Version: Final

EGT2
ENGINEERING TRIPOS PART IIA

Friday 24 April 2015 9.30 to 11.00

Module 3G1

INTRODUCTION TO MOLECULAR BIOENGINEERING

*Answer not more than **three** questions.*

All questions carry the same number of marks.

*The **approximate** percentage of marks allocated to each part of a question is indicated in the right margin.*

Write your candidate number not your name on the coversheet.

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed

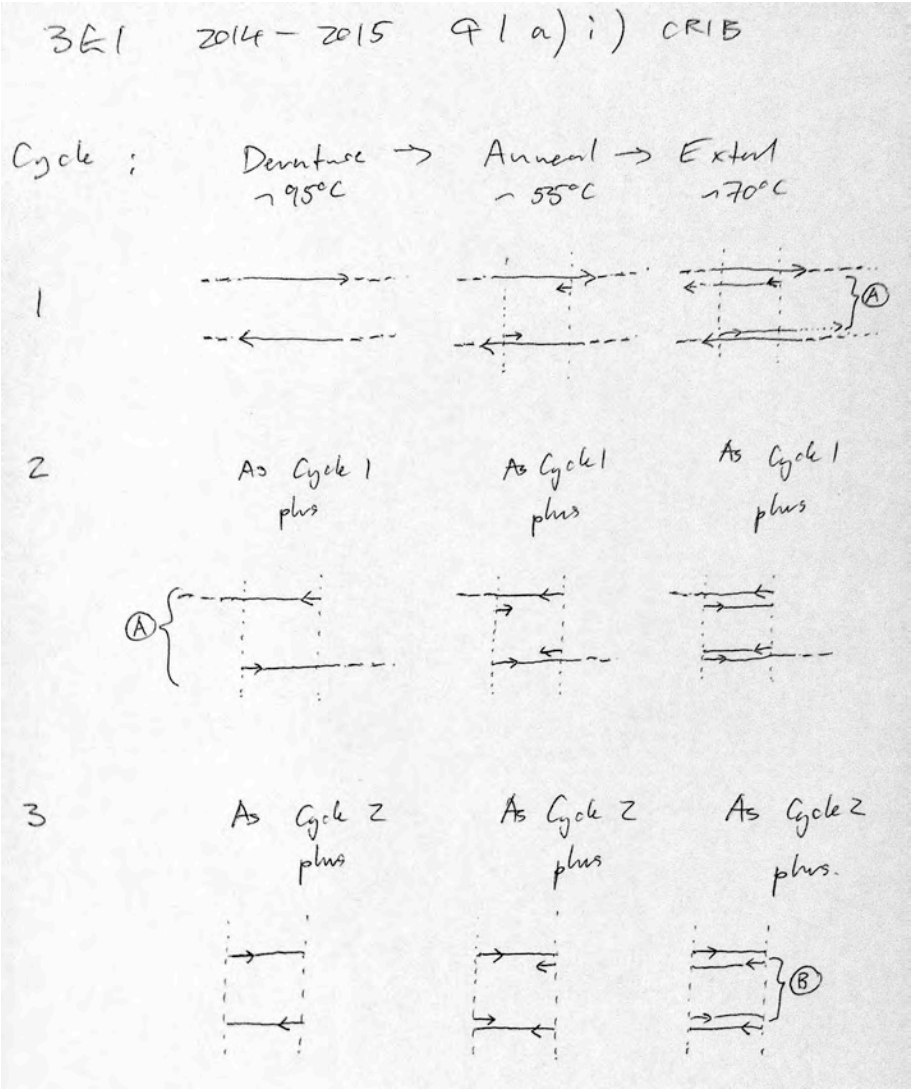
10 Minutes reading time is allowed for this paper

You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.

Version: Final

1 (a) (i) Explain how the polymerase chain reaction (PCR) works, by means of diagrams that show the first three cycles. [30%]

The PCR reaction is composed of template DNA, excess synthetic DNA PCR primers that are complementary to the sequence flanking the region of interest and prime DNA synthesis towards each other, dNTPs, and a thermostable DNA polymerase (e.g. Taq polymerase) all in an appropriate buffer. The reaction proceeds in a thermocycler which alternately denatures the DNA (~95C), allows annealing of the primers (~55C) and allows extension of the primers (~70C). Depending on the application, typically 20-35 cycles of denaturation, annealing and extension are carried out.



(ii) Assuming a typical PCR reaction with 30 cycles, comment on the amplification rates of the different classes of molecules described in (i). [20%]

Cycle one products (A) grow linearly in abundance. The strands generated in cycles 2, 3 that are bounded by the primers (B) grow exponentially in abundance.

(b) Different strains of the same species of plant can produce flowers that are white, pink or red, depending on whether they carry 0, 1 or 2 functional copies respectively of a gene that encodes a pigment-generating enzyme.

PCR primers are designed that can amplify this gene, generating a product of 1000 bases in size. When these primers are used to amplify DNA from a white-flowering strain, a longer product of 1300 bases is observed.

(i) Give a possible molecular explanation for the lack of pigmentation in the white-flowering strain. [10%]

Insertion of some DNA into the enzyme's gene has disrupted its ability to code for the enzyme. The insertion could introduce a premature stop codon, and/or introduce a stretch of amino-acids into the enzyme that are incompatible with its function, or the promoter could be disrupted.

(ii) A pink-flowering and a white-flowering strain are crossed. DNA is prepared from their progeny, and PCR reactions carried out with the above PCR primers. Outline a typical technique used to assess the size of the PCR products, and explain the results expected for white, pink and red flowers. [20%]

Agarose gel electrophoresis of DNA is a molecular sieving technique in which molecules are separated by size when made to migrate through an agarose gel by an electric field. The negatively charged DNA molecules migrate away from the cathode and toward the anode, with smaller molecules moving more rapidly. Samples are loaded after mixing them with a dense loading dye, and DNA is visualised by means of a dye such as ethidium bromide that is included in the gel, and which fluoresces when it is bound to DNA and excited by UV light. A lane of molecular weight markers of known mass are included to allow test sample product sizes to be estimated.

Results:

Red-flowering plant: single 1000 base fragment

White: single 1300 base fragment

Pink: 1000 and 1300 base fragments in equal ratio

(iii) What is the expected ratio of white, pink and red flowers among the progeny and why? [20%]

All progeny inherit a dysfunctional copy of the gene (1300 base size) from the white parent. From the pink parent, half of the progeny inherit a dysfunctional copy of the gene and half a functional copy (1000 base size). Therefore half the progeny have two copies of the dysfunctional gene and are white, and half have one functional gene copy and are pink. None of the progeny will have two functional copies of the gene and therefore none will be red. The ratios are thus 1:1:0 White: pink : red.

An explanation based on a Punnett square is also fine.

Q1. Examiner's Comment:

A very popular question. Across all the candidates it was possible to find very good answers to all parts of the question. However:
- 1a(i) was often answered poorly with candidates failing to indicate clearly the different molecular species produced.
- 1a(ii) was almost never answered fully correctly.
- 1b(1) was answered correctly by few candidates.

2 Antibodies are key molecules in the adaptive immune system.

(a) Sketch the structure of the basic antibody molecule, labelling the key parts. [10%]

Draw a Y-shaped molecule with two tips labeled as antigen binding sites, and the base labeled as the adaptor for immune functions. The two heavy chains and two light chains should be marked.

(b) Briefly explain how it is that the antibody molecule diversity observed in an individual is far higher than the number of protein-coding genes in that individual's genome. [30%]

Both the heavy and light chain genes are assembled from a combinatorial set of parts (V,D,J for heavy; V,J for light). In addition there are two different light chain genes, and further diversity is generated when the parts are fused. Directed mutagenesis further increases diversity in the antigen-binding region. All of the above contribute to allowing an antibody diversity of ~10^{11}, far greater than ~20,000 protein-coding genes directly encoded by the genome. Any given antibody-producing cell only contains one gene rearrangement.

(c) Scientists have made gene fusions between antibody genes and bacteriophage coat protein genes, which result in the antibody fusion protein being expressed on the surface of the bacteriophage. By doing this on a large scale, with many different antibody genes, one can generate a high diversity library. This allows screening to be carried out, for instance by immobilising some target antigen of interest on beads, allowing the bacteriophage library to bind the beads, and then washing away unbound/weakly bound bacteriophage. Tightly bound bacteriophage can then be eluted and used to infect bacteria, allowing generation of further bacteriophage with the same binding specificities.

(i) Outline the steps needed to isolate the antibody genes present in a rat in order to make the above library. [10%]

Take blood from the rat, isolate RNA, make cDNA, PCR the antibody coding regions using antibody gene specific primers.

(ii) What key issue would you consider when designing the fusion between the antibody gene and the coat protein gene? [10%]

Ensure that the gene fusion is made such that the open reading frame is continuous between the two fused genes.

(iii) Briefly outline the approach you might take to fuse the material generated in (i), which contains many antibody genes, with bacteriophage vector molecules, in order to generate an antibody fusion library. [20%]

Explanations based on restriction enzymes, fusion PCR or Gibson Assembly are fine e.g. The PCR primers used to amplify the antibody coding sequences could be terminated in restriction sites that are compatible with the gene fusion required with the bacteriophage vector. Following PCR and digestion of both the bacteriophage and PCR products with the appropriate restriction enzymes, the two could be ligated together with T4 DNA ligase, and transformed into host bacteria, which would produce the antibody fusion bacteriophage.

(iv) The resulting library is screened with a bead-immobilised antigen and the tightly bound bacteriophage particles eluted and used to infect further host bacteria in order to produce more bacteriophage. The screening is repeated several more times, each time using the bacteriophage isolated from the previous cycle of screening. Outline the steps you would take to characterise the diversity present in the final "tight binding" population of bacteriophage. Based on the diversity observed, comment on how you might identify the tightest binding bacteriophage in the population. [20%]

Isolate bacteriophage DNA, then PCR (the variable parts of) the antibody genes, then carry out DNA sequencing on the amplified material. Align the resulting sequences to organise them into related groups (they may not be identical due to sequencing errors). Count the number of members of each group. The most abundant group is likely the tightest binding.

Q2. Examiner's Comment:

A popular question. The question was generally well answered and virtually everyone that answered had a very good knowledge of antibody structure and how natural diversity is generated. This probably reflects the emphasis on this in lecture and the supervision on the topic. The molecular cloning and library analysis was less well answered. Asking the students to use their knowledge from other parts of the lecture material is clearly challenging, as the students' knowledge of DNA sequencing and analysis is very good, yet they did not transfer this knowledge as well as hoped.

3 Assume you are a metabolic engineer working on the pathway shown in Figure 1 (in which reactions (arrows) between intermediate metabolites named A-F are catalysed by enzymes named E_x).

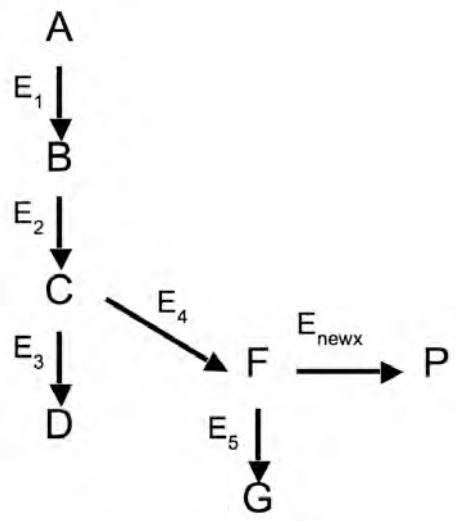


Figure 1.

You are trying to maximize production of the product, P. Enzymes E_1 through E_5 are part of the endogenous pathway native to your host organism, but the organism is not capable of producing P without the introduction of one of several possible foreign enzymes, E_{newx}. You have a collection of DNA parts at your disposal (Table 1) including E_1-E_5, three versions of E_{newx} from different organisms with different Michaelis constants (Km, Table 2) but the same V_{max}, and a collection of promoters including three constitutive promoters with different expression levels (P_{high}, P_{med}, P_{low}) and a repressible promoter (P_{rep}) that can be repressed by addition of the chemical REP. You also have access to the transcription factor TF1 that will up-regulate the expression of E_1-E_5.

(a) You plan to introduce E_{newx} by transforming your host organism with a plasmid built from your collection of parts. Draw that plasmid and explain why you chose the parts that you did. [30%]

The plasmid should contain a promoter (any one can be correct as long as choice is sensibly justified) followed by RBS then E_{new3} then terminator. E_{new3} should be used because its Km is lower than E_5 meaning flux will flow to P rather than G.

(b) Explain how you would go about designing a set of plasmids to maximise the production of P. You can make and experiment with as many plasmids as you like, but in your final system you must use the E_{newx} plasmid designed in (a) plus two additional plasmids, each carrying one gene. [50%]

First, it is likely that up regulating the entire pathway will help greatly so one plasmid should contain TF1 driven by P_{high}. We now have one plasmid left and that should up regulate the enzyme with the most control over the pathway in its current state (with TF1 and E_{new3} overexpressed). To determine this, make a set of plasmids with E_1, E_2, and E_4 separately driven by each of the set of constitutive promoters or with P_{rep}. Measure the flux through the pathway at different expression levels of each enzyme and the one with the greatest increase in flux should be used in the final system.

(c) The host organism is capable of homologous recombination meaning you can replace genomic DNA with parts from your collection. What change(s) might you make to the genome to further maximize production of P? [20%]

E_3 and E_5 are driving flux away from the desired product, so down regulating them will increase production of P. You could replace their promoters with P_{rep} and add REP during growth in order to decrease their expression.

Q3. Examiner's Comment:

A very popular question. The quality of the answers varied widely but a few students did come close to achieving full marks.

- 3a was answered correctly by a number of students with most students choosing the correct enzyme but many describing an incorrect gene structure.
- 3b was almost never answered fully correctly but there were a few answers that were close.
- 3c was answered well by some but very poorly by others.

Part	Function
E_1	Protein coding
E_2	Protein coding
E_3	Protein coding
E_4	Protein coding
E_5	Protein coding
E_{new1}	Protein coding
E_{new2}	Protein coding
E_{new3}	Protein coding
TF1	Protein coding
P_{high}	Promoter
P_{med}	Promoter
P_{low}	Promoter
P_{rep}	Promoter
Ribosome binding site (RBS)	Translation initiation
Terminator	Transcription termination

Table 1: parts

Enzyme	Km (for substrate F, mM)
E_5	20
E_{new1}	60
E_{new2}	30
E_{new3}	10

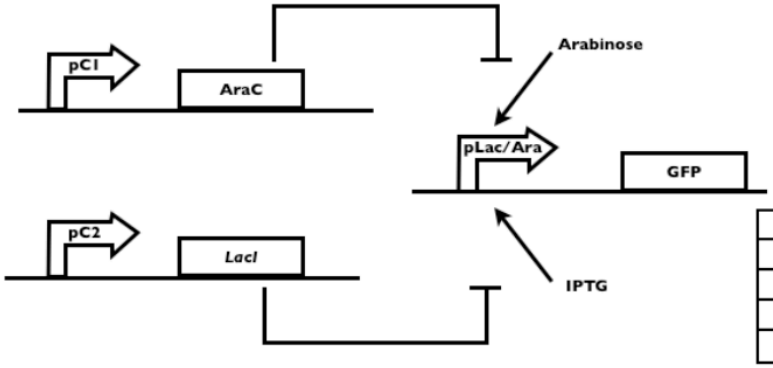
Table 2: Michaelis constants

4 Combinations of inducible repressors and activators can be used to control transcription and build biological logic gates. Table 3 is a registry of biological parts. Note that all promoter parts have ribosome binding sites (RBS) associated with them. In your answers to the following questions you should choose parts from Table 3.

(a) (i) Draw the designs for two different AND gates and the corresponding truth tables. You should minimise the use of the same parts in each of the two designs. [40%]

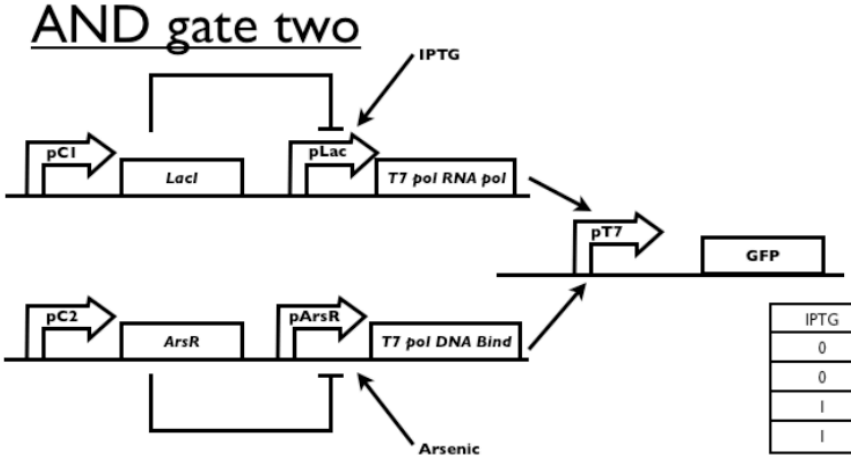
(ii) Describe briefly how each of the designs from (i) work. [30%]

AND gate one



IPTG	Arabinose	GFP
0	0	0
0	1	0
1	0	0
1	1	1

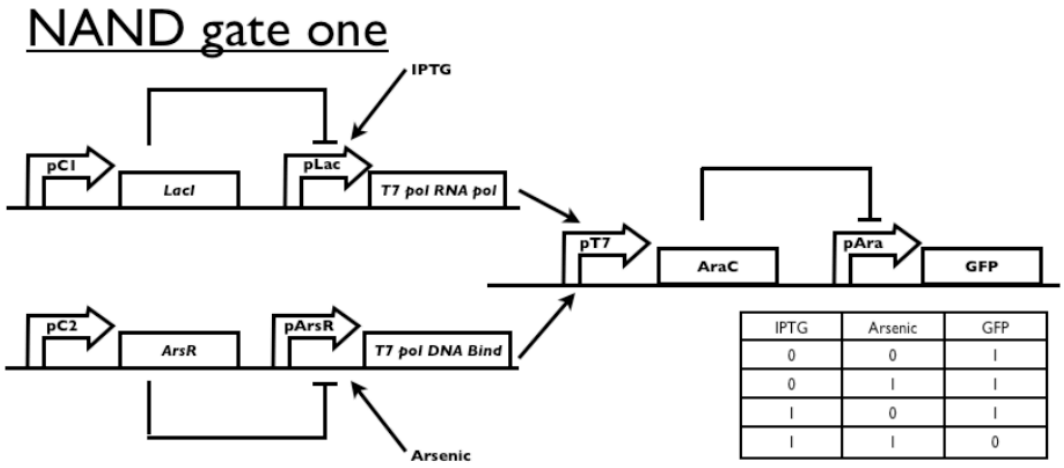
AraC and LacI will be constitutively expressed and repress GFP expression at the pLac/Ara hybrid promoter. In the presence of IPTG **AND** Arabinose, the promoter will be both de-repressed and activated respectively, Allowing GFP to be produced.



IPTG	Arsenic	GFP
0	0	0
0	1	0
1	0	0
1	1	1

ArsR and LacI will be constitutively expressed and repress T7 polymerase RNA polymerase fragment and T7 polymerase DNA binding fragment expression at pArsR and pLac respectively. GFP will only be expressed when both the T7 polymerase fragements are expressed at the same time, i.e. in the presence of Arsenic **AND** IPTG simultaneously.

(b) Modify one of the designs from (a) to make a NAND gate, give the corresponding truth table and explain the modification. [30%]



IPTG	Arsenic	GFP
0	0	1
0	1	1
1	0	1
1	1	0

ArsR and LacI will be constitutively expressed and repress T7 polymerase RNA polymerase fragment and T7 polymerase DNA binding fragment expression at pArsR and pLac respectively. AraC will only be expressed when both the T7 polymerase fragements are expressed at the same time, i.e. in the presence of Arsenic IPTG simultaneously. GFP will be constitutively expressed until both IPTG and Arsenic are present.

Table 3: parts to be used in Question 4 design exercises.

Part	Function	Description
pArsR	Arsenic promoter	Sequence covers RNA polymerase binding site.
pLac	*Lac* promoter	Two LacI binding sites on either side of RNA polymerase binding site, allowing a loop to form preventing RNA polymerase to bind in the absence of IPTG.
pAra	Arabinose promoter	In the absence of arabinose, inhibits transcription by blocking RNA polymerase. In the presence of arabinose, promotes transcription by stabilising RNA polymerase binding.
pLac/Ara	Hybrid *Lac*/*Ara* promoter	The two adjacent upstream sites for AraC remain and the downstream site is eliminated so AraC can only act as an activator. The two sites are in between two LacI binding sites so that RNA polymerase can only bind when AraC is bound and LacI is not bound.
pC1	constitutive promoter	Binds RNA polymerase with high efficiency.
pC2	constitutive promoter	Binds RNA polymerase with high efficiency.
pT7	T7 polymerase promoter	Highly specific for T7 RNA polymerase binding and does not bind cellular RNA polymerase.

Q4. Examiner's Comment:

This was a less popular question, but with a large spread of marks. Three of the candidates did very well, two very poorly and one moderately well. The candidates that got full or nearly full marks clearly understood how genetic parts could be built into AND and NAND gates. One candidate understood the process but made a consistent error in one of the promoters that would change the outcome.

(TURN OVER FOR CONTINUATION OF TABLE 3

Part	Function	Description
LacI	Lac repressor	Protein tetramer forms DNA loop to prevent RNA polymerase binding. When IPTG is bound to the repressor protein, the loop is opened, allowing RNA polymerase to bind.
AraC	Arabinose repressor/activator	Protein dimer forms DNA loop to prevent RNA polymerase binding. When arabinose is bound to repressor protein, the dimer changes conformation and stabilises RNA polymerase to behave as a transcriptional activator.
ArsR	Arsenic repressor	Protein dimer is specific for pArsR and covers RNA polymerase binding site, preventing RNA polymerase binding. When arsenic is bound the repressor is released from the DNA allowing RNA polymerase to bind.
T7 polymerase DNA binding fragment	This protein fragment binds T7 promoter	In conjunction with *T7 polymerase RNA polymerase activity fragment*, will form a complete protein and transcribe from the T7 promoter.
T7 polymerase RNA polymerase activity fragment	This protein fragment has RNA polymerase activity	In conjunction with *T7 polymerase DNA binding fragment*, will form a complete protein and transcribe from the T7 promoter.
GFP	fluorescent reporter protein	The protein fluoresces green under UV light

END OF PAPER

THIS PAGE IS BLANK