PROTEINS

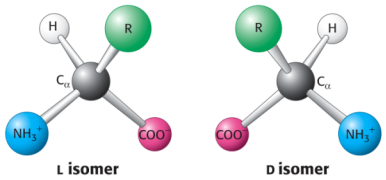
Proteins are the workhorses of the cell

perform > 90% of all the "actions" in the cell

Enzyme catalysis
Storage and transport of small molecules (hemoglobin, ferritin)
Movement (muscles)
Mechanical support (collagen)
Immunity (antibodies)
Signal transduction (cell surface receptors)
Control proteins (hormones, regulation of gene expression)

Proteins are a linear polymer of amino acids

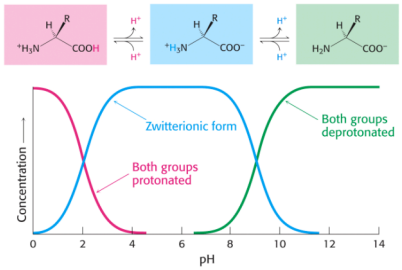
Each amino acid has an amino group, a carboxyl group, and a side chain (and a hydrogen)



Almost all amino acids in proteins are L-amino acids

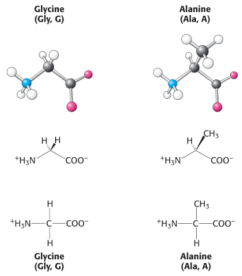
Both the amino group and the carboxyl groups are titratable (ionizable)

zwitterionic form is the predominant form at pH 7 (of a **free** amino acid)



In proteins, the side chain (R-group) is the major determinant of the unique properties of an amino acid

small, aliphatic



note single letter and 3 letter abbreviations

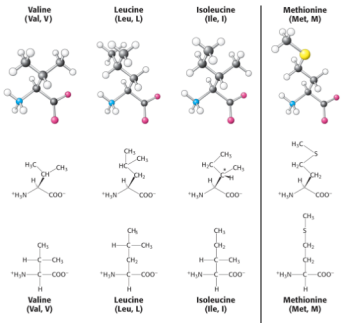
Classifications of the side chain character of amino acids:

hydrophilic versus hydrophobic
(polar versus nonpolar)

ionizable versus nonionizable
(titratable versus nontitratable)
(charged versus uncharged)

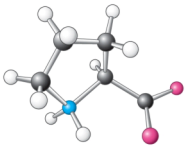
reactive versus unreactive

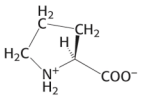
size



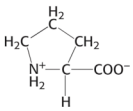
aliphatic

sulfur containing





Proline
(Pro, P)

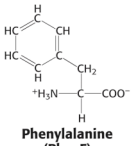


Unique: the amino group is bound to the alpha carbon and the side chain

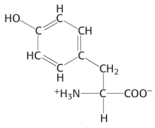
(I.e. the side chain is also the amino group)

pro causes kinks or bends in proteins

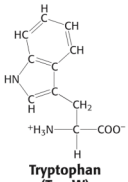
Will not form alpha helixes



Phenylalanine
(Phe, F)



Tyrosine
(Tyr, Y)

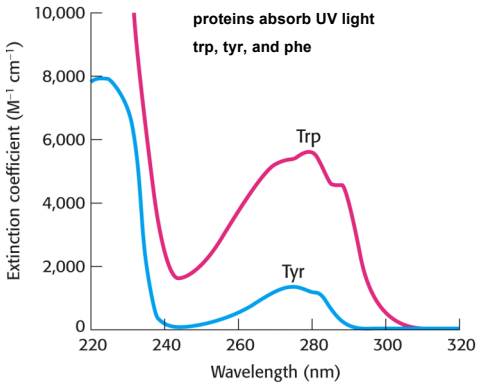


Tryptophan
(Trp, W)

the aromatics

very hydrophobic, non-polar

(tyr has some polar character via its OH group)



aliphatic hydroxyl side chains

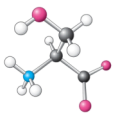
polar,

hydrophilic,

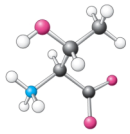
reactive,

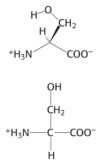
non-titrating (non ionizable)

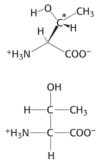
Serine
(Ser, S)



Threonine
(Thr, T)

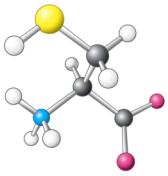


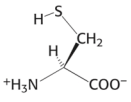




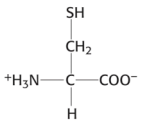
Serine
(Ser, S)

Threonine
(Thr, T)





Cysteine
(Cys, C)



another sulfur containing amino acid

the SH is highly reactive

The basic amino acids

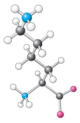
ionizable

positively charged at neutral pH

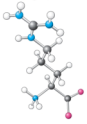
K and R also have the longest side chains

charged amino acids (of either sign) are hydrophilic

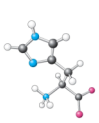
Lysine
(Lys, K)

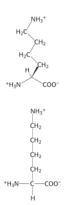


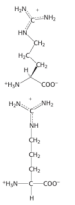
Arginine
(Arg, R)

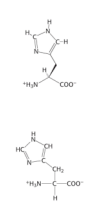


Histidine
(His, H)





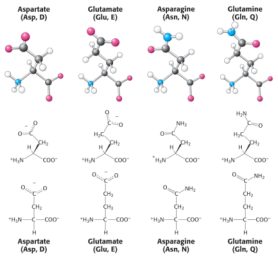




Lysine
(Lys, K)

Arginine
(Arg, R)

Histidine
(His, H)



asp and glu

aspartate/aspartic acid

glutamate/glutamic acid

negatively charged at pH 7

asn and gln

related but not charged or even ionizable (good H-bond formers)

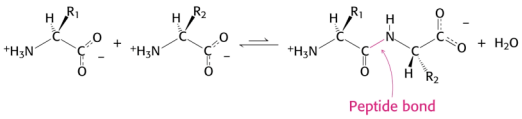
Groups that titrate in a protein:

In a protein, only the side chains and the two ends of the protein can titrate

TABLE 3.1 Typical pK_a values of ionizable groups in proteins

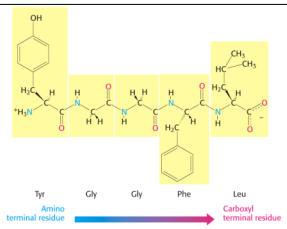
Group	Acid		Base	Typical pK_a*
Terminal α-carboxyl group		⇌		3.1
Aspartic acid Glutamic acid		⇌		4.1
Histidine		⇌		6.0
Terminal α-amino group		⇌		8.0
Cysteine		⇌		8.3
Tyrosine		⇌		10.9
Lysine		⇌		10.8
Arginine		⇌		12.5

*pK_a values depend on temperature, ionic strength, and the microenvironment of the ionizable group.



The amino group of one amino acid reacts with the carboxyl group of another amino acid to form a peptide bond

synthesis vs. hydrolysis



A pentapeptide (5 residues)

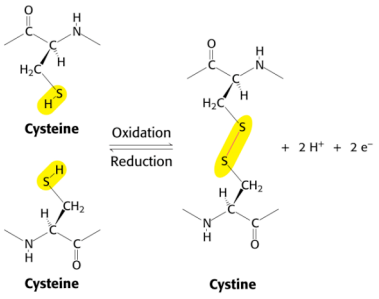
peptides (generally < 50 amino acids)

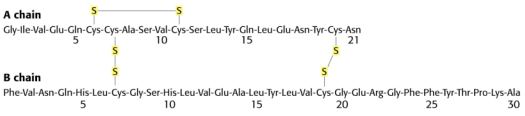
proteins (generally > 50 amino acids)

primary sequence (1° structure)

<u>the properties of a peptide (or protein) are at least the sum of the properties of its amino acids</u>

disulfide bonds: another type of covalent linkage in proteins

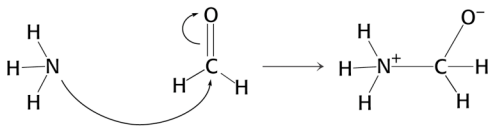




Insulin

the 1st protein sequenced

Fred Sanger, 1953



Covalent bonds

high energy (C-C 85 kcal/mol, C=O 175 kcal/mol)

formation and breaking of covalent bonds are the "work" of enzymes, but they generally use non-covalent bonds to perform these tasks

covalent bonds hold the protein chain together (along with disulfides), noncovalent bonds determine/stabilize the three-dimensional structure

Non-covalent interactions:

stabilize protein structure

are how biological macromolecules do their work

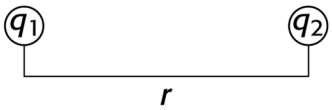
1) Hydrogen bonds (H-bonds)

2) Electrostatic interactions (salt bridges, ion pairs)

3) van der Waals interactions

4) Hydrophobic interactions (hydrophobic bonds, the hydrophobic effect

Electrostatic interactions



the strength of electrostatic interactions have a first power distance dependence

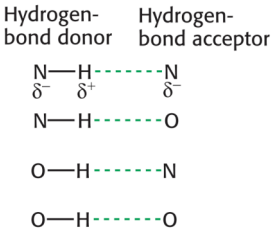
$E = kq_1q_2/\varepsilon\ r$

$E = kq_1q_2/D\ r$

ε= D= the dielectric constant, q = charges, k= proportionality constant

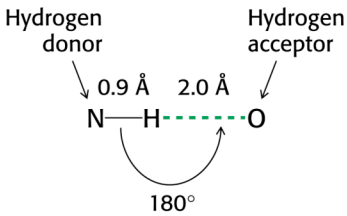
(for Force, the dependence is second order: $F = kq_1q_2/\varepsilon r^2$)

H-bonds



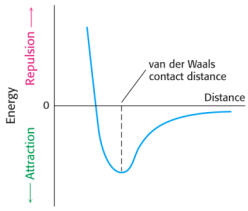
Directionality (unique to hydrogen bonds): H-bonds have a donor and an acceptor (usually N's and O's in biochemistry)

strength 1-7 kcal/mole (usually 1-3)



Hydrogen bonds
strait bonds are stronger than bent ones

van der Waals interactions



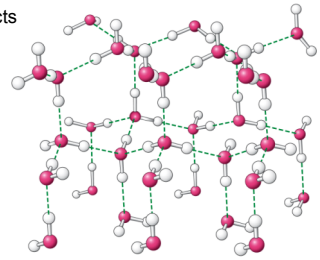
All non covalent bonds have an analogous distance to E relationship

Distance dependence of van der Waals interactions

the van der Waals interatomic distance or

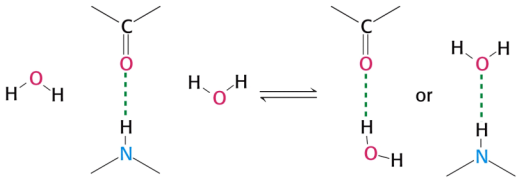
van der Waals radius

Water dramatically effects the strength of all non covalent bonds



ice

water is highly self cohesive – forming multiple H-bonds between molecules in solution and in ice

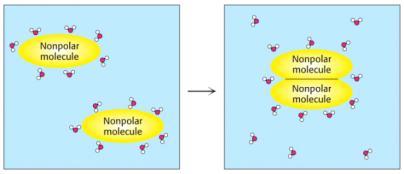


water competes with the normal partners in

H-bonds and electrostatic bonds

ions in solution compete for normal electrostatic bond partners

Hydrophobic interactions



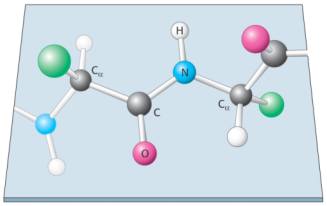
water is semi-structured around a hydrophobic/nonpolar molecule

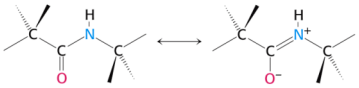
aggregation of hydrophobic molecules decreases their interaction with water, releasing some water to become disordered again (favorable entropy)

the "bond"is not between the 2 hydrophobic groups, it is the result of their solvation properties

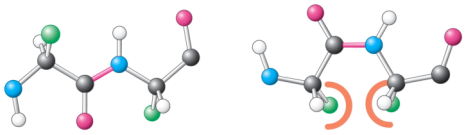
the peptide bond (carbon backbone) is planar due to resonance stabilization

the 2 alpha carbons, the carbonyl carbon, the oxygen, the nitrogen and the hydrogen atoms are all co-planar



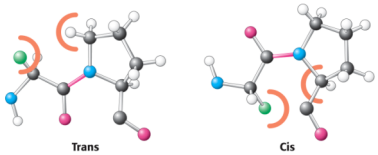


Peptide bond resonance structures

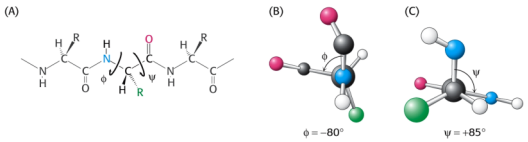


Trans **Cis**

the peptide bond is trans for all amino acids except proline (which can be cis or trans)



Trans **Cis**



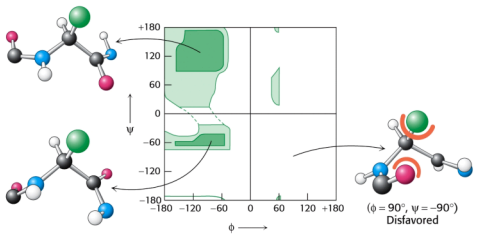
The bonds on both sides of the alpha carbon can rotate

the phi and psi angles

Secondary structure (2° structure) = a localized stretch of specific structure, characterized by certain combinations of phi and psi angles

alpha helix, beta sheet, random coil

beta turns, helical caps



A Ramachandran plot

plots phi angles versus psi angles

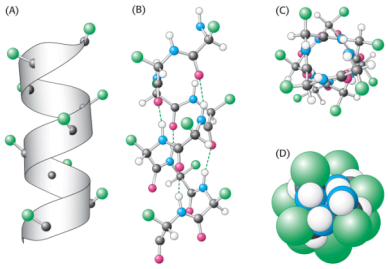
shaded areas are found in real proteins

the alpha helix (a secondary structural element)

Linus Pauling

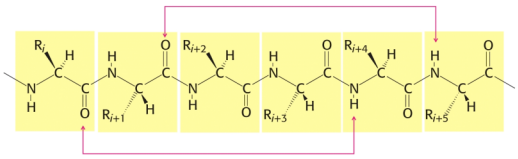
H-bond stabilized

side chains are on the outside

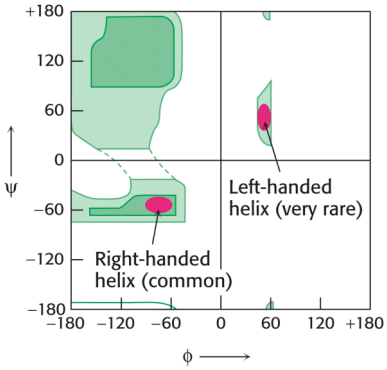


H-bonding in the alpha helix

(C=O of n to NH of n+4)

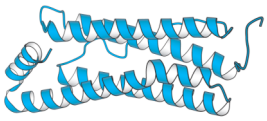


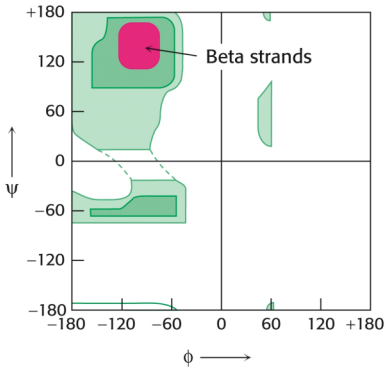
αhelicies can be 10-25 residues long

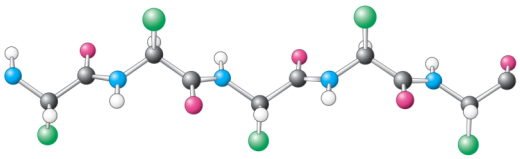


3° structure = how secondary structure elements come together to form the folded polypeptide

a subunit of ferritin, a mostly α helical protein

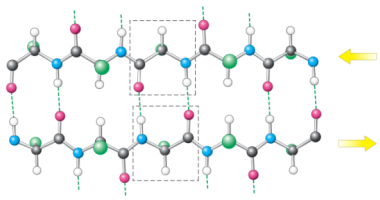






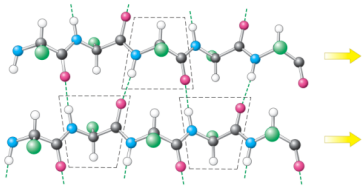
Beta strand, more elongated than an alpha helix

No intrastrand bonds

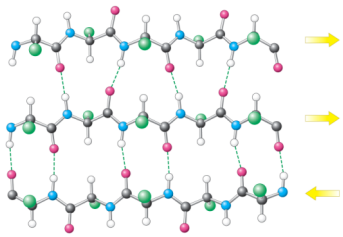


An anti-parallel beta sheet

Beta sheets are stabilized by inter-strand H-bonds



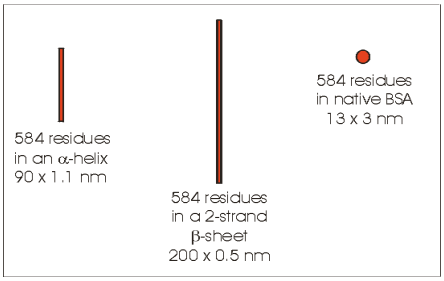
A parallel, 2 stranded beta sheet



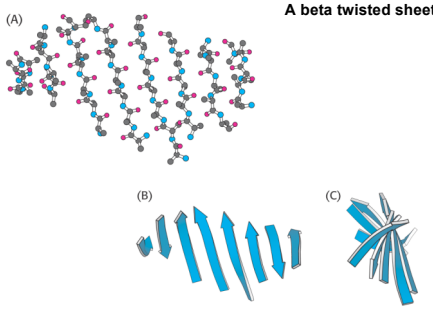
A parallel and antiparallel pair of strands in the same beta sheet

The strands need not follow directly after one another in the protein sequence

Bovine Serum Albumin contains 584 AA Residues



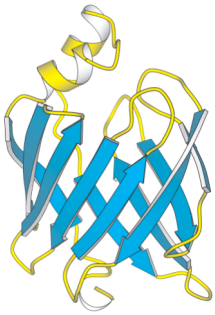
A beta twisted sheet

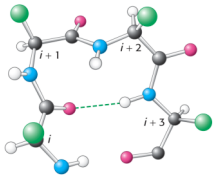


Tertiary structure: the way secondary structure elements fold together

The 3 dimensional topology of the protein

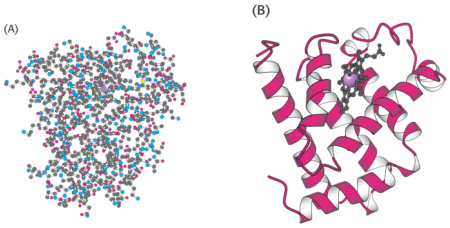
A fatty acid binding protein: mostly β-sheet (a β barrel), plus α helix and random coil (loops)



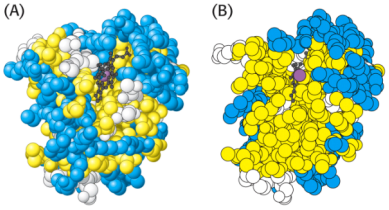


A beta turn: 4 residues, the most compact way a peptide chain can turn

Also stabilized by H-bonding



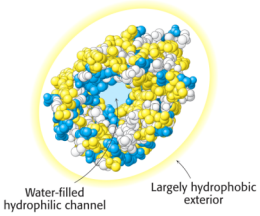
Myoglobin, a mostly α helical protein

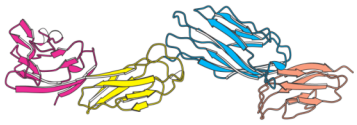


The distribution of polar and nonpolar residues in myoglobin

Most cytoplasmic proteins have a hydrophobic core

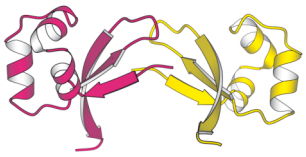
Porin, a membrane protein, has more hydrophobic residues on the outside and more hydrophilic residues on the inside





quaternary structure: the specific association of different, separate polypeptide chains to form a multi-subunit protein

CD4 is a cell surface protein that consists of 4 separate, similar polypeptides



Cro repressor, a gene regulation protein in bacteriophage λ

A dimer of identical subunits

Oligomers, dimers, trimers, tetramers, etc.

The hemoglobin tetramer:

4 separate polypeptides +

4 heme groups (prosthetic groups)

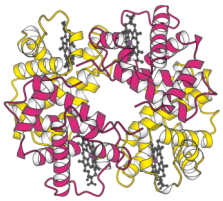
2 alpha chains

2 beta chains

a heme in each chain

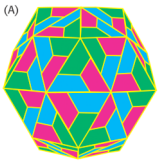
each chain can bind an oxygen:

so the holoenzyme binds 4 O_2's



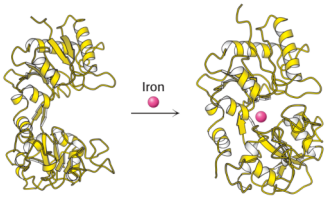
holoenzyme

apoenzyme



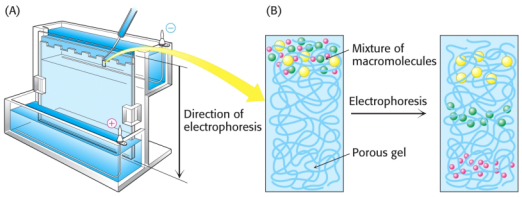
virus capsules are an extreme example of quaternary structure: they can consist of 100's to thousands of separate polypeptide chains

Rhinovirus, a virus that can cause the common cold, consists of 240 subunits (60 copies each of 4 different polypeptides)



Proteins are often flexible, and their flexibility is an integral part of the way they function

<u>conformational changes</u> = shape changes in proteins

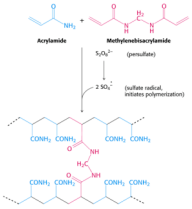


Gel electrophoresis

A separation method based on charge and size

Native versus denaturing gels

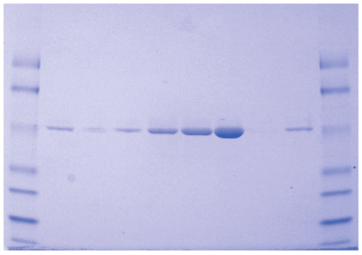
Separation is opposite that of size exclusion chromatography



Altering the concentrations of acrylamide and bis-acrylamide changes the porosity of the gel

SDS is a denaturant, and is negatively charged

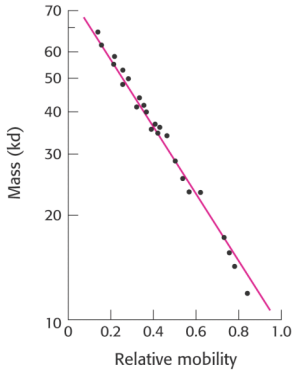
SDS Page



Most proteins need to be stained to be seen by eye

Coomassie blue

Mobility is ~ 1/log MW

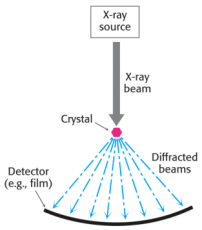


Determining Protein Structure -- Crystallography

Crystallography: the major source of information on protein tertiary and quaternary structure

Starts with formation of crystals from a concentrated solution of the protein

(ammonium sulfate and polyethylene glycol, both remove H_20 from the protein)



X-rays are shined through the protein crystal, and are diffracted by the protein crystal

Just like light is diffracted by any regular lattice

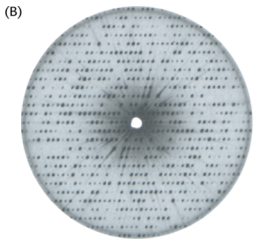
The protein crystal is a three-dimensional lattice

X-rays are used because proteins are much smaller than the resolution of visible light

The diffraction pattern is determined by the structural details of the crystal lattice that is diffracting the X-rays

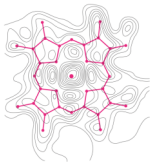
The problem is: there is no lens that can focus X-rays into an image (as could be done with visible light)

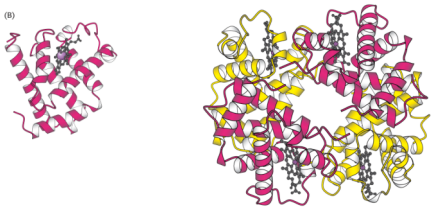
So Fourier transform analysis is used: essentially a mathematical lens that reconstructs the "image" from the diffraction pattern.



The "image" that results from Fourier analysis is an electron density map:

Analogous to a topographical map of the protein, where the regions of highest electron density (where the atoms are) give the strongest signal





Myoglobin (John Kendrew) and hemoglobin (Max Perutz) were the first protein crystal structures solved

Today there are 47,403 structures in the PDB

40,261 in 2006

27,855 in 2004

18,618 in 2002

www.rcsb.org/pdb

Determining Protein Structure -- NMR

Nuclear magnetic resonance (NMR, MRI)

Depends on the intrinsic magnetic properties of certain atomic nuclei

Allows structure determination of proteins in solution

Works for proteins up to about

20 kDa

TABLE 4.4 Biologically important nuclei giving NMR signals

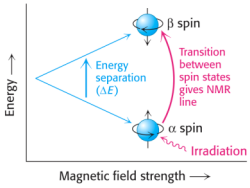
Nucleus	Natural abundance (% by weight of the element)
^{1}H	99.984
^{2}H	0.016
^{13}C	1.108
^{14}N	99.635
^{15}N	0.365
^{17}O	0.037
^{23}Na	100.0
^{25}Mg	10.05
^{31}P	100.0
^{35}Cl	75.4
^{39}K	93.1

Not always the most abundant isotope

Magnetic nuclei can exist in two different spin states

Absorbance of magnetic radiation induces a transition between spin states -- this produces a signal

Just like light has intensity and wavelength, the magnetic field has strength and frequency



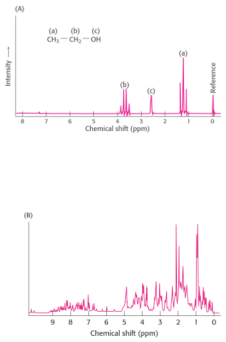
Why people want stronger NMR's

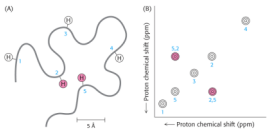
300, 500, 700, 900 MHz

The frequencies at which different nuclei absorb magnetic energy are called "chemical shifts" and are measured in ppm

One dimensional NMR

Nuclei in different environments absorb energy at different frequencies (ppm)





Nuclei that are < 5Å apart influence each other's spin

At each frequency, pulse with every other frequency.

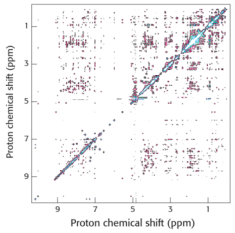
If no atoms are w/in 5Å of each other, then no chemical shifts will change in the 2nd dimension.

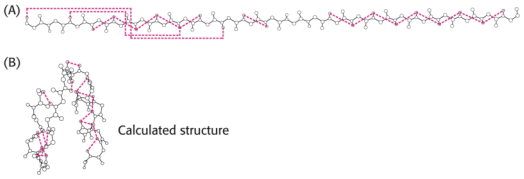
If an there is an atom "Y" within 5Å of the atom absorbing energy in dimension 1 (atom "X"), then atom Y will have a different ppm in the first dimension than in the pulsed 2nd dimension.

The diagonal is the original 1 dimensional NMR spectrum

Assignment of peaks is one of the major difficulties in NMR structure determination

Often done by mutagenesis of each amino acid



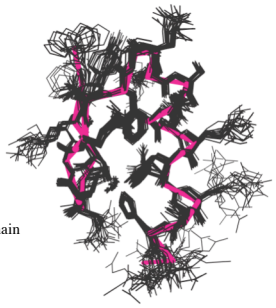


Once you know

1) Which amino acid corresponds to which peaks
2) Which amino acids are < 5Å from each other

Then you can computationally fold the primary sequence into a structure that is consistent with this infomation

NMR always provides a family of structures (usually about 20) which are generally believed to represent the natural breathing (flexibility) of proteins in solution



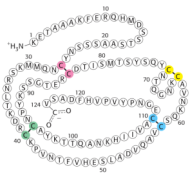
A zinc finger DNA binding domain

How do proteins fold into the proper structure?

Chris Anfinsen's experiments on ribonuclease:

for many proteins, the primary sequence is all the information needed to completely fold the protein

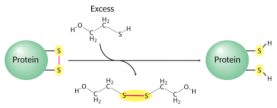
(i.e. to correctly form 2° and 3° and 4° structure).



ribonuclease:

124 amino acids

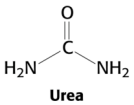
3 disulfide bonds



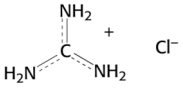
β-mercaptoethanol (β ME) and dithiothreitol (DTT) are two reagents that reduce disulfide bonds in proteins

Large amounts of either reagent completely reduces all the disulfides in a protein

Small amounts of either reagent will allow disulfides to rearrange in a protein (by making it easy for them to reduce and re-oxidize)

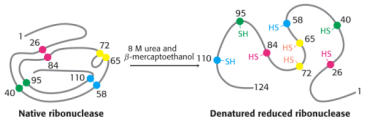


Urea



Guanidinium chloride

guanidine and urea are reagents that cause proteins to denature (unfold)



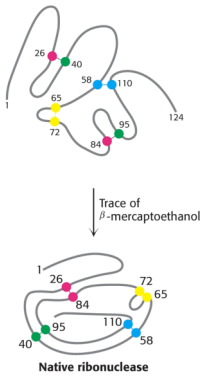
addition of 8M urea and βME to ribonuclease completely denatures (unfolds) it

If the urea, and <u>most</u> of the βME are removed, the protein will spontaneously refold into its correct structure

(incorrect disulfides will rearrange into the correct disulfides)

If all the urea and <u>all</u> the βME are removed, almost all of the protein molecules will refold correctly the first time

(any remaining "scrambled" ribonuclease (with incorrect disulfides) can be corrected using a small amount of βME.



How do proteins spontaneously refold?

What is the "protein folding code" ?

<u>The protein folding problem:</u> how does the primary sequence of a protein dictate its structure? OR How can we predict the structure of a protein from its sequence?

Do they use "random search"?

<u>Levanthal's Paradox:</u>

consider a 100 residue protein,

if each amino acid can be in the alpha helix, beta sheet, or random coil configuration, then there are 3^{100} different possible conformational forms of this protein

$3^{100} = 5x10^{47}$

If each possibility is tried for 0.1 picoseconds ($0.1x10^{-12}$ seconds),it would take $1.6x10^{27}$ years to try all possibilities

This is many times the age of the earth

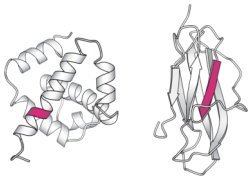
Most proteins completely fold in less than a second

each amino acid has a tendency or propensity toward being in certain types of 2° structures, but these distributions are not extreme enough to allow prediction of a protein's structure from its sequence

TABLE 3.3 Relative frequencies of amino acid residues in secondary structures

Amino acid	α helix	β sheet	Turn
Ala	**1.29**	0.90	0.78
Cys	**1.11**	0.74	0.80
Leu	**1.30**	1.02	0.59
Met	**1.47**	0.97	0.39
Glu	**1.44**	0.75	1.00
Gln	**1.27**	0.80	0.97
His	**1.22**	1.08	0.69
Lys	**1.23**	0.77	0.96
Val	0.91	**1.49**	0.47
Ile	0.97	**1.45**	0.51
Phe	1.07	**1.32**	0.58
Tyr	0.72	**1.25**	1.05
Trp	0.99	**1.14**	0.75
Thr	0.82	**1.21**	1.03
Gly	0.56	0.92	**1.64**
Ser	0.82	0.95	**1.33**
Asp	1.04	0.72	**1.41**
Asn	0.90	0.76	**1.28**
Pro	0.52	0.64	**1.91**
Arg	0.96	0.99	0.88

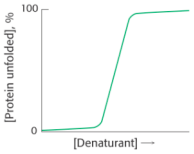
Note: The amino acids are grouped according to their preference for α helices (top group), β sheets (second group), or turns (third group). Arginine shows no significant preference for any of the structures.
After T. E. Creighton, *Proteins: Structures and Molecular Properties*, 2d ed. (W. H. Freeman and Company, 1992), p. 256.



The same 6 residue sequence in two different proteins

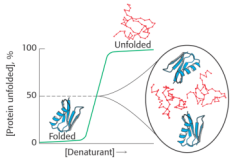
Illustrates that 3° interactions can be strongly involved in determining secondary structure

protein folding is often an equilibrium process (as shown by Anfinsen's experiments), so measuring the energetics of protein folding should provide insights into the process



the experiment: titrate denaturant into a protein solution and monitor the unfolding as a function of denaturant concentration

monitor loss of secondary or tertiary structure (best) or changes in localized structure (such as movement of a trp residue)



At each denaturant concentration one measures the % unfolded protein. This corresponds to a ratio of folded to unfolded protein:

[U]/{F] (or [D]/[N], denatured/native), this is an equilibrium constant

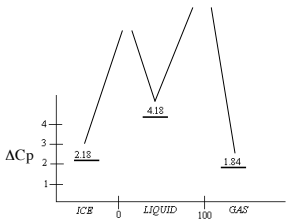
$-RT \ln [U]/[F] = \Delta G_{unfolding}$ at each [denaturant]

graph $\Delta G_{unfolding}$ versus [denaturant], and extrapolate to zero [denaturant]

this gives the $\Delta G_{unfolding}$ of the protein in solution

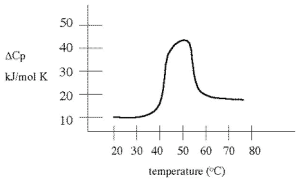
The heat capacity of a substance during a phase change approaches infinity

Shown are the specific heat capacities(J/kg) for different phase states of water



Differential scanning micro-calorimetry can be used to measure the heat capacity of a protein as it "melts" from the folded form to the denatured form.

The center of mass of the peak is the Tm. The difference between the native and denatured state baselines is the Δ Cp. The area under the peak is the ΔH.



DSC and other themal denaturation monitors are increasingly being used as drug screening assays, since binding of anything to a protein will its ΔG of unfolding, and thus could alter the Tm or ΔH of unfolding, or both.

The Laws of Thermodynamics:

1st Law: conservation of energy, the energy of a closed system is constant

2nd Law: the entropy of the universe increases

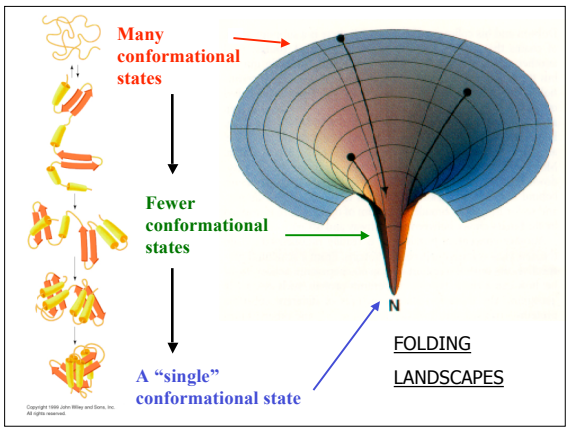
The Guidelines of Biothermodynamics:

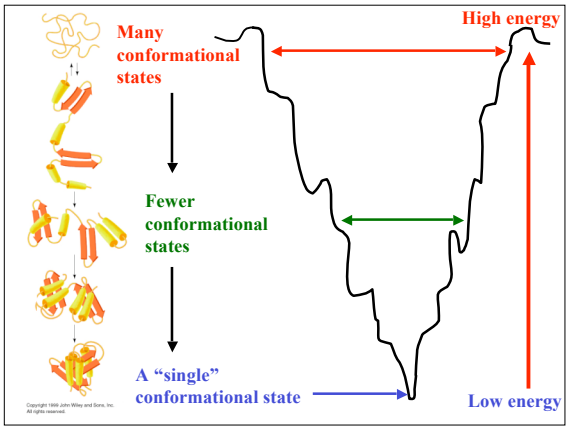
Enthalpy changes (ΔH) generally correspond to changes in non-covalent bonds.

Entropy changes (ΔS) generally correspond to changes in bound water or ions, or changes in conformational flexibility (configurational entropy).

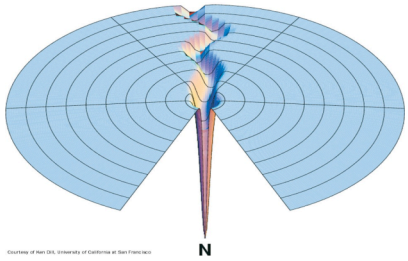
Heat capacity changes (ΔCp) seem to correspond to changes in accessible surface area (ΔASA), as well as changes in conformation (e.g. DNA distortion), and changes in conformational flexibility.

Hypothesis: understanding protein folding thermodynamics will crack the protein folding code.

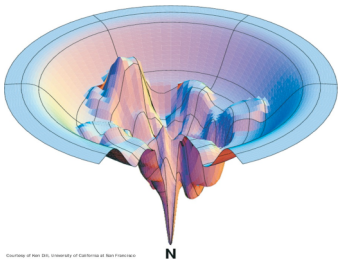




Classic folding landscape (pathway model).



A realistic folding landscape?



Courtesy of Ken Dill, University of California at San Francisco