ENGINEERING TRIPOS PART IIA

Monday May 7 2012 2.30 to 4

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.*

There are no attachments

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS
CUED approved calculator allowed

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

Final version

1 Glucose is the preferred sugar carbon source for the bacteria *E. coli*. However, other sugars can be used when available. Utilization of the sugar lactose is controlled by the *lac* operon. The *lac* operon encodes three genes, two of which are crucial for lactose metabolism. The *lacZ* gene encodes beta-galactosidase, an enzyme with two activities: to convert lactose to allolactose and to cleave lactose and allolactose into glucose and galactose. The *lacY* gene encodes a protein that transports lactose into the cell.

Separate from the operon, the *lacI* gene encodes a repressor protein and is constitutively (constantly) transcribed and translated. The steady state level of the repressor protein is sufficient to restrict transcription of the *lac* operon in the absence of lactose by binding to upstream "operator" sites that flank the "promoter" site. This interferes with the binding of RNA polymerase to promoter DNA and hence interferes with the initiation of transcription.

If lactose becomes available, it can diffuse into the bacterial cell at a low rate. The lactose can be converted to allolactose, where allolactose binds the repressor and releases it from binding the operator sites, allowing transcription of the operon. Transcription of the *lac* operon can be enhanced by the abundance of cAMP, a signaling molecule the concentration of which is inversely proportional to glucose concentration. cAMP binds to a protein that stabilizes the binding of RNA polymerase to the promoter and this in turn enhances the rate of transcription.

(a) How tightly regulated is transcription of the *lac* operon in the absence of lactose and why? [20%]

(b) Positive feedback is a critical mechanism for control of the *lac* operon. Discuss one positive feedback mechanism and any mutations that might affect this mechanism of control. [20%]

(c) Negative feedback is also a critical mechanism for control of the *lac* operon. Discuss one negative feedback mechanism and any mutations that might affect this mechanism of control. [20%]

(d) IPTG is a small molecule that can induce the *lac* operon but does not act as a substrate for the lacZ enzyme. How does this property affect the control of the *lac* operon? [20%]

(cont.

(e) Design and describe a **simple** inducible system for protein production based on control of the *lac* operon. [20%]

2 (a) Give an example of an activated carrier molecule. How is it generated? What is the biochemical purpose of such molecules? Give an example of their use. [30%]

(b) We wish to increase the production of a commercially valuable metabolite that is produced by a biosynthetic pathway.

(i) Give two examples of how the pathway can be manipulated to satisfy this goal. [20%]

(ii) How would you determine which step(s) in the pathway to manipulate? [20%]

(c) Give an example of metabolic engineering, explaining the goal and the strategies used to achieve that goal. [30%]

(TURN OVER

3 (a) When aligning two DNA sequences using dynamic programming, score is gained by aligning identical bases and lost when residues are misaligned. In addition, a penalty is paid for each gap that has to be introduced into the alignment. Explain what would happen if there were no gap penalty. [10%]

(b) Two amino-acid sequence alignments A and B are shown below. Each has the same pair of sequences being aligned. The same number of residues is perfectly aligned in both and there is the same number of gaps. Both alignments are optimal under the scoring system used, which differs only in the way that gaps are penalised.

Alignment A:

```
Sequence 1: APPRQP-RIATELY
              |||| |    ||
Sequence 2: --PRQPER----LY
```

Alignment B:

```
Sequence 1: APPRQP-RIATELY
             | ||| |    ||
Sequence 2: -P-RQPER----LY
```

(i) Explain in terms of gap penalties the difference between the two alignments. [20%]

(ii) Explain which alignment is biologically more realistic and why? [10%]

(c) A blood sample is taken from an individual harbouring a known genetic disease. PCR is used to amplify the final exon of the disease gene, as it is known to commonly harbour the mutation causing the disease. The PCR products are sequenced. Within approximately 20,000 sequencing reads, two versions of the DNA sequence are found: 9900 instances of Type A, and 10050 instances of Type B. An alignment of the two sequence types is shown:

```
Type A: AACTGCAGGATGACGGACGCAGGATCGATAGC
        ||||||||||| ||||||||||||||||||||
Type B: AACTGCAGGATCACGGACGCAGGATCGATAGC
```

(cont.

(i) Explain why two versions of the gene sequence are observed. [10%]

(ii) Using the codon table in Fig. 1, provide three possible molecular explanations for the disease. [30%]

(iii) Another individual with the same genetic disease was sequenced in the same way, but this time only the Type B variant was found. What is a possible explanation for this observation, and what does this imply about the number of copies of the mutation that are needed to cause the disease? [20%]

UUU F	UCU S	UAU Y	UGU C
UUC F	UCC S	UAC Y	UGC C
UUA L	UCA S	UAA *	UGA *
UUG L	UCG S	UAG *	UGG W
CUU L	CCU P	CAU H	CGU R
CUC L	CCC P	CAC H	CGC R
CUA L	CCA P	CAA Q	CGA R
CUG L	CCG P	CAG Q	CGG R
AUU I	ACU T	AAU N	AGU S
AUC I	ACC T	AAC N	AGC S
AUA I	ACA T	AAA K	AGA R
AUG M	ACG T	AAG K	AGG R
GUU V	GCU A	GAU D	GGU G
GUC V	GCC A	GAC D	GGC G
GUA V	GCA A	GAA E	GGA G
GUG V	GCG A	GAG E	GGG G

Fig. 1: *Codon table.* Each codon is followed by the amino-acid that it codes for.

4 A pure yeast strain was spread onto agar nutrient plates and incubated until colonies appeared. Among the 40,000 white colonies observed, a single red colony was found that harboured a mutant gene. The protein generated by this mutant gene was shown to encode an enzyme involved in adenine synthesis. This mutant, called ade2, blocks the adenine synthesis pathway and causes the accumulation of a red metabolic intermediate produced by the preceding enzyme in the pathway. Analysis of the ade2 mutant protein showed that it was truncated compared to the functional protein prepared from the white colonies.

(a) What kinds of gene mutations could lead to truncation of the mutant protein? [20%]

(b) The cells from the red ade2 mutant colony were treated with a chemical mutagen, to induce an elevated rate of mutation, and again spread on nutrient agar plates to form colonies. This time, among 40,000 red colonies, ten white ones were found. It is extremely unlikely that the original ade2 mutation has been reversed. Given this, provide a possible explanation for the appearance of the ten white colonies. [30%]

(c) The ten white mutant colonies from (b) were tested to see whether they could synthesis adenine: nine could not, but one could. DNA sequencing showed that the ade2 mutation was still present throughout this latter colony so somehow the effect of the original mutation had been suppressed. Provide a possible explanation for this observation. [50%]

END OF PAPER