Version SB/final

EGT2
ENGINEERING TRIPOS PART IIA

Friday 3 May 2024 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 You have been provided with two plasmids by a collaborator. Plasmid 1 carries two copies of the promoter P1, with one promoting the expression of a green fluorescent protein (GFP) and the other promoting the expression of a red fluorescent protein (RFP). Plasmid 2 mirrors this arrangement but with two copies of the promoter P2. Subsequently, you have introduced these plasmids into *E. coli* bacteria, resulting in the creation of two distinct strains, Type 1 and Type 2, each carrying the corresponding plasmid.

(a) You imaged 100 cells of Type 1 and 100 cells of Type 2 using the green channel of the microscope. You found that the mean fluorescence intensity was 500 intensity units for both types of cells, while the standard deviation was 200 units for Type 1, and 50 units for Type 2.

(i) Calculate the coefficient of variation (CVs) for each promoter. [10%]

(ii) Would you expect similar results if you imaged the cells in the red fluorescence channel instead? Why? [5%]

(iii) Comment on the relative strengths of each promoter. [10%]

(b) In a second measurement, you imaged 10 individual lineages of the cell Type 1 for 20 generations in the green channel and constructed a distribution of intensities from the individual snapshots. What is the expected mean and standard deviation of this distribution? [10%]

(c) Next you wanted to dig deeper into the sources of variation in expression. You imaged the cells of Type 1 and Type 2 using both the red and green fluorescent channels of the microscope. For each channel, you computed the difference between the red and green channel intensities divided by their sum. We will refer to this quantity as η_1. You found that η_1 was 0.15 for Type 1 and 0.06 for Type 2. What is the biological interpretation of η_1 and why is it smaller for Type 2 compared to Type 1? Let's refer to the remaining variation, $CV^2 - \eta_1^2$, as η_2^2. What is the biological interpretation of η_2? [15%]

(d) The operator next to the promoter P1 has the lacO sequence. The operator next to the promoter P2 has the tetO sequence.

(i) If you move Plasmid 1 into a ΔlacI strain, how do you expect the values of mean expression and η_1 to change? [10%]

(ii) If you introduce a plasmid containing the repressilator, consisting of LacI, TetR, and λcI, how is the value of η_2 expected to change and why? [10%]

(e) Next, you inserted the entire dual reporter cassette for P1 into the chromosomal backbone of the wild-type *E. coli*.

(i) How do you expect the value of η_2 to change? [10%]

(ii) However, when your collaborator inserted it into the chromosomal backbone of the E. coli strain of their lab, they estimated much higher values for both η_1 and η_2. When troubleshooting, you realised that their strain is ΔRecA. From this observation, what can you infer about the function of RecA? [10%]

(f) If you were to introduce a copy of LacI next to the GFP sequence, how would that impact the estimated value of CV from measurements in the green channel? [10%]

2 Your iGEM team is currently working on developing a phage display system for a novel phage virus protein. This specific virus, in its original state, is a lysogenic one, meaning it infects bacterial cells and integrates its chromosomal DNA into the bacterial chromosome backbone. The viral genome is silently maintained and passed on through cell divisions. However, when the bacterial cell encounters a stress signal, denoted by S_0, the virus detects the threat. It then removes its genome from the bacterial chromosome, replicates the genome, and assembles its other components to generate 32 copies of itself. Subsequently, the virus causes the host cell to lyse, releasing the copies into the surrounding environment. This mode of replicating and lysing the host to release copies of itself is called the lytic mode of a virus.

(a) You identified a signalling system in the viral genome, which is involved in the recognition of the stress signal S_0. This signalling system includes a trans-membrane protein H_0 and a response regulator R_0. When you deleted the sequence of H_0 and R_0 from the viral genome, you found that the virus could no longer induce its lytic mode. However, compared to the previous version, this new virus was maintained in the bacterial chromosome for longer periods of time. Explain these observations. [15%]

(b) The viral genome has a sequence of another gene R_1 which together with R_0 constitutes a toggle switch module, such that the system is stably maintained in lysogenic mode in the absence of the signal S_0.

(i) Do you expect R_1 and R_0 to mutually activate each other's expression or repress each other? Why? [10%]

(ii) Why do you think the lytic-lysogenic switch uses a toggle switch module, instead of a simple positive feedback loop? [10%]

(c) If you culture the virus with its prey bacteria in the presence of the signal S_0, the virus stays in the lytic mode, where it keeps infecting new cells and lysing them to produce its copies for a subsequent round of infections.

(i) Assuming the virus takes 10 minutes from the time of infection to lysis, and produces 32 copies of itself, write down the equation of its growth, under the condition of unlimited prey cells. [10%]

(ii) What is the doubling time of this virus? [10%]

(d) The virus spends 5 minutes out of the 10 minute infection window to inject its genomic DNA into the cell. Therefore, it has 5 minutes to make copies of itself.

(i) If the viral DNA polymerase has a replication speed of 100 bp per second, then what is the maximum length of the viral genome? [10%]

(ii) The prey bacterial cell has a genome size of 3 mega base pairs. If the virus has to degrade the genome of its prey cell to generate the precursors for its own genome synthesis, what is the upper limit of the number of viruses produced from a single infection? Use the estimate from the previous question. [10%]

(e) You decided to study the infection process from individual viruses on individual bacterial cells using a fluorescent reporter. You have used a green fluorescent protein reporter next to a promoter of a viral capsid protein in its genomic sequence, such that every time the virus produces a capsid protein, it also produces a copy of the GFP. (Note: A capsid is the protein shell of a virus, enclosing its genetic material).

(i) When you imaged individual cells infected with the virus, you saw that most cells lysed within 10 minutes of infection, but less than 5% of the cells showed a detectable green fluorescence signal. How can you explain this observation? [10%]

(ii) Instead, you decided to fuse the GFP protein to the capsid protein, by placing the GFP gene sequence adjacent and in-frame with the capsid protein sequence. As expected, you found that most viruses were fluorescent in the green channel. However, you found that while the infection to lysis time of this new virus remained the same as the original virus, the probability of infection dropped. You also found that after a day of culturing this virus in the bioreactor described above, most of the viruses did not show green fluorescence. What are the most likely explanations for the two observations? [15%]

3 *Escherichia coli* bacteria are often used as a biofactory to produce protein- and peptide-based drugs, but this can prove challenging when the drug is toxic to cells. For example, certain anti-microbial peptides (AMPs) or certain cancer drugs inhibit core cell functions. One route to synthesise highly cytotoxic proteins is to use a cell-free TX-TL system, where transcription and translation take place in a cell lysate, or in a purified mixture of the cell components that are essential for TX-TL activity. However, these systems are much more expensive for large-scale production and typically have low yields compared to traditional cell-based fermentation.

(a) You decide to adopt an *E. coli* cell-free TX-TL system to trial the production of a highly toxic protein "Protein X" with a poorly-understood mode of action, as part of a drug discovery project.

(i) Describe the process of transcription and translation that will produce your protein from the encoding DNA and name the main components that need to be present for each reaction to proceed. [10%]

(ii) How does transcription differ in prokaryotic compared to eukaryotic cells? [5%]

(iii) How does transcription differ in cell-free compared to cellular systems? [5%]

(b) Your first step is to design a protein expression plasmid for Protein X for use in lysate-based cell-free TX-TL. You will synthesise the DNA *de novo* via a company, to save time with assembly and cloning.

(i) Draw your plasmid, labeling the major genetic elements, and briefly describe their function. [10%]

(ii) One advantage of cell-free TX-TL is that proteins can be expressed from a linear DNA expression construct. You decide to pursue this approach to reduce the time required to produce and purify a sufficient quantity of DNA for cell-free TX-TL expression. Draw a variation of your diagram in (b)(i) to illustrate the required parts of this construct. [10%]

(iii) Next, you use PCR to amplify your linear construct. Use a diagram to illustrate the different stages of a single cycle in a PCR reaction. Describe the reaction conditions and what is happening at the molecular level in each step. [10%]

(c) You perform a gel extraction to purify your linear DNA fragments for expression, add this at 20 nM final concentration into your lysate-based cell-free TX-TL, and incubate at 29 °C for 10 hours.

(i) What strategies and techniques (name at least three) could you use to assess the expression levels of your protein? [10%]

(ii) You do not observe any protein expression and your initial suspicion is that one of either transcription or translation is failing in your TX-TL system. How would you determine which one? [10%]

(iii) Upon scrutinizing the test results, you ascertain that transcription and translation are not the underlying issues. Please propose three alternative explanations for observing no expression and outline the troubleshooting steps you could take for each of these possibilities. [15%]

(d) You solve the issue and achieve successful expression but the observed yield is too low and the cost of producing enough of the cell-free TX-TL system is too high to produce therapeutically relevant doses. You decide to revert to attempting expression in *E. coli* cells. Describe three approaches that you could consider to express a highly toxic protein *in vivo*.

[15%]

4 You have been tasked with establishing an automation pipeline to develop new nanobody-based biologics for cancer therapeutics. Initial candidates will be developed by *in silico* protein design using artificial intelligence (AI). The pipeline should synthesise and screen nanobodies against a biomarker for bowel cancer and then feed the results back to inform further design cycles, with the intention to improve performance during each design-build-test-learn cycle. The pipeline should operate without requiring a "human in the loop", i.e. a bioengineer performing analyses or experiments. This form of automation is also known as a "self-driving lab".

(a) What are nanobodies and what are their advantages compared to monoclonal antibodies? [10%]

(b) You will need to screen a library of nanobodies generated by the AI nanobody design tool. What is genotype-phenotype linkage and why is it important in nanobody library screening? [10%]

(c) You decide to take a phage display approach.

(i) Describe the experimental steps required to go from a library of digital protein sequences produced by the design algorithm to a surface-displayed library on phage. [10%]

(ii) Describe the process of biopanning. [10%]

(iii) You have identified fifteen candidate nanobodies for high-affinity to bowel cancer biomarkers. How do you next scale up production for further characterisation studies? [5%]

(iv) You will further quantify the affinity of the candidate nanobodies to the target antigens. What other characteristics should you assess that would be critical for use of the nanobody as part of a human therapeutic? [10%]

(d) An easily-expressed nanobody with very high affinity to a surface biomarker of bowel cancer is identified at the end of the screening. How could this be incorporated into a bowel cancer therapeutic? [10%]

(e) Both AI and synthetic biology raise concerns about biosafety and biosecurity as well as ethical issues that might arise from their use or misuse.

(i) What broad categories of actors are involved in the governance of synthetic biology? [10%]

(ii) Is new biosafety and biosecurity governance needed for your new self-driving lab? Briefly explain your answer. [10%]

(iii) Describe three considerations beyond biosafety and biosecurity that would be considered a part of undertaking Responsible Research and Innovation. [15%]

END OF PAPER

THIS PAGE IS BLANK