Part 1
X-ray Crystallography
What happens to electron when it is hit by x-rays?

1. The electron starts vibrating with the same frequency as the x-ray beam
2. As a result, secondary beams will be scattered in all directions
Scattering from a molecule

1. Molecule is composed of many electrons
2. Each electron will scatter secondary radiation upon exposure to x-rays
3. The scattered secondary beams will interact and cause interference
4. The scattering from a molecule is dependent upon the number of and distances between electrons
5. In other words, scattering from molecule is dependent on its atomic structure
6. If we would know the amplitudes and phases of scattered molecule, we could calculate the structure of molecule...
In practice...

- Scattering from a single molecule is far too weak to be observed but possible.
- If molecules are all oriented in the same way (like in crystal), the scattering from individual molecules will be multiplied in certain directions
The unit cell and a crystal
The electron density equation

$$\rho(xyz) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} F(hkl) \exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)]$$

- $h, k, l$ – indices of reflections
- $xyz$ – coordinates
- $F$ – amplitude of reflections
- $\alpha$ – phase of reflections
- $V$ – unit cell volume
The Phase Problem

• With detector you can measure only the intensity of reflections
• The information about phases is lost – there is no such thing as “phase meter” because we have no lenses for x-rays.
• This means, you must obtain phase information in some other way
• For small molecules (<100 atoms), direct methods exist. This means, that you can calculate phases from amplitudes without any extra information.
• Proteins are far too big to use direct methods, so other tools are developed
Isomorphous replacement

- By introducing heavy atoms in protein crystal (by soaking), the diffraction pattern can be altered
- It is possible to determine positions of heavy atoms and from them the phases
- One must use at least 2 different heavy atom soaks
- Problems:
  - 1) Unit cell dimensions of crystal might change upon soaking
  - 2) Crystal might get destroyed upon soaking and not diffract at all
  - 3) Heavy atom ions might not bind in well defined places
Molecular replacement

• Currently the most common technique
• Applicable only if a similar structure already exists (at least 25% sequence identity)
• The phases of known structure are combined with intensities of unknown
• Before that, the known model has to be in silico placed in an artificial unit cell in the same orientation and translation from origin as in the structure of interest
• For this, rotation and translation functions exist

Problems:
• May not work, if unknown structure is less than 30% identical to the known structure
• Model bias – what’s that?
Observed amplitudes

Known structure

Manx cat

Phases unknown!

Unknown structure

Cat

Calculated amplitudes and phases

Fourier Manx cat
Observed amplitudes, calculated phases

The tail becomes visible!

FFT
Be aware – this happens, if structures are not similar enough!!

Duck amplitudes + cat phases

Looks like a cat!!
Model building

• Fitting of protein sequence in the electron density
• Easy in molecular replacement
• More difficult if no initial model is available
• Unambiguous if resolution is high enough (better than 3.0 Å)
• Can be automated, if resolution is close to 2Å or better
Refinement

- Automated improvement of the model, so it explains the observed data better
- The phases get improved as well, so the electron density maps get better
Validation

• Assesment of the final(?) model quality
• How the geometry of amino acids look like? (Ramachandran plot)
• Are non-covalently atoms far enough from each other? (no atom bumps)
• Are residues “happy” in their environment? (hydrophobic in core, polar on surface)
• Are the hydrogen donors/acceptors satisfied?
Part 2
NMR Structure Determination
NMR Interactions

**NOE**
- a through space correlation (<5 Å)
- distance constraint

**Coupling Constant (J)**
- through bond correlation
- dihedral angle constraint

**Chemical Shift (δ)**
- very sensitive to local changes in environment
- dihedral angle constraint

**Dipolar coupling constants (D)**
- bond vector orientation relative to magnetic field
- alignment with bicelles or viruses
COSY- COrrelation Spectroscopy

[Diagram of COSY experiment with time axis (t) and frequency axes (ω₁ and ω₂)]

- ω₁ (¹⁵N) in ppm
- ω₂ (¹H) in ppm
- ¹H and ¹³C axes

A: ²H and ¹⁵N planes
B: ¹H and ¹³C planes
**NOE- Nuclear Overhauser Effect**

**Nuclear Overhauser Effect (NOE, $\eta$)** – the change in intensity of an NMR resonance when the transition of another is perturbed, usually by **saturation**.

$$\eta_i = \frac{(I-I_o)}{I_o}$$

where $I_o$ is thermal equilibrium intensity.

**Saturation** – elimination of a population difference between transitions (irradiating one transition with a weak RF field)

Populations and energy levels of a homonuclear AX system (large chemical shift difference)

Observed signals only occur from single-quantum transitions
2D NOESY (Nuclear Overhauser Effect)

Relative magnitude of the cross-peak is related to the distance ($1/r^6$) between the protons ($\geq 5\AA$).

NOE is a relaxation factor that builds up during the “mixing-time (tm)"

![Diagram of 2D NOESY experiment](image)
NMR Structure Determination

Going from NOESY Data to a structure

2D NOESY Spectra at 900 MHz
Once we complete the Assignment of All the Remaining NOEs in an Iterative Process to Obtain the Structure

**Protein Structure Refinement**

**Iterative Cycle**
- Distance Constraints Assignments
- Stereospecific Assignments
- Torsion-Angle Assignments
- 3D Structure Determination

**More Constraints the Better the Structure**
Two Very Important Facts to Remember

- NOEs Reflect the **Average** Distance
- Protein Structures Are **Dynamic**

In reality, protein undergoes wide-ranges of motions (snapshots of 100 BPTI conformations)

We visualize protein structures as a static image
Improving the Quality of NMR Structures

• Stereospecific Assignments
  - Making stereospecific assignments increase the relative number of distance constraints while also tightening the upper bounds of the constraints
  - There is a direct correlation between the quality of the NMR structure and the number of distance constraints
    - more constraints $\rightarrow$ higher the precision of the structure

Increasing Number of NOE Based Constraints
Part 3
Structure Determination by Electron Microscopy
References and other useful material

• Texts
  – Biophysical Electron Microscopy: Basic Concepts and Modern Techniques by U. Valdre (Editor), Peter W. Hawkes (Editor)
  – Three-Dimensional Electron Microscopy of Macromolecular Assemblies by Joachim Frank
  – Negative Staining and Cryoelectron Microscopy: The Thin Film Techniques by Robin J. Harris, James R. Harris

• Reviews

• Web
  – [http://ncmi bcm.tmc.edu/%7Estevel/spintro/siframes.htm](http://ncmi bcm.tmc.edu/%7Estevel/spintro/siframes.htm)
  – [http://cryoem.berkeley.edu/~nieder/em_for_dummies/](http://cryoem.berkeley.edu/~nieder/em_for_dummies/)
Thermionic emission (Shottkey effect)
Electron lenses and “optical path”
Why use electrons....Part 1
### Why use electrons...Part 2

<table>
<thead>
<tr>
<th></th>
<th>Electrons 80-500keV</th>
<th>X-rays 1.5 A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio inelastic/elastic</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Mechanism of damage</td>
<td>$2^{nd}$ e$^{-}$ emission</td>
<td>Photoelectric e$^{-}$ emission</td>
</tr>
<tr>
<td>Energy per inelastic event</td>
<td>20 eV</td>
<td>8 keV</td>
</tr>
<tr>
<td>Energy per elastic event</td>
<td>60 eV</td>
<td>80 eV</td>
</tr>
<tr>
<td>Energy relative to electrons</td>
<td></td>
<td></td>
</tr>
<tr>
<td>inelastic (Compton)</td>
<td>1</td>
<td>400</td>
</tr>
<tr>
<td>elastic (Rayleigh)</td>
<td>1</td>
<td>1000</td>
</tr>
</tbody>
</table>
Negative Stain and Cryo

• Negative stain (usually 0.5% uranyl acetate)
  – Easy to prepare
  – Good contrast
  – Preservation
  – Sample distortion
  – Resolution limited to about 20 angstroms

• Cryo
  – Difficult sample prep
  – Low contrast
  – Best preservation and therefore resolution
Negative staining

Bob Horne (Cambridge)
Cryo prep using holey film

H. Fernandez-moran
B. Glaeser
K. Taylor
J. Dubochet

Aaron Klug
Flash freeze in liquid ethane
Samples

- **Single Particles (Proteins, Ribosome)**
  - No crystallization
  - Weak amplitude, no diffraction, alignment ambiguity, particle flexibility
  - ~7 angstroms

- **Fibers and filaments (tubulin, collagen)**
  - No crystallization, 2D distortion corrections, phase restrictions
  - Weak amplitude, no diffraction
  - ~9 angstroms

- **2D crystals (BR, AQP, LHCII)**
  - Diffraction amplitudes, 2D distortion corrections, crystallographic methods
  - Crystallization, many tilts required, anisotropic data
  - ~3 angstroms

- **Tubular crystals (AchR, Ca^{++}-ATPase)**
  - Crystallization, No diffraction
  - Isotropic data, 3D distortion corrections, phase restrictions
  - ~5 angstroms
Single particles

- Applicable to any protein or protein complex > 50kD
- Most common sample
- Number of software suites available
- Resolution ~9Å (<7 with symmetry)
Fibers and filaments

DNA, collagen, etc
2D Xtals

Henderson and Unwin
Tubular crystals

Rolled 2D xtal
Tubular xtal versus 2D or 3D xtal

Multiple Orientations

Single Orientation
Data collection
Image recording

• Film
  – High density content (~20kx16k pixels)
  – Slow (development time, drying)
  – Requires digitization (scanning takes hours)

• CCD
  – Low density content (4kx4k pixels)
  – Fast (ms to sec)
  – Direct digital
Processing data

- **Single Particles (Proteins, Ribosome)**
  - Pick particles
  - Align
  - Classify, average and reconstruction

- **Fibers and filaments (tubulin, collagen)**
  - Pick segments determine symmetry
  - Align/rotate
  - Average

- **2D crystals (BR, AQP, LHCII)**
  - Process images to achieve phases
  - Process diffraction data for amplitudes
  - Combine and refine as in X-ray

- **Tubular crystals (AchR, Ca^{++}-ATPase)**
  - Determine tube symmetry
  - Pick segments and distortion correction
  - Average and sum segments
Data processing 1: single particle
Mostly swiped from Steve Ludtke’s web site http://ncmi.bcm.tmc.edu/~stevel/EMAN/doc/

Single particle reconstruction

...
Software

• Spider
  - http://www.wadsworth.org/spider_doc/spider/docs/
  - $1500 + $300/yr for updates, with source

• Imagic
  - http://www.imagescience.de/imagic/welcome.htm
  - Commercial package, ~ $6000/yr

• EMAN
  - http://ncmi.bcm.tmc.edu/~stevel/EMAN/doc
  - Free, complete with C++ source
The Reconstruction Process

1. Collect Data & Digitize
2. Select Particles
3. Preliminary 3D Model
4. Refine 3D Model
5. Power Spectrum/CTF Parm.
Pick particles (manual or semiauto)
Evaluate Particles

Looking for astigmatism, drift, charging etc....
Now on to the first model

• First rule of thumb...be cautious...
• How to classify particles
  – Reference free classification and alignment
  – MSA
• Application of symmetry
• Random conical tilt
Reference free classification MSA

Can we tell the symmetry a priori???
MSA......variance....(SD)^2
Random conical tilt

- Image pairs taken of the same sample with an angular tilt applied between them
- Determine particle pairs and construct reference model
Use of common lines to align different orientations

Reconstruction & the Asymmetric Triangle
Refinement - EMAN

Initial 3D Model -> Projections -> New 3D Model

Particles -> Classify Particles -> Align & Average Classes
Class Averages
Multiple rounds of refinement...
Convergence when no improvement in the alignment statistics

Final vs. x-ray
Try different symmetries

Check Other Possibilities

D5

D6

D7

D8
Data processing 2: 2D xtal
Why lattice lines?

Z dimension has an effective real space D of infinity. Hence in reciprocal space the lattice spacing is 0.
**Table 1. Electron crystallographic table**

**A. Crystallographic parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group (layer group)</td>
<td>P3 (p3)</td>
</tr>
<tr>
<td>Lattice constants</td>
<td>( a = b = 62.45 \ \text{Å}, \gamma = 120^\circ )</td>
</tr>
<tr>
<td>Thickness</td>
<td>( c = 100 \ \text{Å} ) (assumed in refinement)</td>
</tr>
<tr>
<td></td>
<td>( 70 \ \text{Å} ) (used in LATLINE)</td>
</tr>
</tbody>
</table>

**B. Electron diffraction (amplitude information)**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of diffraction patterns</td>
<td>339</td>
</tr>
<tr>
<td>Resolution limit used (Å)</td>
<td>3.0</td>
</tr>
<tr>
<td>Maximum tilt angle (deg.)</td>
<td>70.6</td>
</tr>
<tr>
<td>No. of observed reflections</td>
<td>110,812</td>
</tr>
<tr>
<td>Friedel R-factor (%)</td>
<td>17.6</td>
</tr>
</tbody>
</table>

**C. Electron microscopy (phase information)**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of images</td>
<td>181</td>
</tr>
<tr>
<td>Resolution limit used (Å)</td>
<td>3.0</td>
</tr>
<tr>
<td>Maximum tilt angle (deg.)</td>
<td>61.2</td>
</tr>
<tr>
<td>No. of observed reflections</td>
<td>25,225</td>
</tr>
</tbody>
</table>

**D. Merged data**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>3.0</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>6892 (with amplitudes and phases)</td>
</tr>
<tr>
<td>Merging R-factor (%)</td>
<td>31.3</td>
</tr>
<tr>
<td>Phase residual (deg.)</td>
<td>46.8</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>78.4</td>
</tr>
</tbody>
</table>

**E. Refinement**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of degrees of freedom</td>
<td>6672 (without hydrogen atoms)</td>
</tr>
<tr>
<td>R-factor (%)</td>
<td>23.7</td>
</tr>
<tr>
<td>Free R-factor (%)</td>
<td>33.0</td>
</tr>
<tr>
<td>Phase residual (deg.)</td>
<td>54.4</td>
</tr>
<tr>
<td>Free phase residual (deg.)</td>
<td>63.3</td>
</tr>
</tbody>
</table>
A well refined EM map
Resolution and Resolvability

• Single particles, filaments, tubes
  – FSC
    • Which criteria to use (0.5 or 3 sigma)

• 2D xtals diffraction (like X-ray)
  – But anisotropy or point spread function
Point spread function
Resolution vs Resolvability

• Resolution is a calculated value
  – FSC or measured amplitudes above a certain sigma value.

• Resolvability is a perceived value
  – What can I see in the map
    • Is a 4 angstroms map really 4 angstroms if one cannot discern beta sheet structure?