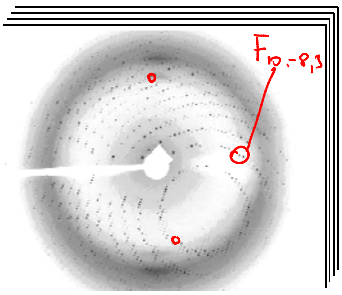
How do we obtain the crystal structure of a protein from the diffraction data $|F_{hkl}|$?

X-ray crystallography course 2007, Karsten Theis, UMass Amherst

Protein structure from X-ray diffraction

Raw data: Diffraction images



Crystallize a protein with known chemical structure:

MSALEFGPSLKMNE...

Conformation, 3D-structure:

```
CRYST1  221.200   73.600   80.900  90.00  90.00  90.00 P 21 21 2     4
ATOM      1  N   ILE A   6      97.764  18.390  39.211  1.00 84.23
ATOM      2  CA  ILE A   6      97.130  18.979  37.983  1.00 84.74
ATOM      3  C   ILE A   6      96.655  17.885  37.031  1.00 84.98
ATOM      4  O   ILE A   6      97.460  17.052  36.605  1.00 85.82
ATOM      5  CB  ILE A   6      98.139  19.855  37.248  1.00 99.99
ATOM      6  CG1 ILE A   6      99.043  18.979  36.389  1.00 99.99
ATOM      7  CG2 ILE A   6      98.984  20.617  38.263  1.00 99.99
ATOM      8  CD1 ILE A   6     100.297  18.614  37.178  1.00 99.99
ATOM      9  N   ALA A   7      95.359  17.896  36.702  1.00 84.56
ATOM     10  CA  ALA A   7      94.757  16.906  35.795  1.00 83.75
ATOM     11  C   ALA A   7      95.816  16.303  34.876  1.00 83.47
.
.
.
ATOM   7604  O2*   G R   3      73.810  43.517  21.774  1.00 62.48
```

Reciprocal space

Real space

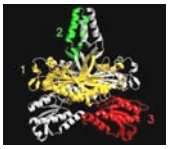
X-ray crystallography course 2007, Karsten Theis, UMass Amherst

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

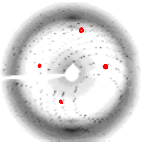
Phase

$k_i - k_{out}$

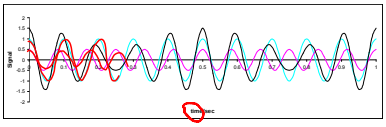
Diffraction: real space vs. reciprocal space

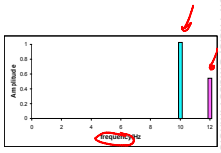


calculate



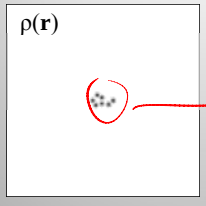
pulsed NMR, FT-IR and other spectroscopic methods: time series vs. frequencies

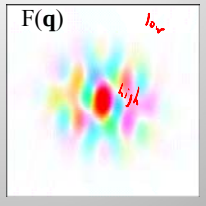




X-ray crystallography course 2007, Karsten Theis, UMass Amherst

Fourier tour in two dimensions





Light/dark: Intensities

Colors: Phases

Molecule e^--density → Fourier transform

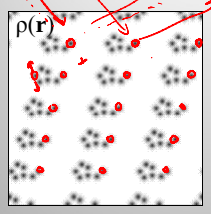
"real space" "reciprocal space"

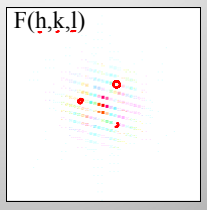
This tour is from Kevin Cowtan's picture book of Fourier at www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html

X-ray crystallography course 2007, Karsten Theis, UMass Amherst

Crystals amplify the scattering signal

Crystal





Fourier transform

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

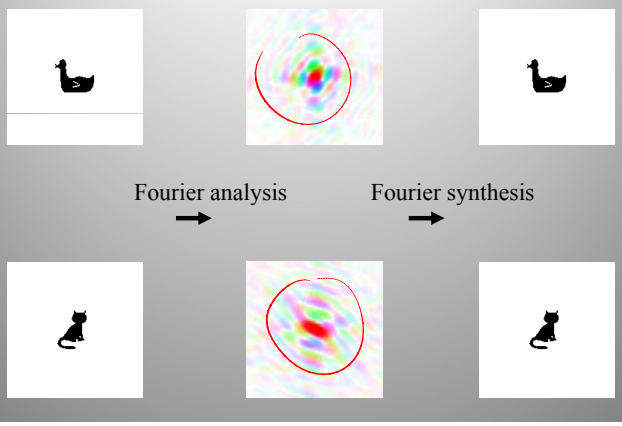
The signals originating from the molecules in the crystal add up because the molecules are in identical orientation

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

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

X-ray crystallography course 2007, Karsten Theis, UMass Amherst

The Fourier transform is reversible



Fourier analysis →

Fourier synthesis →

X-ray crystallography course 2007, Karsten Theis, UMass Amherst

Questions: X-ray diffraction

1) Why don't we see a diffraction pattern in a medical X-ray image?
2) Why can we distinguish between bones and soft tissue in a medical X-ray image?
3) Why don't we see a diffraction pattern when observing crystals under the light microscope?
4) In a protein with 10000 atoms, how many atoms contribute to a given diffraction spot?

X-ray crystallography course 2007, Karsten Theis, UMass Amherst

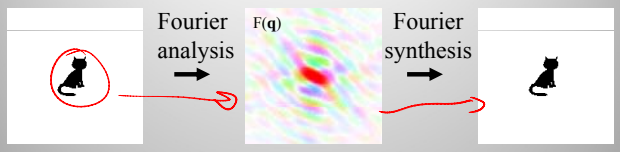
Why is solving and interpreting a crystal structure difficult?

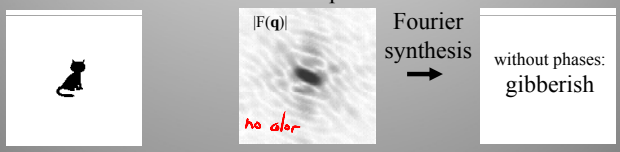
The Phase Problem

Crystals aren't perfect because proteins are flexible

Protein conformations differ to some extent depending on their environment

Problem: without phases, no electron density





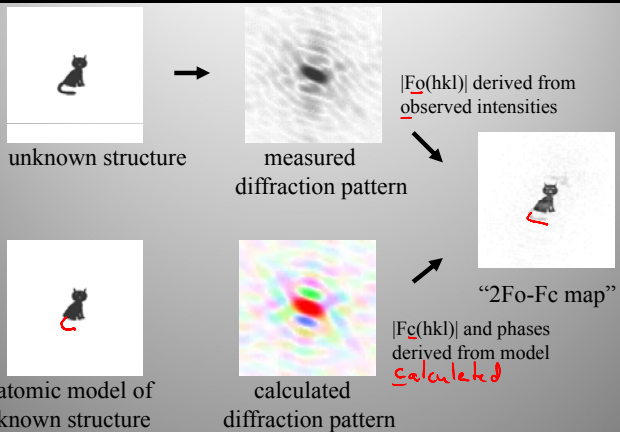
X-ray crystallography course 2007, Karsten Theis, UMass Amherst

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.

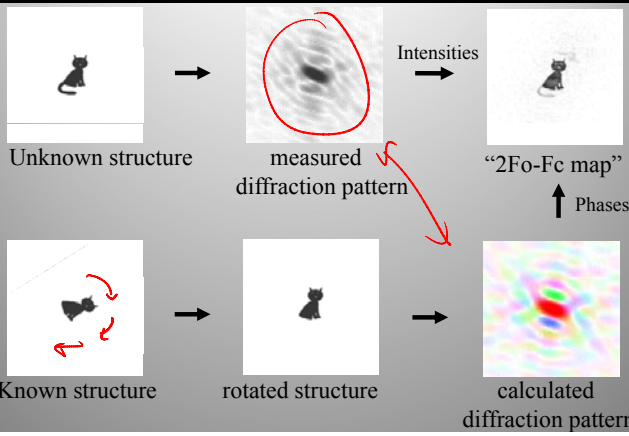
X-ray crystallography course 2007, Karsten Theis, UMass Amherst

Calculating phases from an atomic model



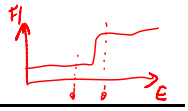
X-ray crystallography course 2007, Karsten Theis, UMass Amherst

Molecular replacement: unknown structure has a distinct crystal packing



X-ray crystallography course 2007, Karsten Theis, UMass Amherst

MIR/MAD method

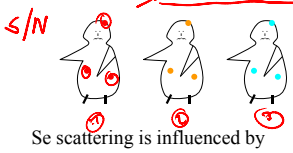


Solve a simpler problem first: just a couple of atoms

Collect data of crystals multiple times, with slight variations

MAD
Multiwavelength Anomalous Dispersion

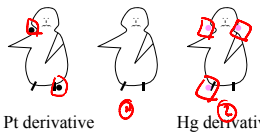
measure identical crystal at 3 different wavelengths near Se absorption edge



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

native

Hg derivative

X-ray crystallography course 2007, Karsten Theis, UMass Amherst

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

X-ray crystallography course 2007, Karsten Theis, UMass Amherst

Crystal imperfection (disorder), not wavelength, limits resolution

The electron density, averaged over time and over the crystal, is more or less "fuzzy" depending on crystal quality

- 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)

X-ray crystallography course 2007, Karsten Theis, UMass Amherst

Disorder in real space

Lower resolution — Higher resolution

File size: 2 Kb

11 Kb

672 Kb

X-ray crystallography course 2007, Karsten Theis, UMass Amherst

B-factors and disorder

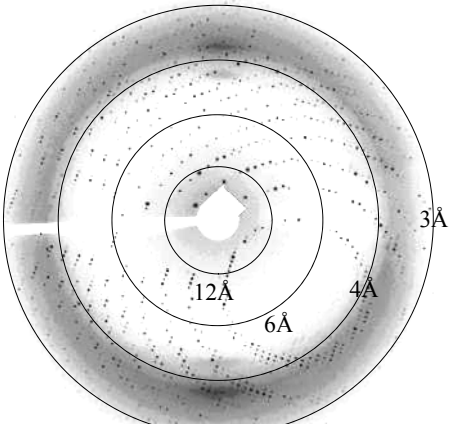
- Synonyms: Temperature factors, atomic displacement factors
- B-factors describe how the electron density of an atom is broadened by static and dynamic disorder in the crystal

Static disorder: distinct atomic positions in different unit cells of the crystal
Dynamic disorder: changes in conformation over time during the measurement

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

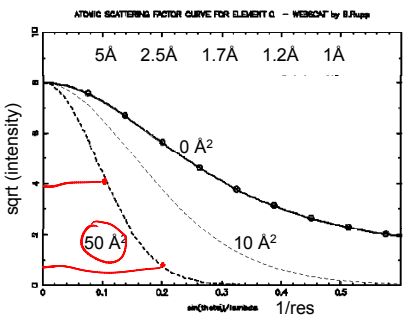
X-ray crystallography course 2007, Karsten Theis, UMass Amherst

Disorder prevents collection of high resolution data in reciprocal space



X-ray crystallography course 2007, Karsten Theis, UMass Amherst

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



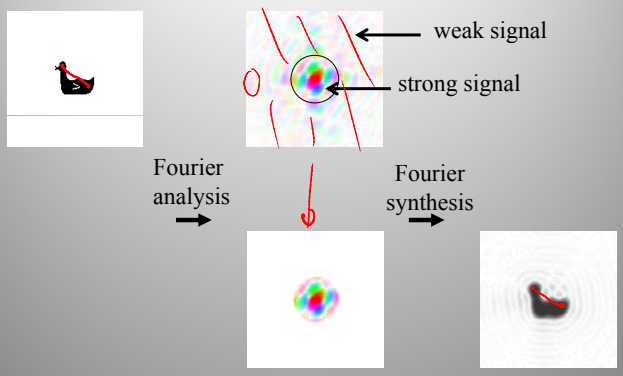
The intensity of the diffraction data decreases with resolution

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

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

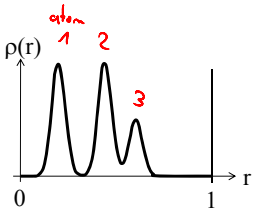
X-ray crystallography course 2007, Karsten Theis, UMass Amherst

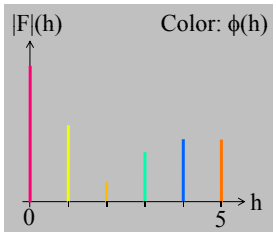
Low resolution data: loss of detail



X-ray crystallography course 2007, Karsten Theis, UMass Amherst

The Fourier transform (in 1D)





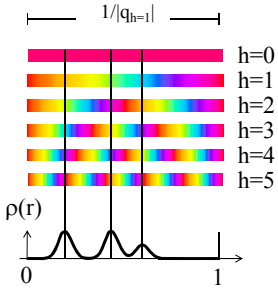
Real space

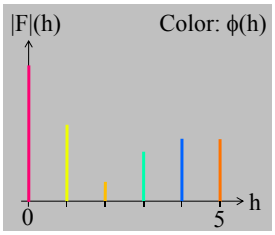
Reciprocal space

Fourier analysis ⇄ Fourier synthesis

X-ray crystallography course 2007, Karsten Theis, UMass Amherst

Fourier analysis





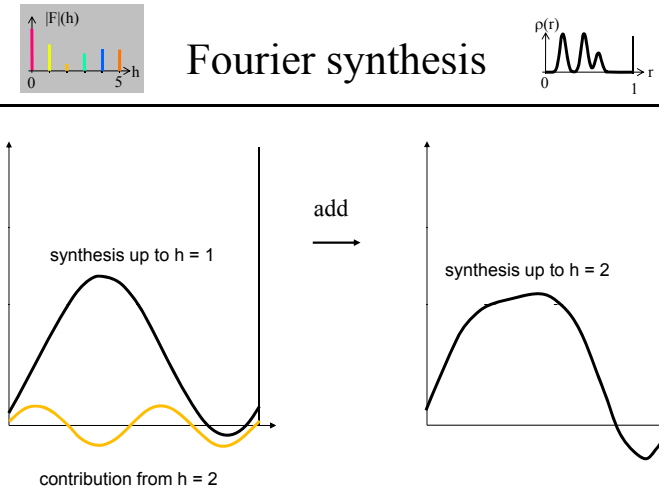
Real space

Reciprocal space

What resolution does h=5 correspond to (please estimate from atomic distances)?

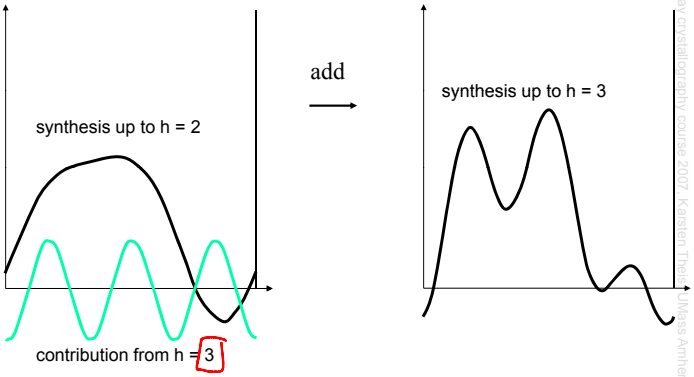
X-ray crystallography course 2007, Karsten Theis, UMass Amherst

Fourier synthesis

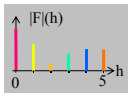


X-ray crystallography course 2007, Karsten Theis, UMass Amherst

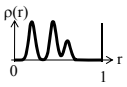
Fourier synthesis

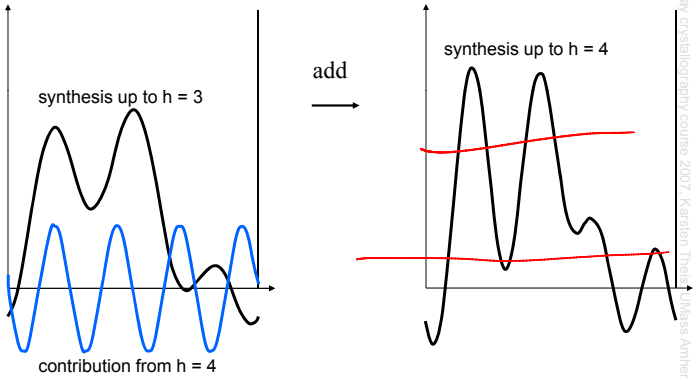


X-ray crystallography course 2007, Karsten Theis, UMass Amherst

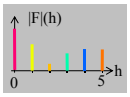


Fourier synthesis

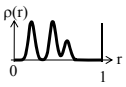


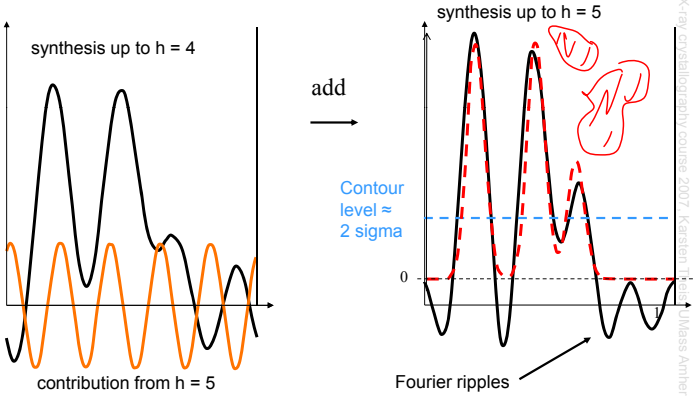


X-ray crystallography course 2007, Karsten Theis, UMass Amherst



Fourier synthesis

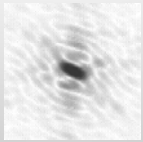




X-ray crystallography course 2007, Karsten Theis, UMass Amherst

Low resolution leads to model bias

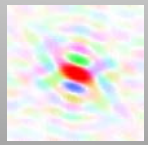
Real structure



measured diffraction pattern

Intensities

Current model



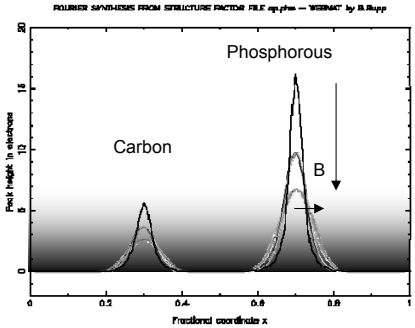
calculated diffraction pattern

Phases

Although the cat has no tail in the real structure, it appears in the model-biased density

X-ray crystallography course 2007, Karsten Theis, UMass Amherst

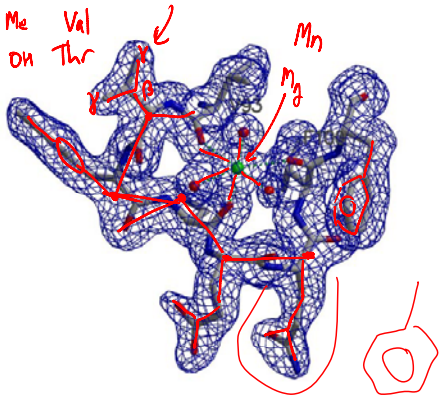
Disorder makes interpretation more difficult



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

X-ray crystallography course 2007, Karsten Theis, UMass Amherst

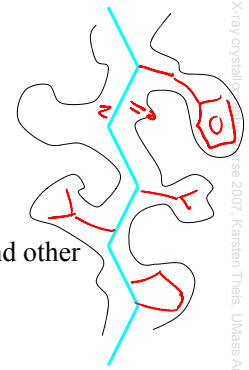
Interpretation of electron density



X-ray crystallography course 2007, Karsten Theis, UMass Amherst

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

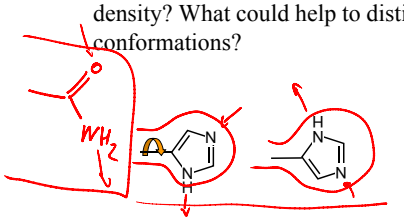
- Protein/solvent regions
- C-alpha trace
- main chain, peptide direction
- sequence assignment
- side chain conformations
- disulfides, metals, glycosylation and other surprises

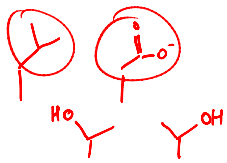


X-ray crystallography course 2007, Karsten Theis, UMass Amherst

Problem set 9: interpreting electron density

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?





X-ray crystallography course 2007, Karsten Theis, UMass Amherst

Assessing overall quality of structures

Quality criteria
- Resolution (related to # of observations per # of atoms)
- R-factor and free R-factor (compares measured intensities to those calculated from the crystallographic model)
- 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

X-ray crystallography course 2007, Karsten Theis, UMass Amherst

Spotting regions that are problematic

trivial ↔ subtle

1) Region is absent from the crystallographic model
   a. unstructured
   b. folded but moves independently of rest of protein
2) Region has high B-factors
3) Region is wrong in the model, i.e. different from true conformation in the crystal (check electron density)
4) Region is ordered and 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

X-ray crystallography course 2007, Karsten Theis, UMass Amherst

Critical use of diffraction data

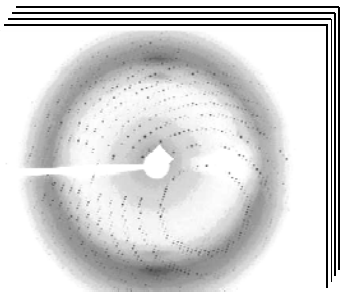
- The atomic model is an interpretation of the electron density that in turn is based partially on the atomic model (model bias)
- High resolution and low R-factors → reliable and accurate structures

- Be aware of possible differences between a protein in solution and in different crystal environments
- Expect structural changes upon ligand binding or other cellular events
  - Hydrophobic cores of structural domains do not change much in different environments
  - Conformations of surface side chains and the relative orientations of structural domains connected by flexible linkers might differ from case to case
- Flexibility is expected in regions with high temperature factors
- Look for structural variablity by superimposing atomic models obtained under different conditions

X-ray crystallography course 2007, Karsten Theis, UMass Amherst

To reciprocal space and back:
For hands-on experience, enroll in the 2008 intensive course

Raw data: Diffraction images



Reciprocal space

Crystallize a protein with known chemical structure:

MSALEFGPSLKMNE…

Conformation, 3D-structure:

```
CRYST1  221.200   73.600   80.900  90.00  90.00  90.00 P 21 21 2     4
ATOM      1  N   ILE A   6      97.764  18.390  39.211  1.00 84.23
ATOM      2  CA  ILE A   6      97.138  18.979  37.983  1.00 84.74
ATOM      3  C   ILE A   6      96.655  17.885  37.031  1.00 84.98
ATOM      4  O   ILE A   6      97.460  17.052  36.605  1.00 85.82
ATOM      5  CB  ILE A   6      98.139  19.855  37.248  1.00 99.99
ATOM      6  CG1 ILE A   6      99.043  18.979  36.389  1.00 99.99
ATOM      7  CG2 ILE A   6      98.984  20.617  38.263  1.00 99.99
ATOM      8  CD1 ILE A   6     100.297  18.614  37.178  1.00 99.99
ATOM      9  N   ALA A   7      95.359  17.896  36.702  1.00 84.56
ATOM     10  CA  ALA A   7      94.757  16.906  35.795  1.00 83.75
ATOM     11  C   ALA A   7      95.816  16.303  34.876  1.00 83.47
.
.
.
ATOM   7604  O2*   G R   3      73.810  43.517  21.774  1.00 62.48
```

Real space

X-ray crystallography course 2007, Karsten Theis, UMass Amherst