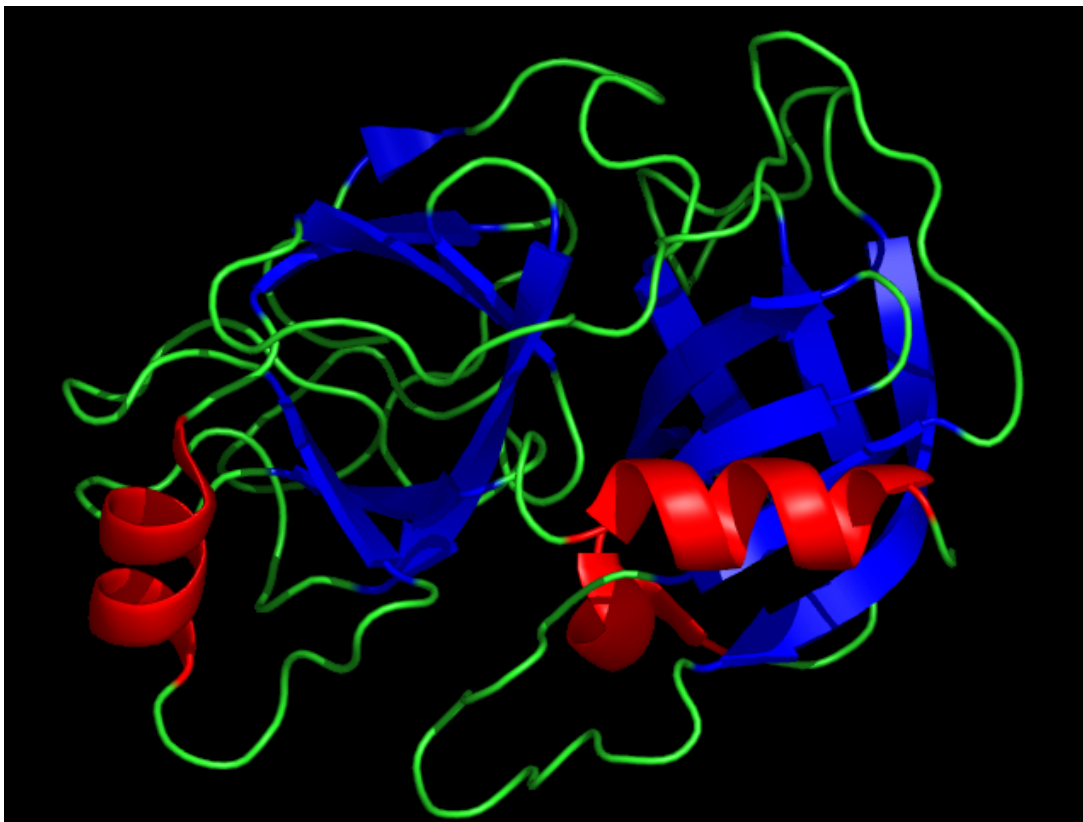


## Biotechnology: Analytical Techniques

[See online here](#)

**Biotechnology is an emerging field that allows scientists to use analytical techniques to study the structure and function of biomolecules. Protein purification is an all but mandatory step for studying macromolecules, but this task is not necessarily easy. Several experimental methods, deductive logic, and even a little bit of luck make this possible. Here you'll read about some of the methods.**



### Introduction

Understanding the structure of a given protein is key to deciphering its function. Like the other major biological macromolecules, nucleic acids and polysaccharides, **proteins are polymers of smaller units**. Unlike many nucleic acids, proteins do not have a uniform, regular structure. This is because the 20 kinds of amino acid residues from which proteins are made have widely differing chemical and physical properties.

The sequence in which the amino acids are strung together to form proteins can be **analyzed directly, or indirectly via DNA sequencing**. In either case, amino acid sequence information provides insights into the chemical and physical properties of proteins. Additionally, their relationships to other proteins, and ultimately their

mechanisms of action in living organisms can be determined.

## Proteomics – Basic Information

**Controlled conditions**, such as pH and temperature, are **required for protein purification**. Failure to control the pH of the solution most often results in the denaturation of proteins. Additionally, the thermal stability of proteins varies depending on the protein being studied. Proteins can be purified on the basis of their charge, polarity, size, binding specificity, and solubility.

Differences in solubility permit proteins to be concentrated and purified by salting out. The phenomenon is primarily a result of the competition between the added salt ions and the other dissolved solutes for molecules of solvent. At very high salt concentrations, so many of the added ions are dissolved that there is significantly less bulk solvent available to dissolve other substances, including proteins.

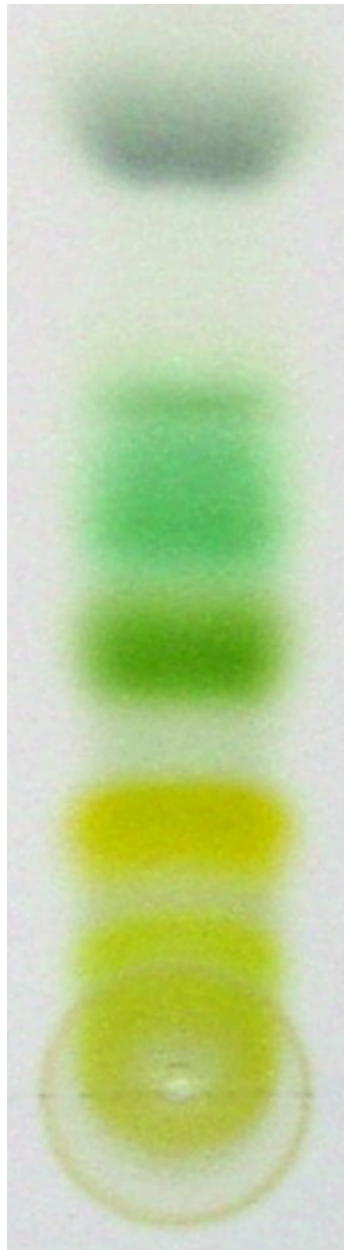


Image: "Chromatography of chlorophyll" by Flo-commonswiki.  
License: [CC BY-SA 2.5](https://creativecommons.org/licenses/by-sa/2.5/)

**Chromatography** separates soluble substances by their rate of movement through an insoluble matrix and is a technique for **purifying molecules by charge** (ion-exchange chromatography), **hydrophobicity** (hydrophobic interaction chromatography), **size** (gel filtration chromatography), and **binding specificity** (affinity chromatography). Binding and elution often depend on the salt concentration and pH.

In **ion-exchange chromatography**, charged molecules bind to oppositely charged groups that are linked to a matrix such as cellulose. Anions bind to cationic groups on **anion exchangers** and cations bind to anionic groups on **cation exchangers**. The most frequently used anion exchanger is a matrix with attached **diethylaminoethyl (DEAE)** groups, and the most frequently used cation exchanger is a matrix bearing **carboxymethyl (CM)** groups.

Binding affinities of proteins depend on the presence of other ions that compete with the protein for binding to the ion exchanger and on the pH of the solution, which influences the net charge of the protein.

Analysis of a protein's sequence begins with end-group analysis, to determine the number of different subunits and the cleavage of disulfide bonds.

**The N-terminal analysis** reveals the number of different types of subunits. Each polypeptide chain (if it is not chemically blocked) has an N-terminal residue. Identifying this "end group" can establish the number of chemically distinct polypeptides in a protein. For example, insulin has equal amounts of the N-terminal residues Gly and Phe, which indicate that it has equal numbers of two nonidentical polypeptide chains.

The N-terminus of a polypeptide can be determined by several methods. The fluorescent compound **dansyl chloride** reacts with primary amines to yield dansylated polypeptides. The treatment of a dansylated polypeptide with aqueous acid at high temperature hydrolyzes its peptide bonds. This sets the dansylated N-terminal residue-free, which can then be separated chromatographically from the other amino acids and identified by its intense yellow fluorescence

**Edman degradation** uses cleaved polypeptides whose fragments are suitable for sequencing. Residues are removed, one at a time, from the N-terminus. Peptides can also be sequenced by mass spectrometry.

Disulfide bonds between cysteine residues must be cleaved to separate polypeptide chains that are disulfide-linked, to ensure that polypeptide chains are fully linear. Disulfide bonds can be reductively cleaved by treating them with **2-mercaptoethanol**.

Polypeptides that are longer than 40–100 residues cannot be directly sequenced and must, therefore, be cleaved, either enzymatically or chemically, to specific fragments that are small enough to be sequenced. **Proteases** differentially cleave either internal peptide bonds or the N/C-terminal residues depending on their cleavage requirements.

The digestive enzyme **trypsin** has the greatest specificity and is, therefore, the most valuable member of the arsenal of endopeptidases used to fragment polypeptides. It cleaves peptide bonds on the C side (toward the carboxyl terminus) of the positively charged residues arginine and lysine if the next residue is not proline. Several chemical reagents promote peptide bond cleavage at specific residues. The most useful of these, **cyanogen bromide (CNBr)**, cleaves on the C side of methionine residues.

A protein's sequence is then reconstructed from the sequences of overlapping peptide fragments and from information about the locations of disulfide bonds. The sequences of

numerous proteins are archived in publicly available databases.

## Proteomics – Worked Examples

### Worked example 1

Find the amino acid sequence of a pentadecapeptide using the following information:

- **CNBr1:** Alanine-Glycine-Leucine-Lysine-Methionine-Proline
- **CNBr2:** Alanine-Arginine-Aspartate-Cysteine-Glutamine-Glycine
- **CNBr3:** Methionine-Phenylalanine-Tryptophan
  
- **Trypsin1:** Alanine-Cysteine-Glutamine
- **Trypsin2:** Alanine-Glycine-Lysine-Methionine-Phenylalanine-Tryptophan
- **Trypsin3:** Arginine-Aspartate-Glycine-Leucine-Methionine-Proline

N-terminal analysis revealed the following:

- **Phenylalanine:** Pentadecapeptide
- **CNBr1 N-terminus:** Glycine
- **CNBr2 N-terminus:** Aspartate
- **Trypsin1 N-terminus:** Cysteine
- **Trypsin3 N-terminus:** Leucine

C-terminal analysis revealed that glutamine was the C-terminus of the pentadecapeptide.

**Step 1: Arrange the given information and change the amino acids to their one-letter abbreviations to make things easier for you.**

(Note that bold letters are information that can be deduced from the given information.)

- **CNBr1 (C1):** Alanine-Glycine-Leucine-Lysine-Methionine-Proline ⇒ **G**-A(orP)-K-L-F(orA)-**M**
- **CNBr2 (C2):** Alanine-Arginine-Aspartate-Cysteine-Glutamine-Glycine ⇒ **D**-G(orA)-R-C-A(orG)-**Q**
- **CNBr3 (C3):** Methionine-Phenylalanine-Tryptophan ⇒ **F-W-M**
  
- **Trypsin1 (T1):** Alanine-Cysteine-Glutamine ⇒ **C-A-Q**
- **Trypsin2 (T2):** Alanine-Glycine-Lysine-Methionine-Phenylalanine-Tryptophan ⇒ **F-W-M**-G(orA)-A(orG)-K
- **Trypsin3 (T3):** Arginine-Aspartate-Glycine-Leucine-Methionine-Proline ⇒ **L**-P(orG)-M-D-G(orP)-R

**Step 2: Find a probable alignment of fragments.**

From the given information, the N-terminus for the uncleaved peptide is phenylalanine (F), and the C-terminus is glutamine (Q), so you can place those in their respective places. Probable alignment of fragments:

- T2-T3-T1
- C3-C1-C2

**Step 3: Align your fragments into an arbitrary pentadecapeptide using your probable alignment.**

— — — — —

**F-W-M-G(orA)-A(orG)-K- L-P(orG)-M-D-G(orP)-R- C-A-Q**

**Step 4: Check your work using what you know about peptide cleavage reagents and enzymes.**

- CNBr cleaves on the C-side of methionine (M) residues
- Trypsin cleaves on the C-side of arginine (R) and lysine (K) residues; if the next residue is not proline (P)

CNBr would cleave at the following sites:

**F-W-M**■G(orA)—A(orG)—K—**L**—P(orG)—M■D—G(orP)—R—**C-A-Q**

Trypsin would cleave at the following sites:

**F-W-M**—G(orA)—A(orG)—K■—**L**—P(orG)—M—D—G(orP)—R■—**C-A-Q**

**Step 5: Ask yourself if this alignment of residues makes sense, and if any of your unknowns conflict with each other.**

Glycine and alanine are both nonpolar residues, and would not interfere with CNBr or trypsin cleavage; however, you know that proline cannot be the next residue on the C-side of a tryptic cleavage. Therefore your most probable sequence of amino acid residues for the pentadecapeptide is:

F-W-M-G-A-K-L-G-M-D-P-R-C-A-Q

## Worked example 2

A heptapeptide is cleaved into a dipeptide and a pentapeptide by cyanogen bromide and into a tripeptide and a tetrapeptide by treatment with trypsin. The tripeptide produced by tryptic hydrolysis, but not the tetrapeptide, absorbs light at 280nm. Given these experimental results, which of the following is the most likely sequence of amino acid residues in the peptide sample?

- Glycine-Methionine-Tryptophan-Lysine-Alanine-Valine-Arginine
- Arginine-Methionine-Alanine-Lysine-Tryptophan-Valine-Glycine
- Glycine-Lysine-Alanine-Tryptophan-Methionine-Valine-Arginine
- Methionine-Glycine-Alanine-Lysine-Tryptophan-Valine-Arginine
- Glycine-Methionine-Alanine-Lysine-Tryptophan-Valine-Arginine

**Step 1: Rule out sample heptapeptides that do not follow the mechanism of CNBr cleavage and form a dipeptide and pentapeptide.**

- Remember that CNBr cleaves on the C-side of methionine residues.
- Glycine-Methionine■Tryptophan-Lysine-Alanine-Valine-Arginine
- Arginine-Methionine■Alanine-Lysine-Tryptophan-Valine-Glycine
- Glycine-Lysine-Alanine-Tryptophan-Methionine■Valine-Arginine
- Methionine■Glycine-Alanine-Lysine-Tryptophan-Valine-Arginine (This peptide is a hexapeptide and a single amino acid residue, so we can rule it out.)
- Glycine-Methionine■Alanine-Lysine-Tryptophan-Valine-Arginine

**Step 2: Rule out sample heptapeptides that do not follow the mechanism of cleavage by trypsin and form a tripeptide and a tetrapeptide.**

Remember that trypsin cleaves on the C side of lysine and arginine residues, except when proline is the next residue.

- Glycine-Methionine-Tryptophan-Lysine-Alanine-Valine-Arginine
- Arginine-Methionine-Alanine-Lysine-Tryptophan-Valine-Glycine (This peptide is two tripeptides and a single amino acid residue, so we can rule it out.)
- Glycine-Lysine-Alanine-Tryptophan-Methionine-Valine-Arginine (This peptide is a hexapeptide and a single amino acid residue, so we can rule it out.)
- Glycine-Methionine-Alanine-Lysine-Tryptophan-Valine-Arginine

**Step 3: Find the peptide that possesses a tripeptide fragment produced by tryptic hydrolysis that would absorb light at 280nm.**

Remember that tryptophan absorbs light at 280nm.

- Glycine-Methionine-Tryptophan-Lysine-Alanine-Valine-Arginine (This peptide can be ruled out because the tetrapeptide absorbs light, and we are looking for a tripeptide.)
- Glycine-Methionine-Alanine-Lysine-Tryptophan-Valine-Arginine (This peptide has the amino acid sequence that follows all of our experimental criteria.)

## Worked example 3

Given the following pKa values: 2.0 for the C1 carboxyl group of an amino acid; 4.0 for a side-chain carboxyl group; 9.5 for an  $\alpha$ -amino group; and 12.5 for a guanidino group - the best estimate for the net charge of the peptide Glutamate-Glutamine-Arginine-Valine at pH = 12.5 is:

**Step 1: Draw out the peptide with the R groups, as well as the N and C-terminal functional groups.**

**Step 2: Assign charges on the R groups and N and C-terminal functional groups based on their pKa's in a pH of 12.5.**

The N-terminus will be in the uncharged  $\text{NH}_2$  form. Glutamate will have a negative charge because its pKa is 4.0. Valine will not have a charge, but the C-terminus will have a negative charge because its pKa is 2.0. Arginine will have a charge of +0.5 because its pKa is equal to the pH of the solution. This means that half of the time it will be protonated (which would give it a +1 charge), and half the time it will not (a neutral charge).

**Step 3: Add up your charges.**

$$^{-}1.0 + ^{-}1.0 + ^{+}0.5 = ^{+}0.5$$

## References

Aebersold, R. and Mann, M. [Mass spectrometry-based proteomics](https://doi.org/10.1038/422198a). Nature 422, 198-207 (2003). Available at: <https://www.ncbi.nlm.nih.gov/pubmed/12634793>

Boyer, R.F. Biochemistry Laboratory: Modern Theory and Techniques. Benjamin Cummings (2006).

Burgess, R.R. and Deutscher, M.P. (Eds.). Guide to Protein Purification. (2nd ed.). Methods Enzymol. 463 (2009).

Findlay, J.B.C. and Geisow, M.J. (Eds.). Protein Sequencing. A Practical Approach. IRL Press

(1989).

Galperin, M.Y. and Cochrane, G.R. [The 2011 Nucleic Acids Research database issue and online database collection: a community of data resources](#). Nucleic Acids Res. 38, Database issue D1–D6 (2011). Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3013748/>

Janson, J.-C. (Ed.). Protein Purification: Principles, High Resolution Methods, and Applications. (3rd ed.), Wiley (2007).

Ninfa, A.J., Ballou, D.P., and Benore, M.. Fundamental Laboratory Approaches for Biochemistry and Biotechnology. (2nd ed.). Wiley (2010).

Simpson, R.J., Adams, P.D., and Golemis, E.A. (Eds.). Basic Methods in Protein Purification and Analysis. A Laboratory Manual. Cold Spring Harbor Laboratory Press (2009).

Tanford, C. and Reynolds, J. Nature's Robots: A History of Proteins. Oxford University Press (2001).

**Legal Note:** Unless otherwise stated, all rights reserved by Lecturio GmbH. For further legal regulations see our [legal information page](#).

Notes