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

$F(q^*) = \int_{\text{unit cell}} \rho(r) e^{2\pi i q^* \cdot r} dr$ : Fourier transform

Diffraction: real space vs. reciprocal space

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

Crystals amplify the scattering signal

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)

The Fourier transform is reversible

Fourier analysis

Fourier synthesis

Protein structure from X-ray diffraction

Raw data: Diffraction images

Crystallize a protein with known chemical structure:
MSALEFGPGLKMN...

Conformation, 3D-structure:

Fourier tour in two dimensions

Molecule $e^2$-density

Fourier transform

"real space"

"reciprocal space"

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

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

<table>
<thead>
<tr>
<th>Unknown structure</th>
<th>measured diffraction pattern</th>
<th>Fourier analysis</th>
<th>Fourier synthesis</th>
</tr>
</thead>
</table>

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

<table>
<thead>
<tr>
<th>unknown structure</th>
<th>measured diffraction pattern</th>
</tr>
</thead>
</table>

| atomic model of known structure | calculated diffraction pattern |

Molecular replacement: unknown structure has a distinct crystal packing

<table>
<thead>
<tr>
<th>Unknown structure</th>
<th>measured diffraction pattern</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Known structure</th>
<th>rotated structure</th>
</tr>
</thead>
</table>

| “2Fo-Fc map” |
|--------------|------------------|

| Intensities |
|------------|------------------|

| “2Fo-Fc map” |
|--------------|------------------|

| Phases |
|--------|------------------|

<table>
<thead>
<tr>
<th>Calculated diffraction pattern</th>
</tr>
</thead>
</table>
MIR/MAD method

Solve a simpler problem first: just a couple of atoms
Collect data of crystals multiple times, with slight variations

MAD
Multiple Anomalous Dispersion

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

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

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)

Disorder in real space

Lower resolution
Higher resolution

File size: 2 Kb
11 Kb
672 Kb

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

Disorder prevents collection of high resolution data in reciprocal space

<table>
<thead>
<tr>
<th>B-factor</th>
<th>Displacement</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 Å²</td>
<td>0.25 Å</td>
</tr>
<tr>
<td>40 Å²</td>
<td>0.51 Å</td>
</tr>
<tr>
<td>80 Å²</td>
<td>1.01 Å</td>
</tr>
</tbody>
</table>
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.

Low resolution data: loss of detail

The Fourier transform (in 1D)

Fourier analysis

Fourier synthesis

Fourier synthesis
Fourier synthesis

\[ \rho(r) \]

X-ray crystallography

synthesis up to \( h = 3 \)

contribution from \( h = 4 \)

add

synthesis up to \( h = 4 \)

contribution from \( h = 5 \)

Fourier ripples

Low resolution leads to model bias

Real structure

no tail

Current model

tail?

Phases

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

Disorder makes interpretation more difficult

Phosphorus noise due to

1) errors in experimental data

2) errors in phases derived from model

Interpretation of 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
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?

Spotting regions that are problematic

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

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 R-factor + other checks)

Who checks the crystallographer?
- The reviewers
- The protein data bank
- Competing labs working on similar structures

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 variability by superimposing atomic models obtained under different conditions

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

Raw data: Diffraction images
Crystallize a protein with known chemical structure:
MSALEFGPSLKMNE...

Conformation, 3D-structure:

<table>
<thead>
<tr>
<th>Residue</th>
<th>Main Chain</th>
<th>Side Chain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.04</td>
<td>38.04</td>
</tr>
<tr>
<td></td>
<td>3.99</td>
<td>37.98</td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>39.00</td>
</tr>
<tr>
<td></td>
<td>3.99</td>
<td>38.99</td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>39.00</td>
</tr>
</tbody>
</table>

Reciprocal space Real space