

Lecture Notes - 3
7.24/7.88J/5.48J
The Protein Folding Problem

- **Reprise RNase refolding**
- **General Features of Globular proteins**
- **Interior Packing**

Rnase Refolding

We can cartoon the experiments:

Assume that reduced form is unfolded and populates an ensemble of unfolded states in 8M urea:

[Unfolded] \longleftrightarrow [Intermediate] \longleftrightarrow [Native]

[U]

Rapidly exchanging statistical ensemble of random coils

[I]

???

[N]

Native fold, \$ S-S bonds, 1 out of 105 possible sets of S-S bonds

[Aggregated]

Non-native (no enzymatic activity), non-native (scrambled) disulfide bonds.

At the end of the reaction we have native state

Soluble inactive – misfolded or perhaps small oligomers

Precipitated: Anfinsen showed that these were S-S bonded network = he called “scrambled”

Critical question is nature of the intermediates as go from here to there:

If there was a sequence in which these bonds formed, and one could figure out the sequence or pathway, it might give you considerable information on the steps in the folding pathway.

So certainly we need to understand native-fold, and interactions that determine it, stabilize it once formed;

However this does not solve problem

We will return to the question after reviewing the general anatomy of globular proteins, which will occupy most of next week.

General features of 3-D structures of solved proteins; Globular, Soluble proteins:

A. Isolation and Crystallization

The first crystals of proteins sufficiently large and ordered to diffract X-rays were prepared of the digestive enzyme pepsin:

Bernal, J. D., and Dorothy Crowfoot. Nature (1934): 133, 794.

This was followed by crystals of insulin, Lactalbumin, hemoglobin, and chymotrypsin:

Bernal, Fankuchen, and Perutz (1938)

The rise to power of the Nazis in Germany and Italy and the outbreak of war brought all these studies to a halt. Bernal led the world scientific community into opposition to the fascists. This was not only with respect to their persecution of Jewish scientists, but that the notion of Aryan science different from the general growth of knowledge, represented a broader attack and the growth and communication of knowledge.

The rebuilding and repopulating of the civilian biomedical infrastructure took about ten years, and re-launched in the 1950's. In England - British Medical Research Council, in US - National Institutes of Health.

Solution of Myoglobin structure by John Kendrew in 1958 and hemoglobin structure in 1962 by Max Perutz - it became clear that alpha helix actually present in globular proteins. With the solution of the lysozyme structure by David Phillips (1966), revealed that anti-parallel beta-pleated sheet also present in globular proteins.

B. Review Criteria (Biases) for the selection of proteins for crystallization and perhaps into the success or failure of crystallization

1. Available in large quantities - usually extracellular protein from extracellular fluids
 - Hemoglobin from blood
 - Myoglobin from muscle
 - Lysozyme from egg white/ tears??
 - Ribonuclease from stomach??
 - Insulin from pancreas??
 - Lactalbumin from milk ??
 - Chymotrypsin; from saliva, stomach??

2. Stable during purification

- Resistant to Residual proteases
- Does not aggregate
- Resistant to Cold denaturation

3. Very soluble in aqueous buffers

This excludes:

- Fibrous proteins; evolved to interact with each other to form structures; keratin, collagen
- Other Structural proteins; ribosome structural proteins, histones, actins, tubulins
- Membrane proteins

4. Must crystallize

- Proteins which behave differently than the dogma; take multiple conformations in solution

Protein database sharply under-represents – fibrous proteins, structural proteins, membrane proteins

C. Structures whose atoms have been located at high resolution

- Globular aqueous soluble: 1,100 domains
- (- Membrane: 12)
- (- Fibrous: 5)

This is the class of proteins for which single crystals have been obtained. Clearly the most extensive database is for proteins soluble in aqueous solutions and which form crystals.

D. General Features of Crystallized Globular Proteins

1. Surface/Volume Relationships

Globular proteins have relatively sharply defined:

- Exterior Surface = H₂O Interface
- Interior volume - Is the interior of a native protein crystalline or oily?

A quantitative answer to this was given by Richards:

"The Interpretation of Protein Structures: Total Volumes, Group Volumes and Packing Density" Frederic M. Richards (1974) J. Mol. Biol., 82, 1 - 14.

Problem comes down to calculating how densely packed atoms are in protein molecules?

We can define the **packing density** of a molecule as:

Volume of its van der Waals envelopes of its constituent atoms / actual volume of the intact folded molecule

This value is dimensionless

- Close packed hard spheres = 0.74

- Close packed infinitely long cylinders = 0.91

- Crystals of various small organic molecules 0.70 - 0.78 values (<0.6 very rare)

The van der Waals radii - from interatomic distances of nonbonded atoms in organic crystals.

However calculation of the volume of irregular structures not trivial. Richards' procedure made use of a geometrical construct called **Voronoi polyhedra**.

- Draw vector from each atom to closest surrounding atoms.
- Draw a plane through it corresponding to intersection of van der waals radii. This will define a polygon.
- When you do this for all atoms space is completely filled.
- Thus obtain geometrically legitimate volume. Can compute this from atomic coordinates.

For the following solved proteins, the average packing density:

- Hen egg white lysozyme = 0.75
- Ribonuclease S = 0.74

Conclusion: The interior of globular proteins is as tightly packed as are amino acids in organic crystals.

Subsequent investigations have supported this general result for that set of proteins included in crystallized set.

Counter argument; rigidity necessary for crystallization is only found in those proteins which are continuously dense and lack hinges.

2. Are there interior cavities or entrained solvent bobs?

Careful examination of a larger set of proteins was carried out by

A.A. Rashin, Michael Iofin, and Barry Honig (1986) "Internal Cavities and Buried Waters in Globular Proteins" Biochemistry, 25, 3619-3625.

These procedures involve in general:

- Construction of Van der Waals radii;
- Find places where these radii not buried or in contact with other radii;
- Use a spherical probe (water model = sphere radius 1.4Å) walk along van der Waals radii with smaller and smaller probe to find places that can walk along

Accessible surface area is defined by center of water molecule rolling over van der Waals surface. Thus there exists inaccessible surface area.

Gives an estimate of volume of cavity and accessible surface area.

Protein	MW	#	VolA3	# H ₂ O
Pancreatic Trypsin Inhibitor	6,000	2	70	4
cytochrome c	11,000	5	34	2
ribonuclease	13,000	5	41	1
egg lysozyme	14,000	8	190	5
erythrocrucorin	14,000	9	401	0
Myoglobin	17,000	23	391	1
trypsin	23,000	13	494	17
chymotrypsin	26,000	26	571	10

- Cavities = less than 2% of the volumes of these proteins.
- No single cavity can host a probe of 2Å or more.
- Almost all cavities become connected to protein surface as probe diameter goes down to 0.4 Å.

What does this suggest?

Suggests that these represent channels or hinges.

Volume of internal water molecule for larger cavities is about 30Å

Van der waals radius of water is about 1.4 Å

Some cavities which appear non-polar can contain a water molecule forming three hydrogen bonds. This is because the hydrogen bond can be longer than the sum of the van der waals radius of a polar atom and a probe., and thus some polar atoms with which water forms hydrogen bonds may not contribute to the accessible surface.

Note: those cavities that seem to be empty could be contributing bound water, but if water crystallographically not fixed, smeared out.

3. Functional Significance of small cavities

Largest empty cavity in their sample located between helices B, E, and G, between two domains of the globin.

Conserved in six other myoglobin type proteins (Very diverse). May be important for relative motions of helices.

More than half of cavities in multi-domain proteins are at domain interfaces.

He considers similar, linkage explanation, rather than packing problem.

Prediction; Those sequences that have evolved can close pack?

Question: WHAT FRACTION OF amino acid sequences is there no close packed globular structure?

For retinal containing pigments models suggest that cavity of 100Å would allow for rotation of chromophore.

Further studies carried out by:

"Interior and Surface of Monomeric Proteins" Susan Miller, Joel Janin Arthur Lesk and Cyrus Chothia (1987) J. Mol. Biol., 196, 641-656.

For example lysines and arginines almost always at surface (<4% buried)

4. Additional Features of Inside Outside

General Features of Globular proteins

- No Knots
- More than half of the polypeptide sequence is organized into regular secondary structure, predominantly alpha helices and beta sheets
- Interiors are densely packed - Extremely rare to bury bulk solvent
- Buried charges very rare
 - In rare cases, ion pair, metal complex, or active site
- Interiors are hydrophobic (carbonyls and amides are H-bonded)
- A large fraction of the accessible surface of the extended polypeptide chain is buried in the protein interior.

E. Conclusions re protein folding

- 1) Buried side chains truly removed from water: Major energetic contribution to protein folding; but driving force, as originally proposed by Walter Kauzmann, is entropic effect on water structure; hydrophobic residues cannot H bond with water; organizes water molecules with respect to own partners.
- 2) These side chains almost always hydrophobic: extremely rare to bury charged side chains; If the inside of a protein is closely packed and excludes water, becomes quite clear that in fact the inside and the outside are very different environments. Therefore the partitioning of residues into the chain between interior and surface positions may be very important in defining fold.
- 3) Contradiction: Burying side chains also requires burying backbone; but backbone has Carbonyl oxygens and amide nitrogens which are hydrophilic, polar groups;
- 4) But in actual fact, largest number of peptide bonds in proteins are part of regular secondary structures, alpha helices and beta sheets: partial charges are neutralized by maximal hydrogen bond formation:
- 5) SO naively folding processes may have at least two level of organization

Formation of secondary structure, docking against each other to form tertiary structure; Framework model:

Now clearly, given the large fraction of chain that is in regular secondary structure, a substantial fraction of the internal packing of globular proteins is the...

So at least two processes that we need to investigate:

- Formation of secondary structure;
- Next level; packing of secondary structure into tertiary structure:

Might be tightly coupled;

F. Packing of Secondary Structures

One of the determinants of the shapes of proteins and the folds is the packing of the units of secondary structure against each other.

Short amino acid sequences = peptides > At room temperature in constant motion; flickering in and out between a variety of conformations;

Homopolymers; In solution also flickering in and out between non-regular conformations and mixture of these with alpha helices;

Two levels of Packing:

- Within units of secondary structure; alpha helix, beta strands
- Between units of secondary structure- helix/helix, helix /sheet
- Packing Interactions critical for formation of native structure.