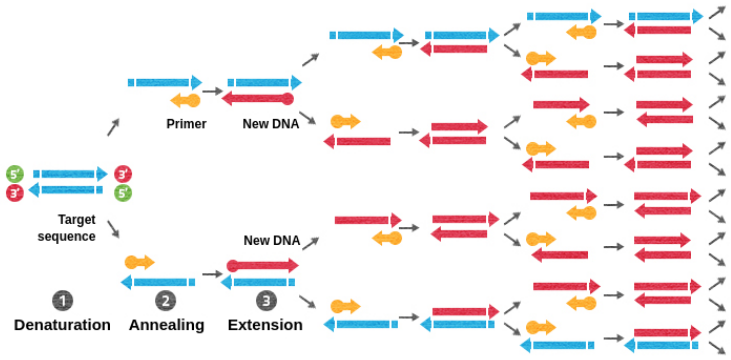
Amplifying DNA: The Polymerase Chain Reaction (review)



- PCR is a cyclical process with three steps: Denaturation, Annealing, and Extension.
- Denaturation occurs at high temperatures (95C), Annealing at low temperatures (around 50-65C), and Extension at moderate temperatures (72C).
- The components of the reaction include template DNA, PCR primers, heat-resistant DNA polymerase, and DNA nucleotides.
- Denaturation separates the double-stranded DNA into two single strands.
- Annealing binds DNA primers and single-stranded DNA.
- Extension with heat-resistant DNA polymerase builds new DNA strands and amplifies the desired DNA.
- The process occurs for multiple cycles, exponentially multiplying the desired DNA.
- Taq polymerase is a thermo-stable DNA polymerase essential for PCR.

Review steps of PCR here: https://dnalc.cshl.edu/resources/animations/pcr.html

Primer design points:

- **Keep the melting temperatures (T_m) of each primer pair within 2°C of one another (ideally).** The T_m can be approximately calculated by the formula T_m = (A+T) x 2 + (G+C) x 4, however, more precise, and elaborate T_m calculation tools are available online. Having a similar T_m between primers ensures that the forward and reverse primers will be bound to their complementary DNA strands at the same time, reducing the chance that the primer with the highest T_m will bind to nonspecific DNA sequences. T_m between 55-80ºC are preferred.

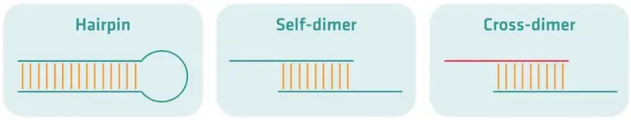
- **Use an annealing temperature (T_a) 5-10ºC below the primer with the lower Tm.** The annealing temperature is the temperature at which the primers will bind to a new template strand. This needs to be lower than the T_m so that the primers can efficiently bind to the target, but not too low that the primers bind to nonspecific targets.

- **Keep primers between 18-22 base pairs long (ideally).** This primer length is long enough to ensure binding specificity while also short enough to keep the T_m within an appropriate range.

- **Design primers with a GC content of 35-65%.** A GC content between 35% and 65% without long stretches (> 4 bases) of the same nucleotide will ensure enough sequence complexity for optimal primer specificity. Primers should end (3') in a G or C, or CG or GC: this prevents "breathing" of ends and increases efficiency of priming, however, runs of three or more Cs or Gs at the 3'-ends of primers may promote miss-priming at G or C-rich sequences and should be avoided.

- **Minimize G/C repeats, especially at the 3' end of the primer.** Cytosine and guanine have stronger binding affinity than adenine and thymine and repeats of more than 4 G or C can bind to many places in the genome with high affinity. If these repeats are at the 3' end of the primer, DNA polymerase can extend amplicons in off-target locations, which can ultimately decrease the PCR efficiency.

- **Check oligonucleotide sequences for hairpins, self-dimers, and hetero dimers.** Avoid using oligonucleotides that have strong interactions with themselves (hairpins or self-dimers) or with other oligonucleotides in the reaction mixture (hetero-dimers), especially if the interaction occurs at the 3' end of the primer. Often, we assess how much primer-primer interactions are tolerated by looking at the Gibbs free energy ΔG of the primer pairs, meaning the amount of energy needed for a primer to form a particular secondary structure with itself. Very low -ve ΔG are not well tolerated for primer design as the primers will be more likely to form dimers/hairpins. Try to aim for ΔG = ⩾ -3 kcal mol-1 for hairpins and ΔG = ⩾ -5 kcal mol-1 for hetero/self-dimers.



- **Check the oligonucleotide sequences (primers) for specificity.** Primers should be specific to the target sequence of interest. To ensure specificity to the right genomic location, you can use NCBI's BLAST tool to cross reference the primer sequence against the entire genome.

- **The reverse primer you use in a PCR reaction must be a reverse compliment of the DNA sequence you are using.**

```
Sequence:

>unk
TATGGTGTGACCGAAAGGTAAGATGATAGTTATGGTTGAGCTGGTAGCAGTAT

This means the DNA looks like:

5'TATGGTGTGACCGAAAGGTAAGATGAATCATCTAGTTATGGTTGAGCTGGTAGCAGTAT'3
3'ATACCACACTGGCTTTCCATTCTACTTAGTAGATCAATACCAACTCGACCATCGTCATA'5

You can design primers to amplify it:

5'TATGGTGTGACCGAAAGGTAAGATGAATCATCTAGTTATGGTTGAGCTGGTAGCAGTAT'3
                                  3'CAATACCAACTCGACCATCGTC'5     (reverse primer)

   5'GGTGTGACCGAAAGGTAAGATG'3                                    (forward primer)
3'ATACCACACTGGCTTTCCATTCTACTTAGTAGATCAATACCAACTCGACCATCGTCATA'5


Forward primer: 5'GGTGTGACCGAAAGGTAAGATG'3
Reverse primer: 5'CTGCTACCAGCTCAACCATAAC'3



TARGET AMPLICON:

5'GGTGTGACCGAAAGGTAAGATGAATCATCTAGTTATGGTTGAGCTGGTAGCAG'3
3'CCACACTGGCTTTCCATTCTACTTAGTAGATCAATACCAACTCGACCATCGTC'5

Number of bp: 53
```

The .txt of this is in Moodle: eg_sequence.txt

Exercise:

You have isolated a bacterium from a patient at a hospital. After sequencing the bacteria, there was one sequence that was unknown…

Find the DNA sequence file on Moodle. It should be called:

Unknown.fasta

FASTA is a file format: FASTA is a text-based, bioinformatic data format used to store nucleotide or amino acid sequences. Here are just some examples:

>Name and or description of the sequence
ACCTTCTTGAAATACTCCAAAAAGAGAAAGTCAACATCAATATTGTTGG

>Name and or description of the sequence
MADQLTEEQAEFKEAFSLFDKDGDGTITELGTVMRSLGQNPTELQDMINEVDAD

Use BLAST to identify what the **Unknown.fasta** sequence is.

BLAST is "Basic Local Alignment Search Tool". BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.

https://blast.ncbi.nlm.nih.gov/Blast.cgi

- You are doing nucleotide to nucleotide Blast.
- Use "Highly similar sequences (megablast)" option

What is the Unknown sequence of?

How confident can we be of this?

Can you figure out what gene this is specifically? What protein is it coding for?

You have a lot of lake water samples that arrived in your lab. These lake samples are from a public beach. You would like to know if that water contains the above sequence. Design primers to run a PCR reaction to test the samples to identify which contain that DNA. Aim for amplicon size between 150-500bp.

Design forward and reverse primers. You will do this with consideration for the above points.

1) Design forward and reverse primers manually by hand.
    a. Forward primer:
    b. Reverse primer:
    c. Amplicon length:
    d. What is the Tm of the primers?
    e. What would the annealing temperature for your PCR reaction?

2) Design forward and reverse primers using the online IDT tool: **PrimerQuest**
   user: ENV103
   password: ENV103epfl!
   Use online primer design tools to design forward and reverse primers using the same DNA sequence above.
    a. Forward primer:
    b. Reverse primer:
    c. Amplicon length:
    d. What is the Tm of the primers?
    e. What would the annealing temperature for your PCR reaction?

3) Design forward and reverse primers using ChatGPT and use ChatGPT to calculate the **c.**, **d.,** and **e..**
    a. Forward primer:
    b. Reverse primer:
    c. Amplicon length:
    d. What is the Tm of the primers?
    e. What would the annealing temperature for your PCR reaction?

Let's check all three sets for:

1) Primer dimers and hairpin formations. You can do this manually, which you hopefully did for the manual portion. But also, we have tools for this: https://eu.idtdna.com/calc/analyzer

2) Specificity. Use NCBI-BLAST to search your primer sequences to see if they are specific to your DNA target sequence. *Use "Highly similar sequences (megablast)" option.

3) Discuss which is a better pair. Choose one.

Resources:

Make complement/reverse compliments: https://www.bioinformatics.org/sms/rev_comp.html
Tm calculator: http://insilico.ehu.es/tm.php?formula=basic