Biomolecular Structure and Mechanics
7th of May 2025

Structural Biology
X-ray Crystallography

Dr. María José Marcaida López

LBM

LAB AI 2351

Content of lectures

- Why x-rays and why crystals?
- Macromolecular crystallization
- Crystal packing
- The diffraction experiment
- The phase problem
- Molecular replacement
- Refinement and model building

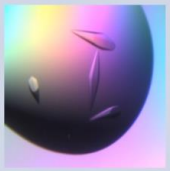
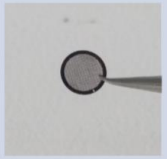
OBJECTIVE
Cover an overview on x-ray crystallography to help you interpret published articles.

References

Rupp, B. (2009) Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology, 1st edition (Garland Science)

Advantages of x-ray crystallography

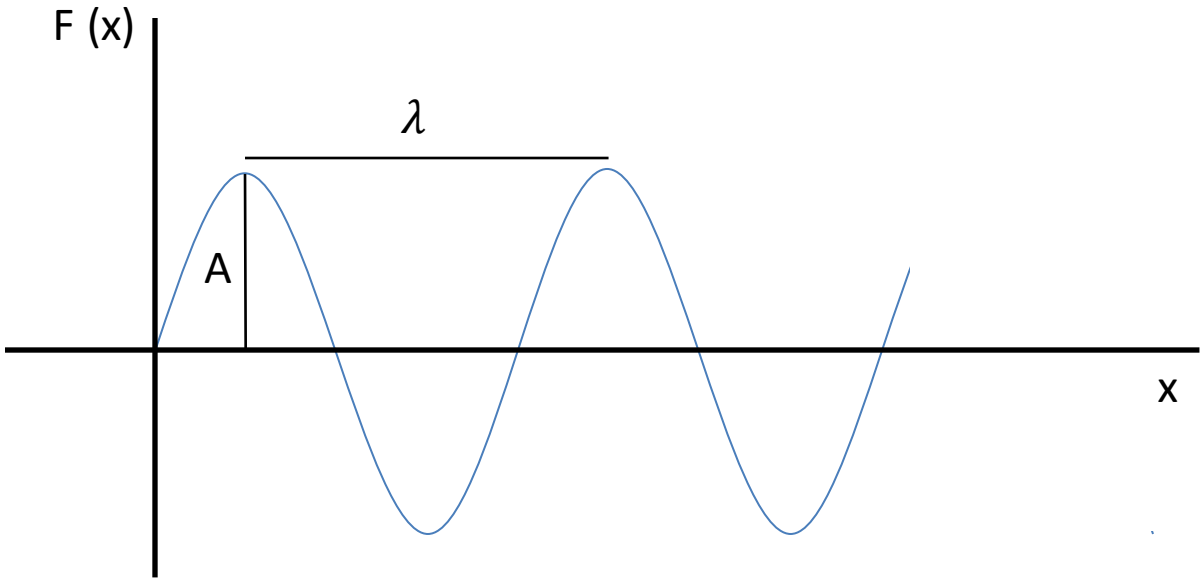
- X-ray crystallography has the power to resolve the 3D structure of all kinds of macromolecules.
- It is the only method that can routinely reach atomic resolution.
- But you need to crystallize the macromolecule.

Techniques	PROS	CONS
X-ray crystallography 	✓ Provide very detailed atomic information ✓ Easy to perform ✓ Not expensive ✓ Software free and user friendly	✓ No size limits ✓ Need to form crystals ✓ High protein quantity ✓ Difficult for membrane proteins
NMR 	✓ Small flexible proteins ✓ In solution ✓ Info on dynamics	✓ Not for big complex. (<40kDa) ✓ Low through-put ✓ High expertise ✓ High protein quantity, labeled
Single-particle EM 	✓ Big complex, membrane proteins ✓ Not much protein needed ✓ Can achieve high-resolution	✓ Not easy for small protein yet (>70kDa) ✓ High expertise ✓ Low Through-put ✓ High-end equipment

Why x-rays?

X-rays are electro-magnetic radiation, consider only the oscillating electric field

X-rays as Waves



$F(x)=A \sin(2\pi\nu x + \alpha)$

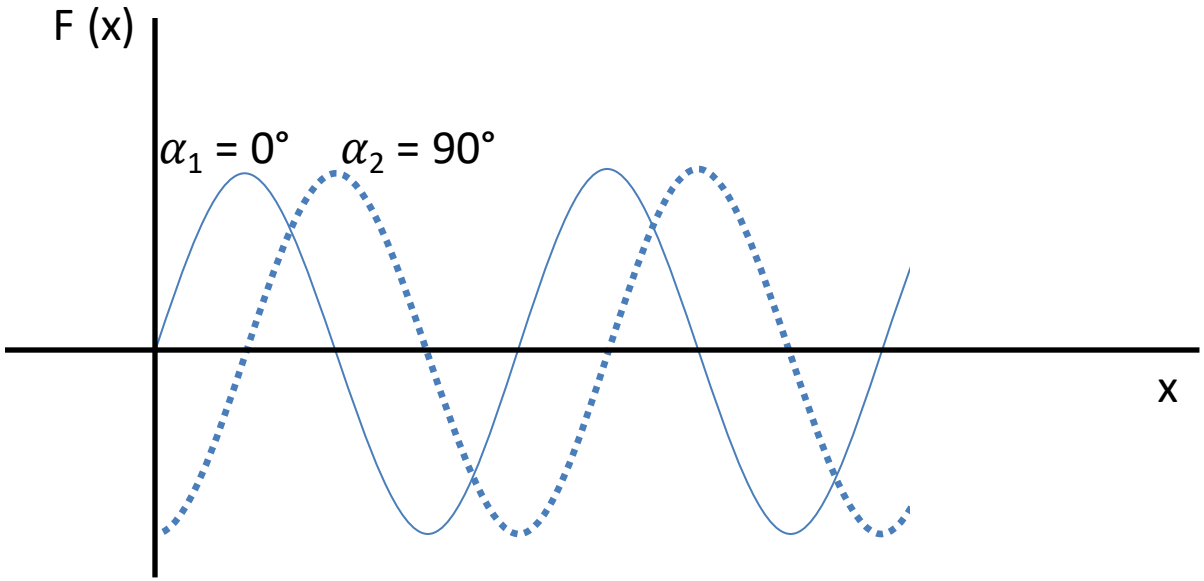
Where:
A: amplitude
$\nu = 1/\lambda$: frequency
λ = wavelength
α = phase

The energy of the radiation is related to its λ like this:

$E=hc/\lambda$

Where h is Plank's constant and c is the speed of light

X-rays as Waves



$F(x)=A \sin(2\pi\nu x + \alpha)$

Where:

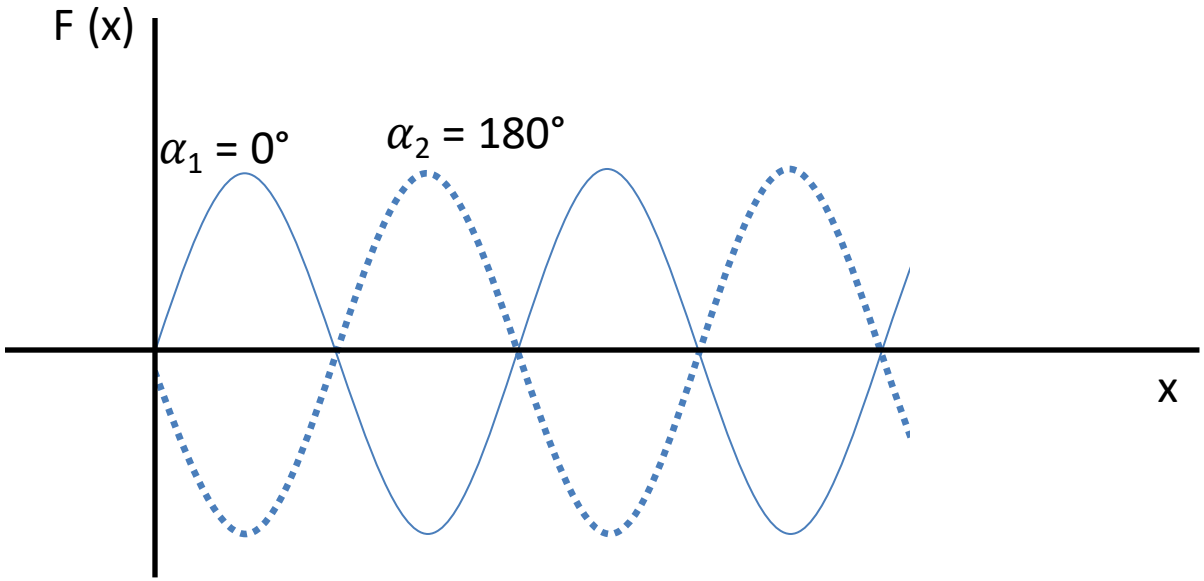
A: amplitude

$\nu = 1/\lambda$: frequency

λ = wavelength

α = phase

X-rays as Waves



$F(x)=A \sin(2\pi\nu x + \alpha)$

Where:

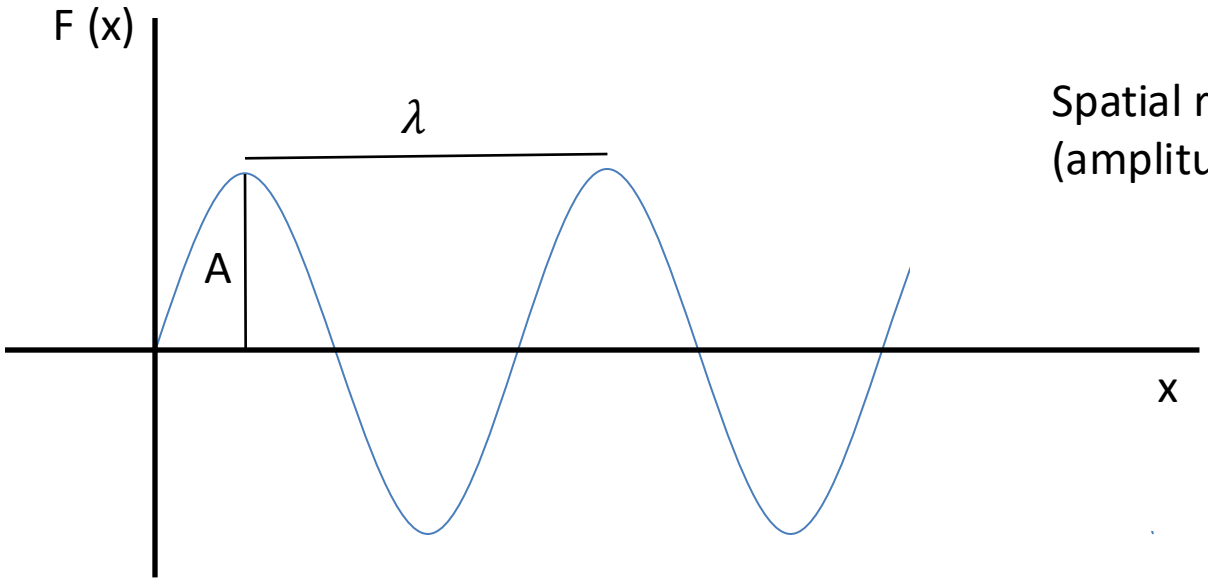
A: amplitude

$\nu = 1/\lambda$: frequency

λ = wavelength

α = phase

Waves



Spatial representation of a wave (amplitude vs x)

$F(x)=A \sin(2\pi\nu x + \alpha)$

Where:

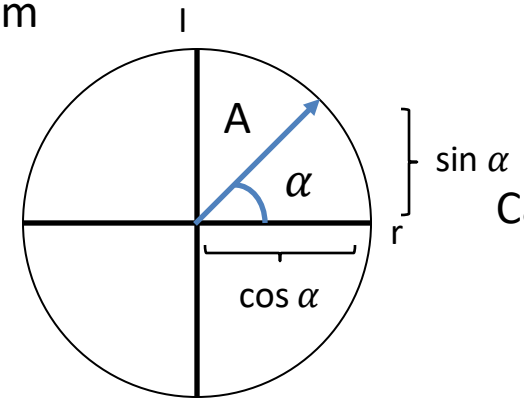
A: amplitude

$\nu = 1/\lambda$: frequency

λ = wavelength

α = phase

Argand Diagram



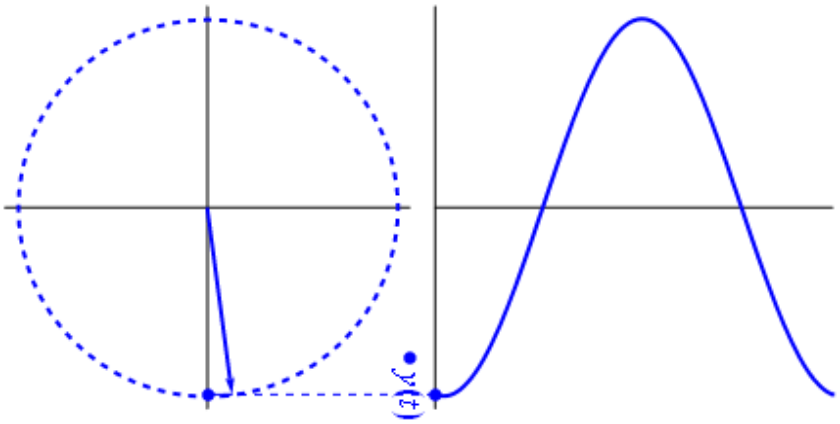
Can be also represented as a complex number where:

real component is $A\cos\alpha$

imaginary component is $A \sin\alpha$

$Ae^{i\alpha} = A(\cos\alpha + i\sin\alpha)$

Waves



$F(x)=A \sin(2\pi\nu x + \alpha)$

Where:

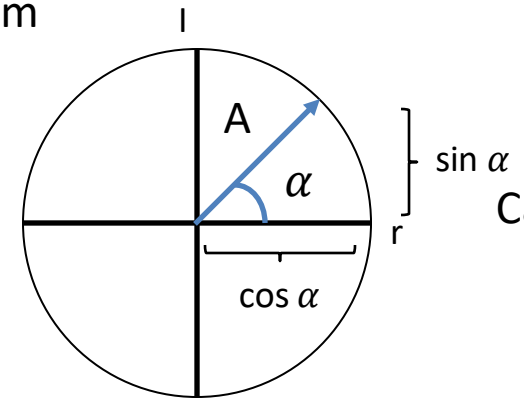
A: amplitude

$\nu = 1/\lambda$: frequency

λ = wavelength

α = phase

Argand Diagram



Can be also represented as a complex number where:

real component is Acos α

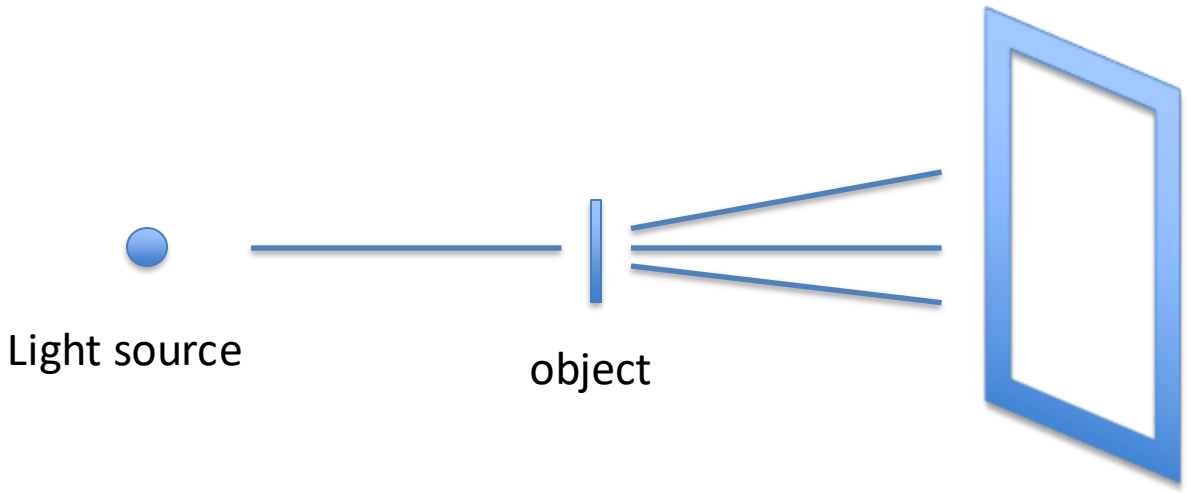
imaginary component is A sin α

$e^{i\alpha} = A(\cos\alpha + i\sin\alpha)$

Why x-rays?

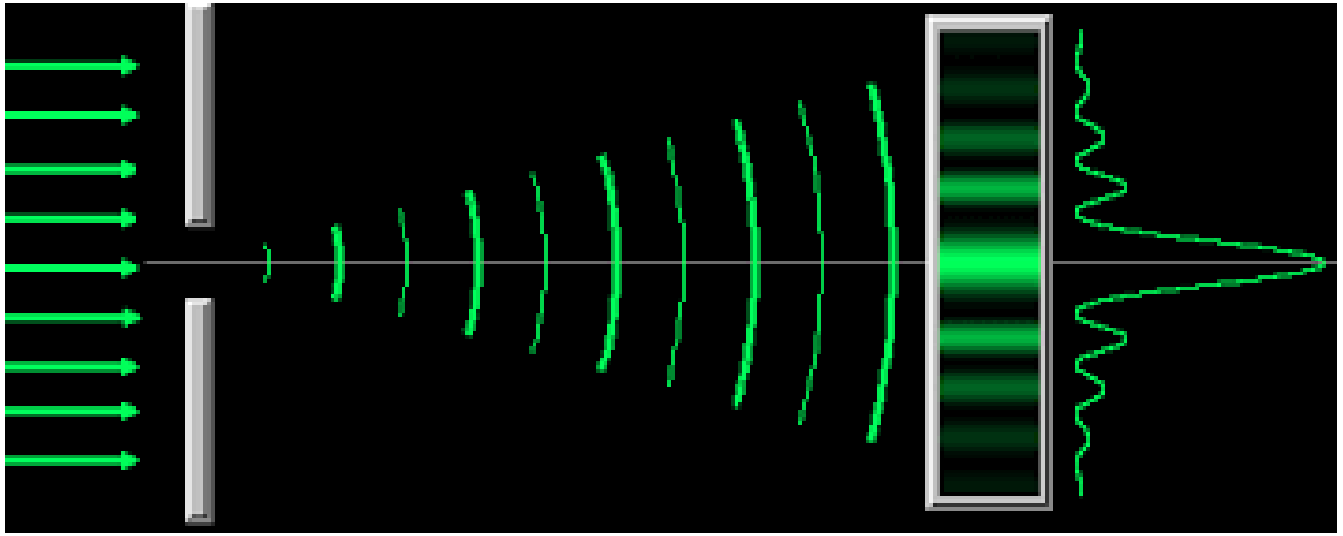
Let’s consider what happens when a light wave encounters an object of the same size as its wavelength…

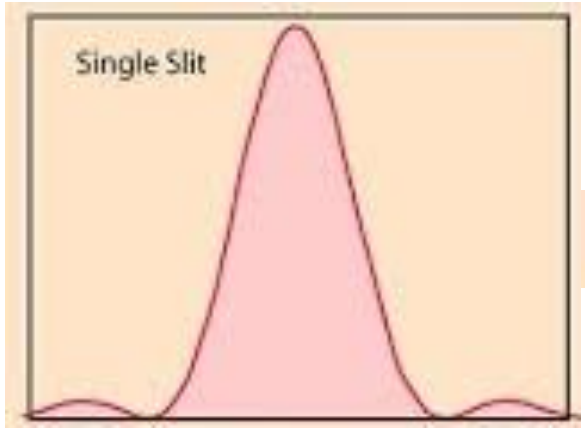
Single slit diffraction pattern of light



T. Young 19th century

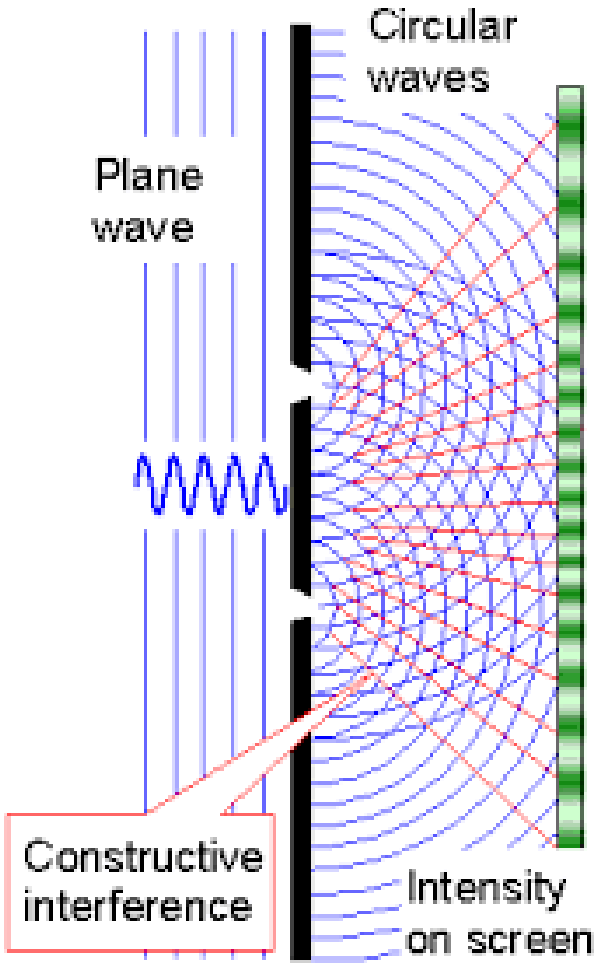
Single slit diffraction pattern of light

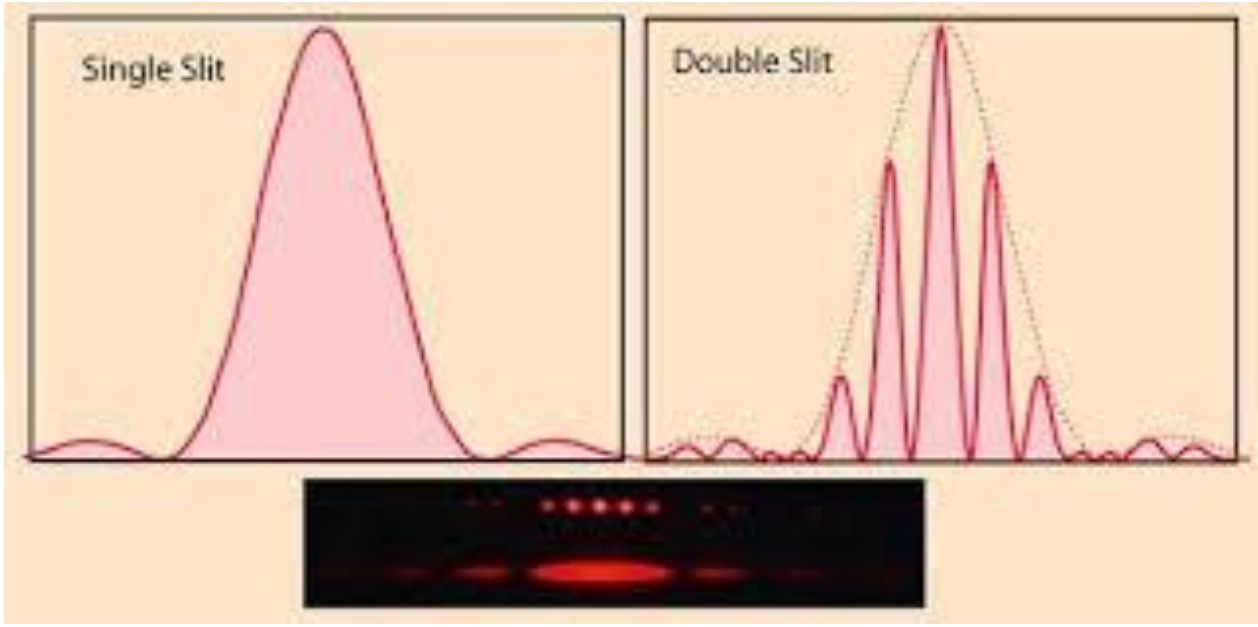




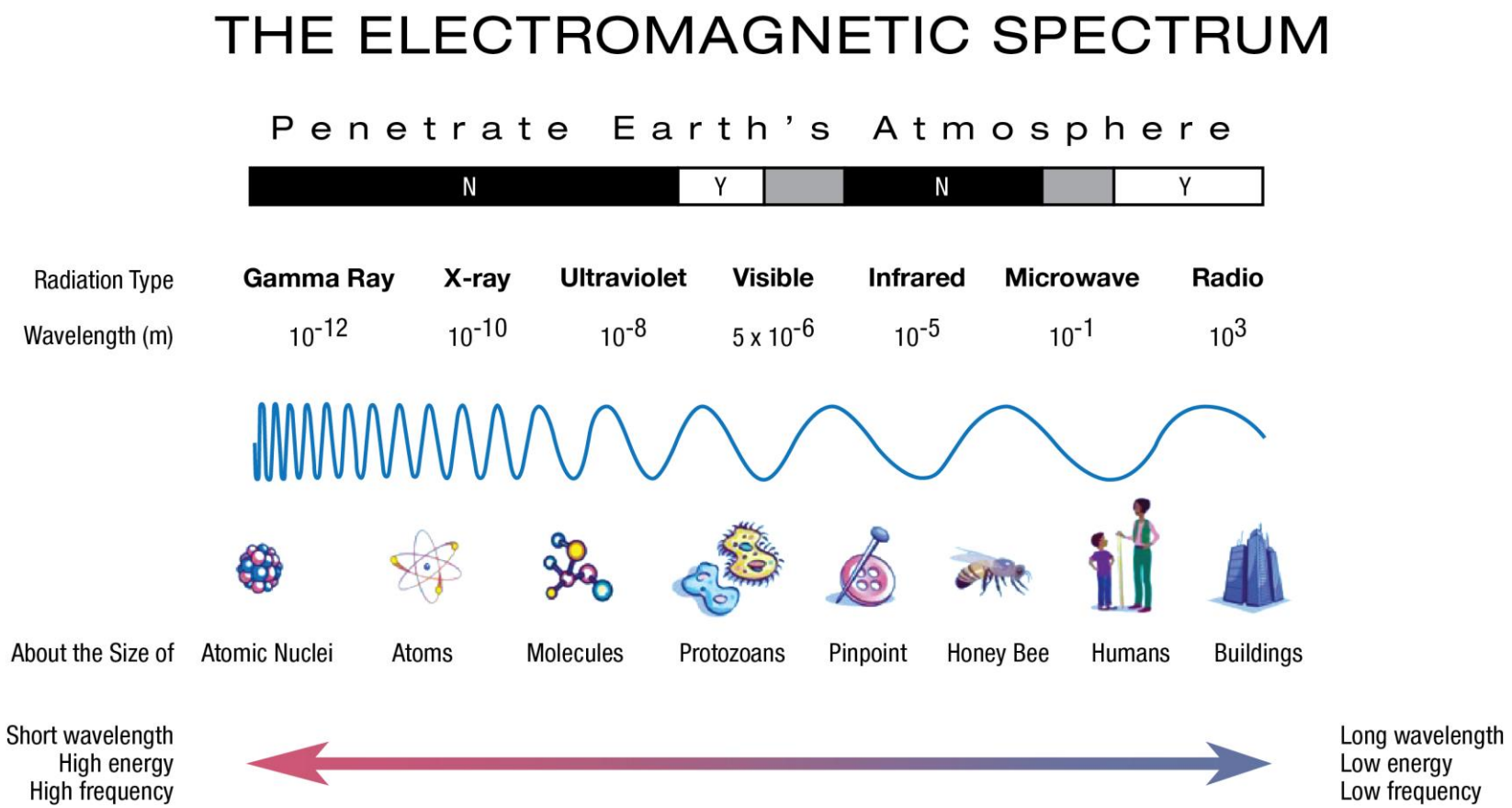
T. Young 19th century

Double slit diffraction pattern of light





Why x-rays?



Wavelenth of the radiation has to be
of the same size as the object we want to observe

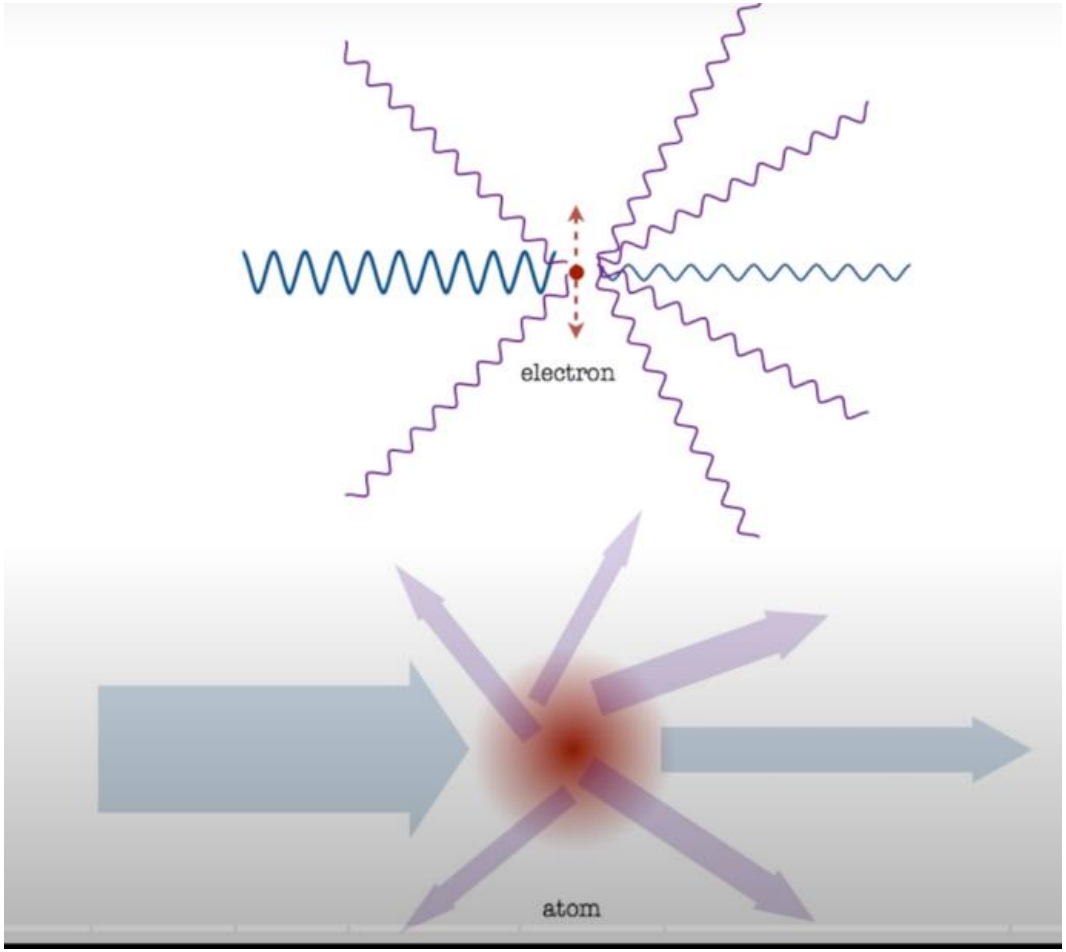
$1\ Å = 1 \times 10^{-10}\ m$

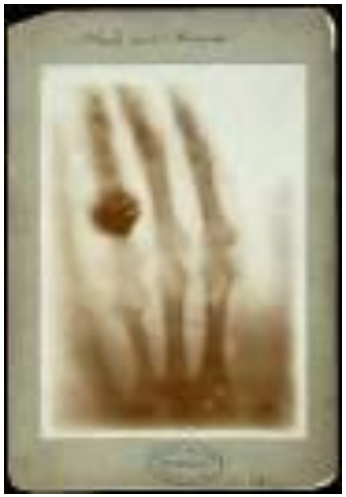
My NASA data

X-ray scattering

electron

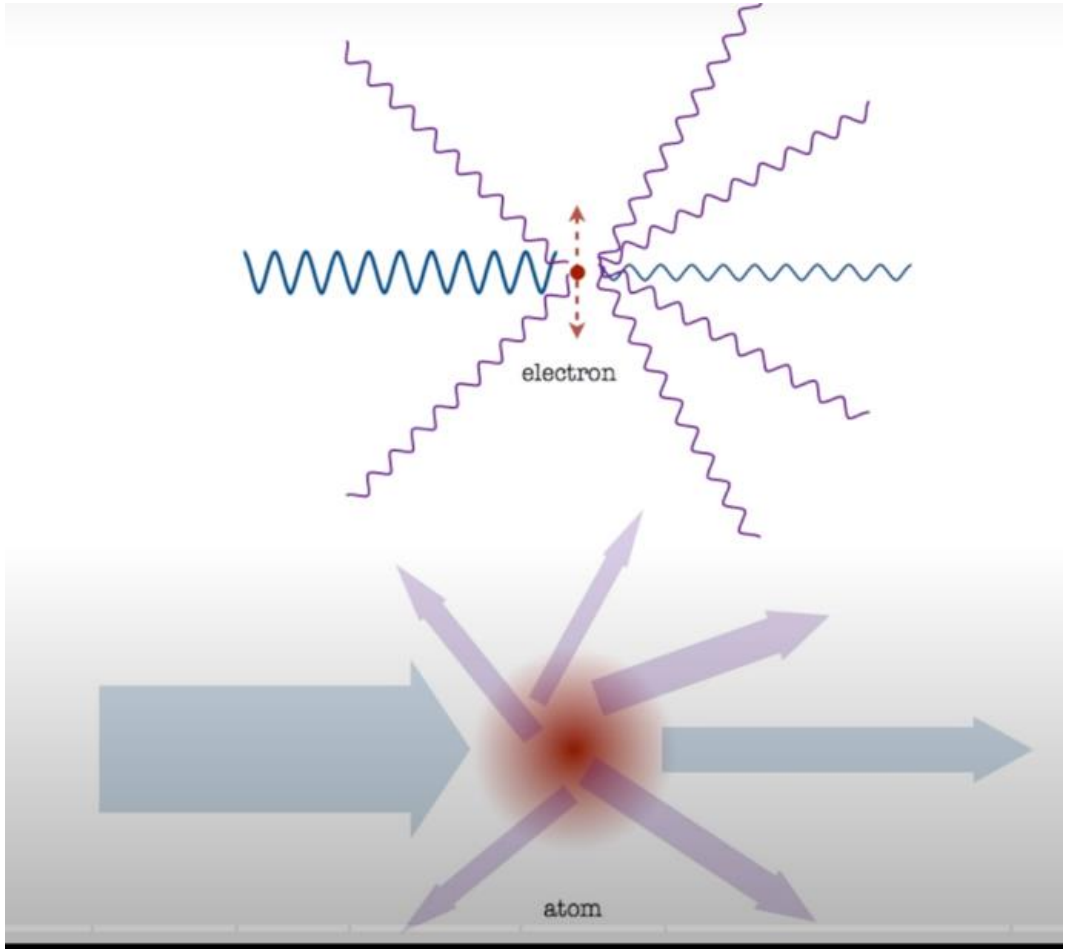
X-ray scattering





Röntgen, 19th century

X-ray scattering



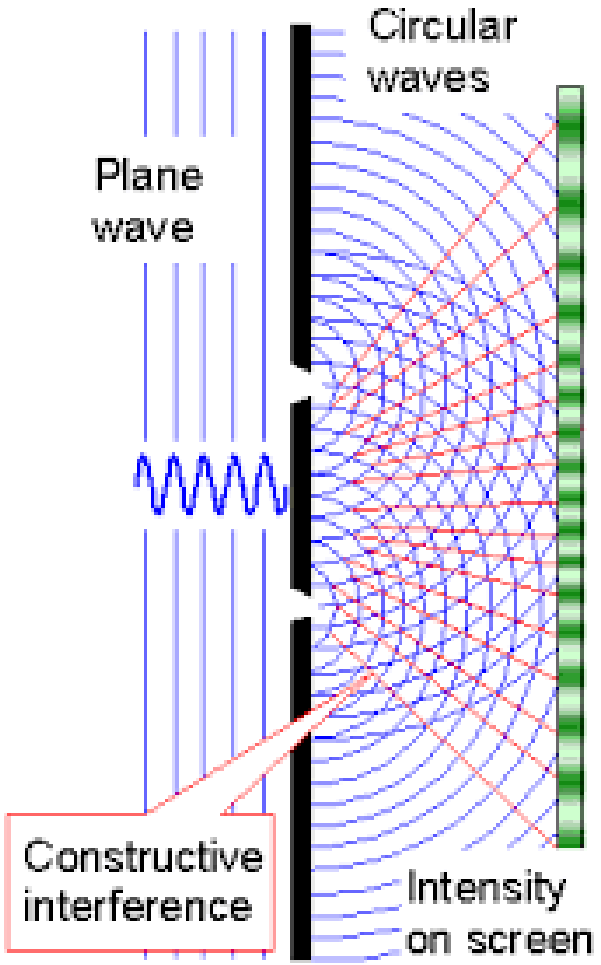
DIFFRACTION experiment

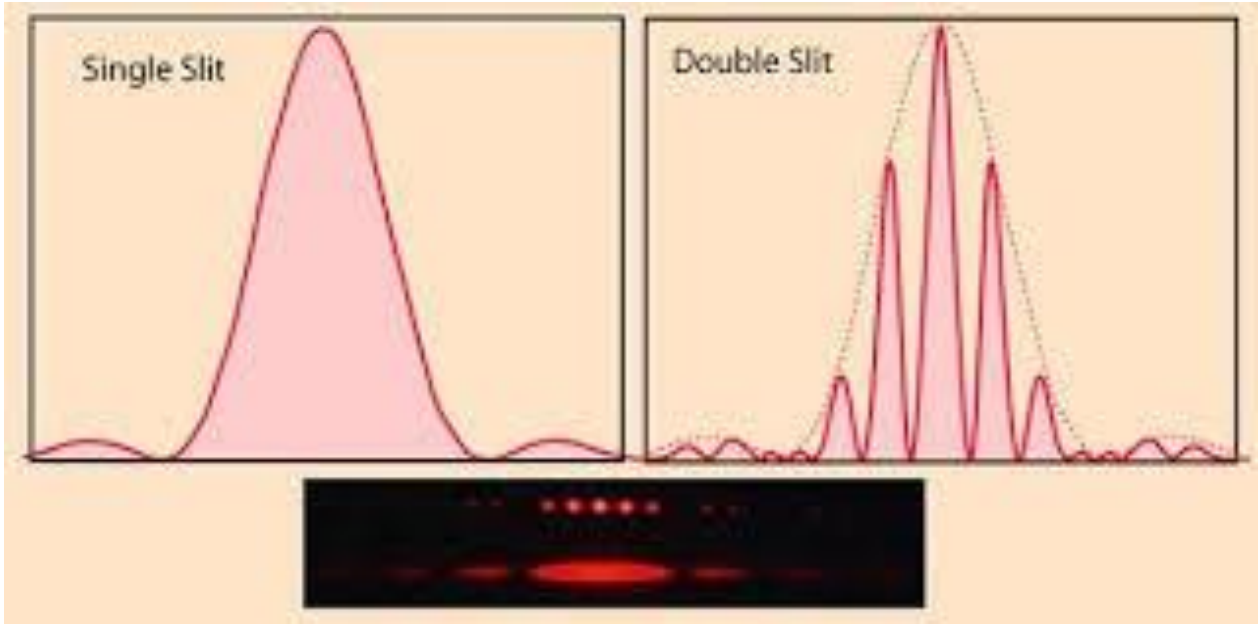
- What we measure is the intensity of the scattered X-rays
  $\lambda out = \lambda in$
- What we calculate is the electron position - electron density - of the object doing the scattering

Why crystals?

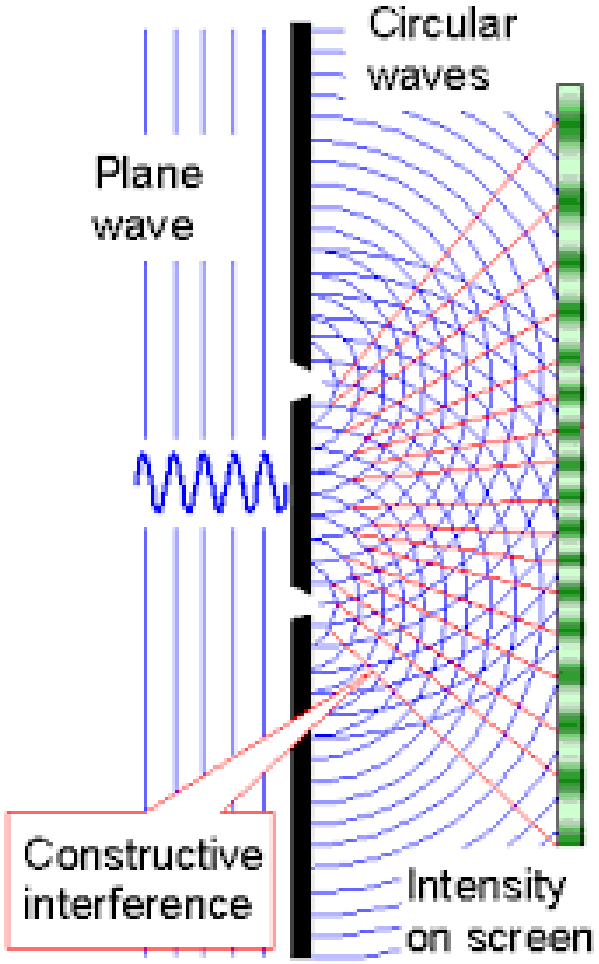
Let’s consider what happens when a light wave encounters **an array** of objects of the same size as its wavelength…

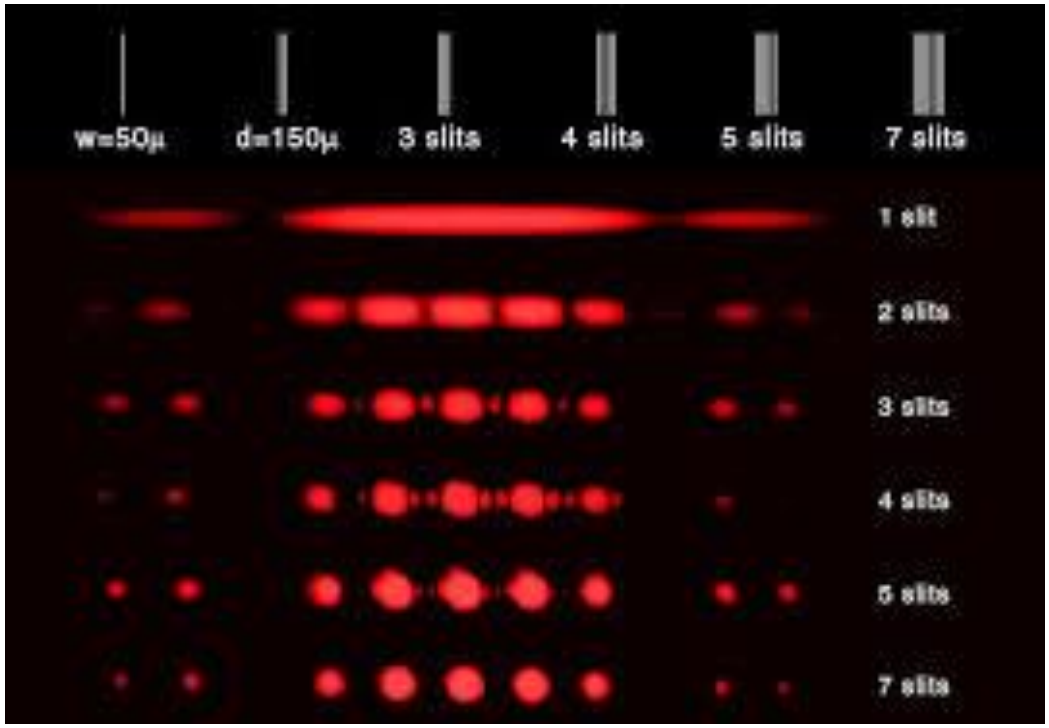
Double slit diffraction pattern of light



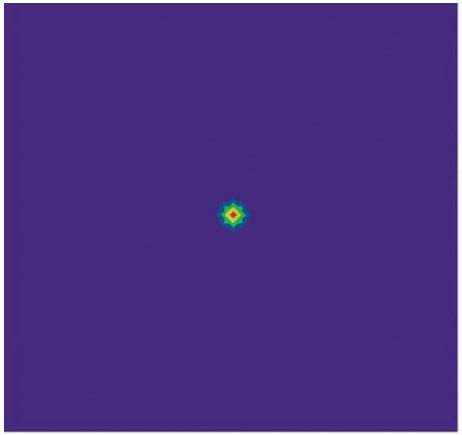


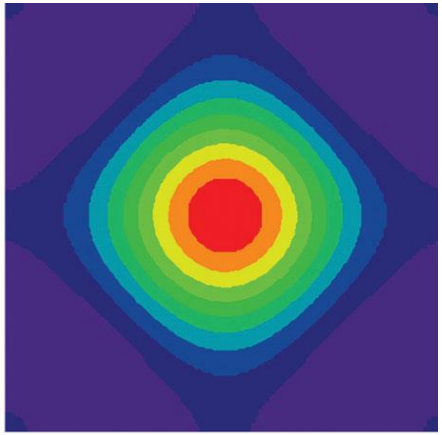
Multiple slit diffraction pattern of light

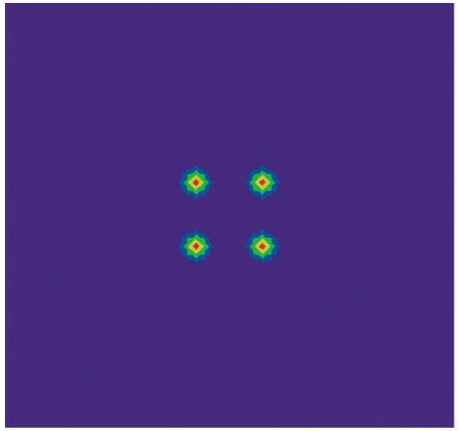


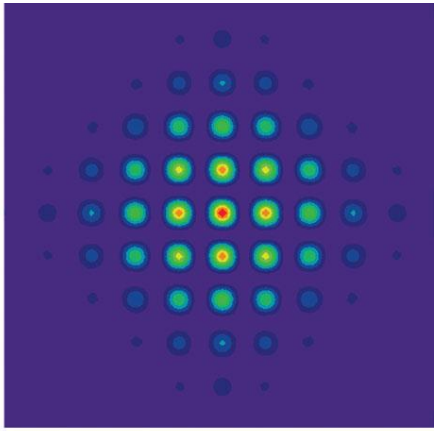


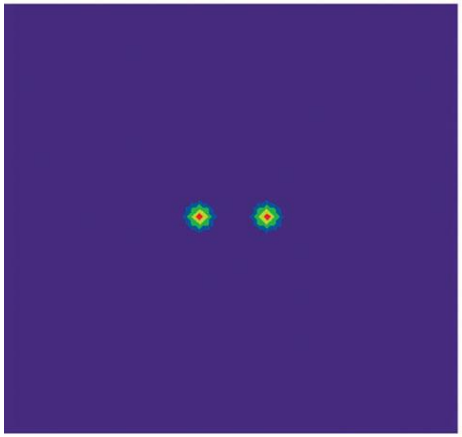
2-D diffraction pattern

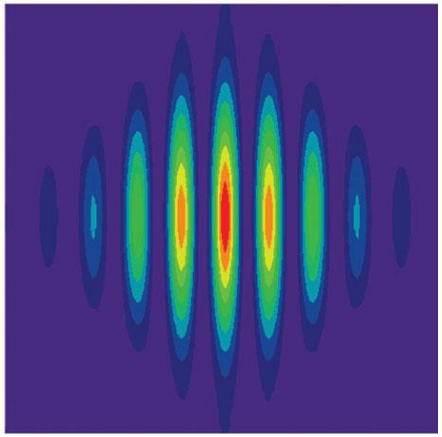


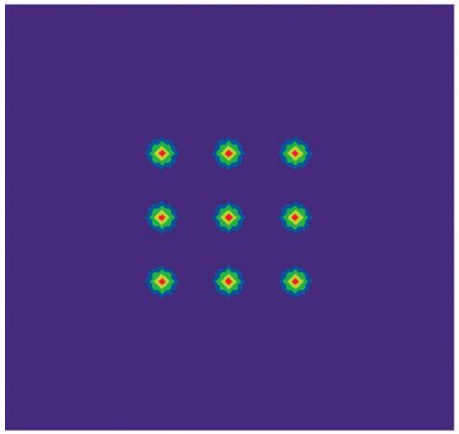


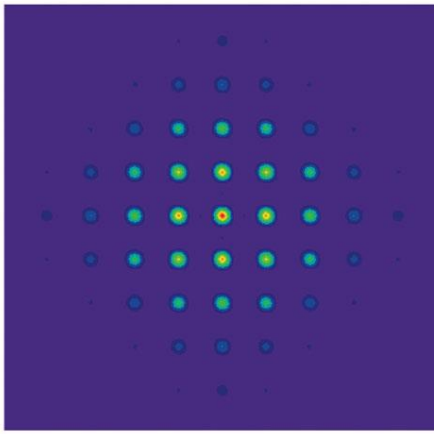












object

Diffraction pattern

object

Diffraction pattern

Why do we need protein crystals?

- X-ray diffraction of a single molecule is very weak and yields limited structural information
- By having protein molecules in repeating units in a 3-D array, scattered X-rays cancel each other out in most directions except for discrete diffraction spots ("amplified signal")

Macromolecular crystallization

- But you need to crystallize the macromolecule.
- For that, you need a relative high amount of pure protein!

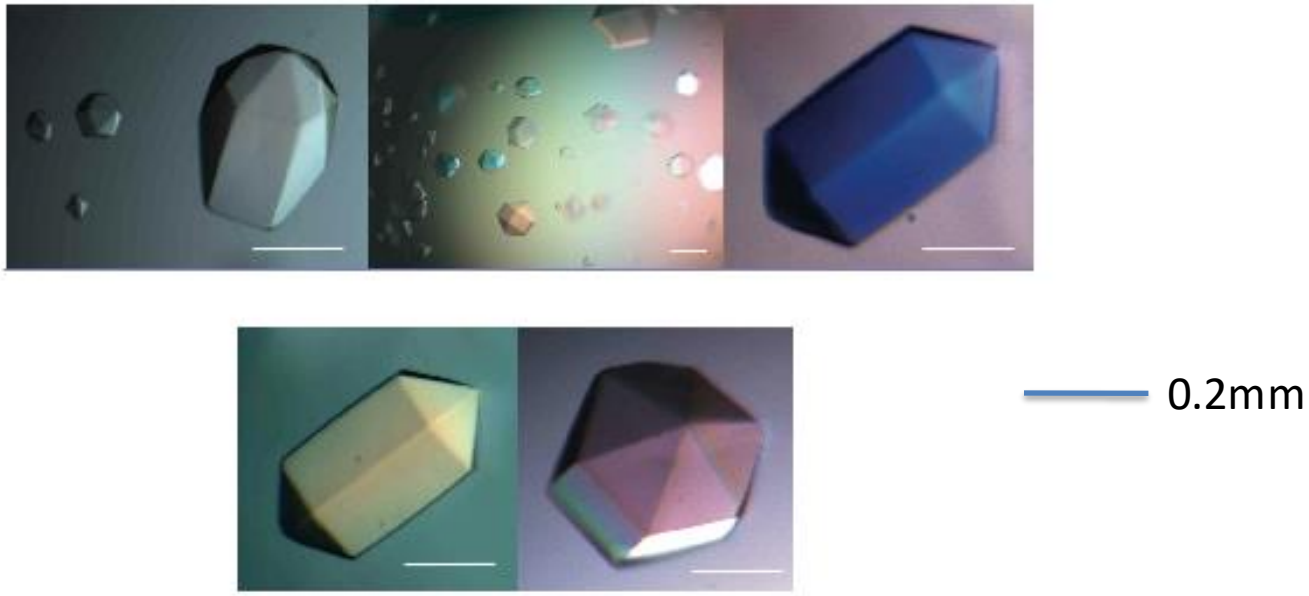
how high and how pure????

Content of lectures

- Why x-rays and why crystals?
- **Macromolecular crystallization**
- Crystal packing
- The diffraction experiment
- The phase problem
- Molecular replacement
- Refinement and model building

Macromolecular crystallization

- Ideally, one needs at least 2 mg of pure sample at high concentrations (5-30 mg ml^{-1}).
  - Good sized macromolecular crystal is 0.2 mm x 0.2 mm x 0.2 mm. Assuming a molecular weight of 50 kDa, it has ≈10^{13} molecules ≈ 1 µg of macromolecule.



Macromolecular crystallization

- Ideally then, one needs at least 2 mg of pure sample at high concentrations (5-30 mg ml^{-1}).
  - Good sized macromolecular crystal is 0.2 mm x 0.2 mm x 0.2 mm. Assuming a molecular weight of 50 kDa, it has ≈10^{13} molecules ≈ 1 µg of macromolecule.

Obtaining these amounts of macromolecule is the first challenge in protein crystallography.

Macromolecule expression and purification

How?

1. Can purify it from natural sources
2. Make it with recombinant methods:
   - can obtain larger quantities of sample,
   - can control precisely which sequence (s) to work on
   - can add affinity tags to aid purification.

Macromolecule expression and purification

- Recombinant methods
  1. Bacterial expression systems, normally *E. coli*
  2. Eukaryotic expression systems
     - Necessary when work on eukaryotic complexes or membrane proteins or when macromolecule needs posttranscriptional modifications
     - more expensive and time consuming
     - Can use yeasts, insect cells or mammalian cells, among others.

Macromolecule expression and purification

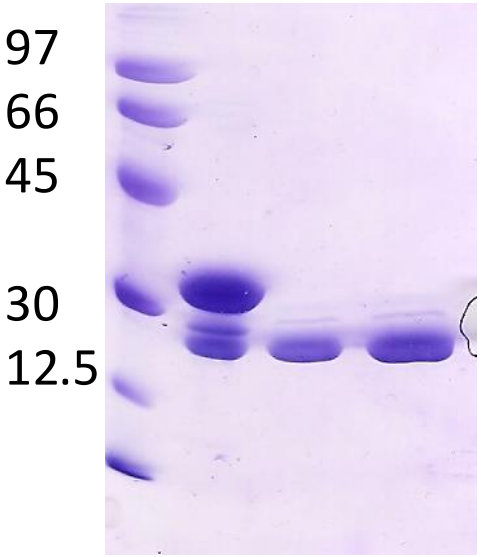
- Trouble shooting
  - Good strategy to test many isoforms, homologs, and variants of the protein or protein complex of interest in parallel.
    - Amino acid composition
    - Flexible parts - conservation
  - Macromolecular complexes are a challenge, specially when individual proteins cannot be purified individually.

Macromolecule expression and purification

- Purity
  - Chemically pure
  - Conformationally pure

Macromolecule expression and purification

- Purity
  - Chemically pure (assessed by gel electrophoresis, protein chromatography and mass spectrometry)



- One type of macromolecule
- Must not contain covalent heterogeneity
- Must not contain mixtures of truncation products

Macromolecule expression and purification

- Purity
  - Conformationally pure (assessed by size exclusion chromatography and light scattering)
    - Soluble (do not want non-specific aggregates)
    - Homogeneous - monodisperse
    - Can avoid flexible parts? (add mutations?)
    - Binding of partners or cofactors (other proteins, nucleic acids, small molecules) may help stabilize macromolecule in one conformation.

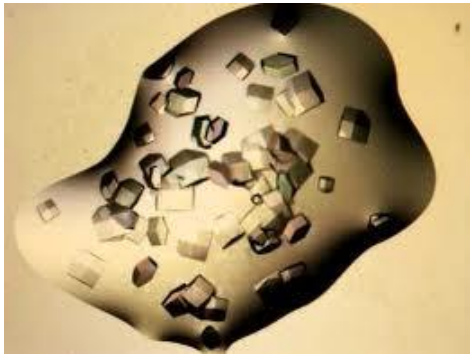
Crystallization

Mix the macromolecule solution with a precipitant mix

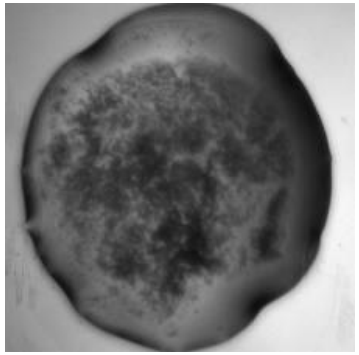
Obtaining crystals is the second (biggest) challenge in protein crystallography.

Intermolecular interactions

crystals



Simple, amorphous precipitation.



Crystallization

- Two ingredients:
  - Macromolecule solution – PURE, find ideal concentration (normally very high!)
  - Precipitant solution

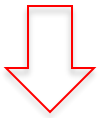
Crystallization

- Precipitant solution:
  - affect macromolecular hydration, molecular crowding, solubility, hydrophobic interactions and electrostatics
  - various sizes of polyethylene glycols (PEGs), alcohols, and salts (successful in the past)

Crystallization

- Parameters that affect crystallization:
  - macromolecule concentration and quality
  - presence of ligands
  - precipitant type, salt concentration, presence of detergent
  - pH, temperature
  - …

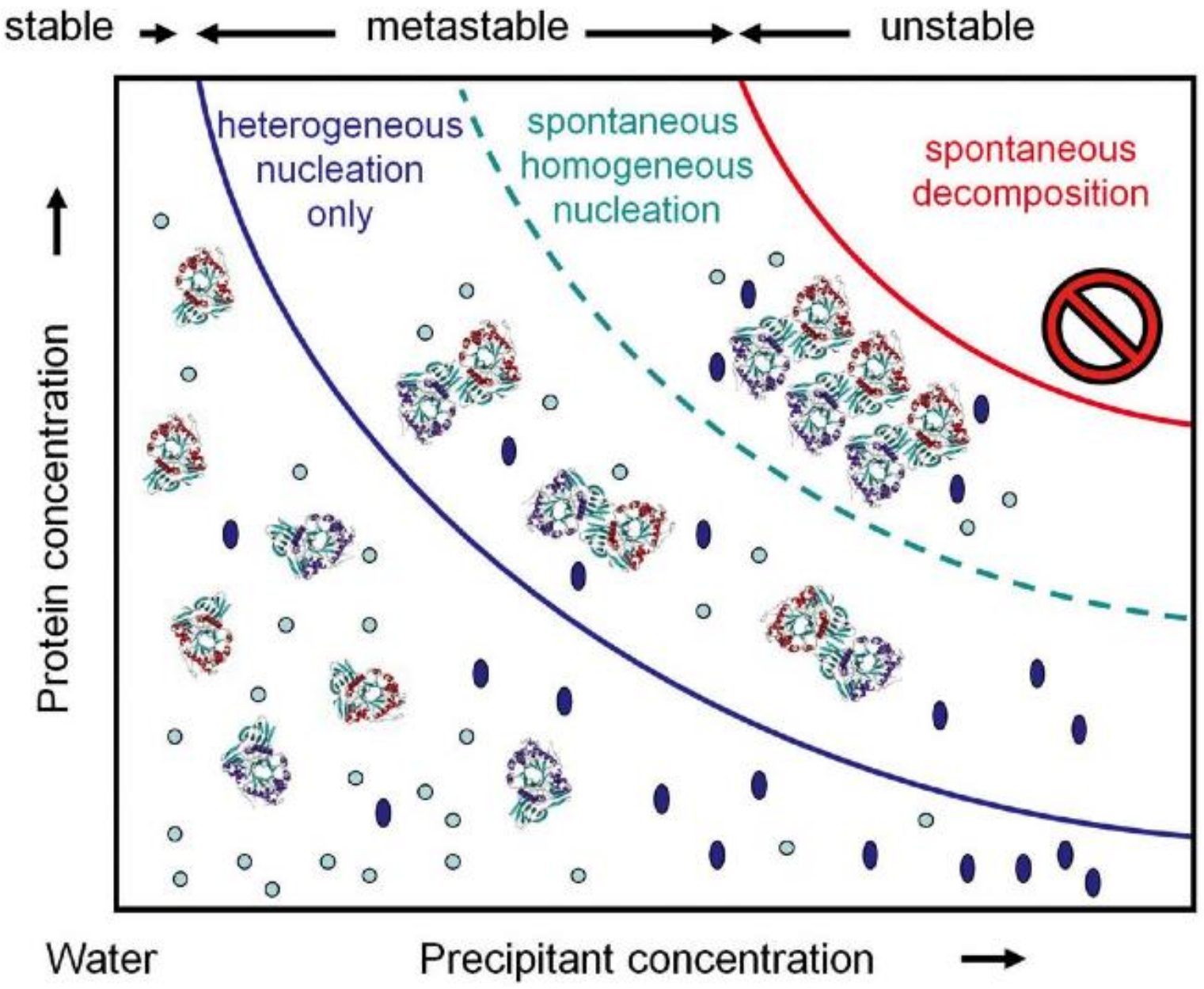
∞ variables!!



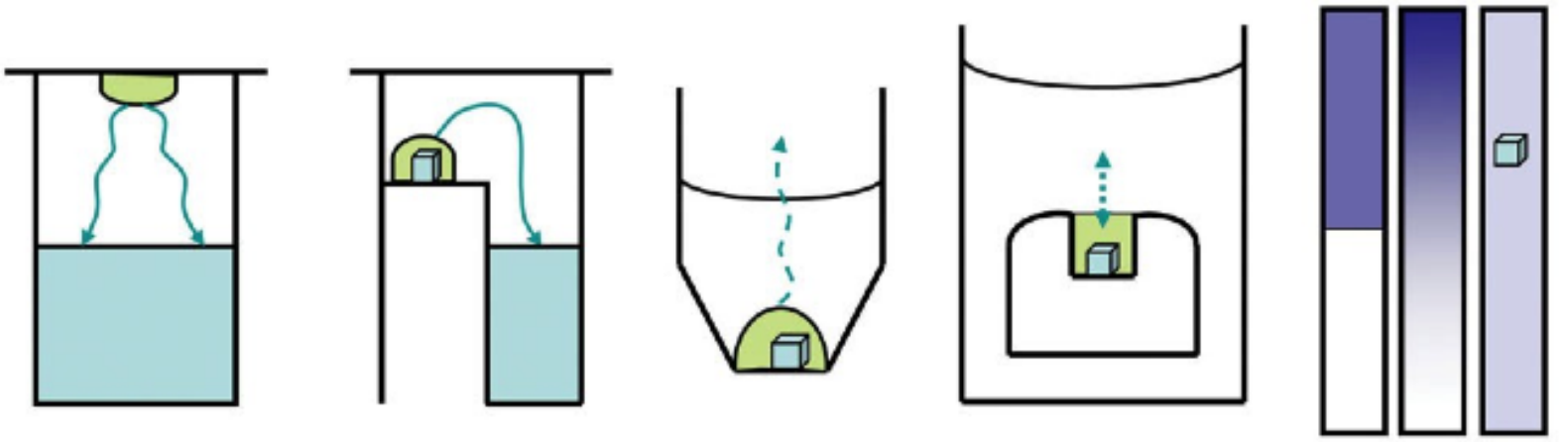
sparse matrix screens and robots

Parameters affecting crystallization

PHYSICAL	CHEMICAL	BIOCHEMICAL
Temperature	pH	Purity of sample
Surface	Precipitant type	Ligands, inhibitors
Method to approach equilibrium	Precipitant concentration	Aggregation state
Gravity	Ionic strength	Source of macromolecule
Pressure	Specific ions	Chemical modifications
Time	Degree of supersaturation	Genetic modifications
Vibrations/mechanical perturbation	Reductive/oxidative environment	Inherent symmetry of macromolecule
Electro/magnetic fields	Concentration of macromolecule	Isoelectric point
Viscosity of medium	Metal ions	Stability of sample
Rate of equilibrium	Polymers	
	Detergents	



Popular crystallization techniques



The classic: hanging-drop vapor diffusion

The variant: sitting-drop vapor diffusion

Micro-batch under oil

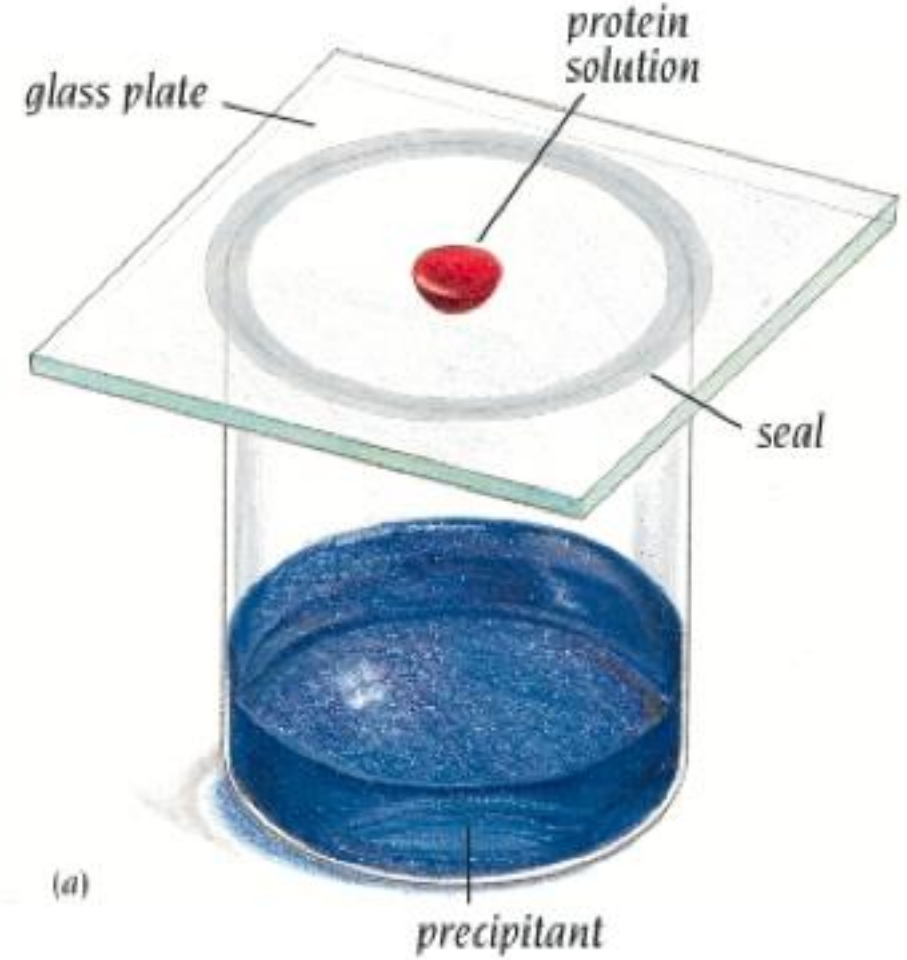
Micro-dialysis

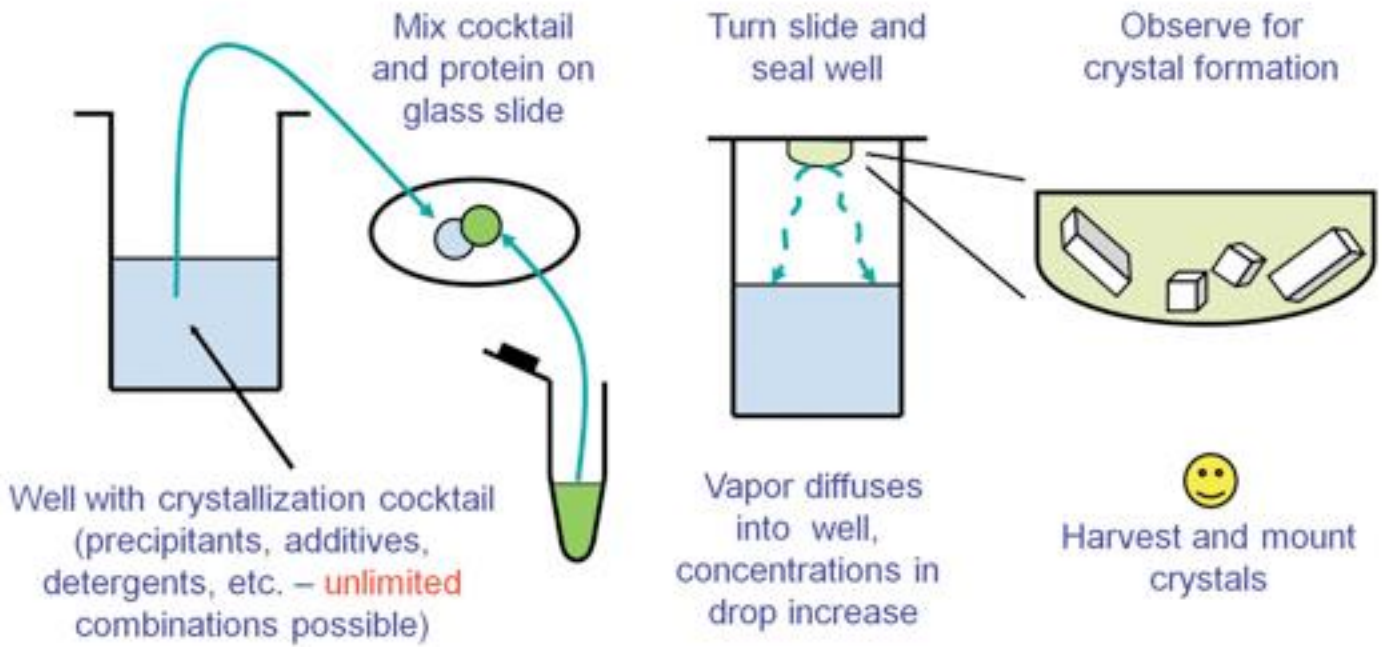
Free-interface diffusion

Lab practical tomorrow Thursday 8th of May in lab AI 2142 2nd Floor
Meet at 8:10 @ SV Reception desk

Crystallization

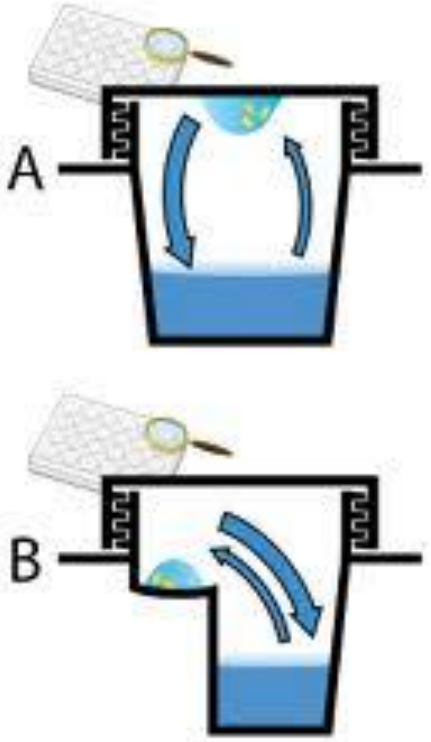
- Methods
    1) Vapour diffusion
        A: hanging drop





Crystallization

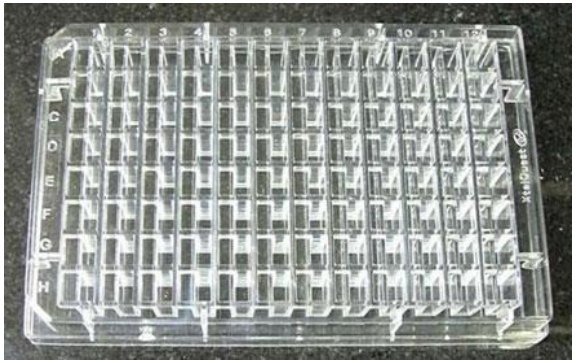
- Methods
    1) Vapour diffusion
        A: hanging drop
        B: sitting drop

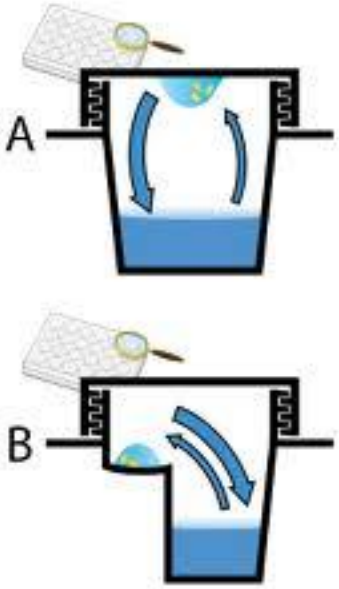


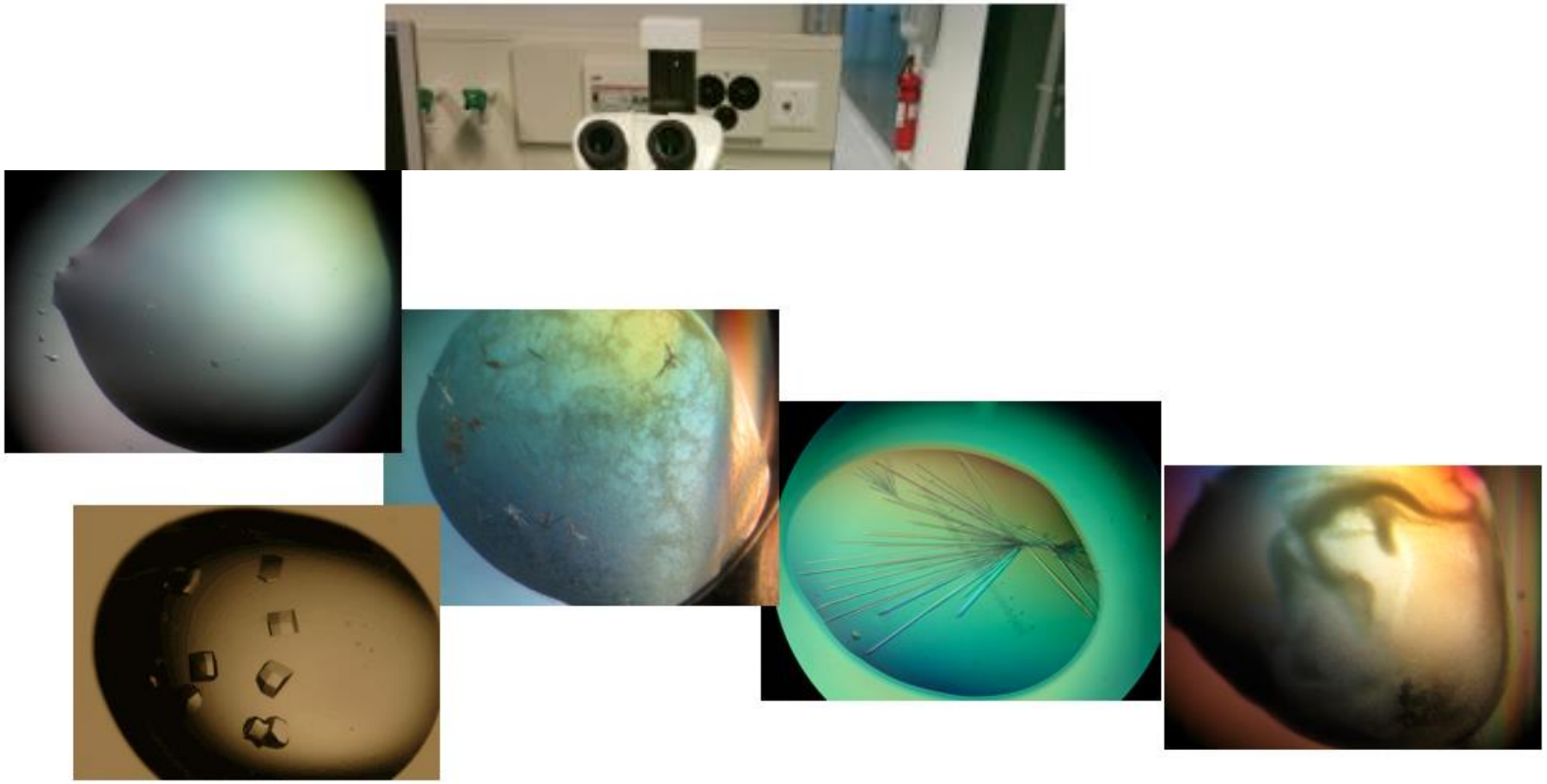
Crystallization

- Methods

  1) Vapour diffusion



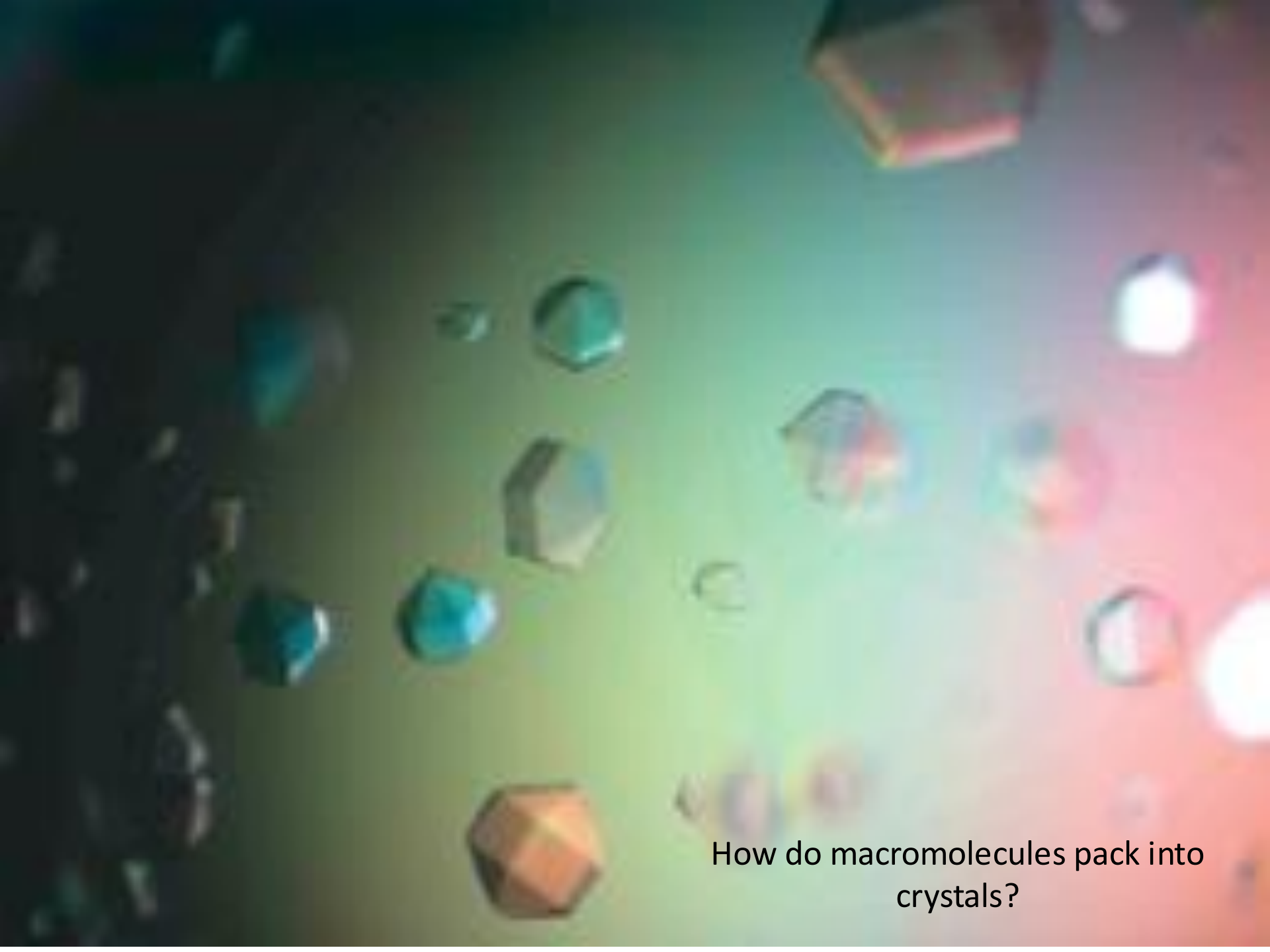




Vary conditions to optimize crystals

Content of lectures

- Why x-rays and why crystals?
- Macromolecular crystallization
- **Crystal packing**
- The diffraction experiment
- The phase problem
- Molecular replacement
- Refinement and model building


How do macromolecules pack into crystals?

The crystals

A crystal is a homogenous repetitive arrangement of atoms.

Two definitions:

- Unit cell
- Asymmetric unit (AU)

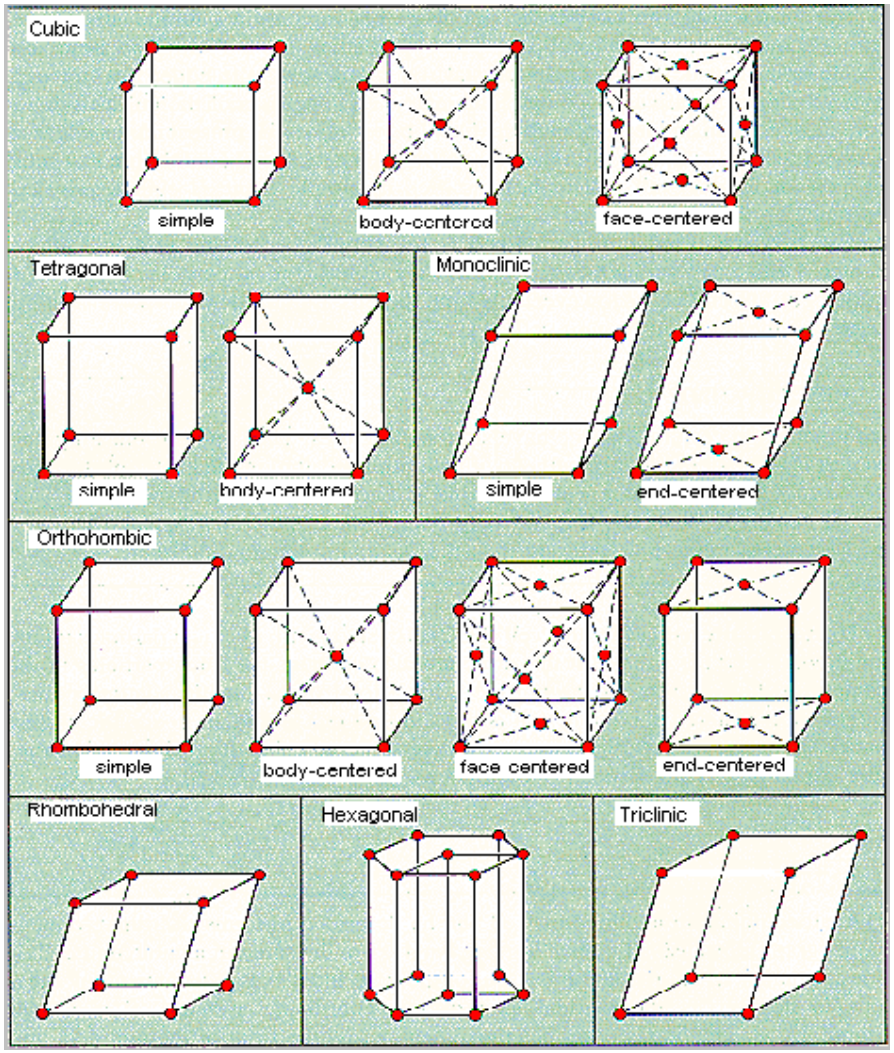
The crystals

A crystal is a homogenous repetitive arrangement of atoms.

Two definitions:

- Unit cell: the repetitive unit that reproduces the crystal by applying the lattice translations

The unit cell



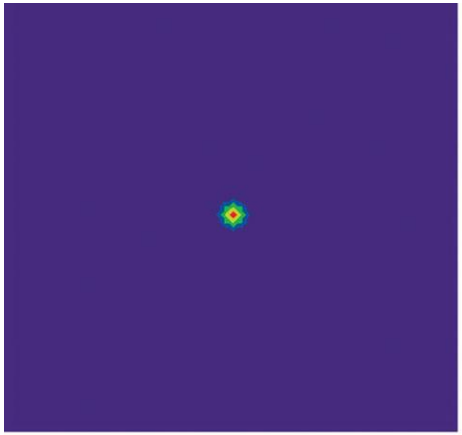
14 types of unit cells

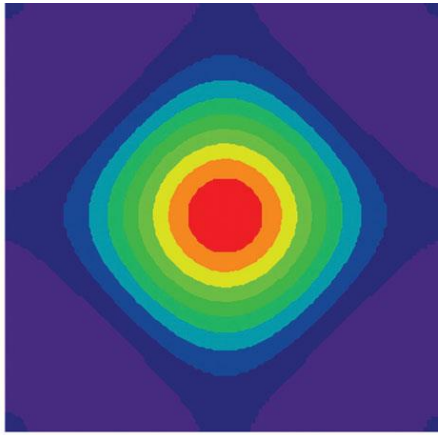
Cell parameters:

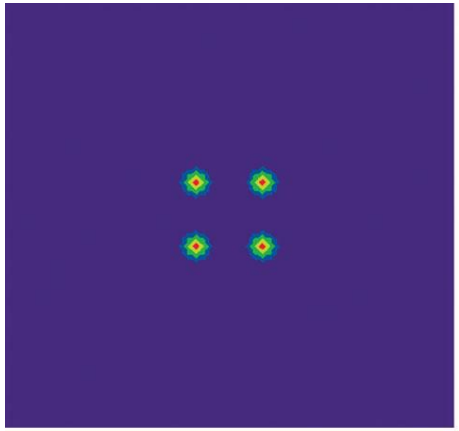
Cell axes lengths: a, b, c

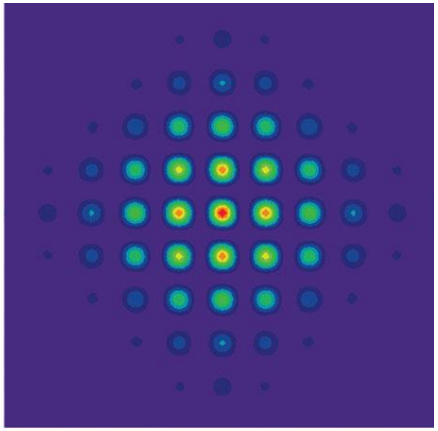
Cell axes angles: α, β, γ

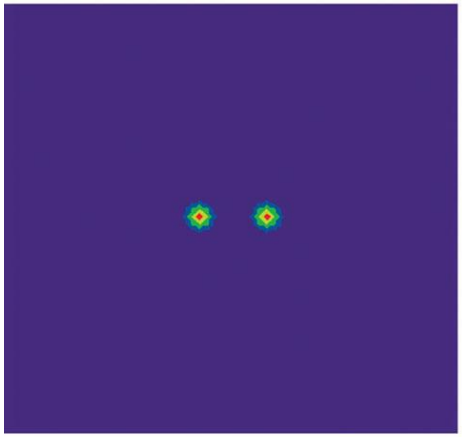
2-D diffraction pattern

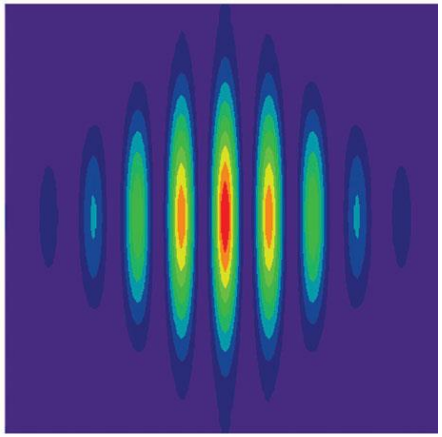


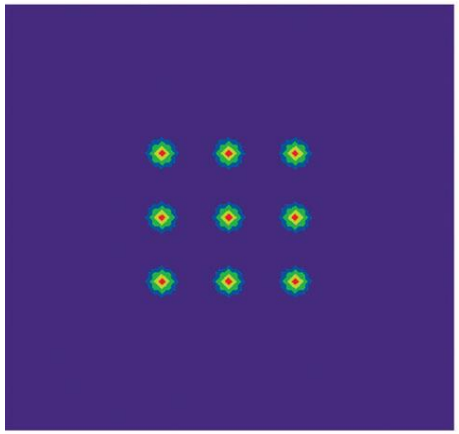


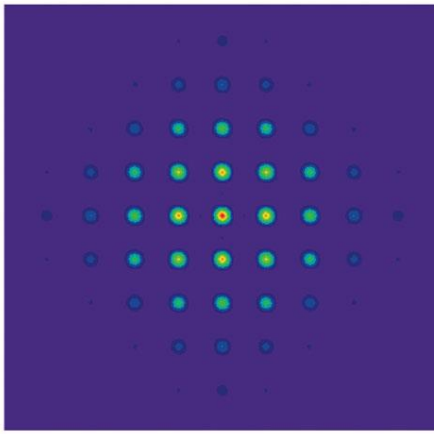












object

Diffraction pattern

object

Diffraction pattern

Symmetry

The unit cell may contain a single copy of the macromolecule, but more often than not, it contains many copies that are related to each other by symmetry operations (translation, rotation and screw axis).

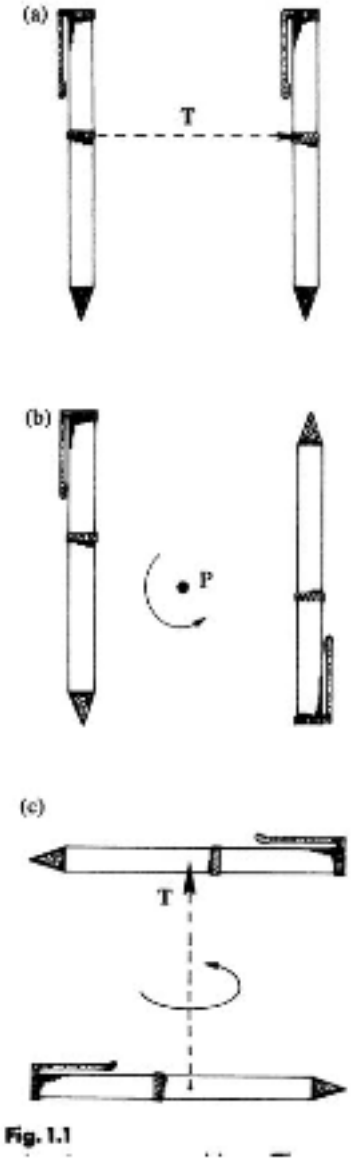


Fig. 1.1

Translation

Rotation

Screw axis

Because proteins are asymmetric molecules composed of chiral structures, the only possible symmetry operations are rotation, translation and screw-axis (combination of rotation and translation)

The crystals

A crystal is a homogenous repetitive arrangement of atoms.

Two definitions:

- Unit cell: the repetitive unit that reproduces the crystal by applying the lattice translations
- Asymmetric unit (AU): is the smallest possible unit that reproduces the cell by applying the symmetry operations.

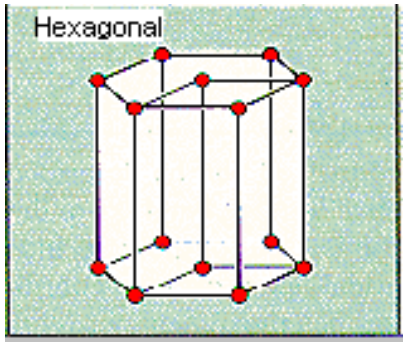
We solve the structure of the asymmetric unit.

Symmetry – space groups

The symmetry operations are defined by the space group.

Therefore, for protein crystals, there are 65 types of crystallographic symmetry (65 space groups).

Example: space groups for the Primitive Hexagonal lattice.



Primitive Hexagonal	P31	143	P3	
		144	P31	(0,0,3n)*
		145	P32	(0,0,3n)*
	P3112	149	P312	
		151	P3112	(0,0,3n)*
		153	P3212	(0,0,3n)*
	P3121	150	P321	
		152	P3121	(0,0,3n)*
		154	P3221	(0,0,3n)*
	P61	168	P6	
		169	P61	(0,0,6n)*
		170	P65	(0,0,6n)*
		171	P62	(0,0,3n)**
		172	P64	(0,0,3n)**
		173	P63	(0,0,2n)
	P6122	177	P622	
		178	P6122	(0,0,6n)*
		179	P6522	(0,0,6n)*
		180	P6222	(0,0,3n)**
		181	P6422	(0,0,3n)**
		182	P6322	(0,0,2n)

Crystal packing vs biological assembly

Asymmetric unit content (Matthews coefficient):

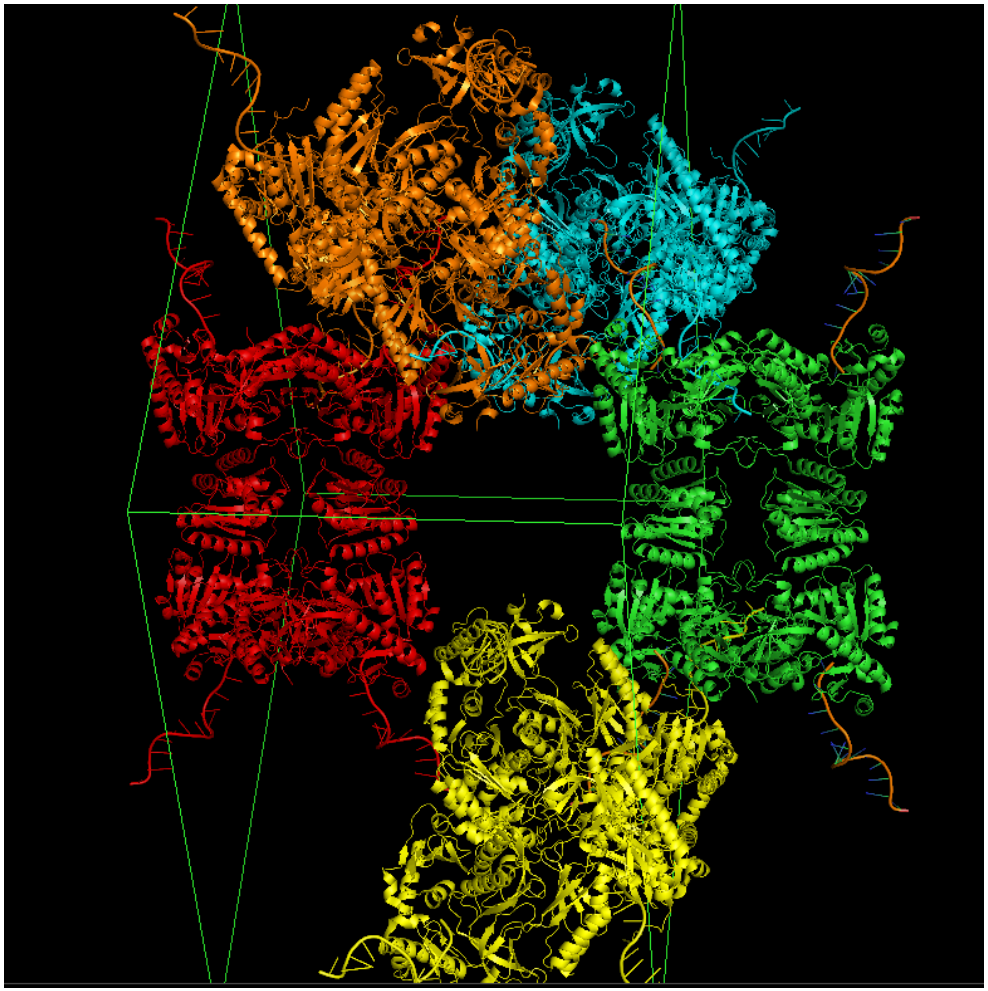
Normal macromolecular crystals are “very liquid”: solvent content between 30-70% - very fragile – low diffraction (compared to salt crystals)

Difference between oligomers formed by crystal contacts (artefacts) or real complexes with physiological relevance.

Sometimes there is space in the crystals for ligands to diffuse inside a pocket or an active site.

Crystal packing vs biological assembly

Look at RNase E, IDmo-I and PREPL structures as examples

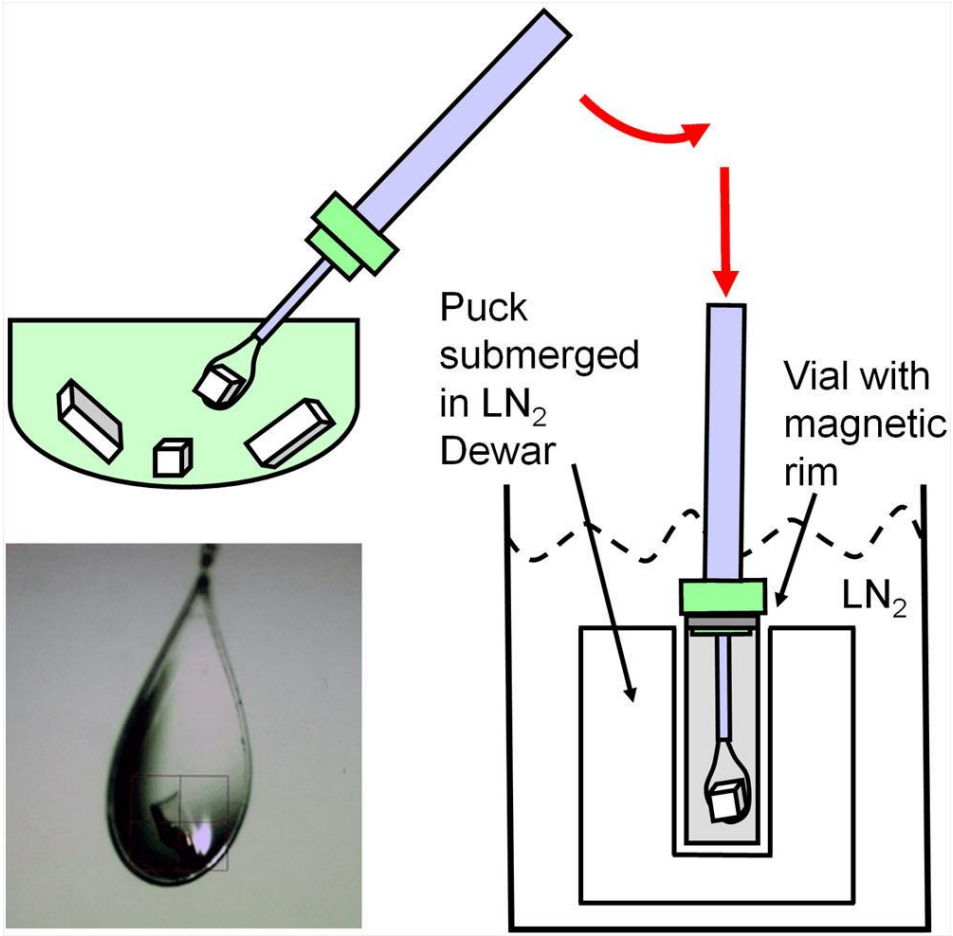


Content of lectures

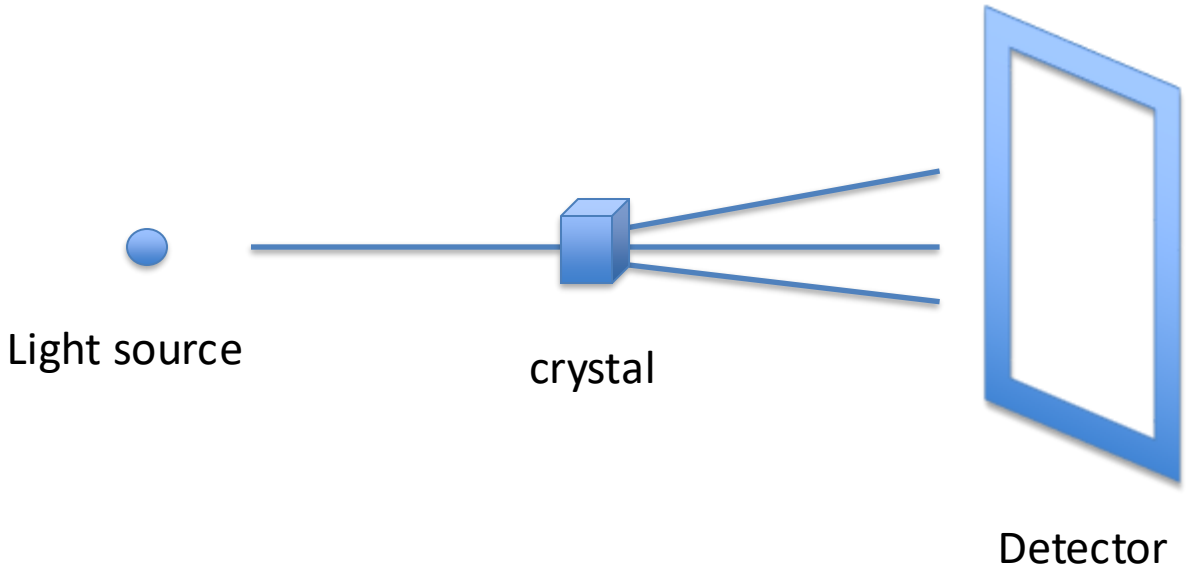
- **Why x-rays and why crystals?**
- **Macromolecular crystallization**
- **Crystal packing**
- The diffraction experiment
- The phase problem
- Molecular replacement
- Refinement and model building

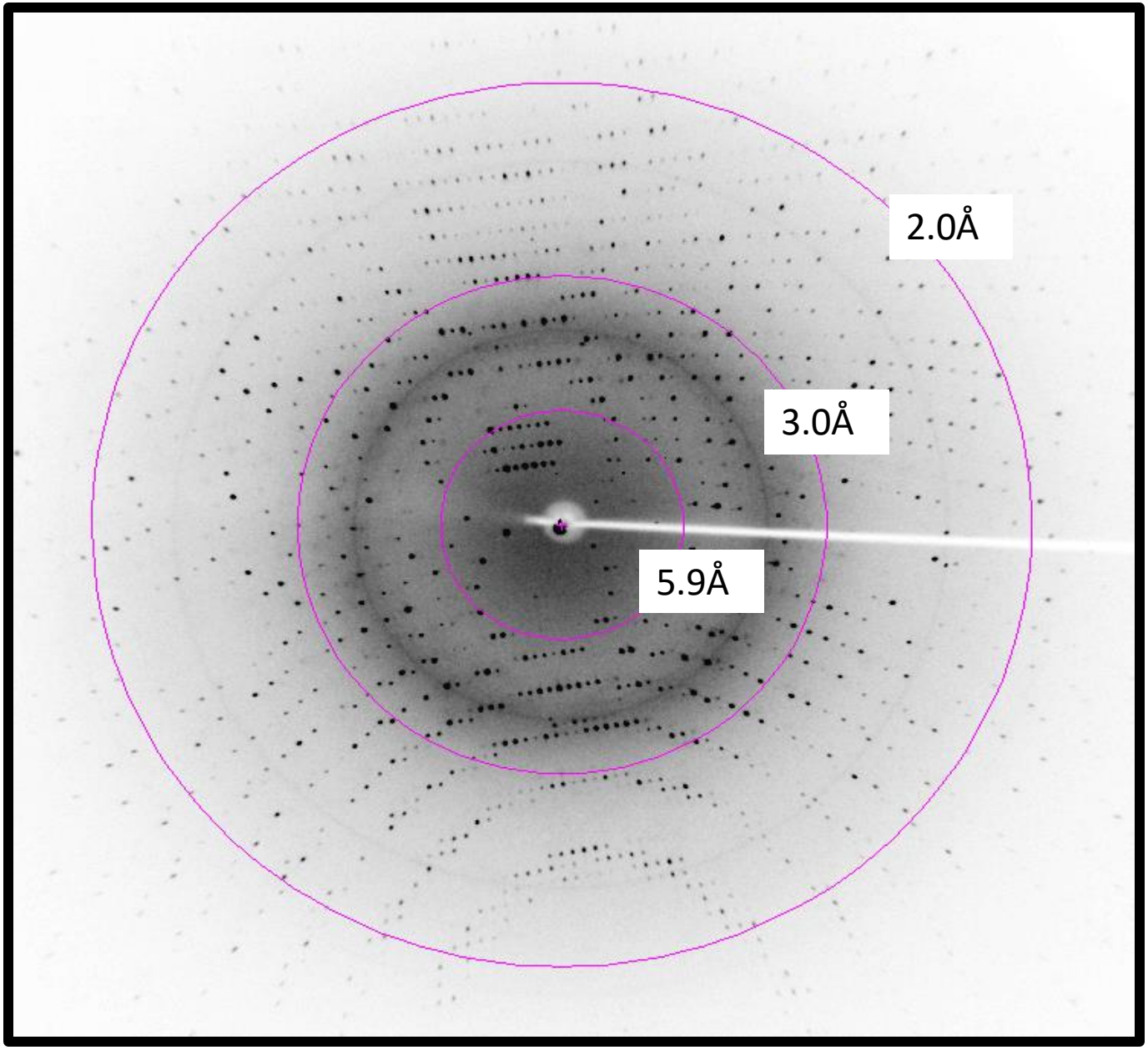
Freezing crystals

- Need to cool crystal at 100K while it is being irradiated with the X rays to avoid radiation damage.
- Care must be taken in the procedure of freezing the crystal to avoid crystal cracking or formation of ice crystals.
- Dip crystal in cryoprotectant solution (containing for example glycerol, low MW PEG's) which freezes without forming ice.

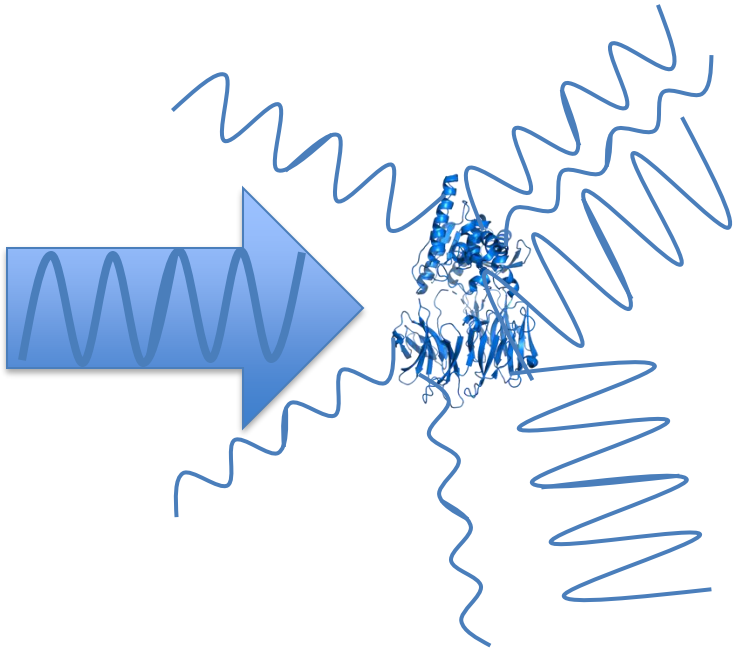


- Radiate the crystals with x-rays (λ=1Å)

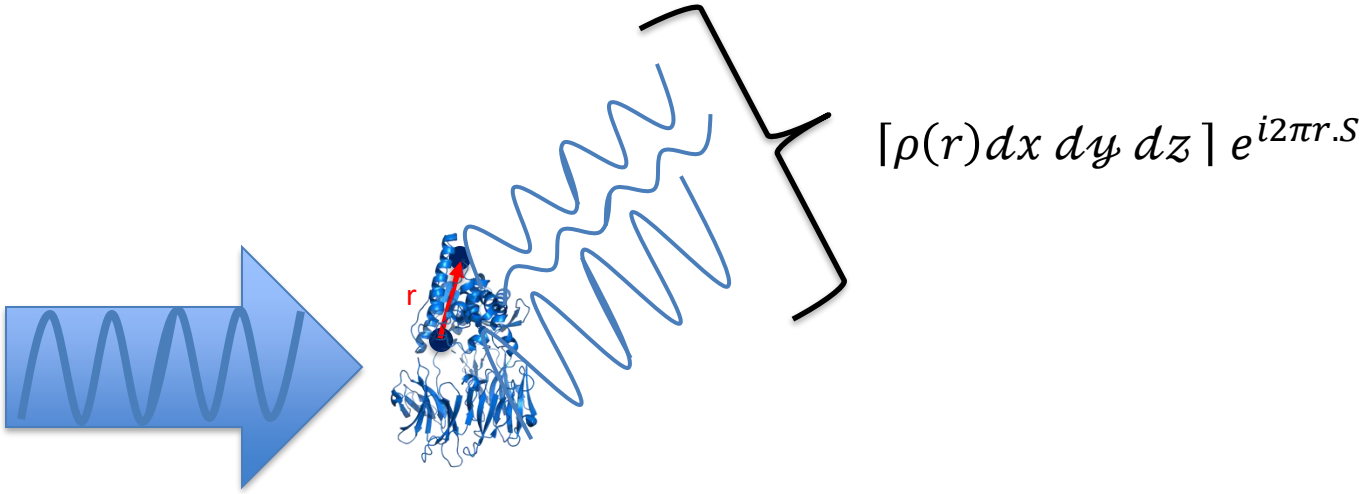



2.0Å
3.0Å
5.9Å

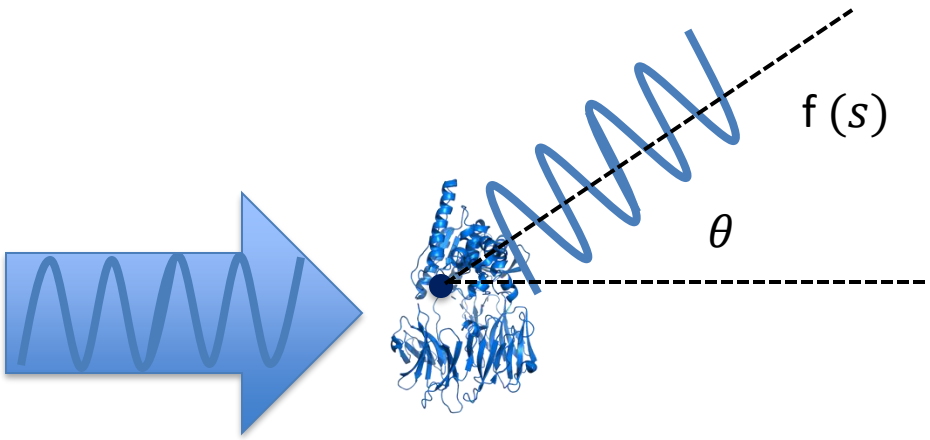
Diffraction from a molecule



Diffraction from a molecule

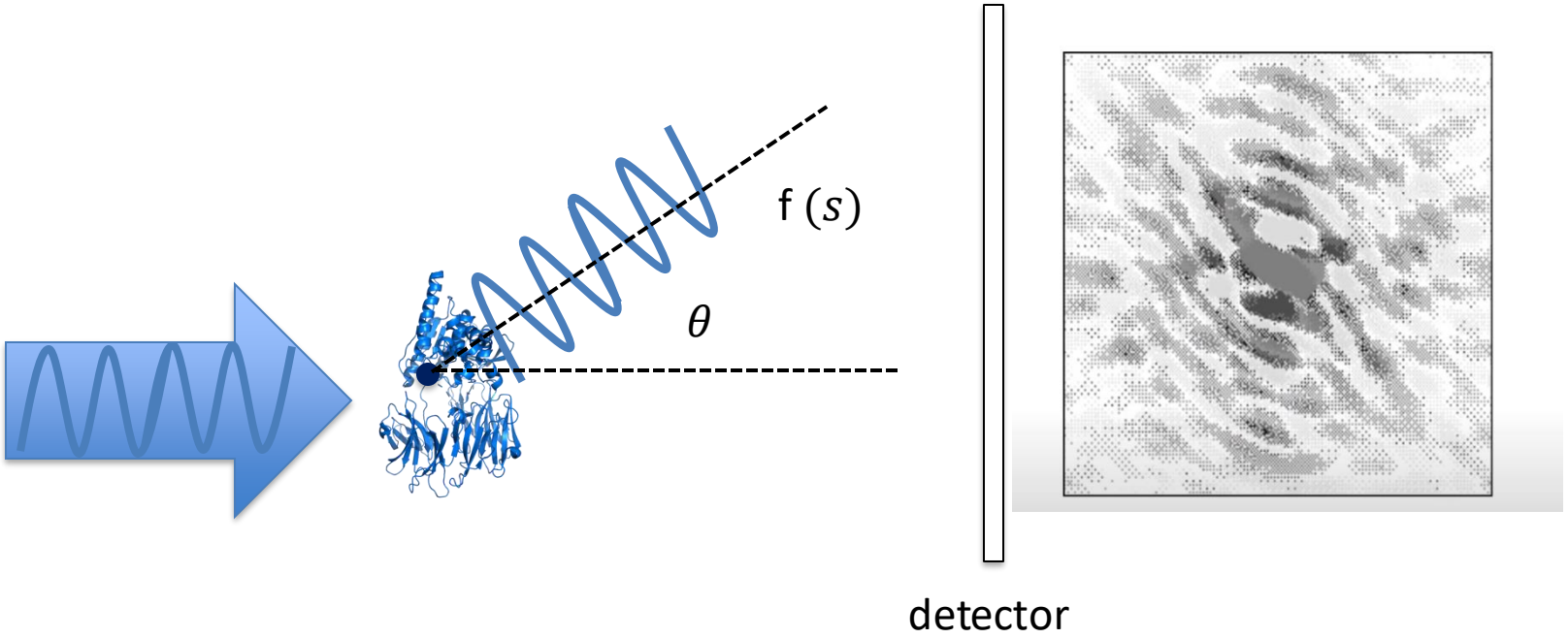


Diffraction from a molecule



$$\mathrm{f}\,(s) = \iiint [\rho(r) dx\, dy\, dz\,]\, e^{i2\pi r.S}$$

Diffraction from a molecule



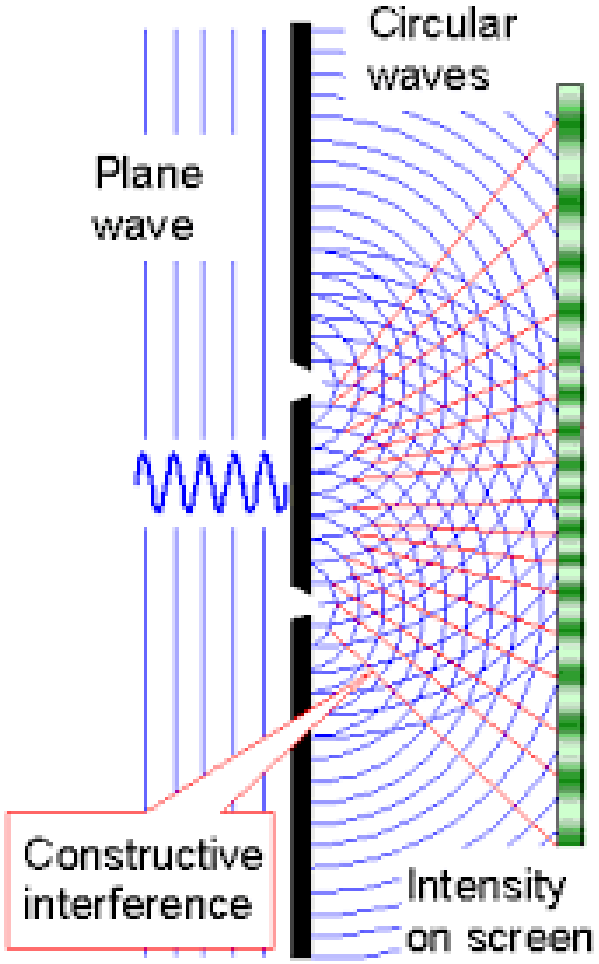
$$\mathrm{f}\,(s) = \iiint \rho(r)\, e^{i2\pi r.S}\, dxdydz$$

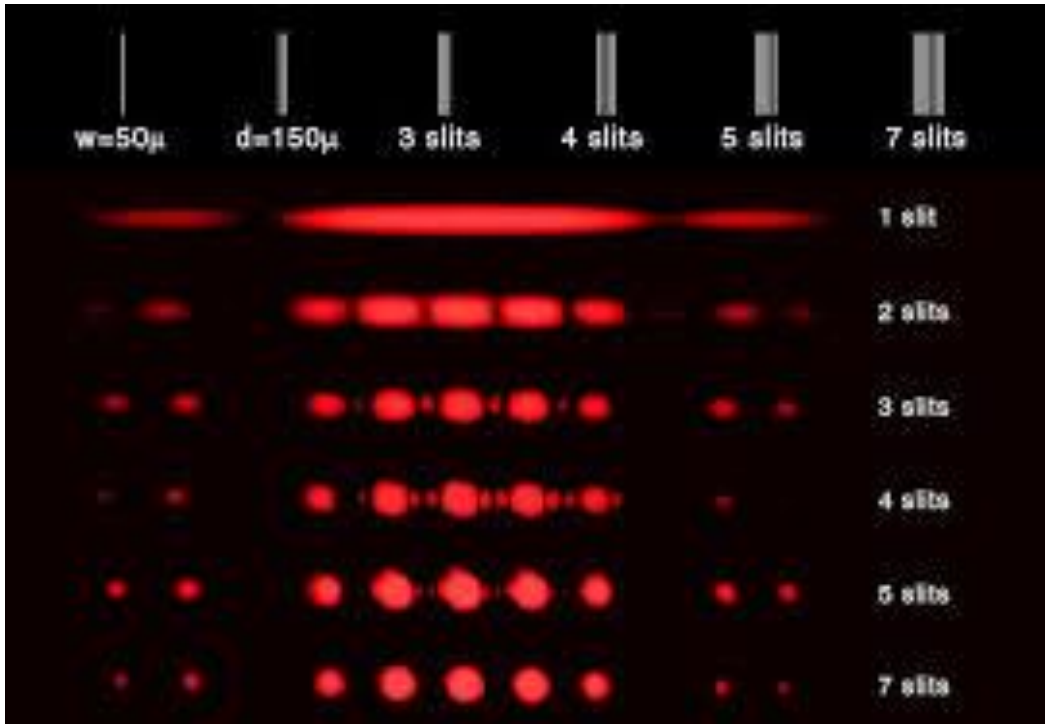
All atoms in the molecule contribute to each diffracted X-ray and to the calculation of each structure factor

Why do we need protein crystals?

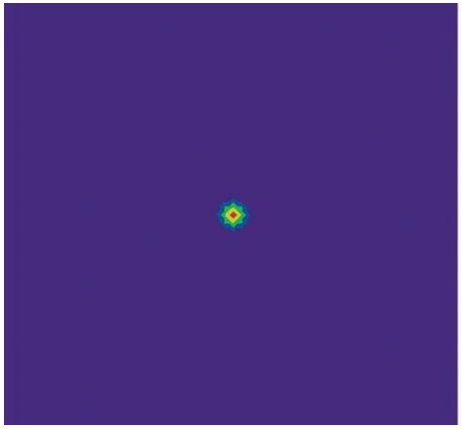
- X-ray diffraction of a single molecule is very weak and yields limited structural information
- By having protein molecules in repeating units in a 3-D array, scattered X-rays cancel each other out in most directions except for discrete diffraction spots ("amplified signal")

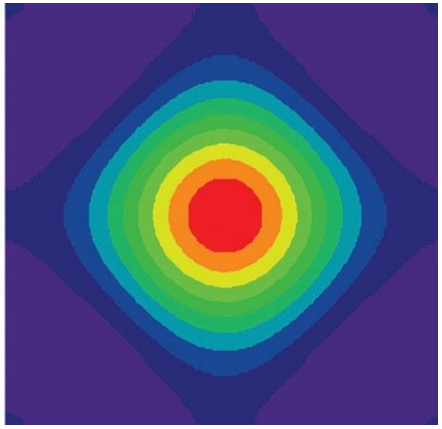
Multiple slit diffraction pattern of light

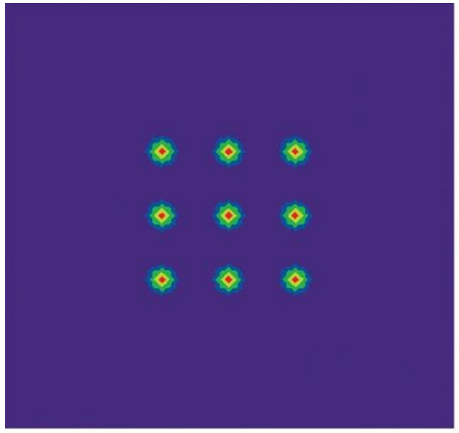


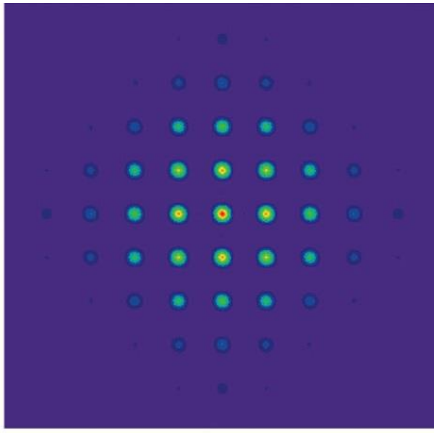


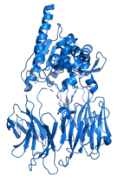
2-D diffraction pattern

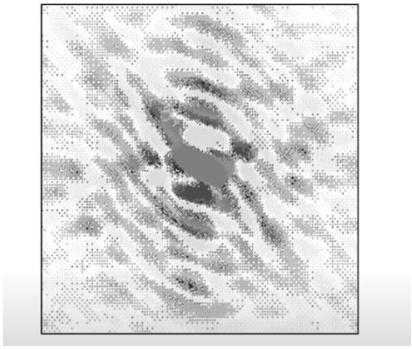


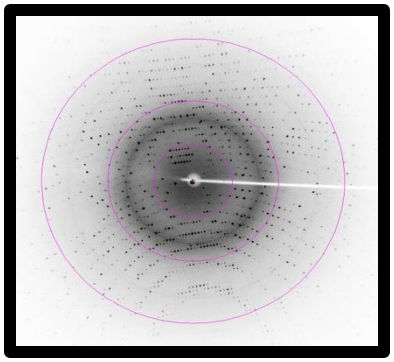










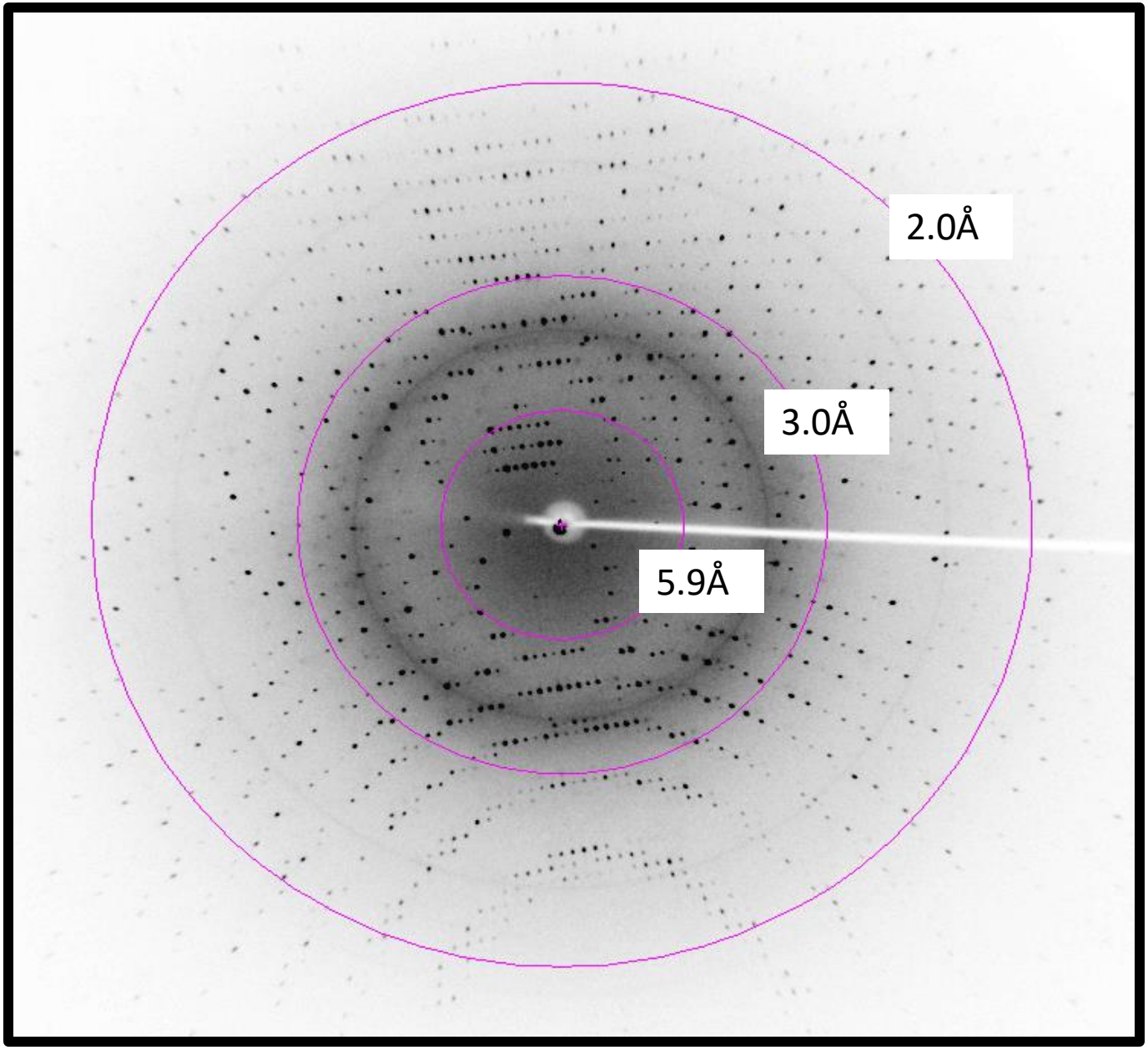


object

Diffraction pattern

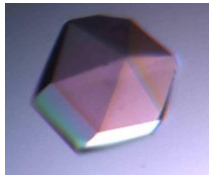
object

Diffraction pattern


2.0Å
3.0Å
5.9Å

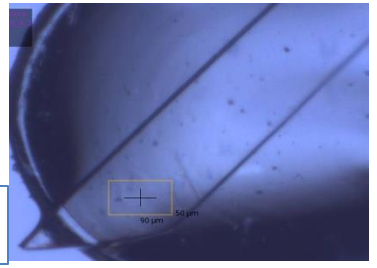
Practical examples from the lab

- 3 Å – Aurora A-inhibitor complex



- 2.1 Å – I-DmoI – DNA complex

- 1.6 Å – SPF- ligand complex



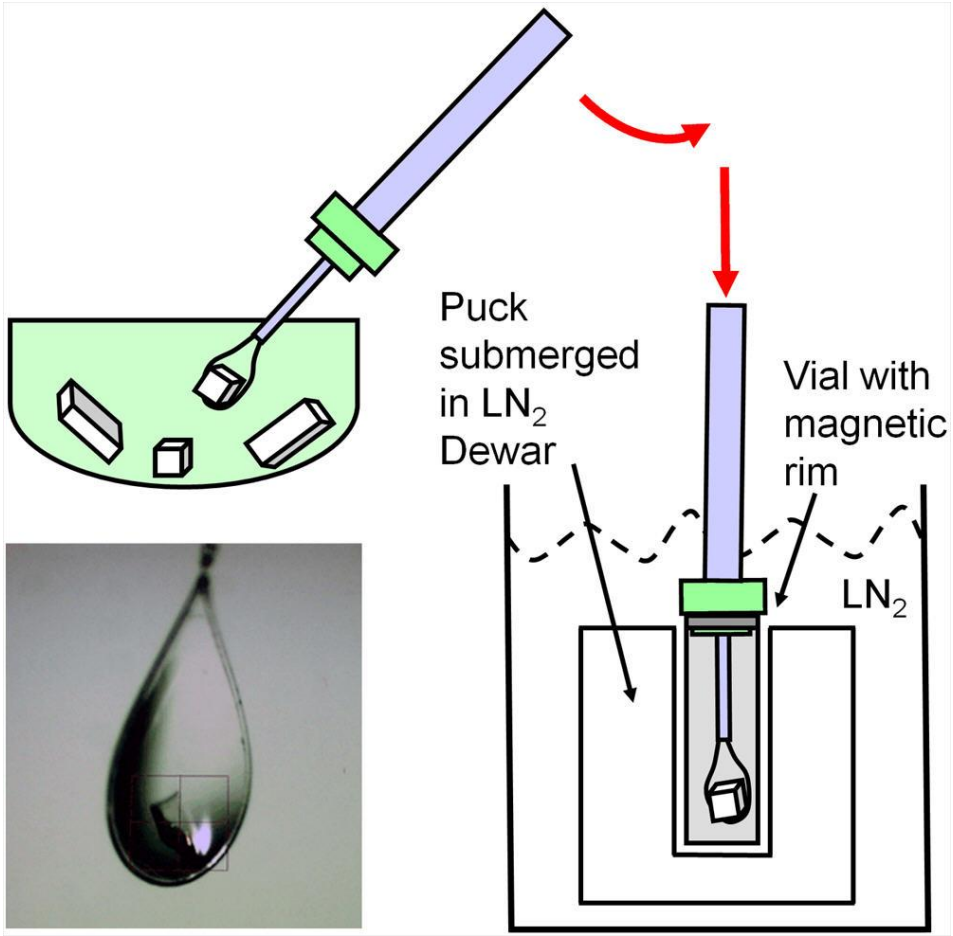
Lab practical tomorrow Thursday 8th of May in lab AI 2142 2nd Floor

Content of lectures

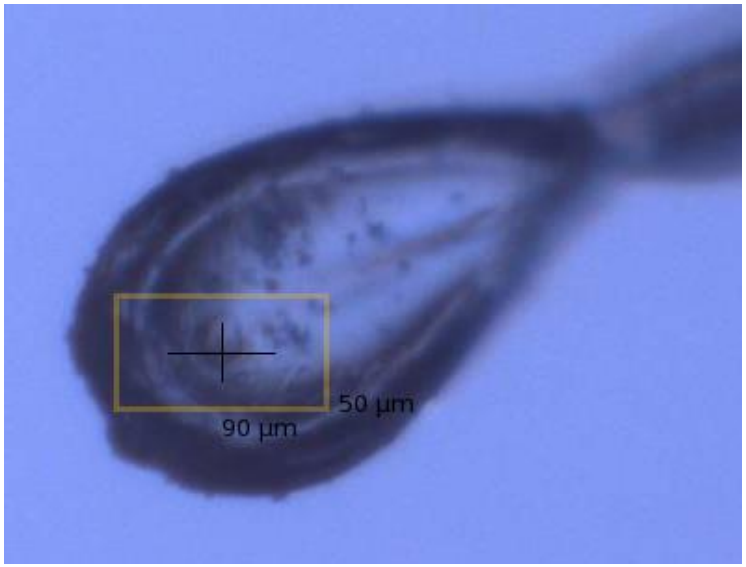
- **Why x-rays and why crystals?**
- **Macromolecular crystallization**
- **Crystal packing**
- **The diffraction experiment**
- The phase problem
- Molecular replacement
- Refinement and validation

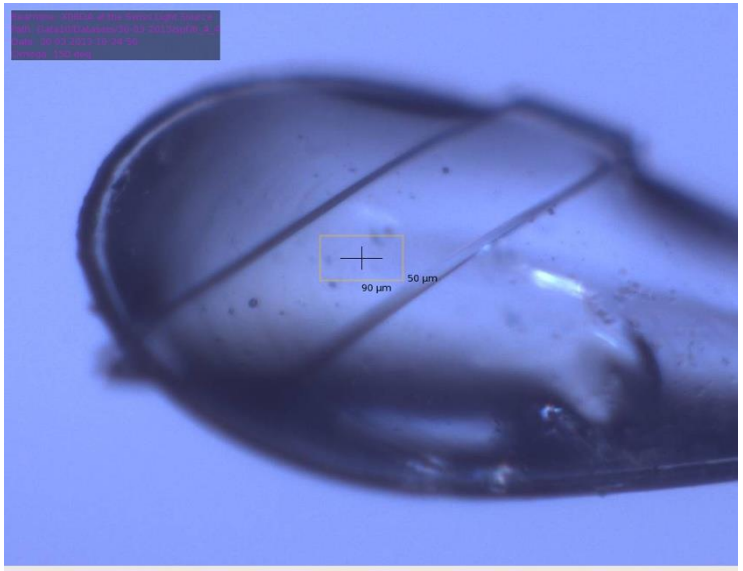
Freezing crystals

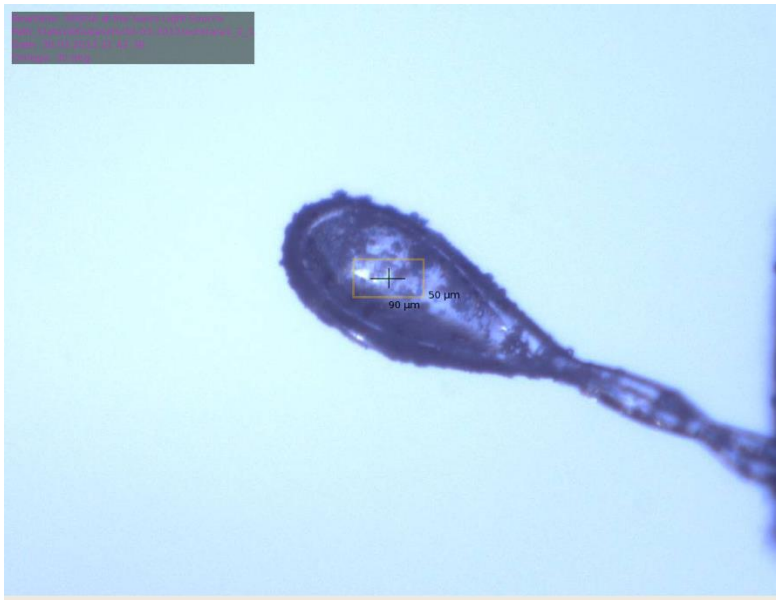
- Need to cool crystal at 100K while it is being irradiated with the X rays to avoid radiation damage.
- Care must be taken in the procedure of freezing the crystal to avoid crystal cracking or formation of ice crystals.
- Dip crystal in cryoprotectant solution (containing for example glycerol, low MW PEG's) which freezes without forming ice.



Examples







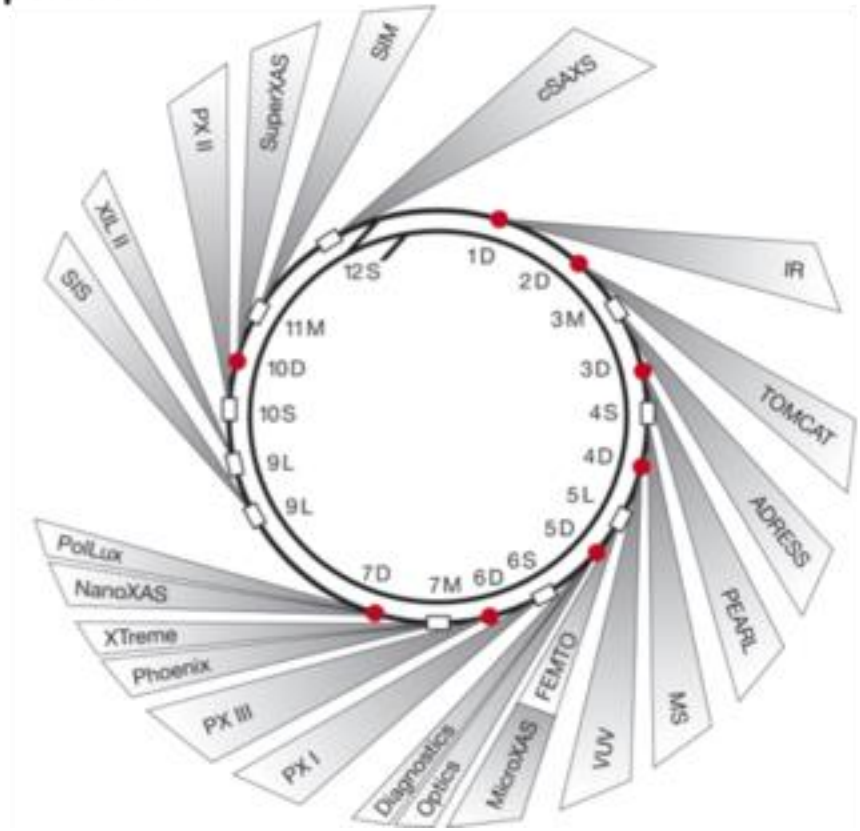
Synchrotrons

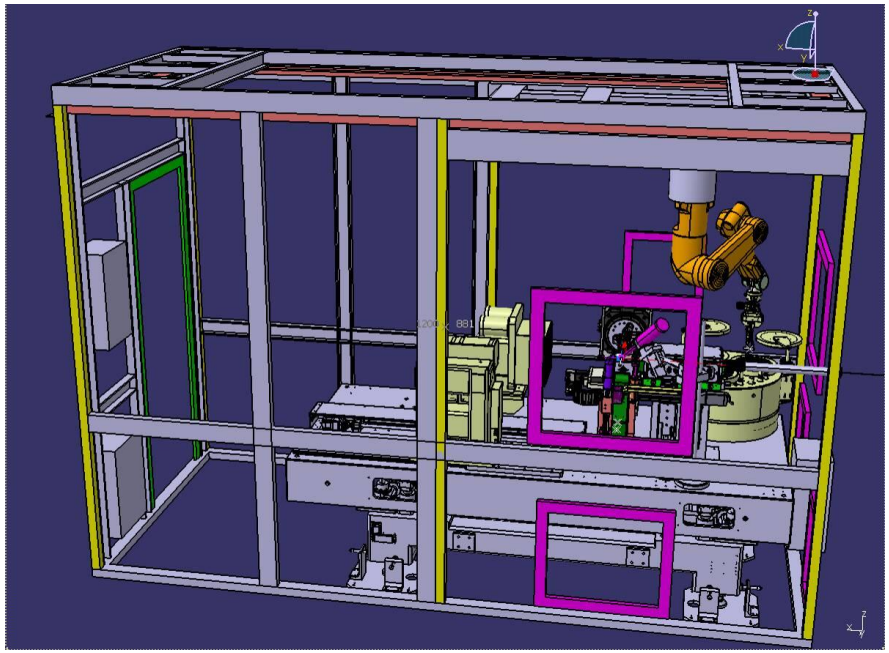
ESRF (Grenoble, FR)

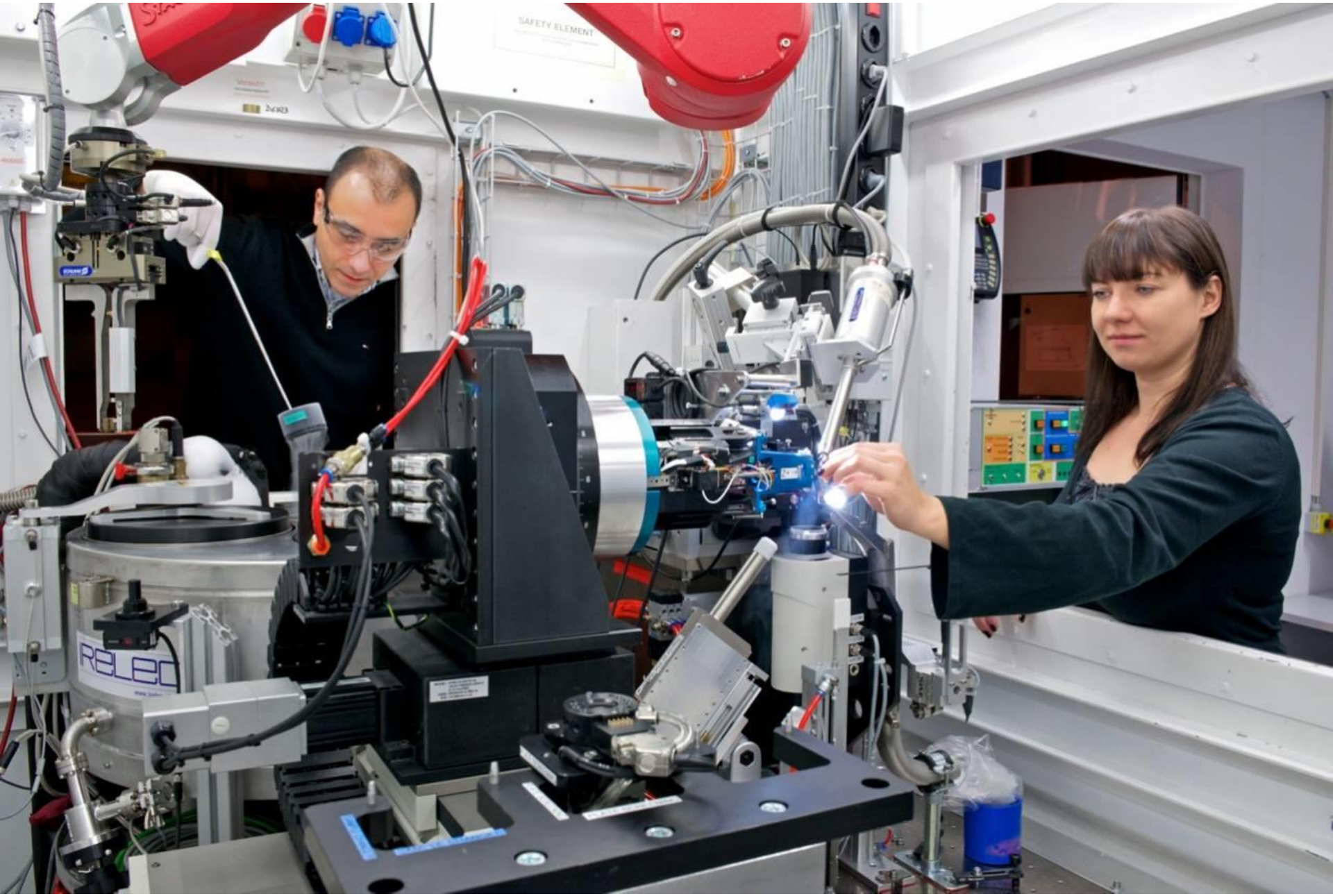
SLS (Villigen, CH)

SLS

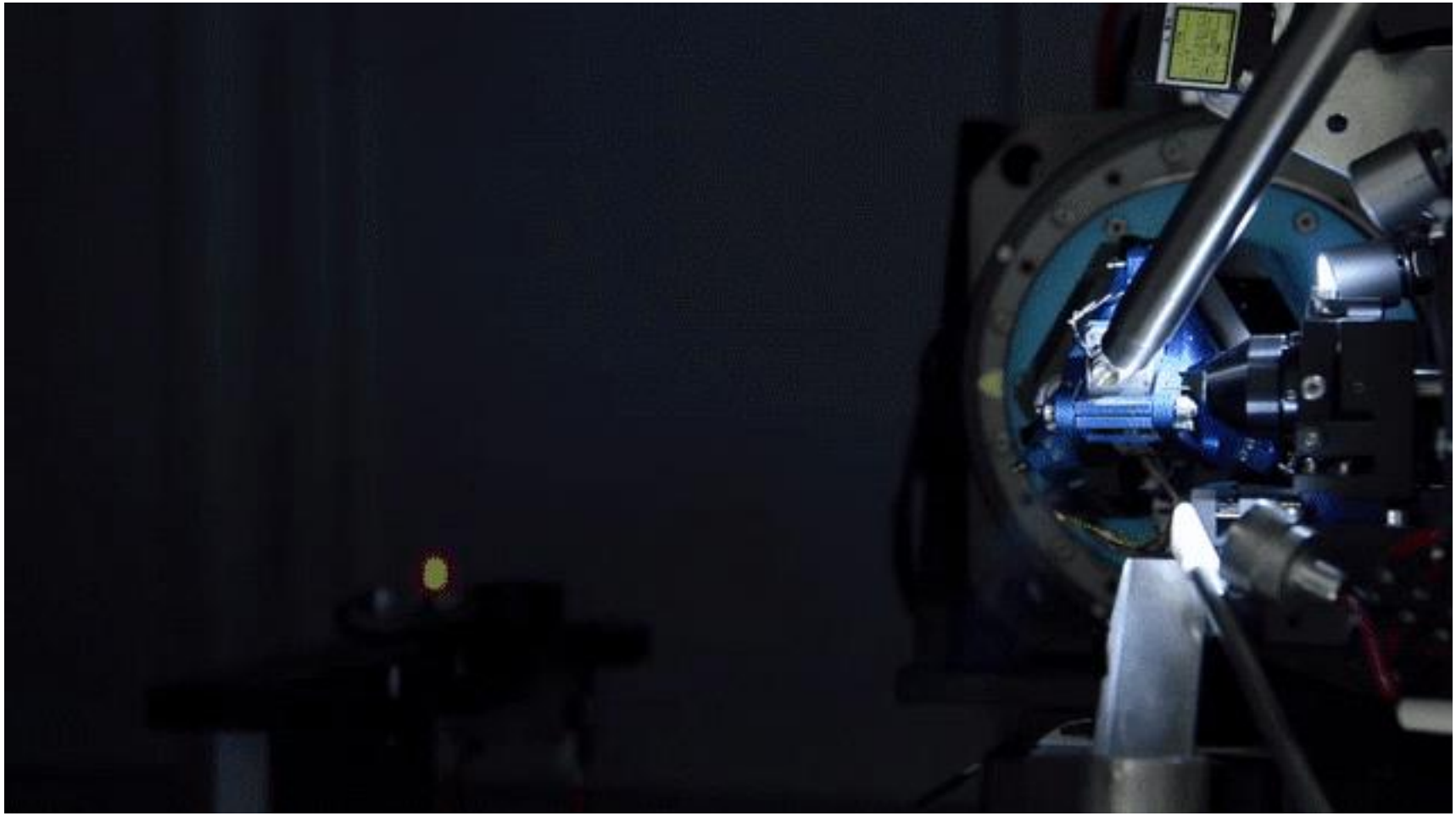
Beamline Map: 16 beamlines are in user operation.



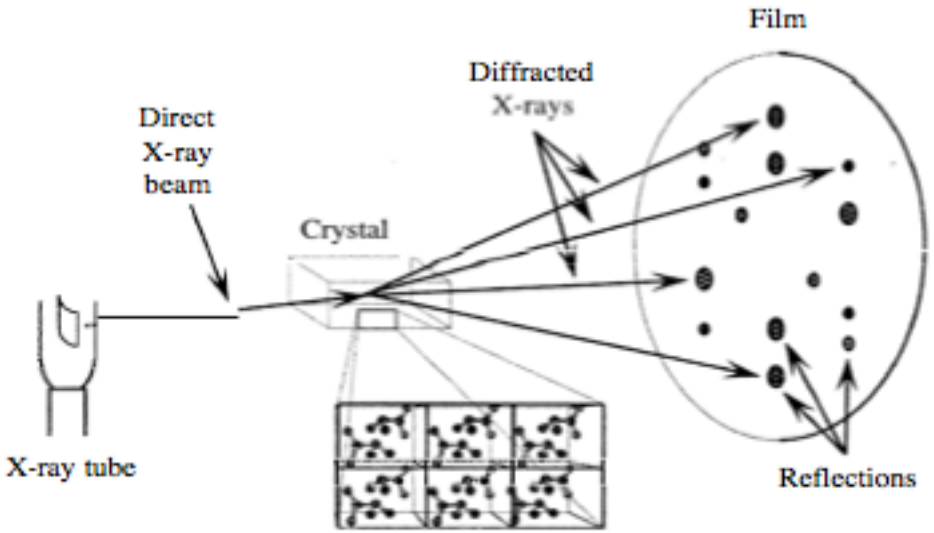



SAFETY ELEMENT

SLS-PXIII



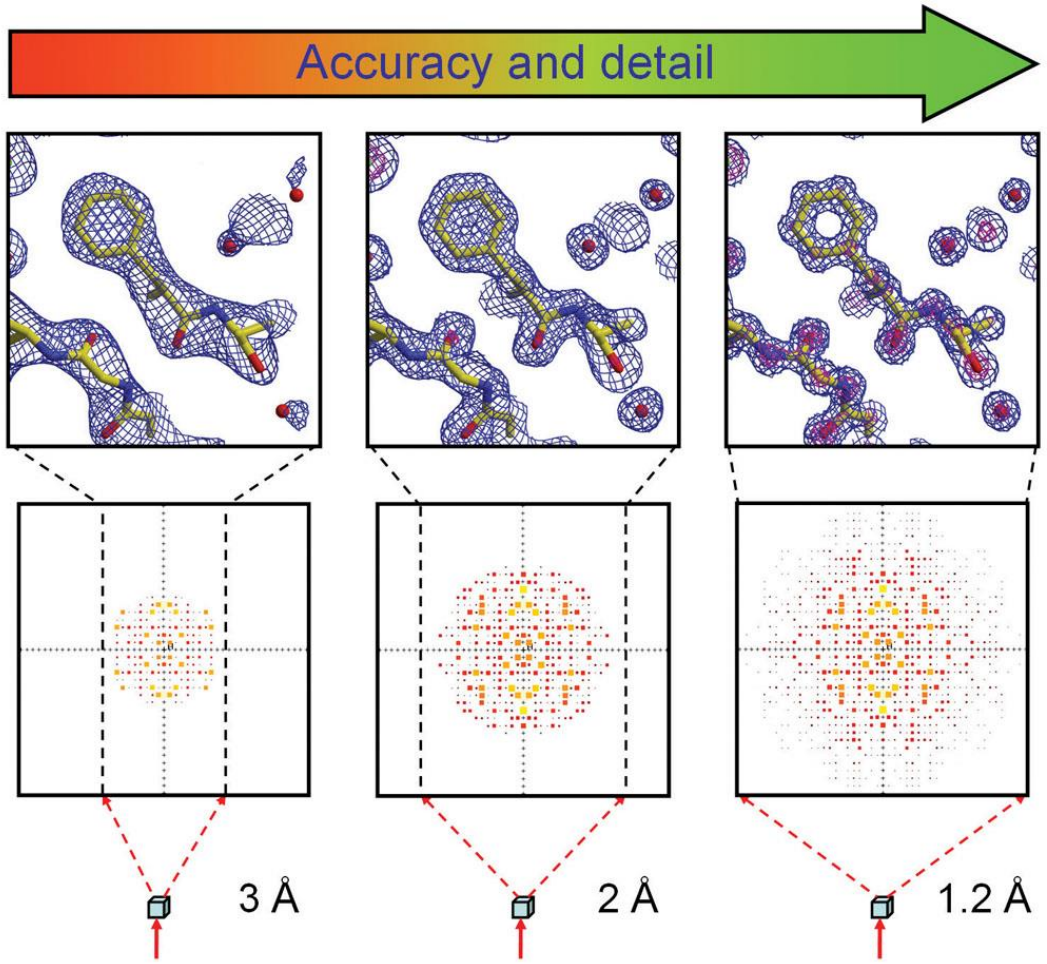
The diffraction experiment– data collection



Rhodes, G. (2006)

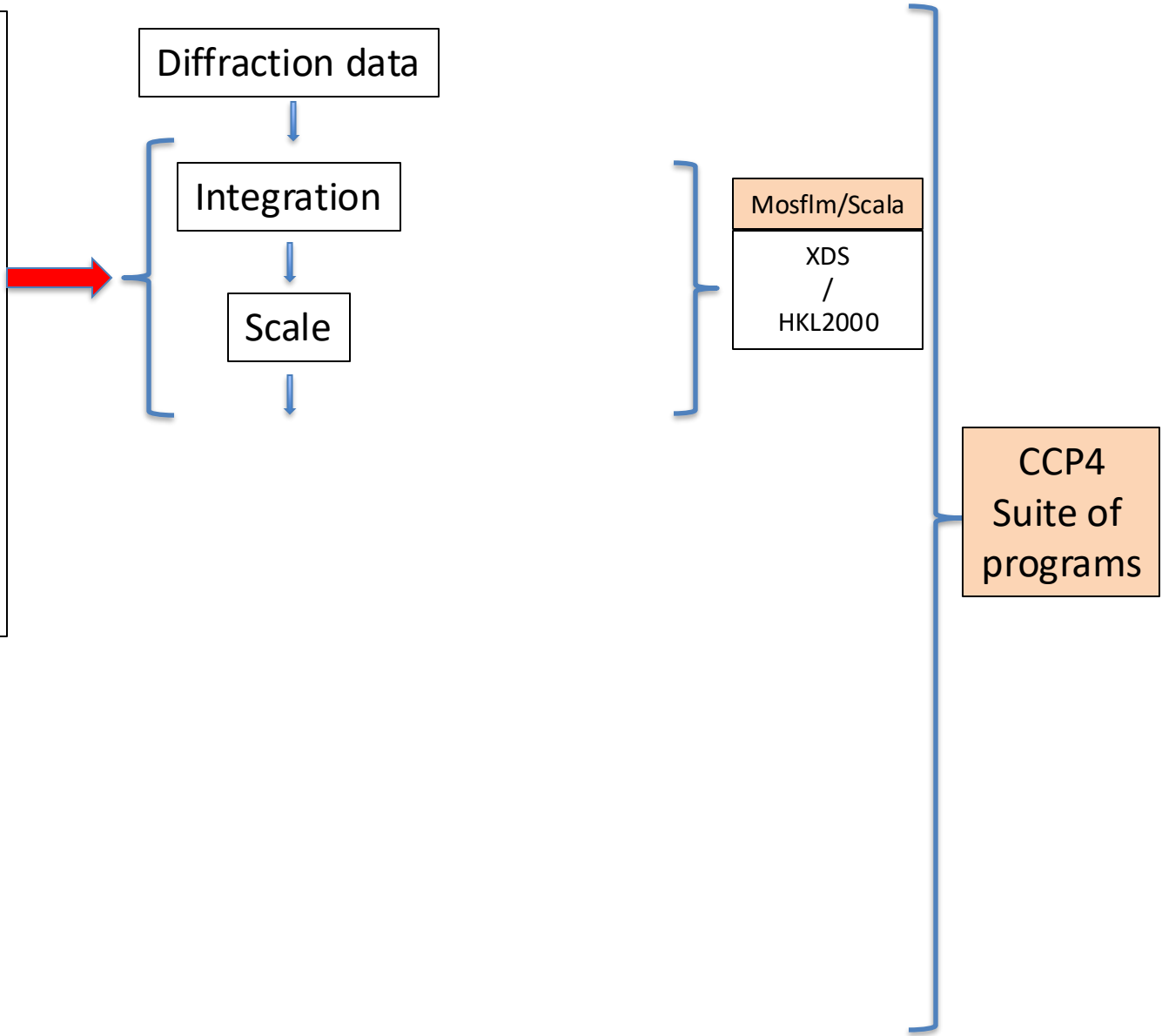
Crystal rotates around itself in order to collect all possible reflections – rotate the Ewald Sphere- collect a full data set
The rotation range depends on the symmetry of the crystal – higher symmetry less rotation range necessary

X-ray diffraction - resolution



Data processing and statistics

-Calculate cell parameters
-Refine cell parameters
-Calculate space group;
-Scale intensities and errors;
-Transform intensities to structure factor amplitudes;
-OUTPUT file: list of measured reflections with their intensities and their errors. 'Mtz file'



Data processing and statistics

X-ray source & Detector
Wavelength (Å)
Space group
Cell (Å °)
Resolution range (Å)
Number of observations
Unique reflections
Completeness
Rmerge
Mean I/σ(I)

Statistics for data set and high resolution shell

Completeness should be around 95% and most of the time 100%
Rmerge should be ≤10%
Mean I/σ(I) (signal to noise ratio) should be ≥2

$$R_{merge} = \frac{\sum_{hkl}\sum_{i=1}^{N}\left| I_{(hkl)i} - \bar{I}_{(hkl)} \right|}{\sum_{hkl}\sum_{i=1}^{N} I_{(hkl)I}}$$

Data processing and statistics

	Aurora A	**SPF**
X-ray source & Detector	SLS PXIII Pilatus	SLS PXIII Pilatus
Wavelength (Å)	1	1
Space group	$P6_122$	$P2_1$
Cell (Å °)	82.3 82.3 169.4 90 90 120	71.9 73.8 94.4 90 96.3 90
Resolution range (Å)	66-2.9 (3-2.9)	30 -1.6 (1.7-1.6)
Number of observations	92281	401202
Unique reflections	8204	128933
Completeness	100%	99%
Rmerge	12.5% (90%)	5% (30%)
Mean I/σ(I)	14.2 (2.5)	11.5 (2.8)