Version GM/final

EGT2
ENGINEERING TRIPOS PART IIA

Friday 6 May 2022 2 to 3.40

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 Consider a cascade circuit, where a transcription activator TA1 activates the expression of a second transcription activator TA2, which in turn activates the expression of a GFP reporter gene. TA1 is constitutively expressed and is activated by signal S1. Actived TA1 causes TA2 to be expressed at the rate of E2, and TA2 is activated by signal S2. GFP expression starts when the level of TA2 exceeds the threshold K2.

(a) Sketch a schematic of the circuit showing the different genetic elements (promoters and coding sequences), how the different signals regulate the transcription factors, and how the activated transcription factors regulate the different genes. [25%]

(b) Sketch a graph of the levels of TA1, TA2 and GFP in the situation where signal S1 appears at t = 0, and signal S2 is present throughout the experiment. [25%]

(c) If the signal S1 occurs as a pulse of duration P and then vanishes, sketch how the levels of TA1, TA2 and GFP change. [25%]

(d) What is the minimum value of P required for the activation of GFP expression? [10%]

(e) Cells carrying the circuit are exposed to the S1 signal for the minimum required pulse duration as above. However, no GFP signal is detected. What is the most likely explanation for this discrepancy? [15%]

2 You have designed a synthetic genetic circuit to produce a substance of value. You know that such a circuit will impose a resource burden on the carrying host. Therefore you intend to quantify the effect of this circuit on host fitness. You design an experiment in which you co-culture the circuit-carrying strain (strain 1) with a circuit-free strain (strain 2), monitoring the culture to see how one competes with the other. To distinguish the two strains by microscopy, you express yellow fluorescent protein (YFP) from promoter P1 in the circuit-carrying strain, and cyan (blue) fluorescent protein (CFP) from promoter P2 in the circuit-free strain. Both of these fluorescent reporter genes are integrated into the bacterial chromosome.

(a) Strain 1 (carrying the genetic circuit on a plasmid backbone and expressing YFP from P1) has a doubling time of 24 minutes, and strain 2 (carrying no circuit and expressing CFP from P2) has a doubling time of 20 minutes, both in rich medium. What is the expected proportion of these two strains after 6 hours of growth in a mixed culture that had equal proportions of these strains at t = 0? [30%]

(b) If you were to perform the experiment from part (a) above five times, would you expect to see blue and yellow cells consistently in this proportion? Explain your answer. [10%]

(c) When you actually carry out the experiment in part (a) above you are surprised that the proportions of colours observed are the opposite of those expected.

(i) Given reasons to explain this observation. [20%]

(ii) How can your explanations be verified? [15%]

(d) The unexpected observation in (c) above suggests that there are problems in the design of this experiment. How could you change the strain construction to avoid these problems? [25%]

3 The bacterial species *E. coli* is commonly used to produce proteins for medical or industrial applications. Plasmids can be used to carry the genetic circuits that allow expression of these proteins. Human insulin was one of the first recombinant proteins expressed in *E. coli*. In humans, insulin is expressed as a single polypeptide that is proteolytically processed such that two polypeptide chains (designated polypeptides A and B), which are bound together by two disulfide bonds, form the mature single protein unit. *E. coli* does not have the cellular machinery to process a single polypeptide in this way. Therefore the A and B polypeptides have to be expressed independently, purified, and then joined via an oxidation reaction.

We wish to make human insulin by synthesising DNA encoding a genetic circuit, rather than cloning the parts from elsewhere.

(a) First, we need to build a plasmid into which this synthetic DNA will later be cloned. Using a diagram, show the key DNA elements that should be present on the plasmid, and describe the function of each of them. [15%]

(b) The original scheme for making human insulin A and B chains was based on making gene fusion polypeptides where, in separate constructs, part of the lacZ protein was fused with either the A or B polypeptides. These lacZ::A and lacZ::B fusion coding sequences were expressed under the control of the lac promoter, and the resulting proteins were purified using antibodies that reacted against the lacZ part of the fusion proteins. After purification, this lacZ protein fragment was cleaved off using proteases.

An adaptation of this approach would be to use a peptide tag such as six consecutive histidine residues (a 'His tag') instead of the lacZ component of the fusion, as such residues allow tight binding to a nickel column facilitating purification. Adjacent to the His tag would be a cleavage site for a specific protease. Draw the full genetic circuit for this approach applied to polypeptide A, showing all the parts and how they interact. [30%]

(c) To put the above scheme into practice, you would have to design a DNA sequence carrying the insulin polypeptide A under the control of the lac promoter. How would your design be influenced by the fact that you are intending to express a human open reading frame in *E. coli*? [5%]

(d) Your designed DNA arrives from the manufacturer as linear double-stranded DNA. Using the plasmid vector from (a), describe the steps required for inserting the synthesised DNA into the vector. [15%]

(e) The circular DNA resulting from (d) above is transformed into *E. coli*. Describe the steps following this that are required to isolate clones, and to check that the DNA constructs carried by the clones are correct. [15%]

(f) *E. coli* cells containing the correct genetic construct identified above were grown by shaking at 37°C in rich medium but do not express polypetide A. You prepare plasmid DNA from such cultures but it is only present in vanishingly low quantities instead of being abundant. Assuming the plasmid preparation was carried out correctly, what is the simplest explanation with regards to the growth conditions? [10%]

(g) You solve the problem in the previous step and repeat the experiment. The expression level of polypeptide A is now measurable but low. Analysis of mRNA shows that the gene is also being transcribed at low levels. What is the simplest explanation with regards to the growth conditions? [10%]

4 In the central dogma of molecular biology, translation is the process by which information encoded in the sequence of four nucleic acid bases (A, C, G and T) determines the sequence of 20 different possible amino acids in a protein.

(a) Explain why a triplet coding system is needed to encode the 20 different amino acids. [10%]

(b) During protein translation it is possible to insert a non-standard fluorescent amino acid at defined locations within an engineered protein by means of an 'orthogonal' aminoacyl-tRNA synthetase (aaRS)/tRNA pair. This aaRS adds ('charges') the fluorescent amino acid to its own target tRNA rather than to the native *E. coli* tRNA molecules. These fluorescently charged tRNAs are able to take part in protein translation. In order to make space in the genetic code the anticodon loop of the fluorescently charged tRNA recognises the least-used stop codon, UAG. Thus, by engineering this codon into the desired location in a protein's gene it is possible to introduce the fluorescent amino acid into a precisely defined location in the protein.

Using an established aaRS/tRNA system, a researcher wants to modify a protein, P, using this approach. They have engineered the gene for P by introducing a 5'-TAG-3' codon at the desired position. This construct has then been introduced into the *E. coli* strain that expresses the appropriate aaRS/tRNA pair. The engineered strain is grown correctly in culture medium containing the non-standard amino acid, X, which is taken up into the cell. However, the researcher is disappointed to find that around half of protein P variants are truncated at the site where amino acid X should be present.

(i) What is the most likely explanation for the observed protein truncation? [10%]

(ii) In addition, the strain grows more slowly than the parent strain. Explain why this is the case. [10%]

(iii) It is possible, but slow and expensive, to resynthesise bacterial genomes completely by progressively replacing segments of the bacterial chromosome in living cells. If time and money were not limiting, how would you redesign the genome to avoid the truncation problem above? [15%]

(c) Short peptide affinity tags are often added to proteins of interest in order to allow purification. For instance, the FLAG tag is an eight amino acid tag that is tightly bound by a particular antibody, which in turn can be attached to a solid substrate, so allowing purification. Explain where in the gene coding for the protein described in part (b) above you would place the tag and why you would choose this location. [15%]

(d) Agarose gel electrophoresis is used at several stages when constructing recombinant DNA. Outline how it works. [15%]

(e) After constructing a clone, it is normal practice to determine its sequence. Outline how Sanger capillary sequencing works when determining the sequence of a 300 base pair fragment inserted into a cloning vector of known sequence. [25%]

END OF PAPER

THIS PAGE IS BLANK