Basic Methods in Cellular and Molecular Biology

(Analyzing Proteins)

ANALYZING PROTEINS

❖ Proteins perform most cellular processes:
they catalyze metabolic reactions,
use nucleotide hydrolysis to do mechanical work,
and serve as the major structural elements of the cell.

❖ The great variety of protein **structures** and **functions** has stimulated the development of a multitude of techniques to study them.

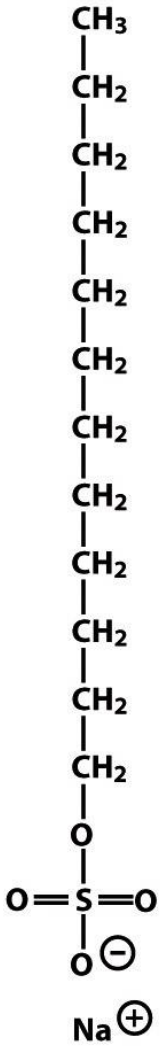
Proteins Can Be Separated by SDS Polyacrylamide-Gel Electrophoresis

Proteins usually possess a net positive or negative charge, depending on the mixture of charged amino acids they contain.
An electric field applied to a solution containing a protein molecule causes the protein to **migrate at a rate that depends on its net charge and on its size and shape**.

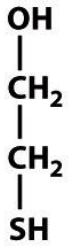
The most popular application of this property is **SDS polyacrylamide-gel electrophoresis (SDS-PAGE).**

It uses a highly cross-linked gel of polyacrylamide as the **inert matrix** through which the proteins migrate.

The gel is prepared by polymerization of monomers; the pore size of the gel can be adjusted so that it is small enough to retard the migration of the protein molecules of interest.
The proteins are dissolved in a solution that includes a powerful negatively charged detergent, **sodium dodecyl sulfate**, or **SDS**.
Because this detergent binds to hydrophobic regions of the protein molecules, causing them to unfold into extended polypeptide chains, the individual protein molecules are released from their associations with other proteins or lipid molecules and rendered freely soluble in the detergent solution.
In addition, a reducing agent such as **β-mercaptoethanol** is usually added to break any S–S linkages in the proteins, so that all of the constituent polypeptides in multisubunit proteins can be analyzed separately.



SDS

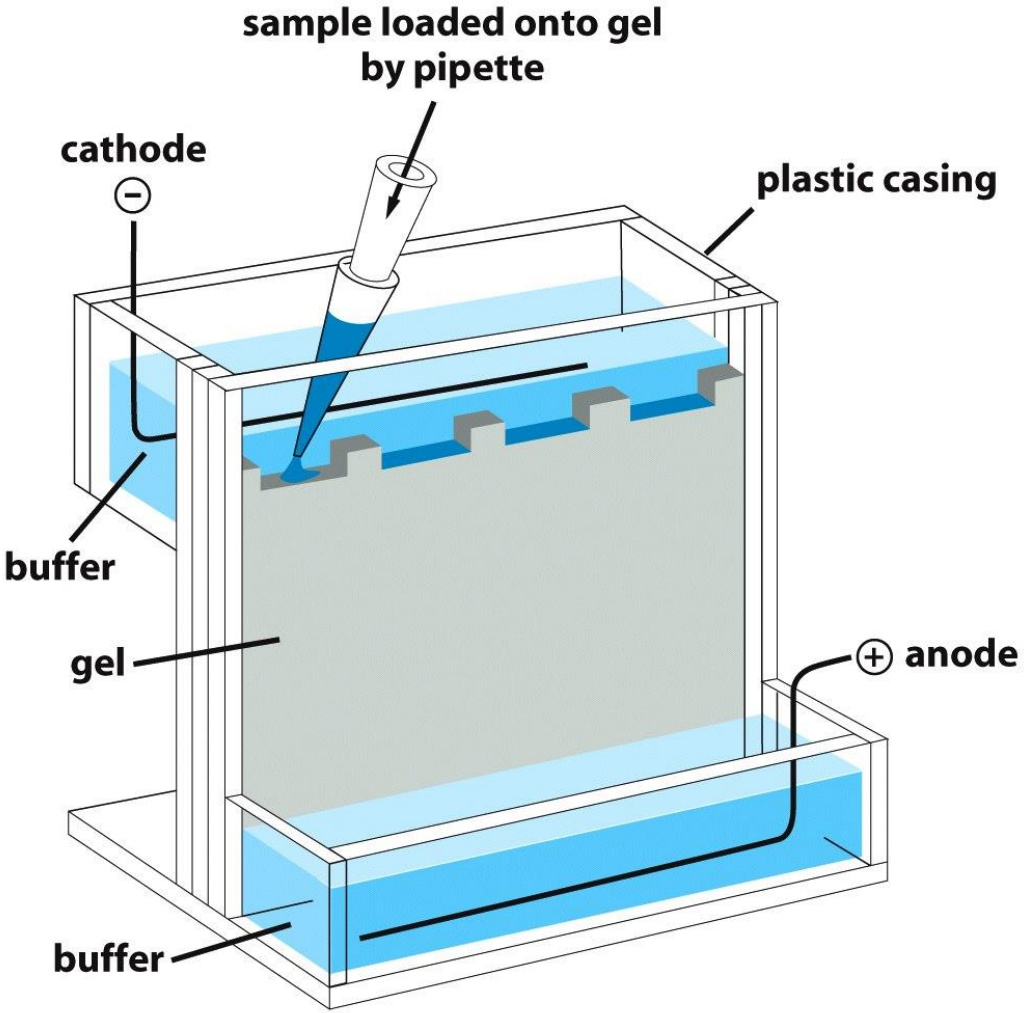


β-mercaptoethanol

Proteins Can Be Separated by SDS Polyacrylamide-Gel Electrophoresis

What happens when a mixture of SDS-solubilized proteins is run through a slab of polyacrylamide gel?

- ✓ Each protein molecule binds large numbers of the negatively charged detergent molecules, which mask the protein's intrinsic charge and cause it to migrate toward the positive electrode when a voltage is applied.

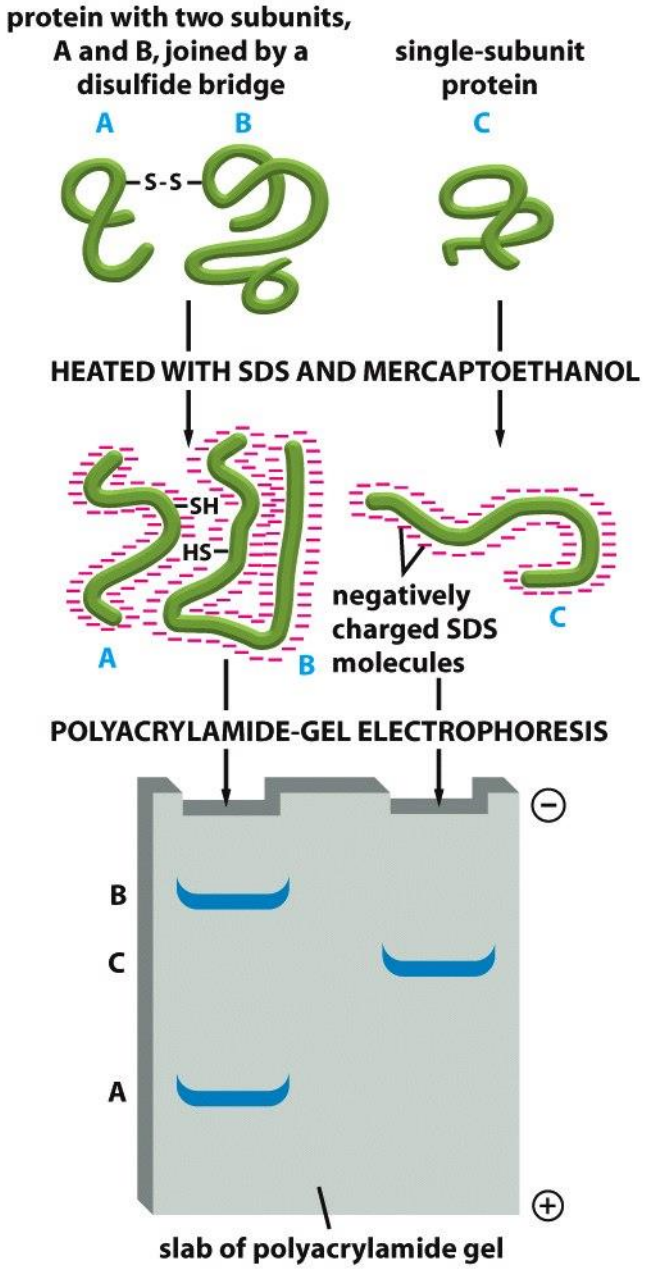


Proteins Can Be Separated by SDS Polyacrylamide-Gel Electrophoresis

Proteins of the same size tend to move through the gel with similar speeds because:
(1) their native structure is completely unfolded by the SDS, so that their shapes are the same, and
(2) they bind the same amount of SDS and therefore have the same amount of negative charge.

Larger proteins, with more charge, are subjected to larger electrical forces but also to a larger drag.

In the mesh of the polyacrylamide gel, which acts as a **molecular sieve**, large proteins are retarded much more than small ones.
As a result, a complex mixture of proteins is fractionated into a series of discrete protein bands arranged in order of molecular weight.

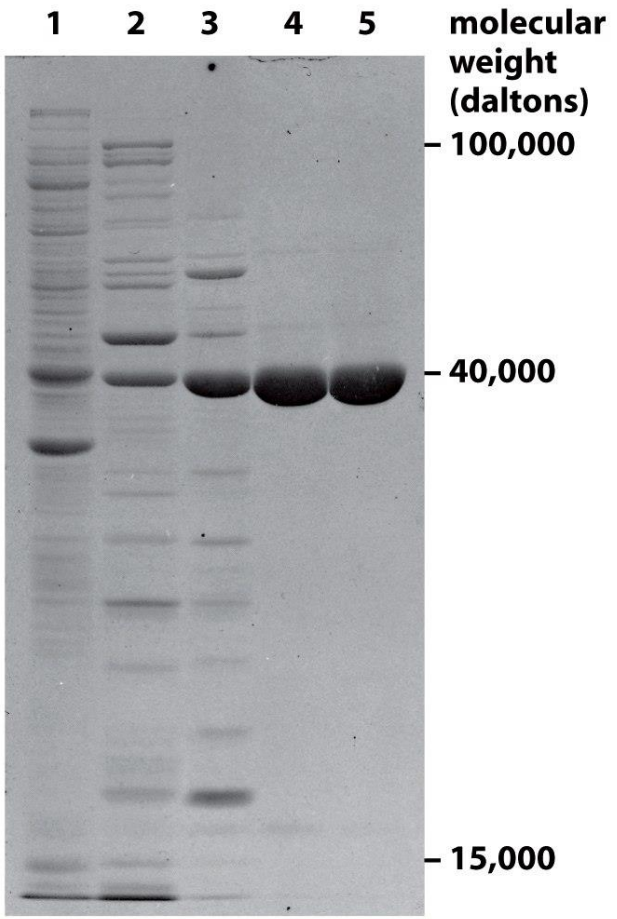


Proteins Can Be Separated by SDS Polyacrylamide-Gel Electrophoresis

-The major proteins are readily detected by staining the proteins in the gel with a dye such as **Coomassie blue**.

-Even minor proteins are seen in gels treated with a **silver stain**, so that as little as 10 ng of protein can be detected in a band.

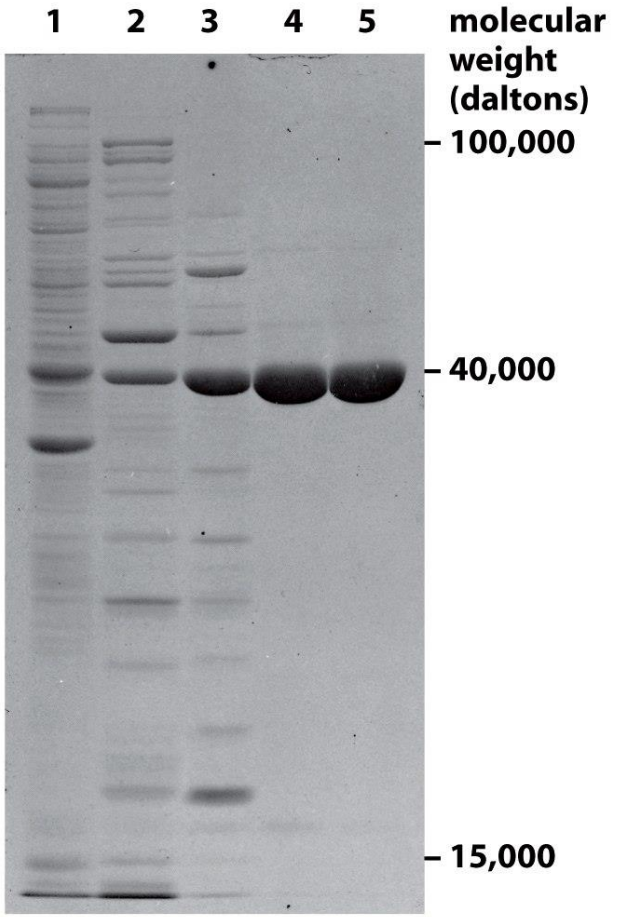
-For some purposes, specific proteins can also be labeled with a radioactive isotope tag; exposure of the gel to film results in an autoradiograph on which the labeled proteins are visible.



Proteins Can Be Separated by SDS Polyacrylamide-Gel Electrophoresis

- SDS-PAGE is widely used because it can separate all types of proteins, including those that are normally insoluble in water—such as the many proteins in membranes.

- And because the method separates polypeptides by size, it provides information about the **molecular weight** and **the subunit composition of proteins**.



Two-Dimensional Gel Electrophoresis Provides Greater Protein Separation

- Because different proteins can have similar **sizes**, **shapes**, **masses**, and **overall charges**, most separation techniques such as SDS polyacrylamide-gel electrophoresis or ion-exchange chromatography cannot typically separate all the proteins in a cell or even in an organelle.

- In contrast, **two-dimensional gel electrophoresis**, which combines two different separation procedures, can resolve up to 2000 proteins in the form of a two-dimensional protein map.

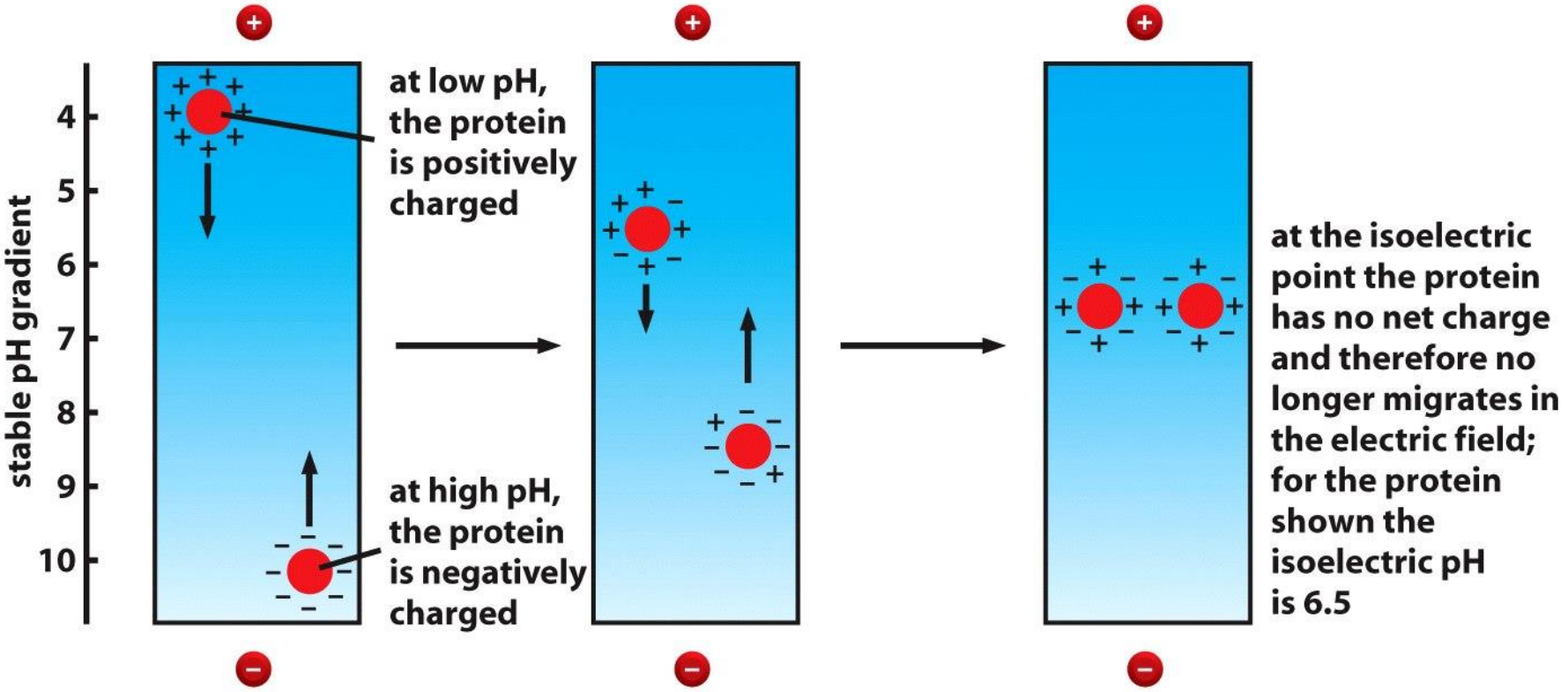
- In the first step, the proteins are separated by **their intrinsic charges**. The sample is dissolved in a small volume of a solution containing a nonionic (uncharged) detergent, together with β-mercaptoethanol and the denaturing reagent urea:
- **This solution solubilizes, denatures, and dissociates all the polypeptide chains but leaves their intrinsic charge unchanged.**

- The polypeptide chains are then separated in a pH gradient by a procedure called **isoelectric focusing**, which takes advantage of the variation in the net charge on a protein molecule with the pH of its surrounding solution.

- Every protein has a characteristic **isoelectric point**, the pH at which the protein has no net charge and therefore does not migrate in an electric field.

- In isoelectric focusing, proteins are separated electrophoretically in a narrow tube of polyacrylamide gel in which a gradient of pH is established by a mixture of special buffers. Each protein moves to a position in the gradient that corresponds to its isoelectric point and remains there. This is the first dimension of two-dimensional polyacrylamide-gel electrophoresis.

Two-Dimensional Gel Electrophoresis Provides Greater Protein Separation



-Separation of protein molecules by isoelectric focusing.
At low pH (high H+ concentration), the carboxylic acid groups of proteins tend to be uncharged (–COOH) and their nitrogen containing basic groups fully charged (for example, –NH3 +), giving most proteins a **net positive charge**. At high pH, the carboxylic acid groups are negatively charged (–COO–) and the basic groups tend to be uncharged (for example, –NH2), giving most proteins a **net negative charge**.

-At its isoelectric pH, a protein has no net charge since the positive and negative charges balance. Thus, when a tube containing a fixed pH gradient is subjected to a strong electric field in the appropriate direction, each protein species migrates until it forms a sharp band at **its isoelectric pH**.

Two-Dimensional Gel Electrophoresis Provides Greater Protein Separation

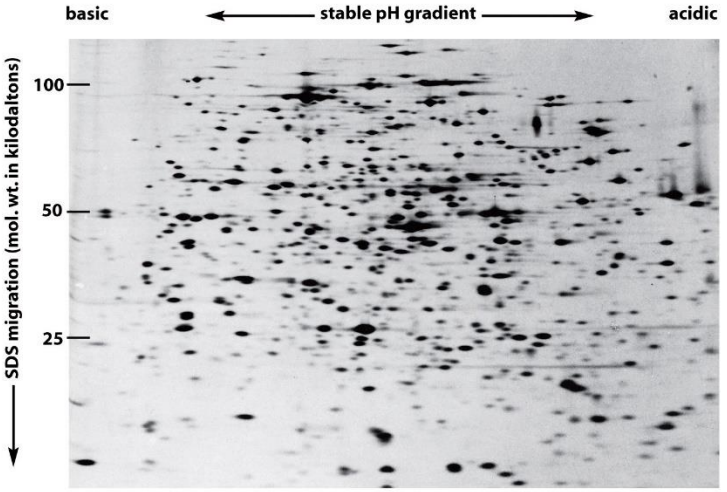
- In the second step, the narrow tube gel containing the separated proteins is again subjected to electrophoresis but in a direction that is at a right angle to the direction used in the first step.

- This time **SDS** is added, and the proteins separate according to their size, as in one-dimensional SDS-PAGE: the original tube gel is soaked in SDS and then placed along the top edge of an SDS polyacrylamide-gel slab, through which each polypeptide chain migrates to form a discrete spot. This is the second dimension of two-dimensional polyacrylamide-gel electrophoresis.

- The only proteins left unresolved are those that have both identical sizes and identical isoelectric points, a relatively rare situation.

- Even trace amounts of each polypeptide chain can be detected on the gel by various staining procedures—or by autoradiography if the protein sample was initially labeled with a radioisotope.

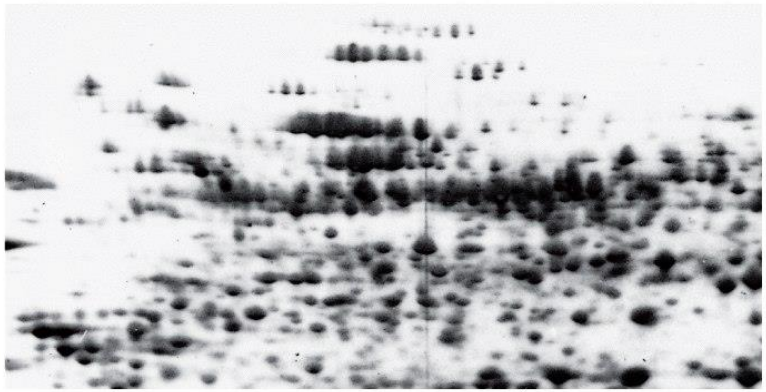
❖ **The technique has such great resolving power that it can distinguish between two proteins that differ in only a single charged amino acid, or a single negatively charged phosphorylation site.**

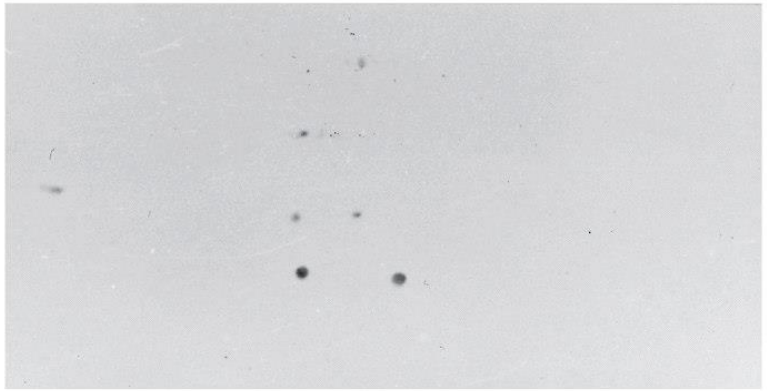


Two-dimensional polyacrylamide-gel electrophoresis. All the proteins in an E. coli bacterial cell are separated in this gel, in which each spot corresponds to a different polypeptide chain.

Specific Proteins Can Be Detected by Blotting with Antibodies

- A specific protein can be identified after its fractionation on a polyacrylamide gel by exposing all the proteins present on the gel to a **specific antibody** that has been labeled with a **radioactive isotope** or a **fluorescent dye**.
- This procedure is normally carried out after transferring all of the separated proteins present in the gel onto a sheet of **nitrocellulose paper** or **nylon membrane**.
- Placing the membrane over the gel and driving the proteins out of the gel with a **strong electric current** transfers the protein onto the membrane.
- The membrane is then soaked in a solution of labeled antibody to reveal the protein of interest.
- This method of detecting proteins is called **Western blotting**, or **immunoblotting**.
- Sensitive Western-blotting methods can detect very small amounts of a specific protein (1 nanogram or less) in a total cell extract or some other heterogeneous protein mixture.
- The method can be very useful when assessing the **amounts of a specific protein** in the cell or when measuring **changes in those amounts under various conditions**.





Mass Spectrometry Provides a Highly Sensitive Method for Identifying Unknown Proteins

A frequent problem in cell biology and biochemistry is the identification of a protein or collection of proteins that has been obtained by one of the purification procedures.

-Because the genome sequences of most experimental organisms are now known, catalogs of all the proteins produced in those organisms are available. The task of identifying an unknown protein (or collection of unknown proteins) thus reduces to matching some of the amino acid sequences present in the unknown sample with known cataloged genes.

-This task is now performed almost exclusively by using mass spectrometry in conjunction with computer searches of databases.

- **Charged particles have very precise dynamics when subjected to electrical and magnetic fields in a vacuum.**

-Mass spectrometry exploits this principle to separate ions according to their **mass-to-charge (m/z) ratio**.

-It is an enormously sensitive technique.

-It requires very little material and is capable of determining the precise mass of intact proteins and of peptides derived from them by enzymatic or chemical cleavage.

-Masses can be obtained with great accuracy, often with an error of less than one part in a million.

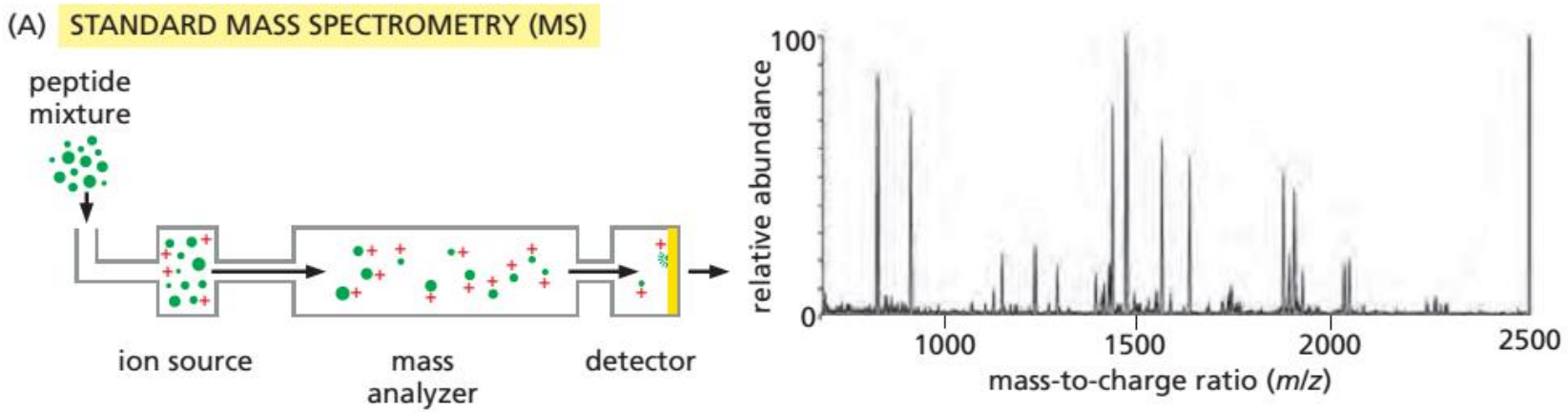
Mass Spectrometry Provides a Highly Sensitive Method for Identifying Unknown Proteins

Mass spectrometry is performed using complex instruments with three major components:

-The first is the **ion source**, which transforms tiny amounts of a peptide sample into a gas containing individual charged peptide molecules.

-These ions are accelerated by an electric field into the second component, the **mass analyzer**, where electric or magnetic fields are used to separate the ions on the basis of their mass-to-charge ratios.

-Finally, the separated ions collide with a **detector**, which generates a mass spectrum containing a series of peaks representing the masses of the molecules in the sample.



Mass Spectrometry Provides a Highly Sensitive Method for Identifying Unknown Proteins

There are many different types of mass spectrometer, varying mainly in the nature of their **ion sources** and **mass analyzers**.

One of the most common ion sources depends on a technique called **matrix-assisted laser desorption ionization (MALDI)**.

-In this approach, the proteins in the sample are first cleaved into short peptides by a protease such as **trypsin**.
-These peptides are mixed with an **organic acid** and then **dried onto a metal or ceramic slide**.
-A brief laser burst is directed toward the sample, producing **a gaseous puff of ionized peptides**, each carrying **one or more positive charges**.

- In many cases, the MALDI ion source is coupled to a mass analyzer called a **time-of-flight (TOF) analyzer**, which is a long chamber through which the ionized peptides are accelerated by an electric field toward a detector.

- **Their mass and charge** determine the time it takes them to reach the detector: large peptides move more slowly, and more highly charged molecules move more quickly.

❖ **By analyzing those ionized peptides that bear a single charge, the precise masses of peptides present in the original sample can be determined.**

- This information is then used to search genomic databases, in which the masses of all proteins and of all their predicted peptide fragments have been tabulated from the genomic sequences of the organism.

❖ **An unambiguous match to a particular open reading frame can often be made by knowing the mass of only a few peptides derived from a given protein.**

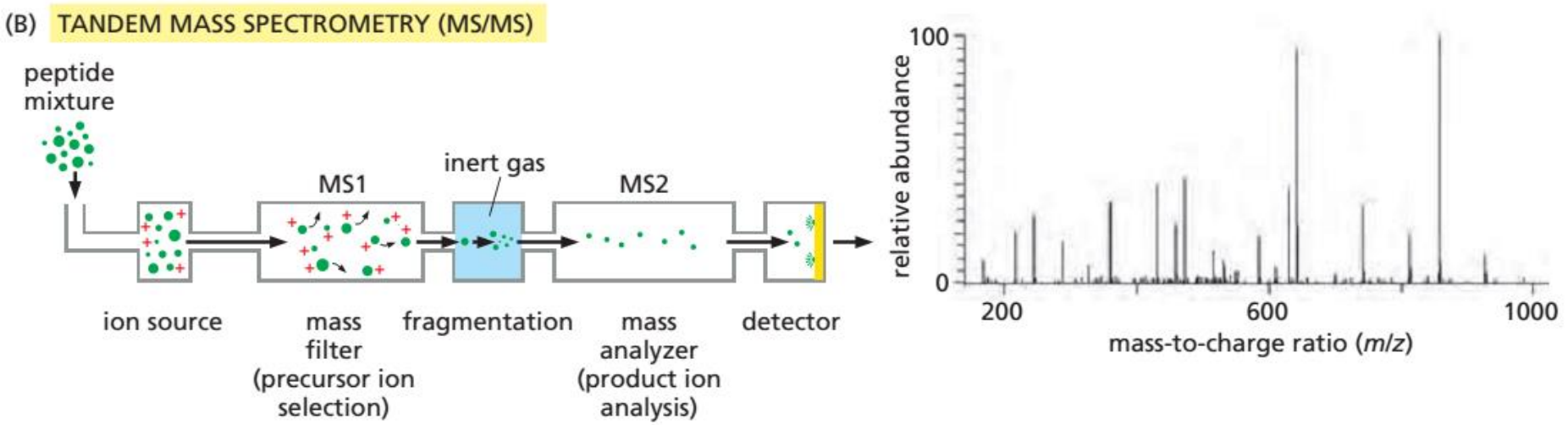
Mass Spectrometry Provides a Highly Sensitive Method for Identifying Unknown Proteins

By employing two mass analyzers in tandem (an arrangement known as **MS/ MS**), it is possible to directly determine the amino acid sequences of individual peptides in a complex mixture.

The MALDI-TOF instrument is not ideal for this method.

Instead, MS/MS typically involves an **electrospray ion source**, which produces a continuous thin stream of peptides that are ionized and accelerated into the first mass analyzer.
The mass analyzer is typically either a quadropole or ion trap, which employs large electrodes to produce **oscillating electric fields** inside the chamber containing the ions.
-**These instruments act as mass filters: the electric field is adjusted over a broad range to select a single peptide ion and discard all the others in the peptide mixture**.

In tandem mass spectrometry, this single ion is then exposed to an **inert, high-energy gas**, which collides with the peptide, resulting in **fragmentation, primarily at peptide bonds**. The second mass analyzer then determines the masses of the peptide fragments, which can be used by computational methods to determine the amino acid sequence of the original peptide and thereby identify the protein from which it came.



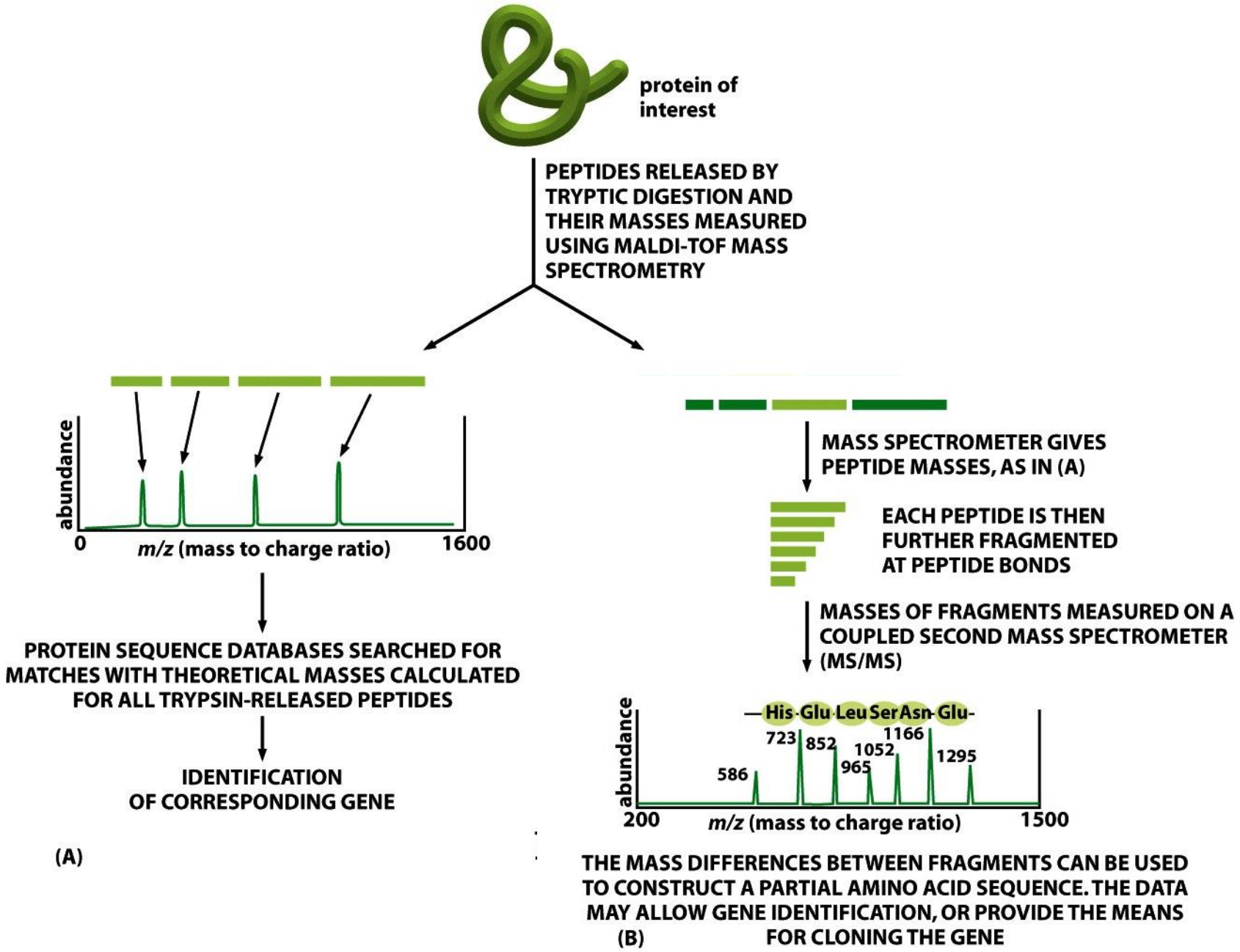


Figure 8-21 *Molecular Biology of the Cell* (© Garland Science 2008)

Mass Spectrometry Provides a Highly Sensitive Method for Identifying Unknown Proteins

- Tandem mass spectrometry is also useful for detecting and precisely mapping post-translational modifications of proteins, such as **phosphorylations** or **acetylations**.

- Because these modifications impart a characteristic mass increase to an amino acid, they are easily detected during the analysis of peptide fragments in the second mass analyzer, and the precise site of the modification can often be deduced from the spectrum of peptide fragments.

- A powerful, "two-dimensional" mass spectrometry technique can be used to determine all of the proteins present in an organelle or another complex mixture of proteins:

-First, the mixture of proteins present is digested with trypsin to produce short peptides.
-Next, these peptides are separated by automated high-performance liquid chromatography (LC).
-Every peptide fraction from the chromatographic column is injected directly into an electrospray ion source on a tandem mass spectrometer (MS/MS), providing the amino acid sequence and post-translational modifications for every peptide in the mixture.

- This method, often called **LC-MS/MS**, is used to identify hundreds or thousands of proteins in complex protein mixtures from specific organelles or from whole cells.
- It can also be used to map all of the phosphorylation sites in the cell, or all of the proteins targeted by other post-translational modifications such as acetylation or ubiquitylation.

Protein Structure Can Be Determined Using X-Ray Diffraction

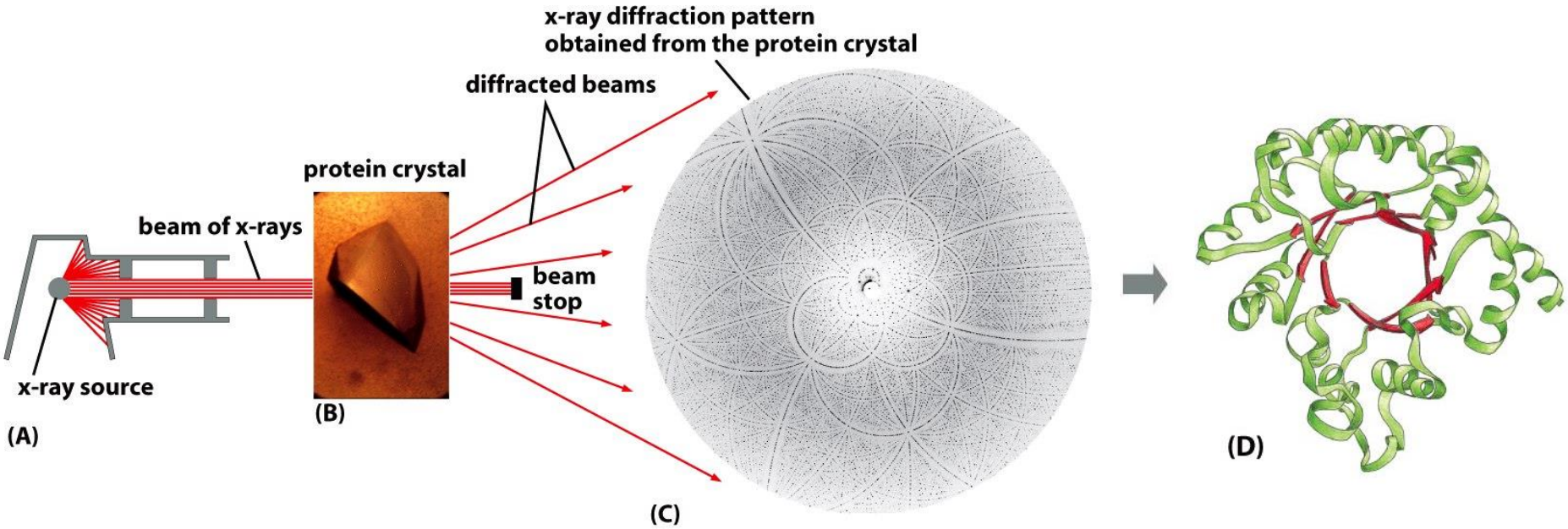
The main technique that has been used to discover the three-dimensional structure of molecules, including proteins, at atomic resolution is **x-ray crystallography**.

X-rays, like light, are a form of electromagnetic radiation, but they have a much shorter wavelength, typically around **0.1 nm** (the diameter of a hydrogen atom).

If a narrow beam of parallel x-rays is directed at a sample of a **pure protein**, most of the x-rays pass straight through it.

A small fraction, however, are scattered by the atoms in the sample.

If the sample is a well-ordered crystal, the scattered waves reinforce one another at certain points and appear as diffraction spots when recorded by a suitable detector.



Protein Structure Can Be Determined Using X-Ray Diffraction

- **The position and intensity of each spot in the x-ray diffraction pattern contain information about the locations of the atoms in the crystal that gave rise to it.**

- Deducing the three-dimensional structure of a large molecule from the diffraction pattern of its crystal is a complex task and was not achieved for a protein molecule until 1960.

- But in recent years **x-ray diffraction analysis** has become increasingly automated, and now the slowest step is likely to be the generation of suitable protein crystals.

- This step requires large amounts of very pure protein and often involves years of trial and error to discover the proper crystallization conditions; the pace has greatly accelerated with the use of **recombinant DNA techniques** to produce pure proteins and **robotic techniques** to test large numbers of crystallization conditions.

- Analysis of the resulting diffraction pattern produces a **complex three-dimensional electron-density map.**

- Interpreting this map—**translating its contours into a three-dimensional structure**—is a complicated procedure that requires **knowledge of the amino acid sequence of the protein**.

- Largely by trial and error, the sequence and the electron-density map are correlated by computer to give the best possible fit.

Protein Structure Can Be Determined Using X-Ray Diffraction

The reliability of the final atomic model depends on the resolution of the original crystallographic data:

0.5 nm resolution might produce a low-resolution map of the polypeptide backbone, whereas a resolution of 0.15 nm allows all of the non-hydrogen atoms in the molecule to be reliably positioned.

- A complete atomic model is often too complex to appreciate directly, but simplified versions that show a protein's essential structural features can be readily derived from it.

- The three-dimensional structures of tens of thousands of different proteins have been determined by x-ray crystallography or by NMR spectroscopy—enough to allow the grouping of common structures into families.

- <u>These structures or protein folds often seem to be more conserved in evolution than are the amino acid sequences that form them.</u>

- X-ray crystallographic techniques can also be applied to the study of macromolecular complexes. The method was used, for example, to determine the structure of the ribosome, a large and complex machine made of several RNAs and more than 50 proteins.

NMR Can Be Used to Determine Protein Structure in Solution

- **Nuclear magnetic resonance (NMR) spectroscopy** has been widely used for many years to analyze the structure of small molecules, small proteins, or protein domains.
- Unlike x-ray crystallography, NMR does not depend on having a crystalline sample.
- It simply requires a small volume of concentrated protein solution that is placed in a strong magnetic field;
- indeed, it is the main technique that yields detailed evidence about **the three-dimensional structure of molecules in solution.**
- Certain atomic nuclei, particularly hydrogen nuclei, have a magnetic moment or spin: that is, they have an intrinsic magnetization, like a bar magnet.
- The spin aligns along the strong magnetic field, but it can be changed to a misaligned, excited state in response to applied radiofrequency (RF) pulses of electromagnetic radiation.
- When the excited hydrogen nuclei return to their aligned state, they **emit RF radiation**, which can be measured and displayed as a spectrum.
- The nature of the emitted radiation depends on the environment of each hydrogen nucleus, and if one nucleus is excited, it influences the absorption and emission of radiation by other nuclei that lie close to it.

NMR Can Be Used to Determine Protein Structure in Solution

- It is consequently possible, by an ingenious elaboration of the basic NMR technique known as two-dimensional NMR, to distinguish the signals from hydrogen nuclei in different amino acid residues, and to identify and measure the small shifts in these signals that occur when these hydrogen nuclei lie close enough together to interact.

- Because the size of such a shift reveals the distance between the interacting pair of hydrogen atoms, NMR can provide information about the distances between the parts of the protein molecule.

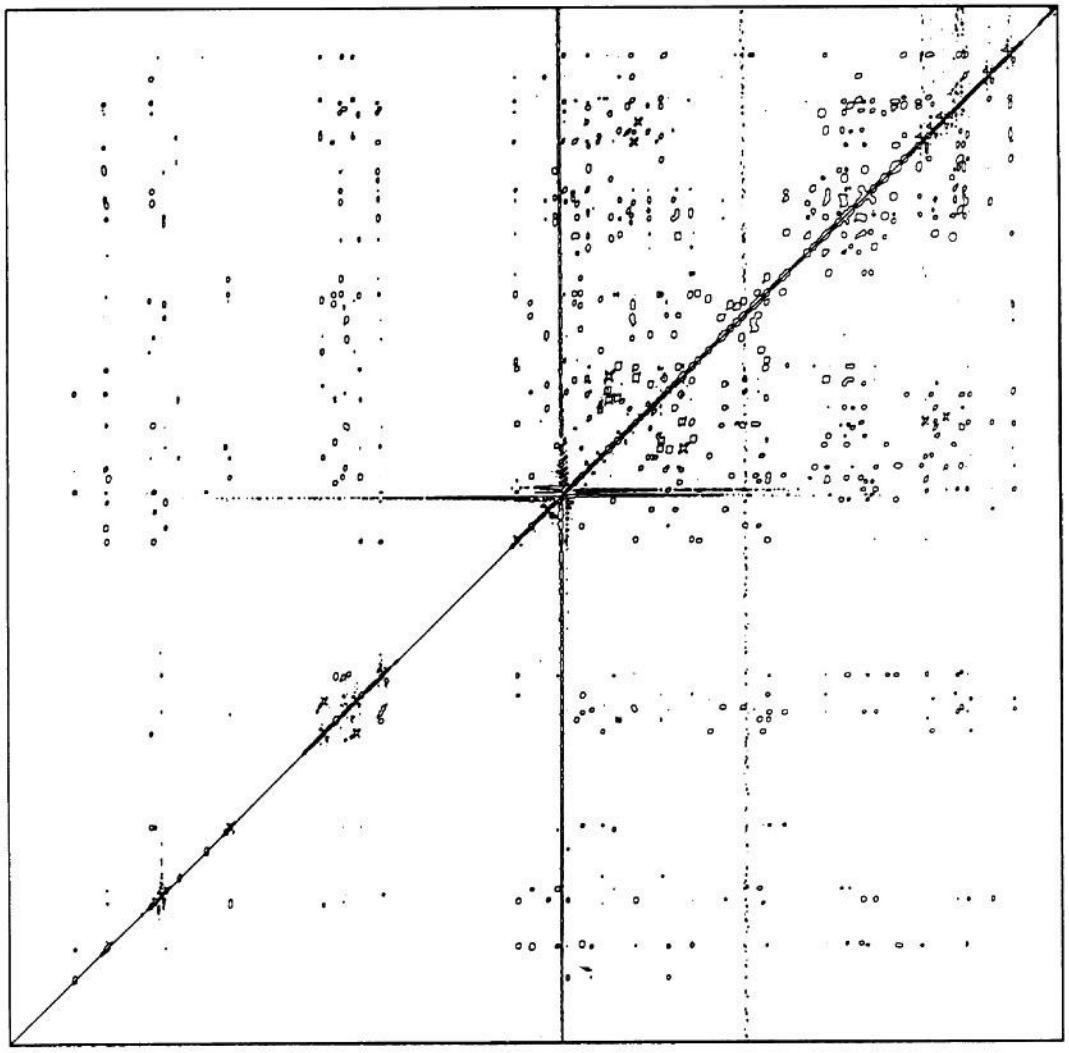
- By combining this information with a knowledge of the amino acid sequence, it is possible in principle to compute the three-dimensional structure of the protein.

- For technical reasons, the structure of small proteins of about 20,000 daltons or less can be most readily determined by NMR spectroscopy.

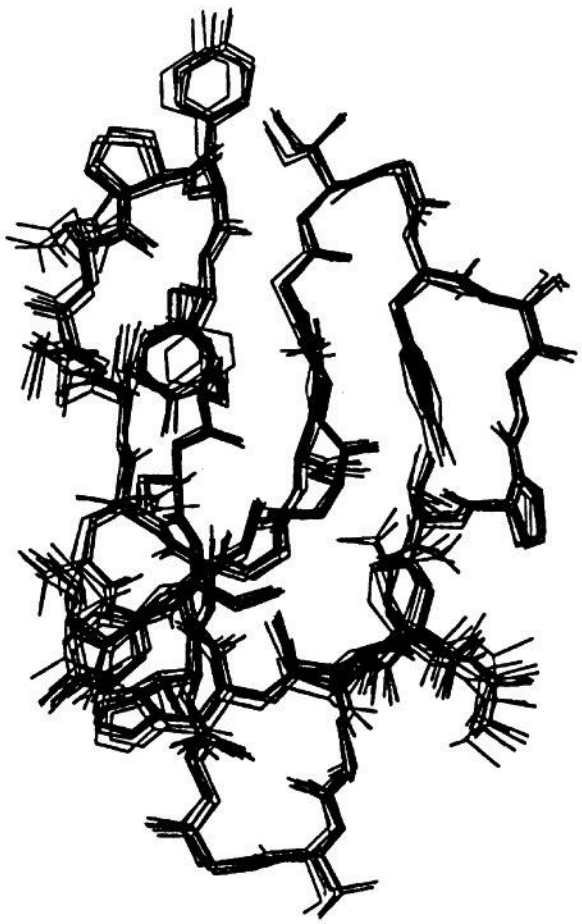
- Resolution decreases as the size of a macromolecule increases. But recent technical advances have now pushed the limit to about 100,000 daltons, thereby making the majority of proteins accessible for structural analysis by NMR.

- Because NMR studies are performed in solution, this method also offers a convenient means of **monitoring changes in protein structure**, for example during protein folding or when the protein binds to another molecule.

- NMR is also used widely to investigate molecules other than proteins and is valuable, for example, as a method to determine the three-dimensional structures of RNA molecules and the complex carbohydrate side chains of glycoproteins.



(A)



(B)