Version GM/final

EGT2
ENGINEERING TRIPOS PART IIA

Friday 7 May 2021 13.30 to 15.10

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 and at the top of each answer sheet.*

STATIONERY REQUIREMENTS
Write on single-sided paper.

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed.
You are allowed access to the electronic version of the Engineering Data Books.

10 minutes reading time is allowed for this paper at the start of the exam.

The time taken for scanning/uploading answers is 15 minutes.

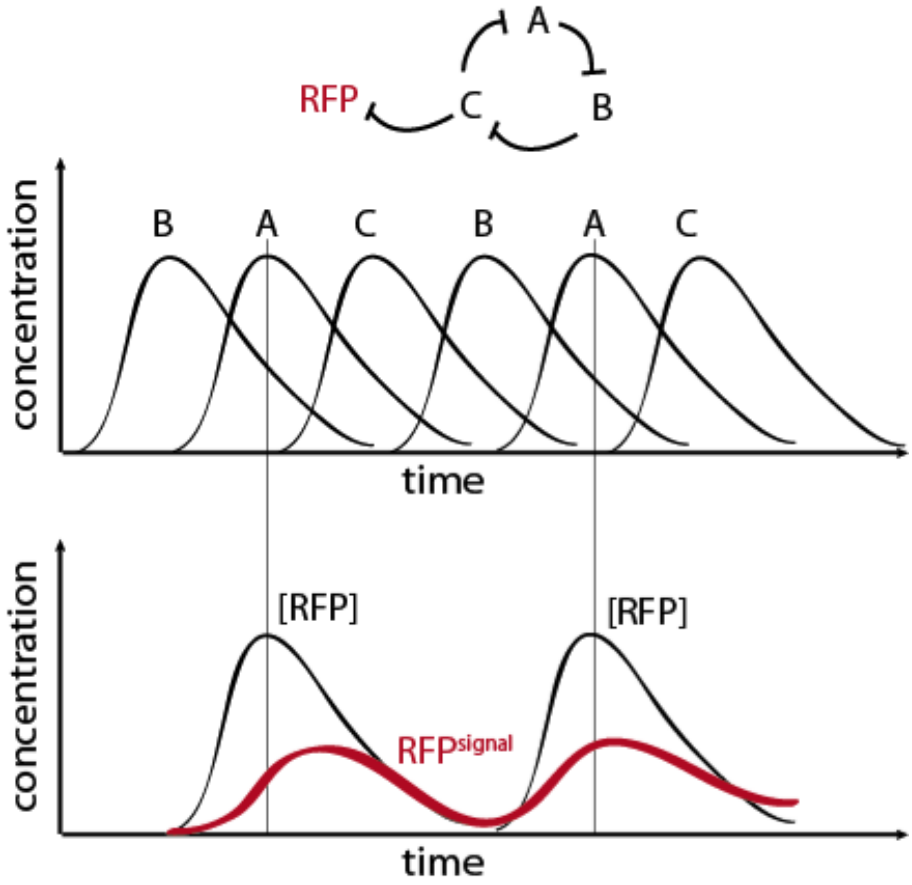
Your script is to be uploaded as a single consolidated pdf containing all answers.

1 Consider a repressilator circuit that has three repressor proteins (A, B, and C) controlling each other's expression in a cyclic manner. Protein C also controls the expression of a red fluorescent reporter protein (RFP, maturation time = 40 min) that can be used to monitor the function of this circuit using fluorescence microscopy.

(a) (i) Sketch the expected changes in the concentration of all of the three repressors over time. On a separate sketch underneath this, using the same time scale, sketch the expected concentration of the RFP reporter and also its signal.

[20%]

Crib:



(ii) Explain what is shown by the plots you have drawn above.

[10%]

Crib:

If the order of repression is A>B>C, the order of the concentration profile should be B>A>C. The profile of individual proteins have a rise part and a dilution part. The dilution is from cell growth and it dictates the period.

The RFP chromophore has to mature to become fluorescent and therefore there is a delay between the actual protein concentration dynamics and the fluorescent signal

(cont.

that one sees. Since the cells are diluting the protein and the signal, the slower maturation relative to the dilution from growth will make only a subset of proteins become fluorescent. Therefore, the amplitude of the concentration profile estimated from the signal will also be lower.

(b) Imagine that all three of the repressors have the same repression coefficient. Imagine also that their peak concentrations are 100 times higher than their repression coefficients. If all of these proteins are only diluted through growth, and the generation time of this kind of bacteria is 20 minutes, what is the expected period of this oscillator? [25%]

Crib:

Consider the time at which one of the repressors is at its peak concentration. Its concentration has to decrease 100-fold before the controlled gene starts to be expressed. Each repressor protein has to dilute 100x for the next repressor to be expressed, which would take roughly 6-7 generations, since $2^6 = 64$ and $2^7 = 128$. Therefore, the repressilator circuit would visit the same state after 18 - 21 generations, three times the time for each of the arms to complete. Since the generation time is 20 min, this corresponds to 6-7 hours.

(c) What would be the expected period of this oscillator if you add a degradation tag to each of these repressors and to the RFP gene as well, which reduces their half-lives to 2 min? [10%]

Crib:

If we add a degradation tag, the signal removal would get much faster, and it would take 10 times faster (assuming same 1st order kinetics) to reach below the repression coefficient. Therefore the period of the repressilator would become 36-42 minutes.

(i) Why can you no longer use the RFP reporter for monitoring the dynamics? [10%]

Crib:

This is comparable or faster than the maturation time of the fluorescent reporter RFP. Therefore, we cannot use the RFP for the monitoring the dynamics.

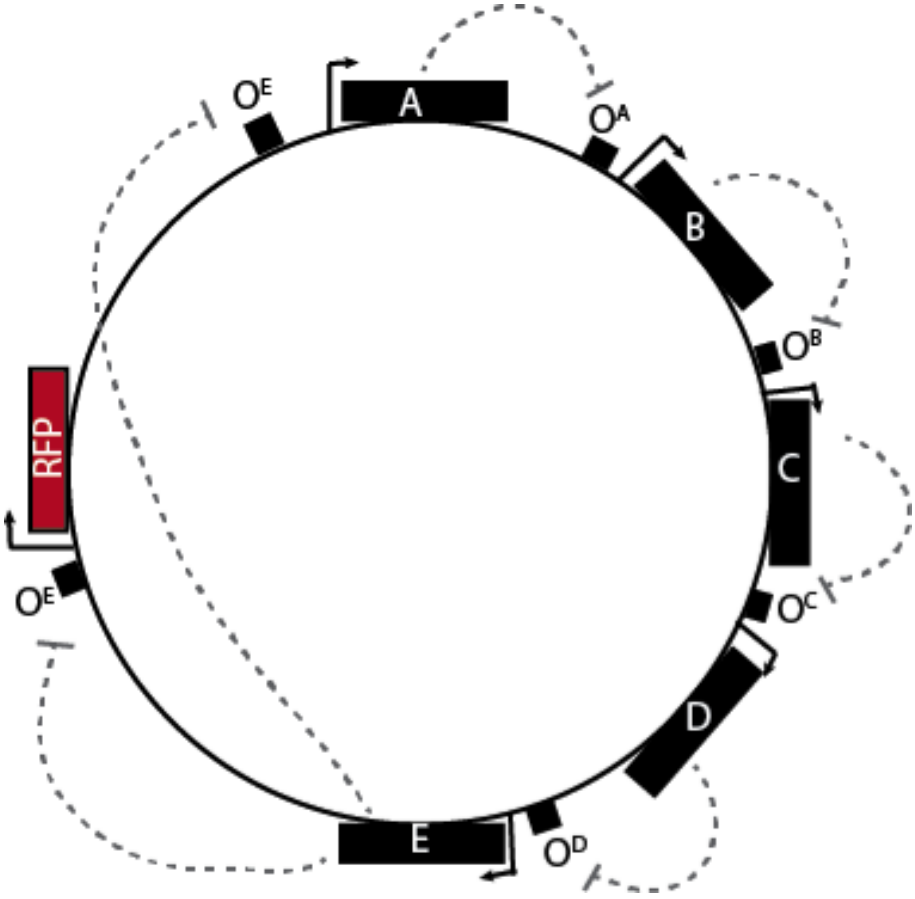
(ii) What modification is needed to this circuit design to observe the dynamics? [10%]

Crib:

We can instead use a fluorescent protein with faster maturation time, and also tag it with the same degradation tag as the other repressors so that it does not accumulate.

(d) To achieve a slower periodicity, we can add two more repressors in the cycle named D and E. Draw the conceptual design of the overall circuit as it would be laid out on a plasmid. [15%]

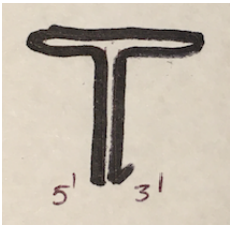
Crib:



2 (a) Sketch and then design the sequence of a single strand of synthetic DNA that will fold to make a double-stranded molecule in the shape of the capital letter T. Take the lengths of both the horizontal and vertical portions as 2k. Ignore the thickness of the double strand itself and assume that a five base loop is needed for a DNA strand to fold back on itself.

[25%]

Crib:



Many solutions are possible but for the most symmetric result the strand should start and end at the base of the T. One such solution: $T_{2k}C_kN_5G_{2k}N_5C_kA_{2k}$

(b) Suppose you wish to make the equivalent molecule but from enzymatically generated RNA and without the use of PCR or cloning.

(i) Describe how you would derive a new design based on the first, and what components you would need in the enzymatic reaction.

[40%]

Crib: Position a T7 RNA polymerase promotor a suitable distance upstream of the original design that transcription starts at the right place. Synthesis a complementary strand to the entire design and anneal the two strands to yield a double-stranded molecule that will act as a template for T7 RNA polymerase.
Transcription would be carried out by including the DNA strand in a suitable buffer for the enzyme, including NTP monomers, and the T7 RNA polymerase itself.

(ii) Considering your answer to part (a), what problem might you have in making the reaction work with the new design?

[15%]

Crib: There may be a tendency for the individual strands to fold up into T-shaped molecules more quickly than they anneal to make a double stranded molecule. Thus the annealed DNA may not be a good substrate for the transcription reaction.

(iii) Without using other enzymes or gel electrophoresis, how can you modify your design to remove the DNA once the RNA synthesis has taken place?

[20%]

Crib:
A biotin molecule could be incorporated into the synthesised DNA molecules (e.g. at

the extreme 5' end), allowing the DNA to be pulled out of solution using streptavidin coated beads.

3 A frog has been discovered that secretes antibiotic from its skin and it is hoped that, if it can be produced in sufficient quantities, it might be of therapeutic value. Characterisation of the antibiotic indicates that it is in fact a mixture of peptides that have antibiotic function. Therefore we would like to clone the genes encoding these peptides in order to allow them to be expressed for bulk production of the peptides. The peptides are analysed by direct peptide sequencing and it is discovered that they share their first 7 amino acids in common and also their last 6 amino acids in common. Between these common sequences are found variable peptide-specific sequences of 5-10 amino acids.

(a) You are provided with the genome sequence of the frog and with the amino acid sequences of the peptides. As they have different alphabets, outline how you would go about converting them in order that you can search for the corresponding genes.

[10%]

Crib:

Could either translate the genome in all 6 frames, or reverse translate the amino acid sequences into (redundant) DNA sequences. Then use sequence comparison programs (e.g. DP) to find matches.

(b) Having found the segments of the genome that correspond to the peptides, you discover that the two common regions are each encoded by a single segment of the genome nearby on the same chromosome. Furthmore, the many variable segments are arrayed in between these two constant segments. You hypothesise that perhaps some process like antibody gene rearrangement leads to the fusing of the terminal common segments with one variable region to yield a mature peptide-producing gene. You consider that perhaps different cells in the frog's skin contain different such rearranged genes, so leading to the observed diversity of antibiotic peptides secreted by the skin.

Based on the above and starting with an mRNA sample derived from the frog skin, outline an approach by which you can generate a library of all the genes that encode these peptides, ready to transform into host cells to generate a clone library.

[50%]

Crib: The common segments of 7 and 6 amino acids are encoded by 21 and 18 base DNA sequences respectively. A reverse complement of the 18 base sequence can be used as a primer to prime a reverse transription reaction yielding a single stranded cDNA. Addition of the 21 base sequence will allow PCR amplification of the cDNA template yielding double-stranded DNA molecules corresponding to the genes. These can be ligated into a suitably cut vector. The ligation mixture represents a library ready to be transformed into host cells.

(c) Having constructed a library as above in (b) using an *E. coli* expression vector, the mixture is transformed into *E. coli*. However the transformation efficiency is extremely poor compared to an identical transformation carried out with the empty vector, and the colonies grow very weakly. What has gone wrong and how might the problem be fixed?

[20%]

Crib: The peptides are being expressed, and the antibiotics are active against the *E. coli*. Using an expression vector with an inducible promotor might solve the problem.

(d) As an alternative to expression in living *E. coli* cells, the library is added to a cell-free transcription/translation system made from *E. coli* cells. It is possible to detect transcription of the genes, but production of the peptides is still weak. Give two possible explanations for these observations.

[20%]

Crib: The genes may need to be codon-optimised to express well in *E. coli*. It is possible that the antibiotic action of the peptides is to inhibit the translation process. Partial credit for suggesting the reaction mixture is lacking sufficient components for translation.

4 A yeast gene knock-out library has been created that is a collection of strains. In each strain a different yeast gene has been knocked out by replacing it with a selectable marker. Next to the marker is a barcode that identifies the knocked-out gene. Across the library, all the barcodes are flanked by the same PCR primers.

(a) If there are 6000 strains in the collection, justify the minimum length of barcode that would be required to distinguish all strains.

[10%]

Crib: $4^x = 6000$. log(6000)/log(4) = 6.28. So require 7 bases.

(b) Why in practice might one want to use a longer barcode sequence than above, and what features might it include?

[20%]

Crib:

To allow for sequencing errors the barcode should be longer to reliably distinguish between strains. From the DNA data store story we know that consecutive bases can be a problem so designing base 3 barcodes that avoid consecutive bases that are the same would be a good idea. (Mentioning the use of error-correcting codings also good, though this was not covered.)

(c) Suppose that we wish to discover whether any members of the gene knock-out library confer resistance to a toxin. Outline an efficient process by which we could do this.

[20%]

Crib: Small samples of all the yeast strains would be pooled in growth medium and treated with the toxin for an extended time to allow growth of any resistant strains. Growth of the culture to high density would indicate that such strains were present.

Alternatively a pool is split in halves to allow a comparison between growth medium alone and growth medium with toxin.

(d) Having carried out the process above, how would the resistant strains be identified?

[20%]

Crib: DNA would be prepared from the culture, PCR would be carried out using the primers that flank the barcodes and the PCR product sequenced. Barcodes present at high abundance would identify the resistant strains.

(e) Curiously, it is found that strains that are resistant to the toxin now also require

uracil to be supplied in their growth medium in order to be able to grow. Outline how this fact can be exploited to identify human genes that carry out the same function as the knocked-out yeast gene.

[15%]

Crib:

Transform the toxin-resistant yeast knock-out line with a human cDNA expression library. Plate the transformants in the absence of uracil and look for growth. Colonies that are resistant should carry plasmids expressing the human gene of interest.

(f) Having isolated a plasmid carrying a 3000 base candidate human gene, justify which method of DNA sequencing you would use to sequence it.

[15%]

Crib:

Any approach can be used if justified: conventional sanger sequencing primed from flanking vector sequences would not cover the whole gene but addition custom primers could be designed. Illumina sequencing would be overkill unless the clone was barcoded and included in a pool: sequencing reads with the barcode used would be extracted, the barcode removed and the sequences assembled. Oxford Nanopore or PacBio would be able to read the whole gene in one read but would require multiple independent reads to build up a low-error consensus.

END OF PAPER

Q1 Repressilator:

The least popular question. A significant fraction of students did not understand question 1 (b), but the ones that understood it, got the calculations right. Students tended to perform well on 1(a) and 1(d). These are more descriptive and required correct intuition. However, a fraction of these students often missed some crucial details in the sketches. Most of the students got 1(c) correct, but some of them misunderstood the impact of delay in observed signal dynamics. Overall, it seems students have good grasp of circuit elements and function, but often lacked the quantitative understanding of circuit properties.

Q2 DNA/RNA structures:

Many candidates missed that they should design a sequence in (a) rather than just sketch its structure. Surprisingly performance on (bi) was poor, with most but not all candidates failing to remember elementary facts about transcription that had been well covered. As a result performance on (bii) was also poor.

Q3 Frog skin antibiotics:

(a) was generally answered well. In (b) many students did not notice that the common regions provide information for PCR priming sites.

Q4 Yeast knockout collection:

The most popular question. Many students had difficulty with the fact the question was about a library of yeast strains rather than a library of clones, even though the construction and use of the yeast library was covered in the lectures. This impacted performance on (c) and (d). (e) was supposed to be hard and many students had difficulty. Nevertheless, high marks were obtained by some.