Protein Structural Models
for CHEM 641 Fall 07

Brian Bahnson
Department of Chemistry & Biochemistry
University of Delaware

Diffraction
- Light Microscope 400-700 nm
  - uses lens to focus diffracted light into an image
- X-ray crystallography uses light of 0.154 nm (1.54 Å)
  - no x-ray lens available,
  - first protein structures solved 1959 by John Kendrew
    and Max Perutz (myoglobin and hemoglobin)
- Need to know wavelength (Cu sources $\lambda = 1.54 \text{ Å}$)
- Need to measure Amplitude of Diffracted Reflections
  - Intensity is proportional to $A^2$
- Need to estimate the relative phase – refinement leads to correct values and a finished structure

Growth of Protein Crystals:
by Vapor Diffusion

- Hanging drop method
- Fast
- Simple
- Inexpensive
  - Test for growth in a variety of conditions

Crystal Nucleation & Growth

Basic concept:
1. Concentrate solution enough so nucleation occurs in only a few cases
2. Initial growth pulls some protein out of solution
3. Reducing [protein] back into metastable range
4. Grow only a few large crystals
Factors effecting protein crystal growth

*Most important

- Ionic Strength*
- Specific Ions (i.e. Ca++)
- Protein Concentration*
- Detergents
- pH*
- Temperature*
- Time

- Monodispersion*
- Vibrations
- Pressure
- Gravity
- Purity Of Protein*
- Ligands
- Binding partners

Overview of X-ray Experiment

Protein crystal in a loop

X-ray Equipment in Delaware

RU-H3R X-ray generator

Close-up of Cryo Crystal

Zoom into a single reflection

Movie of a Data Collection
The Diffraction Condition

- Reflections are the result of constructive interference
  \[
  \sin \theta = \frac{AB}{d}
  \]
  \[
  d \sin \theta = AB
  \]
  \[
  n \lambda = 2d \sin \theta
  \]

Methods to Obtain Phase Estimates

- Multiple Isomorphous Replacement (MIR) - labeled protein with heavy atoms
- Molecular Replacement (MR) - use a structurally homologous protein, (>25% sequence identity)
- Multi-wavelength Anomalous Dispersion (MAD phasing) - the fastest way to solve a structure due to better quality initial phases

Molecular Replacement

- Molecular Replacement (MR) - another method to estimate phases
- use a structurally homologous protein
  >25% sequence identity is sometimes possible
  >50% sequence identity is a safe bet

1) Make search model
   - find structural model of sequence homolog
   - from sequence alignment and homolog structure, create model
   - mutate or trim down to what the two proteins have in common
   - energy minimize to eliminate bad geometry (intro to refinement)

2) Search model defines electron density - “FT the electron density”

\[ F_{\text{obs}} = \iint_{xy} \langle \phi(x,y) \rangle e^{i(\mathbf{k} \cdot \mathbf{r} + \theta)} dx dy \]

Structure Based Sequence Alignments – homology modeling

Multi-wavelength Anomalous Dispersion (MAD phasing)

- Se, Hg, Yb, Ho, etc…
- XAFS - X-ray absorption fine structure, measure by fluorescence intensity off of protein crystal as a function of x-ray energy

MAD phasing - continued

Example: Isocitrate Dehydrogenase (IDH)

- Dimer of 46,500 Da subunits - 9 methionines/subunit
- Selenomethionine expression

\[
\rho(x,y,z) = \frac{1}{V} \sum_{h} \sum_{k} w_{hk} |F_{\text{obs}}| e^{-2\pi i (hx + ky + \ell z - \alpha_{\text{iso}})}
\]

This would have been nicer
One dimer of IDH – MAD phased map with bones built in

Interpret map – find α-helices, β-sheets, trace chain
N→C, add side chains, and refine. This is your initial model!

Obtain $|F_{\text{calc}}|$ and $\alpha_{\text{calc}}$ from Initial Model

Initial model build in theoretical reciprocal space

$|F_{\text{calc}}|$ and $\alpha_{\text{calc}}$

X-ray Refinement

$\Phi = w_F \sum_{hkl} (|F_o| - |F_c|)^2 + w_E E_{\text{total}}$

$|F_{\text{obs}}|$ and $|F_{\text{calc}}|$ observed calculated x-ray term ideal geometry term

Observe Comparison Between $|F_{\text{calc}}|$ and $|F_{\text{obs}}|$ using Electron Density Difference Maps

Shorthand: $F_o - F_c$ difference maps

$+$ where model should be added

$-$ where model should be removed

Active Site and Complete Model of IDH

2$F_o$ – $F_c$ map

How do you know when a structure is done?

R-factor $= \frac{\sum |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum |F_{\text{obs}}|}$

Other criteria:

RMSD $< 0.01$ Å bond lengths

$< 2^\circ$ bond angles

Ramachandran plots of phi and psi

What does electron density look like at different limits of resolution?
Nuclear Magnetic Resonance (NMR)

Atoms observed in proteins: $^1$H, $^{15}$N, and $^{13}$C

- unpaired electron spin
- Spinning nucleus

Natural abundance - okay for $^1$H (99.98%), not okay for $^{15}$N (0.36%), and $^{13}$C (1.11%).

Solution - is E. coli expressed protein with uniformly labeled $^{13}$C-glucose and $^{15}$N labeled NH$_4$Cl.

Limitations - proteins < 30 KDa, must be stable with no aggregation for days, soluble to 15-50 mg/ml.

The energy gap gives rise to a slight excess of the $+1/2$ state

$\Delta E = h\nu$

$N_{\text{upper}} / N_{\text{lower}} = e^{-E/kT}$

Boltzmann Distribution

Atoms observed in proteins - $^1$H, $^{15}$N, and $^{13}$C

The energy gap gives rise to a slight excess of the $+1/2$ state

$\Delta E = h\nu$

$N_{\text{upper}} / N_{\text{lower}} = e^{-E/kT}$

Boltzmann Distribution

Definition of Chemical Shift

$\delta = \left( \nu - \nu_{\text{REF}} \right) \times 10^6 / \nu_{\text{REF}}$

units of ppm

Spin-Spin Coupling - is through bonds

a.k.a. scalar-coupling or J-coupling

$J_{A-B}$

$H_A$ $H_B$

Looking at only $H_A$

If $H_B$ not there

deshielded

$\delta$ (ppm)

NOE Effects - are through space

$\text{NOE} = \frac{1}{\tau_c} f(\tau_c)$

NOEs are highly distant dependent, seen from 2-5 Å

These distances are the constraints that allow a 3D protein structure to be built. Analogous to X-ray diffraction data.
A Protein’s $^1$H-NMR Spectra is Complex

To assign peaks of proteins, this needs to be spread into a 2nd or 3rd dimensions.

Two Papers to read if you’re interested, but don’t worry about this for CHEM 641


The Structure of Calmodulin

Crystal structure

NMR Structure

-bundle of 30 structures