Version SB/final

EGT2
ENGINEERING TRIPOS PART IIA

---

Friday 9 May 2025 9:30 to 11.10

---

**Module 3G1**

**MOLECULAR BIOENGINEERING I**

*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 at the start of the exam.

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

## You may not remove any stationery from the Examination Room.

1 The textile industry is one of the most polluting in the world, leading to recent innovations towards sustainable, bio-based pigment production and dyeing processes. To increase the variety of pigments that can be produced, scientists are sequencing the genomes of more organisms, collecting metagenomic sequences from different environments and trawling existing databases of gene sequences to find new and improved pigment production pathways. Once discovered, these can be engineered for optimal expression of the pigment chemicals in scalable host expression systems.

(a) A new family of enzymes has been discovered in fungi that can produce many coloured pigment chemicals from a single precursor molecule.

(i) An enzyme that catalyzes the transformation of a blue chemical into a red chemical has evolved from an ancestral enzyme that originally produced a green chemical. Describe where you would expect the mutations to arise within the enzyme sequence and how these mutations could become fixed at high prevalence in a fungal population. [15%]

Crib: This would require a mutation or a series of potentiating mutations that produce the phenotype of red pigment production (most likely from a mutation in the enzyme active site). The phenotype then needs to be maintained in the population by selective pressure (an evolutionary force that causes a particular phenotype to be more favorable in certain environmental conditions). It is possible that mutations can become fixed without conferring an advantage, but it is much less likely. If this is suggested, a specific mechanism should be given e.g. population bottleneck, genetic linkage to a different favourable trait.

(ii) Could you use directed evolution to accelerate the production of a specific pigment color in bacteria, similar to how bacteria developed the ability to grow aerobically on citrate in the long-term Lenski experiment? [5%]

Crib: Unlikely. This would require applying a strong selective pressure under experimental conditions on an ecological timeline. Typically this requires a phenotype that confers a very strong survival advantage such as ability to resist antibiotics, use a specific carbon source or survive harsh conditions.

(b) You have identified 114 putative members of the new family of enzymes using bioinformatics techniques on metagenomic datasets, and you generate a DNA library for expression and testing. As 114 is a large number of novel enzymes to investigate, you consider incorporating automation into your experimental plan.

(i) How would you obtain the DNA for these enzyme variants? [5%]



(cont.

Crib: It would be very challenging, if not impossible, to obtain the original organisms or samples in order to PCR out the sequence, given the search methods used metagenomic datasets. De novo synthesis would be much more rapid and efficient and would enable you to codon optimise, domesticate and otherwise improve the sequences for expression in the chosen host organism.

(ii) Once you have the enzyme DNA, list the steps involved in designing and building your DNA library, then inserting it into a host *Escherichia coli* bacteria. For each step, comment on how easy it would be to automate the experimental techniques. [15%]

Crib: The first part of the answer should cover the DNA assembly process and the candidate could describe any valid technique: Gibson, Type IIS, overlap PCR or other. The main steps will involve obtaining the plasmid backbone, restriction digest and ligation for Type IIS, exonuclease digestion, polymerase gap-filling and ligation for Gibson assembly.

Second part should be transformation into competent cells via heat shock or electroporation.

Steps involving solely liquid handling and incubation e.g. digestion and ligation are very amenable to automation. Gel electrophoresis, extraction, electroporation are hard to automate. Heat shock transformation is intermediate difficulty.

(iii) What are three advantages and three disadvantages of incorporating automation into the design-build-test-learn cycle for this experiment? [10%]

Crib: Advantages of Automation in the Design-Build-Test-Learn (DBTL) Cycle: increased throughput for routine experiments; enhanced precision and accuracy (reduced human error); time efficiency (freeing up researchers to focus on tasks for which humans as better); scaling of experiments from small to large datasets; integration with software for real-time data acquisition and analysis.

Disadvantages of Automation in the DBTL Cycle: high cost; high learning curve and investment needed in skills and personnel, high maintenance costs and issues with downtime; reduced flexibility and rapid adaptation "on the fly"; troubleshooting can be complex and require significant expertise; depends on standardized protocols which may limit the ability to explore diverse approaches

(c) In order to verify the sequences, you will sequence all the plasmids in your library using Sanger sequencing. For this, you will need to design DNA primers.

(i) What are primers and what is their function? [5%]

Crib: Short, single-stranded sequences of DNA or RNA (typically 18–30 nucleotides

long) that serve as starting points for DNA synthesis. Bind to a complementary sequence on the target DNA, providing a free 3'-hydroxyl group for DNA polymerase to extend and synthesize the new DNA strand during PCR, sequencing or DNA replication.

(ii) You receive the sequencing results back and realise that not all of the DNA sequences match the original sequence from the bioinformatics analysis. Why might this be? [5%]

Crib: DNA synthesis and sequencing can introduce errors but a likely cause is mutations introduced during a PCR step because DNA polymerases do not have 100% fidelity. This is why we use high fidelity polymerases when amplifying DNA for cloning.

(iii) What types of mutations can arise in DNA that encodes proteins, and how might these mutations affect the protein sequence and function? [10%]

Crib:

Point mutation – a change in one base in the DNA sequence (can be silent or lead to a single amino acid mutation which may or may not affect the protein function)

Substitution – when one or more bases in the sequence is replaced by the same number of bases (can be silent or lead to a single amino acid mutation which may or may not affect the protein function)

Insertion – when a base is added to the sequence (results in frameshift, will almost certainly result in a nonfunctional protein)

Deletion – when a base is deleted from the sequence (results in frameshift, will almost certainly result in a nonfunctional protein)

More complex mutations e.g. copy number variation, gene duplication, inversion are also valid.

(d) You decide to express the enzymes combinatorially in a cell-free TX-TL system to prototype different pigment synthesis pathways, using linear polymerase chain reaction (PCR) amplicons as the input DNA.

(i) What approach would you take to assess the functionality of the enzymes in this assay? [5%]

Crib: Production of pigments with different and unknown colour is being measured, so colorimetry would be the most obvious readout. Ideally mentioning spectrophotomtetry or multi-wavelength imaging with filters to account for identifying a combination of different pigments being produced in the same pathway.



(cont.

(ii) Initially you do not get any output from your assay, but the cell-free TX-TL system is functional because you successfully express GFP from your plasmid DNA positive control. In your experimental assays, there are no bands of the expected size visible on SDS-PAGE and when you perform a reverse transcription quantitative PCR (RT-qPCR) there are no amplicons of the expected size. What is the most likely explanation and what are three ways you could fix this issue? [15%]

Crib: Degradation of the linear DNA. Solutions could include: cell strain that is deficient in exonucleases, "decoy" DNA containing chi sites to saturate RecBCD machinery, adding nuclease inhibitors such as gamS protein, using plasmid DNA instead, any other suggestion that removes the exposed and of the DNA.

(e) After much research and development, your new chemical production pathway is ready to be scaled up and deployed on-site at textile dye houses as a whole-cell biomanufacturing approach. It is crucial for the process to be environmentally sustainable, economically feasible, and safe. Describe three challenges that you might expect when scaling up bioprocessing from lab scale (1-2 L) to production scale (>500 L). [10%]

Crib: There are a variety of challenges related to maintaining process consistency, efficiency, and product quality. Major ones are: 1) ensuring that environmental conditions such as temperature, pH, dissolved oxygen, and mixing are uniform throughout the larger bioreactor, as differences in mass and heat transfer rates can lead to suboptimal or heterogeneous conditions. 2) Shear forces in larger bioreactors may also affect the viability of cells or the integrity of biological products, requiring optimization of impeller designs and agitation speeds. 3) Nutrient delivery and waste removal can be less efficient at larger scales, potentially impacting cell growth and productivity. 4) Equipment costs, operational complexities, and regulatory compliance become more prominent.

2 A new fast-growing eukaryotic microorganism called *Cambridgiensis novus* has been discovered which has the potential to be used as an alternative to the current "workhorses" of recombinant protein expression. You are asked to try expressing a novel anti-HER2 humanized monoclonal antibody called zatuzumab in *Cambridgiensis novus*, because zatuzumab has been challenging to express in existing yeast expression systems. HER2-positive cells are present in a number of human cancers, so this antibody has potential to be used as a component in a number of therapies.

(a) What potential biosafety and biosecurity concerns could arise from the development or use of this new organism? [10%]

Crib: Biosafety: Pathogenicity (whether the microorganism has the potential to cause disease in humans, animals, or plants), ability to persist and spread in the environment, antibiotic resistance, effective inactivation methods. Determine containment measures required, assume worst-case scenario.

Biosecurity: Evaluate whether the microorganism has the potential for dual-use (e.g., use as a biological weapon), follow national and international regulations governing the handling of novel organisms (e.g., the Cartagena Protocol, WHO Biosafety Manual). Consider access controls and secure storage.

(b) What are the advantages of expressing proteins in the following organisms?

(i) *Escherichia coli* bacteria [10%]

Crib: Should cover the majority of the following:

Very rapid growth (doubling time of 20 min); can produce large quantities of diverse types of recombinant protein in a short amount of time; cost-effective, simple media requirements and robust to different conditions (e.g. temperature, pH); can be scaled up from small laboratory volumes to industrial-scale fermenters; well-established protocols exist for transforming plasmid DNA.

Additional options: easy to engineer the genome, well-characterised genetics and metabolism, availability of genetic tools.

(ii) *Pichia pastoris* yeast [10%]

Crib: Should cover the majority of the following:

Rapid growth (but lower than *Escherichia coli*); high levels of recombinant protein (can approach gram-per-liter yields); cost-effective, simple media requirements and robust to different conditions (e.g. temperature, pH); can be scaled up from small laboratory volumes to industrial-scale fermenters; well-established protocols exist for transforming plasmid DNA; recombinant proteins can be secreted into the culture medium; capable of performing eukaryotic-specific post-translational modifications (might mention one or more of these e.g. glycosylation, phosphorylation, and disulfide bond formation).

Additional options: molecular machinery to assist protein folding, reduced endotoxin contamination (compared to *Escherichia coli*). Well-characterised genetics and metabolism, availability of genetic tools.

(c) You have been provided with a protein expression plasmid for zatuzumab and asked to test it for constitutive expression in *Cambridgiensis novus*. Describe the standard protocol for using a plasmid to constitutively express recombinant protein in a host microorganism. [10%]

Crib: Plasmid is introduced into the host microorganism through transformation using chemically competent cells and heat shock, or electrocompetent cells and electroporation. A selectable marker (e.g. antibiotic resistance, metabolite dependency) is used to identify successfully transformed cells. Positive colonies are screened for the presence of the recombinant plasmid by PCR or sequencing. The host cells. Verified colonies are then cultured in media containing antibiotics or other chemicals required to maintain the plasmid. The recombinant protein is harvested by lysing the cells or collecting secreted proteins from the culture medium. There is no need to discuss purification in this answer. Mention of inducing expression e.g. IPTG is incorrect as the questions is specifically about constitutive expression.

(d) You take samples of the *Cambridgiensis novus* and load them into an SDS-PAGE gel, but there is no detectable protein expression.

(i) What does SDS-PAGE allow you to measure, and how does it work? [10%]

Crib: SDS-PAGE allows one to measure protein molecular weight and amount. SDS denatures protein so it is a linear molecule and neutralises the amino acid charges. Polyacrylamide gel then sieves the protein based only on length/molecular weight. Protein is stained in order to to be visualised and position on the gel is compared to a protein ladder of known molecular weights.

(ii) One reason for lack of expression could be that the promoter in the plasmid is not compatible with the RNA polymerase or transcription factors in *Cambridgiensis novus*. Provide three other reasons why the plasmid you have been given might not lead to protein expression in this host. [10%]

Crib: incompatible codon optimisation, RBS is not functional with *Cambridgiensis novus* ribosome, other regulatory elements are not compatible e.g. terminator, 5' or 3' UTR. Other valid reasons: the protein needs chaperones that are not present in the cell line, secondary structures in the mRNA prevent expression.

(iii) In order to expand the genetic tools available to regulate protein expression in *Cambridgiensis novus*, describe an experiment to find and characterise novel promoters from this host. [10%]

Crib: Typically an experiment would combining bioinformatics, molecular biology, and reporter gene assays but the lectures did not cover much bioinformatics. Putative promoter regions could be identified using bioinformatics tools and will typically be upstream of annotated genes (identified using sequence alignment and motif discovery algorithms to find binding sites, TATAA boxes etc). Could also take an experimental approach and PCR amplify promoter regions up to 500 base pairs upstream of coding sequences from genomic DNA.

To test the promoters, they can be cloned into a reporter vector upstream of a reporter gene (e.g., GFP, luciferase, or beta-galactosidase). The recombinant plasmids are transformed into the host and expression of the reporter is measured under different environmental or physiological conditions. Techniques such as fluorescence microscopy, spectrophotometry, or chemiluminescence are used to quantify reporter activity. Additional tests such as qPCR can also be used to measure expression.

(e) After many months of optimisation, you overcome expression challenges and discover that *Cambridgiensis novus* allows for efficient protein secretion, reducing the cost of purification. This allows you to produce enough zatuzumab to begin characterising its function.

(i) What are the main features of humanised monoclonal antibodies? [10%]

Crib: Humanized monoclonal antibodies (mAbs) are engineered antibodies designed to minimize immunogenicity by replacing most of their non-human components with human immunoglobulin sequences, while retaining antigen specificity. The complementarity-determining regions (CDRs) responsible for antigen binding are

derived from a non-human species and grafted onto human framework regions, maintaining high specificity and affinity for the target while reducing the risk of anti-drug immune responses.

(ii) Zatuzumab binds to and inhibits the HER-2 receptor. How could it be engineered to be more effective as an anti-cancer therapeutic? [10%]

Crib: Protein engineering to enhance their specificity, stability and resistance to degradation. Increasing potency, and ability to recruit immune responses e.g glycoengineering to enhance binding. Strategies include: conjugating mAbs with cytotoxic drugs or radioactive isotopes for targeted delivery; carrying immunostimulatory payloads such as cytokines; engineering bispecific antibodies to simultaneously target tumor cells and immune cells like T cells.

(iii) What differences might be introduced by using a new expression host that might affect the therapeutic potential of zatuzumab? [10%]

Crib: Different glycosylation patters and post-translational modifications that reduce the binding affinity of the mAb for the HER-2 receptor and increase immunogenicity. Increased levels of endotoxins in some hosts can trigger severe inflammatory and immune responses in humans.

3 For the iGEM project, your team has designed a synthetic oscillator that uses a dual feedback design (linked positive and negative feedback loops).

(a) Draw this circuit in **SBOL** (Synthetic Biology Open Language) notation to indicate how each genetic part connects. Label the core components (promoters, ribosome binding sites, coding sequences, terminators) and specify which loops provide positive regulation and which loops provide negative regulation. [15%]

Crib: 10 marks for correct sketch and labels on each part. 5 marks if only correct sketch but poor labelling.
The arrow signs should show which is positive and which is the negative feedback. Additional labels are welcome.

(b) Briefly explain how you expect these two feedback loops to work together to generate sustained oscillations. [10%]

Crib: The answer should contain the idea that typically an oscillator has a core motif - a negative feedback with delay. the positive feedback provides the delay.

(c) The circuit is placed in a bacterial strain which undergoes binary fission to reproduce. If the carrier cell has a doubling time of 20 minutes, can growth-based dilution alone be used to achieve a synthetic oscillator with a period faster than 30 minutes? Provide a detailed explanation of why or why not. [10%]

Crib: Since the bacterial cell doubles every 20 mins, to reach half the value of steady state, system would be need 20 mins in each direction. For the oscillator to have a 30 min period, it needs to build up the protein and dilute it within 30 mins period, which seems too short for a growth and division dependent dilution process. So the answer should be no and should contain the a version of the above-mentioned explanation.

(d) Suppose we want to *further shorten* the oscillator period to around 15 minutes. What fundamental design changes (for example, modifications to the circuit design) might be necessary to accomplish this faster oscillatory behaviour? [10%]

Crib: We need to introduce degradation tags on each regulatory protein, so that they are removed much faster.

(e) Compare the resource demands (for ribosomes, RNA polymerases, and ATP) of the original oscillator (longer period) and the redesigned oscillator (shorter period). Which elements of the faster circuit design would likely change the metabolic burden imposed on the carrier cell, and why? [15%]



(cont.

Crib: The ribosome and RNAP demand would increase to increase the production rate for balancing the incresed removal rate. Ribosome and RNAP consume GTP and ATPs. Additionally the degradation process consumes ATP, so the ATP demand would significantly increase.
The plot should be a steady oscillation of the reporter production.

(f) Assume you have two bacterial strains: Strain A carrying the original oscillator and Strain B carrying the new, faster oscillator. You mix these two strains in a single culture to begin a growth experiment. How do you expect the ratio of Strain A to Strain B to change over time if they grow together, and why? Sketch a plot of this ratio over time, assuming unlimited resources. [10%]

Crib: The ration of the two strains will change over time, depending on the difference in their doubling time. Since strain A is likely to be much faster in doubling time, due to lower resource demand from the original oscillator, we expect it to take over the mixed population over time.
The plot of the relative ratio should exponentially increase over time, until the resources becoming limiting. The exponential growth factor is the difference between their growth rates.

(g) Describe how you would use a fluorescent protein reporter to quantify the fast oscillator's behavior in real time. Address the following points:

(i) **Design strategy:** How to adjust the circuit design to use the fluorescent reporter. [10%]

(ii) **Measurement strategy:** Time-lapse fluorescence microscopy or flow cytometry; frequency of data acquisition. [10%]

Crib: An additional ORF+promoter+RBS+terminator sequence needs to be incorporated into the circuit design. The ORF should correspond to the coding sequence for the fluorescence reporter. There should be an operator next to the promoter sequence, which is activated by either of the two regulatory proteins in the dual feedback circuit.
Timelape microscopy is needed as cytometry cannot track cells. Frequency should be at least 7.5 mins to ensure the oscillation period can be reliably estimated.

(h) You are comparing two fluorescent protein reporters: **RFP** (maturation time of ~40 minutes) and **YFP** (maturation time of ~4 minutes). Which reporter would be most suitable for tracking rapid oscillatory behavior, and why? [10%]

Crib: YFP. The maturation time is fast enough for tracking the fast oscillation.

4 A recently analysed novel organism has 200 genes connected by 300 regulatory edges. A single point mutation in the organism's RNA polymerase (RNAP) was identified, which can double the promoter binding rate, potentially altering the network's behaviour. Answer the following questions about the implications and functioning of the network's topology and the effect of the mutated RNAP.

(a) You observe that 35 out of the 300 edges are self-arrows (autoregulation). Are these self-arrows more frequent than expected under a random model? Briefly explain the potential biological meaning of this frequency. [20%]

Crib: The answer should compare the observed number of self arrows (35) with the expected number or frequency of self arrows in random networks, as shown in the class, which should be 1.5 (300/200). The answer should include a calculation of the expected frequency and the underlying logic.

Since the self-arrows appear much more frequently than expected by chance, it suggests that autoregulation is a motif that has been evolutionarily selected in such networks.

(b) Of the 35 self-arrows, 30 represent negative autoregulation. What roles and advantages might negative autoregulation provide, in terms of the organism's fitness or adaptation to environmental changes? [10%]

Crib: Negative autoregulation can help with fast response and stablity towards fluctuations in input parameters.

(c) Sketch a rate plot comparing a simple regulatory mechanism (no autoregulation) to negative autoregulation. Label the axes and indicate the steady state value for each. [15%]

Crib: Rate plot should include amount of the species in the x axis and the rate of production and degradation in the y axis. The degradation curves are same for both cases. The rate of production for no autoregulation is a straight line with slope = 0. In case of the autoregulatory circuit, the production plot is hill curve for repression.

The point of intersections between the production and degradation plots should be indicated as the steady state expression level. x-axis should be labelled as concentration of the species, and y axis should be labelled as the rate or dx/dt.

(i) Briefly discuss how negative autoregulation alters the shape of the production rate curve compared to simple regulation. [5%]

Crib: The answer should include a discussion of the hill type shape of production term in the negative feedback as the gene represses its own production.

(ii) Briefly discuss how these differences in production curves relate to gene



(cont.

expression stability. [10%]

Crib: The answer should discuss how the production curve intersects with the removal curve and how this intersection point is not affected by changes in the promoter strength, which affects the beta term only.

(iii) Explain how negative autoregulation can shorten the response time of gene expression. Why might a faster response time be beneficial for this organism? [15%]

Crib: This answer should ideally include a sketch of how negative autoregualtion causes the response time to be dominated by the production rate, not removal rate. Faster response helps organisms to adapt with changing conditions.

(d) A single point mutation in the organism's RNA polymerase doubles the promoter binding rate. If the wild-type polymerase is replaced by this mutated version, how would you expect the steady-state concentration of the 30 negatively autoregulated genes to change, and why? [10%]

Crib: Since the mutated RNAP will only affect the beta values (promtoer strength) of the expression, we do not expect any changes in the steady state expression levels. The shape of the negative repression curve indicates that the point of intersection with the removal curve is relatively invariant with respect to the changes in the beta values.

(e) Researchers also identified 40 instances of a three-gene motif where: Gene 1 represses Gene 2, Gene 2 represses Gene 3, and Gene 3 represses Gene 1. What is the expected function (or typical behavior) of such a triple-repression motif in gene regulatory networks? Sketch the rate of production of a genetic reporter whose transcription is negatively regulated by Gene 3. [15%]

Crib: We expect each of the three genes to undergo oscillatory expression, and take turns in rock-paper-scissor style. This circuit is the central motif of the repressilator system, discussed in the class.
The plot should be a steady oscillation of the reporter production.

**END OF PAPER**

THIS PAGE IS BLANK