Version GM/Final

EGT2
ENGINEERING TRIPOS PART IIA

Monday 28 April 2014 9.30 to 11.00

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 coversheet.

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed
Supplementary page: one copy of Figure 3 (Question 4)

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

1 (a) (i) Describe the structure of DNA. [30%]

(ii) Draw the generic monomers for RNA and proteins. [10%]

(iii) Show how two adjacent monomers are bonded together in RNA, and in proteins. [15%]

(b) It has been found helpful to describe protein structures and protein complexes in terms of a hierarchy of different structural forms.

(i) Name and describe the different levels of protein structure hierarchy and state what the most important bonding types are for the first two levels. [30%]

(ii) The amino-acid side chains confer the structural and functional properties of proteins: describe three distinct types of property of amino acid side chains. [15%]

(cont.

2 Limonene is a monoterpene with a strong orange smell found in citrus fruits and is used as a renewable solvent in cleaning products. Derivatives, such as perillyl alcohol, have medicinal uses. A recent paper describes the production of limonene and perillyl alcohol in *E. coli* (See Figure 1 below).

(a) In this study, the enzyme that produces limonene from GPP comes from the grand fir (*Abies grandis*):

(i) How might the authors have found this enzyme? Assume that no reports have been published on this exact enzyme. [15%]

(ii) How might this exogenous enzyme be expressed in *E. coli*? [15%]

(b) IPP (along with other compounds in the pathway) is an important metabolite for many processes in *E. coli*.

(i) What potential problems would this raise for the design process? [20%]

(ii) How could these problems be solved? [20%]

(c) You think the authors may not have expressed all the enzymes in the pathways at the optimal levels. How would you go about determining which enzymes would be the best targets for changes in expression level to optimize the pathway? [30%]

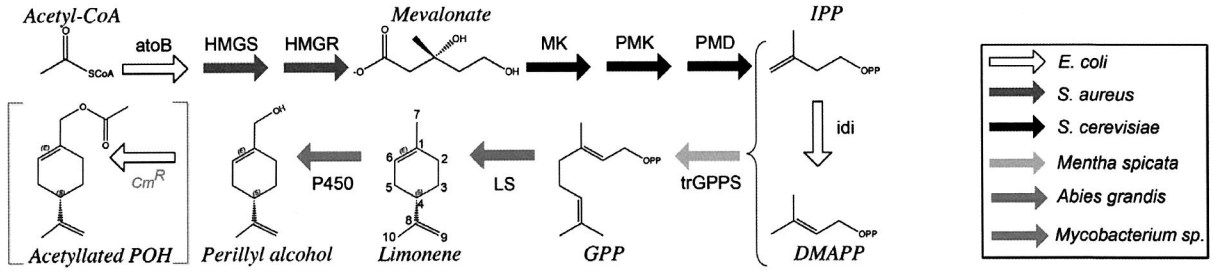

Figure 1: Production of limonene and perillyl alcohol in *E. coli*

(TURN OVER

3 (a) (i) Describe (with a diagram if appropriate) the Sanger dideoxy method of DNA sequencing. [20%]

(ii) What limits the length of sequence that can be read? [5%]

(b) Most current "next generation" sequencing methods require multiple copies of a template molecule for the actual sequencing step. Describe a method by which these copies can be made. [15%]

(c) Some of the latest sequencing technologies, such as the Pacific Biosciences "SMRT" (single molecule real-time) system operate directly on single DNA molecules. What are the potential advantages of truly single-molecule methods, compared to methods that require multiple copies of the template molecule? [15%]

(d) In finding the best global alignment between two DNA sequences, the overall score of the finished alignment is calculated from a similarity matrix which assigns a positive score to each pair of aligned bases which are identical, a lower or negative score to each pair of aligned bases which are not identical, and a negative score (a *gap penalty*) for gaps which are introduced into either sequence.

The following is a typical similarity matrix giving scores for matched and mis-matched pairs of bases:

	Base in first sequence:	A	G	C	T
Base in second sequence:					
A		10	-1	-3	-4
G		-1	7	-5	-3
C		-3	-5	8	0
T		-4	-3	0	9

Using the above similarity matrix, and a linear gap penalty of -5 per nucleotide position, use dynamic programming to prove that the best global alignment between the two sequences `ATAGC` and `ATGC` is:

```
ATAGC
AT-GC
```

[30%]

(e) In the similarity matrix above, the scores on the diagonal (for matched bases) range from 7 to 10. Likewise, the scores for non-matching (non-diagonal) pairs range from -5 to 0. Give possible reasons for this variation amongst the diagonal values and amongst the non-diagonal values. [15%]

(cont.

4 The formation of bacterial biofilms is important to disease pathogenesis. The signalling processes that underlie the formation of biofilms may be understood in terms of 2 dimensional pattern formation. The three plasmids shown in Figure 2 (over page) were made to build up a genetic circuit that is based on acyl-homoserine lactone (AHL) signalling between individual bacteria.

(a) Based on the plasmid diagrams provided in Figure 2, fill in the six promotor and gene names in Figure 3 (provided on the supplementary sheet), as a first step in constructing a genetic circuit diagram. [10%]

(b) On the same copy of Figure 3, complete the genetic circuit diagram by using the information shown in Table 2 (over page), noting that the output of the circuit is the GFP gene product, which is fluorescent. [20%]

(c) A bacterial culture harbouring all the above plasmids is spread on an agar dish onto the centre of which is placed a small paper disc saturated with tetracycline. Taking into account the relationships between component activities shown in Table 1 (below), sketch the resulting pattern of bacteria expressing GFP. [30%]

AHL	CI	LacI[1]	LacI[2]	GFP
++	++	-	++	-
+	+	-	+	+
-	-	++	-	-

Table 1: Relationship between different components of the genetic circuit when in different states.

(d) Explain how the genetic circuit gives rise to the pattern that is shown in your answer to (c). [40%]

(TURN OVER for Figure 2 and Table 2 of Question 4

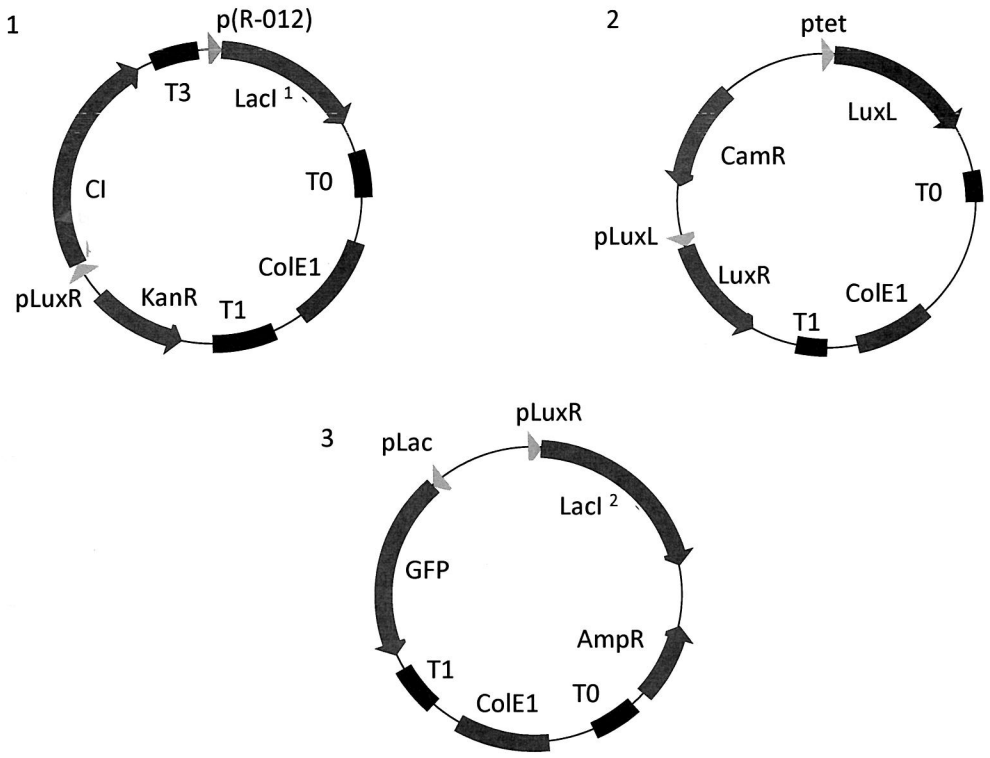

Figure 2: Three plasmids, which together can be used to investigate biofilm pattern formation in bacteria.

(cont.

pLuxL	Promoter activated by LuxL gene product
pLuxR	Promoter activated by LuxR gene product
pLac	Promoter repressible by lacI gene product
Ptet	Tetracycline activated promoter
p(R-012)	Promoter repressible by CI gene product
$lacI^1$	Wild type Lac repressor gene. In the absence of Lactose or IPTG its gene product represses pLac.
$lacI^2$	Mutant Lac repressor gene. Its gene product binds the pLac promoter less efficiently than $lacI^1$. Thus low levels of $lacI^2$ will not repress pLac.
LuxL	Gene product activates the pLuxL promoter.
LuxR	Gene product activates the LuxR promoter.
CI	Gene product represses p(R-012) promoter activity.
AHL	Acyl-Homoserine Lactone, a molecular species responsible for inter-cellular communication
GFP	Green Fluorescent Protein gene.
KanR	Kanamycin resistance gene.
AmpR	Ampicillin resistance gene.
CamR	Chloramphenicol resistance gene.
ColE1	*E.coli* origin of replication

Table 2: list of parts used in the plasmids, and their functions.

END OF PAPER

THIS PAGE IS BLANK

Version GM/Final Candidate Number:

EGT2
ENGINEERING TRIPOS PART IIA
Monday 28 April 2014, Module 3G1, Question 4

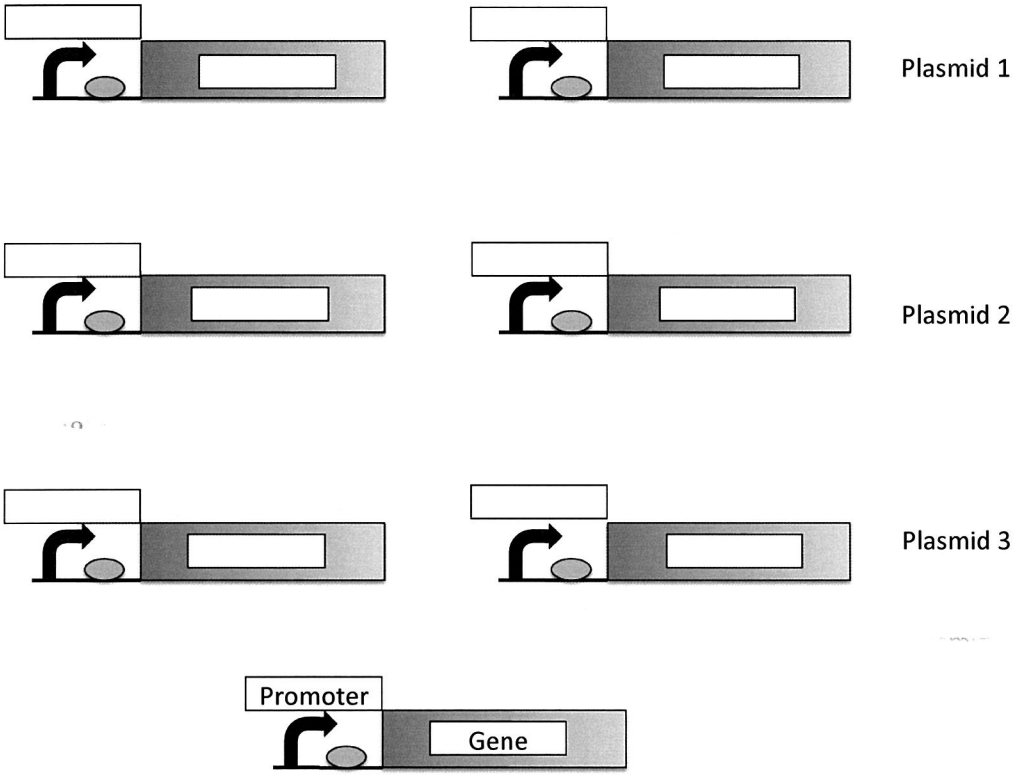

Figure 3: Genetic circuit diagram. Based on the three plasmids shown in Figure 2, fill out the promotor and gene names in the blank boxes, as shown in the example at the bottom.