**Primer Design for Targeted Amplicons**

To select a specific target from a genome requires that the amplification primers be very specific. Thus, the design of the primers is critical to the success of the experiment. Because PCR amplification requires 2 balanced chemical/enzymatic reactions to give a good yield, you must understand the factors that influence accuracy and rate. Because genomes are so large and are not random, you must also understand how to get the specificity you want, or you will get many side-reactions. Primer design requires that you employ computer-based sequence analysis: this tutorial leads you through one possible workflow in carrying out that analysis.

1. **General Considerations in Primer Design**

**Specificity**

PCR is capable of amplifying a single target DNA fragment out of a complex mixture of DNA, if the sequence contains enough unique elements. Identifying the unique elements is the first step in primer design for target specificity. Of course, you might want to amplify a family of genes, in which case you would aim for a different level of specificity. Primers are short single-stranded oligonucleotides (‘primers’) which hybridize (‘anneal’) to one strand of the target DNA (the ‘template’). Base-pairing complementarity leads to a short double-stranded region that serves as a site where a DNA polymerase can bind and begin to copy the template. In order to copy both strands of the target there must be an oligonucleotide primer for each strand – if they are at the right distance apart, the polymerase will be able to copy the complete region between and through the primer-binding sequences. Repeated rounds of such synthesis will achieve geometric amplification of a DNA fragment. The oligonucleotides must be complementary to the templates.

Consider the following fragment of dsDNA (the convention is that the top strand is the one that is 5’-3’). The position where a primer is located 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 they bind only to the intended template DNA, creating binding sites for a thermostable DNA polymerase only at the target region. However, complex mixtures of DNA , like genomic DNA likely have other sequences which are complementary or nearly complementary to the primers. So long as the secondary sites of the two primers are quite far apart or are on the same strand they will not lead to a competing amplification product. When a primer is complementary to a repeated sequence in the genome there is a much greater chance that a competing product will occur.

Two critical issues for specificity:

1. Primers must be complementary to flanking sequences of target region, on opposite strands and less than 2000bp apart (except where long-range PCR is the goal).
2. Primers should not be complementary to many non-target regions of genome – repeated regions and gene families in particular present design challenges.

**Melting Temperature (Tm)**

The *annealing* temperature for a PCR reaction is based on the melting temperature (Tm) of the primers bound to the templates. The Tm is the temperature at which a population of double stranded DNA molecules has half of the molecules in a single stranded state and half in the double stranded state. At temperatures above the Tm more of the DNA molecules will be single stranded form; at temperatures below the Tm more of the DNA is in the double stranded form.

Before the primer can hybridize to the template, the target DNA must be melted into single strands – so there is an initial high-temperature melting step. At the annealing temperature only half of the primer is hybridized to the template, so we usually choose a slightly lower Tm in order to get a better efficiency of annealing. However, if there is a secondary site with a single-base mismatch, we might choose a lower efficiency in order to achieve higher specificity. Typically, the annealing reaction is carried out about 5º below the Tm. If the annealing temperature is too high, the primer will not anneal to the template DNA at all. Since both primers have to be in the same mixture, they experience the same reaction conditions – thus the two primers must have very similar Tm s. Typically the Tm should be within 3-5º of each other. The closer the Tm’s the better.

The Tm of a molecule is dependent on its sequence, but base stacking and the number of hydrogen bonds affects the response, so the relationship is not simple.In general the greater the GC content of DNA the higher its Tm. Most current PCR design programs use the SantaLucia nearest-neighbor parameters to estimate the Tm. The length of the primers need not be identical, it is a balanced Tm. that matters. Generally, for cost reasons, primers are designed to be 17-25nt in length, although there are specific types of experiments that do not follow that rule.

Two issues are critical for Tm.

1. The two primers should have a similar Tm.
2. The Tm of most primers in the above length range will be within55-72º.

**Primer Length - comments**

Primers that are too short are much more likely to be non-specific. A primer only 4 nucleotides long will bind to thousands of sequences in a genome, leading to unwanted amplification. Primers that are too long will diffuse and bind more slowly, so the annealing step of the PCR reaction takes longer. Longer primers are also more likely to have internal structures, like hairpins, that make them less efficient as primers. At best this will lead to a smaller amount of PCR product.

As stated above, for standard PCR oligonucleotides of 17-25 nucleotides are generally chosen. .

**Product Size**

Where primers bind on the template determines the size of the final product. Primers complementary to nearby regions will yield a short amplicon and primers at a distance will yield longer amplicons Most thermostable polymerases are quite good at amplifying 1000-2000nt templates, and some can yield 15,000-30,000bp amplicons. For standard PCR chemistries and enzymes, the amplicon size is generally limited to ~1000bp. For real-time PCR the size is generally limited to <200bp. For longer amplicons the reaction conditions are changed, not the primer design parameters.

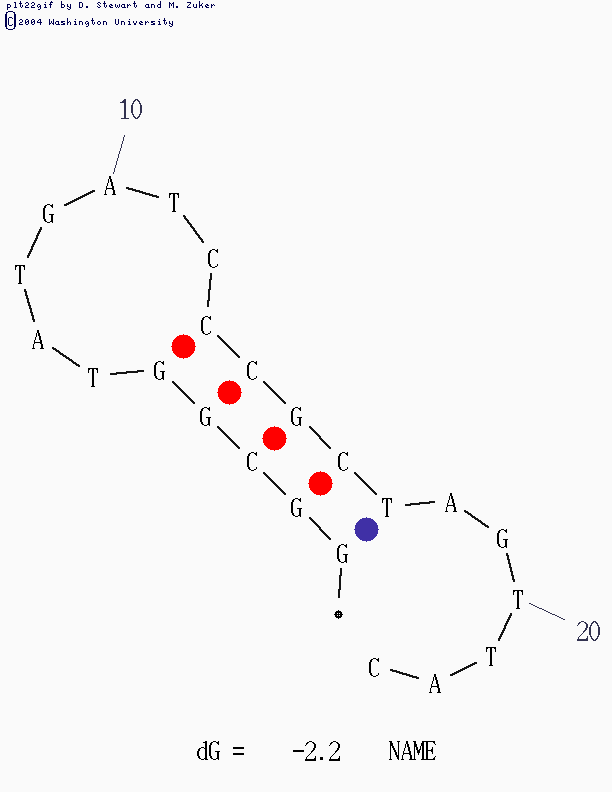
1. **Artifacts Eliminated in the Design Process**

**Primer Dimers**

If the primers have self-complementary sequences the primers, which initially are at quite high concentration, will self-hybridize or cross-hybridize – then they are not available to bind to the target DNA. Effectively, their concentration has been lowered for the desired reaction. There are two types of potential self-complementary sequences, those that lead to hairpins and those that lead to dimers.

Hairpins

*Intra*molecular complementary sequences can lead to base pairing within the same molecule. In the example below, the primer **GGC GGT ATG ATC CCG CTA GTT AC** is shown**.** It can base pair *internally*, forming a ‘hairpin’ structure. A primer that is base pairing with itself cannot base pair with its target DNA. The potential for intramolecular base pairing is usually analyzed using computer programs such as mfold. A rule of thumb is to avoid primers that form a hairpin of more than 6 base pairs.



Primer Dimers

Primers can also participate in *inter*molecular base pairing. This is base pairing between two different primer molecules. They may have the same sequence, or they may be the two different primers for the reaction, usually called the forward and reverse primers, the first being the complement to the top strand of the target and the second the complement to the bottom strand of the target. If the base pairing is between the forward and the reverse primer it is called *hetero*dimer formation. If the base pairing is between two molecules of the same primer it is called self-dimer or *homo*dimer formation.

The example primer shown above can form several types of self-dimers, shown below. The first example is quite stable, even though there are no runs longer than 4bp, there are several runs in close proximity.

5'       GGCGGTATGATCCCGCTAGTTAC

          |||| || || ||||

3' CATTGATCGCCCTAGTATGGCGG

The second example is less stable, but still problematic, mostly because 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, as shown. The addition of nucleotides to the primers will prevent them from base pairing with the target DNA. In both cases the interaction lowers the effective concentration of the primer available to the intended reaction.

5' GGCGGTATGATCCCGCTAGTTAC

       |||     | |     |||

3'     CATTGATCGCCCTAGTATGGCGG

DNA synthesis

5' GGCGGTATGATCCCGCTAGTTACcgcc

       |||     | |     |||

3' ccgcCATTGATCGCCCTAGTATGGCGG

**G/C Content**

In general one aims to design primers that are about 50% G/C and 50% A/T, but this is dependent on the genome and specific targets. However, high GC primers tend to be quite difficult to design without any internal structure. Long stretches of any one base, or of dinucleotides, often lead to mispriming.

**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 – however, too much stability can lead to mis-priming so you should avoid more than 2 such terminal nucleotides in the last 3. This terminus is called a G/C clamp.

1. **Target Considerations**

It should be noted that once the target has been denatured the template strands may also form secondary structures – if a primer binding site falls within a strong hairpin it may not have access to the template at the expected annealing temperature. Thus it is sometimes worthwhile to also fold the template, not just the primer. If there is a problem you can then adjust the primer binding site.

The target sequence against which you design primers is often a single examplar or the gene. Many genes have polymorphisms distributed across the sequence. There are two concerns: one is that the primer, having bound across a polymorphism, will effectively ‘remove’ it, so you will miss that data about existing variance. The second concern is that the primer may not bind as efficiently as you expect – if there are multiple gene sequences available it is often worthwhile comparing them to see where the common sites of variation are found, so you can avoid them in your design.

Finally, don’t forget that in eukaryotes the mRNA does not contain introns (usually) – most often you cannot use the same primer pairs to interrogate mRNA and gDNA.

**IV. Summary**

1. Primers should be 17-25 bases in length;
2. Base composition should be balanced, usually at 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 55-72oC are preferred and primer Tms should balance within 3-5°C;
5. Primer self-complementarity (ability to form 2o 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 so that the polymerase forms very short amplicon products with them.
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.
8. Synthetic chemistry note: it is difficult for the common synthetic oligonucleotide chemistries to efficiently produce more than 4 G’s in a row, so try to avoid such a sequence. This does not have to do with mis-priming, but with organic chemistry.

**Web Based Tools for Primer Design**

There are a number of good Web-based primer design tools. The oldest is called Primer 3. It can be found at a number of sites. One such is the Biology Workbench. The application allows you to cut and paste a desired gene target sequence, and specify certain conditions about the final amplicons, such as length, sequences that must be included. Etc. The application will suggest a number of forward and reverse primer pairs, and allow you to consider product size, primer size, Tm, GC content GC clamps and dimer formation. Another good site is provided by one of the larger oligonucleotide synthesis companies, Integrated DNA Technologies, or IDT. It has facilities for including the analysis of hairpins, homodimers and heterodimers.

One way to find a primer

1. Access Biology Workbench (<http://workbench.sdsc.edu/> Note: you have to register to get an account) and import the region of your gene that you plan to amplify using PCR.
2. Run the Primer3 program in Workbench.
3. Scroll down to primer criteria on the Primer 3 page and change the first two default settings.
   1. Under product size change range from 100-300 to 400-600.
   2. 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.
6. Note: another site where you do not have to register is here: <http://bioinfo.ut.ee/primer3-0.4.0/primer3/>

Checking primer structure with oligocalc.

1. Open a second browser window. Go to the OligoAnalyzer site (<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/> ) .
2. Paste the first left primer in the sequence box on the upper left hand corner.
3. Enter your PCR/hybridization conditions: generally you would use an oligo concentration of 0.1uM, 50mM Na+, 1.5-5mM Mg++, 0.2mM dNTPs.
4. Click Hairpin button. An” mfold” box will appear below the sequence box. Click “Calculate” (or “submit”) on the mfold box to run analysis.
5. An output box appears with structures (usually several, depending on the possible stable tructures that can form). 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.
6. 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.
7. Repeat steps 4 and 5 for the complementary right primer.
8. Click the Heterodimer button. Paste the left primer into the “primary sequence box and the right primer into the secondary sequence box. Click Calculate.
9. 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.
10. Evaluate all primer pairs; choose a primer that best meets criteria outlined.

Modified primers – making fusion amplicons

Note that if you are adding the adaptor sequence for one of the sequencing platforms you need to repeat the modeling process above to check for cross-hybridization and structures.