Question no. 1 (10 p)

a) A tryptic peptide VKEGMNIVEAMER (Monoisotopic molecular mass = 1504.7) was analysed by MALDI MS. What do you expect the mass spectrum to look like? Draw the mass spectrum and indicate the m/z values and the approximated isotope distribution for the peptide ion species.

b) The same peptide VKEGMNIVEAMER was analysed by ESI MS. Draw the mass spectrum in the m/z range 400 to 1700 and indicate the m/z values and isotope distribution for the ion signals that you expect to see in the spectrum.

c) The doubly charged (i.e. doubly protonated) peptide VKEGMNIVEAMER was sequenced by ESI MS/MS. Draw the mass spectrum as you expect it to appear. Indicate the types of fragment ion series that you expect to see. Explain the appearance of the MS/MS spectrum, including assignments of peaks. A table of amino acid residue masses is provided below.

Amino Acid Residues

Amino acid	Three (one)-letter code	Monoisotopic mass	Average mass
Glycine	Gly (G)	57.021	57.052
Alanine	Ala (A)	71.037	71.079
Serine	Ser (S)	87.032	87.078
Proline	Pro (P)	97.053	97.117
Valine	Vat (V)	99.068	99.133
Threonine	Thr (T)	101.048	101.105
Cysteine	Cys (C)	103.009	103.145
Isoleucine	Ile (I)	113.084	113.160
Leucine	Leu (L)	113.084	113.160
Asparagine	Asn (N)	114.043	114.104
Aspartic acid	Asp (D)	115.027	115.089
Glutamine	Gln (Q)	128.059	128.131
Lysine	Lys (K)	128.095	128.174
Glutamic acid	Glu (E)	129.043	129.116
Methionine	Met (M)	131.040	131.199
Histidine	His (H)	137.059	137.142
Phenylalanine	Phe (F)	147.068	147.177
Arginine	Arg (R)	156.101	156.188
Tyrosine	Tyr (Y)	163.063	163.176
Tryptophan	Trp (W)	186.079	186.213
Homoserine lactone	—	83.037	83.090
Homoserine	—	101.048	101.105
Pyroglutamic acid	—	111.032	111.100
Carbamidomethylcysteine	—	160.031	160.197
Carboxymethylcysteine	—	161.147	161.181
Pyridylethylcysteine	—	208.067	208.284

Question no. 2 (10p)

In some cases of *in vitro* fertilisation (IVF) preimplantation genetic diagnosis (PGD) are used.

  a) Briefly explain the technology and the purpose of PGD.
  b) Discuss possible ethical challenges with PGD.

Question no. 3 (10 p)

Several large scale protein-protein interaction studies have been undertaken in *Caenorhabditis elegans*. Briefly describe how you would have set up such a protein-protein interaction experiment. Does *C. elegans* have any advantages compared to other animals if you were to verify some of the protein-protein interactions?

Question no. 4 (5 p)

Genetic studies in *Arabidopsis thaliana* have been instrumental to understand the mechanisms of flower development. Why is *Arabidopsis* particularly suited for this type of studies?

Question no. 5 (5p)

The binding site for a transcription factor is often represented by a position weight matrix (PWM). Briefly describe the individual steps that are needed to make a PWM. No formulas are needed.

Question no. 6 (10p)
Explain what we mean by phylogenetic footprinting for motif discovery in genomic sequences. Explain why phylogenetic footprinting is used.

Question no. 7 (10p)

Briefly describe what considerations should be made when preparing for a microarray gene expression analysis with regards to ensuring high quality RNA for analysis.

Question no. 8 (10p)

Fluorescence Resonance Energy Transfer (FRET) is the non-raditative transfer of photon energy from a donor fluorophore to an acceptor fluorophore. The FRET application can be used to investigate the physical interaction of two proteins in living cells if both proteins are made fluorescent. In most cases this is done by fusing one of the proteins of interest to a fluorescent protein that can act as a donor fluorophore and the other protein of interest to a protein that can act as an acceptor fluorophore.
  a) Give an example of a FRET pair (hint: Both are variants derived from the enhanced green fluorescent protein (EGFP)).
  b) Which of the EGFP variants you answered in **a)** (Question a) can act as a donor fluorophore and which can act as an acceptor fluorophore?
  c) In how close proximity do the two proteins have to be for FRET to occur?
  d) Under what conditions do we use FRET?

Multiple choice questions (only one correct answer).
Correct answer gives 1 p

1.

Mitochondrial, chloroplast and bacterial mRNAs lack poly A tail. If you have to perform a microarray gene expression analysis on samples from one of these origins, which special considerations have to be made during cDNA synthesis?	
A	increase the time of cDNA synthesis to ensure complete synthesis
B	decrease the temperature for better annealing of the cDNA synthesis primer
C	a "random" primer must be used for the cDNA synthesis
D	a oligo dT primer must be used for the cDNA synthesis
E	a modified oligo dT primer must be used for the cDNA synthesis

2.

The statistical power of a microarray experiment can be most efficiently increased by:	
A	increasing the number of technical replicates
B	increasing the number of independent biological replicates
C	increasing the number of microarrays used
D	using factorial designs
E	increasing the number of dependant biological replicates

3.

What information does the Bioanalyzer RIN value provide	
A	the RIN value gives information about genomic DNA content
B	the RIN value gives information about the RNA purity
C	the RIN value gives information about degree of RNA degradation
D	the RIN value gives information about the RNA amount
E	the RIN value is an evaluation of the isolation procedure

4.

A microarray gene expression analysis can be done on tissue or cell culture samples	
Which one of the factors mentioned below contribute to the relative strength of a tissue-based study on cancer as opposed to one in a cancer cell line	
A	The gene expression pattern also reflects tumour-local tissue interaction
B	The gene expression pattern is less complex and easier to understand
C	The gene expression pattern is more robust with regard to RNA quality
D	The gene expression pattern reflects intracellular signalling better

5.

The gene expression in tissue is the sum of gene expressions of several cell types	
It may be necessary to localize the gene expression to a specific cell type. Which one of these methods shows unequivocally from which cell the mRNA derives?	
A	Quantitative RT-PCR on the tissue homogenate
B	Western blot on the tissue homogenate
C	In situ hybridization on tissue sections
D	Immunohistochemistry on tissue sections

6.

When gene expression is analyzed by microarray, the following strategy is used:	
A	probes and samples are hybridized to each other and attached to a surface
B	the samples are attached to the surface and the probes are hybridized to the samples
C	the probes are attached to a surface and the sample is hybridized to the probes
D	the probes are labeled and hybridized to the samples attached to the surface
E	the probes are labeled in a solution and hybridized to a surface

7.

What do we mean by "High Throughput Sequencing" (HTS)?	
A	a "large number" of DNA samples can be sequenced at low cost
B	a "large number" of DNA samples can be sequenced simultaneously at one instrument
C	a "very large number"(e.g. millions) of sequencing reactions are run in parallel at e.g. a solid surface
D	"very long" (more than 5000 bases) sequence reads are obtained in each sequencing run
E	the speed of the sequencing reactions is very short (less than 10 minutes) for each sequencing run

8.

The Illumina microarray platform system is based on:	
A	gene probes spotted on glass
B	photolithograpically generated gene probes
C	microspheres (beads) coated with oligonucleotide probes
D	gene probes spotted on filter
E	real time sequencing of expressed transcripts

9.

Fluorescent detection is based on the Stoke`s shift of the fluorochrome	
What does the Stoke`s shift describe?	
A	Difference in emission wavelength for different fluorochromes
B	The emission light intensity of different fluorochromes
C	The exitation/absorption maximum wavelength for different fluorochromes
D	The difference between exitation and emmision wavelength of a fluorochrome
E	The fluorescent stability of a particular fluorochrome

10.

Green fluorescent protein (GFP) has become an important tool in life science.	
What property of GFP is particularly useful?	
A	The high enzymatic activity that allows sensitive detection
B	The high photo-stability
C	The fluorescent property of GFP fusion proteins
D	The high fluorescent intensity
E	The large Stoke`s shift

11.

Fluorescent imaging can be quantified in terms of differences in relative intensity	
What is this quantification based on?	
A	Relative intensity detected by the grey-scale detector(s)
B	The absolute intensity of the particular fluorochrome
C	The colour of the fluorescent dye
D	The relative intensity detected by the digital colour camera
E	The large Stoke`s shift

12.

Luciferase is used extensively in many different approaches in life science.	
What is luciferase?	
A	A fluorescent protein
B	A fluorescent dye
C	A fluorescently labelled secondary antibody
D	An enzyme
E	A reporter plasmid

13.

Fluorescence recovery after photo-bleaching (FRAP) is a technique to study protein dynamics in live cells.	
What is this technique based on?	
A	The irreversible photo-bleaching of a fluorescent protein
B	The reversible photo-bleaching of a fluorescent protein
C	The intensity of green-fluorescent protein (GFP)
D	The mobility of a particular fusion protein
E	Rapid image acquisition of live cells

14.

DNA and RNA quantity is assessed by photometric measurements at 260 nm	
Additional measurements at 280 nm is performed:	
A	To distinguish DNA from RNA
B	To calculate the molar absorbtivity (extinction coefficient)
C	To determine protein concentration
D	To calculate your DNA/RNA ratio

15.

In 2-D PAGE, proteins are separated based on isoelectric point as well as on size. After separation in the first dimension, how should the separated proteins be treated to ensure migration in the same direction in the second dimension	
A	The proteins should be heated to 95°C to prevent hybridization
B	The proteins should be treated with urea to ensure denaturation
C	The proteins should be treated with DTT to keep them in the reduced state
D	The proteins should be treated with SDS to give a net negative charge

16.

Column chromatography is a common way to separate groups of proteins based on their specific characteristics. You want to separate two proteins, the first having pI =4.8 and MW=34 kDa, and the second having pI=5.1 and MW=98 kDa	
Which of the following techniques would you employ?	
A	Gel filtration chromatography
B	Hydrophobic interaction chromatography
C	Affinity chromatography
D	Ion-exchange chromatography

17.

Trypsin is a protease that is widely used in proteomics because it:	
A	Removes DNA and RNA from complex biological samples.
B	Digests proteins into peptides that are suitable for mass spectrometry analysis.
C	Eliminates phosphorylation from proteins, thereby facilitating mass spectrometry analysis.
D	Solubilizes intact membrane proteins.

18.

Stable isotope labelling is used in quantitative proteomics to determine changes in protein abundances by comparative analysis of two or more samples.	
Protein quantification by stable isotope labelling is based on:	
A	Detection of radioactivity of carbon-14 in over-expressed proteins
B	Incorporation of Carbon-13 and/or Nitrogen-15 into proteins to distinguish them from unlabeled proteins by 2D electrophoresis
C	Incorporation of Carbon-13 and/or Nitrogen-15 into proteins to distinguish them from unlabeled proteins by mass spectrometry
D	Incorporation of Phosphorous-32 to detect and quantify protein phosphorylation

19.

The life cycle of *Caenorhabditis elegans* is divided into specific stages.	
What is the dauer stage?	
A	A pre-hatching stage
B	A developmentally arrested dispersal stage
C	A stage induced by food after L4 molting
D	The first stage after fertilization

20.

Epistasis analyses are often used in molecular genetics research.	
What type of information can be inferred from these types of experiments?	
A	Where a particular gene is located on a chromosome.
B	Where a particular gene is located on a chromosome.
C	Can provide information about the size of a gene
D	It is used to determine a functional order of action of two genes

21.

mRNA co-expression analysis is a popular tool used in *Arabidopsis* research.	
What can these studies reveal?	
A	Provide information about gene function and gene expression patterns
B	Identify evolutionary relationships between expressed genes
C	Help identifying related genes in other plants / organisms
D	Identify genes without coding exons

22.

The major difference between transgenic mice and knockout mice is that:	
A	transgenic mice always employ the use of cloned genes derived from other species.
B	transgenic mice have foreign genes that integrate at targeted loci through homologous recombination.
C	transgenic mice have a functional foreign gene added to their genome
D	knockout mice always have a unique phenotype.

23.

When participating in medical research people have to be informed and sign a consent form.	
Which of the principles in bioethics is reflected in the informed consent?	
A	Justice
B	Non-maleficience
C	Dignity
D	Integrity
E	Autonomy

24.

In bioethics as well as medical ethics there are four major principles.	
Identify these four principles:	
A	Autonomy, dignity, beneficence and non-maleficence
B	Autonomy, integrity, dignity and justice
C	Autonomy, beneficence, integrity and justice
D	Dignity, beneficence, autonomy and justice
E	Beneficence, justice, non-maleficence and autonomy

25.

EYFP, EBFP and ECFP are all spectral variants of the original enhanced green fluorescent protein (EGFP). They have all different excitation and emission spectra's.	
Rank the excitation spectra's including that of EGFP, starting with the one that is exited by the highest energy laser line?	
A	EYFP, EGFP, EBFP and ECFP
B	EYFP, EGFP, ECFP and EBFP
C	ECFP, EBFP, EGFP and EYFP.
D	EBFP, ECFP, EGFP and EYFP
E	EGFP, EYFP, ECFP and EBFP

26.

A mouse has been injected with tumour cells expressing the enhanced green fluorescent protein, EGFP.	
On what principle should the *in vivo* imaging method be based to follow the invasion of the tumor cells in the mouse?	
A	Optical bioluminescence
B	Computer tomography
C	Optical reflectance
D	Magnetic resonance
E	Positron emission tomography

27.

Which reporter gene is the most frequently used for monitoring bioluminescence *in vitro* in mammalian cell culture and mammalian *in vivo* models?	
A	Bacterial luciferase
B	Luciferin
C	Firefly Luciferase
D	Renilla luciferase
E	Oxyluciferin

28.

Transcription factor binding sites can be represented for example by a consensus sequence or a position weight matrix.	
What do we mean by an IUPAC consensus sequence?	
A	A standard sequence recognised by all transcription factors
B	An IUPAC standard for position weight matrices
C	A binding site representation consisting of the set of bases occurring at each position, encoded with IUPAC symbols
D	A binding site representation using only the standard symbols A, C, G and T
E	A set of binding sites in IUPAC file format

29.

Many different methods have been developed for analysing gene regulation by transcription factors. It may be relevant in this context to do a benchmark study.	
What do we mean by benchmarking?	
A	Giving a mark of approval to the best methods
B	Identifying (marking) weak spots in individual software tools
C	Combining predictions from several tools (i.e. using a bench of tools)
D	Comparing several predictions from the same tool (i.e. predictions marked to be identical)
E	Comparing performance of different methods against a common reference, e.g. by using a standard data set

30.

The benchmark study of prediction methods for transcription factor binding sites that was published by Tompa *et al* used three different types of sequences; real, generic and markov	
What is the advantage of the markov sequence set?	
A	It does not contain any unknown real binding sites
B	It represents random (average) performance
C	It represents a prediction without any binding sites
D	It is a well designed data set (done by Markov *et al*)
E	There is no particular advantage, it is just a different data set