Version GM/2

EGT2
ENGINEERING TRIPOS PART IIA

---

2 May 2019 2.00 to 3.40

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 Phage display is a powerful method for isolating antibody genes of interest from high diversity libraries. The power of phage display arises from the fact that the gene encoding a particular displayed protein is carried in the phage and thus there is a physical association between the gene and the corresponding displayed protein. A related powerful technique is called mRNA display, a cell-free system in which DNA libraries are transcribed into a corresponding population of mRNA molecules and these are then translated into proteins. The association between "gene" and protein in mRNA display is made by means of a chemical linkage that fuses the mRNA and the protein.

Screening mRNA display libraries works in much the same way that phage display does. A target of interest, for instance a protein, is immobilised on beads. The target-carrying beads are mixed with the library and displayed proteins are allowed to bind. After washing to remove non-specifically bound displayed proteins, the specifically bound displayed proteins and their associated genetic material are eluted with e.g. high salt.

(a) Give two advantages that mRNA display might offer over phage display. [10%]

(b) Camelids (llamas, dromedaries, camels) have immune systems like that of humans but produce a type of antibody that differs in structure to human antibodies: compared to human antibodies the light chain is missing and thus the mature antigen-binding protein is derived from a single variable immunoglobulin gene analogous to the human heavy chain gene.

(i) Give two reasons why an antibody based on a single gene might be better suited for mRNA display. [20%]

(ii) How would you obtain a set of single domain antibody DNA sequences with sufficient diversity, from an immunised animal? [10%]

(iii) An alternative approach to obtaining a diverse library would be through gene synthesis: to introduce diversity, random sequences of fixed length can be synthesised and inserted into the hypervariable regions of the antibody gene. What possible disadvantage might this approach have? [10%]

(c) Cells express proteins on their surfaces. Some of these proteins are found on all cells and others are specific to sub-categories of cells. Recently it has been discovered that some tumour cells express unique cell surface antigens and these are a route to destroying the cancer, for instance by directing the immune system to recognise these antigens.

You are given both normal and tumour cells from a patient. Describe modifications to the mRNA display process that will allow you to isolate mRNA from the display library

that discriminate between normal and tumour by encoding proteins that bind only to the tumour cells. [30%]

(d) Through your screen you find an mRNA that encodes a protein that strongly binds to tumour cells and not to normal cells. You want to find a way to display this protein on the surface of particular immune cells, which are capable of killing cells to which they tightly bind.

How might you engineer a human cell to display this protein, P, on its cell surface? [20%]

2 You are working with an oceanic microorganism (M), which lives in very low numbers in North Atlantic waters (average temperature < 20°C). It receives carbon-containing molecules (2-6 carbon atoms) from the water and, in the presence of energy provided by sunlight, converts them into useful building blocks. You would like to engineer this microorganism so that it consumes long hydrocarbon chains, hoping that the resulting microbes might be exploited to clear oil spills on the surface of the ocean. A paper has just been published, which describes a novel microbe (D) isolated from a desert environment. (D) can utilise long chain hydrocarbons efficiently as a carbon source, but is able to do this only in the presence of sunlight. You come up with a plan to make use of this microbe to engineer (M).

(a) Explain whether or not (M) becomes a cell factory once it is modified to acquire the desired ability to consume long hydrocarbon chains. [5%]

(b) What problems can be anticipated in utilising the hydrocarbon uptake and utilisation pathways from (D)? Suggest possible methods and approaches to overcome these problems. [20%]

(c) Suppose you would like to run preliminary modelling analyses to evaluate the feasibility of your approach. Assume that the metabolic network of (M) is available, while that of (D) is not. However, kinetic data measured in living cells are available for the hydrocarbon uptake of (D). Explaining your reasoning, suggest a modelling approach that would be useful for this situation, and one that would not be useful. Hint: pH and temperature are important parameters in determining reaction kinetics. [20%]

(d) This project idea is pitched to an expert on metabolism in (M), who praises it by saying: "Very good choice of a host; this microbe can top flocculate when nutrient sources become scarce." Explain why this is relevant to your idea. Hint: Top flocculation occurs when the microbes clump together into a floc on the surface of the liquid environment. [20%]

(e) Why might flocculation not occur and what genetic modification(s) could help ensure that flocculation would take place at the right time? [15%]

(f) Assume that (M) has now been engineered to work in the North Atlantic. In the laboratory it has been tested on oceanic water samples that have been contaminated with oil. After successfully using (M) to clear up the oil contamination, you find that the rest of the marine microbiome has been severely and adversely affected. Explain a possible reason for this observation. [20%]

3 (a) Bacteriophages, also known as phages, are viruses which infect bacteria. Like all viruses, phages commandeer the cellular machinery of their hosts to make copies of themselves. With the aid of a diagram, explain what functional elements encoded in the DNA are needed to encode a phage protein and cause it to be expressed. [5%]

(b) Some phage genomes are composed of RNA instead of DNA, which might seem to conflict with the central dogma of molecular biology that the *E. coli* host follows. Briefly explain two ways that phages with RNA genomes might avoid this problem. [10%]

(c) As a defence against phage infection, many bacteria produce restriction enzymes which cut foreign DNA at particular sequences. For instance, the NdeI restriction enzyme cuts at 5'-CATATG-3'.

(i) Show how often on average it would be expected that an NdeI restriction site will occur in a random DNA sequence with a 60% A-T base composition. [5%]

(ii) A new enzyme is discovered that cuts 5'-CATCGG-3'. How often will this be expected to cut in the same random sequence as above? [5%]

(d) PhiX3G1 is a phage that infects the bacterium *Escherichia coli* and contains a single NdeI site in a gene essential for its ability to carry out infection. Wild type (WT) *E. coli* does not naturally express NdeI and is used as a host to grow a preparation of phage particles. A strain of *E. coli* (Eng) has been engineered to express NdeI along with a methylase that protects the *E. coli* genome from cleavage. Identical quantities of the phage preparation are used to infect a) the WT strain and b) the Eng strain. The efficiency of infection can be estimated by counting the number of phage "plaques", which are colony-like zones of killing formed on a lawn of bacteria.

(i) Explain why the number of plaques is lower when infecting the Eng strain. [15%]

(ii) Phage from many plaques arising from infection of the Eng strain above ("Round 1") are now used to repeat the experiment above ("Round 2"). Explain why there is now no difference in infection efficiency between the WT and Eng strains. [10%]

(iii) 10,000 plaques arising from the Round 2 infection of the WT strain above are now tested individually ("Round 3"), against both WT and Eng strains. 9,999 plaques tested behave as in Round 1 (low efficiency infection of Eng strain) but one behaves as in Round 2 (equally efficient infection of both strains). What molecular events across the three rounds explain the behaviour of the phage in this distinct plaque? [30%]

(e) PhiX3G1 possesses a DNA genome encoding 10 unique protein sequences. The total combined length of these 10 proteins is 1986 amino acids. The genome of PhiX3G1 is 5386 nucleotides in length. Why are these results discordant and what might the explanation be? [20%]

4 (a) Why is it useful to sequence the genomes of organisms? [15%]

(b) As part of the process of sequencing a genome, the genomic DNA is mechanically sheared in order to generate random fragments. Typically this process results in there being single-stranded regions at the ends of each DNA strand and, unless repaired, these would interfere with subsequent processing steps. Thinking of cases in which the single-stranded region has a 5' end, outline the components needed in an enzymatic reaction that could repair this region back into double-stranded DNA. [15%]

(c) Subsequent processing allows the sequencing of the above DNA. The resulting sequencing reads are compared to each other and those that overlap can be assembled to form longer sequences, perhaps ultimately the whole genome. What factors influence the extent to which a full and accurate genome assembly can be made? [20%]

(d) Is it typically easier to find protein-coding genes in prokaryotic or eukaryotic genomes? Explain why, by outlining the way in which such gene structure can be delineated in each case. [30%]

(e) You have isolated and cultured a number of different prokaryotic and eukaryotic microbes that produce coloured pigments only when exposed to ultraviolet (UV) radiation. In each case you have reason to believe that the product of more than one gene is required in each microbe for pigment synthesis. Outline the approach you might take to discover the genes responsible for pigment synthesis. [15%]

(f) How might the genomic distribution of these genes differ in prokaryotes vs eukaryotes? [5%]

END OF PAPER