Module 1: Introduction

Lecture 1 Introduction

Bioanalytical techniques, as the name suggests, are the analytical tools to study the biological molecules; non-biological molecules involved with life, such as drugs; and biological processes. These tools are routinely used to identify, estimate, purify, and characterize the biomolecules. Quantification of molecules in biological samples is at the heart of bioanalysis and is routinely used to diagnose various diseases and metabolic disorders. For example, estimation of thyroxine and triiodothyronine concentrations in blood provides information about the activity of thyroid gland. Home pregnancy test kits look for the human chorionic gonadotropin (hCG) hormone in the urine, presence of which above a threshold concentration is an indicator of pregnancy. Bioanalytical methods are also used to detect drugs and their metabolites in biological samples. Initially, nonspecific assays were used to quantify the drugs in biological samples. Evolution of the existing assays, advancement in instrumentation, and introduction of newer techniques have made it possible to distinguish the drug molecules and their closely related metabolites in complex biological specimens.

Estimation of the analytes

Identification and quantification of analytes is perhaps the most common application of bioanalytical methods. Various diseases and disorders including cancers are diagnosed by estimating the levels of the characteristic biomarkers in a particular tissue or organ. Semenogelase, for example, is a biomarker for prostate cancer, one of the most frequently diagnosed cancers in human males.

Biomarker: In disease and diagnostics, a biomarker is a molecule, presence of which beyond a threshold level is an indicator of the biological state.

In cell biology, a biomarker is a molecule characteristic of a cell type or a group of cells e.g. Oct-4 is a biomarker for embryonic stem cells.
**Qualitative versus quantitative analyses**

A qualitative analysis simply tells about the presence or absence of an analyte in a sample. An absence of analyte, however, may result due to concentrations below the detection level of the bioanalytical technique used. Qualitative analyses are used wherein detection of an analyte is sufficient to take further course of action. For example, identification of a banned performance-enhancing drug in athletics is sufficient enough to determine the qualification of the athlete to participate in the event. In certain cases, however, it is important to estimate the concentration of the analyte. A quantitative analysis would result in the determination of actual amount of the substance present in the sample. Consider a person suspected to be diabetic. A qualitative test for glucose is not good enough to ascertain if the person is diabetic. It is important to accurately determine the concentration of glucose in the blood to arrive at a conclusion. Breath alcohol detectors are used by traffic personnel to quantify the breath alcohol level, which in turn is proportional to blood glucose level and thereby enable them to identify the drunk drivers.
**Accurate and precise determination of analytes**

It is hardly necessary to explain how critical an accurate determination of an analyte is. If a breath alcohol detector is not accurate, a drunk driver may be let off risking the life of others while a sober one may be detained. Unless the concentration of analyte is determined accurately and precisely, it is difficult to make meaningful conclusions.

So, what exactly do the accuracy and precision mean? Accuracy is the measure of how closely the measured values match the true values. Precision tells about the reproducibility of the measurement *i.e.* how closely the measured values are if repeated measurements are made on the sample (Figure 1.1).

![Figure 1.1 Schematic representations of accuracy and precision. Consider the centre of the concentric circles as the true value; the measured values are represented as the black dots. The measured values shown in panel A are close to the true value (accurate) as well as to each other (precise). The measured values in panel B are close to each other (precise) but far from the true value (inaccurate). The individual values in panel C are far away from the true value but randomly distributed about the true value; the average value lies close to the true value (accurate but imprecise). Panel D represents inaccurate and imprecise measurements.](image)

It is easy to imagine the consequences of using an inaccurate equipment; it would give inaccurate results. Imprecise equipments, even if accurate, are problematic as a large number of measurements are required to arrive close to the true value which may take considerable amount of time. An analytical tool therefore has to be both accurate and precise to be used reliably and for faster analysis.
Identification and characterization of molecules

Researchers involved in the discovery of novel bioactive natural products often have to identify the bioactive component present in the crude sample; for example, isolation of novel antibiotics and antimicrobial peptides from various organisms. Individual components in the crude sample are isolated based on the differences in their physical and chemical properties. The bioactive component is identified by testing the activities of these isolated compounds. The bioactive compound is then characterized using various spectroscopic methods to arrive at its structure and function(s). Bioanalytical techniques can typically be classified as shown in Figure 1.2.

![Figure 1.2 Various bioanalytical methods](image)

Spectroscopic tools such as infrared spectroscopy, circular dichroism spectroscopy, and nuclear magnetic resonance spectroscopy can provide structural information about the molecules which in turn provides insights into their functional aspects.

Studying biological processes

Life is an outcome of the complex interplay of biological molecules. These involve interactions between macromolecules (e.g. protein-protein interactions and DNA-protein interactions, RNA-protein interactions); interactions of biomolecules with small molecules (glucose channels, water channels, ligand-binding) and ions (K⁺ channel, Na⁺ and K⁺ pump, Ca²⁺ channels); and interaction of molecules with light
(chlorophyll, photoreceptors). Interactions of the molecules with their receptors/ligands, both in vitro and in vivo, are usually studied using various spectroscopic and microscopic tools. Fluorescence spectroscopy and microscopy are among the most commonly employed tools to study the biological processes. Discovery of the green fluorescent protein (GFP) and subsequent development of its analogs with different spectral properties have revolutionized the area of cellular research. Before discussing in detail the various tools that have gained importance in bioanalytical research, it is worthwhile to take a pause for very quickly reviewing the important structural aspects of major classes of biomolecules.

**Features of major biomolecules**

Classification of biomolecules is largely based on their chemistry. There are four major classes of biomolecules: proteins, nucleic acids, carbohydrates, and lipids.

*Amino acids and proteins*

Proteins constitute the functional machinery in the living systems by carrying out most of the biological reactions. They are the unbranched polymers of L-α-amino acids. D-amino acids do exist in nature, but such molecules are rare. The structure of a typical amino acid is shown in Figure 1.3A.

![Figure 1.3 Structures of amino acids and proteins: structure of a typical L-α-amino acid (A); peptide bond showing the partial double bond character (B); primary and secondary structures (C); and tertiary and quaternary structures (D) formed by proteins.](image)

(A) Carbon group<br>(B) Peptide bond showing the partial double bond character<br>(C) Primary structure<br>(D) Tertiary structure, Quaternary structure
The R group (shown in Figure 1.3A) is what differentiates the 20 standard amino acids present in proteins. During protein synthesis, the amino acids are linked together through an amide bond, called peptide bond (Figure 1.3B). Delocalization of nitrogen’s lone pair of electrons over carbonyl group imparts a partial double bond character to the peptide bond putting severe conformational constraints on the polypeptide backbone (Figure 1.3B). The sequence of amino acids in a polypeptide chain is termed as its primary structure (Figure 1.3C). The linear polypeptide chain can adopt local higher order structures stabilized through hydrogen bonds; these local ordered structures are termed secondary structures (Figure 1.3C). Two such secondary structures found in proteins are: \( \alpha \)-helices and \( \beta \)-sheets. Further folding of the unstructured regions in the polypeptide chain results in a compact structure, termed the tertiary structure, the highest structural level of a single chain protein (Figure 1.3D). Certain proteins function as multimers wherein more than one polypeptide chains assemble together through non-covalent interactions to form what is called a quaternary structure (Figure 1.3D).

**Nucleotides and nucleic acids**

A nucleotide is composed of a pentose sugar, a nitrogenous base, and one or more phosphate groups (Figure 1.4A). The pentose sugar can either be ribose (in a *ribonucleotide*) or 2′-deoxyribose (in a *deoxyribonucletotide*). The nitrogenous bases present in nucleic acids are the derivatives of purine (Adenine, A and Guanine, G) and pyrimidine (Thymine, T; Cytosine, C; and Uracil, U) (Figure 1.4B). Nucleotides, as you may know, are the structural units of nucleic acids. Apart from that, nucleotides such as adenosine triphosphate (ATP), guanosine triphosphate (GTP), coenzyme A (CoA), flavin adenine dinucleotide (FAD), nicotinamide adenine dinucleotide (NAD\(^+\)), and nicotinamide adenine dinucleotide phosphate (NADP\(^+\)), play critical roles in metabolism and intracellular signaling.
Nucleic acids constitute the genomes of living organisms, carry the information in the form of messengers, act as adapters, catalyze the biological reactions, and play regulatory and defense roles as well. Nucleic acids are the unbranched polymers of nucleotides; in a nucleic acid, each nucleotide, except the terminal ones, is linked to two nucleotides through phosphodiester bonds. Nucleic acids can be classified into two categories: ribonucleic acid (RNA; contains ribose sugar) and deoxyribonucleic acid (DNA; contains 2′-deoxyribose sugar). DNA can have A, G, C, and T as its bases while RNA can have A, G, C, and U. In nucleic acids, A can form hydrogen bonds with T and U while C can form hydrogen bonds with G. A-T, A-U, and G-C are said to constitute the complementary base pairs and play crucial roles in processes like replication, transcription, and translation. DNA is usually composed of two strands wrapped around each other in a double-helical fashion (dsDNA) (Figure 1.4C). Single stranded DNA (ssDNA), however, does exist in certain viruses. Similarly, RNA molecules are typically single stranded but certain viruses do have double stranded RNA (dsRNA). Furthermore, RNA molecules can also adopt local double-helical structures to adopt a 3-dimensional structure, e.g. tRNA (Figure 1.4D). In a double helical nucleic acid structure, the bases on one strand form hydrogen bonds with the complementary bases on the other. The bases lie roughly perpendicular to the nucleic
acid backbone and the stacking interactions between them further stabilize the double helical structure.

*Carbohydrates*

Carbohydrates are the polyhydroxy aldehydes or ketones. Aldehyde sugars are often referred to as aldoses while ketone sugars are called ketoses. Monosaccharides such as glucose and fructose are the simplest carbohydrates (Figure 1.5). The monosaccharides can join covalently to give disaccharides, oligosaccharides, and polysaccharides. Unlike proteins and nucleic acid, polysaccharides can be branched. Based on their functions, polysaccharides can be classified as structural polysaccharides (*e.g.* cellulose and chitin) and storage polysaccharides (*e.g.* starch and glycogen).

![Open chain and ring structures of glucose (A) and fructose (B). Notice the two stereoisomers formed during cyclization of the open chain structure. A disaccharide (sucrose) formed from condensation of glucose and fructose (C).](image-url)
Lipids

Lipids are amphipathic molecules with polar head groups and non-polar hydrocarbon region. Like carbohydrates, lipids also have both structural and storage roles in living systems. Apart from these, lipids play important roles in signal transduction pathways inside the cells. In vertebrates, for example, triacylglycerols are stored as fuels in specialized cells called adipocytes. Phosphoinositides, phosphorylated forms of phosphatidylinositol, are involved in cell signaling and membrane trafficking. Structural lipids perhaps constitute the most important class of the lipids because it is the structural lipids that define a cell. Figure 1.6A shows the structure of a glycerophospholipid, the most common structural lipid present in biomembranes.

![Figure 1.6 A glycerophospholipid showing the amphipathic nature of lipids (A). Self-assembly of lipids producing a lipid bilayer (B).](image)

The amphipathic nature of the lipids allows them to self-assemble in aqueous solutions to form a bilayer (Figure 1.6B). The assembled lipid bilayer has polar head groups interacting with water molecules whereas non-polar hydrocarbon regions are buried inside making the membrane core highly hydrophobic that acts as a barrier for polar and charged chemical species. It is this barrier that separates a cell from the external environment. A cell, however, needs to communicate and exchange material with its environment. To accomplish these tasks, biological membranes have acquired specialized proteins.
Having reviewed the structures of the four major biomolecules, we are now ready to look at the bioanalytical techniques that have gained importance in recent times (Lecture 2) and to go through them in detail in the following lectures.
QUIZ

Q1: Molecular weight of alanine is 89.09 Da. What will be the molecular weight of a 10-residue peptide composed entirely of alanine?

Ans: Formation of a peptide bond between two amino acids is accompanied by release of a water molecule. Synthesis of a decapeptide (a 10-residue peptide) involves formation of nine peptide bonds \textit{i.e.} removal of nine water molecules. Therefore, the molecular weight of the peptide will be:

\[
\text{Molecular weight} = (10 \times 89.09) - (9 \times 18) \text{ Da}
\]
\[
= (890.9 - 162) \text{ Da}
\]
\[
= 728.9 \text{ Da}
\]

Q2: How does a partial double bond on peptide bond put constraints on proteins’ conformations?

Ans: Due to partial double bond character, rotation about the peptide bond is not possible therefore restricting the number of conformations a polypeptide chain can adopt. If there were rotation possible about peptide bond, a protein would be able to sample many more conformations.

Q3: Write down the sequence(s) of RNA molecule(s) that the following DNA molecule can generate (assume entire molecule is transcribed to RNA):

5’–TACGCTGAC–3’
3’–ATGCGACTG-5’

Ans: In biological systems, nucleic acids are synthesized from 5’–3’ end. Assuming that both the DNA strands can template the transcription, following RNA molecules will be generated:

From \textit{5’–TACGCTGAC–3’} strand: 5’–GUCAGCGUA–3’
From \textit{3’–ATGCGACTG-5’} strand: 5’–UACGCUGAC–3’

Q4: Which of the following molecules act as enzymes?

Proteins, Nucleic acids, Carbohydrates, Lipids

Ans: Most enzymes are proteins but RNA molecules also catalyze certain biochemical reactions. Therefore, both proteins and nucleic acids can act as enzymes.
Lecture 2 Modern Approaches in Bioanalysis and Bioassays

Initial bioanalytical methods were not highly specific and were relatively insensitive as compared to the modern methods. The assays included colorimetric estimation of the compounds or simple bioassays, such as antibiotic estimation by quantifying their ability to inhibit microbial growth. Development of pharmacokinetics during 1930s started demanding for more specific and sensitive methods to accurately determine the concentrations of drugs and metabolites in biological specimens. Around this time, spectroscopic techniques such as UV/Visible spectroscopy, infrared spectroscopy, and chiroptical spectroscopy were seeing advancement but were largely restricted to the analysis of chemical compounds. Lack of sensitive instrumentation around that time further restricted their applications to biological samples that usually have low concentrations of molecules. Second half of the 20th century saw a rapid development in the instrumentation and development of new methodologies that eventually would find applications in life sciences and medicine. Liquid chromatography turned out to be a major advancement towards achieving sensitivity and power of resolving the closely-related metabolites. Reversed-phase chromatography, for example, has proved to be an excellent tool for resolving and analyzing the small molecules with excellent sensitivity. Electrophoresis is another powerful tool for analyzing and separating biomolecules. It has turned out to be an indispensable tool for analyzing nucleic acids. Integrity of isolated nucleic acids, cleavage of DNA molecules by restriction enzymes, mapping of restriction sites in a DNA molecule, and joining of two or more DNA fragments by ligases are some of the diverse applications of electrophoresis in a molecular genetics laboratory (Figure 2.1). DNA molecules differing in even one base pair can be separated by electrophoresis; this allows sequencing of DNA by Sanger’s method. Electrophoresis is also used to analyze proteins. Electrophoresis allows separation of proteins based on their isoelectric points. SDS-PAGE (Sodium dodecyl – polyacrylamide gel electrophoresis) of proteins separates the proteins based on their size and therefore allows determination of their molecular weights (discussed in lecture 32).
Quantification of an analyte, as has been discussed in the previous lecture, is among the most common applications of analytical tools. You may be familiar with the use of UV/visible light for recording absorption of organic molecules to determine the concentration of the compound. It is therefore clear that light or electromagnetic radiation can interact with the matter providing useful information about it. Interaction of electromagnetic radiation with matter is termed as spectroscopy. Absorption of UV/visible radiation is associated with electronic transitions in the molecules; UV/Visible spectroscopy is therefore also referred to as the electronic spectroscopy. Absorption of ultraviolet and/or visible radiation is the most commonly employed method to estimate the concentration of biomolecules such as proteins, peptides, nucleotides, nucleic acids, carbohydrates, and lipids. Absorption at 260 nm and 280 nm provides information about the nucleic acid contamination in protein preparations. Phenol is commonly used to isolate nucleic acids; $\frac{A_{260}}{A_{280}}$ is used to determine phenol contamination in nucleic acid preparations and has become a routinely used method in molecular biology laboratories. Electronic spectroscopy goes beyond quantification of biomolecules: fluorescence spectroscopy is used to study various biological processes viz. protein folding/unfolding, binding studies, etc. Electronic circular dichroism spectroscopy is a chiroptical method and finds applications in analyzing protein and peptide structures, protein folding/unfolding, binding studies, etc.
Infrared spectroscopy probes the vibrational frequencies in the molecules; the frequency of vibration depends on the strength of the bond and the atoms involved thereby allowing identification of functional groups present in the organic molecules. As the absorption depends on the concentration, infrared spectroscopy can also be utilized for determining the concentrations of the analytes. The vibrational frequencies of the bonds are sensitive to the conformation of the molecule as well as the interactions of the atoms involved. Infrared spectroscopy can therefore provide information about the conformations of the molecules. In fact, infrared spectroscopy is often used to determine the secondary structures of the polypeptides. Advent of nuclear magnetic resonance (NMR) spectroscopy in 1940s revolutionized the analysis of small molecules. When used alongside infrared spectroscopy, NMR spectroscopy can quickly provide the complete structure of the molecules. Advancement in the hardware and development of the experimental methods has made NMR spectroscopy one of the most powerful weapons in a chemist’s and biochemist’s arsenal. NMR is routinely employed to study the structure and dynamics of biomacromolecules. In fact, NMR is the only tool that provides atomic resolution structure of the molecules in solution. This is a big plus for NMR spectroscopy over X-ray crystallography that needs a crystal for determining the atomic resolution structure. Furthermore, solid state NMR spectroscopy can be used to study the solid samples including single crystals. Atomic resolution structure determination requires the biomolecules with very high purity. High purity biological macromolecules are obtained through one or more chromatographic methods. The principle underlying the separation of molecules is their partitioning between a stationary and a mobile phase. The partition coefficient of a molecule depends on its physicochemical properties and molecules can be separated based on their size, charge, hydrophobicity or affinity to a particular ligand. Chromatographic techniques can also provide analytical information, e.g. molecular weight can be determined using size exclusion chromatography wherein there is a relationship between the molecular weight and the elution volume. NMR spectroscopy, however, has come a long way since its discovery and it is now possible to determine the structures of biomolecules in their native milieu i.e. inside the living cells.
Discovery of polymerase chain reaction (PCR) (Figure 2.2) was a major step forward in the biomedical research and diagnostics. Presence of a pathogen inside the body is classically detected using serological methods or culture of the infectious agents. Owing to its excellent sensitivity, PCR can detect the presence of pathogens earlier than the serological tests. Other than infectious diseases, PCR is also used to detect genetic disorders. It is hard to imagine doing research in the areas of molecular genetics without employing PCR.

**Figure 2.2 Principle of PCR amplification of DNA**

Complexity of the biological systems hardly needs any mention. To understand the molecules at function in a living system, it is important to look at the system altogether rather than individual components and processes. Sequencing of complete genomes led researchers to estimate the number of genes a particular organism expresses and further to understand the co-expression of a large number of genes and their role in physiology and pathophysiology. Identification and estimation of the subset of proteins expressed at any instant can provide useful information about the system viz. expression of a gene or a set of genes beyond a threshold level may be a marker of a disease. Intrinsically low levels of a large number of proteins, however, posed a challenge for detecting and identifying them. Application of mass spectrometry to proteins and peptides provided major breakthrough towards achieving
this. Mild ionization techniques such as electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI) could successfully ionize large biomolecules without damaging them. This opened up a plethora of possibilities and resulted in the development of a new research discipline, called proteomics. Proteomics refers to the study of the complete set of proteins expressed by a cell or an organism. The ultimate goal of the proteomic studies is to identify all the proteins present in the specimen; quantify them; and identify the posttranslational modifications, if any. A proteomic approach typically starts with the isolation of the total protein from the sample. Total protein is then resolved into its components using 2-D gel electrophoresis that separates the proteins according to their isoelectric points in one dimension and their molecular weights in the other. The individual protein spots are then cut from the gel and eluted out. The proteins are then analyzed using mass spectrometry either directly or after digesting with a sequence specific protease such as trypsin. The proteins can then be identified either by de novo sequencing or by using databases having sequence information and thereby mass information of the peptide fragments (discussed in lecture 13). Proteomic analysis is useful in identifying the markers for various processes and diseases. For example, comparing the samples from a set of healthy individuals with that of individuals having some disease/disorder can identify if the protein levels go up or down in the unhealthy individuals as compared to the healthy ones. Systematic studies with a large number of individuals are likely to result in identification of biomarkers for the diseases. The need to retrieve and analyze the huge amount of data generated from genome sequencing projects led to the development of another discipline, called Bioinformatics. Bioinformatics utilizes computer science and mathematics to organize and retrieve the biological data. The biological information such as sequences of nucleic acids and proteins, their structures, post-translational modifications of proteins, etc. are organized and stored in the databases. The databases can be accessed to retrieve the required information for analysis.
The role microscopy plays in understanding biological systems and processes hardly needs any introduction. The first uses of microscopes for observing the biological specimens date back to 1660s. It would have not been possible to identify and understand the organization of microorganisms without using microscopy. Light microscopy is used to identify the microorganisms based on their morphology and the specific stains they take up. A routine quantitative application of microscopy is to count the number of different cells per unit volume of blood or any other sample using a hemocytometer. Presence of cells that are not expected in the healthy individuals may be an indicator of anomaly/disease. For example, a simple microscopic analysis of blood sample will identify the sickle cell anemia; presence of pus cells in urine, quantified by microscopy, is an indicator of infection. Light microscopy uses light as the illumination radiation and is perhaps the most familiar form of microscopy. In the simplest microscopic methods, a specimen is illuminated by visible light and observed either against a bright background (bright-field microscopy) or a dark background (dark-field microscopy). Fluorescence microscopy, one of the most commonly used microscopic methods in biological research, has emerged as a very powerful tool for studying molecular processes owing largely to the advancement in optics and discovery of the green fluorescent protein and development of its analogs with different spectral properties (discussed in lectures 15 and 16). Confocal laser scanning microscopy (CLSM) is a type of fluorescence microscopy that allows imaging of the samples at different focal planes i.e. light emitting from below or above the desired focal plane is eliminated. This results in very high lateral resolution and allows determining the spatial localization of the molecules (discussed in lecture 16). Total internal reflection fluorescence (TIRF) microscopy is another type of fluorescence microscopy wherein the optics allows imaging of the molecules that are in close proximity to the microscopic slide (discussed in lecture 15). The resolution of light microscopes depends on the wavelength of the light used. The smaller the wavelength of the light used, the better the resolution obtained. Wavelength of the visible light imposes a resolution limit of ~0.2 μm on the light microscopes (discussed in lecture

**Hemocytometer:** Hemocytometer, also known as Neubauer chamber, is essentially a glass slide which has a counting chamber at the centre. A glass cover is placed on the hemocytometer and the sample is gently introduced into the chamber. The sample chamber has a grid which allows counting of cells in a defined region using a microscope.
14). What it means is that the two point objects closer than ~0.2 μm cannot be resolved using a light microscope.

In electron microscopy (discussed in lectures 17 and 18), the electrons are accelerated by applying a very high accelerating voltage. The wavelength of the electron beam is inversely proportional to the square root of the accelerating voltage, and wavelengths smaller than 0.5 nm can be generated. This provides around three orders of magnitude improvement in resolution. Scanning electron microscopy (SEM) scans the specimen and provides surface information of the specimen. In transmission electron microscopy (TEM), electrons penetrate into the sample and the transmitted electrons generate the image. TEM, therefore, provides information about the internal structures of the specimen. Both SEM and TEM generally require staining of the specimen with a heavy atom. There have been several advancements in transmission electron microscopy, cryo-electron microscopy (Cryo-EM) is perhaps the most noted one. Cryo-EM allows the imaging of hydrated samples, does not require any staining and can provide resolutions between 5-10 Å making the method useful in studying the structures of biomacromolecules. Advent of scanning probe microscopy, especially the atomic force microscopy (discussed in lecture 19), could make it possible doing imaging in solution with resolutions comparable to electron microscope.
QUIZ

Q1: A 500 bp long DNA sample was digested using two restriction enzymes. The digested DNA products were analyzed using gel electrophoresis as shown below:

Map the restriction sites for the two enzymes in the DNA molecule

**Ans:** Digestion with restriction enzyme-1 gives two bands of sizes 200 bp and 300 bp. This results in following two possibilities:

- **(A)**
- **(B)**

Digestion with restriction enzyme-2 gives two bands of sizes 350 bp and 150 bp:

- **(C)**
- **(D)**

Digestion with both restriction enzymes-1 and 2 gives two bands of sizes 200 bp and 150 bp. The possible restriction maps from the given data are: (A+D) and (B+C) *i.e.*

- **(A+D)**
- **(B+C)**

The combinations, (A+B), (A+C), (B+D), and (C+D) do not fit the given data.
Q2: Trypsin is one of the most commonly used proteases for generating peptide fragments for mass spectrometric analysis.

   a) What are the cleavage sites for trypsin in a polypeptide chain?
   b) Name a protease that cleaves at the carboxyl end of the aromatic amino acids.

Ans:

   a) Trypsin leaves at the carboxyl end of arginines and lysines.
   b) Chymotrypsin cleaves at the carboxyl end of the aromatic amino acids.