BIO450 Primer Design Tutorial

The most critical step in your PCR experiment will be designing your oligonucleotide primers. Poor primers could result in little or even no PCR product. Alternatively, they could amplify many unwanted DNA fragments. Either way, it would interfere in subsequence cloning steps. Therefore it is critical that you design your primers carefully. Primer design requires extensive computer-based sequence analysis and this tutorial is designed to lead you through that analysis.

General Considerations in Primer Design

Specificity

PCR is capable of amplifying a single target DNA fragment out of a complex mixture of DNA. This ability depends on the specificity of the primers. Primers are short single-stranded oligonucleotides which anneal to template DNA and serve as a "primer" for DNA synthesis. In order to achieve the geometric amplification of a DNA fragment, there must be two primers, one flanking each end of the target DNA. It is essential that the primers have a sequence that is complementary to the target DNA. Consider the following fragment of DNA. (Remember, that when both strands of DNA are shown the top strand runs 5'-3') The location of two primers is indicated by >>'s.

```
 541 CATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTAC
     GTAATAGTTGTTTTATGAGGTTAACCGCTACCGGGACAGGAAAATGGTCTGTTGGTAATG

 601 CTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTT
     GACAGGTGTGTTAGACGGGAAAGCTTTCTAGGGTTGCTTTTCTCTCTGGTGTACCAGGAA
     >>>>>>>>>>>>>>>>>>

 661 CTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATAA
     GAACTCAAACATTGTCGACGACCCTAATGTGTACCGTACCTACTTGATATGTTTATT
                         <<<<<<<<<<<<<<<<<<<<
```

The forward primer (>>>>>) which will be complementary to the lower strand and must run 5'-3' will have the sequence: 5'-CTGTCCACACAATCTGCC -3'

The reverse primer (<<<<<) which will be complementary to the upper strand and must run 3'-5' and will have the sequence: 3'-GACCCTAATGTGTACCGTAC-5'. However, we always write DNA sequences in the 5'-3' direction so the reverse primer would be written: 5'-CATGCCATGTGTAATCCCAG-3'

Ideally primers would be complementary only to the target sequence. This would ensure that the taq polymerase only copies the target region. However, when using genomic DNA as the template, there will likely be other sequences which are complementary or nearly complementary to the primers. A well designed primer will only be complementary to a few regions besides to target sequence. This secondary binding will not interfere with PCR because if a primer is complementary to an isolated sequence of DNA it will not lead to amplification of that region. Geometric amplification only occurs in those regions in which the annealed primers face one another on opposite strands. However if a primer is designed that is complementary to some repetitive sequence in the genome there is a much greater chance the primers will anneal to two

regions that are complementary to each other resulting in the amplification of non-targeted sequence.

Two critical issues for specificity:

1. Primers must be complementary to flanking sequences of target region
2. Primers should not be complementary to many non-target regions of genome.

Melting Temperature (T_m)

The annealing temperature for a PCR reaction is based on the melting temperature (T_m) of the primers. The T_m is the temperature at which a population of a double stranded DNA molecule is partially denatured such that half of the molecules are in the single stranded state and half are in the double stranded state. At temperatures above the T_m the DNA molecules will be in the single stranded form; at temperatures below the T_m the DNA can form the double stranded form.

In order for the primers to anneal to the target DNA the annealing temperature must be below the T_m of the primers. Typically, the annealing reaction is carried out about 5º below the T_m. If the annealing temperature is too high, the primer will not anneal to the target DNA. If the annealing temperature is too low the primer will mis-anneal to sequences which aren't perfectly complementary. The most important consequence of this is that the two primers designed for a PCR experiment should have very similar T_m's. Typically the T_m should be within 5º of each other. The closer the T_m's the better.

The T_m of a molecule is dependent on its sequence, however the relationship between sequence and T_m is not simple. In general the greater the GC content of DNA the higher its T_m. The Wallace formula can be used to give a rough estimate of T_m.

$$T_m = 2(A+T) + 4(G+C).$$

There are other formulas for estimating T_m such as "nearest neighbor" which may give a more accurate measure of T_m. Additionally, there are a number of web sites that will calculate T_m using these or other formulas.

Two issues are critical for T_m.

1. The two primers should have a similar T_m.
2. The T_m should be within55-72º, around 60º is ideal.

Primer Length

Primer length considerations fit the Goldilocks paradigm; they must be neither too short nor too long. If primers are too short they will lack specificity. For example consider a primer only 4 nucleotides long, GATC. Although this primer may bind to a flanking region of a target sequence, it will also bind to thousands of other sequences on the chromosomes. This could lead to amplification of unwanted sequences. Alternatively, if primers are too long, this affects the rate of annealing. Annealing efficiency is proportional to primer length. Therefore very long primers will not anneal efficiently and this will lead to a reduction in the amount of PCR product produced.

For standard PCR, an oligonucleotide of 18-24 nucleotides is ideal. It is long enough to be specific to the target region, yet short enough to anneal efficiently.

Product Size

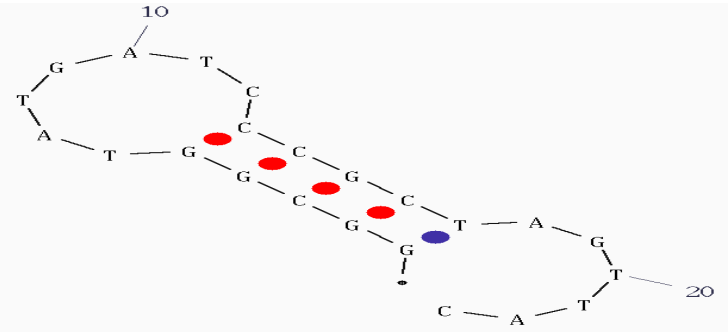
The choice of primers determines the size of the PCR product. If the two primers are complementary to nearby regions on the template DNA, then a small fragment of DNA will be amplified. If the two primers are complementary to regions farther apart, then a larger fragment of DNA will be amplified. Basic taq polymerase can easily amplify fragments up to 1000 to 2000bp. (Special polymerases can be used to amplify larger fragments.) For standard PCR, the primers should be complmentary to regions on the target DNA within 1000bp of each other.

Primer Dimers

If the primers have self-complementary sequences the primers, which are in high concentration, will anneal with themselves. If they anneal with themselves they are not available to bind to the target DNA. There are two types of potential self-complementary sequences, those that lead to hairpins and those that lead to primer dimers.

Hairpins

Intramolecular complementary sequences can lead to base pairing within a molecule. Consider the primer **GGC GGT ATG ATC CCG CTA GTT AC.** It can base pair internally and form the following hairpin structure. A primer that is base pairing with itself cannot base pair with its target DNA. Primers must be designed to minimize intramoleculari base pairing. Intramolecular base pairing is usually analyzed using computer programs. Avoid primers that contain more than a string of 3 intra molecular base pairs.



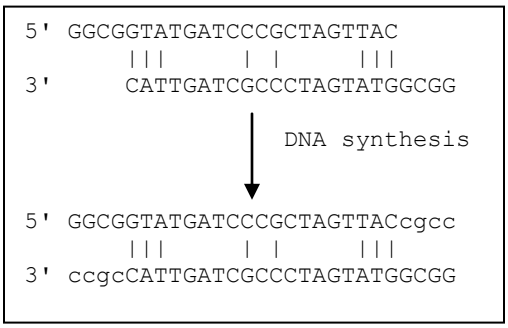
Primer Dimers

Primers can also participate in intermolecule base pairing. This is base pairing between two different primer molecules. If the base pairing is between the forward and the reverse primer it is called heterodimer formation. If the base pairing is between just one of the two primers it is called self-dimer formation.

The example primer used above can form several self-dimers (see the 2 boxes below). Both examples of primer dimer are problematic. The first is a highly stable structure with numerous base pairs. If the primers are base pairing with themselves they cannot base pair with the target DNA.

```
5'       GGCGGTATGATCCCGCTAGTTAC
          |||| || || ||||
3' CATTGATCGCCCTAGTATGGCGG
```

The second is less stable, but still problematic. The 3' end of the primer is involved in the base pairing. When this happens the second primer can be used as a template for DNA synthesis. The addition of nucleotides to the primers will prevent them from base pairing with the target DNA. This may be the single most common problem with PCR reactions.



G/C Content

As discussed for melting temperature, it is important that primers be about 50% G/C and 50% A/T. It is also important that regions within the primer not have long runs of G/C or A/T. A stretch of A/T's might only weakly base pair while a stretch of G/C might promote mis-annealing. It is also useful to avoid a long string of a single nucleotide or even long strings of purines or pyrimidines.

G/C clamp

Stable base pairing of the 3' end of a primer and the target DNA is necessary for efficient DNA synthesis. To ensure the stability of this interaction, primers are often designed ending in either a G or a C. (GC base pairs are more stable than AT base pairs.) This terminal G or C is called a G/C clamp.

Summary

1. primers should be 17-28 bases in length;
2. base composition should be 50-60% (G+C);
3. primers should end (3') in a G or C, or CG or GC: this prevents "breathing" of ends and increases efficiency of priming;
4. Tms between 55-80^{o}C are preferred;
5. primer self-complementarity (ability to form 2^{o} structures such as hairpins or primer dimers) should be avoided;
6. it is especially important that the 3'-ends of primers should not be complementary (ie. base pair), as otherwise primer dimers will be synthesised preferentially to any other product;
7. runs of three or more Cs or Gs at the 3'-ends of primers may promote mispriming at G or C-rich sequences (because of stability of annealing), and should be avoided.

Web Based Tools for Primer Design

This semester we will be using two different internet applications for our primer design. The first is an application in Biology Workbench called Primer3. This application will analyze target regions and recommend forward and reverse primer sequences. Its analysis can be directed to specific target regions of genes and analyzes factors such as product size, primer size, tm, GC content GC clamps and dimer formation. The second application, oligocalc, <http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx> is provided by one of the companies that we order primers, IDT. It has applications for analysis of hairpins, homodimers and heterodimers. We will use it to double check the primers identified by Primer3.

How to find a primer

1. Access biology workbench and import the region of your gene that you plan to amplify using PCR.
2. Run primer3 program in workbench.
3. Scroll down to primer criteria on the primer 3 page and change the first two default settings.
    a. Under product size change range from 100-300 to 400-600.
    b. Change the GC clamp size from its setting of zero to a setting of one.
4. Click Submit to complete the analysis.
5. Primer3's output includes an "optimal" pair of primers. The locations of these primer sequences on the target sequence are reported. Four pairs of alternative primers are also reported.

Checking primers with oligocalc.

1. Open a second browser window. Go to the oligocalc site.
2. Paste the first left primer in the sequence box on the upper left hand corner.
3. Click Hairpin button. An" mfold" box will appear below the sequence box. Click "Calculate" (or "submit") on the mfold box to run analysis.
4. An mfold output box will appear. Scroll down that box to observe structures. Note free energy of structure and number of basepairs that support structure, and location of the 3' end. A good primer will have fewer than 5 base pairs of hairpin structure, a free energy -5kcal/mole and the 3'end will not be stabilized by base pairing.
5. Click the Self-Dimer button. A new window will open with the results. A good primer will have a free energy below – 5 kcal/mole, and the 3' end will not be stabilized by base pairing.
6. Repeat steps 4 and 5 for the complementary right primer.
7. Click the Heterodimer button. Paste the left primer into the "primary sequence box and the right primer into the secondary sequence box. Click Calculate.
8. A new window will open with the results. A good primer pair will have a free energy below – 5 kcal/mole, and the 3' end will not be stabilized by base pairing.
9. Evaluate all primer pairs using 1 through 8. Choose a primer that best meets criteria outlined.

Add T7 Promoter Sequence

1. Add the following sequence onto the 5' end of your best left primer. TAATACGACTCACTATAGGGAGA
2. Repeat analysis for hairpin, self-dimer and heterodimer (using your best right primer, nothing added). If t this sequence generates a problem, you may have to use one of your other primers.

Submit a Report on your primers including the following elements

Name of gene you are studying.

The sequence of the following primers

1. Sequence of best left prime
2. Sequence of best left primer with T7 promoter added onto 5’ end
3. Sequence of best right primer

For each primer report

A. Primer Length
B. GC Content
C. GC Clamp
D. Tm
E. Primer Dimer
    a. Hairpin analysis
        i. Number of base pairs
        ii. Free energy
        iii. 3’ end involvement
    b. Self Dimer analysis

For the best left primer with T7 and best right primer

A. Heterodimer analysis
        i. Number of base pairs
        ii. Free energy
        iii. 3’ end involvement
B. Report size of product