ENGINEERING TRIPOS PART IIA

Friday 3 May 2013   2.00 to 3.30

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.*

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

There are no attachments

STATIONERY REQUIREMENTS   SPECIAL REQUIREMENTS
Single-sided script paper   CUED approved calculator allowed

Final version

1   Making proteins requires substantial energy so the bacterial cell regulates the synthesis of certain proteins such that they are made only when needed. The tryptophan (*Trp*) operon (Fig. 1) consists of a set of genes that encode enzymes required to make the amino acid tryptophan. The *Trp* operon is a negative feedback system mediated by a repressor protein. The *trpR* gene encodes the repressor and is constantly expressed at a low level. The resulting repressor protein monomers associate into a tetramer. In the presence of tryptophan, the tetramer cooperatively binds two tryptophan molecules to form an active repressor complex, i.e. binding the first tryptophan facilitates the binding of the second tryptophan to form the active complex. The active repressor complex binds the operator region of the operon, and blocks RNA polymerase from carrying out transcription. Note that each of these interactions is reversible.

Fig. 1: The *Trp* operon.
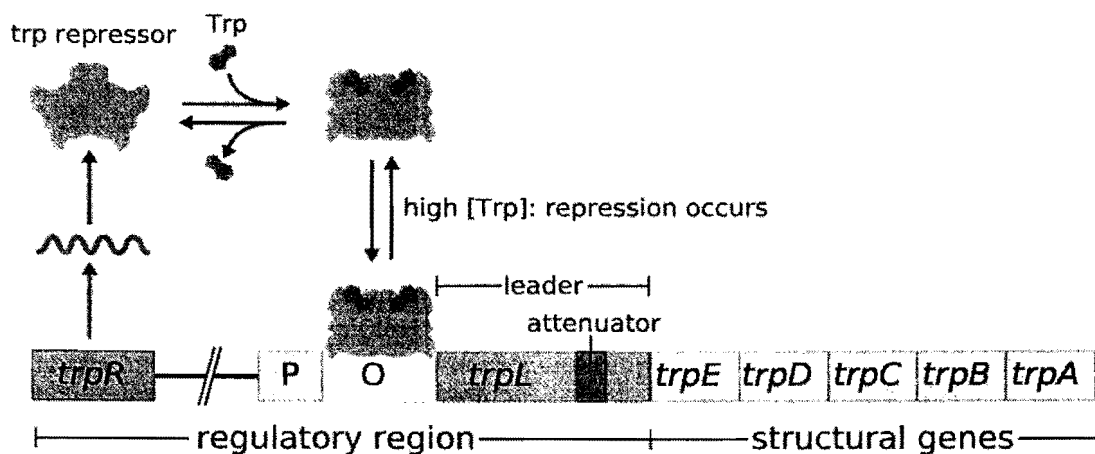
(a)   Suggest a simple way to modify the *Trp* operon to easily measure the level of repression.   [10%]

Insert a reporter gene (e.g. coding for green fluorescent protein) into the operon. A good test would have 3 separate constructs with the reporter placed at the beginning, somewhere in the middle and the end of the operon to control for position effects.

Final version

(b)   Recall that proteins have a half-life, i.e. that they will be degraded at a specific rate. Assuming the translation rate is constant and tryptophan abundance is high, discuss how increasing the rate of transcription of the *trpR* gene might affect regulation of the operon.   [25%]

Increased transcription of trpR in this case would result in a higher abundance of the repressor protein. It probably won't make a lot of difference in high tryptophan abundance, but in moderate to low tryptophan abundance, there will be more tryptophan binding sites available so making active repressor would be of lower probability so the operon may be de-repressed more easily. However, as the binding of tryptophan is cooperative, it may make less of a difference than if tryptophan binding were completely independent events on the repressor tetramer.

(c)   A mutant *trpR* gene is integrated downstream of the functional *trpR* gene and is co-transcribed and translated equivalently such that both genes produce the same amount of monomer repressor. The monomer protein encoded by the mutant *trpR* gene is equally capable of associating with both functional and mutant monomers to form tetramers. In a high abundance of tryptophan, the operon now shows over ten-fold greater expression.

Devise a simple model of how the addition of the mutant *trpR* gene might affect the regulation of the *Trp* operon, and use the model to predict the quantitative impact of the mutation on *Trp* operon expression.   [35%]

In this case, the operon shows over a ten-fold greater expression suggesting that the mutant monomer is "poisoning" any tetramer that it is part of. A simple quantitative model would be that the inclusion of even a single mutant monomer makes the tetramer non-functional so only one in 16 tetramers would be functional. This may lead to significant de-repression because of a greater than 10-fold decrease in functional repressors (assume linear relationship between repressor number and repression).

(d)   The expression of the *Trp* operon is further regulated by attenuation of transcription (Fig. 2 over page). Since translation is initiated as soon as the ribosome binding site becomes available on the mRNA, the speed of the moving ribosome determines the secondary structure of the transcribed mRNA. In the *Trp* operon, the

Final version

leader sequence of the *trpL* gene contains two tryptophan codons so that, when tryptophan is abundant, the ribosome can quickly translate the sequence and move on, whereas when tryptophan is scarce, the ribosomes stall. In high tryptophan abundance, regions 1 and 2 of the mRNA are covered so that regions 3 and 4 form a stem-loop structure forcing the RNA polymerase to fall off, terminating transcription. In low tryptophan abundance, the ribosome stalls at region 1, allowing the more favourable stem-loop structure between regions 2 and 3 to form, so the RNA polymerase can continue transcription.

(i) How might the attenuation of the *Trp* operon be affected if one of the tryptophan codons were replaced by a GGG glycine codon? [15%]

The replacement of a tryptophan codon with a glycine codon would result in less ribosome stalling for a given tryptophan concentration. This would have the effect of allowing the 3-4 stem loop structure to form in lower tryptophan abundance and hence block the operon from being transcribed/translated compared to the normal two tryptophan codon situation.

(ii) How might the attenuation of the *Trp* operon be affected if more DNA were inserted between regions 1 and 2? [15%]

This may have the effect of appearing to "slow the ribosome" allowing the 2-3 stem loop structure to form more efficiently thus permitting the transcription/translation of the operon in higher tryptophan abundance compared to the normal region1-2 spacing. Conceivably, the sequence may generate other stem loop structures with region 2 so the 3-4 stem loop would always form and the operon would never be switched on.

Final version

High level of tryptophan
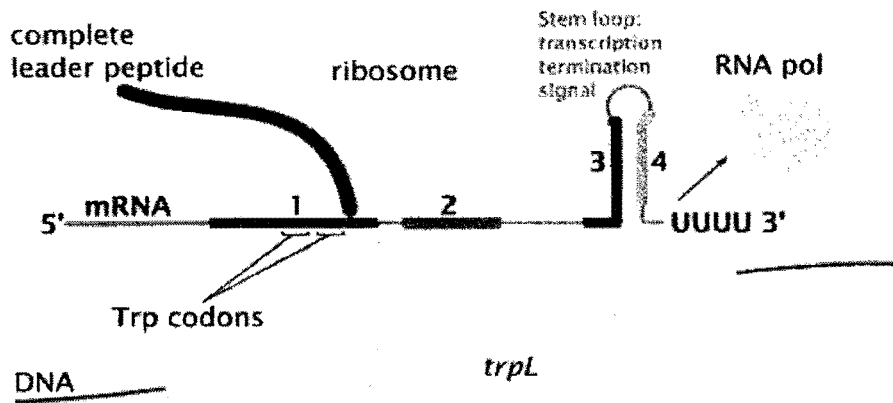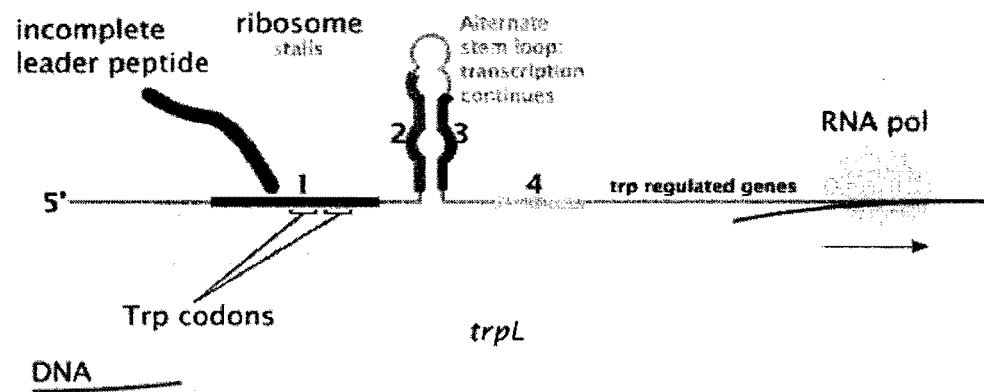

Low level of tryptophan

Fig 2: Transcriptional attenuation in the *Trp* operon. Upper: RNA polymerase dissociates from *trpL* when tryptophan is abundant. Lower: RNA polymerase remains bound and continues transcribing the operon when tryptophan is scarce.

Final version

2 Muconic acid can be used as a precursor for plastics such as polyurethane or polyethylene terephthalate (PET). The following pathway can be used for the production of muconic acid in yeast:
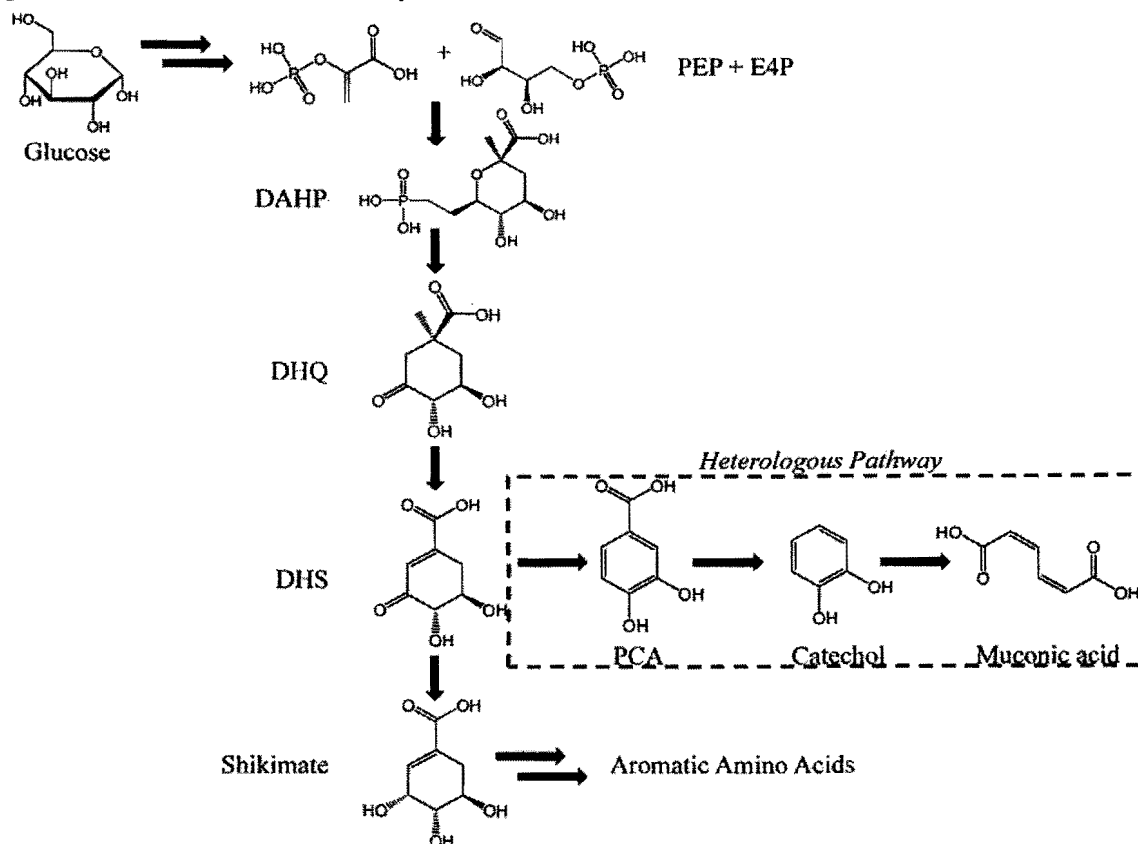

Fig. 3: Pathway for muconic acid production in yeast.

(a) Given that wild type yeast is not capable of producing muconic acid and that you want to maximize production of muconic acid, explain your general strategy for achieving that goal.   [20%]

Find and introduce enzymes from other organisms that catalyze the reactions in the heterologous pathway, maximize flux to the precursor DHS, and minimize flux through the byproduct shikimate pathway.

Final version

(b)   How would you determine which enzymes in the endogenous pathway to manipulate?   [20%]

Perform metabolic control analysis: Manipulate levels of each enzyme of the pathway by, for example, gene dosage, and measure flux through the pathway. Plot flux vs. enzyme level and the slope of the curve gives the control coefficient. Pick enzymes with high control coefficients for manipulation.

(c)   Give examples of strategies used by metabolic engineers for addressing problems like this one, and explain how those strategies might be applied to the above pathway.   [60%]

The Keasling group produced artemisinic acid in yeast by introducing artemisia annua genes into yeast. They maximized flux to precursors by introducing additional copies of upstream genes and upregulating a transcription factor responsible for the expression of the pathway. They reduced flux to byproducts by replacing the promoter for the enzyme at that step with a repressible one. In the muconic acid pathway, then, you would introduce the foreign genes, try to determine the endogenous regulation of the upstream steps with the goal of finding a global transcription factor for the pathway and upregulating it, introduce additional copies of genes with high control coefficients, and downregulate the shikimate pathway by mutating genes or introducing repression.

3   (a)   What is the molecular basis for information transmission between generations in known living systems?   [10%]

The nucleic acids DNA and RNA use four bases to store information digitally. Complementary base pairing allows molecules to be copied and a copy is passed to each member of the next generation.

(b)   One type of information transmitted between generations is the ability to code for proteins. Describe two features of many living systems that help to ensure that protein sequences do not change between generations.   [10%]

Proofreading by DNA polymerase; redundancy of genetic code

Final version

(c)   Outline the important features of the mechanism by which DNA polymerase replicates a double-stranded DNA plasmid.   [40%]

helicase to unwind DNA; RNA primase for DNA polymerase; polymerisation in 5'->3' direction with proof-reading; continuous replication of leading strand; discontinuous replication and ligation of lagging strand.

(d)   It is required to design a new vector for expression of heterologous proteins in *E. coli*.   Outline the vector design issues you would consider for successful expression and purification of target proteins, discussing compromises where relevant.   [40%]

High vs low copy number replication origin: high improves likely yield but greater chance of problems with toxicity. Inducible promoter: lower chance of toxicity as expression limited until induce but greater costs due to need for inducer. Purification: fusion protein with affinity tag (e.g. antigen for antibody, GST etc), or Histidine hexamer for purification with nickel resin.

4   (a)   Viruses are known to replicate their genomes with low fidelity and generate very large numbers of viral particles during the course of an infection.
    (i)   Explain the advantages and disadvantages of poor fidelity replication.   [10%]

Positive: can explore more of available sequence space thus potentially evading host defences e.g. immunes response. Negative: may mutate essential function thus killing off that lineage.

    (ii)   Explain whether the advantages outweigh the disadvantages, and why.   [10%]

Positive is more significant than negative as large numbers of viral particles means that it doesn't matter if some small fraction are unviable.

(b)   A type of virus known as a retrovirus characteristically has long terminal repeats at each end of the linear genome. Suppose we are interested in sequencing a sample of a particular species of retrovirus in order to measure the sequence variation present in the population. This species has imperfect long terminal repeats of 400 bases in length.

Final version

(i)   Explain what type of sequencing method you would use and why.   [10%]

Conventional Sanger sequencing as it has reads 100s of bases long, often >400 bases. This will increase the chance that the sequencing reads can be unambiguously aligned to one or other terminal repeat.

(ii)   Outline the method by which you would identify the locations at which variation is occurring.   [20%]

A DNA sequence alignment algorithm (dynamic programming) would optimally align sequencing reads to the reference sequence for the species. The positions of mismatches, insertions and deletions could indicate sequence variations or possibly sequencing errors. Multiple observations of the same variation(s) increases the chance that they are true sequence variations and not sequencing errors.

(c)   Complementary base pairing between nucleic acids can be used to build structures on the nanometre length scale. For instance one might construct a "cube", admittedly with floppy edges.

(i)   Explain how the corner of a cube (i.e. three edges joined at a corner) could be constructed.   [10%]

Three single stranded molecules would be designed that can base pair with each other in such a way that the start of each molecule is on one edge and the end on another:
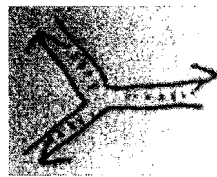
(ii)   Design RNA sequences that would self-assemble into such a corner.   [30%]

RNA_1:	AAAAAAAAAACCCCCCCCCC
RNA_2:	GGGGGGGGGGGAGAGAGAGAG
RNA_3:	CUCUCUCUCUUUUUUUUUUUUU

Final version

(iii)   Give two ways in which such RNA sequences might be produced.   [10%]

Chemical synthesis, or transcription of synthetic genes in a host and subsequent purification.

END OF PAPER