An Introduction to X-Ray Crystallography of Small Molecules and Biological Macromolecules

Single Crystals → X-Ray Diffraction → 3D-Structure of paclitaxel

Burke Group Literature Seminar
Nagarjuna Palyam

20th July 2012
Spectroscopy in Structure Elucidation

All regions are defined as much by the mechanism of the transitions by the frequency or energy (e.g. outer shell electron excitation for UV-VIS & Scattering for XRD)
<table>
<thead>
<tr>
<th>Spectroscopy in Structure Elucidation of Small and Biomolecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Nuclear magnetic resonance (NMR)</td>
</tr>
<tr>
<td>2) X-ray crystallography (XRD)</td>
</tr>
<tr>
<td>3) Mass Spectroscopy (MS)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>NMR</strong></th>
<th><strong>X-ray diffraction</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactions of external magnetic field with the intrinsic magnetic properties of atomic nuclei with a spin angular momentum (Quantum mechanics)</td>
<td>Interactions of X-rays with inner-shell electrons in molecules in a crystal and results in scattering (Braggs equation various theories)</td>
</tr>
<tr>
<td>Identifies the hydrogen and carbon atom environment, It explicitly determines distances between hydrogen atoms</td>
<td>XRD detects all atoms except hydrogen's (only 1e^-), implicitly determines distances between all other atoms; Better diffraction for Fluorine and above</td>
</tr>
<tr>
<td>NMR only determines distances for atoms that are close to each other</td>
<td>Crystallography is best at determining distances between atoms that are far apart</td>
</tr>
<tr>
<td>NMR structures are usually represented as an ensemble. NMR detects structural dynamics in solution 2° and 3° ) and multiple conformations</td>
<td>Crystal structures are usually represented as a single average structure. Studying of motions are not possible</td>
</tr>
<tr>
<td>Effective size limit of 35 kDa (~400 amino acids); Structures are typically less precise (error 0.5-1Å)</td>
<td>No size limit; Structures can be very precise (coordinate error ~0.1 Å); whole 3D structure</td>
</tr>
<tr>
<td>In lots of cases, where from NMR data could predict two or more possible conformations</td>
<td>X-ray crystallography requires crystals (~50% don’t grow)</td>
</tr>
</tbody>
</table>
# Why Use X-rays?

<table>
<thead>
<tr>
<th>Entry</th>
<th>Size (1 Å = 1x10(^{-10}) m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bohr radius (electron to nucleus in Hydrogen atom)</td>
<td>0.529 Å</td>
</tr>
<tr>
<td>H-H bond length</td>
<td>0.742 Å</td>
</tr>
<tr>
<td>C-C bond length</td>
<td>1.54 Å</td>
</tr>
<tr>
<td>C-H, C-O, C-N, C-Sn &amp; Br-Br bond lengths</td>
<td>1.09 Å, 1.43 Å, 1.47 Å, 2.16 Å &amp; 2.28 Å</td>
</tr>
<tr>
<td><strong>X-rays for XRD</strong></td>
<td><strong>0.5 to 1.8 Å</strong></td>
</tr>
<tr>
<td>Diameter of Water molecule</td>
<td>2.82 Å</td>
</tr>
<tr>
<td>Diameter of glucose</td>
<td>10 Å (1 nm; 10(^{-9}) m)</td>
</tr>
<tr>
<td>Diameter of DNA helix &amp; Hb molecule</td>
<td>2 nm &amp; 6 nm</td>
</tr>
<tr>
<td>Size of virus &amp; bacteria</td>
<td>75 nm &amp; 1 µm</td>
</tr>
<tr>
<td>Length of sperm cell</td>
<td>25 µm</td>
</tr>
<tr>
<td>Diameter of a point (.)</td>
<td>300 µm</td>
</tr>
</tbody>
</table>

- Lab X-ray sources use Cu K\(\alpha\)_1 radiation. \(\lambda = 1.5405\) Å or Mo K\(\alpha\)_1 radiation. \(\lambda = 0.75\) Å.  
  \(E = \hbar v = \hbar c / \lambda; \quad E_{\text{x-ray}} = 10^4 \text{ ev}\)  
  or  
- Synchrotron radiation can have wavelengths in the range 0.5 Å – 1.8 Å. *e.g.* (APS) at Argonne National Laboratory near Chicago  

X-Rays have right enough wavelengths to “see” the bonds, atoms and molecules in crystal structure
Time-line of X-ray Crystallography

1895  →  Roentgen, discovers X-rays
       Shadow of Frau Röntgen’s hand (1896)
       1st Nobel prize in physics 1901

1912  →  Laue, invented X-ray diffraction
       CuSO4 Diffraction (1912)

1915  →  Bragg determined structure of ZnS, NaCl
       by XRD and wins Nobel Prize at the
       age of 25 years (youngest till date)

1928  →  Lonsdale, structure of benzene with 6 equal sized bonds
       instead of alternating double/ single bonds

1935  →  Robertson et al. solved structure of phthalocyanins,
       first complex organic molecule by XRD

1953  →  Watson/Crick (Rosalind) structure of DNA

1962  →  Perutz & Kendrew, Hemoglobin, Myoglobin

1964  →  Dorothy C Hodgkin cholesterol (1937, penicillin (1949)
       vitamin B-12 (1957) and insulin (1969).

2000  →  Thomas Steitz, complete atomic structure of
       large ribozome subunit at 2.4 Å resolution
       Nobel in 2009  >100,000 atoms + hydrogens!
Accelerated Electrons “Scatter” Light

Why don’t protons or other nuclei scatter light?
Too small!
Diffraction from a particle and crystal

**Single particle**
- To understand diffraction we also have to consider what happens when a wave interacts with a single particle. The particle scatters the incident beam uniformly in all directions.

**Crystalline material**
- In crystalline material, the scattered beams may add together in a few directions (0 to 180°) and reinforce each other to give diffractions pattern due to interference.

...something we won't see very often (Visible Light)
Diffraction and Bragg’s equation

Scattered beams in phase, they add up

Scattered beams not in phase, they cancel each other

Path difference (BC+CD) = nλ
give constructive interference.

Path difference depends on distance between lattice planes d

Braggs Law: Net path difference is \( n\lambda = 2dsin\theta \)

Sir William Henry Bragg (1862-1942)
William Lawrence Bragg (1890-1971)

The father and son team of Sir William Henry and William Lawrence Bragg were awarded the Nobel prize for physics "for their services in the analysis of crystal structure by means of X-rays" in 1915.
**Basics of X-ray diffraction (XRD)**

**X-ray diffraction** is a phenomenon when X-ray photons bombard with electron cloud to result in a scatter pattern, constructive interference of scattered waves result in XRD.

**X-ray crystallography** is a method of determining the arrangement of atoms within a crystal, in which a beam of X-rays strikes a crystal and causes the beam of light to diffract. The resulting diffraction pattern can produce a 3-D image of the density of electrons within the crystal. From this electron density, the mean positions of the atoms in the crystal can be determined, as well as their chemical bonds, their disorder and various other information.

XRD plays a pivotal role in:
- Elucidate bonding networks interactions
- Identify stereogenic centers and differentiate isomers
- Establish packing framework and structure elucidation
Basics of X-ray Crystallography

- Crystal = trillions of copies of the object
- Diffraction (scattering) pattern — can't be focused
- Electron density map image and model
- 3-D structure

- Small molecules (< 0.83 Å resolution)
- MIDA boroantes (~0.80 Å resolution)
- < 0.5 Å resolution, can see π-orbitals!!
- Proteins/ biomolecules (< 3.0 Å resolution)

* Crystals are rapidly cooled (NOT FROZEN) to near liquid nitrogen temperature to **eliminate radiation damage** and increase resolution

- Reduced thermal vibrations, disorder, merging and scaling errors

θ  Glancing angle
2θ Diffraction angle
α  Aperture angle
nl = 2d sin θ Bragg's law
User Operated - CCD Detector

~$200K
Crystal Quality is Essential

YES

Good crystal XRD, very spherical dots

MAYBE

Marginal crystal XRD, may be twin lattice, shoulder dots

NO

Unusable crystal XRD, polycrystals, lack standard lattice

The “spots” (intensity/reflection) are actually the Fourier transform of the electron diffraction, each spot contains the structural lattice information.
XRD data processing

Structure solution/refinement

SHELXTL
XPREP
XS
XP

P 21/n
a = 23.3481(4)
b = 12.3868(2)
c = 23.8191(1)
beta = 92.756(1)

Unit cell parameters

Refinement/ Data analysis
PLATON, enCIFer
IUCr CheckCIF, publCIF

Chemistry
XL

Molecular graphics
ORTEP 3
Mercury

- ORTEP stands for Oak Ridge Thermal Ellipse Plotting program
- All information from a structure is contained in the crystallographic information file (cif)
XRD: Absolute structure of chiral compound crystals

Question before Bijovet 1951: how to correlate microscopic absolute configurations to macroscopic properties such as the sign of the optical rotation of polarized light.

Emil Fischer: relative system; assign ‘D’ configuration to (+) Glyceraldehyde. His ‘lucky’ choice was later ‘confirmed’ by calculations and physical methods.

Bijvoet used **Anomalous Dispersion (Resonant Scattering)** to Solve the Absolute Structure Problem (“Determination of the Absolute Configuration of Optically Active Compounds by Means of X-Rays” *Nature* 1951, 168, 271)

**Flack parameter x (0 and 1):** The current official method to establish the absolute configuration of a chiral molecule calls for the determination of the Flack x parameter. (Flack, H.D. (1983). *Acta Cryst.* A39, 876-881)

Options for the Absolute Structure Determination of **Light Atom Compounds (C/H/N/O atoms only):**

- Add HBr in case of tertiary N
- Co-crystallize with e.g. CBr₄
- Co-crystallize with compound with known absolute configuration
- Alternative: Statistical analysis of Bijvoet pair differences
## Crystal data & Structure refinement processing (rtf file)

<table>
<thead>
<tr>
<th>Table 1. Crystal data and structure refinement for sq2.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identification code</strong></td>
</tr>
<tr>
<td><strong>Empirical formula</strong></td>
</tr>
<tr>
<td><strong>Formula weight</strong></td>
</tr>
<tr>
<td><strong>Temperature</strong></td>
</tr>
<tr>
<td><strong>Wavelength</strong></td>
</tr>
<tr>
<td><strong>Crystal system</strong></td>
</tr>
<tr>
<td><strong>Space group</strong></td>
</tr>
<tr>
<td><strong>Unit cell dimensions</strong></td>
</tr>
<tr>
<td><strong>Volume</strong></td>
</tr>
<tr>
<td><strong>Density (calculated)</strong></td>
</tr>
<tr>
<td><strong>Absorption coefficient</strong></td>
</tr>
<tr>
<td><strong>F(000)</strong></td>
</tr>
<tr>
<td><strong>Crystal size</strong></td>
</tr>
<tr>
<td><strong>Theta range for data collection</strong></td>
</tr>
<tr>
<td><strong>Index ranges</strong></td>
</tr>
<tr>
<td><strong>Reflections collected</strong></td>
</tr>
<tr>
<td><strong>Independent reflections</strong></td>
</tr>
<tr>
<td><strong>Completeness to theta = 25.40°</strong></td>
</tr>
<tr>
<td><strong>Absorption correction</strong></td>
</tr>
<tr>
<td><strong>Max. and min. transmission</strong></td>
</tr>
<tr>
<td><strong>Refinement method</strong></td>
</tr>
<tr>
<td><strong>Data / restraints / parameters</strong></td>
</tr>
<tr>
<td><strong>Goodness-of-fit on F²</strong></td>
</tr>
<tr>
<td><strong>Final R indices [I&gt;2sigma(I)]</strong></td>
</tr>
<tr>
<td><strong>R indices (all data)</strong></td>
</tr>
<tr>
<td><strong>Largest diff. peak and hole</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Atomic coordinates (x 10⁶) and equivalent isotropic displacement parameters (Å² x 10⁶) for sq2. U(eq) is defined as one third of the trace of the orthogonalized U tensor.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>N(1)</td>
</tr>
<tr>
<td>N(2)</td>
</tr>
<tr>
<td>C(1)</td>
</tr>
<tr>
<td>C(2)</td>
</tr>
<tr>
<td>C(3)</td>
</tr>
<tr>
<td>C(4)</td>
</tr>
<tr>
<td>C(5)</td>
</tr>
<tr>
<td>C(6)</td>
</tr>
<tr>
<td>C(7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3. Bond lengths [Å] and angles [°] for sq2.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>N(1)-C(21)</td>
</tr>
<tr>
<td>N(1)-C(17)</td>
</tr>
<tr>
<td>N(2)-C(26)</td>
</tr>
<tr>
<td>N(2)-C(22)</td>
</tr>
<tr>
<td>C(1)-C(2)</td>
</tr>
<tr>
<td>C(1)-C(2)≠1</td>
</tr>
<tr>
<td>C(1)-H(1A)</td>
</tr>
<tr>
<td>C(2)-C(3)</td>
</tr>
<tr>
<td>C(2)-C(7)</td>
</tr>
<tr>
<td>C(3)-C(4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6. Torsion angles [°] for sq2.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>C(2)#1-C(1)-C(2)-C(3)</td>
</tr>
<tr>
<td>C(2)#1-C(1)-C(2)-C(7)</td>
</tr>
<tr>
<td>C(1)-C(2)-C(3)-C(4)</td>
</tr>
<tr>
<td>C(7)-C(2)-C(3)-C(4)</td>
</tr>
<tr>
<td>C(2)-C(3)-C(4)-C(5)</td>
</tr>
</tbody>
</table>
**UNIT CELL**
A unit cell can be any unit of a lattice array which when repeated in all directions, and always maintaining the same orientation in space, generates the lattice array.

**UNIT CELL TYPES and THE SEVEN CRYSTAL SYSTEMS**

- **Cubic**
  \[ a = b = c. \quad \alpha = \beta = \gamma = 90^\circ. \]

- **Tetragonal**
  \[ a = b \neq c. \quad \alpha = \beta = \gamma = 90^\circ. \]

- **Orthorhombic**
  \[ a \neq b \neq c. \quad \alpha = \beta = \gamma = 90^\circ. \]

- **Monoclinic**
  \[ a \neq b \neq c. \quad \alpha = \gamma = 90^\circ, \beta \neq 90^\circ. \]

- **Triclinic**
  \[ a \neq b \neq c. \quad \alpha \neq \beta \neq \gamma \neq 90^\circ. \]

- **Rhombohedral**
  (or Trigonal)
  \[ a = b = c. \quad \alpha = \beta = \gamma \neq 90^\circ. \]

- **Hexagonal**
  \[ a = b \neq c. \quad \alpha = \beta = 90^\circ, \gamma = 120^\circ. \]
Crystal quality: what goes in is what comes out!

- Use pure and substantial compounds,
- develop solubility profile
- Single crystals & large enough suitable for XRD (0.2 - 0.5mm in 2 of 3 dimensions)

1. Solvent choice

✓ Do
  - aim for moderate solubility
  - remember “like dissolves like”
  - use highly volatile solvents
✓ Typical solvents: MeCN, MeOH, EtOH, iPrOH, ether, DCM, EtOAc, toluene, THF
  • Avoid use of “floppy” solvents” n-pantene

2. Nucleation sites

✓ Crystallization begins at defect sites
  - scratches in glassware
  - dust or lint
  • A few sites are necessary
  • Too many will result in small crystals

3. Mechanics

✓ Crystal growth takes a steady hand!
  - re-dissolve the sample
  - knock off crystallites
  • Avoid areas prone to mechanical vibration
  • Avoid constant “check in” on your samples

4. Time

✓ Crystal growth takes time
  - reduces lattice defects and twins
  - results in larger crystals
  • Best results may appear in 2 days to 2 weeks
  • Sometimes these “rules” are broken
**Crystallization Techniques**

1. **Slow evaporation**
   - Dissolve sample to near saturation
     ▫ use solvents in which sample is only moderately soluble
   - Loosely cover vial
     ▫ 1 dram vials with holes poked in a plastic cap
   - Wait
     ▫ depends on vapor pressure of solvent
     ▫ 2 days to 2 weeks

2. **Slow cooling**
   - Dissolve sample in hot solvent
     ▫ good for material that is insoluble at room temperature
   - Cap off and allow to cool slowly
     ▫ moderate temperature with oven, heating pad, cotton wool, water bath, or a warm spot in the lab

3. **Solvent diffusion**
   - Use two solvents, S1 and S2
     ▫ material is soluble in S1 but not S2
     ▫ S2 is less dense than S1
   - Dissolve in S1 in vial, slowly add S2 to form a layer on top
   - Crystals grow at the S1-S2 interface as solvents diffuse slowly.
   - DCM/Et₂O popular combination

4. **Vapor diffusion**
   - Similar to solvent diffusion, but uses separate vials for S1 and S2
     * dissolve material in S1, in open small vial
     * place small vial in larger vial with S2 and cap off
     * must choose solvents carefully
Crystallography of Biological Macromolecules
Crystallography of Biological Macromolecules

A: Glass plate with protein solution and precipitant

B: Crystals formed

C: X-ray source and crystal with diffraction

D: Image analysis

E: Molecular structure reconstruction
Summary

SELECT A SUITABLE CRYSTAL

A

CRYSTAL SYSTEM
and
UNIT CELL DIMENSIONS

B

FULL DATA SET COLLECTION

C

BRAVAISS LATTICE

D

SPACE GROUP

E

CONSTRUCT AN
ELECTRON DENSITY MAP

F

LOCATE ATOM
POSITIONS

G

STRUCTURE
REFINEMENT

Crystal

Structure

Hemoglobin

Ribozome

Material testing

Tomography
(CAT)

XRD

1.0Å

2.5Å

3.0Å

4.0Å
## X-ray Crystallography: Useful Links

<table>
<thead>
<tr>
<th><strong>UIUC X-ray Facility</strong></th>
<th><a href="http://scs.illinois.edu/x-ray/">http://scs.illinois.edu/x-ray/</a></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reciprocal Net</strong></td>
<td><a href="http://reciprocalnet.scs.illinois.edu/recipnet/index.jsp">http://reciprocalnet.scs.illinois.edu/recipnet/index.jsp</a></td>
</tr>
<tr>
<td><strong>Cambridge Structural Database (CSD)</strong> ~600K organic samples CIF files</td>
<td><a href="http://webcsd.ccdc.cam.ac.uk/">http://webcsd.ccdc.cam.ac.uk/</a></td>
</tr>
<tr>
<td><strong>Protein Data Bank (PDB)</strong> ~25K protein structures</td>
<td><a href="http://www.rcsb.org/pdb/home/home.do">http://www.rcsb.org/pdb/home/home.do</a></td>
</tr>
</tbody>
</table>
| **XRD applets and instructional materials** | http://www.ruppweb.org/Xray/101index.html  
http://www.ruppweb.org/level1/new_cryscomp_applets.htm |
| **Crystal Growing Tips** | http://xray.chem.ufl.edu/html/crystal_growing_tips.htm  
http://imserc.chem.northwestern.edu/IMSERC/X-Ray/XRayHome.html  
| **Diffraction tutorials** | http://www.uni-wuerzburg.de/mineralogie/crystal/teaching/basic.html |
| **Structures of Life-NIH** | http://publications.nigms.nih.gov/structlife/chapter2.html |
| **Absolute structure determination** | http://www.absolutestructure.com/ |