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 (*CV*s) for each promoter. [10%]

Crib:

P1 has a *CV* of 0.4 and P2 has a *CV* of 0.1.

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

Crib:

Yes, as the controlling promoters are the same. If someone answers No, then their second answer should be - the fluorescent proteins have different photophysical properties, such as maturation times.

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

Crib: Since both constructs have the same mean intensity, but the construct with P2 has lower noise level, we expect P2 to be a stronger promoter than P1.

(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%]

Crib:

Since the process is expected to be a stationary process, the average and noise from measurements over time should be same as the noise across the population. Therefore the mean intensity should be 500 units ans the standard deviation should be 200 units.

(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%]

Crib: The value of η_1 quantifies to the intrinsic noise in the expression system. Since P2 is a stronger promoter, the intrinsic noise in the expression from P2 is lower. η_2 refers to the extrinsic noise in the system, which is typically dependent on upstream components.

(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%]

Crib:

Since the promoter P1 is repressed by lacI, moving the plasmid to a lacI- strain will cause the promoter to be derepressed. As a result the mean intensity is expected to go up and the value of η_1 is expected to drop.

(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%]

Crib:

Since the promoter P1 is repressed by lacI, an oscillating production of LacI is expected to introduce extrinsic variability to the expression, which is likely to increase the value of η_2.

(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%]

Crib:

A major source of extrinsic variability is the fluctuation in plasmid copy number. Copy numbers of chromosome is more tightly regulated. Therefore, when the circuit is transferred to the chromosomal backbone, we expect the upstream noise contribution to be reduced, which is expected to decrease the value of η_2.

(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%]

Crib:

RecA's must be related to the chromosome copy number maintenance.

(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%]

Crib:

Introducing LacI next the GFP sequence will introduce negative autoregulation to the system. This is expected to reduce the noise level in gene-expression. So the value of CV is expected to drop.

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%]

Crib:

The signalling system functions by expressing large copies of H_0 to be present in the host cell membrane for detecting S_0 and copies of R_0 to be present in the cell cytoplasm to trigger lytic mode. Removing these genes impairs the detection mechanism. On the other hand, deleting these genes cause the load of the genome to be lower, and therefore the system is maintained for longer. The wild-type cell is expected to select mutants of these genes, in the absence of S, causing the functional system to be lost over time.

(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%]

Crib:

Toggle switch uses mutual repression. So we expect the two genes R_0 and R_1 to repress each other.

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

Crib:

Positive feedback does not produce two stable states, as toggle switch does.

(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%]

Crib:

The equation of growth of number of viruses follows: $n_t = n_0 * 32^{t/10}$

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

Crib:

Since the virus multiplies to produce 32 copies every 10 minutes, which is equivalent to 5 rounds of doubling its number ($2^5 = 32$), the doubling time of the virus is = 10 minutes/5 = 2 mins.

(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%]

Crib: The maximum available time for the viral DNA polymerase to replicate the viral genome is 5 minutes (= 300 seconds). Therefore, the maximum length of genome it can replicate during that period is 30, 000 base pairs.

(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%]

Crib: Since the bacterial genome is 100x larger than the estimated size of viral genome, the virus can only create 100 copies of itself before the genomic DNTPs are exhausted.

(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

(cont.

detectable green fluorescence signal. How can you explain this observation? [10%]

Crib:

Most variant of green fluorescent proteins have a maturation time longer than 5 minutes. This suggests that from the time of their production and folding it takes on an average 5 minutes for GFP proteins to become fluorescent. Since the virus has only 5 minutes to replicate its genome and produce the capsid proteins, before it lyses the cell open, most GFP protein molecules are not expected to become and show any signal during this time.

(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%]

Crib:

Since the viral capsid protein is now fused with a heterologous protein, the assembly of the capsid can be impacted and cause misassembled viruses to form. Such viruses are not likely to be as efficient as the wild-type virus in infecting their prey cell.

This reduction of infection propensities should be under severe evolutionary pressure, so mutants or knockouts where the GFP production is impaired are likely to be selected in the culture.

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%]

Crib:

Transcription is the process by which the information in a strand of DNA is copied into a new molecule of messenger RNA (mRNA). It requires RNA polymerase and ribonucleotides (deoxyribonucleotide is incorrect).

Translation is the synthesis of a protein from an mRNA template where the code in the mRNA is converted into an amino acid sequence in a protein. It requires ribosomes, tRNA's and amino acids.

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

Crib:

Prokaryotic cells lack a true nucleus, only have a single type of RNA polymerase, promoters are quite simple compared to eukaryotic promoters which often rely on many transcription factors, genes are typically arranged in operons and RNA does not undergo processing prior to translation.

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

Crib:

Differences in the concentrations of different components, which are not renewable so they will get depleted. Viscosity and physical properties are different from inside an enclosed cell (e.g. affecting diffusion). There is no genomic DNA present so no dynamic regulation of gene expression by up or down-regulation of other genes.

(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

(cont.

time with assembly and cloning.

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

Crib:

Diagram should be circular with a continuous expression cassette consisting of (from 5' to 3'): 1) promoter (regulates expression and is the binding site for the RNA polymerase and any transcription factors, credit for mentioning that this could be constitutive or inducible); 2) ribosome binding site or Shine-Dalgarno sequence (regulates ribosome binding); 3) Protein X CDS (contains the codons that define the order of amino acids in Protein X, credit for start and stop codons); 4) terminator (terminates transcription). Origin of replication and resistance marker should be elsewhere on the circle and not connected directly to the expression cassette but their exact location is not important.

(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%]

Crib:

Required parts are the promoter, RBS, Protein X CDS and terminator. Origin of replication and resistance marker are not required.

(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%]

Crib:

Diagram should illustrate denaturation of dsDNA at 95 C followed by binding of primers to the ssDNA at 50-60 C, extension 5'-3' by DNA polymerase at 72 C regenerating dsDNA and then back to denaturation at 95 C.

(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%]

Crib:

SDS-PAGE, functionality testing, incorporation of a fluorescent reporter as fusion to Protein X, use of fluorescent amino acids (acceptable although works much better

in PURE TX-TL). Credit for mass spec, FPLC or other relevant chromatography technique, and other valid biophysical methods.

(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%]

Crib:

You could test translation independently of transcription by adding in vitro transcribed RNA rather than DNA into the TX-TL reaction. Another option is to measure RNA and protein expression independently from a control DNA encoding fluorescent protein and an RNA aptamer which generates a different fluorescence wavelength (or another signal) when it binds to its ligand. Credit for other reasonable suggestions.

(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%]

Crib: Three of:

- Insufficient quantity of DNA: add more DNA.
- Insufficiently strong transcription or translation activation: try a stronger promoter or RBS, add more inducer if relevant.
- Protein X is being degraded: adjust reaction parameters such as pH, redox state, check for known proteolysis sites in Protein X, add protease inhibitors, engineer for increased stability.
- Protein X inhibits TX-TL: more complicated, may need to redesign expression approach.

(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%]

Crib:

The exact mechanism of Protein X toxicity is stated as unknown so several options could be viable. Three of the following or other logical and feasible solution:

- Grow cells slowly at low temperatures and with low Protein X expression (low copy number ori, weak promoter/RBS) to avoid toxicity building up.
- Express in periplasm to avoid build up in the cytosol.
- Express as an inactive fusion with a carrier protein that is cleaved in downstream processing.
- Split and express as two or more parts that are later joined.
- Engineer the protein to reduce toxicity in an expression context while retaining it in the end-use context e.g. more specifically toxic to target cells. However, this would typically require knowledge of mechanism.

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%]

Crib:

Nanobodies are fragments derived from the variable region of heavy-chain-only antibodies found in camelids (accept llamas or camels). Monoclonal antibodies have both heavy and light chains. They are smaller, more soluble, can be expressed in bacteria and are very stable and robust to different conditions. It is easier to engineer them for different purposes because they are a fragment of a single homogeneous chain, this includes generating multivalent nanobodies.

(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%]

Crib:

Genotype-phenotype linkage is the connection between the genetic information (genotype) and the observable characteristics (phenotype) of an organism or in this case an expressed construct. Maintaining a physical link between the DNA sequence and the nanobody by compartmentalisation or physical attachment (e.g. in ribosome display) is important because the information in DNA is easily retrieved by DNA sequencing and can be replicated and amplified for further expression/analysis via DNA replication or even transcription. Proteins alone contain information that is difficult to obtain because protein sequencing is expensive and information is lost during translation that cannot be recovered, due to codon redundancy.

(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%]

(cont.

Crib:
Codon-optimised DNA encoding the protein is determined and DNA library is synthesised *de novo* and cloned into a phagemid vector - any assembly method could be used as this is typically only a two-part assembly. Phagemid vectors fuse the nanobody gene to a gene encoding a coat protein of a filamentous bacteriophage (e.g. M13). Vector is transformed via heat shock or electroporation into *E. coli* for phage production. When the phage is produced, the nanobody is displayed on its surface, bound to a coat protein.

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

Crib: The phages are introduced to the target antigen immobilized on a solid support and incubated to allow binding. Several wash steps follow to remove weak binders and enrich strong binders. Strongly bound phages are eluted under mild conditions that disrupt nanobody-antigen binding (e.g. low pH or high salt concentration). Phages reinfect fresh bacteria and high-affinity phages are now enriched. Several more rounds occur.

(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%]

Crib: The selected candidates are cloned out of the phagemid vector and into an expression vector for production e.g. in *E. coli* or yeast and produced using shake flasks, bioreactor or other standard methods.

(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%]

Crib: The nanobody should be specific only to the target antigen and not bind to any other proteins or biomarkers found in humans. The nanobody should not itself be immunogenic and stimulate the patient's immune system to attack or neutralise it.

(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%]

Crib:

Nanobodies can be conjugated to chemical (e.g. chemotherapy drugs, toxins, or radioisotopes) or biological therapeutic payloads and deliver them selectively to cancer cells. They can be coupled to immune cells to bring them into close proximity e.g. Chimeric antigen receptor T cells (CAR-T cells) or Bi-specific T-cell engagers (BiTEs).

They may also operate without further engineering to interfere with signaling pathways

and this nanobody could be combined with others to produce a multivalent construct that is more selective.

(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%]

Crib:

Governments, national and international regulatory bodies (e.g. National Regulatory Agencies), international organisations (eg. different parts of the United Nations (UN), Convention on Biological Diversity Conference of the Parties (COP)), research institutions and scientists, civil society organisations, NGOs, biotechnology companies and other industry stakeholders.

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

Crib:

Most of the risks involved in operation of the self-driving lab described here would be covered by existing governance mechanisms. Credit for suggesting that the implementation of different biosafety and biosecurity measures may need to be handled differently when the experiment designer is an AI system and there is reduced human oversight e.g. cybersecurity could be considered a higher risk and more preventative measures put in place.

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

Crib:

Three of the following (or closely related topics):

- Ethics: maintaining ethical standards in research e.g. by considering and minimising potential harm.
- Societal Engagement: involving the public and stakeholders in the research and innovation process
- Gender Equality: ensuring equal opportunities for all genders and addressing gender biases in research processes and outcomes.
- Open Access: open sharing of research results, data, and knowledge.
- Governance and Accountability: having clear governance structures and mechanisms to ensure accountability in research and innovation.

(cont.

- Environmental Sustainability: considering the environmental impacts of research and innovation activities.
- Inclusion and Diversity: ensuring diverse perspectives, backgrounds, and voices are considered to avoid biases and better address the needs of different communities.

END OF PAPER

THIS PAGE IS BLANK