Hierarchy of Protein Structure

20 Amino Acids – There are 20^n possible sequences for a protein of *n* residues! 100 residue protein has 20^{100} possibilities 1.3 X 10^{130} !

There are ~ 40,000 sequences in the human genome (~100,000 proteins)

primary (1°) structure: the amino acid sequence

secondary (2°) structure: frequently occurring substructures or folds

tertiary (3°) structure: three-dimensional arrangement of all atoms in a single polypeptide chain

quaternary (4°) structure: overall organization of non-covalently linked subunits of a functional protein.

85

Protein and Peptide Structure

UV-Vis: Phe, Tyr, Trp, co-factors; concentration

Fluorescence: Trp, Tyr, covalently attached dyes (Cys)

Circular Dichroism (CD): backbone conformation

Infrared/Raman: characteristic bond vibrations

Electron Paramagnetic Resonance (EPR): environment near unpaired electrons (a radical or paramagnetic metal)

3-D structure determination

X-ray crystallography- method of choice. Major limitation is that the protein must form suitable crystals and the crystal diffraction pattern must be solved

multi-dimensional NMR- technology limited, restricted to peptides and "small proteins" (~30 KD, ~250 AA's)

86

Proteins have a native three-dimensional conformation (folded state)

Denatured: unfolded state of the protein (random coil)

Unfolded (denatured) $\rightleftharpoons$ Native (folded)

$\Delta G \simeq$ -16 - 40 KJ/ mol
(- 4 - 10 kcal/mol)
K_{eq} –~1000 - 10^8

Proteins folds by stabilizing desired conformations and destablizing undesired ones

ΔS is highly negative

Solvation issues
H-bonding
hydrophobic effects

Protein folding is a very complicated problem !!

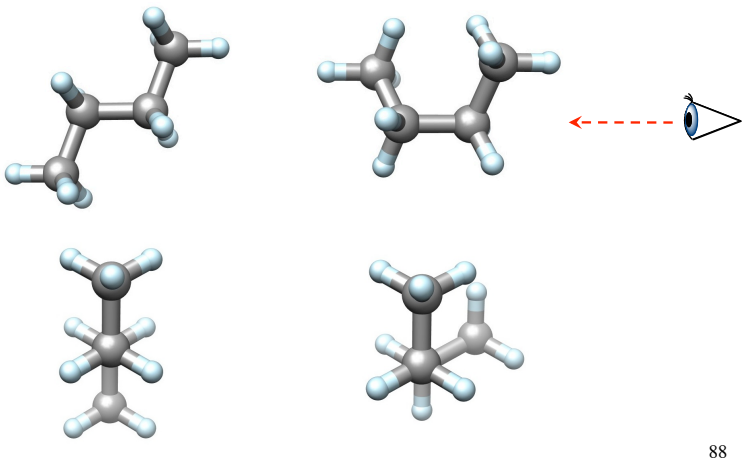
87

Conformation:

bond length
bond angle
dihedral (torsion) angle

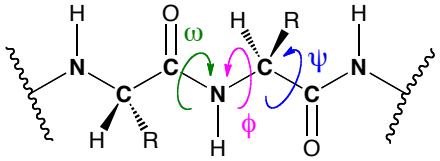
There is a high energy penalty for deforming bond lengths and angles from their ideal values

Butane



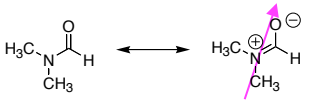
88

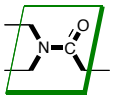
Backbone conformation



ω-angle: Cα-CO–NH-Cα
ϕ-angle: CO-NH–Cα-CO
ψ-angle: NH-Cα–CO-NH

The amide bond and ω-angle





coplanar

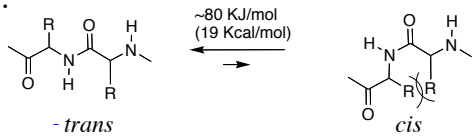
Methyl groups are not equivalent because of restricted rotation about the amide bond

Amide bonds have a large dipole moment ~ 3.5 D.

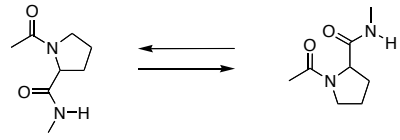
H_2O = 1.85 D
NH_3 = 1.5 D
H_3CNO_2 = 3.5 D

89

The ω-angle (dihedral angle of the Cα atoms) in peptides and proteins is 180° or 0°.

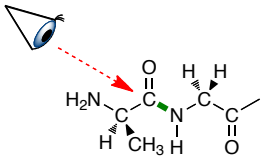


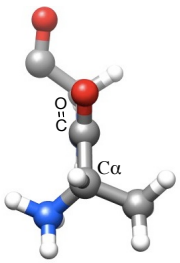
trans *cis*



trans *cis*

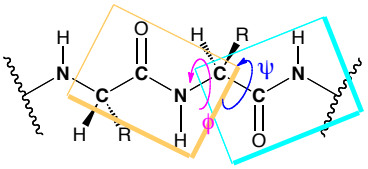
For pro, the cis is slightly favored



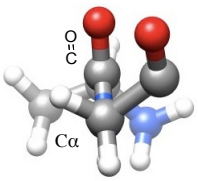


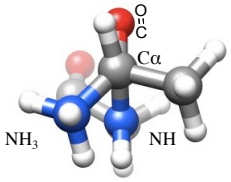
ω = 180°

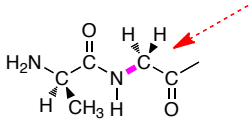
90

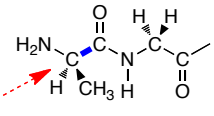


ϕ-angle: CO-NH–Cα-CO ψ-angle: NH-Cα–CO-NH



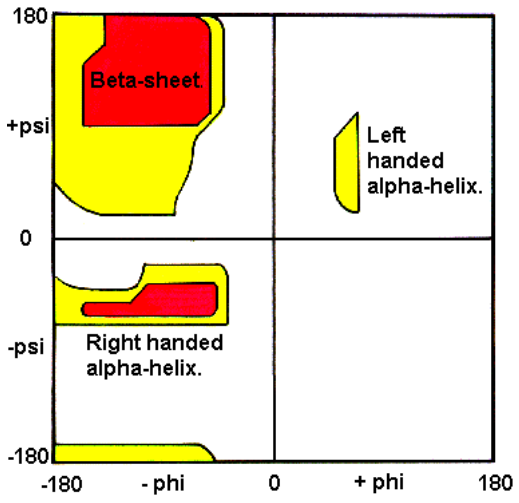






91

Ramachandran Plot



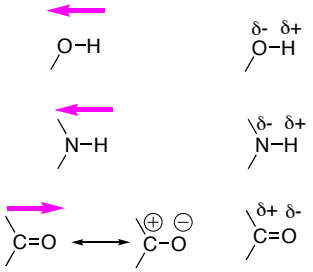
Secondary structure:

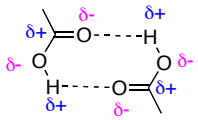
α-helix: ϕ = -60°, ψ = -40°

β-sheet: ϕ = 140°, ψ = -140°

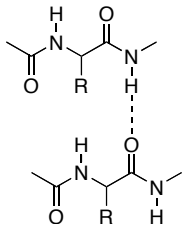
92

Hydrogen Bonding: non-covalent interaction, 4-16 KJ/mol





In solution, carboxylic acids exit as hydrogen bonded dimers



N-O distance
2.85 - 3.20 Å

optimal N-H-O
angle is 180 °

93

Hydrophobic Effects: tendency for non-polar solutes to aggregate in aqueous solution to minimize the hydrocarbon-water interface

Water is a dynamic hydrogen-bonded network.
water molecules around a solute is highly ordered
- ΔS, entropic penalty

Proteins fold to minimize their surface contact with water

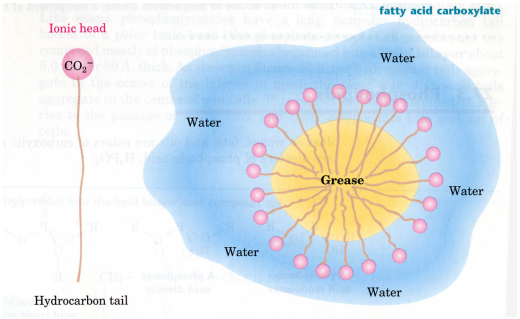
micelle structure: hydrocarbon on the inside, polar group on the outside.

Hydrophobic effects are important in the binding of substrates (ligands) into protein receptors and enzymes

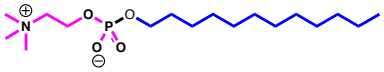
94

Micelles

Steric acid

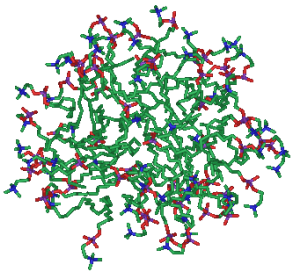


dodecylphosphocholine (DPC)



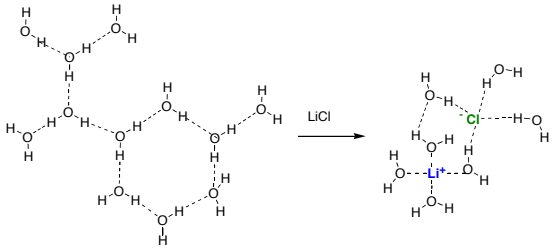
polar head group

hydrophobic tail



95

Salts can modify the hydrophobic effect through the change of water structure



Dissolving LiCl in water causes a net decrease in overall volume, less "cavities" in bulk water structure for solutes. (salting out)

Other salts such as guandinium chloride break up water structure and create more "cavities" or allow "cavities" to form more easily, allowing easier solvation of solutes. (salting in)

Surface tension studies to not support the cavitation theory.

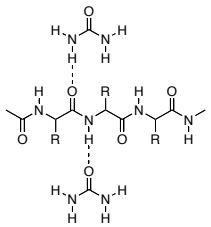
96

Hydrophobic effects are very important in the binding of a substrate into a protein (enzyme or receptor)

Denatured proteins- unfolding of the native three-dimensional structure of a protein by chemical influences such as:

- additives: guandinium salts, urea
- heat
- pH

old idea: denaturants such as urea unfolded proteins by hydrogen-bonding to the amide backbone



Mechanism probably involves better solubilization of the sidechains that are normally folded into the interior of the protein

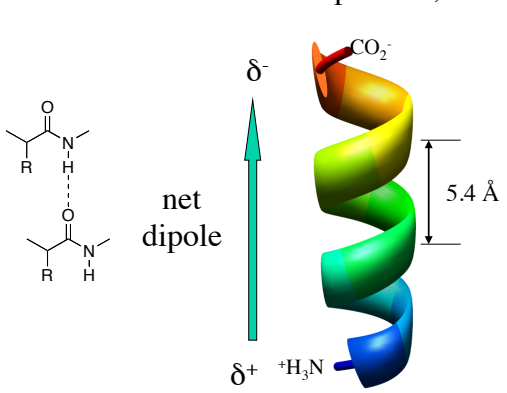
97

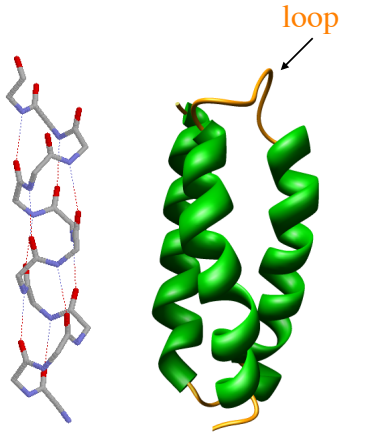
Protein Structure:
primary (1°) structure: the amino acid sequence
secondary (2°) structure: frequently occurring substructures
supersecondary: discrete, commonly occurring combinations of secondary structures (motifs); helix-loop-helix, βαβ
domains: independent folding subunits; β barrel, helical bundle
tertiary (3°) structure: three-dimensional arrangement of all atoms in a single polypeptide chain
quarternary (4°) structure: overall organization of non-covalently linked subunits of a functional protein.

Common secondary structures: α-helix
β-sheet
β-turn
disulfide bonds

98

α-Helix: amino acids wound into a helical structure
3.6 amino acids per coil, 5.4 Å





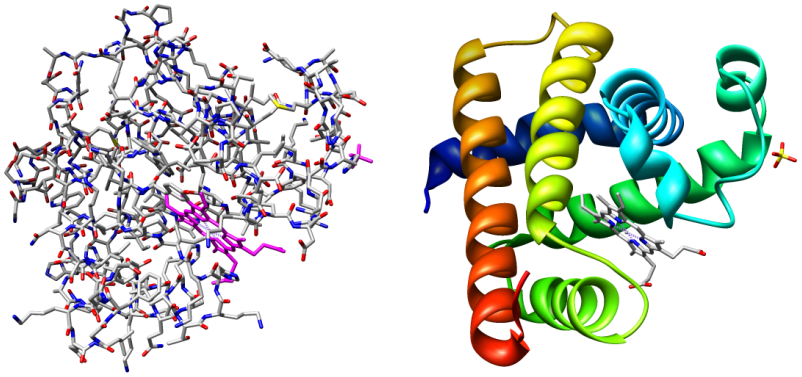
α-helix has a net dipole

pdb code: 2A3D

α-helix are connected by loops

99

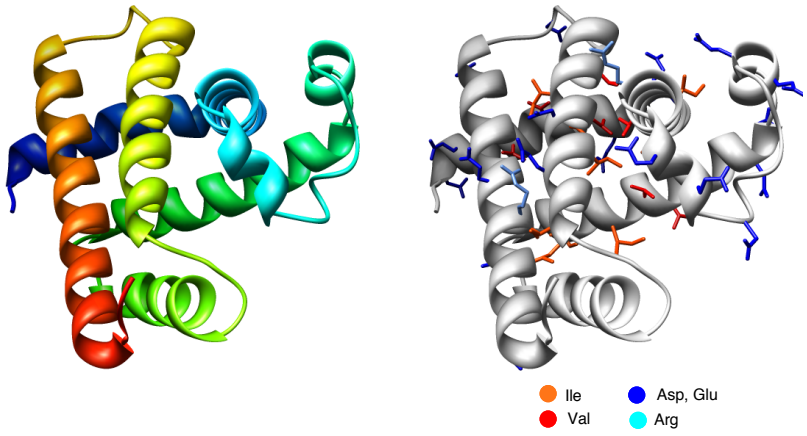
X-ray Structure of Myoglobin



pdb code: 1WLA

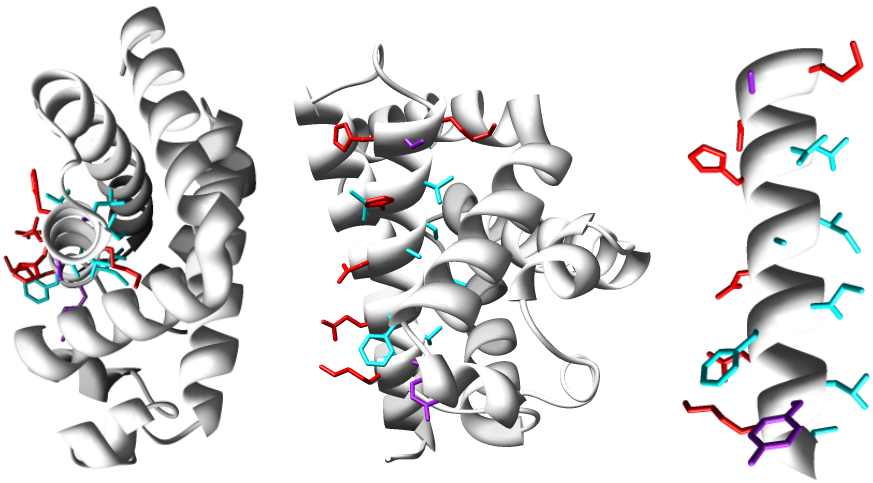
100

Hydrophobic and Hydrophilic Residues of Myoglobin



101

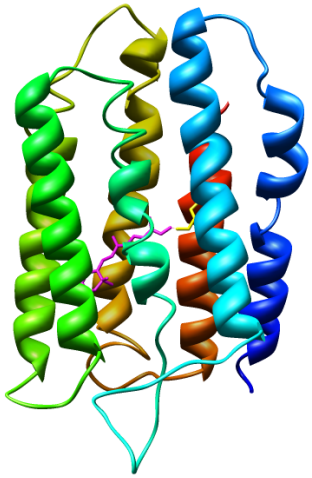
Myoglobin



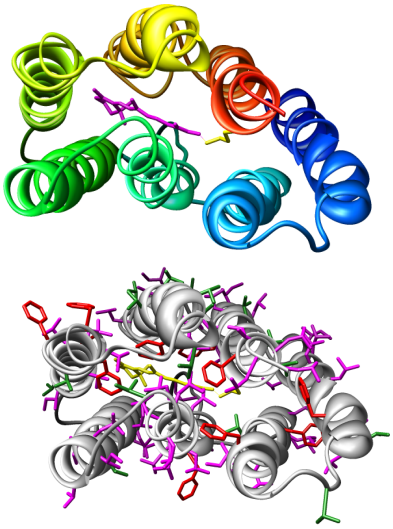
Pro • Ile • Lys • Tyr • Leu • Glu • Phe • Ile • Ser • Asp • Ala • Ile • Ile • His •Val • His • Ser • Lys

102

Bacteriorhodopsin



Schiff base linkage between
Lys-216 and retinal

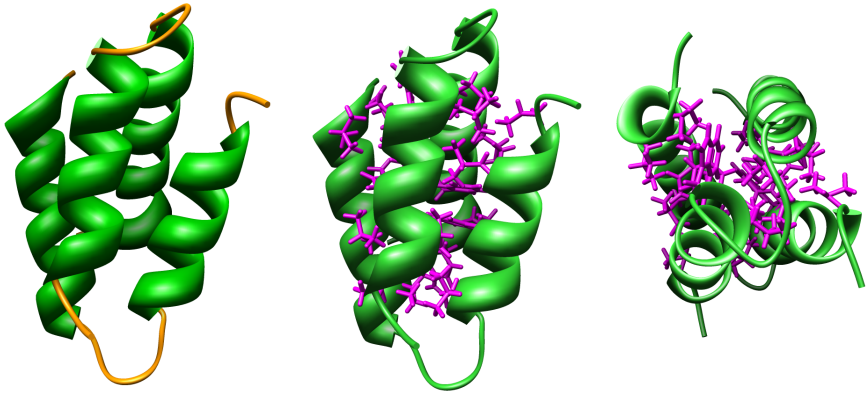


Leu Ile Val Phe

pdb code: 1AP9

103

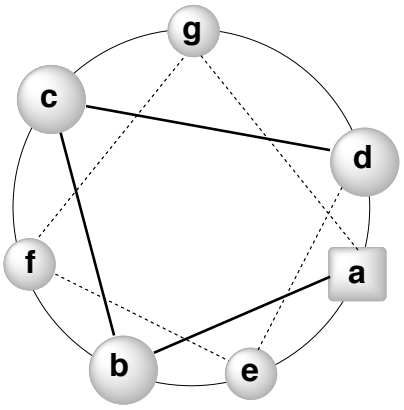
Helical Bundles: hydrophobic sidechains form an interface between α-helices (*de novo* protein design)



GLY GLU VAL GLU GLU LEU GLU LYS LYS PHE LYS GLU LEU TRP LYS GLY PRO ARG ARG
GLY GLU ILE GLU GLU LEU HIS LYS LYS PHE HIS GLU LEU ILE LYS GLY

pdb code: 1qp6

104

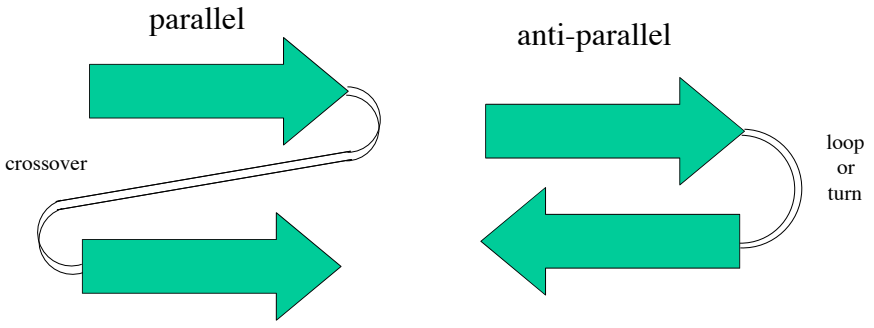


a b c d e f g a b c d e f
GLY GLU VAL GLU GLU LEU GLU LYS LYS PHE LYS GLU LEU TRP LYS GLY PRO ARG ARG

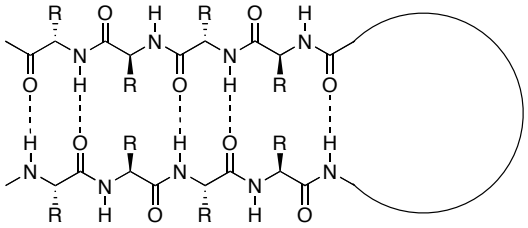
a b c d e f g a b c d e f g a
GLY GLU ILE GLU GLU LEU HIS LYS LYS PHE HIS GLU LEU ILE LYS GLY

105

β-sheets and β-turns



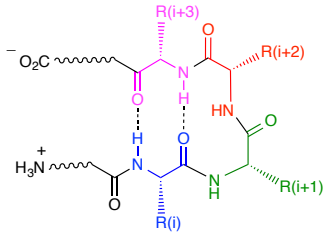
anti-parallel
β-sheet



loop
or
turn

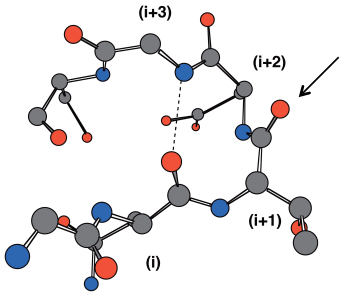
106

β-Turn



H-bond between (i) and (i+3) residues

β-Turn of Lysozyme
(residues: Asn46-Thr47-Asp48-Gly49)

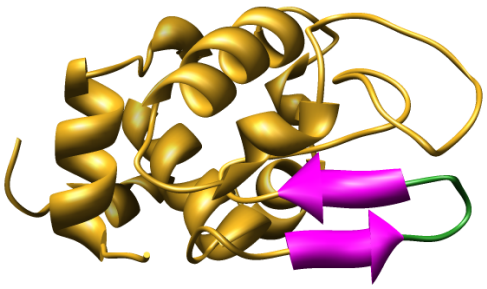


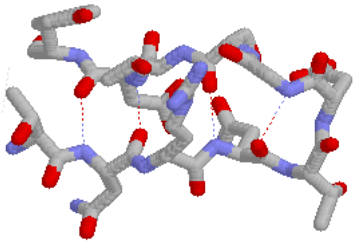
(i+1) carbonyl on the opposite side of the sidechains = Type I β-turn

β-Turn: a region of the protein involving four consecutive residues where the polypeptide chain folds back on itself by nearly 180 °. This chain reversal gives proteins a globular rather than linear structure. (Chou & Fasman *J. Mol. Biol.* **1977**, *115*, 135-175.)

107

β-Turn of Lysozyme





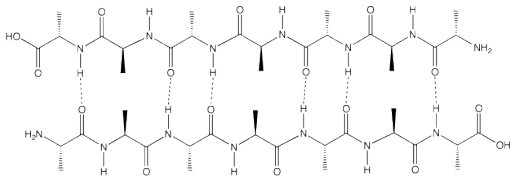
Typ_{53}-Asp_{52}-Thr_{51}-Ser_{50}-Gly_{49}-Asp_{48}

Thr_{43}-Asn_{44}-Arg_{45}---------Asn_{46}-Thr_{47}

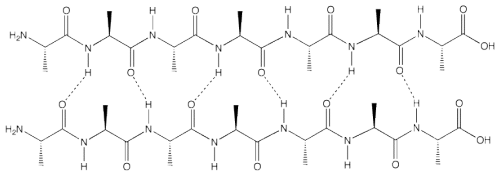
pdb code: 1AZF

108

Anti-parallel

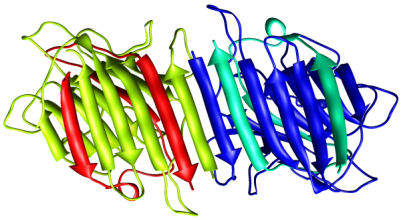


Parallel



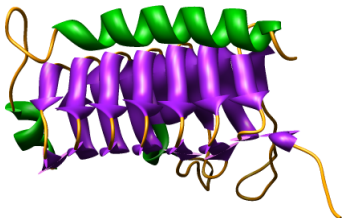
Anti-parallel β-sheets of lectin

pdb code: 2LAL



Parallel β-sheets carbonic anhydrase

pdb code: 1QRM



109

Some amino acids are found more often in certain secondary structures than others.

Chou, P.F.; Fasman, G.D. *Ann Rev. Biochem*. **1978**, *47*, 251-176

α-helix: Met, Glu, Ala, Leu, Gln, Lys, His

β-sheet: Thr, Tyr, Phe, Ile, Val, His

β-turns: Pro > Asn, Ser, Asp, Gly

His: α-helix ≈ β-sheet >> β-turn
Arg: α-helix ≈ β-sheet > β-turn
Cys: α-helix >> β-sheet > β-turn
Trp: β-sheet > α-helix > β-turn

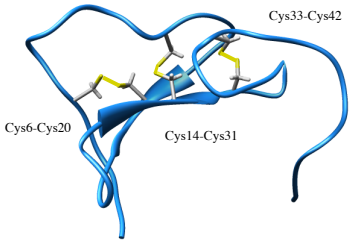
110

Disulfide bonds: covalent structural scaffolds, redox active, reversible

(Cysteine) 2 Cys-SH $\underset{[H]}{\overset{[O]}{\rightleftharpoons}}$ Cys-S-S-Cys (Cystine)

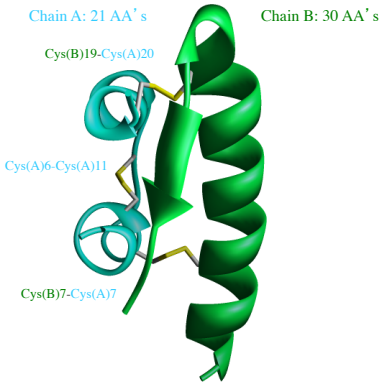
EGF

pdb code: 3EGF



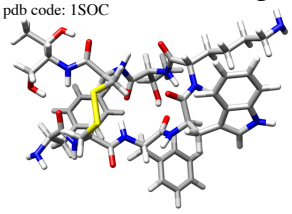
Human Insulin

pdb code: 1XDA



Somatostatin Analog

pdb code: 1SOC

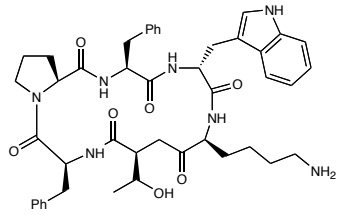


111

Somatostatin:

Ala-Gly-Cys-Lys-Asn-Phe-PheTrp
S
S
Cys-Ser-Thr-Phe-Thr-Lys

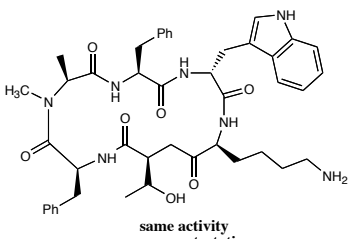
β-turn reponsible for biological activity



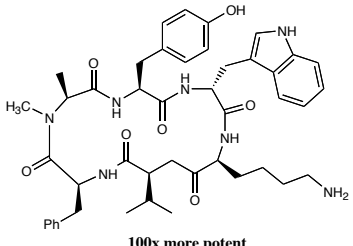
D-stereochemistry at the (i+2) psoition stabilizes the β-turn

remove disulfide

Pro-Phe-D-Trp
Phe-Thr-Lys



same activity as somatostatin

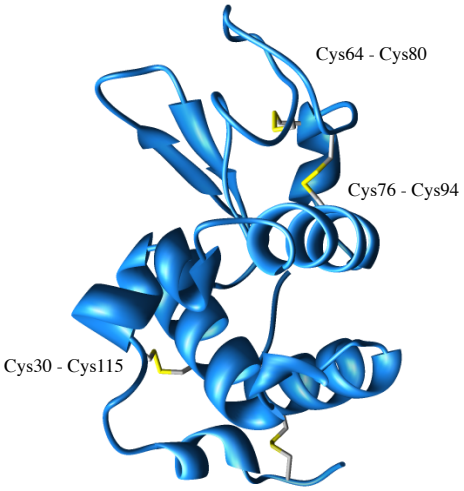


100x more potent than somatostatin

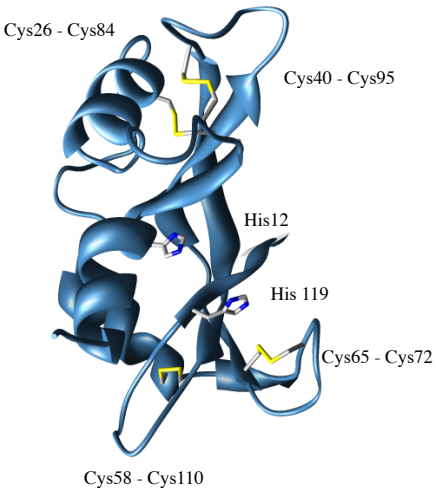
112

Disulfide bonds

lysozyme
pdb code: 1AZF



ribonuclease
pdb code: 1ALF



113

Combinatorial Chemistry: molecular diversity

"Synthesis and Applications of Small Molecule Libraries."
Thompson, L. A.; Ellman, J. A. *Chem. Rev.* **1996**, *96*, 555-600.
"Design, Synthesis, and Evaluation of Small-Molecule Libraries."
Ellman, J. A. *Acc. Chem. Res.* 1996, 29, 132 -143.
Combinatorial Chemistry, Nicholas K. Terrett, Oxford University Press, London, 1998

pharmaceutical industry- drug discovery

Lead Cmpd —optimization (synthesis)→ **Drug** —10 years, > $ 1 Billion→ **$ $ $**

Lead identification: literature (open & patent)
nature (natural products)

Careful optimization of a lead structure via chemical synthesis
"methyl-ethyl-butyl-futile game"

Number of marketable drugs per compounds that undergo preliminary biological testing: $\frac{1}{10{,}000}$ ← Rational drug design ← Combinatorial chemistry

114