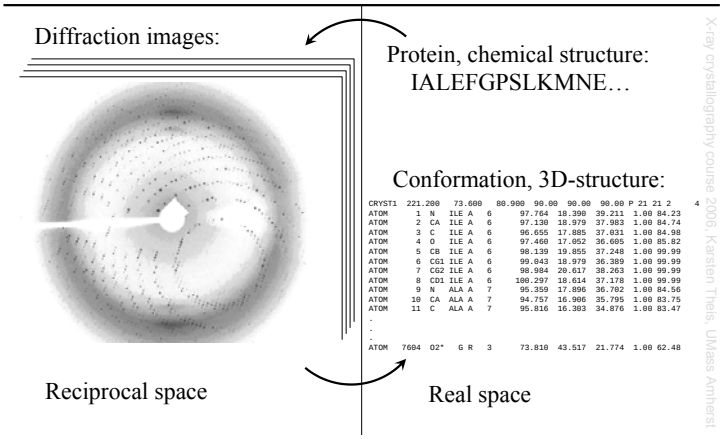
Protein structure from X-ray diffraction



Diffraction data quality

- Because of crystal symmetry and Friedel's law, the measurement of diffraction data is redundant (multiple measurements)
- Reduncancy allows images to be scaled relative to each other, even when X-ray intensity or crystal volume irradiated varies from image to image (other more complicated effects: radiation damage, absorbtion of crystal and air…)
- Multiple measurements are averaged, giving more accurate values and an estimate of the statistical error
- Parameters that are quoted in the literature: Signal/noise (or R_{merge}), completeness, redundancy

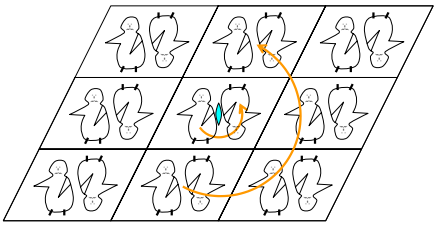
Crystal symmetry

- Symmetry operations
- Unit cell and asymmetric unit
- Symmetry elements
- Exercise: Fishes in different shapes and colors

- Symmetry of reciprocal space
- Friedel's law

Crystal Symmetry Operations

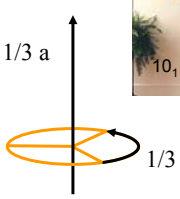
- Crystallographic symmetry operations are valid throughout the crystal
- E.g. rotation by 180º (half turn) around "◊"



Symmetry Elements

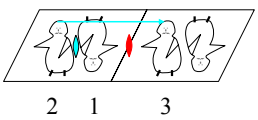
Example for symmetry element: 4-fold axis
Includes the following operations:
Rotation around the axis by 90, 180, 270 degrees

The following symmetry elements are possible for crystals with chiral content like proteins:
2,3,4,6-fold rotation axes
2_1,$3_{1/2}$,$4_{1/2/3/}$,$6_{1/2/3/4/5}$-fold screw axes
Translations a/2, b/2, c/2 in different combinations



Space group: the combined symmetry elements

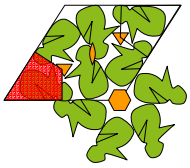
- a 2-fold axis (1◊2) and a lattice translation (2→3) combined result in a 2-fold axis (1◊3)

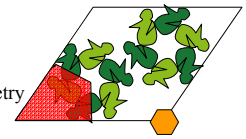


- There are 230 different space groups describing 3D symmetry
- Protein crystals are chiral; a mirror symmetry would change chirality; this limits the number of spacegroups to ~70 (so-called acentric ones)

The asymmetric unit

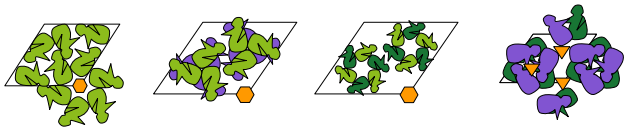
- Crystal symmetry duplicates/triplicates etc. molecules placed into the unit cell
- The unique volume of the unit cell not related by crystal symmetry is called asymmetric unit
- Non-crystallographic symmetry (NCS)
  - the asymmetric unit often contains more than one copy of the protein
  - the operations superimposing these multiple copies are called non-crystallographic symmetry operations
  - NCS-related molecules aren't identical and have different crystal environments





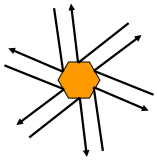
The content of the asymmetric unit

- Matthews coefficient and solvent content
  - V_M=Volume(AU) / Mw (protein in AU)
  - Observed V_M values typically are in the range of 2.0 .. 4.0 Å^3/Da
  - If we assume an average protein density (1.23 g/ml), we can calculate the solvent content from Vm
  - Make assumption of how much protein is in AU and check Vm
- Known subunit structure vs. crystal symmetry
  - e.g. hexagonal or trigonal crystals of hexameric helicase



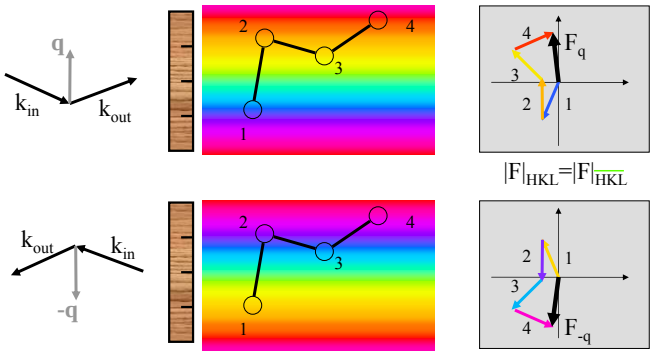
Symmetry of the diffraction data

- If the crystal obeys a certain rotation symmetry, the intensities in reciprocal space will as well. Screw axes in the crystal also result in rotational symmetry in reciprocal space.



- We can distinguish between the presence of pure rotation axes and screw axes in the crystal by so-called systematic absences of axis reflections. Axis reflections have Miller indices h00 (along a*), 0k0 (along b*), and 00l (along c*).

Friedel's law



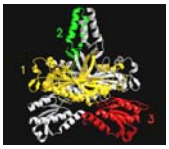
Structure Factors F_{hkl}

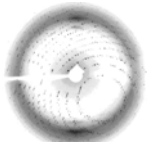
- For each reflection, the structure factor F_{hkl} gives the magnitude and the phase of the wave resulting from interference of all atoms in the unit cell
- The measured intensities I_{hkl} are proportional to the square of the structure factor magnitude $|F|_{hkl}$
- The structure factors give the information about the structure of whatever is inside the unit cell

How do we obtain
the crystal structure of a protein
from the diffraction data $|F_{hkl}|$?

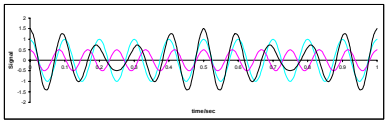
$$F(\vec{q}) = \int_{unit\ cell} \rho(\vec{r}) e^{2\pi i \vec{q}\cdot\vec{r}} d^3r$$: Fourier transform

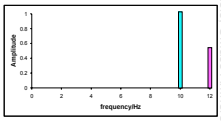
Diffraction: real space vs. reciprocal space



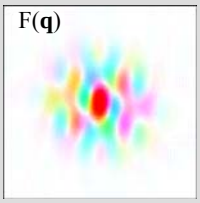


pulsed NMR, FT-IR and other spectroscopic methods: time series vs. frequencies





Fourier tour in two dimensions



Light/dark: Intensities

Colors: Phases

Molecule e⁻-density

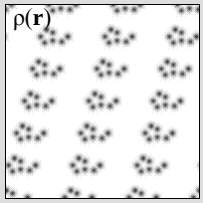
Fourier transform

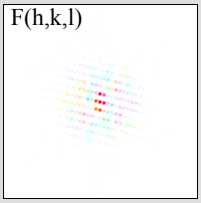
"real space"

"reciprocal space"

Crystals amplify the scattering signal

Crystal





Fourier transform

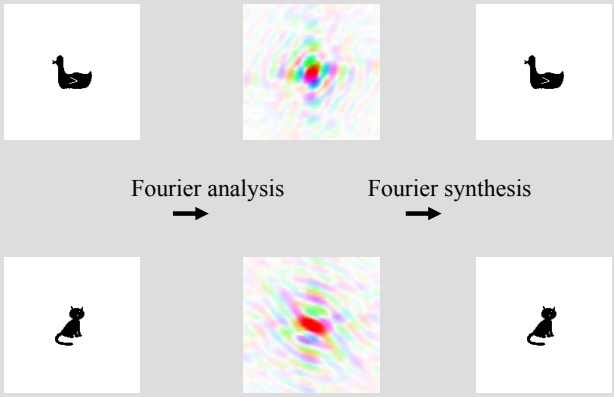
The signal from a single molecule would be much too weak to detect

The signals originating from the molecules in the crystal add up because the molecules are in identical orientation

Scattering by a crystal (called diffraction) results in a pattern with discrete spots and empty areas (another level of signal/noise increase)

Downside: diffraction image represents average structure (over time and crystal)

The Fourier transform is reversible



The Fourier duck lives at www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html

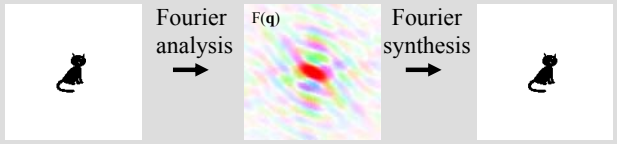
Why is solving and interpreting a crystal structure difficult?

The Phase Problem

Crystals aren't perfect because proteins are flexible

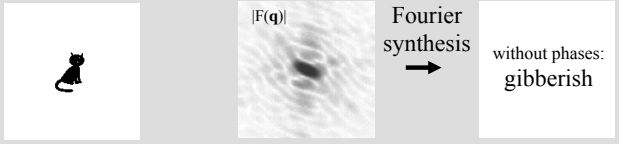
Protein conformations differ to some extent depending on their environment

Problem: without phases, no electron density



Unknown structure

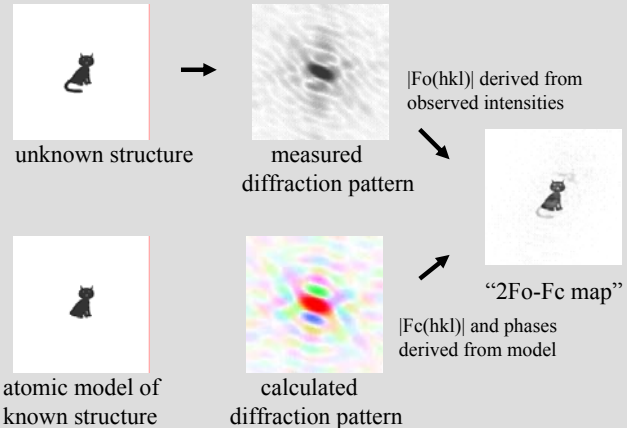
measured diffraction pattern



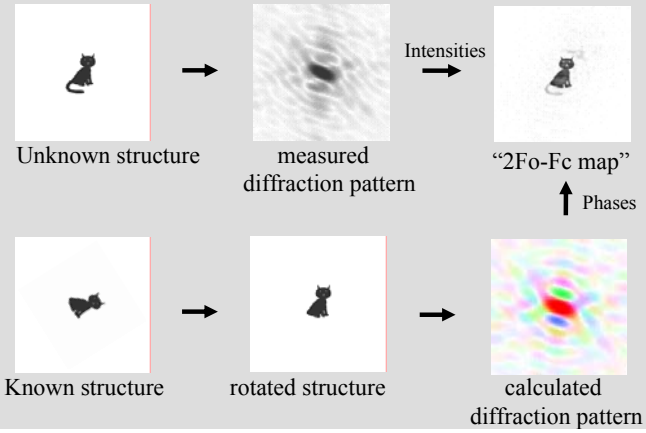
Solving the phase problem

1. Direct methods (guessing the phases)
   Works for small molecules (lots of measurements per atom in structure)
2. Patterson methods (inter-atomic vectors)
   Works for simple structures (4 atom pairs with 2 atoms, 100 atom pairs with 10 atoms
3. Model phases; molecular replacement
   Relies on partial knowledge of the structure
4. MIR/MAD techniques (*ab initio*)
   Prepare crystals that contain Se, Hg, Pt or other heavy atoms at a handful of positions in the crystal and solve that simple structure first using method 1. or 2.

Calculating phases from an atomic model



Molecular replacement: unknown structure has a distinct crystal packing



MIR/MAD method

Solve a simpler problem first: just a couple of atoms

Collect data of crystals multiple times, with slight variations

MAD
Multiwavelength Anomalous Dispersion

measure identical crystal at 3 different wavelengths near Se absorption edge

Se scattering is influenced by choice of wavelength, scattering by other atoms is not

MIR
Multiple Isomorphous Replacement

soak crystals in different substances and measure

Pt derivative | Hg derivative | native

What is the best method?

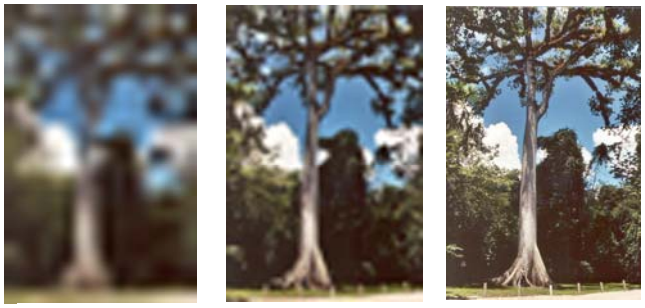
MIR/MAD works for all protein structures
but…
requires additional measurements, experience and lots of work to place atoms into density

Molecular replacement is fast
but…
large model bias in low-resolution structures
- errors are propagated
- new features are overlooked

explanation coming up..

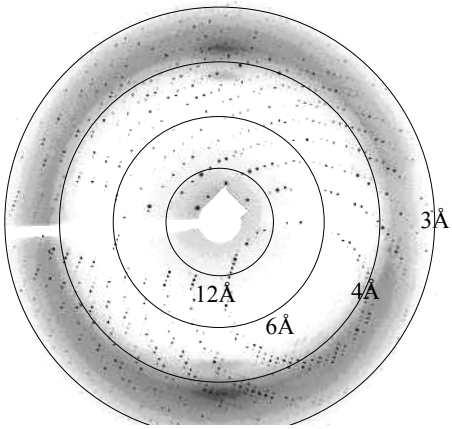
Resolution: real space

Lower resolution | Higher resolution



File size: 2 Kb | 11 Kb | 672 Kb

Resolution: reciprocal space



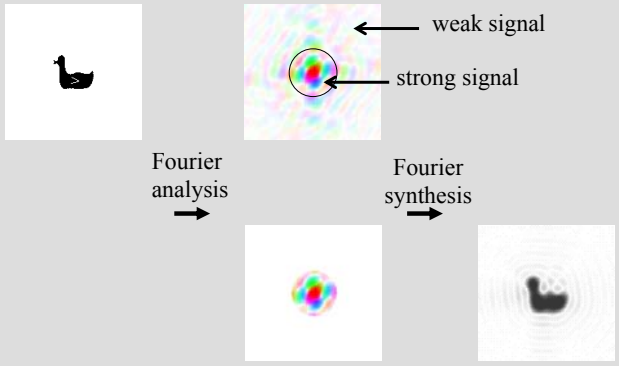
What prevents high resolution?

- electrons have distribution around atom rather than one single location (same for all crystals)
- molecules move as rigid bodies in crystal packing (depends on crystal contacts)
- atom positions change with time (local dynamic disorder)
- atom positions change from one unit cell to another (local static disorder)

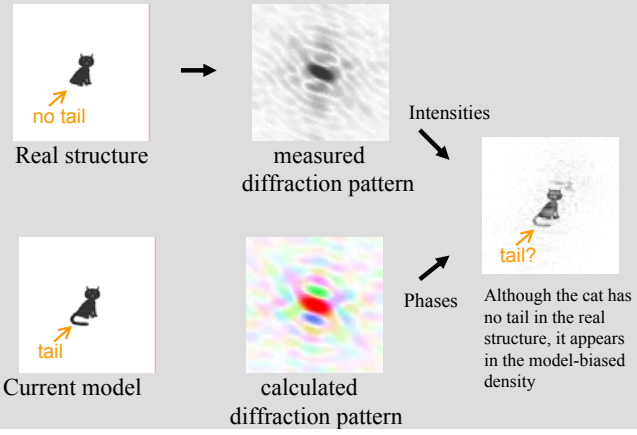
Temperature factors (B-factors) are introduced to account for disorder in the atomic model

B-factor	Displacement
20 Å^2	0.25 Å
40 Å^2	0.51 Å
80 Å^2	1.01 Å

Low resolution data: loss of detail



Low resolution leads to model bias



B-factors and disorder

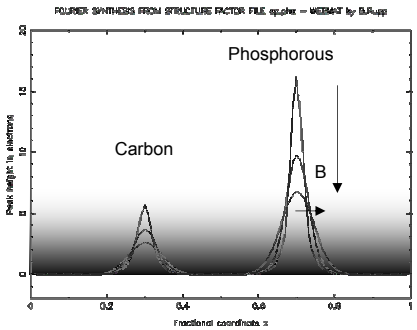
- Synonyms: Temperature factors, atomic displacement factors
- Describe how the electron density of an atom is broadened by static and dynamic disorder in the crystal
- Static disorder: distinct atomic positions in different unit cells of the crystal
- Dynamic disorder: changes in conformation over time during the measurement

Atomic displacement and B-factors

$$\sigma(\text{atomic position}) = \sqrt{[\text{B-factor} / (8\pi^2)]}$$

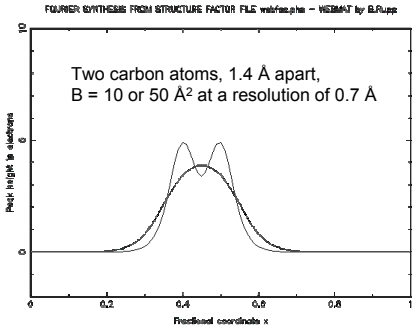
B-factor in Å^2	Atomic displacement in Å
5	0.25
20	0.50
50	0.80
80	1.01

Disorder makes interpretation more difficult

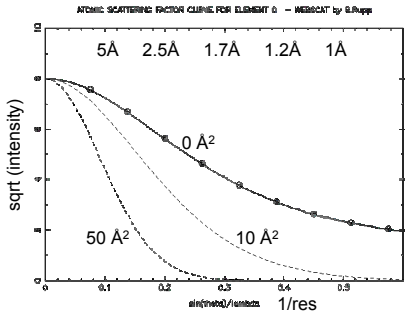


Shading represents noise due to
1) errors in experimental data
2) errors in phases derived from model

Disorder makes interpretation more difficult



Disorder makes data collection (and solving a structure) more difficult



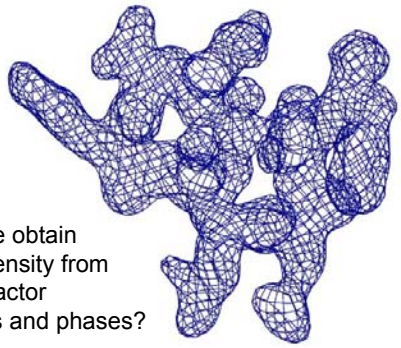
Low resolution → High resolution

The intensity of the diffraction data decreases with resolution

The higher the average B-factor, the faster the drop-off.

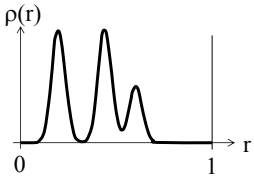
This makes it more difficult (impossible) to collect high resolution data

Electron density

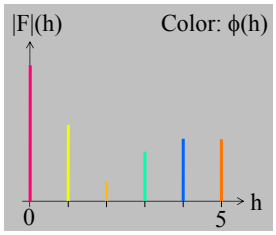


How do we obtain electron density from structure factor amplitudes and phases?

The Fourier transform



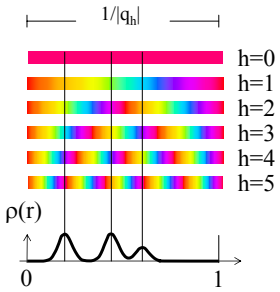
Real space



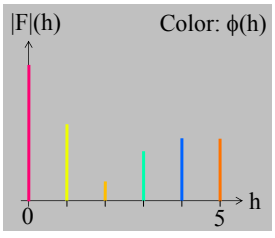
Reciprocal space

Fourier analysis
Fourier synthesis

Fourier analysis

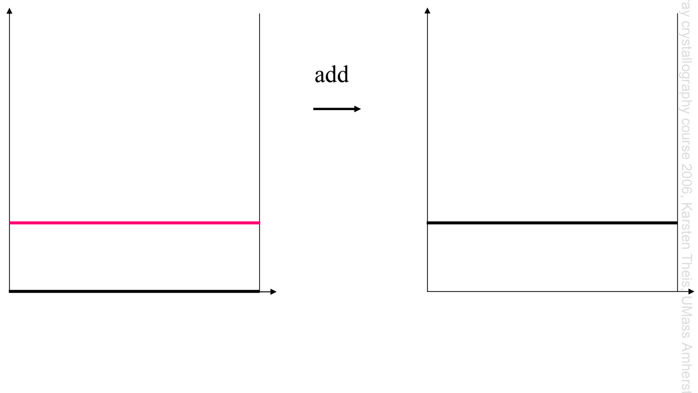


Real space

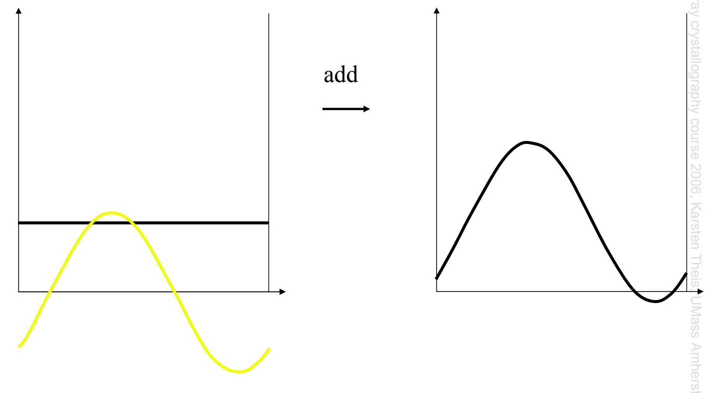


Reciprocal space

Fourier synthesis

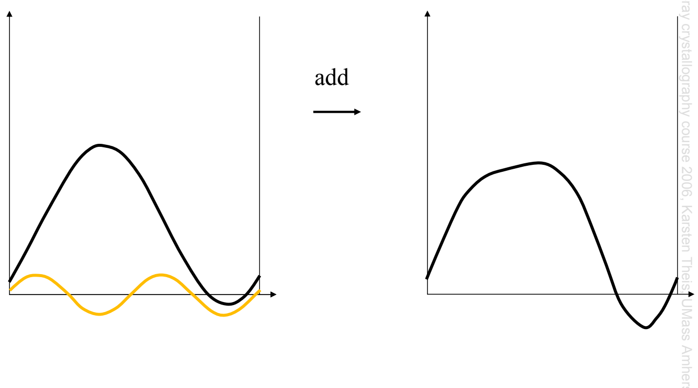


Fourier synthesis



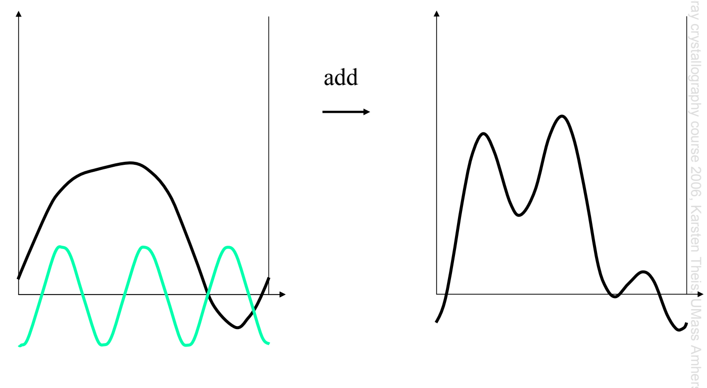
X-ray crystallography course 2006. Karsten Theis UMass Amherst

Fourier synthesis



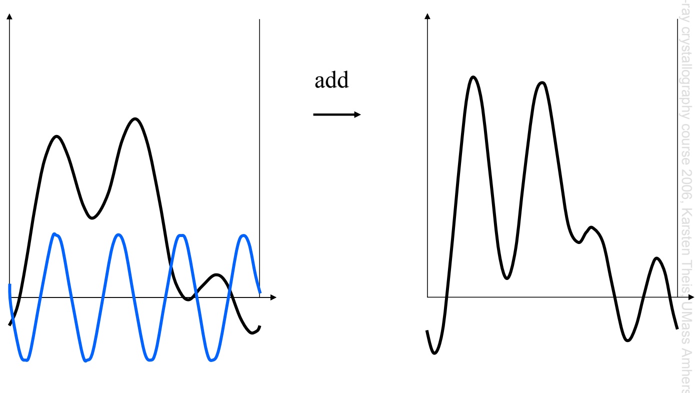
X-ray crystallography course 2006. Karsten Theis UMass Amherst

Fourier synthesis



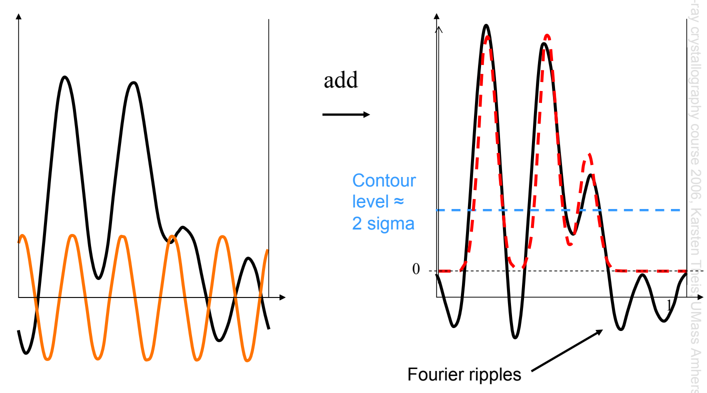
X-ray crystallography course 2006. Karsten Theis UMass Amherst

Fourier synthesis



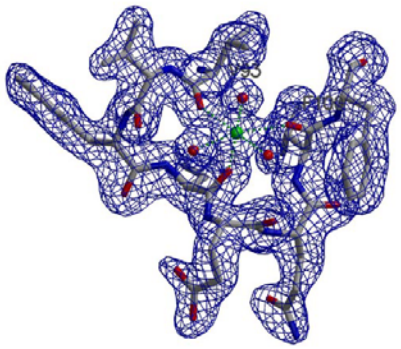
X-ray crystallography course 2006. Karsten Theis UMass Amherst

Fourier synthesis



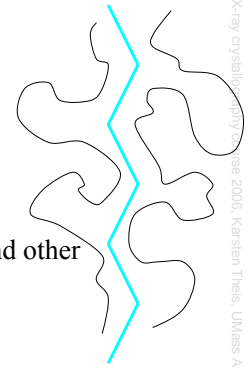
X-ray crystallography course 2006. Karsten Theis UMass Amherst

Interpreted electron density



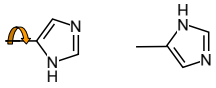
Building a model into the electron density involves interpretation and prior knowledge

- Protein/solvent regions
- C-alpha trace
- main chain, peptide direction
- sequence assignment
- side chain conformations
- disulfides, metals, glycosylation and other surprises



Questions: structure solution

1) Which pairs of amino acids have very similar electron density and are thus difficult to distinguish crystallographically? Asp/Glu ; Thr/Val ; Leu/Ile ; Lys/Met ; Asp/Asn ; Leu/Asp ; Glu/Gln
2) Which amino acids other than histidine have two side chain conformations resulting in almost identical electron density? What could help to distinguish the two possible conformations?



Assessing overall quality of structures

Quality criteria

- Resolution (related to # of observations per # of atoms)
- Rfactor and free Rfactor (compares measured intensities to those calculated from the crystallographic model; <20% is best)
- Geometry (Ramachandran plot, bond lengths and angles)
- Publication date (determines whether certain methods were available, i.e. refinement, free Rfactor + other checks)

Who checks the crystallographer?

- The reviewers
- The protein data bank
- Competing labs working on similar structures

Problematic regions within a structure

1) Region is wrong in the model, i.e. different from true conformation in the crystal (check electron density)
2) Region is absent from the crystallographic model
   a. unstructured
   b. folded but moves independently of rest of protein
3) Region was correctly modeled but…
   a. Conformation in solution is different
   b. Region is flexible in solution but not in the crystal
   c. Conformation changes upon ligand binding