University of Glasgow | School of Life Sciences

Molecular Methods

Laboratory Manual

Session 2014-2015

Table of Contents

Overview of the course 1
Pre-Lab Exercise : Designing a PCR Experiment 4
Part A Identifying Gene Mutations 9
Background to the Lab 10
DAY 1 - PCR 11
DAY 1 - Agarose Gel Electrophoresis And DNA Cleanup 14
DAY 2 - Cloning a PCR product 19
DAY 2 - Transformation of Competent E.coli Cells 23
DAY 3 - Analysing Your Transformed Bacterial Cells 25
DAY 4 – Plasmid Miniprep 30
DAY 4 – Sequencing 34
Exercise 1: PCR And Primer Design 40
Exercise 2: Restriction Mapping 49
Exercise 3: Laboratory Calculations 54
Computer Exercise 1: How to Investigate the Function of a Gene 58
Computer Exercise 2: Sequence Analysis 64
Part B Using PCR and Real-time quantitative PCR to Detect and Quantify HIV 71
Background to the Lab 72
DAY 5 (AM) - Conventional HIV PCR 73
DAY 5 (AM)– Real-Time Quantitative PCR 75
DAY 5 (PM) – Agarose Gel Electrophoresis of conventional HIV PCR 78
Computer Exercise 3: Real-time qPCR Data Analysis 79
APPENDIX 1 - Safety Issues 83
APPENDIX 2 - Using a Gilson pipette 84
APPENDIX 3 - Recipe for agarose gel electrophoresis of DNA 85
APPENDIX 4 - Circular Plasmid Map of pBluescript Vector 86
APPENDIX 5 - Useful Web URLs: 87
APPENDIX 6 - The Genetic Code 88
APPENDIX 7 – How to Use a QR Code 89

OVERVIEW OF THE COURSE

This course is designed to teach you **practical molecular biology** in the way that most graduate students learn it - on the job! To accomplish this, we will firstly replicate an experiment that was actually performed in one of our research labs and secondly, detect and quantify viral infection using a technique that diagnostics labs use. In this way, you will get training in a series of techniques in the context in which they would normally be used, rather than as disjointed academic exercises. At the end of this course, we hope that – at the least – we will have demystified some of these procedures; and that with luck, you will have gained the understanding and confidence to be able to plan and perform your own experiments. You may even find it helpful to use your manual and notes next year for your project, or even for your postgraduate study. There are two main parts to this course:

PART A: Identifying Gene Mutations: The **overall goal** is to **characterize** the **mutation** which causes a **behavioural abnormality** in the fruitfly ***Drosophila melanogaster***. Here is the sequence of events:

Design primers to amplify the relevant sequence of the target gene

↓

Perform a polymerase chain reaction (PCR), using *Drosophila* genomic DNA as template

↓

Check the success of the reaction by agarose gel electrophoresis

↓

Purify and measure the yield of the PCR product

↓

Ligate the PCR product into a prepared plasmid vector

↓

Transform bugs (*Esherichia.coli*) with the vector containing the PCR product

↓

Grow the bugs in liquid culture and harvest and purify the plasmid

↓

Check the identity of the insert by restriction and agarose gel electrophoresis

↓

Use computer based analysis to compare the sequence of the insert with other database entries

PART B: Detection and Quantification of HIV: The **overall goal** here is to detect Human Immunodeficiency Virus (HIV) in patient blood samples using methods that would be utilised in a diagnostics lab. Here is the sequence of events:

Design primers to amplify a relevant HIV gene sequence

↓

Perform real-time quantitative PCR (as well as conventional PCR), using cDNA from patient blood samples

↓

Check success of conventional PCR by agarose gel electrophoresis

↓

Analyse real-time qPCR output data to determine viral load

Aims

This course aims to provide you with the tools for you to understand and perform basic molecular biology techniques. The practical and computer laboratories described in this manual are designed so that you get first-hand experience of the basic procedures used in molecular biology, in a research-based context.

Intended Learning Outcomes

By the end of the course you should be able to:

- Demonstrate a broad-based knowledge of laboratory-based practices in the field of molecular biology
- Describe techniques which are used routinely in molecular biology laboratories
- Describe in detail the theory behind the molecular methodologies that you undertake in the laboratory
- Analyse and perform calculations relating to your lab-based techniques in the field of molecular biology
- Analyse and explain results from experimental data
- Perform basic bioinformatics

Detailed Intended Learning Outcomes

At the end of the course, you should be able to:-

PCR:

- Describe the theory of, and be able to design and perform, a polymerase chain reaction (PCR) reaction
- Understand the principles behind PCR primer design
- Analyse PCR results from gel electrophoresis
- Compare and contrast PCR and real-time qPCR
- Analyse real-time qPCR data

Ligation and transformation:

- Describe the strategies used in cloning DNA into a suitable vector
- Describe, and be able to perform, transformation of *E.coli* with a vector
- Explain the principles and purpose of antibiotic selection and blue/white screening

Plasmid miniprep and restriction digest:

- Describe and perform a small-scale plasmid purification from an overnight culture
- Understand the concept of, and be able to deduce, a restriction map for a plasmid

DNA sequencing:

- Understand the principle of dideoxy sequencing of a plasmid template
- Understand the use of computers in sequence entry and analysis

Computer Exercises:

- Be familiar with using computers and the internet to access biologically relevant data, including genome project data for humans, *Drosophila*, other model organisms and microbes.

Overall:

- Be able to select basic molecular techniques from those used above to provide practical solutions to simple problems in molecular research.
- Be able to relate molecular biology to your individual field of interest

Assessment

Depending on your degree subject, the assessment will be either an open or closed book short-answer test which will take place under examination conditions. In addition, some degree groups will be asked to identify, investigate and produce a written summary or an oral presentation of a piece of current research that relates molecular biology to your degree subject. You will be notified well in advance the details of your degree group assessment.

Every part of the course described above is considered assessable! Check your timetable for the location, time and date of the test.

Course Materials

This is more than just a lab course! Firstly, we do not explain everything exhaustively because we expect you to make use of what you have already learned in the lecture course. Secondly, as in real research, there are a lot of spaces while you wait for things to happen. Your spare time is to be filled in with student-directed learning, either as part of formal exercises, or simply to fill in areas you're not too clear on.

The following resources are all **integral** and **assessable** parts of your course:

The practical work.

This lab book. As well as reading it, note that there are several questions throughout the manual that you should answer. The book contains a lot of useful information, which we expect you to understand.

Your lecture course. We expect you to draw on this to understand concepts such as PCR, plasmid vectors, etc.

Group talks. The lab leader will give formal talks at the start of each lab session. These will contain important safety and organizational information, and will review the skills to be learned in the work to follow. It is not enough to be there and to listen politely; you should take notes, and ask questions if necessary.

Demonstrators. They should be pro-active and come to you to discuss what you are doing, offer help as needed, and make sure you grasp what is going on. For your part, you should pay attention, ask them questions to aid your understanding, and make notes as appropriate.

Exercises. These are included at the back of the manual. You will be invited to do these at appropriate parts of the course, and the results can be discussed with your demonstrators.

Computer skills. You need to have a reasonable grasp of IT skills and will need to be able to "surf the net" using Netscape or similar software.

Moodle

The Molecular methods Moodle site contains lecture notes/lab notes/videos/computing session resource/additional reading/web sites.

Moodle site: Level 3- Molecular Methods-can be accessed from your Moodle page.

Please use this site for revision and background reading. We consider the material on it to be examinable, and we expect you to be fluent in the use of the Internet.

E-learning resources

Articulated Lectures. These short lectures are available via the Molecular Methods Moodle site and provide a revision of background information on Central Dogma concepts that you should already know from first and second year courses.

Videos and Quizzes. These have been developed in-house to aid with various aspects of the course, including primer design and restriction mapping. They can be accessed via QR codes at the relevant page in this manual or via the Moodle site or the app below.

Molecular Methods App. This is a web-based app that you can use as a resource for various aspects of the lab. You will need to log in with a username and password of your choice. You can access the app via this web link - http://molecularmethods.clinmed.gla.ac.uk/ - or via the Molecular Methods Moodle site or use the QR code below. If you are unsure how to use a QR code then refer to the notes found in Appendix 7 or use the weblink.

QR Link to Molecular Methods App

Timetable

Your timetable should indicate the location and times of your lab and computer sessions. This timetable may need to be adjusted as we go, so it's important to be flexible, good-natured and available! Please check Moodle for most up to date timetable.

Laboratory Session Venues: Labs will be held in the **Biotechnology Laboratory (room 624)** on Level 6 of the Boyd Orr Building.

Computer Session Venues: These will be held either in the **Joseph Black Building Undergraduate Teaching Computer Cluster A5-06** , the **West Medical Building Cluster, Room 515** or in **Boyd Orr Building Rm 808** or **Level 9**. Check the location of your sessions in advance!

Safety

Before starting the lab you should read the safety notes in Appendix 1 at the back of the laboratory manual.

PRE-LAB EXERCISE : DESIGNING A PCR EXPERIMENT

Introduction

You should complete this exercise prior to starting in the lab. Ask a demonstrator during the lab session to check this over and he/she can help with any incomplete answers.

During the practical aspect of this course (both parts A and B) you will set up and analyse a polymerase chain reaction (PCR) experiment. PCR has become essential to almost every aspect of molecular research and clinical diagnosis. There are numerous applications of PCR that include genetic testing (detect mutations associated with genetic diseases), the detection of pathogen DNA (e.g HIV), DNA fingerprinting for forensic analysis (isolate DNA from a crime scene), and phylogenetic analysis (evolutionary relationships between species).

The method relies on thermal cycling, consisting of cycles of repeated heating and cooling of the reaction for DNA melting and enzymatic replication of the DNA. There are numerous factors that must be considered when designing a PCR experiment. These include using appropriate PCR conditions, and using appropriate reaction mixture components. In reality, the whole process can become a bit complicated, as there is a lot to think about. Fortunately, there are some general rules which provide an excellent starting point for designing your experiment. If you follow these simple rules, and take a little time to think about your experiment, it will make PCR experiments a lot easier.

Aim

By the end of this exercise you should have a firm understanding of how to set up a PCR

1. PCR Conditions

The reaction conditions of a PCR amplification are composed of the total number of cycles to be run and the temperature and duration of each step in those cycles. The decision as to how many cycles to run is based upon the amount of DNA target material you start with as well as how many copies of the PCR product you want.

Your first task is to consider the PCR conditions. Complete the information in Table 1 using the following resources (you can find the link to these sites on the Molecular Methods Moodle page):

Virtual Lab (PCR) - http://learn.genetics.utah.edu/content/labs/pcr/

YouTube (PCR) - http://www.youtube.com/watch?v=JRAA4C2OPwg&feature=related

Google - http://www.google.co.uk

Table 1. PCR Conditions

Temperature	Time	Step	Description
		Initial denaturation	
		Denaturation	
		Annealing	
		Extension	
		Final Extension	

x 25-35

Q1. What is the connection between the extension time and the size of the PCR product?

Q2. Is an extension time of 1 minute at 72°C sufficient to amplify a PCR product of 750 base pairs?

Q3. At what cycle number does the desired **double-stranded** PCR product appear?

Q4. What will happen during the PCR if the annealing temperature is too low?

Q5. What happens if the extension temperature is too high?

2. Components of a PCR

Now that you have an understanding of the temperature conditions used in PCR, we will turn our attention to the components which make up the reaction. The *Taq* DNA polymerase which you will be using during the lab is called GoTaq® Flexi DNA Polymerase (supplied by Promega); look on the company website (www.promega.com) for more information about the properties and use of this enzyme.

Q6. Locate the 'complete protocol' on the Promega website for 'GoTaq® Flexi DNA Polymerase' and complete Table 2 (search on the Promega site for catalogue # M830a).

Table 2. Main components of a PCR

Taq DNA polymerase
Magnesium

When you come to set up the PCR on day 1 of the practical, all of the above components will be present in the reaction, although many will be present in the PCR 'master mix'. Use your knowledge from lectures, or by google searching, to answer the following questions.

Q7. What is a PCR 'master mix' and what is the advantage of using it?

Q8. What is a primer?

Q9. What organism was *Taq* DNA polymerase originally isolated from?

Q10. What is the significance of using *Taq* DNA polymerase in a PCR?

Q11. What is the function of Magnesium Chloride ($Mg\ Cl_2$) in the reaction?

Q12. What could happen if there is too much magnesium in the reaction?

3. Amplification of a DNA Sequence

Now consider the PCR using an actual DNA sequence.

Below shows a DNA sequence containing the region to be amplified in the shaded box). Note that in 'real life' the amplified sequence is usually several hundred nucleotides long.

```
5' AGGTCAGATACAGATGGATACGCAGTGCAGATCCGATACAGATCA 3'
3' TCCAGTCTATGTCTACCTATGCGTCACGTCTAGGCTATGTCTAGT 5'
```

Cycle 1:

(a) denaturing:

```
5' AGGTCAGATACAGATGGATACGCAGTGCAGATCCGATACAGATCA 3'

3' TCCAGTCTATGTCTACCTATGCGTCACGTCTAGGCTATGTCTAGT 5'
```

(b) annealing of primers:

```
5' AGGTCAGATACAGATGGATACGCAGTGCAGATCCGATACAGATCA 3'
                                  CGTCT 5'

          5' CAGAT
3' TCCAGTCTATGTCTACCTATGCGTCACGTCTAGGCTATGTCTAGT 5'
```

(c) Extension of the new chain:

Q13. Write in the diagram below the sequence of the new DNA strands you might expect in this cycle of the PCR:

```
5' AGGTCAGATACAGATGGATACGCAGTGCAGATCCGATACAGATCA 3'
                                  CGTCT 5'

          5' CAGAT
3' TCCAGTCTATGTCTACCTATGCGTCACGTCTAGGCTATGTCTAGT 5'
```

Cycle 2:

Now you have to think about the second cycle of the PCR process. Following the denaturing stage, the DNA from the end of the previous cycle will once again be single-stranded.

Choose one of the new strands from the first PCR cycle above, write it out again below and 'attach' the appropriate primer to it (annealing stage) then once again create the new strand you would expect from the extension phase.

What you created at the end of this cycle is a DNA fragment that is identical to the target sequence. This then becomes the template for the next cycle of the PCR.

Cycle 3:

Write out the new strand below, and 'attach' the appropriate primer to it (annealing stage), then create the new strand you would expect from the extension phase.

PART A
IDENTIFYING GENE MUTATIONS

BACKGROUND TO THE LAB

Drosophila melanogaster, otherwise known as the fruit fly, provides an excellent model organism for studying behaviour. The flies demonstrate complex behaviours, particularly during courtship and mating. Males perform an intricate series of behaviours, including species-specific songs generated by wing vibrations, in order to attract a mate. Virgin females are generally receptive to males, but after mating the females become unreceptive and fend off males who court them. Several mutations have been isolated that affect the courtship behaviour of males and/or females, for example, chaste, dissatisfaction, paralytic and spinster. In this practical you will attempt to discover the molecular basis for an X-linked mutation, icebox, which was first characterized as causing homozygous females to be unreceptive to the advances of males. Males who carry the mutation take longer to complete courtship. By genetic mapping the mutation was localized to a specific region of the Drosophila X chromosome. This chromosomal region contained several genes, including the neuroglian gene (Nrg), which encodes a cell adhesion molecule that plays a role in brain development; thus Nrg was a **candidate gene** for the icebox mutation. To test whether the icebox mutation was indeed an **allele** of Nrg, genetic tests were carried out between icebox and known alleles of Nrg. In other words, if female flies carry the icebox mutation on one X chromosome and an Nrg mutation on the other X chromosome, are they normal in phenotype (which would suggest that icebox is in a different gene from Nrg) or mutant in phenotype (telling us that icebox is an allele of the Nrg gene)? The studies indicated that icebox was indeed an allele of Nrg.

Q14. Why is *Drosophila melanogaster* used as a model organism?

Q15. What is meant by the term 'allele' of a gene?

Q16. What is the difference between the 'phenotype' and 'genotype' of an organism?

The next question is to discover what the exact DNA change is in the *icebox* mutant, and to try to establish how this mutation results in the observed phenotype. Various tools can be used to identify which region of a gene sequence carries a mutation, for example, SSCP (single strand conformation polymorphism) and DGGE (denaturing gradient gel electrophoresis). Analysis of *icebox* DNA indicated one or more differences in the sequence encoding the start of exon 3. This is the region of *Nrg* that you are going to clone and sequence from *icebox* mutant flies, to try to characterize the mutation that leads to the *icebox* phenotype.

If all goes well, you will actually analyse on the last day, the sequence of the **same** piece of DNA you amplified on the first day. However, it would be **really exceptional** for something like this to run without a hitch in the real world (ask one of the demonstrators!), so we will have the relevant materials ready to "fix" you up, if your particular experiment fails at any time. Good luck!

Additional Resources

KMC O'Dell (2003) The voyeurs' guide to *Drosophila melanogaster* courtship. Behavioural Processes **64(2):** 211-223.

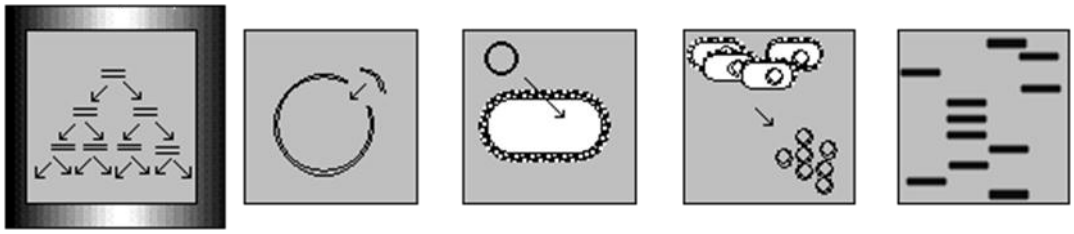
M Hortsch (1996) The L1 family of neural cell adhesion molecules: old proteins performing new tricks. Neuron **17:**587-593.

From Genes to Genomes: Concepts and Applications of DNA Technology. Dale, Von Schantz and Plant. Wiley-Blackwell (2012).

An Introduction to Genetic Engineering. Nicholl. Cambridge Press University (2008).

Please note that the above books can be borrowed from Lab 624 in the Boyd Orr overnight.

DAY 1 - PCR



Introduction

Reading: Dale et al., 3rd Ed: p109–119.

In this lab, you will set up and analyse a **Polymerase Chain Reaction (PCR).** PCR, invented in 1984, has become essential to almost every aspect of molecular research and clinical diagnosis. The *Drosophila* genome project, like that for humans, is completed and we therefore have access to the complete sequence of the *Nrg* gene from normal flies. We have designed primers for you that will anneal to the sequence either side of our target region and amplify specifically this DNA. You will have a go at designing PCR primers while your reactions are in the PCR machine.

To perform PCR, the following are placed in a plastic eppendorf tube:

- a DNA template (which can be either genomic or cDNA, derived from RNA)
- a FORWARD primer, which complements the reverse strand
- a REVERSE primer, which complements the forward strand
- excess nucleotide monomers ("dNTPs")
- a suitable buffer
- *Taq* DNA polymerase

The ingredients are mixed and placed in a computer-controlled heating block. The reaction runs over approximately 3 hours.

Today, PCR is so widely performed that the basic reagents can be purchased ready-mixed. Your PCR "Master Mix" already contains the dNTPs, buffer and *Taq* polymerase, and so only requires primers and template to be added. It also contains a green dye as a visual aid. You will set up a PCR, and while this is running, you will attempt to design some primers. When the PCR has finished, you will run the products out on a gel, quantify the product and purify the DNA. If the reactions succeed we will save the DNA for next time, when we will attempt to clone it into the plasmid cloning vector pBluescript.

Method

1. Setting up the PCR

Before starting, make sure you know how to dispense liquids **accurately** with a Gilson pipette (Appendix 2)! It is important that you work cleanly, or you will end up doing PCR with *human* genomic DNA as your template!

Each pair should set up **two** PCRs. Make sure you use thin-walled PCR tubes for this:

Tube 1: *Drosophila* genomic DNA template

Tube 2: No DNA negative control (adjust the final volume with **sterile** distilled water as required).

The ingredients and volumes are listed in the table below:

PCR Recipe:		Tube 1	Tube 2
	PCR Master Mix (dNTPs + *Taq* pol + buffer + dye)	25 µl	25 µl
	Template (*Drosophila* genomic DNA)	2 µl	
	Forward primer (10nM)	2 µl	2 µl
	Reverse primer (10nM)	2 µl	2 µl
	Sterile distilled water	19 µl	21 µl
	Final volume	**50 µl**	**50 µl**

Close the tubes, and write your initials on the top of each using a marker pen. Make sure you also distinguish which contains the template and which is the control. Mix the contents by flicking the tube and shake down or spin **very briefly at low speed** (less than 4K for a few seconds only) to collect the contents to the bottom of the tubes. Put the tubes into PCR machine when you are ready.

2. Running the PCR

Load the tubes into the PCR machine, around the temperature sensor tube. Your demonstrator will start the reaction, which will last around 3 hours. You can watch the progress of the temperature cycling on the display panel of the PCR machine

For this reaction, we will use the following programme:

Start		95°C	5 min	to ensure the genomic DNA is fully denatured
Then 5 X	denaturing	95°C	1 min	these 5 cycles have an extended denaturing as the original
	annealing	60°C	30 sec	genomic DNA will take a long time to denature after each
	extension	72°C	1 min	synthesis step
Followed by 25 X	denaturing	95°C	30 sec	in these 25 cycles the denaturing time has been reduced
	annealing	60°C	30 sec	as by now "short", easily denatured PCR products will
	extension	72°C	1 min	represent significant amounts of the available template
And finally		72°C	5 min	to ensure that all products are completely extended

While you are waiting for the PCR to finish, answer the questions at the end of this section and then work through '**Exercise 1: PCR and Primer Design** '.

3. After the PCR

We will now:-

A. Run some of the PCR product and the control reaction on an agarose gel to check the success of the PCR.

B. Save 5µl of the PCR product for running on a gel on a later day of the lab.

C. Purify the remaining PCR product from the other reagents, using a commercial DNA clean–up kit.

Each of these steps is described in the next section.

Questions

Q17. The template used in your PCR reaction can either be genomic DNA (gDNA) or complementary DNA (cDNA). What is the difference between these?

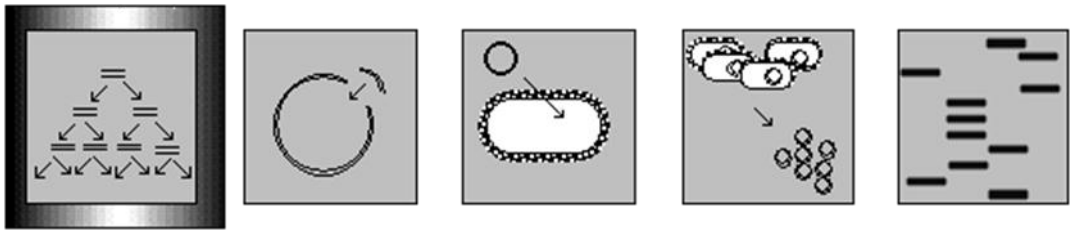
Q18. What does *Taq* DNA polymerase do in a PCR?

Q19. Thin-walled tubes cost ten times as much as conventional eppendorfs. Why do you think they are important for PCR?

Q20. Use the sequence and primer information of the gene (see **Exercise 1: PCR and Primer Design**, Figure 2) to calculate the expected size of the PCR fragment.

Q21. What is the function of the heated lid in the PCR machine?

DAY 1 - AGAROSE GEL ELECTROPHORESIS AND DNA CLEANUP



Introduction

Reading: Dale et al., 3rd Ed: p33-34.

If the PCR has worked, you will have millions of copies of the fragment of DNA amplified. The fragment size is predictable from the expected distance spanned by the primers we used in the PCR (see **Exercise 1: PCR and Primer Design** for more information on primer design). So to find out whether we have successfully amplified the target sequence we want, we need to detect the presence of a characteristic size of DNA fragment in our tube. This is accomplished by **running a gel**, and is one of the most everyday actions in a molecular biology lab.

An appropriate DNA ladder (i.e. DNA fragments of known size) should always be run in parallel to enable calculation of the size of your fragments.

Method

1. Preparation of the Agarose Gel

The gel mix will be prepared for you; but the recipe is detailed in Appendix 3, should you subsequently need to make your own. The gel requires approximately an hour to set solid, so should be prepared sufficiently in advance of the time it is required. Make sure the demonstrators have added 'SYBR Safe' to the molten agarose solution. This allows fluorescent staining of your DNA, which you will see later.

- Seal the edges of a gel tray and pour in the molten agarose. If there are any air bubbles push these to the side using a pipette tip (these would distort the running of the gel), add the comb at the top of the gel, and allow to cool to room temperature.
- When the gel has set, flood the gel tank with the TAE buffer (there should be a depth of about 5 mm TAE above the gel). Once you have removed the comb, you are ready to load your samples.

2. Preparation of PCR Samples for Electrophoresis and Running the Gel

In some PCR samples it is necessary to add 'loading buffer' which contains glycerol, and dyes that separate during electrophoresis to monitor the migration of DNA. Fortunately, the Mastermix you used in the PCR already contains a blue and yellow dye (known as 'Green GoTaq® Flexi Buffer'), therefore, there is no need to add additional loading buffer. The yellow dye migrates at a rate faster than primers.

- If you have not loaded a gel before you should ask a demonstrator to show you how and you can also practise loading samples into a well using loading dye.
- You should have your two PCR samples (one containing *Drosophila* genomic DNA, and the other a negative control). Load one sample onto the gel per person – 7μl of each will be sufficient. The demonstrators will load a DNA ladder onto the gel to show you how to load successfully.
- Remember to write your name on the loading sheet so you know where your sample is on the gel.
- Finally check with a demonstrator how to place the safety lid on the gel kit and start it running at around 100 V. The gel will take approx 1 hour to run.

CAUTION: ***potentially lethal voltages. Do not handle gel kits unsupervised, and if you notice a problem in a running gel (like a leak, steam or smoke), seek help but do not touch.***

3. Storing a Sample of your PCR for Future Reference

Place 5µl of your genomic DNA PCR product into a fresh tube (labelled) and place in box at the front of the lab. This will be stored in the freezer so that you can use it as a positive control in an agarose gel on a later day of the lab.

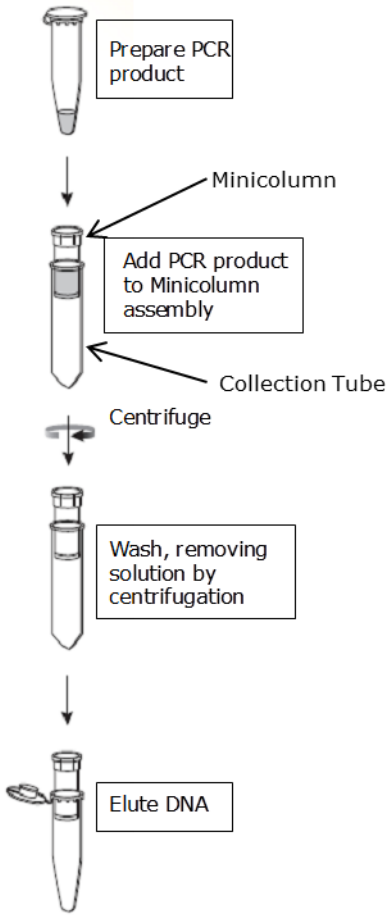
4. DNA Cleanup

We will clean up **only** the **genomic DNA** PCR product, using a commercial DNA cleanup kit. PCR products are commonly purified to remove excess nucleotides and primers. The principle is that DNA sticks strongly but reversibly to silica, which will be present in a Minicolumn. The PCR product is first mixed with a solution of guanidium isothiocyanate, which denatures and destroys any protein (including enzymes) to aid in their removal from the DNA.

1. To prepare the PCR product for the clean-up procedure, add an equal volume of Membrane Binding Solution to the remaining PCR reaction (should be about 38 µl but use a Gilson pipette to estimate the volume).

CAUTION: The guanidium isothiocyanate in the Membrane Binding Solution can denature the proteins in your skin and eyes if it comes into contact with them.

Wear gloves and goggles and work carefully! Splashes of this solution into eyes should be rinsed immediately with plenty of eyewash.



2. Clearly label the minicolumn assembly with a sample identifier and your initials.
3. Using a P200 Gilson pipette, transfer the prepared PCR product to the Minicolumn assembly and incubate for 1 minute at room temperature. Place the samples in the centrifuge and make sure these are balanced correctly in the rotor.

Ask a demonstrator if you're uncertain how to balance a centrifuge.

4. Centrifuge the Minicolumn assembly at maximum speed (13,000rpm) for 2 minutes. Remove the Minicolumn from the assembly and discard the flowthrough in the Collection tube into the disinfectant on your bench. Put the Minicolumn back into the Collection Tube.
5. Wash the column by adding 700 µl of Membrane Wash Solution to the Minicolumn. Centrifuge the Minicolumn assembly for 2 minutes at maximum speed.
6. Empty the Collection Tube as before and place the Minicolumn back in the Collection Tube. Repeat the wash with 500µl of Membrane Wash Solution and centrifuge the Minicolumn assembly for 7 minutes at maximum speed.
7. Remove the Minicolumn assembly from the centrifuge, being careful not to wet the bottom of the column with the flowthrough. Empty the Collection Tube and recentrifuge the column assembly for 2 minutes at maximum speed.

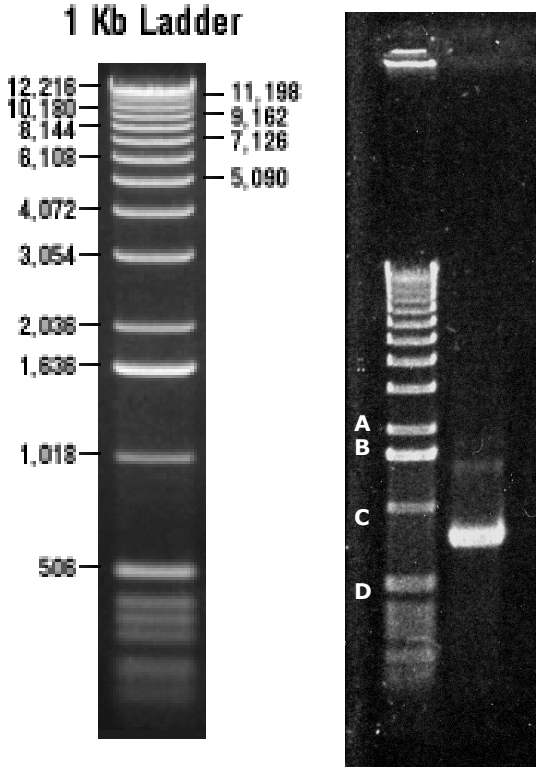
NOTE 1: The most common cause of failure in a DNA cleanup is leaving residual ethanol (contained within the Membrane Wash Solution) in the column at this stage; this may affect further stages of the practical.

NOTE 2: The second most common cause of failure results from not reading the next sentence!

8. Discard the Collection Tube and carefully **transfer** the Minicolumn to a clean 1.5ml microcentrifuge tube (**cut off the lid of the tube**). Add 40 µl* TE buffer directly to the centre of the column using a P200 Gilson (* note that the TE volume for the elution step will vary when you use this technique on later days). Incubate at room temperature for 1 minute. Put the column/tube assembly in the centrifuge and spin for 2 minutes at maximum speed. The DNA should now have **eluted** from the column into the tube.
9. Label the tube and put a lid on. This *should* contain the PCR product you will use for the rest of your labs! You should make sure your PCR product tube is fully labelled with the contents and your initials and then placed in the appropriate place to be stored until next time. Check with a demonstrator if unsure.

5. Interpreting the PCR gel

Once the gel run is complete, switch off the gel kit and carefully remove the gel tray. The demonstrators will help you get an image of the DNA on your gel using a gel imaging system. You should now able to interpret the results of your PCR.

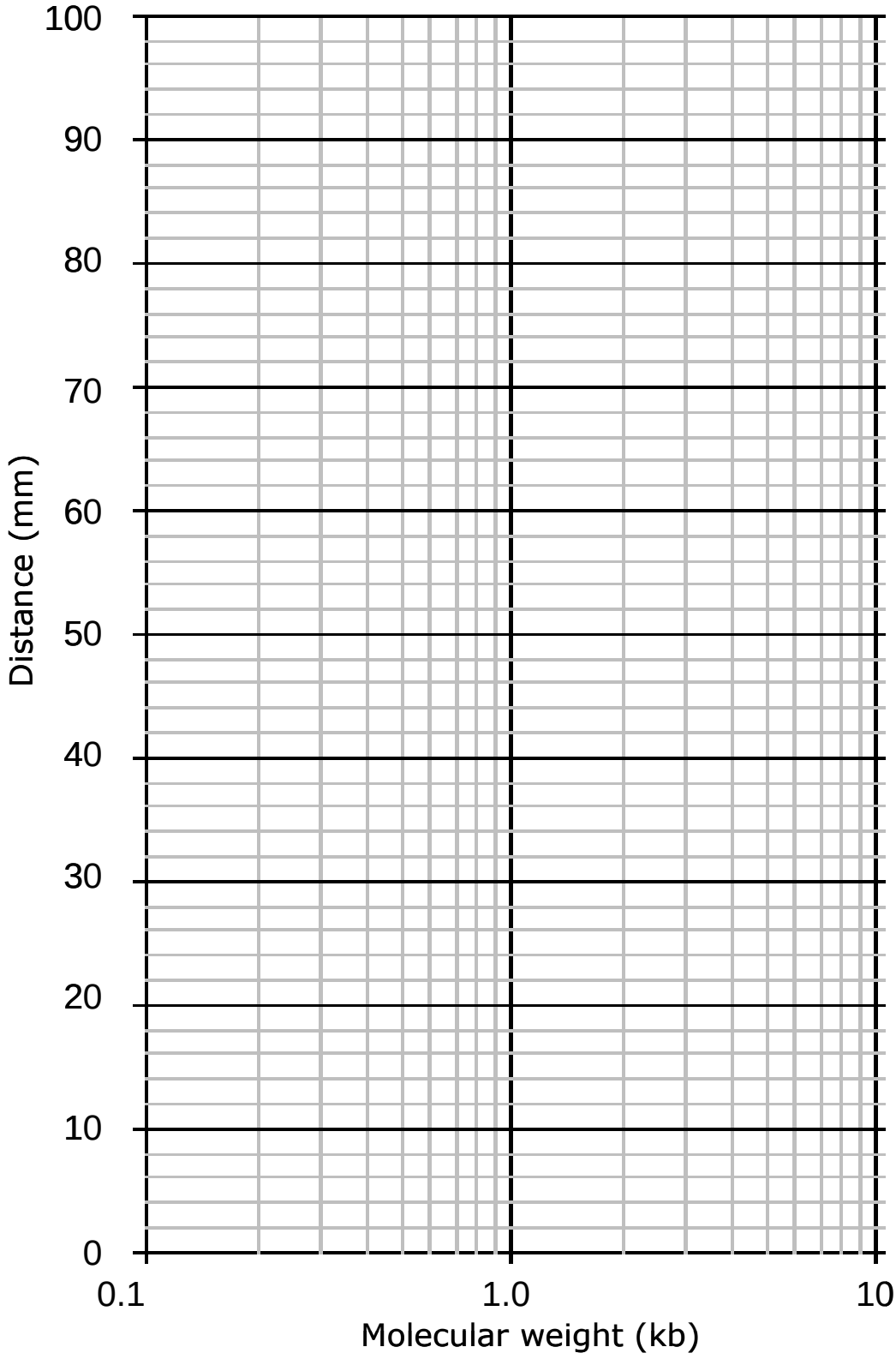


Paste your PCR gel here!

Figure 1 DNA Ladders. Here is a photograph (centre panel) of a successful PCR, showing a product from the primers and template you have used. The ladder, which you will be using today, is the Life Technologies™ 1kb DNA ladder (it's called a 1kb ladder because the high molecular weight fragments increase in size by 1kb increments, although on the 1% agarose gel the fragments >4kb are not well resolved); the fragment sizes in the 1kb ladder are shown in the left panel.

Ladder band sizes in bp (from top):
12216 / 11198 / 10180 / 9162 / 8144 / 7126 / 6108 / 5090 / 4072 / 3054 / 2036 / 1636 / 1018 / 506 / 396 / 344 / 298

- Using the log paper on the next page, plot the migration of the markers (linear, y-axis) against relative molecular weight in kb (logarithmic, x-axis). Depending upon the range of markers that you have used, you should then be able to construct a straight line fit through the data points. Finally, measure the migration of your PCR product and from the graph you have just constructed, estimate the relative molecular weight of the product in kb.


100
90
80
70
60
50
40
30
20
10
0
Distance (mm)
0.1
1.0
10
Molecular weight (kb)

Questions

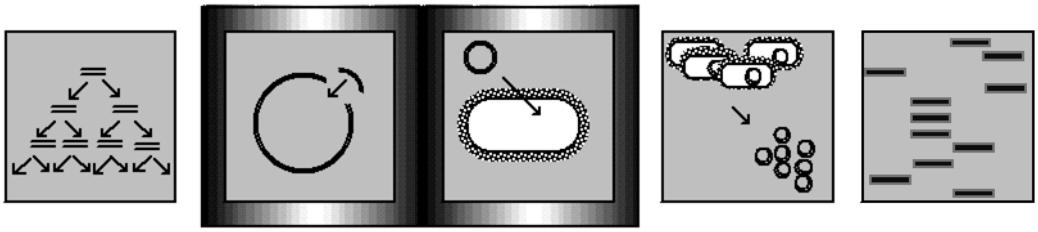
Q22. What is the purpose of the DNA cleanup stage of the process?

Q23. How "pure" was your reaction? Did the PCR pull out other, spurious bands? Why might this have happened?

Q24. The fragments in the DNA ladder have the following sizes (in base pairs) but all may not be easily resolved: 12216, 11198, 10180, 9162, 8144, 7126, 6108, 5090, 4072, 3054, 2036, 1636, 1018, 517, 506, 396, 344, 298 and some smaller pieces of DNA. Use this information to identify what size bands A-D are. Do the same with your own gel – can you identify the marker sizes?

Q25. We can "eyeball" the product and by comparison with the markers, estimate its size. How does this compare with the predicted size of the product, which you calculated in a previous question?

DAY 2 - CLONING A PCR PRODUCT

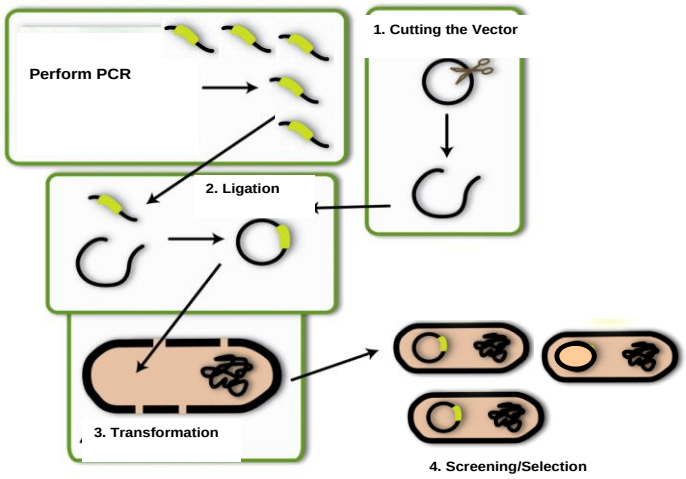


Introduction

Reading: Dale et al., 3rd Ed: p25-29.

Molecular cloning refers to the process of making multiple copies. Cloning is a commonly used method to amplify DNA fragments (such as a PCR product). You will be cloning your PCR product by **ligating** it into a plasmid vector. This process involves several steps:

1. **Breaking apart a plasmid vector** – a plasmid vector is a circular piece of DNA that has several key properties (see Appendix 4 for a diagram of the pBluescript vector you will be using in the lab).
2. **Ligation** – 'gluing' together the PCR product into the cut plasmid vector.
3. **Transformation** – inserting the newly formed pieces of DNA into bacterial cells. These transformed cells are then allowed to grow and multiply.
4. **Screening/selection** – selecting out cells that were successfully transformed with the new vector.



Although PCR is quick and straightforward, cloning PCR products can be a little problematical. Getting a PCR product to join into a plasmid vector is an inefficient and difficult process but there are several techniques that can be used to try to optimise the procedure. One of the most common methods of cloning involves the use of restriction enzymes (or to use the correct term, restriction endonucleases), which cut the DNA at a specific nucleotide sequence, known as a restriction site. In this lab, we used PCR primers that had restriction sites added near their 5' ends (a KpnI site for the forward primer, and a BamHI site for the reverse primer – see **Exercise 1: PCR and Primer Design** for more detail). The introduction of these specific changes into a PCR product during the reaction is called **PCR mutagenesis**, although the changes will be, by definition, at the ends of the PCR product. The use of different restriction sites at each end of the PCR product provides an important advantage when cloning.

The choice of restriction enzyme sites for this cloning experiment was governed by (i) cost: the enzymes should be cheap; (ii) compatibility: the sites should be present in multiple cloning sites of common vectors; (iii) ease: the enzymes should use compatible buffers, so that we can perform a double-digest; and (iv) site usage: the enzymes should not cut within the product we're trying to clone. For (iv), we can use computer programmes to search for restriction sites within a fragment of DNA of known sequence.

Restriction enzymes cut DNA at specific sites. They are derived from different bacteria (where they participate in a kind of defence system, chopping up foreign, but not "self" DNA), so they require individual reaction conditions to work optimally. The abilities to restrict **digest** DNA at **particular sites**, to **ligate** it into a **plasmid vector**, and to **transform bacteria** with it in order to amplify it, are central to **recombinant DNA technology**.

Method

1. Setting up Restriction Digests

Measure the volume of your cleaned–up PCR product by setting a Gilson P200 to the volume you think you have and sucking up the sample. If you fail to suck up the whole sample, or if you are sucking up air at the end, you need to eject the sample back into the tube, change the volume setting up or down respectively and try again.

Assemble the following components, **adding the enzymes last, on ice**:

NOTE: restriction enzymes are **thermolabile**, and **must be kept on ice** when not frozen.

Your cleaned PCR product (check what volume you have first)	up to 41 µl
10x restriction buffer	5 µl
KpnI enzyme	2 µl
BamHI enzyme	2 µl
Sterile distilled water	x µl
Final volume	**50 µl**

Seal. Mix by flicking the bottom of the tube, and then centrifuge for a few seconds to spin the liquid back down.

NOTE: restriction enzymes, like all enzymes in molecular biology will not thank you for vortexing them. This "tube flicking" is a gentler and thorough alternative.

Label the top of the tube clearly, and transfer it to a 37°C water bath float for at least 1 hour.

2. DNA Cleanup

The reaction conditions required by DNA ligase are very precise, and differ from those needed by KpnI or BamHI. So we must clean up the DNA again.

Use *all* of your digest (~50 µl) and **remember** to add an equal volume (~50 µl) of Membrane Binding Solution. Refer back to the "DNA cleanup" section on DAY 1 for the method, but **at the last stage elute with 25 µl TE buffer**.

3. Quantification of DNA

You now have to measure the amount of DNA you have obtained following the DNA cleanup. To do this you will use the NanoDrop spectrophotometer in the lab, which allows accurate and reliable quantitation of nucleic acids using a very small volume of your DNA sample. You will also be able to tell how pure your sample is and whether you have any contaminants present. If sample is pure (i.e. without significant amounts of contaminants such as any enzyme left from the restriction digest), you can use the spectrophotometer to measure amount of UV irradiation absorbed by the bases.

For quantifying DNA, absorbance readings should be taken at wavelengths of 260 nm and 280 nm (A_{260} and A_{280}). The reading at 260 nm allows calculation of the concentration of nucleic acid in the sample. The relationship between DNA concentration and A_{260} is as follows:

A_{260} of 1.0 for double-stranded DNA = 50 ng/µl of dsDNA

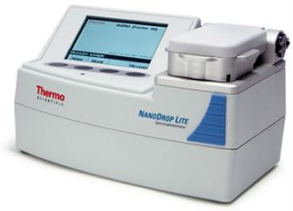
The reading at 280 nm gives the amount of protein in the sample. Pure preparations of DNA have a A_{260}/A_{280} value of 1.8 to 2.0. If there is contamination with protein, this ratio will be significantly less than the values given above, and accurate quantitation of the amount of nucleic acid will not be possible. So, for DNA quantification the formula becomes:

DNA concentration = A_{260} x 50 ng/µl (**Note**: if you have to dilute your DNA sample then the dilution factor will have to be taken into account in this calculation)

Fortunately for you, NanoDrop does all the measurements and calculations for you and all you need to do is read the final concentration!

To use NanoDrop, you will have to (with the help of a demonstrator!):

- Take your DNA from the cleanup over to the NanoDrop and, using a P2 gilson pipette, take 2 μl and place this carefully on the lower optical surface of the machine




- Select the 'Measure' function and take a note of the concentration reading below:

DNA concentration =

4. Setting up a Ligation Reaction

NOTE: DNA ligase and ligase buffer are **particularly heat labile**, so should be kept on ice.

Assemble the following**, adding the enzyme last,** in 3 eppendorfs labelled with your initials:

	Tube 1 (plasmid + insert)	Tube 2 (plasmid only)	Tube 3 (ligase control)
Double Cut Plasmid (200ng/μl) (pBluescript **plasmid** precut with *Bam* HI and *Kpn*I)	200ng	200ng	-
Your *Bam* HI /*Kpn*I cut PCR product (the "insert")	100ng	-	-
Single Cut Plasmid (200ng/μl) (pBluescript plasmid precut with *Kpn*I)	-	-	200ng
5x ligase buffer	4μl	4μl	4μl
Sterile distilled water	to 19μl	to 19μl	to 19μl
T4 DNA ligase enzyme	1μl	1μl	1μl
Final volume	**20μl**	**20μl**	**20μl**

- The final volume in all tubes should be 20 μl. **Add the ligase last.**
- Incubate the ligation reaction at room temperature for at least 1 hour.

Questions

Q26. Given the relative sizes of pBluescript and the PCR product, what insert:vector molar ratio have we set up in tube 1?

Q27. What is the advantage of having different restriction sites at either end of our PCR fragment?

Q28. What is the purpose of the DNA ligase in the ligation reaction?

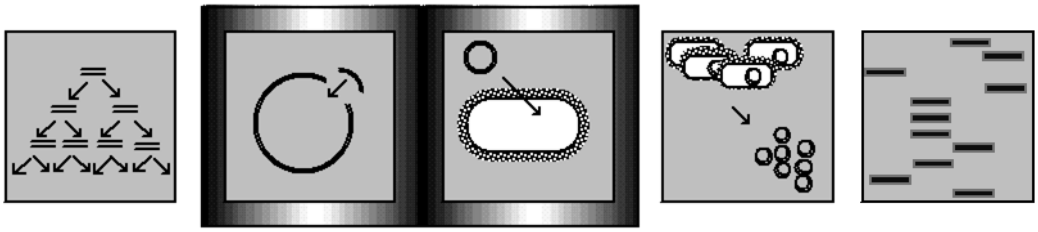
Q29. What would happen if DNA ligase was not used in the creation of the recombinant plasmid?

Q30. If you had a purified stock of DNA and you dilute 1μl of this purified DNA sample into a total volume of 50μl distilled water. You then measure the absorbance of this diluted sample in a NanoDrop at 260 and 280 nm and obtain the following readings: A_{260} = 0.550; A_{280} = 0.364.

a) What is the DNA concentration of the stock? Express your answer as μg/μl.

b) What is the A_{260}/A_{280} ratio of the purified DNA? Comment on the purity of the sample.

DAY 2 - TRANSFORMATION OF COMPETENT E.COLI CELLS



Introduction

Reading: Dale et al., 3rd Ed: p59-61.

By now, we should have ligated (or glued) our PCR product into a plasmid vector. Next we have to put it into a prepared bacterial *Esherichia.coli* "**host"** strain to grow it up: this process is called **transformation**. Since both the ligation we have just done and the transformation we are about to perform are not very efficient, we will use two methods to **select** only those bugs that contain a PCR-product containing plasmid. Selection is an important concept in this work!

NOTE: Be scrupulously clean, or some other bug may land in your cultures and contaminate them.

Method - Transformation of Competent Bacteria

NOTE: When adding DNA to the cells, make sure you have actually taken up liquid into the pipette tip and make sure that this liquid is pipetted directly into the cell suspension (not onto the side of the tube).

1. **Competent bacteria** are prepared by chilling in the presence of divalent cations such as Ca^{2+} (as in $CaCl_2$), which makes the cell membrane more permeable. You will be given -on ice- aliquots of competent *E. coli* strain DH5α which have been prepared for you and resuspended in 100mM of $CaCl_2$.

NOTE: Competent cells are very fragile. Pipette very gently; mix very gently; keep on ice and avoid temperature changes until you deliberately heat shock them.

2. Take 5 aliquots of 50 µl competent cells in screw-top vials and thaw on ice. Label as per the samples below:
3. Add to your 50 µl aliquots of competent cells:
    1. 2 µl of your "Plasmid + insert" ligation (Tube 1)
    2. 2 µl of your "Plasmid only" ligation (Tube 2)
    3. 2 µl of your "ligase control" ligation (Tube 3)
    4. 2 µl of a "teacher's ligation" of PCR product plus plasmid which is known to have worked
    5. Each pair should also set up ***one*** of the following, such that each control is carried out by your bench as a whole:
    (a) Uncut pBluescript (2 µl of 10 pg/µl stock)
    (b) Cells only (without DNA)
    (c) Double cut, unligated plasmid (2 µl of pBluescript cut with BamHI & KpnI)

The significance of these important transformation controls will become clear later.

4. Tap the tube gently to mix the DNA and cells, then leave 30 min on ice.
5. Meanwhile, label 5 selective plates (L-agar which contains 50 µg/ml ampicillin, 50 µg/ml IPTG and 50 µg/ml X-gal), one for each reaction, with your initials and sample details. Also label the control, L-agar only plate, with your initials. NOTE: always label petri dishes neatly on the side of the base, not the lid.
6. Heat-shock the transformation mixtures in a 42°C water bath for EXACTLY 1 min.
7. Put the tubes straight back onto ice for 2 min.
8. Add 200 µl sterile L-Broth to each tube. Seal. Incubate at 37°C for 30 min. This is called the **expression** step.
9. Pipette 100 µl of each transformation onto the ampicillin plates and spread with a sterile plastic spreader. For the transformation of your own plasmid + insert ligation, pipette 50 µl onto the control L-agar plate (without ampicillin) and spread. **This plate represents a further control (d).**
10. Incubate all the plates upside down in the 37°C incubator overnight.

NOTE: Plates are usually incubated upside down.

Questions

Q31. What is meant by 'competent' cells?

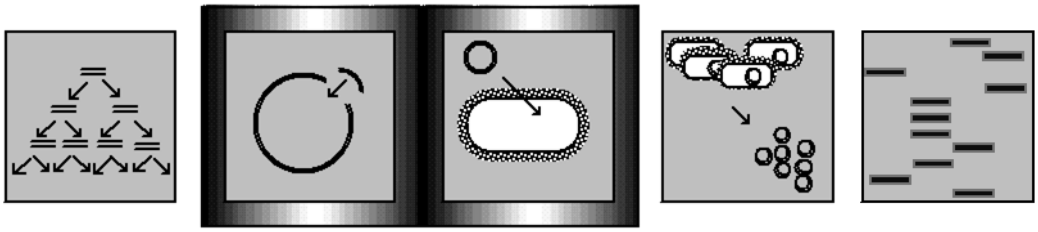
Q32. The agar plates should be labelled on the base and NOT the lid (see step 5 in the above method). Why is this?

Q33. Why are the plates incubated upside down (see step 10 in the above method)?

Q34. Why is the expression step (see step 8 in the above method) important?

Q35. You set up a further control whereby your cells were plated out onto a plate without ampicillin (control (d) - see step 9). What do you expect to see on this plate tomorrow and why is this control important?

DAY 3 - ANALYSING YOUR TRANSFORMED BACTERIAL CELLS



Introduction

In the process described in the previous section, transformation efficiency is very poor and only a small percentage of the bacteria will actually take up the plasmid. Bacteria that have been transformed must be selected, or isolated. The plasmid is too small to be seen, so it is useful for it to contain a gene that expresses a characteristic that can be seen or interpreted, for example, antibiotic resistance. Plasmids can have multiple genes inserted and these can be used to select for bacterial cells containing a plasmid. In the case of this lab, the cells containing plasmid with your ligated gene will grow on agar plates that contain the antibiotic ampicillin and a sugar known as X-Gal. The colonies of bacteria on your plates should be either white or blue in colour.

1. Analysing your transformations

With luck, you should have colonies on at least some of your plates. This section will guide you through their significance.

Q36. Briefly, explain the significance of the difference between a blue and a white colony.

Before analysing your own plates, think about the following table, considering the **expected results** first.

Q37. Complete the Table below, explaining what you expect to see, and what you actually see, on each of the plates. Score as "none", "few" (1-10), "some" (10-100) or "lots" (100+).

Plate Description	Expected Result		Your Result		Overall Class Result	
	Blue	White	Blue	White	Blue	White
1) plasmid: PCR product ligation						
2) Plasmid only ligation						
3) Ligase control						
4) Teacher's ligation						
5) (a) Uncut pBluescript						
(b) Cells only (without DNA)						
(c) Double cut, unligated plasmid						
(d) Cells only – no ampicillin selection						

Discuss the following questions, **explaining how you arrive at your conclusions**:

Q38. Which of the control plate(s) tell you that the antibiotic selection worked? Did this selection work on your own plate(s) and for the majority of the class?

Q39. Did the blue/white screening work in your plates?

Q40. Note the appearance of control plate grown in absence of ampicillin (d). You should have seen a bacterial lawn, or that the entire plate was covered with bacteria. Why does a bacterial lawn appear on this plate while other plates with growth have distinct, punctate colonies?

Q41. Which of the control plate(s) tell you that the cells survived the transformation procedure? Did your own cells survive this process?

Q42. One of the controls you set up for the ligation reaction was to test whether the DNA ligase was working or not. What was your result for this control, following the transformation reaction? Comment on the importance of including this control as one of the ligation reactions you carried out.

Q43. In a restriction digest it is not uncommon for a few DNA molecules to escape digestion (cutting). One of the transformation controls ((c) double cut, unligated plasmid) set up by the class will tell us whether our plasmid preparation contained any residual circular (uncut) plasmid, which had not been cut by either enzyme. Was the pre-cut pBluescript plasmid vector completely cut – comment on your own result and the class as a whole? If not, can you think why that might be?

2. Calculating Transformation Efficiency

Aim

To determine how well your *E.coli* competent cells took up and expressed plasmid vector.

Background

Your next task is to determine the extent to which you genetically transformed *E.coli* bacterial cells. This quantitative measurement is referred to as the **transformation efficiency**. In many experiments it is important to genetically transform as many cells as possible. The transformation efficiency is calculated to help scientists determine how well the transformation is working.

Transformation efficiency is a number. It represents the total number of bacterial cells that express a plasmid vector, divided by the amount of DNA used in the experiment. It tells us the total number of bacterial cells transformed by one microgram of DNA. To work this out in this lab, use the control plate that was transformed with uncut pBluescript plasmid (control plate (a)).

To calculate the transformation efficiency, you require to know the following information:

10pg/μl = concentration of uncut pBluescript plasmid DNA

2μl = volume of the above plasmid solution used for transformation

100μl = volume spread on the agar plate

250μl = total volume of transformation reaction (50μl of bacterial cells plus 200μl L-Broth)

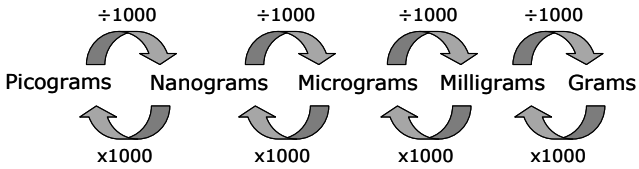
1. Determine the Concentration of DNA in μg

In molecular biology you often use extremely small measures, such as picogram or nanogram amounts of DNA. It is useful to be able to convert between different units of measurement.

Remember:

NAME	ABBREVIATION	EQUIVALENT
1 gram	g	1000mg
1 milligram	mg	1000μg
1 microgram	μg	1000ng
1 nanogram	ng	1000pg
1 picogram	pg	1/1000ng

To convert from smaller units to larger ones, we need to divide by 1000, or to convert from a larger unit to a smaller one, we need to multiply by 1000.



As seen in the table above, each smaller unit has 1000 of them to equate to the larger unit. For instance, 1 mg is equivalent to 1000μg. To convert to a smaller unit you need more of them so multiply.

For example:

Micrograms to *picograms (μg to pg):* 1,000,000 x μg = pg

For example, if μg = 143 then pg = 1,000,000 x 143 = 143,000,000pg or 1.43×10^8 pg

Picograms to *micrograms (pg to μg): pg* ÷ 1,000,000 = μg

For example, if pg = 143 then μg = 143÷ 1,000,000 = 0.000143 μg or 1.43×10^{-4} μg

Q44. Given the concentration of uncut pBluescript plasmid DNA is 10pg/μl, convert this to μg/μl

2. Determine the Total Number of Transformed Cells

Have a look at a control plate representing the cells transformed with uncut pBluescript plasmid (control plate (a)). Each colony on the plate can be assumed to be derived from a single cell. As individual cells reproduce, more and more cells are formed and develop into what is termed a colony. The most direct way to determine the total number of cells containing uncut pBluescript is to count the colonies on the plate. Enter the number of colonies below:

#Colonies =

3. Determine the Fraction of pBluescript DNA in the Bacterial Cells spread on the L-agar plate.

A. Total amount of pBluescript plasmid DNA:

(DNA in μg) = (concentration of DNA in μg/μl) x (volume of plasmid in μl)

B. Fraction of pBluescript plasmid DNA (in the bacteria) that actually got spread onto the L-agar plate:

Fraction of DNA used = Volume spread on L-agar plate (μl) ÷Total volume of the transformation reaction (μl)

Therefore, pBluescript DNA spread (in μg) = A x B

4. Transformation Efficiency

Based on the calculations from Step 2 and Step 3, determine the transformation efficiency:

Transformation Efficiency = Total number of colonies growing on the agar plate ÷ amount of DNA spread on the agar plate (in μg)

Enter that number →

Transformation efficiency = transformants/μg

5. Transformation Efficiency – How Successful was Your Transformation?

Q45. The transformation protocol you used generally has a transformation efficiency of $2x10^8$ to 10^{11} transformants/μg plasmid DNA. How does this compare with your own result?

Q46. Note below the transformation efficiency of several others in the class – how does your result compare with theirs?

Q47. Thinking about the transformation protocol, what factors do you think might influence transformation efficiency?

3. Setting up an 'Overnight' Culture

Introduction

E.coli grows well either on nutrient agar (as last time) or in vigorously shaken liquid culture. It is this latter form of bacteria that we need for tomorrow. Liquid cultures are used to prepare large amounts of recombinant plasmid DNA, and until the advent of PCR, this was the *only* way of amplifying DNA. Because it is cheaper, can handle larger DNA fragments and is less prone to introducing sequence errors, it is still the method of choice for many applications. In this section, we choose single colonies from your plates, and add them to nutrient L-broth containing antibiotic, and grow them overnight with shaking.

CAUTION: ***If the experiment has succeeded, you are handling recombinant bacteria. Observe good microbiological practice. Dispose of all solutions into disinfectant. Clear up and disinfect any spills.***

Method

Set up 2 overnight cultures:

1. You are provided with 20 ml L-broth, to which you must add ampicillin to a final concentration of 50 μg/ml (stock solution is 50 mg/ml). To do so, use the following calculation:

$$\frac{\text{Final concentration}}{\text{Stock concentration}} \text{ X final volume} = \text{volume to add to solution}$$

- Check with a demonstrator that you have calculated the amount correctly.
- Put approx 10 ml into each of 2 culture tubes, labelled clearly on the tube, **not the lid**, with your initials and "blue" or "white".

2. Inoculate a blue colony as a control into the first tube (these bacteria should carry plasmid vector with no insert). To pick the bacteria, use the sterile plastic loop provided and touch an isolated, single blue colony, making certain that some of the cells have been transferred to the loop. Drop it into the liquid L-broth and shake the loop a bit. Replace the tube's cap as soon as the inoculation is complete.
3. Inoculate a white colony (which should be bacteria carrying recombinant plasmid [vector + insert]) into the second tube using the same method as above. If you didn't get any white colonies, the demonstrator will provide you with a plate.

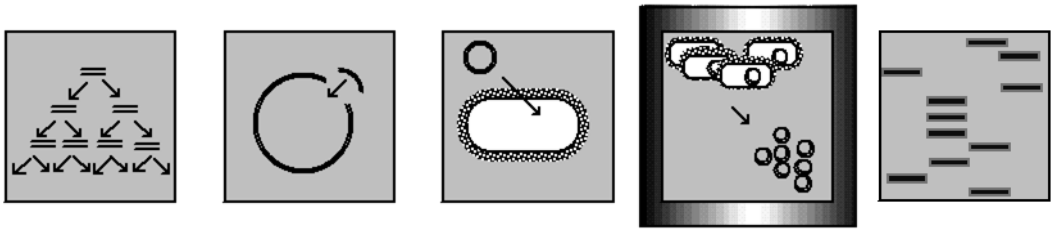
The cultures will be grown overnight in a shaking incubator at 37°C.

Questions

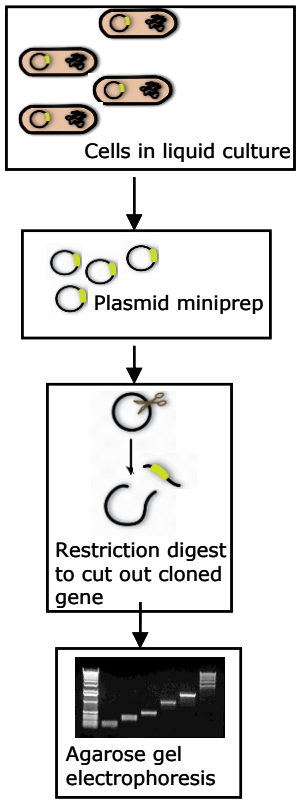
Q48. Suppose you have a stock concentration of ampicillin at 100mg/ml and you want to make 5ml of L-broth having 50μg/ml. What volume of stock ampicillin should you add?

Q49. If ampicillin was prepared as a 500X stock solution, how many μl would you have to add to 25ml to make it a final 1x concentration?

DAY 4 – PLASMID MINIPREP



Introduction



We hope that the plasmid isolated from the white colony will be a recombinant plasmid which carries our PCR product. There are several ways to check this, for example, by PCR, restriction digest or sequencing. We will be doing a restriction digest to determine if the plasmid carries the *Nrg* gene amplified DNA. First, we will harvest plasmid DNA from the bacteria grown overnight then use restriction enzymes (BamHI and KpnI) to cut out the cloned PCR product from the plasmid (this should only happen in bacteria from the white colony). Finally, we will run the samples on an agarose gel to check the size of the resulting DNA. This should allow us to see if the "white" plasmid does indeed contain an insert of about the correct (ie PCR product) size.

We start by harvesting the cells grown in liquid culture by spinning them down to a pellet, then extracting the plasmid. There are several ways of doing this, the simplest is to resuspend the pellet, and lyse the cells in a mixture of alkali and detergent. When the alkali is then neutralised, almost everything *except* the plasmid is precipitated down (unlike the proteins and chromosomal DNA, plasmid DNA can renature under these conditions).

These experiments are performed on various scales, and are named on the basis of the volume of the "overnight". Usually, a set of plasmid minipreps is produced from 5–10 ml overnight cultures, and the plasmids are digested or sequenced to establish which contain the insert of interest. If more DNA is required a midiprep (about 50 ml) or a maxiprep (about 400–1000 ml) can also be prepared.

1. Plasmid extraction

A brief outline of the procedure shown here:

Remember that you are working with **recombinant bacteria**

Spin down cells. Resuspend in buffer

↓

Lyse cells with detergent. Neutralise to precipitate proteins & genomic DNA

↓

Spin and save supernatant. Collect plasmid on a Minicolumn. Wash with alcohol.

↓

Elute DNA with aqueous buffer

Now collect your overnight cultures and mix briefly to resuspend the bacteria; each person should do one miniprep, noting which is which! If your overnights failed, borrow someone else's - there will be spare liquid.

1. Your aim is to produce a pellet corresponding to 4.5 ml culture in an eppendorf tube that only holds 1.5 ml. For each culture:
   - Put 1.5 ml (2 x 750 µl) culture in a labelled eppendorf. Centrifuge at 10,000rpm for 2 min. Decant the liquid supernatant into disinfectant.
   - Add a further 1.5 ml of the same culture to the pellet in the same tube (be careful not to mix up the "blue" and "white" cultures), spin down cells as above and remove supernatant by pouring or with a pipette.
   - Repeat the above step. You should now have a single cell pellet from 4.5 ml of bacterial culture.

Get rid of **all** the supernatant after the last spin (any medium remaining can be briefly spun to the bottom of the tube and removed using a P200 Gilson). *Any contaminants could inhibit restriction enzymes later on.*

2. Add 250 µl Cell Resuspension solution. Mix thoroughly by pipetting up and down with a P200 (plus yellow tip). *Any lumps of cells that remain will not lyse in the next step, reducing your yield.*
3. Add 250 µl Cell Lysis solution. Invert tube four times to mix; don't vortex. *This lyses the cells in a mixture of alkali and detergent.* Liquid should go visibly "gloopy" as the chromosomal DNA is released. Incubate at room temperature for about 2 min until the cell suspension clears.
4. Add 10 µl Alkaline Protease Solution and mix by inverting the tube 4 times. Incubate at room temperature for 5 min. Alkaline protease inactivates endonucleases and other proteins released during the lysis of the bacterial cells that can adversely affect the quality of the isolated DNA. **Do not exceed 5 minutes of incubation with this enzyme, as nicking of plasmid DNA may occur!**
   From now on, be careful and don't treat the DNA violently. The separation of chromosomal DNA from plasmid is based on size, and if you shear the chromosomal DNA into smaller pieces, it will copurify with the plasmid.
5. Add 350 µl ice-cold Neutralisation solution. Invert the tube sharply four times to mix thoroughly. You may see a white flocculent (flaky) precipitate. *The rapid neutralization causes the protein and chromosomal DNA to precipitate, the plasmid however is able to renature under these conditions.*
6. Spin down at maximum speed for 10 minutes.

Plasmid DNA will now be purified from the bacterial lysate using microcentrifugation to force the cleared lysate through a Minicolumn and the plasmid DNA, which has bound to the column, will be washed.

7. Label the Minicolumn with your initials. Transfer about 800 µl of the cleared lysate to the Minicolumn. *Take care not to transfer any of the white precipitate; this would reduce the purity of the product.*
8. Centrifuge the supernatent at maximum speed for 1 minute. Remove the Minicolumn from the tube and discard the flowthrough from the Collection Tube. Reinsert the Minicolumn into the Collection Tube.
9. Add 750 µl of Column Wash Solution to the Minicolumn. Centrifuge for 1 minute as before. Remove the Minicolumn from the tube and discard the flowthrough. Reinsert the Minicolumn into the Collection Tube.
10. Repeat step 9 using 250 µl of Column Wash solution and centrifuge this time for 2 minutes.
11. Place the Minicolumn into a labelled 1.5ml eppendorf (remove the lid), being careful not to transfer any of the Column Wash Solution.
12. Elute the DNA by adding **40 µl TE buffer** directly into the column using a P200 Gilson pipette, then spinning for 1 minute. After eluting the DNA, discard the Minicolumn.
    Again, note that residual alcohol in the column is the main reason for poor yield.
13. When finished, place a lid on the tube and **label it carefully** with your initials and "blue plasmid" or "white plasmid" and keep on ice. You are now ready to set up your restriction digests.

2. Digestion (cutting) of your plasmid DNA with KpnI and BamHI

You should now have plasmid DNA purified from (i) a white colony and (ii) a blue colony. You will now digest a sample of each plasmid with BamHI and KpnI. In **fresh eppendorfs**, set up the following reaction, **one digest for each of your plasmids**, as follows (40 µl final volume):-

- 22 µl water
- 4 µl 10x restriction buffer
- 10 µl of your plasmid DNA preparation, then add
- 2 µl KpnI
- 2 µl BamHI

Flick the tube and spin down briefly in the microfuge. Incubate ≥ 1 hour at 37°C. **This will be your "cut" plasmid**. Meanwhile**,** store the remainder of your plasmid preparation on ice **(this is your "uncut" plasmid).**

3. Agarose gel electrophoresis

Prepare your samples for gel electrophoresis as follows:

- To the tube containing your 5 µl sample saved from the original PCR, add 15 µl water plus 6 µl loading dye.
- For each "uncut" plasmid, pipette 3 µl into a small eppendorf and add 17 µl water and 6 µl loading dye.
- For each "cut" (ie restriction digested) plasmid, pipette 20 µl into a small eppendorf and add 6 µl loading dye.

On a 1% agarose gel in TAE buffer, run pairs of "uncut" and "cut" samples of each of your plasmids and the PCR sample (with the **PCR next to the "white cut"** sample). **Load the whole volume for each sample**. Your gel should also have a **DNA ladder**.

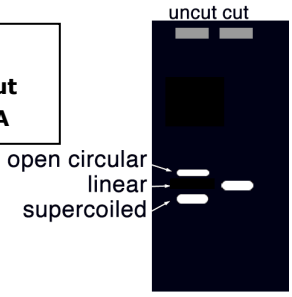
More than one group will be using the gel. Make sure there is at least one well left empty between your samples and those belonging to the next group. Mark on the sheet provided which samples are in which lanes.

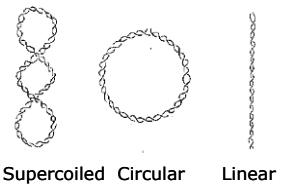
Run then photograph your gel.

4. Interpreting Your Gel

Interpreting the gels showing plasmid DNA is not always straightforward and requires practice. In order to interpret your results, first study the following and read the notes to help you understand.

Representation of an Agarose Gel showing Cut and Uncut Plasmid DNA





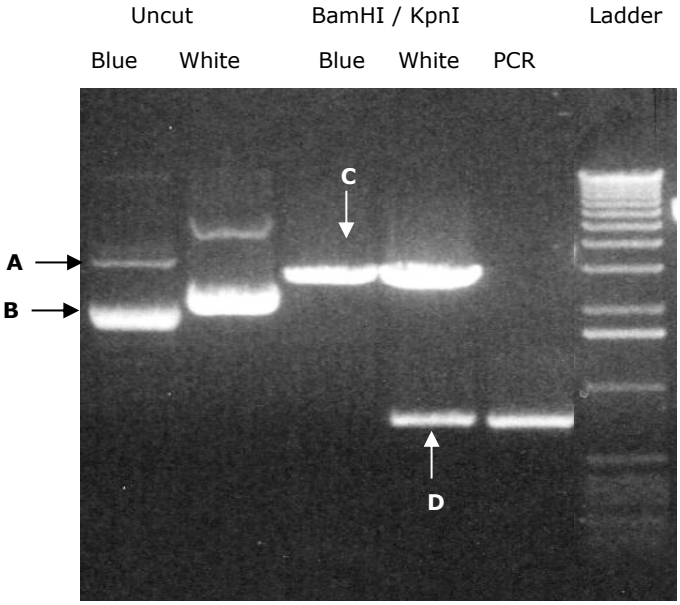
Representation of Plasmid DNA Conformations

(a) **Movement of DNA fragments.** Pieces of DNA, for example, DNA fragments cut by restriction enzymes or PCR products (like the gel you ran on Day 1), are **linear** pieces of DNA. They separate in an agarose gel according to size. The smaller pieces travel the fastest and farthest through the gel matrix. In the 'cut' lane of the above representation of a gel, this shows plasmid DNA that has been cut open with a restriction enzyme and is therefore now in a linear form. It has run as a single band on the gel.

(b) **Conformations of uncut plasmid DNA**. Unlike linear DNA, uncut plasmid DNA has several distinct conformations that can be identified when the uncut plasmid is electrophoresed in an agarose gel.

- **Supercoiled DNA** is the fastest conformation of the uncut plasmid. It can wind up into a compact structure. Impact taking a circle of string and rolling it around in your hands until it forms a little ball. Because of its compact shape, supercoiled DNA is the fastest moving conformation in the gel.
- **Circle DNA** is also called relaxed DNA. This has a more like a floppy circular conformation and therefore is the slowest moving conformation in an agarose gel.

Now, have a look at your own gel and also a similar one shown below, then answer the questions.



The gel opposite shows a typical result. Samples are run against a DNA ladder (remember, this contains **linear DNA fragments of known sizes**).

The first two lanes are uncut plasmids from a blue colony and a white colony, respectively.

The second two lanes show BamHI+KpnI digested plasmids from the blue and white colonies. The white colony plasmid is seen to have an insert of about the same size as the original PCR product in lane 5.

Questions

Answer the following questions using both the gel photograph shown on the previous page and your own gel photograph.

Q50. Why can we not calculate accurate sizes for the **uncut** plasmids from this gel?

Q51. On the gel photograph, why does the "cut" pBluescript plasmid in lane 3 (labelled 'C') run more slowly than the major band of "uncut" pBluescript plasmid in lane 1 (labelled 'B')?

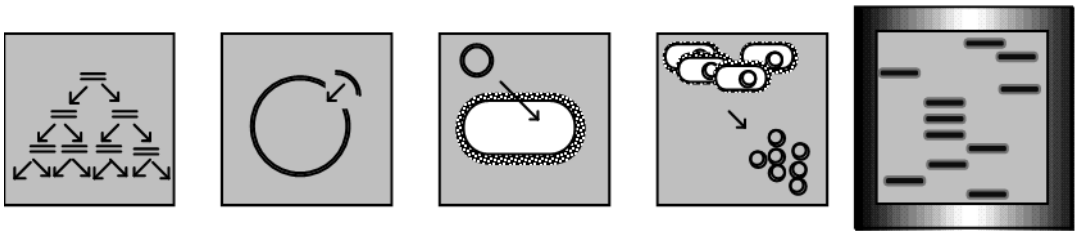
Q52. Why are there multiple bands in the "uncut" pBluescript lane (labelled 'A' and 'B')?

Q53. Following on from the above question, why do we not see multiple bands in the "cut" pBluescript lane?

Q54. Is there evidence for any contamination with *E. coli* chromosomal DNA or with RNA? Where would you expect these nucleic acids to appear on the gel?

Q55. How does the size of the insert (labelled 'D') compare with that of the original PCR product on your gel photo?

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.

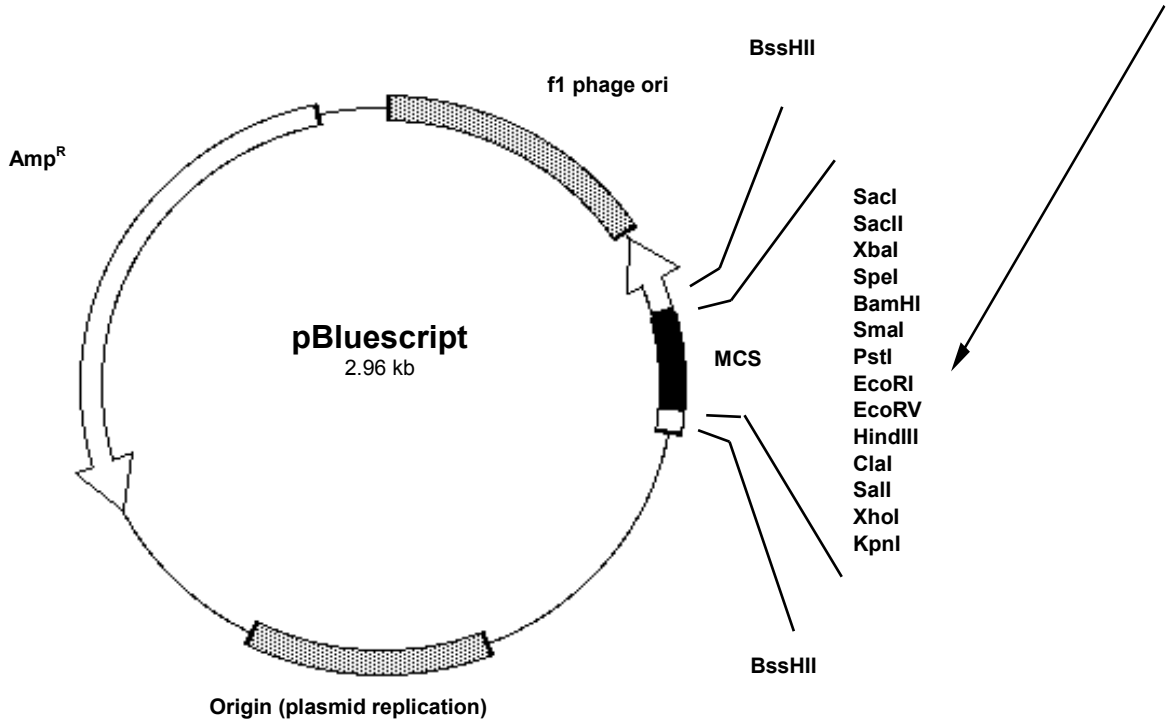
The **multiple cloning site** (MCS) present in the pBluescript vector you have been using is indicated below. The forward and reverse primer sequences are highlighted in bold:

Forward DNA sequencing primer

```
5'GTAAAACGACGGCCAGT3'

                  BssHII          T7 promoter                    SacI    SacII          XbaI  SpeI  BamHI SmaI  PstI  EcoRI
600                 |                                            |       |              |     |     |     |     |     | 71
2
  GTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATT
C
  CATTTTGCTGCCGGTCACTCGCGCGCATTATGCTGAGTGATATCCCGCTTAACCTCGAGGTGGCGCCACCGCCGGCGAGATCTTGATCACCTAGGGGGCCCGACGTCCTTAA
G


EcoRV     ClaI    HincII
|   HindIII |      SalI   XhoI          KpnI                              BssHII
713   |     |        |     |             | 760                             |       800                      82
6
GATATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCC
CTATAGTTCGAATAGCTATGGCAGCTGGAGCTCCCCCCCGGGCCATGGGTCGAAAACAAGGGAAATCACTCCCAATTAACGCGCGAACCGCATTAGTACCAGTATCGACAAAGG

                                                  T3 promoter
                                                                                  3'GTACCAGTATCGACAAAGG5
'
```

This is the sequence of the MCS region of the plasmid

Reverse DNA sequencing primer

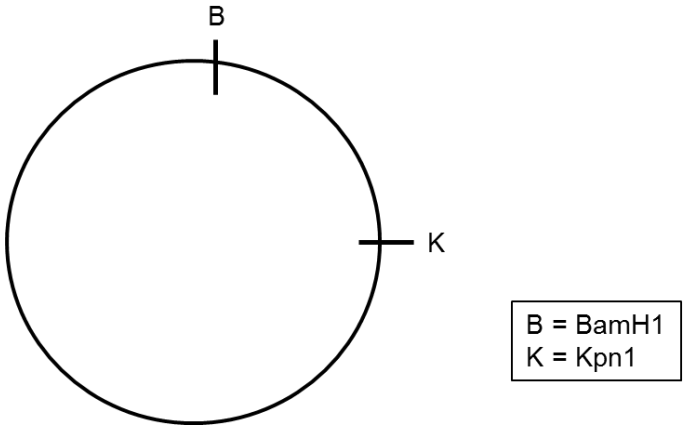


Q56. Below is a schematic diagram representing the pBluescript plasmid cloned with the *Nrg* gene sequence that you amplified by PCR. Can you label this diagram with the following (also state what size in bp you think these are):

- 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



Q57. 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.

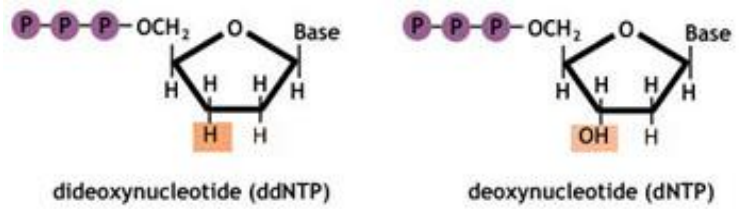
Components of a Sequencing Reaction:

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.

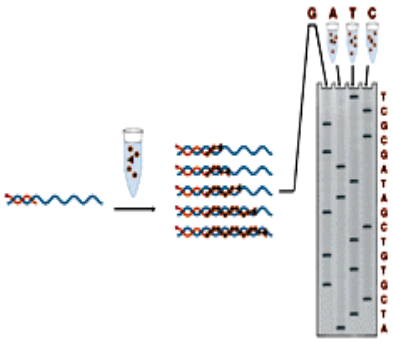
The vector containing the *Nrg* gene that you cloned will be sequenced by automated sequencing. In order to understand the biology behind the technology you will first do an exercise on manual sequencing.

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

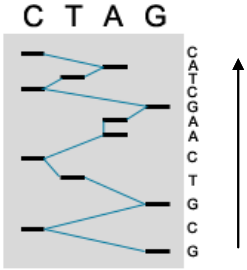
- The ddNTPs lack a 3' hydroxyl group, so when they are incorporated by chance into a growing DNA strand, further extension of that strand becomes impossible. The ddNTPs are radiolabelled so can be detected by autoradiography.

- 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 filn

- 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'
… but if ddTTP was included in the reaction mixture at a low concentration, these termination products would also result -		5'	GCGCCCT	3'
		5'	GCGCCCTGAACT	3'
		5'	GCGCCCTGAACTCCT	3'
		5'	GCGCCCTGAACTCCTT	3'

Questions

Q58. 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.

Q59. 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

Q60. 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.

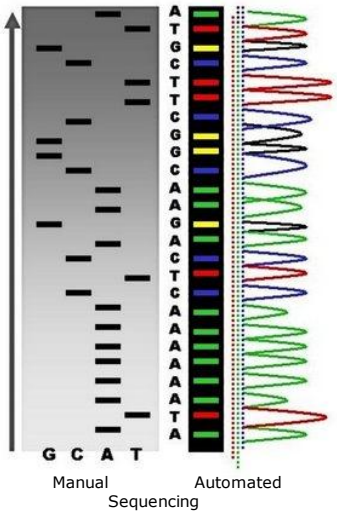


Figure 1 Manual and Automated Sequencing

- 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

Q61. Why in Figure 1 are there different coloured fragments in the middle lane of this diagram? How does this differ from manual sequencing?

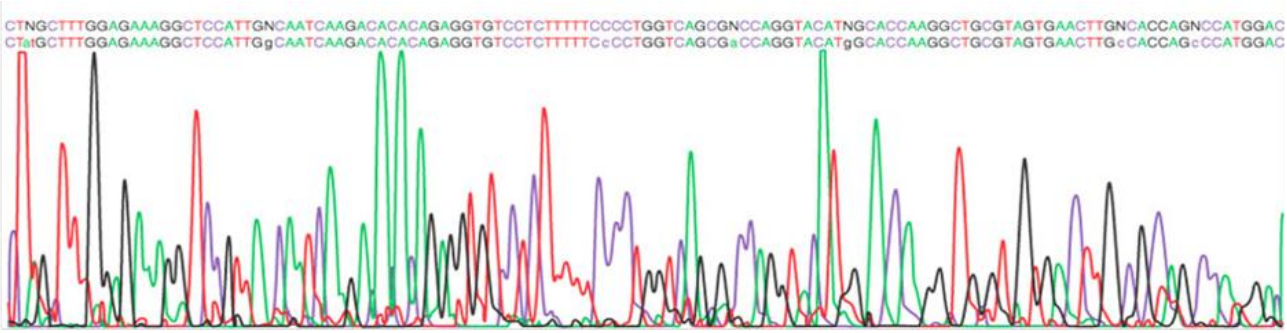


Figure 2 Example of a chromatogram resulting from an automated sequencing reaction

Q62. 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?

Q63. Further along on Figure 2 locate the sequence CCATT. Is this the correct sequence?

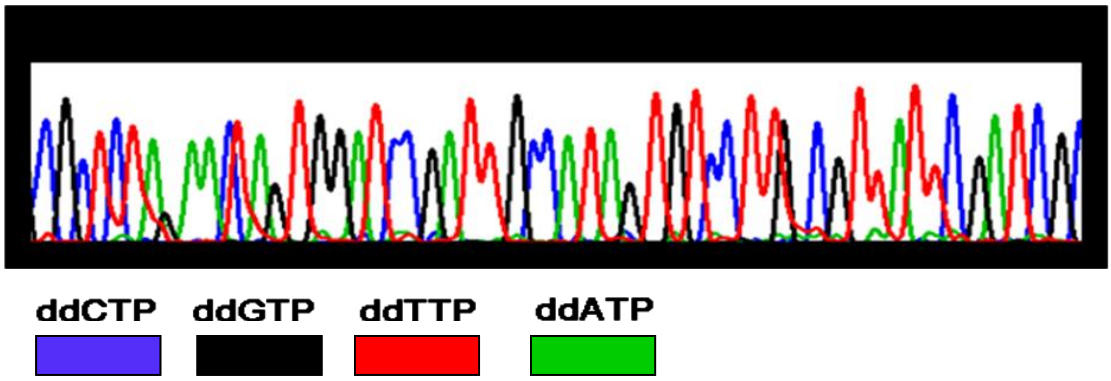


Figure 3 A chromatogram for you to interpret

Q64. Once you are confident about how to collect and interpret sequence data in this way, gather sequence from the chromatogram shown in Figure 3.

Q65. 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.

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.

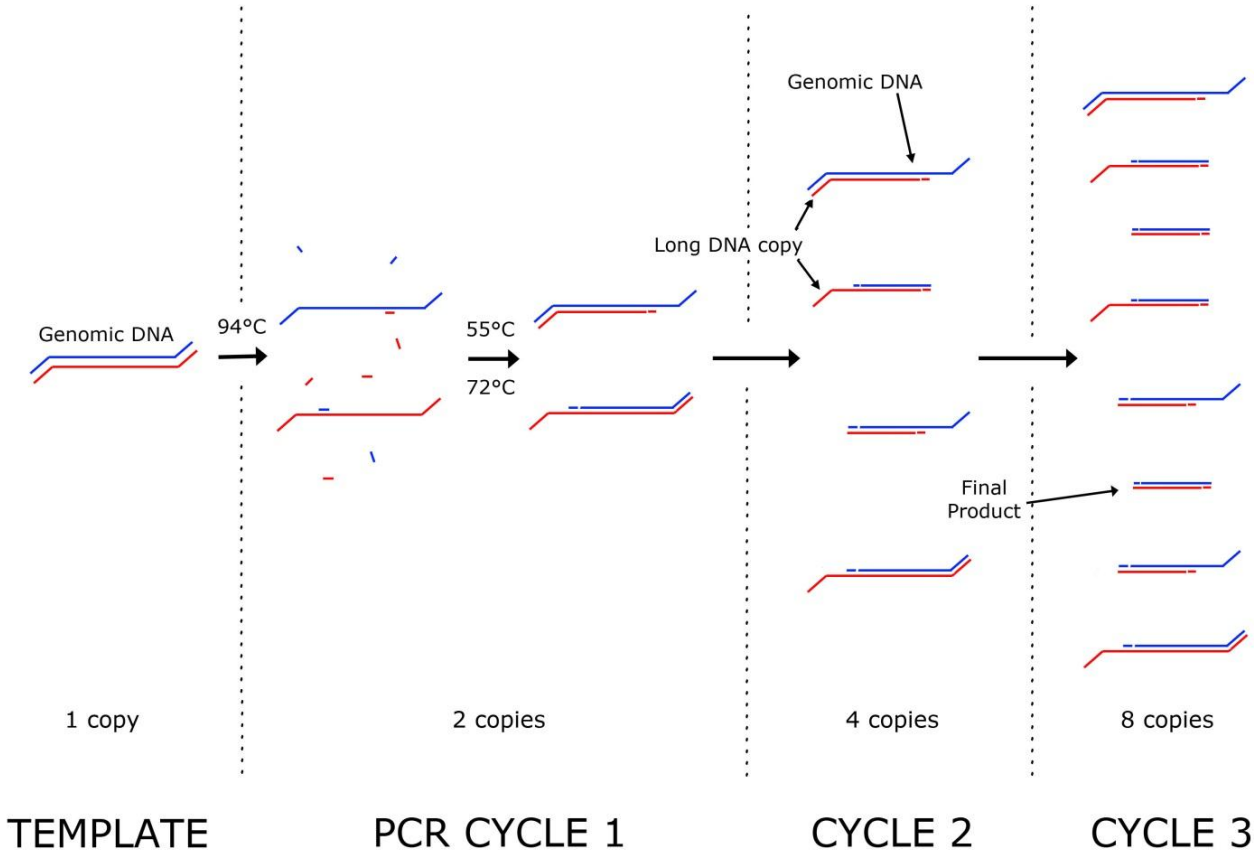


Figure 1 First Three Cycles of a Polymerase Chain Reaction

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.

Q66. **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.

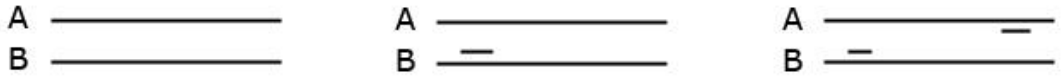
Cycle 3: The process is repeated, and primers anneal to the newly synthesised strands. Taq polymerase synthesises complementary strands, producing double-stranded DNA fragments that are identical to the target sequence. The process is repeated and the number of target fragments doubles for each subsequent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

Q67. 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



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						

Q68. 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 send 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 send 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.

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′ AGGTCAGATA**CAGATGGATACGCAGTGCAGA**TCCGATACAGATCA 3′

REVERSE PRIMER **3′ CGTCT 5′**

FORWARD PRIMER **5′ CAGAT 3′**

3′ TCCAGTCTAT**GTCTACCTATGCGTCACGTCT**AGGCTATGTCTAGT 5′

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

Forward (or Left) Primer: CAGAT

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

Q69. 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.

GCAAACTATT ATGTGTATTA GTATTCTATT GGTATTCATT ATCGACTGAC ACCCCCCACT ATTTCTACTT

TTAATCTACA G**ATATAGTTG GATCAAGAAC GGCAAGAAGT TCGATTGGCA GGCGTACGAT AACCGCATGC**

TGCGGCAGCC AGGACGTGGC ACCCTGGTGA TCACCATACC CAAGGACGAG GATCGCGGCC ACTATCAGTG

CTTTGCGTCC AATGAATTCG GAACGGCCAC CTCGAACTCA GTATATGTGC GTAAGGCCGA GCTGAATGCC

TTCAAGGATG AGGCGGCCAA GACACTGGAG GCCGTCGAGG GTGAGCCCTT TATGCTGAAA TGTGCCGCAC

CCGATGGTTT TCCCAGTCCG ACAGTCAACT GGATGATCCA GGAGTCCATC GATGGCAGCA TCAAGTCGAT

CAACAACTCT CGCATGACCC TCGATCCTGA GGGTAATCTC TGGTTCTCGA ATGTTACCCG TGAGGATGCC

AGCTCCGATT TCTACTATGC CTGCTCGGCC ACCTCGGTGT TTCGCAGTGA ATACAAGATT GGCAACAAGG

TGCTCCTCGA TGTCAAACAG ATGGGCGTTA GTGCCTCGCA GAACAAGCAT CCGCCCGTGC GTCAATATGT

TTCCCGTCGC CAGTCCTTGG CGTTGCGTGG CAAGCGAATG GAACTGTTTT GCATCTACGG TGGAACACCG

CTGCCGCAGA CCGTGTGGAG CAAGGATGGC CAGCGTATAC AGTGGAGCGA TCGAATAACG CAAGGACACT

ATGGCAAATC ACTGGTCATT CGGCAGACAA ATTTCGATGA TGCCGGCACA TACACCTGCG ACGTGTCCAA

CGGTGTGGGC AATGCCCAAT CCTTCTCCAT...

Figure 2 Part of the Neuroglian 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	• 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 that others and must be avoided.
Length	Optimal primer length is 18-30 nucleotides
GC content	40-60%
T_m	• Melting temperature (T_m) between 55°C and 65°C • T_m = 2°C x (A+T) + 4°C x (C+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.

Q70. 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).

Q71. 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: $T_m = 2(A+T) + 4(G+C)$. 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.

Q72. 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.

Q73. 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!

Q74. What would happen if, say, the forward primer had the sequence 5'GTTCGCATTCGAATGCGAAC3'? 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.

Q75. 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.

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.

Q76. Using the above information together with the selected forward primer sequence indicated on the sequence on Figure 2, write out a suitable primer (5' to3') which will incorporate a KpnI site into the PCR product (the KpnI recognition sequence is GGTACC) and allow cleavage of the end of the PCR product by KpnI.

Q77. Draw a diagram, with all 5' and 3' ends marked, showing how this left primer, with the additional sequence, will anneal to the *Nrg* template

Primer Design Helper

http://goo.gl/uroojK

The following part of this exercise is not done by all degree groups. Your lab leader will inform you if you are doing the next section.

Using RT-PCR to characterize products of alternative splicing

It has been estimated that approximately 60% of human genes exhibit alternative splicing of exons, with an average of about 3 alternatively spliced transcripts per gene, and that 70% of these alternative splicing events generate different versions of the protein product. Some genes generate large numbers of alternate transcripts, and thus obtaining the primary sequence of a gene is only the start of the task of characterizing how that gene functions. To see the alternate transcripts we must obviously analyse the mRNA, and one method is to employ reverse transcriptase PCR (RT-PCR).

Alternate splicing produces two versions of the neuroglian protein: there are two versions of the last exon, exon 7: 7a and 7b. mRNAs with exon 7a encode the shorter, non-neuronal neuroglian (167kDa) and mRNAs with exon 7b encode the longer, neuronal neuroglian (180kDa).

CGGATGGTTC ATTGGCATGA TGCTGGCCCT GGCCTTCATC ATCATCCTCT TCATCATCAT CTGCATTATC CGACGCAATC
GGGGCGGAAA GTACGATGTC CACGATCGGG AGCTGGCCAA CGGCCGGCGG GATTATCCCG AAGAGGGCGG ATTCCACGAG
TACTCGCAAC C**/5..66nt..6/**GTTGGATAA CAAGAGCGCT GGTCGCCAAT CCGTGAGTTC AGCGAACAAA CCGGGCGTGG
AAAGCGATAC TGATTCGATG GCCGAATACG GTGATGGCGA TACAG**/6..1556nt..7a/**GCATG AATGAAGATG
GATCCTTTA TTGGCCAATA TGGACGCAAA GGACTTTGAT TTAATTAGTA AGCAGCGCAC CGCAACAGCA ACTCAAAAAT
AATATCGAAA CCGAGCCCTT AACCCCAAAA ATCAAAAAAT CAACAAGACC AAACACCATC ACAGCAGAAA AATGAAAAAA
TTAATGAAAA TAATAGTAGC CTACATTTTA TTCGACTATA AGTGCAAACA CCACGACTAA TTTAAAGTAT ATATAAAAAT
AGAGGTTTTA TATATAACTA TTAAAATCTT AAAATGTGTA AAAAAAAAAA CAAACAA**/7a...1356nt...7b/**GACAA
TTTACCGAGG ATGGCTCCTT CATTGGCCAA TATGTTCCTG GAAAGCTCCA ACCGCCGGTT AGCCCACAGC CACTGAACAA
TTCCGCTGCG GCGCATCAGG CGGCGCCAAC TGCCGGAGGA TCGGGAGCAG CCGGATCGGC AGCAGCAGCC GGAGCATCGG
GTGGAGCATC GTCCGCCGGA GGAGCAGCTG CCAGCAATGG AGGAGCTGCA GCCGGAGCCG TGGCCACCTA CGTCTAAGAG
GCGTGGCTGG GATTCACTTG CCCCATTGTT CTCCTGATTT TCTACCAAAC GATTCAAACG CCTCTTAAAC AAAAAAGAAA
CTGTGTAATT CTATGTGTAA AACGAAAACT GCTTTAAGTG TTCTGCAAAA AA**/7b**

The coding sequences for *Drosophila* neuroglian exons 5 (3' end only), 6, 7a and 7b. Splice junctions are shown (/), and intron length is indicated. Stop codons are underlined.

Q78. Explain how RT-PCR might be utilized to determine which *Drosophila* tissues express the non-neuronal neuroglian transcript. Mark on the sequence regions which would generate suitable primers for this, bearing in mind the rules for primer design.

Q79. How could you design your primers so that they will be specific for cDNA and not anneal to any genomic DNA which has contaminated your mRNA preparation?

From protein to primers

The similarity between genes may be rather limited, particularly if the shared function has emerged independently through convergent evolution. Can we use similarity at the protein level as the basis for the design of PCR primer? The answer is "yes" but as we'll see, the design of primers in this way demands careful thought.

Alignment of the protein sequence of *Drosophila* Nrg with other entries on the protein database reveals that the amino acid sequence FNEDGSFIGQ is well conserved amongst neural cell adhesion molecules from a range of species. Imagine we wanted to isolate part of the cDNA for an *Nrg* homologue from an animal about which little was known at the genetic level. Could we exploit the conservation of the amino acid sequence in designing a PCR primer? The deduction of the DNA consensus is not as straightforward as you might think. Although any given codon always codes for a given amino acid, a given amino acid may have more than one possible codon. So although any DNA sequence codes for a specific amino acid sequence, a given amino acid sequence can be coded by many different DNA sequences. This is known as **degeneracy**. You will see from the genetic code at the end of the manual that different amino acids can be coded by 1, 2, 3, 4 or even 6 codons. So choosing a **well-conserved** amino-acid stretch is not enough - one with **low degeneracy** (ie as few different ways of coding for the amino acids as possible) is necessary for successful primer design.

Q80. Why do we want to minimise degeneracy in designing a PCR primer?

Using the table below, evaluate if the nominated sequence would be useful for primer design. In Row 2 of the table, below each amino-acid in your consensus, write its degeneracy (ie the number of different codons for that amino acid), **then select the run of 8 amino-acids** with the lowest degeneracy. You will need to use the Table on codon usage (below) to help with this.

Q81. What would be the total degeneracy of your primer needed to code for these 8 aa (i.e. the number of different DNA sequences that could encode this same amino-acid sequence)?

Eg: H	1. Consensus (single letter aa code)	F	N	E	D	G	S	F	I	G	Q
2	2. Degeneracy										
CAC	3. Likeliest codon										
CAT	4. Second likeliest codon										
0.68	5. Prob (likeliest codon)										
0.32	6. Prob (second likeliest codon)										
1.0	7. Combined probability										
$CA\frac{C}{T}$	8. Degenerate DNA sequence										

Table of codon usage in 44 nuclear genes from our system of interest:

AA	Codon		Fraction
A	Ala	GCA	0.13
	Ala	GCC	0.57
	Ala	GCG	0.11
	Ala	GCT	0.19
R	Arg	AGA	0.06
	Arg	AGG	0.07
	Arg	CGA	0.06
	Arg	CGC	0.61
	Arg	CGG	0.07
	Arg	CGT	0.13
N	Asn	AAC	0.78
	Asn	AAT	0.22
D	Asp	GAC	0.68
	Asp	GAT	0.32
C	Cys	TGC	0.89
	Cys	TGT	0.11
*	STOP	TAA	0.40
	STOP	TAG	0.20
	STOP	TGA	0.40
Q	Gln	CAA	0.14
	Gln	CAG	0.86

AA	Codon		Fraction
E	Glu	GAA	0.13
	Glu	GAG	0.87
G	Gly	GGA	0.26
	Gly	GGC	0.49
	Gly	GGG	0.03
	Gly	GGT	0.22
H	His	CAC	0.68
	His	CAT	0.32
I	Ile	ATA	0.04
	Ile	ATC	0.71
	Ile	ATT	0.25
L	Leu	CTA	0.04
	Leu	CTC	0.16
	Leu	CTG	0.62
	Leu	CTT	0.03
	Leu	TTA	0.01
	Leu	TTG	0.14
K	Lys	AAA	0.14
	Lys	AAG	0.86
M	Met	ATG	1.00
F	Phe	TTC	0.77
	Phe	TTT	0.23

AA	Codon		Fraction
P	Pro	CCA	0.20
	Pro	CCC	0.49
	Pro	CCG	0.21
	Pro	CCT	0.10
S	Ser	AGC	0.21
	Ser	AGT	0.05
	Ser	TCA	0.05
	Ser	TCC	0.38
	Ser	TCG	0.24
	Ser	TCT	0.07
T	Thr	ACA	0.09
	Thr	ACC	0.65
	Thr	ACG	0.16
	Thr	ACT	0.10
W	Trp	TGG	1.00
Y	Tyr	TAC	0.77
	Tyr	TAT	0.23
V	Val	GTA	0.07
	Val	GTC	0.33
	Val	GTG	0.51
	Val	GTT	0.09

The next trick is to allow for ***codon preference*.** Not all the possible codons are equally likely. Some codons are used preferentially for a given amino-acid, even though others are theoretically possible. These codon preferences are species-specific. The codon preference table for our target species (above) will allow us to deduce the most likely codons to use. For each amino-acid it shows the frequency of usage for each of the possible codons.

For your chosen stretch of 8 amino acids, in Row 3 of the table, for each amino acid, write the preferred codon, and below that, in Row 5, the probability of its use.

Q82. Calculate the probability that this sequence (ie, using only the likeliest codon for each amino-acid) corresponds to the sequence that would be used in our target system to code for your chosen 8 amino-acid sequence, by multiplying the probabilities together (eg. 0.72 x 1.00 x 1.00 x 0.90 = 0.66).

You will find that the probability of even the most likely primer being a perfect match is extremely low. You can make this probability higher by making your primer **degenerate**. In other words, you could make primers with alternate possibilities at some positions. Fortunately, alternate codons for a given amino acid usually differ only at the third base, so you need only offer a choice of bases at one position in each triplet; we are going to limit the choices to two alternates.

For each amino-acid, write the second preferred codon in Row 4, together with its probability in Row 6, and the total probability that one or other of them will be used (by *adding* the individual probabilities) in Row 7.

Q83. Calculate the probability that a degenerate primer using both the likeliest and second most likely codons will contain the exact sequence used in our target system to encode this 8 amino-acid sequence. How big is the improvement compared to use of likeliest codon only?

Write your degenerate primer sequence in Row 8, placing degenerate bases one above the other, eg. ...AC$\begin{matrix}T\\C\end{matrix}$ ACG...

Q84. What is the degeneracy of your primer now? (i.e. how many different DNA sequences will be present in your primer mix?)

You now have a primer but would it conform to the basic guidelines that we considered earlier? This is one reason why designing primers from protein sequences requires more skill than using nucleic acid sequences!

Q85. Assuming that the primer had to be designed using the amino acid sequence that we have considered, suggest a simple improvement that could be made to its 3' terminus.

EXERCISE 2: RESTRICTION MAPPING

Introduction

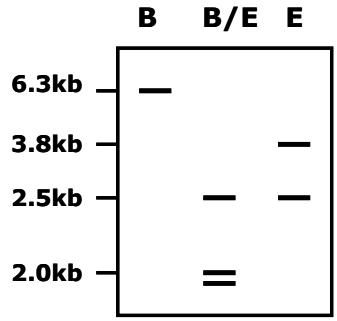
In the early days of molecular biology, scientists commonly used restriction mapping to understand the genes they were studying. In modern research labs, restriction mapping is not as widely used because sequencing has become more generally available and more economical than in the past. However, there are cases in which a restriction map may be adequate for a researcher to perform their investigations. The overall approach to restriction mapping involves the use of digestive enzymes to break down physically a sample of DNA. Once you measure the products of this digestion, you can "reassemble" the pieces and deduce the original sequence of DNA. Restriction mapping, therefore, allows you to determine where restriction sites are in your DNA sequence. If you have a look at appendix 4, you will see the pBluescript vector has many restriction sites annotated. Here, we will show you the basic principles of restriction analysis, then give a series of confidence-building exercises to give you practice. Please check Moodle for many more examples.

Basic principles

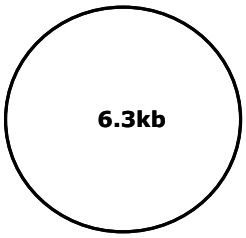
To build a restriction map, you need to estimate the sizes of DNA fragments produced on an agarose gel from a plasmid cut singly or doubly, using a range of restriction enzymes.

There are a few simple rules for mapping and the best way to explain is with the aid of an example.

Imagine in the lab you have taken a plasmid vector and digested this with single restriction enzymes, either BamHI (B) or EcoRI (E) alone and then performed a double digest with the two enzymes together. If you ran the resulting DNA fragments on an agarose gel, it might look like this:



This information can then be 'mapped' onto a circular plasmid (such as seen here) using some rules that will be described below:



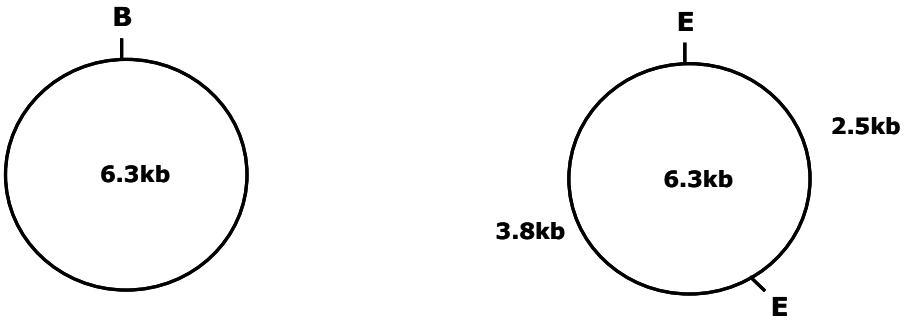
(a) **The sum total of the fragment sizes in every lane should always add up to the same value – the size of the uncut DNA.** In the example above, we know the size of the plasmid is 6.3kb, since the total of DNA fragments in each lane of the gel comes to this amount ie:

BamHI – 6.3

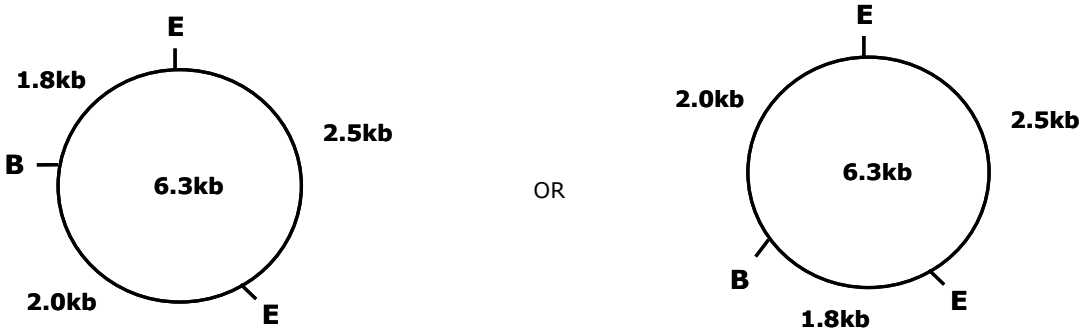
EcoRI - 3.8, 2.5

BamHI and EcoRI – 2.5, 2.0, 1.8

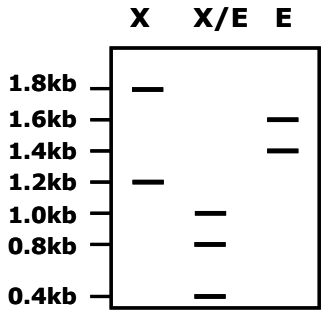
(b) **Digests using a single enzyme shows how many times a particular enzyme cuts the plasmid**. In the example above, when the plasmid is cut with BamHI this yields a single fragment, therefore, there must only be one BamHI site in this plasmid. On the other hand, when the plasmid is cut with EcoRI, this results in two DNA fragments being produced. These two pieces of information cannot be drawn on a single plasmid map until the result from the double digest is considered. Therefore, at the moment we can only place the single cuts on two separate maps as follows:



(c) **Cutting with two enzymes shows where the sites are relative to each other. The key to mapping is to identify the bands that change between the single digests and the double digests.** The double digest produced three fragments - 2.5, 2.0, and 1.8 kb long. A little bit of logic helps to order the sites relative to each other. Notice in the B/E double digest that the 2.5kb band seen in E alone is still present but the 3.8kb EcoRI band is gone and two new bands appear of 2kb and 1.8kb (=3.8kb). Therefore, there must be a BamHI site within the 3.8kb EcoRI fragment. The BamHI site is 1.8kb from one EcoR1 site and 2 kb from the other EcoR1 site. In this particular example, two possible maps could be drawn and both would be correct, since there is not yet enough information to place the enzymes exactly. The final maps could be one of the following:



(d) **Watch out for DNA fragments that appear as the same size on an agarose gel!** Consider the following result from a plasmid digest using two restriction enzymes, XbaI (X) and EcoRI (E). The resulting gel might look like the following:



Using the rules set out previously, **can you first of all work out the size of the DNA fragments** in each lane of the gel and, hence, the total size of the plasmid:

XbaI –

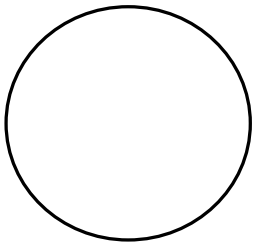
EcoRI -

XbaI and EcoRI –

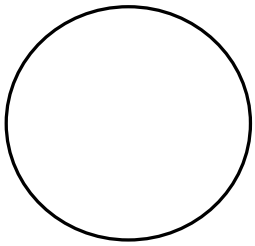
Does the size of the plasmid appear the same in each lane, and if not, why not?

What would the plasmid maps look like from a digest with each single enzyme?

XbaI:

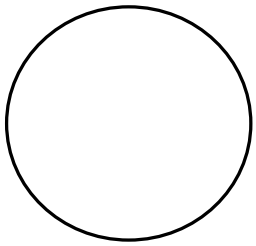


EcoRI:



What would the final restriction map look like when all the information is brought together (Hint: there may be more than one orientation so your answer may not appear the same as your neighbour!)?

XbaI and EcoRI:



Restriction Mapping Helpers

A. Principles of Restriction Mapping

http://goo.gl/MN6Cxi

B. Restriction Mapping Examples

http://goo.gl/shVtW2

Restriction mapping confidence builders

Below are a few examples of restriction mapping problems. Draw restriction maps alongside, for each of the following examples (of increasing complexity). We have not given you the gel images for these digests, but provided you with the length of fragments once digested.

Question 1

Uncut plasmid (circular)	7.0kb
EcoRI	7.0 kb
XhoI	5.5, 1.5 kb
EcoRI+XhoI	3.0, 2.5, 1.5 kb

Question 2

Uncut plasmid (circular)	6kb
EcoRI	1.0, 2.0, 3.0 kb
XhoI	6.0 kb
EcoRI + XhoI	1.0, 1.2, 1.8, 2 kb

Question 3

Uncut plasmid (circular)	7.0 kb
EcoRI	7.0 kb
XhoI	6.0, 1.0 kb
EcoRI + XhoI	3.0, 1.0 kb

Question 4

Uncut plasmid (circular)	7.0 kb
EcoRI	5.0, 2.0 kb
XhoI	6.0, 1.0 kb
EcoRI + XhoI	3.0, 2.0, 1.0 kb

Question 5

Uncut plasmid (circular)	6.0 kb
EcoRI	2.8, 3.2 kb
XhoI	1.3, 4.7 kb
EcoRI + XhoI	0.6, 0.7, 2.1, 2.6 kb

Question 6

Uncut plasmid (circular)	7.0 kb
EcoRI	5.0, 1.2, 0.8 kb
XhoI	4.0, 3.0 kb
EcoRI+XhoI	3.0, 2.0, 1.0, 0.2, 0.8 kb

Question 7

A PCR product was prepared by digestion with PstI and XhoI and was cloned into pBluescript, the vector shown in Appendix 4 of this manual. The following bands were obtained after a single and double digestion of the recombinant plasmid. Draw the restriction map and indicate where the pBluescript vector sequence is (the vector is 3kb in size).

PstI	3.5, 1.8, 1.7
XhoI	3.6, 3.4
PstI + XhoI	3.0, 1.8, 1.1, 0.6, 0.5

Post-Lab Online restriction mapping resources

If you want further examples of restriction mapping then have a look at the Molecular Methods Moodle site – go to 'Other Supporting Resources' and then 'Restriction Mapping Examples'.

There are now many online tools freely available for you to identify restriction sites you are interested in within a particular sequence. Say for example, you wanted to find out if the BamHI and KpnI restriction enzymes were going to cut your cloned *Nrg* gene in half, making it impossible to clone using these enzymes. How would you find out?

See the following exercise on our Molecular Methods Moodle site – Restriction Mapping Examples - Restriction Mapping Online Resource Exercise.

EXERCISE 3: LABORATORY CALCULATIONS

Introduction

Laboratory calculations and making solutions are some of the most important procedures in a laboratory. If solutions are not prepared accurately, experiments may fail. Biological experiments are particularly sensitive to alterations in chemical composition of solutions. For example, enzymatic reactions are extremely sensitive to alterations in pH, and salt concentrations; incorrectly prepared media may inhibit bacterial growth; DNA migration in agarose gels can be altered by inexact calculation.

Therefore, knowledge of how to solve "dilution problems" is immensely important for students in all biological fields. Often you will encounter these kinds of problems in future science courses as well as in future place of employment. For example, you may be asked to prepare a solution of a particular concentration or you may be asked what the concentration of a solution is if you added a known amount of water to a given solution. There are various solutions used throughout this practical and these have been prepared for you but it is important that you would know how you would go about doing this yourself.

If you are confident about calculations that allow laboratory solutions to be prepared then go straight to the questions throughout this exercise. If you are unsure, or need a reminder then read on for some hints and tips. Below you will find worked examples to show you how to do various calculations. However, if you are still unsure, please ask a demonstrator to help, as different people think about these things differently!

Quantitative Expressions of Concentration

Units of mass and volume:

Mass units:

1 gram (g)	= 10^{-3} kg
1 milligram (mg)	= 10^{-3} g
1 microgram (μg)	= 10^{-6} g
1 nanogram (ng)	= 10^{-9} g
1 picogram (pg)	= 10^{-12} g

Volume units:

1 millilitre (ml)	= 10^{-3} L
1 microliter (μL)	= 10^{-6} L

1. Concentrations based on molarity (mol/L)

Molarity is the number of moles of solute in a litre of solution and this is the most common way of expressing the concentration of solution. To calculate the molarity of a solute in a solution we need to know:

- The moles of solute present in the solution.
- The volume of solution (in litres) containing the solute.

One mole of any substance equals its molecular weight (MW) in grams.

A molar solution = A solution containing one mole of solute / litre (1M/L, or 1 mol/L).

Example: How can you prepare exactly 0.5L (500ml) of a 0.25M NaCl solution in water? The MW of NaCl is 58.5g.

Method - you could remember a general formula for this approach:

g = volume wanted (L) x M (moles) x MW, remembering that you always have to put the volume wanted in litres

= 0.5 x 0.25 x 58.5

= 7.31g

Remember that if you are making 500mls of this solution, you are not adding 7.31g to 500mls of water, but you putting 7.31g **into a final volume of 500mls**

Now try the following questions:

Q86. Solutions of Tris and EDTA are commonly used in buffers in the laboratory. How would you prepare 500mls of each of these stock solutions at 0.5M Tris and 100mM EDTA concentrations? molecular weight of Tris - 121.14g/mol, molecular weight of EDTA – 372.2g/mol.

Q87. You use competent bacteria in the lab for the process of transformation. These cells are typically prepared using a $CaCl_2$ method, which creates pores in bacterial cells by suspending them in a solution containing high concentration of calcium. The concentration of $CaCl_2$ used is 100mM. Given the molecular weight of $CaCl_2$ is 110.98 g/mol, how would you prepare one litre of a 100mM solution?

Q88. Bacterial cells that have been transformed with DNA are plated onto L-agar which contains 50 µg/ml ampicillin, 50 µg/ml IPTG and 50 µg/ml X-gal. The following is the basic recipe to prepare 500ml of this L-agar:

5g Tryptone

2.5 g Yeast extract

2.5 g NaCl

1.5 % Agar (7.5g in 500 ml)

Distilled water to 500 ml.

Given the molecular weight of NaCl is 58.44 g/mol, what is the molar concentration of this in the above recipe?

Q89. In the PCR clean-up protocol, the PCR product is prepared for the clean-up procedure by mixing it with 'Membrane Binding Solution', which consists of 4.5M guanidine isothiocyanate and 0.5M potassium acetate (pH 5.0). Given that the molecular weight of these is 118.16g/mol and 95.15 g/mol, respectively, how would you prepare 100mls of membrane binding solution and ensure that the final pH was correct?

2. Concentrations based on %

Calculation of concentration based on % is more straightforward that molarity. Weight/volume (w/v) or volume/volume (v/v) can determine these percentages.

Example: How would you prepare 100mls of a solution containing 10% NaCl?

Method - 10% (w/v) solution = 10g NaCl/100mls

10g NaCl would be dissolved in water and the volume adjusted in a measuring cylinder to 100mls

Example: How would you prepare 150mls of a solution containing 1% HCl (this is a liquid)?

Method - 1% (v/v) solution = 1.5ml/150mls

1.5mls would be made up to the 150mls with water using a measuring cylinder.

Now try the following questions:

Q90. How can the following solutions be prepared?:

(a) 200mls of a 40% (w/v) solution of polyethylene glycol (PEG)

(b) 80mls of a 7% (w/v) solution of NaCl

(c) 200mls of a 95% (v/v) solution of ethanol.

Q91. If you have a stock solution of 10% (w/v) SDS and you wish to prepare 50mls of a solution containing 0.02% SDS, how would you use your stock solution to prepare this?

3. Dilutions from Stocks

In a laboratory, you often do not weigh out chemicals every time you want to prepare a solution, for two reasons; (i) inconvenience, and (ii) sometimes it is impossible to weigh out small enough amounts to make a particular solution. Certain commonly used reagents are prepared as a stock solution at a high concentration, and then this would be diluted to the desired concentration. The Master equation for doing this type of calculation would be as follows:

$$\frac{\text{What you want}}{\text{What you have}} \text{ X final volume} = \text{volume to add to solution}$$

Note: **what you want** = the final concentration you want – this can be expressed as %, M, or mg/ml

what you have = the concentration of the stock solution you have – this can be expressed as %, M, or mg/ml

'what you have' and 'what you want' must be expressed in the same units!!

Final volume = means the amount you want to prepare: if you use litres in your answer will be in litres, if you use mls in your units then your final answer will be in mls.

Example: If you want to make 100mls of 1M salt from a 5M stock, you can solve this as follows:

$$\frac{\text{1M (What you want)}}{\text{5M (What you have)}} \text{ X 100mls} = \text{20mls of 5M stock and add 80mls } H_20 \text{ to give a final volume of 100mls}$$

Example: If you want to make 1L of 100mM Tris from a 1M stock, you can solve this as follows:

$$\frac{\text{100mM (What you want)}}{\text{1000mM (What you have)}} \text{ X 1L} = \text{0.1L of 1M Tris and add 0.9L } H_20 \text{ to give a final volume of 1L}$$

Example: If you want to make 40mls of a 0.5% glucose solution from a 20% stock:

$$\frac{0.5\%\ \text{(What you want)}}{20\%\ \text{(What you have)}} \times 40\text{mls} = 1\text{ml of 20\% glucose and add 39mls } H_20 \text{ to give a final volume of 40mls}$$

Now try the following questions:

Q92. If the stock concentration of the forward and reverse primer in a PCR is 10nM and the final concentration you want of these is 0.4nM. What volume of each primer would be added to the PCR tubes (final volume 50μl)?

Q93. TAE buffer is a buffer solution containing a mixture of Tris base, acetic acid and EDTA. This was used in the preparation of your agarose gel mix (see Appendix 3). In molecular biology it is used in agarose electrophoresis typically for the separation of nucleic acids such as DNA. Assume in the lab you are provided with a 50X stock solution of TAE. How would you prepare 150ml of a 1X TAE gel mix containing 1% agarose?

Q94. To prepare the ligation reaction mix (DAY 2), you are provided with pBluescript plasmid precut with the restriction enzymes BamHI and KpnI. The final concentration of the pBluescript you require for this is 200ng/μl. A stock of plasmid has already been made for you at a concentration of 1mg/ml. How much of this stock would be required to prepare 2mls of plasmid vector ready for use in the reaction?

Q95. Assume that you have stock concentrations of ampicillin prepared at 60mg/ml. How much would you need to add to 500ml of L-agar to have the correct final concentration of 50μg/ml ?

4. Mixed Solutions

In a laboratory, you often have to prepare a solution that is a mixture of many components. Some of these components may have stock solutions, others may have to be weighed out, etc. You may therefore need to use a mix of the above calculations to work out how much of each component is required to prepare the final solution.

Q96. In this lab, you have to carry out a restriction digest a couple of times and this is a common procedure for further DNA analysis. Imagine that you want to digest 2μg of DNA with BamHI; your DNA stock is 1.2mg/ml. In this digest, use 1μl of BamHI and get your final volume to be 30μl. You need to have a final concentration of buffer to be 1X but its' stock is at 10X. How would you prepare this restriction digest mix?

Q97. During the plasmid miniprep two of the solutions you use are called the 'Cell Lysis Solution' (consists of 0.2M NaOH and 1% SDS) and 'Cell Resuspension Solution' (consists of 50mM Tris-HCl pH7.5, 10mM EDTA and 100μg/ml RNase A). Given the following information, how would you prepare 50ml of these two solutions for use in the miniprep?

molecular weight of NaOH is 40 g/mol; Stock solutions of 0.5M Tris-HCl, 10% SDS, 0.1M EDTA, 100mg/ml RNase A

COMPUTER EXERCISE 1: HOW TO INVESTIGATE THE FUNCTION OF A GENE

Aims

As part of your degree at the University of Glasgow, or in your future career, you may be given a gene name and asked to find out more about the DNA sequence or the protein it encodes. You can do this through experimentation but you would be surprised at the huge amount of time and money that this actually takes. It is much easier to look at the web to see if someone else has done the work for you! The aim of this exercise is to show you how you can discover more about a gene. Most of the questions we usually ask ourselves can be solved using the National Centre for Biotechnology Information website (NCBI), although the European Bioinformatics Institute (EBI) can also help (Appendix 5). For some questions, Wiki or Google offer solutions, though we do want accuracy!

An ever–increasing amount of DNA sequence information is available for an ever–increasing number of organisms. The complete genome sequence is already available for many organisms, including E. coli, Saccharomyces cerevisiae (yeast), and Caenorhabditis elegans (a nematode worm). A near–complete version of the Drosophila genome was published in March 2000, and in February 2001 the completion of the first draft of the human genome was announced. Of course getting a complete genome sequence is only the first step. The raw sequence information needs to be analysed to delineate both genes already known from pre–existing sequence information and, using gene–finding computer programmes, new genes by predicting potential protein coding exons and any intervening introns. We need to know how all the genes function and how they are regulated. Even for the bacterium E. coli, roughly 40% of its 4288 genes are of unknown function. For higher eukaryotes the situation is far more complex, with current estimates of about 30,000 to 40,000 genes in the human genome. There is a vast amount of information out there and this exercise will help you understand and access relevant information. You will only scrape the surface of the available information – you could probably spend days investigating the available resources more fully!

Although you can use this exercise to find out more about the *Nrg* gene, you can also use it as an opportunity to choose a different gene that may be of particular interest to yourself.

Choose a Gene

You may have read of a gene in the general press, in a research paper or perhaps you were inspired by one mentioned in a lecture! You may wish to concentrate on that gene OR, alternatively, you can pick one of these –

1) - Neuroglian gene (*Nrg*) gene - *Drosophila melanogaster*.
2) - Serine proteinase inhibitor (SERP1) gene - *Myxoma* virus that causes myxamatosis.
3) - Homeobox1 (AmphiHox1) gene - amphioxus (*Branchiostoma floridae*).
4) – *Plasmodium falciparum* erythrocyte membrane protein 1(PfEMP1) – *P. falciparum* causes malaria.
5) - Breast cancer type 1 susceptibility protein (BRCA1) - Humans.

What does your gene do?

To find a brief overview of the gene, Wikipedia or Google may be the first step. These would hopefully give a brief, understandable summary of the gene.

Q98. What does the gene you have selected do and why is it important to the organism?

Finding a gene on NCBI

To explore a gene further we will concentrate on what NCBI has to offer. The GenBank sequence database is an open access, annotated collection of all publicly available nucleotide sequences and their protein translations and this database is maintained by NCBI. GenBank contains mRNA sequences with coding regions, fragments of genomic DNA with a single gene or multiple genes, and a variety of other sequences such as regulatory sequences and repeats such as SINE, satellite DNA. It also contains whole genomes and DNA from an enormous range of species. NCBI provides search engines allowing you to locate a sequence and with that, a range of data that relates to it. GenBank has both nucleotide and protein sequences but we will generally concentrate on the protein that the gene is producing.

- To obtain the protein sequence of your gene from NCBI open their homepage: www.ncbi.nlm.nih.gov.
- Type the name of the gene and the organism into "All databases" search box.
- Click "Search" and a results page comes up showing a selection of options (see Fig 1 for an example).
- Click on the protein option (shown in Fig 1) and when the results appear, click on your gene in the appropriate organism. The information about the gene will look similar to Fig 2.
- This will be referred to as the '**PROTEIN PAGE'** and we shall come back to it often so keep a tab of this page available.

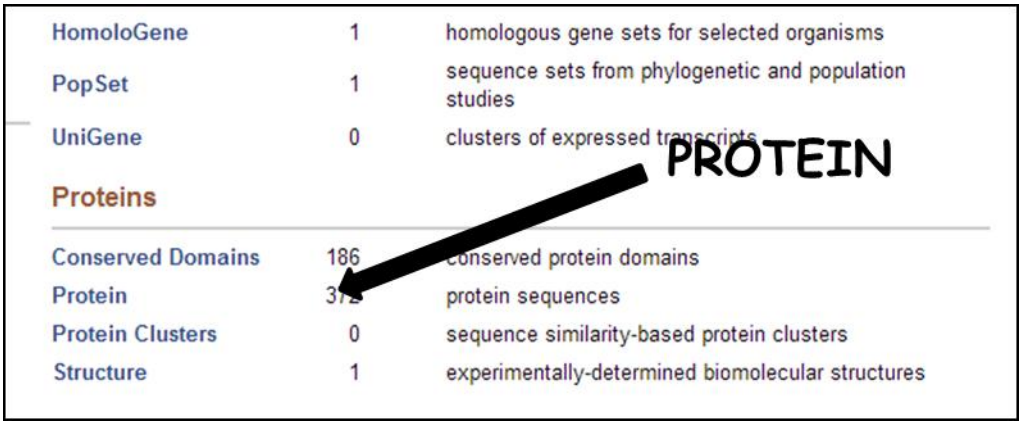


Fig 1. NCBI Gene Search. This screen shot shows part of a results page from a gene search featuring access to information about protein sequences, conserved domains, structure, etc. and they are presented in different sections which can be accessed individually.

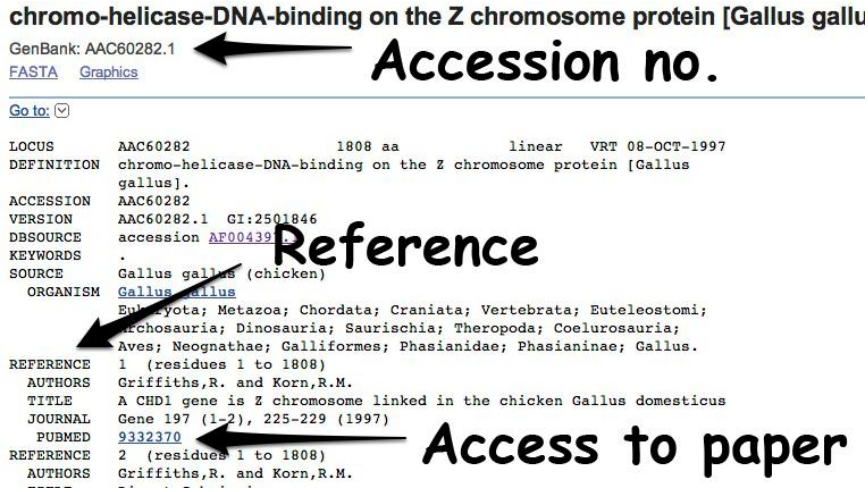


Fig 2. NCBI 'Protein' PAGE. Information about the protein you have selected

Interpreting the information about the protein

- Although the bottom of this page provides the protein sequence, a summary of information about your protein is shown at the top.

Q99. How many amino acids are there in the protein? If you find you are looking a 'partial sequence' (see Fig. 3) this is a fragment of the gene and you have picked the wrong one! Go back a page and pick a sequence featuring the whole gene.

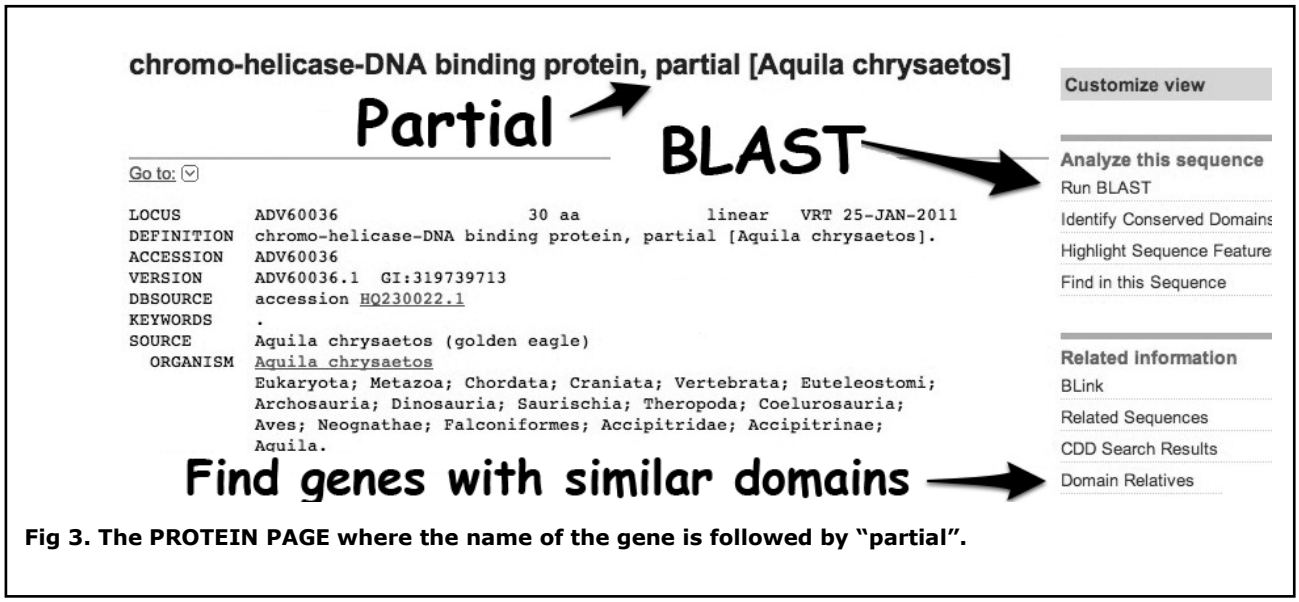


Fig 3. The PROTEIN PAGE where the name of the gene is followed by "partial".

Your initial search has almost inevitably found several sequences that all feature the same gene. To identify each sequence individually they are given an accession number (Fig 2). This is the number that is provided in research papers and allows you to find the right gene quickly on-line.

Q100. What is the accession number for your gene?

Q101. Using the 'All databases' search on the NCBI homepage that you first accessed, for which genes are these accession numbers: (a) M10090 and (b) AF022648

- Go back to the 'protein page' of your gene and down the left of the page there is nearly always a subtitle "REFERENCE" (see Fig 2), for which there may be several research paper citations. The last reference is usually from the group that submitted the original sequence. The references above that usually refer to other papers that have researched the gene. In some cases, where it is a partial sequence or an entire genome, it is unlikely that the reference will be specifically about your gene but may still be of interest. If the reference features an interesting paper you can click on the number in "PUBMED" which takes you to the abstract of this paper (Fig 2).

Q102. Note down the full details of one of these references. There may be more than one reference, in which case just write the top one down.

How to find similar sequences

Gene duplication is a key mechanism in evolution. After a gene is duplicated, the once-identical genes can undergo changes and diverge to create two different genes. This can happen in many ways but it means that instead of having, for example, a single Hox2 gene you actually have two of them Hox2a and Hox2b. What can then happen is that Hox2a diverges in function from Hox2b, whilst retaining a similar sequence. This is the start of a **gene family**. The same genes also turn up in different organisms so both mice and humans have a Hox2. Consequently, information about similar sequences may give you information about both what your gene does and its evolution.

- In the column to the right of the 'protein page' you will be able to carry out further analysis. We shall use the BLAST search which uses the protein sequence to scan GenBank to look for similar sequences.
- Click on "Run BLAST" (Fig 3) *in a new tab*.
- On a new page it loads the accession number of your protein sequence. Just click the blue BLAST button at the bottom of the page. The search often takes a couple of minutes to run.

The results of a BLAST search

BLAST searches GenBank and looks for sequences that are similar to your gene. In the BLAST results it presents the same data in three formats. In each it describes the same group of DNA sequences but in a different way:

1) A graph: where the red line at the top, marked "query", is your gene. The lines below are different sequences that are aligned to your gene. The lines are coloured red, pink, green, blue and black in descending order of homology. So red has close similarity and black more distant.

2) A list of each of the aligned sequences, starting with the most similar one first (Fig 4). It tells you their names and how closely they align. A measure of their similarity is the E-value or expected-value. You have to be more clever than Einstein to understand the E-value! However, if the E-value is greater than 1, then the aligned sequence not that similar to your gene. If the E-value is between 1 and 10e-6 it is uncertain if they are related. If E-value is less than 10e-6 it suggests the sequences are related and the smaller the E-value the closer they align.

3) It then goes through each sequence and shows exactly how it aligns with you your sequence. This shows you the fine details.

Looking through the results that BLAST presents may provide useful information. If the graph shows only pink lines there are a lot of similar sequences in GenBank. In the list of sequences it may refer to lots of copies of the same gene showing it has been heavily researched. There could be sequences such as Sox1, Sox 2, Sox 3 or Sox-like suggesting your gene is part of a gene family (see Fig 4).

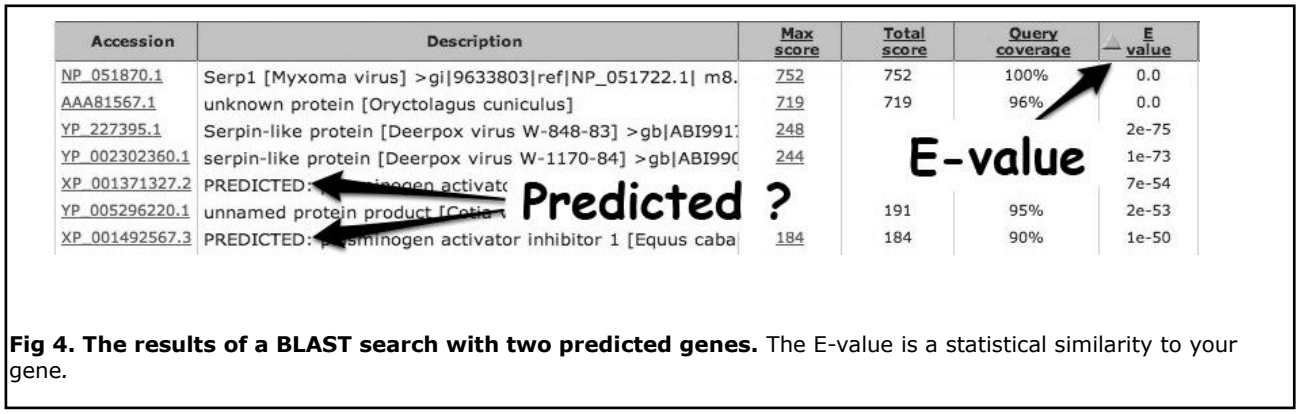


Fig 4. The results of a BLAST search with two predicted genes. The E-value is a statistical similarity to your gene.

Q103. If you look at the graph from your BLAST search, do all the aligned sequences have close similarity to your gene and how do you now this?

Q104. A homologous gene has a similar sequence to your gene so it appears to share common ancestry. Look carefully down the names of the aligned genes in the Blast search, if you find a homologue from a different species, write down the name of the gene and organism.

Q105. From your BLAST search, are there any genes that are 'predicted' (as in Fig 4)? What does this mean?

Q106. What does an E-value of 1e-99 mean (see Fig 4; have a look on You Tube http://goo.gl/xMqk7)?

Q107. At the bottom of the list is the protein with lowest similarity to your gene. What is the E-value? Does that mean it has a significant alignment with your protein (look at You Tube again)?

Domains: functional segments of the protein

Some parts of the gene are more essential than others. When translated into amino acids the important sections fold into a 3D structure that carries out the protein's function. These regions are termed functional domains. A protein may have a one or several domains all contributing to the operation of the proteins. What is more, the same domain can be found in several different genes. For example, the EF-hand domain that binds calcium, is found in at least 66 subfamilies of genes.

Outside the domains, the rest of the gene is usually less important and generally evolves more quickly.

To find if your gene contains a domain and, if it does, find how common your domain is in other genes we shall carry out a cDART search.

- Return to the PROTEIN PAGE (Fig 2).
- Click open "Domain relatives" *in a new tab* (shown in Fig 3; this maybe further down the page on the right). This searches GenBank for genes with similar domains to your protein. In may take a couple of minutes.

At the top of the picture is your gene illustrating the domain(s) it contains. Below this are different genes that also contain this same domain (Fig 5). They may also have the domain from your gene and add some completely different domains. The arrangement of domains is called architecture

- When you hover your arrow over the domains (*^~ in Fig 5) their name and summary of their function appears.

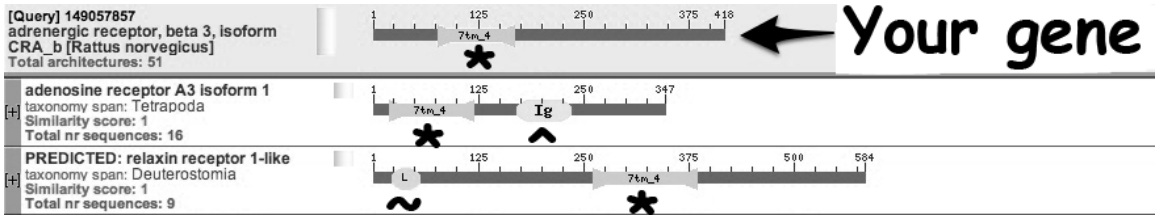


Fig 5. A "Domains relatives" or cDART search. This has found one olfactory receptor domain in this gene (*). This olfactory domain occurs in two other groups of genes. In one it is alongside the Ig (^) the other has an LDLa (~) domain.

Q108. Does your protein have any domains and what do they do? Does this agree with the function of the protein that you gave in the first question of this exercise?

Q109. Name two other proteins with that domain.

Q110. What does one of these two genes do and does it differ in function from your gene?

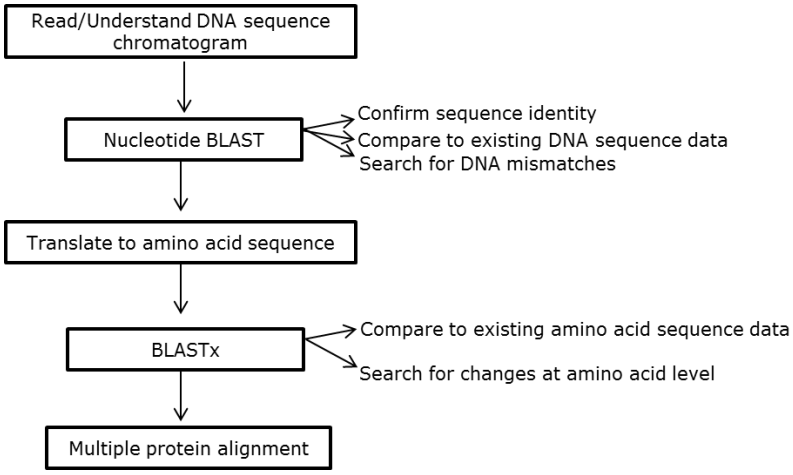
Q111. Have a think and describe a possible reason why you think cDART may not identify a domain in your gene?

Q112. Fill in the following summary table about your gene:

Q	A
1) Do you know what your gene does?	
2) How big is the your protein?	
3) Its accession number is:	
4) Are there closely related sequences on GenBank?	
5) Have you found any homologues?	
6) How many domains does your protein have?	
7) Do other proteins share this/these domains?	
8) Do you know a lot more about this gene in only 1 hour?	

COMPUTER EXERCISE 2: SEQUENCE ANALYSIS

The overall aim of this exercise is to analyse sequence data obtained from sequencing of your cloned PCR product. Briefly, you will:



To begin this exercise, you will use Chromas Lite, which allows you to view chromatograms from automated sequencing runs. First of all, click on 'START' (bottom left of your computer), locate the P-drive and then click on the 'Chromas' file. Open the Chromas application and finally click 'RUN' to launch the program.

Now, navigate to the Molecular Methods Moodle page and under 'Computer Exercises', click on 'Computer Exercise 2'. Right click on the file called '**Cloned nrg product from icebox flies'** and 'save link as' in a folder of your choice.

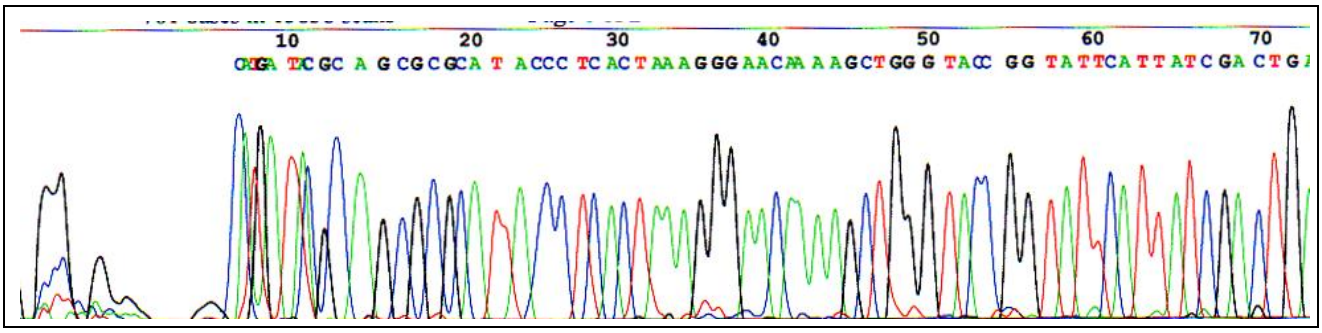
Reading chromatograms from automated sequencing

Return to Chromas Lite and then use **Open** to locate and open the chromatogram file you downloaded from our Moodle site. The chromatogram was obtained from a sequencing reaction run with DNA prepared in exactly the same way as you have done on the Molecular Methods course - a PCR product amplified with primers shown on the previous page was cloned into pBluescript using KpnI and BamHI. The sequencing reaction was run with the pBluescript **reverse sequencing primer** (see multiple cloning site DNA sequence for detail).

Q113. Which restriction site will mark the boundary between vector sequence and the start of the insert when sequencing from the reverse sequencing primer? Which of your cloning primers (used in your PCR) contain this restriction site?

Take a closer look at the chromatogram. Initially, the data has some "noise" - peaks that are broad and look unconvincing - but if you move downstream the signal should settle down to produce a series of sharp, evenly spaced peaks with low background. Locate the restriction site that marks the boundary between vector and insert. You can use the **FIND** tool to locate the sequence of the expected restriction site.

Q114. Note down the restriction site marking the vector/insert boundary in the diagram below:

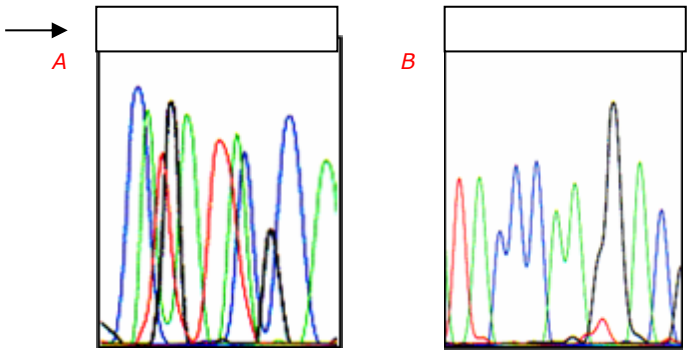


The first 20 bases or so of the insert should match the sequence of the relevant PCR primer used (previous section). The peaks should be clear and evenly spaced, matching the computer-generated sequence that lies above the trace.

Automated sequencing is powerful, but don't accept it uncritically. As you move through the sequence, does the sequence "called" by the computer tally with your interpretation of the data? Does it match what you would expect from the sequence of *Nrg* (you'll find this in Figure 2 of the PCR/Primer Design exercise)? You may find the occasional position where the spacing of peaks is too close for the computer to resolve satisfactorily as the example below. Equally, in regions that are GC rich, the peaks may merge together and the number of nucleotides may have to be inferred from the spacing, rather than the appearance of discrete peaks (see below). If you come across areas like this, your judgement might be better than the computer-called sequence!

For example:

Q115. Write down what you think should be the sequences below (green, A; red, T; blue, C; black, G)



Check your answers by looking at the region on the chromatogram around bases 1-15 for A and 200-210 for B.

Q116. Look carefully at the peaks from around 630 nucleotides. You will notice that the peaks on the chromatogram start to overlap and your faith in the accuracy of the sequence called by the computer may start to wane. From this initial inspection, you should be able to assess how much useful sequence data you have to work with. Subtract the position of the boundary between vector and insert. What are you left with?

Q117. We know that the insert should be 750 nucleotides long. Is it therefore likely that the sequencing reaction has covered the full length of the cloned *Nrg* product?

Q118. If the reaction has sequenced all of the insert and beyond, what features would you expect to see in the data at its furthest point?

Using BLAST searches on the Genbank database

With a bit of patience and an eye to detail, you should be able to confirm that the sequence data has probably come from the *Nrg* gene. The first 20 bases or so should match one of the primers you used for the PCR and beyond that, the data should run into the *Nrg* sequence (shown in Figure 2 of the PCR and Primer Design Exercise).

As you can appreciate, checking by eye is tedious and error-prone - quite often when you have sequenced a piece of DNA it may not be immediately apparent that it has come from the gene of interest or if the sequence is of value to your investigation. One way to assess this is to submit its sequence for comparison with many other known sequences at the GENBANK database. GENBANK is updated nightly, and anyone in the world can send a real or imagined sequence to be analysed for free, using the WWW.

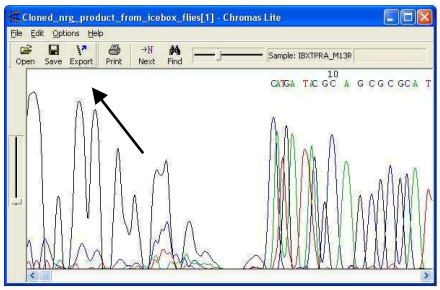
Go back to the web browser and head for the National Center for Biotechnology Information (NCBI) website:

- **http://www.ncbi.nlm.nih.gov/**
- When the page has loaded click on the BLAST button on the right hand side. BLAST stands for Basic Local Alignment Search Tool and it finds regions of similarity between biological sequences.

- Select the **"Nucleotide blast**". This sets up the software to receive a DNA (or RNA) sequence and compare it with other nucleic acid entries on the database. Under section "**Choose Search Set**" select **"others (nr etc)"** and at "**Program Selection**" click on "**highly similar sequences (megablast)".**

At this stage, you need to be able to enter the DNA sequence easily. You could (with patience) type in the data from the chromatogram, base by base but the accuracy will be better if we copy directly from the computer. To do this, select the **Export** button in Chromas Lite (arrowed in the picture below).

Select the folder on the computer where you are accumulating data from this exercise and enter an appropriate file name ("Raw sequence data" would describe it). The data - just the computer-called sequence, not the chromatogram - can then be saved in FASTA format.



Now open a simple text editing program - Notepad would be fine.

This can be found in the Start menu,

- "All programs" and then "Accessories"
- "Open", change "Files of type" from "Text documents (*.txt)" to "All files" or the software won't be able to see the FASTA file that you've just saved!

You should see a header field followed by the sequence taken from your chromatogram - just as bases, not coloured peaks. From your earlier analysis, you should be able to locate where vector sequence ends and the PCR product cloned into Bluescript begins.

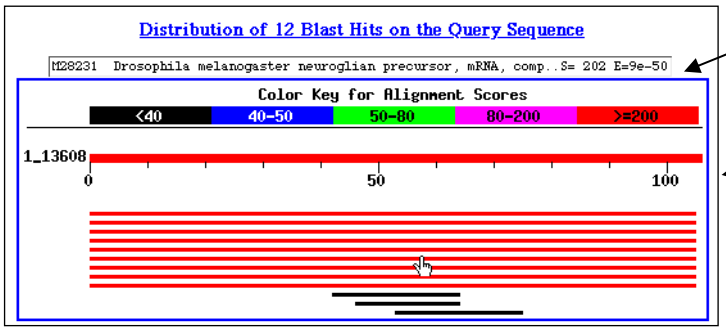
EDIT YOUR SEQUENCE

(1) Delete the vector sequence.

(2) Delete the last 3 or 4 lines of sequence to ensure that the data is of good quality for the next stage of the exercise.

- Now **save** the edited file back to your designated folder using "Save as". Choose a sensible name and save as a text file.
- Click and drag across the edited sequence file and copy. If you return to the NCBI BLAST page, you can now paste the data into the "Search" box. To compare your sequence against the standard database of all nucleotide sequences click the "**Blast!**" button. If you are lucky your results will come back very quickly. If the server is busy you will then get a message which says the page will be updated in X seconds. Use the time you are waiting to read ahead.

When you get your results, scroll down the page and you will see something like this:



This box displays the identity of any sequence match. Simply place the mouse pointer over any of the sequence matches below.

Thick red bar – linear representation of your submitted sequence.

Sequence matches: coloured lines show extent and significance of matches to your sequence found in the database

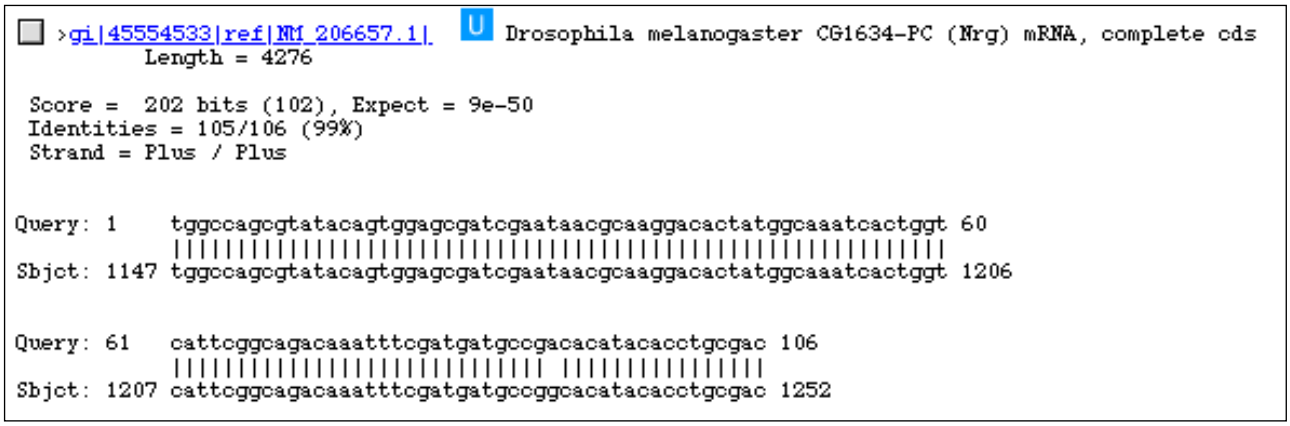
The red bar with numbers in the middle of the frame shown above represents your submitted sequence (we submitted 106 bases of sequence in this example). Below this are a series of lines of varying length. These represent sequences in the database which match your submitted sequence. The extent of the line tells you which part of your sequence is matched by the database sequence, and the colour of the line tells you how good the match actually is. The scale (as shown under "Color Key for Alignment Scores") runs from black, which means "pretty awful", to red which means "extremely good match". The scoring has two main components: the length of the matching sequence and the number of mismatches in the aligned part of the sequence. Thus a sequence of 300 bases which matches your submitted sequence with 10 mismatches between the sequences would score more highly than a 300 base match with 30 mismatches. Both of these would score higher than a perfectly matched alignment of only 20 bases in length.

If you hold the mouse pointer over a match line (as in the example above), the name of the matching sequence will appear in the box above the Color Key; in our example the match is to the "*D. melanogaster* neuroglian precursor

mRNA". Moving the mouse pointer across some of the other matches you will see displayed the identity of the other matching sequences. Further down the page you will see the names of all the matching sequences, together with the probability of the match being a mere coincidence.

Q119. Are you satisfied that the sequence you have gathered has actually come from the Drosophila *Nrg* gene?

Scroll down even further and you'll see a section that shows the actual alignment for each match, as in the following example:



Here you will see a "query" sequence (in this example, a neuroglian mutant called *Nrg3*, which is a temperature sensitive recessive lethal) aligned with the matching "subject" sequence found in the database (in this case, the *Nrg* mRNA, as it says at the top). In the example shown here, the alignment is between nucleotides 1 to 106 of our query sequence and 1147 to 1252 of the neuroglian mRNA. The two sequences are virtually identical but you will see that there is a mismatch between the two sequences, a G > A change at position 1236 in the neuroglian mRNA. Perhaps this nucleotide change could be responsible for the *nrg3* mutant phenotype. We shall pursue a similar analysis to try and see if we can identify mutations that might explain the behaviour of *icebox* flies.

If you scroll down you will come to neuroglian sequences named 'transcript variants'. Choose one of the transcript variants A, B or C to see if there are any mismatches which might be the *icebox* mutation. The other sequences represent, for example, genomic DNA whereas the transcript variants show the mRNA sequence.

Note the position of any nucleotide change(s) together with a six to eight bases on either side so that you can locate this region later. Annotate your notes with the sequence of the "subject" sequence at these points of mismatch.

Q120. Just because there is a DNA sequence change does not mean that the amino acid sequence will change – why?

Translation of DNA sequence

ExPASy is one of many useful sites where analysis tools for molecular biology are freely available. With this **Translate** software, you can paste DNA sequence into the open box and the software will generate all possible protein coding sequences for you.

Set up a new browser window, and go to

- **http://www.expasy.ch/tools/dna.html** (you can go to this web site from the Molecular Methods Moodle site – Computer Exercise section)
- Go back to Notepad (or whatever text editor you have been using) and open up the *edited* sequence file (*not the original, unprocessed data!*). Click, drag and copy the data.
- Paste in the data from the file with our sequenced clone DNA sequence. From the drop-down menu beneath your pasted sequence, select **Includes nucleotide sequence**. Now click **TRANSLATE SEQUENCE**.

The software takes the DNA sequence and translates it in all forward and reverse reading frames, producing aligned protein and nucleic acid sequences. **'Stop' or '–'** marks potential stop codons for each of the reading frames. Remember that our PCR product covers part of an intron as well as exon 3. The natural *start* codon will not therefore be present and stop codon(s) might be present in sequences from the intron. Nevertheless, the most likely reading frame should be free of stop codons.

Q121. What is meant by the term 'reading frame'?

Q122. Which reading frame looks most likely to encode the Nrg protein from the icebox flies?

Translated Blast Searches

The BLAST software at NCBI has been adapted for a range of jobs. We have made DNA/DNA nucleotide comparisons using "**blastn**". Another useful search tool, "**blastx**", assumes that your sequence encodes a peptide, and it compares all 6 possible reading frames of your sequence with the deduced peptides of all known sequences in the database. We now know that we have a potential reading from in the sequence data - we can now use blastx to see if there are any amino acid changes as a result of nucleotide substitutions.

The different 'flavours' of blast. Each has a use – ask you demonstrator if you would like to know more.	
blastn	Search a nucleotide database using a nucleotide query
blastp	Search protein database using a protein query
blastx	Search protein database using a translated nucleotide query
tblastn	Search translated nucleotide database using a protein query
tblastx	Search translated nucleotide database using a translated nucleotide query

To carry out a "**blastx**" search:

- go back to the BLAST front page and this time select "**blastx**".
- Submit your *icebox* mutant DNA sequence as before; your results page will be colour coded, similar to the results from blastn, but note that even a black–coloured match from blastx does indicate a fairly good relationship.

Here is our result with the *Nrg3* mutant sequence, the example that we used before:

```
>gi|24640619|ref|NP_727274.1|  CG1634-PB [Drosophila melanogaster]
gi|14286138|sp|P20241|NRG_DROME   Neuroglian precursor
gi|22831957|gb|AAN09236.1|   CG1634-PB [Drosophila melanogaster]
          Length = 1302

 Score = 76.6 bits (187), Expect = 2e-13
 Identities = 34/35 (97%), Positives = 34/35 (97%)
 Frame = +2

Query: 2   GQRIQWSDRITQGHYGKSLVIRQTNFDDADTYTCD 106
           GQRIQWSDRITQGHYGKSLVIRQTNFDDA TYTCD
Sbjct: 284 GQRIQWSDRITQGHYGKSLVIRQTNFDDAGTYTCD 318
```

The *Nrg3* mutant sequence has a change of G to D at amino acid 313. Looking at the DNA sequence data from blastn a couple of pages back in the manual and using the genetic code table on the last page of the manual, we can tell that this change arose from GGC (encoding glycine) changing to GAC (aspartic acid). This alters the smallest possible amino acid side chain (–H) for a larger, acidic side chain ($-CH_2COOH$), and therefore is a candidate for causing the *nrg3* phenotype. We can write it using standard notation as: G313D (GGC>GAC).

Look at your own blastx results. You may see your sequence aligned with several different versions of the *Drosophila melanogaster* neuroglian sequence – these are sequences submitted to the database by different labs, and you may find that there are differences between the sequences from the database, even though they are all from wild-type flies.

Q123. Suggest two reasons why there might be differences between the neuroglian sequences from the database.

Q124. What is the common domain that is present in all of the proteins discovered in your BLAST search (Nrg/hemolin/L1CAM/neurofascin)?

Focus on any differences between your sequence and the database subjects that are consistent for all of the *D. melanogaster* hits.

Q125. Which amino acid changes might be responsible for the icebox phenotype? Write down the change(s), their position(s) in the protein sequence and a few residues at either side to aid location of the altered amino acid(s).

Multiple Protein Alignment

Multiple protein alignments are important tools in studying proteins. The basic information they provide is identification of conserved sequence regions. This is very useful in designing experiments to test and modify the function of specific proteins, in predicting the function and structure of proteins and in identifying new members of protein families. In order to help us do this, we can export the protein files and do a protein alignment using software called ClustalX.

Once you have the sequence of the same protein from several species, there are several ways to get information from this. One of the most straightforward is a multiple sequence alignment. To understand how this works, think about American versus English spellings:

E: Note the catalogue number, your favourite colour, and send a cheque with your order.
A: Note the catalog number, your favorite color, and send a check with your order.

These sentences can be aligned, so that the equivalent bits coincide, by artificially introducing gaps (-)

E: Note the catalogue number, your favourite colour, and send a cheque with your order.
A: Note the catalog-- number, your favo--rite colo-r, and send a check-- with your order.

Thus we can align functionally equivalent bits of protein sequence, even though evolution has introduced many changes in sequence composition and length.

We have downloaded and aligned the following sequences:

Cloned Nrg	Nrg sequence that you cloned in the lab
Drosophila WT Nrg	Nrg protein from wild type *Drosophila melanogaster*
Anopheles Nrg	Nrg from *Anopheles gambiae* (mosquito)
Moth Hemolin	Hemolin from *Pseudoplusia includes* (moth)
C. elegans L1CAM	L1CAM (a gene of the same family) from *Caenorhabditis elegans* (worm)

- Go back to your Moodle page and right click on '**Nrg alignment file'** and save to a folder of your choice.
- You should have **ClustalX2** installed on your computer. Locate the P-drive as before, then open the ClustalX2 file. Click on the Clustal application and then 'run' to launch the programme.
- In ClustalX go to **FILE** then open the saved aligned file (**Nrg sequences of Nrg.aln**)

Q126. Can you identify the 2 mutations that are consistently different between your cloned icebox Nrg sequence and wild type Nrg from Drosophila, Mosquito and Moth? Does this provide you with any information which might be useful in deciding if the mutation you identified might be responsible for the *icebox* mutant phenotype?

Q127. Why is one amino acid mutation highlighted in the same blue colour? What does this represent?

Q128. L1CAM is a human gene of the same gene family as *Drosophila* neuroglian. Mutations in this gene lead to what disease phenotype? (To find out Google L1CAM)

Q129. It is relatively quick and easy to make *Drosophila* transgenic. How could you establish beyond doubt that a mutation you have identified, and not one of the other changes, is actually responsible for the icebox phenotype?

PART B
USING PCR AND REAL-TIME QUANTITATIVE PCR TO DETECT AND QUANTIFY HIV

BACKGROUND TO THE LAB

Human Immunodeficiency virus (HIV-1) has caused a global epidemic of concern to us all and in the thirty years since its discovery, has resulted in the death of an estimated 30 million people. Approximately 40 million individuals world-wide are currently infected with this virus and many of them will also die of AIDS, the syndrome that results when the virus has ablated their CD4+ T cells. Progression to AIDS is influenced by a variety of host and viral factors, including the copy number of virus particles in the blood of an infected individual. In general, the higher the number of virus particles, the faster progression occurs. The level of virus in the blood can be measured using antibody- and PCR-based methods, especially real-time qPCR.

Real-time qPCR techniques have revolutionised diagnostic virology labs. This method is similar to conventional PCR, which you will also have the opportunity to perform and compare, in that it amplifies specific regions of DNA from a sample of nucleic acid. The product is monitored as the reaction proceeds. During a real-time qPCR, the PCR product is quantified using a fluorescent dye called SYBR Green, which binds to double stranded DNA. Therefore, the intensity of the fluorescence increases as more PCR product is produced with each cycle of the PCR. You can therefore accurately measure viral copy number within a sample.

You will be using conventional and real-time qPCR techniques in this lab to help diagnose whether any of the given patient samples are infected with the HIV virus. This combination of techniques is useful to perform and provide the basis for their comparison. These techniques are based on methods currently used in diagnostic labs to detect this virus and should allow you to transform PCR from a theoretical concept into a real life application.

Lab Plan

Briefly, you will be doing the following:

AM

- Introductory talk
- Set up conventional PCR
- Set up real-time qPCR

PM

- Run agarose gel using PCR samples

Results from the real-time qPCR run are gathered and collated to be analysed in a computing session.

Additional Resources

Suggested reviews:

Wong M.L. and Medrano J.F. (2005) Real-time PCR for mRNA quantitation. *Biotechniques* **39**, 1-10

Fraga D. *et al*. (2008) Real-time PCR. *Current Protocols Essential Laboratory Techniques* **10.3.1-10.3.34.** (http://onlinelibrary.wiley.com/doi/10.1002/9780470089941.et1003s00/pdf)

Valasek M.A. and Repa J.J. (2005) The power of real-time PCR (2005) *Advan. In Physiol. Edu.* **29**, 151-159.

DAY 5 (AM) - CONVENTIONAL HIV PCR

The DNA sequence below shows a partial sequence of the HIV-1 envelope glycoprotein (env) gene. The forward and reverse primers were designed against the underlined sequences. A successful PCR in the presence of HIV-1 env template sequence should amplify the region in bold and the amount of product generated will be directly proportional to the concentration of HIV in the original sample.

ATATAAAGTA ATAAAAATTG AACCATTAGG AATAGCACCC ACCAAGGCAA AGAGAAGAGT
GGTGCAGAGA GAAAAAAGAG CAGTGGGAAT AGTAGGAGCT ATGTTCCTTG GGTTCTTGGG
AGCAGCAGGA AGCACTATGG GCGCAGTGTC ATTGACGCTG ACGGTACAGG CCAGACAATT
TTTGCTGAGG GCTATTGAGG TCTGTTGCAA CTCACAGTCT GGGGCATCAA GCAGCTCCAG
GCAAGAGTCC TGGCTGTGGA AAGATACCTA AGGGATCAAC AGCTCCTAGG GATTTGGGGT
TGCTCTGGAA AACTCATTTG CACCACTGCT GTGCCTTGGA ATGCTAGTTG GAGTAATAAA
TCTCTGGAAG ACATTTGGGA

Figure 1 Part of the HIV-1 envelope glycoprotein (env) sequence. The region in bold will be amplified during the PCR in the lab.

Reagents

You will require the following reagents (with red dots) to run this reaction. Check everything that you need is present before you start.

- 1 x 500µl Qiagen Sterile Nuclease Free Water
- 1 x 15ul PCR HIV F-primer (10µM)
- 1 x 15ul PCR HIV R-primer (10µM)
- 1 x 10ul PCR Blood cDNA Sample 1
- 1 x 10ul PCR Blood cDNA Sample 2
- 1 x 10ul PCR positive control cDNA
- 1 x 150ul PCR Master Mix (green reagent)

Method

1. Check that everything has defrosted. You may have to briefly centrifuge your tubes to ensure all the liquid is at the bottom.
2. Make up a PCR Master Mix for 5 reactions in a 1.5ml eppendorf:

Reagent	**Master Mix for 5 reactions (µl)**
PCR Master Mix	125
10µM PCR HIV F-primer	10
10µM PCR HIV R-primer	10
H_2O	95

3. Using a marker pen, write your initials on the top of each PCR tube and label them 1-4 so that you can distinguish which tube contains which sample. Place 48 µl of the Master Mix into each PCR tube and then add 2µl of the following into the appropriate tube:

   Tube 1: PCR Blood cDNA sample 1
   Tube 2: PCR Blood cDNA sample 2
   Tube 3: PCR positive control cDNA
   Tube 4: Water (Negative control)

4. Close the tubes, mix the contents by flicking the tube and spin **very briefly at low speed** (less than 4K for a few seconds only) to collect the contents to the bottom of the tubes.

5. Place the tubes in the rack at the front of the lab to be put into the PCR machine. The reaction conditions will be as follows:

Start		95°C	5 min
5 X	denaturing	95°C	1 min
	annealing	70°C	30 sec
	extension	72°C	1 min
25 X	denaturing	95°C	30 sec
	annealing	70°C	30 sec
	extension	72°C	1 min
1 X		72°C	5 min

Once all the class samples are in the PCR machine, the demonstrator will start the cycling reaction. This will take around 2 hours. You will run these samples on an agarose gel in the afternoon.

Questions

Q130. The template you will be using in the PCR is in the form of cDNA. Briefly, describe how cDNA is generated?

Q131. Why is it important to use cDNA for these experiments?

Q132. The primers for this reaction have been designed to the sequences underlined in the sequence shown in Figure 1. They are referred to in this lab as HIV F-primer (forward primer) and HIV R-primer (reverse primer). Using your knowledge of PCR primer design, write out below what the forward and reverse sequences would be (hint: write out the sequence 5' to 3').

HIV F-primer:

HIV R-primer:

Q133. What size is the PCR product you would amplify and expect to see if you ran this on a gel?

Q134. Calculate the T_m of these primers (T_m = 2°C x (A+T) + 4°C x (C+G)).

Q135. Look carefully at the chosen sequence of both primers. Do you think the choice is a good one? Qualify your answer using your knowledge of primer design.

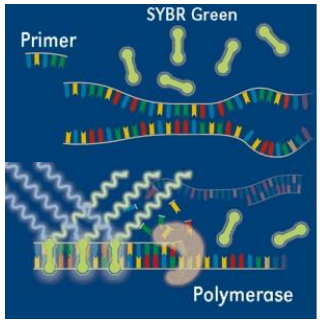
Q136. When designing PCR primers for use in a real-time qPCR, the final PCR product size should be between 50 and 150 nucleotides. Why do you think this might be?

DAY 5 (AM)– REAL-TIME QUANTITATIVE PCR

Introduction

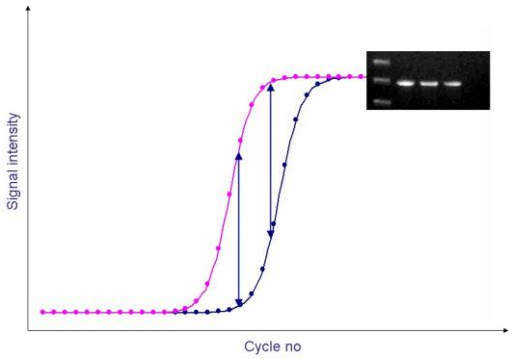
Reading: Dale et al., 3rd Ed: p123–127.

Real-time quantitative PCR (qPCR) is highly suited for a wide range of applications, such as gene expression analysis, determination of viral load, detection of genetically modified organisms (GMOs), SNP genotyping, and allelic discrimination. Real-time qPCR allow accurate quantification of starting amounts of DNA, cDNA, and RNA targets. PCR products can be detected using either fluorescent dyes that bind to double-stranded DNA or fluorescently labelled sequence-specific probes. Fluorescence is measured during each cycle and the amount of this fluorescence is proportional to the amount of PCR product.



The principle of SYBR-based real-time qPCR, which you will be using in this lab practical, is a standard PCR carried out in the presence of a dye, SYBR, which fluoresces when intercalated in the DNA helix. The fluorescence will increase as the amount of the PCR product increases and is quantified after each completed PCR cycle.

The cycle at which the fluorescence exceeds a detection threshold, the **Ct (threshold cycle)** correlates with the number of target cDNA molecules present in the added cDNA. Therefore, by comparison to a calibration curve, it is possible to quantify in absolute amounts the number of target molecules in cDNA samples.



You will be provided with two cDNA samples, each derived from the blood of individuals who are being tested for HIV infection. Using the reagents provided you will be able ascertain whether they are HIV positive or not and assess the copy number of HIV genomes. Quantitative standards have already been run through the real-time qPCR machine for you. Using these, you can compare the results you obtain with your cDNA samples with known concentrations of virus. This will allow you to accurately quantify your samples for viral load.

Reagents

You will require the following reagents (with blue dots) to run this reaction. In addition, you will require 4 real-time fast reaction PCR tubes with caps (both found in strips in a Universal container) and a pink rack, which is used to hold these real-time PCR tubes. Check everything that you need is present before you start.

- 1 x 500µl Qiagen Sterile Nuclease Free Water
- 1 x 20 µl qPCR HIV F-primer (2µM)
- 1 x 20 µl qPCR HIV R-primer (2µM)
- 1 x 10 µl qPCR Blood cDNA sample 1
- 1 x 10 µl qPCR Blood cDNA sample 2
- 1 x 10 µl qPCR positive control cDNA
- 1 x 80 µl SYBR Select Master Mix (in a dark brown Eppendorf)

Method

1. Considerations before starting:

a. Please take your time pipetting and do so as accurately as possible!

b. Real-time qPCR is VERY sensitive to contamination, therefore:
   i. Clean bench area with detergent first
   ii. Use gloves to handle everything and change tips as necessary
   iii. **Use sterile p2, p20 and p200 FILTER TIPS**

c. SYBR Select Master Mix is light sensitive! You must work efficiently, but accurately.

2. Setting up real-time qPCR:

1. Make a qPCR Master Mix for 6 reactions in a sterile 1.5 ml Eppendorf on ice as follows:

Reagent	**qPCR Master Mix for 6 reactions (µl)**
2x qPCR SYBR Select Master Mix	60
2µM qPCR HIV F-primer	12
2µM qPCR HIV R-primer	12
H_2O	24

2. Gently mix the Master Mix tube by flicking the tube and briefly centrifuge the tube for 10 secs at high speed.

3. Place the strip of 4 real-time qPCR tubes in the pink tube holder. The tubes have numbers on them which will be either 1-4 or 5-8. Mark the strip of tubes so the number 1 or number 5 tube has a dot on the side of the tube. **NOTE: Do NOT label the top of the tube**.

4. Accurately pipette 18 µl of qPCR Master Mix into each **real-time qPCR tube** and then add 2µl of the following into the appropriate tube in the correct orientation:

   I. qPCR Blood cDNA sample 1
   II. qPCR Blood cDNA sample 2
   III. qPCR positive control cDNA
   IV. Water (negative control)- NTC (no template control)

The **orientation of the tubes** is as follows:

Tube number	1 or 5	2 or 6	3 or 7	4 or 8
Sample	**Sample 1**	**Sample 2**	**+ve control**	**NTC**

Press the lids on firmly.

5. Take your sample to your lab leader who will position your tubes into the real-time qPCR plate. Please record the number that the lab leader gives you for your analysis.

6. The samples from the group will be run in the ABI Step One Plus Real-time Machine.

The following conditions will be used during the real-time qPCR:

Stage	**Temperature °C**	**Time (min)**	**Cycle number**
Melt/Hot start	95	3 min	1
Amplification	95	3 secs	35
	60	30 secs	
Melt curve	95	15 secs	1
	60	1 min	
	60-95	0.3 secs per temp	

Questions

Q137. Why is it important to prepare a Master Mix?

Q138. Why were you told not to label the top of the real-time qPCR tubes?

Q139. Comment on how real-time qPCR conditions compare to conventional PCR conditions.

Q140. What other types of experiments do you think real-time qPCR could be used for?

DAY 5 (PM) – AGAROSE GEL ELECTROPHORESIS OF CONVENTIONAL HIV PCR

1. Preparation of the Agarose Gel

The gel mix has been prepared for you: note, here you are using a 2% agarose gel. A demonstrator will add 2µl of gel-red into your gel.

- When the gel has set, flood the gel tank with the TAE buffer (there should be a depth of about 5 mm TAE above the gel). Once you have removed the comb and blocks, you are ready to load your samples.

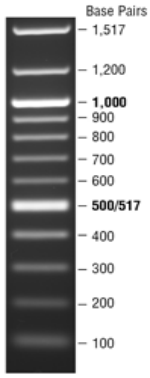
2. Preparation of PCR Samples for Electrophoresis and Running the Gel

- There is one agarose gel for three pairs.
- Collect your 4 PCR samples from the PCR machine. Load 10µl of each sample onto the gel. There is also a DNA Ladder to load – one per gel (15 µl).
- Remember to write your name on the loading sheet so you know where your sample is on the gel.
- After running the gel, please take it to be visualised.

CAUTION: ***potentially lethal voltages. Do not handle gel kits unsupervised, and if you notice a problem in a running gel (like a leak, steam or smoke), seek help but do not touch.***

Interpreting your gel

You used the 100bp DNA ladder below on your gel. Correlate the molecular weight of the bands below with the ladder you ran on your own gel.



Questions

Q141. Why would you use a 2% agarose gel to run this PCR product on?

Q142. Does the size of your PCR product reflect what you predicted?

Q143. Which of your blood cDNA samples is showing a positive PCR result for the HIV *env* gene?

Q144. Have your positive and negative controls worked? If not, what would you do differently next time you ran this experiment?

COMPUTER EXERCISE 3: REAL-TIME QPCR DATA ANALYSIS

Introduction

The real-time qPCR machine will make amplification plots based on the intensity of fluorescence from the Sybr Green from each sample as the amount of PCR product increases with cycle number. The amount of DNA present in a sample is inversely proportional to the cycle number at which it starts to amplify. Therefore, the greater the amount of DNA in your starting sample, the earlier it will start to amplify within the reaction (Figure 1).

The Ct (cycle threshold) is defined as the number of PCR cycles required for the fluorescent signal generated for a specific sample to cross the threshold which exceeds background level (Figure 2). Ct values of <29 are positive reactions indicative of abundant target nucleic acid in the sample. Ct values of >29 indicates small amounts of nucleic acid in the sample.

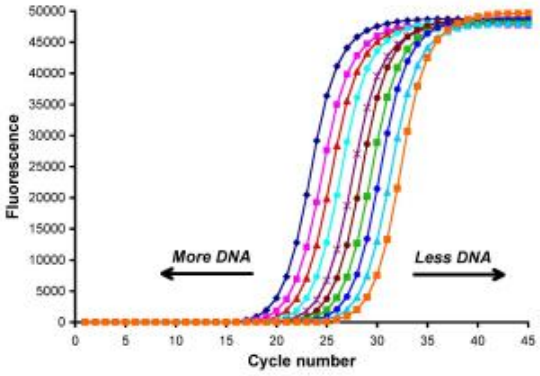


Figure 1. Amplification plots in a real-time qPCR.

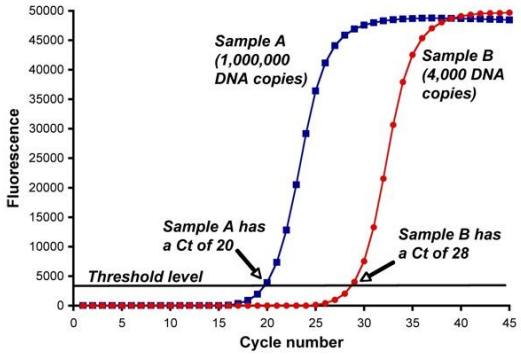


Figure 2. Ct values in a real-time qPCR (example only).

Standard Curve

Data has been generated for you using increasing known concetrations of control cDNA. The same experimental conditions as your own qPCR analysis were utilised for this purpose.

You will use this data to produce your own standard curve and from this calculate virus copy number in the two blood cDNA samples and the positive control.

Questions

Q145. Open the Standard Curve Excel file on Moodle, and fill in the appropriate values for the Standard Curve in Table 1.

Use Excel, a calculator or the equation below to calculate Ct mean and Ct SD

Sample number	Sample (copies/ml)	Ct	Ct Mean	Ct SD
A1	1,000,000			
A2	1,000,000			
A3	1,000,000			
B1	100,000			
B2	100,000			
B3	100,000			
C1	10,000			
C2	10,000			
C3	10,000			
D1	1000			
D2	1000			
D3	1000			
E1	100			
E2	100			
E3	100			

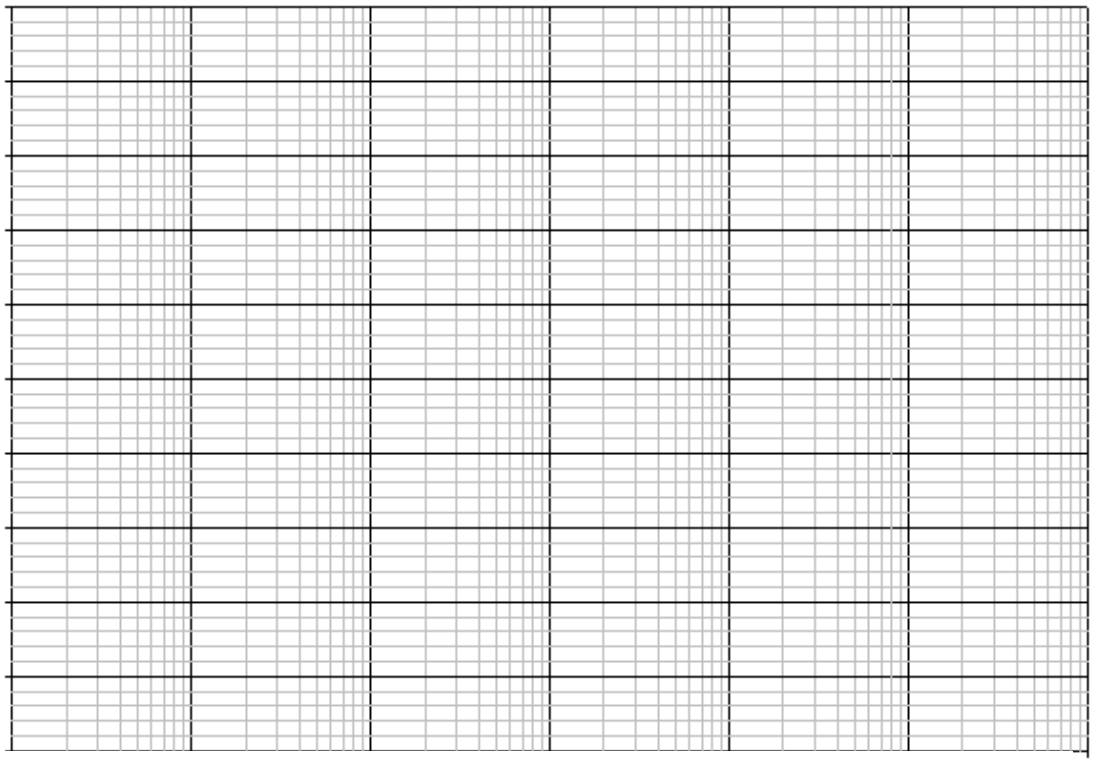
Table 1. Ct values for standards.

Standard Deviation Equation

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

s=standard deviation
x=each value in the data set
$\bar{X}$=mean of values in data set
Σ=sum of
n=number of values in the data set

Q146. Draw a line of best fit using the data obtained in Table 1 on the log paper provided [concentration of standard, x-axis; Ct mean, y-axis].



Experimental Data
Your data is on the Molecular Methods Moodle site. Once you have retreived your real-time qPCR Ct values:

- Note your own Ct values in Table 2 for the reaction that you set up.
- Examine the class data and use as if they are experimental replicates.

Q147. Calculate the mean and standard deviation of the class data and fill in Table 2.

Q148. Plot the Ct values onto your standard curve and calculate the viral load of your unknown patient samples (copies/ml) using the class data.

Sample name	Own expt Ct	Class Ct Mean	Class Ct SD	Copies/ml (class Ct)
Blood cDNA sample 1				
Blood cDNA sample 2				
Positive control cDNA				
Negative control (water)				

Table 2 Real-time qPCR data from Blood cDNA samples

Q149. Look at the Ct values of the samples that you set up. Does your data correlate with the rest of the class? If not, how can you explain the error?

Q150. Look at the calculated Standard Deviation of the class data. What does this tell you about the class data?

Q151. Can you come to a conclusion as to which patient sample has HIV?

Q152. Looking at the table below, what information could you give the individuals who had this test?

Viral load (copies/ml)	Descriptor	Notes
~100,000	High viral load: antiviral treatment recommended	
~10,000	Low viral load: further investigation required in order to determine treatment option, eg $CD4^+$ count/clinical presentation	
<50	Undetectable	Current tests can only reliably detect >50 copies/ml

Q153. Compare your results from your PCR and real-time qPCR experiments. Do these results correlate? If not, what do you think has happened? What additional information has the real-time qPCR expt provided?

APPENDIX 1 - SAFETY ISSUES

Government legislation requires that all procedures in companies of more than five employees have a written risk assessment, and where appropriate a written scheme of work. This statement summarizes the risk assessment, and the lab manual constitutes a written scheme of work. This means that you are obliged to read, understand and abide by this manual in the performance of your lab work.

The lab course is designed to minimize both the severity of the hazards, and the risk that something will go wrong. In summary, this course presents a low risk to you when you perform the skills as directed by this manual.

If something goes wrong, you should seek a demonstrator's help immediately.

Major points

Wear a lab coat and gloves at all times. Change them immediately if they become contaminated.

Be familiar with the use of the eyewash station, and be prepared to assist others to use it.

Some of the **chemicals** are irritants, poisons, or possible carcinogens. You are substantially protected by the small volumes and great dilutions provided, but you should nonetheless work carefully and precisely; dispose of everything as suggested, and seek advice if you spill anything.

You will be handling **recombinant bacteria** containing transgenes. This has previously been assessed as required by the regulations concerning genetic manipulation. In essence, the bugs are genetically crippled to make it unlikely that they would survive outside the laboratory environment; the gene fragment we are working with is not thought to be hazardous, and does not contain a complete reading frame; and we are not attempting to drive expression to high levels. This means that the work receives the lowest possible risk classification, and can be performed on the open bench, subject to the provisions of "Good Microbiological Practice".

Other than described above, the most important requirement of good microbiological practice is to dispose of all materials exposed to recombinant DNA to bleach or to autoclave bags, and to have bleach available to decontaminate spills.

You must label everything clearly with the supplied indelible marker. It must be possible to identify not only what is in a tube, but also whom it belongs to. Quite apart from safety issues, you will lose your samples in a sea of similar tubes very quickly if you don't label them carefully!

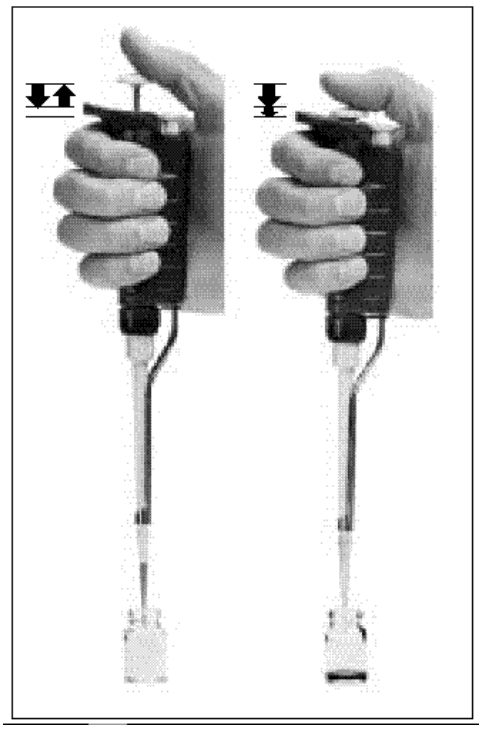
There are also important **physical hazards** associated with the work. In the main this refers to exposure to lethal **voltages** when handling electrophoresis apparatus. Never work with this apparatus except under the direct supervision of a demonstrator.

APPENDIX 2 - USING A GILSON PIPETTE

Molecular biology does *not* require fantastic dexterity, intelligence or skill to perform, but relies **critically** on your ability to dispense small volumes of liquid accurately and consistently, and without cross-contamination. Although **you** all **think** you know what you're doing, **we know** that when you assemble your first PCR reaction in your tube, your versions of 50 µl final volume can vary by a factor of as much as **10 times**!

So take the time to check this guide first.

- Graduated pipettes, like the Gilsons we use here, have maximum volumes (marked on the top of the plunger).
- Never try to force a pipette above its maximum volume, or below zero.
- Choose a Gilson appropriate to the volume you want to measure; don't try to use a Gilson for less than 10% of its maximum volume. e.g. the minimum volume you can accurately dispense with a 20 µl pipette (or "P20") is 2 µl.
- Use a tip appropriate for the pipette. Tips come in three colours: blue, yellow and white. Choose the tip that matches the top of the Gilson plunger. For you, this means yellow tips for everything except the 1000 µl pipette (with the blue top); this takes blue tips.
- **Hold the Gilson pipette as shown opposite**, with your **thumb** (not a finger) on the plunger to regulate the dual thumb pressure system.
- The **graduated** volume is to the **first** pressure (photo on left).
- The second pressure is used only to **expel** all the liquid (right photo).
- Make sure you can tell the difference before you start.

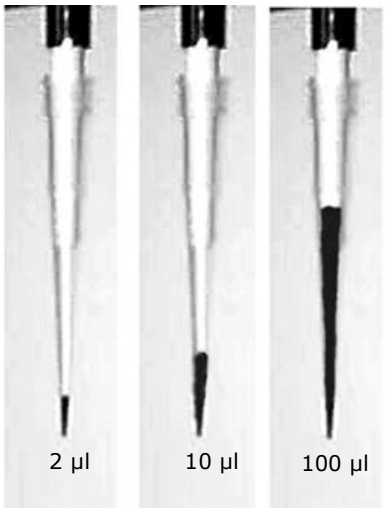


Always draw up liquid by **raising your thumb slowly**. If you simply take your thumb off, you will suck in air and contaminate the barrel of the pipette with spray. If you find air bubbles in the column of liquid, this is a sure sign that you're letting your thumb up too fast.

Small volumes are easily lost. Check every pipetting operation by eye. You should be satisfied that:

- You have taken liquid up into the tip
- It looks about the right amount
- That it all goes into its destination tube

You should have a feel for volumes in these tips. In a yellow tip, a 2 µl aliquot is a column approximately 2 mm long. As a practice, draw up 2 µl, 10 µl and 100 µl into a yellow tip using the appropriate pipette, and memorise the appearance of the volume in the tip (as shown below). This is your best insurance against subsequent pipetting errors!



P1000 for volumes 201 µl to 1000 µl	**1** **0** **0**	**= 1000 µl (1 ml)**
P200 for volumes 21 µl to 200 µl	**2** **0** **0**	**= 200 µl (0.2 ml)**
P20 for volumes 1 µl to 20 µl	**2** **0** **0**	**= 20 µl (0.02 ml)**
P2 for volumes 0.1 µl to 2 µl	**2** **0** **0**	**= 2 µl (0.002 ml)**

APPENDIX 3 - RECIPE FOR AGAROSE GEL ELECTROPHORESIS OF DNA

Linear DNA can be resolved by size using agarose gels of various concentrations. The greater the percentage of agarose, the smaller the linear DNA that can be resolved. Since you have amplified different PCR products during the conventional and real-time PCR protocols, you require to use a different % of agarose to resolve these.

The size of the PCR product you amplify in **Part A** of this manual is 750bp and therefore, the product was resolved using 1% agarose solution. The following describes how this was prepared:

In a conical flask, place 1g agarose + 100 ml TAE buffer. (TAE buffer contains Tris and sodium acetate to buffer the pH; and EDTA to inactivate nucleases which all require divalent ions [these ions are "chelated" by EDTA]).

Microwave until dissolved (about 1 min). Allow to cool to 60 °C.

CAUTION: ***Melted agarose is very hot and sticky, and can cause nasty burns.***

Add an appropriate amount of a DNA intercalating dye. Many labs will use ethidium bromide, which is a suspected carcinogen and needs to be handled with care. In this lab, we shall use a non-toxic compound called SYBR Safe. This has properties similar to ethidium bromide but presents no hazard to you!

Seal the edges of a gel tray and pour in (with comb in place) the melted agarose to a depth of about 4 mm. Remove any air bubbles, as these would distort the running of the gel.

Allow to cool to room temperature. When set, remove comb and sealing tape/blocks, and place in the gel tank.

Flood with TAE buffer. There should be a depth of about 2mm TAE above the gel.

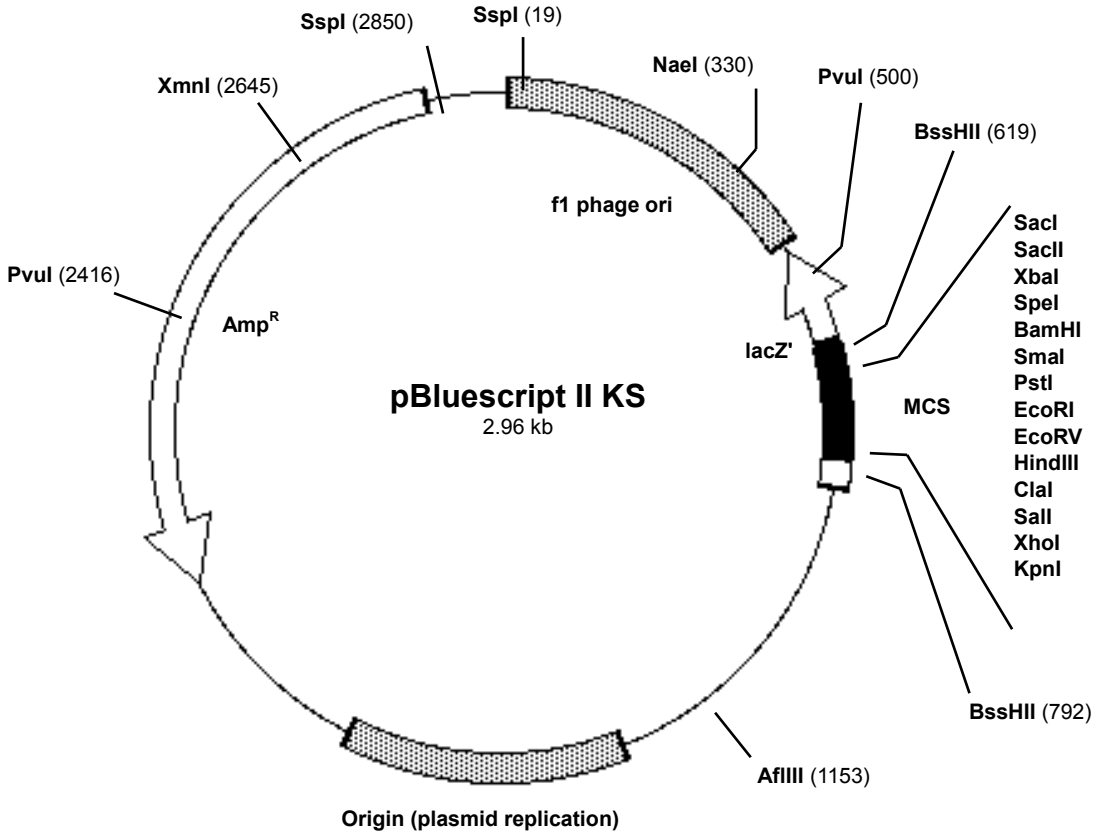
Add loading dye (50% glycerol, 0.01% bromophenol blue) to your sample at a ratio of 1:4. The glycerol causes the sample to sink to the bottom of the wells during loading and the bromophenol blue dye provides a marker for the progress of electrophoresis.

An appropriate DNA ladder (i.e. DNA fragments of known size) should always be run in parallel to enable calculation of the size of your fragments.

Run at 100 V for around 1–2 h, or until the dark blue loading dye is about 1 cm from the bottom of the gel.

NOTE: In **Part B** of this manual, the PCR product is <150bp. This smaller product was separated using TAE containing 2% agarose solution. In addition, a different intercalating stain was utilised. GelRed™ is a new generation of fluorescent nucleic acid gel stain designed to replace the highly toxic ethidium bromide. In addition, it has higher sensitivity than SYBR Safe and is better at visualising smaller PCR products on a gel.

APPENDIX 4 - CIRCULAR PLASMID MAP OF PBLUESCRIPT VECTOR



Multiple Cloning Site (MCS) DNA Sequence

Forward DNA sequencing primer

```
5'GTAAAACGACGGCCAGT3'

                  BssHII      T7 promoter                SacI    SacII          XbaI  SpeI  BamHI SmaI  PstI  EcoRI
600               |           ------------>                |       |              |     |     |     |     |     | 71
2
  GTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATT
C
  CATTTTGCTGCCGGTCACTCGCGCGCATTATGCTGAGTGATATCCCGCTTAACCTCGAGGTGGCGCCACCGCCGGCGAGATCTTGATCACCTAGGGGGCCCGACGTCCTTAA
G


EcoRV     ClaI    HincII
|   HindIII |      SalI   XhoI          KpnI                                BssHII
713   |     |        |     |             | 760                              |       800                     82
6
GATATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCC
CTATAGTTCGAATAGCTATGGCAGCTGGAGCTCCCCCCCGGGCCATGGGTCGAAAACAAGGGAAATCACTCCCAATTAACGCGCGAACCGCATTAGTACCAGTATCGACAAAGG
                                                                  <------------
                                                                  T3 promoter                    3'GTACCAGTATCGACAAAGG5
'
```

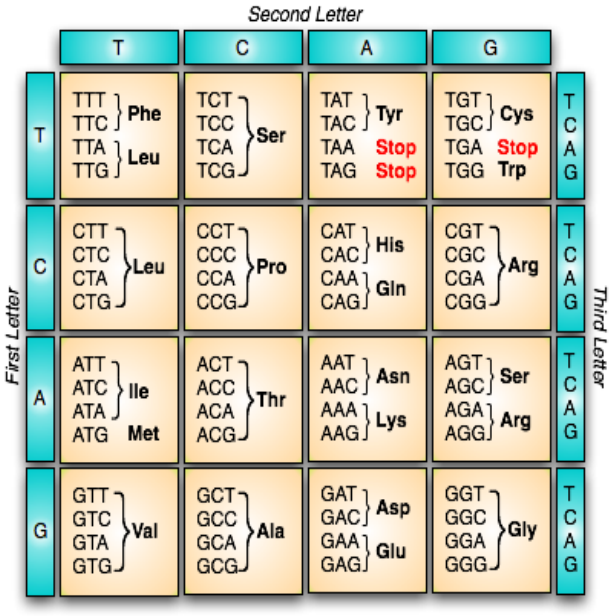
Reverse DNA sequencing primer

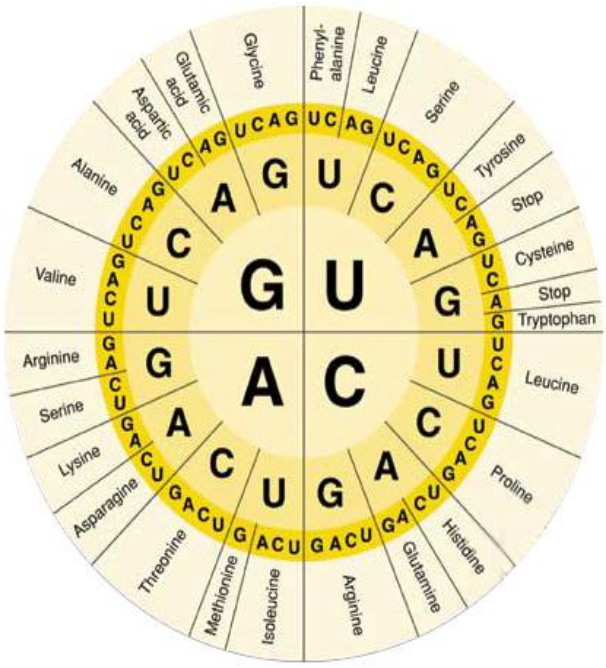
APPENDIX 5 - USEFUL WEB URLS:

Ensembl - online genome databases for vertebrates and other eukaryotic species	http://www.ensembl.org/
European Bioinformatics Institute - an academic research institute located on the Wellcome Trust Genome Campus near Cambridge (UK), which provides a range of bioinformatics services.	http://www.ebi.ac.uk/
FlyBase – a database of *Drosophila* genes and genomes	http://flybase.org/
Gene - integrates information from a wide range of species. A record may include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources worldwide.	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
GeneCards - a searchable, integrated, database of human genes that provides genomic related information, on all known and predicted human genes.	http://www.genecards.org/index.shtml
HUGO Gene Nomenclature Committee - approved gene nomenclature and associated resources including links to genomic, proteomic and phenotypic information, as well as dedicated gene family pages.	http://www.genenames.org/
Human Gene Mutation Database - includes the first example of all mutations causing or associated with human inherited disease	www.hgmd.org
NCBI homepage (access to OMIM, Entrez, Genome Biology resources, etc)	http://www.ncbi.nlm.nih.gov/
NCBI Taxonomy Browser - direct links to some of the organisms commonly used in molecular research projects	http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/
National animal genome research program – data on the genome (DNA) of some animal species	http://www.animalgenome.org/

APPENDIX 6 - THE GENETIC CODE

	T		C		A		G		
T	TTT	Phe (F)	TCT	Ser (S)	TAT	Tyr (Y)	TGT	Cys (C)	**T**
	TTC		TCC		TAC		TGC		**C**
	TTA	Leu (L)	TCA		TAA	STOP	TGA	STOP	**A**
	TTG		TCG		TAG	STOP	TGG	Trp (W)	**G**
C	CTT	Leu (L)	CCT	Pro (P)	CAT	His (H)	CGT	Arg (R)	**T**
	CTC		CCC		CAC		CGC		**C**
	CTA		CCA		CAA	Gln (Q)	CGA		**A**
	CTG		CCG		CAG		CGG		**G**
A	ATT	Ile (I)	ACT	Thr (T)	AAT	Asn (N)	AGT	Ser (S)	**T**
	ATC		ACC		AAC		AGC		**C**
	ATA		ACA		AAA	Lys (K)	AGA	Arg (R)	**A**
	ATG	Met (M)	ACG		AAG		AGG		**G**
G	GTT	Val (V)	GCT	Ala (A)	GAT	Asp (D)	GGT	Gly (G)	**T**
	GTC		GCC		GAC		GGC		**C**
	GTA		GCA		GAA	Glu (E)	GGA		**A**
	GTG		GCG		GAG		GGG		**G**





APPENDIX 7 – HOW TO USE A QR CODE

A QR code (Quick response code is a one (or two) dimensional barcode that has data embedded into it. This code is readable by QR scanners which can be downloaded onto any smart phone or mobile device (e.g. tablet.)

There are a number of apps available for reading a QR code and can easily be installed on a smart phone and/or mobile device. Once the app is installed, the user opens app and point the camera at the centre of the QR code, which then reads it and automatically takes it to the website, if the QR code contains a website URL.

A number of free apps are given as under:

A) Apps for Android
The apps below are available on the Android's built in Google Play App Store and can be downloaded for free.
1. QR Droid (By Droidla)
2. Scan (By Scan.Inc)

B) Apps for iPhone
These apps are available on App store and can be downloaded for free.
3. QR Reader (By TapMedia Ltd)
4. Bakodo-Barcode Scanner and QR code (By Bakodo)

C) Blackberry
This free app can be downloaded by the blackberry users.
5. QR Code Scanner Pro (Jared Company)
More information on QR codes and its uses can be found on this link and papers.
http://en.wikipedia.org/wiki/QR_code

Diazgranados, M. & V. A. Funk, (2013) Utility of QR codes in biological collections. PhytoKeys: 21-34.

Hoy, M. B., (2011) An introduction to QR Codes: linking libraries and mobile patrons. Med Ref Serv Q 30: 295-300.