Version JWA/2

EGT2
ENGINEERING TRIPOS PART IIA

Tuesday 26 April 2016 14.00 to 15.30

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 cover sheet.*

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.

1 You have been given an agar plate with a very small bacterial colony, composed of only a few cells. These bacteria have been transformed with a plasmid vector containing the gene coding for the green fluorescent protein (GFP), flanked by two restriction sites, *Hin*dIII and *Bam*HI.

(a) What are restriction enzymes, what is their original function in bacteria, and why are they useful in modern bioengineering? [10%]

Enzymes, isolated from bacteria, which cut dsDNA at specific sequences, often palindromic. The sequence is specific for each type of RE. Originally part of the bacterial defense response against invading viruses (bacteriophages). The target sequence in the bacterial genome is usually protected by methylation. Useful to cut pieces of DNA from different sources which can be later ligated together to create recombinant DNA molecules.

(b) You are asked to carry out a DNA digestion using these two restriction enzymes (*Hin*dIII and *Bam*HI) to isolate the gene encoding GFP, to be used in subsequent experiments. Knowing that the small colony on your agar plate will not yield enough DNA to carry out the restriction digest, how could you obtain a sufficient amount of DNA to carry out the restriction digest? [10%]

Need a lot of DNA for a restriction digestion, so need to amplify the DNA: grow bacteria in the agar plate or in liquid culture until obtain enough cells to extract the DNA.

(c) You are also interested to study the cellular localisation of a special type of γ-tubulin from the fruit fly *Drosophila melanogaster*, and decide to make a chimeric DNA molecule by joining your target gene (γ-tubulin) with the reporter gene (the GFP) to make a single open reading frame. First you decide to perform a polymerase chain reaction to amplify the target gene (γ-tubulin), which is 2 kb long. Describe the reagents you should mix in your tube and what happens during each step in the thermo-cycler (the machine performing the amplification reactions). [30%]

To perform PCR: buffer solution with target DNA (few copies from the genome of the fruit fly), DNA polymerase, synthetic oligos (primers), dNTPs, cations.

Steps: cycles of denaturation, usually at 94˚C (the complementary DNA strands are dissociated), annealing, temperature is dependent on the primer sequences (the primers anneal on the original DNA in a position corresponding to the

complementary sequence), extension, usually 72˚C (the DNA polymerase polymerises (synthetises) a complementary strand from the primers along the original strand).

(d) Your supervisor then informs you that he wants to use a collaborator's strain that already expresses GFP for the study, so you have to PCR-amplify another reporter gene, a red fluorescent protein (RFP) to tag γ-tubulin. You then decide to use isothermal Gibson assembly to join your gene of interest (γ-tubulin) and the new reporter gene (RFP) in this order (γ-tubulin - RFP) to a plasmid. Describe the necessary PCR reactions, showing diagrams for each (include primers, DNA template and post-PCR fragments). [30%]

The left fragment (γ-tubulin) and the right fragment (RFP) have been amplified by PCR, and you know that part of the forward primer used to amplify the RFP contains the same sequence as the 3' end of the PCR product coding for γ-tubulin.

Gibson assembly: put the two PCR products in a tube with 3 enzymes and use one single temperature: the 5' exonuclease partially digests the 5' ends of all strands, so the DNA fragments can anneal due to the complementarity from the primers used in the previous steps, then the DNA polymerase elongates the strands, and the DNA ligase closes the nicks.

You have to pay attention to the right end of the PCR product coding for γ-tubulin because you need this region to anneal to the left end of the PCR product coding for RFP. In this way, after the partial digestion by the exonuclease, the two PCR products can act as giant primers on each other.

(e) You now want to carry out the isothermal Gibson assembly to join up the γ-tubulin gene, the RFP gene and plasmid backbone. Briefly describe which reagents you have to mix, the overall protocol, what happens during the reaction, and why your primers had to be designed the way they did for the reaction to work. [20%]

Cut the cloning vector and the foreign DNA with compatible restriction enzymes (should have considered this when amplifying the fragments with PCR, but ok if not mentioned)

- isolate fragments of interest via gel electrophoresis
- ligate foreign DNA into the vector with DNA ligase (or by Gibson assembly)
- To check the sequence:
- Transform the resulting mixture (the recombinant DNA) into host cells (*E. coli*)

- Select for growth: select the cells which have taken up the plasmid (select the transformed bacteria)
- Screen for clones containing the correct construct
- Sequence the vector with Sanger sequencing (not next-gen sequencing)

2 Lactate can be used as a precursor for many useful molecules. Yeast does not naturally produce lactate but can be made to do so by introducing a highly expressed lactate dehydrogenase (LDH) with very high Vmax (the activity of an enzyme at the maximum substrate concentration) and deleting the genes encoding GPD1 and PDC1 that produce glycerol and ethanol, respectively (Figure 1). NDE1 is a cytosolic NADH dehydrogenase that oxidizes NADH, functioning alongside glycolysis and fermentation to provide additional redox balance.

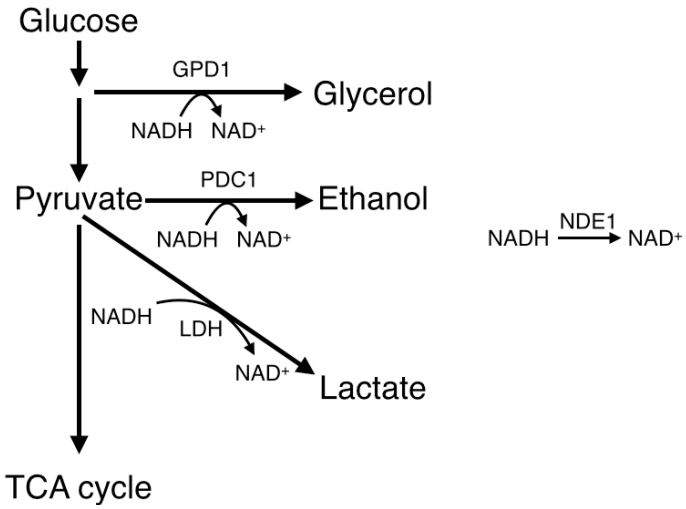


Figure 1.

(a) Why do organisms produce lactate and/or ethanol from pyruvate and what conditions should you grow your yeast in to make sure lactate is produced? [20%]

Ethanol and/or lactate are produced in order to oxidize NADH produced during glycolysis to maintain redox balance. You should grow your yeast in anaerobic conditions to prevent pyruvate from entering the TCA cycle which functions in the presence of oxygen.

(b) Why would deleting the genes encoding GPD1 and PDC1 improve lactate production? [10%]

These enzymes causes flux to be diverted toward glycerol and ethanol instead of lactate. Deleting genes for their production prevents this.

(c) The strategy of simply introducing a gene encoding LDH and deleting the genes encoding GPD1 and PDC1 only produced a small amount of lactate and led to unhealthy yeast. What is a likely cause of this problem and how might you fix it?

Assume that the activity of LDH is much higher than the activity of GPD1 or PDC1 would be if they were present. [30%]

By increasing flux toward lactate, you have also increased the oxidation of NADH, creating a redox imbalance in the cell. To fix this, you could mutate Nde1, lowering or abolishing its activity to rebalance the ratio of NADH to NAD+.

(c) What are some other strategies you might use to increase the production of lactate? [40%]

You could upregulate enzymes involved in glycolysis by expressing an extra copy of a transcription factor that increases their expression on a plasmid. You could directly express extra copies of those enzymes on a plasmid or under a highly expressed promoter in the genome. You could use molecular scaffolds to increase the local concentration of intermediate compounds and optimize stoichiometry. You could attempt to engineer a version of LDH with a higher Vmax or lower Km through mutagenesis and screening. You could use MAGE to screen many variants of the promoters of the enzymes involved in glycolysis to optimize flux through the pathway.

3 The five-gene violacein operon (*vioA*-*vioE*) from *Chromobacterium violaceum* (a gram negative bacterium) is constitutively expressed under its native $_{P}$vio promoter. Expression of the operon results in the production of a purple pigment, visible to the naked eye with an absorption wavelength of 575 nm. The violacein production pathway is shown in Figure 2.

This entire sequence (shown in Figure 3), was directly inserted into the *Bacillus subtilis* (a gram positive bacterium) genome at the *amyE* locus, a well documented site for gene integration. Although molecular analysis confirmed that the sequence had been integrated successfully with no mutations, there was no violacein pigment expressed.

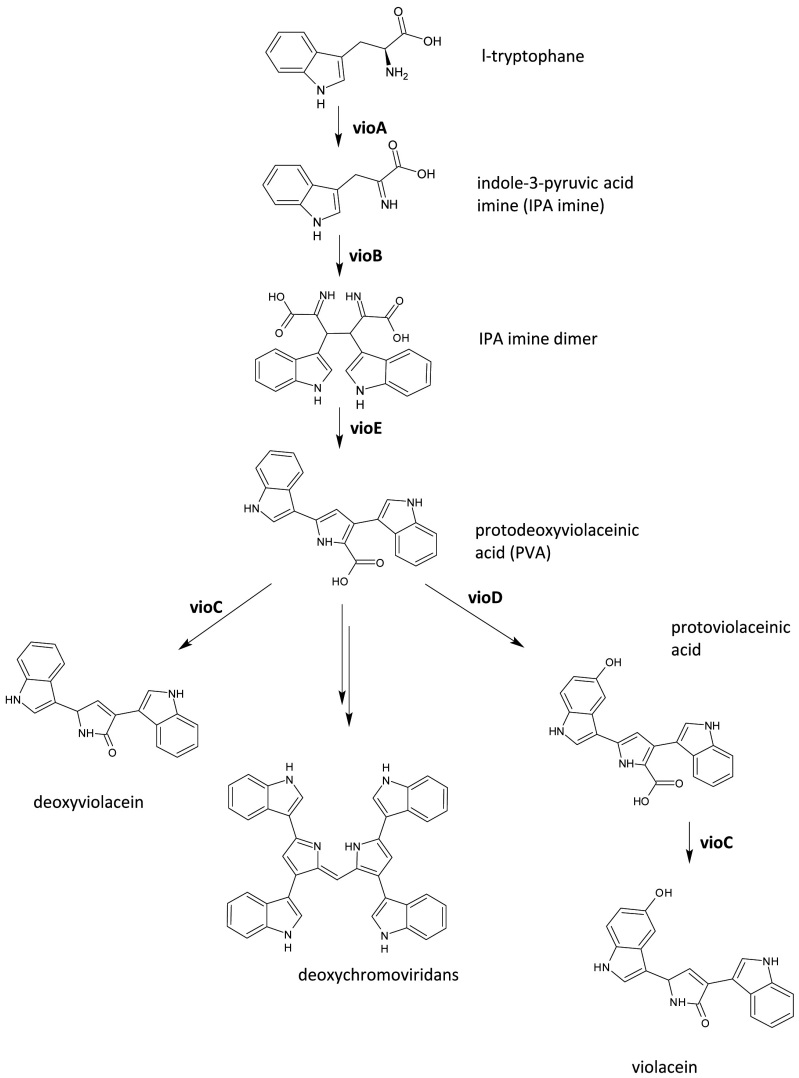


Figure 2.

(cont.

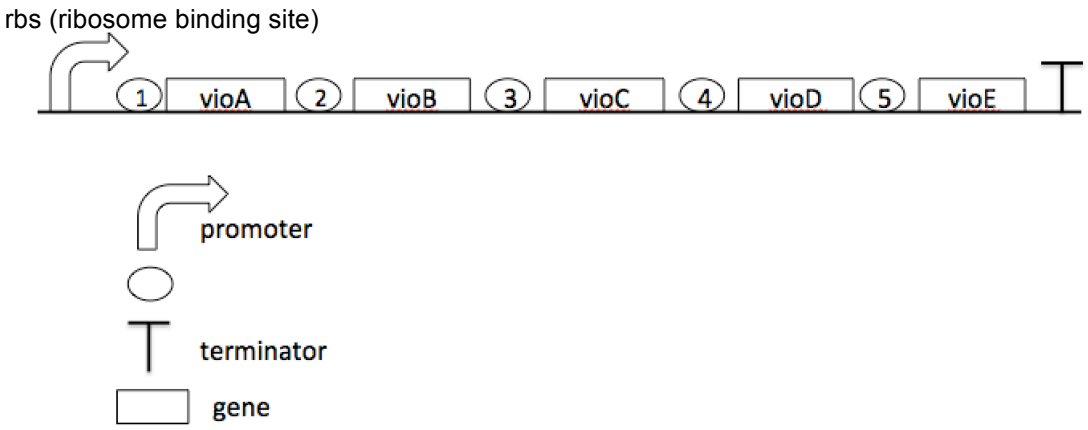


Figure 3.

(a) Describe what mechanisms of gene expression may have contributed to the failure of pigment production in the modified *Bacillus* strain. [15%]

Incompatibility/inefficiency of genetic elements from *C. violaceum* for expression in *B. subtilis*:
Weak promoter: inefficient transcription.
Weak RBS: inefficient translation.
Codon usage: inefficient translation.
Differences GC content may also have an effect.

(b) How can we manipulate these mechanisms to fix this problem? [20%]

To improve the efficiency of transcription and translation, the corresponding sequences can be modified by mutation/cloning to sequences that are known to function well in *B. subtilis*.
The promoter could be changed to that of a strong constitutive *B. subtilis* promoter, or a compatible inducible promoter.
The RBS sequences could be changed to strong *B. subtilis* RBS sequences.
The codon usage of ORFs could be modified to reflect codon usage of highly expressed gene of *B. subtilis*.

(c) After optimizing transcription of the operon, you find that you get very little pigment production. The *Bacillus* colonies grow very slowly and are generally smaller than they are when expressing your control constructs which express GFP.

(i) What could explain this phenomenon? [15%]

It is possible that one of the gene products of the violacein operon, or one of the metabolic intermediates produced by their activity, is toxic to *B. subtilis*.

(ii) How would you test to find what part of the operon is causing the problem? [20%]

Express every combination of the *vio* genes to determine whether specific combinations result in slow grow.

For a more efficient search, one could start by 1) expressing each *vio* gene alone, 2) using the known violacein metabolic pathway, express sets of *vio* genes to test whether one of the violacein metabolic intermediates is toxic (e.g. vioA, vioAB, vioABE, vioABEC, vioABED).

One could add purified forms of the metabolic intermediates to *B. subtilis* cultures to test for toxicity.

(iii) How would you fix the problem so that you get good growth and pigment production? [30%]

The growth defect (toxicity) would be expected to be dose-dependence with respect to levels of *vio* gene expression. Once the minimal set of *vio* genes causing toxicity has been determined, or a specific metabolic intermediate(s), the toxicity (does-response curves) of these factors could be determined by titrating their expression/concentration.

One could titrate the level of gene expression using either inducible promoters or multiple constructs with differing strengths of RBS/promoter sequences.

If the toxicity is due to a metabolic intermedate, then it may be possible to achieve high levels of violacein production by engineering high (relative) activity of metabolic steps downstream of the toxic intermediate (i.e. achieving metabolic high flux, but low concentration of the toxic intermediates).

If the toxicity is due to off-targets activities of one of the enzymes, then one could try to find/evolve/engineer a more specific enzyme.

4 Figure 4 shows a system you are building that has two fluorescent reporters being induced by two separate inducible promoters. Assume that every open reading frame has an rbs (ribosome binding site) preceding it. The genetic circuit described in Figure 4 was expected to yield a yellow output when both inducers are present at the same concentration (0.1 mM), but the output is orange. Table 1 is a list of genetic parts. Table 2 is a reference list of colours.

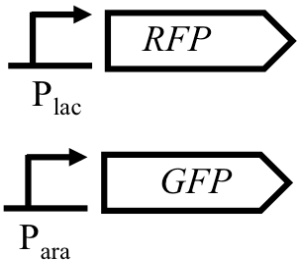
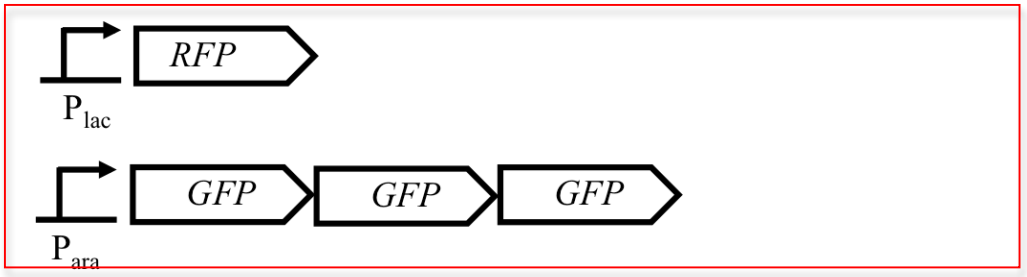


Figure 4.

(a) How would you most effectively fix this system using the parts provided in Tables 1 and 2? Explain why your solution to this problem is appropriate. [20%]

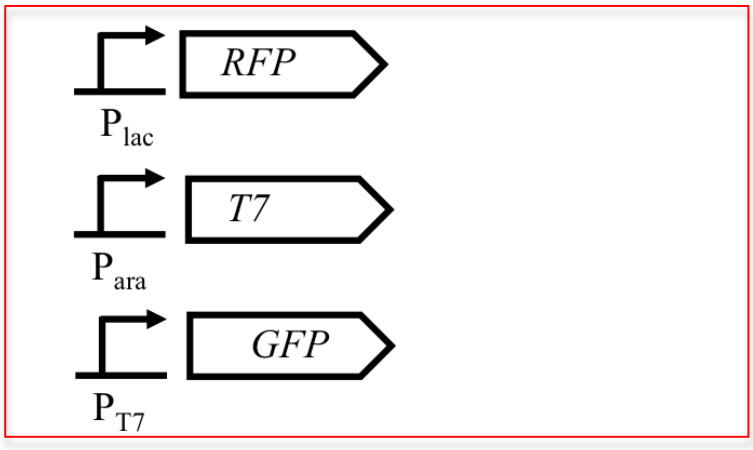
Essentially you need achieve triple the output of GFP or one third the output of RFP. Since there are no clear mechanisms to reduce RFP by that amount, the simplest solution is to triple the number of copies of GFP being produced by tripling the number of copies of the gene.



(b) You now want to use your system at different concentrations of inducer, but yield an orange output in the same way that the original system was providing. In this case, your inducers are present at 0.1 mM IPTG (as before) and 0.001 mM of Arabinose.

(i) Using the list of parts in Tables 1 and 2, how would you achieve this? Explain why your solution to this problem is appropriate. [40%]

As you now have a 100 fold reduction in GFP induction, it is necessary to amplify the output by 100x. In tables 1 and 2 the T7 system is said to provide that level of amplification. By placing the GFP under pT7 and T7 under the arabinose inducible promoter the system should perform as required.



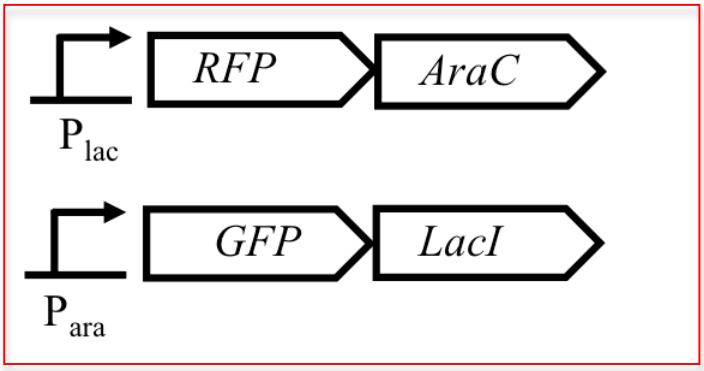
(c) You want to use your system to produce a single output of red or green when one inducer is present at 0.1 mM and the other inducer is absent. In its current configuration, the system is only giving brown or pink outputs.

(i) Why is the system not performing as it should? [10%]

(ii) How would you redesign the system to give you a single output using parts form Tables 1 and 2? Explain why your solution is appropriate. [30%]

(i) this is most likely due to leaky expression in the circuit as the repressor is not being actively produced as part of either system.

(ii) By having activity of one part of the system actively repress the other, equilibrium will shift and provide a binary output in the presence of either inducer (yielding red or green output respectively)



(cont.

Table 1.

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	This protein binds and activates the T7 promoter and works as an RNA polymerase	The protein binds to the T7 promoter (pT7) and acts as a highly active RNA polymerase, yielding 100x mRNA transcripts per T7 polymerase unit compared to native polymerases.
RFP	Fluorescent reporter protein	The protein fluoresces red under blue light
GFP	Fluorescent reporter protein	The protein fluoresces green under blue light
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.

(cont.

Table 2.

Visible Colour	**Ratio (% fluorescence units):**
Red	98% FU RFP 2% FU GFP
Pink	85% FU RFP 15% FU GFP
Orange	75% FU RFP 25% FU GFP
Yellow	50% FU RFP 50% FU GFP
Brown	25% FU RFP 75% FU GFP
Green	2% FU RFP 98% FU GFP

END OF PAPER