

ISC Class 12 Biotechnology Specimen Paper 1

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) In the double helix of DNA the two strands are held together by:
(a) phosphodiester bonds (b) hydrogen bonds between the bases
(c) peptide bonds (d) disulphide bridges
- (ii) Name the enzyme which relieves the supercoiling ahead of the replication fork in a bacterium.
- (iii) The enzyme which lays down a short RNA stretch to start a new DNA strand is:
(a) DNA ligase (b) helicase
(c) primase (d) DNA polymerase I
- (iv) State how many codons of the genetic code are stop codons and name any one of them.
- (v) The recognition sequence of the restriction endonuclease EcoRI is:
(a) 5'-AAGCTT-3' (b) 5'-GGATCC-3'
(c) 5'-CTGCAG-3' (d) 5'-GAATTC-3'
- (vi) Name the enzyme used to make a DNA copy of a messenger RNA molecule.
- (vii) The two antibiotic resistance genes carried by the plasmid pBR322 confer resistance to:
(a) ampicillin and tetracycline (b) kanamycin and streptomycin
(c) chloramphenicol and neomycin (d) penicillin and rifampicin
- (viii) Name the method by which DNA is shot into intact plant cells on the surface of fine metal particles.
- (ix) The plant tissue culture medium most widely used for the culture of the cells of higher plants is:
(a) Dulbecco's modified Eagle's medium (b) Murashige and Skoog medium
(c) Luria-Bertani medium (d) nutrient agar
- (x) Name the enzyme used to strip the cell wall from a plant cell so that a protoplast is obtained.
- (xi) A cell line which has been so altered that it will divide without limit in culture is said to be:
(a) primary (b) senescent

(c) immortalised (d) anchorage dependent

(xii) Name the simple text based format in which a nucleotide or protein sequence is held for submission to a sequence database.

(xiii) The programme used to search a database for sequences similar to a query sequence is:

- (a) BLAST (b) PyMOL
(c) RasMol (d) Cn3D

(xiv) Name the international database in which the three dimensional structures of proteins are deposited.

SECTION B (26 Marks)

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

Question 2 [26]

- (i) A double stranded DNA molecule carries 1200 base pairs, of which 360 are guanine-cytosine pairs. Calculate the number of adenine bases and the percentage of thymine in the molecule. [2]
- (ii) State the two features of DNA replication which the term semiconservative records, and name the two scientists who established it. [2]
- (iii) Write the sequence of the messenger RNA transcribed from the template strand 3'-TACGGCAATCGA-5' and give the first two amino acids it codes for. [2]
- (iv) A linear DNA molecule carries three EcoRI sites. State the number of fragments obtained when it is completely digested, and state how many would be obtained if the same molecule were circular. [2]
- (v) Distinguish between a sticky end and a blunt end, and name one restriction enzyme which leaves each. [2]
- (vi) Explain why a restriction enzyme does not cut the DNA of the bacterium which produces it. [2]
- (vii) A sample holding 5 copies of a target DNA sequence is put through 25 cycles of the polymerase chain reaction with an efficiency of 100 per cent. Calculate the number of copies of the target present at the end. (Take $2^{25} = 3.355 \times 10^7$.) [2]
- (viii) State two properties which a plasmid must possess if it is to serve as a cloning vector. [2]
- (ix) Explain why calcium chloride is used before a plasmid is introduced into Escherichia coli, and give the term for a cell which has been made ready in this way. [2]
- (x) A 1 cm^3 sample of a bacterial culture is diluted one hundred thousand fold, and 0.1 cm^3 of the dilution spread on an agar plate gives 64 colonies. Calculate the number of colony forming units per cm^3 of the original culture. [2]
- (xi) State the function of the laminar flow cabinet and the function of the carbon dioxide incubator in an animal cell culture laboratory. [2]
- (xii) Explain what totipotency is and state why it matters in plant tissue culture. [2]
- (xiii) Distinguish between a pairwise and a multiple sequence alignment, and state what each is used for. [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 transcription of a gene in a prokaryote, naming the enzyme and the signals at which it begins and ends. [3]
- (b) Draw and label a diagram of a transfer RNA molecule in its cloverleaf form, showing the four arms and the two sites at which it does its work. [3]

- (c) Describe the structure of the lac operon of *Escherichia coli*, naming its parts, and explain how lactose switches the operon on. [4]

Question 4 [10]

- (a) A linear DNA fragment 5.0 kb long carries one *EcoRI* site and one *HindIII* site. Digestion with *EcoRI* alone gives fragments of 1.5 kb and 3.5 kb, digestion with *HindIII* alone gives fragments of 0.8 kb and 4.2 kb, and digestion with both together gives fragments of 0.8 kb, 0.7 kb and 3.5 kb. Work out the position of each site measured from the left hand end and draw the restriction map. [3]
- (b) Describe the three steps of one cycle of the polymerase chain reaction, giving the temperature at which each is carried out and what happens at it. [3]
- (c) Describe how a gene of interest is cloned in a plasmid vector, taking the account from the cutting of the two DNA molecules to the selection of the recombinant colonies. [4]

Question 5 [10]

- (a) State three ways in which a medium for the culture of animal cells differs from a medium for the culture of plant cells. [3]
- (b) Describe how a primary culture of animal cells is raised from a piece of tissue, and distinguish between a primary culture and a cell line. [3]
- (c) Describe the preparation of a monoclonal antibody by the hybridoma method, naming the two kinds of cell fused, the agent used to fuse them and the medium in which the hybrid cells alone survive. [4]

Question 6 [10]

- (a) Explain what a sequence database is, name two primary nucleotide sequence databases, and distinguish a primary from a secondary database. [3]
- (b) Explain what a BLAST search returns, and state what the E value of a hit tells the user. [3]
- (c) (1) Explain what is meant by the annotation of a genome. (2) Explain what an open reading frame is and how it is recognised in a sequence. (3) Explain what homology modelling of a protein is and what it needs in order to work. (4) Name two uses of bioinformatics in the design of a drug. [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) hydrogen bonds between the bases. Two hydrogen bonds join adenine to thymine and three join guanine to cytosine. The phosphodiester bonds run along each strand and do not hold one strand to the other.

Correct option 1.

- (ii) DNA gyrase, a type II topoisomerase, which introduces negative supercoils and so takes out the strain built up ahead of the fork as the helix is unwound.

Correct answer 1.

- (iii) (c) primase. DNA polymerase cannot start a strand from nothing and must have a free 3' hydroxyl group to add to, so primase first lays down a short RNA primer.

Correct option 1.

- (iv) Three codons are stop codons. They are UAA, UAG and UGA, and any one of them is accepted.

Three, with one stop codon correctly named 1.

- (v) (d) 5'-GAATTC-3'. The sequence is a palindrome, reading the same on both strands in the 5' to 3' direction, and EcoRI cuts between the G and the A on each strand to leave a four base 5' overhang.

Correct option 1.

- (vi) Reverse transcriptase, an RNA dependent DNA polymerase, which gives complementary DNA, that is cDNA.

Correct answer 1.

- (vii) (a) ampicillin and tetracycline. The two genes are bla, which codes for the beta lactamase that destroys ampicillin, and tet.

Correct option 1.

- (viii) Biolistics, also called the gene gun or the particle gun method, in which DNA coated tungsten or gold particles are fired at the cells at high velocity.

Correct answer 1.

- (ix) (b) Murashige and Skoog medium, usually written MS medium.

Correct option 1.

- (x) Cellulase, which digests the cellulose of the wall. Pectinase, or macerozyme, is used along with it to loosen the cells from one another, and either enzyme is accepted.

Correct answer 1.

- (xi) (c) immortalised. Such a cell line has escaped the limited number of divisions a normal cell is allowed and will go on dividing indefinitely, so it is called a continuous or established cell line.

Correct option 1.

- (xii) The FASTA format, in which a single header line beginning with a greater than sign carries the name of the sequence and the lines below it carry the sequence itself in single letter code.

Correct answer 1.

- (xiii) (a) BLAST, which stands for Basic Local Alignment Search Tool.

Correct option 1.

- (xiv) The Protein Data Bank, usually written PDB.

Correct answer 1.

Question 2 [26]

- (i) Of the 1200 base pairs, 360 are G-C pairs, so the remaining $1200 - 360 = 840$ are A-T pairs. Each A-T pair holds one adenine base, so the number of adenine bases is 840. Each pair holds one thymine as well, so

there are 840 thymine bases. The total number of bases in the double stranded molecule is $2 \times 1200 = 2400$. Percentage of thymine = $(840/2400) \times 100 = 35$ per cent. By Chargaff's rule the percentage of adenine is also 35, and guanine and cytosine are each 15 per cent.

Adenine bases 840 1. Thymine 35 per cent 1.

- (ii) The first feature is that the two strands of the parent molecule separate and each acts as a template, so that each of the two daughter molecules carries one whole strand which came unchanged from the parent. That is the conserved half. The second is that the partner of that old strand is built fresh from free nucleotides, so each daughter molecule is half old and half new, and no daughter molecule holds two old strands or two new ones. Semiconservative replication was established by Matthew Meselson and Franklin Stahl, who grew *Escherichia coli* in a medium holding the heavy isotope of nitrogen and followed the density of the DNA through successive generations on a caesium chloride gradient.

Each daughter molecule shown to carry one parental strand and one new strand 1. Meselson and Stahl named 1.

- (iii) The messenger RNA is built antiparallel to the template and complementary to it, with uracil in place of thymine. Reading the template 3'-TACGGCAATCGA-5' gives the transcript 5'-AUGCCGUUAGCU-3'. The first codon is AUG, which codes for methionine and is also the start codon, and the second codon is CCG, which codes for proline. The first two amino acids are therefore methionine and proline.

Correct mRNA sequence 5'-AUGCCGUUAGCU-3' 1. Methionine and proline 1.

- (iv) A linear molecule has two free ends. Each cut separates the molecule into one more piece than before, so three cuts on a linear molecule give $3 + 1 = 4$ fragments. A circular molecule has no free end, so the first cut merely opens the circle into one linear piece and only the cuts after it add fragments. Three cuts on a circular molecule therefore give 3 fragments. In general a linear molecule gives $n + 1$ fragments and a circular molecule gives n fragments from n cuts.

Linear molecule 4 fragments 1. Circular molecule 3 fragments 1.

- (v) A sticky end is produced when the enzyme cuts the two strands of its recognition sequence at points which are not opposite one another, so that each fragment is left with a short single stranded overhang. Because the recognition sequence is a palindrome, the overhangs on any two fragments cut by the same enzyme are complementary, so they pair by hydrogen bonding with one another and the fragments can be joined. EcoRI leaves sticky ends, cutting 5'-G^AAATTC-3' to leave a four base 5' overhang, and BamHI, HindIII and PstI are also accepted. A blunt end is produced when the enzyme cuts both strands at the same position, straight across the axis, so that no overhang is left and the end is fully double stranded. SmaI leaves blunt ends, cutting 5'-CCC^AGGG-3', and AluI and HaeIII are also accepted. Blunt ends will join to any other blunt end, which is convenient, but the joining is far less efficient than the joining of matched sticky ends.

Sticky end defined as a staggered cut leaving a complementary overhang, with a correct enzyme named 1. Blunt end defined as a straight cut, with a correct enzyme named 1.

- (vi) The bacterium protects its own chromosome by marking the recognition sequences in it. A modification enzyme, a methylase, which is produced together with the restriction enzyme and recognises the same sequence, adds a methyl group to a particular base within every copy of that sequence in the bacterium's own DNA. The restriction enzyme no longer recognises the methylated site and so cannot cut it. The DNA of an invading bacteriophage carries no such methyl groups, so its sites are cut and the phage is destroyed. The restriction enzyme and its methylase working together are called a restriction modification system, and this is how the cell tells its own DNA from foreign DNA.

Methylation of the recognition sites in the host's own DNA by a methylase 1. Unmethylated foreign DNA cut, so the host DNA is spared 1.

- (vii) Each cycle doubles the number of copies of the target, so after n cycles the number of copies is $N_0 \times 2^n$. Here $N_0 = 5$ and $n = 25$, so the number of copies = $5 \times 2^{25} = 5 \times 3.355 \times 10^7 = 1.678 \times 10^8$, that is about 168 million copies. In practice the yield is lower, because the efficiency falls below 100 per cent as the reagents are used up and the reaction reaches a plateau.

Correct use of $N_0 \times 2^n$ with $n = 25$ 1. Number of copies 1.678×10^8 1.

- (viii) First, it must carry an origin of replication, so that once it is inside the host the plasmid is copied by the machinery of the host and is passed on to the daughter cells. A high copy number is an advantage, since it raises the yield of the cloned gene. Second, it must carry a selectable marker, most often a gene for resistance to an antibiotic, so that the few cells which took up the plasmid can be picked out from the very many which did not, simply by growing the whole lot on a medium holding that antibiotic. Any two of the following are accepted in place of these. It must be small, since a small plasmid enters a cell more readily and is easier to handle. It must carry unique recognition sites for a number of restriction enzymes, so that it can be opened at one point only and the foreign DNA put in there. It should carry a means of telling a recombinant plasmid from one which took up no insert.

Origin of replication 1. Selectable marker, or any other correct property 1.

- (ix) The cell wall and the membrane of Escherichia coli will not let a molecule as large as a plasmid through, and both the plasmid and the surface of the cell carry a negative charge, so they repel one another. The cells are therefore suspended in a cold solution of calcium chloride. The divalent calcium ions screen the negative charges and bridge between the phosphates of the DNA and those of the membrane, so the plasmid is held against the surface, and the treatment with the cold salt solution also makes the wall more porous. The cells are then mixed with the plasmid and given a brief heat shock at about 42 degrees Celsius, which carries the DNA across. A cell treated in this way, and therefore able to take up foreign DNA, is said to be competent.

Calcium ions screen the negative charges and make the wall permeable to the plasmid 1. The cell is termed competent 1.

- (x) Each colony on the plate has grown from one colony forming unit in the volume spread. Number of cfu in the 0.1 cm^3 spread = 64. Number of cfu per cm^3 of the dilution = $64/0.1 = 640$. The dilution was one hundred thousand fold, that is a factor of 10^5 , so the original culture held $640 \times 10^5 = 6.4 \times 10^7$ cfu per cm^3 , that is 64 million colony forming units per cm^3 .

Correction for the 0.1 cm^3 plated, giving 640 cfu per cm^3 of the dilution 1. Multiplication by the dilution factor 10^5 to give 6.4×10^7 cfu cm^{-3} 1.

- (xi) The laminar flow cabinet provides the sterile working space. A blower draws room air through a high efficiency particulate air filter, which holds back bacteria, fungal spores and dust, and delivers it as a smooth unmixing stream of clean air across the work surface and outward towards the worker. Because the stream does not swirl, nothing from the room can drift back on to the open culture, so flasks may be opened and pipetting done without the culture being contaminated. The carbon dioxide incubator holds the culture at the conditions of the body. It keeps the temperature at 37 degrees Celsius, the humidity high so that the medium does not dry out or grow concentrated, and the atmosphere at about five per cent carbon dioxide. That carbon dioxide works with the bicarbonate of the medium as a buffer and holds the pH of the medium near 7.4, which the cells need if they are to grow.

Laminar flow cabinet described as giving a sterile HEPA filtered air stream over the work 1. Carbon dioxide incubator described as holding 37 degrees Celsius and about five per cent carbon dioxide to buffer the pH 1.

- (xii) Totipotency is the capacity of a single living cell to divide and differentiate so as to give rise to the whole organism, that is to produce every kind of cell the organism is made of. In plants most living cells keep this capacity through life, which animal cells do not. It matters in plant tissue culture because it is what the whole method rests on. A small explant, or even one cell, placed on a suitable nutrient medium can therefore be made to form a callus and then, as the balance of auxin to cytokinin is changed, to form shoots and roots and so a complete plantlet. Without totipotency no amount of nutrient would turn a piece of leaf or stem into a plant, and micropropagation, the raising of large numbers of genetically identical plants, would not be possible.

Totipotency defined as the capacity of one cell to give the whole organism 1. Its importance, that a small explant can be made to regenerate a complete plant 1.

- (xiii) A pairwise alignment compares exactly two sequences, laying one above the other and inserting gaps where necessary so that the matching residues line up. It yields a score and a percentage identity, and it is used to measure how closely two sequences are related, to find the region two sequences share, and in a database search, where the query is aligned pairwise with every sequence in turn. A multiple sequence alignment compares three or more sequences together in one alignment. It is used to find the positions which

have been conserved across a whole family, since a residue which is the same in every member is likely to matter for the structure or the function, to identify the motifs and domains which define the family, to assign a newly found sequence to a family, and to build the phylogenetic tree which shows how the members are related. ClustalW and Clustal Omega are the programmes commonly used for it.

Pairwise alignment defined with a correct use 1. Multiple alignment defined with a correct use such as conserved motifs or phylogeny 1.

Question 3 [10]

- (a) The enzyme is DNA dependent RNA polymerase, and in a prokaryote one such enzyme carries out the whole of transcription. The three stages are initiation, elongation and termination. In initiation the holoenzyme, which is the core enzyme together with its sigma factor, recognises and binds the promoter, a sequence lying upstream of the gene. In *Escherichia coli* the promoter carries two conserved stretches, one about ten bases and one about thirty five bases before the point at which transcription starts. The sigma factor is what gives the enzyme its ability to find the promoter. The enzyme then unwinds a short length of the helix to expose the template strand. In elongation the sigma factor leaves and the core enzyme moves along the template in the 3' to 5' direction, so that the RNA it builds grows in the 5' to 3' direction. Ribonucleotides are added opposite the template bases, uracil being put in wherever the template carries adenine, and the enzyme needs no primer. The helix closes again behind it. In termination the enzyme reaches a terminator sequence. The RNA transcribed from it folds back on itself into a hairpin followed by a run of uracils, which weakens the hold of the enzyme, and the enzyme, the RNA and the DNA come apart. At some terminators the protein rho is needed to bring this about. In a prokaryote the transcript needs no processing and translation may begin on it while transcription is still going on.

DNA dependent RNA polymerase named and initiation described at the promoter with the sigma factor 1. Elongation in the 5' to 3' direction along the 3' to 5' template 1. Termination at the terminator sequence 1.

- (b) The cloverleaf should be drawn as a single RNA chain folded back on itself, with four arms arranged round a central point. Each arm is drawn as a stem of paired bases carrying a loop at its end, except the first. The acceptor arm, or amino acid arm, is drawn at the top with the two ends of the chain in it. It has no loop, and the 3' end projects as the unpaired bases CCA, to which the amino acid is attached. This is the first of the two working sites, the amino acid attachment site. The D arm, or DHU arm, is drawn on the left, its loop holding the modified base dihydrouridine. The anticodon arm is drawn at the bottom, and its loop carries the anticodon, the three bases which pair with the codon of the messenger RNA. This is the second of the two working sites. The T arm, or T ψ C arm, is drawn on the right, its loop holding thymine, pseudouridine and cytosine. A small variable loop may be drawn between the anticodon arm and the T arm. The labels asked for are the four arms, the amino acid attachment site at the 3' CCA end and the anticodon.

Four arms correctly drawn and named 1. Amino acid attachment site shown at the 3' CCA end of the acceptor arm 1. Anticodon shown in the loop of the anticodon arm 1.

- (c) The lac operon is the stretch of the *Escherichia coli* chromosome which carries the genes for the use of lactose, grouped so that all of them are transcribed together as one messenger RNA under one control. Upstream of the operon lies the regulator gene *i*, which is not part of the operon and is transcribed on its own. It codes for the repressor protein. Within the operon come the promoter *p*, at which RNA polymerase binds, and then the operator *o*, a short sequence overlapping the start of transcription to which the repressor binds. After the operator come the three structural genes in order. Gene *z* codes for beta galactosidase, the enzyme which splits lactose into glucose and galactose. Gene *y* codes for lactose permease, which carries lactose across the membrane into the cell. Gene *a* codes for transacetylase. When no lactose is present the repressor made by gene *i* is free and active. It binds tightly to the operator, and because the operator lies next to the promoter the bound repressor blocks the path of RNA polymerase, so no messenger RNA is made and none of the three enzymes appears. The operon is switched off, and because it is held off until something switches it on it is an inducible operon. When lactose is supplied, a little of it is carried in by the few permease molecules always present and a little is converted by beta galactosidase into allolactose. Allolactose is the true inducer. It binds to the repressor protein at a second site, which changes the shape of the repressor so that it can no longer hold the operator. The repressor falls away, the operator is cleared, RNA polymerase transcribes the three genes as one messenger, and the three enzymes are made so that the cell can use the

lactose. The inducer has therefore worked not by acting on the genes but by taking the repressor off them. When the lactose is used up, no allolactose is left, the repressor becomes active once more and binds the operator again, and the operon is switched off.

Regulator gene i with its repressor, and the promoter and operator named 1. Structural genes z, y and a named with the product of each 1. Repressor bound to the operator blocking RNA polymerase when no lactose is present 1. Allolactose as the inducer, binding the repressor and lifting it off the operator 1.

Question 4 [10]

- (a) Take the left hand end as position 0 and the right hand end as position 5.0 kb. The EcoRI digest gives 1.5 kb and 3.5 kb, so the EcoRI site is either 1.5 kb from the left end or 3.5 kb from the left end. The HindIII digest gives 0.8 kb and 4.2 kb, so the HindIII site is either 0.8 kb or 4.2 kb from the left end. The double digest settles the matter. It gives 0.8 kb, 0.7 kb and 3.5 kb. Take the EcoRI site at 1.5 kb and the HindIII site at 0.8 kb. The two cuts then divide the molecule at 0.8 and at 1.5, giving pieces of 0.8 kb from 0 to 0.8, 0.7 kb from 0.8 to 1.5, and 3.5 kb from 1.5 to 5.0. These are exactly the three fragments observed, so this is the map. The other arrangements may be ruled out. With EcoRI at 1.5 and HindIII at 4.2 the fragments would be 1.5, 2.7 and 0.8 kb, with EcoRI at 3.5 and HindIII at 0.8 they would be 0.8, 2.7 and 1.5 kb, and with EcoRI at 3.5 and HindIII at 4.2 they would be 3.5, 0.7 and 0.8 kb. The last of these gives the same three sizes and is the mirror image of the accepted map, read from the other end, so it is the same map and is accepted. The map should be drawn as a horizontal line 5.0 kb long, marked 0 at the left and 5.0 at the right, with an arrow at 0.8 labelled HindIII and an arrow at 1.5 labelled EcoRI, and the three intervals marked 0.8 kb, 0.7 kb and 3.5 kb.

Two possible positions written down for each site from the single digests 1. Double digest used to place HindIII at 0.8 kb and EcoRI at 1.5 kb 1. Map drawn to scale with both sites and the three intervals marked 1.

- (b) One cycle has three steps. The first is denaturation, carried out at about 94 to 95 degrees Celsius for half a minute to a minute. At this temperature the hydrogen bonds between the two strands of the target DNA break and the double helix comes apart into two single strands, each of which is now free to act as a template. The second is annealing, or primer binding, carried out at about 50 to 65 degrees Celsius for half a minute. The reaction is cooled so that the two primers, which are short single stranded oligonucleotides complementary to the sequences flanking the target on the two strands, can pair with those sequences. The exact temperature is chosen to suit the primers, since too high a temperature leaves them unbound and too low a temperature lets them bind at the wrong places. The third is extension, or elongation, carried out at about 72 degrees Celsius for a minute or so, this being the temperature at which Taq polymerase works best. The enzyme adds deoxyribonucleotides to the 3' end of each bound primer and builds a new strand along each template, so that the region between the two primers is copied. At the end of the cycle each original molecule has become two, and the three steps are repeated, the number of copies of the target doubling at every cycle.

Denaturation at about 94 degrees Celsius separating the strands 1. Annealing at about 50 to 65 degrees Celsius binding the primers 1. Extension at about 72 degrees Celsius by Taq polymerase 1.

- (c) The gene of interest and the plasmid vector are cut with the same restriction endonuclease, one which has a unique site in the vector. Because both molecules are cut by the same enzyme, the sticky ends left on the gene are complementary to those left on the opened vector, so the two will pair. The two sets of fragments are then mixed and treated with DNA ligase. The complementary overhangs come together by hydrogen bonding, and the ligase seals the nicks in the sugar phosphate backbone to give a single closed recombinant plasmid which now carries the foreign gene. The recombinant plasmid is next introduced into the host, usually *Escherichia coli*. The host cells are made competent by treatment with cold calcium chloride, mixed with the plasmid and given a brief heat shock, and a small proportion of them take the plasmid up. This step is transformation. Finally the recombinants are selected. The whole mixture is spread on a medium holding the antibiotic to which the vector's marker gene gives resistance. Only cells which took up a plasmid grow, and all the untransformed cells die. Among the survivors, those carrying an insert must still be told from those whose plasmid closed up again empty, and this is done by the second marker. If the gene was inserted into a second antibiotic resistance gene, that gene is broken and the recombinants have lost resistance to the second antibiotic, which is insertional inactivation. If it was inserted into the gene for beta galactosidase, the recombinant colonies cannot break down the chromogenic substrate in the medium and appear white, while the colonies with no insert appear blue. The colonies selected are then grown in quantity, and the cloned gene

or its product is harvested.

Gene and vector cut with the same restriction enzyme to give matching sticky ends 1. Ligation by DNA ligase to give the recombinant plasmid 1. Transformation of competent host cells 1. Selection on antibiotic, with insertional inactivation or the blue and white test to identify recombinants 1.

Question 5 [10]

- (a) First, the carbon and nitrogen source. A plant medium supplies sugar, usually sucrose at two to three per cent, as the only organic requirement, because a plant cell can make its own amino acids and most of its vitamins from inorganic nitrate and that sugar. An animal cell cannot, so an animal medium must supply glucose together with a full set of amino acids and vitamins ready made. Second, the serum. An animal medium is ordinarily supplemented with about five to twenty per cent of serum, most often foetal bovine serum, which brings the growth factors, hormones and attachment proteins that animal cells will not grow without. No serum is used in plant culture. Third, the growth regulators. A plant medium carries auxins and cytokinins, and the ratio between them decides whether the tissue forms callus, shoots or roots. Animal media carry no such regulators, their place being taken by the growth factors of the serum. Any three of the following are also accepted. A plant medium is set with agar for solid culture, while animal cells are grown in liquid on a treated plastic or glass surface to which they must attach. A plant medium is held at a pH of about 5.7 and at 25 degrees Celsius, an animal medium at a pH of about 7.4 and at 37 degrees Celsius. An animal medium needs a bicarbonate buffer worked against five per cent carbon dioxide and usually carries phenol red as a pH indicator. An animal medium is far richer and more complex and is therefore the more readily spoiled by contamination, so antibiotics are often added to it.

One correct difference, fully stated 1. A second correct difference 1. A third correct difference 1.

- (b) A small piece of tissue is cut from the organ under sterile conditions and washed in a balanced salt solution to remove blood and debris. It is then chopped fine and the cells are freed from one another. This is done by digesting the material that holds them together with an enzyme, trypsin being the one commonly used and collagenase being used where there is much connective tissue, and the digestion is stopped by adding serum, which inhibits the trypsin. The suspension of single cells so obtained is counted, mixed with the complete medium and put into a culture flask or dish, and the flask is placed in the carbon dioxide incubator at 37 degrees Celsius. Most animal cells are anchorage dependent, so they settle, attach to the floor of the vessel and then spread and divide until they cover it as a single layer, a monolayer. This first culture, made straight from the tissue of the animal, is the primary culture. It holds a mixture of the cell types that were present in the tissue, its cells carry the normal chromosome number, and it will divide only a limited number of times. A cell line is what is obtained when the primary culture is subcultured, that is when the monolayer is detached with trypsin and the cells are divided among fresh vessels with fresh medium. From the first such passage the culture is a cell line. It is more uniform than the primary culture, because the cell type which grows best comes to predominate, and it can be propagated, named, characterised and stored frozen. A finite cell line dies out after a certain number of passages. A continuous or established cell line has been immortalised, by spontaneous change or by transformation, and will divide indefinitely.

Tissue dissociated with trypsin or collagenase into a cell suspension 1. Cells seeded in medium and incubated at 37 degrees Celsius to form a monolayer, this being the primary culture 1. Cell line defined as the culture obtained on subculturing, more uniform and capable of being propagated 1.

- (c) A monoclonal antibody is wanted because it is a single kind of antibody, made by one clone of cells, directed against one site on one antigen and therefore uniform from batch to batch. The difficulty is that the cell which makes an antibody of one specificity, the B lymphocyte, will not grow in culture for long, while a cell which grows indefinitely does not make the antibody wanted. The hybridoma method joins the two abilities in one cell. An animal, usually a mouse, is first immunised with the antigen chosen and given time to respond. Its spleen is then removed and the B lymphocytes are taken from it, for among them are cells making antibody against that antigen. These are mixed with cells of a myeloma, a tumour of the antibody forming line, which divide without limit in culture. The myeloma line used is chosen to be one which has lost the enzyme hypoxanthine guanine phosphoribosyl transferase and which makes no antibody of its own. The two kinds of cell are fused using polyethylene glycol, and Sendai virus is also accepted. The mixture is then grown in HAT medium, which holds hypoxanthine, aminopterin and thymidine. The aminopterin blocks the main route by

which a cell makes its nucleotides, so a cell can survive only by the salvage route, for which it needs the enzyme the myeloma lacks. The unfused myeloma cells therefore die for want of the enzyme. The unfused B lymphocytes have the enzyme but cannot divide for long, so they die out of their own accord. Only the fused cells, the hybridomas, survive and grow, since they have the enzyme from the lymphocyte and the immortality from the myeloma. The surviving hybridomas are diluted out so that single cells are grown separately, each colony being the clone of one cell. The clones are screened and the one secreting the antibody wanted is selected and grown in bulk in culture or in the peritoneal cavity of a mouse, and the antibody is harvested from the medium or the ascitic fluid.

B lymphocyte from the spleen of an immunised animal, and the myeloma cell, named as the two cells fused 1. Polyethylene glycol named as the fusing agent 1. HAT medium named, with the selection explained through the enzyme the myeloma lacks 1. Cloning of the hybridomas and screening for the antibody wanted 1.

Question 6 [10]

- (a) A sequence database is an organised and computer searchable collection of nucleotide or protein sequences, each entry carrying the sequence itself together with the information that goes with it, such as the name of the organism, the name and the function of the gene or protein, the literature reference and the accession number which identifies the entry for all time. Because the entries are held to a common format and are indexed, a worker can retrieve a sequence by name or by accession number, or can search the whole collection for sequences resembling one in hand. Two primary nucleotide sequence databases are GenBank, maintained by the National Center for Biotechnology Information in the United States, and the EMBL nucleotide sequence database of the European Bioinformatics Institute. The DNA Data Bank of Japan is the third of that group, and the three exchange their data daily so that each holds the same sequences. A primary database holds the experimental data as they were submitted, that is the sequences themselves, with no interpretation added beyond the submitter's own annotation. GenBank, EMBL, DDBJ and the Protein Data Bank are primary databases. A secondary, or derived, database is built by analysing the contents of the primary databases and holds the result of that analysis rather than raw data, such as the conserved patterns, motifs, domains and families drawn out of the sequences, or entries which have been curated and corrected by hand. PROSITE, Pfam and InterPro are secondary databases, and Swiss-Prot, being a curated protein database, is usually counted here as well.

Sequence database defined with the sequence and its associated information and accession number 1. Two correct primary nucleotide databases named 1. Primary distinguished from secondary, with an example of a secondary database 1.

- (b) A BLAST search takes the sequence the user supplies, the query, and compares it with every sequence in the database chosen, looking not for a match over the whole length but for stretches of local similarity, since two sequences may share only one domain. What comes back is a list of the database sequences which were found to be similar, the hits, ranked with the most significant at the top. Each hit is given with the name and accession number of the sequence, a score, an E value, the percentage identity, the length of the region over which the alignment was made, and the alignment itself written out so that the query and the hit may be compared residue by residue and the gaps and the matched stretches seen. From these the user judges whether the query is related to anything already known, and so can assign a function to a new sequence, find the same gene in other organisms or identify the family it belongs to. The E value, or expectation value, is a measure of how far a hit may be trusted. It is the number of alignments scoring at least as high as the one reported that would be expected to turn up by chance alone in a database of that size. A small E value therefore means that chance is a poor explanation for the similarity and the hit is likely to reflect a real relationship, while a large E value means an alignment of that quality could easily arise by accident. An E value of the order of 10^{-50} points to a close relationship, a value below about 10^{-5} is usually taken as significant, and a value near or above 1 is not worth pursuing. The E value depends on the size of the database searched, so the same alignment is reported as less significant when the database is larger.

BLAST described as a local alignment search returning ranked hits with their alignments 1. Score, percentage identity and E value named among what is returned, with the use of the result 1. E value explained as the number of hits of that score expected by chance, a small value meaning significance 1.

(c) (1) Annotation is the work of attaching biological meaning to a raw genome sequence, which by itself is only a long string of letters. Structural annotation marks out the features along the sequence, that is where the genes lie, where each begins and ends, where the exons, introns, promoters and other control regions are, and where the repeats and the non coding RNA genes are. Functional annotation then says what each of those features does, the function being assigned largely by comparing the predicted gene or protein with sequences of known function in the databases, and this is where a BLAST search is used. (2) An open reading frame is a stretch of a nucleotide sequence which could be read as a continuous set of codons coding for a protein, that is a stretch which begins with a start codon, runs on in the same frame without meeting a stop codon, and ends at a stop codon. It is recognised by having the computer read the sequence in each of the possible frames, three on each strand, so six in all, and report every stretch which runs from an ATG to one of the stop codons TAA, TAG or TGA in that frame. A long open reading frame is unlikely to arise by chance, so a long one is taken as evidence of a real gene, and the prediction may be supported by looking for a promoter upstream of it, by checking whether the codon usage matches that of the organism and by finding a matching messenger RNA or protein. (3) Homology modelling, also called comparative modelling, is the prediction of the three dimensional structure of a protein whose structure has not been determined by experiment, by building it on the known structure of a related protein. It rests on the fact that proteins whose sequences are similar fold in much the same way, so the backbone of the known structure can serve as a framework for the unknown one. What it needs is a template, that is at least one protein of sufficiently similar sequence whose structure has been solved and is deposited in the Protein Data Bank, and a reliable alignment of the target sequence against that template. The higher the identity between the two, the better the model, and below about thirty per cent identity the method becomes unreliable. (4) Bioinformatics is used in drug design to identify and validate the target, the databases of genome and protein sequence being searched for a protein which the disease or the pathogen depends on and which the host does not have or does not need. It is used in virtual screening, in which very large libraries of compounds are docked against the structure of that target in the computer so that only the few which are predicted to bind are made and tested in the laboratory. Any two are accepted, and the following are also correct. It is used to model the structure of a target for which no experimental structure exists, by homology modelling. It is used to predict how a candidate compound will be absorbed and broken down in the body and whether it is likely to be toxic, so that poor candidates are dropped early. It is used in pharmacogenomics, to relate the differences in a patient's genome to the way the patient will respond to a drug.

Annotation explained, with structural and functional annotation distinguished 1. Open reading frame defined and its recognition in the six reading frames explained 1. Homology modelling defined, with the need for a solved template of similar sequence 1. Two correct uses of bioinformatics in drug design 1.