CRiB

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.*

Answers to questions in each section should be tied together and handed in separately.

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%]

The control must be a little leaky in order to produce enough lacZ gene product to convert lactose to allolactose so that the operon can be de-repressed (the allolactose binds to the repressor and inactivates it)

(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%]

The production of the lacY gene product transports lactose into the cell producing a higher concentration of lactose than by diffusion alone. This will accelerate the de-repression of the operon. Any mutation that disruptions the production of lacY protein (e.g. a stop codon or frameshift mutation early in the gene) will eliminate the positive

feedback loop. N.B. a promoter mutation would also affect LacZ and so would not be an appropriate answer.

(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%]

By default the inducer of the lac operon, allolactose, will be cleaved into glucose and galactose by beta-galactosidase. High concentrations of glucose prevents maximal expression of the operon by causing low concentrations of cAMP. This in turn means there is no CRP-mediated stabilisation of RNA polymerase at the promoter. Mutations: various answers possible e.g. mutations in CRP that do not respond to cAMP and constitutively bind the promoter.

(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%]

As IPTG is not degraded, it can continuously induce the operon. As it is not degraded into glucose, there is no negative feedback on the operon.

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

Various answers possible for instance: the lac operon promoter drives expression of the gene of interest. In the absence of induction this is kept off by a constitutive promoter driving expression of lacI gene which codes for the lac repressor. Addition of IPTG will then induce the production of the protein of interest by de-repression.

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%]

ATP, NADH, NADPH.
Oxidative phosphorylation produces ATP by coupling energetically favourable reactions to drive the phosphorylation of ADP to ATP.
The purpose is to drive otherwise energetically unfavourable reactions.
DNA synthesis consumes ATP, regeneration ADP in the process.

(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%]

Alter the expression levels of the enzymes; alter the catalytic activity of one or more of the enzymes (for instance an enzyme may operate on many substrates and so the rate constant could be improved for one particular substrate of interest).

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

Perform metabolic control analysis: for each step in the pathway, measure the incremental change in flux in the pathway associated with an incremental change in enzyme activity.

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

Artemisinic acid production in yeast; hydrocortisone production in yeast. Goal is to maximise specific yield of the product. Strategies are: to introduce exogenous enzymes; to increase the flux through the pathways leading to product; to divert flux away from unwanted by-products.

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%]

There would be a tendency to produce nonsense alignments in which all residues are perfectly aligned, but many gaps have been inserted to allow this perfect alignment.

(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%]

Alignment A uses affine gap penalties and Alignment B uses linear gap penalties. Under the affine system a penalty is paid for the first gap and a lower penalty for subsequent adjacent gaps. In the linear system the same penalty is paid for all gaps whether adjacent or not. Under the linear system the score is equally good whether the first residue in sequence 2 is aligned with the 2nd or 3rd residue in sequence 1. In contrast, under the affine system, aligning residue 1 of sequence 2 to the 2nd residue of sequence 1 would incur a larger overall penalty, as the larger gap open penalty would be paid twice.

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

Alignment A is better because the gaps are more consolidated. This reflects the observation that, as proteins evolve, locations that can tolerate changes in length often change by more than one residue.

(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
```

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

Humans have two copies of each gene, except on the sex chromosomes. One is probably the mutant sequence and the other normal.

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

Depending on the (unspecified) reading frame:
G->C mutation: stop codon read-through causes disease.
C->G mutation: premature stop codon causes disease.
Either of above mutations cause amino-acid change causing disease.

(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%]

The type B variant is the mutant allele and the individual is homozygous for it. The patient in (c)i is heterozygous (carries one mutation and one normal copy of the gene) and so it seems that having a single mutant copy of the gene is sufficient to cause the disease. (Formally, it is also possible that both Type A and Type B represent mutations and that a single copy of the normal gene would in fact protect against disease).

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%]

A nonsense mutation, in which an amino-acid encoding codon is replaced with a stop codon, or a frameshift mutation in which the translation reading frame is altered leading to premature translation termination when a stop codon is reached.

(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%]

They harbour mutations in steps earlier in the pathway to ade2 so that the red metabolic intermediate no longer accumulates.

(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%]

The ade2 mutation could be a nonsense mutation, and the suppressing mutation could be a tRNA gene mutation that creates a tRNA that recognises stop codons and inserts an amino-acid during translation. This would mean that, instead of the stop codon terminating protein translation of an ade2 mutant transcript, an amino acid is inserted and the full length protein can be made, so rescuing the original defect.

END OF PAPER