Phase Calculation

March 3, 2014
Solvent Content

- Fraction of the crystal volume occupied by solvent can be calculated from the crystal density and the partial specific volume.

- Partial specific volume of all proteins is close to 0.74 cm$^3$/g or 1.23 Å$^3$/Da.

- Matthew’s coefficient is the ratio of protein volume to solvent volume in the crystal.

- The volume of the asymmetric unit is estimated from the unit cell dimensions and the space group.

- Solvent content is usually between 40-60% (Matthew’s coefficient between 1.5 and 0.66).
Calculation of Atomic Coordinates

- For each reflection has an amplitude and phase
- Amplitude depends on the scattering factor
- Phase depends on the position of the atoms
- \( F(hkl) = \sum_{j=1}^{n} f_j e^{i\alpha_j} \)
- \( F(hkl) = \) structure factor for reflection \( h,k,l \)
- \( f_j = \) atomic scattering factor
- \( \alpha = \) phase factor
The Phase Problem

- The diffraction patterns provide the intensity of each reflection but lost the phase information for the waves.
- Intensity of the reflection = the structure factor for the molecule squared.
- \( I(hkl) = F(hkl)^2 \)
- Phases need to be experimentally determined or calculated.
Methods to Determine Phases

- There are several methods to determine the phases.
  - Direct methods
  - Isomorphous replacement
  - Multiwavelength Anomalous Diffraction - Single wavelength anomalous diffraction
  - Molecular Replacement
Patterson Function

- Can be used to deduce unit cell’s components
- Patterson function $P(uvw)$ does not rely upon phase information.
- Fourier transform of the intensity of structure factors, with their phases set to zero

$$P(uvw) = \frac{1}{V} \sum_h \sum_k \sum_l \exp[-2\pi(hu + kv + lw)]$$

- Patterson map’s origin therefore contains the vector of each atom with itself, and the peaks throughout the map represent interatomic vectors of the crystal.
Direct Methods

\[ F_{(h,k,l)} = \sum_{j=1}^{\text{atoms}} f_{(j)} \exp[2\pi \cdot i(hx_{(j)} + ky_{(j)} + lz_{(j)})] \]

- Guess the phases and try to calculate electron density. If you get negative values for electron density or random distribution it is wrong. Only certain values will give positive density.

- If there are 100 atoms then there is 300 unknowns (100x3 for X,Y,Z of each atom).

- Another way - if we know the composition of the molecule then we can try a conformation, calculate the diffraction pattern and determine how well the calculated and experimental data correlate.

- Routine method for small molecules (less than 200-300 hundred nonhydrogen atoms).

- This becomes increasingly difficult as the size of molecule increases.

- Computationally intensive.
Multiple Isomorphous Replacement (MIR)

- Historically the most common technique for determining phases
- Attachment of heavy metals at specific locations in the crystal
- The scattering contribution to the INTENSITY is the square of the atomic scattering factor (scattering factor increases with atomic number). Just a few atoms can have a profound difference on the intensities
- Isomorphous replacement means that the protein in the native and derivative crystals has not changed much and the only difference is the addition of the heavy metal
- By comparing the different intensities between native and derivative data sets we can determine the position of the heavy metal sites.
- From these sites we can calculate the phases for the heavy metals
- These phases can be used to calculate phases for the entire molecule
Disadvantages of MIR

- Require to have multiple (at least two) derivative data sets and a native data set
- Binding of heavy metals can be very difficult and result in nonisomorphous
- Too many binding sites cause changes to the protein
- Many partially occupied sites
- No binding
Multiwavelength Anomalous Diffraction

- The newest method to experimentally determine phases.
- Need to have a sample that is derivatized with a heavy metal
- Use incident wavelength of X-rays that are near to the absorption of the heavy atom
- This will cause the intensities to change depending on the wavelength of X-rays used.
- This change in intensities can be used to determine the sites of heavy metal attachment
- Atoms with atomic numbers greater than 20 (Ca and above) have absorption in the range of 0.3 to 3.0Å (wavelength)
- Absorption will create fluorescence which can be measured.
• Collect 3 data sets at different wavelengths
MAD vs MIR

- MIR requires multiple crystals
- MAD can be considered a special case of MIR
- MAD requires one crystal and data is collected at several wavelengths.
- Treat a data set where the atom does not absorb X-rays as “Native” and the data sets where the absorption occurs as “Derivatives”
- Since the data is collected from a single crystal, all data is isomorphous
- MAD data collection needs to be done at a synchrotron since it requires using many different wavelengths
Molecular Replacement

- Know the structure of part or most of molecule
- Similar in many ways to direct methods
- Translation is done first to localize the center of mass, then rotation to get the correct orientation
- Calculate phases using the known structure
- Usefulness - if you have small molecule bound to a protein
- Disadvantage - There can not be large changes in the structure that you are using.
Electron Density

- Once you have determined the phases then use this information to calculate electron density.
- Then build into electron density.
- Atomic resolution requires about 1-1.2Å resolution.
- Most protein crystals do not diffract to this resolution.
- For protein and DNA we do know the structure of the individual amino acids and nucleotides.
- We can approximate their position in the electron density.
Electron Density
Dramatic improvements in the overall structure are likely to result from better definition of disordered regions regardless of resolution.