNA.

Storage Conditions: 50 mM KCl, 10 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 1 mM dithiothreitol, 200 µg/ml BSA and 50% glycerol. Store at –20°C

Diluent Compatibility: Diluent A, see p. 237.

Heat Inactivation: 65°C for 20 minutes.

Note: EcoN I produces DNA fragments that have a single-base 5′ extension which are more difficult to ligate than blunt-ended fragments. More efficient ligation can be achieved by using the Quick Ligation Kit (NEB #M2200).

Not sensitive to *dam*, *dcm* or mammalian CpG methylation.

#R0521S 1,000 units $55
#R0521L 5,000 units $220

5′... CCTNNNNAGG... 3′
3′... GGANNNNTCC... 5′

Eco0109 I

Source: An *E. coli* strain that carries the cloned Eco0109 I gene from *E. coli* H709c (I. Orskov)

Reaction Conditions: NEBuffer 4 + BSA
50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM dithiothreitol (pH 7.9 @ 25°C). Supplement with 100 µg/ml BSA. Incubate at 37°C.

Ligation and Recutting: After 20-fold overdigestion with Eco0109 I, > 95% of the DNA fragments can be ligated and recut.

Concentration: 20,000 units/ml. Assayed on λ DNA (Hind III digest).

Storage Conditions: 20 mM Tris-HCl (pH 8.2), 50 mM NaCl, 0.1 mM EDTA, 10 mM 2-mercaptoethanol, 200 µg/ml BSA and 50% glycerol. Store at –20°C.

Diluent Compatibility: Diluent A, see p. 237.

Heat Inactivation: 65°C for 20 minutes.

Note: Eco0109 I is an isoschizomer of Dra II.

Blocked by overlapping *dcm* methylation (see p. 253).

#R0503S 2,000 units $55
#R0503L 10,000 units $220

5′... PuGGNCCPy... 3′
3′... PyCCNGGPu... 5′

EcoR I

Source: An *E. coli* strain that carries the cloned EcoR I gene from *E. coli* RY13 (R.N. Yoshimori)

Reaction Conditions: NEBuffer EcoR I
50 mM NaCl, 100 mM Tris-HCl, 10 mM $MgCl_2$, 0.025% Triton X-100 (pH 7.5 @ 25°C). Incubate at 37°C.

Ligation and Recutting: After 100-fold overdigestion with EcoR I, > 95% of the DNA fragments can be ligated and recut.

Concentration: 20,000 and 100,000 units/ml. Assayed on λ DNA.

Storage Conditions: 300 mM NaCl, 10 mM KPO_4 (pH 7.5), 0.1 mM EDTA, 10 mM 2-mercaptoethanol, 0.15% Triton X-100, 200 µg/ml BSA and 50% glycerol. Store at –20°C.

Diluent Compatibility: Diluent C, see p. 237.

Heat Inactivation: 65°C for 20 minutes.

Note: Cleavage of mammalian genomic DNA is impaired by overlapping CpG methylation (see p. 252).

For performing double digests with EcoR I, see p. 236.

Conditions of low ionic strength, high enzyme concentration, glycerol concentration > 5%, or pH > 8.0 may result in star activity (see p. 245).

#R0101S 10,000 units $50
#R0101L 50,000 units $200

5′... GAATTC... 3′
3′... CTTAAG... 5′

EcoR V

Source: An *E. coli* strain that carries the cloned EcoR V gene from the plasmid J62 pLG74 (L.I. Gladman)

Reaction Conditions: NEBuffer 3 + BSA
100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol (pH 7.9 @ 25°C). Supplement with 100 µg/ml BSA. Incubate at 37°C.

Ligation and Recutting: After 10-fold overdigestion with EcoR V, > 95% of the DNA fragments can be ligated and recut.

Concentration: 20,000 and 100,000 units/ml. Assayed on λ DNA.

Storage Conditions: 50 mM NaCl, 10 mM Tris-HCl (pH 7.4), 0.1 mM EDTA, 1 mM dithiothreitol, 200 µg/ml BSA and 50% glycerol. Store at –20°C.

Diluent Compatibility: Diluent A, see p. 237.

Heat Inactivation: 80°C for 20 minutes.

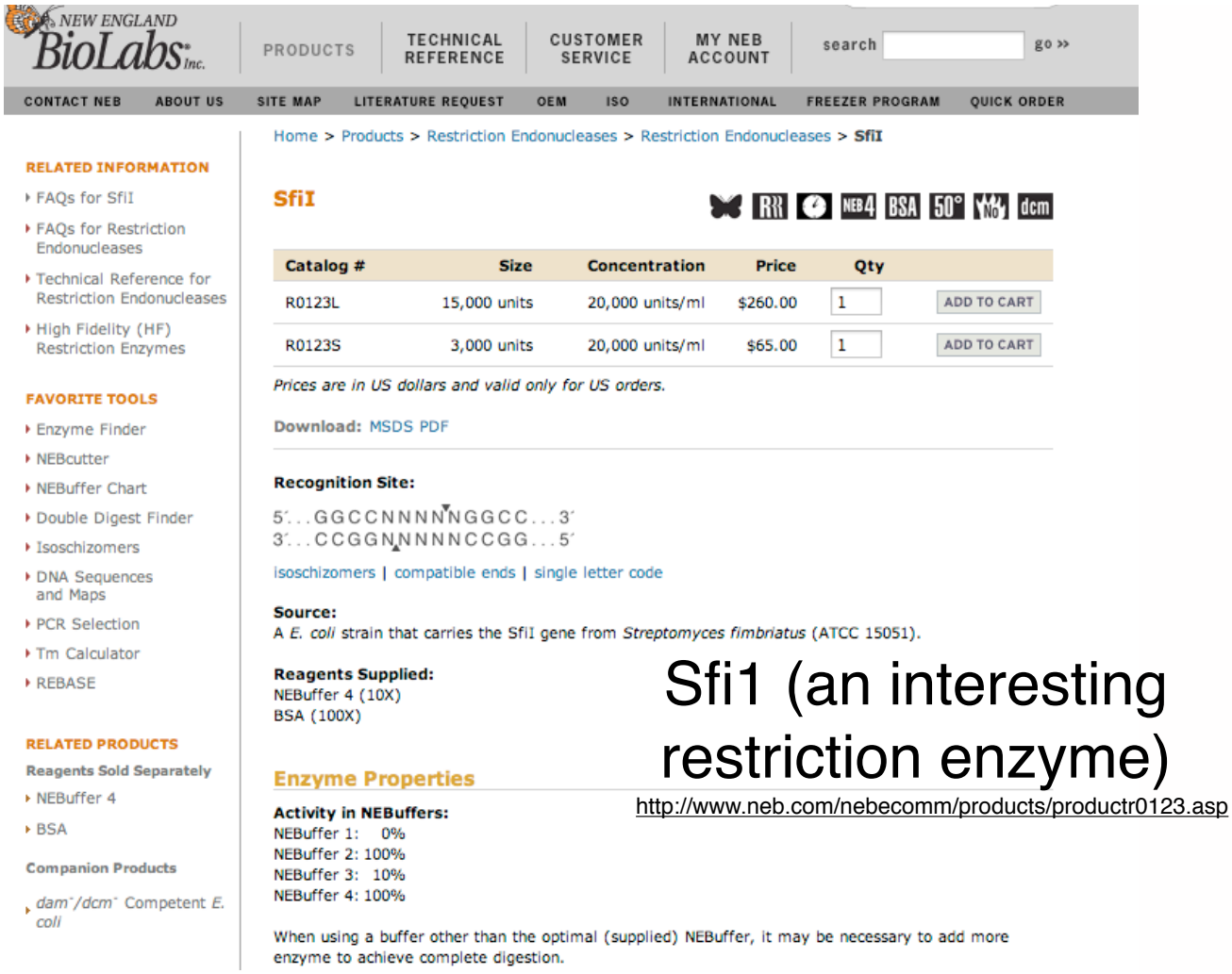
Note: Cleaves to leave a ligatable blunt end in the tetracycline resistance gene of pBR322.

Cleavage of mammalian genomic DNA is blocked by overlapping CpG methylation (see p. 252).

#R0195S 4,000 units $50
#R0195L 20,000 units $200

5′... GATATC... 3′
3′... CTATAG... 5′

- 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.
- *Q: I have found that you need some additional DNA sequence between the restriction site and and the end of the PCR product, but I am not sure how to come up with these sequences*
- *A: I didn't mention it in class, but restriction enzymes often have less activity when they are cutting substrates at the very end of 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.*

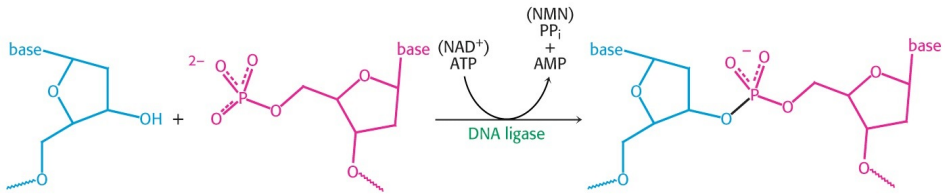


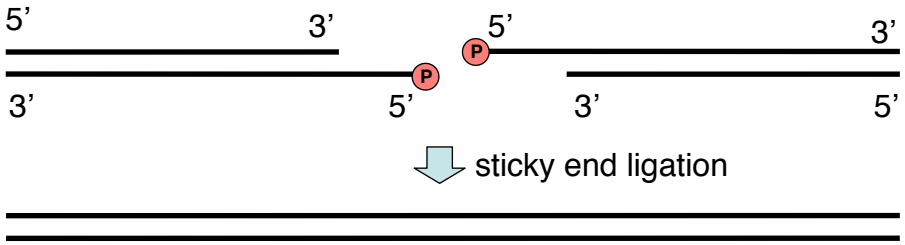
Sfi1 (an interesting restriction enzyme)

http://www.neb.com/nebecomm/products/productr0123.asp

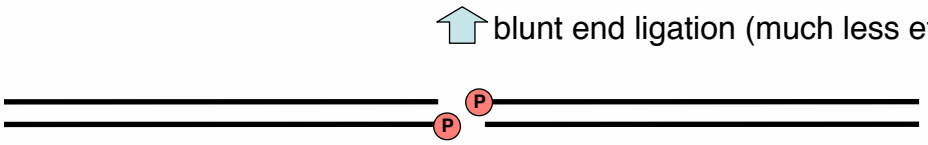
- Note that Sfi1 is an interesting restriction endonuclease that has the following recognition site: 5' GGCCNNNN/NGGCC 3' where N = any nucleotide. What might be an advantage of using such an enzyme?
- The advantage is that you can design the sticky end to be whatever you want. In the phage display example, 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.
- The function of these sticky ends is the same as for any sticky end: to allow the DNA to be ligated to another piece with the same sticky end. You could use two different restriction sites and two different restriction enzymes to achieve the same thing. However, Sfi1 simplifies the process a bit since you only need to add one enzyme to the digestion reaction, yet you get two different sticky ends.

Overall mechanism of DNA ligation





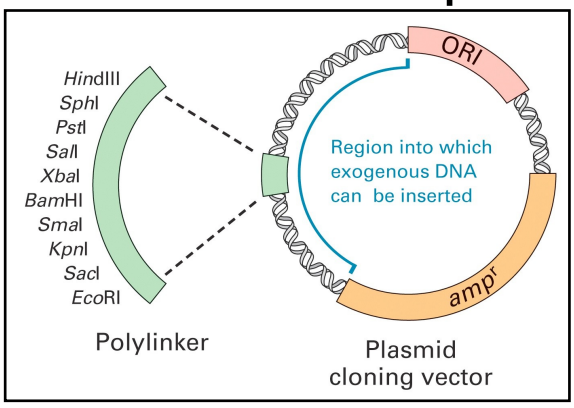
sticky end ligation



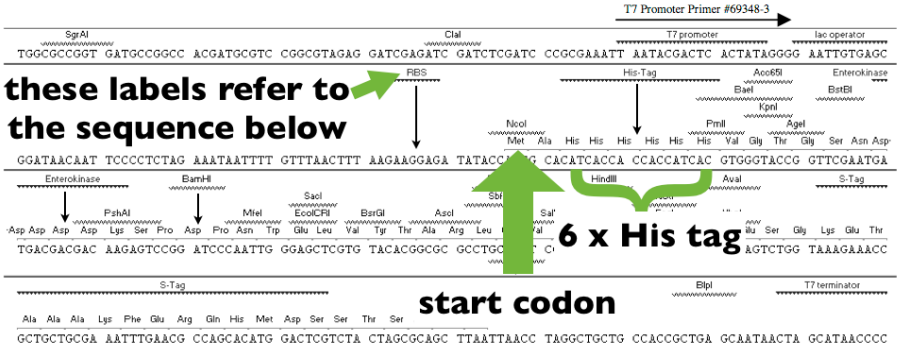
blunt end ligation (much less efficient)

- As we have seen earlier int 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 archaebacteria 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.
- *Q: 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?*
- *A: 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.*
- *Q: I did not totally understand blunt ends and how they are ligated.*
- *A: Ligation is most efficient when there are sticky ends that match up with each other. However, ligation can also happen with blunt ends. A blunt end means that there is no overhang at all. Any two blunt ends can be ligated together, as long as there is at least one 5' phosphate. When primers are purchased, you can choose to get a 5' phosphate added. The default is that there is no 5' phosphate, so it costs a bit more to have one added on.*

Restriction enzymes plus ligases allow us to to manipulate genes and introduce them into bacteria in the form of plasmids

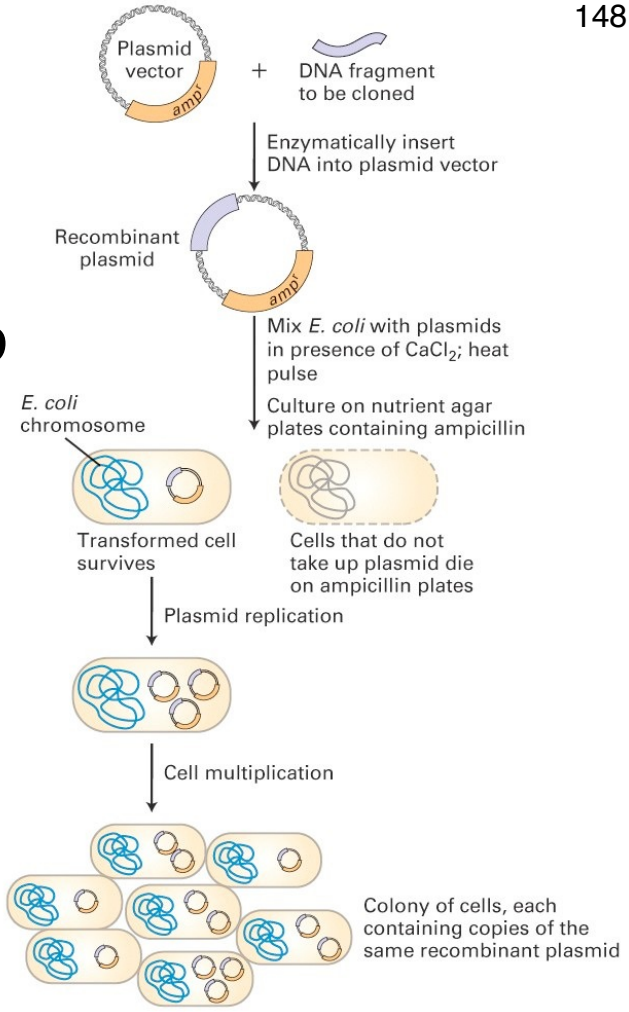


Example of polylinker site for plasmid pET-45b(+)



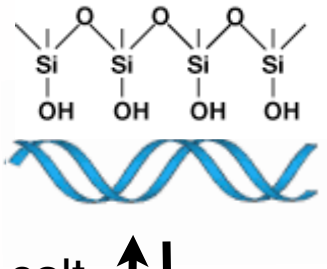
- Plasmids are circular, double-stranded DNA (dsDNA) molecules that are separate from a cell's chromosomal DNA. These extra-chromosomal DNAs, which occur naturally in bacteria, yeast, and some higher eukaryotic cells, exist in a parasitic or symbiotic relationship with their host cell. Plasmids range 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 (polylinker) in which exogenous DNA fragments can be inserted.
- If there is a promoter and a ribosome binding site at the 5' end of the polylinker, 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.

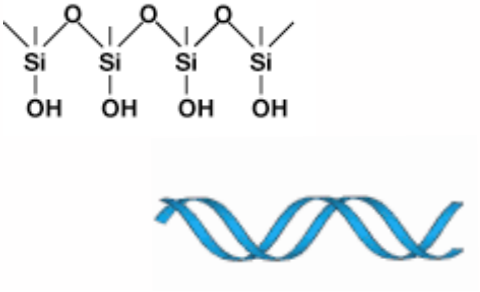
Plasmids introduced into bacteria are copied and can be amplified many times over

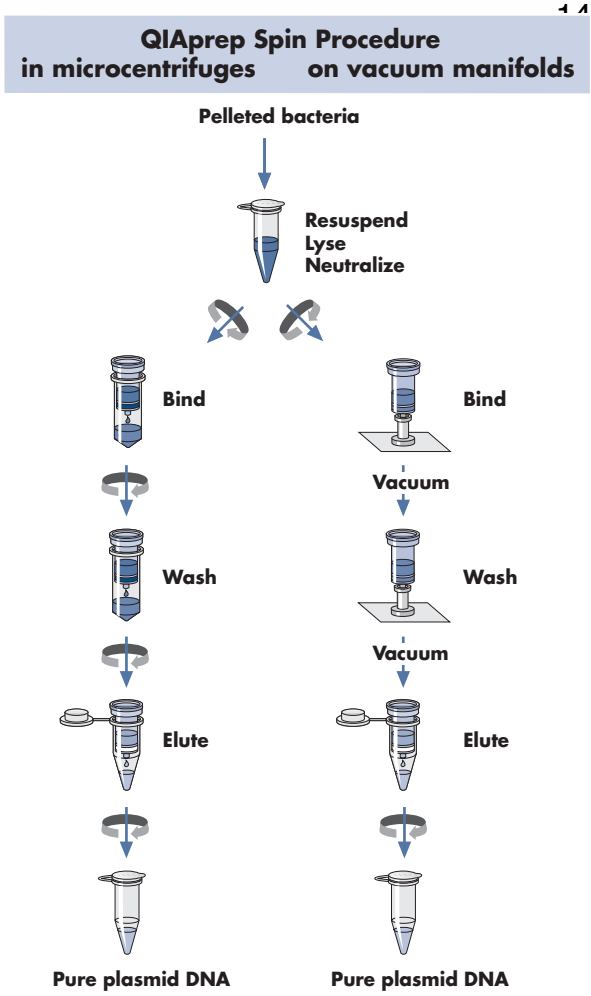


- *Q: Why can a linear DNA molecule not be used to transform bacteria?*
- *A: I am not sure about the exact mechanism involved, but linear DNA is generally not effective for transforming E. coli. I suspect it is recognized as foreign (identified as potentially being phage) and so the cell does not methylate 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.*
- *Q: Could you explain the difference of plasmid replication vs cell multiplication for DNA recombination?*
- *A: Once inside of a bacteria such as E. coli, plasmids are copied by DNA polymerase. Different types of plasmids have different 'copy numbers' which refer to the number of copies of the plasmid (typically 10-100s) inside of each bacteria. As plasmids are being copied inside of E. coli, the E. coli itself is dividing and replicating. So starting with just one E. coli with one plasmid, eventually there will be billions of bacteria per millilitre, each of which contain a number of plasmids equal to the copy number.*

Rapid purification of 10-20 μg plasmid DNA from 2-3 ml of bacterial culture



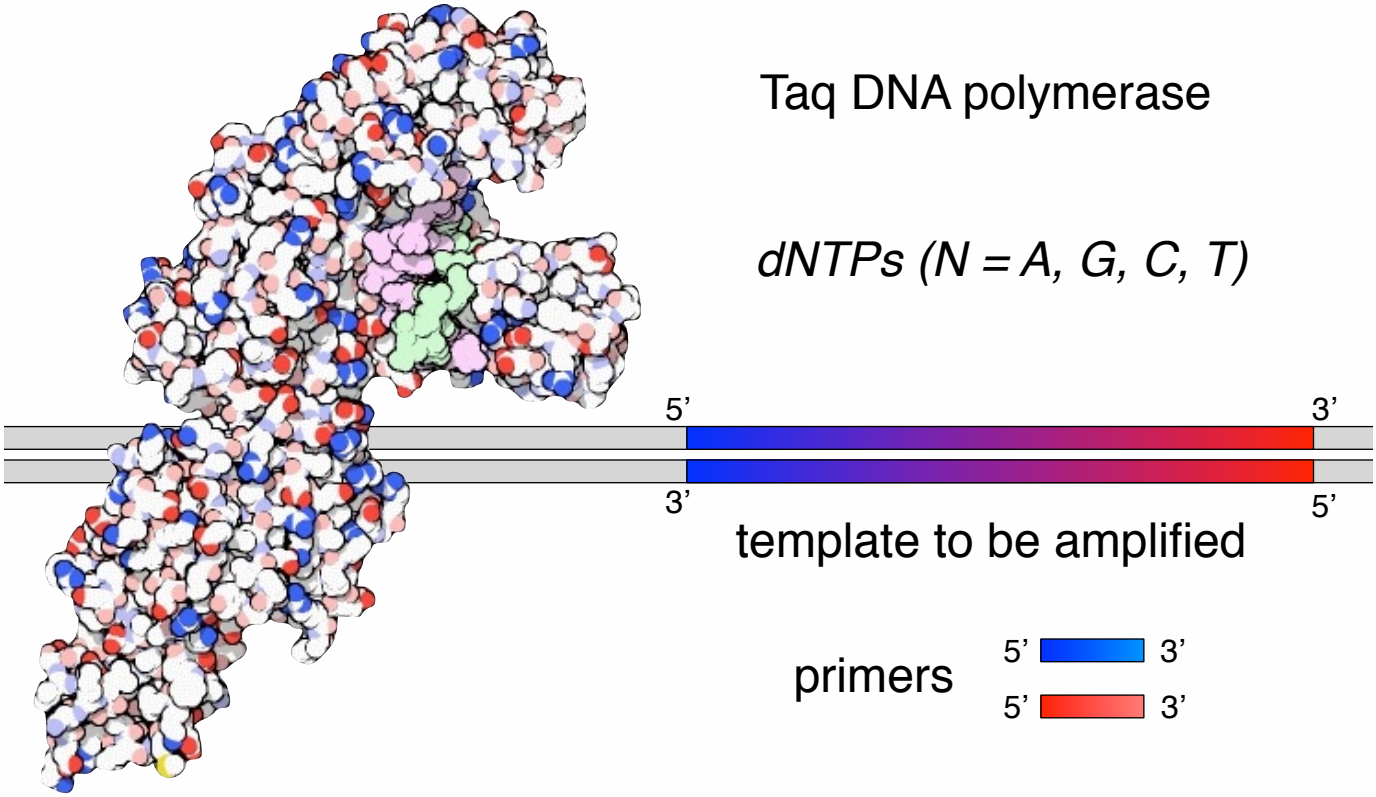




- Many different methods and technologies are available for the isolation of genomic DNA. In general, all methods involve disruption and lysis of the starting material followed by the removal of proteins and other contaminants and finally recovery of the DNA. The choice of a method depends on many factors: the required quantity and molecular weight of the DNA, the purity required for downstream applications, and the time and expense.
- A variety of vendors (one of which is Qiagen) sell kits for purification of plasmid DNA from bacterial culture. You might think that the obvious way of purifying DNA would be to use ion exchange resins, since DNA is so highly negatively charged due to the phosphate backbone. You are right, this is one way of purifying DNA and resins (based on diethylaminoethanol (DEAE)) are available for just this purpose. One disadvantage though is that you need to elute with very high salt concentrations that would interfere with later manipulations. To get rid of the salt the DNA must be precipitated with isopropanol and redissolved in water.
- The preferred method for purification of small quantities of DNA is to take advantage of the fact that DNA binds with high affinity to silica in the presence of chaotropic salts such as guanidinium hydrochloride. The exact nature of this binding does not seem to be well understood. A typical procedure is shown on this slide. The buffers used at each step include:
    - Lyse with: 200 mM NaOH, 1% SDS
    - Neutralize with: 4.2 M Guanidinium-HCl, 0.9 M potassium acetate, pH 4.8
    - Spin to pellet protein precipitate and membranes then bind to column
    - Wash column with: 10 mM Tris-HCl pH 7.5, 80% ethanol
    - Elute with water

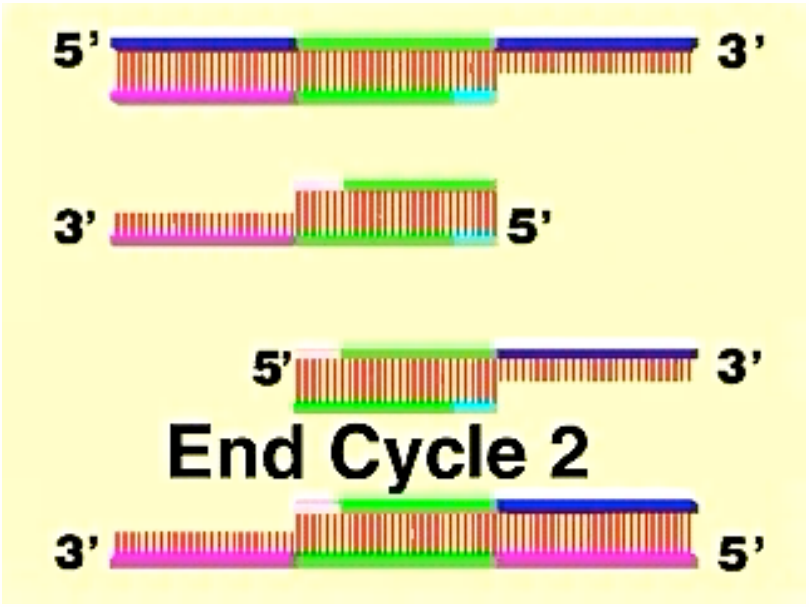
- http://www1.qiagen.com/resources/info/qiagen_purification_technologies_1.aspx
- http://www.madsci.org/posts/archives/2005-04/1113575855.Mb.r.html
- http://www.bio.net/bionet/mm/methods/1996-July/047188.html
- http://www.pnas.org/content/76/2/615.full.pdf

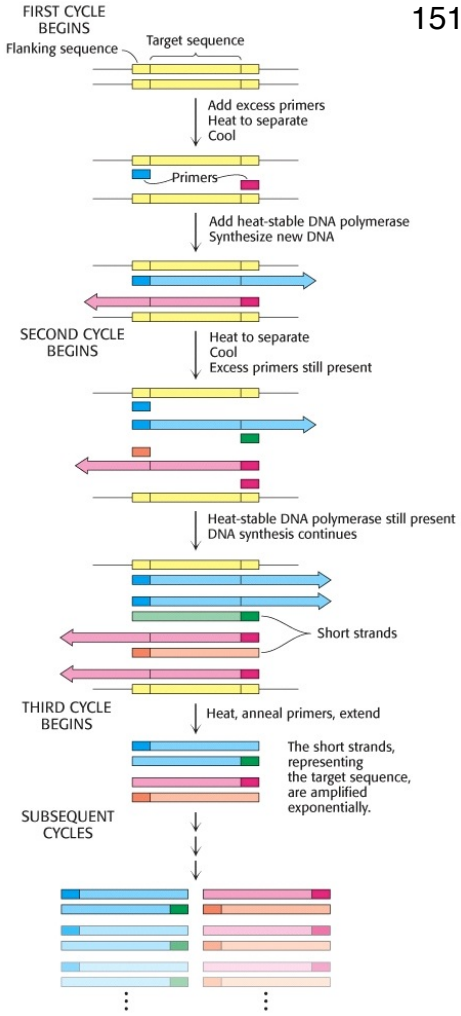
Components of the Polymerase Chain Reaction (PCR)



- 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 (annealing), ~70 °C (extension), and ~95 °C (denature) 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/
- The key to a successful PCR reaction is having primers that bind to the template at the annealing temperature. The Tm that matters is for the binding that will occur in the very first cycle and you should only calculate Tm based on the region that is complementary to the template. Extra restriction sites or added regions should not be included in the Tm calculation if they are not present in the template. In later cycles the length of the complementary sequence will increase to include the whole primer and the Tm will increase.
- *Q: I don't quite understand what the forward and backward primers are. Would you please give me a brief description?*
- *A: 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.*

Polymerase Chain Reaction (PCR)



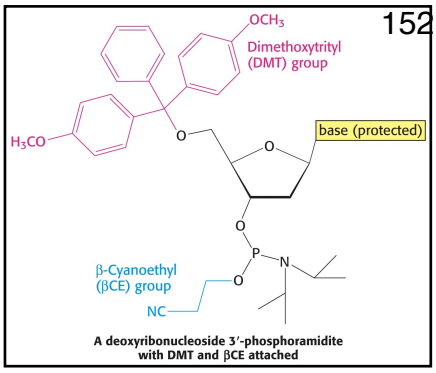


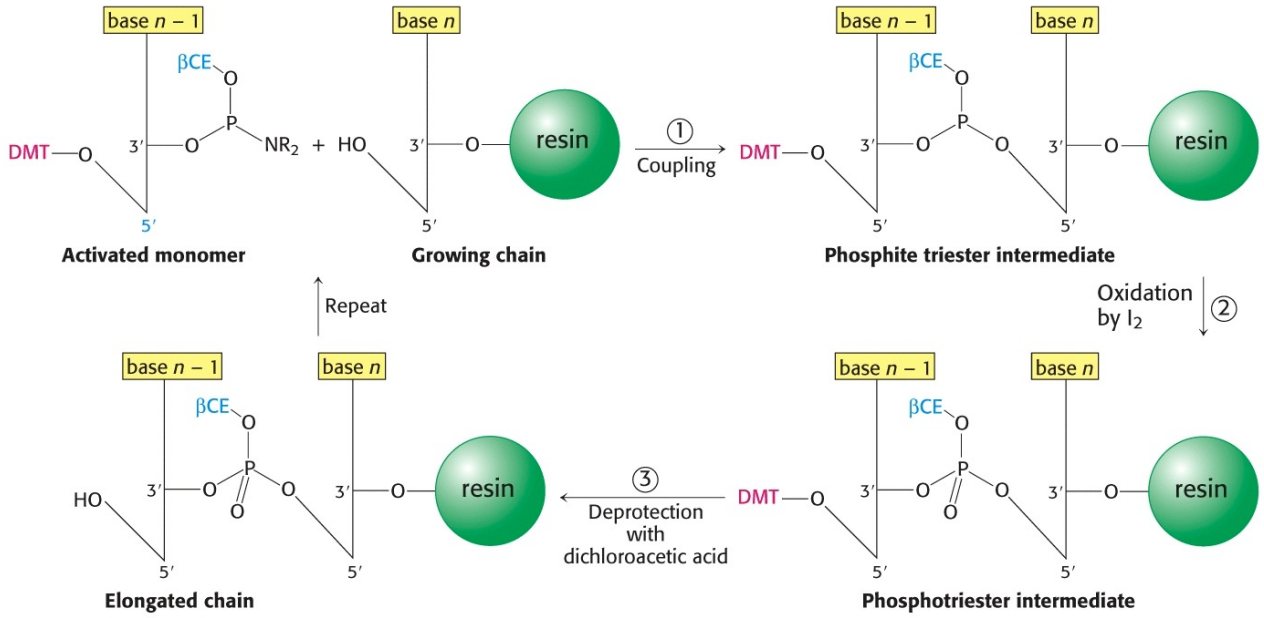
PCR video on youtube: http://www.youtube.com/watch?v=y0INSw9dCTg

- 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
- 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.
- *Q: In PCR, when the temperature is heated up for DNA synthesis, will the primer be separated from the template again? You know, this temperature is higher than the melting temperture.*
- *A: DNA polymerase is probably a little bit active at the annealing temperature, and continues to get more active as the temperature warms up to the extension temperature. Probably this activity is enough to extend the primer such that it does not dissociate from the template. When the reaction warms up to 95, all of the primers (now fully extended for the length of the template) will melt.*
- *Q. One of the components to carry out PCR is "a pair of primers that hybridize with the flanking sequences of the target". What do you mean by that? From my understanding from the movie, the primer hybridize to the target sequences instead of the flanking sequences for the 1st cycle.*
- *A. Your understanding is correct, and this is just a minor issue of word choice. You could say that the target sequence is everything between the two primers, and this is more or less correct, since primers are small relative to the size of a gene. On the there other hand the part of the template where the primer anneals also ends up being included in the PCR product, so it could be considered part of the target sequence.*

Oligonucleotides can be made by chemical synthesis

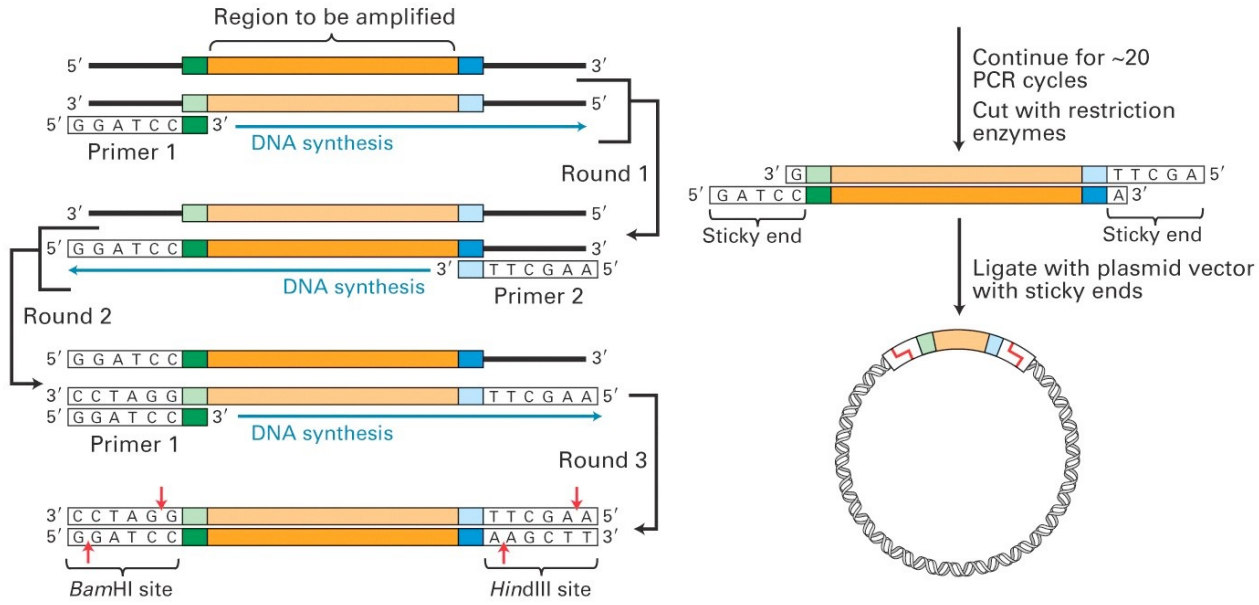
- http://www.invitrogen.com/
- https://www.sigma-genosys.com/
- http://www.idtdna.com/





- DNA strands can be synthesized by the sequential addition of activated monomers to a growing chain that is linked to an insoluble support.
  - In step 1, the 3'-phosphorus atom of this incoming unit becomes joined to the 5'-oxygen atom of the growing chain to form a *phosphite triester*. The 5'-OH group and the 3'-phosphoryl group of the activated monomer are rendered unreactive by attachment of protecting groups.
  - In step 2, the phosphite triester (in which P is trivalent) is oxidized by iodine to form a *phosphotriester* (in which P is pentavalent).
  - In step 3, the DMT protecting group on the 5'OH of the growing chain is removed by the addition of dichloroacetic acid, which leaves other protecting groups intact. The DNA chain is now elongated by one unit and ready for another cycle of addition.
- Each cycle takes only about 10 minutes and elongates more than 98% of the chains.
- At the end of the synthesis, NH_3 is added to remove all protecting groups and release the oligonucleotide from the solid support.
- The sample can be purified by high-pressure liquid chromatography or by electrophoresis on polyacrylamide gels.
- Groups other than nucleotides can be attached to the DNA. An oligonucleotide labeled at one end with ^{32}P or a fluorescent tag can be used to search for a complementary sequence in a very long DNA molecule or even in a genome consisting of many chromosomes.

PCR frees us from the *tyranny of restriction sites*



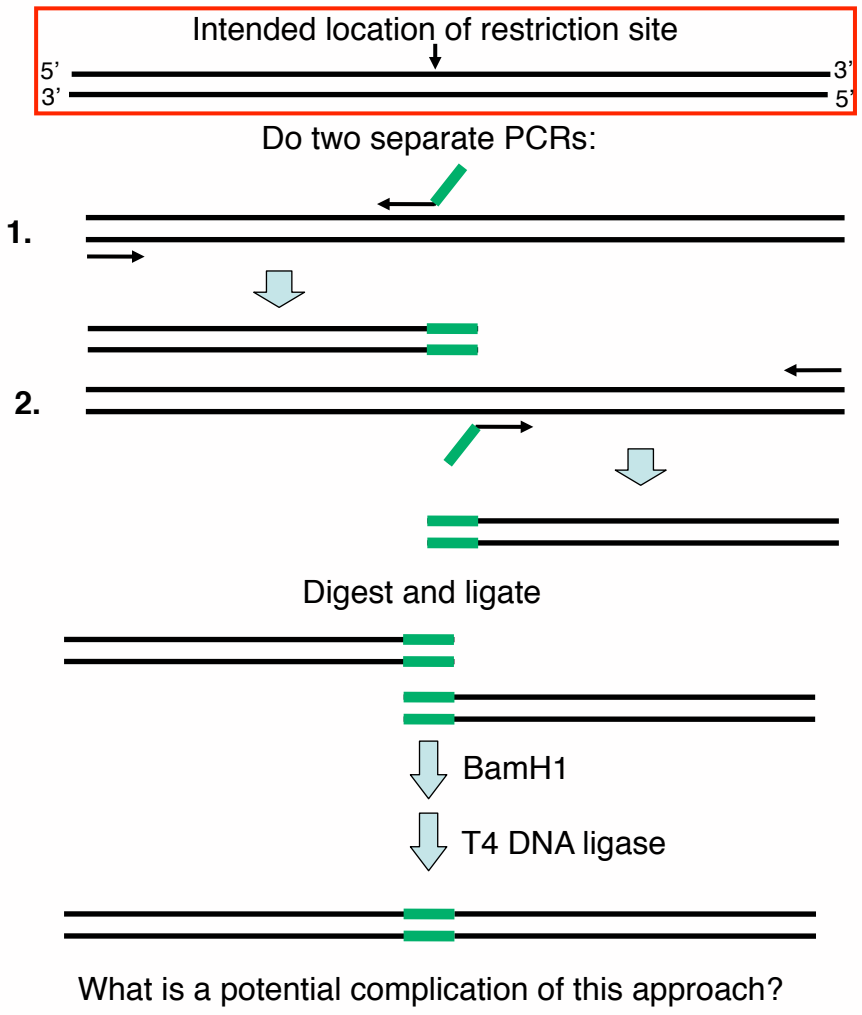
Three approaches to ligating PCR products into plasmids:

- 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.
- *Q: What is the TA overhangs interaction like between plasmid and inserted DNA section? Is there always 3'T overhangs in the plasmid? How is it generated in the plasmid?*
- A: *When Taq is used in PCR, the products always have a 3' A overhang. This is like a short sticky end that could be used to ligate into a plasmid that has 3' T overhangs. Companies sell plasmids that are already cut and have the 3'T overhang already, so you can just mix the PCR product with the purchased cut plasmid, and perform the ligation. This is good for the companies since people need to keep buying new cut plasmid. There are two ways to prepare the cut plasmid. The first would be to first cut the plasmid with an enzyme that makes a blunt end. Then treat it with Taq and only dTTP (not the other 3 nucleotides). When only one nucleotide is present, Taq will add it to the blunt end to create a 3' overhang. It prefers A, but can use the other nucleotides if no A is present. The other approach is to use a restriction enzyme that leaves a 3'T overhang. The only example of an enzyme that can do this is XcmI (http://www.neb.com/nebecomm/products/productr0533.asp).*
- *Q. What does "insert into Bamh1/HindII" site means? Does one side of the DNA conjugates on the BamH1 site and other on the HindIII site ?*
- *A. Recall that we are talking about a circular plasmid DNA. The plasmid will be cut with BamH1 and HindIII to give sticky ends, removing a small piece of the plasmid and effectively breaking the circle. The new piece of DNA will be ligated (inserted) into the cut site to take the place of the piece of DNA that was formerly between the BamH1 and HindIII sites and thereby reforming the circle. This is what is meant by 'inserting into the BamH1/HindIII site'.*

Molecular Biology applications: insertion of a restriction site

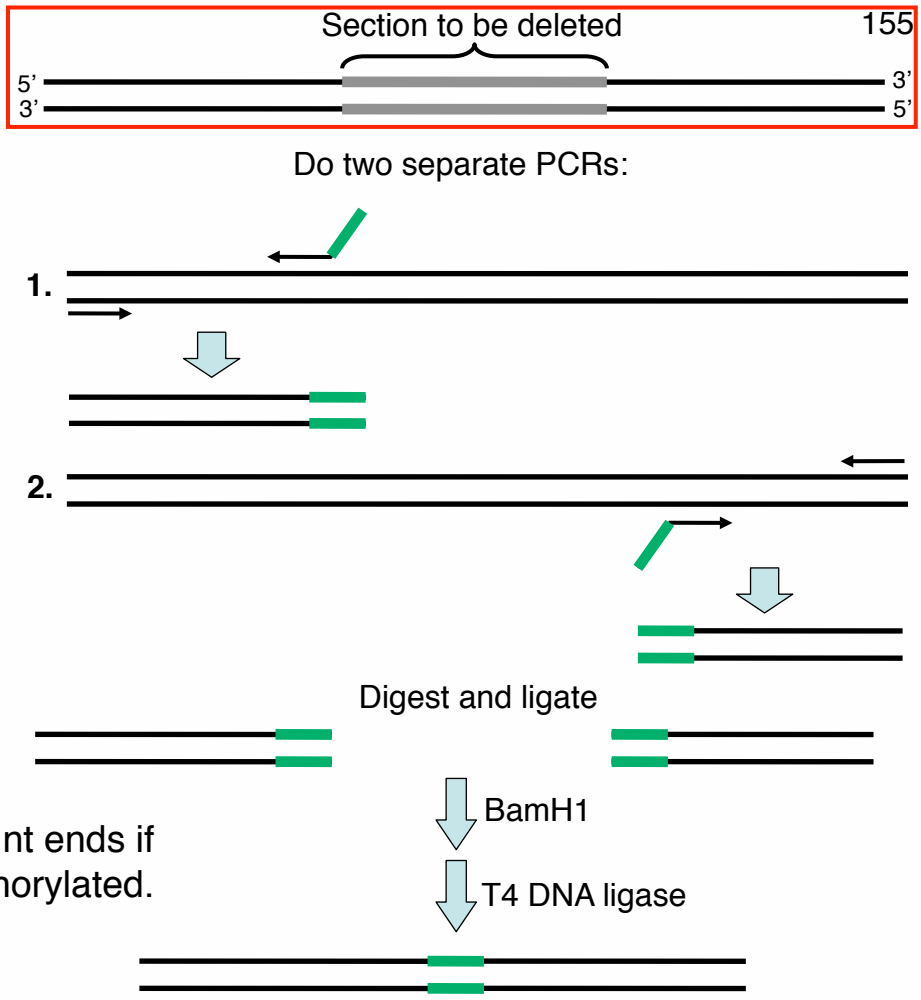
Note that this strategy could also be used to delete a section of DNA, if the internal primers were moved further apart.



What is a potential complication of this approach?

- 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.
- *Q: 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?*
- *A: 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).*
- *Q: What is a potential complication?*
- *A: The complication is that each fragment can ligate to itself. If the two fragments were labeled A and B, we would get a mixture of AA, AB and BB products (in a 1:2:1 ratio).*

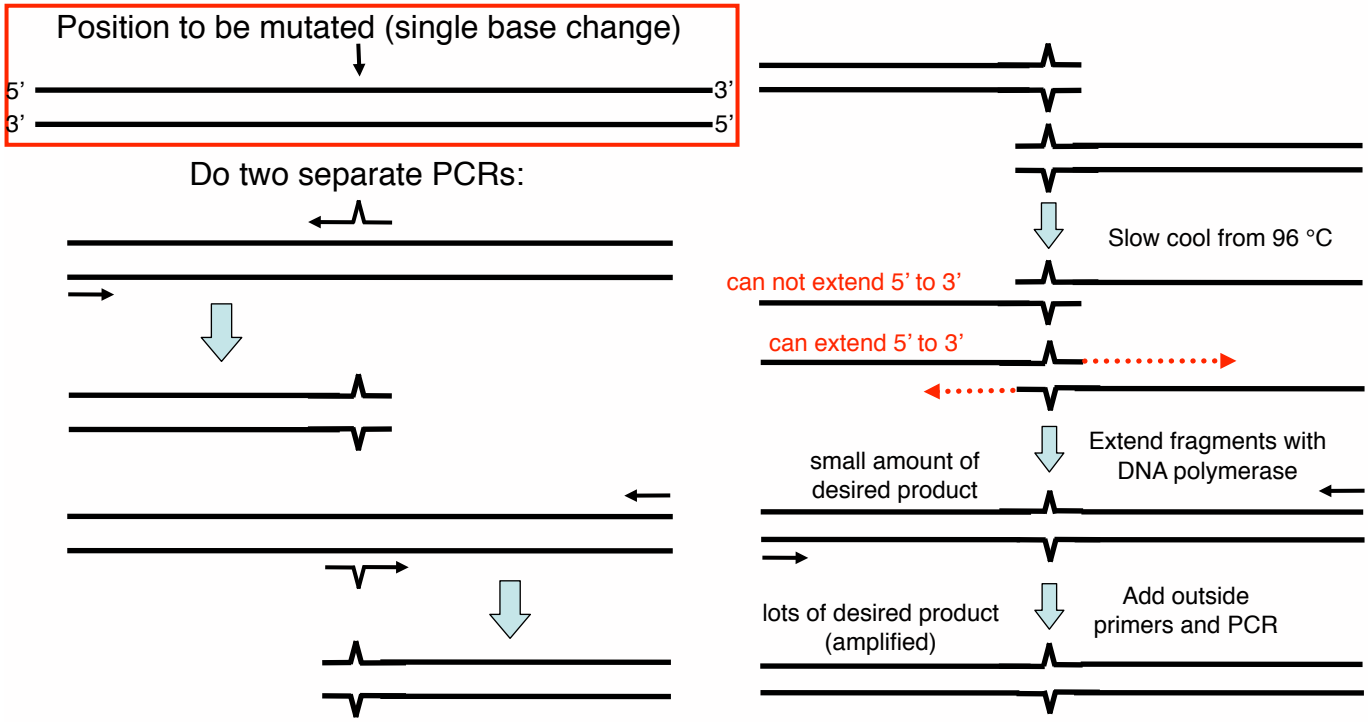
Molecular Biology applications: deletion of a section of DNA



Can also use blunt ends if primers are 5’-phosphorylated.

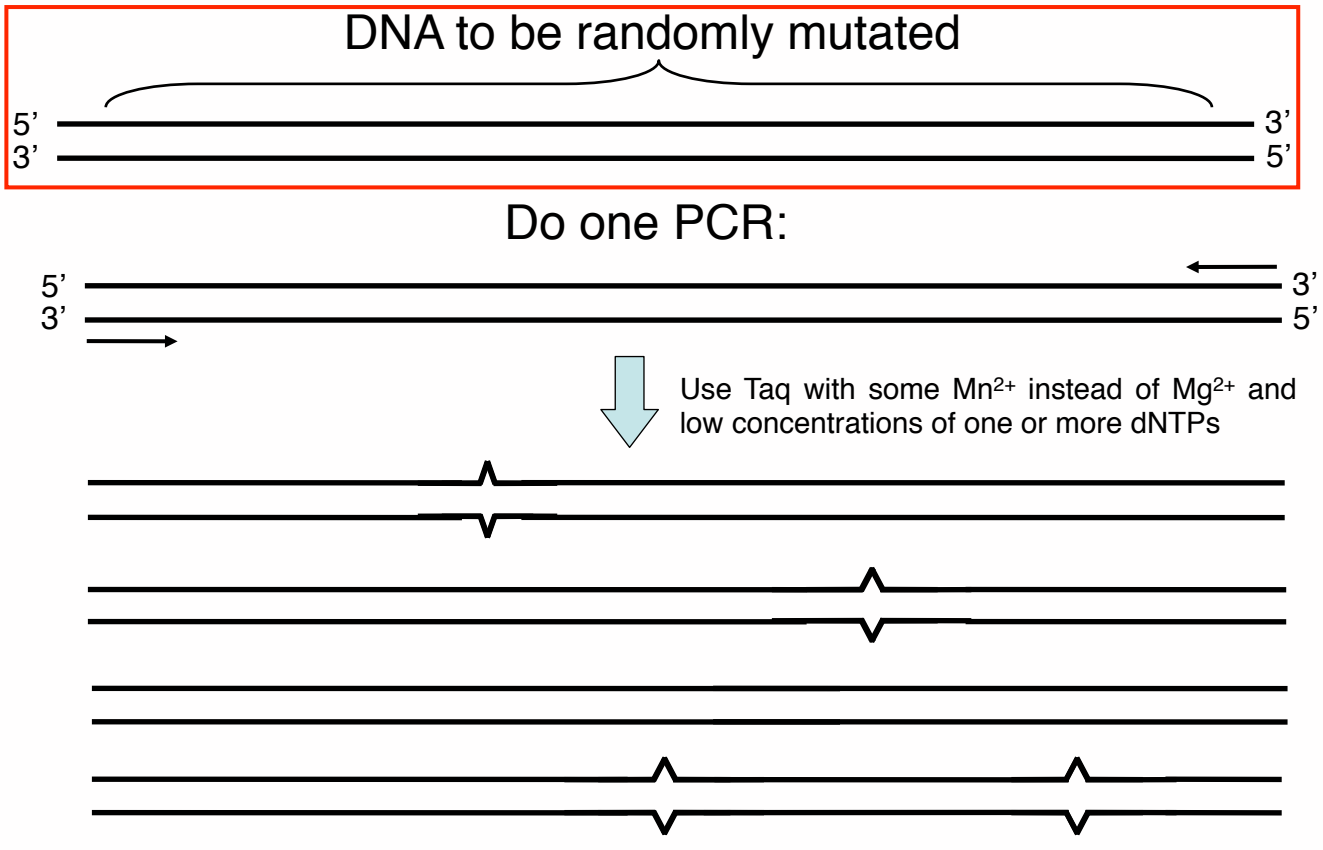
- Ligation is most efficient when there are sticky ends that match up with each other. However, ligation can also happen with blunt ends. A blunt end means that there is no overhang at all. Any two blunt ends can be ligated together, as long as there is at least one 5' phosphate. When primers are purchased, you can choose to get a 5' phosphate added. The default is that there is no 5' phosphate, so it costs a bit more to have one added on.
- *Q. In oder to keep the Tm of the two primers as close as possible, the length(or the no. of nucleotides) of overlap region of the two primers might be different. Is it important to keep the length of overlap region the same for the two primers?*
- *A. Keeping equal length is not important. Only the melting temperature matters. Within 5 degrees is probably fine, but closer is better.*
- *Q. Are all three stop codons suitable to use in designing primer? or UAG is particularly preferred?*
- *A. Any of the three stop codons is fine.*
- *Q. I’ve read that Tm of 55-75°C are preferred, does it refer to the Tm for the complementary sequence between primer and template or the Tm of the primer with primer itself?*
- *A. The only Tm that matters is the Tm for the complementary sequence between the primer and the original template. In the first cycle of PCR, this is the key annealing event. In later cycles, the vast majority of the "template" will be the newly synthesized synthesized DNA which will be completely complementary to the primer. Accordingly, the relevant Tm will now be higher, but this is typically not an issue.*
- *Q. If the Tm for the complementary sequence between primer and template are preferred to be 55-75°C, it is difficult to keep the primer strand within 17-24 bases (tips from http://www.easyprotocols.com/good-primer-design-tips/), as the overlap region of the primer(excluding restriction sites) should have already 17-24 bases to have Tm of 55-75°C. Please advise.*
- *A. The '17-24 base' recommendation is referring to the overlap region. Adding the restriction site and any additional necessary nucleotides will result in longer primer. This is not a problem, as primers can be practically any length, in principle. As they get longer, the price goes up since additional purification is necessary to remove the versions that are shorter due to the <100% efficiency of each synthetic step. Synthetic primers of up to 100 nucleotides are entirely feasible, and some companies are pushing beyond that.*

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).
- *Q. For this slide, why is it possible for only one of the two pairs to extend 5' to 3'?*
- *A. DNA must always be extended from the 3' end. For the top pair (the one that is not extended), the 3' ends are sticking off into space and there is no template for extension.*
- *Q. Is the product of the PCR linear or circular?*
- *A. The product of a PCR reaction must always, always, always, be linear. Quickchange produces circles, but it is not PCR. Of course, you could make a PCR product circular by ligating blunt ends together.*
- *Q. Is there a limit to the number if nucleotides required to keep the primer in frame?*
- *A. There's no limit, but there's no advantage of using more than 0,1, or 2. That is, 0 is the same as 3, 6, 9 etc; 1 is the same as 4, 7, 10 etc; and 2 is the same as 5, 8, 11 etc. If you use more than 0, 1, or 2 you are just inserting extra amino acids between the restriction site and the start codon.*
- *Q. Is it important to add overhangs in the primers?*
- *A. Yes. Please see this table: https://www.neb.com/tools-and-resources/usage-guidelines/cleavage-close-to-the-end-of-dna-fragments*

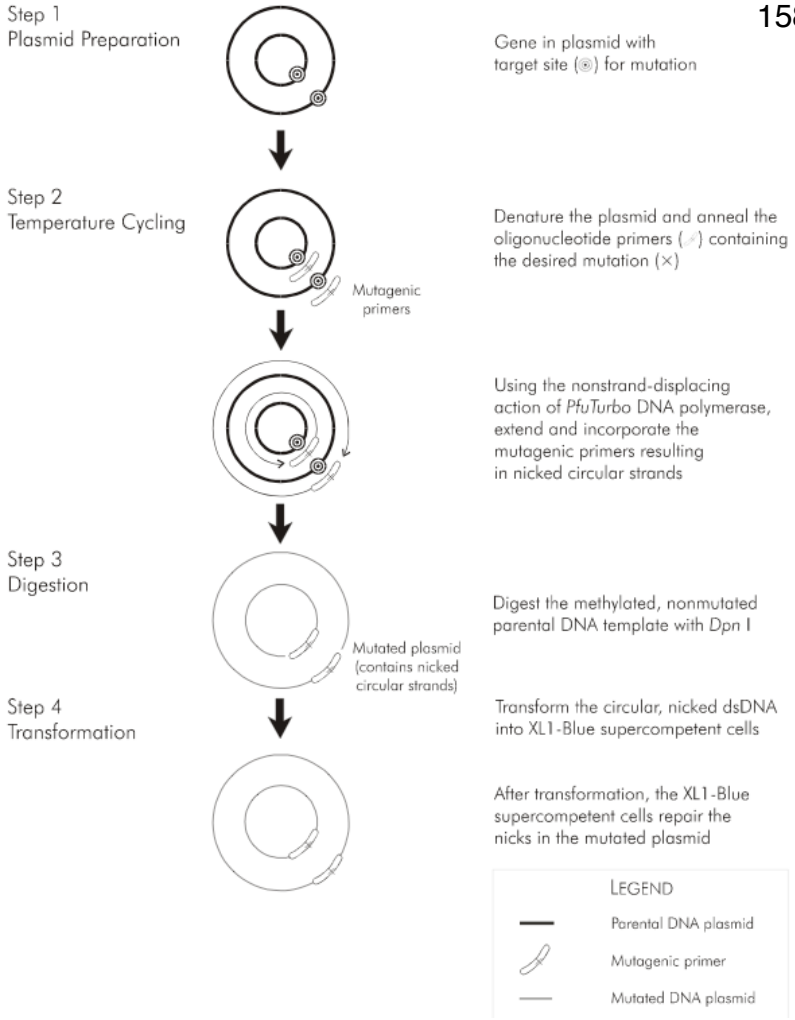
Molecular Biology applications: random mutagenesis



- 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

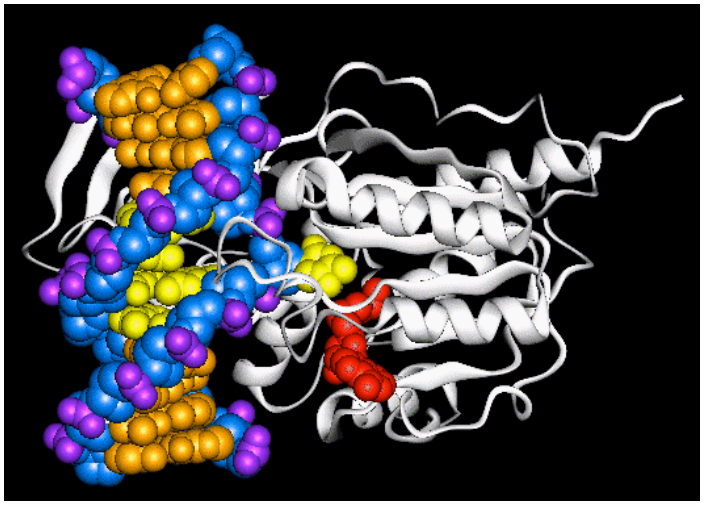
- Why not use Taq?
- Is it exponential or linear amplification during thermocyling?
- Why do you need a kit?

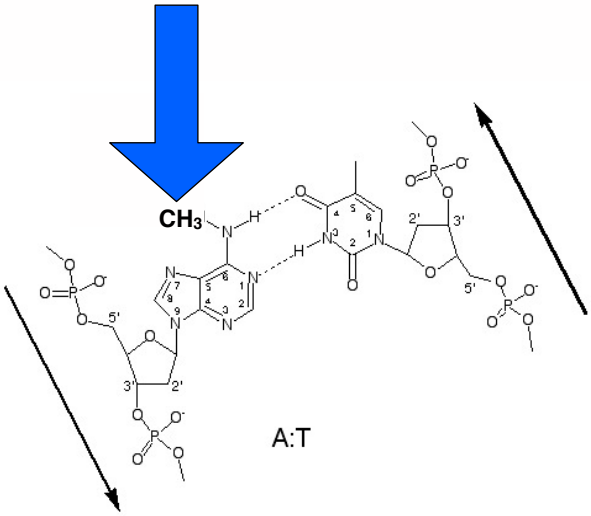


http://www.stratagene.com/

- The next slide shows this process is somewhat more detail.
- *Q: With Quikchange, after the digestion, it seems that are you left with a lot of single strand mutated plasmids because during the thermocyling the new mutated strand can't be used as a template, just wondering if that is indeed what happens and then you can cool it down so that the ss mutated plasmids will bind with its complementary strand to form the ds mutated plasmids?*
- *A: After 18 cycles, there should be 18 times as many mutated strands as there are unmutated (original) strands. When you cool it down for the last time, complementary strands will pair up in all possible combinations of mutated and unmutated strands. The majority will have both strands mutated, simply because this is what there is the most of.*
- *Q: "Why do you need a kit?", I don't quite understand the question.*
- *A: It was meant as a rhetorical question. Kits are convenient, but typically expensive. If you know how the process works, you can purchase the necessary enzymes and components and do the reaction yourself.*
- *Q: In Quikchange, will the newly synthesized nicked strand become linear rather than circular when heated to denaturate dsDNA? And if yes, when they are cooled down, how can they form circle again and then pair up to get dsDNA?*
- *A: Yes. When denatured, all of the products from the Quikchange reaction will be linear. When cooled down, they will form circles for the same reason any two pieces of DNA anneal. That is, there are complementary ssDNA regions with melting temperatures higher than the temperature of the solution following cooling.*
- *Q: In Quikchange ,when E. Coli can fix the nicks why it could not use as template?*
- *A: After E. coli has fixed the nicks, the plasmid could be used as a template. However, the Quikchange reaction takes place in a test tube and there is no enzyme present that would allow the nicks to be fixed. After the plasmid has been introduced into E. coli and repaired by the bacterial machinery, it could be purified as a fully intact plasmid and used for any purpose.*
- *Q.When we choose bacterial machinery to fix the nicks from Quikchange, but not just use ligase?*
- *A. In normal Quikchange this wouldn't work since the primers are not phosphorylated. If they were phosphorylated you could use a ligase to fix the nicks. However, this would add substantial expense and so it is cheaper and easier to just have E. coli do the repair.*

How does Quikchange work? *E. coli* has enzymes that methylate DNA



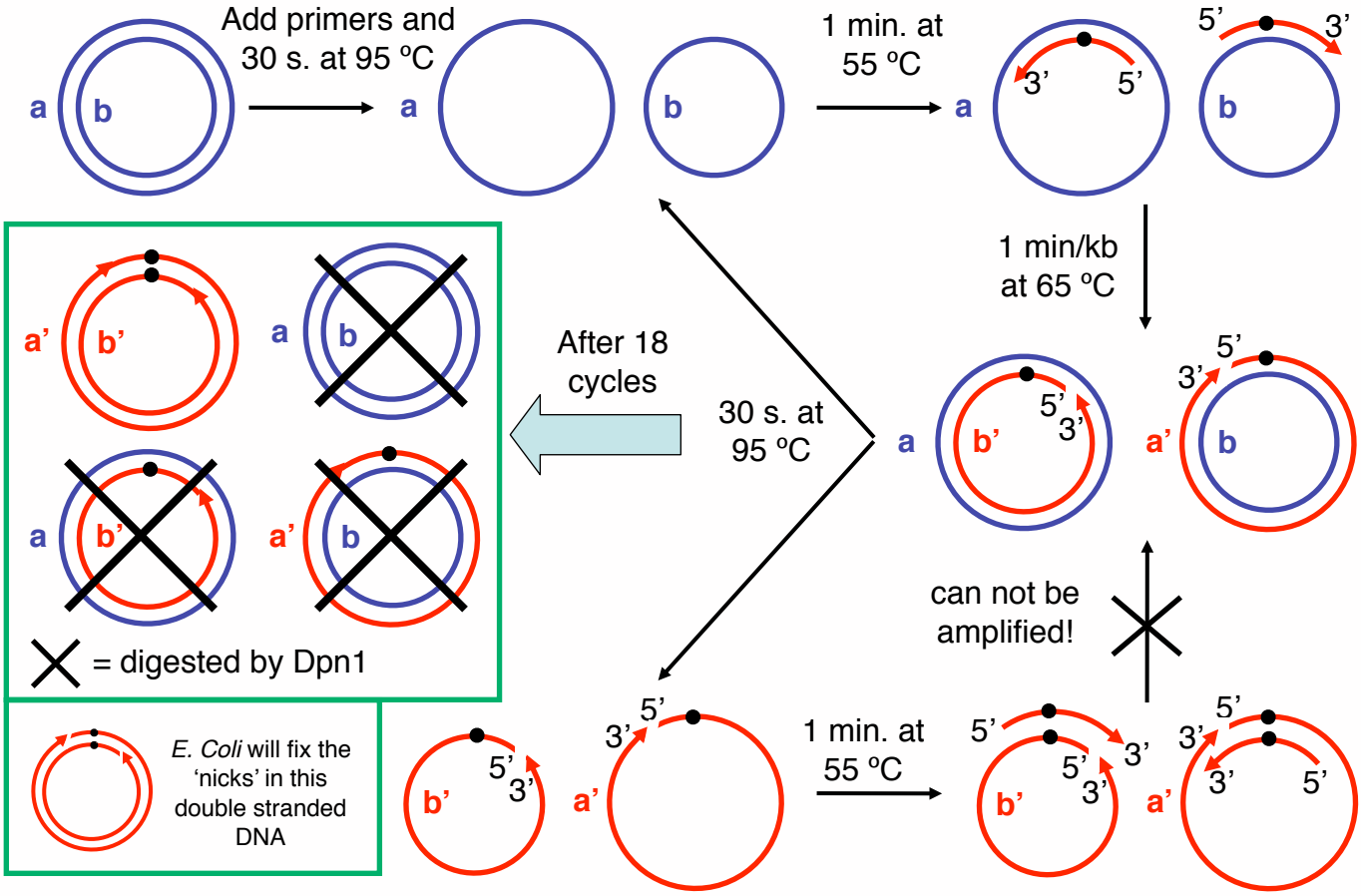


Dam methylase = adenine N6 methyltransferase
Dpn1 = enzyme that cuts at GA*TC, only when A methylated

http://dnamethsoc.server101.com/

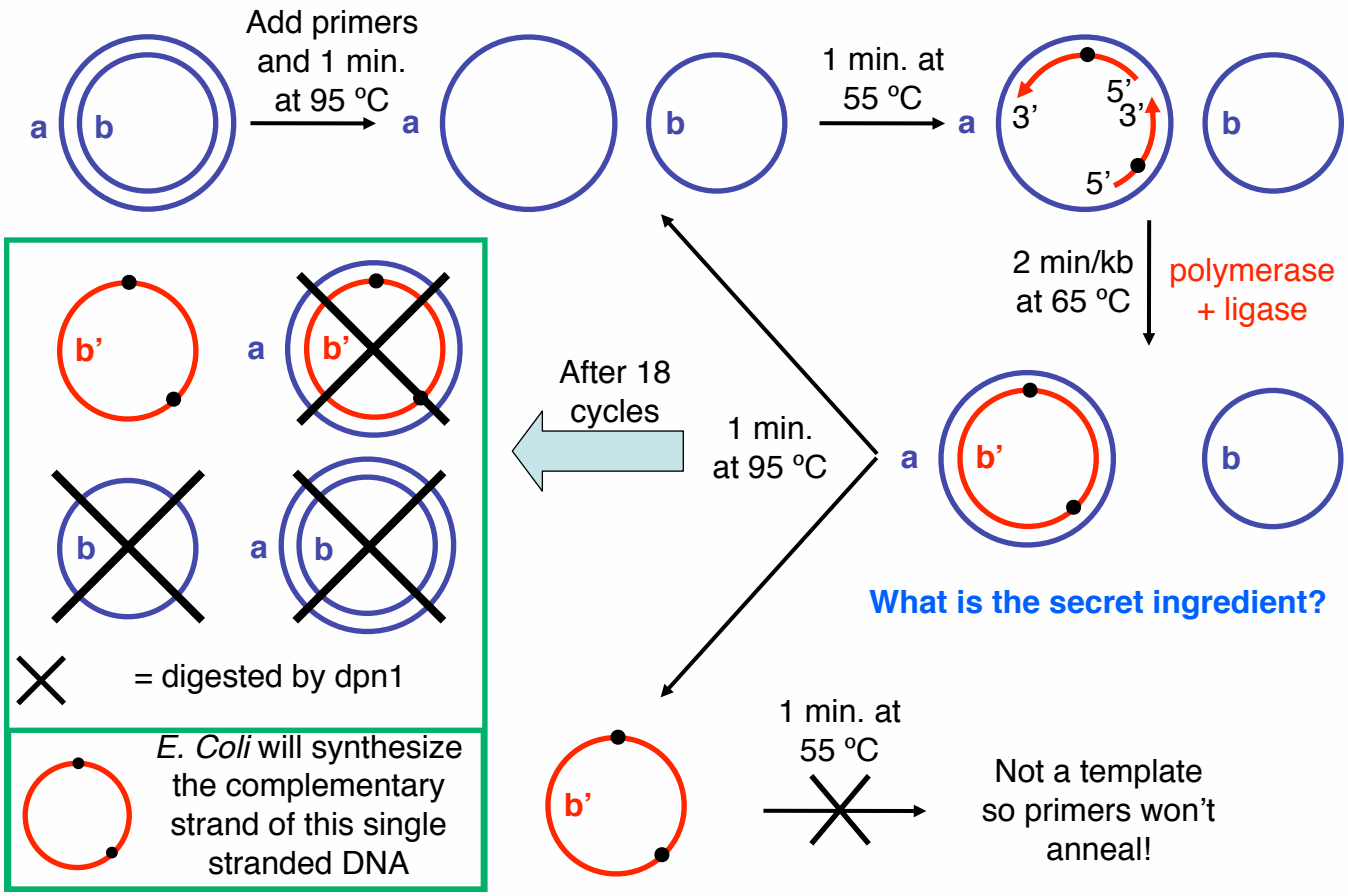
- In prokaryotes, MTases have most often been identified as elements of restriction/modification systems that act to protect host DNA from cleavage by the corresponding restriction endonuclease. Most laboratory strains of *E. coli* contain three site-specific DNA methylases.
  - Dam methylase–methylation 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.
- *Q: 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?*
- *A: 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).*
- *Q: If the plasmids are not methylated why do they do not get chewed up when they are incorporated into Ecoli?*
- A: *Good question. You would think that, since E. coli methylates it's DNA, it probably has a restriction enzyme that digests the unmethylated DNA to protect itself from viruses. It turns out this is not the case, and E. coli actually methylates it's DNA for a different reason. The purpose of methylation is to keep track of which is the template strand (methylated) and which is the newly formed strand (not methylated) following replication. If there is a mismatch in the DNA immediately after replication, the repair machinery will recognize the methylated strand as the 'true' template and use it to repair the mismatch. The Dpn1 restriction enzyme used in Quikchange comes from a different species, and I'm not sure what it's precise biological purpose is.*
- from the NEB technical resource

Quikchange is not PCR



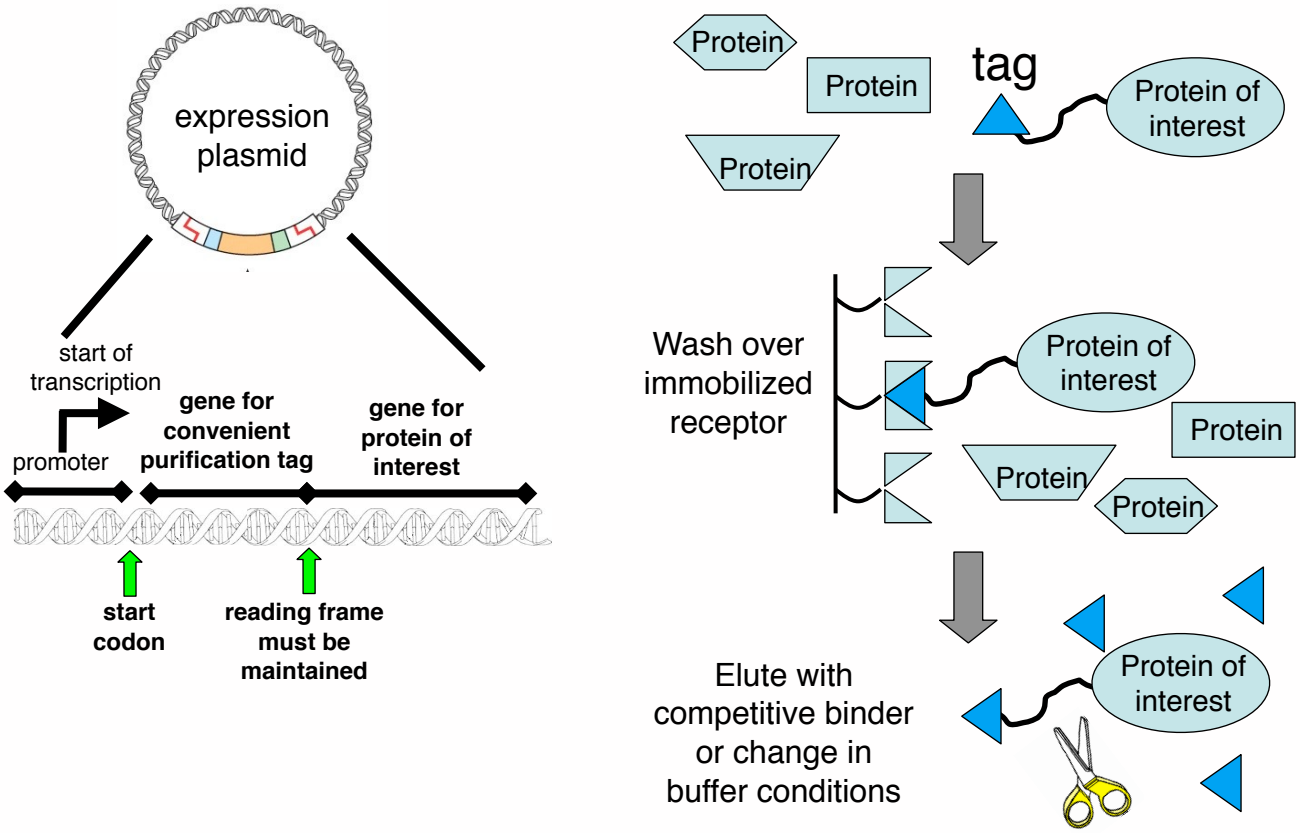
- Strands **a** and **b** are the original complementary strands of the double stranded plasmid DNA
- Heating them at 95 ºC melts the DNA and gives the individual single strands
- Cooling to 55 ºC allows the mutagenic primers that are mostly complementary to template plasmid, with the exception of the designed mutationed
- Heating to 65 ºC 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.
- Taq can not be used for this reaction since it adds a 3' A to the end of newly synthesized DNA
- *Q. The primers used in quickchange for site directed mutagenesis are complementary to each other and easily for a dimer, how is this circumvented?*
- *A. There is a large excess of primers relative to template. Almost certainly most of the primers do end up dimerized upon annealing. However, some of them will bind to the template and enable the DNA extension by polymerase to proceed. Keep in mind that, when using a double stranded template, there is always the possibility of dimerization competing with annealing of primer. The reason it isn't usually a problem is because there is such a large excess of primer that it has a kinetic advantage for binding to the template.*
- *Q. I am not quite sure whether I am right at the design of a primer that could cause a site-directed mutagenesis. For example, I have a DNA sequence 5'GCATGCACGCATGCATGCATGACCTAATTGC.....3' and I want to change T with C, what would the primer be?*
- *A. The forward primer would be something like: 5'- CGCATGCACGCATGACC -3'*
- *The reverse primer would be something like: 5'- GGTCATGCGTGCATGCG -3'*
- *These primers are likely too short. The formula for calculating the melting temperature of mutagenic primers is: Tm =81.5+0.41(%GC) –675/N–%mismatch (from QuikChange® Site-Directed Mutagenesis Kit)*
  where N = primer length (For the primers given here, 17) and %mismatch in this case is 1/17 = 5.9%. Accordingly, for these primers the Tm would be 81.5 +0.41(64.7) - 675/17 - 5.9 = 81.5 + 26.5 - 39.7 - 5.9 = 62.4 degrees (Stratagene recommends Tm > 78 C). There are no restriction sites in mutagenic primers.

Quikchange **Multi**: introduce several mutations at once



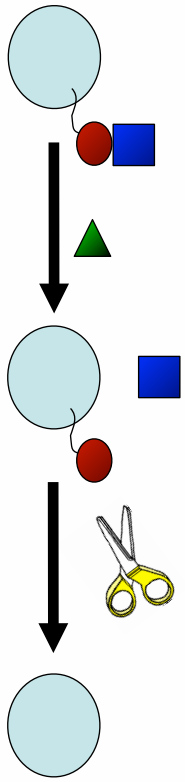
- The key to the Quikchange Multi procedure is that the newly synthesized DNA fragments are getting joined together during the thermocycling reaction. This is sort of like how Okazaki fragments must be joined together during lagging strand DNA synthesis. Accordingly, the 'secret ingredient' must be a ligase. However, it can't be any old ligase, since the mixture is being thermocycled in order to get linear amplification of one strand of the DNA. I strongly suspect that the reaction contains a thermostable ligase, such as Taq ligase.
- *Q: 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?*
- *A: 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.*
- *Q: I did not understand one sentence "It will either be linear or non-covalently circularized due to hydrogen bonds between overlapping sticky ends (as in the Quikchange rxn)". Any overlapping sticky ends will be generated during the Quickchange reaction? I did not understand well about how the nicks disappear by ligation?*
- *A: The idea here is that it is not possible to PCR amplify circular DNA to make a new complete circular DNA with all of phosphodiester bonds intact. In the case of the Quickchange reaction, the product actually does have sticky ends. This is a unique feature of this protocol, since PCR products normally have blunt ends (or just the 3'A overhang with Taq). The reason for the sticky ends in that the DNA polymerase bumps into the 'back-end' of primer bound to the template strand. Try drawing out the product for yourself, both as a circle and in a linear form, and you should see what I mean. The sticky ends are quite long for Quickchange (the whole length of the primer), and the melting temperature for this region is high enough that the plasmid essentially always exists in a circular form, even though there is one phosphodiester bond missing in both strands. If the primer had a 5'phosphate, the addition of a ligase would close the nicks by forming the missing phosphodiester bond. Alternatively, E. coli will fix the nicks itself after the plasmid is introduced into the bacteria (transformation).*
- *Q. I don't understand why the newly synthesized DNA (the small red circle) in Quikchange Muti could not be a template for next run? Unlike the previous Quikchange method, there is no nick in this DNA.*
- *A. There is no nick, but it is same strand (not the complement) as the primers. Accordingly, the primers can not bind to it.*

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).

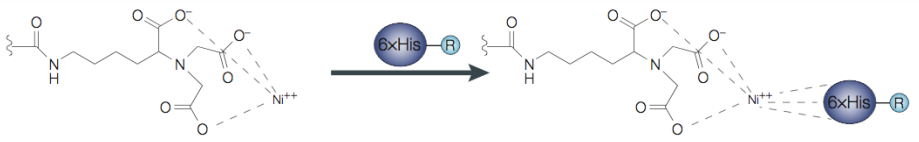
Purification in the age of molecular biology



Affinity tag ●	Matrix ■	Elution condition ▲
Poly-Arg	Cation-exchange resin	NaCl linear gradient from 0 to 400 mM at alkaline pH>8.0
Poly-His	Ni^{2+}-NTA, Co^{2+}-CMA (Talon)	Imidazole 20–250 mM or low pH
FLAG	Anti-FLAG monoclonal antibody	pH 3.0 or 2–5 mM EDTA
Strep-tag II	Strep-Tactin (modified streptavidin)	2.5 mM desthiobiotin
c-myc	Monoclonal antibody	Low pH
S	S-fragment of RNaseA	3 M guanidine thiocyanate, 0.2 M citrate pH 2, 3 M magnesium chloride
HAT (natural histidine affinity tag)	Co^{2+}-CMA (Talon)	150 mM imidazole or low pH
Calmodulin-binding peptide	Calmodulin	EGTA or EGTA with 1 M NaCl
Cellulose-binding domain	Cellulose	Family I: guanidine HCl or urea>4 M Family II/III: ethylene glycol
SBP	Streptavidin	2 mM Biotin
Chitin-binding domain	Chitin	Fused with intein: 30–50 mM dithiothreitol, β-mercaptoethanol or cysteine
Glutathione S-transferase	Glutathione	5–10 mM reduced glutathione
Maltose-binding protein	Cross-linked amylose	10 mM maltose

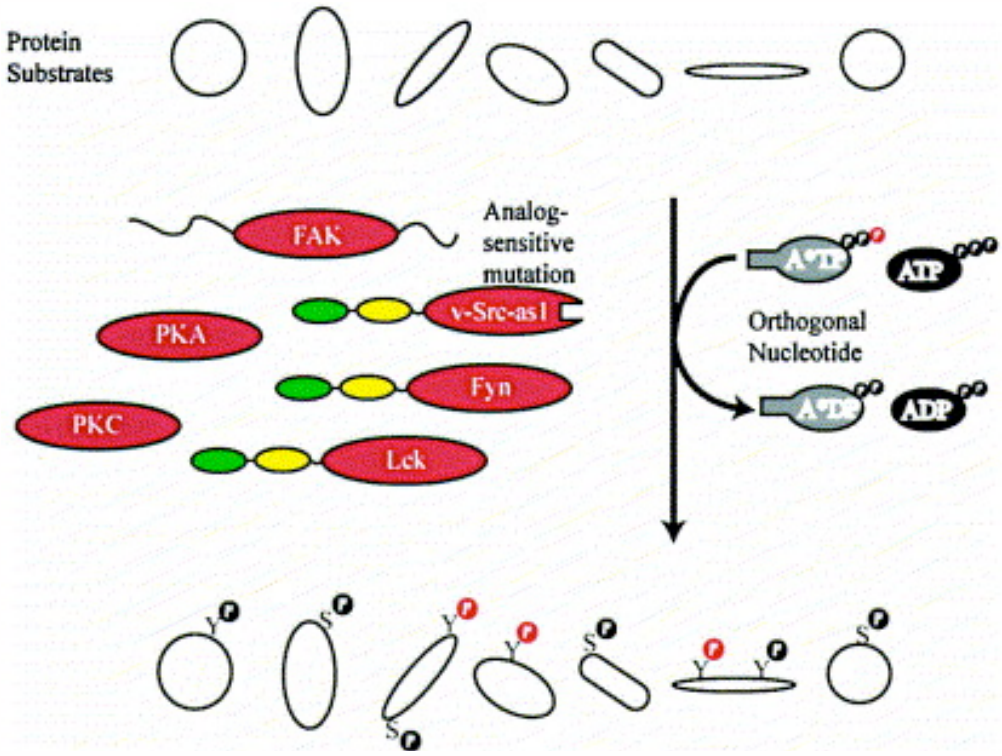
Table 2 Sequence and size of affinity tags

Tag	Residues	Sequence	Size (kDa)
Poly-Arg	5–6 (usually 5)	RRRRR	0.80
Poly-His	2–10 (usually 6)	HHHHHH	0.84
FLAG	8	DYKDDDDK	1.01
Strep-tag II	8	WSHPQFEK	1.06
c-myc	11	EQKLISEEDL	1.20
S-	15	KETAAAKFERQHMDS	1.75
HAT-	19	KDHLIHNVHKEFHAHAHNK	2.31
3x FLAG	22	DYKDHDGDYKDHDIDYKDDDDK	2.73
Calmodulin-binding peptide	26	KRRWKKNFIAVSAANRFKKISSSGAL	2.96
Cellulose-binding domains	27–189	Domains	3.00–20.00
SBP	38	MDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP	4.03
Chitin-binding domain	51	TNPGVSAWQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEPSNVPALWQLQ	5.59
Glutathione S-transferase	211	Protein	26.00
Maltose-binding protein	396	Protein	40.00



- 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.
- *Q. Do we need to remember the most frequently used Tags' name, elution condition and the specific sequence that they could recognize?*
- *A. Yes, you should have a rough idea about the most frequently used tags. For example, poly-His, FLAG, Glutathione S-transferase, and maltose binding protein. You should also know the important characteristics of the matrix and elution conditions. I think that these should be fairly self-evident (i.e., maltose binding protein can be eluted with excess maltose).*

A Chemical Biology/Molecular Biology approach to determining v-Src substrates



Ulrich et al., Tetrahedron, Volume 56, Issue 48, 2000, Pages 9495–9502

- Key reference: Shah, K., Liu, Y., Deirmengian, C. & Shokat, K. M. Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. *Proc. Natl Acad. Sci. USA* 94, 3565–3570 (1997).
- Kevan Shokat of UCSF devised one of the most clever ways of determining the physiologically relevent substrates of a specific kinase in a living cell.
- The challenge is that there are many kinases in one cell, adding phosphate groups to many different proteins. How do you know which proteins are being phosphorylated by one specific kinase? The key is to make something different about the phosphate groups added by the kinase of interest. Specifically, to make them radioactive using ^{32}P. If you can achieve this, and then separate the proteins (using methods we will learn about later), you will find the natural substrates for the kinase since they will be the only proteins with radioactive labels.
- Of course, if you just add radioactive ATP to the cell, all of the kinases will use it and all of the phosphorylated proteins will be radioactive. They trick is to engineer the kinase of interest to use a modified version of ATP that can not be used by other kinases (and has a radioactive label)

A v-Src variant that uses uniquely employs an ATP analog (with a ^{32}P)

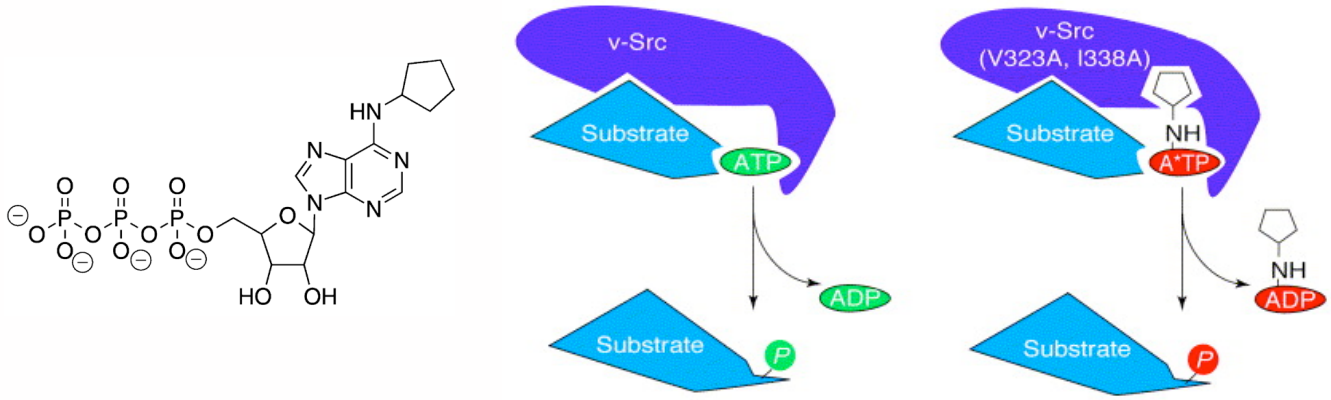


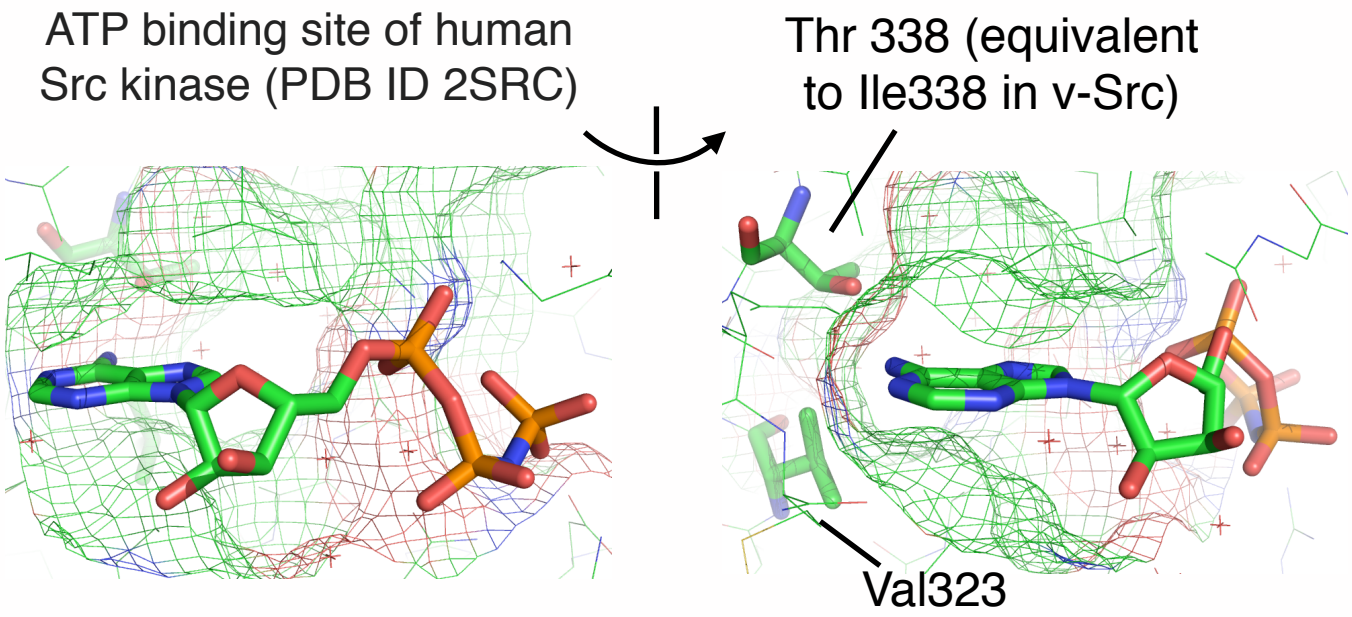
Table 1. GST–XD4 and GST–XD4(V323A, I338A) catalyzed RR-Src peptide phosphorylation kinetic constants

	GST-XD4			GST-XD4 (V323A, I338A)		
Nucleotide	k_{cat}, min^{-1}	K_m, μM	k_{cat}/K_m, $min^{-1}\cdot M^{-1}$	k_{cat}, min^{-1}	K_m, μM	k_{cat}/K_m, $min^{-1}\cdot M^{-1}$
ATP	2 ± 0.5	12 ± 3	1.6×10^5	0.8 ± 0.2	150 ± 20	5.3×10^3
N^6-(cyclopentyl) ATP		>2000 (K_I)		$(5 \pm 2) \times 10^{-2}$	15 ± 3	3.3×10^3

Ulrich et al., Tetrahedron, Volume 56, Issue 48, 2000, Pages 9495–9502
Crews and Splittgerber, Trends in Bioch. Sci. Volume 24, Issue 8, 1 August 1999, Pages 317–320
Witucki et al. Chemistry and Biology, Volume 9, Issue 1, January 2002, Pages 25–33

- Shokat first demonstrated this approach with v-Src in the late 1990s. He first synthesized a modified ATP that had a bulky cyclopentyl group attached to N6 of the adenine ring. None of the normal kinases in the cell can recognize this as a substrate since the bulky group does not fit into the ATP binding site. Note that all kinases are very similar in structure, so it is not surprising that this is a universal feature.
- Shokat then engineered v-Src to have a hydrophobic cavity adjacent to the N6 position of the adenine ring. He did this by mutating two residues that line the binding site into different residues with smaller side chains.
- After testing a variety of analogs of ATP, Shokat found that ATP with a cyclopentyl group attached to N6 was not a substrate for normal v-Src (called XD4 in the table), but could be used by v-Src V323A, I338A.

Engineering v-Src to accept unnatural ATP analogous with a 'bump' on N6



The V323A plus I338A mutations in v-Src create a new hydrophobic pocket near N6 of the adenine base

Xu et al. (1999) Mol.Cell **3:** 629-638

- Here's a closeup of the ATP binding site of Src. This structure is actually of the human enzyme, but it is very similar to that of v-Src.

Molecular biology procedures for making v-Src V323A, I338A

Protein Expression and Purification. Overlap extension PCR was used to make glutathione *S*-transferase (GST)–XD4(V323A, I338A) (23). Pfu polymerase (Stratagene) was used in the PCRs according to the manufacturer's protocol. Six synthetic oligonucleotides were used: primer 1 (5′-TTT*GATCC*ATGGGGAGTAGCAAGAGCAAG); primer 2 (5′-TTT*GAATTC*CTACTCAGCGACCTCCAACAC); primer 3 (5′-TGAGAAGCTGG**C**TCAACTGTACGCAG); primer 4 (5′-CTGCGTACAGTTGA**G**CCAGCTTCTCA); primer 5 (5′-CTACATCGTC**GC**TGAGTACATGAG); and primer 6 (5′-CTCATGTACTCA**GC**GACGATGTAG).

Primer 1 contains a *Bam*HI site, and primer 2 contains an *Eco*RI site (italics). Primers 3 and 4 contain the nucleotide sequence changes to introduce the V323A mutation (mutations shown in bold). Primers 5 and 6 contain the I338A mismatch. The XD4 gene from YEp51–XD4 plasmid (a gift of B. Cochran, Tufts Medical School) was amplified with primers 1 and 2. The PCR product was digested with *Bam*HI and *Eco*RI and ligated into similarly digested pGEX–KT and then transformed into the *Escherichia coli* strain DH5α. The GST–XD4(V323A) was constructed using primers 1, 2, 3, and 4 with the GST–XD4 plasmid as the template. The resulting PCR product was digested with *Bam*HI and *Eco*RI, ligated into similarly digested pGEX–KT, and transformed into DH5α *E. coli* cells. GST–XD4(V323A, I338A) was made in the same manner using primers 5 and 6 with GST–XD4(V323A) as the template. Expression and purification of kinases were carried out in DH5α as described by Xu *et al.* (24), with the exception that the cells were stored at 4°C overnight before centrifugation and lysis by French press (overnight storage is essential for producing highly active kinases).

Primer 1 (start of gene + BamH1)

```
tttggatccatggggagtagcaagagcaag
          M  G  S  S  K  S  K
```

Primer 2 (end of gene + EcoRI; reverse complement shown here)

```
gtgttggaggtcgctgagtaggaattcaaa
 V  L  E  V  A  E  -
```

Primer 3 (V323A forward)

```
          gtt (encode V)
tgagaagctggctcaactgtacgcag
  E  K  L  A  Q  L  Y  A
```

Primer 4 = reverse complement of 3

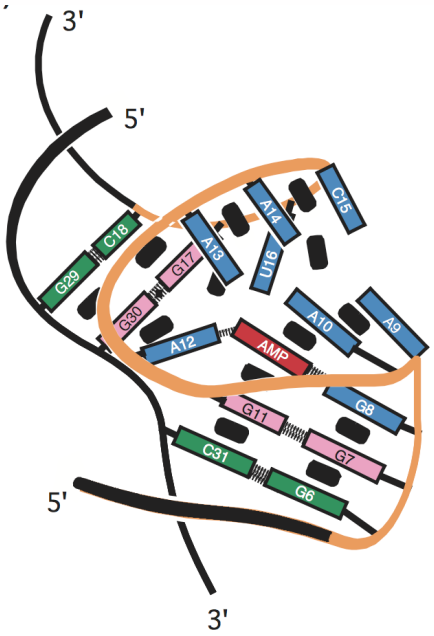
Primer 5 (I338A forward)

```
          att (encodes I)
ctacatcgtcgctgagtacatgag
  Y  I  V  A  E  Y  M
```

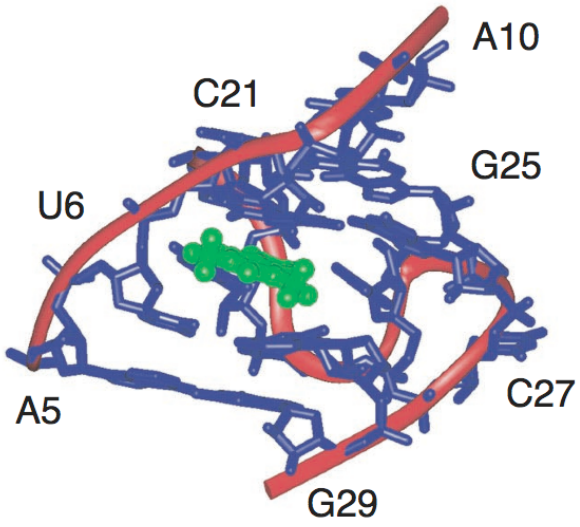
Primer 6 = reverse complement of 5

- You should be able to understand the molecular biology protocol provided in the original paper.
- Briefly, the authors performed overlap extension to introduce the mutations. They had to do the twice, since just one mutation can be inserted in one round of overlap extension. In principle, they could have done a Quikchange procedure with these primers, but this was right around the time when this procedure was first introduced (~1996).
- The authors inserted the mutated gene into the pGEX-KT plasmid which allows inserted proteins to expressed in frame with glutathione-S-transferase. This would enable convenient affinity purification.

RNA aptamers are highly folded structures with distinct binding pockets for their ligands



ATP/AMP RNA aptamer with K_d = 1 μM



Theophylline RNA aptamer with K_d = 0.1 μM

- 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
  - could induce an immune response
- 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.

Applications of molecular biology: systematic evolution of ligands by exponential enrichment (SELEX)

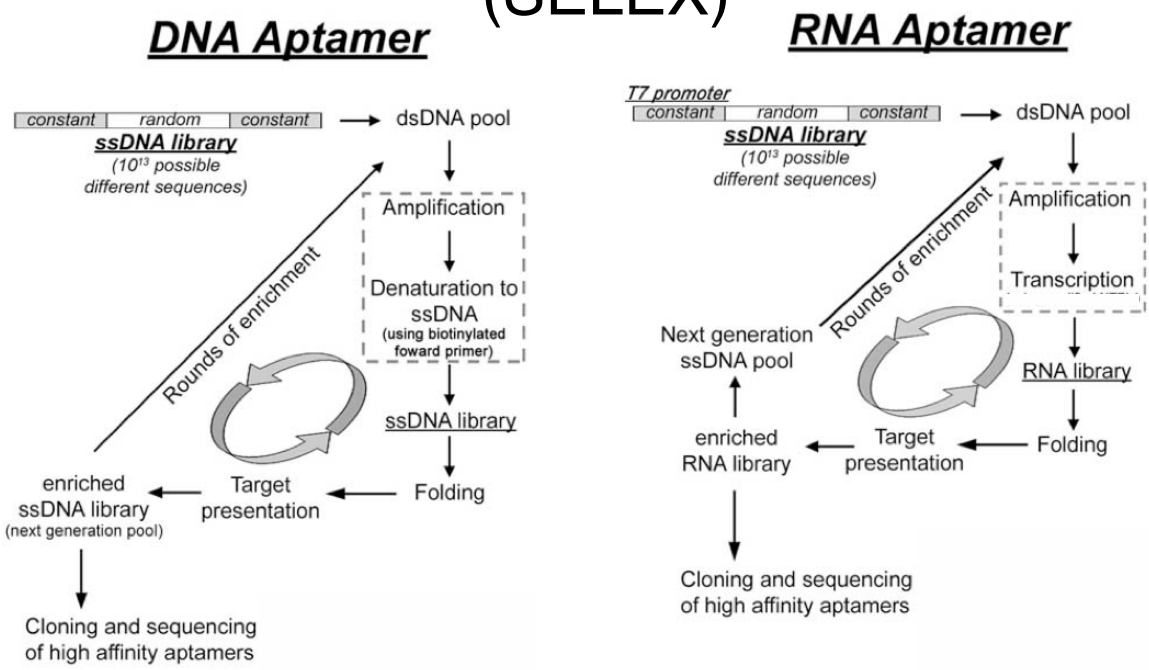


Figure from: Ulrich et al. *Combinatorial Chemistry & High Throughput Screening*, 2006, ***9***, 619-632.

- 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 denaturated 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: David S. Wilson and Jack W. Szostak, Annu. Rev. Biochem. 1999. 68: 611–647.
- *Q. At the end of the 2 cycle, what do you mean by "cloning and seqeuncing of high affinity aptamers"?*
- *A. This means that the dsDNA encoding the high affinity aptamers could be ligated into a plasmid. This would have the advantage of being a convenient way to store the sequence and distribute it to other labs. It would also make it easier to perform DNA sequencing to figure out what the actual sequence of the aptamer is. We will learn more about DNA sequencing towards the end of the course.*
- *Q.What does a biotinylated primer do in SELEX, since only the unbiotinylated sense strand is used in next cycle.*
- *A. It is critical to obtain ssDNA for this application, and the easiest way to separate the strands and keep only one is to use one biotinylated primer. Biotin binds very strongly to streptavidin coated beads. The dsDNA can be attached to the beads and then heated up to melt the duplexes. The DNA that elutes will be single stranded and can be used during the target presentation step.*
- *Q. There is one thing I don't quite understand about DNA aptamer. In the slide, it states biotinylated forward primer is used. From my understanding, biotinylated forward primer will end up produce biotinylated sense strand. However, only unbiotinylated sense strand can be purified and used for the next SELEX cycle. I don't understand how can unbiotinylated sense strand be produced if biotinylated forward primer is used. Can you please explain to this?*
- *A. By choosing whether the forward or reverse primer is the one with the biotin, you could choose which strand is the one that is eluted and used for target presentation. That is, the strand that is used for target presentation is the reverse complement of the one formed from the biotinylated primer. As long as the same primers are used in every round of selection, there should be no problems about getting the 'wrong strand'.*

phage display

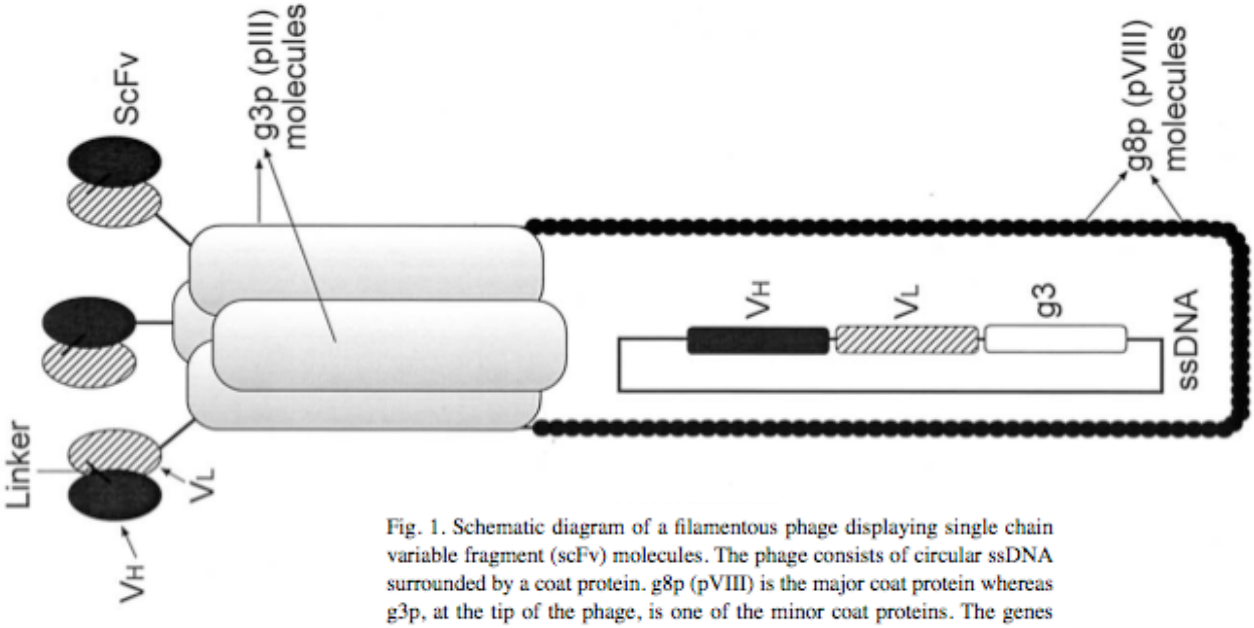


Fig. 1. Schematic diagram of a filamentous phage displaying single chain variable fragment (scFv) molecules. The phage consists of circular ssDNA surrounded by a coat protein. g8p (pVIII) is the major coat protein whereas g3p, at the tip of the phage, is one of the minor coat proteins. The genes encoding the variable domains of the scFv and a linker are fused to gene III (g3) in the genome of the filamentous phage. Consequently, the scFv is displayed as a fusion to g3p (pIII) protein at the tip of the phage. In reality, the scFv is not fused to all g3p protein molecules, and therefore the phage retains its ability to infect bacteria. Four g3p molecules are illustrated in the figure, three of which display scFv molecules.

H.M.E. Azzazy, W.E. Highsmith Jr. / *Clinical Biochemistry* 35 (2002) 425– 445

- There are now a handful of in vitro methods for engineering antibodies with specific binding functions. These methods do not require that an animal be immunized. CDR diversity can be generated using synthetic portions of cDNA that are inserted into the gene encoding an ScFv.
- One of these techniques is known as phage display. A 'phage' or bacteriophage is a virus that infects bacteria. Accordingly, they are harmless to all animals including humans. Phage display uses a particular type of phage called a filamentous phage (typically M13). The reproduction of most viruses is more or less similar, so you can think about the life cycle of M13 as being similar to that of HIV. One important difference is that the circular genome of M13 does not need to incorporate into the genomic DNA of the bacteria host. It is kept separate inside the bacteria.
- The basic principle of phage display is that the gene library (encoding variants of a particular protein) is inserted into a plasmid such that it is fused in frame to a phage coat protein. This plasmid gene library is used to transform *E. coli*. The *E. coli* is infected with a phage that will reproduce inside the bacteria making more copies of itself. These new copies will do two important things:
    1. They will package the plasmid (actually called a phagemid) inside their filamentous structures
    2. They will incorporate some of the phage coat proteins that are fused to the variant protein (which is in turn encoded by the phagemid packaged inside the phage).
- A typical library size is ~10^8 different phage displaying different proteins. The phage that display a variant that happens to bind to the target can be enriched in a process called 'bio-panning'.
- This processes can be repeated several times in order to dramatically enrich for binding domains with the desired specificity
- For those that are interested, see also the related techniques of ribosome display and mRNA display. (Binz, H. K., Amstutz, P., and Plückthun, A. (2005). Engineering novel binding proteins from nonimmunoglobulin domains. Nat. Biotechnol. 23, 1257-1268.)
- *Q: The VH and VL gene are inserted into a plasmid, and then the plasmid makes replication of a ssDNA, which will be packaged in the phage. I wonder how is the phage protein gene g3 inserted into the plasmid containing the genes for VL and VH as shown in the Figure?*
- *A: It's actually the other way around. The VL and VH genes (actually an ScFv gene that contains both) are inserted into the phage genome, which is itself a circular piece of DNA (sometimes referred to as a phagemid). The phagemid is sort of like a plasmid, but there are some key differences. The main one is that it is single stranded and not double stranded. The phagemid is produced inside of E. coli and packaged into the virus when it is being produced in the bacteria using the bacteria's own transcriptional and translational machinery.*

phage display

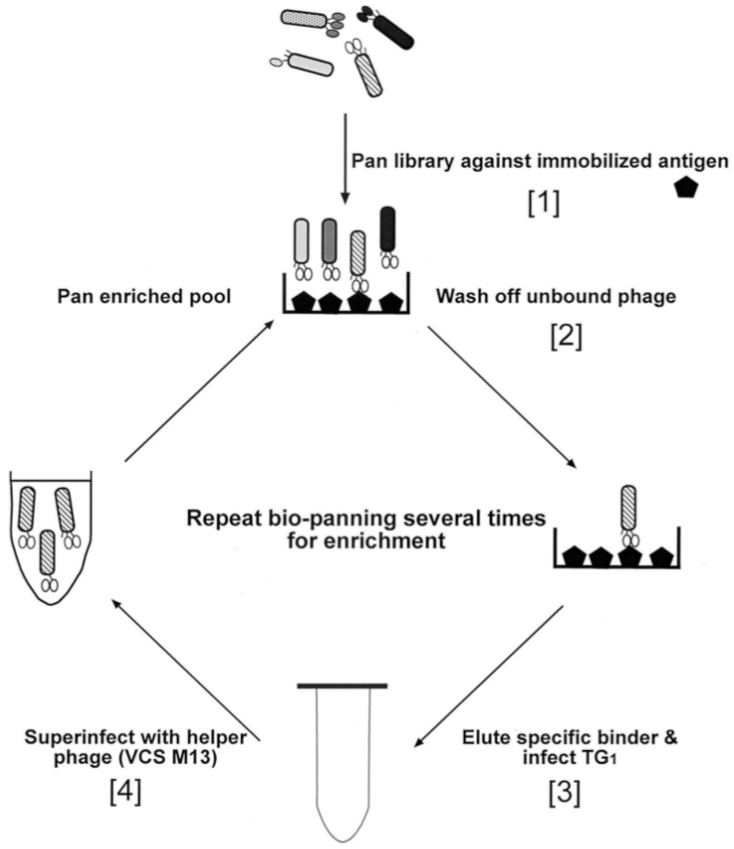
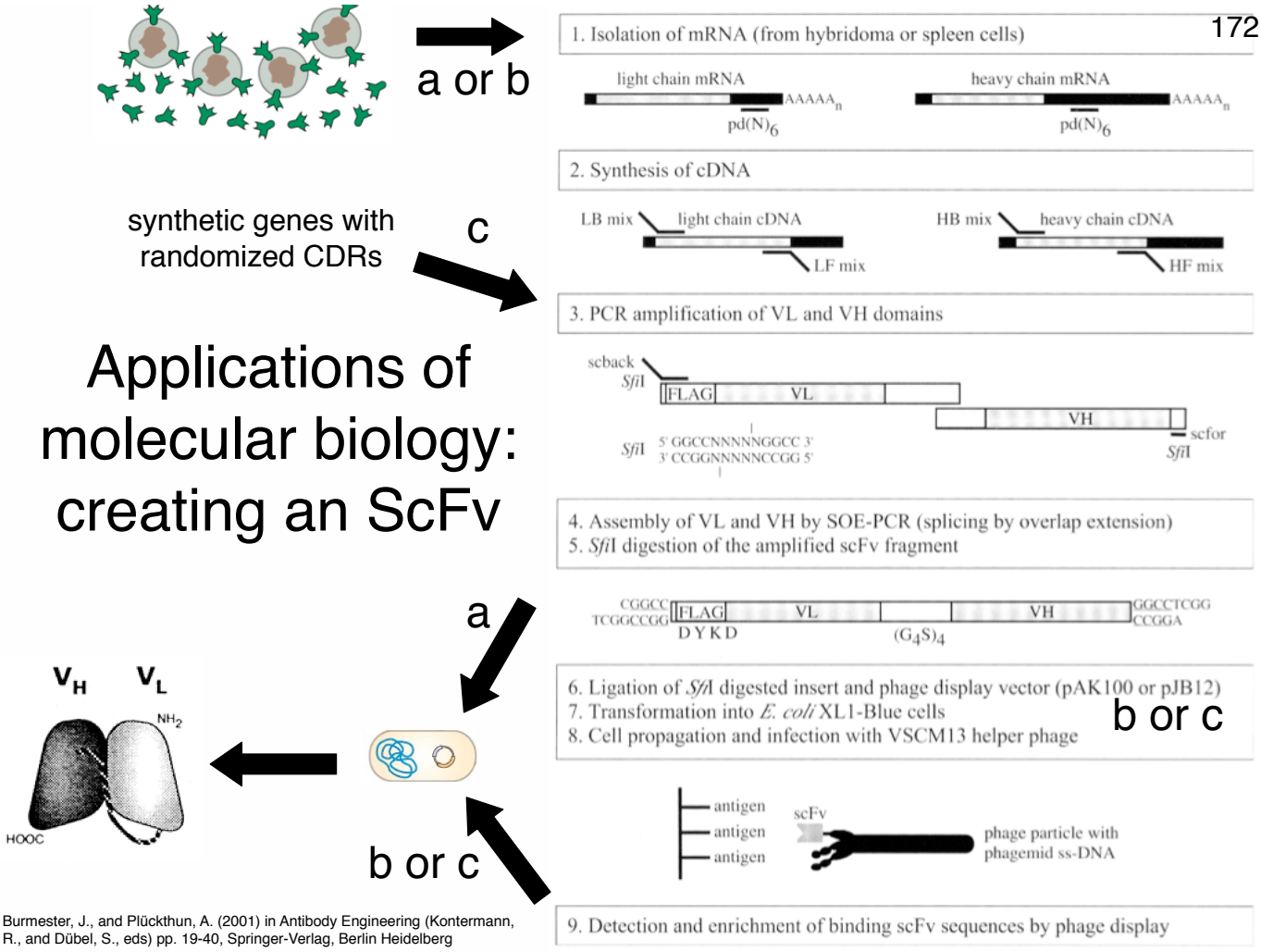


Fig. 4. Bio-panning of phage-display library. The library is screened in four steps: (1) binding of phages to the target antigen (in this case antigen immobilized on a solid support), (2) washing to remove unbound phage, (3) dissociation to recover antigen-specific phage, and (4) amplification of the antigen-specific phage by infection of *E. coli*. Caution must be exercised during the binding and washing steps to avoid low-ionic strength or other conditions that may favor adsorption of phage directly onto plastic or other matrix.

H.M.E. Azzazy, W.E. Highsmith Jr. / *Clinical Biochemistry* 35 (2002) 425– 445

- *Q: In the example of inserting the scfv genes into a phage, how do we go from dsDNA from the library to ssDNA in phage.*
- *A: The trick is that that there is a site in the plasmid called an f1 origin of replication. DNA polymerase can start from this site and just proceed around the plasmid in one direction. This is a so-called 'rolling circle' amplification where DNA polymerase just keeps going around and around making one new new strand of DNA. This single strand of DNA can be cut at appropriate sites and circularized to produce the circular ssDNA that gets packaged into the phage.*
- *Q: I am just wondering that is phage display used to generate human antibody? We can not inject antigen into human body, right? If so, how can we get the biased library that you mentioned for genes of scfv. And what do you mean by cycles for enrichment? Does that means we should repeat all the steps(from prepare gene library) or only repeat affinity chromatography?*
- *A: It's true that you couldn't normally make a biased library of human ScFvs. Though you could imagine a situation where an individual happens to have acquired HIV but fortunately developed a particularly powerful antibody against the virus. In principle, you could take some B cells from that person and then isolate the antibody by phage display. Generally speaking, the biased library would be for non-human antibodies.The cycles of enrichment mean taking the phage isolated after affinity purification, reinfecting E. coli, harvesting the phage, and then repeating the affinity purification. In principle, you could introduce a bit more randomization of the gene with each cycle to create a new library based on the best variants from the previous round. These cycles need to be done because the process is not perfect (i.e., non-specific binding of phage) and you are trying to discover a potentially very rare variant (1 in a million). By repeating these cycles you are enriching for the variant with the desired binding, so it will make up a large percentage of the isolated phage.*

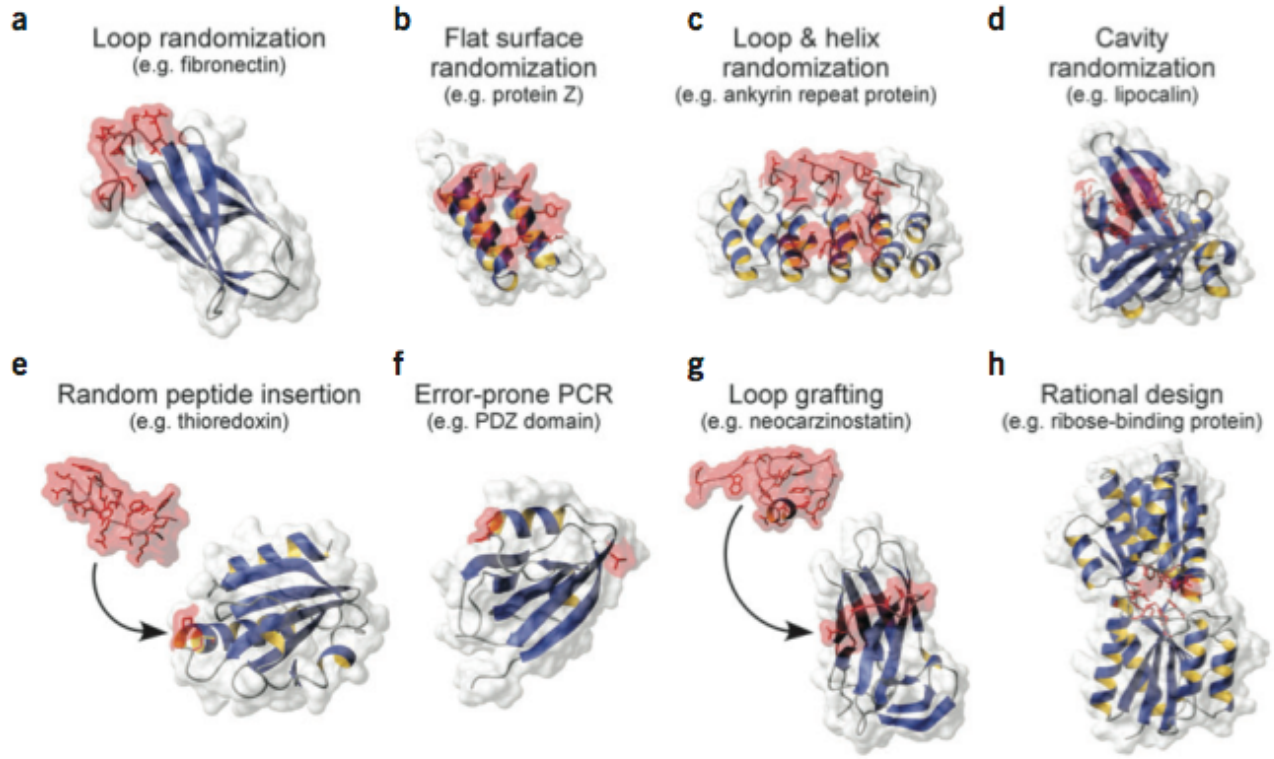
Applications of molecular biology: creating an ScFv



Burmester, J., and Plückthun, A. (2001) in Antibody Engineering (Kontermann, R., and Dübel, S., eds) pp. 19-40, Springer-Verlag, Berlin Heidelberg

- Three scenarios:
  - a) starting with hybridoma expressing monoclonal antibody
  - b) starting with a population of B-cells
  - c) starting with synthetic gene diversity
- *Q: During the enrichment, a bacteria will be infected with the specific phage, does the phage just use the bacteria's tools to manufacture more phage and when that is done it will make slightly different copies of itself which could be used to find a phage with a stronger affinity of the specific antigen? How is the phage eventually released from the bacteria?*
- *A: Just like HIV provirus needs to use the transcription and translation machinery of a human cell to make more copies of itself, a phage needs to use the bacteria's machinery to make more copies of itself. Generally speaking the new phage that are produced will be exact copies of the original phage that infected the bacteria, since all of the normal error-checking/ proofreading mechanisms of the bacteria are at work here. To do something like somatic hypermutation with phage display, the experimenter would need to collect the DNA and purposely introduce some random errors. This could be done by doing a PCR step on the DNA isolated from phage under conditions where the polymerase makes more mistakes than usual (known as error prone PCR). Phage can be grouped into two types: lytic and non-lytic. Lytic means that the host cell bursts and virus particles are released. The type of phage used in phage display are non-lytic, which means they can get out the cell without bursting it. They do this by assembling their coat protein shell in the membrane and sort of extruding themselves out of the cell.*
- *Q: Could you explain what you mean by synthetic genes with randomized CDRs that is used for creating scFV?*
- *A: Rather than PCR a gene from DNA (made by reverse transcription of RNA), it is now possible to chemically synthesize a gene containing up to many hundreds of base pairs. You could easily synthesize a gene for a ScFv in which some subset of the codons, such as those that occur in the CDRs, were randomized. This would provide you with a gene library that you could screen by phage display. There are many companies that do this now, but one of the higher profile ones is Blue Heron Biotech. (http://www.blueheronbio.com/).*

non-immunoglobulin scaffolds enable different types of binding sites to be exploited



- But if we can do phage display, why use antibodies at all?
- We can use many different types of proteins as the scaffold for binding.
- Some will have advantages such as being small relative to antibodies (may be able to access sterically hindered epitopes). Some may fold in cytoplasm (no disulfide bonds)
- However, they will not be as well suited to therapeutic applications.
- Figure Legend. *"Binding-site engineering strategies used with different alternative scaffolds. (a –h) In combinatorial engineering approaches, sequences of a scaffold can be diversified at specified positions by means of defined randomized codons (e.g., in loops (a), flat surfaces (b), combinations of loops and helices (c), or cavities (d)), or a random peptide sequence is inserted into the scaffold (e), usually at a loop, or the scaffold sequence is randomized at undefined positions (f). Target-binding variants of the resulting libraries are subsequently isolated using selection or screening technologies. In rational engineering approaches, preexisting binding sequences (e.g. loops) have been grafted onto a novel scaffold (g), or binding sites have been engineered de novo into a suitable scaffold (h). The different engineering possibilities are illustrated by alternative binding molecules where the engineering in question has been applied: loop randomization (fibronectin), flat surface randomization (protein Z), loop and helix randomization (ankyrin repeat protein), cavity randomization (lipocalin), random peptide insertion (thioredoxin), error-prone PCR (PDZ domain), loop grafting (neocarzinostatin) and rational design (ribose-binding protein). Many other permutations of randomization strategies and scaffolds are conceivable; this figure illustrates each strategy with one published example."* From: Binz, H. K., Amstutz, P., and Plückthun, A. (2005). Engineering novel binding proteins from nonimmunoglobulin domains. Nat. Biotechnol. 23, 1257-1268.