Version GM/Final

EGT3
ENGINEERING TRIPOS PART IIA

Wednesday 3 May 2017 14.00 to 15.30

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.

You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.

1 Suppose we instruct a DNA synthesis company to synthesise a population, P, of single-stranded DNA molecules of the form A(N)20B:

```
          A                B
5' --------->(N)20---------> 3'
```

A and B are distinct sequences, TAATGTGCAATGTTCTTTATCCCCCC and CCCCCCAGAATATACAGAAGGCAGAC respectively. (N)20 means that the DNA synthesis machine adds an equal mixture of all four bases twenty times in a row.

Sequences within this random (N)20 region are able to fold up through complementary base pairing and adopt specific shapes. It is known that the right shapes can bind other molecules tightly. We are interested in searching for such folded sequences that are able to bind to the molecule theophylline, which we have covalently attached to polymer beads. Therefore we carry out the following cycle of steps to enrich for sequences that bind to theophylline:

Step 1) the population P is made up in binding buffer.

Step 2) binding buffer containing P is added to the theophylline beads and incubated with end-over-end rotation for 1 hour at room temperature in a plastic test tube.

Step 3) the beads are washed several times: each time by letting the beads settle to the bottom of the tube, carefully removing the overlying liquid and then resuspending the beads in DNA-free binding buffer.

Step 4) the beads are allowed to settle and the buffer replaced again but this time the tube is heated to 95^{o}C for 5 minutes so that any bound DNA is released from the theophylline beads. The buffer containing released DNA is then removed from the settled beads for use in the next step.

Step 5) The released DNA is amplified using PCR and (using a further step the details of which are not important) the single strands corresponding to the starting material, AN(20)B, are prepared.

This single-stranded material can again be bound to theophylline beads. Thus steps (2) through (5) are repeated another two times in succession so that three cycles in total of enrichment have been carried out. With each cycle we expect to progressively enrich for the specific sequences capable of binding the theophylline beads most tightly.

(cont.

(a) In Step (4) why does heating to 95^{o}C cause the DNA to be released? [5%]

(b) Paying careful attention to the sequences of A and B, write down the sequences of two primers that can be used in a PCR reaction to successfully amplify the sequences released in Step (4). Explain your design. [20%]

(c) It is likely that, in addition to sequences that can bind theophylline, the above procedure will select those sequences capable of binding the polymer beads themselves, and this is undesirable. Suggest and explain an extra step in each cycle of the above procedure that could reduce the abundance of such polymer bead-reacting sequences. [20%]

(d) Following the three cycles through the enrichment procedure described above it is likely that the diversity of sequences present is still high. Justify which sequencing method you would choose to most easily investigate the relative amounts of the most abundant (and thus presumably the tightest binding) sequences in the population. [10%]

(e) Describe how to prepare material from Step (4) so that it can be sequenced using the method you proposed in answer to (d). [5%]

(f) After sequencing you count the number of times each distinct sequence is found. Perhaps the independent sequences that have been selected share some sequence similarity responsible for theophylline binding, and so you are curious to see whether the different sequences are related. Outline how you would approach this challenge. [25%]

(g) Among your sequences you observe the following:

```
ACTATACCTATGGTATGACT (observed 200 times)
ACTATACCTGTGGTATGACT (observed once)
```

Provide three possible explanations for the origin of the sequence that is observed once. [15%]

2 Examine the reaction pathways shown in Fig. 1 below while reading the following: Organism 1 naturally produces the commodity chemical, C, using chemical A as a substrate. The pathway of interest is a linear pathway with two branches; one splitting off from intermediate B1 producing metabolite F through a series of reactions shown by a dashed arrow, and another branch splitting off from intermediate B4 similarly producing metabolite G. The yield of C on substrate A, $Y_{C/A}$, is 0.4g C per gramme of A. Enzyme *x3* has the highest flux control coefficient (FCC) in the linear pathway, with value 0.67. All reactions are non-equilibrium reactions with directionality indicated by the direction of the arrow. The enzymes catalysing each step are denoted in italic fonts as *x1* - x*5, y1, z1, n1* - *n2*. For Organism 1, the numbers in parentheses indicate the number of carbon atoms contained within substrate A, intermediate metabolites B1 – B4, and the product of interest, C. The arrow from B3 to *x1* signifies a positive regulatory action, not a metabolic reaction. Similarly, the blunt-headed arrow from B2 to *x1* indicates an inhibitory regulatory action.

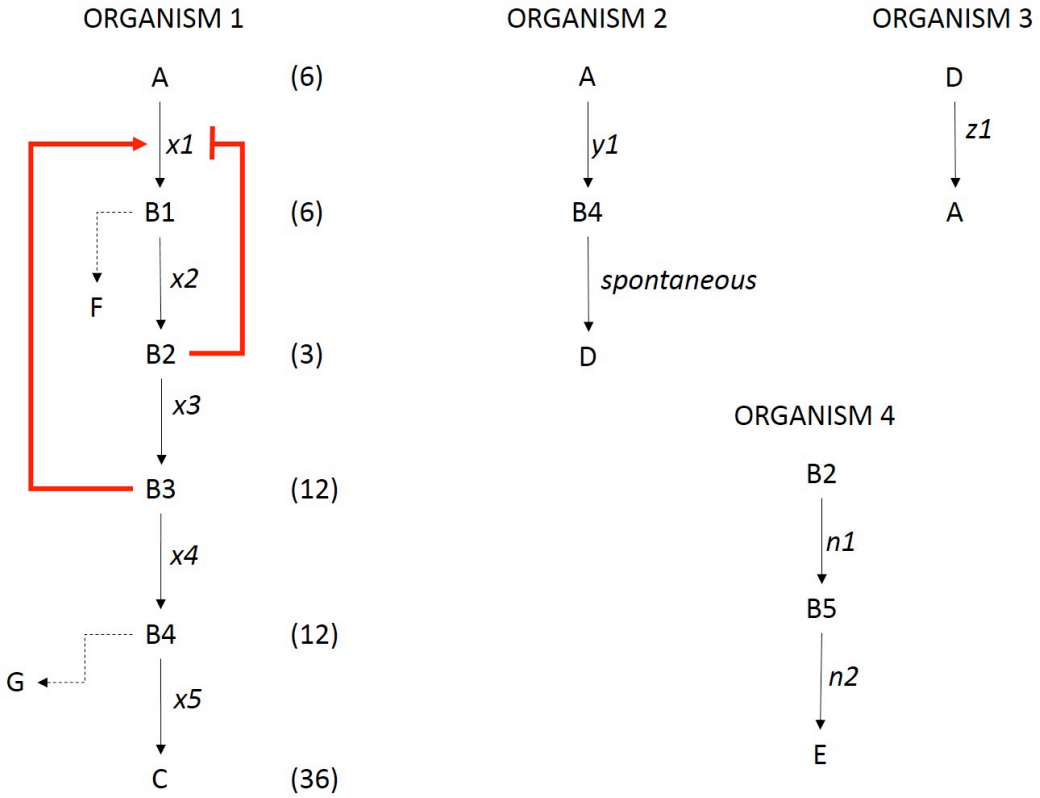


Fig. 1

 (cont.

(a) Explain which enzymatic steps in the linear pathway in Organism 1 are more likely to be anabolic processes and which ones to be catabolic processes. [10%]

(b) Identify the different types of metabolic regulation on this linear pathway. [5%]

(c) You are consulted as a metabolic engineer to suggest alternative genetic engineering strategies to improve $Y_{C/A}$. Digging into the literature, you identify Organism 2 and Organism 3 containing potentially useful pathways, which are also shown in Figure 1.

Describe possible bottlenecks limiting production yield in the linear pathway in Organism 1, and explain whether or not an alternative strategy can be implemented with the help of genes from Organism 2 and Organism 3. [20%]

(d) You investigate further and come across another pathway from Organism 4 (see Figure 1), which could also be useful. You are informed that enzyme *n1* has very high elasticity with respect to its substrate. Explain whether this pathway can be useful to increase the yield of compound C, and outline potential drawbacks or limitations. [15%]

(e) The pathway from Organism 4 is inserted into Organism 1 by genetic modification. The new $Y_{C/A}$ is 0.56g chemical C/ g of chemical A. The new FCCs are calculated and *x3* no longer has the highest FCC. How do you think the metabolism re-wired itself? [20%]

(f) Suggest three alternative methods to improve the yield of chemical C. [15%]

(g) Is it possible to calculate the new yields without proceeding with the genetic modifications and running experiments? What additional information is needed for this technique? What are the possible drawbacks? [15%]

3 (a) Proteins that recognise specific DNA sequences are fundamental tools in Synthetic Biology. Give three examples of such proteins and brief descriptions of their molecular functions. [20%]

(b) Serine-type DNA recombinases catalyse unidirectional inversion or excision of DNA flanked by pairs of asymmetric recognition sites. Excision occurs where the orientations of the recognition sites are aligned. Inversion occurs where the recognition sites point towards each other. How might the essential properties of Serine-type DNA recombinases be used to control bacterial transcription? [20%]

(c) Using only components from Table 1 design a genetic AND gate based on DNA recombination. Draw a diagram to illustrate your design and explain how the gate functions. The inputs should be arabinose and IPTG. The output should be GFP fluorescence. [30%]

(d) How would you expect the behaviour of an AND gate based on serine DNA recombinases to differ from an AND gate based only on transcriptional repressor protein activity, such as *LacI* and *AraC* acting at the same promoter? [30%]

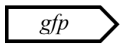
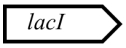
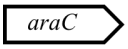
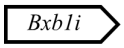
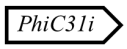
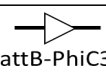
Abbreviation	Class	Description	Symbol
P_C	Promoter	Constitutively active promoter.	
P_{ara}	Promoter	AraC binding represses transcription.	
P_{lac}	Promoter	LacI binding inhibits transcription.	
RBS1	Ribosome binding site	Ribosome binding site. Responsible for recruitment of a ribosome during initiation of translation.	
Term1	Transcriptional terminator	Bidirectional transcriptional terminator. Triggers release of mRNA from the RNA polymerase transcriptional complex. Function is independent of the direction from which RNA polymerase approaches.	
gfp	Protein coding sequence	Codes for Green Fluorescent Protein (GFP), which fluoresces green under 395 nm wavelength light.	
lacI	Protein coding sequence	Codes for LacI protein. LacI binds to the Plac promoter as a tetramer, forming a DNA loop that prevents RNA polymerase from binding. When IPTG is bound to the repressor, the loop is opened, allowing RNA polymerase to bind.	
araC	Protein coding sequence	Codes for AraC protein. AraC binds to the Para promoter as a dimer, forming a DNA loop that prevents RNA polymerase from binding. When araboinose is bound to the repressor, the loop is opened and the AraC dimer stabilises RNA polymerase binding to the promoter.	
Bxb1-i	Protein coding sequence	Codes for Bxb1i DNA recombinase. Bxb1i binds specifically to attachment sites attB-Bxb1 and attP-Bxb1. Depending on the orientation of the attachment sites, Bxb1i catalyses inversion or excision of DNA sequences flanked by these attachment sites.	
PhiC31-i	Protein coding sequence	Codes for PhiC31-i DNA recombinase. PhiC31-i binds specifically to attachment sites attB-PhiC31 and attP-PhiC31. Depending on the orientation of the attachment sites, PhiC31-i catalyses inversion or excision of DNA sequences flanked by these attachment sites.	
attP-Bxb1	Integrase attachment site	Attachment site for Bxbi.	
attB-Bxb1	Integrase attachment site	Attachment site for Bxbi.	
attP-PhiC31	Integrase attachment site	Attachment site for PhiC31i.	
attB-PhiC31	Integrase attachment site	Attachment site for PhiC31i.	

Table 1: Genetic parts for use in Question 3. Parts may be used more than once.

4 Antibodies form a fundamental part of the mammalian immune system. The average human will develop over 10 billion antibodies over their life time, each capable of detecting a distinct epitope.

(a) Describe the structure and function of the different regions of antibodies. [20%]

(b) Describe the mechanisms that achieve such a high level of diversity. [20%]

(c) Current therapeutic use of antibodies requires the use of a single highly specific antibody that must be produced monoclonally. Describe the process by which monoclonal antibodies are produced, starting from a mouse inoculated with a specific antigen. [20%]

(d) Expressing antibodies in bacteria is a very attractive proposal but has so far proven incredibly difficult. What challenges can you identify with attempting this? [20%]

(e) A bacteriophage expression library has been made that expresses single-chain antibodies on its surface. It is screened against immobilised antigen and four clones each with high affinity for the antigen have been identified. In addition to high affinity, what other criteria have to be satisfied in order that these antibodies can be used as a therapy for humans? [20%]

END OF PAPER