Chapter 4
3D structure of proteins

Problems: 1, 2, 3, 4, 6, 8, 9 (May do 10 in lab)

4.0 Intro

Last chapter 1° structure, - what are the chemical bonds
This chapter 2°,3°,4° structure what is arrangement of atoms in space

This is incredibly important . Function depends on 3D arrangement .
Protein denaturation - remove natural activity
how is this done?
Simplest is to heat it up.
Do NOT change a single chemical bond
All you do is change 3D arrangement
Protein losses all function
Protein may even have completely different physical properties
Cooked or uncooked egg white

4.1 Overview

Spatial arrangement of atoms in protein called **conformation**
A conformation change is any change that changes 3D arrangement without making or breaking chemical bonds
simplest conformation change, rotate around a bond
How many bonds can rotate in a protein? How many conformations?
Proteins in functional, folded state are called **native**
generally thought that native state is most stable, most thermodynamically favored conformation
Is true for many proteins, may not be true for all proteins, controversial issue
Will now study what determine a proteins conformation

A. Proteins conformation stabilized largely by weak interaction

Stability - tendency to maintain native conformation
Native proteins only marginally stable
E needed to destroy conformation 20 -65 kJ/mol
200-450 kJ needed to break a covalent bond
So 1/10 E needed to break a chemical bond very weak
So between native and non-native structure will not mess with chemical bonds
What interaction did we talk about in Gen Chem that govern how molecules interact with each other?
Strongest - Charge/Charge
Medium H bonds
Lowest E Van Der Waals/London forces

These are all being used to make a protein conformation #1 force
talked about Chapter 3, the hydrophobic force, the entropic effect that makes oil drops. Thus, the protein interior tends to be filled with hydrophobic (oil like) residues
All of these forces are used and summed together to determine the protein's overall stability.
As the book points out, its actually more complicated. Than this because you have too look at the forces in the folded protein and compare them to forces that occur when a protein is unfolded

For instance, you would think putting an NH of one peptide close to a CO of another peptide in a folded conformation would give you lots of stability because you can form a hydrogen bond. Think again. What would happen when the protein was unfolded? NH would be in water, can it form an H bond there ? YES so did you really gain anything? NO. In fact you might have lost! In water you have 1 H bond fo the NH and one for the CO, in the interior you only have one H bond.

Bottom line, is complicated have a few general rules
Bury hydrophobic in interior
expose charge and polar on exterior
if you bury a + charge you have to bury a - charge next to it and vice versa
if you bury a H bonding group, you have to bury it in a conformation that gives it something to H bond to

Note these rules change if insoluble protein or membrane protein, then environment is different so forces are different

Now have overall principle, lets look at the details

B. Peptide bond is rigid and planar

Figure 4-2
talked last chapter about peptide bond
special chemical properties O is elector negative pulls electrons
N is electron rich so can donate
Gives CN bond a partial double bond character
Makes bond planar
Makes protein structure simpler, not as many bonds to rotate
Most peptide bonds are trans (a few cis are seen in proline)

Look at a dipeptide 2D model
see how have simplified folding each AA only has two free rotations in backbone

Things get even simpler
- Atoms are bulky
- Can occupy same space
- So not all conformations are possible!
- By convention call N-Ca bond phi φ
- Ca to CO bond psi χ
- Can plot allowed and unallowed regions
  - Called Ramachandran plot after inventor
  - Figure 4-3
- Can see that conformation space is restricted
- (Convention of 0 and + rotation given in fig 4-2 not needed here)

4.2 Protein Secondary Structure (2°)

2° structure refers to local conformations in some small region of protein
There are just a few of easily recognized and understood structure: alpha helix and beta sheet, turns, and loops. Anything that doesn't fit into these categories is called random

A. Alpha Helix

First hypothesized by Pauling & Corey 1930's
At that time knew basic dimensions of peptide bond and that was planar
Using X-ray diffraction saw that certain proteins(hair and porcupine quills containing protein alpha keratin, had a distance of 5.15-5.2A in them
with basically that info, came up with model Figure 4-4
- Right handed helix
- 3.6 residues/turn
- 5.4 A / turn
- Stabilized by backbone H bonds
- Side chain stuck away from axis
- In Ramachandran plot Phi -45 to -50, psi = -60
  - So in allowed region

We Now know that about 1/4 of all protein structure is alpha helix, but in any given protein can vary between 0 and 100%

B. Sequence and helical stability

One big question is can you predict structure from sequence
Alpha helix, so have been trying to predict from sequence for about 20 years
- 80-90% can get it correctly. Here are some of the reasons
- Since side chains of AA stick out, room for most AA's in helix but:
  - Proline Rigid ring, no H for H- bond makes a kink
  - Glycine - less structurally constrained so tend to break structure
- Often structural interactions over 3-4 residues are key

Poly Glu and lys at low pH and high pH
+ and negative AA's make ion pair Figure 4-5
Aromatic residues make hydrophobic interaction
Also end to end influences
Each AA has a dipole
All dipoles add up
Net dipole for helix + toward amino, - toward COOH
So put - residues at beginning and + toward COOH end

C. The beta conformation

Beta Sheet originally found in beta keratins - silk, spider webs
This is another structure that Pauling and Corey correctly predicted
structure is extended and H bonds form from one strand to another
To from pleated sheet

strand can be parallel or antiparallel Figure 4-6
Both cases side chains with lots of room above and below
Just a little difference in linearity of H bonds

Ramachandran is upper left (figure 4-8a)

D. Beta Turns

Globular proteins, by def, globular in shape
Peptide chain needs to turn a corner at end of structure
A variety of turns and loop
beta turns are most common and most regular

180° turn accomplished in 4 residues
H bonds forms between CO of residue 1 and nH of residue 4
6 slightly different types, but type I&II Figure 4-7 most common
glycine is frequently found in these turns because so small is easier to fit
proline also found because proline can have a cis- bond which also fits better in turn than normal trans peptide bond

E. Omega Loops

F. Secondary structure characteristics

each particular structure has characteristic phi-psi can examine on Ramachandran plot to see where real structure lie Figure 4-8b
Glycine only one that doesn't follow why???

As mentioned with helix, each structure has a preference for certain AA's, sometimes in certain positions. Can define a probability of finding a specfic AA in a particular structure

Not all biases are understood
Used to try to predict if when a sequence is in one structure or another
(Will try this in lab)
Has been basis of methods used to try to predict protein structure from sequence (See previous chapter)

Only slightly successful

G. Secondary Structure can be assessed by Circular Dichroism

Circular Dichroism - CD we have such a device
Difference in absorption between right and left circularly polarized light
Can use to probe conformation of a protein (See figure 4-9)

4.3 Protein 3° & 4° structure

3° 3D fold of chain overall structure of a single peptide chain
4° how multiple proteins interact in 3D in multi protein complexes

In looking at protein structure useful to classify proteins into 2 classes

Globular - chain folded into a globular shape - usually water soluble
Fibrous - chains folded into fiber or sheets - usually not soluble in water
Other traits

Fibrous - usually only 1 type of secondary structure, used for support, shape, external protection in vertebrates

Globular - often >1 secondary structure, includes most enzymes and regulatory proteins

A. Fibrous proteins Table 4-2

include a keratin, collagen, silk fibroin
used to give strength and/or flexibility - hydrophobic inside and outside for interaction with other proteins
In large supramolecular structures

I. Alpha keratin

Strength - hair-wool, nails, claws, quill horn, hooves, outer layer of skin
Part of broad family of intermediate filament (IF) proteins
Other IF proteins in cytoskeletons of cells

Right handed alpha helix
But two coils around each other like a rope (4° structure)
Gives more strength
This reason didn't quite fit X-ray data
Twisted to left

Has hydrophobics in interface to hold together
Also R groups mesh and interlock for added strength
Overall organization for hair given in 4-10
Other organization for hoof and horn
Added strength using disulfides at various levels in structure
Used for a permanent in hair styling Will smell that smell again!

II. Collagen Figure 4-11

Also for strength
But keratin could be stretched because can pull helix from helix to extended
Collagen is strength that cannot be stretched
Used for tendons & cartilage
Left handed helix
But has 3 extended strand of protein so can't stretch
In his triple helix the 3rd residue of each strand in on close contact on inside
This is usually a gly because so tight
Also need pro or Hypro for kink to make helix
Net 35% gly, 21% pro & Hypro, 11% ala

Typical collage molecule is 3 strand of 1000 AA
Gives a size of 3,000 b 15A
Has a strength > steel wire (can you figure out why)
In collagen fibril stagger collagen molecules and crosslink together (Figure 4-12)
Crosslinks are unusual Lys-lys crosslink
Text page 128
Each tissue has different need, so different amount of crosslink, different overall between fibrils

As get older get more crosslinks, gets more rigid & brittle
Some genetic diseases involve mutation where gly is replaced by another AA
Collagen and specific connective tissues poorly formed
Can be fatal

III. Silk Fiberoin Figure 4-13

Mostly Ala & Gly in antiparallel sheet
Previous editions Alas lock together on one side and gly on other
This edition - Ala gly alternate on a side for interdigitation
Does not stretch because extended
Flexible because sheet held by van der Waals interaction instead of covalent crosslinks

B. Globular proteins

structure backbone fold back on itself, to make roughly globular shape
incredible diversity because most if not all enzymes and regulatory proteins are globular
Known protein structure in 1,000' and doubling every 2 years
(Check in PDB to see what it is this week)
Was 22,611 in 2003, 32,727 in 2005
As get bigger and bigger data base now recognizing many common structural features
Can see genetic relationship in 3D structure

C. Myoglobin (an example) Figure 4-15

One of the first determined
Kendrew in the 1950's
O_2 binding protein of muscle
153 AA
Single iron protoporphyrin ring (heme group)
Is closely related to hemoglobin
Look at C. Look at surface can't see much, just a bunch of atoms
For visualization get rid of atoms, look at main chain, use spirals to emphasize helix (will see flattened arrows for strands of sheets)
Now can see is helical, can see turns at end of helixes
(70% helical , 8 helices total)
Can see how wrap around heme
Look at D
Color hydrophobic blue
Can see are mostly in interior
Confirmed in e where show surface again
Almost as dense as an organic crystal
London forces holding together
Other features observed that follow our earlier discussion
4 pros, 3 at turns, 1 is a kink in a helix
Bend tend to have ser thr + asn
Incompatible with helix
Have extra H bond for end or start helices or interact with water
Binding of O_2
Figure 4-16
Look at heme, polar or nonpolar?
Held noncovalently in hydrophobic pocket
One side of Fe coordinated to His 93
Other side coordinated to O_2
Fe is in +2 state
In water, in presence of O_2 would oxidize to Fe^{3+}
Reason for hydrophobic pocket it to prevent this

(Fe^{3+} does not bind O_2)

D. Others

There are now thousand of proteins, could do detailed analysis of each structure

Lets look for general ideas

Pages 136 & 137 for a few

All have hydrophobic interiors

Smaller proteins -harder to hide hydrophobic- less stable?

Tend to have more disulfides

Amount of helix or sheet varies (see table 4-3)

Ionic bonds and H bonds are used as expected

E. Common structural patterns

In large proteins tend to see structural pieces over and over

Supesecondary structure, or motifs, or folds

See figure 4-21 for some examples

Can often see protein fold into two or more stable globular units

These are called **domains** See figure 4-18

Domain not same as structural motif

Will fold into correct structure even when separated from rest of protein

Study both fold and domains and protein several rules

1. As expected, hydrophobic interaction is dominant core
   Need at least 2 layer of structure to form a hydrophobic core
   One on each side, hydrophobic sealed in middle
2. Helices and sheets tend to be in different layers regions
   H-bond patterns not compatible
3. Usually if close in sequence close in 3D structure
4. Tight crossing or knots not observed
5. In beta conformation there is usually a twist
   Twist is important in overall structure

F. Protein motifs are basis for structural classification

As observe same motifs over and over can begin to pick out in new proteins and use to predict structure in novel proteins

SCOP data base (may use in lab??)

Structural Classification Of Proteins

1st divided into 4 major classes

Alpha, beta, A+B (separate domains)

A/B (mixed in a single domain)

Then divided and subdivide like genetic tree

Thought to be less that 1000 folds

Protein with similar sequence or function are can be considered part of a **family** this family is usually related through evolution

Sometimes two families with similar structure but totally different sequence or function . Then called a **super family**

Hard to say if revolutionarily related but so far back sequence has diverged or true convergent evolution

G. Quaternary structure

- Many proteins multiple peptide subunits
  - Many times different units used for different functions
  - Often used in regulatory role
    - Binding of ligand to one subunit
    - Changes structure
      - That changes structure around it
      - That changes activity
      - First oligomeric protein determined was hemoglobin
        - 2alpha and 2 beta
  - Multi subunit protein called a multimer
  - Individual unit called a monomer or protomer
    - Although protomer can also refer to a group of identical proteins in a larger multimer
  - Can be a few proteins or 100's or proteins
    - If only a few called a oligomer
  - If monomer are different, get unique, complicates structure
  - If monomer are identical usually get symmetric structures
    - 2 main symmetries
      - Rotational arranges about a plane or sphere
        - Several kinds, two commonly observed in proteins
          - Cyclic
          - Dihedral (see figure 4-23)
      - Helical proceeds down an axis

H. Limits to size of proteins

- Low end, need a hydrophobic core
- High end
  - Practical observation, most >100,000 (>1,000 AA's) are multimers
  - Why?
    - More efficient to make many copies of a smaller protein than make one copy of a large protein
    - Synthesis error about 1/10,000
    - So as make large protein more change to make it wrong

I. How do we know what protein structures are? BOX 4-5

X-ray crystallography

In gen chem talked about small molecules, how can shine X-rays at crystals, and see spots , spots related to distance between atoms

Proteins same idea but at another level
need crystals of proteins to shine X-ray at

get pattern in 2D (single photoplate) actually in 3D(rotate around crystal)
distance between bright spots tells about lattice
intensify of spots tells about electron density inside crystal
Fit model to density
make pretty picture
(Figure 1 page 132-133)

NMR

^{1}H NMR
hundred of resonance
sort out with 2D NMR
Hit twice with E, see how E flows from one resonance to another
E can flow along bond or through space
Use E flow along bond to identify
Use E flow through space to figure out who is close to whom
Build models with constraints
FIGURE 2 and 3 page 133 &134 or bring in

Two techniques somewhat complementary
NMR solution, X-ray crystal
NMR smaller proteins X-ray all sizes

4.4 Protein Denaturation and Folding

Proteins made on ribosome in linear sequence
following sequence must fold to native structure
native structure is only marginally stable
small changes can disrupt easily
lets explore transition between folded and unfolded

A. Loss of Structure yields loss of function

protein evolved to work in a particular environment
mess with that environment and you mess up protein structure
loss of 3D structure sufficient to cause loss of function called **denaturation**
Denature does not always mean completely randomized structure

may be only a small piece out of place may be part of a family of slightly unfolded structures

One way is heat
as heat up, make things jiggle, tend to break H bonds
se protein lose structure over a very narrow rang in T
cooperative - lose structure in one part and it all blows apart
are proteins that are heat stable (special for PCR reactions)
    Look at sequence can't tell why yet
    Can be very lucrative (enzymes in detergents)
Also use pH extremes, miscible organic solvents, urea and guanidine HCL and detergents
    Organic solvents, detergents, urea disrupt hydrophobic core
    pH changes in ionization and Charge/charge interactions
thus denatures state NOT equivalent

B. AA sequence determine structure
Said this before, how do we know?
often the above denaturation is completely reversible
    (Thus don't need cell to rebuild)
1st demonstration by Anfinsen in 1950's with Ribonuclease
    This contains 4 disulfides. Even when disulfides removed, still came back to proper structure (figure 4-26)

C. Peptides Fold Rapidly by a stepwise- Process
folding in cells extremely rapid
E coli makes a 100 residue protein in 5s
if did this randomly say each AA has 10 different conformations
    10*10... 10^{100}
    Shortest time to sample a conformation 10^{-13}
    Make total time to structure 10^{87}s or 10^{77}year!!
    So protein can't be try all possible structure
    Called Levinthal paradox after Levinthal who pointed this out in 1968

So folding must follow some pathway
    This is current research
    1 theory
        Hierachic folding
        Helices-sheets first
        Then fold together for higher structure
        Then for higher
    Theory 2
        Hydrophobic collapse
        Hydrophobic interactions drives to compact

But not native so have 'molten globule'
MB may contain secondary structure, but not necessarily in right place
Then reshuffles in collapsed state to find final state

Most likely explanation is somewhere in between, with lots of different intermediates
thought of as a funnel process (figure 4-28) where thousands of structures are whittled down to 1 lowering E every step of the way

D. Some proteins undergo assisted folding in cell

Not all proteins fold correctly in cell
is special mechanism to help fold
action of special proteins **Molecular chaperones**
MC. Interact with misfolded or partially folded proteins and correct folding
Found in organisms from bacteria to humans
2 classes
Hsp70 a family of chaperonens
About 70,000 MW
Abundant in cells stressed by high temp (Heat Shock Proteins)
Figure 4-29
Bind to exposed hydrophobic regions of
Misfolded proteins
Protein still on the ribosome and actively folding
Protein with hydrophobic that are going to a membrane
Prevents aggregation of these hydrophobic with other hydrophobic that would start denaturation and precipitation
Other names DNA K DNA J in E coli
chaperonins
Figure 4-30
Large molecular complex to hold protein in till it folds right
10-15% of all E coli proteins need this
Up to 30% when heat stressed
Proper folding of proteins is linked to large changes of structure in chaperonin proteins and hydrolysis of ATP, but actual mechanism is still unknown

Two other proteins known to help folding
Protein disulfide isomerase (PDI)
Shuffles incorrect disulfides
Peptide prolyl cis-trans isomerase (PPI)
Changes between cis and trans proline
Can be a slow step in some protein folding

E. Folding defects my be responsible for may human genetic disorders

Single change in DNA sequence makes single change in protein sequence

Protein doesn't fold correctly

See box 4-6 for some further insight (kuru, Mad Cow, Althizmiers)

Audrey Gable Creutzfeldt-Jakob