Peptides & Proteins

(thanks to Hans Börner)

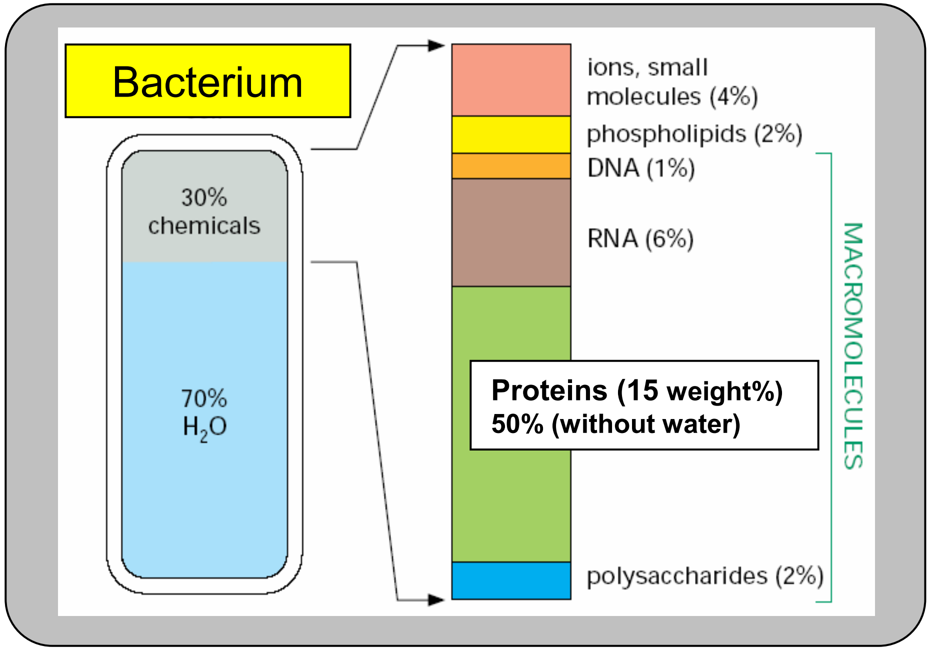
Proteins & Peptides

<u>Proteuos</u>: Proteus (Gr. mythological figure who could <u>change form</u>)

proteuo: **„"first, → ref. the basic constituents of all living cells"**

peptos: **„Cooked → referring to digestion"**

Proteins essential for: Structure, metabolism & cell functions



Bacterium

30% chemicals

70% H_2O

ions, small molecules (4%)

phospholipids (2%)

DNA (1%)

RNA (6%)

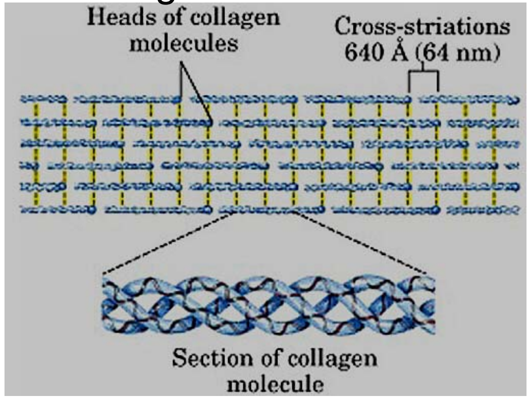
Proteins (15 weight%)
50% (without water)

polysaccharides (2%)

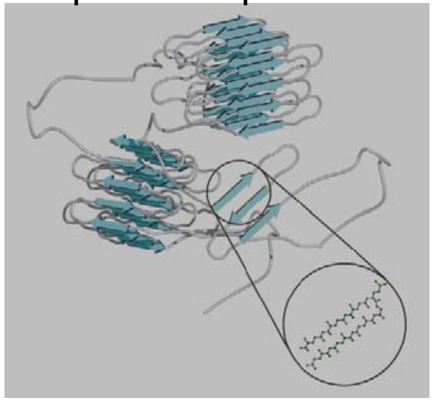
MACROMOLECULES

- Construction materials

Collagen



Spider silk proteins



Proteins

Structural proteins

structural role & mechanical support

"cell skeleton“: complex network of protein filaments.

muscle contraction results from action of large protein assemblies

Myosin und Myogen.

Other organic material (hair and bone) are also based on proteins.

Collagen is found in all multi cellular animals, occurring in almost every tissue.

- It is the most abundant vertebrate protein
- approximately a quarter of mammalian protein is collagen

Proteins

Storage

Various ions, small molecules and other metabolites are stored by complexation with proteins

- hemoglobin stores oxygen (free O_2 in the blood would be toxic)
- iron is stored by ferritin

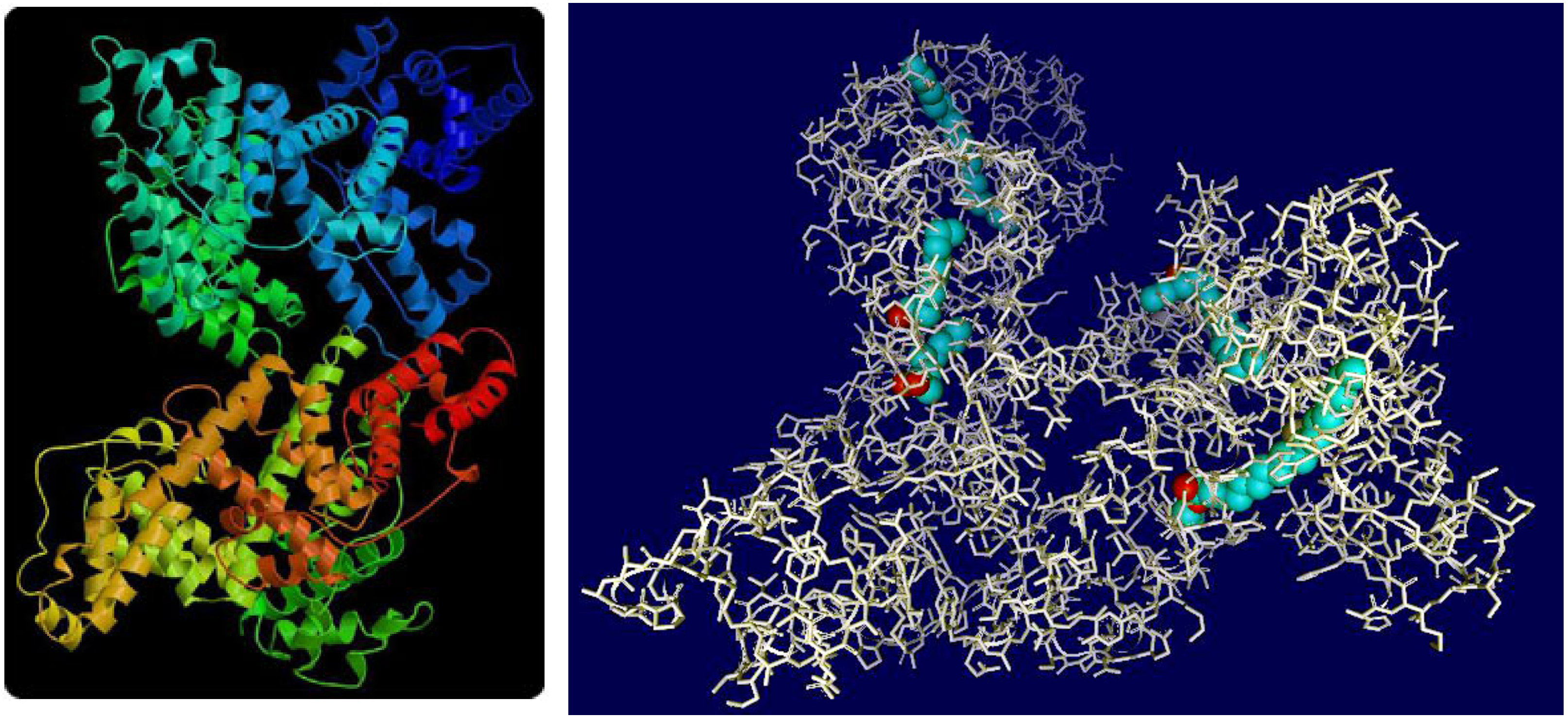
Transport

Proteins are involved in the transportation of particles ranging from electrons to macromolecules.

- Iron is transported by transferrin
- Oxygen via hemoglobin.
- Some proteins form pores in cellular membranes through which ions pass; the transport of proteins themselves across membranes also depends on other proteins.

Beside: Regulation, enzymes, defense, functional properties

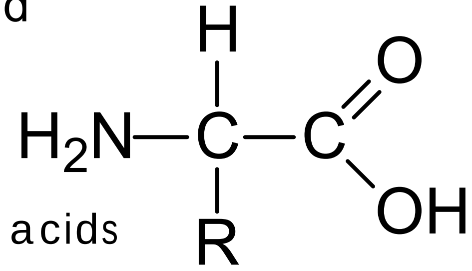
Our universal container system: Albumines



6 zones: IA, IB, IIA, IIB, IIIA, IIIB; IIB and IIIA hydrophobic

structure

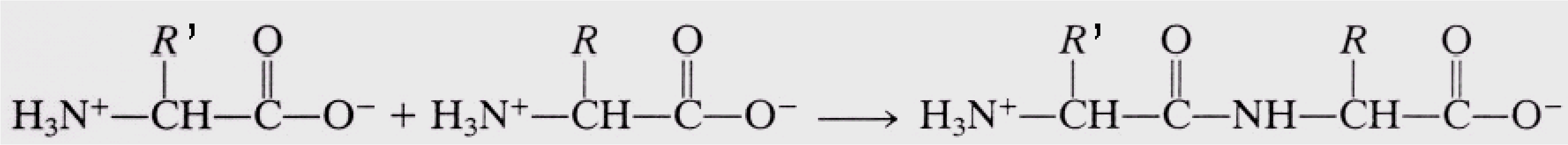
Amino acid



$$H_2N{-}\overset{H}{\underset{R}{C}}{-}C(=O)OH$$

α-amino acids

Dipeptide (2 Aa → condensation)



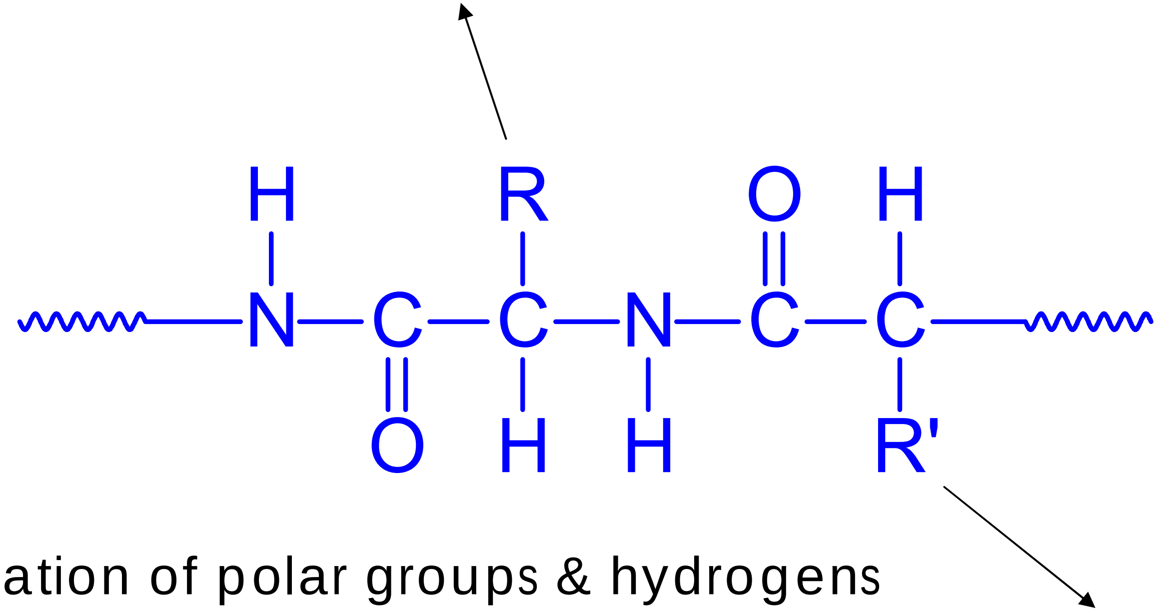
$$H_3N^+{-}\overset{R'}{CH}{-}\overset{O}{\overset{\|}{C}}{-}O^- + H_3N^+{-}\overset{R}{CH}{-}\overset{O}{\overset{\|}{C}}{-}O^- \longrightarrow H_3N^+{-}\overset{R'}{CH}{-}\overset{O}{\overset{\|}{C}}{-}NH{-}\overset{R}{CH}{-}\overset{O}{\overset{\|}{C}}{-}O^-$$

Tripeptide

Peptide: less than 100 amino acid monomers

Proteins: more than 100 amino acid monomers

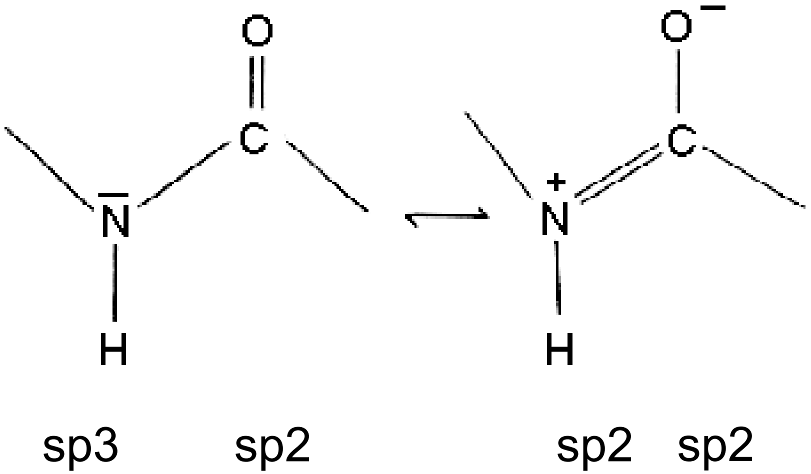
the peptide bond:



side chains

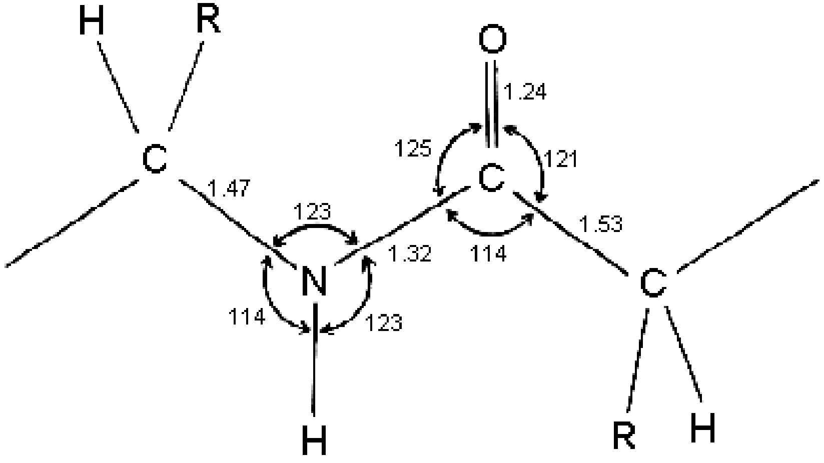
combination of polar groups & hydrogens

side chains



→ as the peptide bond is a hybrid of these resonance forms, these six atoms must all lie in the same plane

Peptide structure (*Linus Pauling* and *Robert Corey* analysis of geometry and dimensions in the crystal structures)



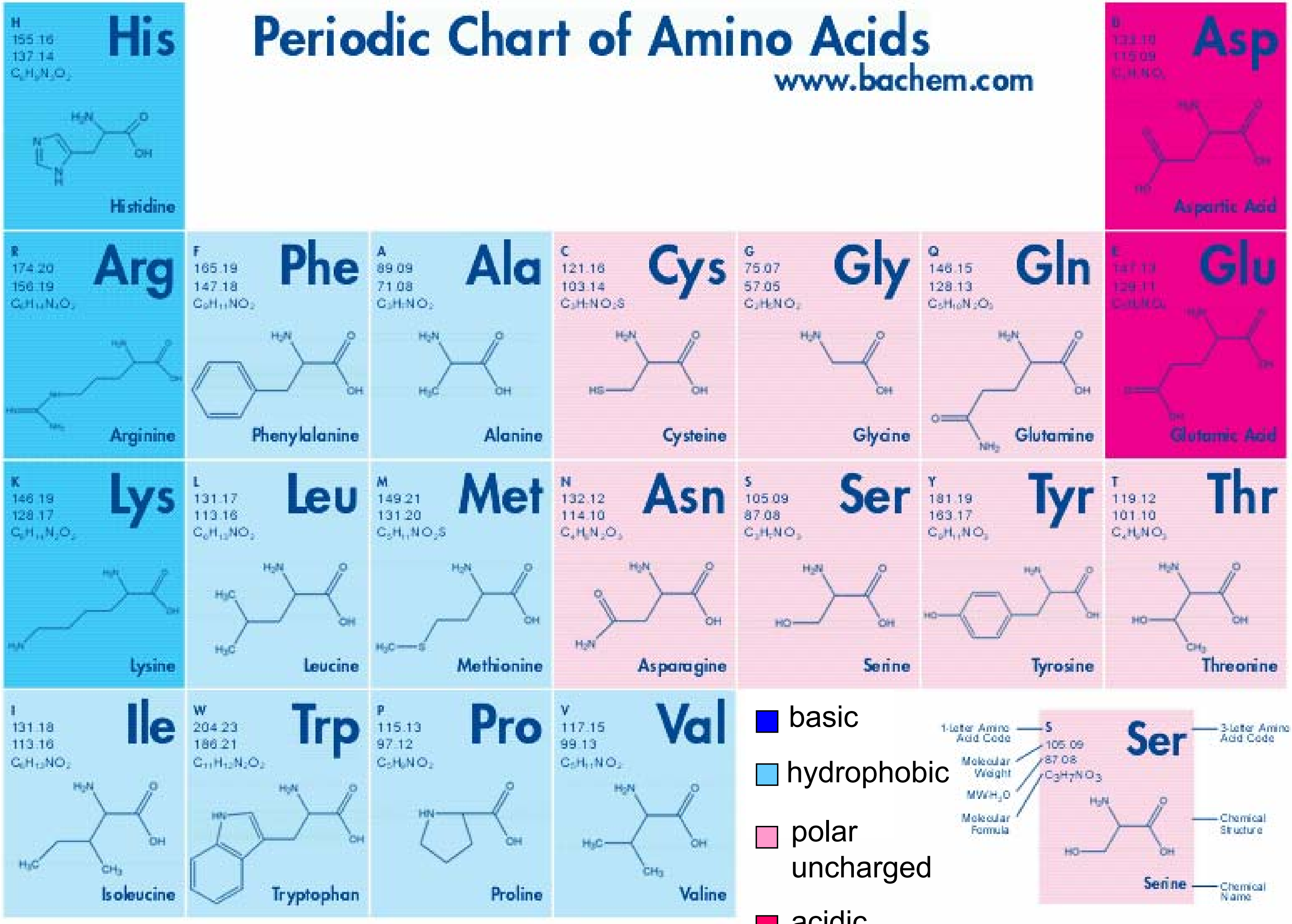
summarised results
(bond lengths in Angstrom units;
bond angles in degrees)

characteristic bond lengths:

•N-CO-bond is shortened: 127 pm (N=C) < **132 pm** < 149 pm (N-C)
(C-N bond in peptide is shorter than in usual C-N amines
partially double bond character of the N-CO-bond! (~40 %)

- 6 atoms of the peptide bond are in similar plan:
  - 2 $C_{a(i), (i+1)}$; 1 >C=O ; 1 N-H **(N- is sp^2-hybrid)**

•rotation is hindered
  - **(rotation around C_a-CO und C_a-NH is possible)**



🟦 basic

🟦 hydrophobic

🟪 polar uncharged

🟥 acidic

Proteins

Proteins are monodisperse:

all molecules of a particular protein possess the same composition, structure, and properties (difference to synthetic polymers!!).

- micro heterogeneity in glycoproteins.

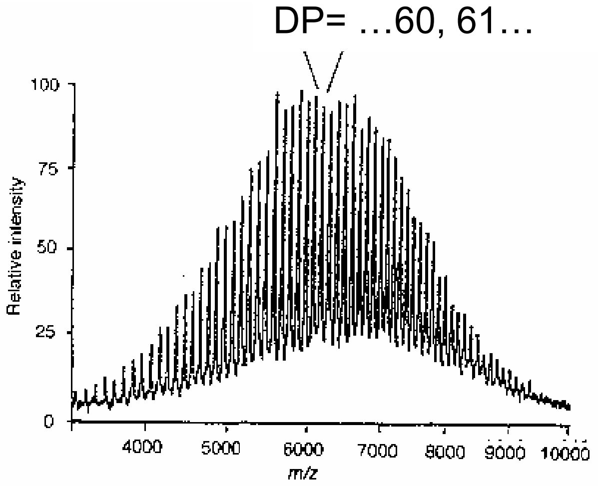

DP= …60, 61…

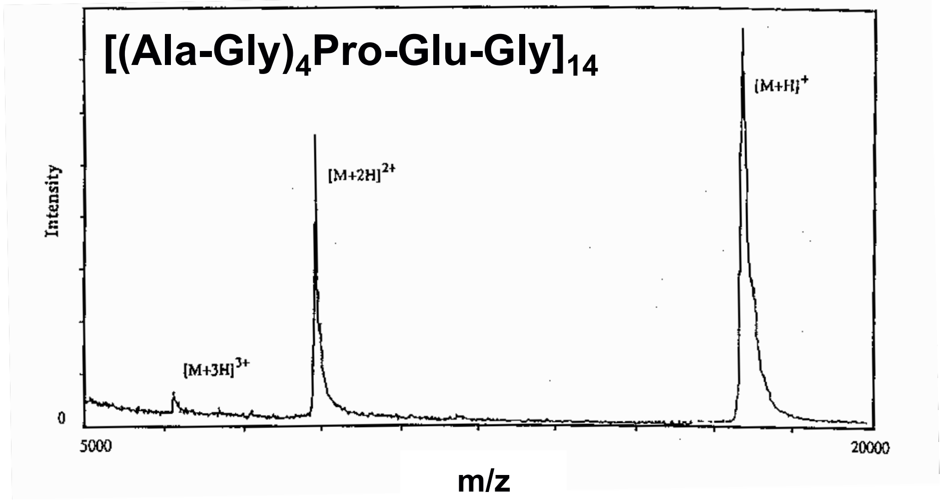
Figure 17-2. Matrix-assisted laser desorption mass spectrum of an approximately 6500 molecular weight poly(methyl methacrylate) with a polydispersity index of 1.03. [Reproduced with permission from Bahr et al. (1992).]

Polydisperse System

PMMA **M_n = 6500**

M_w/M_n = 1.03

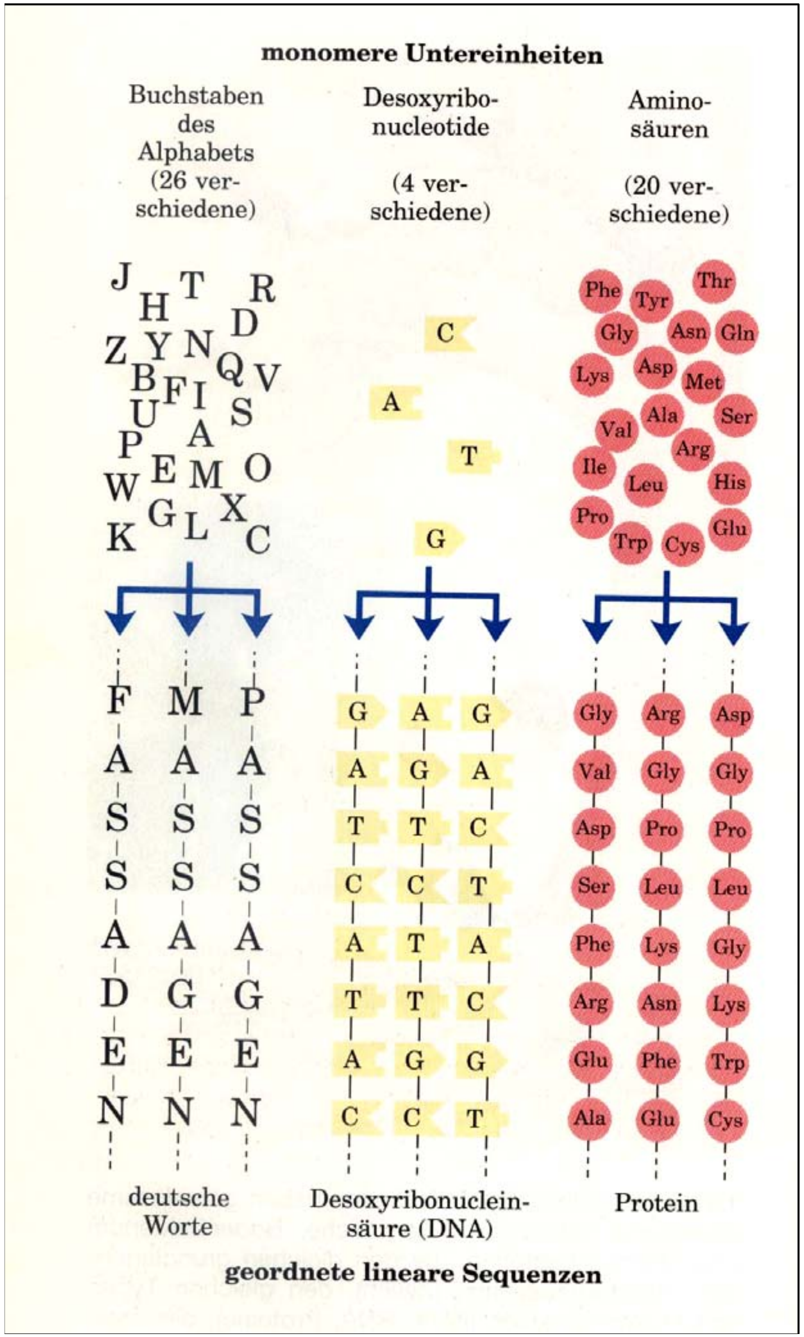
Polypeptid:


$[(\text{Ala-Gly})_4\text{Pro-Glu-Gly}]_{14}$

Homodisperse System:

- **identical macromolecules**
- **defined MW, stereo chemistry, composition und monomer sequence**

➔ defined property profile



- For a unit consisting of 8 subunits:

  Alphabet: $26^8 = 2.1*10^{11}$

  DNA: $4^8 = 65536$

  AA: $20^8 = 2.6*10^{10}$

Proteins can consist of 1000 subunits, DNA of several millions

Almost infinite number of combinations

➔ BUT: solubility !!!

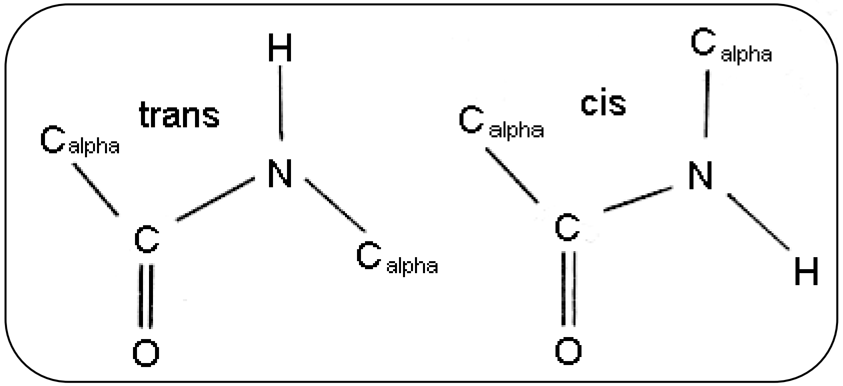
enormous structural variety
polypeptides for many different tasks:
e-coli cell: 3.000 different proteins
human body: more than 150.000

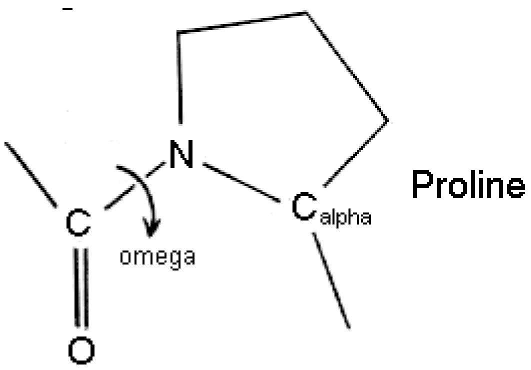
Folding towards superstructures: why do peptides fold at all

Rigidity of the peptide bond
→ reduction of degrees of freedom of the polypeptide during folding.

<u>Peptide bond nearly always has the *trans* configuration</u>

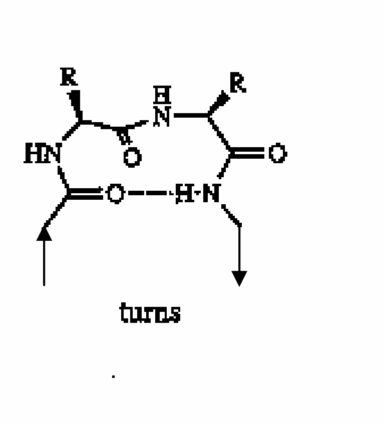
more favourable than *cis (in average ≈ 0,05% cis)*
Exception *Aa-Pro* ≈ 6,5% cis *in average*

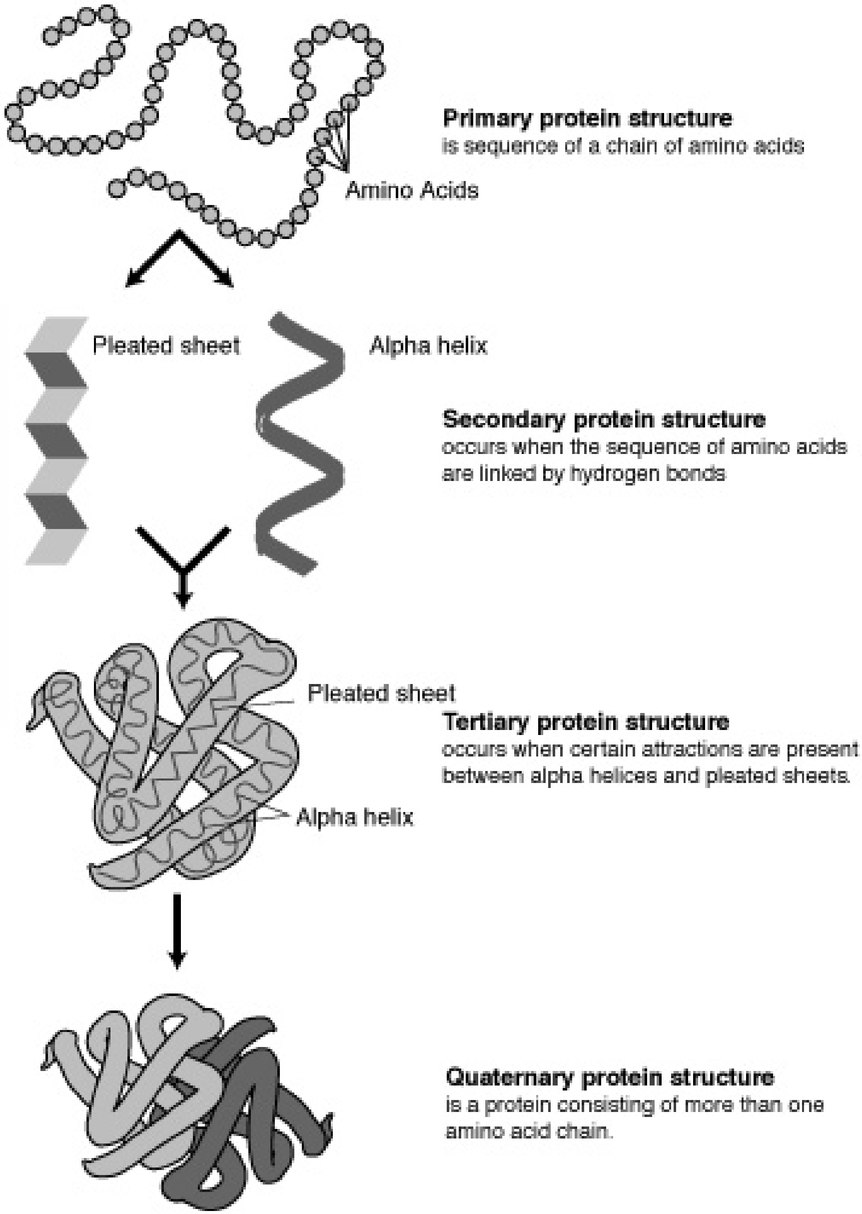




***cis* configuration**:
<u>higher sterical hindrance</u> between the functional groups attached to the C_{alpha} atoms

peptide folding





Courtesy of NHGRI (http://www.nhgri.nih.gov/DIR/VIP/Glossary/)

Amino acids: importance of the different letters

Groups of amino acids:

- aromatic, aliphatic

- polarity and hydrophobicity are depending on the side chain:

→ hydrophobicity index negative: non-polar AA's

→ hydrophobicity index positive: polar AA's

no charge contribution: Gly = 0

polar: Ser, Thr, Asp, Asn, Glu, Gln, His, Lys, Arg

non-polar: (Gly), Ala, Val, Leu, Ile, Pro, Phe, Trp, Met, Cys

not all organisms can synthesize all amino acids:

essential for humans: Ile, Leu, Lys, Met, Phe, Thr, Try, Val

stereochemistry: usually L configuration (occurrence of D)

The letter code: are we really free?

Size (general principle concerning packing in proteins)
folded protein can be compared with a three-dimensional ***puzzle***

Packing is due to optimised van der Waals surface contacts

folding of the polypeptide(s)
→ filling up most of the space in the interior.
→ close fitted-packing
(Interiors of proteins - similar packing density to organic solids)

The letter code: are we free??

Hydrophobicity (Very important driving force!!! → ref hydrophobic effect)
The aliphatic side chains Ala, Val, Leu and Ile (and Gly) (no polar atoms)
→ interact less favorably with water than with other apolar groups.
only van der Waals-interactions possible (temporary dipoles)

➔ A general feature of globular proteins: **hydrophobic** residues are found in the protein interior, while **polar** residues occur on the surface.
(similar phenomenon: lipid micelles – exclusion of water contact)

Complex folding process due to polypeptide chain and divers driving forces.
- **hydrophobic** side chains
- **charged** and **neutral polar** side chain
- **polar** main chain atoms (peptide bond);
- **hydrogen bonding** capacity;

…..covalent side chain bonds, ion bridges…etc.

Hydrophobicity is a very important factor in protein stability; "hydrophobic effect" is believed to play a fundamental role in the spontaneous folding processes.

Amino acid properties important for proteins

Charges

Asp, Glu (one negative charge), Lys and Arg (one positive) ionized under most physiological conditions

'ion pair' - salt bridge:

A specific type of interaction
(statistically 1 ion pair per approximately 30 peptide residues)

Amino acid properties important for proteins

Polarity: Charged and neutral polar side chains participate in hydrogen bonds, both with each other, with the main chain polar atoms and with solvent.

- Ser and Thr: hydroxyl group → act as a donor in one hydrogen bond, and as an acceptor in two.

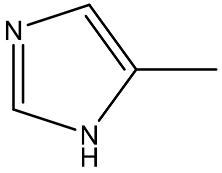
- Asp and Glu: carboxyl group; each oxygen can two hydrogen bonds and donor one hydrogen.

- Asn and Gln have a carbonyl oxygen (C=O bond), which can act as an acceptor in two hydrogen bonds, while the amide nitrogen can donate each of the two hydrogens in a hydrogen bond. nitrogen is a poor acceptor due to delocalized sp2 character.

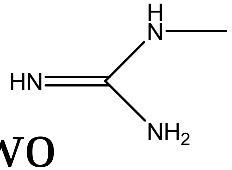
Amino acid properties important for proteins

Polarity

- His has two imidazole nitrogens, either or both of which is protonated. Each of these can act as an acceptor in a single hydrogen bond if it is unprotonated, or as a donor in a single hydrogen bond if it is protonated.

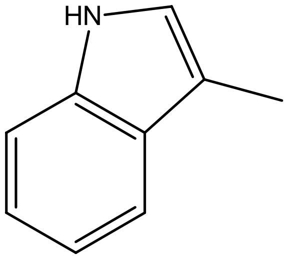


- Arg has a guanidinium group, which is usually protonated, and planar: the carbon atom is *sp2*-hybridized. Each of the two $-NH_2$ groups can donate two hydrogens, and the –NH- group one.



- Lys is usually protonated and donate three protons in hydrogen bonds

- Trp can donate a hydrogen in a single hydrogen bond. Its nitrogen atom is *sp2*-hybridized.



Amino acid properties important for proteins

Aromaticity

Delocalized electrons in aromatic side chains can participate in electrostatic interactions (relatively weak)

However, in the context of proteins, there is a tendency for aromatic side chains to be ‘grouped'

Conformationally Unusual Side Chains

Steric hindrance (Val, (Thr), Pro)

high flexibility (Gly)

play special roles in polypeptide conformation.

hierarchical structure of peptides:

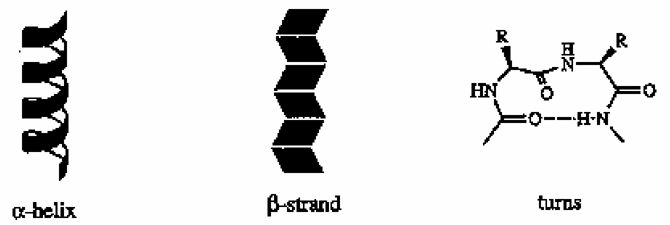
(caused by the various interactions and restrictions
due to the peculiarities of the amid bond and side chain functionalities)

primary structure: nothing more than the sequence of amino acids

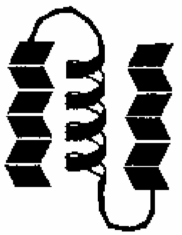
secondary structure: interactions along the polymeric chain
(hydrogen bonding of the peptide bond)
Pauling, Corey; 1951

tertiary structure: secondary structure elements are observed to
combine in specific geometric arrangements
known as motifs or supersecondary structures
arrangement of AA in 3D-space (intermolecular)

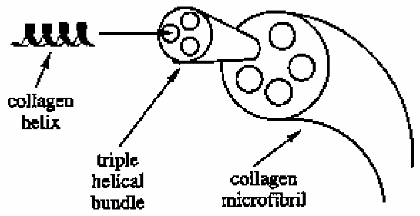
quaternary structure: several protein chains are linked by primarily interactions
between hydrophobic substitutents in the chains



secondary structures



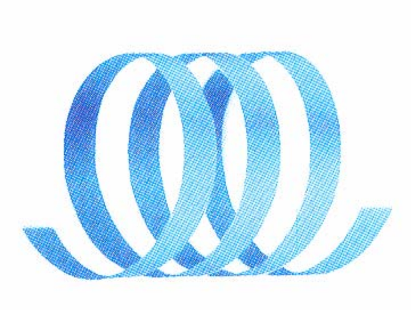
β-strand-α-helix-β-strand
tertiary structure

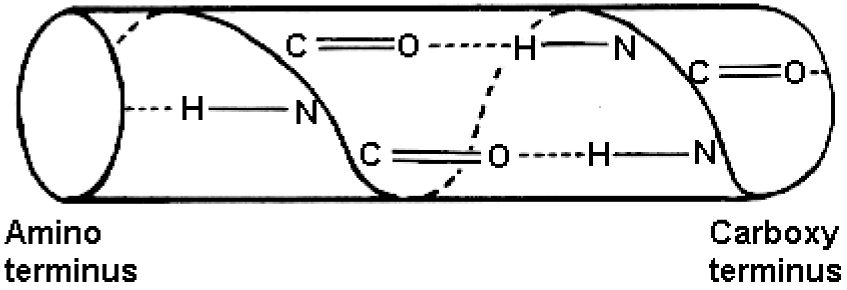


self-assembly of collagen fibrils
quaternary structure

Properties of the alpha-helix.

Helical conformation: Pauling, Corey; 1951*

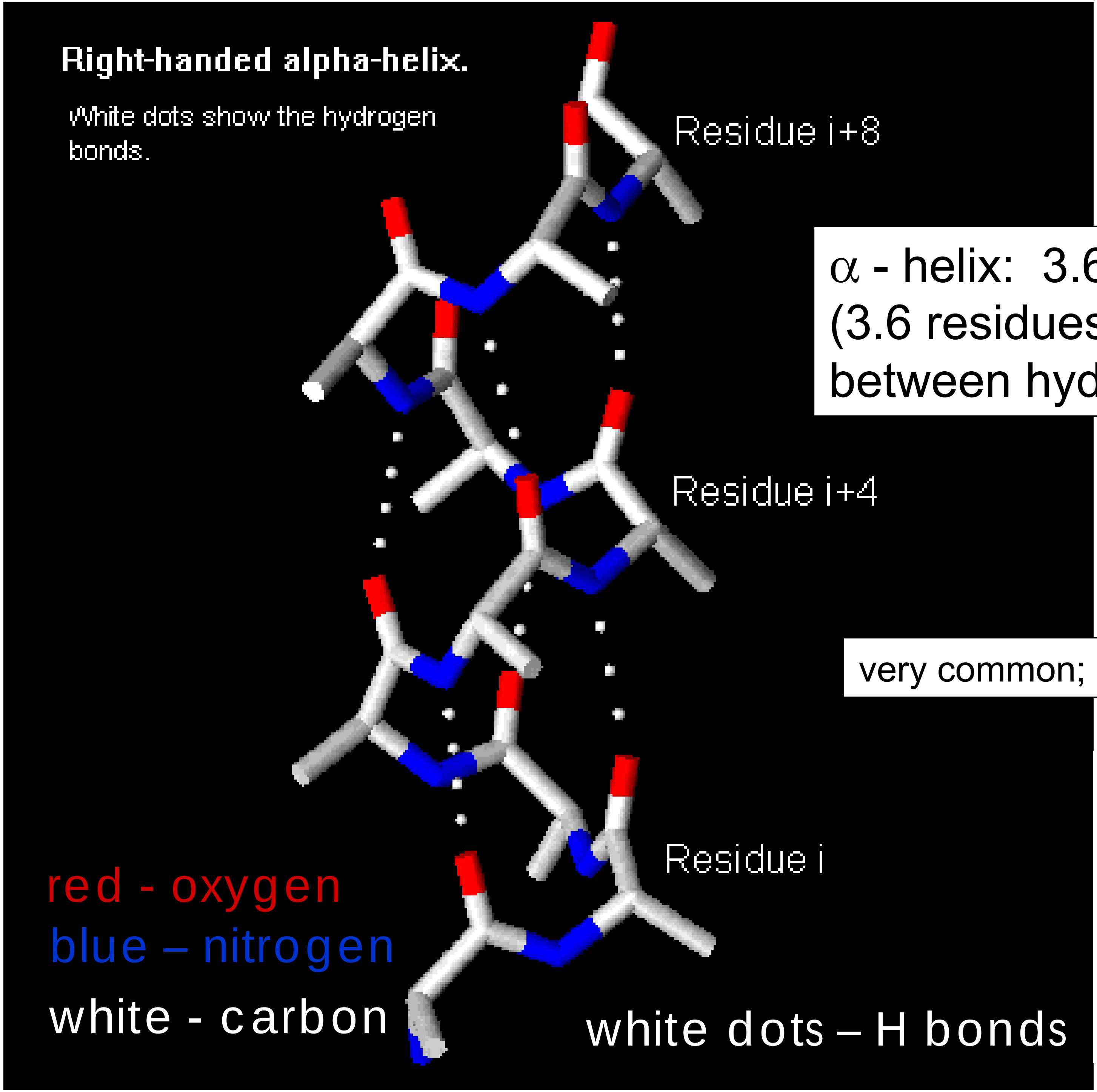




1. # The **structure repeats** itself every **5.4 Angstroms** along the helix axis (the alpha-helix has a pitch of 5.4 Angstroms)
   # **3.6 amino acid residues per turn** (36 amino acids would form 10 turns)
   # **alpha-helix has a rise 1.5 Angstroms per residue** (5.4/3.6)

2. Every main chain C=O and N-H group is hydrogen-bonded to a peptide bond 4 residues away (ie O(i) to N(i+4))
   This gives a very regular, stable arrangement.

3. **Peptide planes are roughly parallel with the helix axis**
   → **dipoles** within the helix **are aligned**, ie all C→O groups point in the same direction and all N←H groups point the other way.
   Side chains point outward from helix axis
   (generally oriented towards its amino-terminal end)



red - oxygen

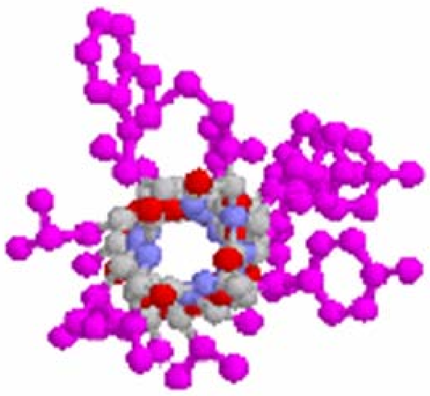
blue – nitrogen

white - carbon

white dots – H bonds

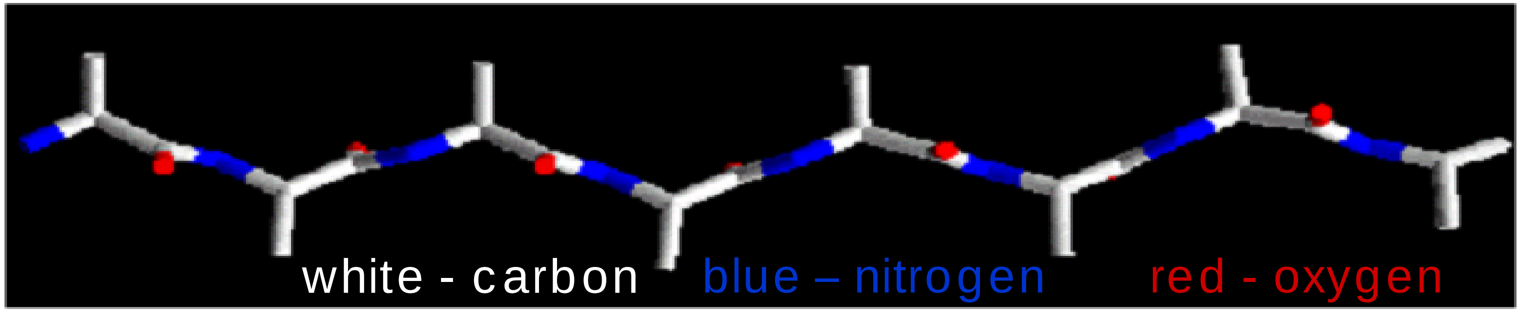
α - helix: 3.6_{13} helix
(3.6 residues/turn, 13 atoms between hydrogen bonds)

very common; ϕ = -57°; ψ = -47°



The Beta-Sheet. beta-conformation

Pauling / Corey derived a model for the conformation of fibrous proteins (beta-keratins)

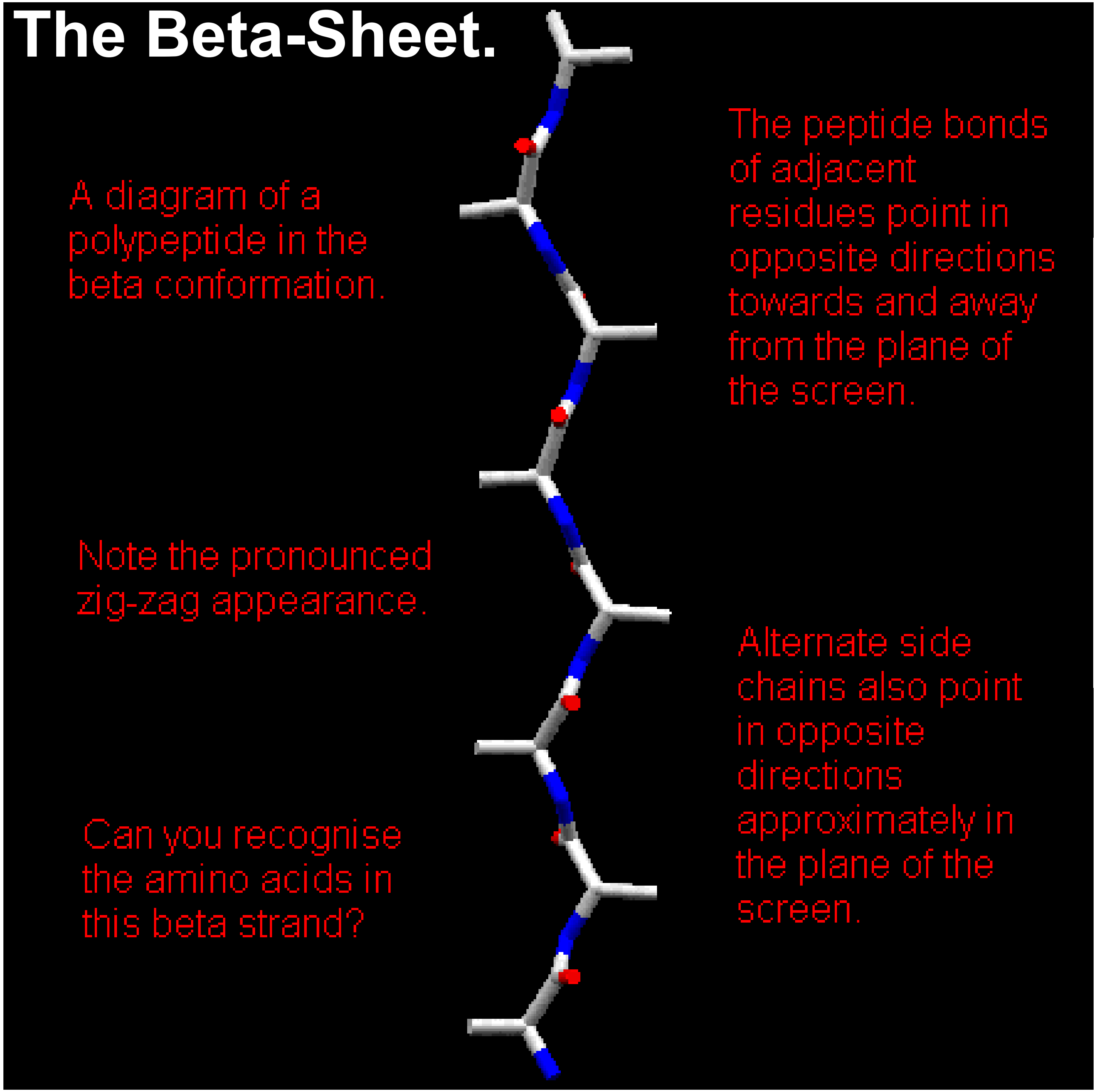


- polypeptide conformation forms more extended zigzags
- **negative phi** angles and **positive psi** angles are in the have.
- Typical values are phi = -140° and psi = 130°

section of polypeptide with residues in the beta-conformation →**is a beta-strand**
strands can associate by main chain hydrogen bonding interactions → **beta sheet**

beta-sheet: two or more polypeptide chains run alongside each other and are linked in a regular manner by hydrogen bonds between the main chain C=O and N-H groups.

The R-groups (side chains) of neighbouring residues in a beta-strand
→ point in opposite directions



The Beta-Sheet.

A diagram of a polypeptide in the beta conformation.

Note the pronounced zig-zag appearance.

Can you recognise the amino acids in this beta strand?

The peptide bonds of adjacent residues point in opposite directions towards and away from the plane of the screen.

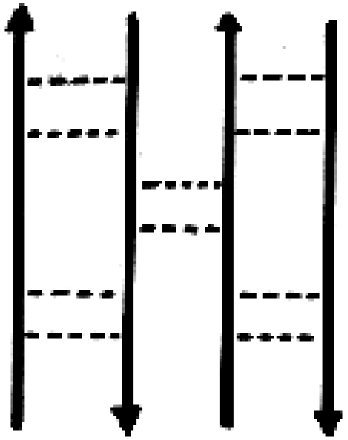
Alternate side chains also point in opposite directions approximately in the plane of the screen.

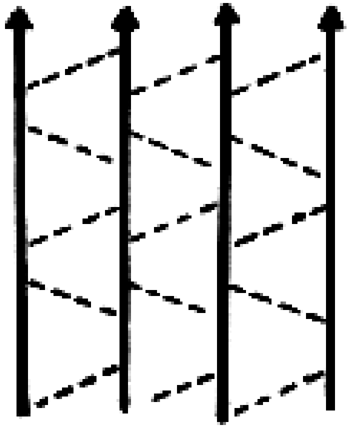
Axial distance between residues is 3.5 A
(alpha-helix: axial distance 1.5A)

Two residues per repeat unit
→ beta-strand pitch 7 A.

secondary structure (b-sheet): strands direction

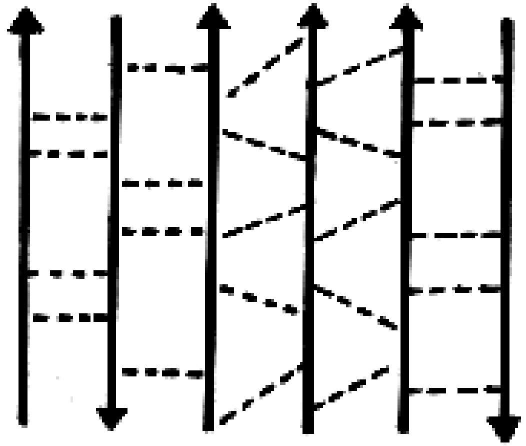
Antiparallel beta-sheet





Parallel beta-sheet

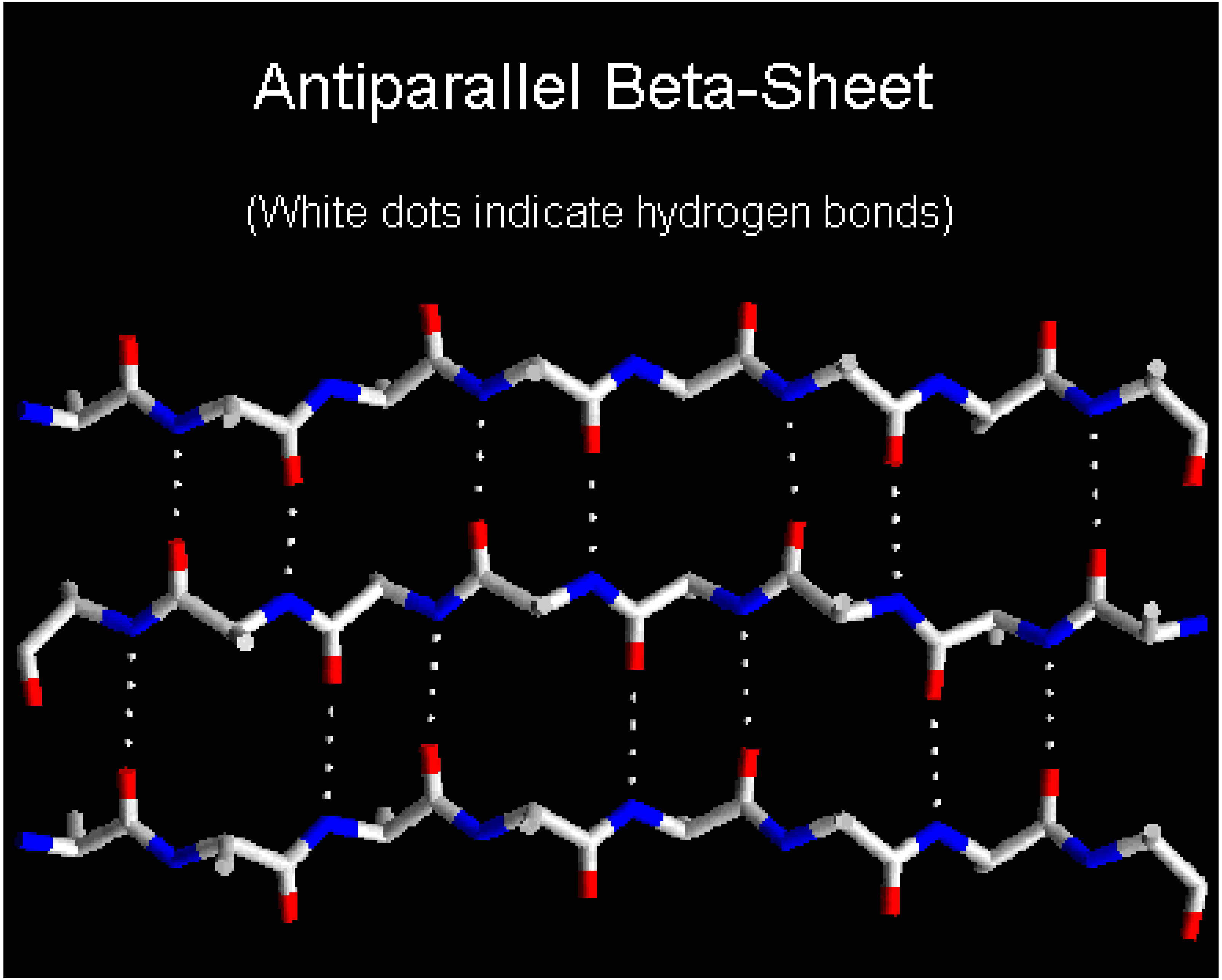
The different types of beta-sheet. Dashed lines indicate main chain hydrogen bonds.



Mixed beta-sheet

parallel: phi = -119°, psi = +113°

antiparallel: phi = -139°, psi = +135°



Antiparallel Beta-Sheet

(White dots indicate hydrogen bonds)

secondary structure: β-turns

are always made by four amino acid residues
→ necessary for sharp changes in overall protein structure
very common is glycine because of rotational freedom
residue i+3 is important for stabilization of the conformation
(Leu > Ala > Ile > Phe)

several types are known, most important:

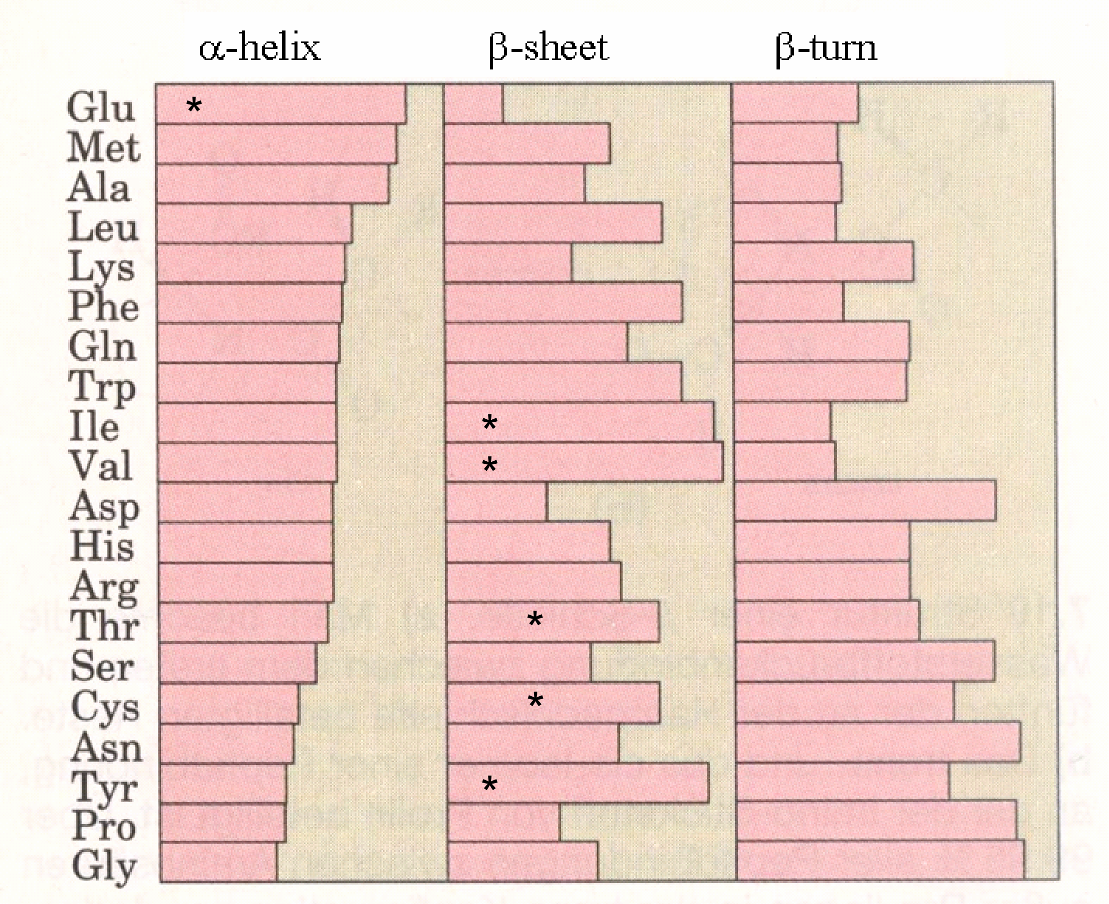
Type I: $\varphi_{i+1} = -60°$ $\psi_{i+1} = -30°$ $\varphi_{i+2} = -90°$ $\psi_{i+2} = 0°$
residue i+1 cannot be proline, all other amino acids are possible

Type II: $\varphi_{i+1} = +60°$ $\psi_{i+1} = +30°$ $\varphi_{i+2} = +90°$ $\psi_{i+2} = 0°$
residues i+1 and i+2 are glycine, also asparagine is common

Type III: $\varphi_{i+1} = -60°$ $\psi_{i+1} = +120°$ $\varphi_{i+2} = +80°$ $\psi_{i+2} = 0°$
all amino acids are possible; corresponds to 3_{10} helix

secondary structure: comparison

relative probabilities of different amino acids to occur in one of the three most common secondary structures:

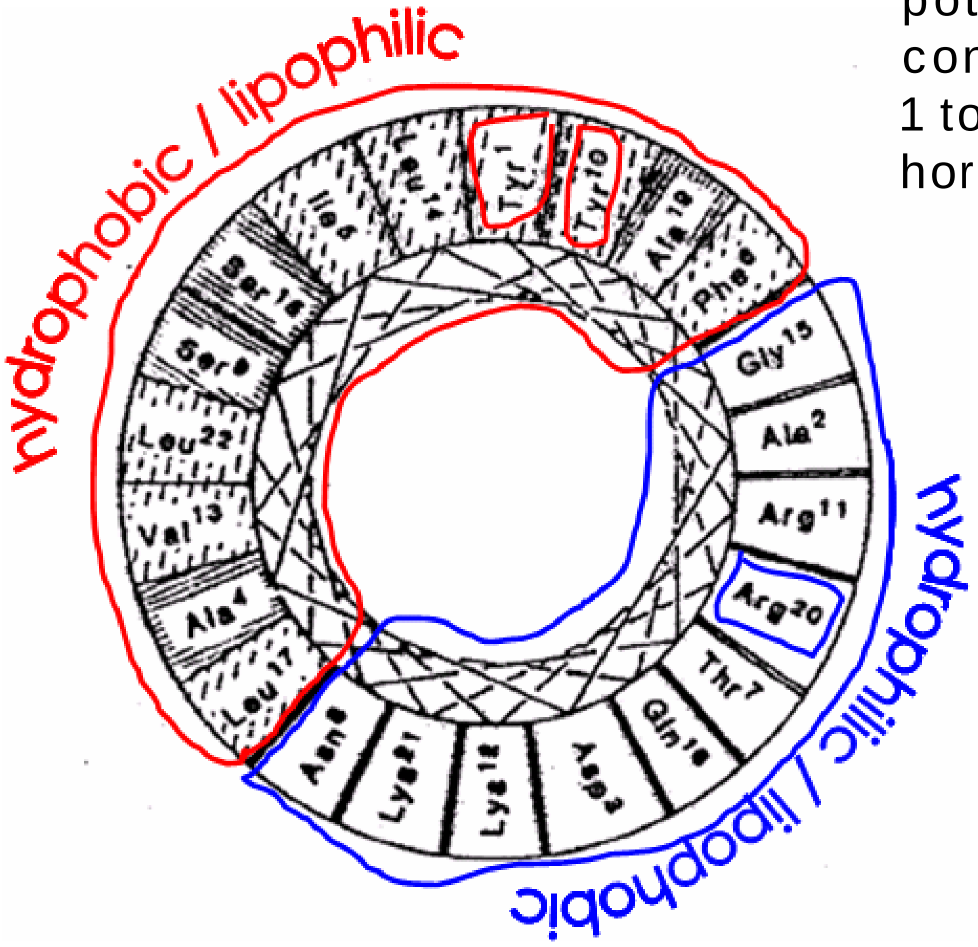


Covalent bond distances and torsion angles:

- are the major properties of the covalent bonds hold proteins together
- particularly, the bond angles between two adjacent bonds on either side of a single atom, or the dihedral angles between three contiguous bonds and two atoms control the geometry of the protein folding
- this is determined by the structure of the amino acids

secondary structure: amphiphilic structures

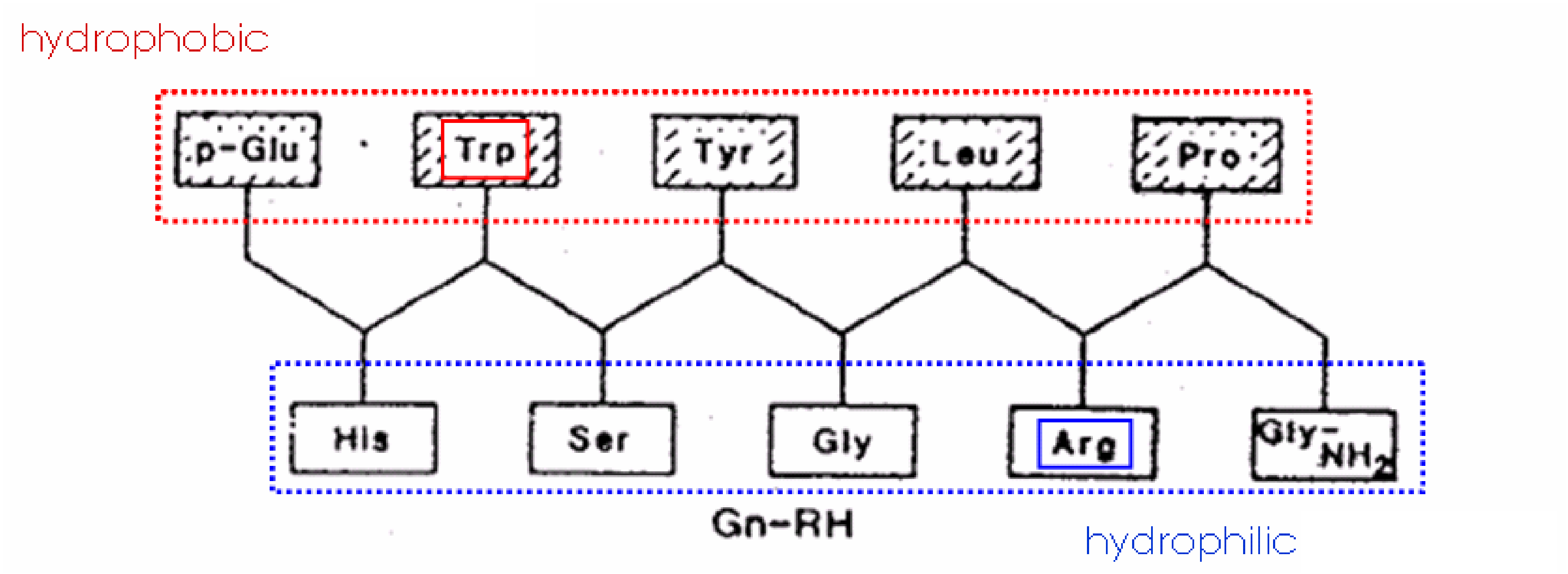
ampiphilic helix



axial projection of a potential π-helical conformation of residues 1 to 22 of human growth hormone releasing factor

L-tyrosine and L-arginine are the most hydrophobic and hydrophilic amino acids, respectively

secondary structure: amphiphilic structures



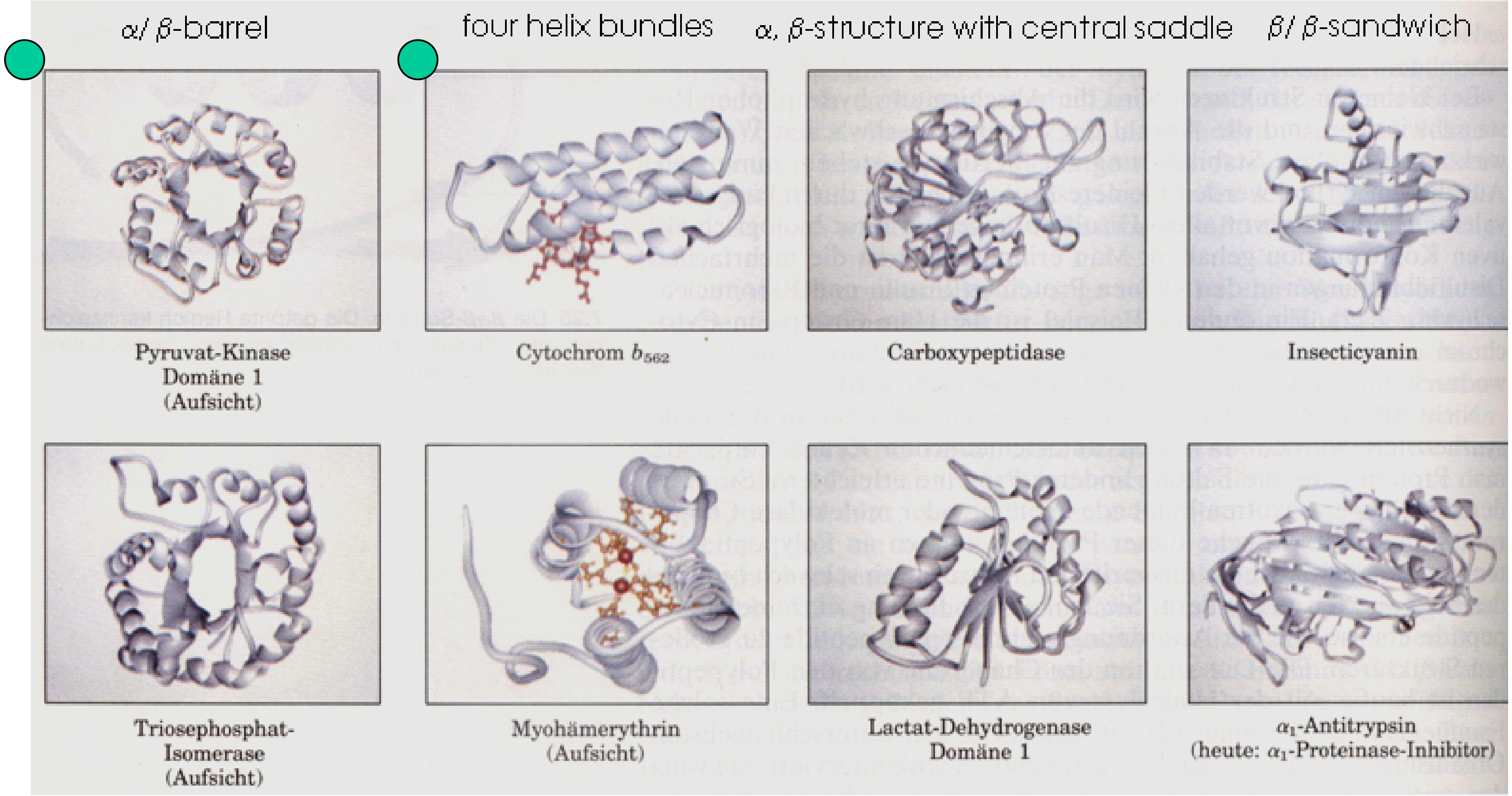
β-pleated sheet conformation of gonadotropin releasing hormone

L-tryptophan and L-arginine are the most hydrophobic and hydrophilic amino acids, respectively

Tertiary Structure

common structural motifs in proteins:

(supersecondary structures)



α/ β-barrel

four helix bundles

α, β-structure with central saddle

β/ β-sandwich

Pyruvat-Kinase
Domäne 1
(Aufsicht)

Cytochrom b_{562}

Carboxypeptidase

Insecticyanin

Triosephosphat-
Isomerase
(Aufsicht)

Myohämerythrin
(Aufsicht)

Lactat-Dehydrogenase
Domäne 1

α_1-Antitrypsin
(heute: α_1-Proteinase-Inhibitor)

Quaternary Structure:

quaternary structure refers to the association of multiple individual protein chains into a single protein with multiple subunits

the arrangement of the subunits (identical or different) gives rise to a stable structure

when they are different, each subunit tends to have a different function

(a common shorthand for describing such proteins is to use Greek letters for each type of subunit, and subscript numeral to specify numbers of units)

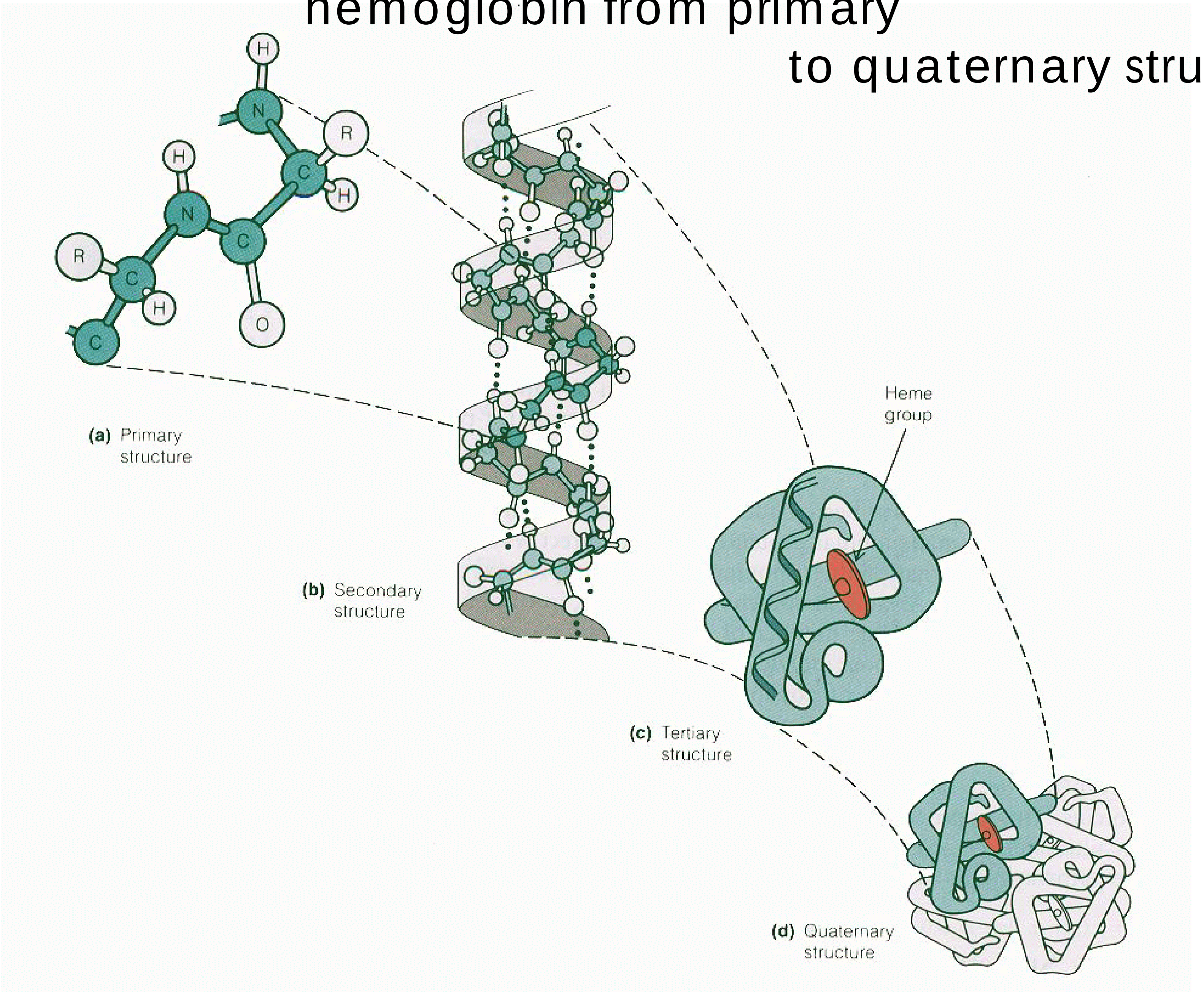
Thus, a protein designated $\alpha_2\beta\gamma$ consists of two α units and one each of β and γ;
the subunits usually are held together by hydrophobic interactions, the clustering serving to reduce exposure of hydrophobic side chains to the solvent; (occasionally, ionic interactions between carboxylate and amino side chains may contribute)

most protein multimers have significant rotational symmetry in the placement of the subunits

the quaternary structure is that level of form in which units of tertiary structure aggregate to form homo- or hetero- multimers

this is found to be remarkably common, especially in the case of enzymes

hemoglobin from primary to quaternary structure

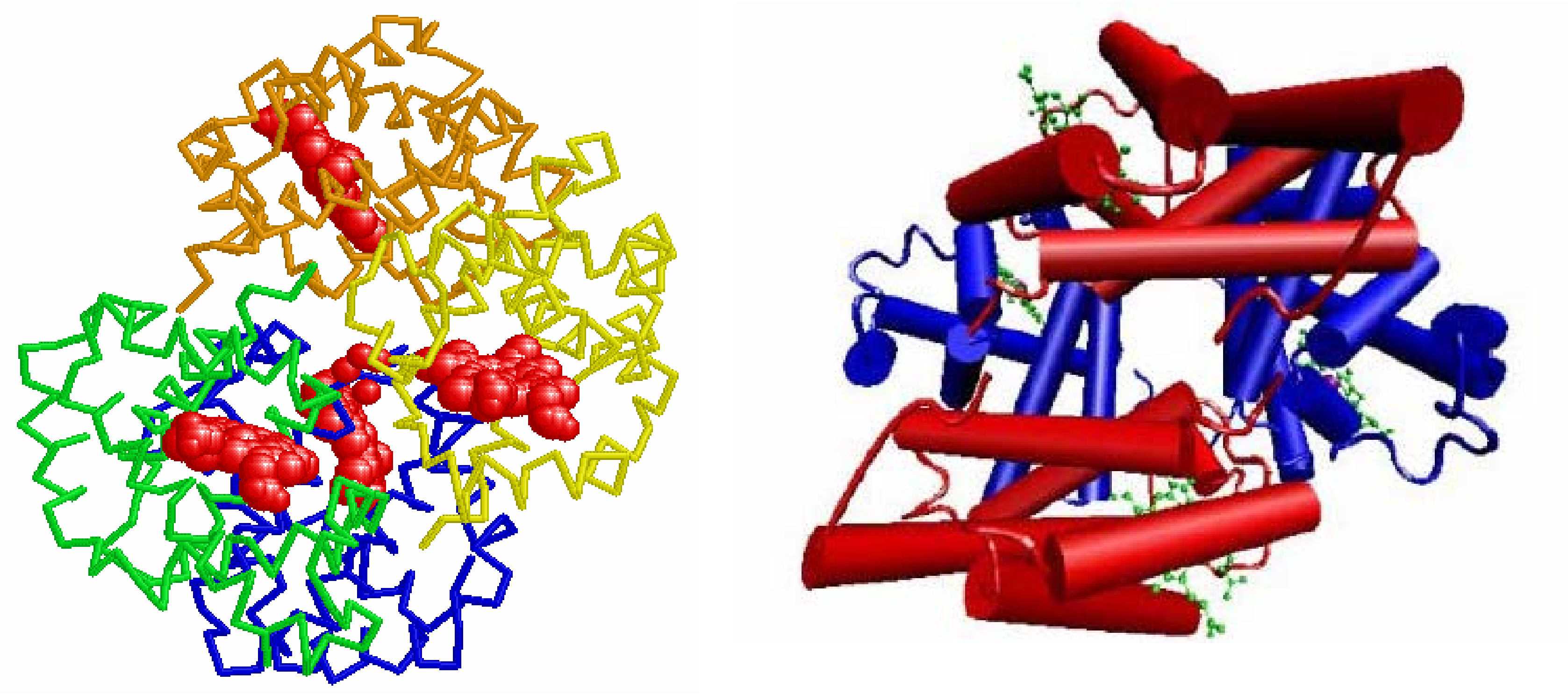


hemoglobin molecule ($\alpha_2\beta_2$)

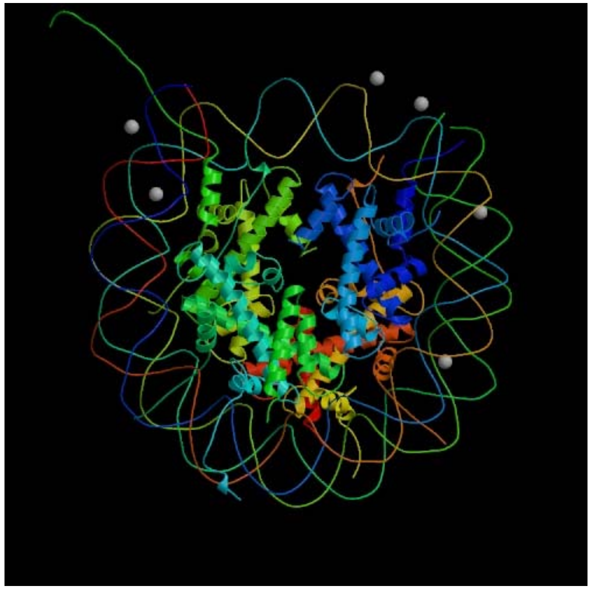
consists of four polypeptide chains: two α-chains, each with 141 amino acids and two β-chains, each with 146 amino acids

the protein portion of each of these chains is called "globin“; the α and β globin chains are very similar in structure
(In this case, α and β refer to the two types of globin; students often confuse this with the concept of α helix and β sheet secondary structures)

both the α and β globin chains contain primarily a helix secondary structure with no b sheets.

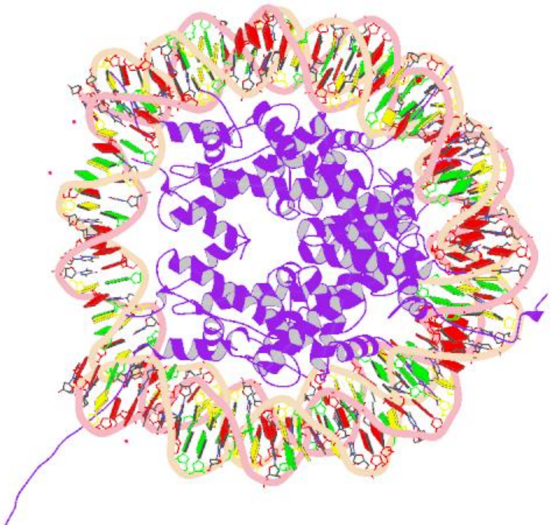


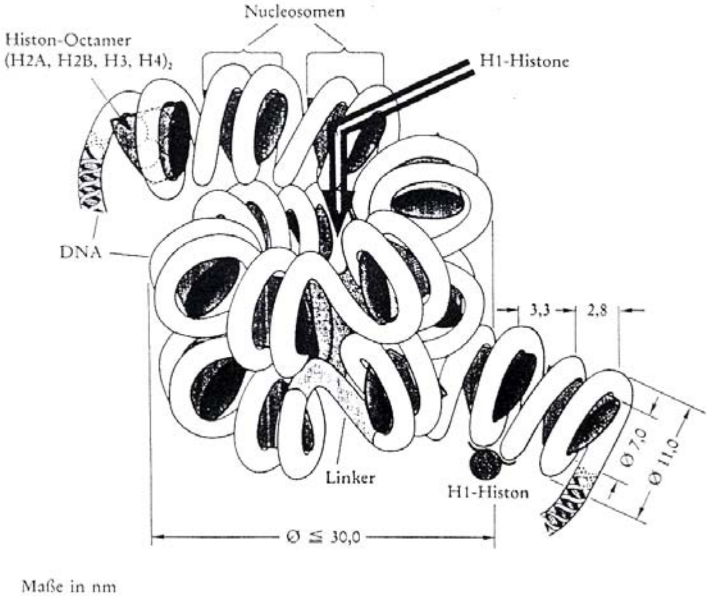
Chromatin scaffold for Info storage



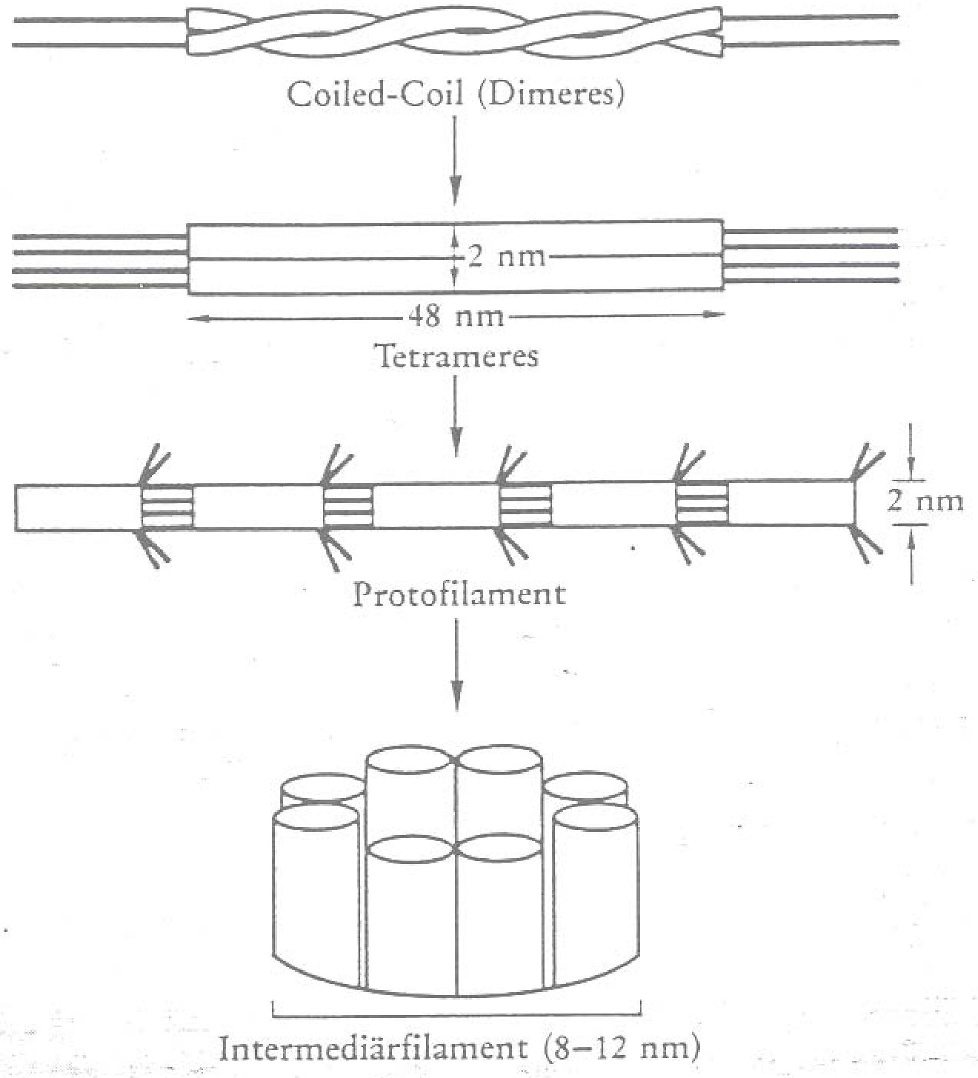
Model complex

(crystallized)

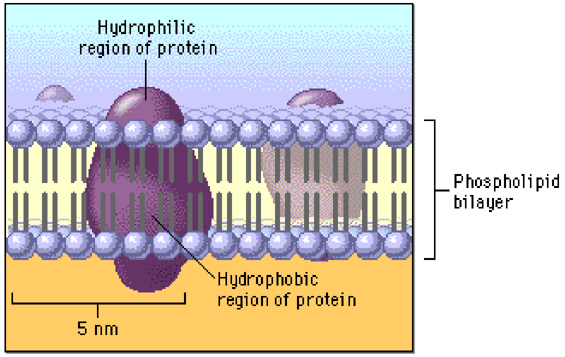




Superstructure of hair



Aqua porin



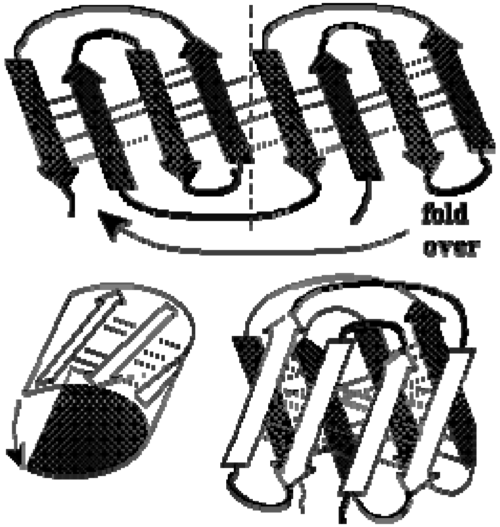
Function of porins

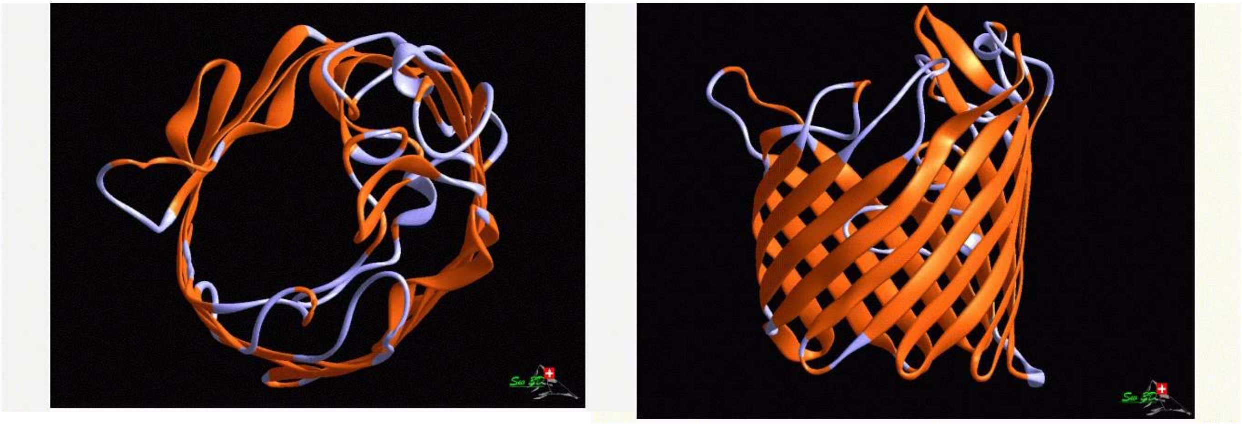
water filled tubes with a diameter of about 1 nm.

OmpF porin: Transmembrane homo-trimer, nonspecific diffusion of ions and molecules up to 600 Da

The diffusion speed depends: on concentration gradient and the molecular weight.

The passing of ions may be regulated by membrane potential (voltage gating).





β-barrel superstructure

Solid-phase supported synthesis (1963)

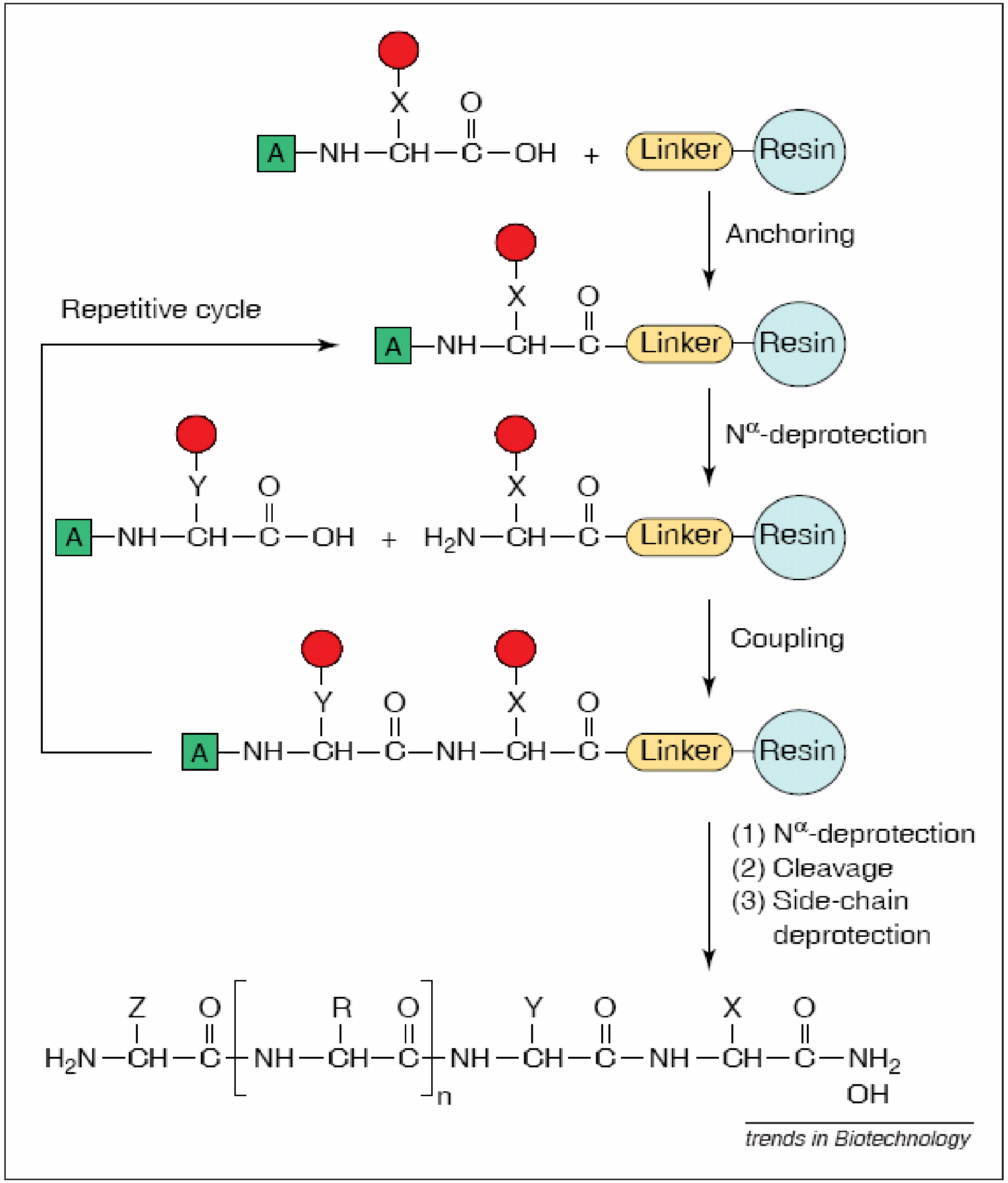
- advantages
  - ease of purification
  - high reaction rates (high concentrations)
  - broad methodology developed
- disadvantages
  - limited rates (solid phase reactions exhibit slower rates)
  - interactions with the support
  - limited solvents (PS resin, but also new developments)
  - orthogonal processes (synthesis and liberation from the support)

Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, 85, 2149. (tetra peptide 80% yield, 66% purity);
1963 JACS Nona peptide, 63% yield, 95% purity
1969 automated synthesis

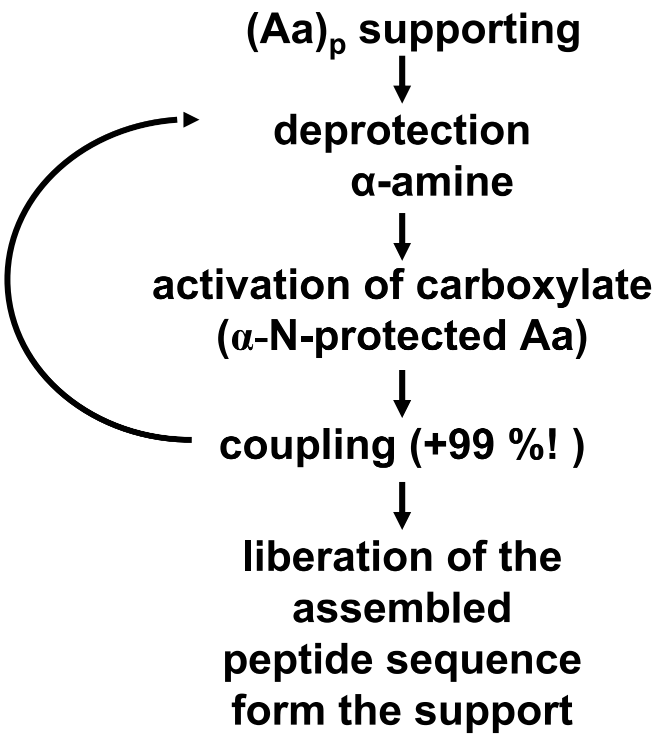
1984 R. B. Merrifield Nobel Prize in Chemistry

Gregg B. Fields, *Trends in biotechnology* 18 227-275 2000

Synthesis of polypeptides: Sequence controlled synthesis / forced step growth process.



trends in Biotechnology



$(Aa)_p$ supporting

deprotection α-amine

activation of carboxylate (α-N-protected Aa)

coupling (+99 %!)

liberation of the assembled peptide sequence form the support

Customized oligopeptide synthesis

Price per amino acid for peptide of 1-30 amino acids:

Purification \ Quantity	Crude	Desalt	>70%	>75%	>80%	>85%	>90%	>95%	>98%
1-4 mg	$4.80	$6.40	$9.60	$12.80	$14.40	$16.00	$19.20	$20.80	$32.00
80-100 mg	$14.40	$16.00	$28.00	$40.00	$44.00	$48.00	$57.60	$64.00	$96.00
1000 mg	$51.20	$52.80	$86.40	$120.00	$132.00	$144.00	$172.80	$192.00	$288.00

Price per amino acid for peptide of 41-50 amino acids:

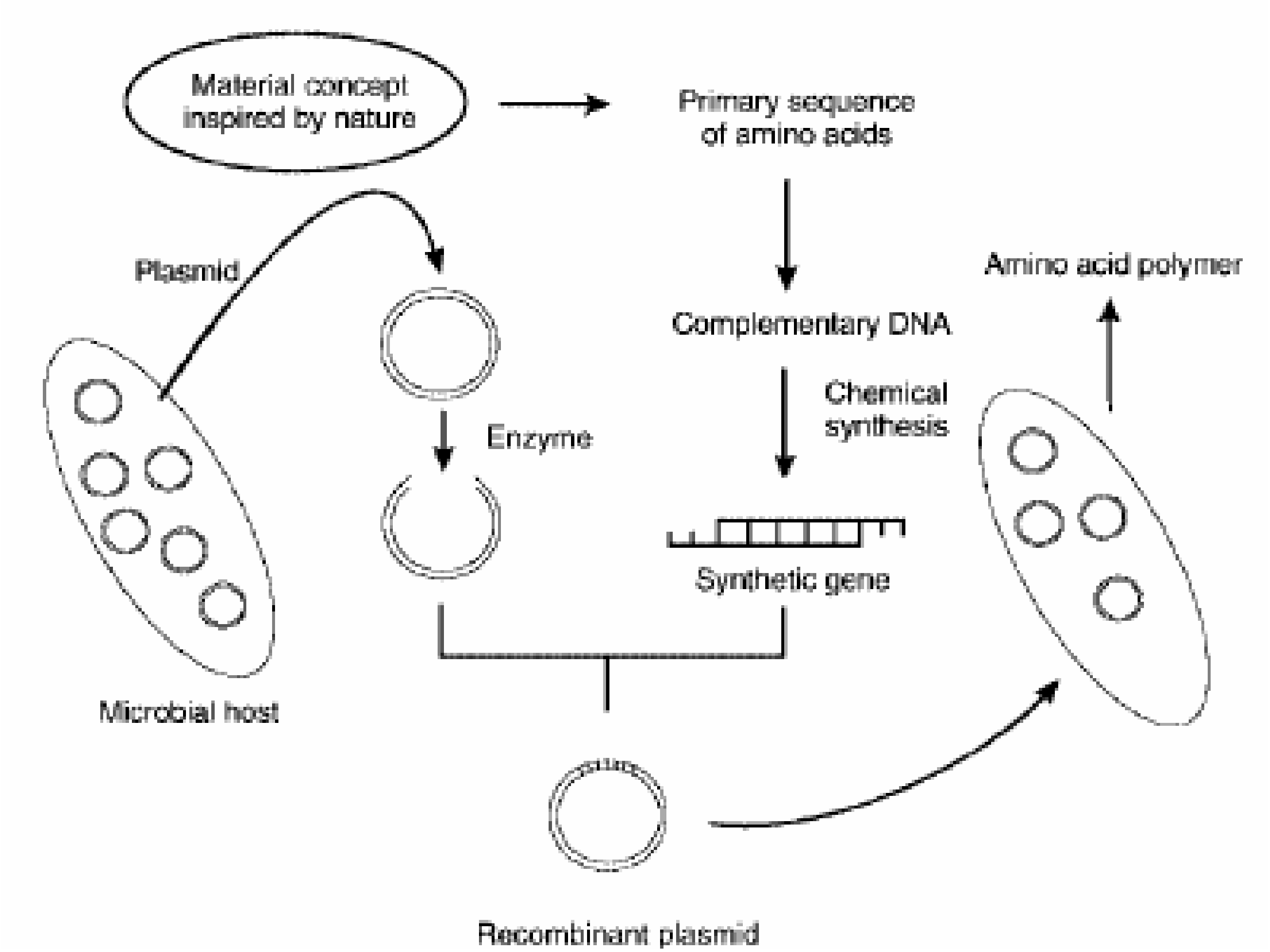
Purification \ Quantity	Crude	Desalt	>70%	>75%	>80%	>85%	>90%	>95%	>98%
5-9 mg	$8.96	$11.20	$15.68	$20.16	$24.64	$26.88	$31.36	$33.60	$51.52
80-100 mg	$20.16	$22.40	$39.20	$56.00	$61.60	$67.20	$80.64	$89.60	$134.40
1000 mg	$71.68	$73.92	$120.96	$168.00	$184.80	$201.60	$241.92	$268.80	$403.20

Solid phase peptide synthesis

GenScript Corporation

Synthesis of non-native proteins using microbial hosts

microbial Host: E. Coli



- **limited to L-Aa isomers**
- **unnatural Aa limited due to tolerance of biological system** (mainly the tRNA)

Recombinant plasmid:
- synthetic gene
- antibiotic resistance
- expression switches

Bombyx mori silk (silkworm moth)

β-Silk (fibroin): composition:

43 % Gly
30 % Ala
12 % Ser
4.8 % Tyr

Thus, 90 % of the silk contains only four AA´s.

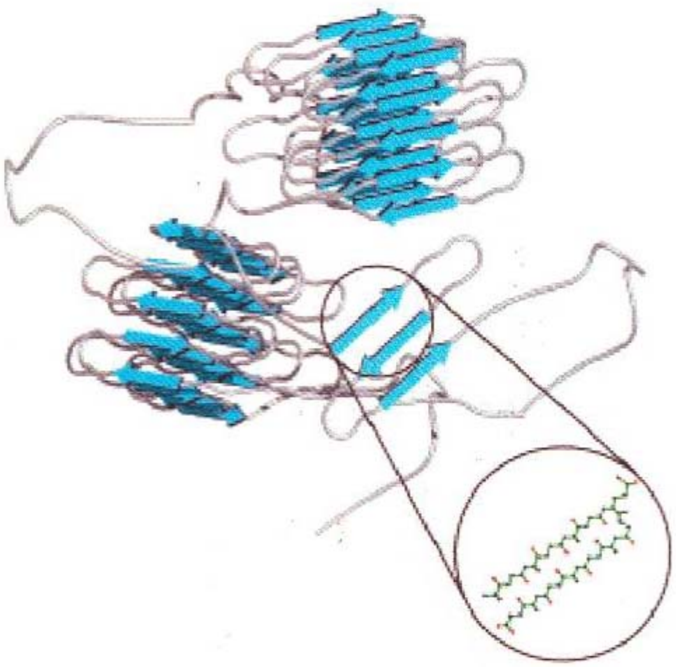
Crystalline regions are formed by repetitive sequences of (Ser-Gly-Ala-Gly-Ala-Gly)$_8$. Tyr occurs at transitions from crystalline to amorphous regions.

The packing of the protein in the crystalline regions is very dense.

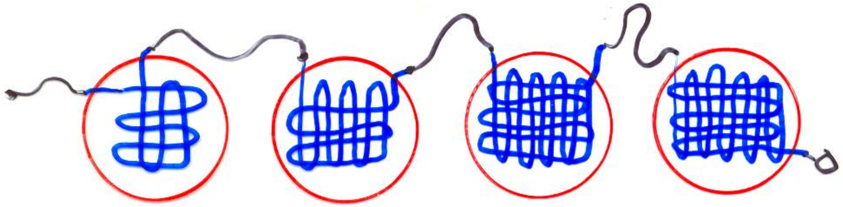
Two subunits, cross linked by disulfide bridges: H-chain: Mw 350,000, L-chain: Mw 25,000 Da.

The secret of spider-silk

Material	Strength (N/m^2)	Energy to break (J/kg)
Dragline silk	$1*10^9$	$1*10^5$
Kevlar	$4*10^9$	$3*10^4$
Rubber	$1*10^6$	$8*10^4$
Tendon	$1*10^9$	$5*10^3$



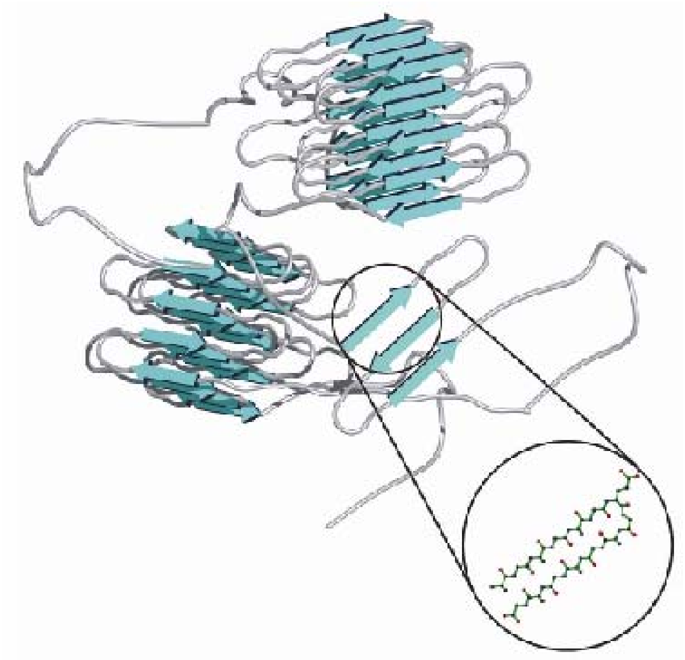
the secret: **„sacrifice-structure“**



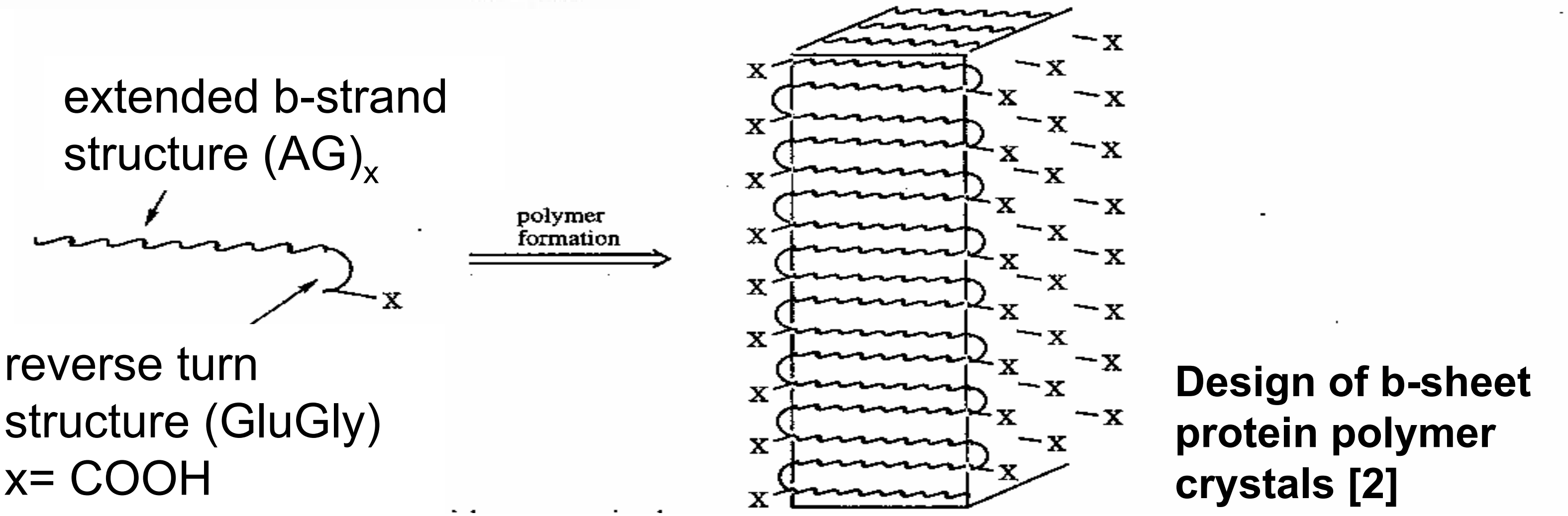
chemistry, lenghtscales & processing!

Artificial Spider silk

David A. Tirrell *Chem. Commun*. 2001, 1897–1904
David A. Tirrell *Adv. Mater.* 1997, 9, 302



Schematic representation of the organization of amorphous and crystalline domains in silk fibers.
Silk has inspired protein engineers to construct b-sheet materials.
A detail of a silk-like peptide sequence ((Ala-Gly)3Glu-Gly)$_n$ is encircled.



extended b-strand structure $(AG)_x$

reverse turn structure (GluGly)
x= COOH

lamellar crystal

Design of b-sheet protein polymer crystals [2]

Silk-like protein constructed by genetic engineering

Nexia (spider goad)

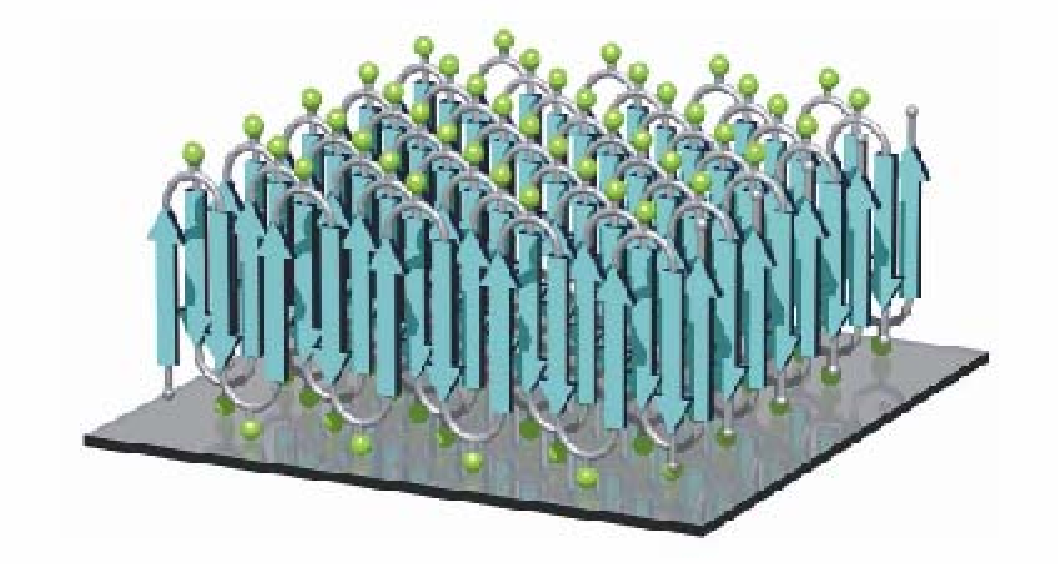
Potatoes (spider potatoes)

nexia
BIOTECHNOLOGIES

Kaplan, D.L. Spiderless spider webs. *Nat Biotechnol* **20**, 239-240 (March 2002).

Lazaris, A. *et al*. Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. *Science* **295**, 472-476 (January 18, 2002).

Design of protein based materials



Schematic representation of the lamellar crystalline phase formed by the peptide sequence $((AG)_3FG)_n$, in which phenylalanine is replaced by p-fluorophenylalanine.
The green spots indicate amino acid side chains with unnatural functionality, in this case fluorine, at the lamellar surface.

soft, ordered catalyst support

etc…pp

David A. Tirrell *Chem. Commun*. 2001, 1897–1904
David A. Tirrell *Adv. Mater.* 1997, 9, 302

Cyclic peptides