

Bioinformatics III

Analysis and prediction of 3D macromolecule structures

Experimental methods for 3D
structure determination

Saulius Gražulis
2024 m.

Methods to obtain structural information

- **X-ray neutron crystallography**
- **NMR**
- **CryoEM (Cryo Electron microscopy, multiple particle analysis, electron tomography)**
- Electron diffraction
- EM (Electron microscopy, negative staining)
- SAXS/SANS – Small Angle X-ray/Neutron Scattering
- EXAFS, XANES
- ...

CryoEM

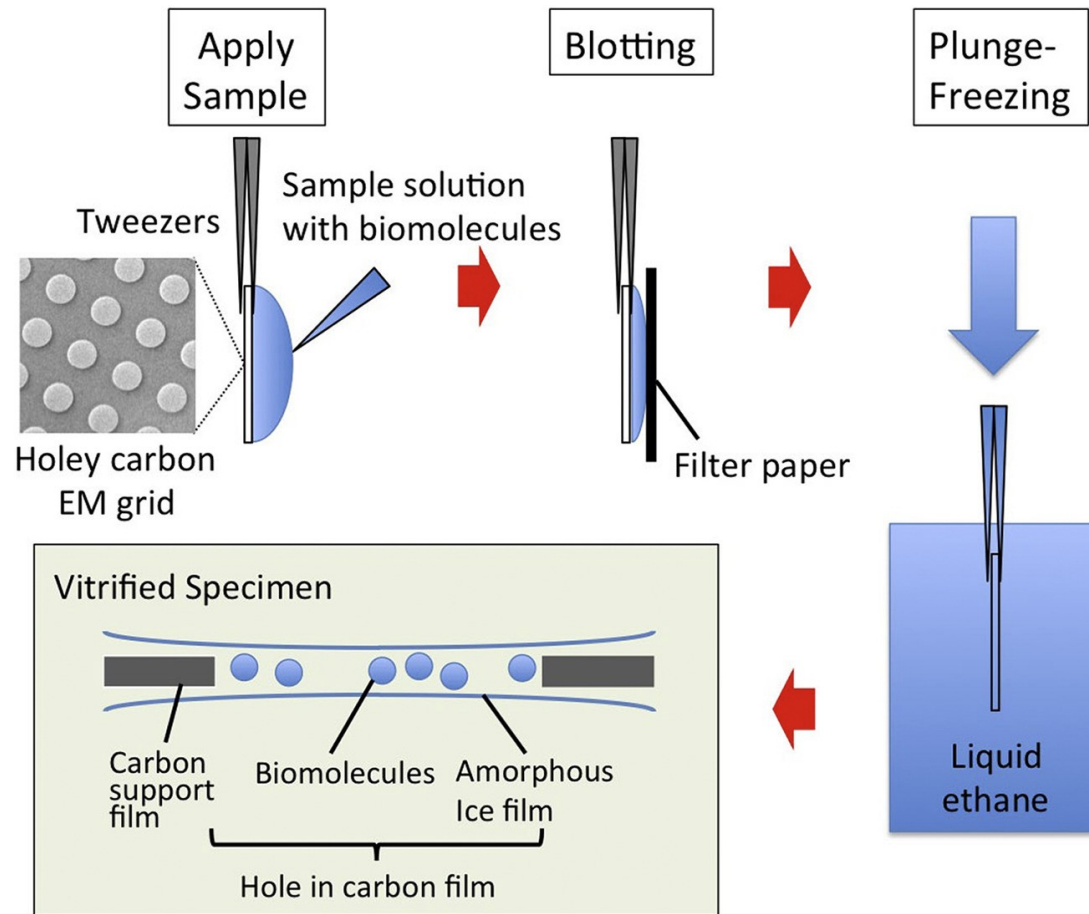
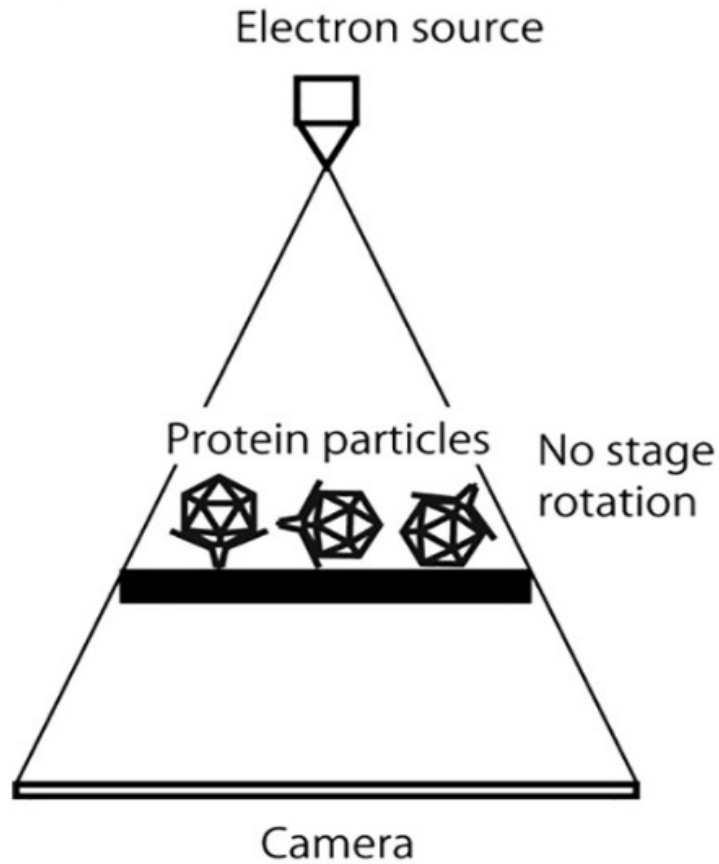
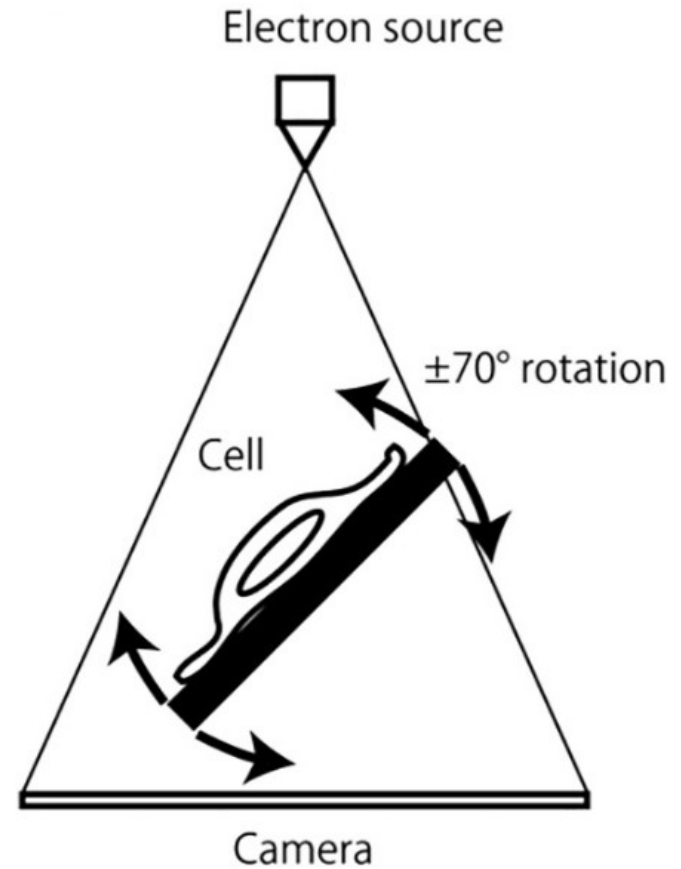


Image taken from: ([Murata & Wolf, 2018](#)) CC BY

SPA & CryoET



SPA (Single Particle Analysis)



ET (Electron Tomography)

CryoEM electron density

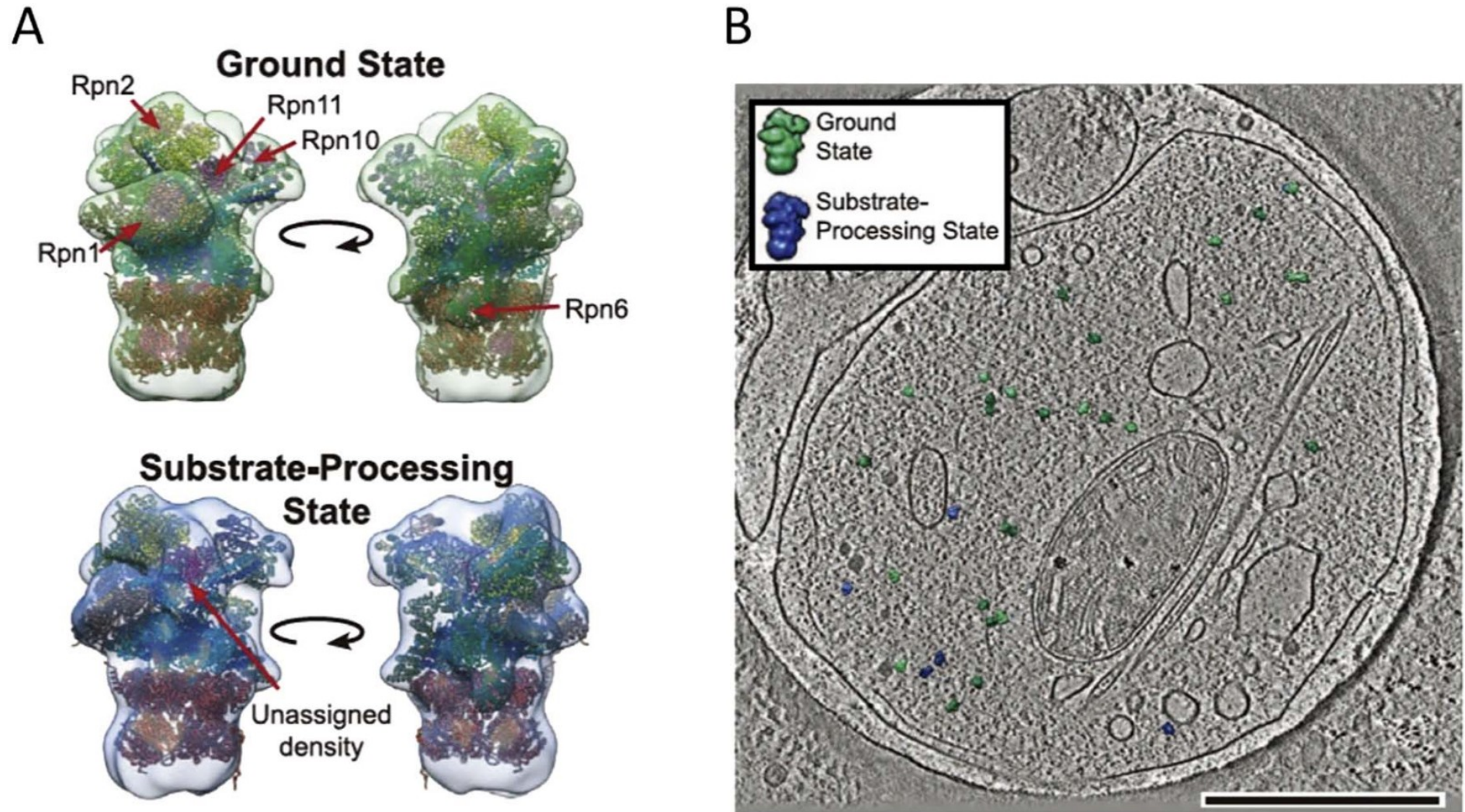
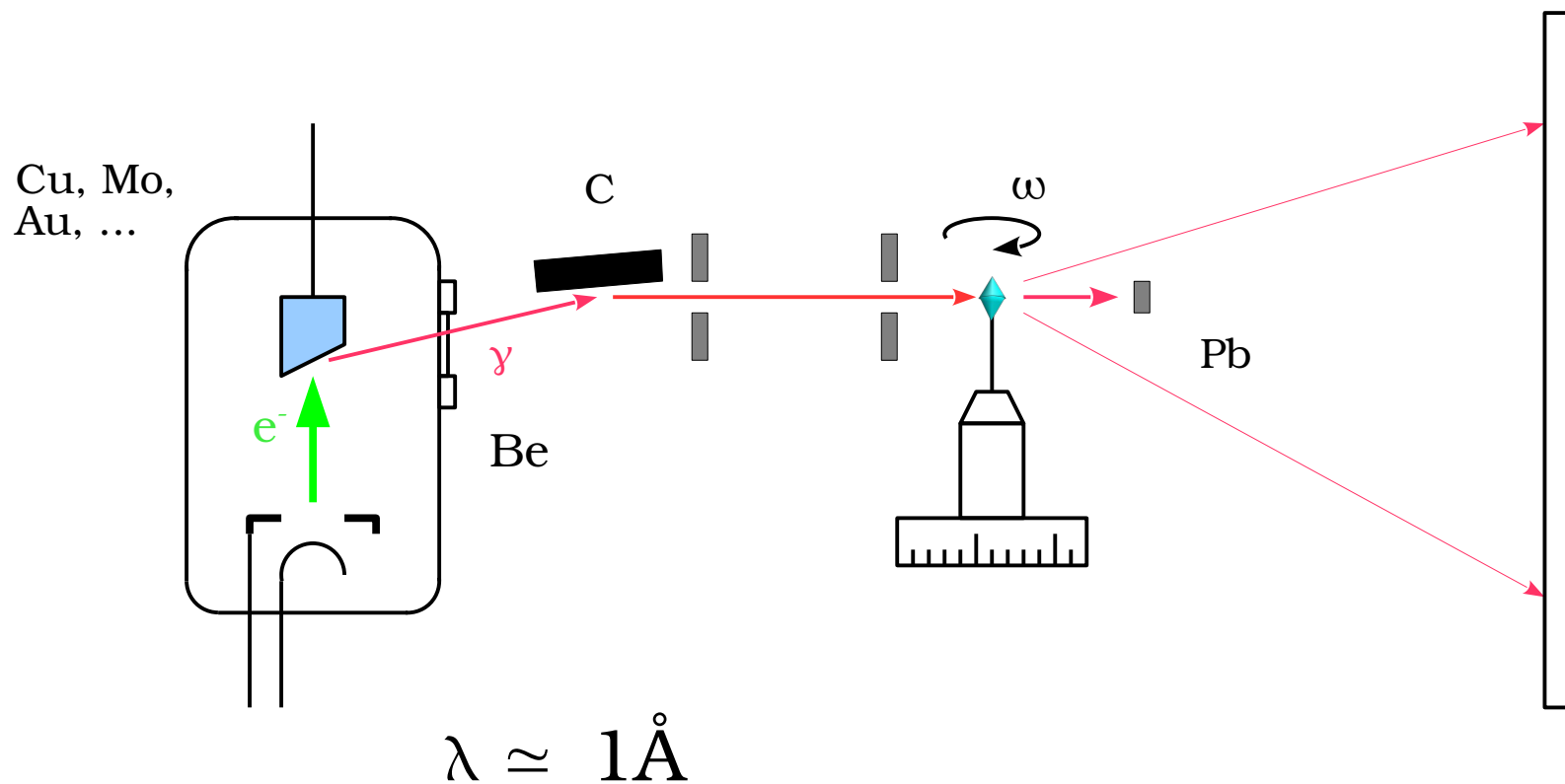


Image taken from: ([Murata & Wolf, 2018](#)) CC BY

Diffraction experiment

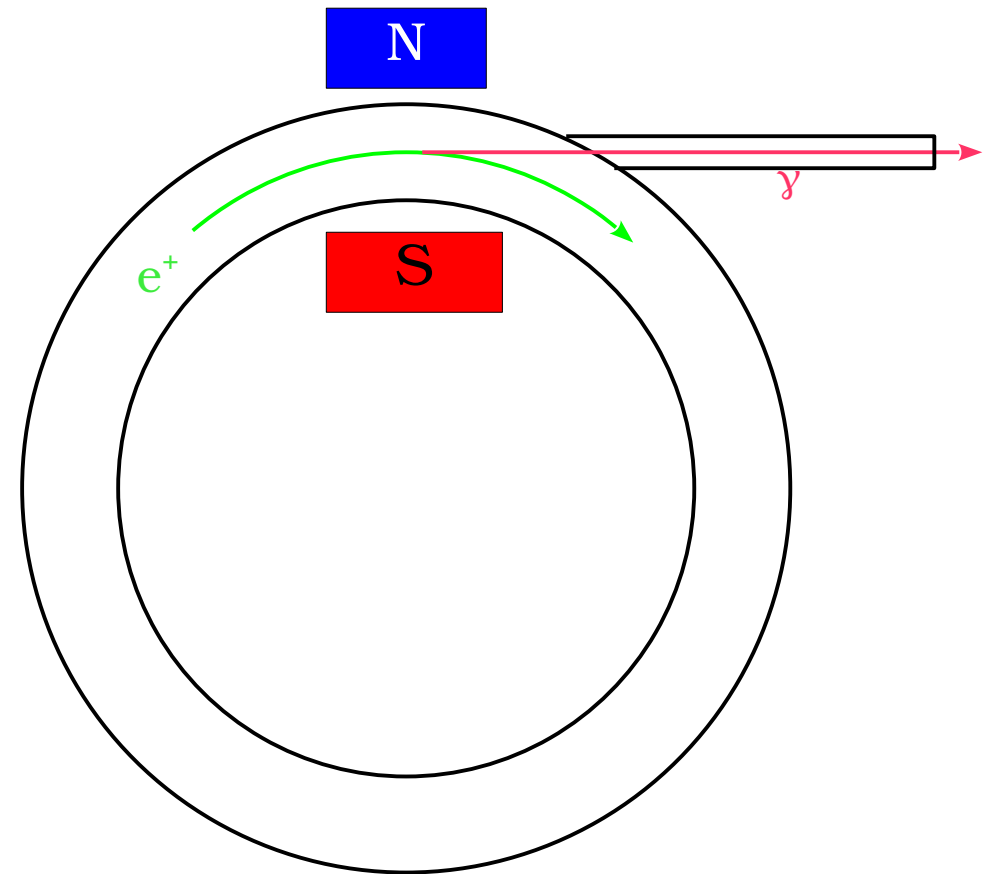
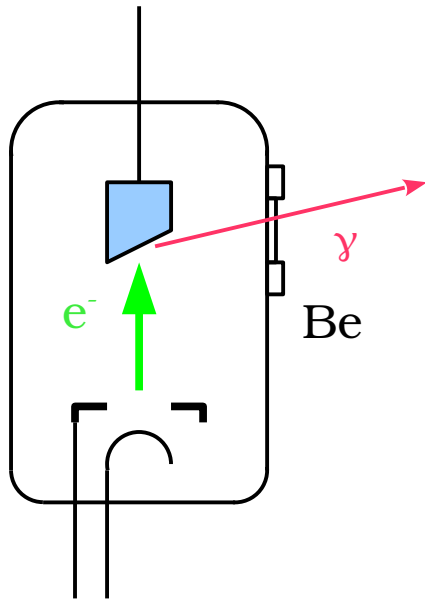


X-ray sources

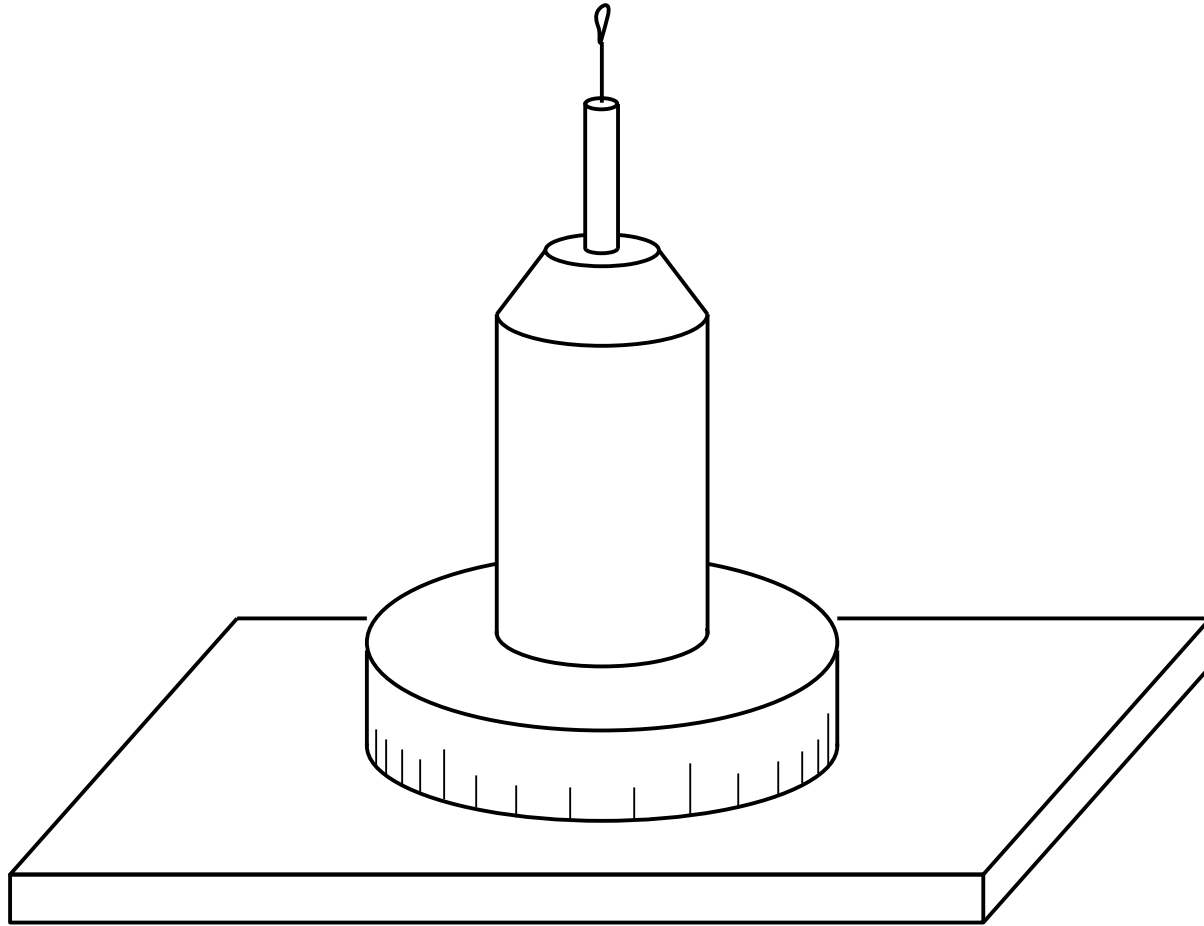
Synchrotrons (particle accelerators)

X-ray tubes

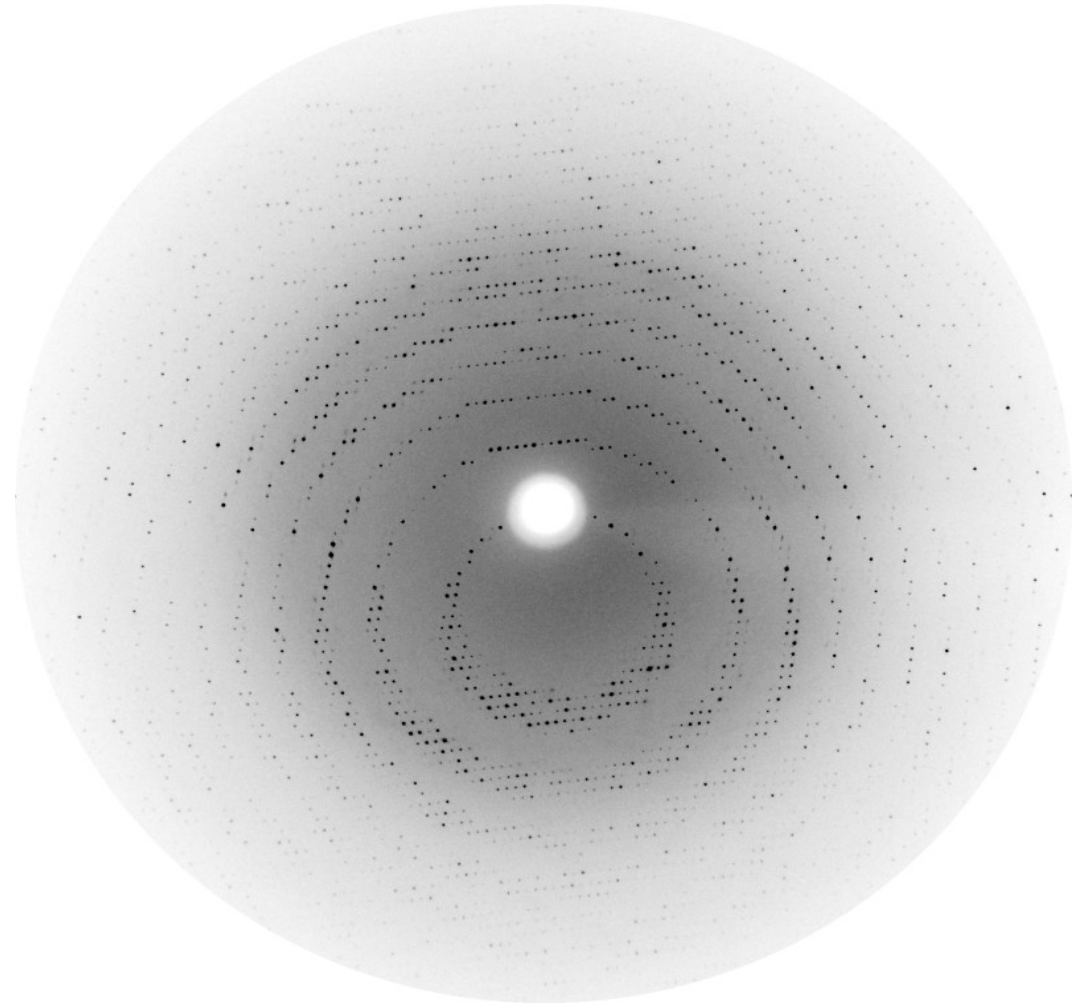
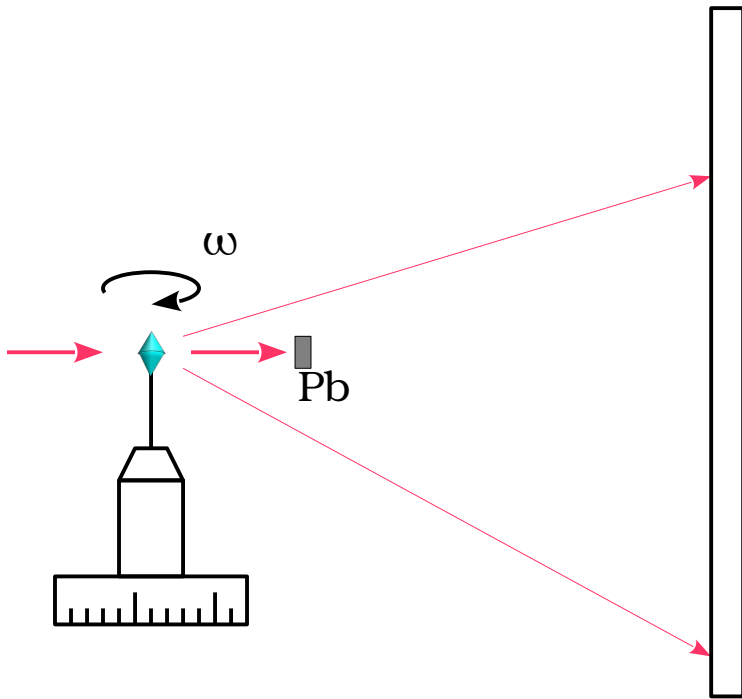
Cu, Mo,
Au, ...



Goniometer head



A diffraction image



Properties of reflections

- Each reflections has:
 - ***an amplitude*** (of the electric field);
 - and ***phase*** (delay with respect to other reflections);
- Both of these parameters depend on a ***structure factor***.

Properties of reflections

- The reflection **positions** (i.e. reflection angles) depend only on the lattice parameters;
- The intensities of the reflections depend on the **contents** of the unit cell.

A complete dataset

When a crystal is rotated, some reflections appear, and some others vanish.

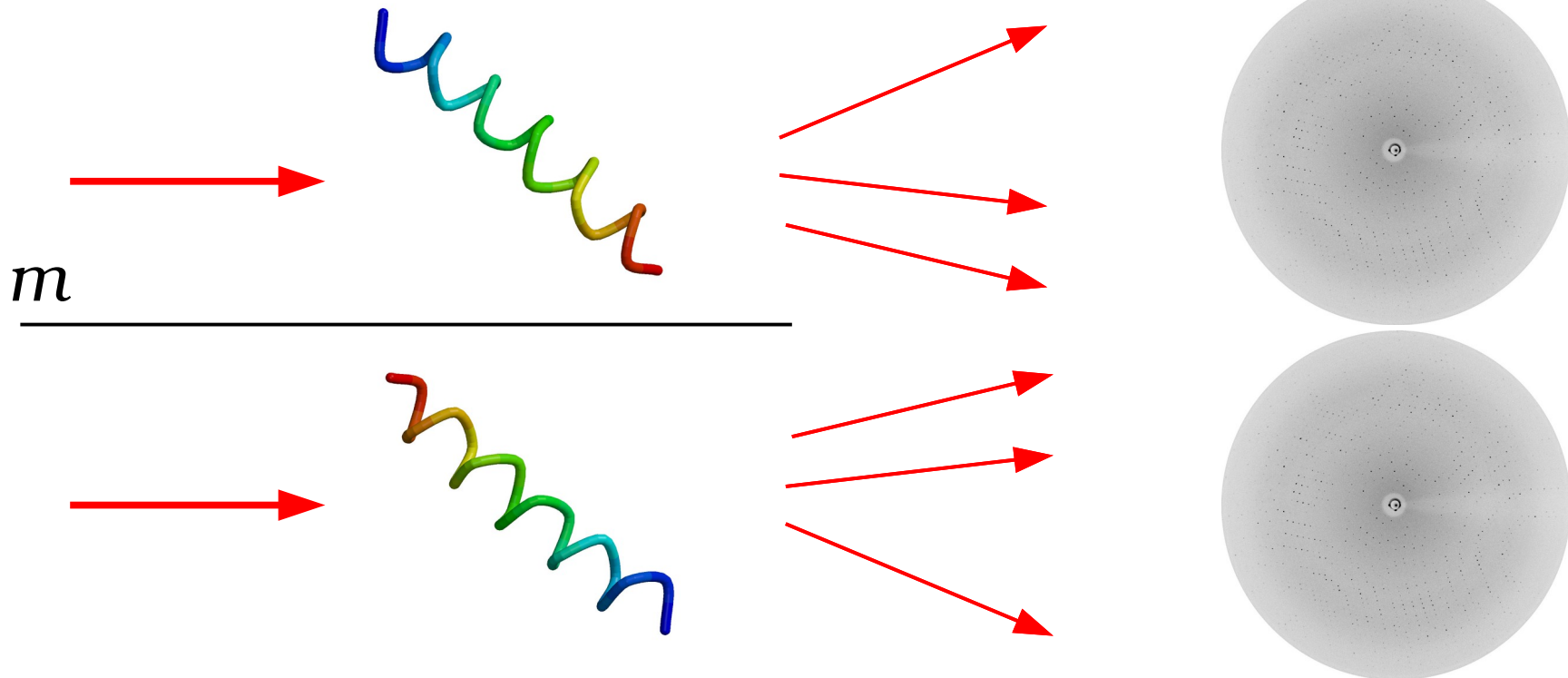
After max. 360° rotation reflections start repeating themselves (maybe sooner, especially if the crystal is highly symmetric).

Properties of a dataset

- Resolution
- Completeness
- Multiplicity
- Signal-to-noise ratio (R_{merge} , I/σ_I)

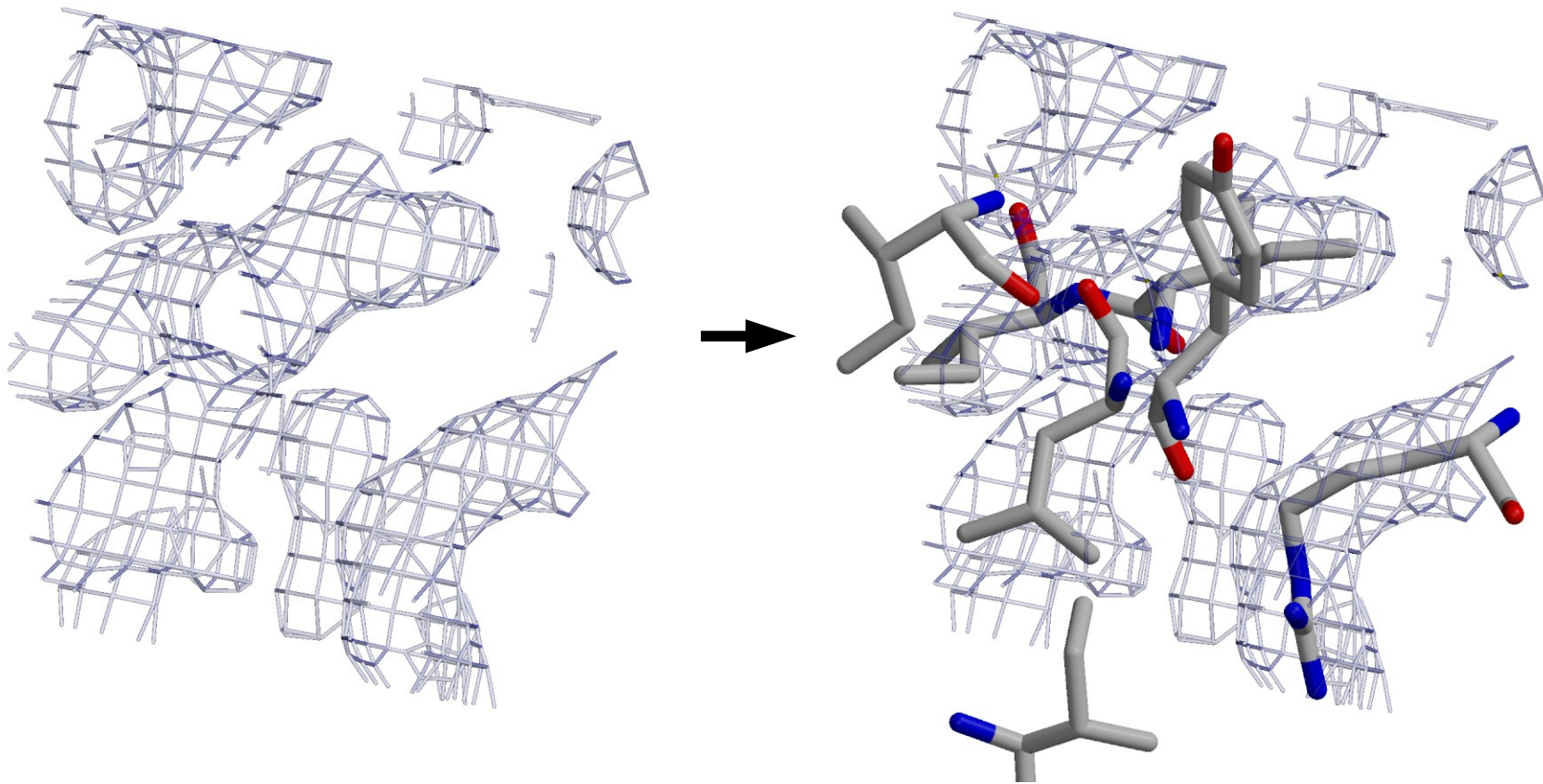
A model might not be unique!

Example:



A mirror image of a structure has (nearly) identical diffraction pattern (identical if there is no anomalous scattering)

Electron density to model



Rankinis modelio konstravimas

The image displays a screenshot of a molecular modeling software interface, likely CCP4, running on a Linux system. The main window shows a 3D wireframe model of a protein structure, with various atoms and bonds represented by blue and green lines. The interface includes several menus and command windows:

- Object_menu**: A list of options for the current object, including BUILD (on), MYS1 (off), MYS2 (off), MINI (off), REFMAC (off), WAT (on), NEG (on), HG (on), NA (off), NB (off), AONB (off), BONA (off), UNREF (off), OLD (off), GLYC (on), TRIS (on), MES (on), SRT (on), SST (on), RRT (on), DELFWT (on), and FWT (on).
- User Menu**: A list of user-defined macros and commands, including Centre_ID, Clear_ID, Clear_flags, @remap, @delpos, @delneg, @mir, @mod, @aomit, @omit, @make, @remake, @regenerate, @next-ca, @prev-ca, @next-O, @prev-O, Move_zone, Move_atom, Dial_next, Dial_previou, Yes, Refi_zone, Lego_Side_Ch, RSR_rigid, RSR_rotamer, @store, @ci, @rcsdiff, Dist_define, @on_startup, @ldmask, @ldomac, @reload, @ldbuild, and @ldmaps.
- Terminal**: A command window on the left showing the user's interactions with the software, including commands like Mol>, Sam>, Map>, and As2>.
- Bottom Panel**: A status bar at the bottom showing the current map format (red/blue) and the current map name (blue).

The interface also features a top menu bar with options like Controls, Display, Graph, Rebuild, Bones, Density, and Menus. A status bar at the bottom indicates the current map format (red/blue) and the current map name (blue).

Prior information is essential!

- Number of molecules in the asymmetric unit
- Protein size and sequence
- Structures of amino acids ;-)
- Composition of crystallisation conditions (ions, buffers)
- Protein production and purification method

Structure refinement

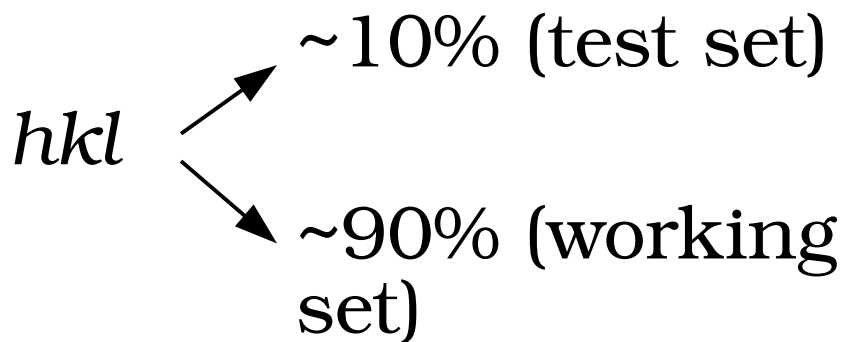
- Least squares method
- Maximum likelihood methods
- Bayesian methods

Refinement criteria

Crystallographic R-factor:

$$R_{cryst} = \sum_{hkl} |F_{hkl}^{obs} - F_{hkl}^{calc}| / \sum_{hkl} |F_{hkl}^{obs}|$$

R-free:



$$R_{free} = \sum_{hkl \in Test\ set} \dots$$

$$R_{cryst} = \sum_{hkl \in Working\ set} \dots$$

Parameters

- Atomic coordinates
 - Thermal displacement parameter(s) (B-factors)
 - Occupancies
-
- Important to know: the observation-to-parameter ratio

Structure quality indicators

- Resolution (Å)
- R factor (full, crystallographic), R_{cryst}
- „Free“ R factor, R_{free}
- $R_{\text{free}} - R_{\text{cryst}}$, $R_{\text{cryst}}/R_{\text{free}}$
- R_{cryst} and R_{free} in the outer resolution shell
- Average error of bond lengths and angles

Non-polymeric molecules

- Solvent (water!)
- Ions
- Ligands, cofactors
- Modified amino acids and/or DNA/RNA bases

Evidence of chaos

- Zero occupancies
- Huge B-factors
- Residues without the side chains

Resources in the Net

Protein Data Bank (PDB)

<http://www.rcsb.org/pdb/>

Macromolecular Structure Database

<http://www.ebi.ac.uk/msd/>

International Union of Crystallography

<http://www.iucr.org/>

Crystallography Open Database (COD)

<http://www.crystallography.net/>