Version GM/3

EGT2
ENGINEERING TRIPOS PART IIA

---

2 May 2018 14.00 to 15.40

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 at the start of the exam.

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

1 The open reading frame from a human gene has been expressed in *E. coli* and the purified protein used to immunise a mouse. The resulting polyclonal antibody preparation, PAb-1, recognises different-sized proteins, derived from differentially spliced transcripts, in human tissue samples: in heart a 270-amino-acid form is present, while in liver there is both a 270-amino-acid form and a larger 300-amino-acid form. It is known that the N-terminal 100 amino acids and the C-terminal 120 amino acids are common to both the 270 and 300-amino-acid forms and each of these regions are derived from single exons.

(a) Explain the difference between polyclonal and monoclonal antibodies. [10%]

(b) PAb-1 is used to test brain tissue samples and in all cases a 220 amino-acid form is observed. What is the simplest explanation for this observation? [10%]

(c) Monoclonal antibodies, MAb-1, MAb-2 and MAb-3, are derived from the immunised mouse and tested against tissues as shown in Table 1 below. Explain, with the use of a diagram, possible mRNA splicing variations that could give rise to the observations in Table 1. [40%]

Table 1: Length in amino acids of proteins recognised by various antibody preparations. NR means nothing is recognised.

Antibody	Heart	Liver
PAb-1	270	270, 300
MAb-1	270	270
MAb-2	NR	300
MAb-3	270	270, 300

(d) Patients with a genetic liver disorder are tested with each of the antibodies and with the same results as given in Table 1, except that MAb-2 no longer recognises the 300-amino-acid form that it recognises in normal tissue samples.

Give a possible explanation for this observation. [20%]

(e) Outline what would be necessary in order to use the mouse monoclonal antibodies for human therapy and why this is important. [20%]

2 As part of an investigation, bacterial cells are broken open, and their contents extracted and purified to prepare a "cell-free" extract. These cellular contents are separated to remove the bacterial DNA, mRNA, cell membranes and the cell wall. Therefore the cell-free extract is composed of the remaining enzymes and tRNAs, which can be concentrated and remain fully functional. This mixture is not capable of sustaining transcription and translation until certain missing components, lost during the extraction process, are added as supplements.

(a) (i) What enzyme in the cell-free extract catalyses transcription? [5%]

(ii) In order to carry out transcription, what missing components must be added to the above cell-free extract and what roles do they play? [20%]

(b) Assuming you added the critical molecules for transcription, a messenger RNA (mRNA) in principle can be made.

(i) Along with the ribosomes, tRNAs and enzymes already present in the cell-free extract, what molecules must be added for translation also to occur, and what is their role? [15%]

(ii) What must be encoded on the mRNA for translation to take place? [15%]

(c) Bacterial cells can detect arsenic at concentrations as low as 10 parts per billion (ppb) using the arsenic (Ars) operon. The Ars operon is regulated by a simple repressor molecule (ArsR) that binds to the promotor region (pArsR) in order to prevent transcription. If the concentration of arsenic reaches 10 ppb then ArsR no longer binds to the promotor and the operon is transcribed.

The cell-free extract was used to make a biosensor in order to detect arsenic. Unfortunately ArsR is lost during purification of the cell-free extract.

Design a simple genetic circuit that will detect arsenic at levels similar to the Ars operon and explain how it works. The gene encoding Green Fluorescent Protein (GFP) should be used to report the presence of arsenic. [15%]

(d) When you test your ciruit, the signal-to-noise ratio isn't high enough to easily detect arsenic at 10 ppb. The noise is due to the random unbinding of the repressor, which generates a low but detectable background signal. Discuss approaches you could take to avoid this problem. [30%]

3 The schematic representation of a hypothetical metabolic pathway with two branches is given below. Substrate A is converted into products E and G. The uppercase letters denote metabolites and lowercase letters denote enzymes catalysing the adjacent reaction step. Route I denotes the pathway converting A to E, and Route II denotes the pathway converting A to G. ADP: Adenosine diphosphate, P: phosphate, ATP: adenosine triphosphate.

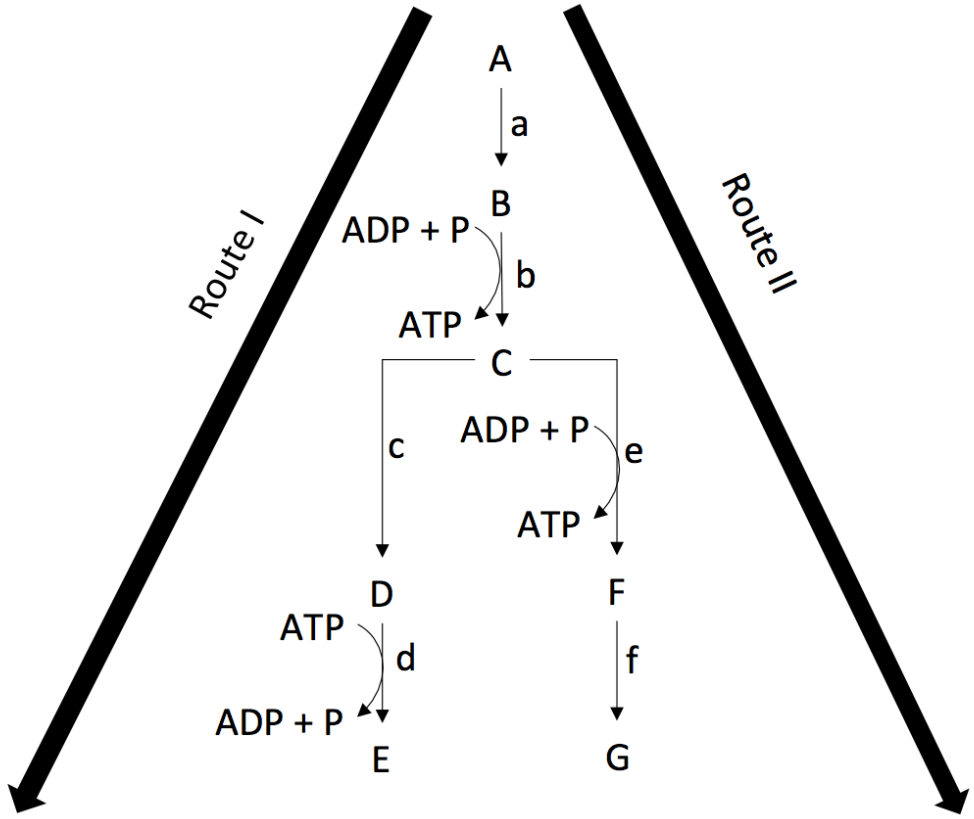

Fig. 1

(a) Are the reactions catalysed by the enzymes b, d, and e likely to represent catabolic or anabolic processes? Explain your answer. [10%]

(b) Explaining your answers, identify which metabolite in each pair below is expected to have a greater molecular weight:

B or C ;

D or E ;

C or F . [10%]

 (cont.

(c) What does it mean for a reaction to have a low saturation constant (K_m)? [10%]

(d) Suppose that metabolite G is the main precursor for a product, which is of commercial interest. You are given the task of improving the metabolic capability of this pathway for producing G (and consequently the product). The overall saturation coefficient (K_m) for Route I and Route II are provided in the table below (Table 2) under aerobic and anaerobic conditions. Assuming that these enzymes are inefficient, would you choose to operate this system under aerobic or anaerobic conditions? Explain your answer. [15%]

Table 2: K_m values

CONDITION	K_M FOR ROUTE I	K_M FOR ROUTE II
AEROBIC	0.06mM	0.3mM
ANAEROBIC	0.9mM	0.1mM

(e) The Response Coefficient (R) is defined as the measure of how the pathway flux (J) responds to an effector (P). Mathematically it is defined as:

$$R_p^{J_{ydh}} = \frac{\delta J_{ydh}}{\delta P} \cdot \frac{P}{J_{ydh}} = \frac{\delta ln J_{ydh}}{\delta ln P}$$

where *ydh* represents any pathway.

Identifying oxygen as an effector of this pathway, how do the pathway fluxes in Route I and II respond to varying oxygen levels? [20%]

(f) Suggest an experimental metabolic engineering approach, or a combination of approaches, to improve the selective production of the precursor metabolite G. [15%]

(g) You are now given the following information. Enzyme h from another organism is functionally equivalent to enzyme f in Fig. 1. Replacing the gene encoding enzyme f by a gene encoding enzyme h changes the overall K_m for Route II. The saturation coefficient for Route II becomes 0.2 mM under aerobic conditions and 0.5 mM under anaerobic conditions. Assuming that the K_m values for Route I remain unchanged, is this a feasible genetic modification for improving the selective production of the precursor metabolite G? Explain your answer, referring to Table 2 as required. [20%]

4 Plants produce 200 billion tons of non-food lignocellulosic (woody) biomass each year, with huge potential as a carbon-neutral fuel source. The microbes that live in the gut of the termite, the termite microbiome, are responsible for digesting the cellulose-rich wood that termites consume. Motivated by this, you extract DNA from the termite microbiome and generate a metagenomics dataset of billions of sequencing reads using the Illumina platform, with the hope of discovering the enzymes responsible for cellulose digestion.

(a) (i) Sequence analysis shows that the extracted DNA is dominated by sequences with high similarity to plants. What has happened and what simple modification to the experiment might avoid this problem? [10%]

(ii) Furthermore, you also find sequences with high similarity to humans within the dataset. What is the likely explanation for this observation and how might the experiment be repeated to avoid this problem? [10%]

(b) Having repeated the experiment and produced a DNA sequence dataset free of plant and human sequences, you carry out sequence assembly. This is done by comparing the sequences and combining those with extensive overlaps of near-identical sequence to form longer sequence *contigs*.

(i) Representatives of which kingdoms of life may reasonably be present in the termite gut microbiome? [15%]

(ii) Thus, when examining the assembled sequence contigs to find genes, what features would you expect the protein-coding genes to have? [20%]

(iii) Examining the length distribution of the sequence contigs you observe that longer contigs are rarer, shorter ones more abundant, and unassembled individual sequencing reads most abundant of all. Discuss reasons for these observations and approaches to increasing the fraction of long contigs. [30%]

(iv) An alternative approach to extracting and sequencing DNA from the termite gut microbiome would be to extract RNA, reverse transcribe it to DNA and sequence that. What possible advantages and disadvantages might this have compared to the previous approach? [15%]

END OF PAPER