

Handout 10: Structure Refinement & Interpretation

Completing the structure and evaluating how good
your data and model agree

Chem 6850/8850

X-ray Crystallography

Department of Chemistry & Biochemistry

cora.lind@utoledo.edu



THE UNIVERSITY OF
TOLEDO

Why you should refine a structure

- We have considered how atoms are located by Patterson, direct methods or specialized methods, as well as from Fourier difference maps
- The atomic positions extracted from these methods are close to the correct values, but very rarely in exactly the right place
- During a process called *refinement*, the starting atomic positions are optimized
 - the goal is to get an as good fit between your model and your data as possible

Possible refinement methods

- You can modify your model in order to obtain better agreement between observed and calculated Fourier maps
- Alternatively, you can modify your model to get better agreement between observed and calculated structure factor magnitudes ($|F_o|$ and $|F_c|$)
- For most small molecule structure solutions, the latter process is used

Letting the computer do the work

- **The refinement process can be described as a minimization process**
 - minimization of the difference between $|F_o|$ and $|F_c|$
 - easily automated
- **The best model is obtained by minimizing the following expression:**

$$R' = \sum_{hkl} w_{hkl} (|F_o| - k|F_c|)^2$$

- called a *least squares minimization*
- w_{hkl} are weighing factors related to the quality/reliability of the data
- k is a scale factor (inverse of scale factor that would need to be applied to $|F_o|$)

Least squares minimization

- Imagine a function F that is linearly dependent on a set of parameters, x_j
 - $F(x_1, x_2, \dots, x_n) = p_1 x_1 + p_2 x_2 + \dots + p_n x_n$
- Suppose we make m independent measurements of F for different values of x_j
 - we want to get the parameters p_j
 - if $m = n$, we just have to solve a set of simultaneous equations
 - if $m > n$, the system is overdetermined

Crystallography and least squares

- The crystallographic function $F(\mathbf{hkl})$ (with the variables: atomic coordinates and displacement parameters of each atom) is not a linear function of the model parameters

$$F(\mathbf{hkl}) = \sum_{j=1}^N f_j \exp[2\pi i(hx_j + ky_j + lz_j)]$$

- The equation cannot be solved for the correct parameters in one go
- Use an iterative process instead

Refinement iterations

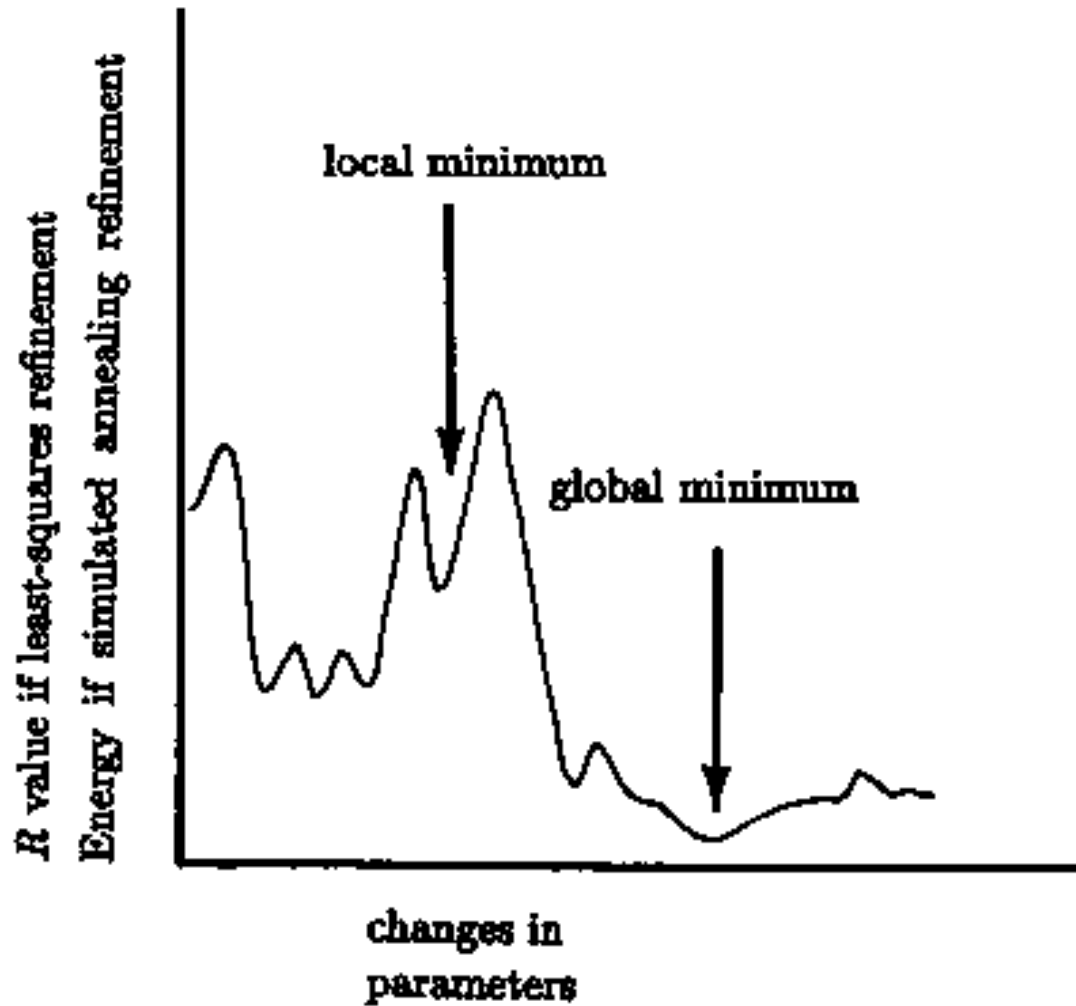
- Our goal is to minimize $R' = \sum_{hkl} w_{hkl} (|F_o| - k|F_c|)^2$
- We can use an iterative least squares approach to give parameter shifts that will lead to an improved agreement between $|F_o|$ and $|F_c|$
- The process is repeated until the suggested shifts are insignificantly small
 - usually considered to be achieved when parameter shift \ll standard deviation
 - at this minimum, the derivative of R' with respect to each parameter (x_j, y_j, z_j, B_j) should be zero



Convergence

- **A refinement process is usually considered as finished when *convergence* is reached**
 - when all parameter shifts \ll standard deviation
- **However, the process only works well if the starting model is sufficiently good**
 - many local minima in which the refinement could “get stuck”
- **Other methods offer better chances to find the absolute minimum from a bad starting model**
 - simulated annealing, random walk...

False minima



Errors

- **The errors of all observations are included in the refinement process via the weight factors**
 - weak or uncertain reflections will have less weight than strong reflections
- **The least squares output will provide esds (estimated standard deviations) for all refined parameters**
- **Watch out for high correlation coefficients**
 - correlation coefficients tell you whether two model parameters really are independent



Judging the refinement

- Statistical values are used to judge the goodness of a refinement
- The level of agreement between observed and calculated structure factors is often indicated by R factors and Goodness of Fit (GooF) values

$$R_F = \frac{\sum_{hkl} (|F_o(hkl)| - k|F_c(hkl)|)}{\sum_{hkl} |F_o(hkl)|}$$

$$wR_{F^2} = \left\{ \frac{\sum_{hkl} [w(|F_o(hkl)|^2 - |kF_c(hkl)|^2)^2]}{\sum_{hkl} [w|F_o(hkl)|^2]} \right\}^{1/2}$$

$$GooF = S = \left\{ \frac{\sum_{hkl} [w(|F_o(hkl)|^2 - k|F_c(hkl)|^2)^2]}{(n - p)} \right\}^{1/2}$$

- for n observations and p parameters



R factors

- For a good small molecule refinement, the final R_F values are expected to be $\sim 0.02-0.08$
- Placing random atoms in a unit cell is expected to give R factors of 0.83 and 0.59 for centric and acentric space groups, respectively
- Obtaining $R_F < 0.20$ usually means that the structural model has no major errors in it

Refinement strategy

- **Unit cell constants need to be refined**
 - original indexing just gives approximate values
- **Atomic positions obtained by Patterson searches, direct methods or other approaches should be refined**
 - look at interatomic distances to judge whether refined positions make sense
 - atoms in special positions cannot move in all directions
- **Atomic displacement parameters should be varied**
 - can indicate wrong atomic weight: $Z_{\text{model}} > (<) Z_{\text{real}}$ leads to large (small) ADPs
 - isotropic overall temperature factors can be obtained from Wilson plots as a first approximation

Estimating the overall temperature factor

- We know that observed structure factor magnitudes are smaller than real values because of thermal motion and scaling issues

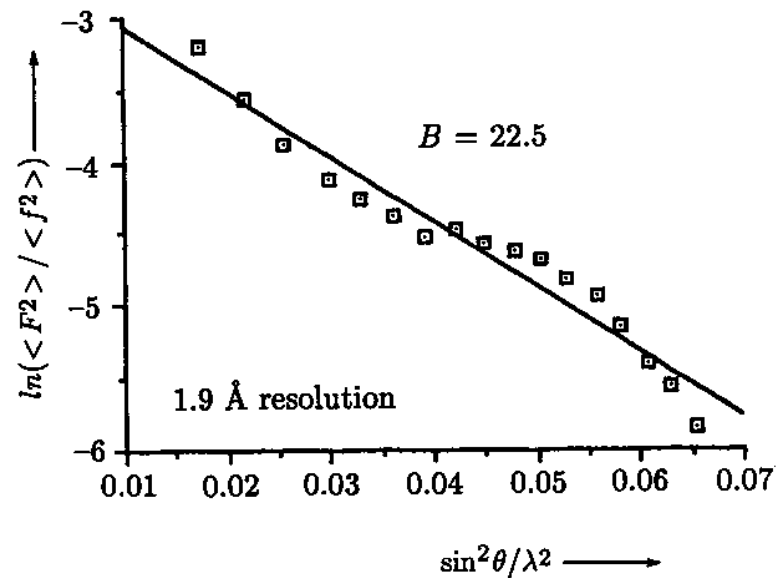
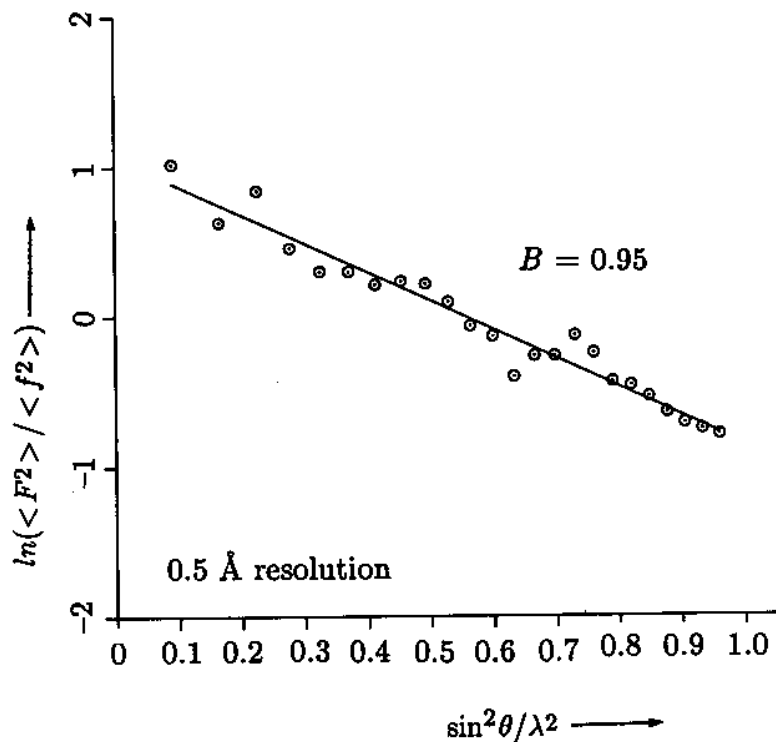
$$K|F_o(hkl)|^2 \approx |F(hkl)|^2 = \sum f^2 e^{-2B \sin^2 \theta / \lambda^2}$$

$$\ln[\sum |F_o(hkl)|^2 / \sum f^2] = \ln K - 2B \sin^2 \theta / \lambda^2$$

- for random atom placement in the unit cell
- Both K and B can be obtained from a Wilson plot



Wilson plots



"Crystal Structure Analysis for Chemists and Biologists", Glusker, Lewis and Rossi, VCH, 1994.

FIGURE 7.22. Two Wilson plots. (a) Diffraction data for a small structure, sodium citrate. (b) Diffraction data for macromolecule, D-xylose isomerase. Note that all the data for the macromolecule in (b) are within the value of $\sin \theta / \lambda$ for the first point for the small structure in (a). (Courtesy H. L. Carrell)



Restraints and constraints

- **It is possible to impose certain restrictions on the refinement due to some knowledge that is not inherent in the diffraction data**
 - restraints and constraints
- **Chemical knowledge, such as bond distances and angles, can be used as a *restraint* in a minimization procedure**
 - a restraint will make certain moves unfavorable, but will not prohibit them
- **Knowledge about molecular connectivity can be used**
 - e.g., a rigid aromatic ring can be described by three positional and three rotational parameters instead of $4n$ parameters (n = number of atoms)
 - this would be a *rigid body constraint*



When to use restraints and constraints

- **Restraints and constraints can improve your data to parameter ratio by reducing the number of parameters or “adding” observations**
 - very useful if your data are limited or of low quality
- **They improve the convergence properties of your refinement and may allow a refinement to converge to the correct answer even if the starting model is poor**
 - they make potentially disastrous parameter changes (e.g., one carbon atom moving so far that the aromatic ring will no longer be connected) unfavorable or prohibit them altogether

What to make of the outcome of your refinement

- **There are several ways of judging whether the outcome of your refinement is any good**
 - low R-factors
 - residual electron density map should have no large positive or negative peaks
 - difference Fourier map should be relatively flat
 - the atomic arrangement should be chemically and physically sensible
- **All of this information can be found in a *.cif file, which you will usually be asked to submit with any crystal structure publications**
 - cif stands for crystal information file
 - you need to know what the entries in it mean to make sense of it

Interpretation of atomic coordinates

- **Atomic coordinates are usually listed as *fractional coordinates***
 - an x-coordinate of 0.6 means: translation of 0.6 of the unit cell along the a-axis
 - makes comparison of isostructural materials easy
 - no need to worry about angles between unit cell axes when describing atom locations
- **To judge whether a structure is sensible, you need to interpret the atomic connectivity (bonding) and packing**
 - you need to calculate interatomic distances
 - structural drawings
- **Watch out: Occasionally, you will come across files that list Cartesian coordinates instead of fractional coordinates!**

Atomic connectivity

- **The presence of bonds is usually judged by interatomic distances**
 - especially for molecular compounds, chemical bonds result in well-defined interatomic distances
 - $\text{C}(\text{sp}^3)\text{-C}(\text{sp}^3) \sim 1.54 \text{ \AA}$
 - $\text{C}(\text{sp}^2)\text{-C}(\text{sp}^2) \sim 1.32 \text{ \AA}$
 - more difficult to judge for extended solids, as bond distances can vary with packing and oxidation state
- **Distances can be calculated from the atomic fractional coordinates**
- **Bond angles are also important**
 - for example, you know that a benzene ring should be planar with 120° angles



Calculation of bond lengths

- The location of each atom relative to the origin of the unit cell can be described as
 - $\mathbf{r} = x \cdot \mathbf{a} + y \cdot \mathbf{b} + z \cdot \mathbf{c}$
- The vector between two atoms is
 - $\Delta \mathbf{r} = \Delta x \cdot \mathbf{a} + \Delta y \cdot \mathbf{b} + \Delta z \cdot \mathbf{c}$
- The bond length can then be calculated as
 - $r = (|\Delta \mathbf{r}|^2)^{1/2}$
- You must take into account symmetry equivalent atoms when calculating bonds!

Error estimate on bond lengths

- From your refinement, you get error estimates on x, y and z
- This information can be used to calculate the errors of the interatomic distances

- Use error propagation:

$$\sigma^2(q) = \sum_{i=1}^N \left[\frac{\delta q}{\delta p_i} \sigma(p_i) \right]^2$$

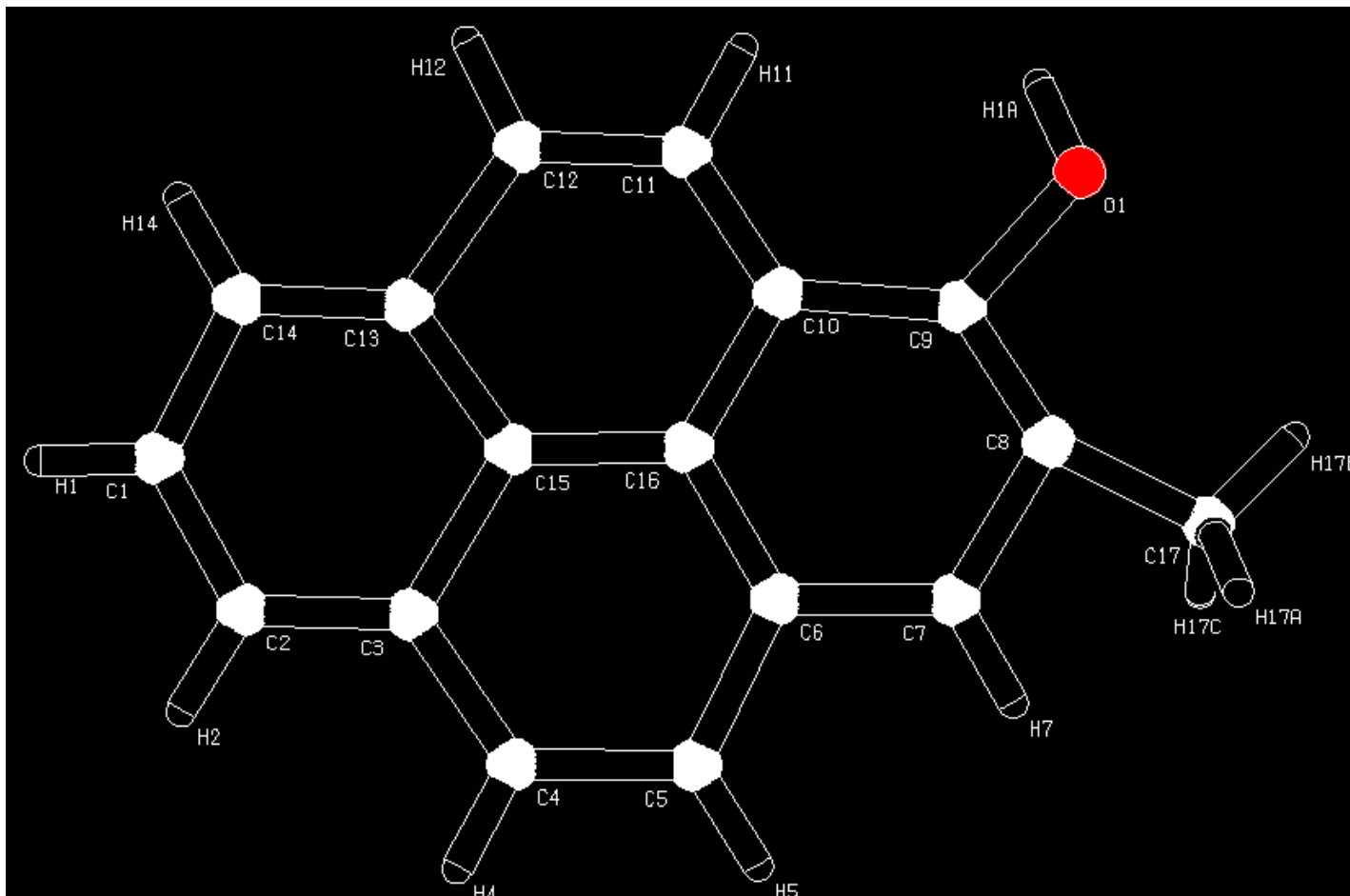
- For example, for two atoms with uncertainties of 0.005 Å in their atomic position,
 - $\sigma = [0.005^2 + 0.005^2]^{1/2}$ Å
 - this treatment assumes no correlation between the atoms

Looking at the structure

- The easiest way to figure out whether your results make sense is by looking at a structural diagram
- There are many different ways of representing a structure
 - ball and stick (PLUTO)
 - thermal ellipsoids (ORTEP)
 - space filling
 - coordination polyhedra (STRUPLO)
- There are also different styles
 - projection
 - perspective
 - stereo

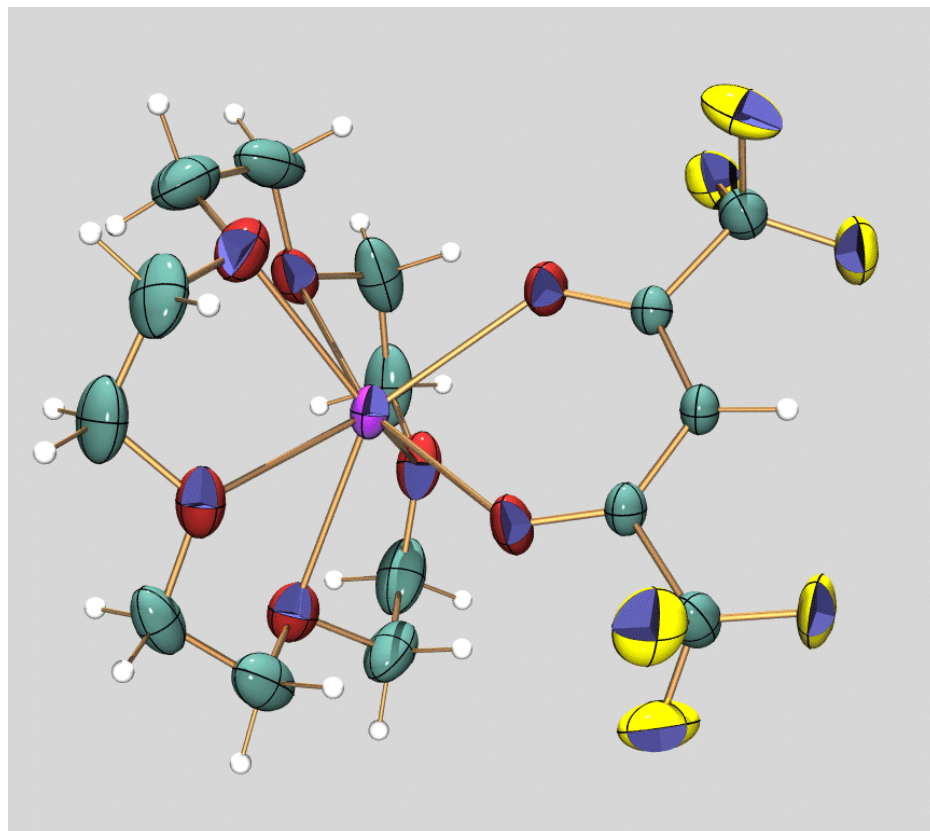
Ball and stick representation

- Very simplified, features can be clearly seen



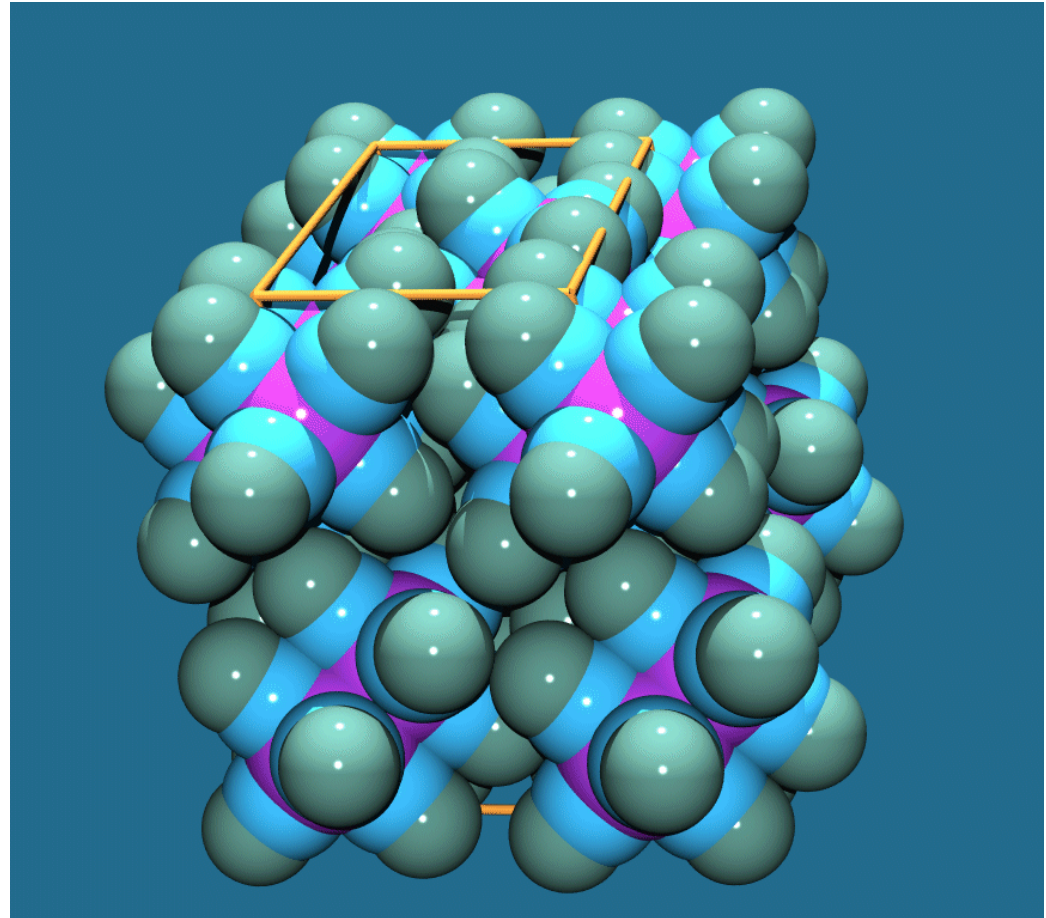
ORTEP (Oak Ridge Thermal Ellipsoid Plot)

- Standard representation for small molecules
- Orientation and size of thermal ellipsoids can be interpreted as thermal motion, static disorder or errors in the model

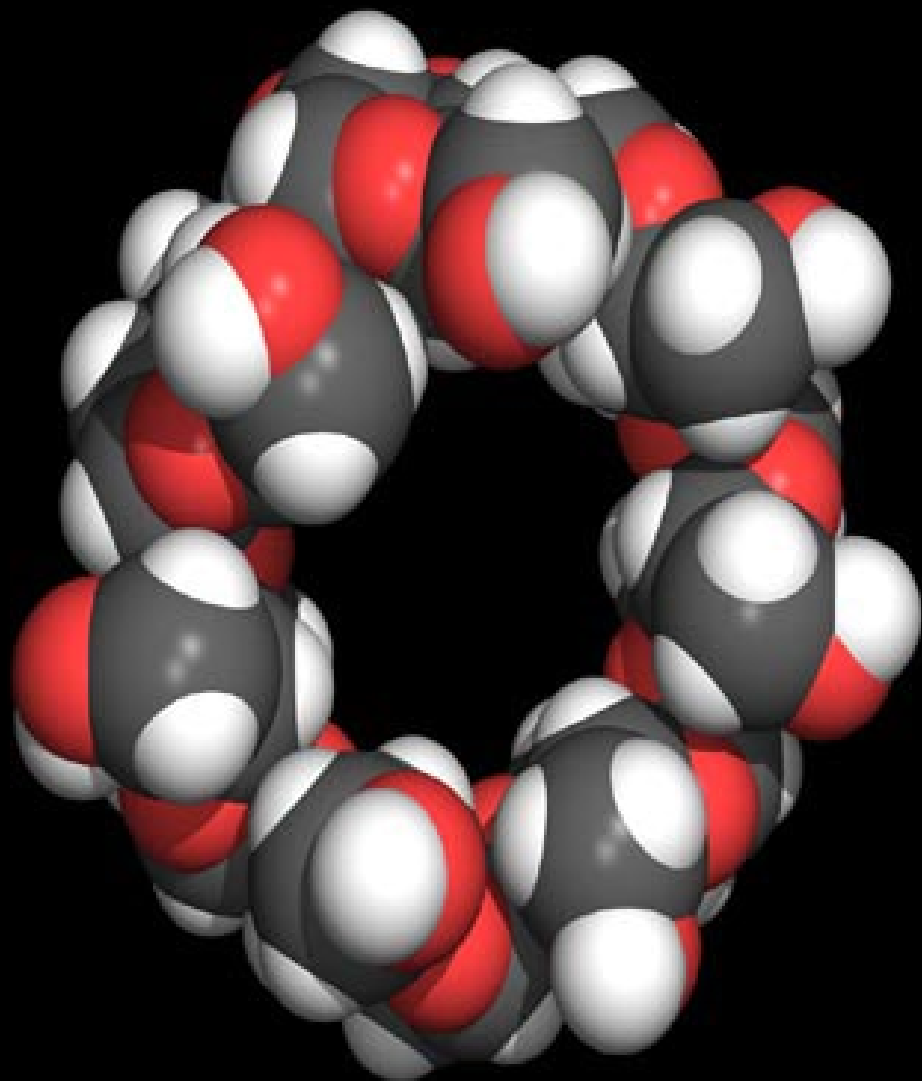
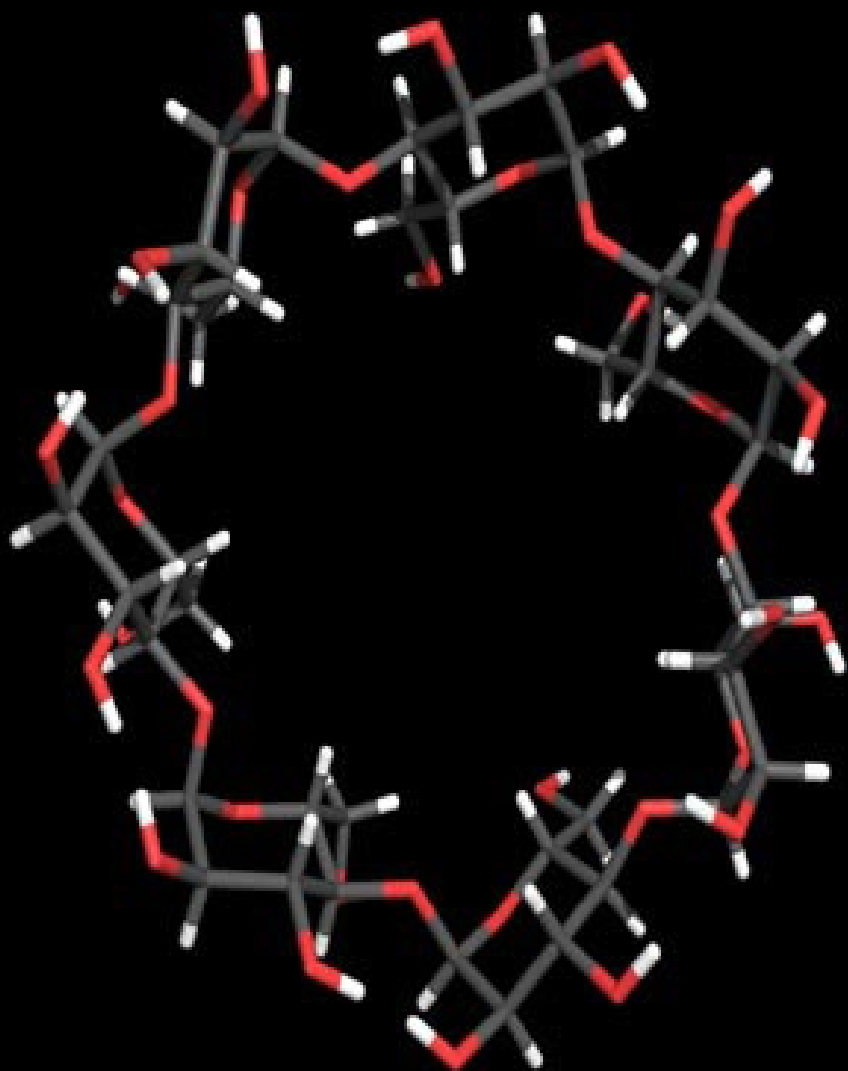


Space filling

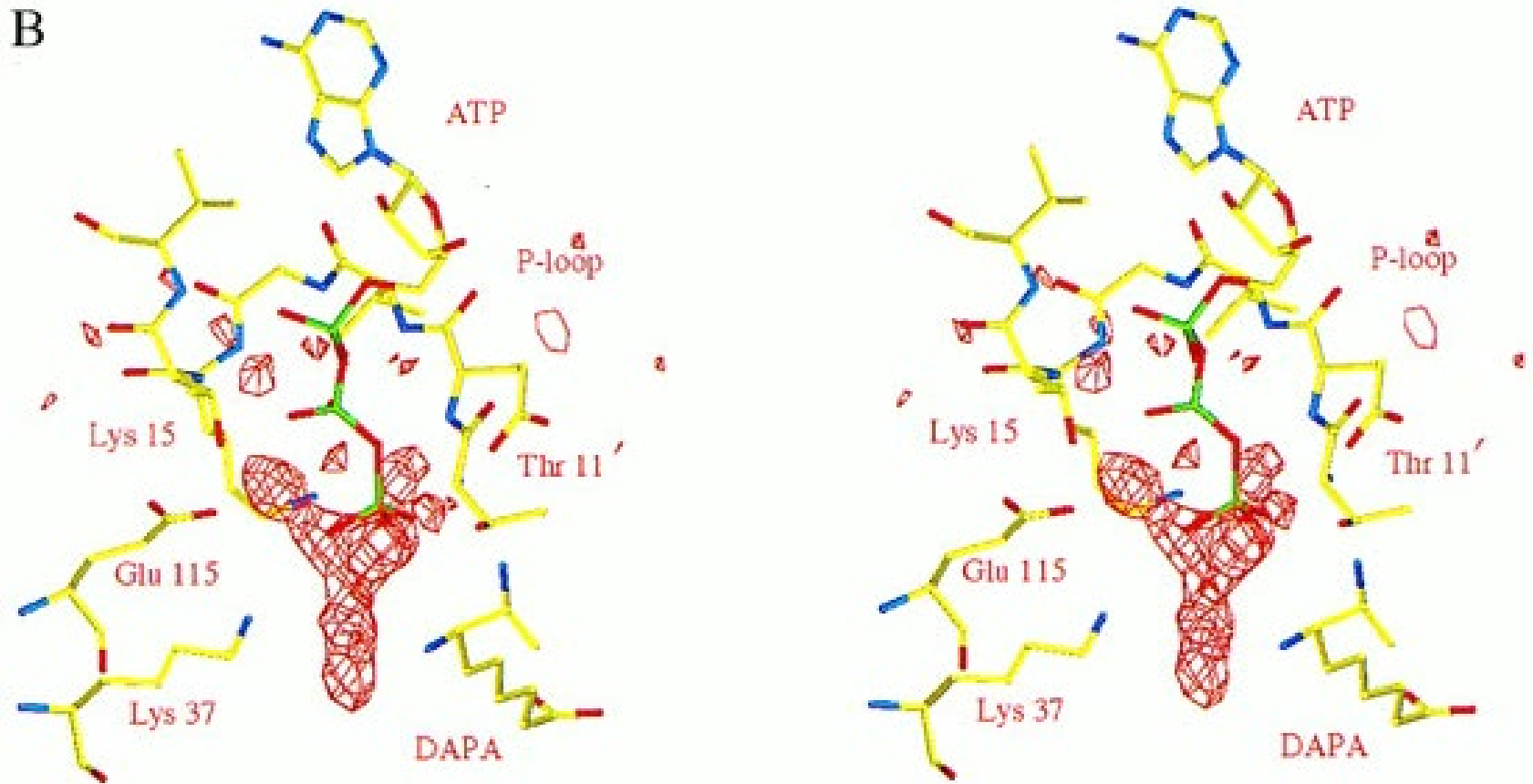
- Useful to get an idea of how densely packed a structure is



When to use which style



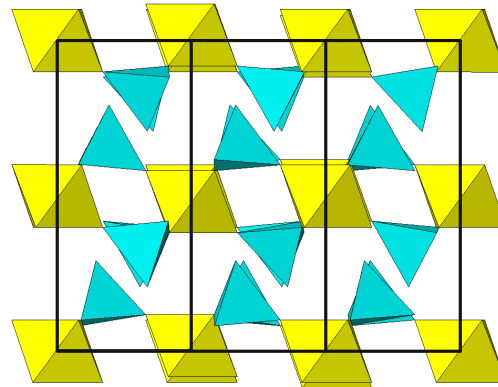
Stereo views



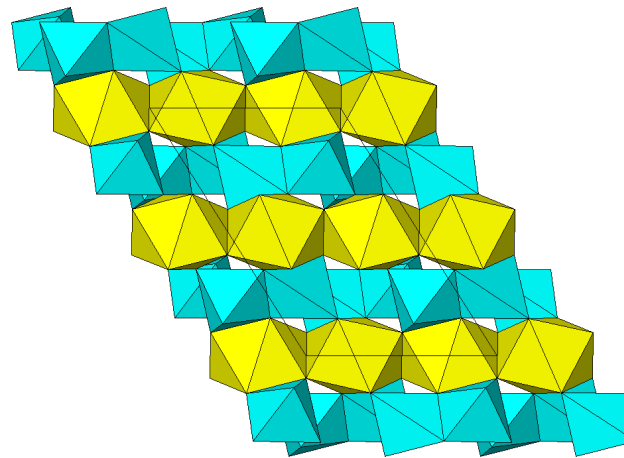
<http://www.pnas.org/content/95/10/5495/F3.large.jpg>

Coordination polyhedra

- Most common representation of extended solids
- Metals sit at center of polyhedra, anions at corners



Trigonal ZrMo_2O_8



Monoclinic ZrMo_2O_8



Atomic connectivity in extended solids

- Extended solids do not consist of a molecular moiety that is packed within the unit cell
- All you know is the chemical stoichiometry (if you did your analysis properly) and the total number of atoms in the unit cell (from estimating Z)
- The same metal atom can form a different number of bonds to the same anion in different structures
 - see the different ZrMo_2O_8 polymorphs
- You must have some idea of reasonable bond distances to interpret the Fourier difference map correctly!

Coordination and bond strength

- For most metals, there is a correlation between bond length and bond strength
- The anions will be more tightly bound the higher the oxidation state of the cation is
- The anions will be less tightly bound when more of them coordinate the same cation
- Pauling recognized that for certain structures each bond can be assigned a “valence” based on its length, and that the summation over all bonds from a metal center should give its valence

Bond valence sums

- In 1977, I.D. Brown introduced the bond valence sum concept
- Relationships between bond length and bond valence for given pairs of atoms were determined by analysis of databases
- **Bond valence: $s_{ij} = \exp((r_0 - r_{ij})/B)$**
 - r_0 and B are parameters deduced from database analysis
 - r_{ij} is the experimentally determined interatomic distance
- **Atomic valence: $V_i = \sum s_{ij}$**



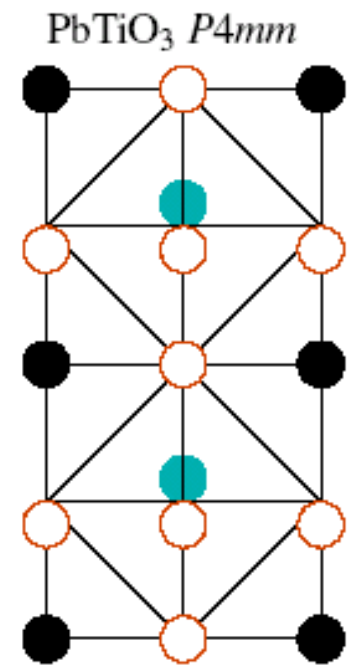
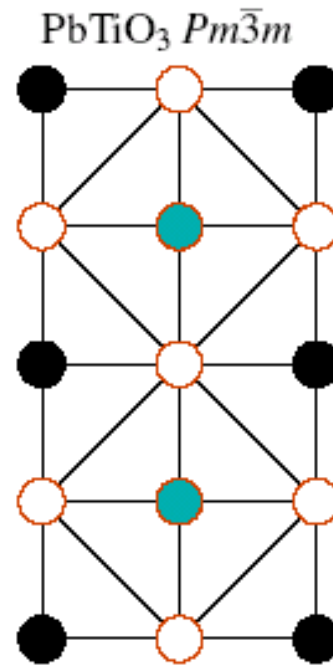
Using bond valence sums

- **Bond valence sums can be used to check the correctness of your crystal structure**
- **You will often know the average oxidation state of your metals**
 - from stoichiometry
 - crystal color (colorless materials usually contain fully oxidized metals)
 - other experiments like ESR
- **You can use bond valence sums to locate missing atoms**
 - your calculated metal valence will be too low if you are missing an anion
- **For this reason, they can be very useful aids in structure solution**



Polar materials

- If the positive and negative dipole moments in a crystal are crystallographically separated, it is called a *polar material*
 - results in positive and negative end of the crystal
- Many solids have a polar axis
 - PbTiO_4 , ZnS , ...



Chiral materials

- **Optically pure chiral molecules will crystallize to give a chiral solid**
 - they don't have a choice, as the mirror image does not exist
- **Racemic mixtures of chiral molecules can crystallize as chiral solids**
 - in each crystal, only one enantiomer exists
 - you get “enantiomorphous” crystals
- **Achiral structural building blocks can assemble to form a chiral solid**
 - quartz consists of achiral SiO_4 subunits

Space groups and chirality

- **Chiral solids can only crystallize in a limited number of space groups**
 - NO mirror planes
 - NO inversion centers
 - NO glide planes
- **Space groups that contain only rotation and screw axes are chiral**
 - any solid crystallizing in such a space group is chiral
- **Some space groups are enantiomorphous**
 - e.g., $P4_1$ and $P4_3$, $P6_1$ and $P6_5$, etc.

Determining absolute structures

- When you have a chiral solid, which enantiomer do you have?
- When you have a polar material, which way does the polar axis point?
- Answering both of these questions would be considered a problem of absolute structure
- Determining the absolute structure of a material is challenging
 - usually not possible with chemical or spectroscopic methods
- Crystallography is THE most important tool for establishing absolute structure!

Absolute structure and anomalous scattering

- Anomalous scattering and the resulting breakdown of Friedel's law can be used for determining absolute structure
- When anomalous scattering occurs, the atomic scattering factor must be expressed as
 - $f = f_0 + \Delta f' + i\Delta f''$
- This results in a modified structure factor expression
 - $F(hkl) = A + iB + (\Delta f' + i\Delta f'')(A_a + iB_a)$
- Inversion of the absolute structure (x, y, z goes to $-x, -y, -z$) gives
 - $F(hkl) = A - iB + (\Delta f' + i\Delta f'')(A_a - iB_a)$
- The magnitudes of these structure factors are *not* the same!

Absolute structure determination in everyday life

- You may work with chiral materials on a daily basis, but you are unlikely to have access to a tunable radiation source all the time
- For Friedel's law to break down, data should be collected “sufficiently near to an X-ray absorption edge”
 - for maximum impact, you want an X-ray energy very close to an absorption edge (e.g., a few eV below the edge)
 - with high accuracy data, there is still a small effect for compounds containing only C, H and O when collecting with Cu $K\alpha$

Other strategies

- The best way to determine absolute structure is the high accuracy measurement of Friedel pairs
- However, there are also other approaches
- Many least squares programs have an absolute structure parameter that can be refined
- Flack or Rogers parameters are the most commonly used absolute structure parameters
- Hamilton's R-ratio can also be used

Hamilton, Flack and Rogers parameters

- **Hamilton's R ratio** makes use of the small differences in $I(hkl)$ and $I(\bar{h}\bar{k}\bar{l})$ that occur even if data are collected not very close to an absorption edge
 - the ratio of the R-values obtained with the trial structure and its mirror image is calculated, showing which enantiomer gives the better fit
- **Johnson first introduced a variable for the $\Delta f''$ term**
 - method was rediscovered by Rogers, and refined by Flack
- **Rogers or Flack parameter: $f = f_0 + \Delta f' + i\eta\Delta f''$**
 - if η approaches +1 or -1, the correct absolute structure can be deduced