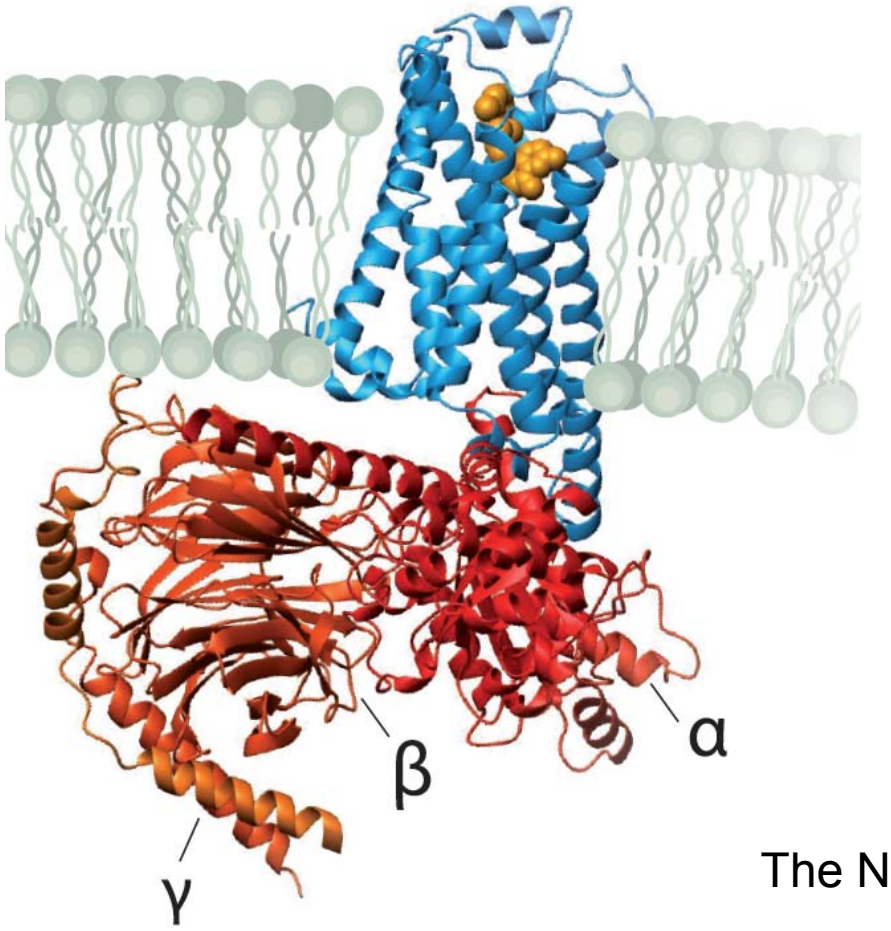
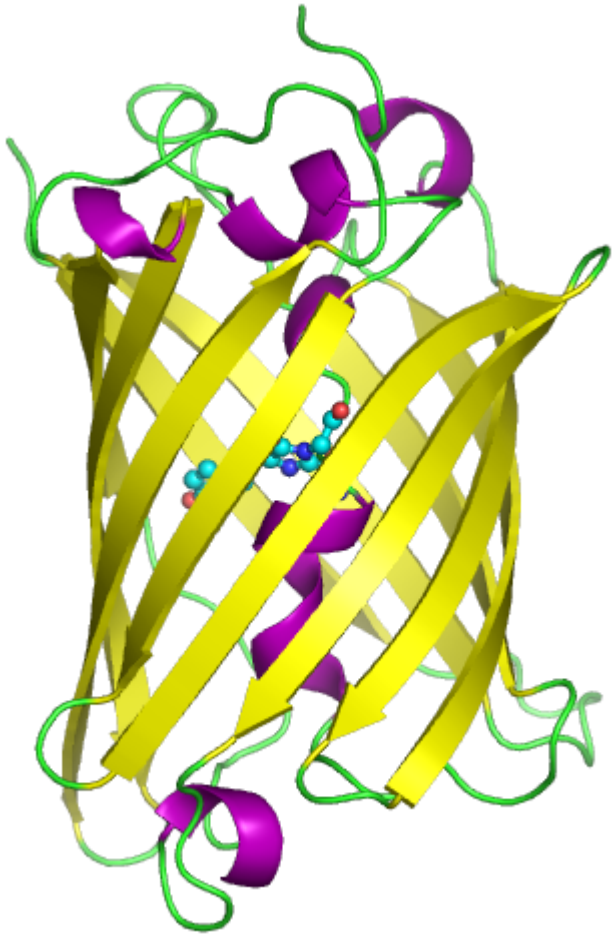
Proteins

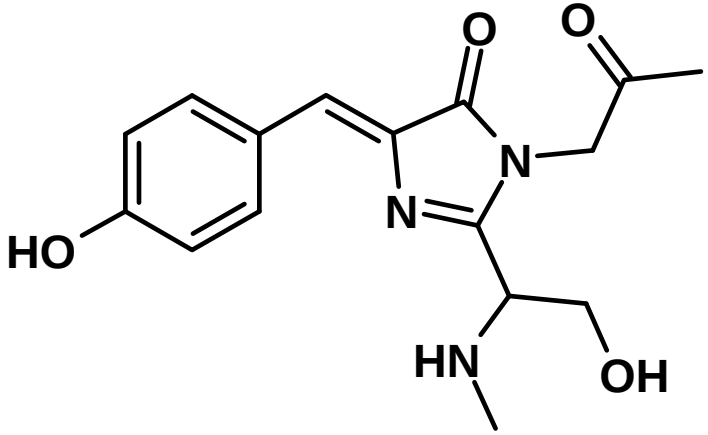
Amino acids, structure and function



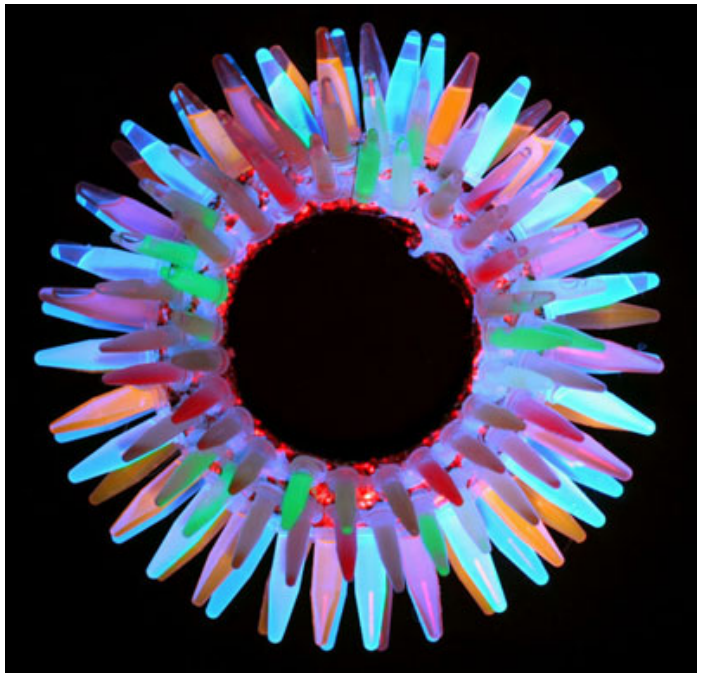
The Nobel Prize in Chemistry 2012

Robert J. Lefkowitz
Brian K. Kobilka





Ser65-Tyr66-Gly67



Variations in GFP structure result in a rainbow of fluorescent proteins

The Nobel prize in chemistry 2008
Osamu Shimomura, Martin Chalfie, Roger Y Tsien
- green fluorescent protein (GFP)

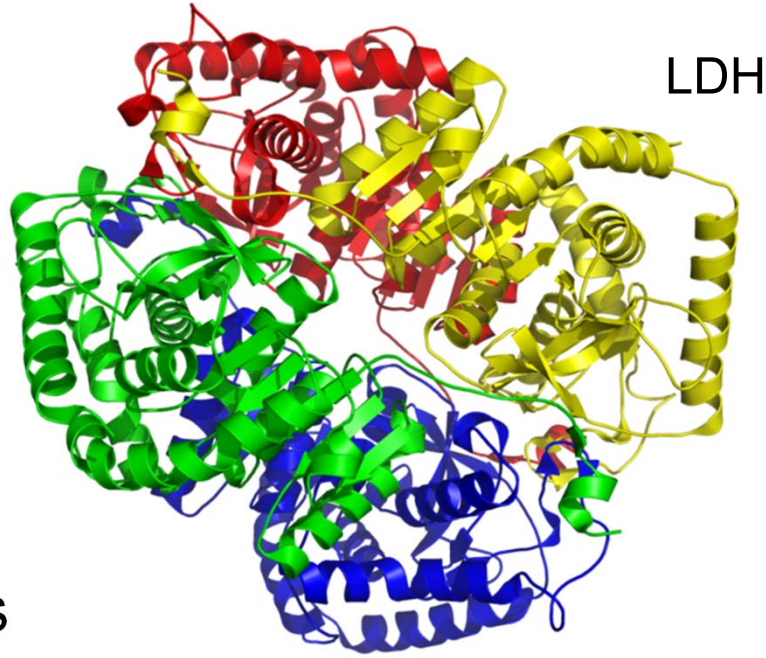
What are proteins doing in the cell?

Proteins have **many functions** in the living cell:

- Catalysis (enzymes)
- Transport and storage
- Mechanical support
- Immune system (antibodies)
- Movements
- Nerve signals
- Control of growth and differentiation

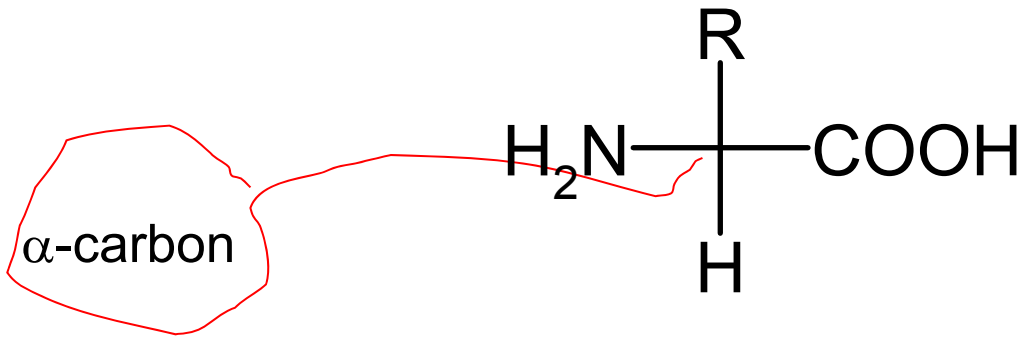
This is possible because:

- Proteins fold into complex **three dimensional structures**
- Proteins have **functional groups** (e.g. -OH, -SH, -COOH, $-NH_2$)
- Proteins are adaptable
- Proteins can **interact** with other molecules



LDH

Proteins are built from 20 different L-amino acids



α-amino acid

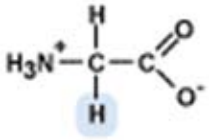
There are 4 groups on the chiral α-carbon:

- H
- COOH (COO^-)
- NH_2 (NH_3^+)
- R = side chain

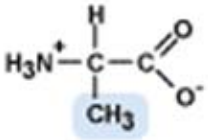
The character of the R-group will give amino acids with different:

- size
- shape
- charge
- hydrophobicity
- hydrogen bonding capacity
- chemical reactivity

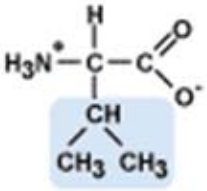
NONPOLAR



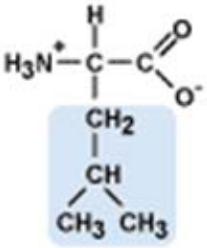
Glycine (Gly)



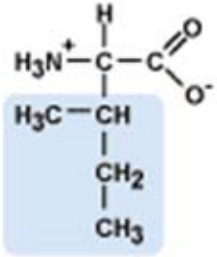
Alanine (Ala)



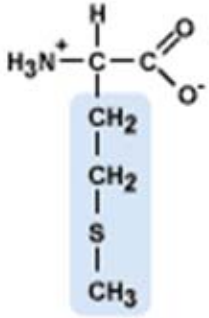
Valine (Val)



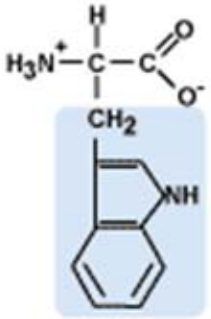
Leucine (Leu)



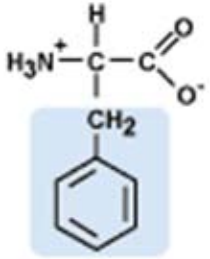
Isoleucine (Ile)



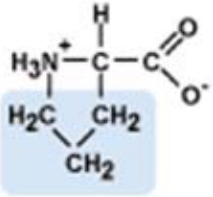
Methionine (Met)



Tryptophan (Trp)

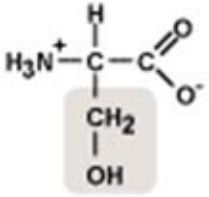


Phenylalanine (Phe)

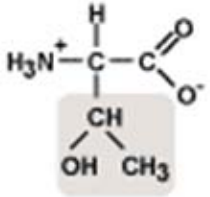


Proline (Pro)

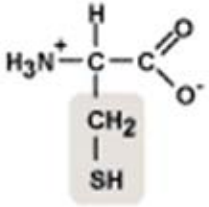
POLAR



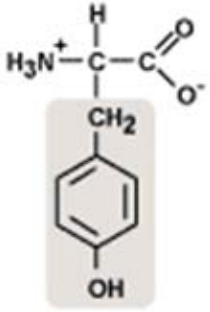
Serine (Ser)



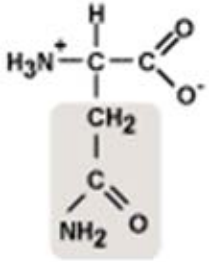
Threonine (Thr)



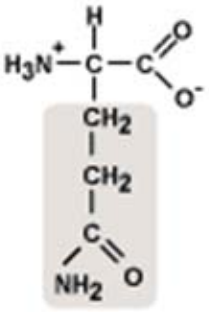
Cysteine (Cys)



Tyrosine (Tyr)

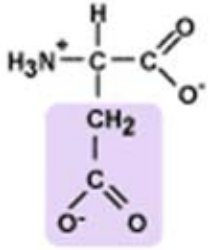


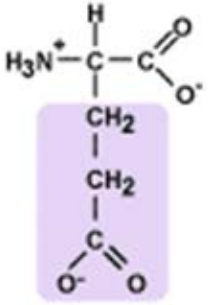
Asparagine (Asn)



Glutamine (Gln)

Electrically Charged

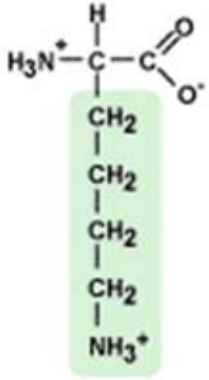




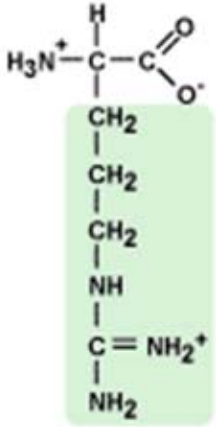
Acidic

Aspartic Acid (Asp)

Glutamic Acid (Glu)

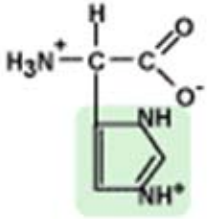


Lysine (Lys)



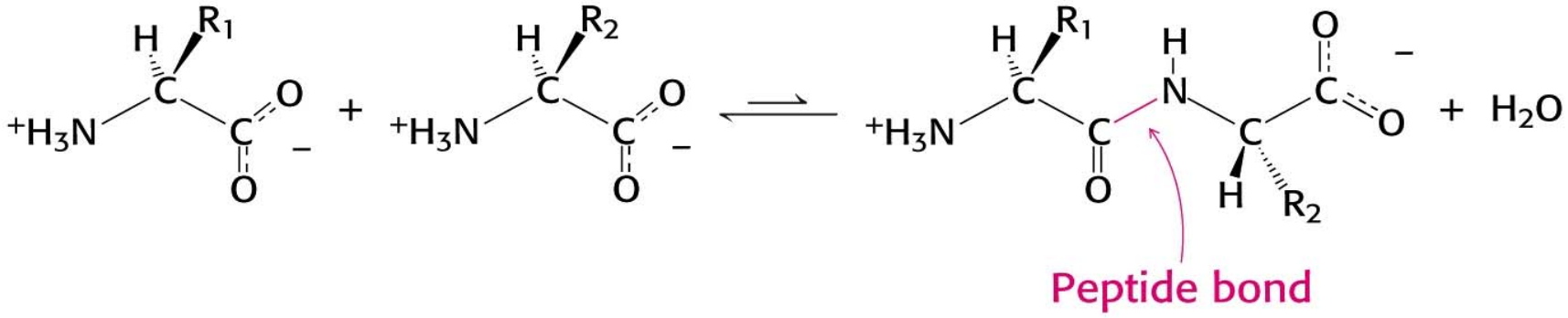
Basic

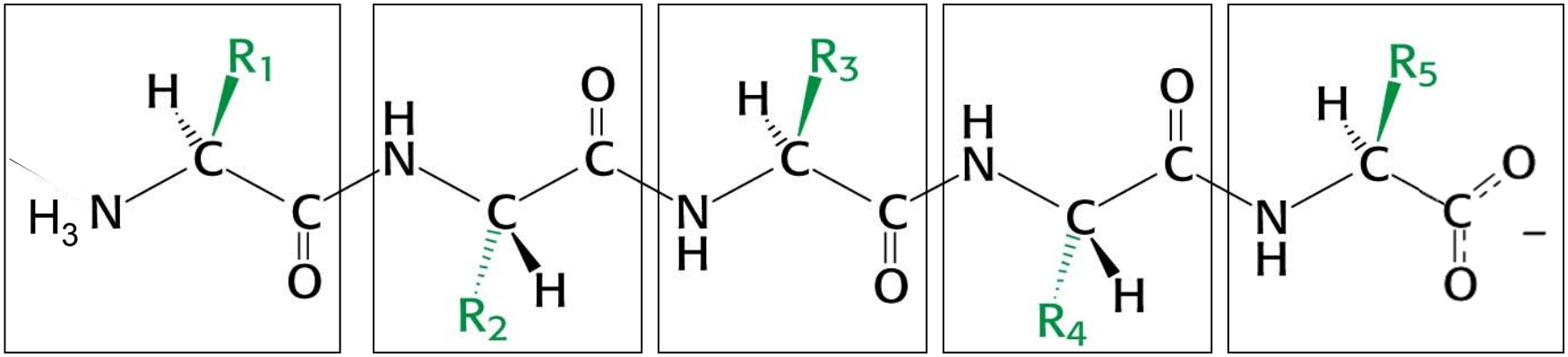
Arginine (Arg)



Histidine (His)

Peptides - polypeptides



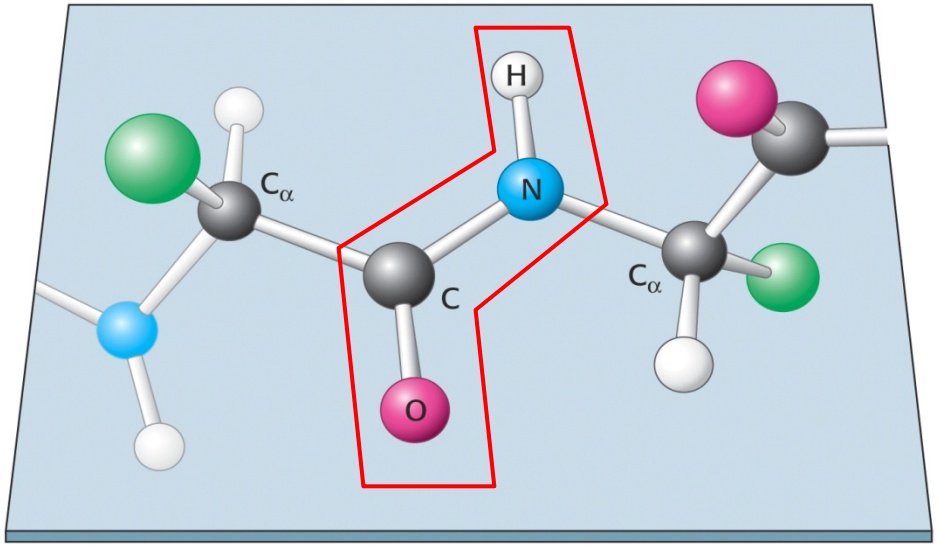


Amino acid 1 | Amino acid 2 | Amino acid 3 | Amino acid 4 | Amino acid 5

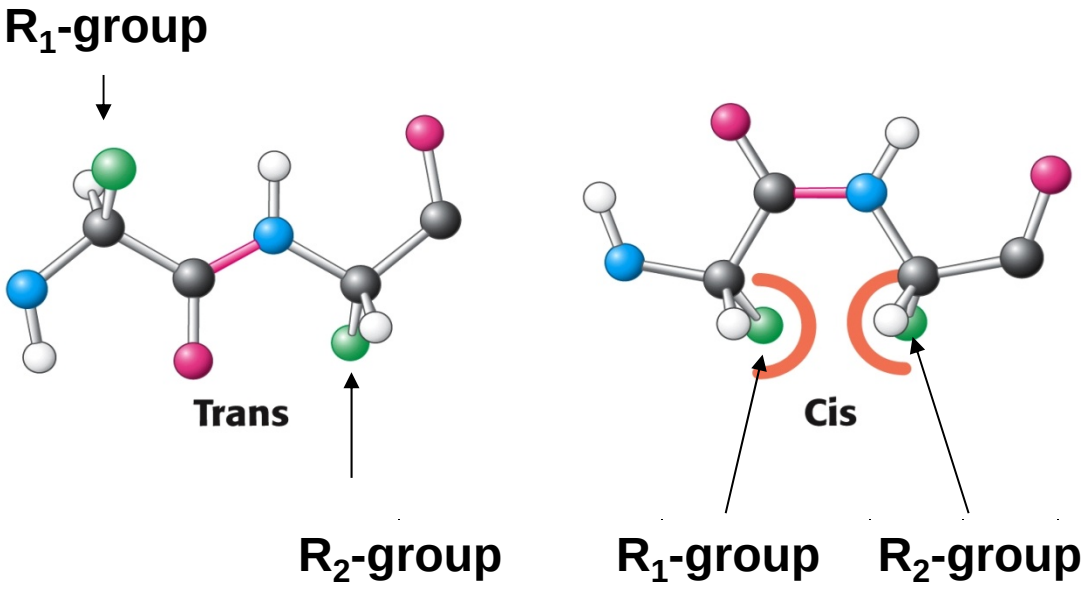
Amino-terminal ***Carboxy-terminal***

The order of amino acids = primary structure

The peptide bond



The peptide bond (CO-NH) is planar and stiff (double bond character)



The peptide bond is almost always in TRANS – configuration because of steric hindrance

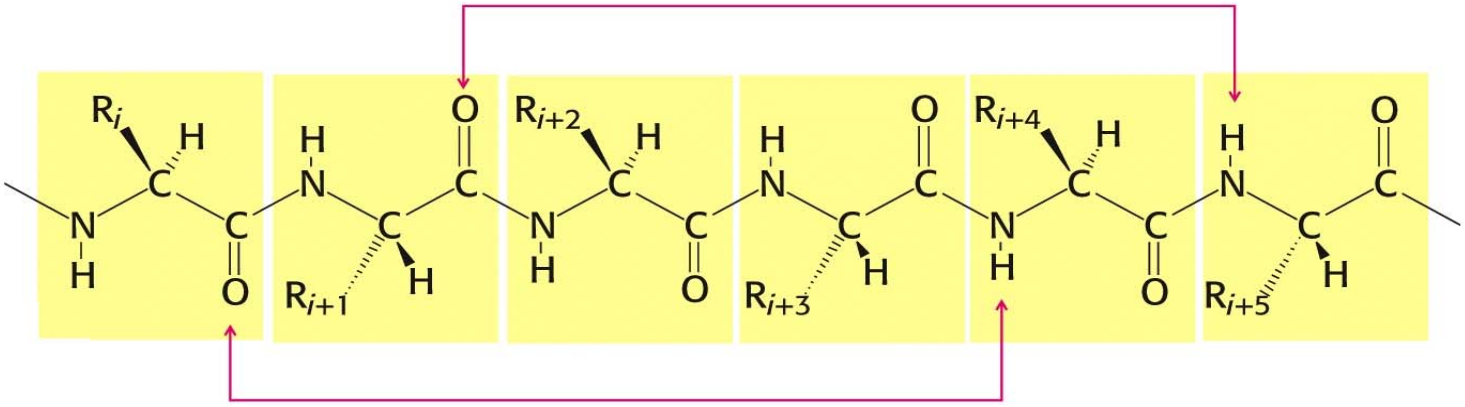
The polypeptide spontaneously folds into regular structures

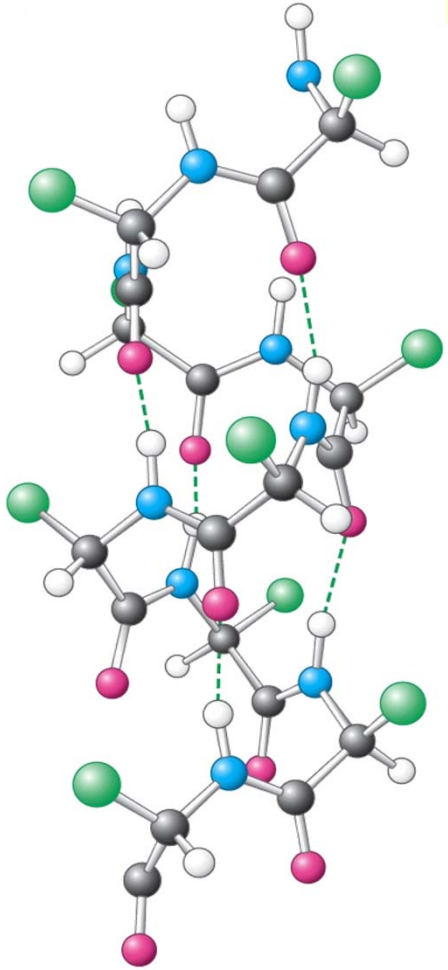
The mixture of stiff elements (the peptide bond) and flexible elements in the polypeptide chain enables the formation of regular structures:

- α-helix
- β-sheet
- β-turn

These are examples of secondary structures

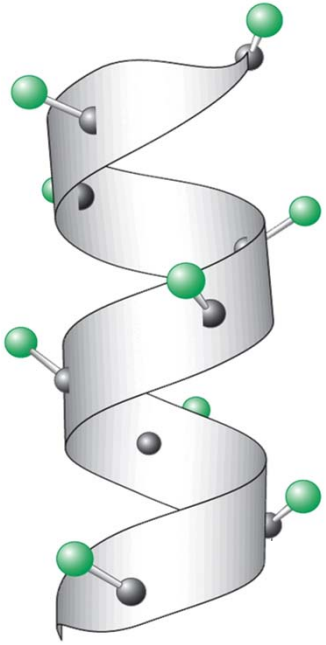
α-Helix





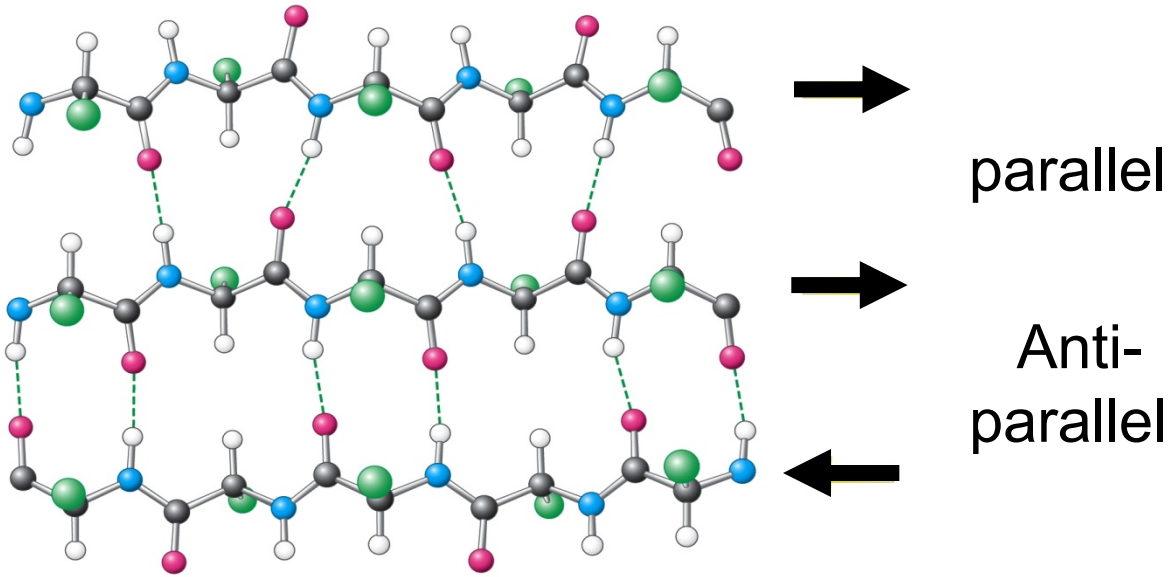
- All NH and CO in the main chain are engaged in hydrogen bonds running parallel to the helix

- 3,6 Amino acids per turn

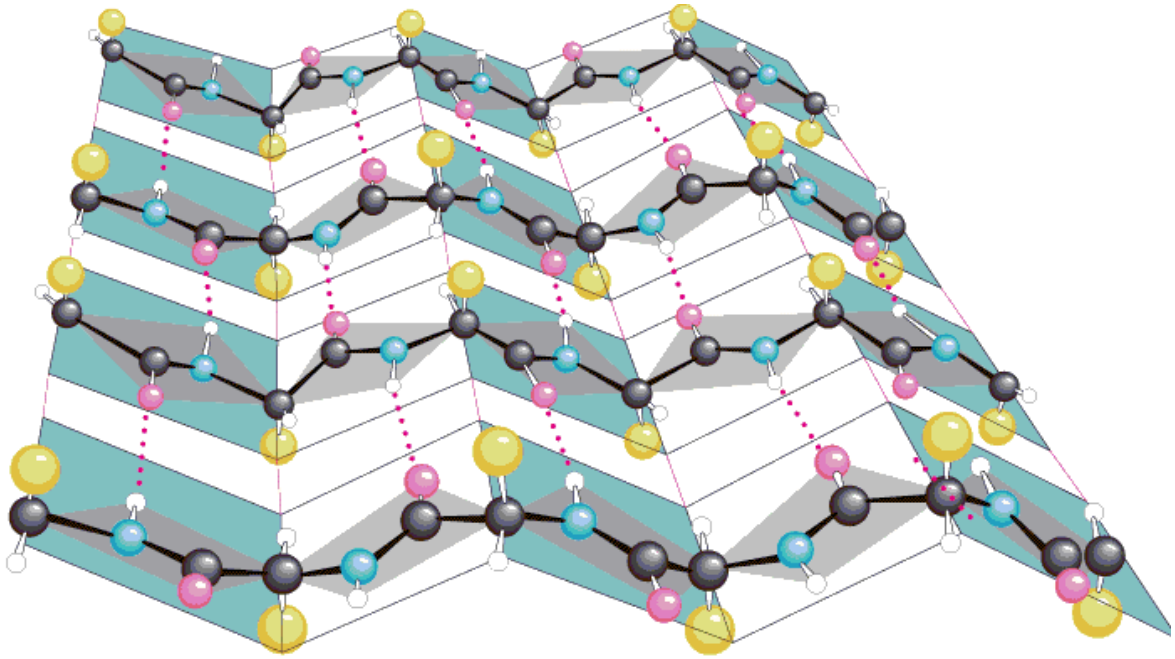


“PILLAR”

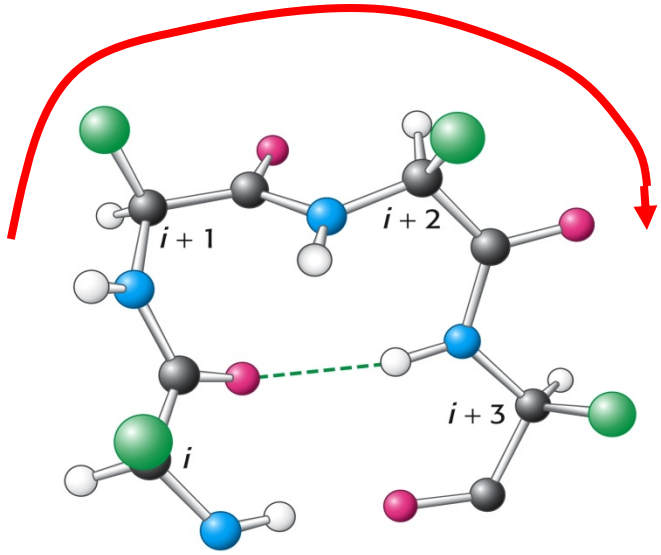
β-sheet (β pleated sheet)



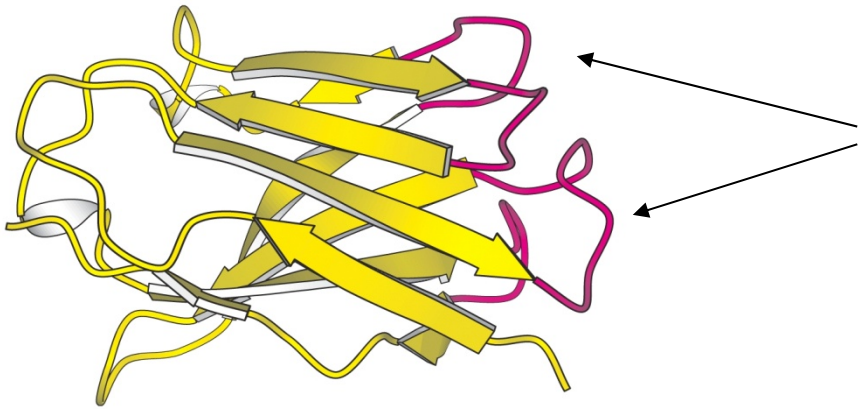
Hydrogen bonds between the stretched strands



β-Turn and loops



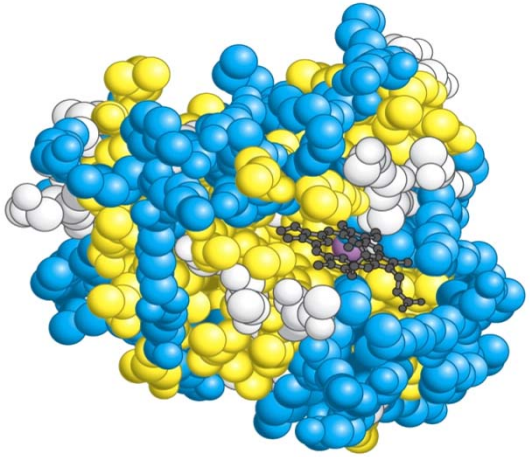
β-Turn has a hydrogen bond between the CO of aa_i and the NH of aa_{i+3}

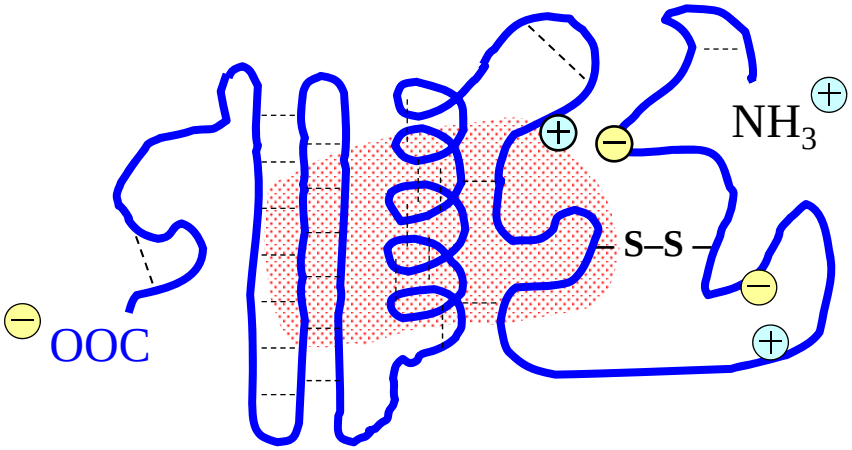


Loops do not have exact regular structures but are stable and well-defined anyway

What keeps the protein together in a complete three-dimensional structure?

Myoglobin



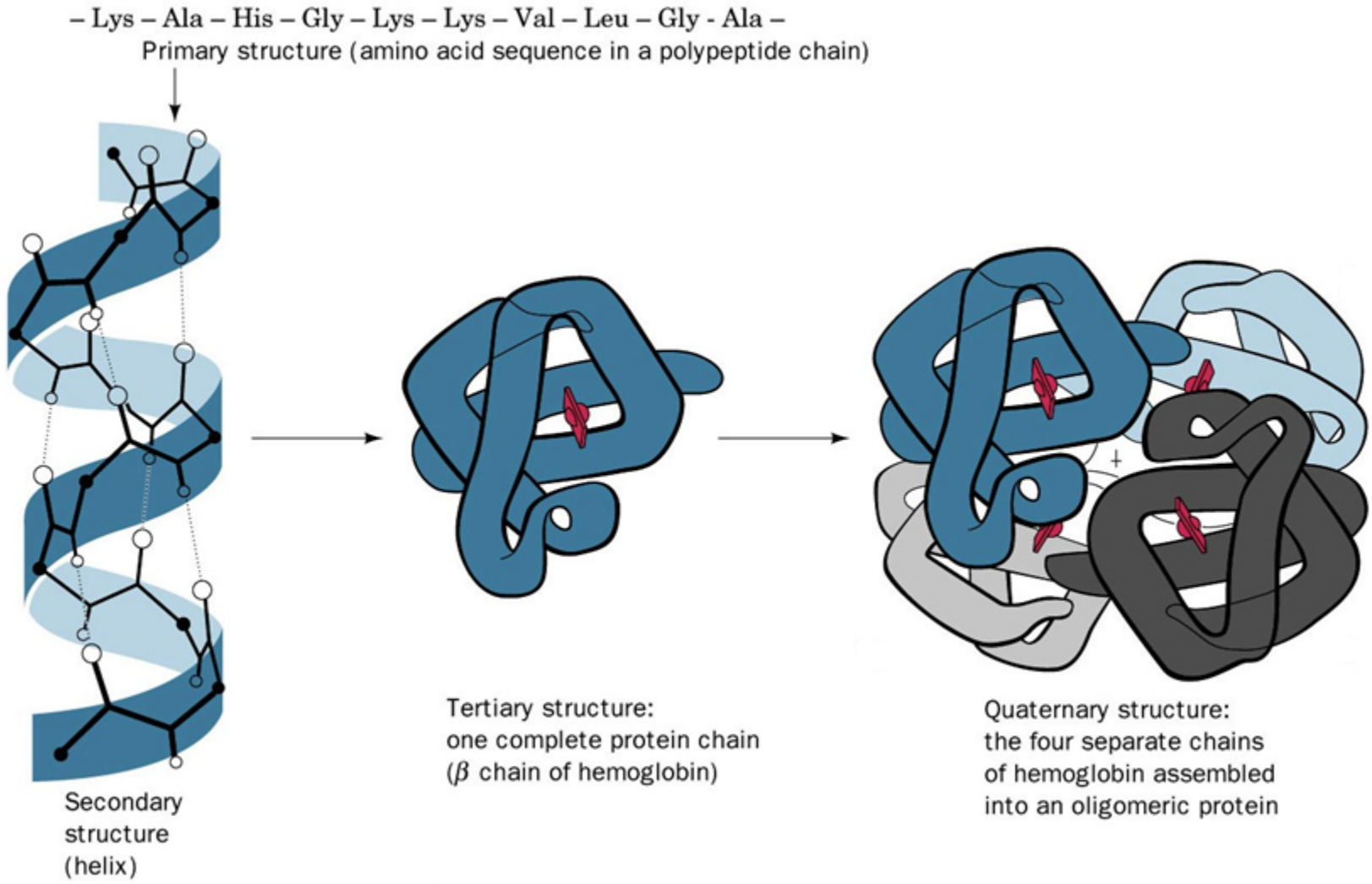


- the secondary structures of α-helix, β-sheet and β-turn
- ionic bonds between side chains
- hydrogen bonds between side chains
- S-S bridges (between two SH-groups of adjacent cysteines)
- the inside of the protein is predominately hydrophobic

Common terms used to describe protein structure

Primary structure:	The sequence of amino acids in polypeptide chain
Secondary structure:	The interactions between amino acids close in the sequence (example: α-helix, β-sheet, β-turn and loops)
Tertiary structure:	Interactions between amino acids distant in the sequence, “The structure at large"
Quaternary structure:	The arrangement of the subunits in a multimeric protein
Subunit:	A part (= a complete polypeptide) of a multimeric protein
Domain:	Some proteins have compact parts with a specific function, e.g. a catalytic domain. Is often coded for by an exon.
Random coil:	Unfolded polypeptide
Native protein:	The correctly folded polypeptide structure
Denatured protein:	The protein has lost its natural folding (e.g. by extensive heating)

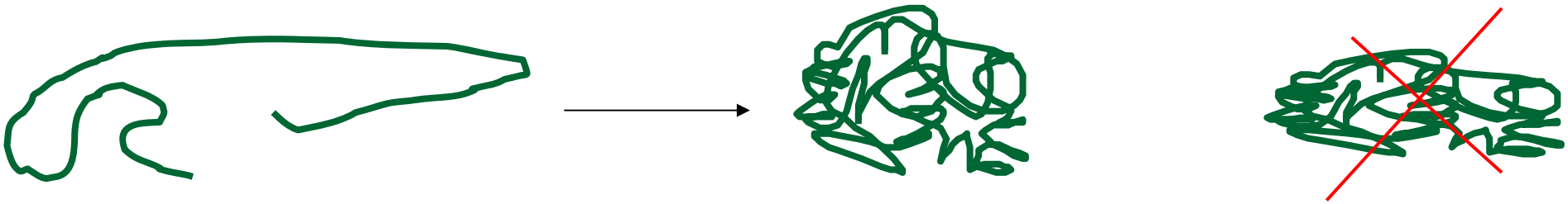
In some proteins several polypeptide chains are joined together to multimeric units



The structure is determined by the sequence

A certain polypeptide sequence always leads to the same structure

Why?



Random coil

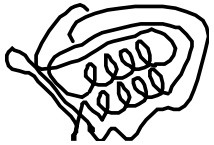
Native protein
(low energy state)

disfavored

The folding process is very rapid

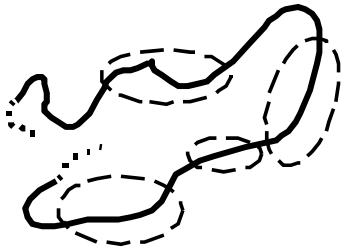
Random coil 

1 s
(10^{27} year)



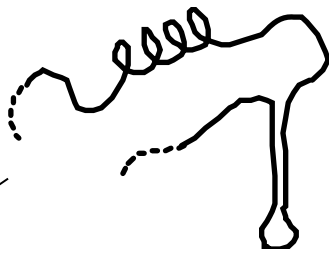
Native protein
= finished, correctly folded protein

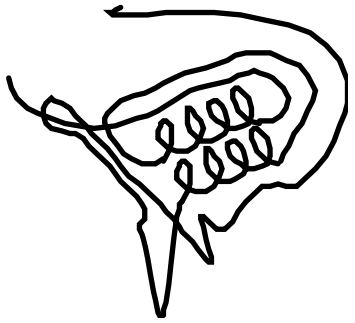
Local folding and keeping of successful intermediates. "Easiest parts first"



Random coil

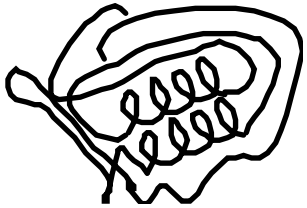
Amino acids have different propensity to participate in α-helix, β–sheet and β-turn





Molten globule:
Compact structure with most parts in place. Hydrophobic amino acids in the centre. Completed

The final adjustment is relatively slow (1 s)



Native protein

The living cell has proteins which assist protein folding

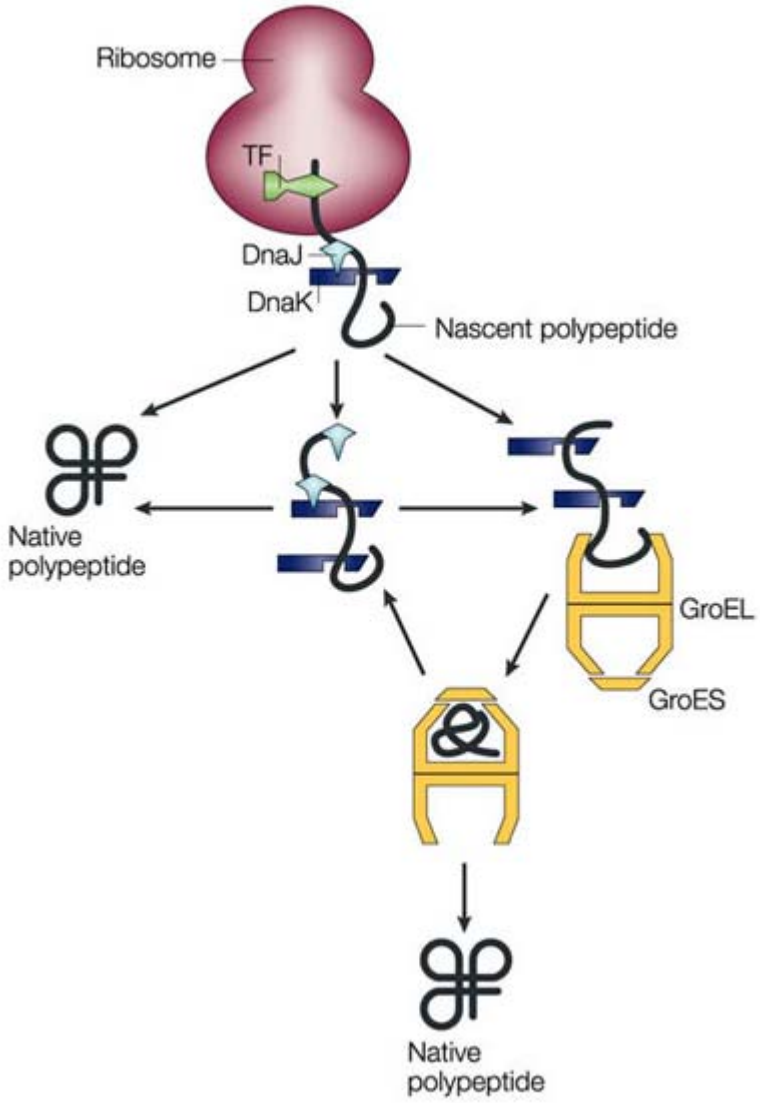
In vitro the protein folding is often easy (dilute solution)

In living cells, high protein conc. can cause problem:

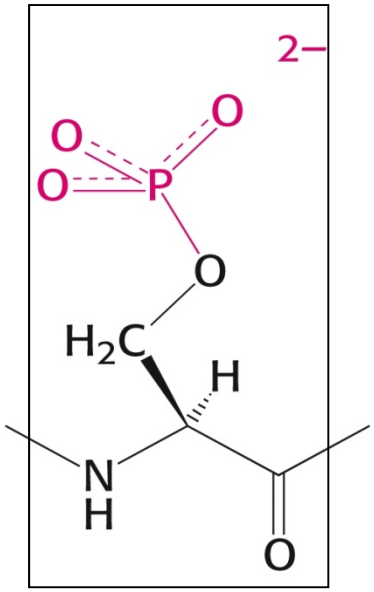
- aggregation
- erroneous S-S bonds

The living cell has solution:

- **Chaperones** (proteins that assist protein folding)
- **Disulphide isomerases** (break up the wrong S-S bonds to allow correct recombination)



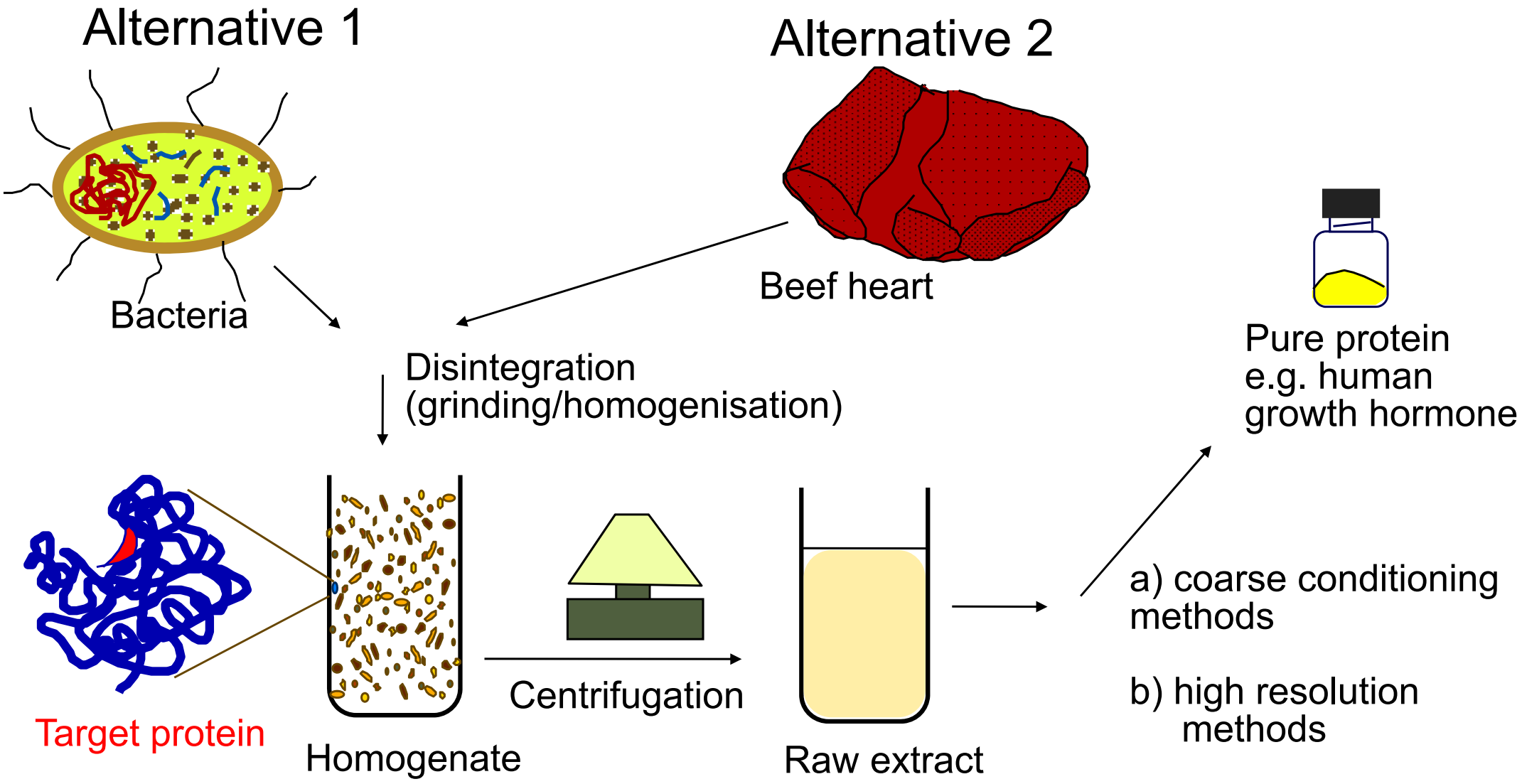
Completed proteins can get new properties by modification



Phosphoserine

- Proteins are built by 20 amino acids in the ribosome
- Some amino acids can be modified later e.g.:
  - **Ser** and **Thr** are phosphorylated (often as a way to control protein activity)
  - **Acetylation** of the amino terminal (stabilised against degradation)
  - **Carbohydrates** can be bound to till **Asn** (a more hydrophilic protein)
  - **Fatty acids** can be bound to e.g. **Cys** (a more hydrophobic protein)
  - **Pro** in collagen can become hydroxylated (the structure is stabilised)
- Sometimes a part of the protein is **removed** (the protein/the enzyme is activated)

Protein purification – Overview



How can we separate different proteins?

Property	**Method**
Size	Dialysis, Gel filtration
Density	Centrifugation
Charge	Ion exchange chromatography, Electrophoresis
Hydrophobicity	Hydrophobic chromatography
Specific affinity	Affinity chromatography
Solubility	Salt precipitation

Important to check the result!

Cells
↓
Homogenate
↓
Raw extract →
↓
Salt precipitation
↓
More….
↓
Pure protein?

- Take samples
- Run SDS-PAGE
- Measure protein content
- Measure enzyme activity

→

The process should lead to:

- Fewer **number of** protein band
- **Higher specific activity** (Units/mg protein)

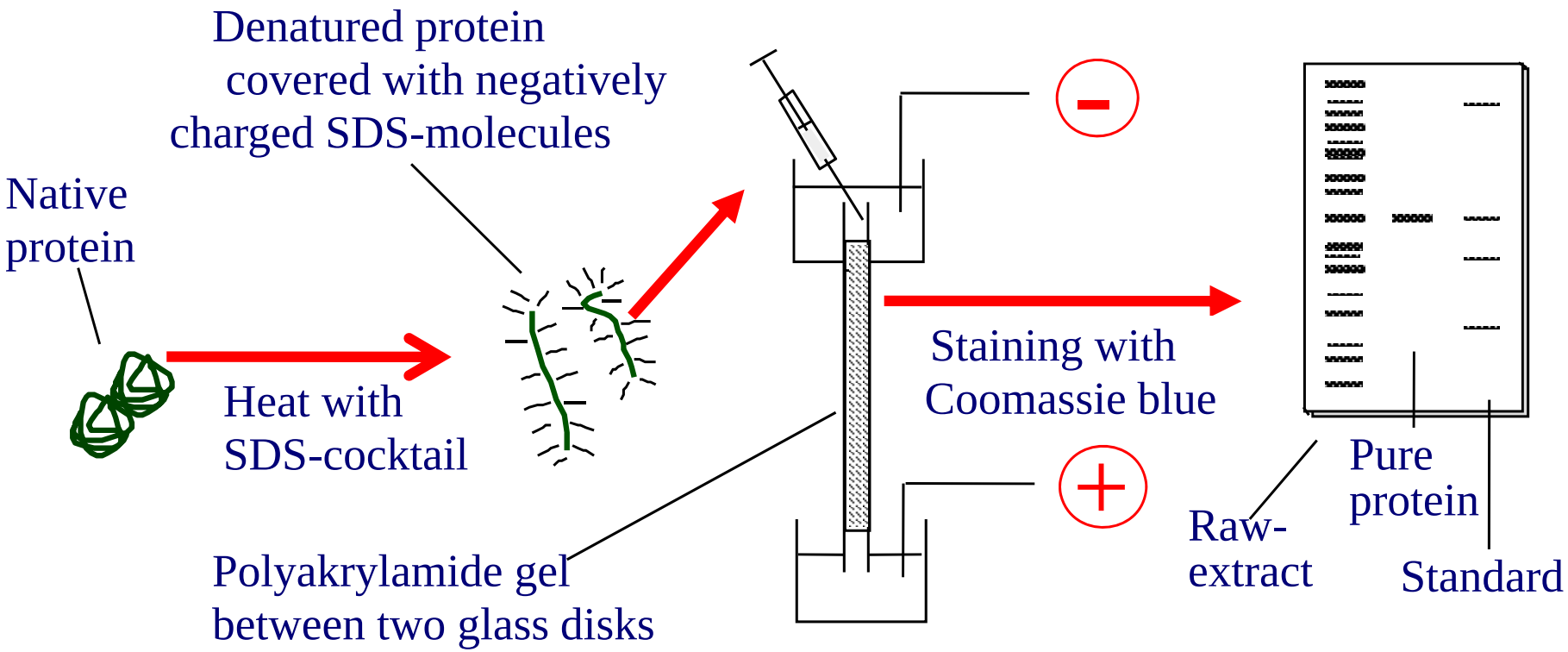
Quantification of a purification protocol for a fictitious protein

Step	Total protein (mg)	Total activity (units)	Specific activity, (units mg^{-1})	Yield (%)	Purification level
Homogenization	15,000	150,000	10	100	1
Salt fractionation	4,600	138,000	30	92	3
Ion-exchange chromatography	1,278	115,500	90	77	9
Molecular exclusion chromatography	68.8	75,000	1,100	50	110
Affinity chromatography	1.75	52,500	30,000	35	3,000

SDS PAGE

(Sodium Dodecyl Sulphate PolyAcrylamide Gel Electrophoresis)

A very common **analytic** separation method



Protein characterization – Overview

Molecular size

- Gel filtration
- SDS PAGE
- Mass spectrometry

Amino acid sequence

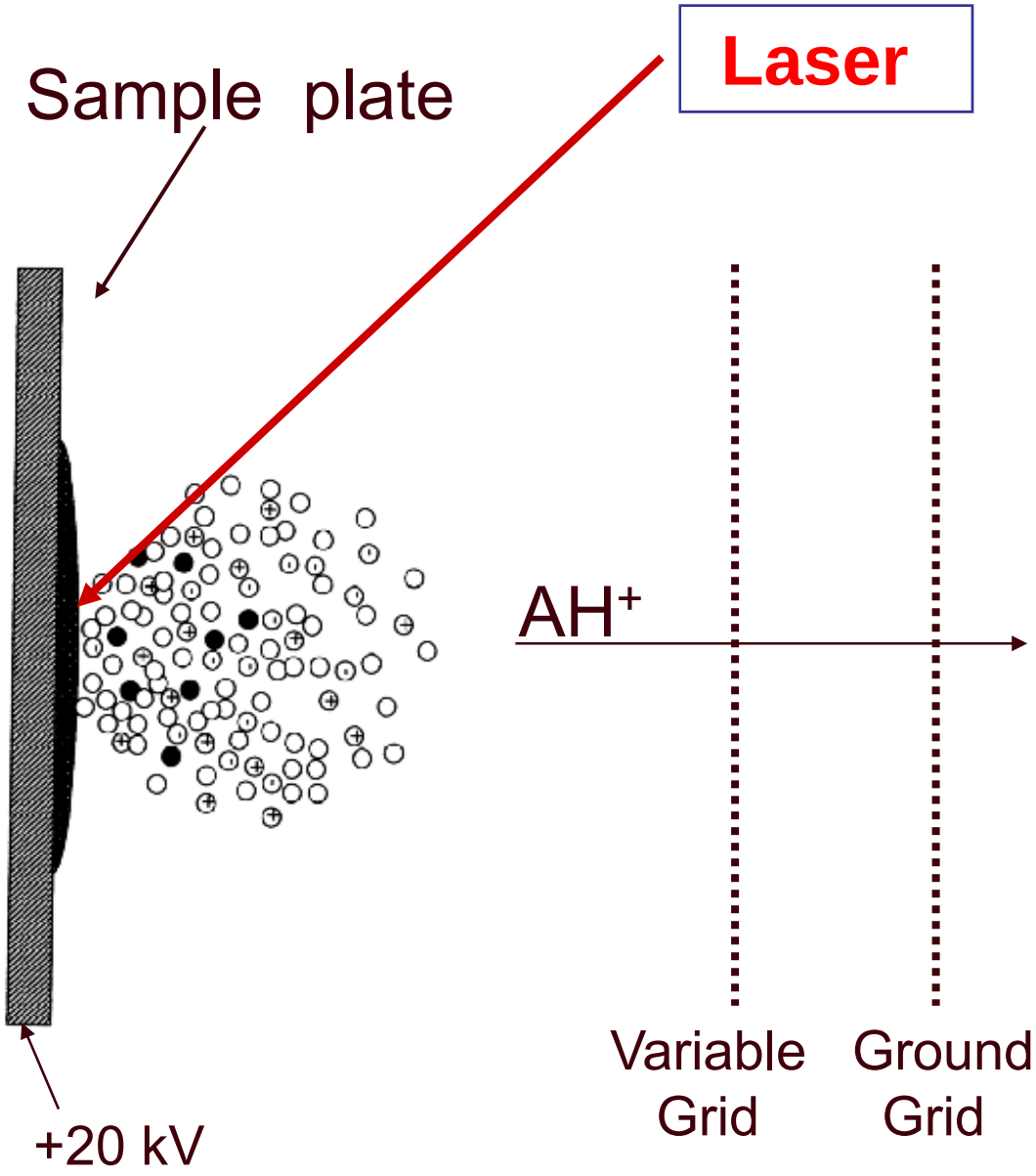
- Edman degradation

3-D structure

- NMR
- X-ray crystallography

Enzyme activity

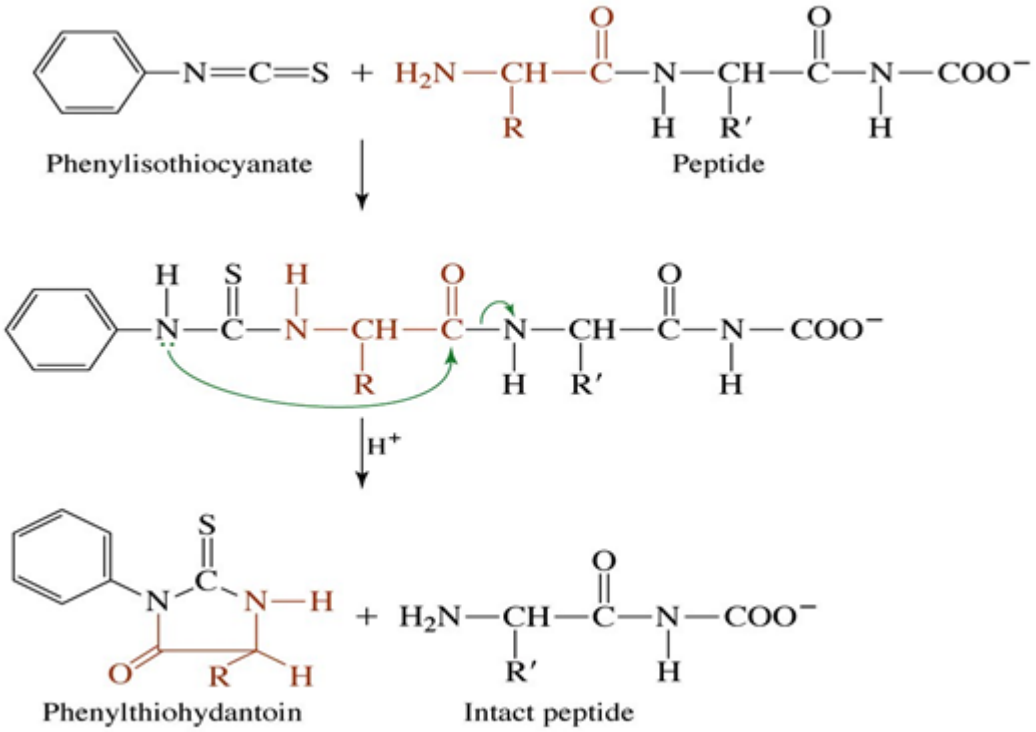
MALDI: Matrix Assisted Laser Desorption Ionization



1. Sample (A) is mixed with excess matrix (M) and dried on a MALDI plate.
2. Laser flash ionizes matrix molecules.
3. Sample molecules are ionized by proton transfer from matrix:

   $MH^+ + A \rightarrow M + AH^+$.

Edman degradation



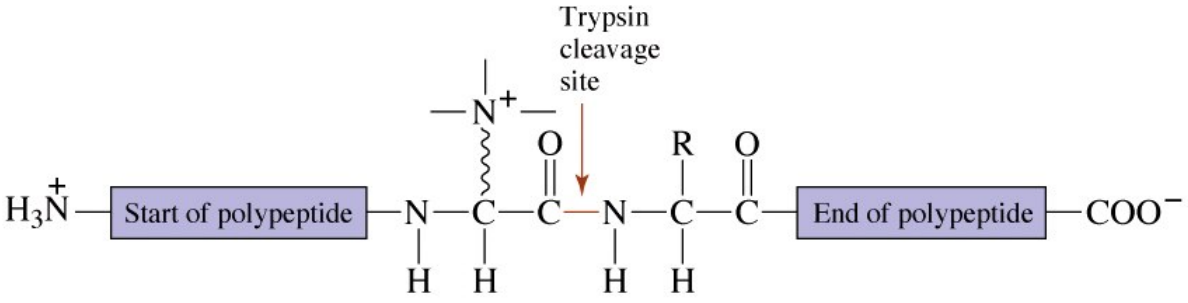
Pehr Victor Edman
(1916-1977)

ACTA CHEMICA SCANDINAVICA 4 (1950) 283–293

Method for Determination of the Amino Acid Sequence in Peptides

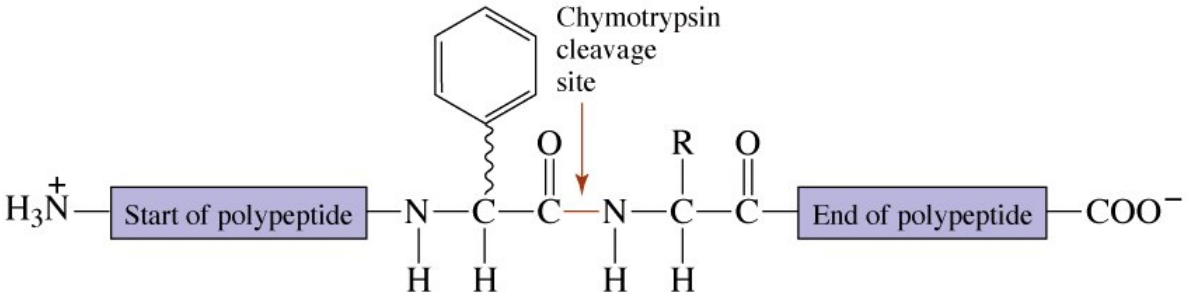
PEHR EDMAN

Department of Physiological Chemistry, University of Lund, Lund, Sweden



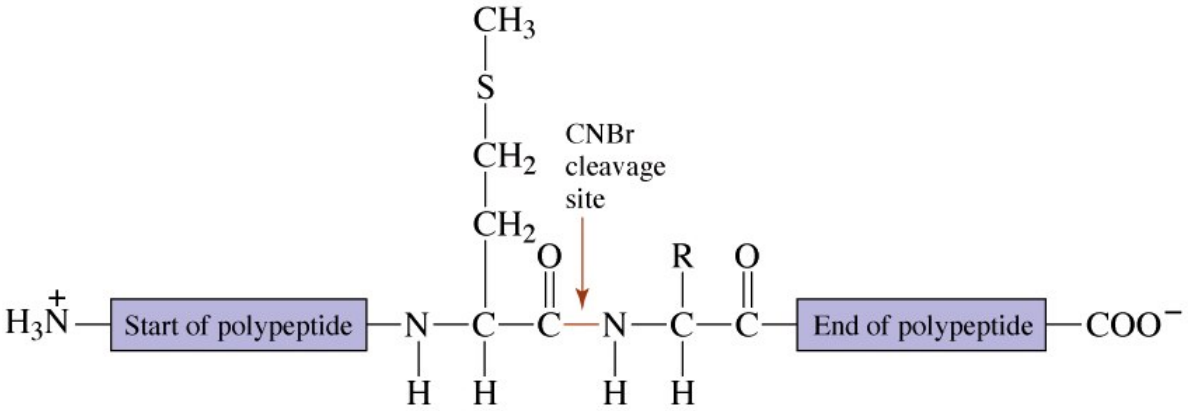
Arginine (Arg) or lysine (Lys)

(a)



Phenylalanine (Phe), tyrosine (Tyr), or tryptophan (Trp)

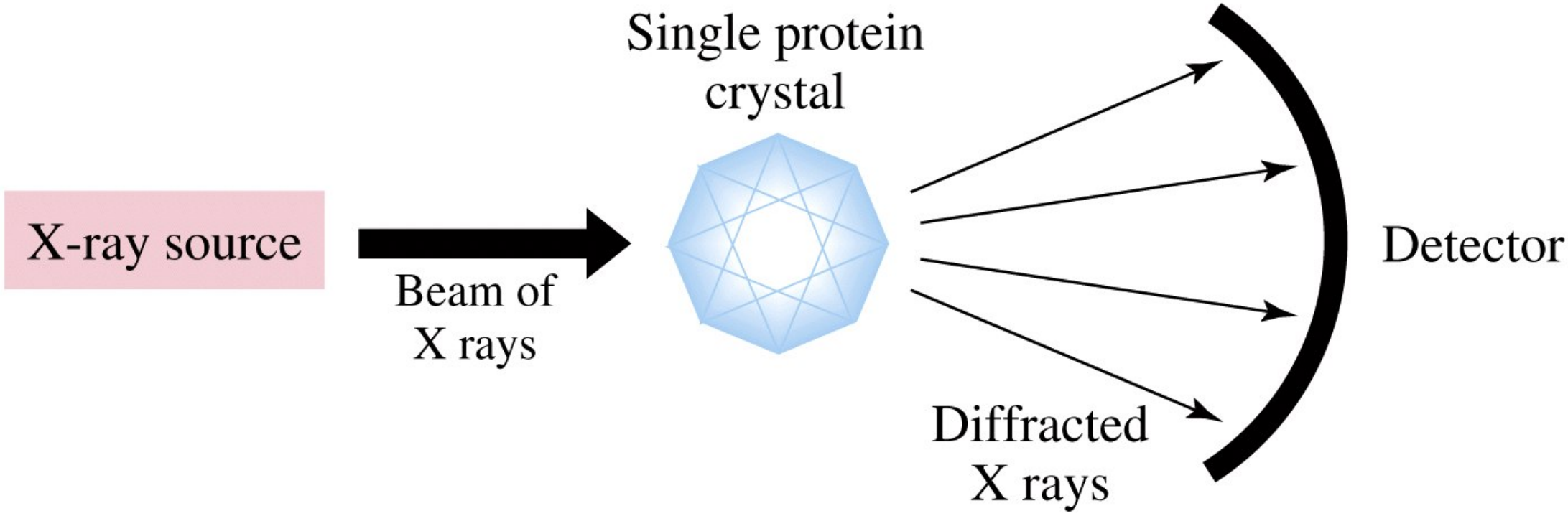
(b)

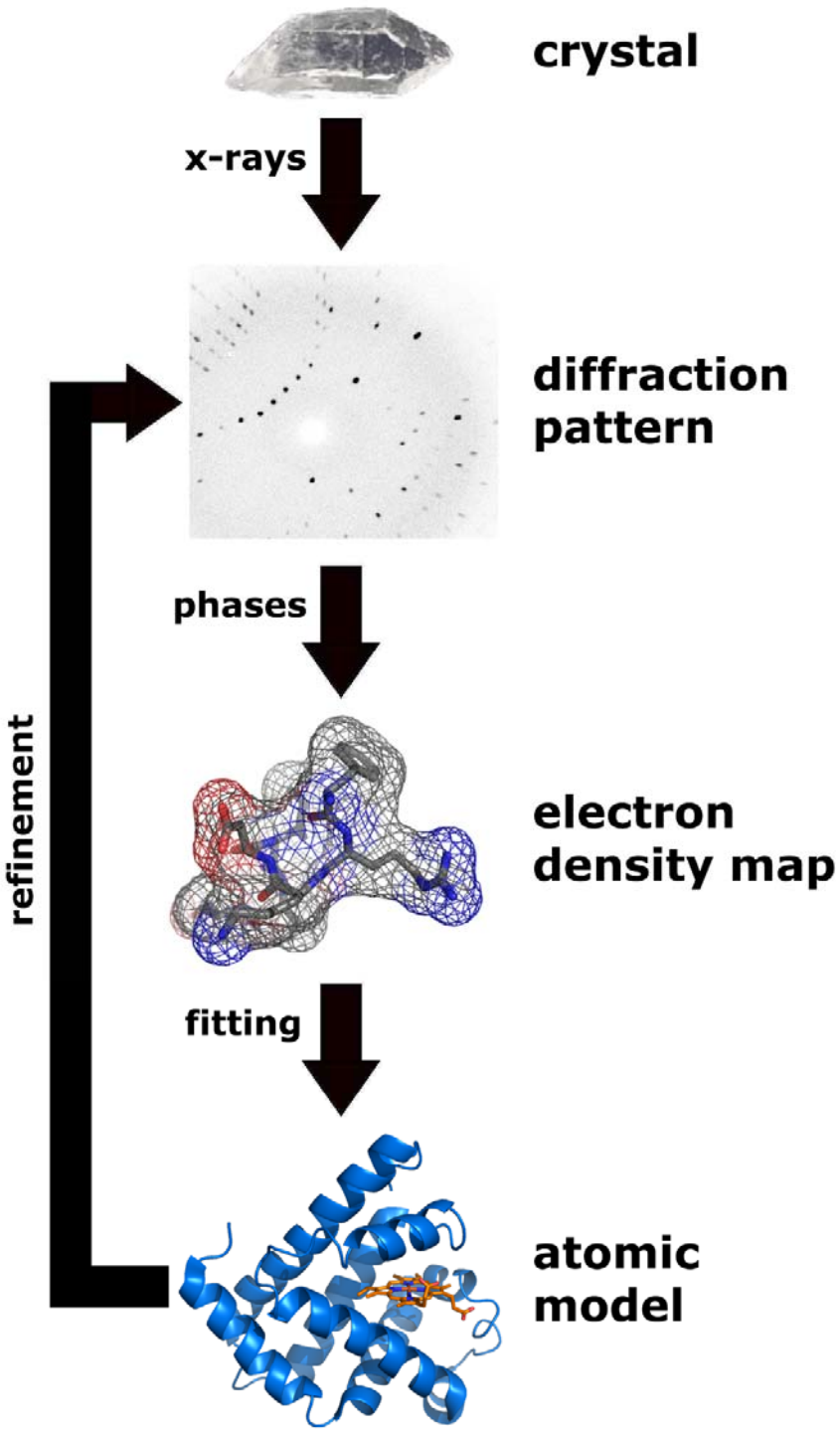


Methionine (Met)

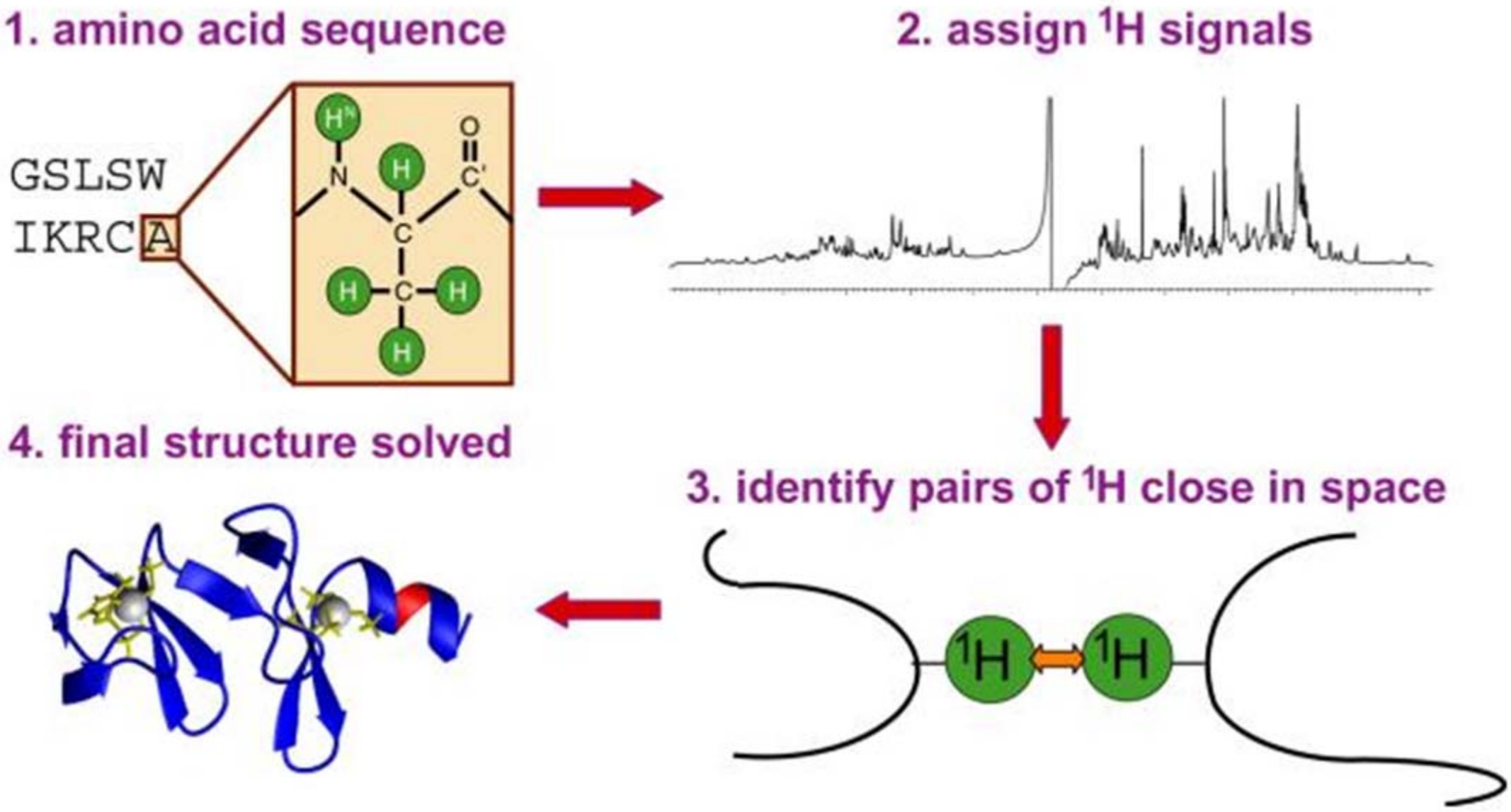
(c)

Structure determination: X-ray crystallography



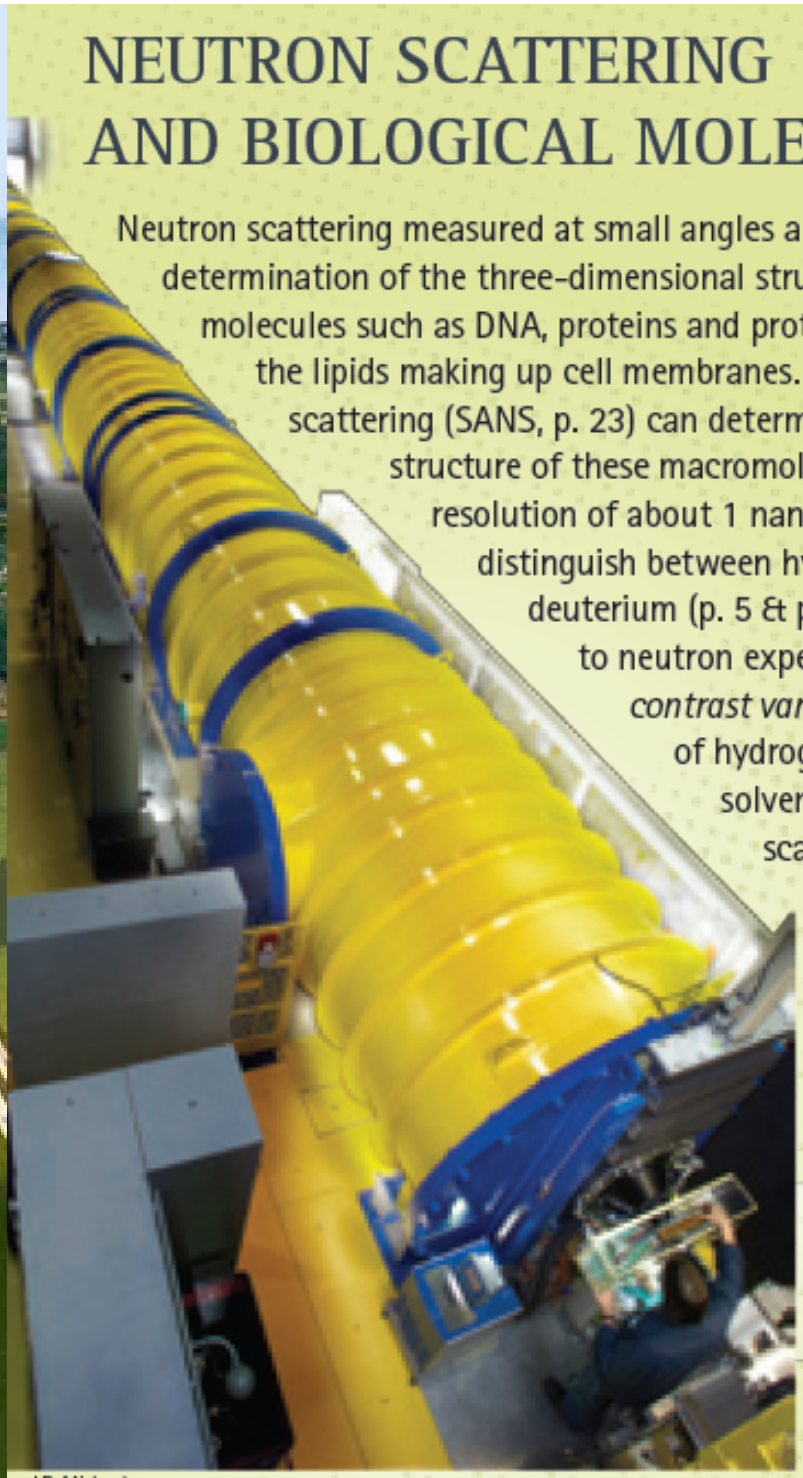

crystal
x-rays
diffraction pattern
phases
electron density map
fitting
atomic model
refinement

Structure determination: NMR



NEUTRON SCATTERING AND BIOLOGICAL MOLECULES

Neutron scattering measured at small angles allows the determination of the three-dimensional structure of large biological molecules such as DNA, proteins and protein complexes, and the lipids making up cell membranes. Small-angle neutron scattering (SANS, p. 23) can determine the solution structure of these macromolecules down to a resolution of about 1 nanometre. Neutrons can distinguish between hydrogen and its isotope deuterium (p. 5 & p. 15). A feature unique to neutron experiments is the use of *contrast variation* in which a mixture of hydrogenated and deuterated solvent is prepared so that its scattering strength matches that of selected components in a structure. This renders them 'invisible' so that other components are then highlighted.



The D11 SANS instrument at the Institut Laue-Langevin

gue/ D. Michon)

Summary

- Proteins are polymers constructed from 20 α–amino acids.
- The 20 amino acids differ in their side chains: polarity and charge
- Peptides and proteins are formed by linking amino acids together through peptide bonds.
- Protein structure: primary, secondary, tertiary and quaternary
- Protein sequence analysis: Edman degradation
- Protein purification: affinity chromatography