Mark Scheme

VICTORIA JUNIOR COLLEGE
BIOLOGY DEPARTMENT
JC2 PRELIMINARY EXAMINATIONS 2015
Higher 3

CANDIDATE NAME

CLASS | | | | | | INDEX NUMBER | |

BIOLOGY 9815/01

28 September 2015

Additional Materials: Answer Paper **2 hours 30 minutes**

READ THESE INSTRUCTIONS FIRST

Write your CLASS/ INDEX no. and name on all the work you hand in.
Write in dark blue or blue pen.
You may use a soft pencil for any diagrams, graphs or rough working.
Do not use any staples, paper clips, highlighters, glue or correction fluid.

Section A
Answer **all** questions.

Section B
Answer **three** out of **four** questions.

Section C
Answer the question.

At the end of the examinations,
1. Fasten all your work securely;
2. Circle the number of the section B question you have answered in the grid opposite.

For Examiner's Use	
Section A	
1	
2	
3	
4	
5	
Section B	
6	
7	
8	
9	
Section C	
10	
Total	

The number of marks is given in brackets [ ] at the end of each question or part question.

Section A

Answer all questions in this section.

Key: (;) means ½ mark and (;;) means 1 m

1 (a) **Fig.1** shows GDP in the binding site of a G protein.

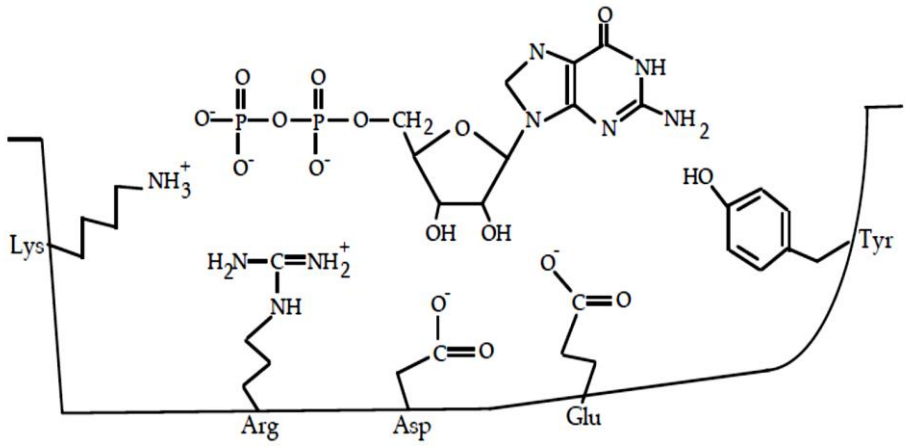


Fig. 1

(i) With reference to **Fig. 1**, describe the types of interactions between GDP and the amino acids that make up the binding site. **[3]**

- Ionic interaction – between side chains of Lys and Arg and the phosphate group of GDP;;
- Hydrogen bonds – between side chains of Asp and Glu with the ribose sugar of GDP;;
- Hydrogen bond – between side chain of Tyr with the $–NH_2$ group of guanosine/ nitrogenous base of GDP;;

(ii) A mutation results in the change of the amino acid from Arg to Asp at the binding site. A similar change occurs within the protein core of G protein.

Discuss the difference, if any, the effect of these mutations on the function/ activity of the protein. **[4]**

2 marks for binding site

- Ability of GDP to bind may be lost or decreased but not totally lost;
- Arg is positively charged whereas Asp is negatively charged/ Mutation results in an amino acid that is of a different charge;
- This will repel the negative charge of the phosphate group which may result in misalignment of the GDP with other amino acids in the active site;
- But the R group/ side group of Asp is small and hence the effect may not be significant i.e. results in slight distortion in alignment;

2 marks for protein core

- Buried charged amino acids tend to occur as "ion pairs" where oppositely charged amino acids interact to form the protein core which is hydrophobic;;
- A change of charge would destabilise the protein as the two negatively charged amino acids will repel each other;;

(b) A **protein family** is a group of proteins that have descended from a common ancestor. Members of the same family typically have similar three-dimensional structures, functions and significant sequence similarity especially in the binding or catalytic sites.

Many protein families belong to a protein superfamily which is the largest grouping (clade) of proteins for which common ancestry can be inferred. Members belonging to the same superfamily share structural similarity but not necessarily sequence similarity.

(i) Explain how it is possible for two proteins to share structural similarity without sequence similarity. **[2]**

- 3-dimensional structure is a result of **interactions between R groups of key amino acids** ► result in the **similar coiling and folding** of the polypeptide;;
- Clusters of similar aa that are responsible for structural folds ► occur as **motifs** (i.e. short sequences of less than 10 aa) while **overall sequence homology is low**;;

(ii) On the other hand, proteins with similar primary sequence can differ in their 3-dimensional configuration.

Explain how this is possible. **[2]**

- Depends on the environment the protein exists in:
  (i) **aqueous** environment favours protein to fold in such a way that the **hydrophilic R groups will be outside** while **hydrophobic R groups** will be located **within the core**;
  (ii) whereas a protein with similar primary structure but located in a **non-aqueous/ organic** solvent will fold differently with **hydrophobic R groups on the surface** of the protein while the **hydrophilic aa will be found in the interior**;
- Presence of **prosthetic group**: this may bind to specific amino acid residues resulting in them not being available for further interaction with other amino acids;;

Accept: **different post translational modifications** (PTMs) affect **3D conformation**;;

[Total: 11]

2 The antibody immunoglobulin G (IgG) is a type of globulin protein that is found in blood or other bodily fluids of vertebrates, and is used by the immune system to identify and neutralize foreign objects, such as bacteria and viruses. It is produced by a type of white blood cell called a B cell.

Fig. 2.1A is a schematic diagram of IgG while **Fig. 2.1B** shows the ribbon model of the same IgG molecule.

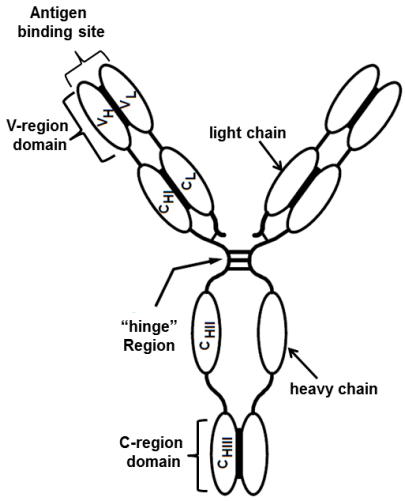


Fig. 2.1A

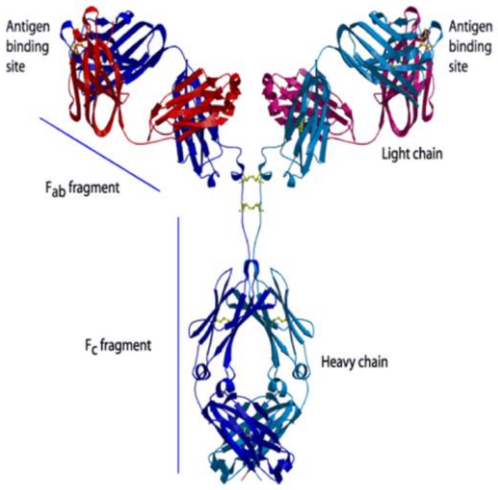


Fig. 2.1B

(a) With reference to **Fig. 2.1**, explain the term "domain". **[2]**

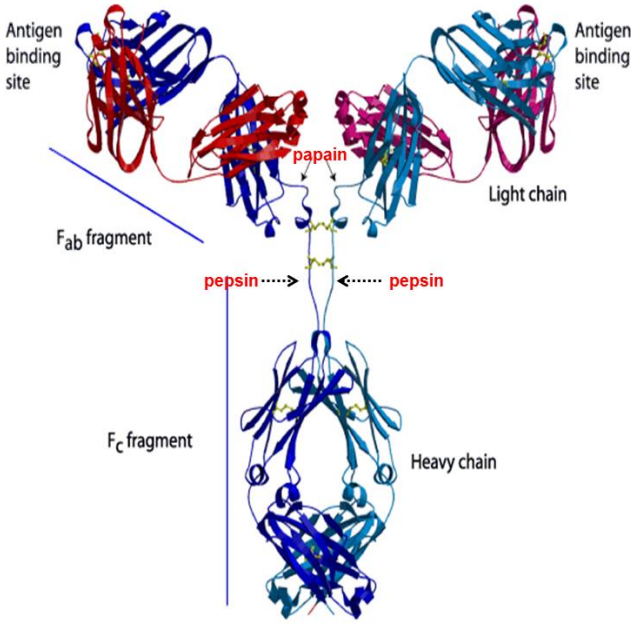
- Protein domains are defined as **compact, folded** structures within a polypeptide chain;;
- That corresponds to a certain **function or binding property**;;
- **Can fold independently** to the rest of the polypeptide chain;;

(b) Account for how the four IgG domains at the antigen-binding sites of the IgG molecule are different from the rest of the IgG domains, with reference to the function of the IgG. **[3]**

- These four IgG domains have **hypervariable** loops or **complementarity-determining regions (CDRs)** (the other IgG domains do not have them);;
- CDRs **vary greatly in their amino acid sequence**;;
- Allow antibodies to be **diverse in their specificity** thus able to **bind to a vast array of antigens**;;

(c) The IgG molecule can be broken down into smaller fragments by proteolytic enzymes such as papain, and pepsin. Treatment with papain yields two kinds of fragments, a pair of F(ab) fragments, and one Fc fragment while treatment with pepsin yields a single fragment, the $F(ab')_2$ but degrades the Fc region.

Indicate, on **Fig. 2.1B**, the site(s) where papain and pepsin are likely to act on IgG. **[1]**



@ ½ m per protein cleavage

(d) Predict the antigen-binding and agglutination capabilities of the resulting fragments that result from separate treatment by these enzymes. **[2]**

Papain	Pepsin
F(ab) fragments can bind since they contain the antigen binding sites;	$F(ab')_2$ fragments can bind antigen since they contain the antigen binding sites;
does not show agglutination;	can show agglutination since agglutination this function requires antigen binding site and bivalent binding for cross-linking;

(e) Papain is found as a component of extracellular latex of *Carica papaya* and is synthesized as an inactive precursor that becomes active in response to wounding of the plant. It is synthesised as a precursor protein called prepropapain that undergoes two cleavage steps, an initial cleavage of the amino acid signal sequence followed by further cleavage of the proregion, to produce the mature peptide.

Explain the importance of the signal sequence to the function of papain. **[2]**

- Allows the synthesis of the enzyme to be localized/ targeted to the ER, which is part of secretory pathway/ endomembrane system;;
- Subsequently modified, sorted and packaged by GA into secretory vesicles to be released to the outside of the cell as component of latex;;

[Total: 10]

3 Large-scale transposon insertional mutagenesis in mice has made significant advances through the development of an efficient *PiggyBac* (PB) transposition system. Two transgenic lines are required in this system. One is a PBase-expressing transgenic line that carries the gene for transposase and a reporter gene for coat colour. The other line carries a non-autonomous PB transposon.

Mating the two transgenic lines produces offspring that can be identified visually by coat colour for subsequent mutant screens. Additionally, the PB system is versatile since the transposon can be modified for different applications.

(a) With reference to the PB system, describe how transposition works and how it can be controlled. **[3]**

- The PB system is composed of two elements: a **transposase** and **PB transposon** whereby
- the **enzyme mobilizes** the transposon (by a cut and paste mechanism) by binding to recognition sites flanking the DNA sequence which then **inserts randomly elsewhere in the genome**;;
- **Transposition is controlled by separating the transposase from the transposon** by **generating two transgenic mouse strains**: one expressing the transposase gene; the second strain carrying the non-autonomous transposon;;
- **Mating brings both elements together in the offspring ►** PBase excises the PB transposon/ **the transposon is mobilized in the soma of the resulting offspring ►** the transposon can then **reinsert at another location** in the genome **► insertional mutagenesis ensues**;;

(b) Outline how disrupted genes can be identified. **[2]**

- Identify **offspring with coloured coat** and **defect/ phenotype of interest ► isolate genomic DNA** of mutant animal;
- **Perform PCR using primers** that anneal to inserted PB transposon but **pointing outwards** so that **sequences flanking insertion element are amplified**;
- **Sequence** flanking sequences;
- Identifying gene by using **BLAST analysis** of **mouse genome database** for **homology** to flanking sequences;

(c) The availability of complete genome sequences from humans and model organisms has allowed us to exploit various reverse genetic approaches.

Outline how RNAi can be used in mutant screens. **[3]**

- *in vivo* loss-of-function/ **knock down** study by the **selectively depressing specific protein activity** followed by **phenotypic analysis compared to controls**;;
- by **silencing gene expression** (in a sequence-specific manner) via **RNA-based mechanism**

- **Long dsRNA is processed to short interfering RNAs (siRNAs) by** the action of a dsRNA-specific endonuclease known as **Dicer**;
- The resultant siRNAs, 21–24 nt in length, are double stranded, and have 3; overhangs of 2 nt, are **incorporated into the RNA-induced silencing complex** (RISC);
- Within RISC complex, a helicase in RISC **unwinds the duplex siRNA**,
- unwound **antisense strand pairs with messenger RNAs** (mRNAs) that bear a high degree of **sequence complementarities/ complementary** to the siRNA;
- **Degrades mRNA using Argonaute proteins**;

(d) Describe one advantage and one disadvantage each of (i) RNAi and (ii) insertional mutagenesis. **[2]**

One advantage + one disadvantage @1m

Advantages of RNAi:

- Does not require the generation of targeting vectors since **small interfering RNAs (siRNA) can be synthesised *de novo*,** therefore **simplifying the process** of generating loss-of function;;
- Does not requires years of work to isolate and characterize the mutants;;

Disadvantages of RNAi:

- **Efficiency** of RNAi can **vary between individual genes**;;
- **Phenotype** may develop as the result of **'off-target' effects** caused by the **cross–reaction** of the small interfering RNA (siRNA) **to other mRNAs** with sequence homology to the candidate gene;;
- **Incomplete gene loss of function**;;
- **RNAi**-induced **phenotype in organism/** cell can **differ from the genetic null phenotype**;;
- dsRNAs/ siRNAs introduced into cells/ organisms **does not elicit effective and specific RNAi response in all situations** e.g. success *in vitro* in cell line may not translate into *in vivo* success in whole organism;;
- Toxicity due to ''oversaturation'' of the dsRNA pathway e.g. mouse model of liver disease led to high death rates in the experimental animals;;

- **Advantage of insertional mutagenesis**:
- **Insertion element of known sequence** into the host cell genome ► **used as a molecular tag** for **mapping the insertion sites ► quick and simple way for tracing candidate genes**;;
- **Disadvantage**:
- Time consuming;;
- AVP;;

[Total: 10]

4 Human lactase belongs to the family of glycosyl hydrolases. This enzyme enables infants to break down lactose, the main sugar in milk. The gene that codes for this enzyme is expressed in the small intestine of humans and other mammals, and remains active throughout the lives of people who are lactose tolerant. On the other hand, it is turned off after infancy in people who are lactose intolerant.

These adults who consume significant amounts of lactose in dairy products may experience symptoms that include abdominal bloating and cramps, flatulence, diarrhoea, nausea or vomiting. To manage lactose intolerance, these individuals often avoid lactose-rich food. However, there may be occasions where lactose avoidance is not possible.

Imagine you are a scientist in a pharmaceutical company who wants to produce lactase in tablet form to help these lactose-intolerant individuals.

(a) (i) Outline how the molecular weight of a purified polypeptide of lactase can be determined by mass spectrometry. Assume ions formed in a mass spectrometer only have a single charge. **[2]**

- Convert lactase into a **gas-phase ion** in an **ion source (e.g. MALDI, ESI, etc);**
- Accelerate the gas phase ion through a **mass analyzer**;
- by an electric field. Upon reaching the **detector;**
- the molecular weight of the enzyme is derived from the **largest peak** in a **mass spectrum** [1/2] since each vertical bar / peak represents an ion with a **specific mass-to-charge ratio;**

(ii) **Fig. 4.1** shows the processing and localization of human lactase, while **Fig. 4.2** shows three distinct peaks in the hypothetical mass spectrum of various translational products of the lactase gene.

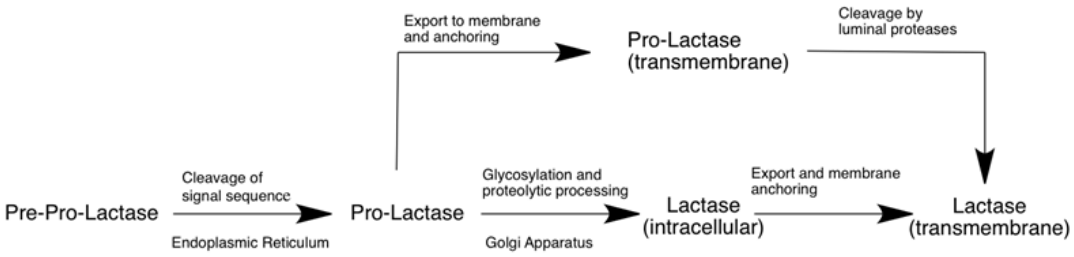


Fig. 4.1

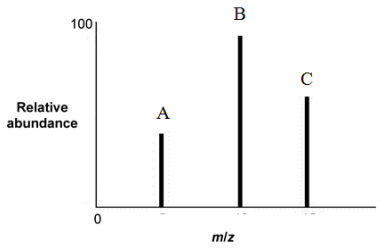


Fig. 4.2

With reference to **Fig. 4.1** and **4.2**, identify which peak (**A**, **B**, or **C**) was likely to be the molecular weight of lactase. Explain your answer. **[2]**

- Peak **A;**
- Upon translation, **lactase undergoes post-translational modification / cleavage to become its functional state;**
- **The three peaks correspond to pre-pro-lactase (largest m/z), pro-lactase and the intracellular / transmembrane lactase (smallest m/z);;**

(b) (i) Explain how ultracentrifugation and a specific column chromatography may be employed to obtain a fraction from the cell extract containing the 160 kDa enzyme. **[2]**

- High speed centrifugation can remove **larger organelles and other dense structures** as they form the **pellet;**
- while the **soluble enzyme** remains in the **supernatant**;

- Since the molecular weight of the polypeptide is known, **size exclusion chromatography / gel filtration;**
- can be used to obtain the cell fraction consisting of the **enzyme and other proteins of around 160kDa** [1/2] since it separates mixture according to **size**;

(ii) The quantity of lactase present in the cell fraction can be detected by incubation with its substrate lactose and performing the Benedict's test followed by comparing to a colour standard.

State another method that can detect and quantify the enzyme accurately. **[2]**

- **Enzyme-linked immunosorbent assay (ELISA)**;;

- ELISA makes use of an **antibody coupled to an enzyme**;
- to bind to the lactase (direct detection) or to another antibody bound to lactase (indirect method). The **chromogenic or fluorogenic substrate** catalyzed by this coupled enzyme produces a **measurable product;**

(c) Describe one similarity and one difference between the specific column chromatography and SDS PAGE in the separation of proteins. **[2]**

Similarities: any one below:

- Both separates proteins base on size;;
- Both involves the use of porous particles to separate the proteins;;

Differences: any one below:

Gel filtration	SDS-PAGE
Proteins need not be negatively charged	All proteins are negatively charged
Proteins may be in their native 3D conformation	All proteins are linear
Does not require electric current; the movement of proteins through the column is based on gravity	Involves electric current where the negatively charged proteins migrate toward the positive electrode
Smaller proteins move slower through the column later than the larger proteins	Smaller proteins move faster through the gel matrix than larger proteins / reach positive electrode faster

[Total: 10]

5 The electron transport chain of the mitochondria is organised into four complexes. Complex III is known as coenzyme Q reductase. It consists of two b-type cytochromes and one c-type cytochrome (**Fig. 5.1**).

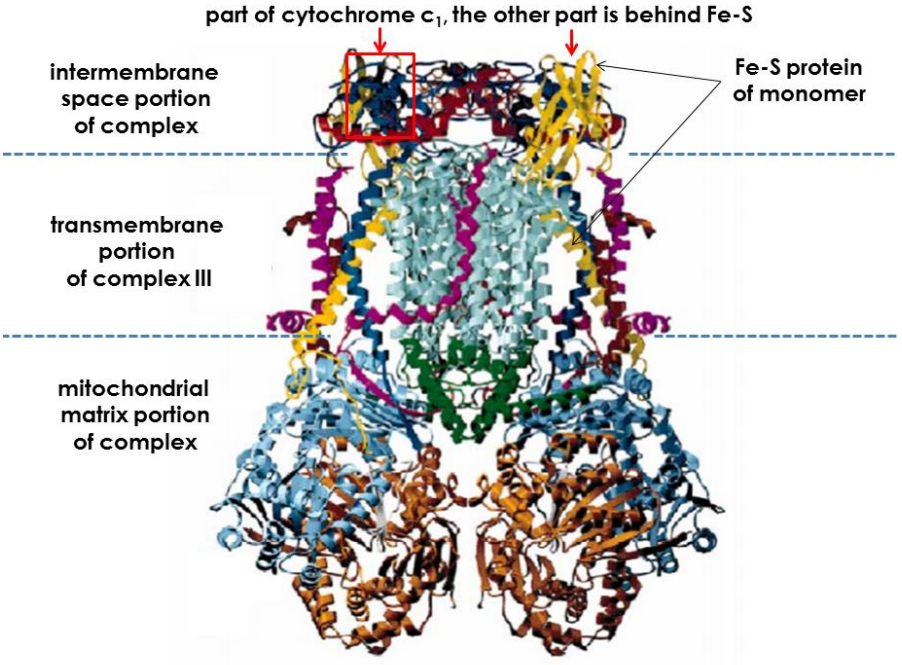


Fig. 5.1
(adapted from https://www.tamu.edu/faculty/bmiles/lectures/electrontrans.pdf)

In a deficiency in the mitochondrial respiratory chain, the level of core proteins (Fe-S protein) and subunit VI of complex III were greatly diminished in the patient's mitochondria. Cytochrome c_1 polypeptide was found at normal levels but was sensitive to proteolysis by trypsin. These results show that complex III is not assembled in the patient's mitochondria.

(a) Explain the role of complex III in the respiratory electron transport chain. **[2]**

- It catalyses **electron transfer** from **ubiquinol to a soluble cytochrome c** with the generation of a **proton gradient across the mitochondrial membrane;;**
- Electron transfer is **coupled to translocation of two protons** across the inner mitochondrial membrane **per oxidized ubiquinone**;;

(b) Describe the advantage of complex III being a multi-enzyme complex. **[3]**

- Complex III is made up of three subunits, cytochrome b (Cb), cytochrome c_1 and the Fe-S protein with active redox centres in **close proximity;;**
- Allows the **rapid shuttling of electrons from ubiquinone to cytochrome c/** for efficient transfer of electrons within the complex and **coupling electron transfer with translocation of H^+ into intermembrane space**;;
- because these active redox centres are associated in the same multi-protein oligomer, the **redox potentials of prosthetic groups increase** from one to another redox group is **energetically favourable**;;

(c) Describe the role of iron-sulfur in the Fe-S protein. **[2]**

- iron-sulphur is the prosthetic group and **highly mobile/ flexible**;;
- **enables the transfer of electron from ubiquinol to cytochrome *c_1***;;

Extra info
X-ray structures of the complex III reveal two different positions for Fe-S protein. One location is close to the low potential heme of Cb, and the other is close enough to cytochrome c_1 suggesting that it may possibly be moving around a pivot joint during the catalytic cycle of the functioning bc_1 complex, shuttling electrons from ubiquinol to cytochrome c_1.

(d) State a post-translational modification and suggest how it can prevent cytochrome c_1 polypeptide from undergoing proteolysis by trypsin. **[2]**

- **N-terminal acetylation**;;
- It **stabilizes** cytochrome c_1 polypeptide, causing it to be **more resistant to degradation**;;

[Total: 9]

Section B

Answer 3 out of the 4 questions in this section.

6 *c-Myc* codes for the transcription factor c-Myc, a nuclear phosphoprotein that plays a role in cell cycle progression, apoptosis and cellular transformation. A mutated version of *Myc* is found in many cancers, which leads to constitutive (continuous) gene expression. Overexpression of the oncoprotein leads to the unregulated expression of many target genes, some of which are involved in cell proliferation. The result is the formation of cancer.

(a) Briefly discuss one possible benefit and one possible limitation to studying the nucleotide sequences of the *Myc* as compared to studying the amino acid sequence of the protein. **[2]**

Benefit @ 1m:
- More informative to study the types genetic mutations that result in cancer;;

Limitation @ 1m:
Studying nucleotide sequences alone does not reveal protein structure and its interactions with other molecules, which is useful for finding anti-cancer solutions/ targets;;

(b) A student purified the protein from a bacterial expression system. After precipitating the abundant protein out of solution, Instead of resuspending the purified protein in protein storage buffer, he added an excess of dilute 0.1M HCl to Myc protein.

Explain what would happen to the structure of the protein. **[2]**

- Myc has alpha-helix as its secondary structure;
- Disruption of H-bonds;
- Formed between the carbonyl (C=O) group of the peptide bond of n^{th} amino acid and the N-H group of the peptide bond $n+4^{th}$ amino acid;
- Alpha-helix disrupted / Single-stranded protein unravels;

(c) **Fig. 6** shows the structure of the Myc protein. In this particular instance, it is shown to be dimerized with its partner Max, another transcription factor.

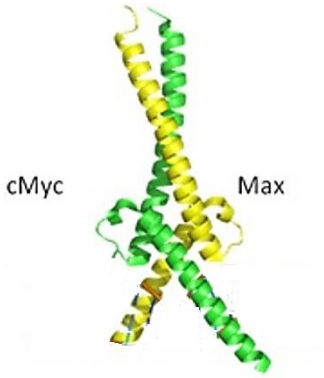


Fig. 6

Account for at least two domains that should be present on Myc protein so that it can carry out its function. **[2]**

- protein-binding domain;
- to bind to Max;
- DNA-binding domain;
- it is a transcription factor, it needs to bind to enhancer sequences;

(d) *c-Myc* protein is believed to regulate expression of 15% of all genes through binding enhancer sequences and by facilitating chromatin remodelling.

Explain how Myc protein is able to regulate gene expression

(i) by binding to enhancer sequences. **[1]**

- It recruits other specific transcription factors / causes the DNA to bend or loop upon itself;
- To stabilize transcription initiation complex ► upregulates gene expression;

(ii) by facilitating chromatin remodelling. **[3]**

- By recruiting histone acetlytransferases (HATs);
- Histones get acetylated/ transfer negatively charged acetyl groups to histone tails e.g. positively charged lysine residues ;
- Net positive charge on histones is less;
- Less affinity to negatively-charged DNA;
- Decondensation of chromatin/ becomes less condensed;
- Promoters or enhancers sequences of genes exposed to TFs and RNA polymerase ► gene expression upregulated;

[Total: 10]

7 (a) (i) Outline how Edman degradation may be employed to determine the amino acid sequence of a protein. **[2]**

- Amino acids at the **N-terminus** of polypeptide are **labeled;**
- and **cleaved off sequentially**;
- **without disrupting other peptide bonds;**
- between other amino acid residues. The amino acids are then identified using **(reverse-phrase) chromatography**;

(ii) Imagine you are a scientist who has previously isolated the 160 kDa human lactase enzyme from a cell extract. There are 1927 amino acids in the human lactase enzyme.

Explain why this Edman degradation sequencing technique is unlikely to be the method of choice for the sequencing of the protein and state a method of choice. **[1]**

- Edman degradation is used to sequence **not more than 50 amino acid residues**;

- **Tandem mass spectrometry**;

(b) (i) You have obtained the amino acid sequence of lactase. Comment on the significance of this sequence if you were to make use of yeast or bacterial cells for mass production of the protein. **[2]**

- **Useful** because the **DNA sequence** can be determined from the amino acid sequence;;
- Clone into an **expression vector** containing an appropriate **yeast / eukaryotic promoter**;
- Introduce the recombinant vector into the host cells for **over-expression** of the gene;

(ii) Assuming that no antibody is available that binds specifically to your protein of interest, describe how the protein can be extracted from the host cells in a purified form using a specific column chromatography. **[5]**

- Affinity chromatography;;
- **Affinity purification tags** attached to the **N- or C-terminals** of the protein of interest;
- **Any example**: glutathione-S-transferase (GST), polyhistidine-tags, biotin tags, etc.;
- **Over-expression** of the **recombinant gene** in suitable cells;
- **Cell lysis** followed by **flowing through affinity matrix/ resin**;
- **Selective protein purification;** *by any one below:*
- GST tags → affinity for immobilized glutathione-covered matrices;
- Biotin tags → avidin / streptavidin;
- Polyhistidine tags → immobilized (divalent) metal ions (e.g. nickel);
- **Washing and elution** of purified proteins;

[Total: 10]

8 In high throughput analysis of gene function, genes were obtained from a library of open ORFs that code for proteins with a variety of functions (kinases, transcription factors and cell surface receptors). The genes were amplified by PCR and then placed into Gateway destination N- and C-terminal GFP fusion vectors. These constructs were introduced into HEK293T cells and GFP fluorescence was visualized by confocal microscopy.

All C-terminal fusion proteins localized to cellular compartments in accordance with previous studies and bioinformatic predictions, while less than half of the N-terminal fusion proteins localized correctly.

(a) (i) What conclusions can be drawn from these observations about tagging at N-terminal and C-terminals protein tagging? **[1]**

- **N-terminal tagging** with GFP adversely **affects the protein localization**/ leads to **incorrect subcellular compartments**;

- Tagging **at the C-terminal** is generally better in **for correct localization of the native protein**;

(ii) N-terminal signal peptides are found in the N-terminal GFP fusion proteins that did not localise correctly. Suggest a cause for the incorrect localisation. **[1]**

- Signal sequence at the N-terminus to be **masked by tag**;;

(b) Other than fluorescent tags, state another way of tagging proteins that can be used for tracking their movement.

Describe one advantage and one disadvantage of each type of protein tags. **[3]**

- **Epitope** tagging (addition of small peptides of 3–14 amino acids e.g. His6 and c-myc) for detection by **immunohistochemistry**;;

One adv + one disadv = 1m for each approach

Advantages of using epitope tags for protein localisation (any one point):
- Due to small size, epitope less likely to disrupt the normal function or localization of the protein;
- Stable and not sensitive to photobleaching;

Disadvantages (any one point):
- Immunohistochemical detection is time-consuming due to complex processing steps;
- Immunohistochemistry can only be performed on fixed cells;
- Difficult to visualize rare targets and highly abundant targets in the same slide;
- Not possible to discern targets in different concentrations since subtly mixed colours from chromogenic dyes are hard to distinguish;

Advantages of fluorescent tags for protein localisation (any one point):
- Can be visualized easily via standard confocal or fluorescent microscopy;
- Can be performed and detected in living cells;
- Fluorescent dyes allow for separate identification of targets;
- Possible to discern targets of very different concentrations on the same slide;

Disadvantage (any one point):
- Fluorochromes susceptible to deterioration at room temperature even when stored away from light;
- Quenching of fluorescent signal when fluorochromes are in close proximity;
- Photobleaching of dyes when subjected to light ► only fluoresce for a limited time;
- Increased background staining due to autofluorescence;

To study protein localization in budding yeast, Erin O'Shea and co-researchers created a GFP-tagged yeast strain collection and database. They reported new information on 70% yeast proteins whose subcellular localisations were previously unknown, providing a better understanding of the processes of transcriptional co-regulation, as well as protein interactions that occur within organelles in eukaryotic cells.

Fig. 8.1A illustrates how they constructed the GFP-tagged library. **Fig. 8.1B** shows distinct subcellular localization of some of these yeast proteins.

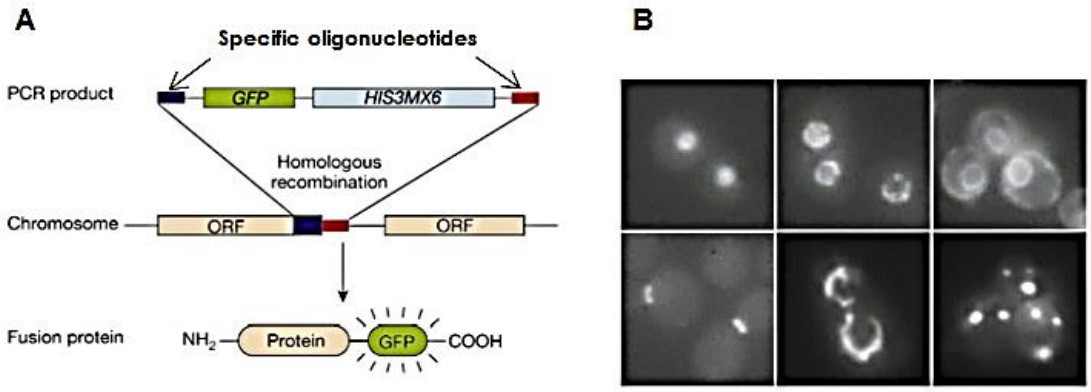


Fig. 8.1 (A) tagging of an ORF in its chromosomal location and **(B)** subcellular localisations of different GFP-tagged proteins (top panel from left: nucleus, nuclear periphery, endoplasmic reticulum; bottom panel from left: bud neck, mitochondrion, lipid globules).

(c) Yeast strains can be mated. Use this information and **Fig. 8.1** to outline the major steps of an experiment to investigate if two proteins colocalize. **[3]**

- **Construct a second library tagged by a variant of GFP (e.g. RFP);**
- PCR products are generated that contain e.g. **RFP tag**, a selectable marker gene (HIS3Mx6) and flanked by specific oligonucleotide primers;
- After **introduction into yeast cells**, **insertion of PCR product occurs by homologous recombination** at 3' end of corresponding ORF in its chromosomal location;
- **Mate the GFP-tagged strain with strain expressing RFP-tagged gene ►** GFP and RFP tagged protein products will be expressed within same cell;
- Analyse yeast cells for positive GFP and RFP signals **separately (i.e. green and red fluorescence, respectively);**
- Then analyse for **simultaneous/ merged signals** of overlap (**i.e. yellow**) by **fluorescence microscopy;**

(d) Aberrant protein localization is a prominent feature of many human diseases. Aberrant folding of proteins with important cellular roles can cause diseases due to loss of the protein function or result in proteins that retain intrinsic function yet cease to function normally due to mislocalization.

State the name(s) of experiment(s) that you will use to confirm the second outcome.

[2]

- Structure not altered confirmed by **XRC** between correctly and misfolded protein ► **same 3D structure**;;
- **Immunohistochemistry** with HA-tagged if specific antibody is not available/ fusion protein with **GFP fluorescence microscopy** to show **subcellular localization is altered**;;

[Total: 10]

9 Phosphorylation of protein kinases (PKs) plays a central role in the signal transduction of cells.

(a) Describe the type of protein-protein binding that allows a protein kinase to recognize the proteins that it will phosphorylate. **[2]**

- **Surface string Interaction** where a portion of the **surface of one protein contacts an extended loop** of polypeptide chain (a "string") on a second protein;;
- Named example to illustrate:
- Occurs between a peptide containing a **phosphotyrosine** residue and the **SH2 domain of Src** and other related proteins.
- Takes place between the **kinase fold** of Src and part of the protein that it phosphorylates.

(b) Explain the effects of phosphorylation of a named protein kinase of your choice. **[4]**

- **Receptor tyrosine kinase**;;
- Causes **conformational change to PK** due to **phosphate group** being negatively charged;;
- **Phosphorylated tyrosine** side chains recognised by other **proteins containing the SH2 domain**, leading to their **activation**;;
- Allows the **transduction of extracellular signal to inside of the cell by phosphorylation of intracellular proteins**;;

(c) Previous studies have shown that Mg^{2+} ions are required at the active site of PKs. Explain the role of Mg^{2+} ion in the function of PK. **[4]**

- PK transfers a **phosphate group from adenosine triphosphate (ATP)** to the side chains of particular serine, threonine or tyrosine residues of substrates;;
- Mg^{2+} ion **functions as a cofactor;;**
- Presence in the **active site** could **enhance the binding affinities of ATP to the PKs;;**
- Thus crucial for the **formation of transition state** of phosphoryl transfer reaction;;

[Total: 10]

Section C

Answer the question in this section.

10 (a) A team of researchers have recently isolated microbial organisms (termed extremophiles) from regions on earth with extreme conditions. They hope that these microbial genomes will contain novel genes that allow these microbes to adapt to life in inhospitable environments. These genes are of interest in biotechnology as the high stability of their protein products can be exploited commercially in industrial processes.

You have received the sequenced genomes of 6 such microbial isolates. Outline the major bioinformatics steps you will take to investigate (i) if these genomes contain potentially novel genes and (ii) the potential functions of these genes. **[8]**

1) Look for **homologous genes in other organisms** and translate into amino acid sequence

- Select **genome/ nucleotide sequence databases** e.g. **GenBank** or **Entrez** to look for homologous genes in other organisms
- Key in microbial sequence (query sequence) in **FASTA format** and using **BLAST alignment tool**, BLASTn to compare nucleotide homology;;

2) Look for **homologous proteins** in other organisms and **phylogenetic relationships** based on protein homology (protein families) after **translating nucleotide to amino acid sequence using tBLASTn**

- Use **ClustalW** alignment tool for multiple sequence alignment to look for homology between query protein and other organisms;;

- Obtain **phylogenetic trees** to view the evolutionary relationships between query protein and other proteins or genomes;;

3) Use **ScanProsite** programme and databases e.g. **SwissProt** to see what certain parts of the protein is similar to, identify protein structure motifs and **to predict protein functional domains and protein function and its mechanism;;**

4) Search **PDB** (database for 3D structures) for similar proteins and use software that will **predict 3D structure** of query protein (3m max):

- **Homology modelling** (use existing homologous proteins as templates, mutate each amino acid in the solved structure computationally into the corresponding amino acid from the unknown structure ► build 3D structure model of extremophile protein)
- **Threading** (if overall homology is low, compare novel aa sequence against a library of structural templates and use scoring function to assess best fit of novel sequence to a given fold) and
- ***Ab initio*** methods (to determine protein structure from scratch based on energy function algorithms for the most thermodynamically stable form of extremophile protein;;

(b) Briefly describe the significance of protein glycosylation. Using your knowledge of analytical techniques, Suggest and outline one method that you would use to detect and quantify glycosylated proteins. You can choose an antibody-based method or mass spectrometry. **[6]**

Significance @ 2m

Protein glycosylation is the attachment of a saccharide moiety to a protein.

- Glycosylation is one of the most abundant posttranslational modifications of proteins.
- Has a role in protein folding, interaction, stability, and mobility (i.e. protein intracellular trafficking and secretion), signal transduction.
- By regulating protein activity, glycosylation is involved in the normal functioning of the cell and in the development of diseases.

Methods for the identification and quantification of oligosaccharides, specifically N- and O-glycosylated proteins include SDS-PAFE gel staining, glycan-specific antibody staining of gels, MS.

Any one method below @ 4m

1a) Protein glycosylation can be resolved on SDS-PAGE ► stain the gel for glycoproteins. Most gel-staining procedures are based on the periodic acid-Schiff (PAS) reaction, in which periodic acid oxidizes two vicinal diol groups to form an aldehyde, which reacts with the Schiff reagent to give a magenta colour. This chromogenic gel staining is most commonly performed with acid fuchsin which can be detected fluorescently at 535 nm and is thus two- to fourfold more sensitive than the visible staining method.

1b) Proteins separated by 2D-PAGE ► blotted onto PVDF membranes can easily be tested for glycosylation by using commercially available carbohydrate specific staining methods.

2) (O-β-GlcNAc) may be analyzed in a method that is specific to this unique glycan moiety, since primary antibodies, for example, commercially available CTD 110.1 and RL2, have been developed against it. The cross-reaction of these antibodies with secondary antibodies conjugated to enzymes that catalyze a color- or luminescence-producing reaction can be exploited in procedures such as blotting to detect O-GlcNAcylated proteins.

3) The identification of glycosylated proteins derived from 1D or 2D gels is generally done by MALDI-MS fingerprint analysis or by fragmentation spectra received by MALDI-PSD or ESI-MS/MS analysis. This identification in combination with database information lead to a first hint about a potential glycoprotein.

4) Glycoproteins are first digested into glycopeptides and either analyzed directly by liquid chromatography and tandem mass spectrometry (LCMS/ MS) or first deglycosylated and then enriched for glycans or peptides prior to analysis (see figure below). MS involves the conversion of the sample into gaseous ions, with or without fragmentation, which are then characterized by their mass to charge ratios (m/z) and relative abundances.

(c) Proteins exert their biological functions within the spatiotemporal context of an intact cell. A protein needs to be at the right place at the right time to gain access to appropriate molecular interaction partners.

A cDNA library from metastatic human breast cancer tissue is available to you. One of these cDNA clones codes for Molecule **C**, which is hypothesised to be a major component of a complex pathway that promotes formation of new blood vessels in metastatic tumours.

Describe an *in vivo* experiment that can be performed to identify all potential interacting partners of Molecule **C** in this pathway. **[6]**

- **Yeast two hybrid** assay/ system to **identify interacting partners of** a given protein, Molecule **C**;;
- **Interaction** between (Molecule **C** and its partner is measured by the **reconstitution of a functional transcription factor** and the **concurrent activation of reporter genes** (e.g. β-gal as a colorimetric reporter gene, since it is very robust and a variety of qualitative and quantitative assays are available);;

- E.g. in a GAL4-based Y2H system, the **transcription factor GAL4 consists of DBD (DNA-binding domain) and AD domain (activation domain);**

- cDNA coding for Molecule **C** is used as **bait protein** is fused to DNA coding **DNA-binding domain** of transcription factor Gal4;
- Each **cDNA of human cancer** library is **ligated to the AD sequence**;
- The **two expression vectors**, one containing GAL4-DBD (i.e. **bait plasmid**) and the other containing GAL4-AD (i.e. **prey plasmid**) are **simultaneously introduced into the mutant yeast strain** (unable to synthesize a certain amino acids or nucleic acid) and will not survive when grown on media that lacks these nutrient);
- **If the bait and prey proteins interact, transcription of reporter gene can occur → a change in cell phenotype such as synthesis of essential nutrient**;;

Extra point (to link to relevant phenotype when interaction takes place) :

- Example of one reporter genes such as HIS3, URA3, ADE2 or lacZ confer ability to grow on plates lacking histidine, uracil, adenine or a blue colony in presence of X-Gal, respectively;

[Total: 20]