

EXERCISE 1: PCR AND PRIMER DESIGN

Theory of PCR and primer design

Polymerase chain reaction (PCR) is a widely used technique for the selective amplification of particular DNA sequences, such as individual genes. You specify the sequence with two short "primers", which flank the region of DNA to be amplified. Primers are DNA oligonucleotides around 20-30 bases in length and the design of these is vital to the success of the PCR. PCR primers are important as they are complementary to the beginning and end of the DNA fragment of interest which one needs to amplify. Any sequence a primer is "complementary" to is repeatedly replicated by DNA polymerase in the reaction. In principle, a single molecule of DNA can be amplified to detectable levels. PCR is widely used in forensic medicine and clinical diagnostics. It can even be used to amplify DNA from the hides of stuffed museum specimens, to "bring back to life" genes from extinct species!

This exercise should aid in the understanding of this technique and be a guide for the design of PCR primers.

A. Understanding PCR

DNA polymerase uses single-stranded DNA as a template for synthesis of a complementary new strand. Both DNA strands of a double helix can serve as templates for synthesis, provided a primer is supplied for each strand. The PCR starts with heating the DNA to separate the two strands, the primers then anneal to their complementary binding sites, and new chains are synthesised. The cycle of heating, primer binding and extension is repeated many times to yield enough DNA that it can be easily visualised on an agarose gel (as you did on DAY 1). The following figure shows a diagrammatic representation of the PCR process, allowing you to visualise the exponential increase of DNA copies.

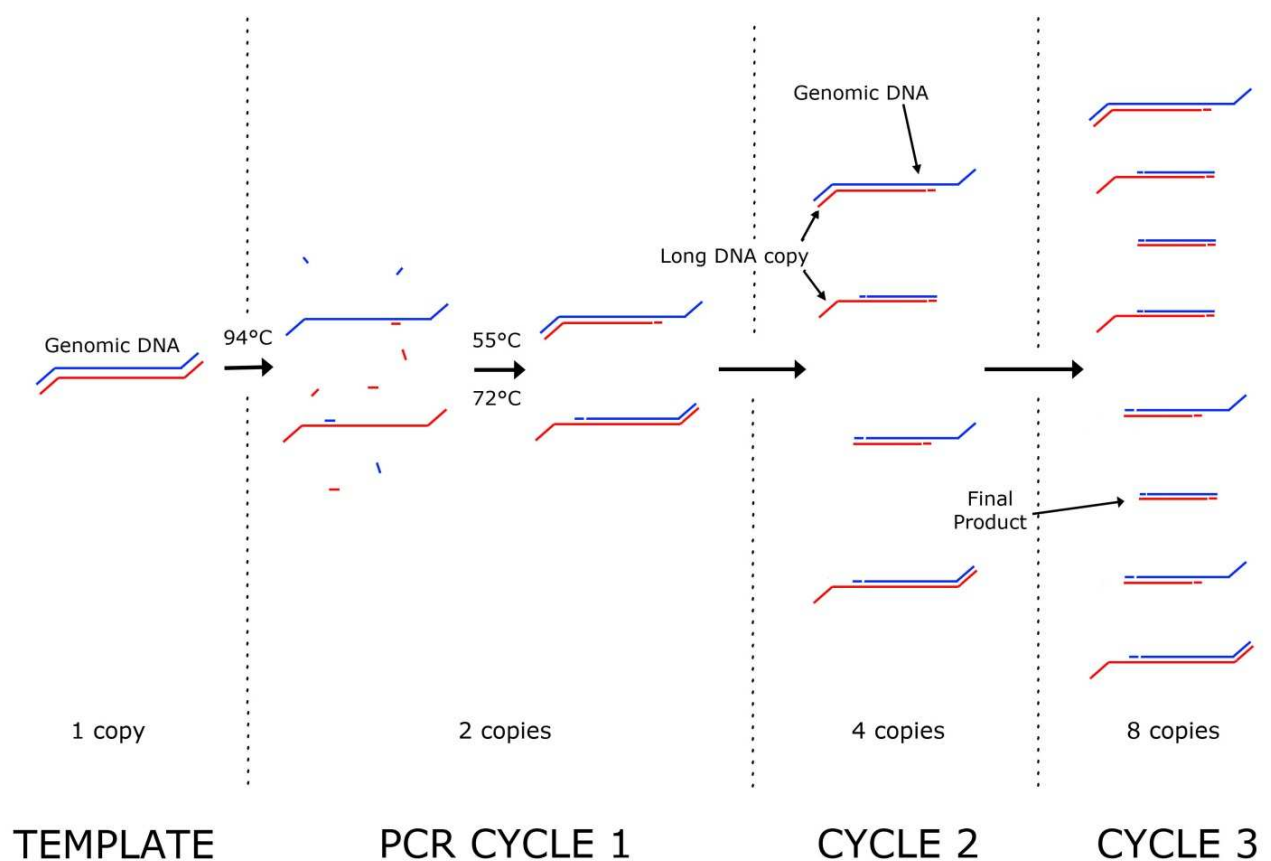


Figure 1 First Three Cycles of a Polymerase Chain Reaction

Cycle 1: After heating to 94°C, the double-stranded DNA dissociates. On cooling quickly to the annealing temperature (typically 40-55°C), the primers anneal to their complementary sequence flanking the target region. *Taq* DNA polymerase clamps on to the 3' end of the primer annealed to the DNA template and extends it. So, on heating to the extension temp of 72°C, both our pieces of DNA will be extended, typically for about a kilobase. There are now 2 double-stranded copies containing the region between the primers.

- Q1.** In cycle 1 does the *Taq* polymerase extend beyond the position of the primer on the opposite strand. Why is this the case?

Cycle 2: The reaction mixture is heated again; the original and newly synthesised DNA strands separate. Primers will bind again at the appropriate annealing temperature and then *Taq* polymerase synthesises new complementary strands. The extension of these chains is limited precisely to the target sequence. The two newly synthesised strands thus span exactly the region specified by the primers.

Cycle 3: The process is repeated, and primers anneal to the newly synthesised strands. *Taq* polymerase synthesises complementary strands, producing double-stranded DNA fragments that are identical to the target sequence. The process is repeated and the number of target fragments doubles for each subsequent cycle of the reaction.

Typically, a PCR programme continues for 30 cycles. Because the replicated strands subsequently act as templates in subsequent cycles, the number of copies increases exponentially - hence the name "chain reaction".

What happens when there is a match for only one primer on the template DNA?

Imagine that you have 3 pieces of double-stranded DNA. One piece of DNA contains no matches to either primer, one contains a match for just the forward primer, and the third piece of DNA contains a match for the forward primer on one strand and the reverse primer on the other strand.

- Q2.** Calculate, for the cycle numbers shown below, how many copies (including the original) of strand A and strand B for each DNA exist at the end of each cycle?

A _____
B _____

A _____
B —————

A —————
B —————

After PCR cycle...	No match for either primer		Match for left primer only		Match for both primers	
	A	B	A	B	A	B
1						
2						
3						
4						
5						
30						

- Q3.** If 30 cycles of the PCR are carried out on a sample containing 200 dsDNA templates, how many dsDNA copies are made after (a) 3 PCR cycles, and (b) 30 cycles?

B. PCR primer Design

To design primers we need some sequence information, and now that several genome projects have been completed, we can get the information we require from online databases. Once a suitable sequence has been chosen the normal practice is to order the required primer from a supplier – you tell them the sequence you want made and they synthesise it using chemical technology. For about 25p per nucleotide (so roughly £6 per primer) you can get enough of a primer to carry out hundreds of reactions. But you have to make sure that the sequence you request is in the correct orientation, ie, 5' to 3', and this means that, for a reverse (or right) primer you need to deduce the complementary strand sequence. **There is a YouTube video accessible through the Molecular Methods app, which explains primer design** - https://www.youtube.com/watch?v=c-f1H07D_70.

In a PCR, there are two primers and you must order them in the correct orientation from 5' to 3' – see below for an example:

5' AGGTCAGATACAGATGGATACGCAGTGCAGATCCGATACAGATCA 3'

REVERSE PRIMER

3' CGTCT 5'

FORWARD PRIMER

5' CAGAT 3'

3' TCCAGTCTATGTCTACCTATGCGTCACGTCTAGGCTATGTCTAGT 5'

Reverse (or Right) Primer: TCTGC (reverse of CGTCT)

Forward (or Left) Primer: CAGAT

* **ADD** arrows to the above example to indicate the direction the sequence will extend from the primers.

To see if you understand here are a couple of primers to design. We use the Neuroglial genomic DNA sequence below which shows part of intron 2 followed by the start of exon 3 (**exon sequences are in bold face**).

- Q4.** Deduce the forward and reverse primer sequences, correctly orientated (5' to 3') to the underlined nucleotides below. Remember DNA is double stranded and we have only shown one of the strands to save space. It may help if you write out the underlined areas as double strands.

```
GCAAAC TATT ATGTGTATTA GTATTCTATT GGTATTCATT ATCGACTGAC ACCCCCCACT ATTTCTACTT
TTAATCTACA GATATAGTTG GATCAAGAAC GGCAAGAAGT TCGATTGGCA GCGGTACGAT AACC GCATGC
TGCGGCAGCC AGGACGTGGC ACCCTGGTGA TCACCATAACC CAAGGACGAG GATCGCGGCC ACTATCAGTG
CTTTGCGTCC AATGAATTCG GAACGGCCAC CTCGAACTCA GTATATGTGC GTAAGGCCGA GCTGAATGCC
TTCAAGGATG AGGCGGCCAA GACACTGGAG GCCGTCGAGG GTGAGCCCTT TATGCTGAAA TGTGCCGCAC
CCGATGGTTT TCCCAGTCCG ACAGTCAACT GGATGATCCA GGAGTCCATC GATGGCAGCA TCAAGTCGAT
CAACAAC TCT CGCATGACCC TCGATCCTGA GGGTAATCTC TGGTTCCTGA ATGTTACCCG TGAGGATGCC
AGCTCCGATT TCTACTATGC CTGCTCGGCC ACCTCGGTGT TTCGCAGTGA ATACAAGATT GGCAACAAGG
TGCTCCTCGA TGTCAAACAG ATGGGCGTTA GTGCCTCGCA GAACAAGCAT CCGCCCGTGC GTCAATATGT
TTCCCGTCGC CAGTCCTTGG CGTTGCGTGG CAAGCGAATG GAACTGTTTT GCATCTACGG TGGAAACCCG
CTGCCGCAGA CCGTGTGGAG CAAGGATGGC CAGCGTATAC AGTGGAGCGA TCGAATAACG CAAGGACACT
ATGGCAAATC ACTGGTCATT CGGCAGACAA ATTTGATGA TGCCGGCACA TACACCTGCG ACGTGTCCAA
CGGTGTGGGC AATGCCCAAT CCTTCTCCAT...
```

Figure 2 Part of the Neuroglial genomic sequence. This shows part of intron 2 followed by the start of exon 3 (in bold).

C. Rules of Primer Design

Good primer design is essential for successful PCR. The important design considerations are summarised below and are key to specific amplification with a high yield.

Sequence	<ul style="list-style-type: none">• Avoid long runs of a single base eg. ACTGGGGGGGGCA• Have a G or C at the 3' end• Avoid primer secondary structure eg self-annealing and hairpin loop• Primer sequence must be unique to the DNA template. Actual DNA sequences are not random so some are more common than others and must be avoided.
Length	Optimal primer length is 18-30 nucleotides
GC content	40-60%
T_m	<ul style="list-style-type: none">• Melting temperature (T_m) between 55°C and 65°C• $T_m = 2^{\circ}\text{C} \times (\text{A}+\text{T}) + 4^{\circ}\text{C} \times (\text{C}+\text{G})$

Primer Length

Primers should ideally be 20 to 25 nucleotides long (range 18 to 30 nucleotides). This ensures that they will anneal to a unique sequence and therefore be specific. At the start of a PCR there will be very few template molecules, but sufficient primer to generate billions of copies of the template; in other words there will be billions of primer molecules. All these primers will be trying to locate a sequence to anneal to, and, since there will not be enough target DNA template, we have to ensure that there are unlikely to be other matches, or close matches, for our primer elsewhere in the template. This is ensured by using a sequence long enough that it has a very good chance of having a unique annealing site within the template. But we don't want the primer to be too long as efficiency of annealing decreases with increased primer length.

Q5. How many times are you likely to find (a) a particular 12 nucleotide sequence, and (b) a particular 18 nucleotide sequence within the human genome of 3.3×10^9 nucleotides (for the purposes of the calculation, assume a GC:AT ratio of 1:1).

Q6. What does the result from the previous question tell you about the importance of primer length?

GC content and T_m

The GC content should be 40 to 60% to ensure good annealing of the primer.

The GC content and melting temperature are clearly related: GC pairs have three H-bonds, whereas AT base pairs only have two H-bonds. So GC-rich sequence anneals more strongly, and we want our PCR primers to anneal efficiently.

The formula to determine the T_m of a primer uses nearest neighbour thermodynamic calculations, requiring enthalpy, entropy and molar gas constants to be considered. Fortunately, there are computer packages which will carry out the calculation for you! For sequences of the length of most PCR primers a reasonable approximation of the T_m can be generated by the equation: $T_m = 2(\text{A}+\text{T}) + 4(\text{G}+\text{C})$. Most importantly, it is crucial that the two primers have very similar annealing temperatures. If the T_m for one primer is, say 10°C lower than the T_m for the second primer, then at the relatively low annealing temperature required for the first primer it is very likely that the second primer will be annealing to additional, close matches in the template DNA. For every PCR we need to choose an annealing temperature which is low enough to allow the primers to anneal, but high enough to prevent mismatches in primer annealing. This temperature is usually about 5°C below that calculated by the above equation, and the best temperature to use for a primer pair is often determined experimentally.

- Q7.** Using the above equation, work out the T_m for the sequences chosen for the forward and reverse primers which you designed to the neuroglian sequence shown in Figure 2 (see previous question).

3' end

The 3' end base should be a G or a C, or even better, GC or CG.

- Q8.** Why is it best to have a G or a C at the 3' end of the primer?

Sequence self-complementarity

There should be no intra-primer or inter-primer homology, particularly at the 3'ends. The 3' end bases of the two primers must not be complementary, and in other regions of the primer, there should be no homologies longer than 3 base pairs. As mentioned earlier, there will be billions of copies of primer and few copies of template during early reaction cycles. This means that the most likely molecule a primer will meet in the reaction tube is another primer molecule!

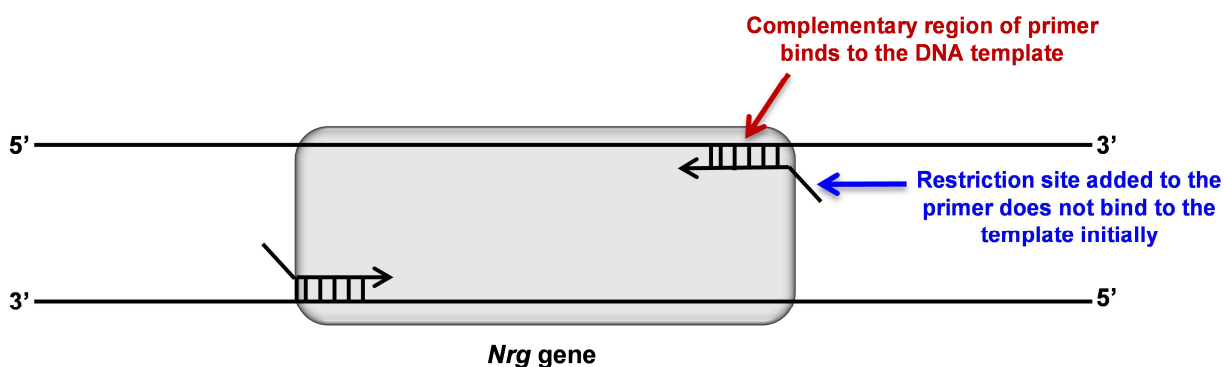
- Q9.** What would happen if, say, the forward primer had the sequence 5'GTTCGCATTCTGAATGCGAAC3'? Draw below the structure a single primer molecule could form, and the structure formed by two of these primers together and suggest why a PCR using this sequence for one of the primers will be likely to fail.
- Q10.** What would happen if the forward primer had the sequence 5'GTTCGTCTCAACGAAGTC3' and the reverse primer had the sequence 5'GGAAGACATCTGGTCGAC3'?

Polypurine and polypyrimidine sequences

These are to be avoided. Runs of Gs or Cs will generate primers which have a high propensity to anneal to GC-rich regions of the template DNA and therefore mis-prime. Runs of As or Ts will promote dissociation of that region of primer from the template due to the weaker H-bonding. Polypurine (A,G) and polypyrimidine (C,T) tracts should also be avoided. Aim for a pretty much random distribution of nucleotides.

D. Designing PCR primers which generate restriction sites at the ends of the PCR product

Cloning of PCR products is complicated by the fact that *Taq* polymerase tends to generate single base overhangs, usually (but not always) A, at the 3' end of the product. Some kits for cloning PCR products actually take advantage of this, and produce cut vector, ready for use, that has a single T overhang at each 3' end. However, this is expensive, so you are using the alternative strategy, which is to introduce restriction enzyme sites into the ends of your PCR product by adding the appropriate sequence to the 5' end of your primer. By doing this, you produce a primer which is deliberately mismatched with the template at the 5' end. But this is OK – the only position where mismatches are not tolerated by PCR is at the very 3' end of the primer.



What happens when a primer mismatches the template at the 5' end? When the primer anneals to the original template, the mismatched region will not pair with template as it will not be complementary. But each strand we make in PCR becomes a template for all the subsequent rounds of PCR (see the diagrams at the beginning of this exercise), and so the "mismatched" sequence of the primer is copied faithfully into every subsequent PCR product.

- Q11.** Using the above information together with the selected forward primer sequence indicated on the sequence on Figure 2, write out a suitable primer (5' to 3') which will incorporate a KpnI site into the PCR product (the KpnI recognition sequence is GGTACC).
- Q12.** Now do the same for the reverse primer sequence but this time incorporate BamHI (recognition sequence is GGATCC) at the 5' end of the reverse primer, which will incorporate a BamHI site into the PCR product.

The only other point we need to bear in mind is that only some restriction enzymes will cut close to the end of a DNA molecule, and the ones that do generally require at least 3 or 4 bases **to the 5' end** of the recognition site in order to cut the DNA efficiently. This is often referred to as the '**stuffer**' region.

- Q13.** In the primer sequences you have designed in the above two questions now add in a stuffer at the 5' ends of both primers which will allow cleavage of the ends of the PCR product by restriction enzymes.
- Q14.** Annotate Figure 3 to indicate where you think the 5' and 3' ends of the primers would be and also specify where the BamHI and Kpn1 sites would be in the primers.