Version SB/final

EGT2
ENGINEERING TRIPOS PART IIA

Friday 5 May 2023 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 Your synthetic biology project team is designing a genetic circuit that needs to rapidly respond to the presence of a toxic chemical in an environment. You collected some sample from the target environment and tested the growth of your chassis bacteria in it. The chassis bacterial cells doubled every two hours in this media.

(a) If you use a transcription repressor, which undergoes an allosteric change in activity when bound to this chemical, to control the expression of a fluorescent protein reporter, what is the expected response time of your designed system (time to reach 50 % of the steady-state level). [15%]

Crib:

The response time for activation is dictated by the removal rate of the reporter protein. As the reporter protein is only removed through the dilution from cell division, the rate of removal is equal to the rate of cell doubling. Therefore, the response time to reach 50 % of the steady state level is equal to the doubling time of the cell, which is two hours.

(b) After some initial testing you realised that the design is not sensitive to low levels of the toxic chemical. Your team lead suggested you try out three more repressors that are sensitive to this chemical. These new repressor candidates have different cooperativities (n_1, n_2, and n_3) and repression coefficients (K_1, K_2, and K_3). If $n_1 > n_3 > n > n_2$ and $K_2 > K > K_1 > K_3$, where n and K are the corresponding values for the original repressor, which of the three candidates would you pick for your new design and why? [15%]

Crib:

We should pick repressor 3.

The detection threshold is determined by the repression coefficient (K) of the repressor protein. As repressor 3 has the lowest value of the repression coefficient, it is expected to be sensitive to the lowest levels of the chemical.

(c) Since the response time of the above-mentioned design seemed too slow for your application, you decided to use a degradation tag on the reporter protein to speed up the response time.

(i) What other changes do you need in your circuit design to achieve the same steady-state response level? [10%]

Crib:

Since the steady-state level is determined by a balance of production and removal rate, an increase in the rate of removal using a degradation tag needs to be compensated by a proportional increase in the rate of production. This could be achieved by the use of a stronger promoter and a stronger ribosomal binding site.

(cont.

(ii) If the degradation tag makes the fluorescent reporter degrade at 0.2 min^{-1} rate, what is the new response time of the designed system? [10%]

Crib:

The new response time should be log(2)/0.2 min = 3.45 min

(iii) Which of these two versions is expected to be more energetically costly for the chassis cell and why? [10%]

Crib:

The new design with the degradation tag is expected to be energetically more costly, as (1) it requires production of more proteins through the use of stronger promoter and RBS and (2) uses ATP for the degradation of these proteins.

(iv) Could you think of an alternative design for speeding up the response time, which might be energetically more efficient compared to the previous two designs? [10%]

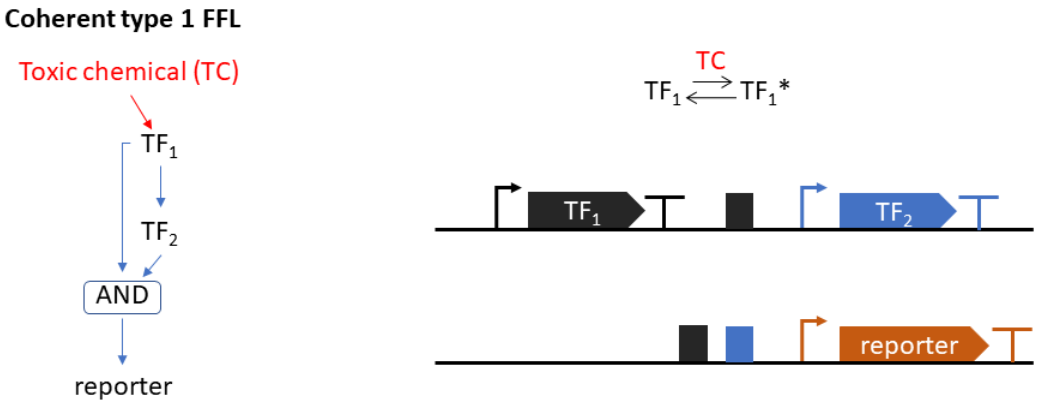
Crib:

We could use a negative autoregulation motif to speed up the response time.

(d) After testing the various designs, you realised that all the circuits are very sensitive to brief pulses of the toxic chemical. You want to avoid that. So you decided to use a feedforward design to ensure that the circuit detects persistent signals only. Draw the detailed schematics of such a feedforward circuit and the corresponding SBOL diagram. [30%]

Crib:

We need to use a coherent type 1 feedforward circuit to make sure that the circuit only responds to persistent presence of the toxic chemical.



2 You are using a bacterial plasmid-based expression system for your bioproduction project. Your plan is to produce a fluorescent protein, which is known to have no effect on the growth of the host bacterial cell. You picked a low copy plasmid as the vector for this work, where it is known that on average every newborn cell has 5 copies of the plasmid.

(a) What is the main source of variance of plasmid numbers across the population, if the replication is tightly controlled? [10%]

Crib:

The main source of variation is partitioning of the integer number of plasmids into two daughter cells during cell division.

(b) What is the expected variance of plasmid copy number distribution in the population and what is the expected fraction of plasmid-free cells in the population? [20%]

Crib:

For binomial partitioning with equal probability of partitioning into the two daughters, the variance in the distribution is given by: $np(1-p)$, where $p = 0.5$. The number of plasmids per cell just before the division is 10. So the expected variance is 10 x 0.5x 0.5 = 2.5.

The probability of finding a cell with zero plasmids is $P_0 = \binom{10}{0} \cdot 0.5^0 \cdot 0.5^{10}$ $= 0.000976 \approx 0.001$

(c) If each plasmid imposes a 5% demand on the cellular ribosome for the protein production and causes the cell to grow slower proportionally, how would that affect the distribution of plasmid-copy numbers in the population? [20%]

Crib:

Every additional plasmid costs the cell 5% of its ribosomes, which are otherwise used for cell growth and maintenance. Therefore, cells with higher copy-number of plasmids are expected to be growing slower compared to cells with lower copy-number of plasmids. Plasmid-free cells are expected to be the fastest growing cells. Therefore, over time the population distribution of plasmid copy number will shift towards the lower copy numbers, as those cells will disproportionately increase in number due to faster proliferation. A significant fraction of cells will have no plasmids, which will increase with time.

(d) To eliminate any unnecessary burden from the protein production, you have used a transcription activator to turn the protein production on and off as needed. Explain with

reason whether or not you expect a plasmid-free and plasmid-containing cell to have the same growth rate, if the protein production is kept off. [15%]

Crib:

No. The plasmid containing cells may still grow slower compared to the wild-type cell without plasmid, as replicating the plasmid will take time. Also the transcription factor is expected to be leaky and will therefore cause baseline expression of the protein product. Both of these will cost the cell its cellular resources and will make it grow slower.

(e) During experiments, you realised that the chemical inducer for the transcription activator seems to affect the growth of the chassis cell. As a result, you sought to avoid keeping the inducer in the growth-medium throughout the bioproduction experiment. Describe a genetic circuit design in which the inducer is added only briefly but transcription is maintained in the activated state for an extended period of time. [10%]

Crib:

You could use the genetic toggle switch design.

(f) To avoid the takeover by plasmid-free cells, you have also introduced an antibiotic-resistant cassette in the plasmid, and the cells are always grown in the presence of the corresponding antibiotic. Explain with reason if you expect this new system (bacteria containing the newly designed plasmid) to grow faster than the original design. [10%]

Crib:

No. It will grow slower. Constitutive expression of the antibiotic resistant protein will also impose metabolic burden on the cell, making it grow slower.

(g) After running an extended bioproduction experiment in a culture containing antibiotics, you decided to image the cells under a microscope. Since the cells were grown in the presence of the antibiotic, it is expected that all cells carry the plasmid. However, in your images, you found that a significant number of cells don't show any detectable signal from the fluorescent protein. How would you explain this observation? [15%]

Crib:

Since the expression of the fluorescent protein product is not beneficial for the chassis cell, mutations with reduced or impaired expression of the protein product could have growth advantage, without any effect on the expression of the antibiotic resistant protein. These cells will not express any fluorescent protein but will survive the antibiotic. Due to their faster growth rates, such mutant cells will increase in number over time.

3 (a) The restriction enzyme BamHI cuts the DNA motif shown below, with the cuts on each strand being shown with an apostrophe:

```
5'  G'GATC C  3'
3'  C CTAG'G  5'
```

In the list of restriction enzymes below, only the top strand is shown.

```
BamHI   G'GATC C
BglII   A'GATC T
BsrGI   T'GTAC A
KpnI    G GTAC'C
Sau3AI   'GATC
```

Fill out the empty cells of the table on the attached supplementary sheet to indicate which enzyme combinations produce compatible sticky ends. [20%]

Crib:

	BamHI	BglII	BsrGI	KpnI	Sau3AI
BamHI		X			X
BglII					X
BsrGI					
KpnI					
Sau3AI					

(b) Which of the above restriction enzymes will in general be less useful for building DNA constructs and why? [15%]

Crib: Sau3AI has a four base recognition motif so it will, within a sequence of equal base frequency, cut every 4^4 or ≈250 bases. Cloning vectors and many genes are several 1000 bases long so Sau3AI will tend to cut them into several small fragments, which in general is not useful.

(c) We would like to make a protein A/protein B fusion. The sequence of DNA that codes for the C-terminus of protein A is shown below, followed by three versions of the DNA sequence of Protein B. The DNA bases are grouped into triplets corresponding to the reading frame. The EcoRI restriction enzyme cuts `G'AATTC` and is to be used to make the fusion.

```
Protein A C-terminus:  ... TCC TGC TAT GTG AAT TCA TAA

Protein B N-terminus version 1:   GAATTC ATG AAA CCC TTT GGG CAC ...

Protein B N-terminus version 2:  GAATTCA ATG AAA CCC TTT GGG CAC ...

Protein B N-terminus version 3: GAATTCAA ATG AAA CCC TTT GGG CAC ...
```

Which of the three alternative versions for the Protein B N-terminus will yield the desired protein fusion? [15%]

Crib: version 2 is correct fusion (yielding ...SCYVNSMKPFGH...)

(d) For the two incorrect versions of the Protein B N-terminus from part (c), describe in detail the consequence of the fusion with reference to the codon table given at the end of the question and comment on the relative lengths of the products of translation. [20%]

Crib:

1) ...SCYVNS* : The stop codon truncates the fusion so shorter than (3).

3) ...SCYVNSNETLWA... : The fusion in the wrong frame will continue to be translated

(e) If only the protein A gene and version 3 of the Protein B sequence are available as clones, explain how PCR could be used to create the correct fusion. [10%]

Crib: Protein A would be amplified by conventional PCR. Sequence 3 of protein B would make a fusion in the wrong phase. The solution is to design a fusion PCR primer that misses out the extra base and use this with a protein B reverse primer to amplify protein B. Finally the fusion is carried out by mixing the products of the first two reactions and amplifying them with protein A forward, protein B reverse primers.

(f) There is insufficient information supplied to allow the design of all the primers needed. Give the sequence of those that can be designed with the information given. [20%]

Crib: Fusion primer: TCC TGC TAT GTG AAT TCA ATG AAA CCC TTT GGG CAC
Reverse primer for protein A: TGA ATT CAC ATA GCA GGA

Codon Table

Each codon is followed by the corresponding
single-letter amino-acid code.
* indicates a stop codon.

TTT F	TCT S	TAT Y	TGT C
TTC F	TCC S	TAC Y	TGC C
TTA L	TCA S	TAA *	TGA *
TTG L	TCG S	TAG *	TGG W
CTT L	CCT P	CAT H	CGT R
CTC L	CCC P	CAC H	CGC R
CTA L	CCA P	CAA Q	CGA R
CTG L	CCG P	CAG Q	CGG R
ATT I	ACT T	AAT N	AGT S
ATC I	ACC T	AAC N	AGC S
ATA I	ACA T	AAA K	AGA R
ATG M	ACG T	AAG K	AGG R
GTT V	GCT A	GAT D	GGT G
GTC V	GCC A	GAC D	GGC G
GTA V	GCA A	GAA E	GGA G
GTG V	GCG A	GAG E	GGG G

4 Write short answers to the following questions.

(a) In broad terms, how does a gene drive system work? [10%]

Crib:

Eukaryotes typically have two copies of each chromosome. A gene drive system works by copying itself from one copy of a chromosome to the other, so that when the organism reproduces all children receive the drive system.

(b) Why does exchange of genetic information enable evolution to proceed faster? [5%]

Crib:

Without exchange of genetic information (via horizontal gene transfer in eukaryotes, or sexual reproduction in eukaryotes) evolution cannot easily test the advantages of combinations of beneficial mutations that have occurred in different lineages.

(c) What is codon optimisation and what problem does it seek to avoid? [10%]

Crib:

Codon optimisation is the resynthesis of an open reading frame to reflect the codon-usage of the host organism in which is will be expressed. This is to avoid inefficient translation and thus low protein levels that could otherwise occur.

(d) Describe the concept of a library in molecular bioengineering giving a specific example. [10%]

Crib:

A library is a collection of molecules or clones. Many examples are possible e.g. a population of molecules prepared for sequencing is a sequencing library; a population of cloned gene fragments is a gene library; a cDNA libary represents the transripts expressed in a sample.

(e) Why are hybridomas important? [5%]

Crib:

Hybridomas are immortal cells lines that can produce a monoclonal antibody in unlimited quantities.

(f) Why is it necessary to humanise therapeutic antibodies? [5%]

Crib:

If antibodies from another organism as used for therapy in humans, then the human immune system recognises them as foreign and attacks them.

(g) What is a substitution matrix as used in protein sequence alignment? [10%]

Crib: A substitution matrix provides a measure of the similarity between two aligned residues: more strongly similar residues have more strongly positive scores, more strongly dissimilar ones have more strongly negative scores.

(h) Why do affine gap penalties give better protein sequence alignments? [10%]

Crib: Gaps tend to appear in runs when protein sequences are aligned and the gap penalty is a measure of how surprising it is to have to insert a gap. Therefore if a gap has been inserted it is less suprising to see another one and correspondingly the penalty should be smaller.

(i) What issues need to be taken into consideration when using DNA as a data store? [10%]

Crib: The fact that the writing (synthesis) and reading (sequencing) steps are noisy so error-correcting coding schemes need to be used. Also that there is a limit to the length of DNA that can be synthesised so it is necessary to break the information into separately indexable blocks.

(j) The three stop codons are TGA, TAG and TAA. Design a 12 base sequence that has a stop codon in all three reading frames on both strands. [5%]

Crib:

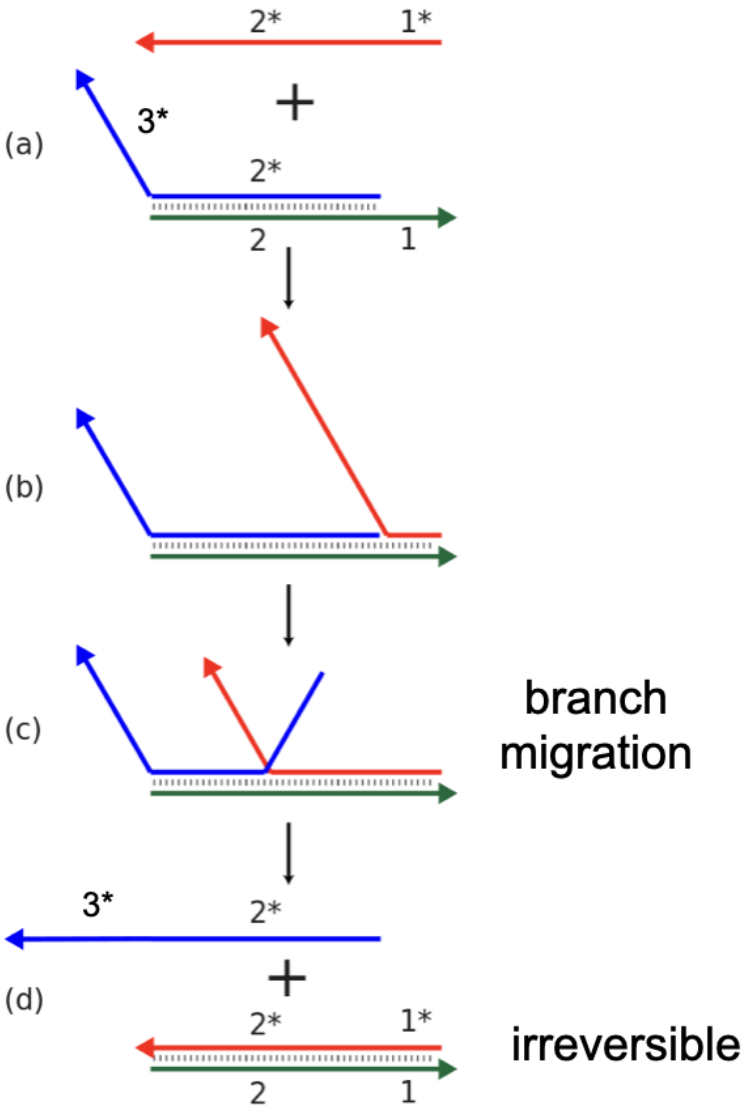
```
5' TTAATTAATTAA 3'
3' AATTAATTAATT 5'
```

(k) Describe the relationship between genome size and complexity of organisms. [5%]

Crib: Although prokaryotes are simple and have small genomes, and eukaryotes are more complex and have larger ones, in general there is not a good relationship between genome size and complexity of the organism for eukaryotes. This is due to most eukaryotes containing large but different amounts of repetitious DNA.

(l) Sketch the various steps in an irreversible strand exchange reaction. [10%]

Crib:



(m) What change has to be made to the system in part (l) above to enable the reaction, once completed, to be reversed? [5%]

Crib: The sequence of the lower strand of the product should be 321 instead of 21, with 3 therefore able to bind 3* of the blue sequence to initiate the reverse reaction.

END OF PAPER

THIS PAGE IS BLANK