

ISC Class 12 Biotechnology Specimen Paper 2

Written by the Future Topper subject team on the CISCE examination pattern. Not an official CISCE paper.
Found a mistake? Write to contactfuturetopper@gmail.com

Maximum Marks: 70 | Time Allowed: Three hours | Reading Time: 15 minutes

Paper 1 (Theory) carries 70 marks. Sections A and B are compulsory. Attempt any three questions from Section C.

General Instructions

Write your answers on the separate answer paper provided. The first 15 minutes are for reading the paper, and you may not write during that time. Section A and Section B are compulsory, and from Section C you must attempt any three of the four questions. Marks for each question, or part of a question, are shown in brackets. No diagrams, maps or gel photographs are printed in this paper. Where a question asks you to draw a diagram, a flow chart or a restriction map you are expected to draw it yourself in pencil and label it. Nucleotide sequences must be written with the 5' and 3' ends marked. All working for numerical questions must be shown on the same sheet as the answer, and the final answer must carry its unit. This paper carries 70 of the 100 marks for the subject, the remaining 30 being for the practical work, the project and the practical file.

SECTION A (14 Marks)

Attempt all parts of Question 1.

Question 1 [14]

Each part carries 1 mark. For the parts that offer four choices, write only the option you choose. For the rest, answer in a single sentence or give the single word, name or value asked for.

- (i) The two strands of a DNA double helix are described as antiparallel because:
- (a) they carry the same base sequence
 - (b) they run in opposite directions, one 5' to 3' and the other 3' to 5'
 - (c) they are of unequal length
 - (d) they are held by covalent bonds
- (ii) Name the short stretches of DNA built on the lagging strand during replication, and name the scientist after whom they are called.
- (iii) The strand of a gene which carries the same sequence as the messenger RNA, with thymine in place of uracil, is called:
- (a) the template strand
 - (b) the antisense strand
 - (c) the coding strand
 - (d) the lagging strand
- (iv) Name the site on a prokaryotic messenger RNA at which the small subunit of the ribosome binds.
- (v) The number of nucleotides in the recognition sequence of most of the restriction enzymes used in gene cloning is:
- (a) two or three
 - (b) twenty to thirty
 - (c) more than one hundred
 - (d) four to eight
- (vi) Name the part of the Ti plasmid of *Agrobacterium tumefaciens* which is carried into the plant genome, and state what is put in its place in a vector built for gene transfer.
- (vii) In agarose gel electrophoresis the fragments of DNA move towards the anode because:
- (a) they carry a net positive charge
 - (b) they carry no charge
 - (c) the sugar phosphate backbone gives them a net negative charge
 - (d) the gel pushes them that way
- (viii) Name the dye used to stain DNA in an agarose gel and state the light under which the bands are then seen.

- (ix) A plant raised from a single cell or a small explant by tissue culture, and genetically identical with the parent plant, is called:
 (a) a hybrid (b) a somaclone
 (c) a mutant (d) a haploid
- (x) Name the plant hormone which, when present in the medium in a high proportion to auxin, makes a callus form shoots.
- (xi) The process by which a culture vessel full of animal cells is divided among fresh vessels with fresh medium is:
 (a) cryopreservation (b) transfection
 (c) subculturing (d) fusion
- (xii) Name the substance added to a suspension of animal cells before they are frozen in liquid nitrogen, and state what it prevents.
- (xiii) The programme commonly used to build a multiple sequence alignment is:
 (a) ClustalW (b) BLAST
 (c) FASTA (d) Entrez
- (xiv) Give the term for the branch of bioinformatics which studies the complete set of proteins expressed by a genome.

SECTION B (26 Marks)

Attempt all parts of Question 2. Each part carries 2 marks.

Question 2 [26]

- (i) A sample of double stranded DNA is found to hold 22 per cent adenine. Calculate the percentage of each of the other three bases in it, and name the rule you have used. [2]
- (ii) Explain why one of the two new strands at a replication fork is built continuously while the other is built in pieces. [2]
- (iii) Write the sequence of the messenger RNA transcribed from the coding strand 5'-ATGTCCAAAGGGTAA-3', and state the number of amino acids in the peptide it codes for. [2]
- (iv) Distinguish between the exonuclease and the endonuclease activity of a nuclease, and state which of the two a restriction enzyme shows. [2]
- (v) The restriction enzyme BamHI recognises the sequence 5'-GGATCC-3' and cuts between the first and the second G on each strand. Write the two ends produced, marking the 5' and 3' ends, and state the kind of end it is. [2]
- (vi) A plasmid cut with a single restriction enzyme and run on a gel gives two bands, of 4.0 kb and 1.2 kb. State the number of sites the enzyme has in the plasmid and calculate the size of the plasmid in base pairs. [2]
- (vii) Explain what complementary DNA is and state one advantage of cloning it in place of genomic DNA when a eukaryotic protein is to be made in a bacterium. [2]
- (viii) Explain what a shuttle vector is and state why it is useful. [2]
- (ix) Explain what somatic hybridisation is and name the agent commonly used to fuse two protoplasts. [2]
- (x) A culture is seeded with 2.0×10^4 cells and reaches 1.6×10^5 cells in 36 hours of exponential growth. Calculate the number of population doublings and the doubling time of the culture. [2]
- (xi) Explain why animal cells must be given a surface to attach to while most plant cells need none, and give the term used for cells which have this requirement. [2]
- (xii) Explain what the contact inhibition of a cultured cell line is and state what its loss indicates. [2]

- (xiii) State what the accession number of a database entry is, and explain why a sequence is cited by its accession number rather than by its name. [2]

SECTION C (30 Marks)

Attempt any three questions from this Section. Each question carries 10 marks.

Question 3 [10]

- (a) Describe the three stages of the translation of a messenger RNA on the ribosome, naming the sites of the ribosome at which each step occurs. [3]
- (b) Draw and label a diagram of a replication fork in a bacterium, showing the parent strands, the leading strand, the lagging strand with its fragments and the direction in which the fork is moving. [3]
- (c) Describe the three kinds of processing which the primary transcript of a eukaryotic gene undergoes before it leaves the nucleus, and explain what alternative splicing allows a cell to do. [4]

Question 4 [10]

- (a) A circular plasmid 8.0 kb long carries two HindIII sites and one EcoRI site. Digestion with HindIII alone gives fragments of 5.0 kb and 3.0 kb, and digestion with both HindIII and EcoRI together gives fragments of 4.0 kb, 3.0 kb and 1.0 kb. Work out where the EcoRI site lies in relation to the two HindIII sites and draw the circular map. [3]
- (b) Describe how a gene is introduced into a dicotyledonous plant using *Agrobacterium tumefaciens*, from the building of the vector to the raising of the transgenic plant. [3]
- (c) Describe the Southern blotting technique and state one use to which it is put. [4]

Question 5 [10]

- (a) Describe the three phases of the growth curve of a batch culture of animal cells, stating what the cells are doing in each. [3]
- (b) Describe how a suspension culture of plant cells is raised and maintained, and state two uses to which such a culture is put. [3]
- (c) (1) Explain what anther culture is and state what is obtained by it. (2) Explain how a virus free plant may be raised from an infected stock by meristem culture. (3) Explain what a bioreactor does and name two conditions it must control. (4) Explain what cryopreservation is and state the temperature at which the cells are held. [4]

Question 6 [10]

- (a) Explain what a phylogenetic tree is, state what the branch points and the branch lengths of such a tree stand for, and state what is needed before a tree can be built from sequence data. [3]
- (b) Explain what the Human Genome Project set out to do, state two of its findings, and name two of the ways in which its data are now used. [3]
- (c) (1) Explain what a DNA microarray is and what it measures. (2) Explain what gene prediction software has to do in a eukaryotic sequence that it need not do in a bacterial one. (3) Explain what molecular docking is. (4) State two problems which the growth of biological databases has created for the people who use them. [4]

ANSWER KEY AND MARKING SCHEME

Full working is given for every numerical. Marks shown in italics are the step marks an examiner would award. Accept any correct alternative method which reaches the same answer, accept a sequence written without the end labels provided the direction is clear, and accept a correctly drawn diagram, flow chart or restriction map in any reasonable style.

Question 1 [14]

- (i) (b) they run in opposite directions, one 5' to 3' and the other 3' to 5'. The two sugar phosphate backbones are therefore of opposite polarity, which is why the 5' end of one strand lies against the 3' end of the other at each end of the molecule.

Correct option 1.

- (ii) They are the Okazaki fragments, named after Reiji Okazaki.

Fragments named with Okazaki 1.

- (iii) (c) the coding strand. It is not transcribed, but since it is complementary to the template strand it carries the same sequence as the transcript, and it is therefore the strand written down when a gene is quoted. It is also called the sense strand.

Correct option 1.

- (iv) The ribosome binding site, also called the Shine-Dalgarno sequence, which lies a few bases upstream of the start codon and pairs with a complementary stretch in the 16S ribosomal RNA of the small subunit.

Correct answer 1.

- (v) (d) four to eight. Most are six, EcoRI, BamHI, HindIII and PstI among them, and a shorter site would occur so often in the DNA that the enzyme would cut it into useless pieces.

Correct option 1.

- (vi) The T DNA, which is the region of the plasmid transferred into the plant chromosome. In a vector built for gene transfer the tumour causing genes of the T DNA are removed, the plasmid being disarmed, and the gene of interest together with a selectable marker is put in their place between the border sequences.

T DNA named, with the gene of interest and a marker replacing the tumour genes 1.

- (vii) (c) the sugar phosphate backbone gives them a net negative charge. Each phosphate group carries a negative charge at the pH of the buffer, so every fragment moves to the positive electrode, and the gel then separates them by size, the smaller fragments travelling further.

Correct option 1.

- (viii) Ethidium bromide, and the bands are seen as orange fluorescence under ultraviolet light.

Dye named with ultraviolet light 1.

- (ix) (b) a somaclone.

Correct option 1.

- (x) Cytokinin. A high ratio of cytokinin to auxin in the medium turns the callus towards shoot formation, while a high ratio of auxin to cytokinin turns it towards root formation.

Correct answer 1.

- (xi) (c) subculturing, also called passaging.

Correct option 1.

- (xii) A cryoprotectant, dimethyl sulfoxide being the one usually used and glycerol being the alternative. It prevents the formation of ice crystals within the cells, which would otherwise pierce the membranes and kill the cells as they freeze.

Cryoprotectant named, with ice crystal formation prevented 1.

- (xiii) (a) ClustalW. Clustal Omega, MUSCLE and T-Coffee are other programmes used for the same purpose.

Correct option 1.

- (xiv) Proteomics.

Correct answer 1.

Question 2 [26]

- (i) In double stranded DNA adenine pairs only with thymine and guanine only with cytosine, so the amount of adenine equals the amount of thymine and the amount of guanine equals that of cytosine. Thymine is therefore 22 per cent. Adenine and thymine together account for $22 + 22 = 44$ per cent, so guanine and cytosine together account for $100 - 44 = 56$ per cent. Since the two are equal, each is $56/2 = 28$ per cent. The answer is adenine 22, thymine 22, guanine 28 and cytosine 28 per cent. The rule used is Chargaff's rule of base equivalence.

Thymine 22 per cent, with the pairing reason 1. Guanine and cytosine each 28 per cent, and Chargaff's rule named 1.

- (ii) The two parent strands are antiparallel, so at a fork the two templates run in opposite directions, yet DNA polymerase can add a nucleotide only to the free 3' hydroxyl of a growing chain and so can build a strand only in the 5' to 3' direction. On one template the 5' to 3' direction of growth happens to point the same way as the fork is moving. The polymerase can therefore follow the fork as it opens and build that strand in one unbroken piece, and it is called the leading strand. On the other template the 5' to 3' direction of growth points away from the fork. The polymerase cannot run backwards, so it must wait until a stretch of that template has been exposed, then start afresh near the fork and work away from it, and repeat this for each new stretch uncovered. The strand is therefore built as a series of short pieces, the Okazaki fragments, each begun on its own RNA primer, and it is called the lagging strand. The primers are afterwards removed, the gaps filled by DNA polymerase I and the pieces sealed together by DNA ligase, so that one continuous strand results.

Antiparallel templates with synthesis possible only 5' to 3' 1. Leading strand continuous and lagging strand in Okazaki fragments later joined by ligase 1.

- (iii) The messenger RNA has the same sequence as the coding strand, with uracil in place of thymine, so from 5'-ATGTCCAAAGGGTAA-3' the transcript is 5'-AUGUCCAAAGGGUAA-3'. Reading it in codons from the 5' end gives AUG, UCC, AAA, GGG, UAA, that is five codons in fifteen bases. AUG is the start codon and codes for methionine, UCC codes for serine, AAA for lysine and GGG for glycine, while UAA is a stop codon and codes for no amino acid, the chain being released at it. The peptide therefore holds four amino acids, methionine, serine, lysine and glycine.

Correct mRNA sequence 5'-AUGUCCAAAGGGUAA-3' 1. Four amino acids, the fifth codon being a stop codon 1.

- (iv) An exonuclease attacks a nucleic acid only at an end. It removes the nucleotides one at a time from the 3' or the 5' terminus and works its way inwards, so it needs a free end and cannot touch a circular molecule, and it shortens a molecule rather than dividing it. An endonuclease attacks the chain at a point within it, breaking a phosphodiester bond in the interior and so cutting one molecule into two, and it can act on a circular molecule. A restriction enzyme shows endonuclease activity, which is why it is properly called a restriction endonuclease. It cuts within the molecule, at a specific recognition sequence, and so divides the DNA into defined fragments.

Exonuclease and endonuclease correctly distinguished by where they cut 1. Restriction enzyme named as an endonuclease 1.

- (v) The recognition sequence is palindromic, so the double stranded site is 5'-GGATCC-3' paired with 3'-CCTAGG-5'. The cut falls between the first and the second G on each strand, that is G[^]GATCC read from the 5' end of either strand, and because each strand is cut at that same point in its own direction the two cuts are not opposite one another but are staggered by four bases. Writing the left hand fragment with its top strand above its bottom strand, it is 5'-G-3' over 3'-CCTAG-5', so its bottom strand runs on four bases past the end of its top strand. Reading those four unpaired bases from their own 5' end gives GATC. The right hand fragment is 5'-GATCC-3' over 3'-G-5', so here it is the top strand which runs on four bases, and those bases are again GATC read from the 5' end. Each of the two fragments therefore ends in a single stranded stretch of four bases, GATC, carried on the strand whose 5' end is free, which is a four base 5' overhang, or protruding end. This is a sticky end, also called a cohesive end. Since every end cut by BamHI carries the same GATC overhang, the overhang on any one fragment is complementary to that on any other and any two such ends will pair and can then be sealed by ligase.

Staggered cut written out with the four base overhang GATC shown on each end 1. Named as a sticky end with a 5' overhang 1.

- (vi) A plasmid is circular, and a circular molecule gives as many fragments as there are cuts, since the first cut only opens the circle. Two bands therefore mean two fragments and two sites for the enzyme in the plasmid. The whole circle has been accounted for by the two pieces, so the size of the plasmid is the sum of the two fragments, $4.0 + 1.2 = 5.2$ kb. One kilobase is 1000 base pairs, so the plasmid is $5.2 \times 1000 = 5200$ base pairs.

Two sites, with the reason that a circular molecule gives n fragments from n cuts 1. Size 5.2 kb, that is 5200 base pairs 1.

- (vii) Complementary DNA, written cDNA, is a DNA copy made from a mature messenger RNA by the enzyme reverse transcriptase, the RNA of the hybrid formed being then removed and the second DNA strand built to give a double stranded molecule. Because it is copied from the mature messenger, it carries only the sequence which is actually translated. The advantage in making a eukaryotic protein in a bacterium is that a eukaryotic gene in the chromosome is interrupted by introns, and a bacterium has no spliceosome and so cannot remove them. If the genomic gene were cloned, the bacterium would transcribe the introns along with the exons and translate the lot, and the protein would be wrong. The introns have already been spliced out of the messenger, so the cDNA made from it is a continuous coding sequence which the bacterium can transcribe and translate correctly. A second advantage, also accepted, is that a cDNA library made from a particular tissue holds only the genes which that tissue is expressing, and in proportion to how strongly each is expressed, so a gene abundant in that tissue is far easier to find in it than in the whole genome.

cDNA defined as a DNA copy of mRNA made by reverse transcriptase 1. Advantage, that the introns are already removed so a bacterium can express it 1.

- (viii) A shuttle vector is a vector which carries two origins of replication and two selectable markers, one set recognised by one host and the other set by a second host, so that the same recombinant molecule is able to replicate and be selected in either. A common pairing is an origin that works in *Escherichia coli* together with one that works in yeast. Its usefulness lies in letting the two stages of the work be done where each is easiest. The construct is built, multiplied in quantity, checked and stored in *Escherichia coli*, which grows fast, is easy to transform and gives a high yield of plasmid. The very same molecule is then moved into the second host, the eukaryote, where the gene can be expressed in a cell which folds the protein properly and adds the sugar groups a bacterium cannot add. Nothing has to be rebuilt between the two stages, and the construct can be taken back to the bacterium at any time.

Shuttle vector defined as carrying origins and markers for two different hosts 1. Use, that the construct is built and multiplied in the bacterium and then expressed in the eukaryote 1.

- (ix) Somatic hybridisation is the production of a hybrid plant by fusing together the protoplasts of two somatic cells, usually from two different varieties or species, in place of the fusion of gametes. The walls are first stripped from the two sets of cells with cellulase and pectinase so that naked protoplasts are obtained, the two kinds are mixed and made to fuse, and the fusion product, the hybrid protoplast, is cultured so that it builds a new wall, divides to form a callus and is then regenerated into a whole plant, the somatic hybrid. Its value is that it crosses barriers which sexual crossing cannot, since the two parents need not be sexually compatible and need not even flower, and the pomato raised from the potato and the tomato is the familiar example. The agent commonly used to fuse the protoplasts is polyethylene glycol, which brings the membranes into close contact. Electroporation, in which a brief electrical pulse is applied, is the other method in common use, and a solution of sodium nitrate or a high pH calcium solution is also accepted.

Somatic hybridisation defined as the fusion of protoplasts of two somatic cells, regenerated into a plant 1. Polyethylene glycol, or electroporation, named as the fusing agent 1.

- (x) The fold increase in cell number is $1.6 \times 10^5 / 2.0 \times 10^4 = 8$. The population doubles each time it is multiplied by two, and $8 = 2^3$, so the number of population doublings is 3. The doubling time is the time taken for one doubling, that is the total time divided by the number of doublings, $36/3 = 12$ hours. In general the number of doublings n is found from $N_t = N_0 \times 2^n$, so $n = \log_2(N_t/N_0) = \log_2 8 = 3$, and the doubling time is t/n .

Fold increase 8, giving 3 population doublings 1. Doubling time 12 hours 1.

- (xi) Most animal cells taken from a solid tissue are anchorage dependent. In the body they were held in place against a basement membrane or a framework of extracellular matrix, and they need such a surface before they will spread and divide. A cell held in suspension cannot form the attachments through which it senses that support, so it neither spreads nor passes the checkpoint which allows it to divide, and it may die. In culture the attachment proteins of the serum coat the floor of the flask and the cell binds to that coating, spreads over the glass or treated plastic and then divides to give a monolayer. A plant cell has no such requirement. It is held and supported by its own rigid cellulose wall, which does the work the matrix did for the animal cell, so a plant cell will divide freely while suspended in a liquid medium with nothing to attach to. The term used for cells which must have a surface is anchorage dependent, and the cells which do not need one, among them the blood cells and many tumour lines, are called anchorage independent.

Animal cells need the attachment the extracellular matrix gave them, while the plant cell is supported by its own wall 1. Term anchorage dependent 1.

- (xii) Contact inhibition is the behaviour by which a normal cell in culture stops dividing once it is surrounded on all sides by its neighbours. The cells spread over the floor of the vessel until they meet, they then stay as a single layer one cell thick, and when that monolayer is complete and the surface is covered, division ceases even though medium and nutrients are still present. The culture is then said to be confluent, and it must be subcultured if the cells are to go on dividing. Loss of contact inhibition means the cells no longer stop when they touch one another. They continue to divide after the surface is covered, pile up in several layers, form dense heaps and grow over one another, and they will often grow without attaching to any surface at all. This is one of the marks of a transformed or malignant cell, so its loss indicates that the line has been transformed and has acquired the behaviour of a cancer cell. Such cells also lose the limit on the number of divisions they may make and become immortal.

Contact inhibition explained as the stopping of division at confluence in a monolayer 1. Its loss indicates a transformed or malignant cell, which piles up in layers 1.

- (xiii) The accession number is the unique and permanent identifier which a database gives to an entry when that entry is first deposited, such as a GenBank number of the form AF123456 or a Protein Data Bank code of four characters. It belongs to that record for good. It is never reassigned to anything else, and if the entry is corrected or extended the accession number is kept while a version number attached to it is advanced, so the earlier version can still be retrieved. A sequence is cited by its accession number rather than by its name because a name is not a reliable handle. The same protein goes by several names in the literature, different genes in different organisms have been given the same name, names are changed as nomenclature is revised, and a name may be spelt or abbreviated in more than one way. A search by name may therefore return the wrong record, several records or none. The accession number points to one record and one only, so a reader who is given it can retrieve exactly the sequence the author worked on, which is what makes the work reproducible. Journals accordingly require the accession number to be quoted.

Accession number defined as the unique permanent identifier of an entry 1. Reason for citing it, that names are ambiguous and change while the number points to one record 1.

Question 3 [10]

- (a) The three stages are initiation, elongation and termination. The ribosome has two sites at which a transfer RNA may sit, the P site, or peptidyl site, and the A site, or aminoacyl site, with a third site, the E site or exit site, from which the spent transfer RNA leaves. In initiation the small subunit of the ribosome binds the messenger RNA, in a prokaryote at the ribosome binding site just upstream of the start codon, and the initiator transfer RNA carrying methionine pairs with the start codon AUG. The large subunit then joins, and the initiator transfer RNA is placed in the P site, leaving the A site empty over the second codon. Initiation factors and GTP are needed. In elongation the transfer RNA whose anticodon matches the codon now in the A site enters that site with its amino acid. The enzyme peptidyl transferase, which is a ribozyme and belongs to the large subunit, then forms a peptide bond between the amino acid in the A site and the growing chain held in the P site, so the chain is handed over to the transfer RNA in the A site. The ribosome then moves along the messenger by one codon, which is translocation, so that the transfer RNA now carrying the chain is shifted into the P site, the empty transfer RNA passes to the E site and out, and the A site comes to lie over the next codon. The cycle repeats, one amino acid being added each time. In termination the A site comes over a stop

codon, UAA, UAG or UGA. No transfer RNA matches it, a release factor binds instead, the bond between the chain and the last transfer RNA is hydrolysed, and the finished polypeptide is set free. The two subunits then separate from the messenger.

Initiation described with the initiator tRNA at the start codon in the P site 1. Elongation described with entry at the A site, peptide bond formation and translocation 1. Termination at a stop codon with the release factor and the freeing of the polypeptide 1.

- (b) The fork should be drawn as a Y lying on its side. The two parent strands come in from the right as a double helix drawn as two parallel lines, and they part at the fork and run off to the left as two separated templates, one above and one below. Mark the direction in which the fork is moving with an arrow pointing to the right, into the unopened duplex. Mark the polarity of each template, one running 3' to 5' towards the fork and the other 5' to 3' towards it, and label both as parent strands. On the template whose 3' end lies at the fork, draw the new strand as one long unbroken line following the fork closely, with an arrowhead at its 3' end pointing towards the fork, and label it the leading strand. On the other template draw the new strand as a series of short separate pieces, each with its own arrowhead pointing away from the fork, and label the strand the lagging strand and the pieces the Okazaki fragments. A short RNA primer may be drawn at the 5' end of each fragment, a gap left between neighbouring fragments where the ligase has yet to act, and the helicase shown as a wedge at the fork itself.

Fork drawn with the two parent strands separating, their polarities marked, and the direction of movement shown 1. Leading strand drawn continuous towards the fork 1. Lagging strand drawn as Okazaki fragments pointing away from the fork, both strands labelled 1.

- (c) The primary transcript of a eukaryotic gene, the pre messenger RNA, is altered in three ways while it is still in the nucleus. The first is capping. While transcription is still going on, a modified guanine nucleotide, 7-methylguanosine, is added to the 5' end of the transcript in an unusual linkage. The cap protects that end from the exonucleases of the cell, and it is also what the small subunit of the ribosome recognises when the messenger is later to be translated, so without it translation does not begin properly. The second is tailing, or polyadenylation. The transcript is cut a little beyond a signal sequence near its 3' end and a tail of about two hundred adenine residues is added to the cut end by poly-A polymerase, no template being used. The tail protects the 3' end from degradation, helps carry the messenger out of the nucleus and adds to its working life in the cytoplasm. The third is splicing. The coding stretches of a eukaryotic gene, the exons, are separated by non coding stretches, the introns, and both have been transcribed. The spliceosome, a complex of small nuclear RNAs and proteins, recognises the sequences at the boundaries of each intron, cuts the transcript at both ends of the intron, releases the intron as a loop and joins the two flanking exons together. The result is a continuous coding sequence, and only then is the molecule a mature messenger RNA fit to leave the nucleus. Alternative splicing is the joining of the same set of exons in more than one combination, one or more exons being left out of a given mature messenger. It allows one gene to give rise to several different messengers and so to several related but distinct proteins, called isoforms, and which combination is made can be settled according to the tissue or the stage of development. A cell therefore gets far more proteins than it has genes, which is one reason why the number of human genes turned out to be so much smaller than the number of human proteins.

Capping at the 5' end with 7-methylguanosine and its function 1. Polyadenylation at the 3' end and its function 1. Splicing out of the introns by the spliceosome 1. Alternative splicing explained as yielding several proteins from one gene 1.

Question 4 [10]

- (a) The plasmid is circular and 8.0 kb round. The two HindIII sites cut it into two arcs, and the HindIII digest shows these to be 5.0 kb and 3.0 kb, which together make up the whole 8.0 kb as they must. Call the two sites H1 and H2, so that going one way round from H1 to H2 is 5.0 kb and going the other way is 3.0 kb. Adding EcoRI gives three fragments, 4.0 kb, 3.0 kb and 1.0 kb, and these also sum to 8.0 kb. The single EcoRI site must lie within one of the two arcs and must divide that arc into two of the three fragments, leaving the other arc whole as the third fragment. The 3.0 kb fragment appears in both digests unchanged, so the 3.0 kb arc has not been cut and the EcoRI site is not in it. The EcoRI site therefore lies in the 5.0 kb arc, and it has divided that arc into 4.0 kb and 1.0 kb, which is consistent since $4.0 + 1.0 = 5.0$. The site lies 1.0 kb from one

of the two HindIII sites and 4.0 kb from the other, measured along the 5.0 kb arc. The map should be drawn as a circle marked 8.0 kb. Place H1 at the top, labelled HindIII. Going clockwise, mark the EcoRI site 1.0 kb from H1, then H2 a further 4.0 kb on, labelled HindIII, and the remaining 3.0 kb of the circle runs from H2 clockwise back to H1. Mark the three intervals 1.0 kb, 4.0 kb and 3.0 kb. The mirror image, with the EcoRI site 1.0 kb anticlockwise from H1, is the same map viewed from the other side and is equally correct.

The two HindIII arcs identified as 5.0 kb and 3.0 kb, summing to the whole plasmid 1. The 3.0 kb fragment unchanged in both digests, so the EcoRI site lies in the 5.0 kb arc and splits it into 4.0 and 1.0 kb 1. Circular map drawn with both HindIII sites, the EcoRI site and the three intervals marked 1.

- (b) *Agrobacterium tumefaciens* is a soil bacterium which naturally infects a wound on a dicotyledonous plant and causes a crown gall tumour there. It does so by transferring a piece of its large Ti plasmid, the T DNA, into the chromosome of the plant cell, where the genes of that piece are then expressed. The bacterium is therefore used as a natural vector, and the first step is to build the vector. The tumour inducing genes are cut out of the T DNA, which disarms the plasmid so that no gall is formed, and the gene of interest is put in their place together with a promoter that works in plants and a selectable marker, usually a gene for resistance to an antibiotic such as kanamycin. Only the short border sequences at the two ends of the T DNA are needed for the transfer, so everything between them may be replaced. In practice a binary system is often used, the engineered T DNA being carried on one small plasmid while the genes needed to drive the transfer stay on a second, and the construct is first built and multiplied in *Escherichia coli*. The engineered plasmid is then put into *Agrobacterium*. Small pieces of plant tissue, commonly leaf discs, are cut and incubated with the bacterial culture so that the bacteria attach to the cut edges and transfer the T DNA, carrying the gene of interest, into the chromosomes of the plant cells. The explants are then moved to a culture medium holding an antibiotic which kills the remaining bacteria along with the selection agent, so that only the plant cells which took up the T DNA and so carry the marker can grow. Those cells divide to form a callus, and the callus is transferred to media whose auxin and cytokinin balance is adjusted to make it form first shoots and then roots. The plantlets are hardened and moved to soil, and they are tested to confirm that the gene is present and is being expressed. The plants so raised are transgenic, and the gene is passed on to their seed.

Ti plasmid disarmed and the gene of interest with a promoter and marker put between the T DNA borders 1. Plant explants co-cultivated with the Agrobacterium so that the T DNA is transferred into the plant genome 1. Transformed cells selected on the marker and regenerated through callus into whole transgenic plants 1.

- (c) Southern blotting is a method for finding a particular DNA sequence among the many fragments of a digested sample, and it is named after Edwin Southern. The DNA is first cut with one or more restriction enzymes and the fragments are separated by size on an agarose gel by electrophoresis. The gel itself is a poor support for the next steps, so the DNA is then transferred out of it. The gel is treated with alkali, which separates the two strands of every fragment, and a sheet of nitrocellulose or nylon membrane is laid on the gel with buffer drawn through the whole sandwich by capillary action, or by a vacuum, or by an electric field. The single stranded fragments are carried out of the gel on to the membrane and stick to it in exactly the positions they occupied in the gel, so the pattern of the gel is preserved, and they are fixed there by baking or by ultraviolet light. The membrane is then hybridised. It is incubated in a solution holding a labelled probe, which is a single stranded piece of DNA or RNA complementary to the sequence being looked for, the label being a radioactive phosphorus atom or a non radioactive marker. The probe pairs by hydrogen bonding with any fragment on the membrane that carries the complementary sequence and not with the rest, and the unbound probe is then washed away. Finally the position of the bound probe is revealed, by laying the membrane against an X ray film so that the radioactivity marks it, which is autoradiography, or by a colour or chemiluminescent reaction. A band appears only where the sequence sought was present, and its position on the membrane gives the size of the fragment carrying it. The technique is used in DNA fingerprinting, where the probe is directed at the repeated sequences whose pattern differs between individuals, and it is used to detect a gene or a mutation in a patient's DNA, to count the copies of a gene in a genome, and to check that a transgene has integrated into a transformed organism. Any one such use is accepted.

DNA digested and the fragments separated by size on an agarose gel 1. Denatured fragments transferred to a nitrocellulose or nylon membrane in the same pattern and fixed there 1. Hybridisation with a labelled complementary probe and the washing away of the unbound probe 1. Detection by autoradiography or an equivalent, with one correct use of the technique 1.

Question 5 [10]

- (a) The three phases are the lag phase, the log or exponential phase and the stationary or plateau phase. In the lag phase, which follows the seeding of the vessel, the number of cells scarcely rises. The cells are recovering from the trypsin treatment that detached them, settling on to the floor of the vessel, attaching and spreading, and restoring the cytoskeleton and the surface proteins they lost. They are also adjusting to the fresh medium and building up the enzymes and the materials needed before they can divide, so the work of this phase is preparation rather than multiplication. In the log phase the cells divide steadily and the number rises exponentially, doubling at a constant interval which is the doubling time of the line. Nutrients are plentiful, no waste has yet accumulated, there is still free surface to spread on, and the proportion of cells actually dividing is at its highest. This is the healthiest and most uniform state of the culture, and it is the phase from which cells are taken for an experiment or for subculturing. In the stationary phase the rise levels off and the number of cells stays nearly constant. The surface is now fully covered, so the cells are confluent and contact inhibition stops further division, while the nutrients are running low, the medium is becoming acid and toxic products have accumulated. Division has not stopped altogether but the rate at which cells divide is matched by the rate at which they die. If the culture is left longer it enters a decline, the cells detaching and dying, so a stock must be subcultured into fresh medium before that point.

Lag phase described as attachment and adjustment with little increase in number 1. Log phase described as exponential division with a constant doubling time 1. Stationary phase described as confluence with nutrients exhausted and division balanced by death 1.

- (b) A suspension culture is a culture in which single cells and small clumps of cells grow freely dispersed in a liquid medium. It is raised from a callus. A friable callus, that is one which crumbles readily, is first obtained on a solid medium, and pieces of it are transferred to a flask of the same medium made up without agar. The flask is placed on a rotary shaker and agitated continuously at about a hundred revolutions a minute. The shaking breaks the clumps apart, keeps the cells dispersed and in contact with the medium, and sees that oxygen reaches them, since a settled clump would soon be starved of it. The culture is kept at about 25 degrees Celsius. It is maintained by subculturing at regular intervals, a measured volume of the suspension being transferred into fresh medium before the nutrients are used up, and it may also be run as a continuous culture in which medium is added and withdrawn steadily so that the cells are held in the exponential phase. For a larger volume the shake flask is replaced by a bioreactor with a stirrer, a supply of air and controls for temperature and pH. Such a culture is used for the production of secondary metabolites, that is the alkaloids, pigments, flavours, enzymes and medicinal compounds which the plant makes, since the cells can be grown in bulk in a vessel all the year round in place of harvesting the plant. It is also used as the starting material for the isolation of protoplasts, for raising a large number of single cells for mutant selection and transformation, for somatic embryogenesis and the production of synthetic seed, and for studying the metabolism and the growth of plant cells under conditions which can be controlled exactly. Any two of these are accepted.

Friable callus transferred to a liquid medium without agar 1. Continuous agitation on a shaker to disperse the cells and supply oxygen, with subculturing to maintain it 1. Two correct uses, such as the production of secondary metabolites and the isolation of protoplasts 1.

- (c) (1) Anther culture is the culture of an immature anther, or of the pollen grains taken out of it, on a nutrient medium, so that the pollen grain is diverted from its ordinary course. Instead of giving rise to a pollen tube it divides repeatedly and forms a callus or an embryoid, which is then regenerated into a plant. Because the pollen grain is a product of meiosis and carries only one set of chromosomes, the plant obtained is haploid. Such a plant is sterile, and its chromosome number is therefore commonly doubled with colchicine to give a homozygous diploid, a double haploid, which a breeder values because every gene in it is in the homozygous state and a recessive character cannot be hidden. (2) A plant infected by a virus is usually not infected in every cell. The apical meristem has no vascular connection of its own and divides very fast, and the virus spreads chiefly through the vascular tissue, so the virus does not keep pace with the dividing cells at the very tip and the youngest cells of the dome are often free of it. A very small piece of that apical meristem is therefore dissected out under the microscope and cultured on a nutrient medium, where it grows into a shoot and is then rooted to give a whole plant. The plant so raised is free of the virus although the parent stock was infected, and it can be multiplied to give virus free planting material. Heat treatment of the stock before the

meristem is taken improves the chance of success. (3) A bioreactor, or fermenter, is a closed vessel in which cells, or an enzyme, are held under controlled conditions so that a reaction can be carried out on a large scale and a product obtained in quantity. It supplies the cells with the medium, mixes the contents so that the cells, the nutrients and the oxygen are evenly distributed, and holds the conditions steady while the product forms, and it is built so that it can be sterilised and the product harvested without contamination. Two conditions it must control are the temperature and the pH, and the dissolved oxygen, the rate of stirring and the rate at which the substrate is fed in are also accepted. (4) Cryopreservation is the storage of cells, tissues, embryos, seeds or whole organs at a very low temperature, at which all the metabolism of the cell is brought to a stop so that the material can be kept alive without change for years and revived when it is wanted. Before freezing, a cryoprotectant such as dimethyl sulphoxide or glycerol is added and the cooling is done slowly, so that water leaves the cells and ice crystals, which would pierce the membranes, do not form within them. The material is held in liquid nitrogen at minus 196 degrees Celsius, or in the vapour just above it at about minus 150 degrees Celsius.

Anther culture explained as giving haploid plants from pollen 1. Meristem culture explained, the apical cells being free of the virus 1. Bioreactor defined with two conditions it controls 1. Cryopreservation defined with the temperature of liquid nitrogen, minus 196 degrees Celsius 1.

Question 6 [10]

- (a) A phylogenetic tree is a branching diagram which sets out the inferred evolutionary relationships among a group of genes, proteins, populations or species, drawn so that the arrangement of the branches shows which of them are more closely related to one another. The organisms or sequences compared are placed at the tips of the branches. A branch point, or node, stands for the last common ancestor of everything that lies beyond it, that is the point at which one ancestral lineage divided into two, so the order of the nodes gives the order in which the lineages separated. The length of a branch, in a tree drawn to scale, stands for the amount of change along that lineage, which is the number of substitutions per site, and where the rate of change can be taken as roughly constant it stands in proportion to the time elapsed. A tree drawn without scaled branch lengths, a cladogram, shows only the pattern of branching and nothing about amount or time. Before a tree can be built from sequence data, the sequences to be compared must first be aligned, by a multiple sequence alignment, so that the positions set against one another in the comparison are positions which correspond to one another, having descended from the same position in the common ancestor. Everything that follows is computed from the differences at those aligned positions, so an unreliable alignment gives an unreliable tree whatever method is then used. The sequences must also be homologous, that is genuinely descended from a common ancestral sequence, or the tree means nothing.

Phylogenetic tree defined as a branching diagram of inferred relationships 1. Node as the common ancestor and branch length as the amount of evolutionary change 1. A multiple sequence alignment of homologous sequences needed first 1.

- (b) The Human Genome Project was the international programme which set out to determine the complete sequence of the three thousand million base pairs of the human genome, to identify all the genes within it and to map their positions on the twenty three pairs of chromosomes. It also undertook to store the information in databases open to everyone, to develop the tools needed to analyse it, to sequence a set of model organisms for comparison, and to examine the ethical, legal and social questions the information would raise. It ran for over a decade, and the finished sequence was published at the end of it. Among its findings are these. The human genome holds roughly twenty to twenty five thousand protein coding genes, far fewer than the hundred thousand which had been expected, and one gene therefore yields several proteins through alternative splicing and later modification. Less than two per cent of the genome actually codes for protein, the rest being regulatory sequence, non coding RNA genes and a very large amount of repeated sequence. The genes are distributed unevenly, chromosome 1 carrying the most and the Y chromosome the fewest, and about half the genome consists of repetitive elements. Any two findings are accepted, and the following are also correct. The sequences of any two unrelated people are more than ninety nine per cent identical, the differences being largely single nucleotide polymorphisms, of which millions have been mapped. Many human genes have close counterparts in other organisms, and the function of a great many was assigned by that resemblance. The data are now used in the diagnosis of inherited disease and in genetic testing and counselling, the gene

responsible being known and its sequence available. They are used in the identification of targets for new drugs and in pharmacogenomics, where a patient's own variants are used to predict how that patient will respond to a drug and at what dose, which is the basis of personalised medicine. Any two uses are accepted, and the following are also correct. They are used in comparative genomics and in the study of human evolution and migration. They are used in forensic identification. They underpin the whole of modern bioinformatics, since the human sequence is the reference against which new sequence data are read.

Aim stated as sequencing the whole human genome and identifying and mapping its genes 1. Two correct findings 1. Two correct uses of the data 1.

- (c) (1) A DNA microarray is a small solid support, a glass slide or a silicon chip, on which very large numbers of single stranded DNA probes have been fixed in an ordered grid, each spot in the grid holding many copies of the probe for one particular gene and the position of every spot being known. Labelled nucleic acid prepared from a sample is washed over the array and hybridises to the spots whose probes it matches, and the fluorescence at each spot is then read by a scanner. What the array measures is the amount of the nucleic acid in the sample which corresponds to each probe. In its commonest use, labelled cDNA made from the messenger RNA of a tissue is applied, so the brightness of a spot measures how strongly that gene was being expressed, and thousands of genes are measured in one experiment. Two samples labelled with different dyes may be compared on one array, which shows at a glance which genes are turned up and which turned down between a diseased tissue and a healthy one. Arrays are used in the same way to type single nucleotide polymorphisms and to detect mutations. (2) In a bacterium the coding sequence of a gene is continuous, so a long open reading frame running from a start codon to a stop codon is itself close to being the gene, and prediction is largely a matter of finding such frames. A eukaryotic gene is broken up by introns, so the coding sequence is in pieces and the reading frame is interrupted. The software must therefore find the exons separately and work out which of them belong to one gene, which means it must recognise the splice signals at the start and the end of each intron, the donor and the acceptor sites, settle the boundary of each exon, and then join the exons in the right order and the right frame to give the predicted protein. It must also allow for one gene yielding more than one transcript through alternative splicing, and it must contend with genes which are few and far apart in a very large genome holding much repeated sequence, where a bacterial genome is densely packed with genes. (3) Molecular docking is the computational prediction of how one molecule will fit against another, most often how a small molecule, a possible drug, will sit in the binding site of a target protein. The programme is given the three dimensional structure of the target and of the candidate, places the candidate in the site in a great many orientations and conformations, and scores each arrangement by estimating the strength of the interaction, the hydrogen bonds, the electrostatic attraction and the shape complementarity, taken together. It returns the most favourable pose and a score which estimates the affinity. Its value is that libraries of many thousands of compounds can be docked in the computer, so only the few that score well need be synthesised and tested, which saves a great deal of laboratory work. (4) The volume of data has grown faster than the means of handling it, so a user needs computing power, storage and some skill in programming merely to work with a whole genome, and the simple searches which once sufficed now return more hits than can be examined. The quality of the data is uneven. A large part of the annotation in the databases was assigned not by experiment but by resemblance to another entry, so an error once made is copied from entry to entry, and a user cannot always tell which assignments were tested and which inferred. Any two are accepted, and the following are also correct. The data are scattered across a great many separate databases which use different formats, different identifiers and different names for the same thing, so bringing them together is laborious and redundant entries for one gene are common. Entries are added, corrected and withdrawn continually, so a result obtained today may not be reproducible against the database as it stands next year unless the version used is recorded.

DNA microarray described as an ordered grid of probes, measuring gene expression by hybridisation 1. Eukaryotic gene prediction must find exons and splice sites because the reading frame is interrupted by introns 1. Molecular docking defined as the computational fitting and scoring of a ligand in a binding site 1. Two correct problems created by the growth of the databases 1.