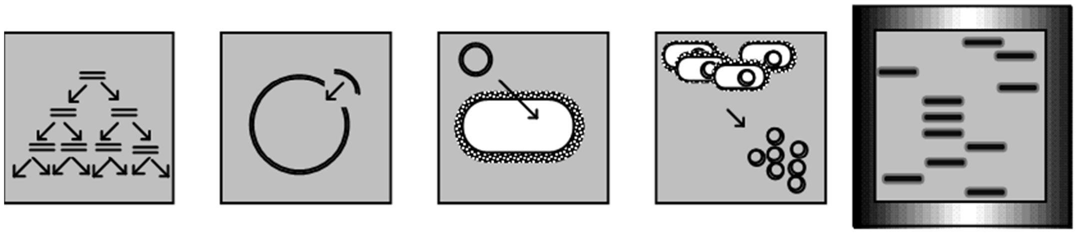
DAY 4 – SEQUENCING



By cloning the PCR product into a plasmid, we can store it indefinitely in a stable form, going back as often as we like to produce large amounts of it. One of the many benefits of this approach is that it offers us the chance to determine the sequence of our PCR product. You have cloned the *Nrg* gene from fruit flies with the icebox phenotype into a plasmid vector which allows you to sequence the gene for further investigation. The dideoxy sequencing method developed by Sanger allows us to obtain several hundred bases at a time, moving from a location where our sequencing primer anneals to the vector, to the gene you have cloned into the vector. Using this technology we can discover if the *Nrg* DNA sequence you have cloned has any DNA mutations in comparison to a wild type *Nrg* gene.

The **multiple cloning site** (MCS) present in the pBluescript vector you have been using is indicated below. The forward and reverse primer sequences are highlighted in bold:

Forward DNA sequencing primer

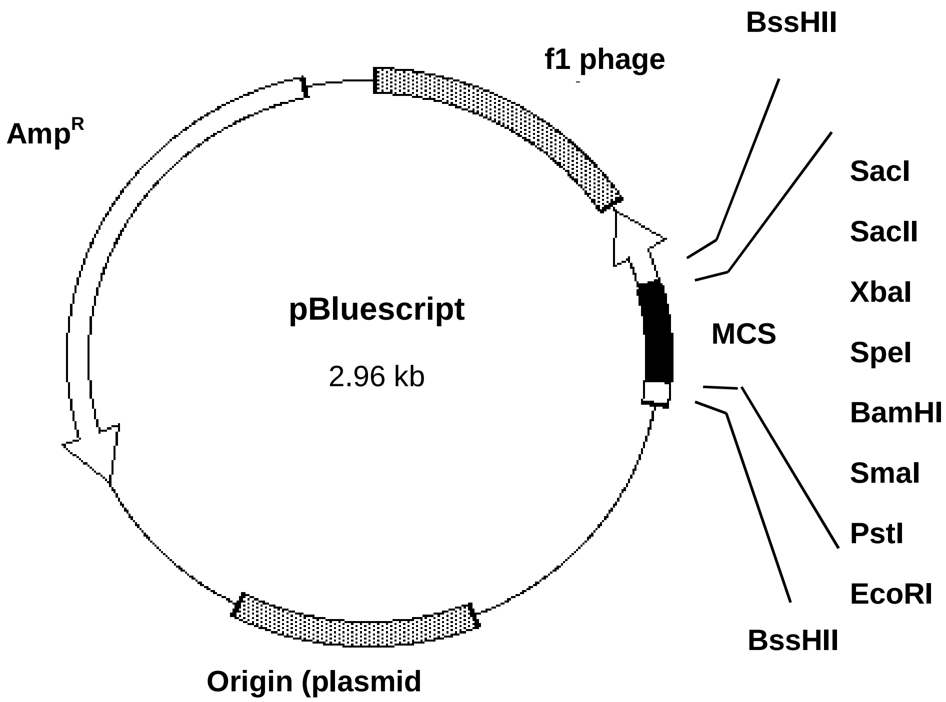
5'**GTAAAACGACGGCCAGT**3'

```
                  BssHII     T7 promoter        SacI    SacII           XbaI  SpeI  BamHI SmaI  PstI  EcoRI
600                 |      ------------->         |       |               |     |     |     |     |     |  712
  GTAAAACGACGGCCAGTGAGCGCGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTC
  CATTTTGCTGCCGGTCACTCGCGCATTATGCTGAGTGATATCCCGCTTAACCTCGAGGTGGCGCCACCGCCGGCGAGATCTTGATCACCTAGGGGGCCCGACGTCCTTAAG

EcoRV     ClaI    HincII
|   HindIII |       SalI   XhoI           KpnI                                BssHII
713   |     |         |     |              | 760                               |       800                     826
GATATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCC
CTATAGTTCGAATAGCTATGGCAGCTGGAGCTCCCCCCCGGGCCATGGGTCGAAAACAAGGGAAATCACTCCCAATTAACGCGCGAACCGCATTAGTACCAGTATCGACAAAGG
                                                                  <-------------
                                                                    T3 promoter               3'GTACCAGTATCGACAAAGG5'
```

Reverse DNA sequencing primer

This is the sequence of the MCS region of the plasmid

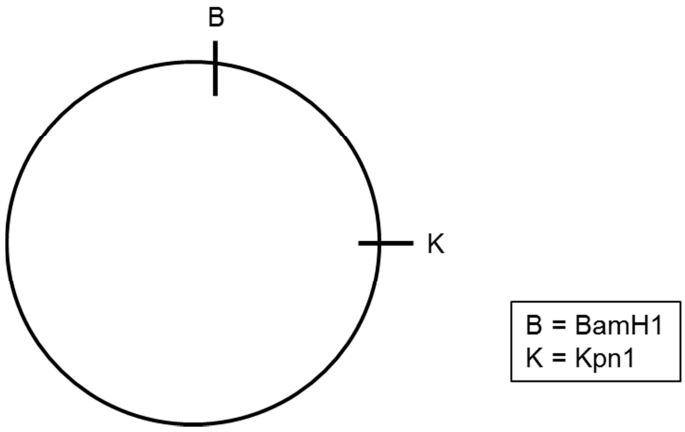


Q1. Below is a schematic diagram representing the pBluescript plasmid cloned with the *Nrg* gene sequence that you amplified by PCR. Can you label this diagram with the following (also state what size in bp you think these are):

- the cloned *Nrg* PCR product
- pBluescript plasmid

In addition, indicate on the diagram where you think the locations of the following might be:

- the forward and reverse PCR primer sequences
- the forward and reverse sequencing primer



Q2. Write down the first 10 nucleotides which you might expect to be sequenced from the forward and reverse primers:

Forward primer: 5' GTAAAACGACGGCCAGT.....
Reverse primer: 5' GGAAACAGCTATGACCATG.....

Sequencing reactions

Although manual dideoxy sequencing was the norm for a number of years, it has now been largely replaced by automated sequencing. Exciting technology has also been developed for large-scale sequencing projects, referred to as next generation sequencing, in which millions of sequence can be processed in parallel. The principle behind these reactions remains the same. They are enzymatic reactions that synthesizes DNA *in vitro*. The synthesized DNA is complementary to the template DNA. By determining the nucleotide sequence of the synthesized DNA, we can deduce the sequence of the template DNA.

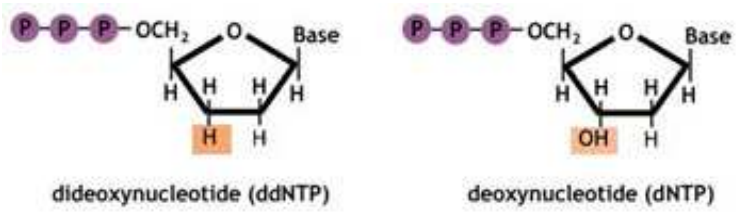
Components of a Sequencing Reaction:

Template: single-stranded DNA that you want to sequence. It can be a PCR product, genomic DNA, or as in your case, a cloned fragment.

Primer: a short fragment of DNA that binds to one end of the template DNA. The primer provides specificity to the sequence reaction and also serves as the anchor to which nucleotides are added.

Deoxynucleotides (dNTPs): extend the primer, forming a DNA chain. All four nucleotides (A,T,G,C in deoxynucleotide form) are added to the sequencing reaction.

Dideoxynucleotides (ddNTPs): another form of nucleotide that inhibit extension of the primer. Once a ddNTP has been incorporated into then DNA chain, no further nucleotides can be added.



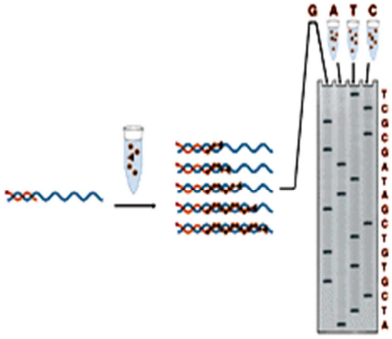
DNA polymerase: incorporates the nucleotides and dideoxynucleotides into the growing DNA chain.

Buffer: stabilizes the reagents and products in the sequencing reaction.

The vector containing the *Nrg* gene that you cloned will be sequenced by automated sequencing. In order to understand the biology behind the technology you will first do an exercise on manual sequencing.

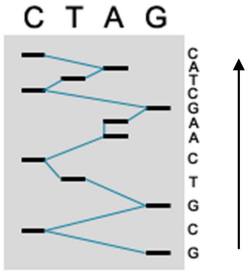
Overview of Sanger dideoxy manual sequencing

- Four reaction tubes are assembled to include all the above components, and each will have one of the four specific ddNTPs (ddATP *or* ddCTP *or* ddGTP *or* ddTTP). These are added at a lower concentration than the dNTP (say, 1 : 100 ddNTP : dNTP).
- The sequencing reaction begins with annealing of the sequencing primer to a specific binding site on a DNA template that has been converted to single stranded DNA.
- DNA polymerase can extend the primer, incorporating dNTPs to complement the template strand and synthesising new complementary DNA.



Sanger Sequencing

- The ddNTPs lack a 3' hydroxyl group, so when they are incorporated by chance into a growing DNA strand, further extension of that strand becomes impossible. The ddNTPs are radiolabelled so can be detected by autoradiography.
- Each lane of the gel shows the products generated from a single termination reaction (doping with ddATP *or* ddCTP *or* ddGTP *or* ddTTP), so sequencing of a DNA template requires *four lanes* on a gel.
- Each time the ddNTP included in that reaction is incorporated, a proportion of the reaction is terminated. Termination products build up: some will have incorporated the added ddNTP early in the reaction and will therefore be short; others will have incorporated a ddNTP later and will be rather longer.
- Using a large, high quality polyacrylamide gel, it is possible to separate DNA fragments that differ in size by just ONE nucleotide. The fragments can then be detected by autoradiography as the radioactive emission fogs the corresponding part of the x-ray film.
- We can decipher the sequence of the DNA by working up the image from the bottom (smallest termination products) to the top (largest), noting the order in which the bands appear, and in which reaction they have arisen.



Example of a Manual Sequencing Gel

Imagine a single stranded template that includes the following sequence (note the polarity!):

	3'	CGCGGGACTTGAGGAAGTCC	5'
Primer extension through this region would create →	5'	GCGCCCTGAACTCCTTCAGG	3'
	5'	GCGCCCT	3'
… but if ddTTP was included in the reaction mixture at a low	5'	GCGCCCTGAACT	3'
concentration, these termination products would also result -	5'	GCGCCCTGAACTCCT	3'
	5'	GCGCCCTGAACTCCTT	3'

Questions

Q3. Based upon your experience from this course, can you suggest one approach that could be used to convert double stranded plasmid DNA to a single stranded DNA.

Q4. Write out below the sequence of termination products that would arise if the sequencing reaction was doped with ddATP. What would happen if ddCTP or ddGTP were added instead?

DNA template	3'	CGCGGGACTTGAGGAAGTCC	5'
Un-interrupted extension product	5'	GCGCCCTGAACTCCTTCAGG	3'

Products with ddATP

Products with ddCTP

Products with ddGTP

Q5. Write out the sequence of the DNA arising from this autoradiograph:

Overview of automated sequencing

Automated sequencing is based on manual Sanger sequencing, but there are important differences to note:

- All *four* ddNTPs are included in a *single* sequencing reaction.
- Each of the four ddNTPs carries a different fluorescent dye that that can be distinguished using laser illumination and detection equipment. Effectively the termination products become colour-coded, one wavelength of fluorescence showing termination through incorporation of ddATP, a different signal for ddCTP, unique third and fourth emissions for ddGTP and ddTTP.
- The reaction is run through multiple cycles of heating and cooling. This allows the single sequencing primer many opportunities to anneal to the template, extend and generate termination products, increasing the sensitivity of the reaction.

- The sequencing reaction is run through a polyacrylamide gel in glass capillaries to separate the termination products according to size.

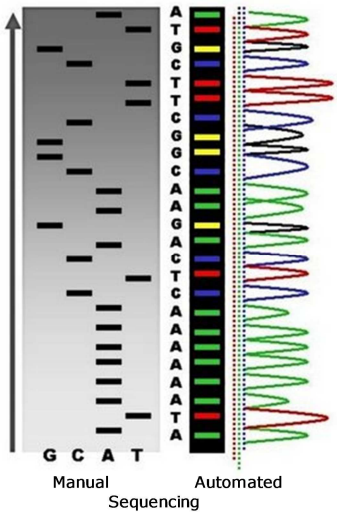


Figure 1 Manual and Automated Sequencing

- The gel is run for a long period of time and as each DNA fragment passes through a detection system near the bottom of the gel, its fluorescent signature is measured and recorded by computer. The computer analyses the intensity and wavelength of the emitted light and from this data it attempts to identify which ddNTP is responsible for termination at that position on the template (this process is termed "base calling").
- Data from an automated sequencing run is often delivered to the investigator as a chromatogram. Here, the intensity of signal from each terminator is plotted out as the DNA fragments pass through the detection system linked to the analytical gel.
- We can work out the sequence of the DNA template if we note the order in which they appear left to right.
- The computer's "base calling" - the most likely base at each point in the analysis - is shown over the chromatogram (top line). In most instances, the sequence is clear but here and there, the data is ambiguous and the skill and judgement of the human investigator is required.

Questions

Q6. Why in Figure 1 are there different coloured fragments in the middle lane of this diagram? How does this differ from manual sequencing?

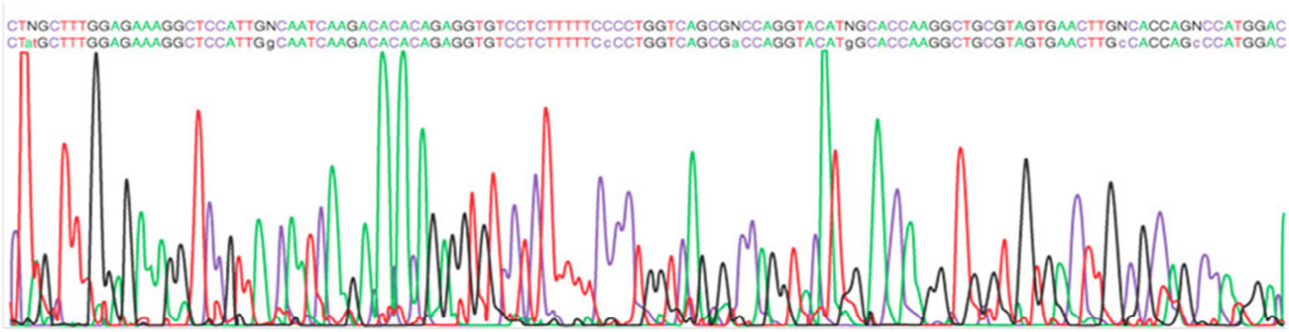


Figure 2 Example of a chromatogram resulting from an automated sequencing reaction

Q7. The third base in the sequence shown in Figure 2 is recorded as "N". Why do you think this is? Shortly after, the sequence AAA appears. Can *you* see three clear A peaks or is it the spacing of the peaks that suggests a third A residue is present?

Q8. Further along on Figure 2 locate the sequence CCATT. Is this the correct sequence?

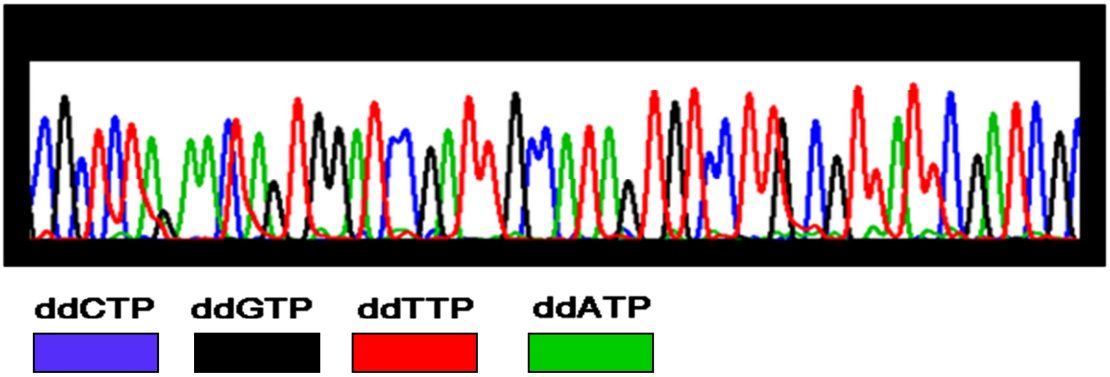


Figure 3 A chromatogram for you to interpret

Q9. Once you are confident about how to collect and interpret sequence data in this way, gather sequence from the chromatogram shown in Figure 3.

Q10. The chromatogram in Figure 3 was taken from a real sequencing reaction of the *Nrg* gene. The DNA template was pBluescript carrying the insert ligated into the BamHI /KpnI sites. The sequencing reaction was carried out with a Forward sequencing primer.

Can you identify the BamHI restriction site (GGATCC) on the chromatogram in Figure 3? Mark this on the sequence you have written down in the above question and indicate where the vector and insert sequences are in relation to this.