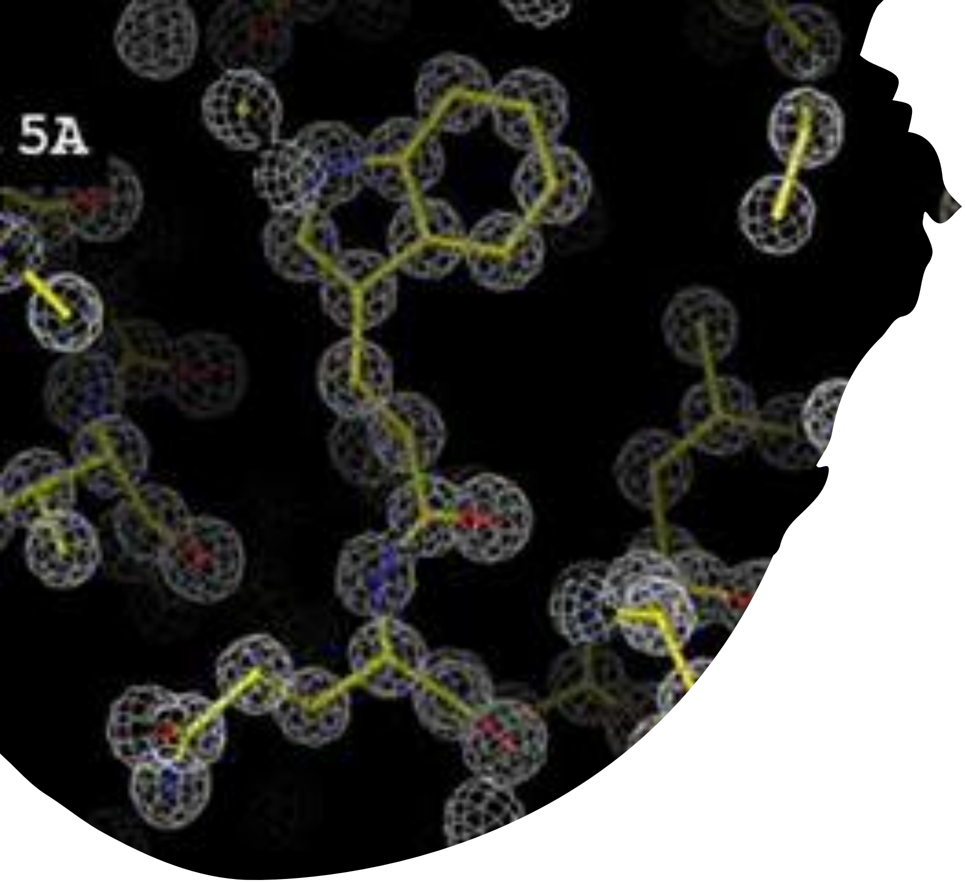
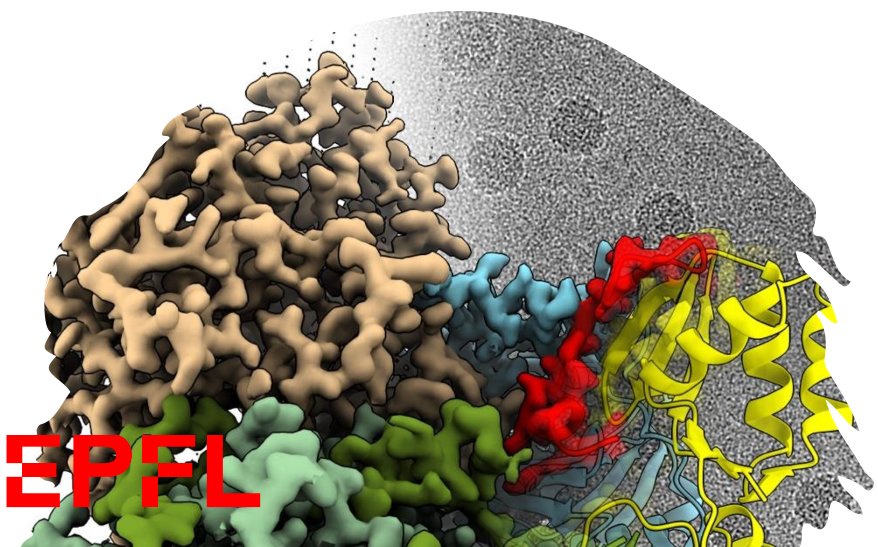




EDMS BIO-643: Integrative Structural Biology for Life Sciences

- Luciano Abriata
- Kelvin Lau
- Yoan Duhoo
- Anna-Sophia Krebs
- Jonathan Schneider
- Florence Pojer
- Guest: Henning Stahlberg

Protein Production and Structure Core Facility, SV-PTPSP

Fall Semester 2025

EPFL

EDMS BIO-643

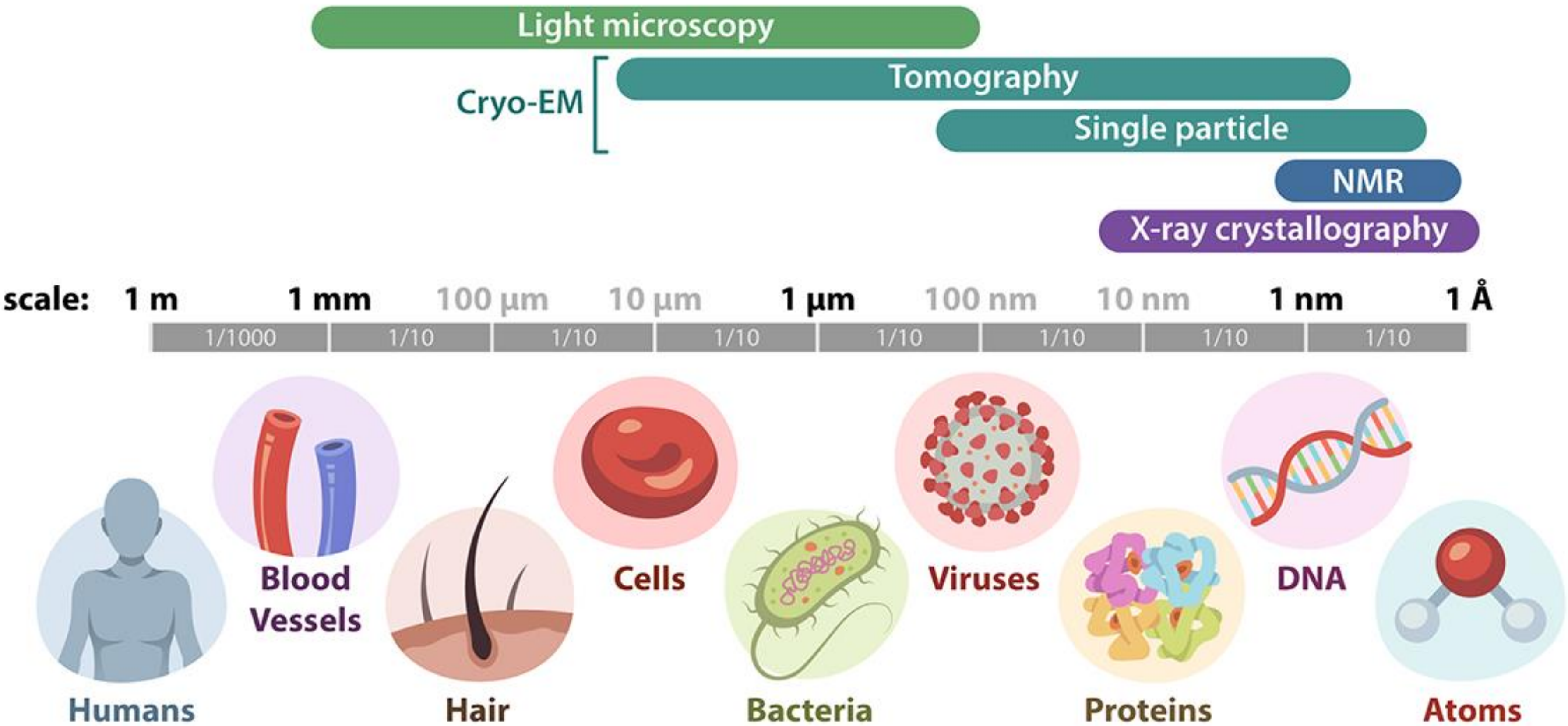
Integrative structural biology for Life sciences

Fall semester 2025

Date/Time: Thursdays 3:15 to 5pm

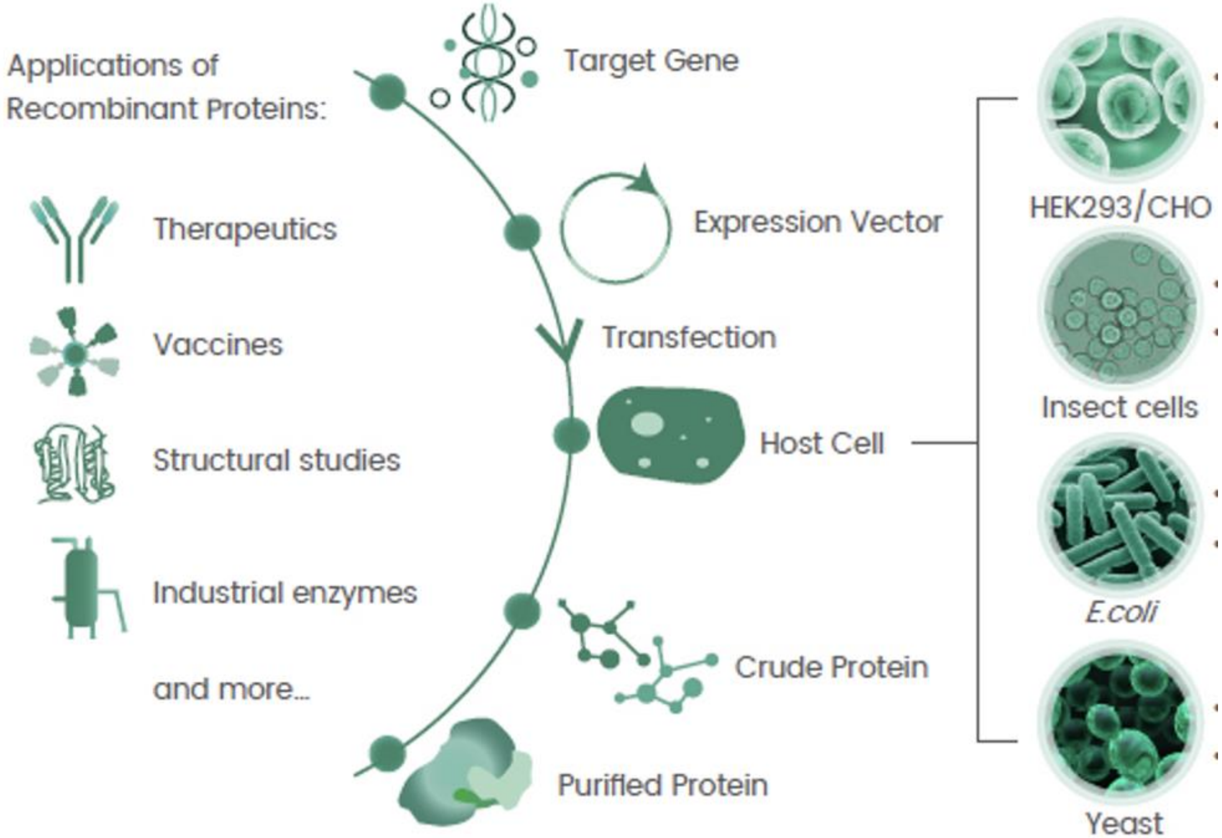
Location: AI 3142

Date	Topics	Speakers
Sept 11th	Course introduction	All teachers
Sept 18th	Modeling tools	Luciano
Sept 25th	Modeling tools -visualization in ChimeraX	Luciano, Yoan, Kelvin
Oct 2nd	X-ray crystallography theory + software	Florence/Kelvin
Oct 9th	no class	
Oct 16th	X-ray software	Kelvin/Yoan
Oct 23rd	no class (EPFL break)	
Oct 30th	X-ray software	Kelvin/Florence
Nov 6th	cryoEM theory	Henning
Nov 13th	cryoEM software	Yoan, Anna-Sophia, Jonathan
Nov 20th	cryoEM-ET software	Yoan, Anna-Sophia, Jonathan
Nov 27th	cryoEM-ET software + visit DCI	Yoan, Anna-Sophia, Jonathan
Dec 4th	Bio-NMR theory + software + visit	Luciano, Kelvin
Dec 11th	Students presentations	All teachers
dec 18th	Students presentations	All teachers



https://morgridge.org/feature/cryo-em-primer/

Prerequisite for all techniques: a good sample



https://www.sinobiological.com/news/recombinant-protein-expression

Useful resources in protein sciences

❑ UniProt

Very informative and up to date: Function, Names & Taxonomy, Subcellular location, Pathology & Biotech, Interaction, Structure, Family & Domains, Sequence, …
AlphafoldDB included now in UniProt – models are to take with caution

❑ Protparam

Essential for biophysical parameters: Number of amino acids, Molecular weight, Theoretical PI, Amino acid composition, Atomic composition, Extinction coefficients.
https://web.expasy.org/protparam/

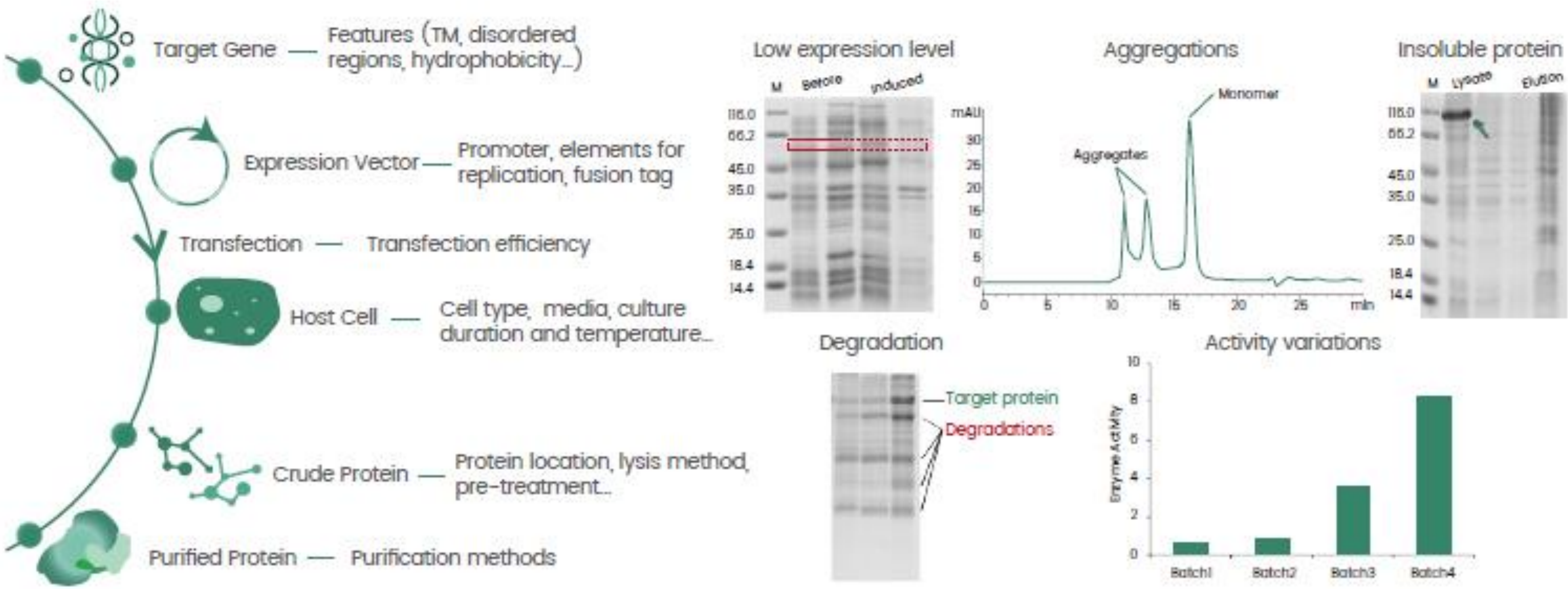
❑ Alphafold, SWISSMODEL and I-TASSER

Useful for quick 3D modelling of any proteins based on their AA sequence
https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb
https://swissmodel.expasy.org
https://zhanglab.ccmb.med.umich.edu/I-TASSER/

❑ HHPRED:

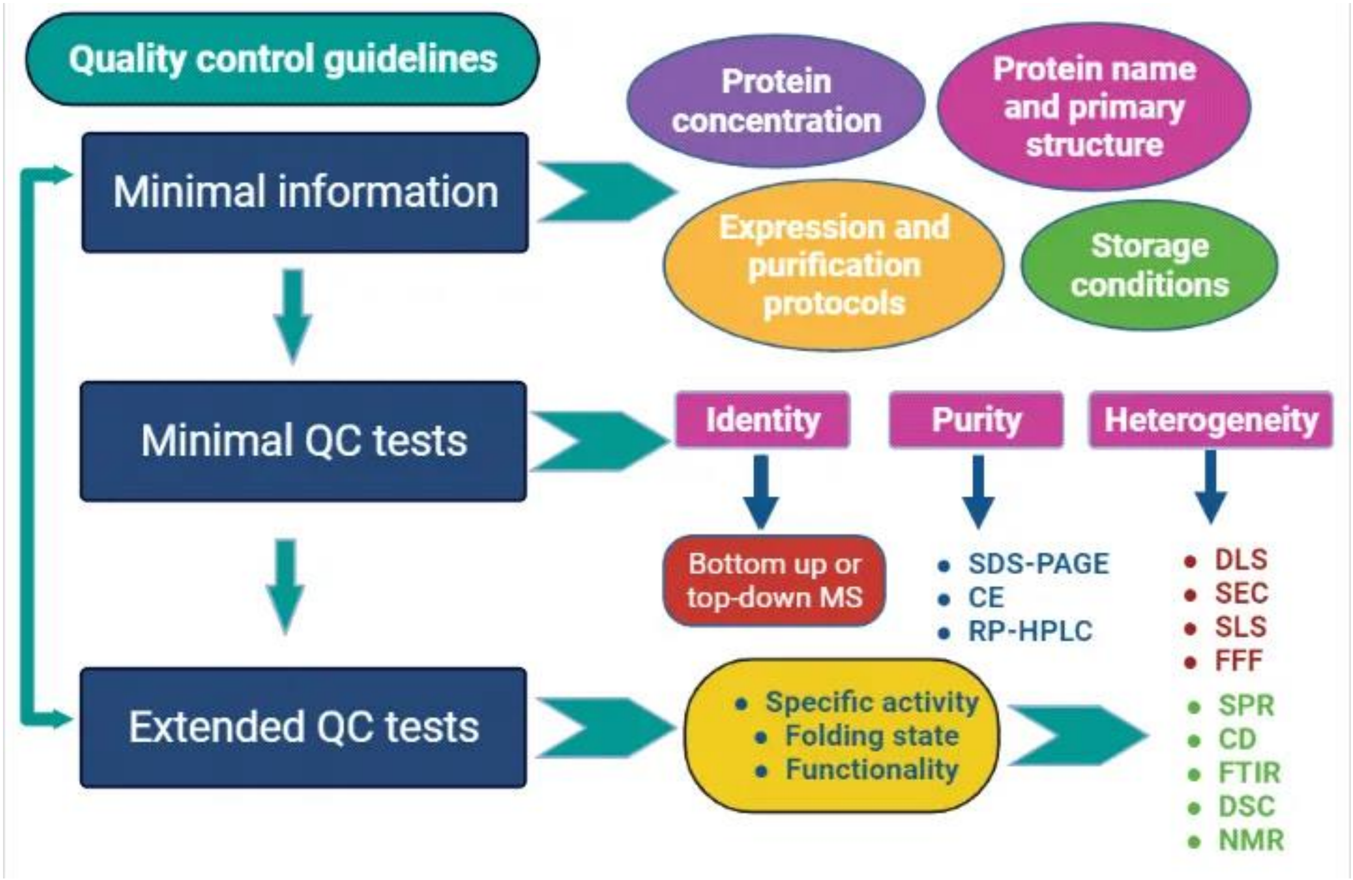
Useful for unknown function. It is a sequence database searching and structure prediction:
https://toolkit.tuebingen.mpg.de/tools/hhpred

Challenges of a good sample



https://www.sinobiological.com/news/recombinant-protein-expression

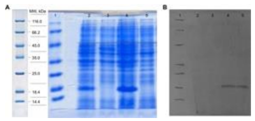
Essentiel to perform Quality Controls (QC)



https://www.kbdna.com/improved-reagent-qc-for-proteins

Tools for Biophysical characterization at PTPSP

(SDS)-PAGE + Western blotting



Oligomeric state, purity, identity

Aktas (Cytiva)

Oligomeric state

DLS Stunner (Unchained Bio)

Concentration, oligomer state, Purity Free

Mass photometer (Refeyn)

Purity, homogeneity, Oligomeric state

Chirascan V100 CD

Secondary structure, folding and stability

SPR: Biacore 8K (Cytiva)

Binding assays and kinetics

MicroCal PEAQ-ITC (Malvern)

Binding assays and thermodynamics

BLI (GatorBio)

Binding assays and kinetics

WAVE system (Creoptix)

Grating-coupled interferometry (GCI)

Uncle (Unchained Bio)

Stability, screening

Fida-1 (Fidabio)

Binding assays and kinetics

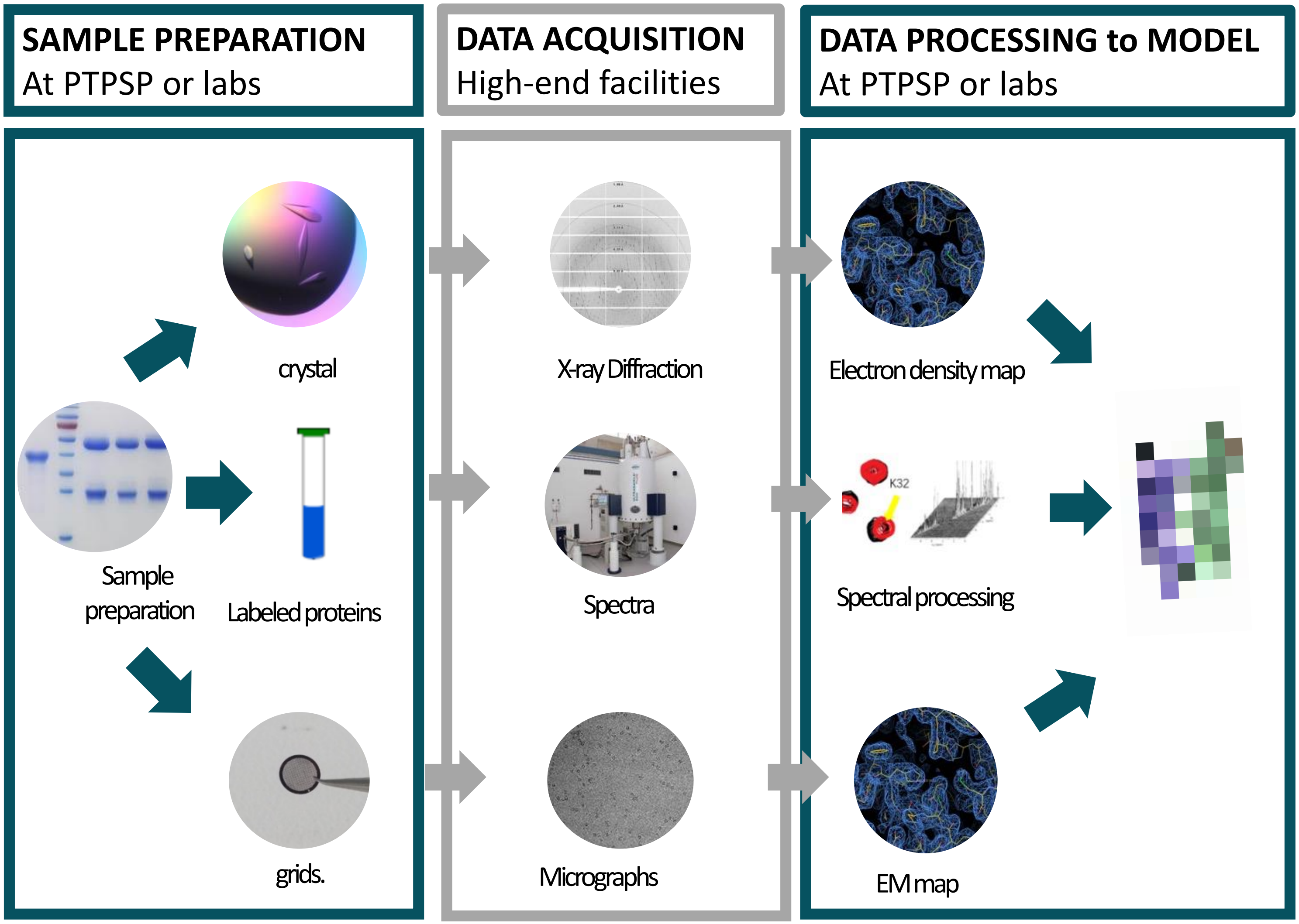
Quality control

Protein-protein/drug interactions for drug discoveries

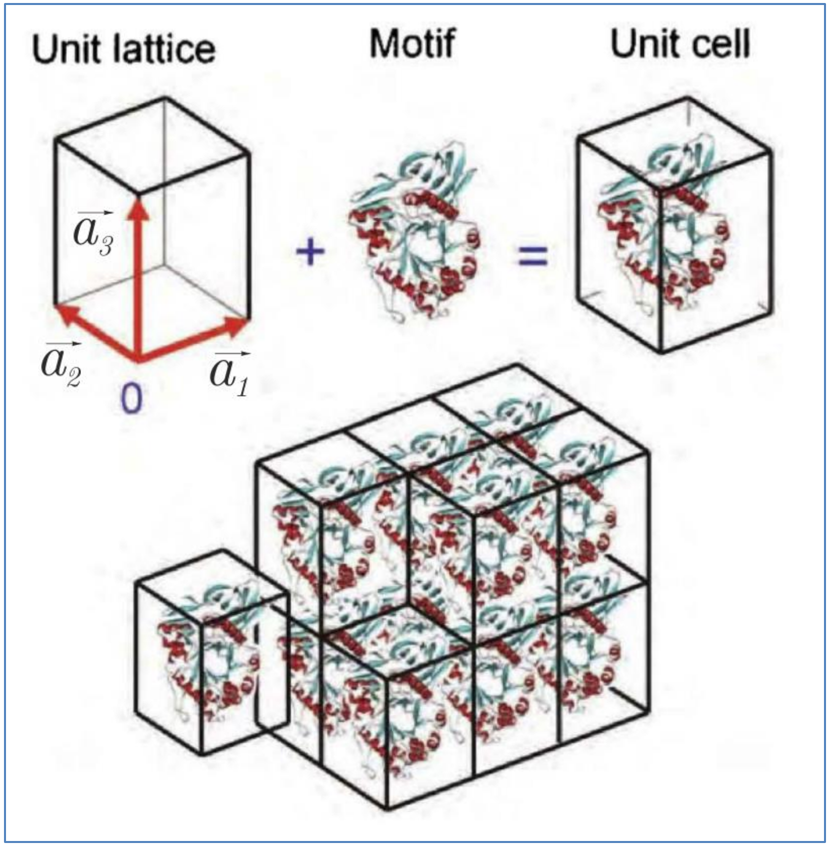
- **Dates are available 4-6 times per year upon registration on the PTPSP website**

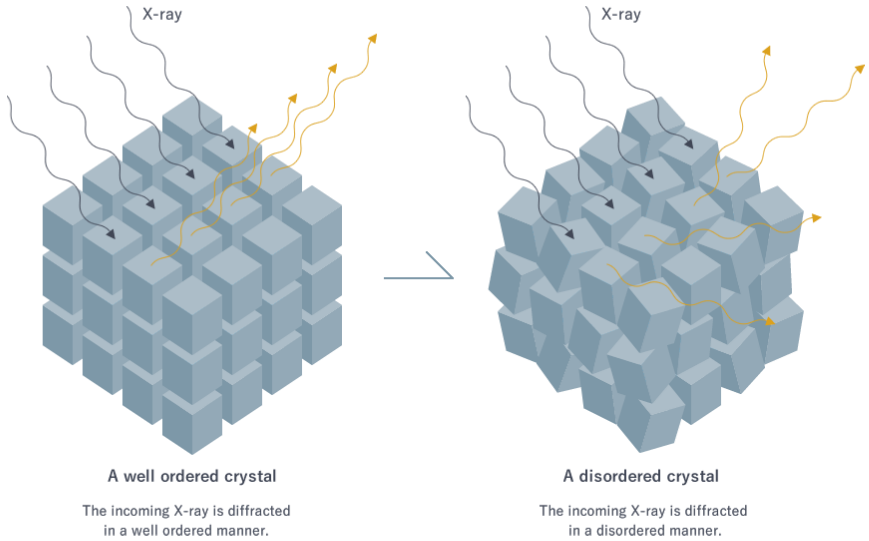
- **Private lab tours can be arranged upon request**

Experimental techniques for atomic structural studies



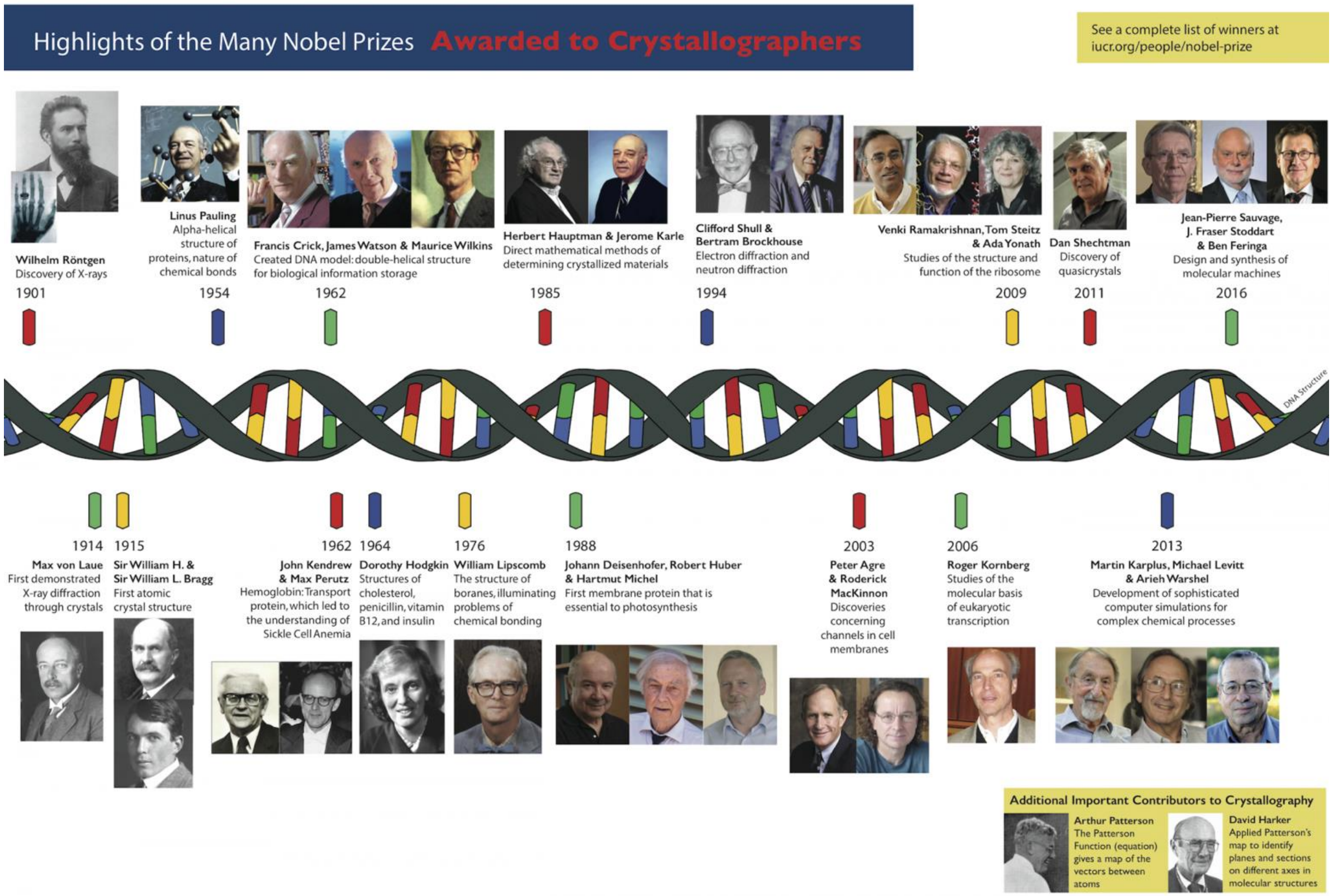
Sample quality is key to form crystals and for a good diffraction





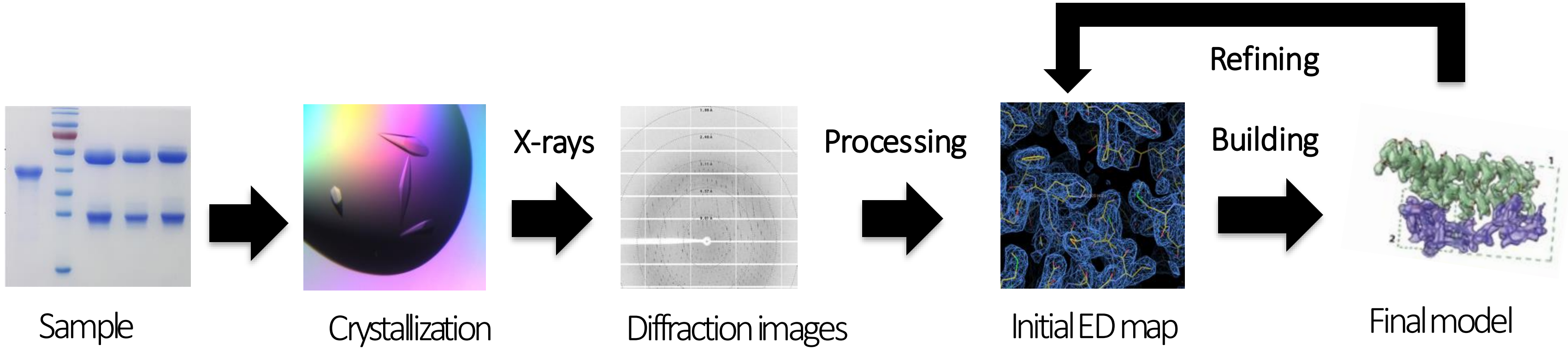
https://humans-in-space.jaxa.jp/protein/en/public/about/structural_analysis.html https://pho.rs/p/115

Success of X-ray crystallography



PDB database In Oct 2025:

198 063 models obtained by X-ray crystallography out of 242 874 models (cryoEM: 29 460 models)



Prerequisites in sample preparation:

- Essential to design suitable construct, to obtain stable and homogeneous protein or complex (e.g.; add binding partners, co-factors, small molecules Or remove flexible domains Or focus on certain domains)

- High-purity and quantity of your protein of interest or complex: Minimum two steps purification (e.g. affinity followed by size exclusion chromatography)

- Need around 120µl at 10mg/ml to screen around 400 conditions
- No phosphate buffer, to reduce formation of salt crystals

Steps for crystallization:

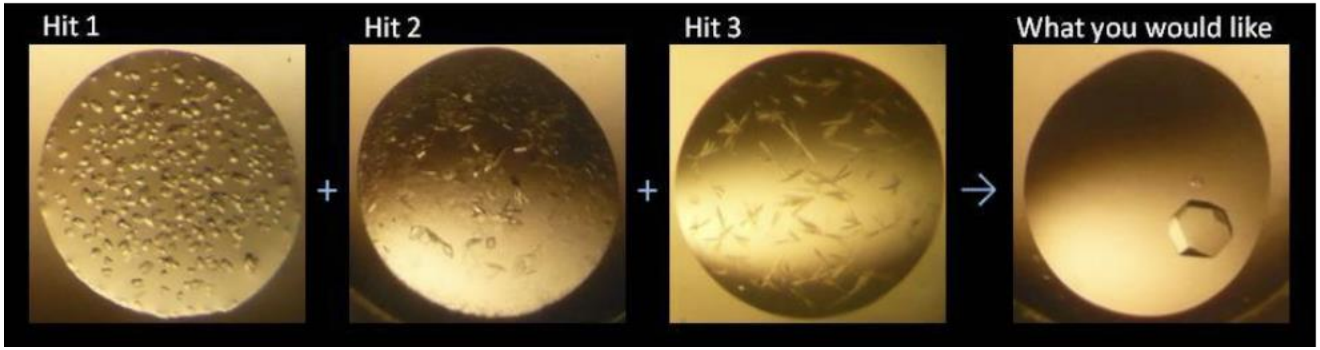
- Screening of crystallization conditions using commercial kits with robots (variation of salt, precipitant, pH, temperature, protein concentration,…)

- Incubation of plates at 18C or 4C

- Checking of crystals formation under light microscope

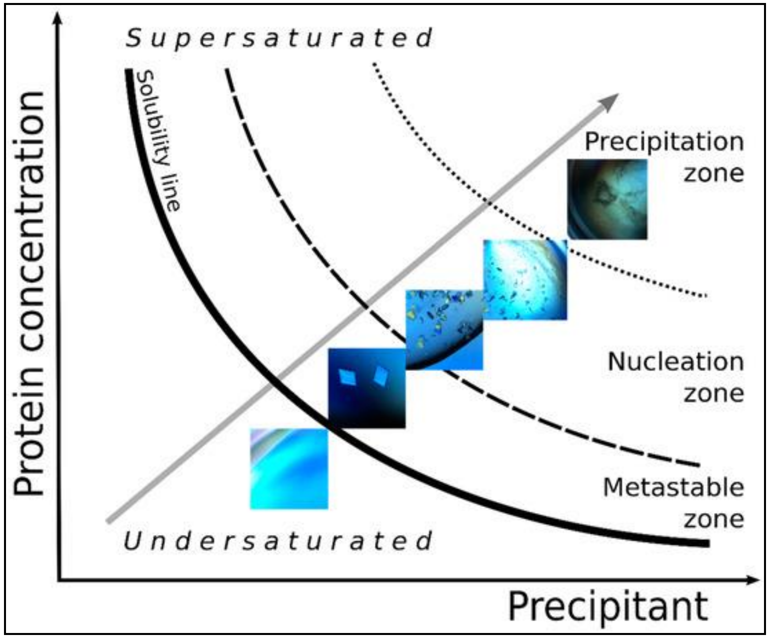
Crystals form in days to months; screening success rate is 0-10% depending on proteins

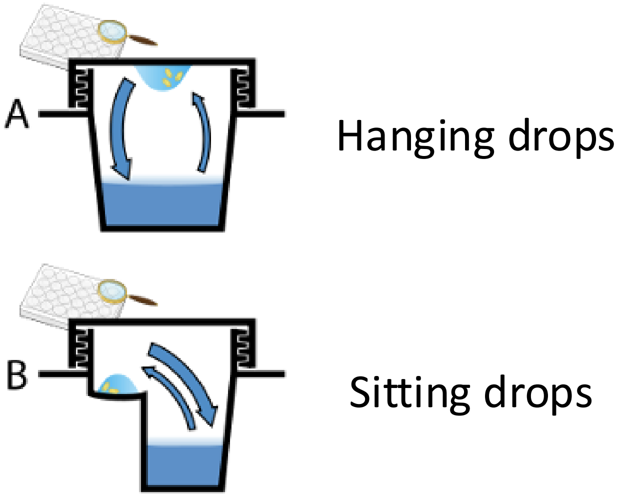
- Optimization of crystals to single crystals to obtain suitable diffraction pattern and improve resolution

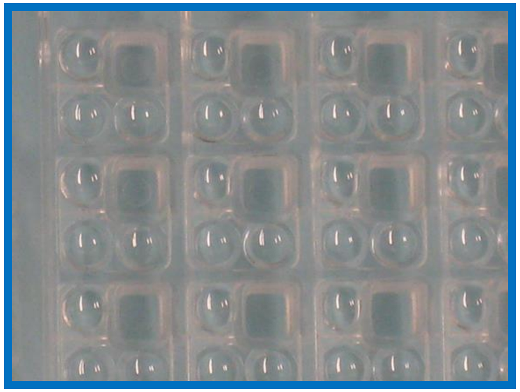


https://www.douglas.co.uk/CombinatorialNewslet.htm

Vapor diffusion techniques

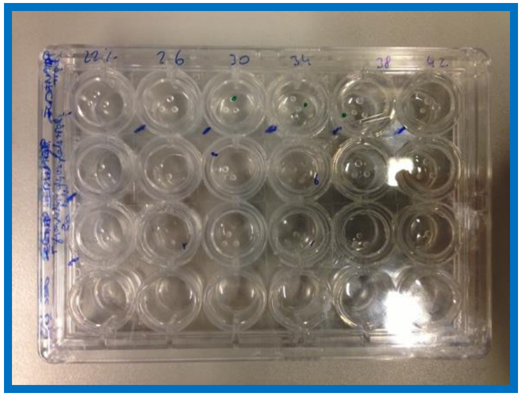




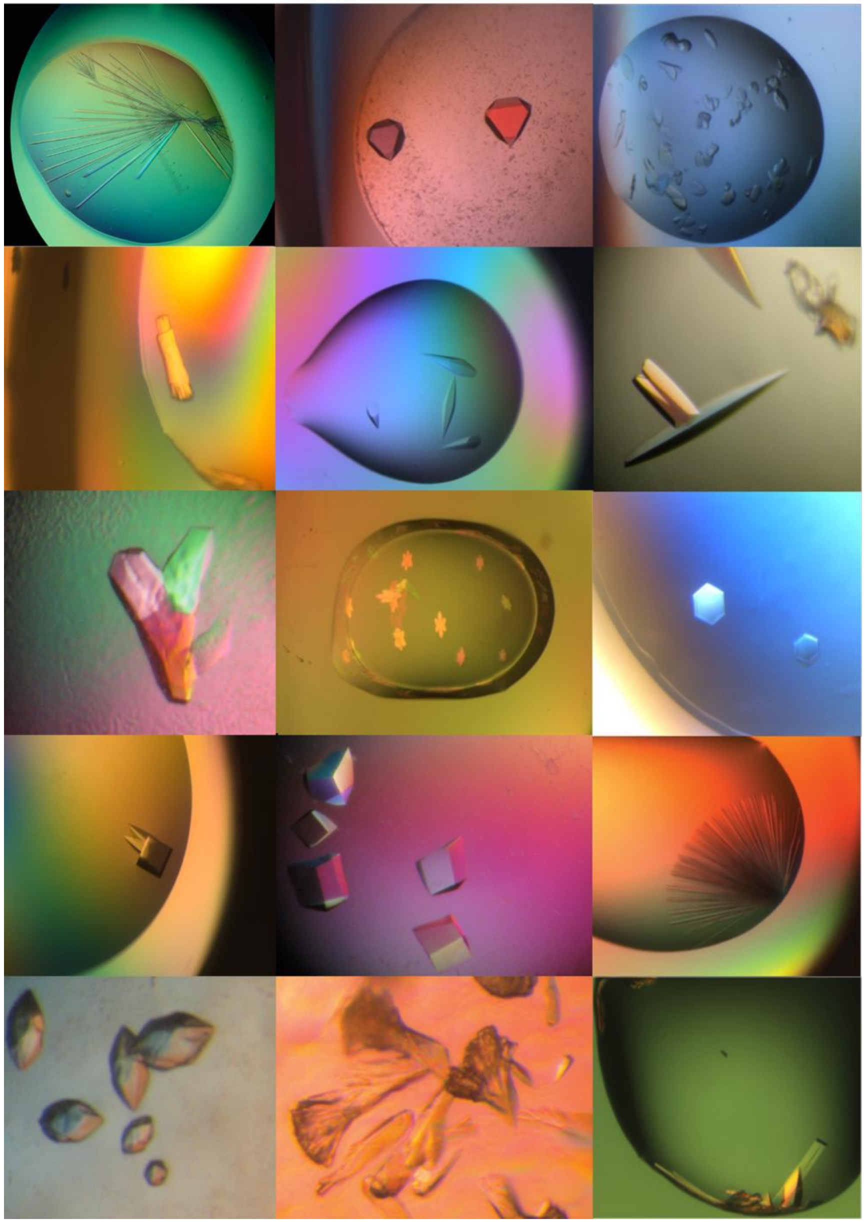


Sitting drops vapor diffusion
Mix of precipitants and protein
(total of 0.2ul to 1ul)

Mosquito (STP labtech)

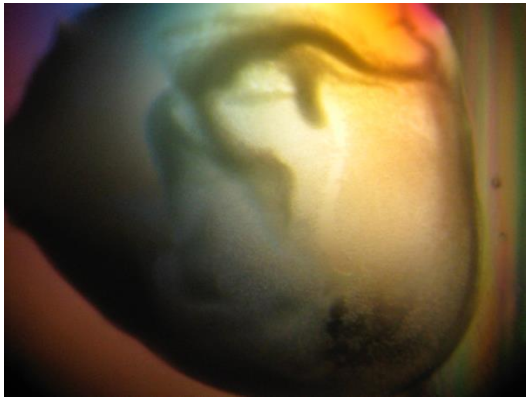


Hanging drops vapor diffusion
mix precipitants and protein
(total of 1ul to 4ul)



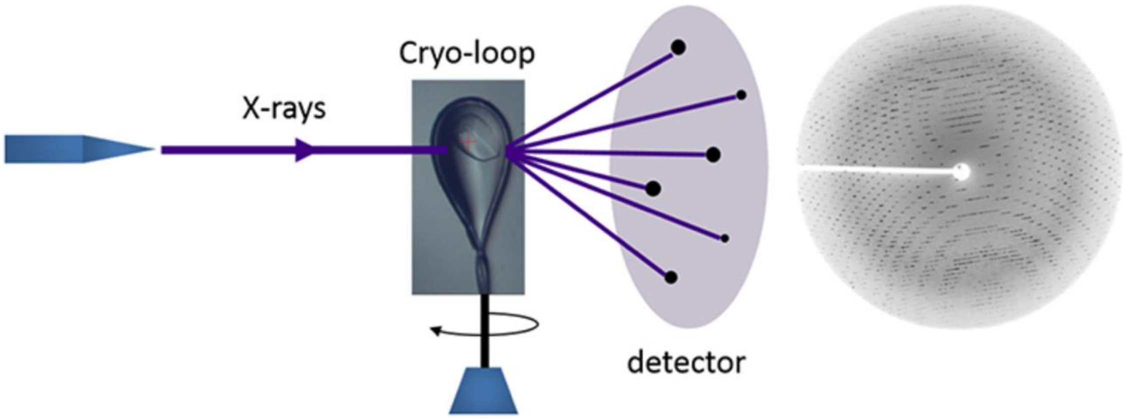
Crystals can have many shapes and sizes

Observation under microscope
(Automated instrument on market also)



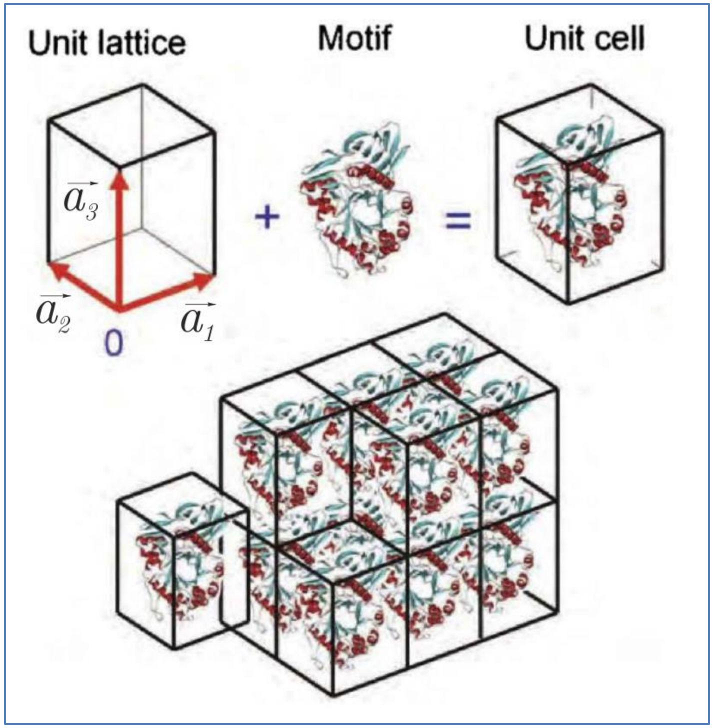
Heavy protein precipitation

Why X-rays?



Why X-rays? Their wavelength is of the order of the angström and are comparable to the distances between atoms. This wavelength allows the X-rays to interact with the crystal's atomic structure, producing diffraction patterns that provide information about the precise arrangement and spacing of the atoms.

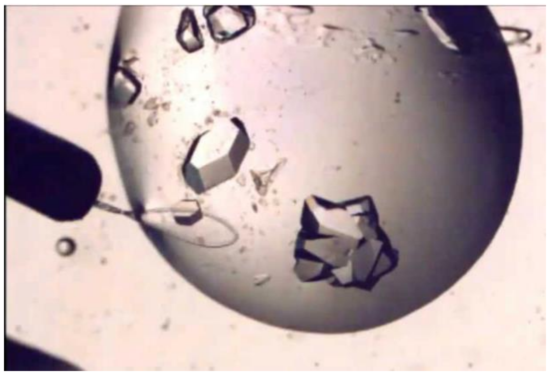
Why forming crystals?



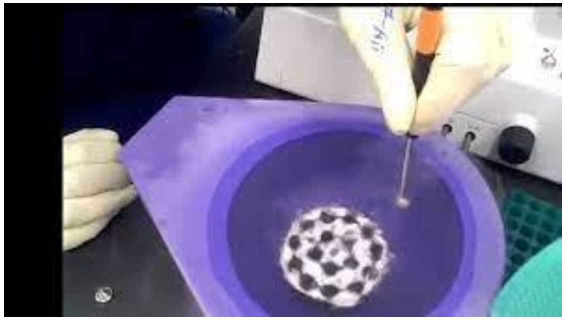
Why forming crystals? A crystal arranges huge numbers of molecules in the same orientation, so that scattered X-ray waves can add up in phase and raise the signal to a measurable level.

A crystal acts as an amplifier.

Preparation of crystals prior data collection
Cryoprotecting and Freezing



Add 25% glycerol (or other protectants) to drop before fishing

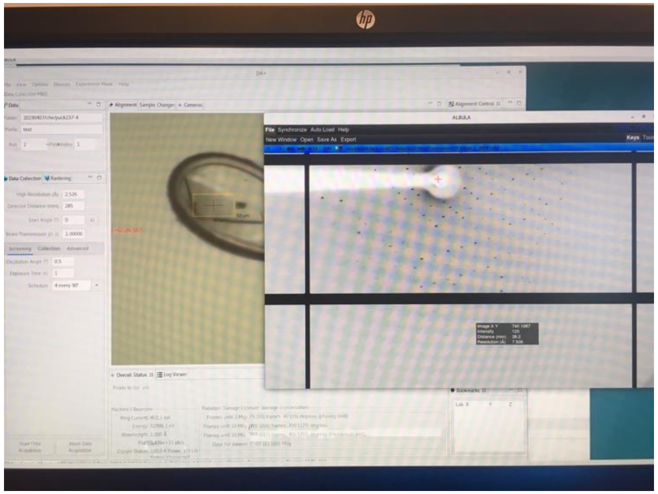


16 crystals per puck- 7 pucks per dewar- Total of 112 crystals per dewar

Data collection at synchrotrons

Data collected at Swiss Light Source (SLS-PSI, Villigen) or ESRF (Grenoble)

Beamtime is free for Academic labs: application once per year for beamtime via an academic proposal

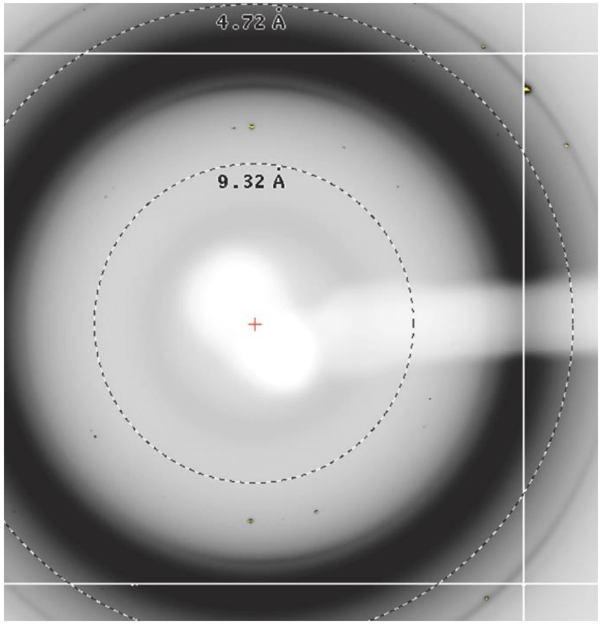


Data are collected automatically and often also pre-processed

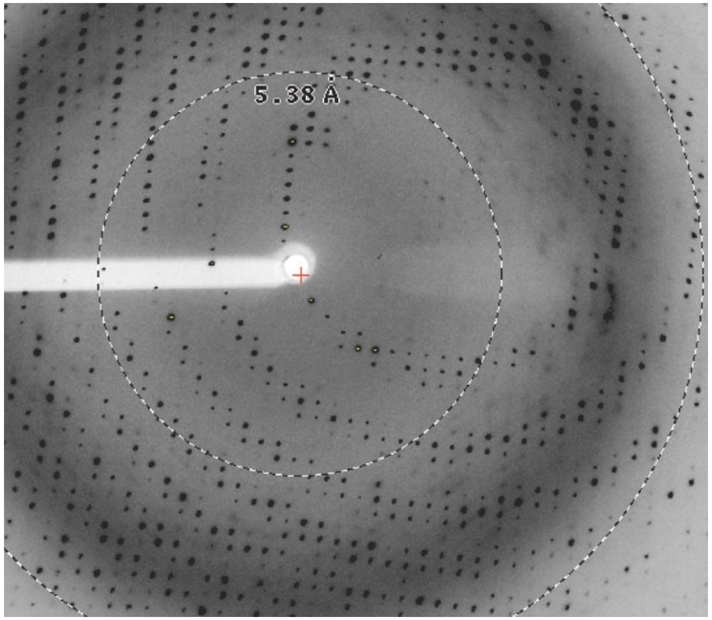
Data can also be collected remotely

Data collection at synchrotrons

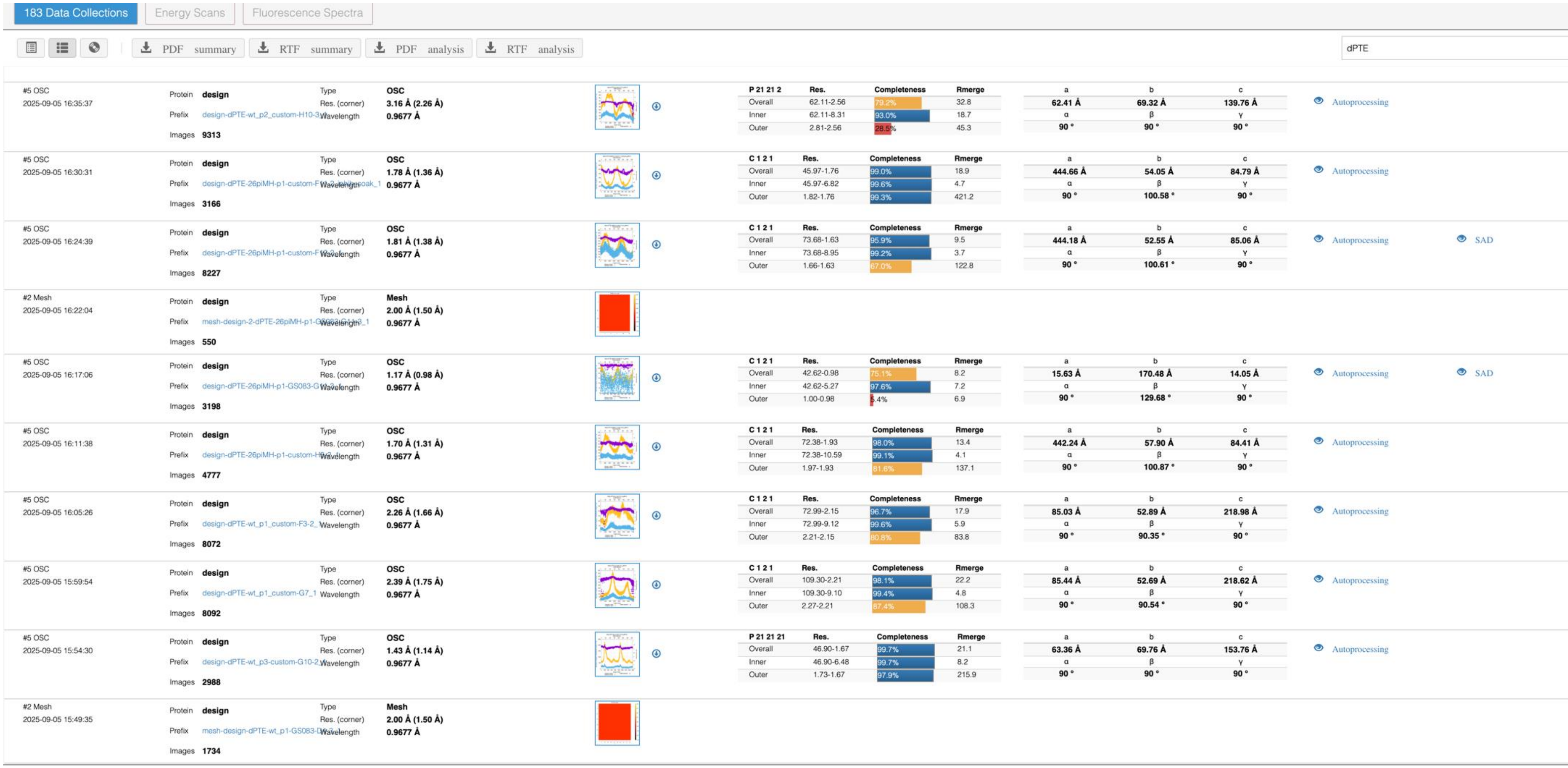
- For a full data collection, images are collected by rotating the crystal (slices of 0.1-0.2° for a total of 360° for crystal with low symmetry)
- The total number of images to collect depends on the symmetry of molecules in the crystal
- A complete data collection takes only 2-3 minutes (no strategy needed)
- High thought-put method



X-ray diffraction of crystal of salt!

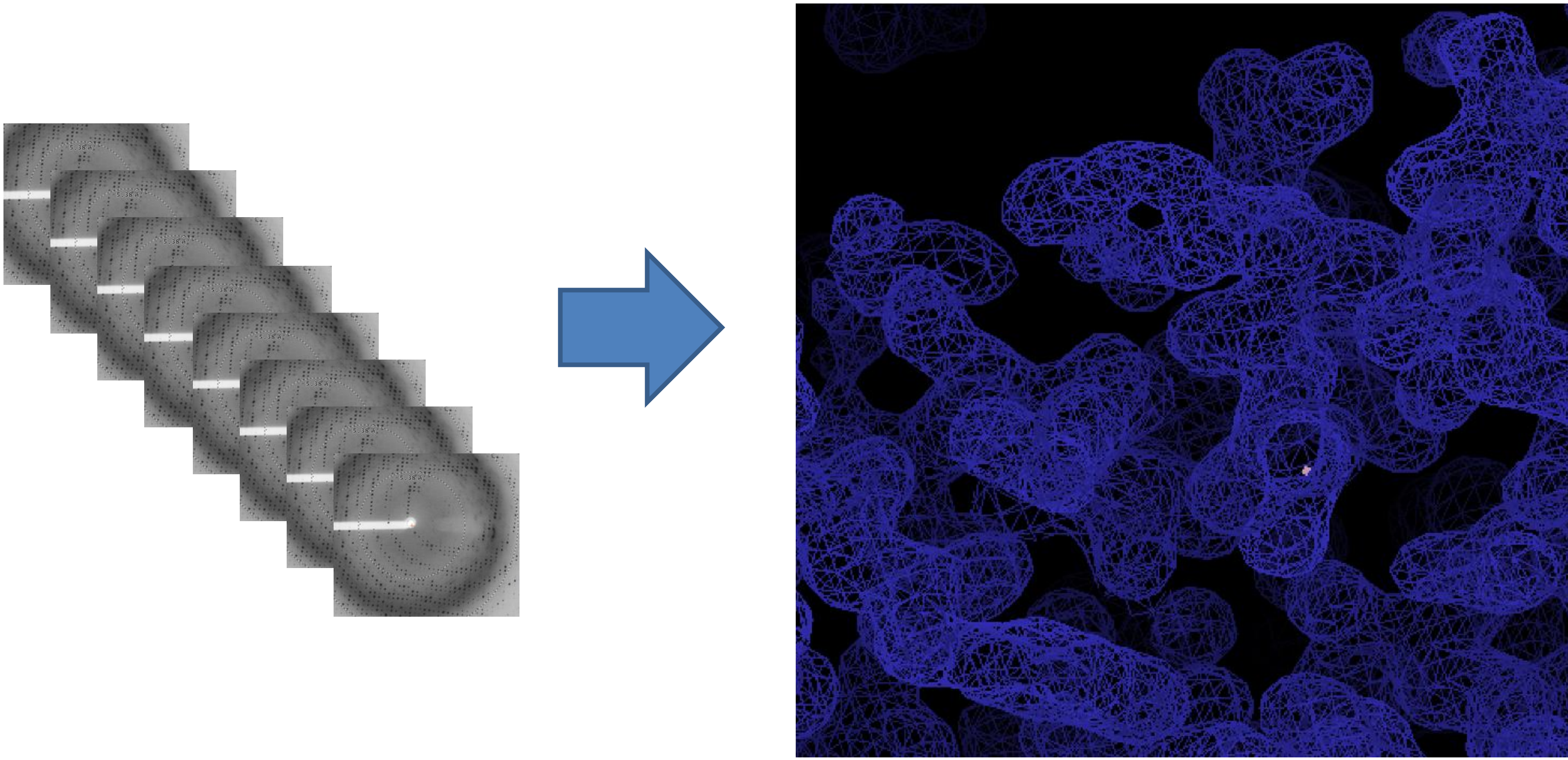


X-ray diffraction of protein crystal (one 2D image)

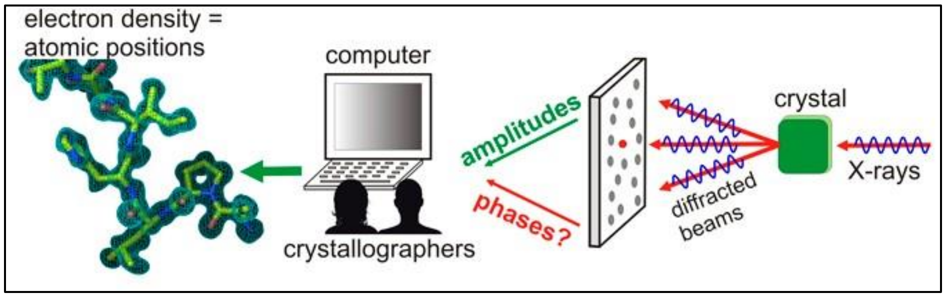


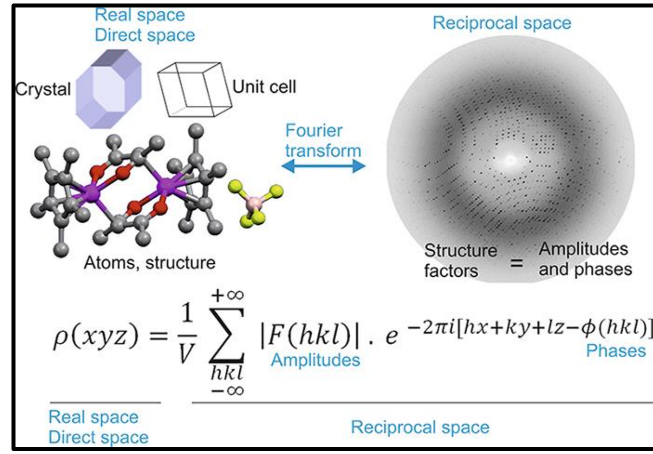
Gui at ESRF beamline:
Highly automated and easy to look and download collected data

Data Processing and phasing:
from 2D Images to first electron density map



Data Processing and phasing: from 2D Images to first electron density map

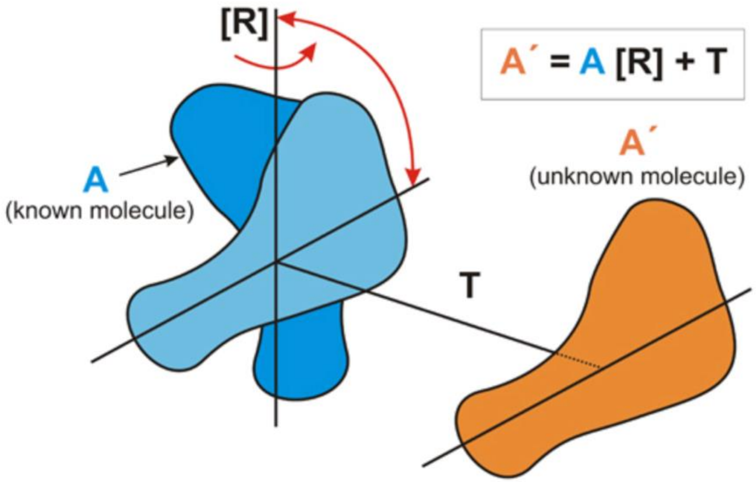




- Software, such as XDS and Mosflm, converts 2D diffraction images in an unique reflection file, to generate a first electron density map.
- Data processing steps include spots finding and autoindexing; Integration (measurement of spot intensities); scaling and merging into one reflection file.
- The phases are still missing and can be estimated by Molecular Replacement or other methods.

https://www.rigaku.com/newsletters/mabu/june2018/rigaku.journal_01.pdf https://www.creative-biostructure.com/

Methods to provide estimates of phases

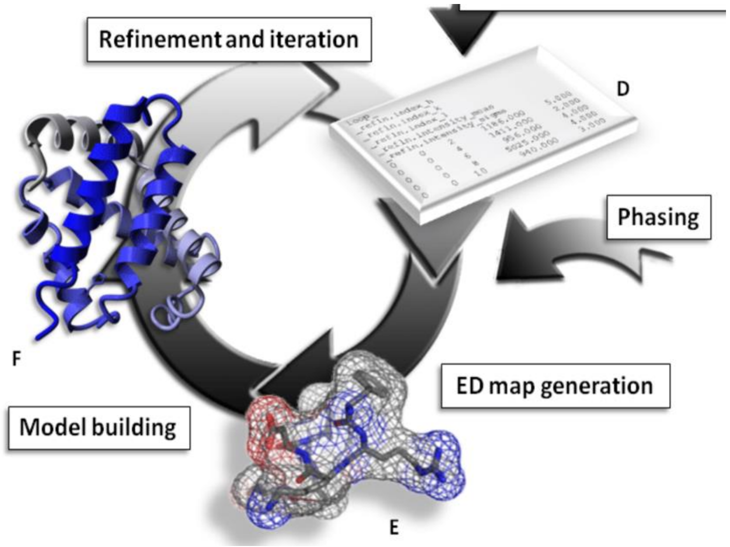


- Molecular replacement (MR): uses a known, similar model to determine an unknown one by finding the best fit through rotation and translation against experimental data. Tools like AlphaFold now assist in this process and are integrated into X-ray crystallography software.

- Multi-wavelength Anomalous Dispersion (MAD) or SAD: Heavy atoms are electron dense and give rise to measurable differences in the intensities of the spots in the diffraction pattern
- Native SAD: take use of natural sulfur atoms present in proteins

https://www.xtal.iqf.csic.es/Cristalografia/parte_07-en.html

Building and refining of the initial model

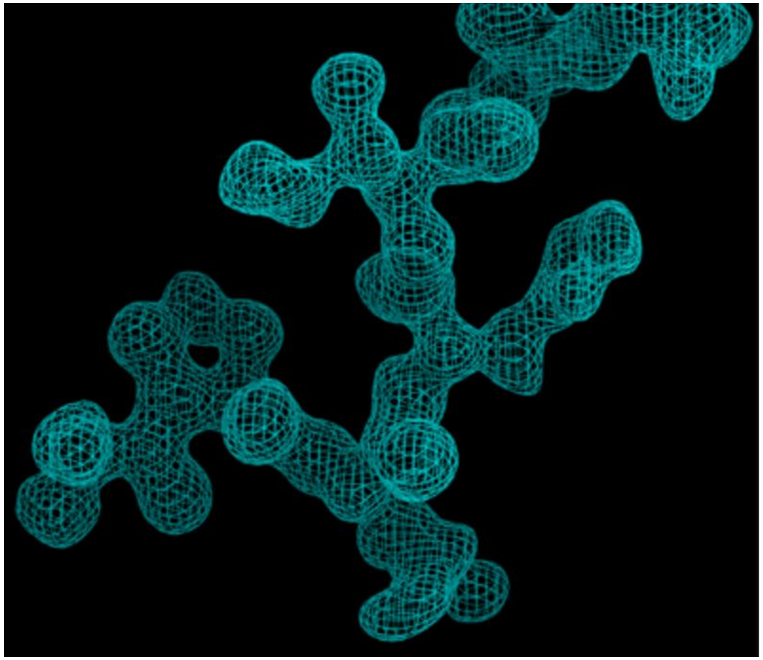


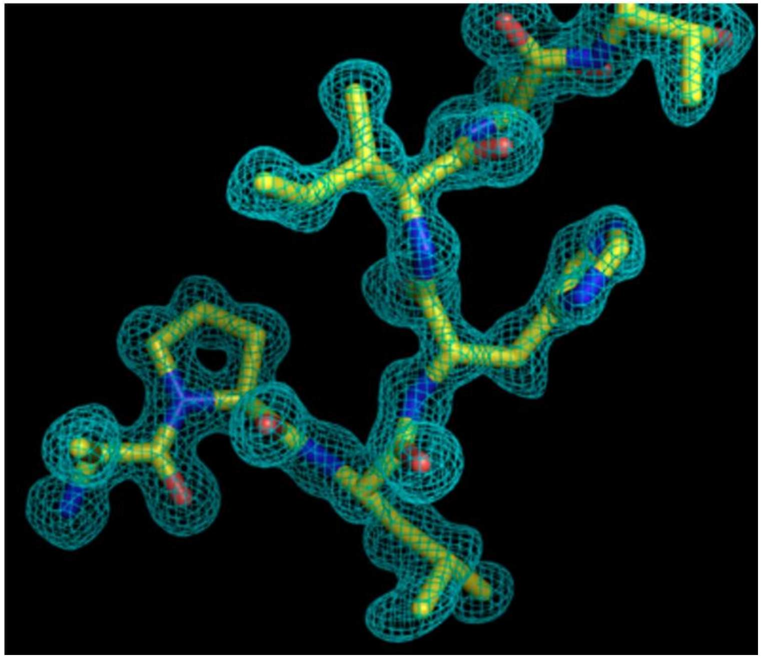
The initial model is usually suboptimal so the model has to be refined. Cycle of <u>manual</u> model building, refinement and new map generation is iterated until the model is considered correct

Phenix CCP4

Software: Phenix, CCP4i, Coot, Pymol, ChimeraX

Goal: the model should fit well the data

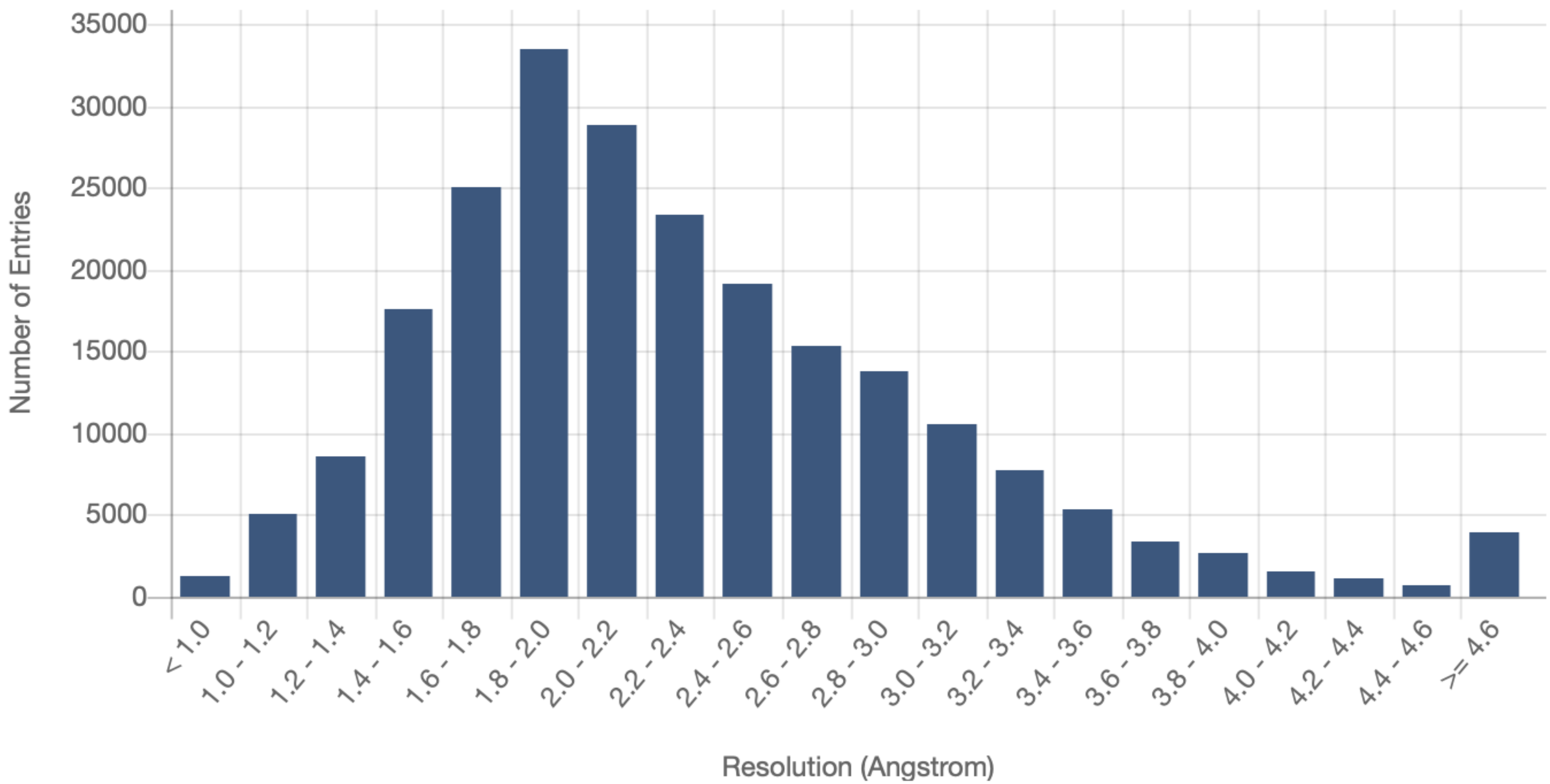




- The process can take hours to days or months depending on initial quality of the map and of the model
- Each cycle of manual model building and refinement contributes to the progressive optimization of the electron density map
- Importance of resolution in model building

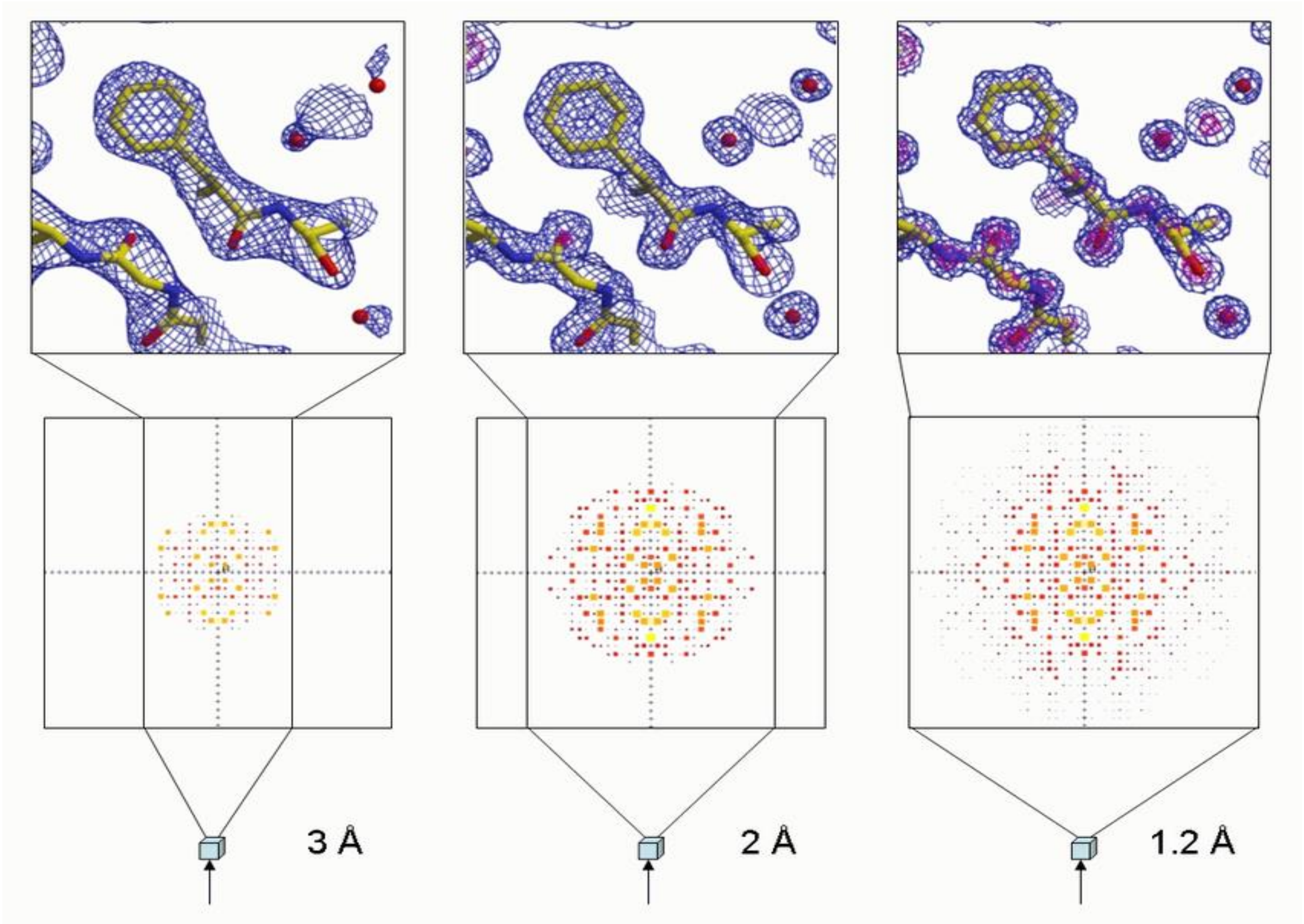
PDB Statistics:

PDB Data Distribution by Resolution



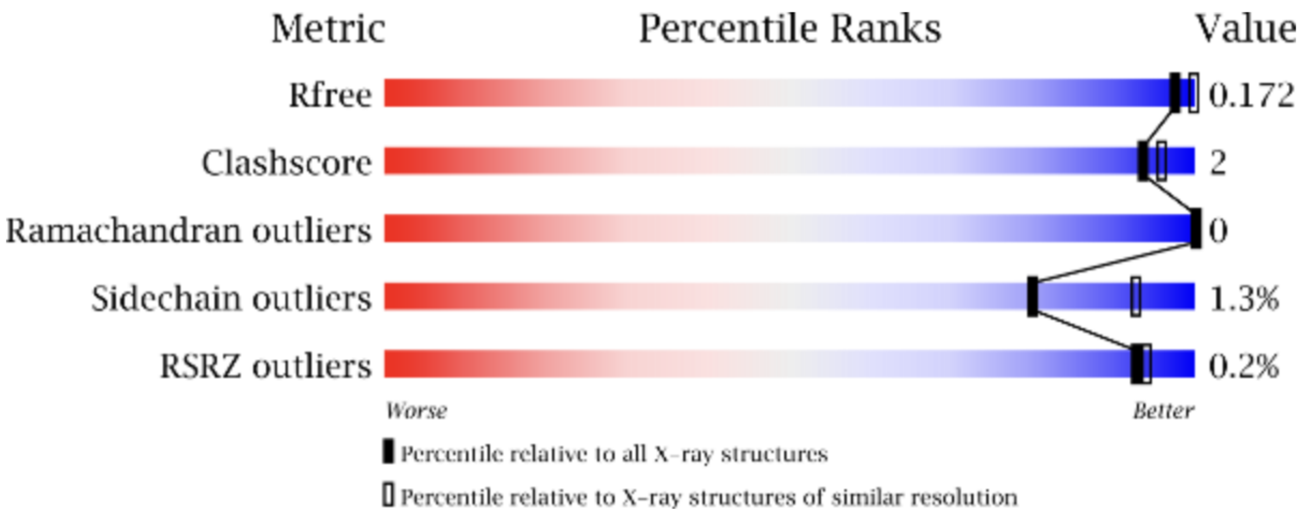
Electron density maps in function of resolution

movie



http://www.ruppweb.org/Xray/resolution.html

Validation of final model with metrics (can be very tedious)



An example slider graphic for a relatively good structure.

Rfree This is a measure of the fit of the model to a small subset of the experimental data, which was not used in model refinement (**Brünger, 1992**). The value quoted is from a recalculation by the **DCC** program (**Yang et al., 2016**). Further information can be found in the *Data refinement and statistics* section of the report, **as described below**.

Clashscore This score is derived from the number of pairs of atoms in the model that are unusually close to each other. It is calculated by **MolProbity** (**Chen et al., 2010**) and expressed as the number or such clashes per thousand atoms. Further information can be found in the *Close contacts* section of the report, **as described below**.

Ramachandran outliers A residue is considered to be a **Ramachandran plot** outlier if the combination of its ϕ and ψ torsion angles is unusual, as assessed by **MolProbity** (**Chen et al., 2010**). The Ramachandran outlier score for an entry is calculated as the percentage of Ramachandran outliers with respect to the total number of residues in the entry for which the outlier assessment is available. Further information can be found in the *Torsion angles, Protein backbone* section of the report, **as described below**.

Sidechain outliers Protein sidechains mostly adopt certain (combinations of) preferred torsion angle values (called rotamers or rotameric conformers), much like their backbone torsion angles (as assessed in the Ramachandran analysis). **MolProbity** considers the sidechain conformation of a residue to be an outlier if its set of torsion angles is not similar to any preferred combination. The sidechain outlier score is calculated as the percentage of residues with an unusual sidechain conformation with respect to the total number of residues for which the assessment is available.

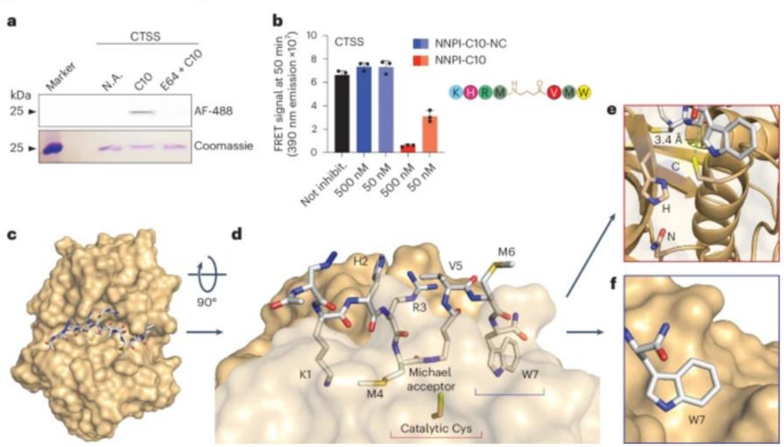
Interpretation and making sense of observations
Deposition in database and publication

> Nat Chem Biol. 2024 Sep;20(9):1188-1198. doi: 10.1038/s41589-024-01627-z.
Epub 2024 May 29.

Antibody-peptide conjugates deliver covalent inhibitors blocking oncogenic cathepsins

Aaron Petruzzella [1 2], Marine Bruand [1 2], Albert Santamaria-Martínez [1 2],
Natalya Katanayeva [1 2], Luc Reymond [3], Sarah Wehrle [4], Sandrine Georgeon [4], Damla Inel [5 6],
Floris J van Dalen [7 8], David Viertl [9 10], Kelvin Lau [11], Florence Pojer [11],
Margret Schottelius [9 12], Vincent Zoete [2 5 6], Martijn Verdoes [7 8], Caroline Arber [5 6],
Bruno E Correia [13], Elisa Oricchio [14 15]

Fig. 2: Covalent binding mode of NNPI-C10.



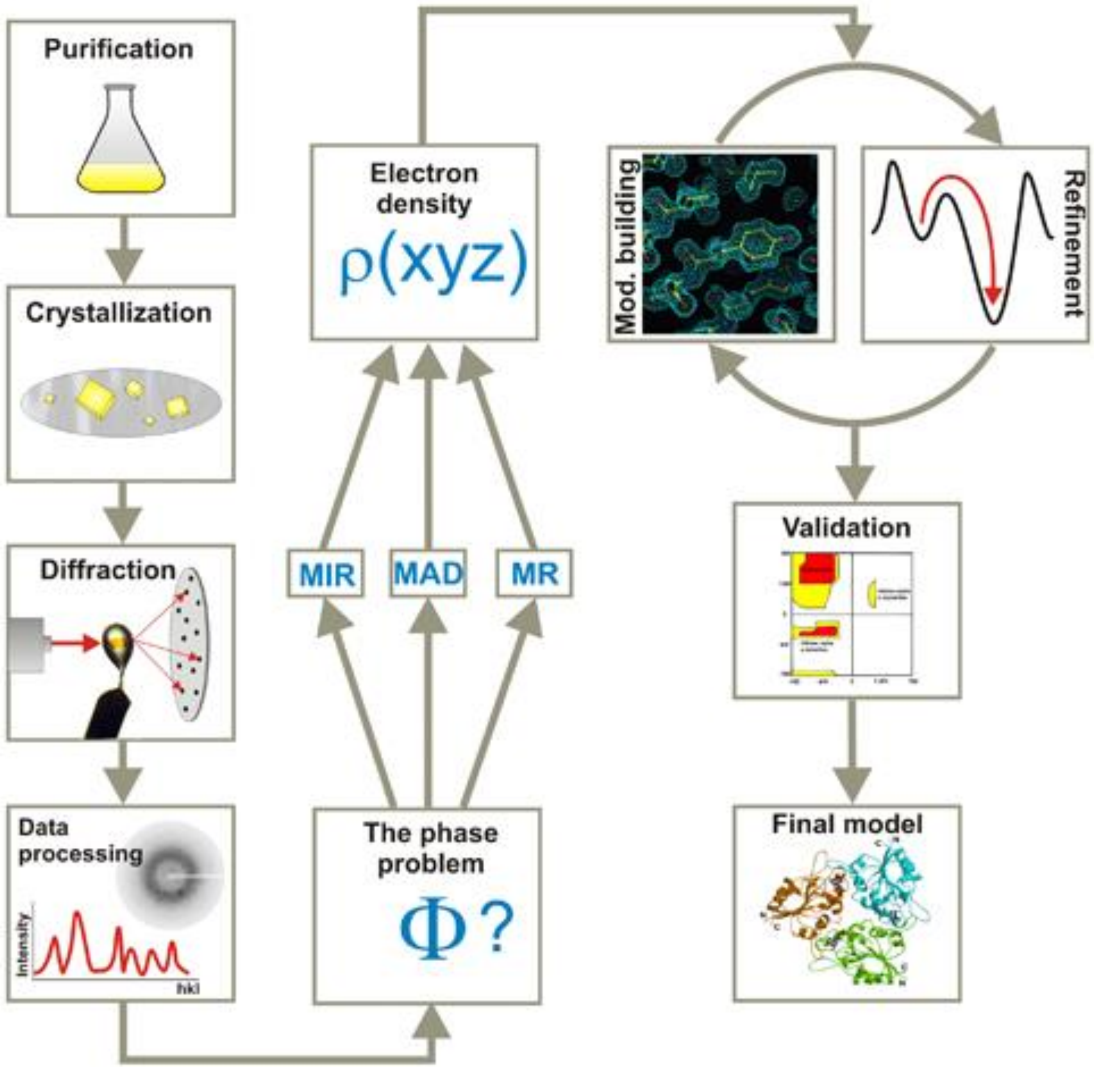
Supplementary Table 1. X-ray data collection and refinement statistics.

	Cathepsin S mutant + NNPI-C10	Cathepsin S WT + NNPI-C10
Data collection		
Space group	*P* 42 21 2	*P* 21 21 2
a, b, c (Å)	87.71, 87.71, 69.44	44.63, 129.22, 37.67
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	43.85-1.73 (1.82-1.73)	42.19-1.56 (1.64-1.56)
R_{meas}	0.12 (2.71)	0.143 (1.46)
Mean I/σI	13.1 (0.74)	8.67 (1.31)
Completeness (%)	99.9 (99.3)	99.8 (99.7)
Multiplicity	8.5 (8.1)	5.7 (5.4)
CC1/2	0.99 (0.324)	0.997 (0.665)
Refinement		
Resolution (Å)	43.85-1.73	42.19-1.56
Total reflections	460653	344424
R_{work}[#]	0.1761 (0.3881)	0.1708 (0.3365)
R_{free}[#]	0.2059 (0.4062)	0.2251 (0.3663)
Number of non-hydrogen atoms	1928	1946
macromolecules	1781	1775
ligands	82	61
solvent	65	110
protein residues	226	226
RMS deviations (bonds)[#]	0.008	0.006
RMS deviations (angles)[#]	0.90	0.81
Ramachandran favored (%)[#]	97.73	97.73
Ramachandran allowed (%)[#]	2.27	2.13
Ramachandran outliers (%)[#]	0	0
Rotatmer outliers (%)[#]	1.59	2.13
Clashscore[#]	4.10	4.00
Average B-factor[#]	38.85	27.96
macromolecules	38.08	26.79
ligands	55.38	59.07
solvent	39.11	29.56
PDB	**8PI3**	**8RND**

Data availability

The atomic models related to this study have been deposited to the PDB under accession codes 8PI3 and 8RND. Source data are provided with this paper.

Summary Pipeline of X-ray structure determination

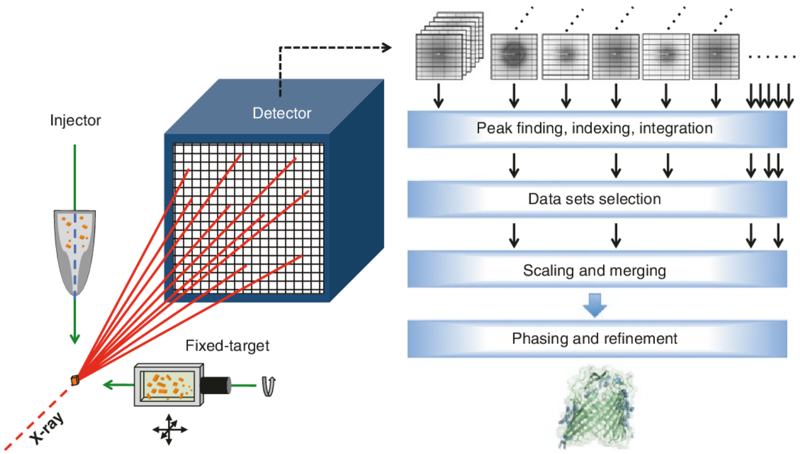


Once an initial model is placed in electron density, some additional steps (**manual building** of the detailed model, **refinement** and **validation**) are carried out to obtain the final model

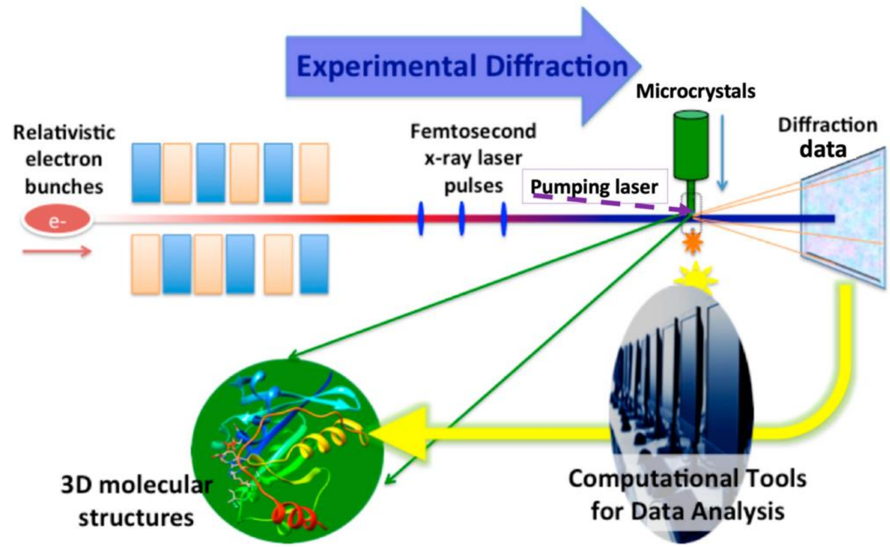
https://www.xtal.iqfr.csic.es/Cristalografia/parte_07-en.html

Other techniques on the X-ray side

SSX: Serial Synchrotron X-ray crystallography



XFEL: X-Ray Free-Electron Laser



Kay Diederichs, Meitian Wang (2017) Serial Synchrotron X-Ray Crystallography (SSX)

Lui et al (2019) The XFEL Protein Crystallography: Developments and Perspectives

Time to do some Hands-on work

- **Data collection and processing**

Look at diffraction pattern of different crystals; and see how to go from 2D images to the data file (called reflection file) with XDS software

- **Solving the structure by Molecular Replacement**

Using Phaser in Phenix suite software

- **Manual model building and refinement**

Using coot and phenix-refine

- **Validation of model and deposition**

In coot and PDB database