Version SB/final

EGT2
ENGINEERING TRIPOS PART IIA

Friday 9 May 2025 9:30 to 11.10

Module 3G1

MOLECULAR BIOENGINEERING I

*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.

You may not remove any stationery from the Examination Room.

1 The textile industry is one of the most polluting in the world, leading to recent innovations towards sustainable, bio-based pigment production and dyeing processes. To increase the variety of pigments that can be produced, scientists are sequencing the genomes of more organisms, collecting metagenomic sequences from different environments and trawling existing databases of gene sequences to find new and improved pigment production pathways. Once discovered, these can be engineered for optimal expression of the pigment chemicals in scalable host expression systems.

(a) A new family of enzymes has been discovered in fungi that can produce many coloured pigment chemicals from a single precursor molecule.

(i) An enzyme that catalyzes the transformation of a blue chemical into a red chemical has evolved from an ancestral enzyme that originally produced a green chemical. Describe where you would expect the mutations to arise within the enzyme sequence and how these mutations could become fixed at high prevalence in a fungal population. [15%]

(ii) Could you use directed evolution to accelerate the production of a specific pigment color in bacteria, similar to how bacteria developed the ability to grow aerobically on citrate in the long-term Lenski experiment? [5%]

(b) You have identified 114 putative members of the new family of enzymes using bioinformatics techniques on metagenomic datasets, and you generate a DNA library for expression and testing. As 114 is a large number of novel enzymes to investigate, you consider incorporating automation into your experimental plan.

(i) How would you obtain the DNA for these enzyme variants? [5%]

(ii) Once you have the enzyme DNA, list the steps involved in designing and building your DNA library, then inserting it into a host *Escherichia coli* bacteria. For each step, comment on how easy it would be to automate the experimental techniques. [15%]

(iii) What are three advantages and three disadvantages of incorporating automation into the design-build-test-learn cycle for this experiment? [10%]

(c) In order to verify the sequences, you will sequence all the plasmids in your library using Sanger sequencing. For this, you will need to design DNA primers.

(i) What are primers and what is their function? [5%]

(ii) You receive the sequencing results back and realise that not all of the DNA

sequences match the original sequence from the bioinformatics analysis. Why might this be? [5%]

(iii) What types of mutations can arise in DNA that encodes proteins, and how might these mutations affect the protein sequence and function? [10%]

(d) You decide to express the enzymes combinatorially in a cell-free TX-TL system to prototype different pigment synthesis pathways, using linear polymerase chain reaction (PCR) amplicons as the input DNA.

(i) What approach would you take to assess the functionality of the enzymes in this assay? [5%]

(ii) Initially you do not get any output from your assay, but the cell-free TX-TL system is functional because you successfully express GFP from your plasmid DNA positive control. In your experimental assays, there are no bands of the expected size visible on SDS-PAGE and when you perform a reverse transcription quantitative PCR (RT-qPCR) there are no amplicons of the expected size. What is the most likely explanation and what are three ways you could fix this issue? [15%]

(e) After much research and development, your new chemical production pathway is ready to be scaled up and deployed on-site at textile dye houses as a whole-cell biomanufacturing approach. It is crucial for the process to be environmentally sustainable, economically feasible, and safe. Describe three challenges that you might expect when scaling up bioprocessing from lab scale (1-2 L) to production scale (>500 L). [10%]

2 A new fast-growing eukaryotic microorganism called *Cambridgiensis novus* has been discovered which has the potential to be used as an alternative to the current "workhorses" of recombinant protein expression. You are asked to try expressing a novel anti-HER2 humanized monoclonal antibody called zatuzumab in *Cambridgiensis novus*, because zatuzumab has been challenging to express in existing yeast expression systems. HER2-positive cells are present in a number of human cancers, so this antibody has potential to be used as a component in a number of therapies.

(a) What potential biosafety and biosecurity concerns could arise from the development or use of this new organism? [10%]

(b) What are the advantages of expressing proteins in the following organisms?

(i) *Escherichia coli* bacteria [10%]

(ii) *Pichia pastoris* yeast [10%]

(c) You have been provided with a protein expression plasmid for zatuzumab and asked to test it for constitutive expression in *Cambridgiensis novus*. Describe the standard protocol for using a plasmid to constitutively express recombinant protein in a host microorganism. [10%]

(d) You take samples of the *Cambridgiensis novus* and load them into an SDS-PAGE gel, but there is no detectable protein expression.

(i) What does SDS-PAGE allow you to measure, and how does it work? [10%]

(ii) One reason for lack of expression could be that the promoter in the plasmid is not compatible with the RNA polymerase or transcription factors in *Cambridgiensis novus*. Provide three other reasons why the plasmid you have been given might not lead to protein expression in this host. [10%]

(iii) In order to expand the genetic tools available to regulate protein expression in *Cambridgiensis novus*, describe an experiment to find and characterise novel promoters from this host. [10%]

(e) After many months of optimisation, you overcome expression challenges and discover that *Cambridgiensis novus* allows for efficient protein secretion, reducing the cost of purification. This allows you to produce enough zatuzumab to begin characterising its function.

(cont.

(i) What are the main features of humanised monoclonal antibodies? [10%]

(ii) Zatuzumab binds to and inhibits the HER-2 receptor. How could it be engineered to be more effective as an anti-cancer therapeutic? [10%]

(iii) What differences might be introduced by using a new expression host that might affect the therapeutic potential of zatuzumab? [10%]

3 For the iGEM project, your team has designed a synthetic oscillator that uses a dual feedback design (linked positive and negative feedback loops).

(a) Draw this circuit in **SBOL** (Synthetic Biology Open Language) notation to indicate how each genetic part connects. Label the core components (promoters, ribosome binding sites, coding sequences, terminators) and specify which loops provide positive regulation and which loops provide negative regulation. [15%]

(b) Briefly explain how you expect these two feedback loops to work together to generate sustained oscillations. [10%]

(c) The circuit is placed in a bacterial strain which undergoes binary fission to reproduce. If the carrier cell has a doubling time of 20 minutes, can growth-based dilution alone be used to achieve a synthetic oscillator with a period faster than 30 minutes? Provide a detailed explanation of why or why not. [10%]

(d) Suppose we want to *further shorten* the oscillator period to around 15 minutes. What fundamental design changes (for example, modifications to the circuit design) might be necessary to accomplish this faster oscillatory behaviour? [10%]

(e) Compare the resource demands (for ribosomes, RNA polymerases, and ATP) of the original oscillator (longer period) and the redesigned oscillator (shorter period). Which elements of the faster circuit design would likely change the metabolic burden imposed on the carrier cell, and why? [15%]

(f) Assume you have two bacterial strains: Strain A carrying the original oscillator and Strain B carrying the new, faster oscillator. You mix these two strains in a single culture to begin a growth experiment. How do you expect the ratio of Strain A to Strain B to change over time if they grow together, and why? Sketch a plot of this ratio over time, assuming unlimited resources. [10%]

(g) Describe how you would use a fluorescent protein reporter to quantify the fast oscillator's behavior in real time. Address the following points:

(i) **Design strategy:** How to adjust the circuit design to use the fluorescent reporter. [10%]

(ii) **Measurement strategy:** Time-lapse fluorescence microscopy or flow

(cont.

cytometry; frequency of data acquisition. [10%]

(h) You are comparing two fluorescent protein reporters: **RFP** (maturation time of ~40 minutes) and **YFP** (maturation time of ~4 minutes). Which reporter would be most suitable for tracking rapid oscillatory behavior, and why? [10%]

4 A recently analysed novel organism has 200 genes connected by 300 regulatory edges. A single point mutation in the organism's RNA polymerase (RNAP) was identified, which can double the promoter binding rate, potentially altering the network's behaviour. Answer the following questions about the implications and functioning of the network's topology and the effect of the mutated RNAP.

(a) You observe that 35 out of the 300 edges are self-arrows (autoregulation). Are these self-arrows more frequent than expected under a random model? Briefly explain the potential biological meaning of this frequency. [20%]

(b) Of the 35 self-arrows, 30 represent negative autoregulation. What roles and advantages might negative autoregulation provide, in terms of the organism's fitness or adaptation to environmental changes? [10%]

(c) Sketch a rate plot comparing a simple regulatory mechanism (no autoregulation) to negative autoregulation. Label the axes and indicate the steady state value for each. [15%]

(i) Briefly discuss how negative autoregulation alters the shape of the production rate curve compared to simple regulation. [5%]

(ii) Briefly discuss how these differences in production curves relate to gene expression stability. [10%]

(iii) Explain how negative autoregulation can shorten the response time of gene expression. Why might a faster response time be beneficial for this organism? [15%]

(d) A single point mutation in the organism's RNA polymerase doubles the promoter binding rate. If the wild-type polymerase is replaced by this mutated version, how would you expect the steady-state concentration of the 30 negatively autoregulated genes to change, and why? [10%]

(e) Researchers also identified 40 instances of a three-gene motif where: Gene 1 represses Gene 2, Gene 2 represses Gene 3, and Gene 3 represses Gene 1. What is the expected function (or typical behavior) of such a triple-repression motif in gene regulatory networks? Sketch the rate of production of a genetic reporter whose transcription is negatively regulated by Gene 3. [15%]

END OF PAPER