The Phase Problem

- Data collection → |F|
- Map calculation requires vector F
  - direction or phase offset
- Phases can not be measured directly

Methods to be covered

- Direct methods - briefly
- Ab initio - skip
- Molecular Replacement
- Isomorphous Replacement
- Multi-wavelength Anomalous Diffraction

Phase determination - Direct Methods

- Statistical interdependence of structure factors
  - $P(u) = f(|F_{21}|, |F_{22}|, \ldots)$
- Apply constraints
  - E.g. atomicity
  - Spheres uniform density
  - Separated by vacuum
- Nobel Prize
  - Hauptman & Karle
- Applies to "small" molecules
  - Salts
  - Organic molecules
  - Small proteins
  - "Shake-N-Bake"
  - Hauptman & Weeks; Sheldrick
  - ~1000 atoms
- Heavy atom "sub-structures"
  - Derivatives
  - SeMet

Overview

- Quickest method
- When related "probe" structure is known
- Requirement
  - Know how to superimpose probe structure
  - On unknown structure
  - In a different unit cell
  - (Before unknown structure is known)
  - How to:
    - Orient - 3 angles - "Rotation Function"
    - Place - 3 position vector components
      - "Translation function"
- Method not without its difficulties

Part 2:

Molecular Replacement
When Related Structure Known
How related must the probe structure be?
- No hard & fast rules - but empirical bottom line
- To get an interpretable map
  - >70% structure needs to be approximated
  - Atoms say w/in 2 Å
- Sometimes can combine probes, sum to >70%
- Difficult to figure orientation / translation
- Methods improving...

Determination of the Orientation
- Patterson synthesis
  - \( P(x) = \sum_h |F_h|^2 \cos(2\pi hx) \)
  - No phases
  - Auto-correlation
  - Vectors between atoms
- Compare
  - Vectors w/in molecule
  - Not between molecules
  - "Self-vectors" shorter
- Patterson depends on molecular orientation

Orientation from Patterson Overlap
- Rotate Probe model coordinates
- Calculate Patterson
- Assess overlap
- Compare to observed Patterson
- Step over 3 angles
- At which orientations are observed and calculated Pattersons well correlated?

Challenges of Rotation Function
- Many solutions look equally good.
- The highest scoring is not always correct
- Correct could be 30th... or worse

Patterson vectors that determine orientation
- Patterson contains
  - Peaks for all molecules
  - Peaks between neighbors - w/in & between unit cells
- Red Patterson peaks are from single molecule

Patterson vectors that determine orientation
- If consider only peaks close to origin
  - More are self peaks (red)
  - Less likely to have spurious solution
  - "Integration radius"
  - Impossible to completely separate
    - Self vs. cross peaks
    - Noise in rotation function
    - Perhaps some spurious solutions

\[
R(C) = \int P(u)P_C(u) du
\]
Care needed with rotation functions
- Most sensitive to...
  - Large reflections - \(|F|^2\)
  - Make sure all large \(F\) have been measured
  - Higher resolution data - say 3 to 5 Å
  - Check that RF not sensitive to exact limits
- Very noisy
  - Rank according to signal / noise
  - Correct solution is often the 5th, sometimes the 30th peak.
  - Continue structure determination with several solutions - which works out best?

Translation functions
- Position w/in unit cell when orientation known
- Greatest challenge of Molecular Replacement
- What position most consistent w/ diffraction data?
- Translation function: \(T(t) = \int P_{1,2}(u,t) P(u) \, du\)
  - \(P_{1,2}\) are Patterson vectors between molecules related by crystal symmetry
  - \(P(u)\) is observed Patterson
- Patterson Correlation, \(\text{Corr}(t) = \frac{\sum_n (F_u^2 - <F_u^2>)(F_{c2}^2 - <F_{c2}^2>)}{\left(\sum_n (F_u^2 - <F_u^2>)^2 \sum_n (F_{c2}^2 - <F_{c2}^2>)^2\right)^{1/2}}\)

Translation Functions are Challenging
- Patterson functions intrinsically noisy
- Translation functions sensitive to exact orientation
  - Slight orientational error \(\Rightarrow\)
  - May miss correct position
- Techniques to improve your chances
  - Combine with other information
    - Packing analysis - molecules overlap?
  - Refine orientation - Patterson correlation function

Solving Molecular Replacement
- Two steps: (a) Orientation (RF); Position (TF)
- Several packages that combine them
  - Explore several possible RF solutions
  - Reduce errors due to differing conventions
- Programs: Phaser (Max. likelihood); AMoRe; GLRF
- Model \(\rightarrow\) \(F_{\text{calc}}\): \((F_{\text{calc}}, \Phi_{\text{calc}})\)
  - Combine w/ data: \((F_{\text{calc}}, \Phi_{\text{calc}}) \rightarrow \text{hybrid map}\)
  - Remodel \(\rightarrow\) better \(\Phi_{\text{calc}}\) \(\rightarrow\) better map \(\rightarrow\) model...
- Success judged by agreement between \(F_{\text{calc}}\) & \(F_{\text{obs}}\)
  - ... and ability to improve it with refinement
  - Expected (new) features in map, e.g. sequence
  - Need for caution

Confusing Names
- Uses Heavy Atoms, but \textit{not} "Heavy Atom Method"
- Adds atoms rather than \textit{replacing} them
  - Historically - based on methods where replaced
  - Isomorphous - protein must remain in same conformation after heavy atoms added
  - or almost so

Part 3: ISOMORPHOUS REPLACEMENT
CLASSIC APPROACH W/O RELATED STRUCTURE
Phase Det. – Isomorphous Replacement
1. Collect "native" data set: |F_P|A
2. Attach heavy atom(s) to protein
3. Collect "derivative" data set: |F_PH|
4. Solve heavy atom positions from (F_PH – F_P)

Heavy Metals
- Few atoms bound
  - Need to be able to solve as small molecule
  - Need to be able to detect
- High atomic number - f^2 = \sum Z^2.
  - Contribution \propto Z^2.
- Hg, Pt, Pb, Au, U...
- > 200 reagents, e.g.: K_2PtCl_4, HgAc_2, p-chloromurcuribenzoic acid, UO_2(NO_3)_2, PbAc_2
  - Try a wide selection
- Covalent binding to 1º amines:
  - K_2PtCl_4, K_2AuCl_4...
  - Charged interaction also possible, e.g. K_2AuCl_2
- Electrostatic binding
  - E.g. PbAc_2, uranyl acetate & carboxylates

Heavy Metal - Chemistry
- Hg binds covalently to Cys
  - Great if works
  - Sometimes reduces essential disulfides
  - Denatures protein
- Covalent binding to 1º amines:
  - K_2PtCl_4, K_2AuCl_4...
- Charged interaction also possible, e.g. K_2AuCl_2
- Electrostatic binding
  - E.g. PbAc_2, uranyl acetate & carboxylates

Why particular reagents may not work
- Conformational change
  - Denaturing
  - Subtle non-isomorphism
- Binds at too many sites (to determine positions)
- No binding sites - reactive sites occluded
- Buffer interactions
  - PtCl_4^{2-}, AuCl_4^{2-} react w/ amino "Good" buffers
  - Reagent precipitated
  - Buffers containing PO_4, SO_4 precipitate Hg^+, Hg^{2+}, Pb^{2+} etc..

Searching for derivatives
- Typically have to test dozens of reagents
  - Sometimes hundreds
  - Each at several concentrations
- Excellent guidelines for efficient searches:
  - Petsko, G. Methods in Enzymology 114
  - Chemical series - try most reactive, then least
    - E.g. PtCl_4^{2-}, AuCl_4^{2-}
  - But... Differ in "hardness", lability
  - Ionic vs. covalent interactions
  - Try examples of "soft" & "hard" species

Derivatives - the bottom line
- Diffraction / phasing power
- Days of work, each test
- Data set
  - Quality of diffraction
  - Are the intensities changed?
  - Determine sites
  - Phases - good enough?
Screening tests – eliminate candidates

- Does it precipitate?
  - Mother liquor - no need to waste protein!
- Does it react?
  - Colored compounds
    - Some change color w/ valency e.g. Pt(II) \(\rightarrow\) Pt(IV)
    - E.g. PtCl\(_4^2-\), AuCl\(_4^2-\)
    - Others - color should concentrate in crystal
  - Non-colored
    - Does overdose crack a crystal?
      - No: probably not reacting
      - Yes: reacting or osmotic shock?
    - Does it change the diffraction pattern?
      - E.g. PtCl\(_4^2-\), AuCl\(_4^2-\)
      - Others – color should concentrate in crystal

How much should the diffraction be changed?

- Maximize heavy atom signal w/o changing protein
- Measure \(\Delta F = \sum |F_{PH} - F_P| / \Sigma F_P\)
  - Above 30% - usually non-isomorphous
  - Below 12% - barely detectable
- Want
  - Small number of binding sites (1 to 6)
    - Complete reaction at these sites
      - Full "occupancy"
    - Check w/ Patterson or Difference Fourier (later)
  - Usually need to optimize concentration, soak time

Frustations of Screening

- Can fail at a number of stages
- Final tests require substantial investment of work
  - Careful preliminary tests!
- May need to try many compounds
- May need to transfer to more favorable buffer
- Will need ~ three derivatives
  - Couple of months \(\rightarrow\) a year or two

From heavy atoms to phases... (overview)

- For each reflection...
- Solve for \(\alpha_P\) by triangulating: \(F_{PH} = F_P + F_H\)
- Need \(\alpha_H\), calculated from positions in unit cell.
- Determination of positions
  - Difference Fourier if preliminary phases
  - Difference Patterson w/o phases

Meaning of the Patterson

- \(P(u) = \int \rho(x)\rho(x-u)dx = \Sigma |F_i|\cos2\pi(hx)\)
- Let \(\rho(x) = 0\), except at atom positions
- \(P(u)\) is zero except when \(x\) & \(x-u\) are atoms
- Peaks in \(P(u)\)
  - When \(u\) is an inter-atomic vector
    - Height = \(\rho(\text{atom1}) \times \rho(\text{atom2}) = Z_1 \times Z_2\)
    - Number = \(N^2\), \(N\) at origin
  - Blurred according to resolution - overlapped
  - Interatomic vectors \(\rightarrow\) solve small structure
  - Large structure - Patterson too complicated
  - Difference Patterson \(|F_{PH} - F_P|\) approx heavy atoms

Patterson \(\rightarrow\) Atom positions: Harker Sections

- Patterson peaks a.k.a. "vectors"
- Crystal symmetry \(\rightarrow\) concentration in planes
- Example 2-fold along b:
  - \((x,y,z) = (-x, y, -z) \rightarrow \text{vector} = (2x, 0, 2z)\)
  - Harker section \((u,v,w) = 0; u=2x; w=2z\)
- Example 2, along b:
  - \((x,y,z) = (-x, y+\frac{1}{2}, -z) \rightarrow \text{vector} = (2x, \frac{1}{2}, 2z)\)
  - Harker section \((u,v,w) = \frac{1}{2}; u=2x; w=2z\)
  1. Search (Harker sections) for peaks
  2. Find \((x,y,z)\) consistent w/ peaks
     - Educated guesswork
     - Systematic computational searches
Difference Pattersons Full of Error

- Crude approximation
  - Heavy atom vectors: $\sum_{h} |F_{PH,h}|^2 \cos 2\pi (hx)$
  - "P" for protein; "PH" for protein + heavy atom
  - Can only calculate: $\sum_{h} (|F_{PH,h}| - |F_{P,h}|)^2 \cos 2\pi (hx)$
  - Many background peaks
  - Small (20%) difference between 2 exptl values
  - Then squaring the difference!
- Very sensitive to
  - Errors in intensity data
  - Missing reflections
  - Some prove intractable

What to do when Patterson insoluble?

- Put aside
- Find another derivative
- Use 2nd derivative to calculate approx phases
- Calculate difference Fourier using 1st derivative amplitudes and 2nd derivative phases
- $p(x) = \frac{1}{V} \sum_{h} (|F_{PH,h}| - |F_{P,h}|) \exp(-2\pi ihx)$
  - Coefficients are not squared - less error
  - N peaks for N sites

Using heavy atom positions...

- From Difference Patterson / Fourier
- Calculate $F_H$ vector = $\sum_{fh} \exp{2\pi i h \cdot x}$
  - W/ measured $|F_P|$ & $|F_{PH}|$ amplitudes
  - Using cosine rule:
    - $|F_{PH}|^2 = |F_P|^2 + |F_H|^2 + 2|F_P||F_{PH}|\cos(\alpha_P - \alpha_H)$
  - $\alpha_P = \alpha_H \pm \cos^{-1}{(|F_{PH}|^2 - |F_P|^2 - |F_H|^2) / 2|F_P||F_{PH}|}$
- Symmetry of cosine: 2 angles have same cosine
- Two phase angles are equally probable
- (Note convention of plotting negative $F_{PH}$)

Single Isomorphous Replacement Phase Ambiguity

- $\alpha_P = \alpha_H \cos^{-1}{(|F_{PH}|^2 - |F_P|^2 - |F_H|^2) / 2|F_P||F_{PH}|}$
  - Symmetry of cosine: 2 angles have same cosine
  - $\alpha_P = \alpha_H \pm \text{something}$
  - Two phase angles are equally probable
  - (Note convention of plotting negative $F_{PH}$)

Multiple Isomorphous Replacement (MIR) to Resolve this Ambiguity

- 2nd derivative w/ heavy atoms in different places
- Different $F_H$
- Only one solution same for both derivatives
- Or nearly so...

Effect of Errors

- Consider small error in $|F_P|$
  - Changes intersection point
  - Changes protein phase
- Measure particular $|F_P|$
  - "Real" value + random error
  - $P(|F_P|)$ is distribution
  - $\Rightarrow$ Distribution of $\alpha_P$
  - "Phase probability distribution"
- Remember 2 possible phases
  - $\Rightarrow$ Bi-lobed distribution
  - $P(\alpha)$
Types of Errors
- $|F_P|, |F_{PH}|$ experimental measurement error
- $|F_H|$ if heavy atom model is incomplete/inaccurate
  - Heavy atom refinement methods
  - Maximum Likelihood vs. Least-Squares
- Lack of closure, $\varepsilon$
  - Errors $\rightarrow$ triangle $F_{PH} = F_P + F_H$ should not close
- Other errors contribute to $\varepsilon$
  - Non-isomorphism
  - Protein changed
  - Derivative not protein + heavy atoms
  
MIR & Phase Probability Distributions
- Each derivative $\rightarrow$ probability distribution
- How to combine the information?

MIR Phase probability distributions
- Derivative 1
- Derivative 2
- Derivative 3...
- Combined by product

Use of Phase Probabilities
- Updated as new phase information added
- Modified according to constraints
  - Non-crystallographic symmetry
  - Solvent flattening, etc.,
- Map calculation
  - One phase for each reflection
  - Which one?

Best & Most Probable phases
- $P(\alpha)$
- $\alpha_{best}$

Uncertainty in the Best Phase
- More confident of phase if
  - One peak dominates $P(\alpha)$
  - Peak is sharp
- Different reflections may have phases determined w/ more or less confidence
- Can we use this information to give maps of minimal error?
- More emphasis to well-determined reflections.
- Weights - a.k.a. “figure of merit”
MIR - Conclusion

**Advantages**
- Prior structure not required
- Requires only standard laboratory x-ray equipment
- Errors are random and systematic
- Use other methods when appropriate
- MIR is a robust method of last resort

**Disadvantages**
- A lot of work
- Large random errors

---

Part 4: ANOMALOUS DIFFRACTION - MAD PHASING

**Anomalous Diffraction**
- SIRAS - A way of resolving the phase ambiguity
  - Sometimes
- Multiwavelength Anomalous Diffraction (MAD)
  - Powerful new method for accurate phases

---

**Review - Scattering by a Free Electron.**
- Electromagnetic radiation = oscillating field.
- Field accelerates a charged particle with frequency \( \nu \).
- At max (or min) of field, \( E_i \).
  - Force on charged particle is greatest
  - Acceleration is greatest
  - Electron displacement \( \pi/2 \) from \( E_i \).
- The accelerating orbital electron initiates a second electromagnetic wave with 2nd phase change of \( \pi/2 \).

**When an Electron is Not Free**
- As nucleus becomes larger & more +ve...
- Electrons increasingly tethered
- Scattering from dipoles with natural oscillation frequency \( \nu_n \).
- Compared to a free electron, scattering is
  \[ f_n = \frac{\nu^2}{\nu^2 - \nu_n^2 - i \kappa_n \nu} \]
  - Forced, damped oscillator
  - \( \nu = \) frequency of incident radiation
  - Changes magnitude
  - Note also complex
  - Phase lag, dependent on damping constant, \( \kappa_n \)
  - Phase difference (scattered-incident) > 2\( \pi \).

**Imaginary component of scattering factors**
- \( f \) used more than \( \Delta f \), but also used for \( f + \Delta f \).
  - \( \Delta f \) will be used to avoid confusion
  \[ f_{\text{anom}} = f + \Delta f + f' \]

---

(c) Michael S. Chapman
Effect on Heavy Atom Structure Factors
- Imaginary $f''$ rotates structure factor anti-clockwise
- $F_{PH}(+h) = F_{PH}(-h)$
- Different directions
- $F_{PH}(+) = F_{PH}(-)$
- Friedel's law breaks
- Can use $|F_{PH}(+)|$, $|F_{PH}(-)|$ as 2 derivatives

Precise Data Needed
- Anomalous scattering is small
  - ~6% for Hg atom & CuK radiation
  - Can increase by changing $\lambda$
    - Needs synchrotron source w/ tunable wavelength
  - Precisely measured data to be able to detect anomalous signal

When are Anomalous Effects Significant?
- $f_a = v^2 / (v^2 - v_n^2 - i\kappa v)$
- Limit: $v \gg v_n \Rightarrow f_a = 1$
  - Scattering from free electron
- Limit: $v \ll v_n \Rightarrow f_a = 0$
  - No Scattering
- Significant when $v = v_n$
  - $v_n$ are the absorption edges: K, L ...

Anomalous scattering $N^o$ absorption edge
- Se K edge
- 30 electrons
- Max 30% change

Two Strategies for Phasing with Anomalous Diffraction
**With tunable x-ray source**
- MAD method
  - Collect at 3 wavelengths
    - Maximize $|\Delta F| - \lambda_1$
    - Maximize $f'' - \lambda_2$
    - Far from edge - $\lambda_3$
    - Treat $F(\lambda_2)$ as ~ native
    - No need for another crystal
    - $F(\lambda_2)$, $F(\lambda_3)$ like 2 derivatives

**With Fixed wavelength**
- SIRAS / MIRAS
  - Collect native + derivative
  - Primary phasing from SIR / MIR
  - Collect both $F(+)$, $F(-)$
  - Differences in $F(\lambda_1)$, $F(\lambda_3)$
  - Supplementary phase information
  - Breaks ambiguity
  - (Determines hand)

Theory - Anomalous Diffraction $\Rightarrow$ Phases
- $\alpha_P(+) = -\alpha_P(-)$
- Correct solutions must be mirror images about the Real axis
- (Dotted line)
A Trick to Simplify
- Plot mirror of \( F_H(-) \)
- Solutions now superimpose

Mirror image changes direction of rotation
- \( f^* \) rotates \( F_H \) anticlockwise
- \( f^*_{\text{mirror}} \) rotates \( F_H \) clockwise

SIRAS
- Resolve phase ambiguity with single derivative
- Based ~ 6% differences between \( F_{PH}(+) \) & \( F_{PH}(-) \)
- Can only be exploited w/ excellent data
- \( \alpha_P(+) \) and \( \alpha_P(-) \)
  - Likely not exactly the same
  - Approximately at best
- Maps rarely interpretable until phases refined

SIRAS & MIRAS
- SIRAS
  - Modest supplement to SIR phasing
- MIRAS
  - Modest supplement to MIR phasing

Multiwavelength Anomalous Diffraction
- MAD Phasing

MAD
- Principles exactly the same as SIRAS
- but... Tune \( \lambda \) to maximize the anomalous effects
- Change \( \lambda \) to mimic isomorphous replacement
  - MIR: Change protein & collect diffraction
  - MAD: Same protein & change wavelength
    - Protein must contain an anomalous scatterer
    - "Derivative" is isomorphous - by definition
    - Eliminate major source of error
    - MAD can \( \Rightarrow \) very precise phases
Anomalous Scatterers

- Natural atom
  - Fe proteins etc.
- Isomorphous atom substitution
  - Lanthanide for Ca++, etc.
  - Se for S
    - Express in bacteria that require Met
- Replacement of Met in media by seleno-Met
  - Expression can be a challenge.
- When all else fails:
  - Make derivative - solve derivative not native

Picking wavelengths

- $\lambda_1$ absorption edge
  - $\Delta$F
- $\lambda_2$ far from edge
  - Little effect
- $F(\lambda_1) - F(\lambda_2)$:
  - Large change in $|F_H|$ magnitude
  - Little change in direction ($f^*$)
- $\lambda_3$ -- Max $f^*$
  - Max Bijvoet difference: $||F_{PH(+)}| - |F_{PH(-)}||$

Processing MAD Data

- Start as in SIR - determine heavy atom sites
- Then calculate phases...
- Several methods
  - All fundamentally like MIRAS
  - Where do the magnitudes of $F(\lambda_1)$, $F(\lambda_2)$... intersect?
  - Known Magnitudes and directions for
    - $F_A = F_H$, $\Delta f$, $f^*$

MAD Algorithms

- Hendrickson & Smith - deterministic method
  - Calculate $F_A$, $\Delta f$, $f^*$ from 1st principles
  - Phase determined geometrically
    - 2 wavelengths enough (if no exptl. error)
    - 3rd $\to$ Least squares $\to$ best solution
- Pseudo MIR - pretend each $\lambda$ is a derivative
  - Statistics through phase probability distributions
- Now -- Maximum likelihood methods
  - SHARP: Maximum likelihood refinement of MIR / MAD parameters (Bricogne & Colleagues)
  - SOLVE / RESOLVE: Maximum likelihood MAD $\to$ auto-building (Terwilliger & Colleagues)

Isomorphism in MAD

- All data from one crystal
  - "Native" + "Derivative"
- Data sets are isomorphous by definition
- Eliminate big source of error in phasing
- Surprising how much one can do w/ a little anomalous signal
  - If perfectly isomorphous

Why’s everyone MAD about MAD

- No derivatives required
  - Seleno-Met expression or metalloprotein
- At most one derivative required
- Most accurate experimental phases possible
- If strong anomalous scatterer
  - Mannose Binding Protein A / Ho$^3+$ (Burling & Brünger)
Phase Determination → Phase Refinement

- Phase determination is approximate
  - Molecular replacement:
    - known model is not unknown structure
  - Isomorphous replacement:
    - Small differences between $F_{\text{iso}}$ & $F_{\text{p}}$
    - Assumes heavy metals do not change protein structure
- Phases may need refining
- Maps will have much error

Role of Phase Refinement

- Occasionally, 1st map → good model
- Atomic refinement converges easily
- Little/no need for phase refinement
- Sometimes, 1st map is not interpretable
  - Some can be modeled
  - None can be modeled
  - Phase refinement attempts to improve it

Information that can be used

- Partial model
- Constraint that two identical subunits should have same electron density
  - When not related by crystallographic symmetry
- Map features common to all protein crystals
  - Solvent regions flatter
  - Expected shape of density
  - Histogram of density levels

Density Modification and More

- Averaging, solvent flattening are examples of "Density modification"
- Something gained by merely modifying map
  - Symmetry averaging reduces noise
- More gained by requiring phases to be consistent with the constraint

Phase changes

- Consider:
  - Fourier transform: $F, \phi \rightarrow$ map
  - Inverse transform: map $\rightarrow$ same $F, \phi$.  
    (Not doing anything)
- Now Consider:
  - Fourier transform: $F, \phi \rightarrow$ map
  - Modify map $\rightarrow$ map' (symmetry, solvent flatten)
  - Inverse transform: map' $\rightarrow$ $F, \phi'$ (changed)
    - FT again: $F', \phi' \rightarrow$ map
    - Map would fit constraints exactly
    - (But actually, can do a lot better...)
  - Note that both $F$ & $\phi$ have changed
- Expected $\phi$ to change
- $F$ was observed – probably should not be changing
Phase combination

- New Regime:
  - Fourier transform: $F, \phi \rightarrow \text{map}$
  - Modify map $\rightarrow \text{map}'$ (symmetry, solvent flatten)
  - Inverse transform: $\text{map}' \rightarrow F', \phi'$ (changed)
  - Discard $F'$.
  - Use original $|F|$ w/ modified $\phi'$.
  - FT: $|F|, \phi' \rightarrow \text{map}''$
  - Fits constraints better than map, but not like $\text{map}'$.
  - Inverse transform again: $\text{map}'' \rightarrow F'', \phi''$.
  - Have further improved the phases.
- Cycle until no further change in phases.

End Point of Phase Refinement

- Map consistent with:
  - Constraints
    - Symmetry, solvent flattening, partial model...
  - Observed amplitudes

Phase Refinement by Density Modification

Constraints that are commonly imposed:
- Solvent flattening / flipping
- (Histogram matching)
- Symmetry averaging

Density modification 1 - Solvent Flattening

- Solvent molecules more motile
  - Smeared at high resolution
  - Solvent regions should be ~ featureless = "flat".
  - Phase errors $\rightarrow$ errors in all parts of map
  - Solvent regions may not start flat
  - How can we change phases to maximize the flatness?

Solvent Flattening B.-C. Wang implementation

- Determine solvent region in map
- Change density to average
- FT-invert map $\rightarrow |F_{\text{map}}|, \phi_{\text{map}}$
- Discard $|F_{\text{map}}|$: Combine $\phi_{\text{map}}$ with $|F_{\text{o}}|, \phi_{\text{experimental}}$
- Calculate a new map
  - Flatter, but not flat
  - Repeat the process

How to determine solvent region -- Premise

- Need to know which areas to flatten.
- Solvent electron density
  - Few features
  - Some density everywhere
  - Low average value
- Protein regions
  - Very High where protein atoms
  - Lower than solvent between protein atoms
  - Average higher than solvent
Determination of Protein-Solvent Boundary

- Wang (1985)
  - User defines "solvent fraction", S.
  - Locally average density
  - Weighted average
  - Smeared over 10Å radius
  - Designate lowest S fraction as solvent

- Leslie (1987)
  - Smearing density is a convolution with weighting function.
  - Scalar product in reciprocal space.
  - Weighting function is centrosymmetric
  - Convolution is scalar multiplication - simple
  - Attenuate |F|'s
  - FT → smeared map
  - Then like Wang (1985)

Solvent Flattening - Summary

- Can be applied to all proteins
- Sometimes ambiguous map → interpretable.

Symmetry Averaging

A powerful form of density modification

Source of the Information - Redundancy!

- Diffraction = continuous molecular transform sampled at lattice points
- ½ information to reconstruct - missing phases
- 2nd crystal:
  - Transform sampled @ different pts.
  - Information to calculate phases
    - in principle
  - Multiple crystals = internal symmetry
  - Multiple copies of molecule in crystal a.u.:
    - Unit cell bigger ⇒ more reflections
    - Same information needed to solve unique part

History

- Reciprocal space methods developed by Rossmann, Blow, Crowther, Main et al., 1960's
- Potential realized when a real-space equivalent was formulated (Bricogne, 1976)
- Slow realization - multiple copies advantageous
  - 1980's: more structures determined w/ NCS
  - 1990's: many determinations w/ multiple crystals

Basic real-space algorithm

- Experimantal Amplitudes
- Initial phases
- Bricogne, 1976
- Weights
- Map
- Modified map
- Back-transformation
- Phases
- Calculated Amplitudes

Recombine?
Averaging Prerequisites
- Initial phases
  - "Envelope" – which part of unit cell to average
  - Orientation of the symmetry
  - Position (origin) about which to rotate
- Usual methods
  - Rotation and Translation functions

Nomenclature
- Due to central importance of Rotation & Translation functions, often see reference to
  - "Phase refinement by Molecular Replacement"
- Confusing - Prefer
  - "Molecular replacement" for
    - use of homologous known structure for phasing
  - "Symmetry averaging" for
    - Use of symmetry redundancy for phase improvement

Envelope defines regions to average
- Average protein w/ same bit of protein
  - Not solvent, some other part of protein...
- General case - define individual protein

The Envelope Challenge
- Requires electron density map
- May start very poor
- Recognizing solvent protein boundary not trivial
  - Solvent flattening methods may help
- Distinguishing proteins near guess-work
- Need enough good guess to start
- Structure determination often blocked by poor starting envelope – envelope definition is often the most challenging step in structure determination.

Automatic Envelope Determination
- Solvent boundary à la B.C. Wang
- Trial & error
  - For each region in map...
    - Apply symmetry operator
    - If density not similar, might not be protein
- Smoothing, Overlap trimming
- Programs use one or more of these tricks
- May be able to improve envelope after some initial cycles of averaging

Current Programs do more
- Rave, DM, Solomon, Squash, Solve/Resolve
  - 2nd generation programs
- Important aspects more & more similar
- User-friendliness, portability
  - Averaging, FT's phase combination all in one program
- Incorporation of:
  - Other density modification, e.g. solvent flattening
  - Multiple crystal forms
  - Sophisticated envelopes
Power of Symmetry Averaging

- Most powerful type of phase refinement.
  - Final maps can be excellent
- Power $\propto \sqrt{\# \text{ equivalents}}$
- Phase Extension
  - Generate phases for reflections that have no phase
  - When many equivalents
  - Phases for reflections near those already phased
  - 1 or 2 lattice units
  - Extend very slowly in resolution

Summary

- Phase refinement is often required to get an interpretable map
- Maps are also improved with phases calculated from a preliminary model, but
  - 1st have to be able to build a model
  - Will consider $\phi_{\text{calc}}$ maps later
- Next workshop – building an initial model