TMM 3102: Protein Structure, Function and Disease

- Structural Biology Methods: X-ray Crystallography (September 30th, 2021)

Jyh-Yeuan (Eric) Lee, Assistant Professor, BMI

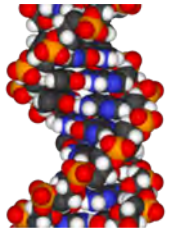
(Partially adopted from former lectures by Dr. John Baenziger)

A major key concept in human physiology (or any organism) is how all the biological matters work in the bodies. From systems (such as gastrointestinal), organs, tissues, to cells, all comes down to operations of biological macromolecules, such as DNA or proteins.

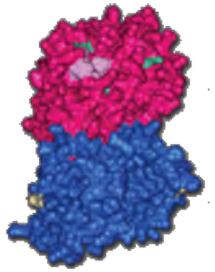
As we discuss the molecular interactions, we are looking at reactions that happen among thousands of atoms that make up individual macromolecules.

This is the spirit of this course. We are looking at how proteins work in our bodies and how they contribute to the physiological functions at "**atomic**" resolution.

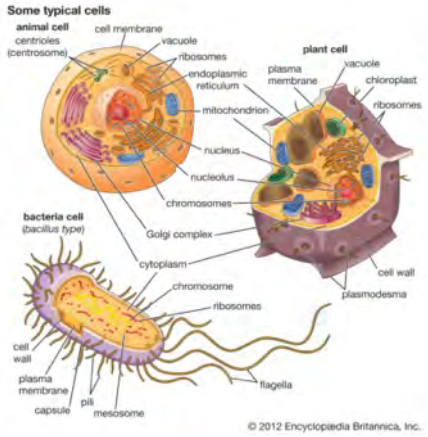
Nucleotides (DNA)

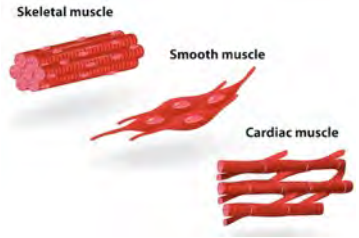


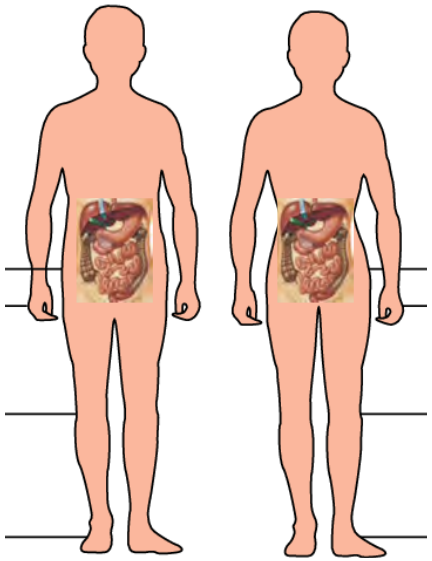
Atoms (C, H, O, N, P)

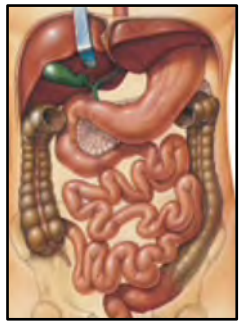


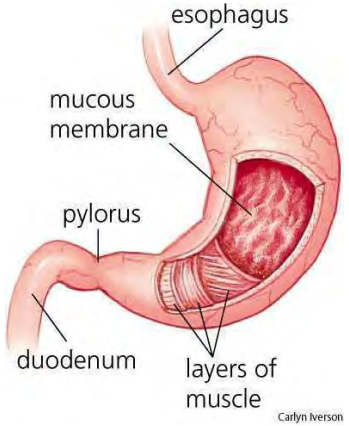
Proteins (tubulin/microtubule)











Structural Biology

- Understanding biology by examining three dimensional (3-D) molecular architectures and their changes.

- Learning life in action with the eyes of atoms: chemical and physical properties of biological matters.

- Structures of biological molecules determine their functions.

Central dogma:

Sequence ➜ Structure ➜ Function

Making Structural Biology Possible

1915: X-ray Crystallography

The Nobel Prize in Physics 1915

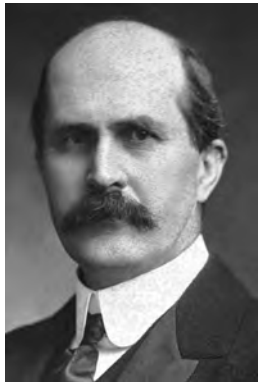


Photo from the Nobel Foundation archive.

Sir William Henry Bragg

Prize share: 1/2

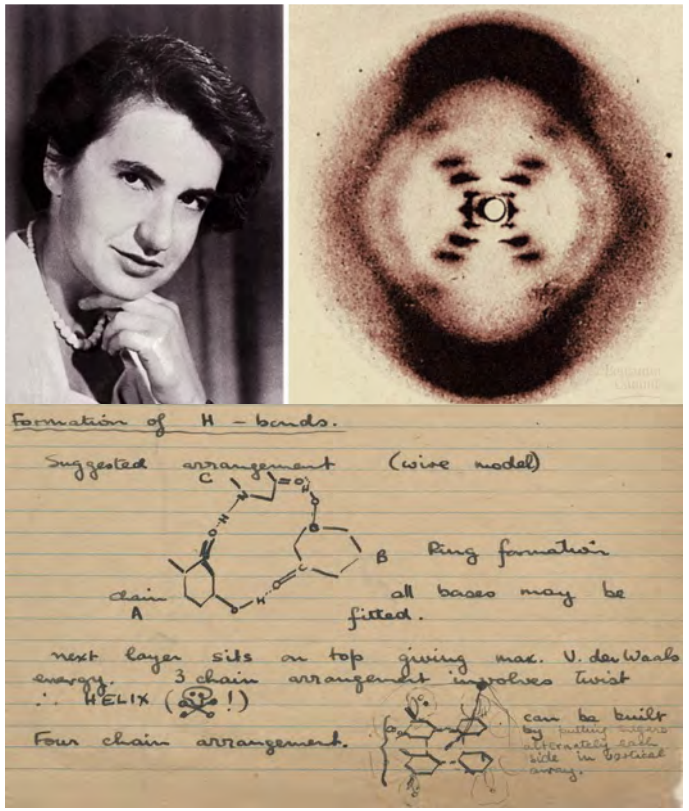
Photo from the Nobel Foundation archive.

William Lawrence Bragg

Prize share: 1/2

1952: DNA Double-Helix Structure

Rosalind Franklin



(1962: Nobel Prize to Watson/Crick/Wilkins)

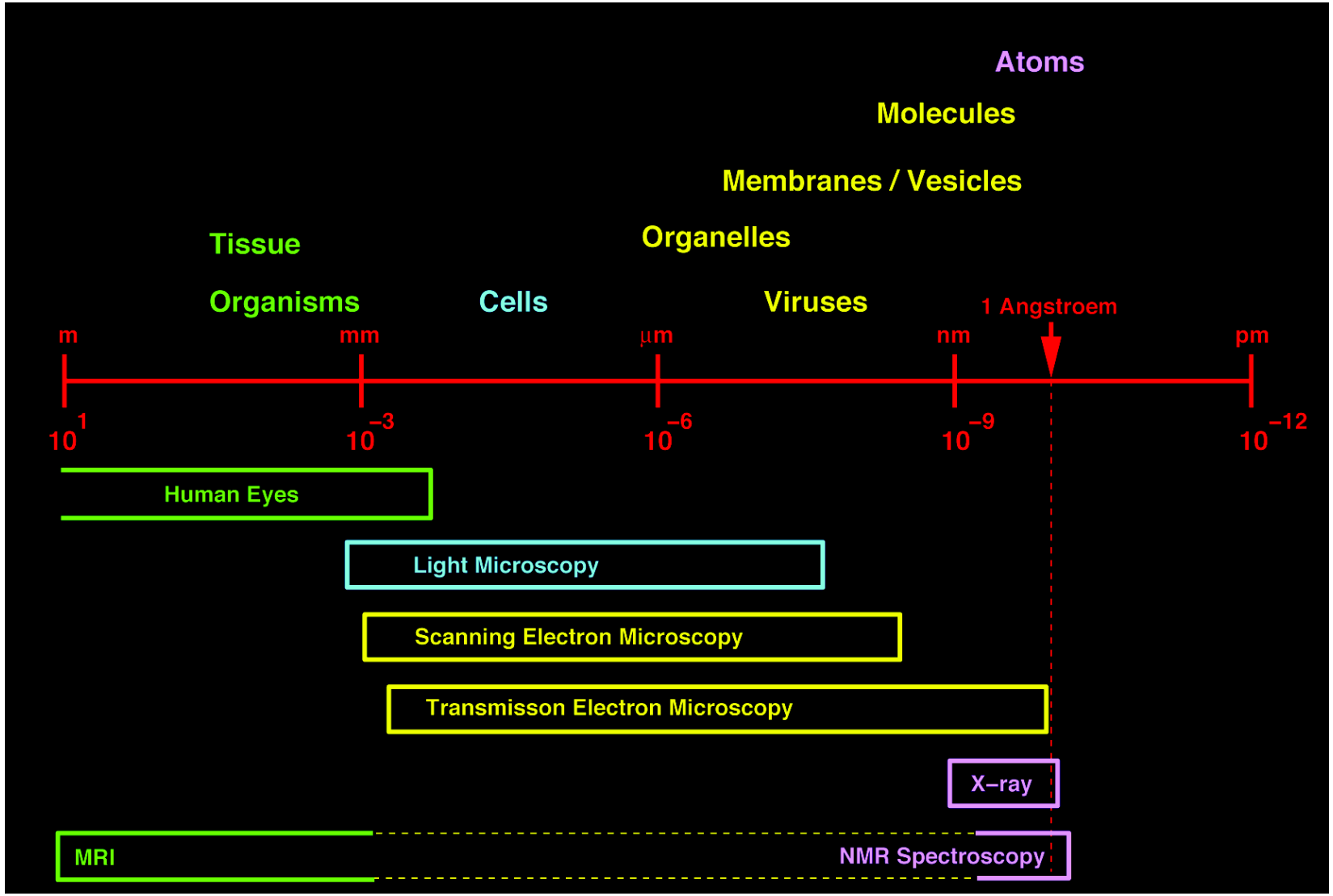
How ”Tiny” Can We See?

From human’s eyes to analytical instruments, we are all limited to how small objects we can see.

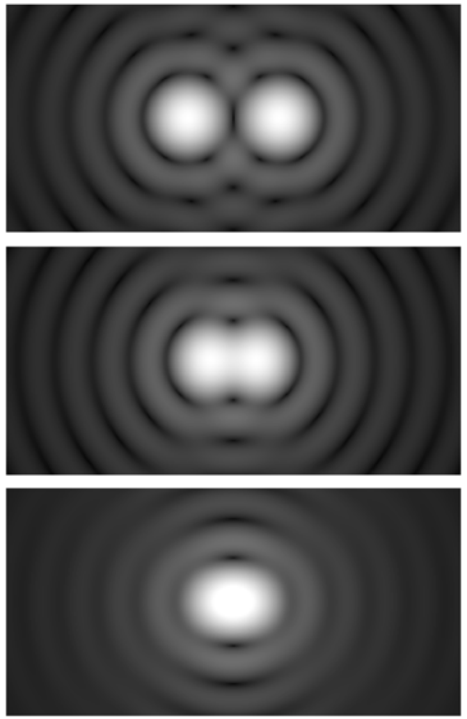
For cells, we can easily observe under a light microscope and with more detailed information using electron microscopes.

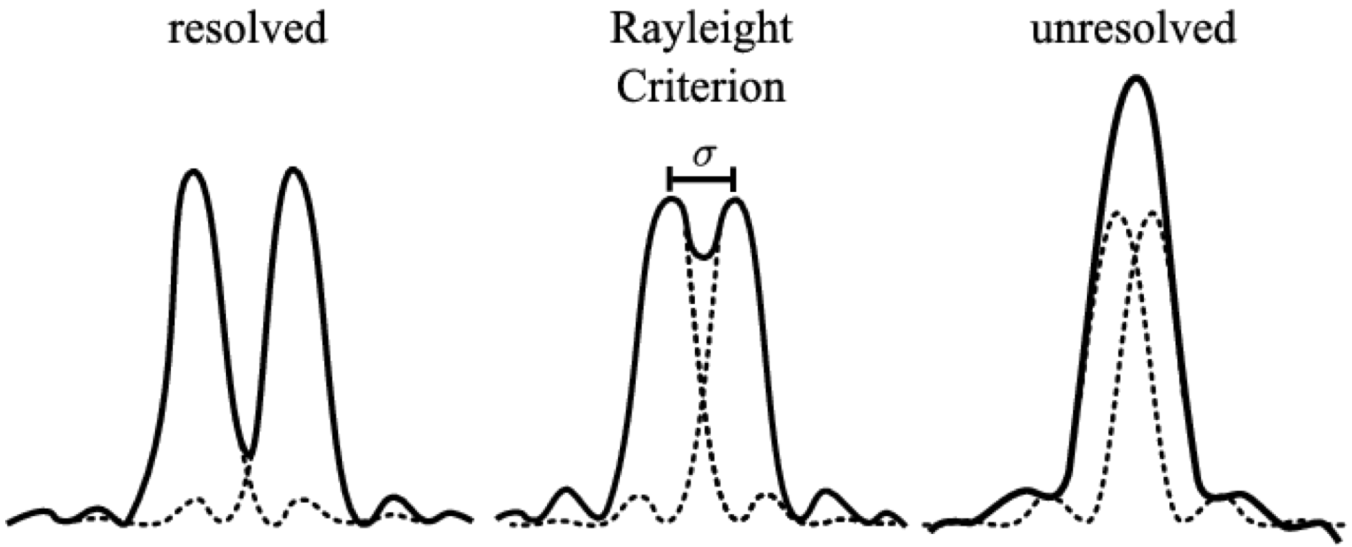
To see objects at atomic resolution, so far, we know X-ray crystallography, transmission electron microscopy, and NMR spectroscopy can enable such high-resolution imaging.

This course will selectively focus on these three methodologies that enable vast protein structure-function studies so far.



Resolution: Rayleigh Criterion

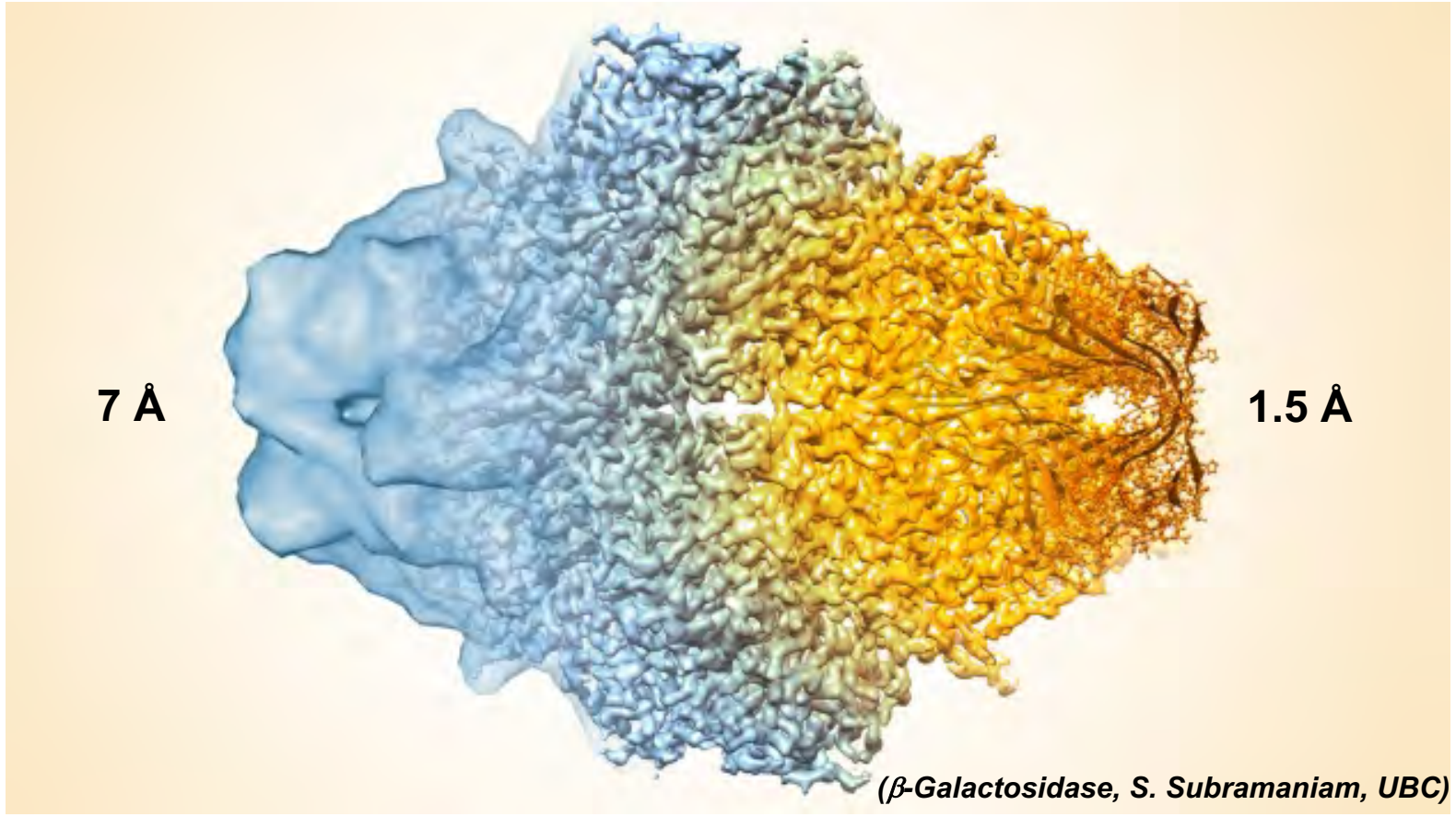




"Two points or two spectral lines of equal intensity are just resolved by an optical instrument when central maximum of the diffraction pattern of one falls on the first minimum of diffraction pattern of the other."

Resolution in the context of protein structure

Resolution is the minimum distance between two points that allows each point to be distinguished as a separate entity”

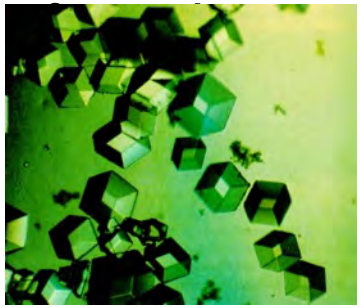


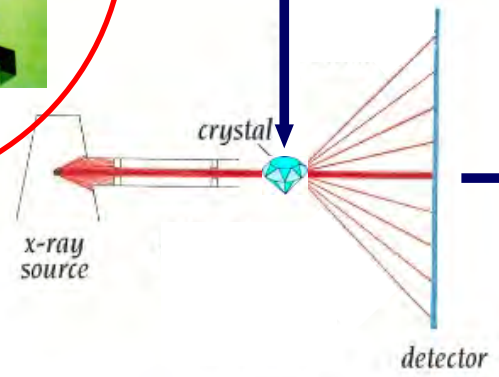
Increasing resolution

X-ray crystallography

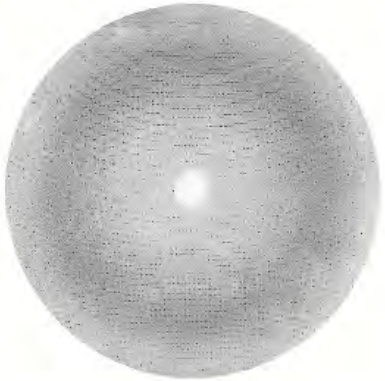
X-ray Crystallography involves 3 main steps:

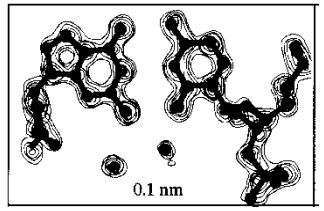
Step 1: protein crystallization



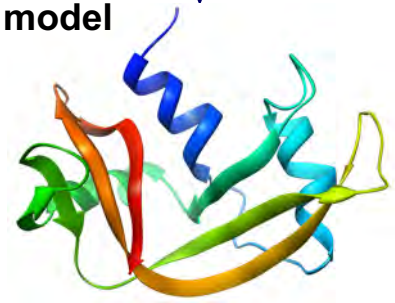


Step 2: collection and interpretation of diffraction data

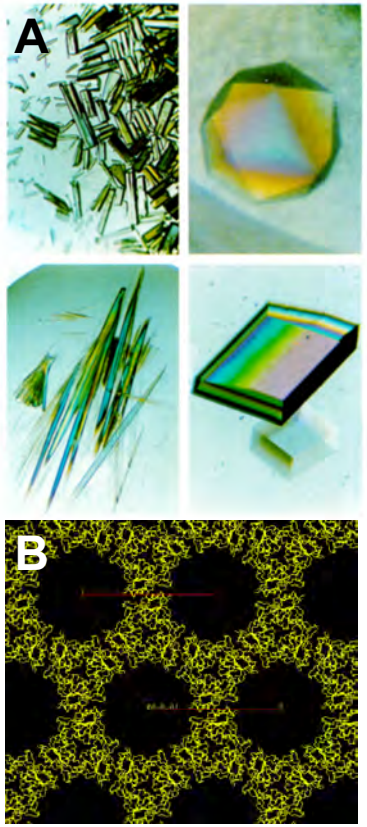




Step 3: build and refine a structural model



Protein crystals



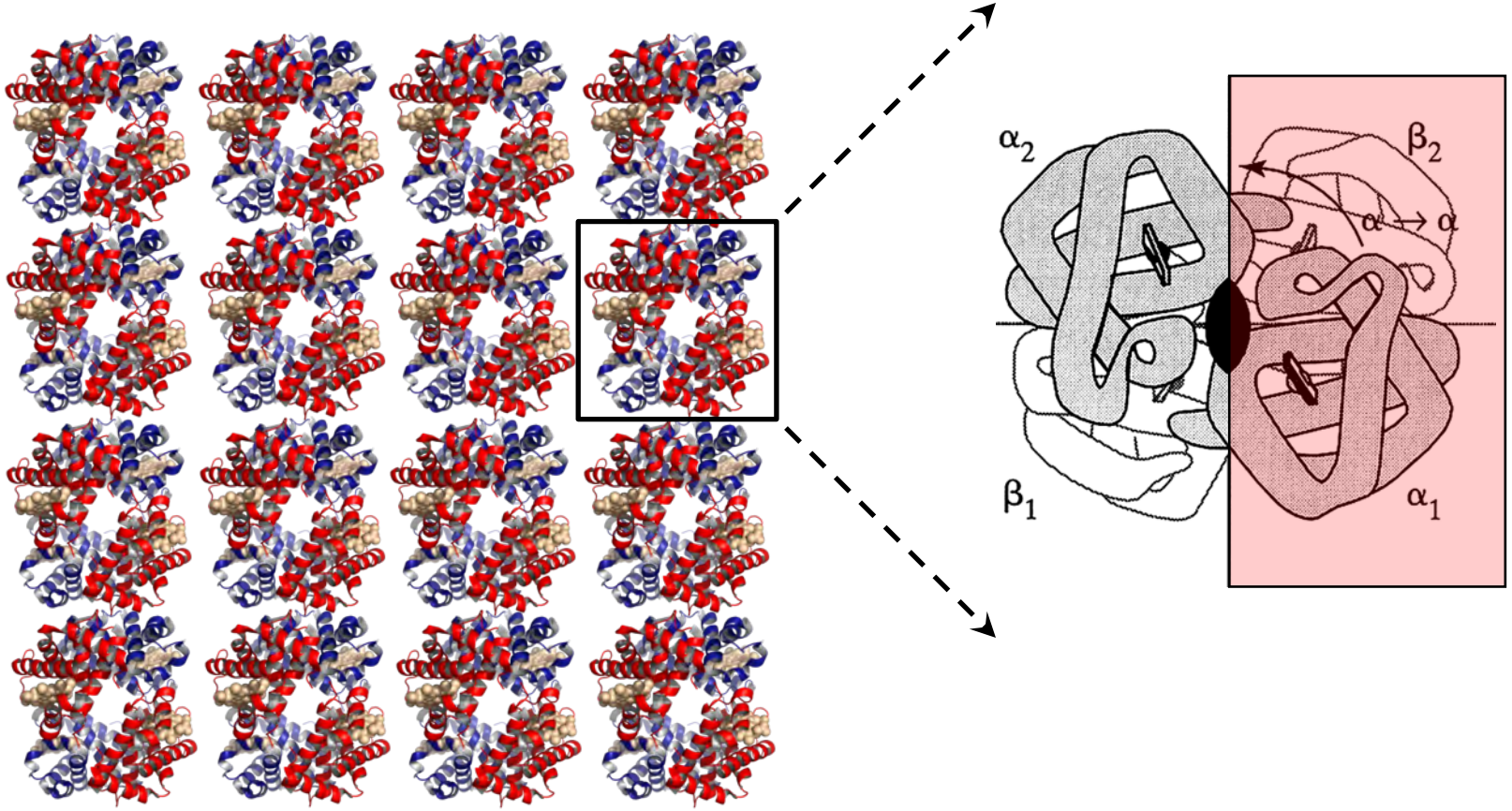
Properties of protein crystals:

1. regular repeating 3-dimensional lattice of protein molecules
2. differ from precipitates by their regular, repeating nature
3. differ from solids in that when broken, they break into smaller replicas of the original crystal
4. interact in a unique fashion with visible light and X-rays
5. vary tremendously in terms of size, shape, and morphology
6. often loosely packed with proteins leaving a lot of empty space for solvent

Protein crystal morphology (how proteins pack in the crystal) dictates how a crystal interacts with X-rays, so we must understand crystal morphology to understand X-ray crystallography.

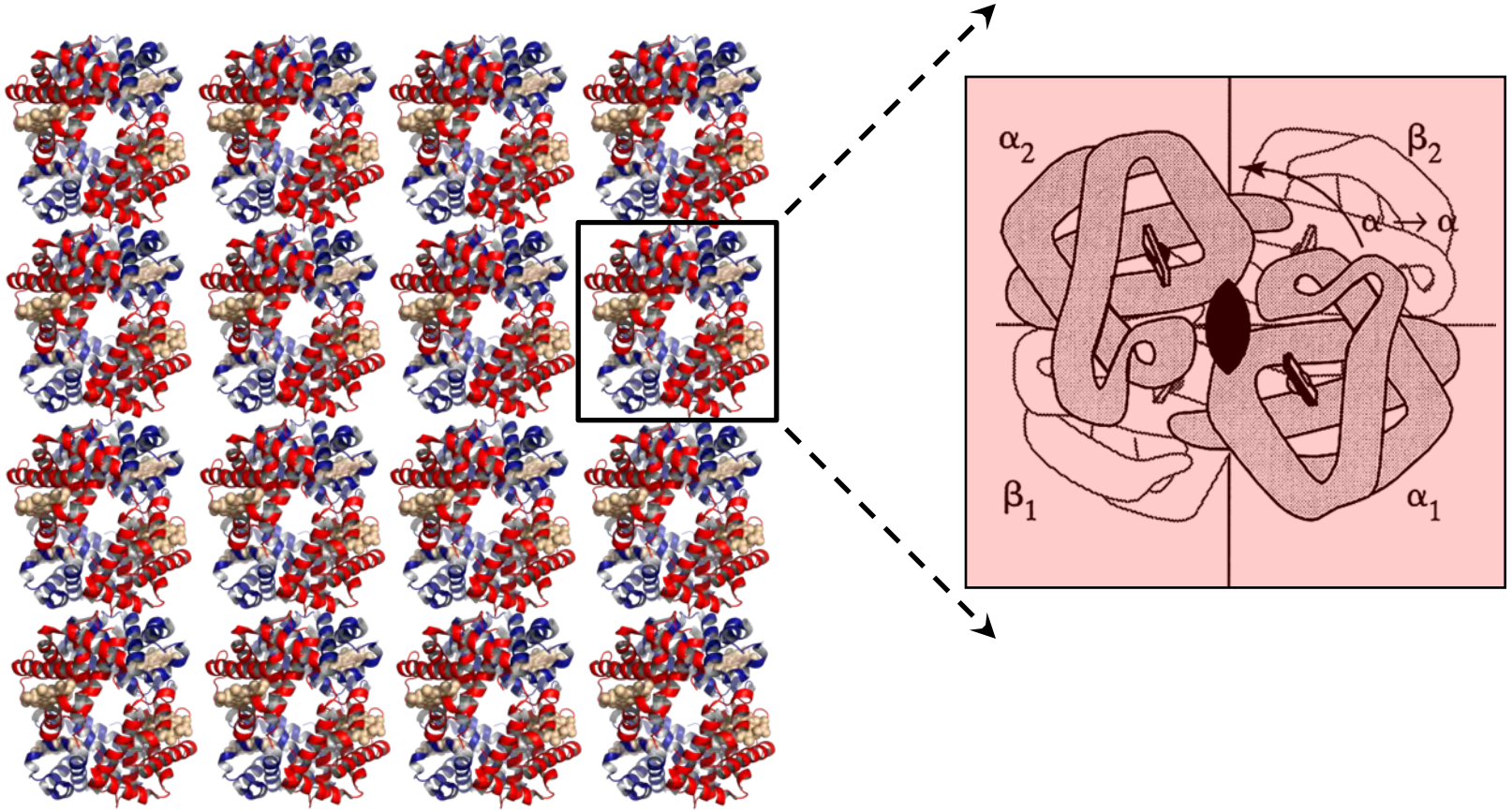
Protein crystal morphology

1. **Asymmetric Unit:** the smallest unit that is repeated in the crystal by rotation and/or translation



Protein crystal morphology

2. Lattice Motif: the smallest unit that is repeated translation alone



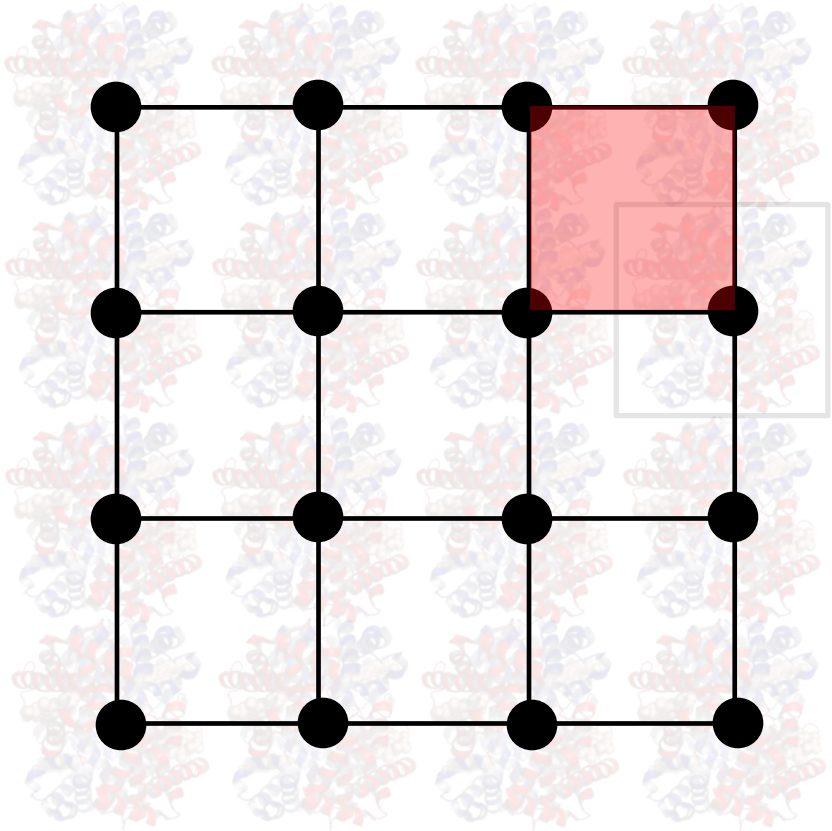
Protein crystal morphology

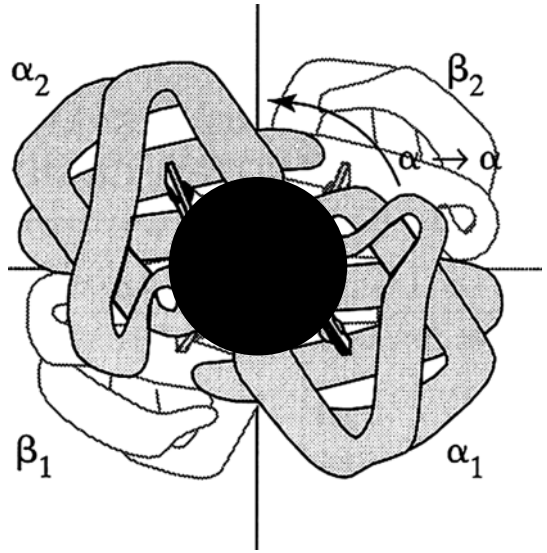
3. Unit Cell:

(i) assign a dot to the center of each lattice motif

(ii) draw lines between adjacent dots

(iii) each box represents a unit cell, but…





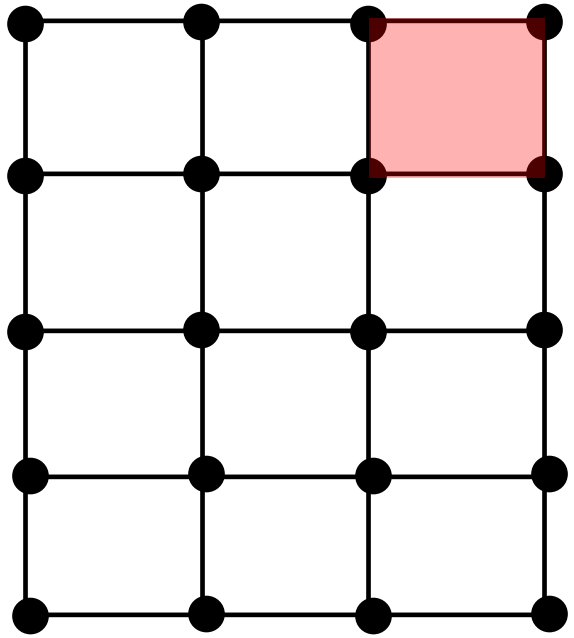
Protein crystal morphology

3. Unit Cell:

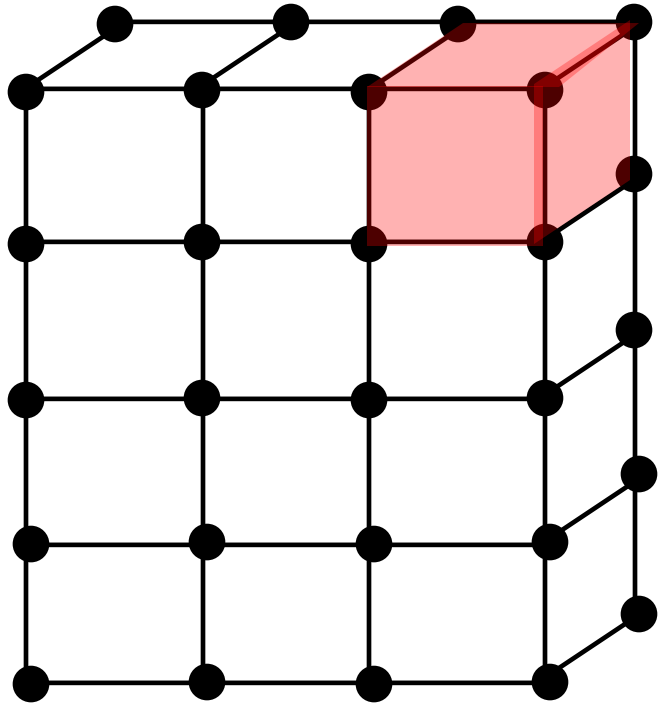
(i) assign a dot to the center of each lattice motif

(ii) draw lines between adjacent dots

(iii) each box represents a unit cell, but a unit cell has three dimensions



2D representation of
a crystal and a unit cell

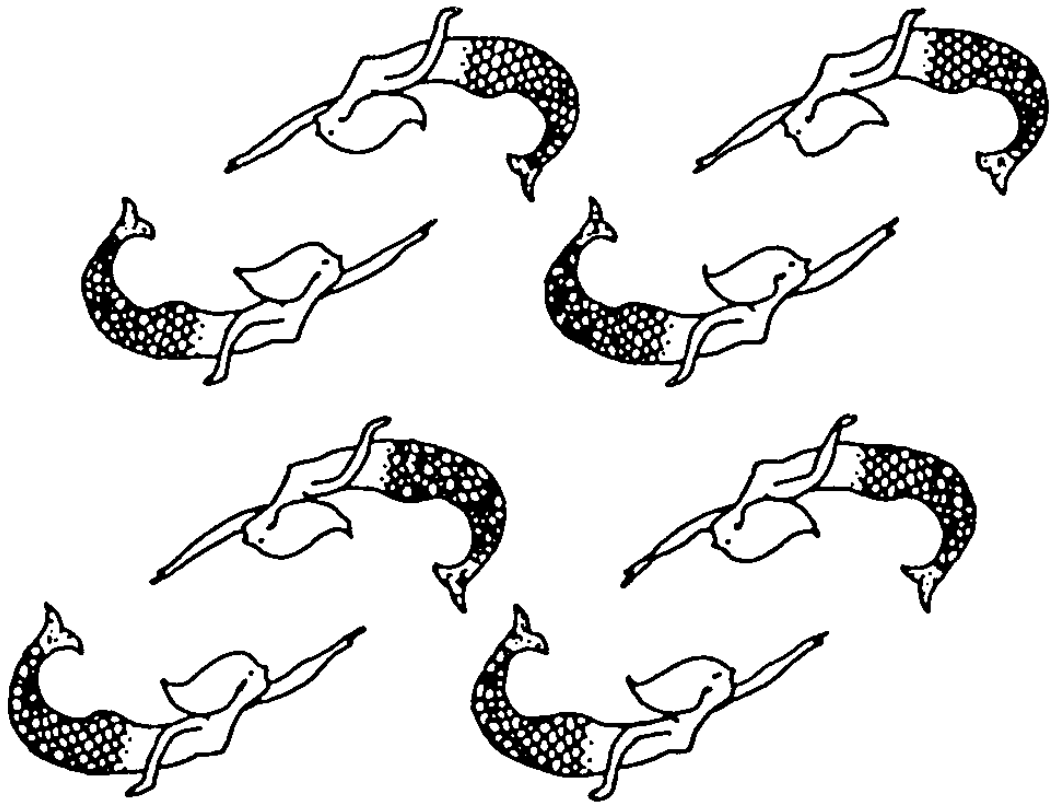


3D representation of a
crystal and a unit cell

Protein crystal morphology

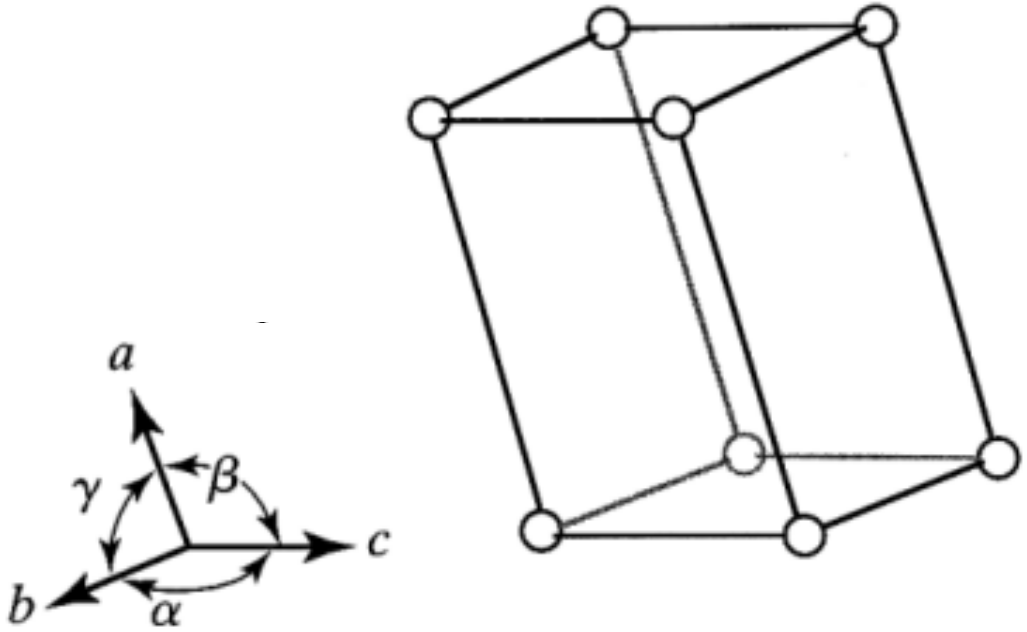
3. Unit Cell:

Can you identify the asymmetric unit, the lattice motif, and the unit cell in this crystal?



Protein crystal morphology

4. **Bravais Lattice:** each unit cell is characterized by:
   1) the **lengths** of its axes [(**x, y, z**) or (**a, b, c**)]
   2) the **angles** between the three axes (α, β, γ)

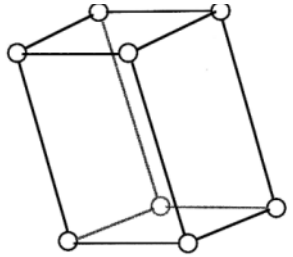


There are 14 different types of unit cells called the *Bravais Lattices*

Bravais Lattices

Triclinic

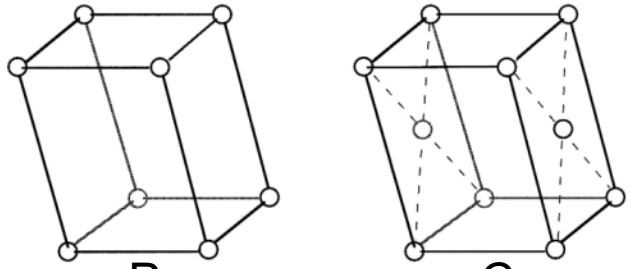
$\alpha \neq \beta \neq \gamma \neq 90°$



P

Monoclinic

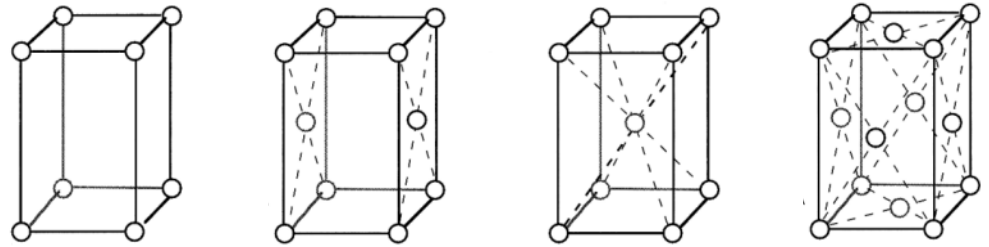
$\alpha = \gamma = 90°, \beta \neq 90°$



P C

Orthorhombic

$\alpha = \beta = \gamma = 90°$



P C I F

Location of Lattice Points

P = primitive

I = body-centered

F = face-centered

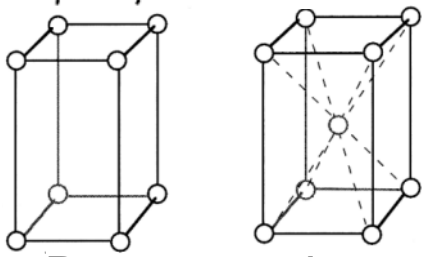
C = base-centered

R = rhombohedral

Bravais Lattices (continued)

Tetragonal

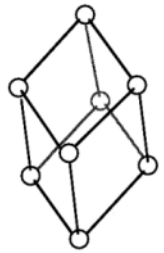
$\alpha = \beta = \gamma = 90°$



P I

Trigonal

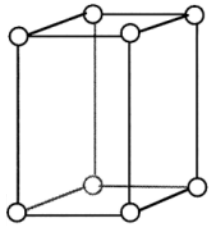
$\alpha = \beta = 90°, \gamma = 120°$



P

Hexagonal

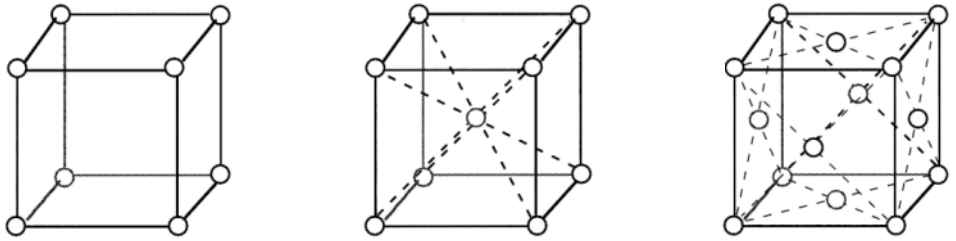
$\alpha = \gamma = 90°, \beta = 120°$



P

Orthorhombic

$\alpha = \beta = \gamma = 90°$



P I F

Location of Lattice Points

P = primitive

I = body-centered

F = face-centered

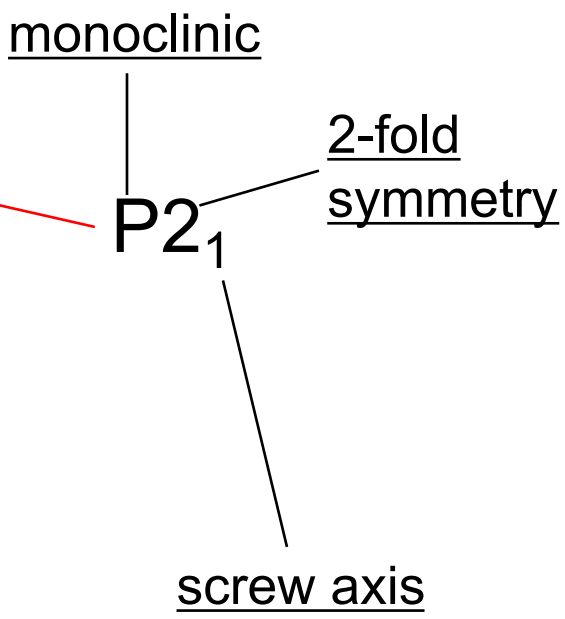
C = base-centered

R = rhombohedral

Protein crystal morphology

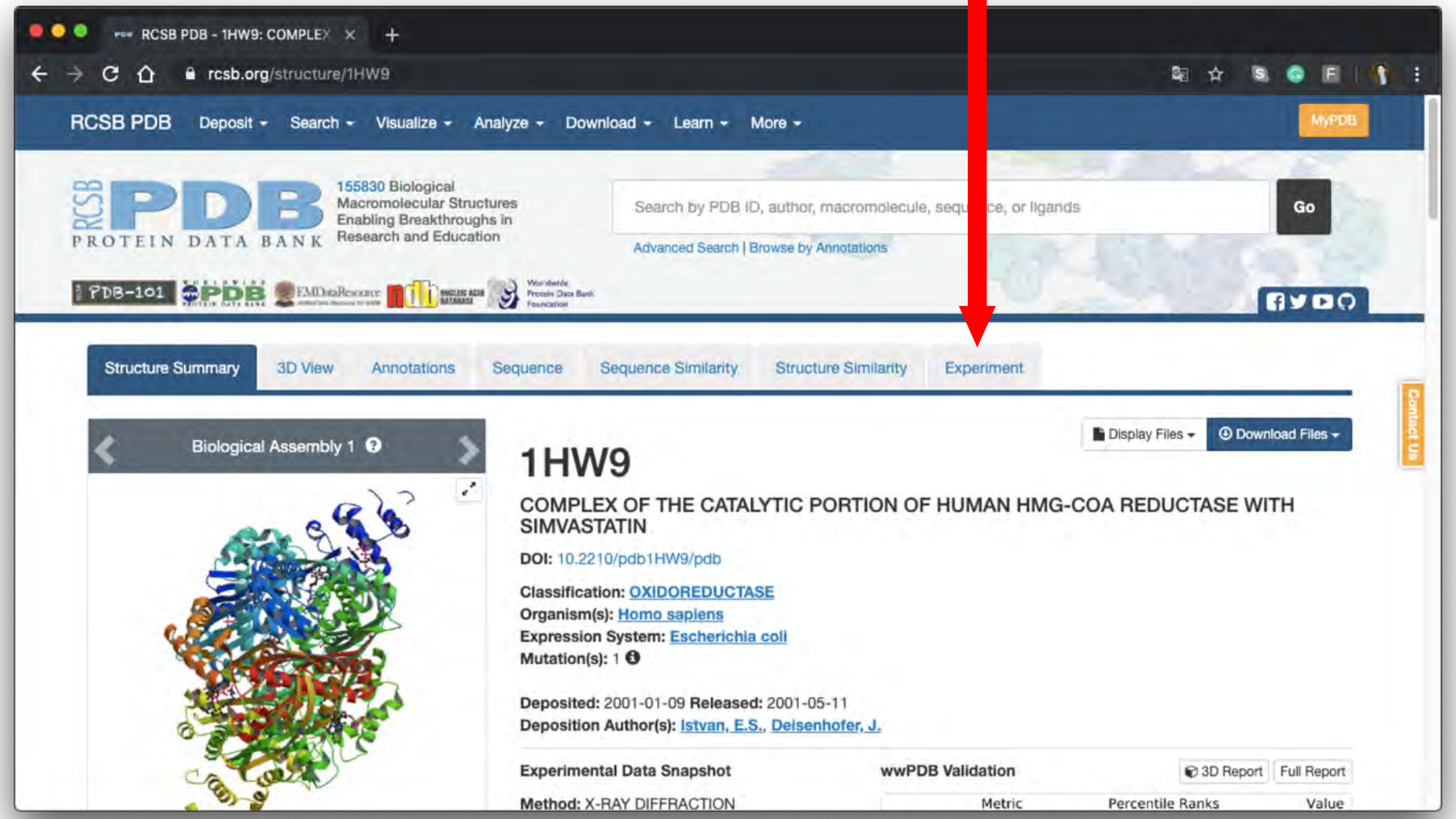
5. Space groups: for asymmetric molecules, such as proteins, there are 65 ways to arrange them in the 14 Bravais Lattices – each is called a space group

Lattice Type	Possible Bravais Lattices	Crystal Shape	Possible Space Groups
Triclinic	P	$a \neq b \neq c$ $\alpha \neq \beta \neq \gamma \neq 90°$	$P1$
Monoclinic	P, C	$a \neq b \neq c$ $\alpha = \gamma = 90°, \beta \neq 90°$	$P2$, $P2_1$, $C2$
Orthorhombic	P, C, I, F	$a \neq b \neq c$ $\alpha = \beta = \gamma = 90°$	$P222$, $P2_12_12_1$, $P2_12_12$, $P222_1$, $C222$, $C222_1$, $F222$, $I222$, $I2_12_12_1$
Tetragonal	P, I	$a = b \neq c$ $\alpha = \beta = \gamma = 90°$	$P4$, $P4_1$, $P4_2$, $P4_3$, $I4$, $I4_1$, $P422$, $P42_12_1$, $P4_122$, $P4_12_12$, $P4_222$, $P4_22_12$, $P4_32_12$, $P4_322$, $I422$, $I4_122$
Trigonal	P	$a = b \neq c$ $\alpha = \beta = 90°, \gamma = 120°$	$P3$, $P3_1$, $P3_2$, $P321$, $P312$, $P3_112$, $P3_121$, $P3_212$, $P3_221$,
	R (Rhombohedral)	$a = b = c$ $\alpha = \beta = \gamma < 120° (\neq 90°)$	$R3$, $R32$
Hexagonal	P	$a = c \neq b$ $\alpha = \gamma = 90°, \beta = 120°$	$P6$, $P6_1$, $P6_2$, $P6_3$, $P6_4$, $P6_5$, $P622$, $P6_122$, $P6_322$, $P6_322$, $P6_422$, $P6_522$
Cubic	P, I, F	$a = b = c$ $\alpha = \beta = \gamma = 90°$	$P432$, $P4_132$, $P4_232$, $P4_332$, $F432$, $F4_132$, $I432$, $I4_132$



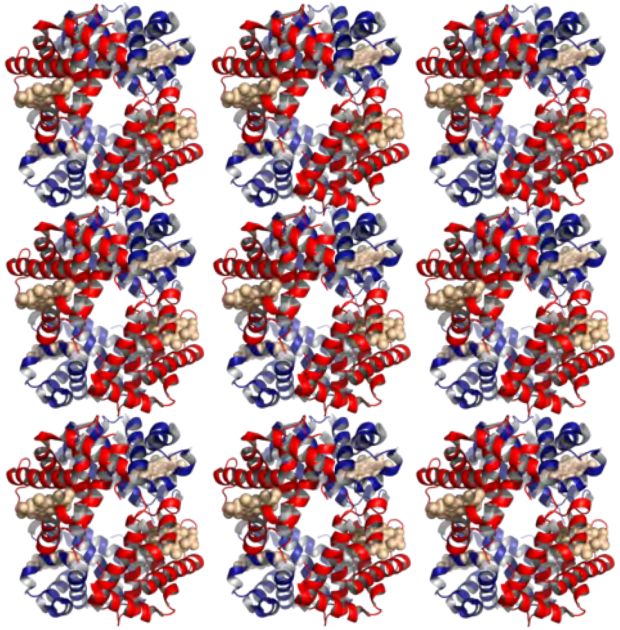
Protein crystal growth: idea

click here

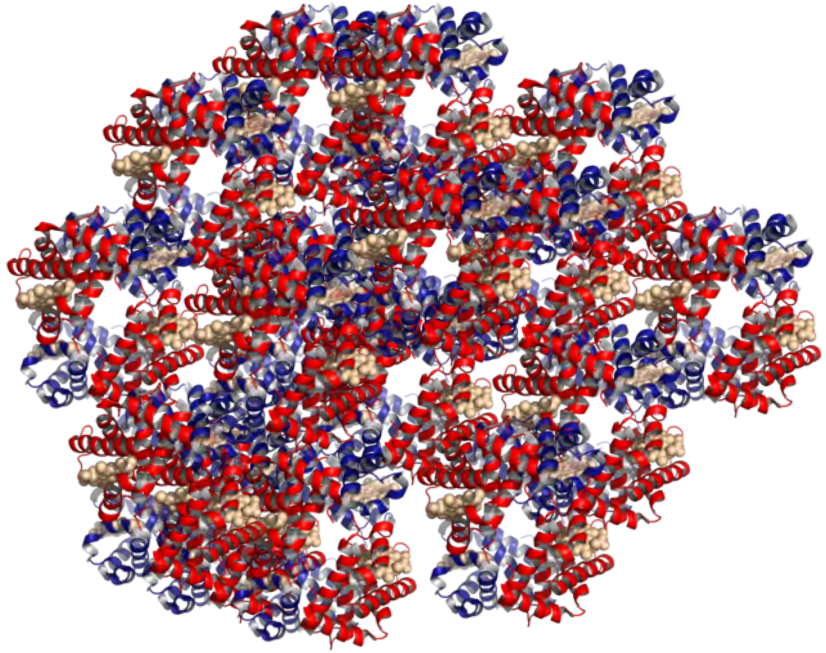


Protein crystal growth: idea

Question: how do you get a protein crystal?
Answer: Basically, you precipitate the protein out of solution, but…



protein crystal



amorphous precipitate

Protein crystal growth: idea

Crystallization is a slow controlled precipitation:

1) Select a target
2) Express and purify large amounts of protein target (>95% pure)
3) Prepare a concentrated solution of pure protein target
4) Slowly increase the concentration of a protein in the presence of a precipitant until a supersaturated solution is obtained
5) Allow protein to crystallize over a period of hours, days, weeks or months

Proteins will hopefully nucleate and then a crystal will hopefully grow.

(Protein crystallization is **Science and Art**.)

Protein crystal growth: idea

Phase diagram tells us what conditions we would expect during protein crystallization.

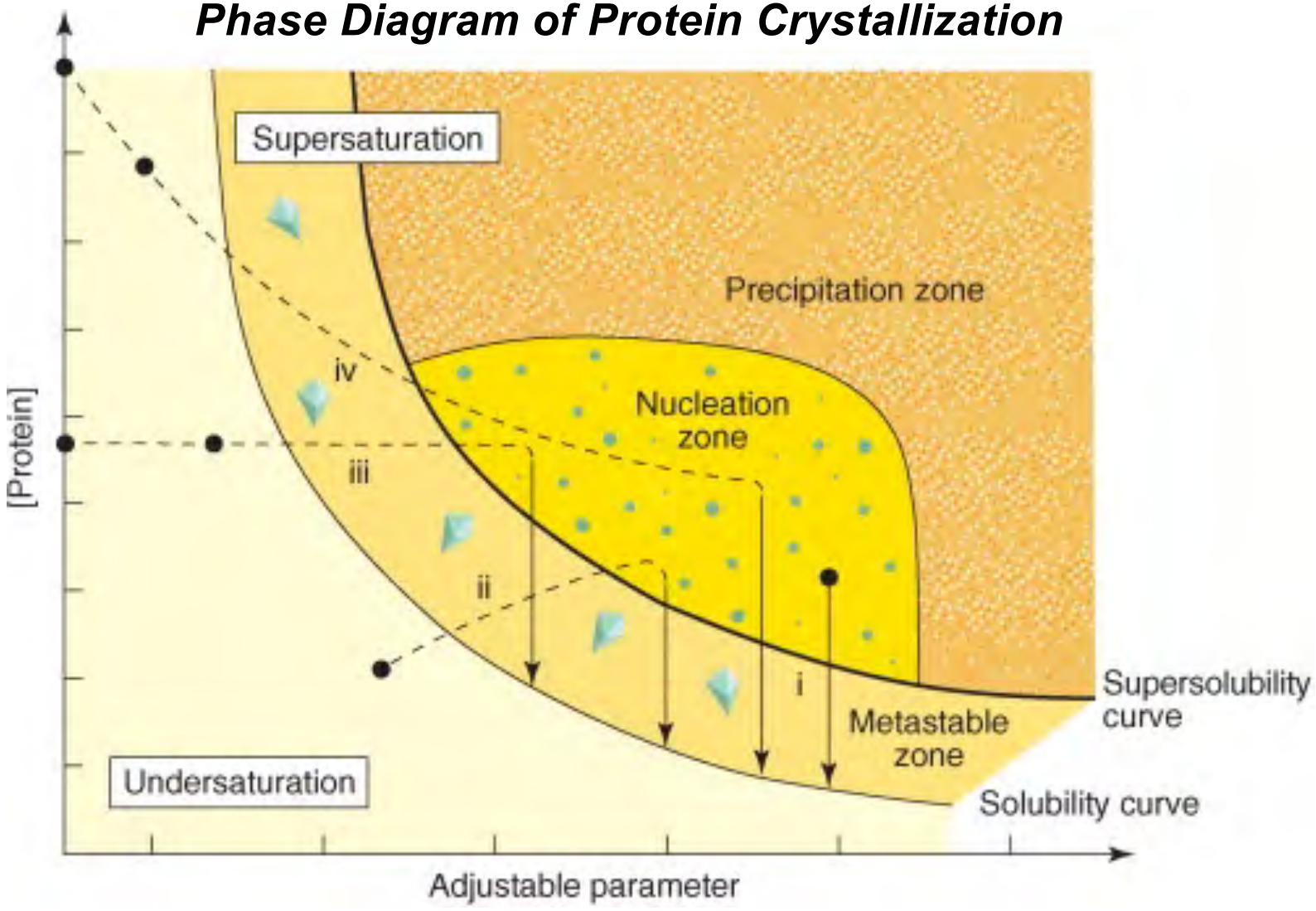
Nucleation zone:
Protein aggregation starts showing signs of becoming crystalline. Sometimes it's called phase separation.

Metastable zone:
Mature protein crystals are being observed.

Precipitation zone:
Irregular protein aggregation becomes irreversible and makes proteins denatured or misfolded.

Undersaturation:
Crystallization never happens.

Phase Diagram of Protein Crystallization



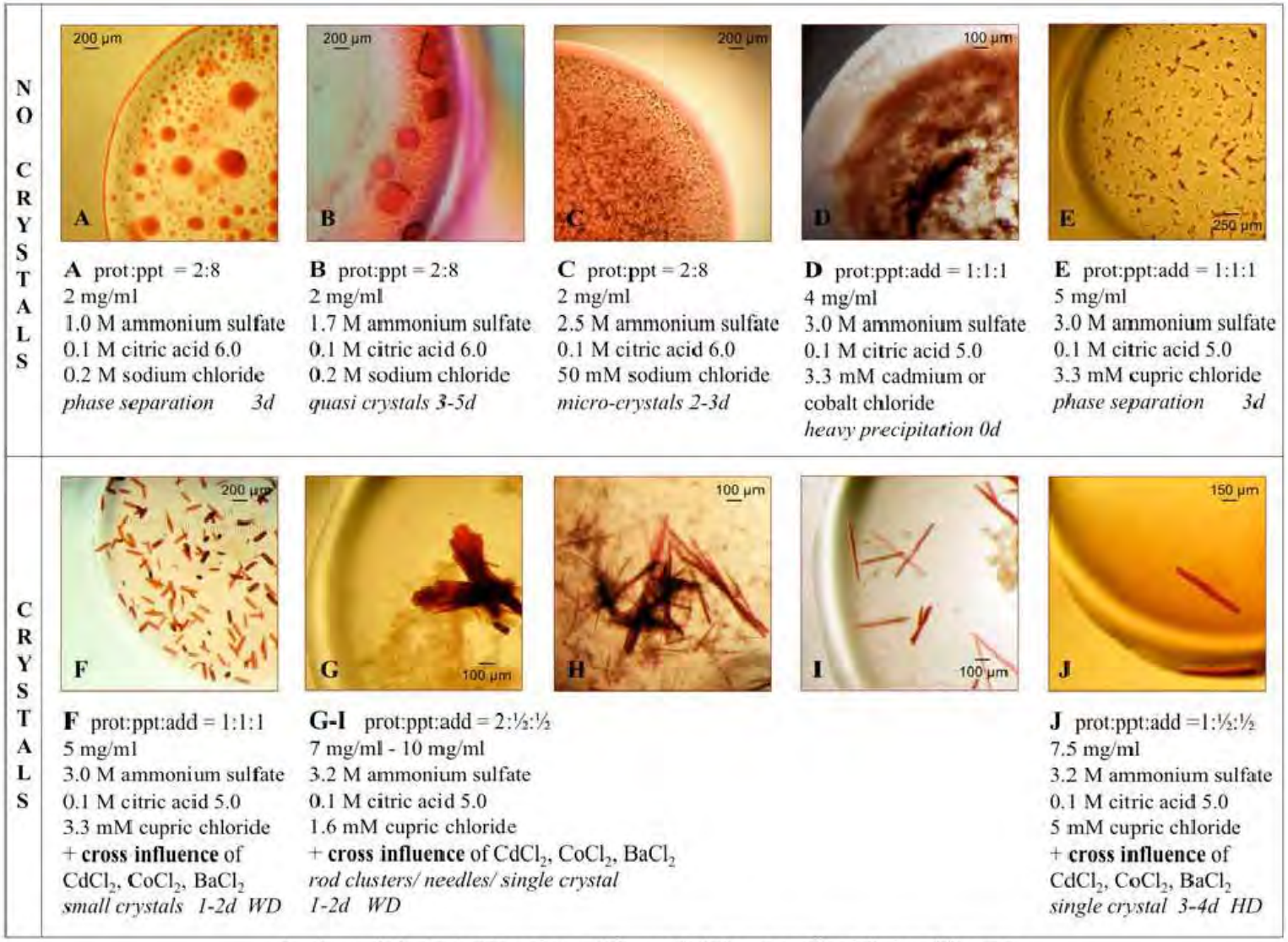
(Naomi Chayen, Curr Opin Struct Biol, 2004)

Protein crystal growth: idea

What do we expect to see during protein crystallization experiments?

No crystals:
Protein precipitation (irreversible), D
Phase separation , A/E
Microcrystals or quasicrystals, B/C

Crystals:
Small crystals, F
Needles, H/I
Rod clusters, G
Single big crystal, J



(Nemčovičová & Smatanová, Chemistry, 2012)

Protein crystal growth: ways

What to consider before crystallizing proteins?

Table 1. Biochemical & Chemical variables that could or do affect protein crystal growth	
Purity of the sample	Genetic modifications
Conformational flexibility of the sample	Symmetry of the molecule
Homogeneity of the sample	Stability and level of denaturation of the sample
pH and buffer	Isoelectric point
Type and concentration of the precipitant (reagent)	His tags and other purification tags – presence or absence
Concentration of the sample	Thermal stability
Purity of the sample	pH stability
Additives, co-factors, ligands, inhibitors, effectors, and excipients	History of the sample
Chaotropes	Proteolysis
Detergents	Microbial contamination
Metals	Storage of the sample
Ionic strength	Handling of the sample and associate cleanliness
Reducing or oxidizing agents	Anion and cation type and concentration
Source of the sample	Degree of relative supersaturation
Presence of amorphous or particulate material	Initial and final concentration of the reagent
Post-translational modifications	Path and rate of equilibration
Chemical modifications	

Table 2. Physical variables that could or do affect protein crystal growth	
Temperature	Electric and magnetic fields
Rate of equilibration	Surface of the crystallization device
Method of crystallization	Viscosity of the reagent
Gravity, convection, and sedimentation	Heterogeneous and epitaxial nucleants
Vibration and sound	Geometry of crystallization device
Volume of the sample and reagent	Time
Pressure	Dielectric property of the reagent

(Crystal Growth 101, Hampton Research, 2019)

Protein crystal growth: ways

General crystallization approaches

Table 3. Crystallization Methods – Achieving Supersaturation	
Vapor Diffusion (Sitting, Hanging, Sandwich)	Sequential Extraction
Batch (Microbatch with or without oil)	pH Induced
Dialysis (Microdialysis)	Temperature Induced
Free interface diffusion (Counter diffusion, liquid bridge)	Effector Addition (Silver Bullet)
Controlled Evaporation	

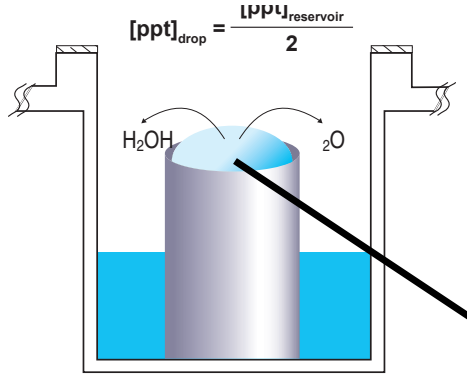
(Crystal Growth 101, Hampton Research, 2019)

Protein crystal growth: ways

General crystallization approaches

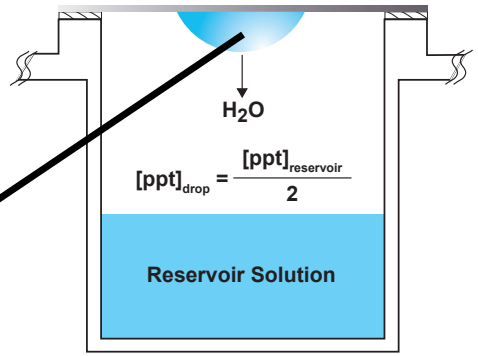
Vapor diffusion: sitting-drop

Figure 1
Process of Vapor Diffusion



Vapor-diffusion: hanging-drop

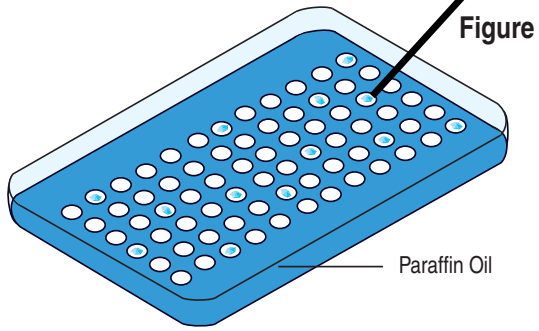
Figure 1
Process of vapor diffusion



Water evaporation to concentrate proteins in the crystallization drops.

Micro-batch

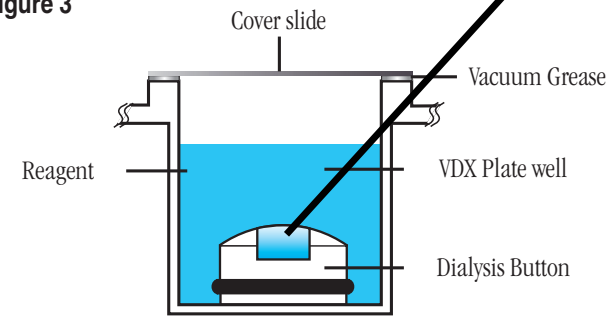
Figure 1



Buffer exchange to gradually precipitate proteins in the dialysis cassette.

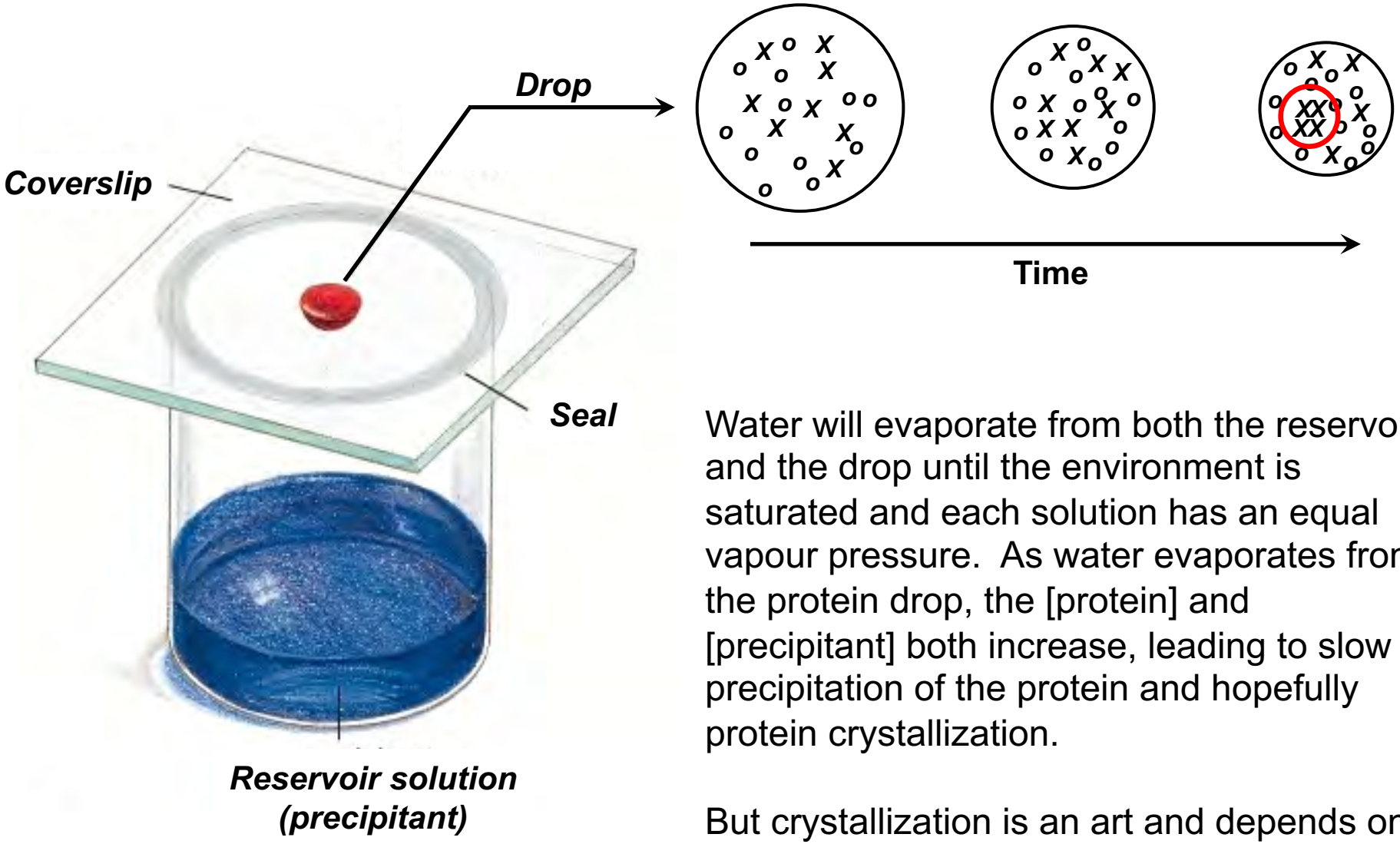
Micro-dialysis

Figure 3



Protein crystal growth: ways

Vapor diffusion



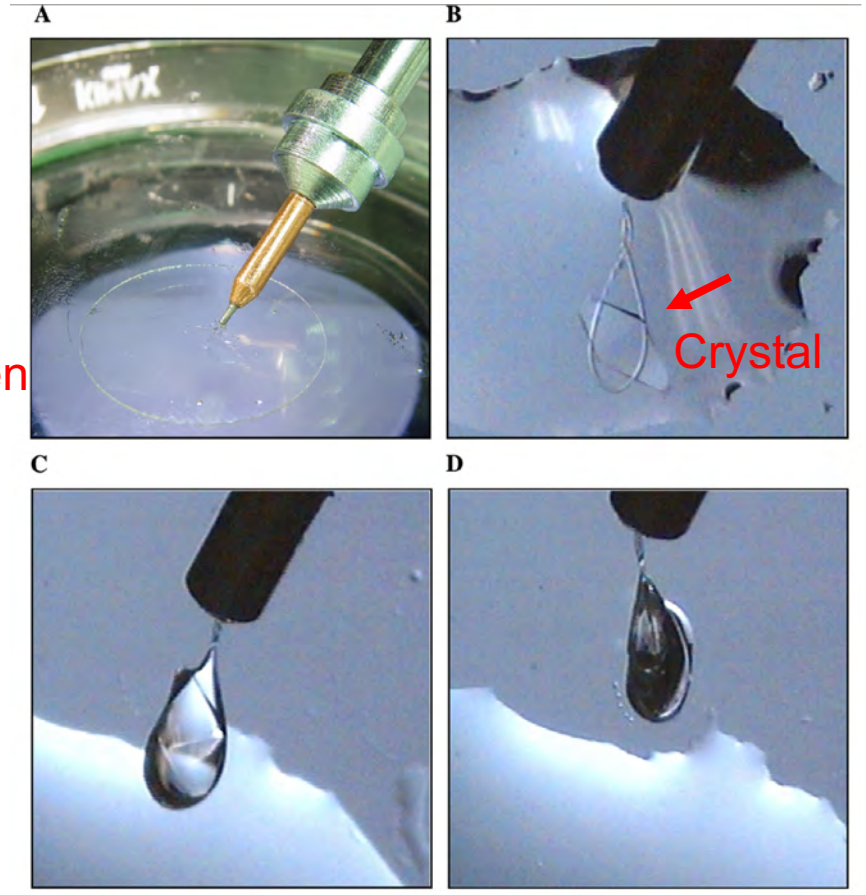
Water will evaporate from both the reservoir and the drop until the environment is saturated and each solution has an equal vapour pressure. As water evaporates from the protein drop, the [protein] and [precipitant] both increase, leading to slow precipitation of the protein and hopefully protein crystallization.

But crystallization is an art and depends on numerous factors…

Protein crystal growth: preparation for data collection (harvest)

Cryo-cooling

Liquid nitrogen ($N_{2(l)}$)



(Pflugrath, Methods, 2004)

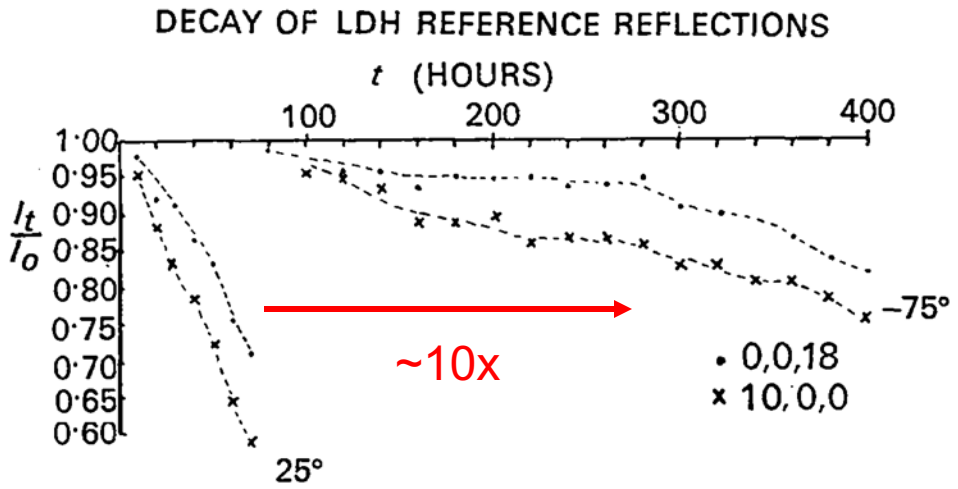


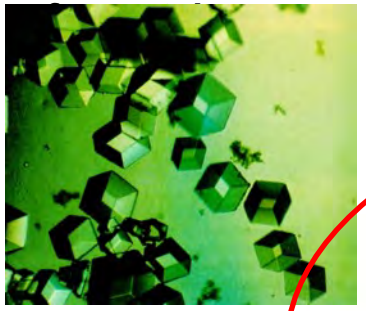
Fig. 5. The ratio I_t/I_0 for two reference reflections plotted as a function of exposure time for a typical native and frozen crystal. I_t represents the intensity at time t. Results for 0,0,18 and 10,0,0 are shown with dots and crosses respectively.

(Haas & Rossmann, Acta Crryst, 1970)

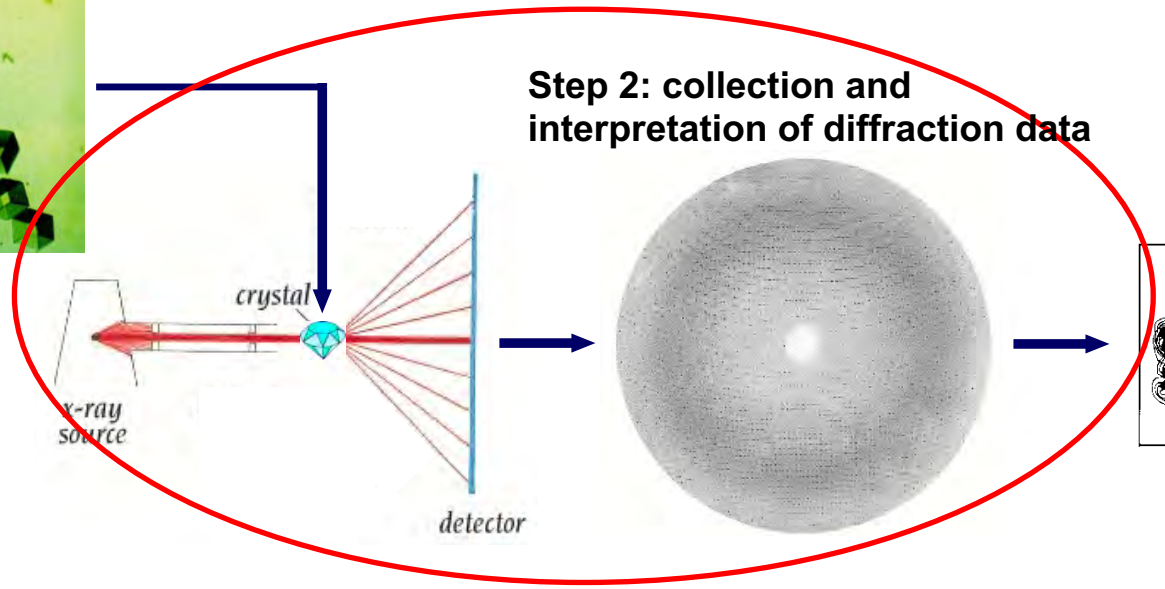
X-ray crystallography

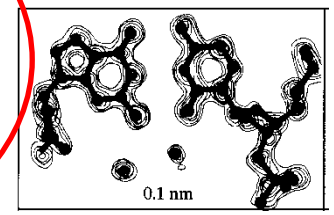
X-ray Crystallography involves 3 main steps:

Step 1: protein crystallization

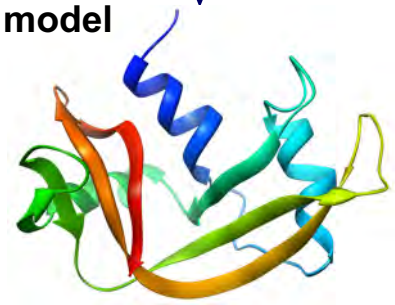


Step 2: collection and interpretation of diffraction data

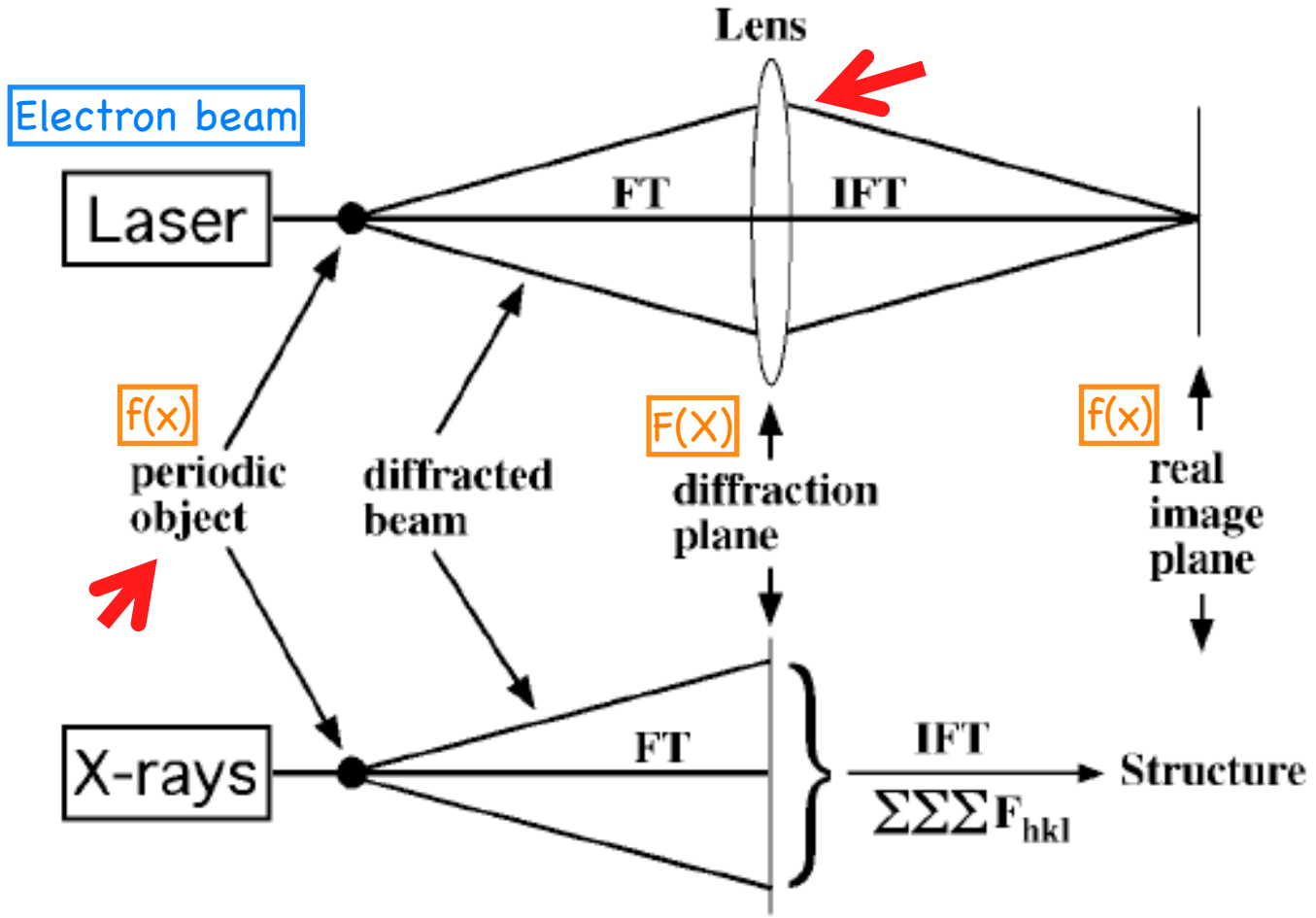




Step 3: build and refine a structural model



Optical & X-ray Diffractions

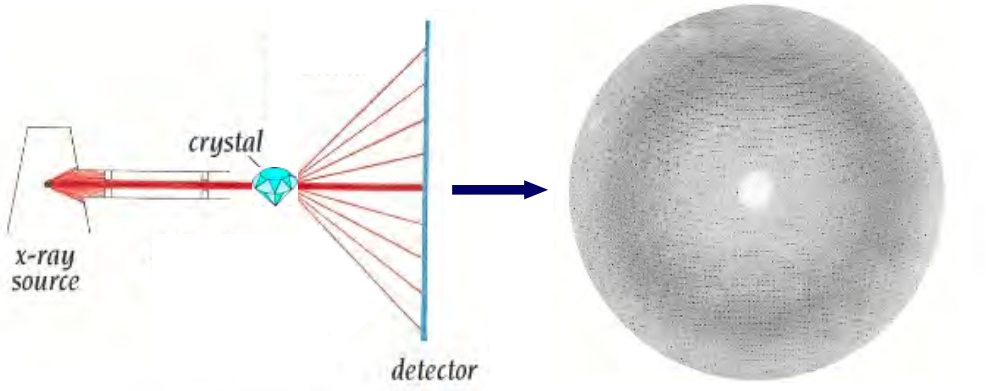


If F(X)=FT[f(x)], then f(x)=IFT[F(X)]
where FT=Fourier transform & IFT=Inverse Fourier transform.

Note: important in X-ray crystallography and 3-D reconstruction algorisms.

X-ray crystallography

Data collection: principles



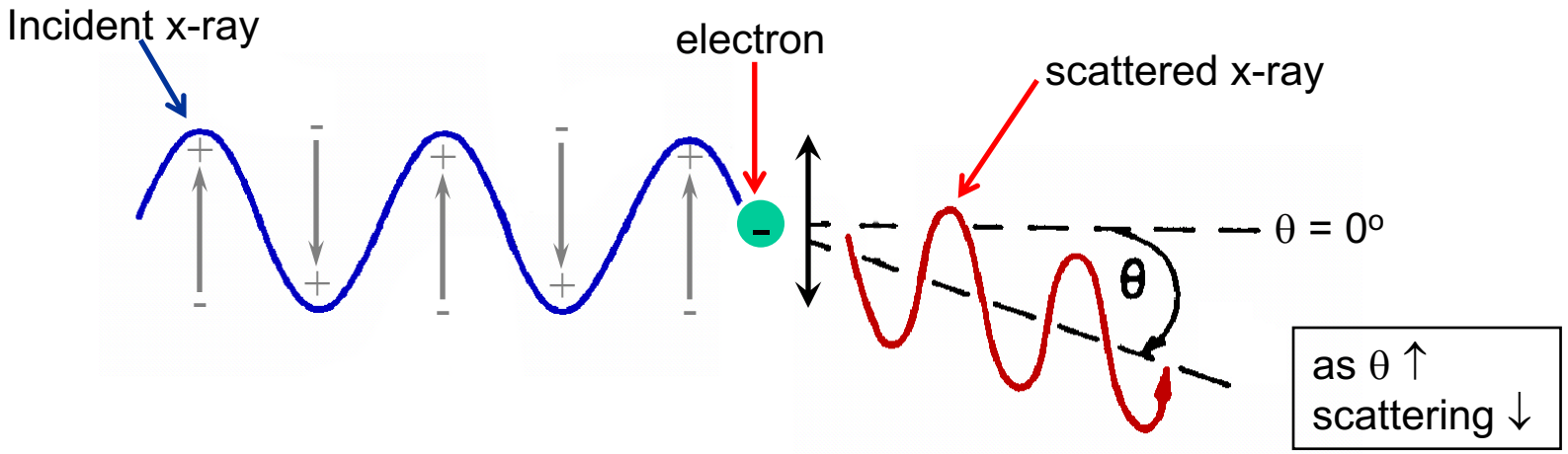
1. Scattering of X-ray
2. Constructive vs Destruct interference
3. **Bragg's Law**

A protein crystal is exposed to X-rays. Most x-rays travel right through the crystal, but some are *scattered* by the electrons surrounding each atom (electron density). Although scattering occurs in all directions most scattered X-rays *interfere destructively* – so no X-ray is detected. Some scattered X-rays *interfere constructively leading to a relatively intense "diffracted" X-ray* (red lines). Scattering and diffraction are at the heart of x-ray crystallography.

To understand x-ray crystallography, we must first understand basics of **how x-rays are scattered and interfere with each other.**

Diffraction data

1. Scattering of x-rays

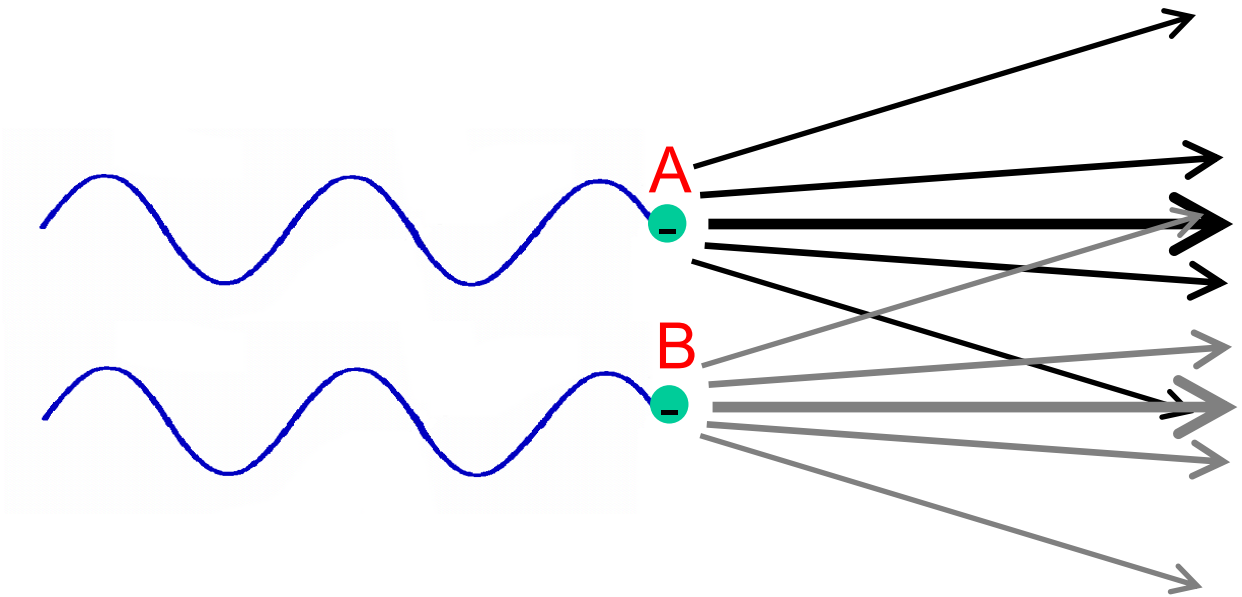


X-rays can be thought of as having wave properties where the electric field of the radiation oscillates as it travels through space. The oscillating electric field of the incident X-ray (I_o) causes the negatively charged electrons in a molecule to oscillate at the same frequency. The oscillating electrons can then emit x-rays of the same wavelength, but the x-rays are emitted in every direction. These emitted x-rays are called scattered x-rays.

Note that x-rays are scattered in the forward direction more intensely than in other directions (as θ increases, the scattering intensity decreases).

Diffraction data

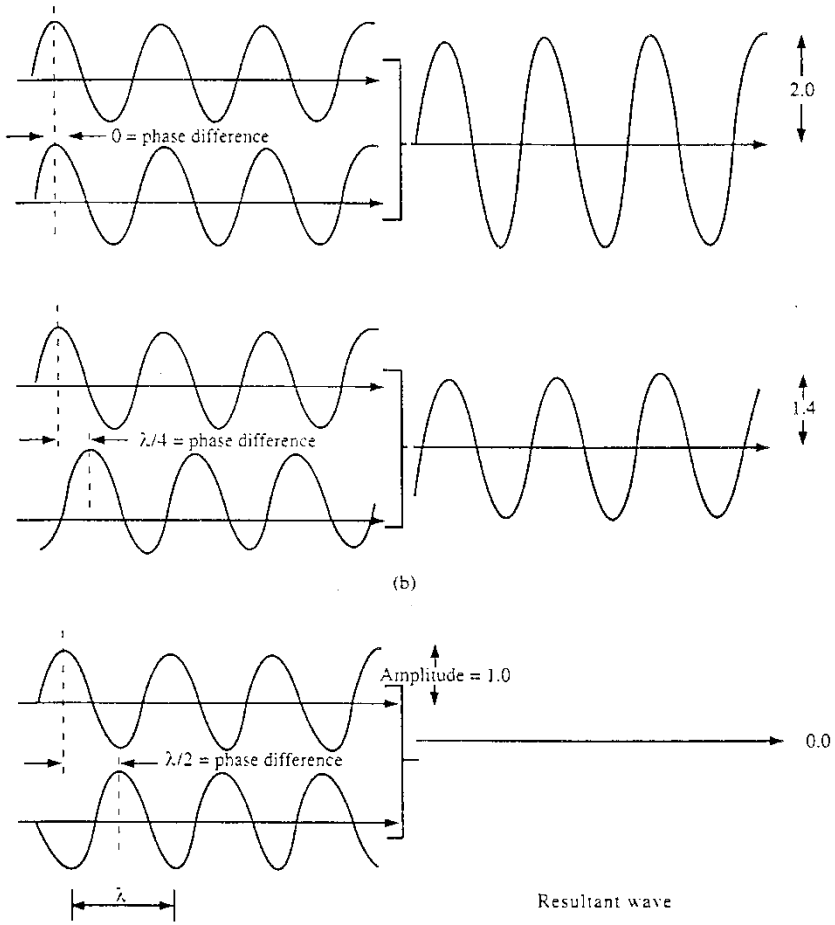
2. Constructive versus destructive interference: what happens if we have more than one electron?



The x-rays scattered from atom A (black) will overlap and thus interact with the x-rays scattered from atom B (gray), etc., a process referred to as “interference”. Interference can be either constructive or destructive.

Diffraction data

2. Constructive versus destructive interference:



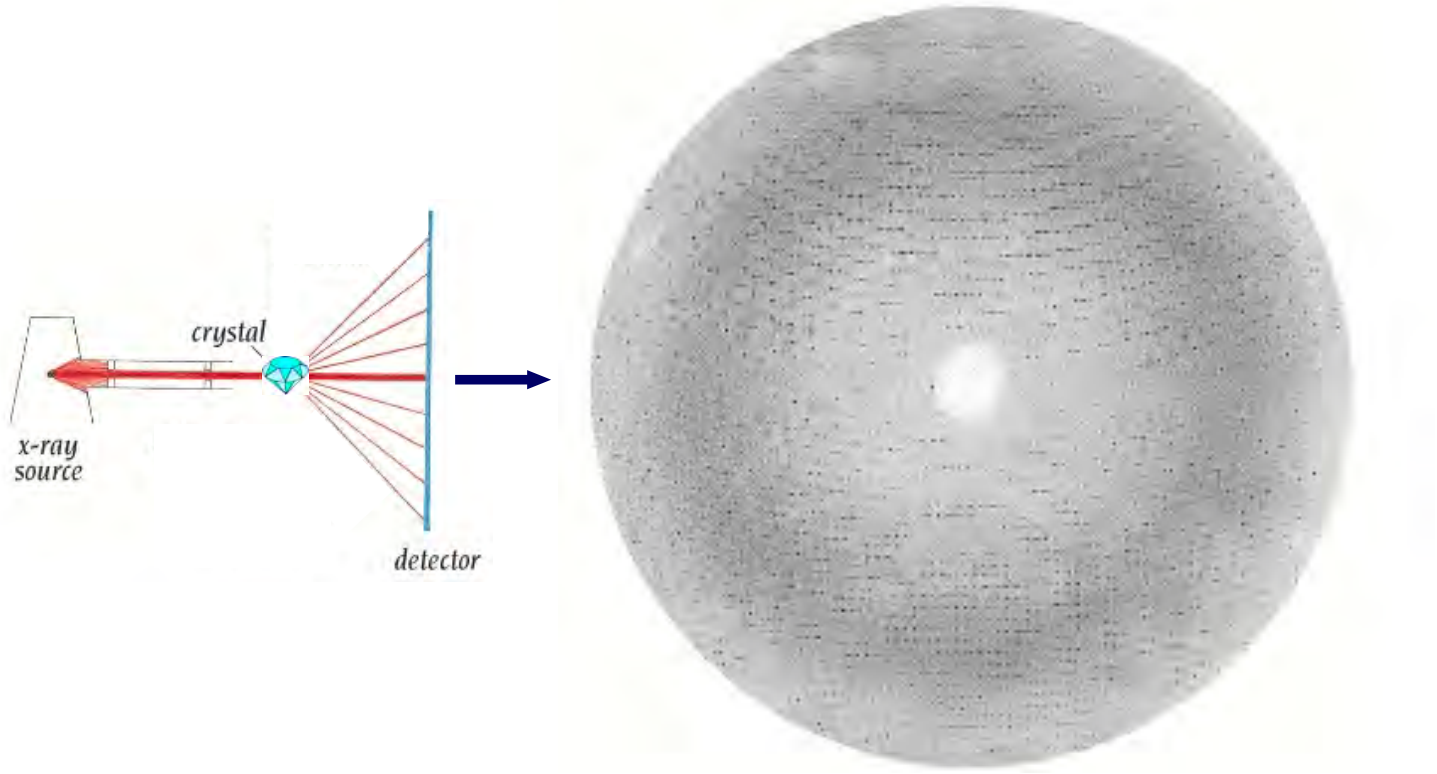
1) if the scattered x-rays are “in phase” (*i.e. they reach their max and min at same point in space*), they add together to increase the intensity – *constructive interference.*

2) if you add 2 waves that oscillate only slightly out of phase, the resultant X-ray is not as intense as above.

3) if you add 2 waves that are completely out of phase (*one reaches a max while the other is at its minimum*), they cancel each other – *destructive interference.*

Diffraction data

2. Constructive versus destructive interference:

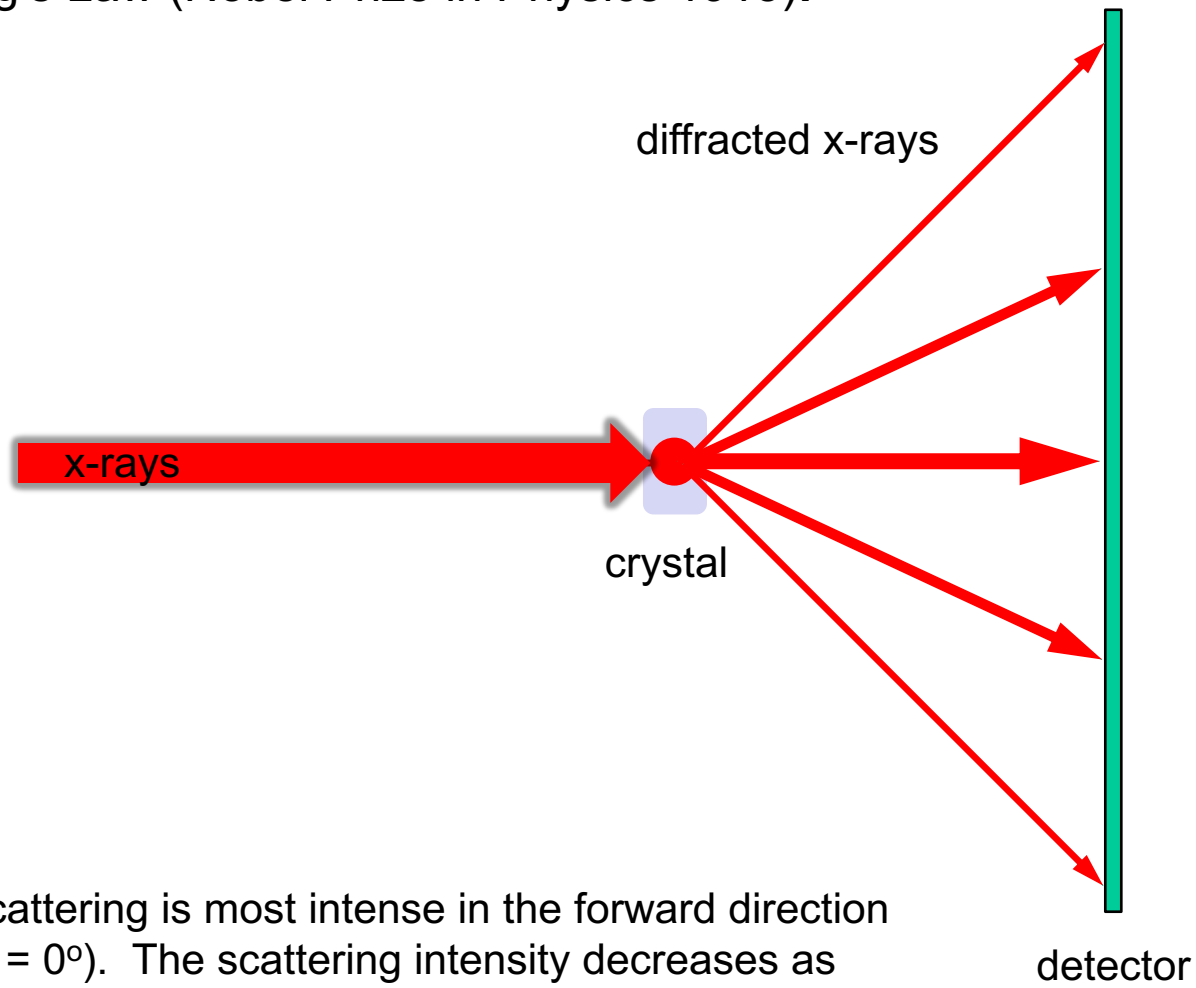


diffraction pattern

The scattered x-rays that emerge in phase interfere constructively. These “enhanced” x-rays are called diffracted x-rays, and they lead to on the x-ray detector. The resulting pattern of spots is called a diffraction pattern. *What determines whether the scattering of x-rays leads to diffraction?*

Diffraction data

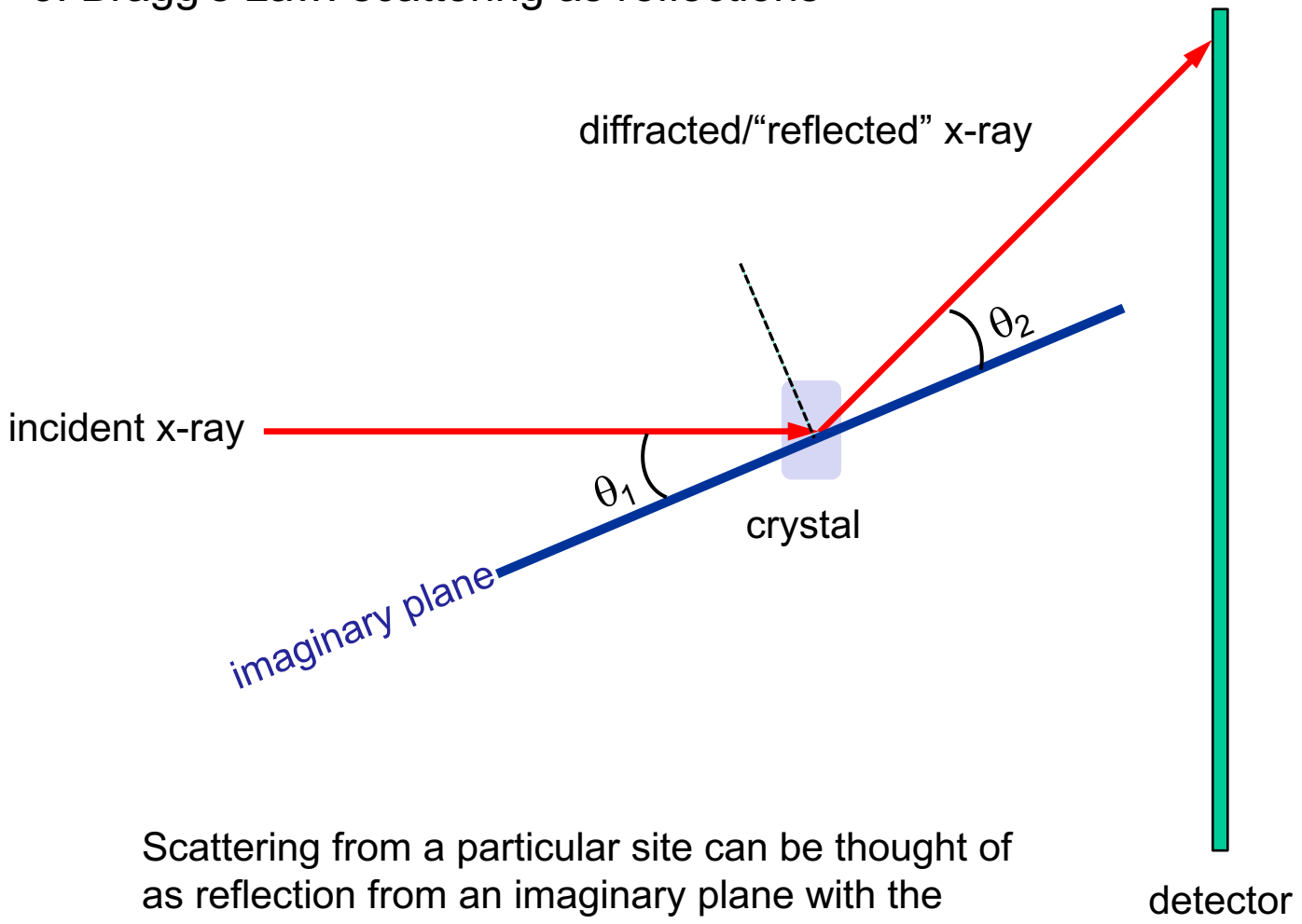
3. Bragg's Law (Nobel Prize in Physics 1915):



Scattering is most intense in the forward direction ($\theta = 0^o$). The scattering intensity decreases as you move away from θ.

Diffraction data

3. Bragg's Law: scattering as reflections

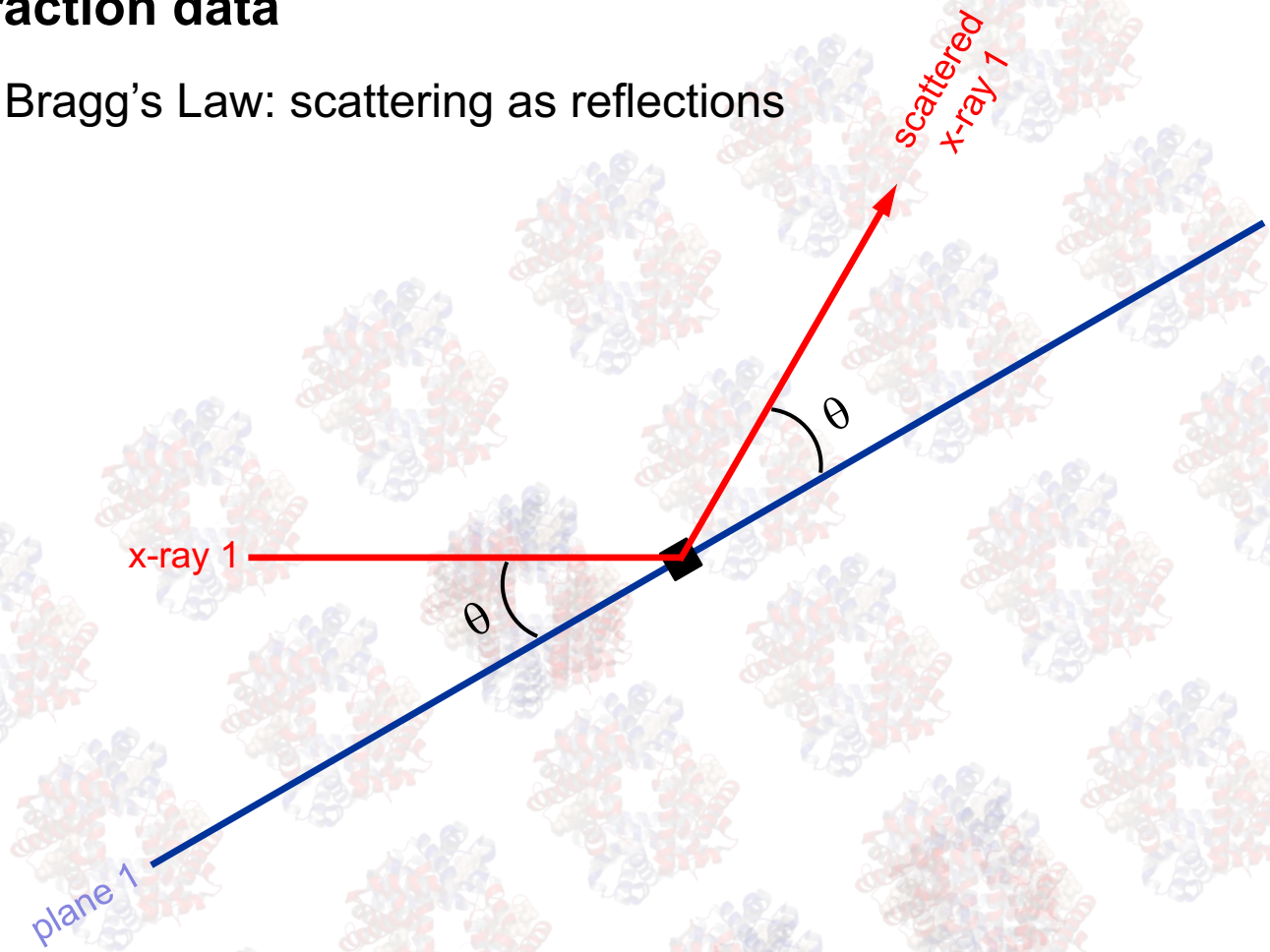


Scattering from a particular site can be thought of as reflection from an imaginary plane with the plane drawn so that the angle of incidence, θ_1, is equal to the angle of reflection, θ_2.

In the next slide we zoom in on the crystal – shown as the blue rectangle

Diffraction data

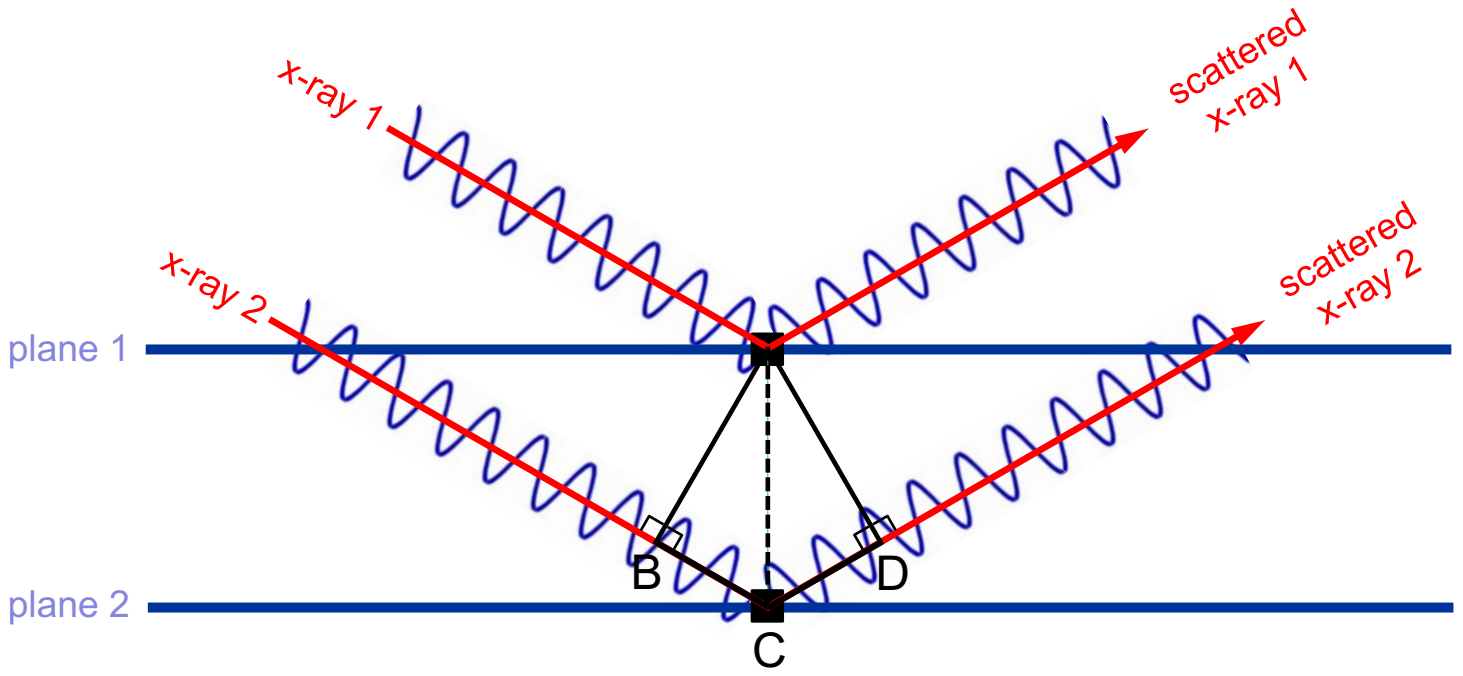
3. Bragg's Law: scattering as reflections



Scattering of x-rays from a single atom, A_1, in a single protein within the crystal (the diagram situates A1 in the unit cell.

Diffraction data

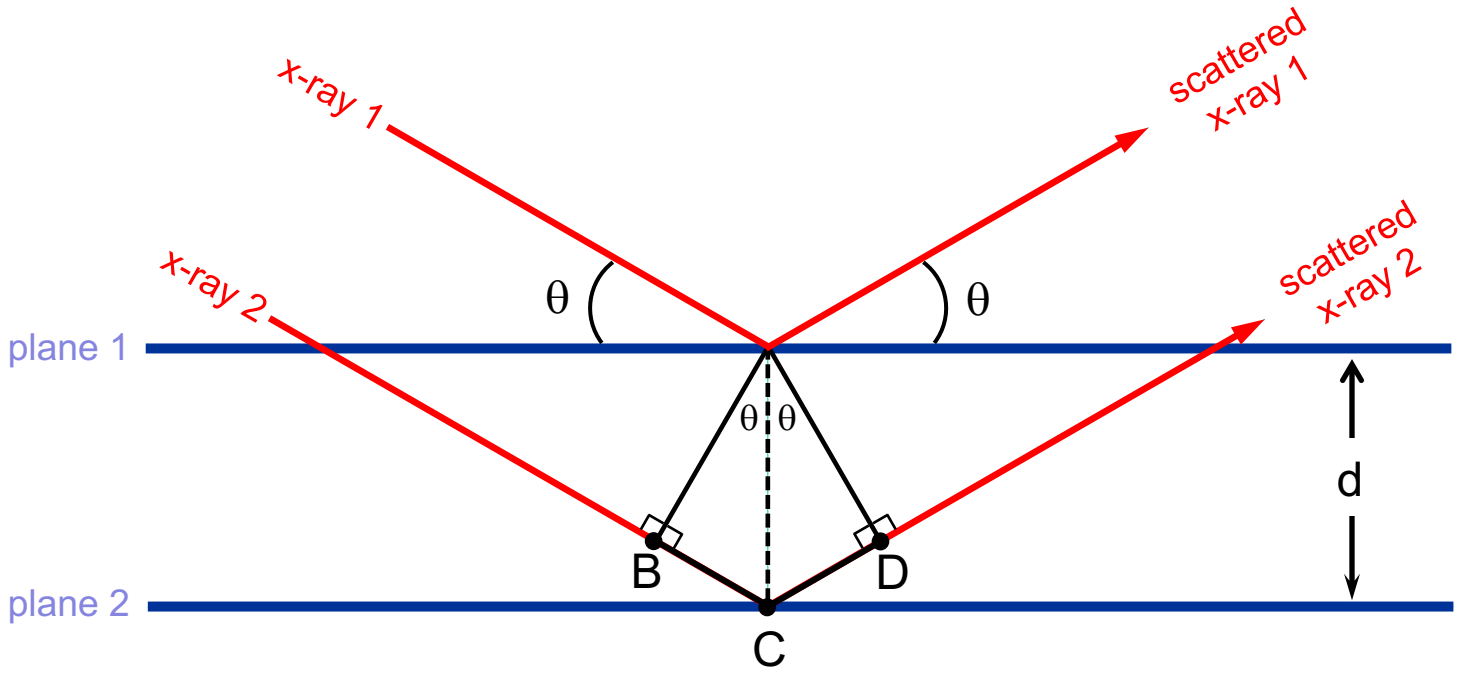
3. Bragg's Law: scattering adjacent planes and interference



Given that x-ray 1 and x-ray 2 are from the same source, they have the same frequency and phase. Bragg noticed that x-ray 2 travels a longer distance before it hits A_2 than x-ray 1 travels until it hits A_1 Scattered x-ray 2 also travels a similar distance until it catches up to scattered x-ray 1.

Diffraction data

3. Bragg's Law: scattering adjacent planes and interference



$$BC = CD = d \cdot \sin\theta$$

X-ray 2 travels a farther distance (BC + CD) than x-ray 1. The two x-rays will only emerge in phase if the extra distance traveled by x-ray 2 is an integer multiple of the wavelength of the x-ray (λ). Bragg's Law states that diffraction will occur when:

$$n\lambda = 2d \cdot \sin\theta \ (n = 1, 2, 3\ldots)$$

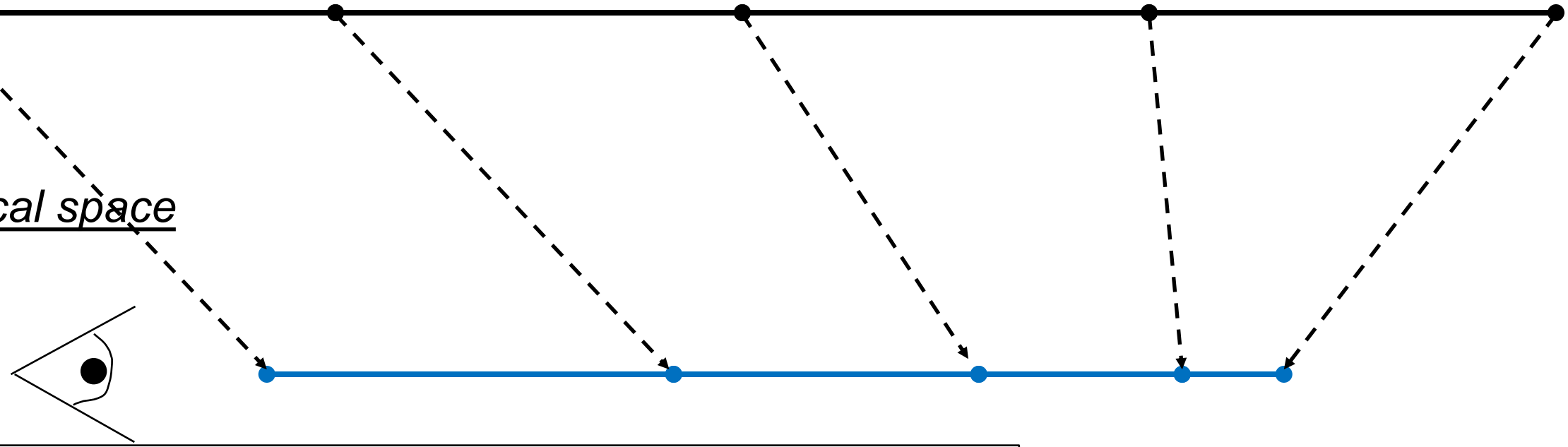
Reciprocal Space

Real space & reciprocal space

Highway mileage marks

Real space

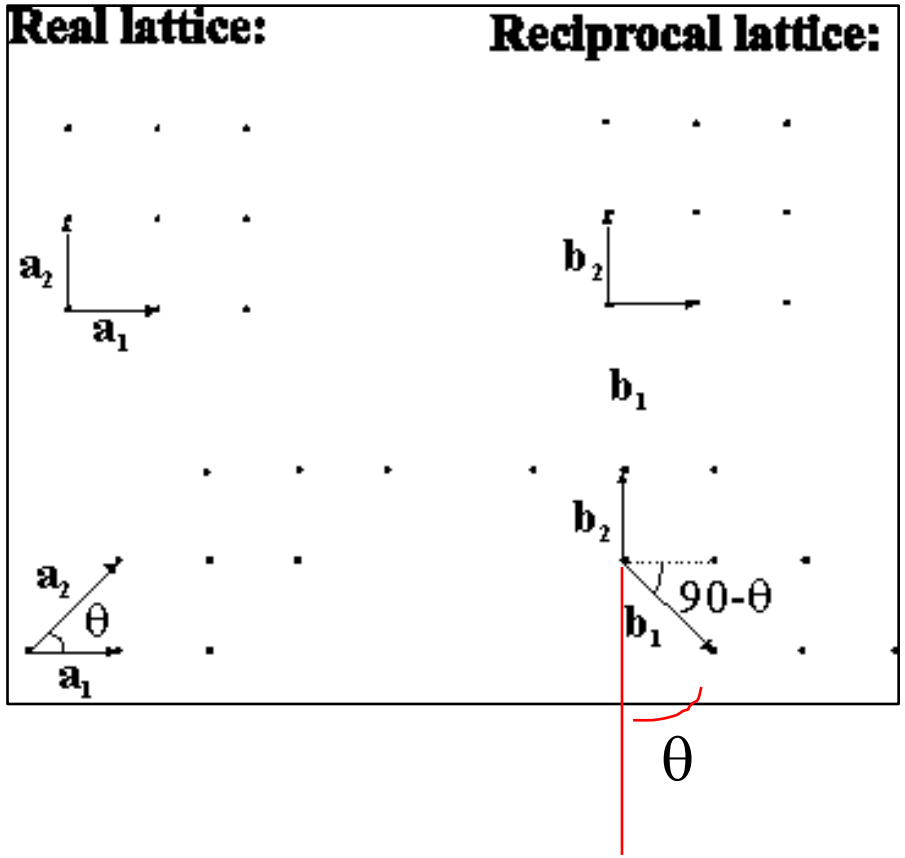
Reciprocal space

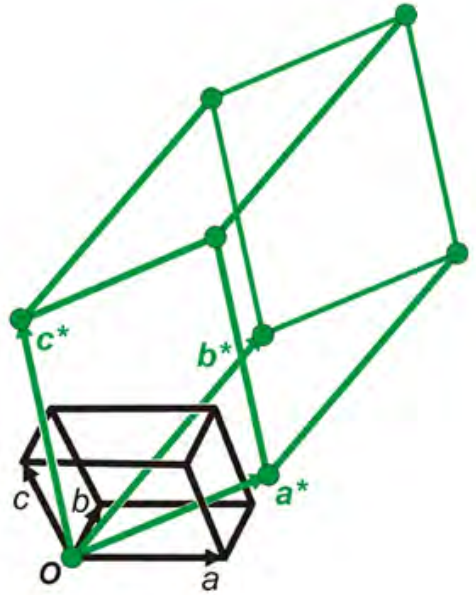


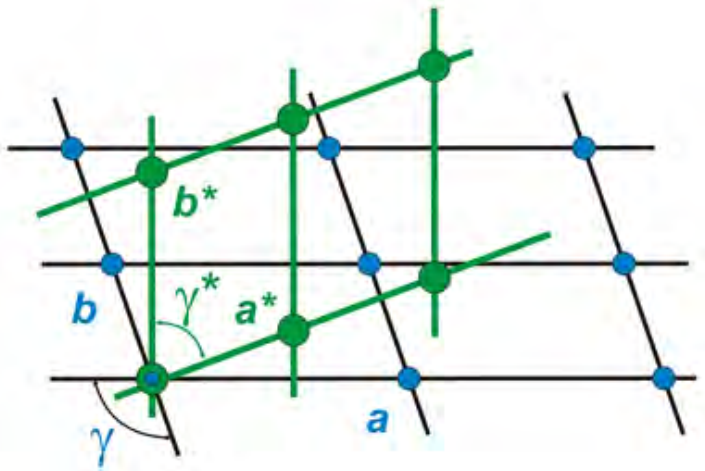
Perspective when looking at the mileage marks from one end of the road.

Reciprocal Space

Real space & reciprocal space

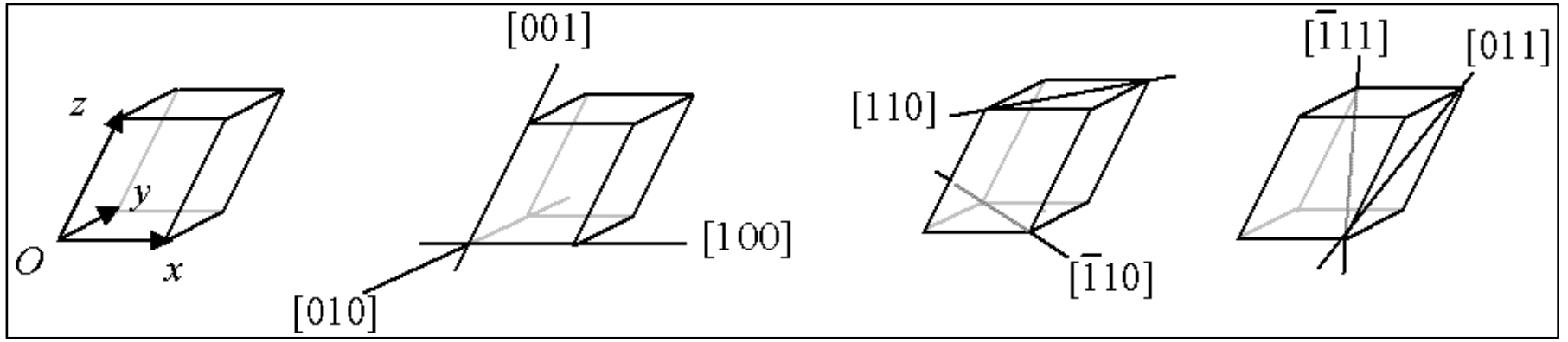


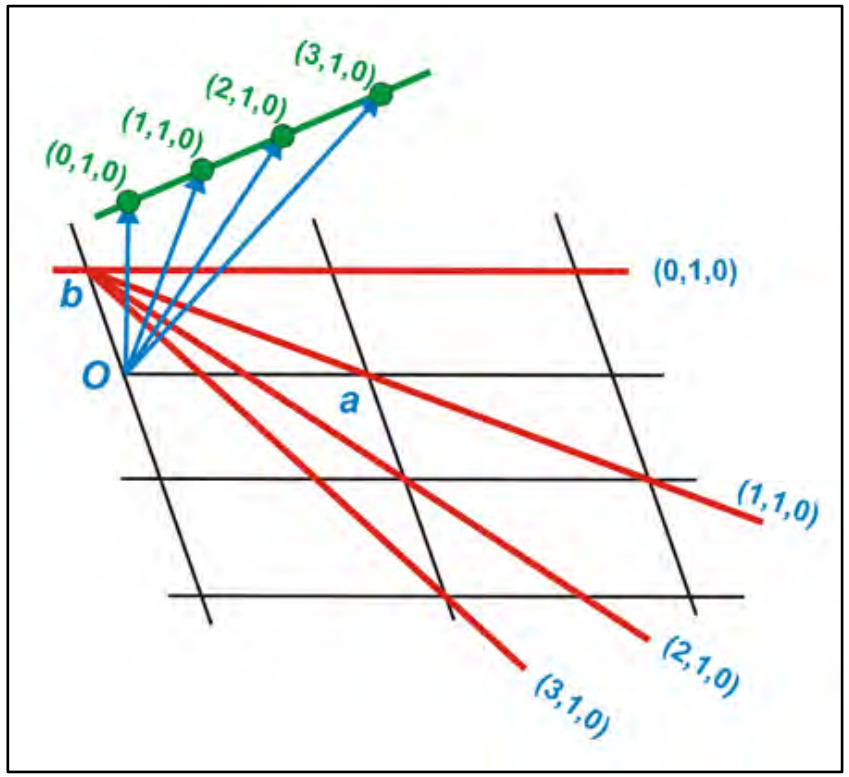


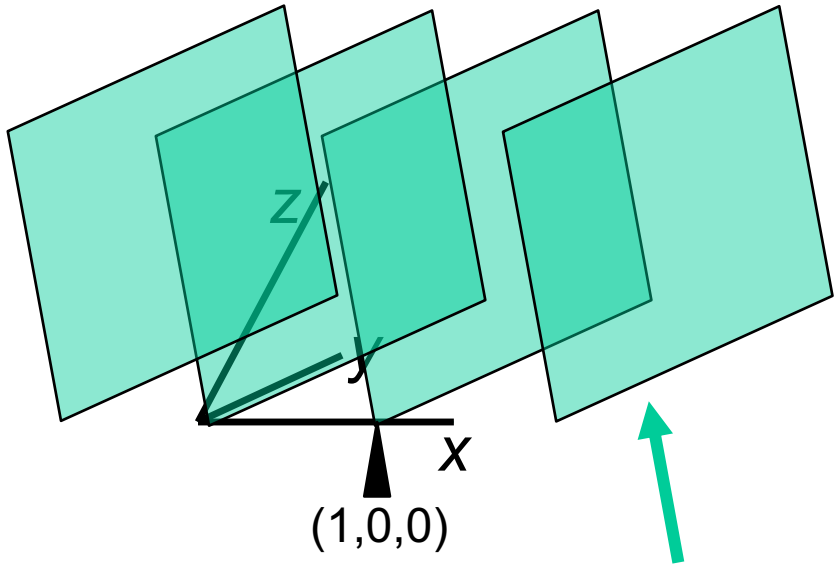


Reciprocal Space

Miller index



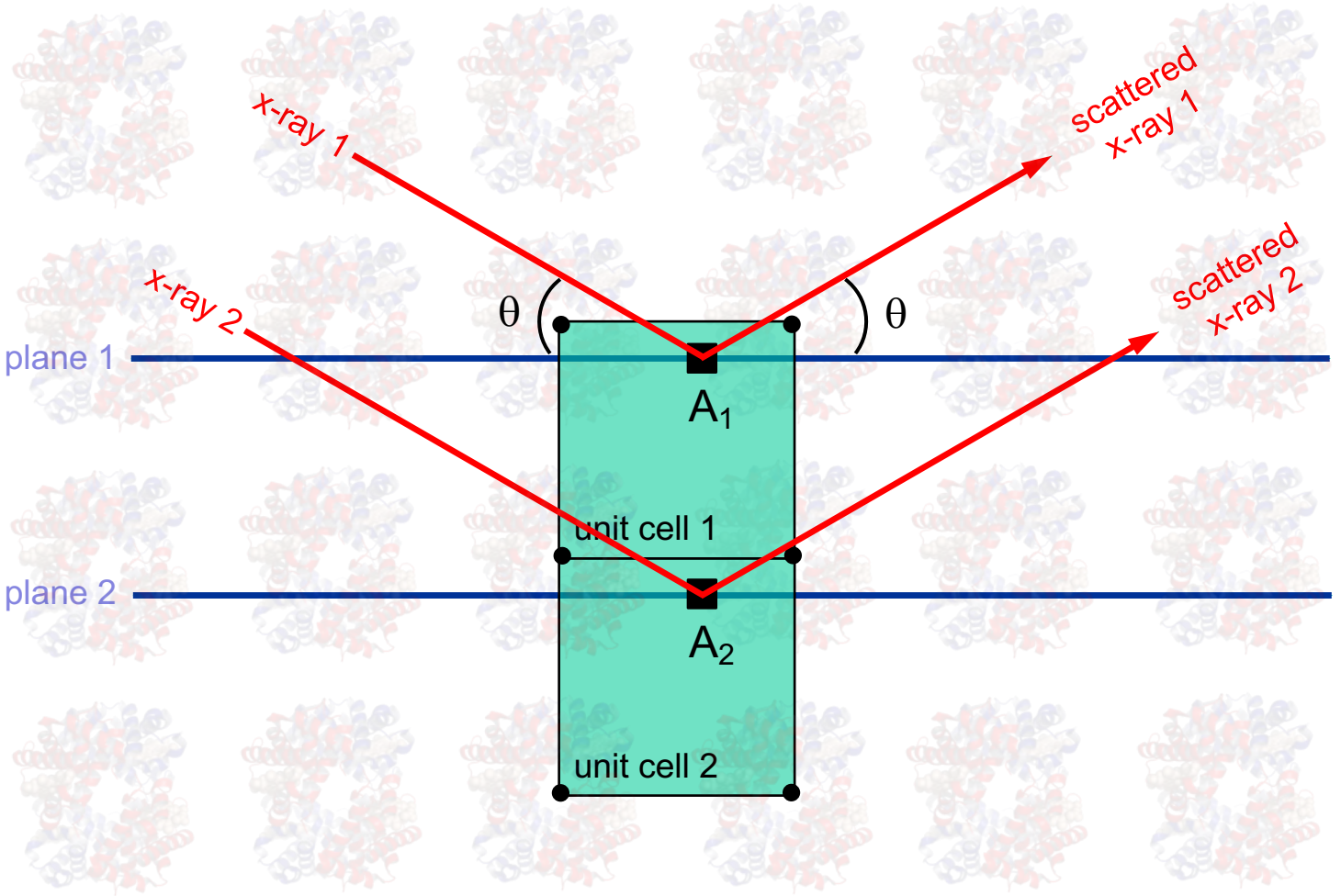




Miller plane: [100]

Diffraction data

3. Bragg's Law: scattering adjacent planes

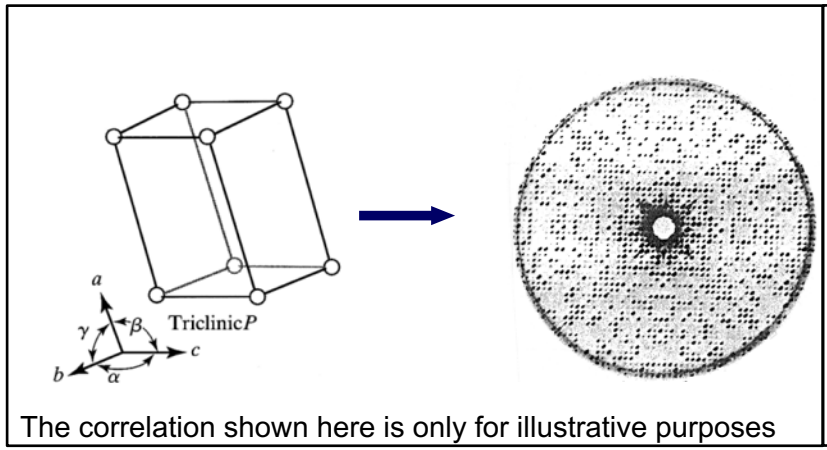


Scattering of x-rays from atoms A_1 and A_2. Bragg asked the simple question. What will determine whether scattered x-ray 1 and scattered x-ray 2 emerge in phase to interfere constructively?

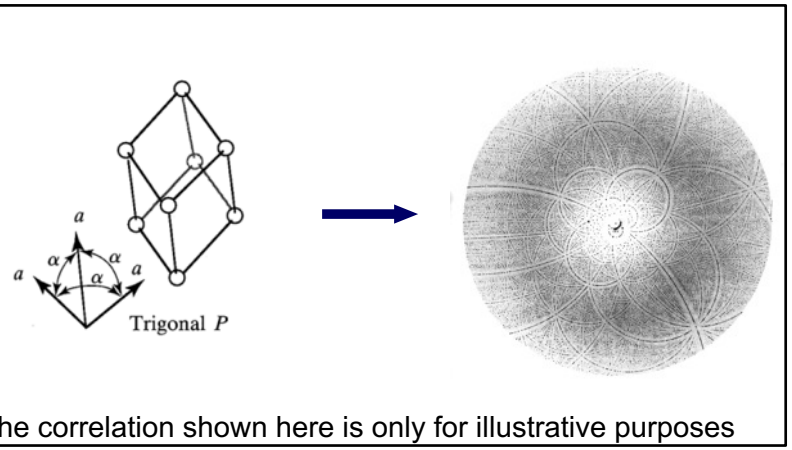
Diffraction data

3. Bragg's Law: the unit cell dictates the diffraction pattern

If the x-rays scattered in one direction from two identical atoms in adjacent unit cells interfere constructively, then all x-rays scattered that direction from identical atoms in adjacent unit cells will interfere constructively. The distance between these atoms depends on the type and size of the unit cell.



The correlation shown here is only for illustrative purposes

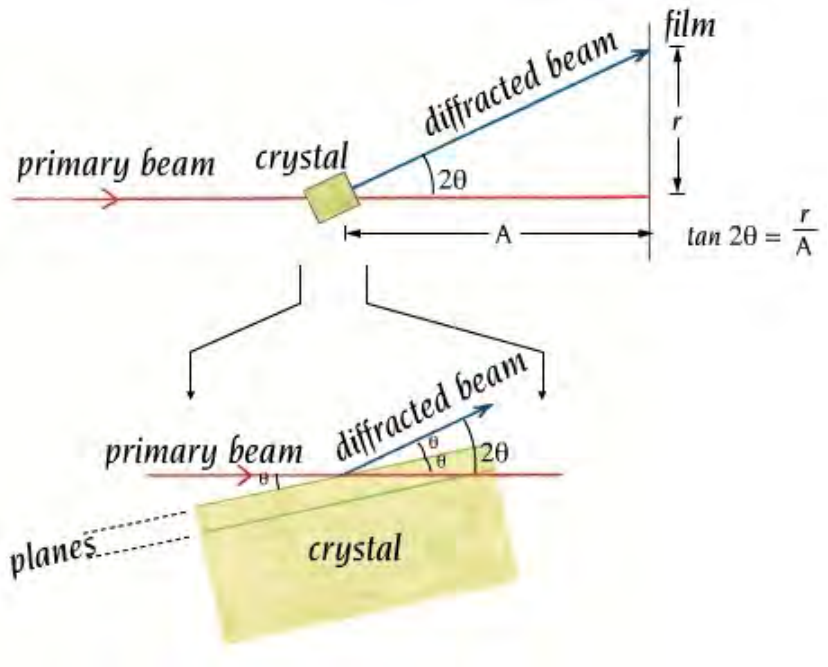


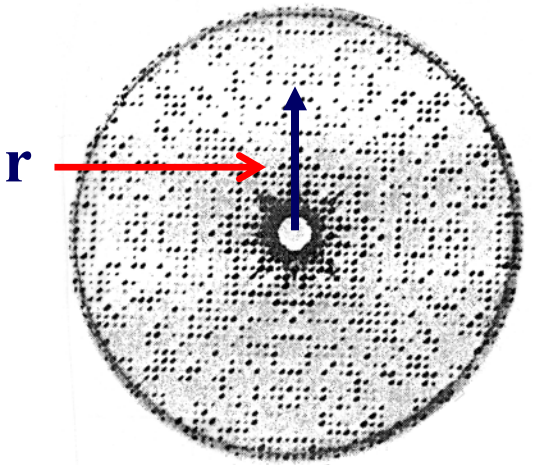
The correlation shown here is only for illustrative purposes

From the pattern of diffraction spots, you can determine the type and size of the unit cell - this is the first step in solving a structure!

Diffraction data

3. Bragg's Law: the unit cell dictates the diffraction pattern





The pattern of spots dictates the type of unit cell.

The distance from the center of the diffraction pattern to the location of the spot, **r**, is related to unit cell size. You can calculate the "reflection angle", θ, using the formula:

$$\tan 2\theta = r/A$$

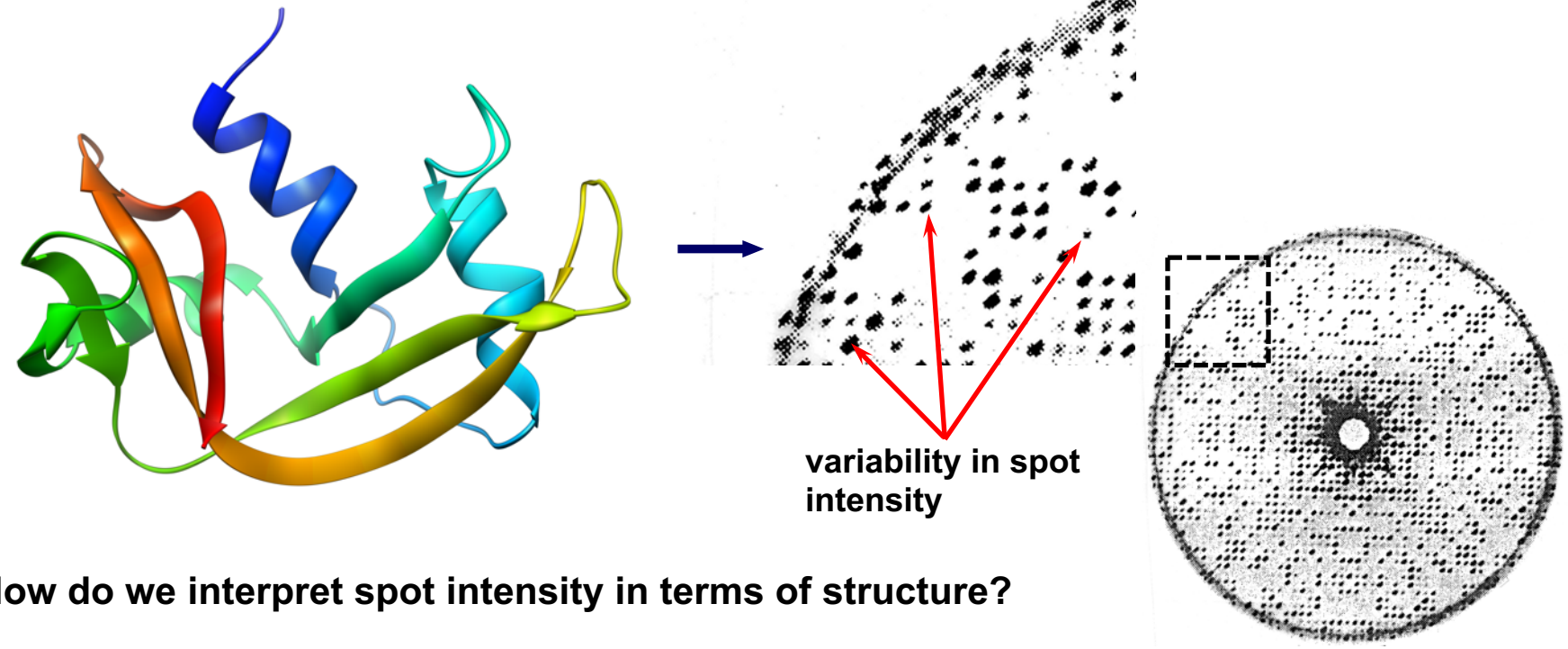
where **A** is the distance from the crystal to the detector.

Diffraction data

3. Bragg's Law: the unit cell dictates the diffraction pattern

Because distances between identical atoms in different unit cells, **d**, are repeated over and over (it is a crystal – a regular repeating array of protein molecules!), these distances (i.e. the unit cell) govern whether diffraction will occur or not and thus the pattern of spots in the diffraction pattern.

The x-rays scattered from different atoms within the unit cell also interfere with each other, but as these differences between atoms in the unit cell are highly variable (i.e. they are governed by the 3D protein structure!), they only modulate the intensity of the diffraction spots. ***So the intensity of each spot tells you about protein structure!***

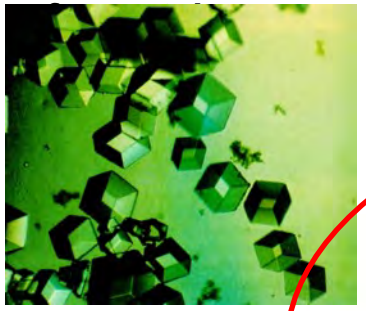


How do we interpret spot intensity in terms of structure?

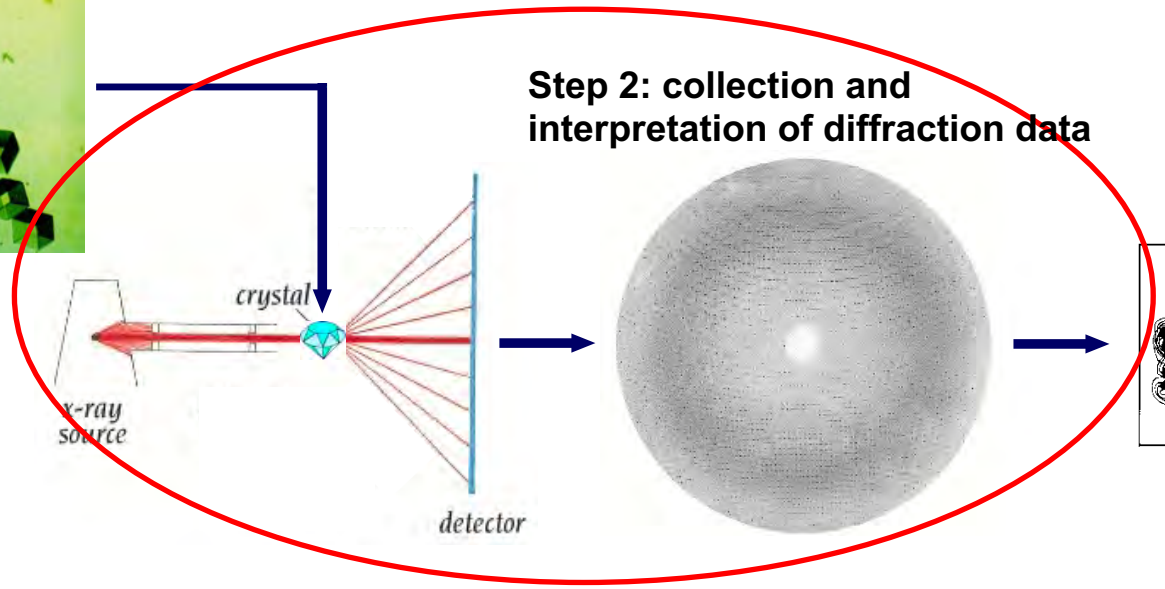
X-ray crystallography

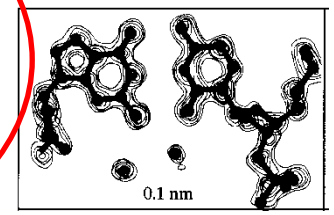
X-ray Crystallography involves 3 main steps:

Step 1: protein crystallization

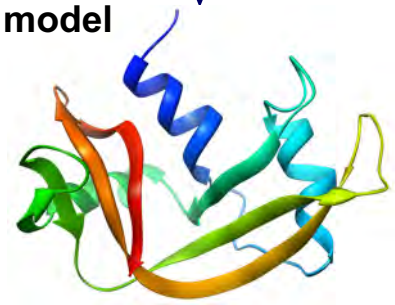


Step 2: collection and interpretation of diffraction data



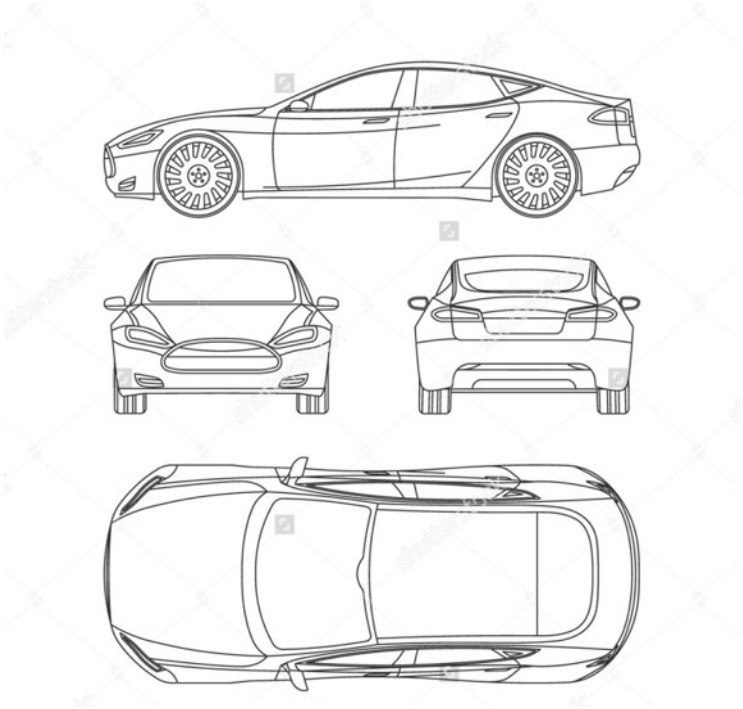


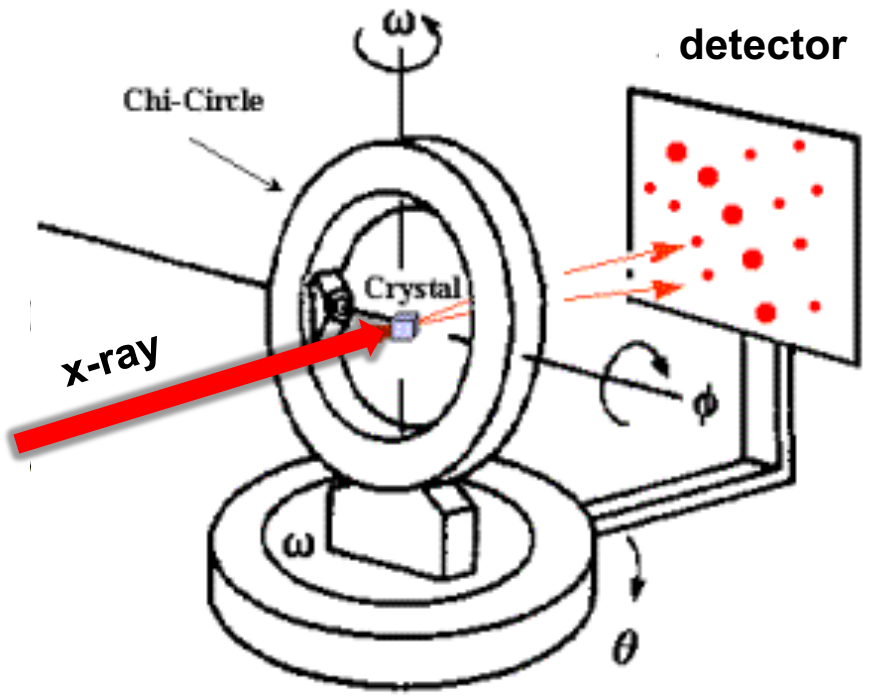
Step 3: build and refine a structural model



Diffraction data

From diffraction to electron density

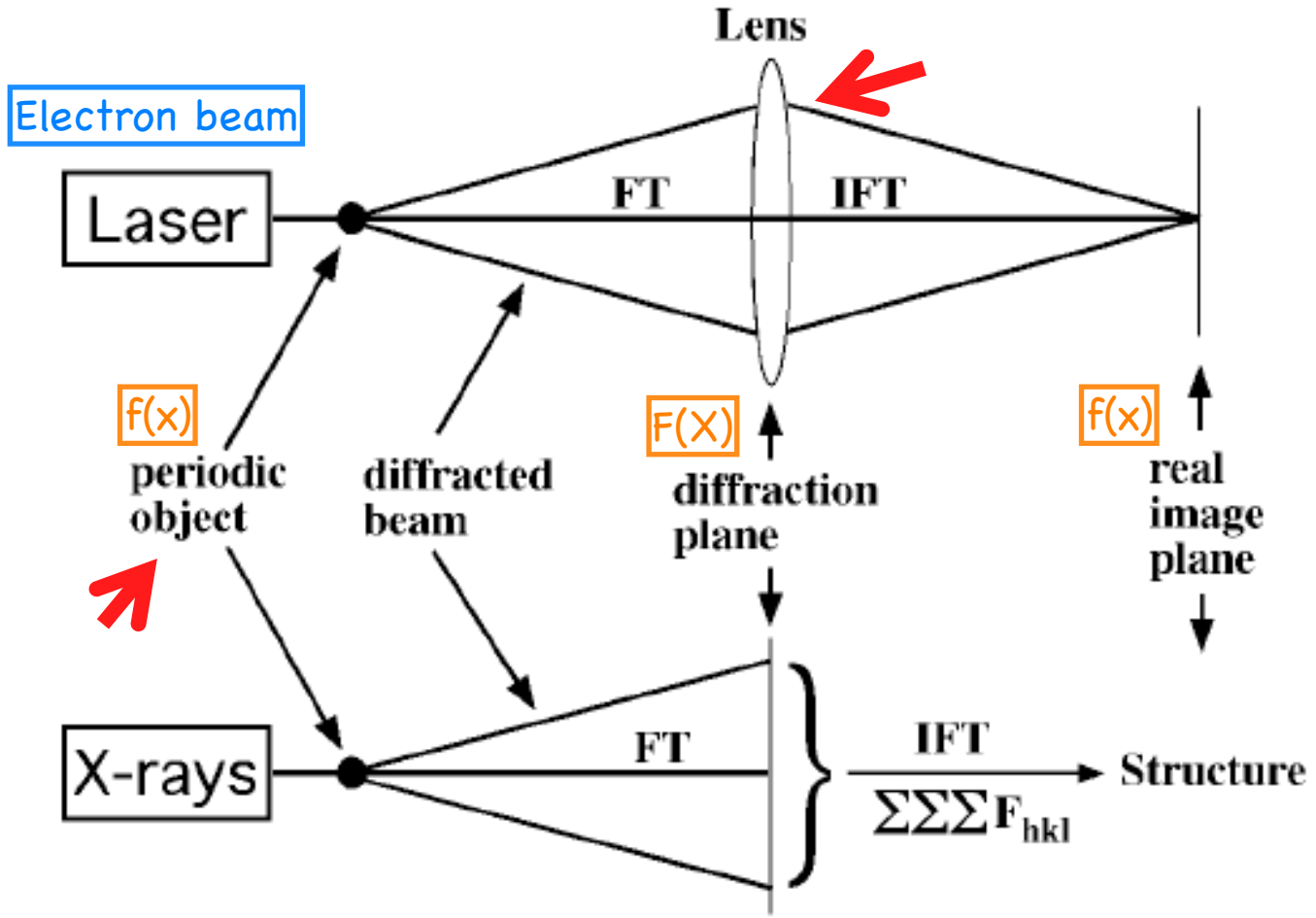




To fully describe any object in 3D, you must have “pictures” from every angle.

The same goes for a protein in a crystal. To solve the protein structure, we must record diffraction patterns (i.e. take “pictures”) from every angle. To do this, the protein crystal is placed in a goniometer – a device that allows one to record a diffraction pattern, then rotate the crystal, then record another. This allows one to get a 3D diffraction pattern.

Optical & X-ray Diffractions



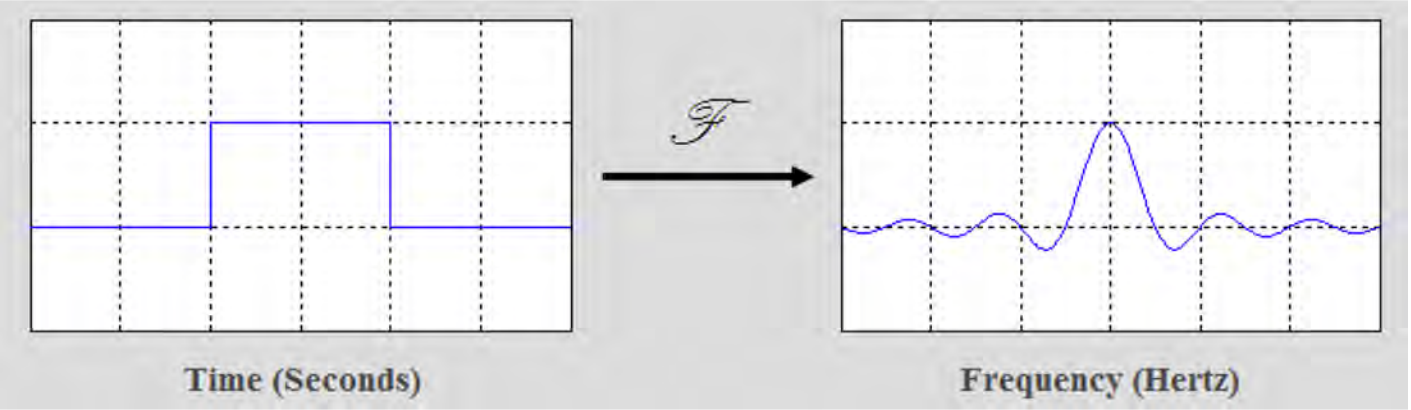
If F(X)=FT[f(x)], then f(x)=IFT[F(X)]
where FT=Fourier transform & IFT=Inverse Fourier transform.

Note: important in X-ray crystallography and 3-D reconstruction algorisms.

Fourier Transform

$$\mathcal{F}\{g(t)\} = G(f) = \int_{-\infty}^{\infty} g(t)e^{-i2\pi ft}\,dt$$

$$\mathcal{F}^{-1}\{G(f)\} = g(t) = \int_{-\infty}^{\infty} G(f)e^{i2\pi ft}\,df$$



The Fourier Transform is a tool that breaks a waveform (a function or signal) into an alternate representation, characterized by sine and cosines. The Fourier Transform shows that any waveform can be re-written as the sum of sinusoidal functions.

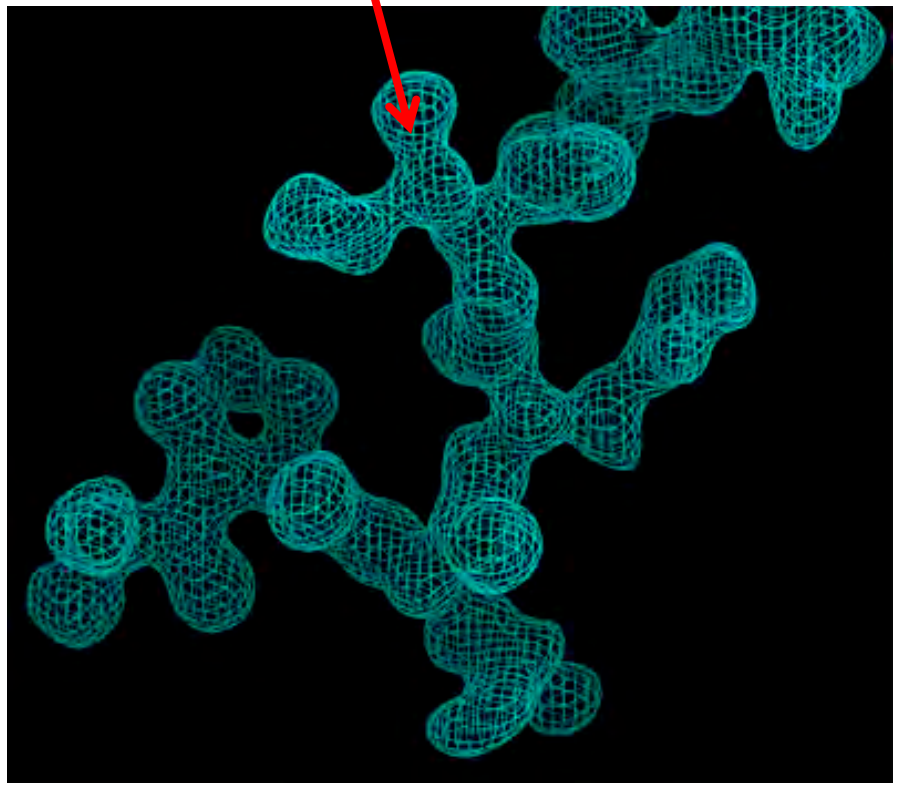
Diffraction data

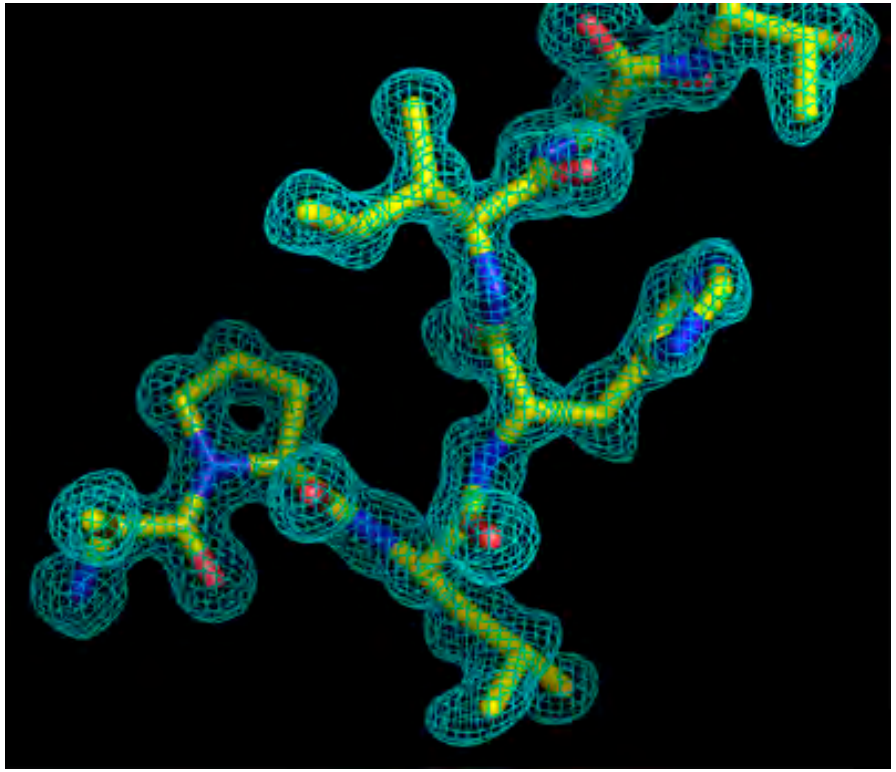
From diffraction to electron density

$$\rho(xyz) = \frac{1}{V} \sum_{hkl}^{+\infty}_{-\infty} |F(hkl)| \, . \, e^{-2\pi i[hx+ky+lz-\phi(hkl)]}$$

Amplitudes

Phases?





Diffraction data

From diffraction to electron density – the “phase” problem

Importance to deal with phase problem in X-ray crystallography:

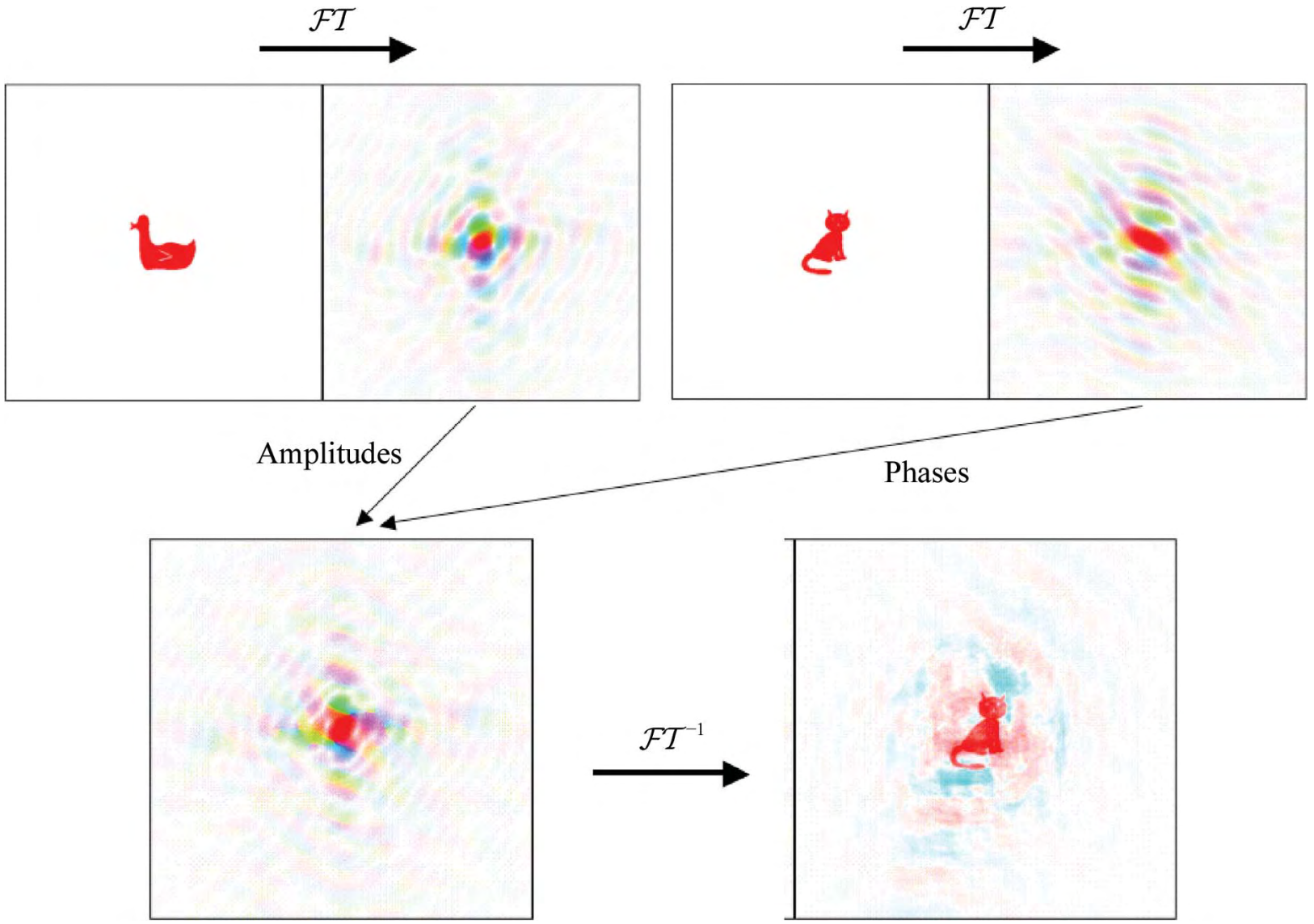
Duck: (top left)
An image of duck is take and converted into its reciprocal image by Fourier transform.

Cat: (top right)
An image of cat is take and converted into its reciprocal image by Fourier transform.

Image reconstruction: (bottom)
When using the amplitude information from “duck” and the phase information from “cat”, such hybrid operation generates an artificial reciprocal image (bottom left).

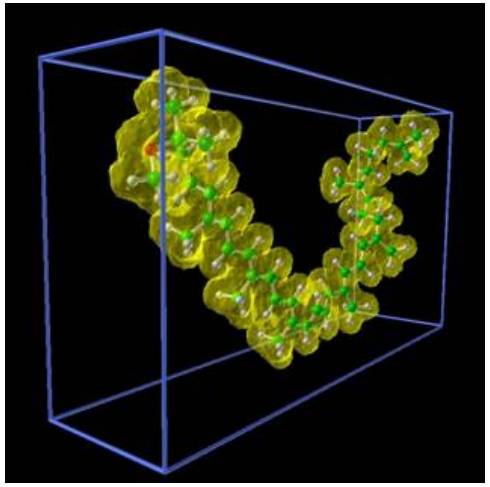
Via an inverted FT (FT^{-1}), the reconstruction results in a “cat”.

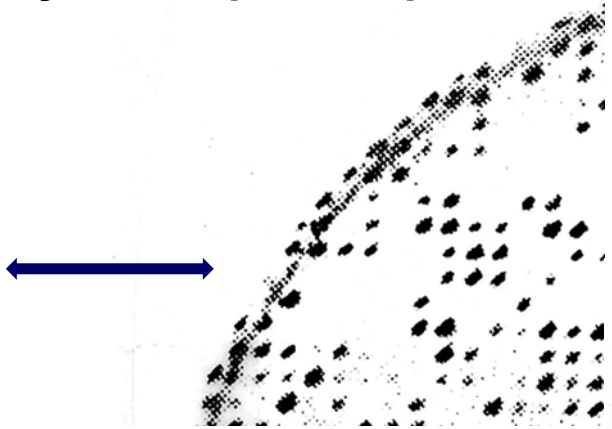
This illustrates why the “phase” information is very important in construct a model from X-ray data.

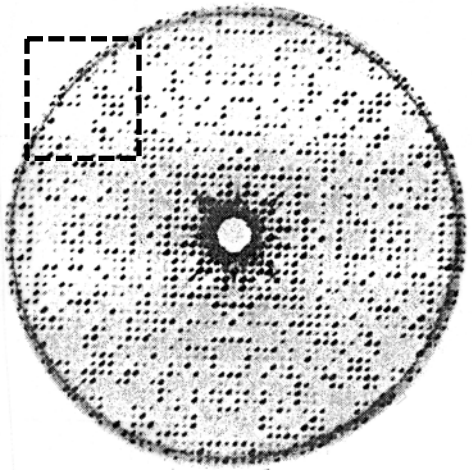


Diffraction data

From diffraction to electron density – the "phase" problem







$$\rho(xyz) = \frac{1}{V} \sum_{hkl}^{+\infty} |F(hkl)| \, . \, e^{-2\pi i[hx+ky+lz-\phi(hkl)]}$$

Amplitudes — Phases?

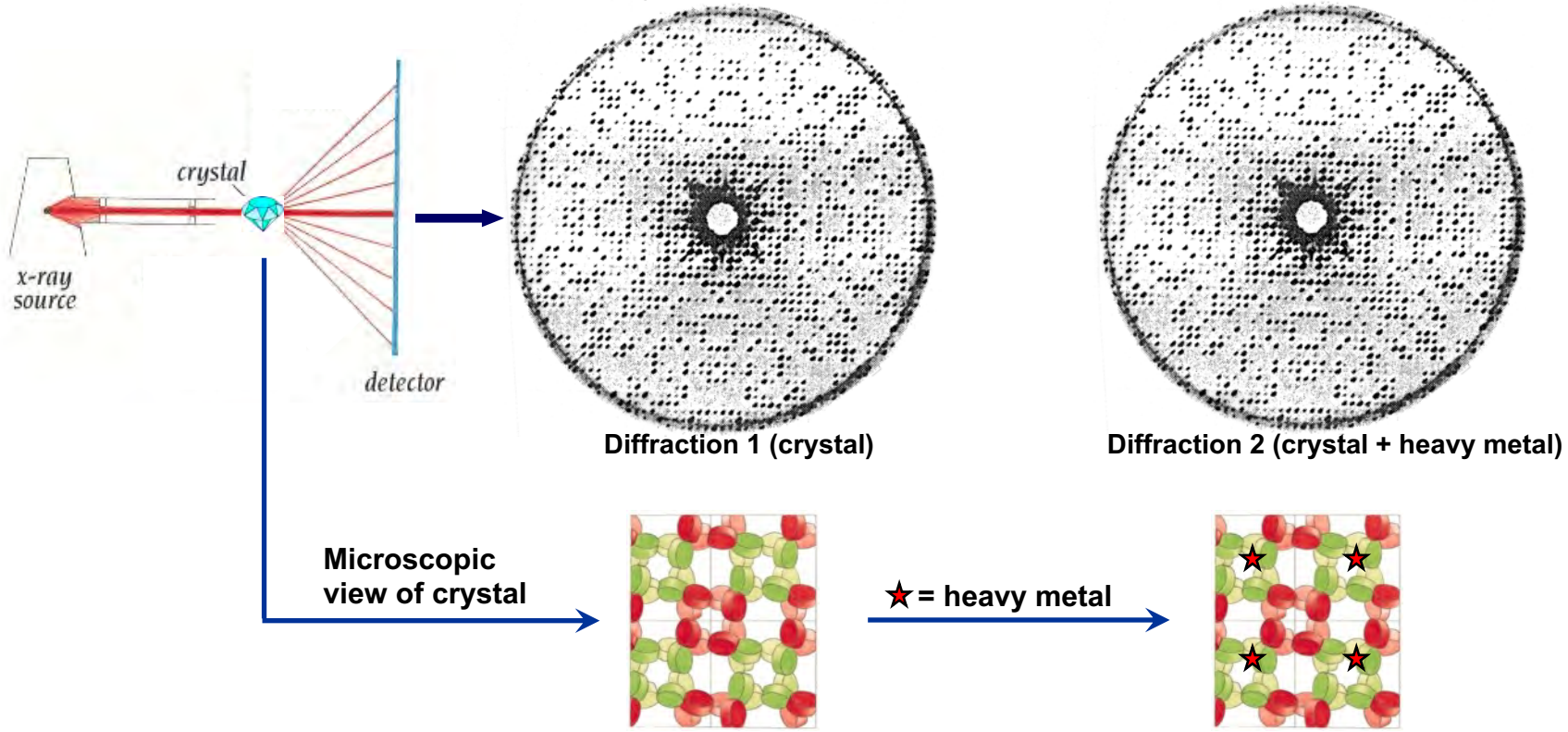
If we can define all the structure factors, $F_{(h,k,l)}$, in terms *of their λ, intensity, and phase, then we can use a mathematical formula – the Fourier transform – to transform the data into a 3D electron density map.*

We know the λ (from the x-ray source). We can define the intensity of each diffracted x-ray, or structure factor, from the intensity of the spot in the diffraction pattern.

A major problem is how to define the phase.

Diffraction data

From diffraction to electron density – the "phase" problem

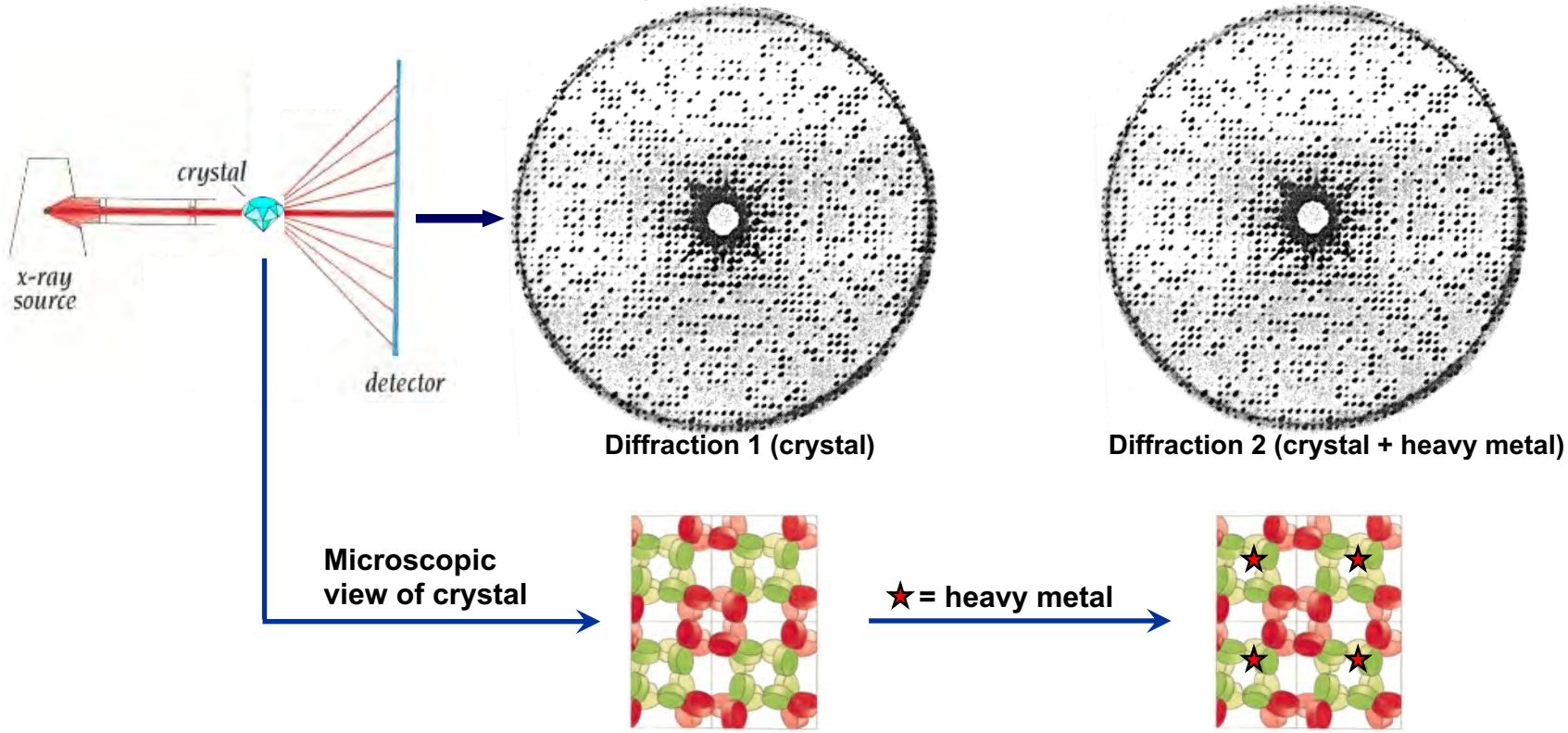


The measurement is called Isomorphous Replacement (Perutz/1956 & Kendrew/1958):

1. Record a complete diffraction data set for the protein crystal.
2. Soak the crystal in a liquid containing a heavy metal. Need the heavy metal to diffuse into the crystal and to select reproducible sites in each unit cell.
3. Record a new set of diffraction data.

Diffraction data

From diffraction to electron density – the "phase" problem

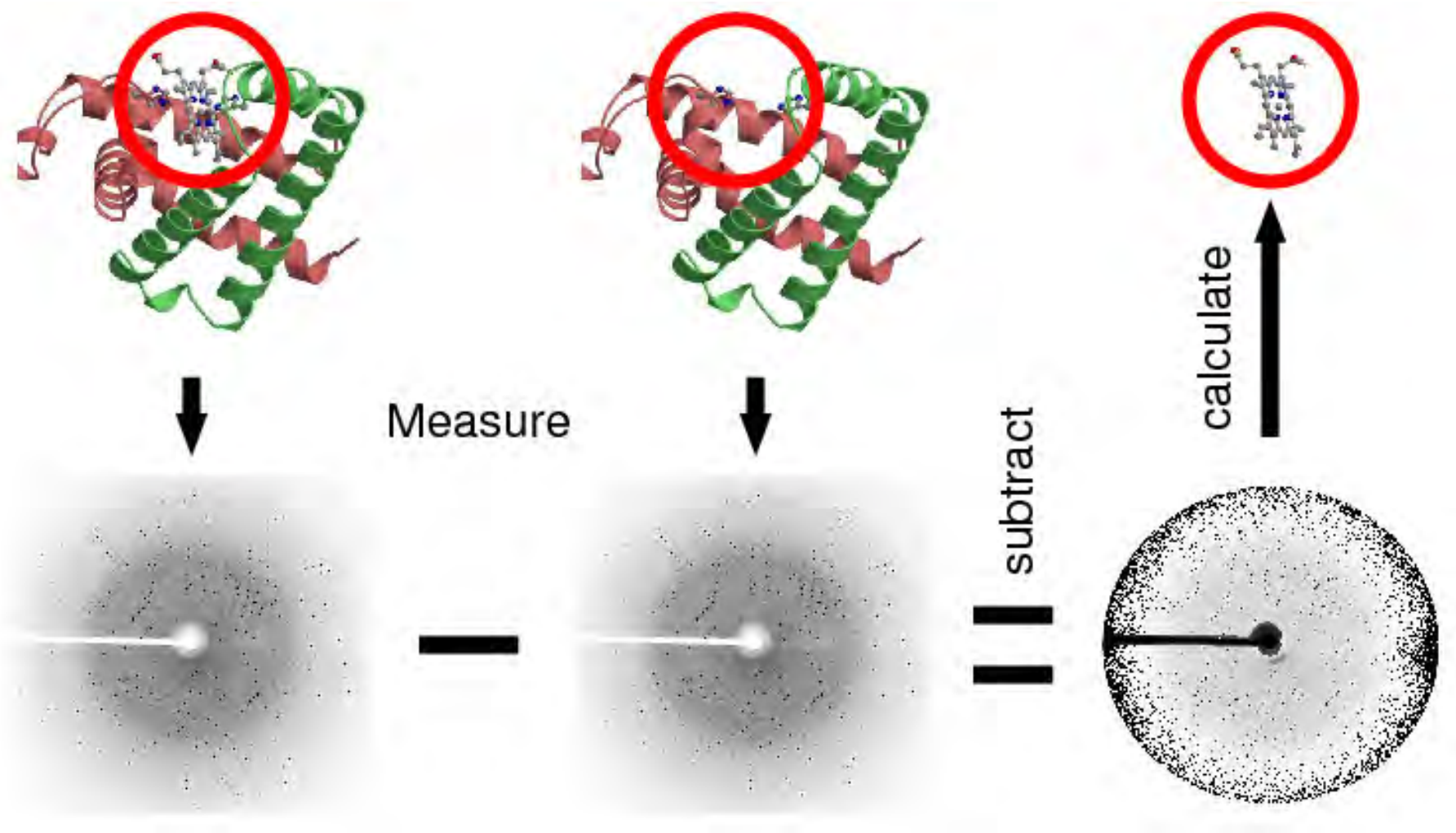


What next?

If you subtract Diffraction 1 (protein) from Diffraction 2 (protein + heavy metal), you will get the diffraction pattern for the heavy metal alone. As there is only (hopefully) one or two metals per unit cell, we can use mathematical tricks to define the phase of the heavy metal. Additional tricks can then be used to define the phase of the x-rays diffracted from the protein alone.

Diffraction data

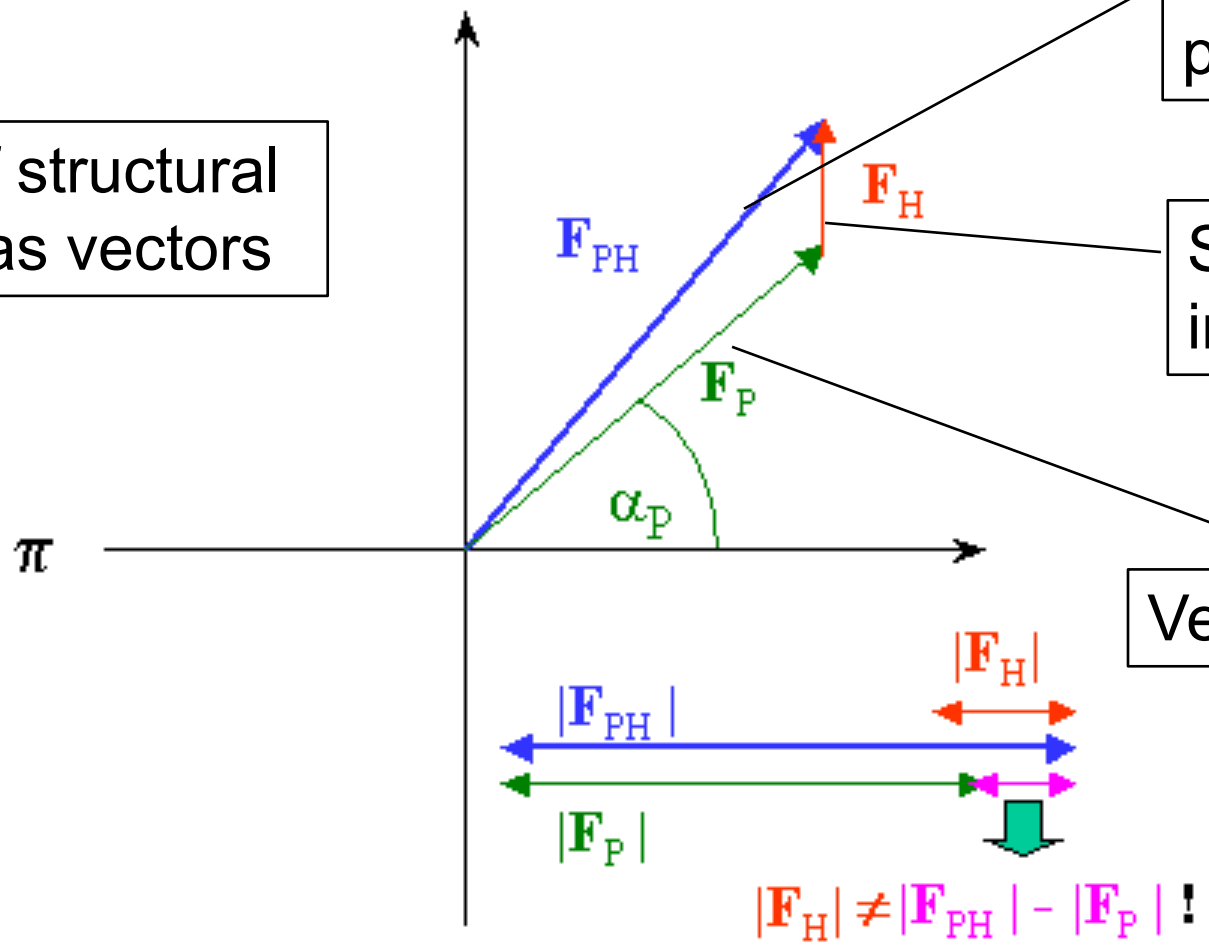
From diffraction to electron density – the “phase” problem



Diffraction data

From diffraction to electron density – the "phase" problem

Think of structural factors as vectors



Vector of protein crystals in the presence of some heavy metals

Subtracted vector representing information from heavy metals

Vector of protein crystals alone

Diffraction data

From diffraction to electron density – the "phase" problem

1) Direct method (*ab initio*): Direct methods are based on the positivity and atomicity of electron density that leads to phase relationships between the(normalized) structure factors. The requirement of what is for proteins very high-resolution data (<1.2 A) has limited the usefulness of ab initio phase determination.

2) Molecular Replacement (MR): if we know the structure of a homologous protein, we can use this information to calculate approximate phases of scattered x-rays and thus solve the phase problem.

3) Isomorphous Replacement: Soak a crystal in a solution containing a heavy metal (**SIR**) and collect a second set of diffraction patterns. Repeat with other heavy metal derivatives (**MIR**). Mathematical tricks are then used to solve the phase problem.

4) Anomalous Dispersion (or Scattering): Uses single (**SAD**) or multiple (**MAD**) wavelengths of X rays and heavy atom labeled proteins (either with a selenomethionine instead of methionine or chemical labeling) to solve phase problem. When used in conjunction with isomorphous replacement, **SIRAS** or **MIRAS**.

Acta Crystallographica Section D
Biological Crystallography

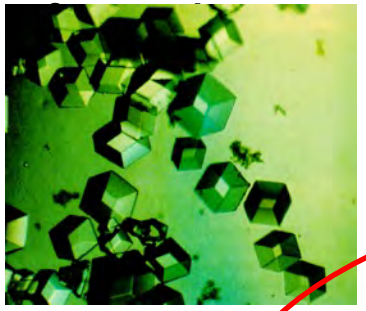
Introduction to phasing

(Garry Taylor, 2010)

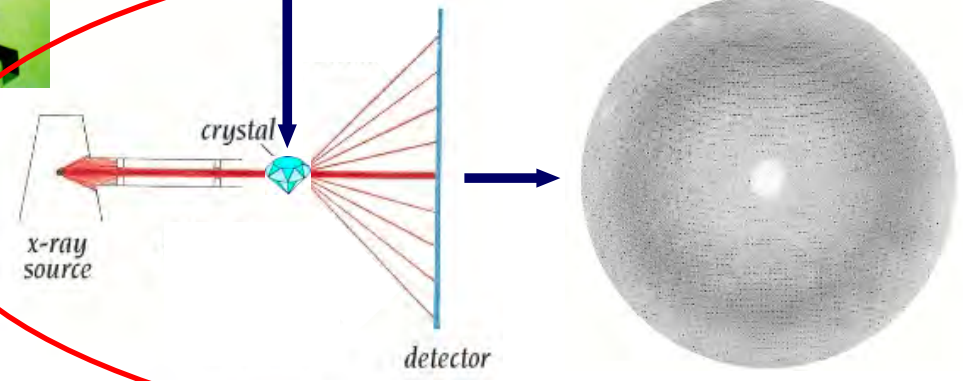
X-ray crystallography

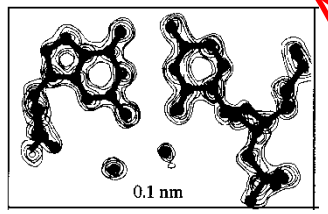
X-ray Crystallography involves 3 main steps:

Step 1: protein crystallization

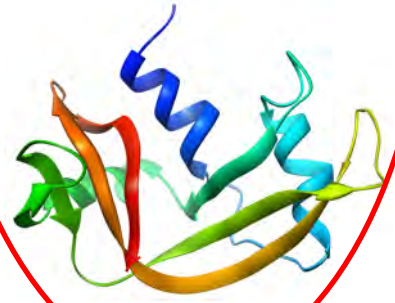


Step 2: collection and interpretation of diffraction data



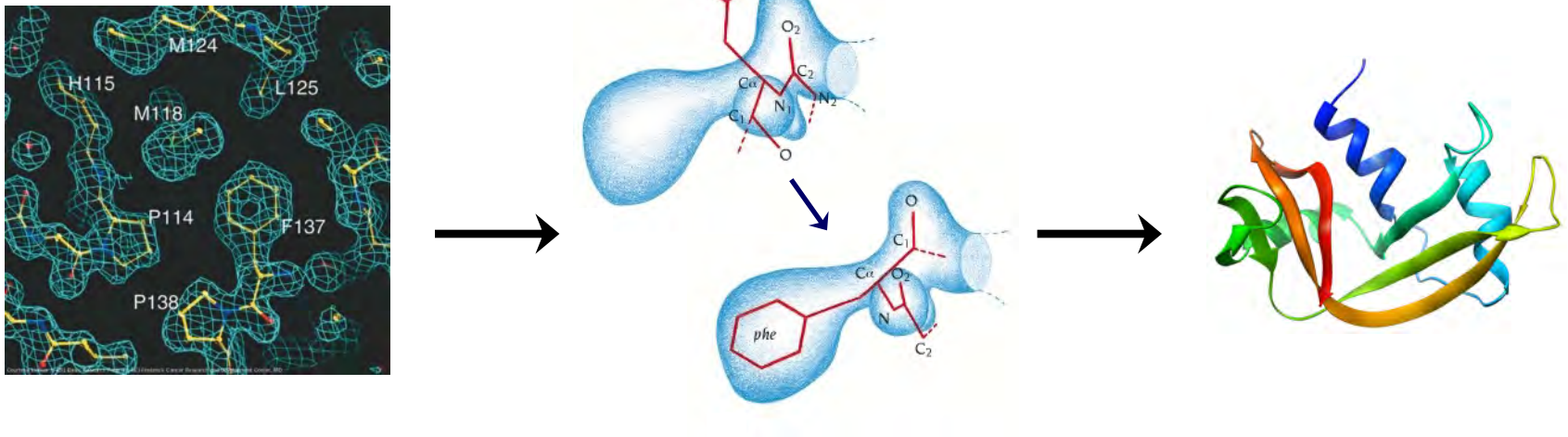


Step 3: build and refine a structural model



A structural model

From electron density to an atomic model



Electron Density Map

Fitting the aa sequence into the map

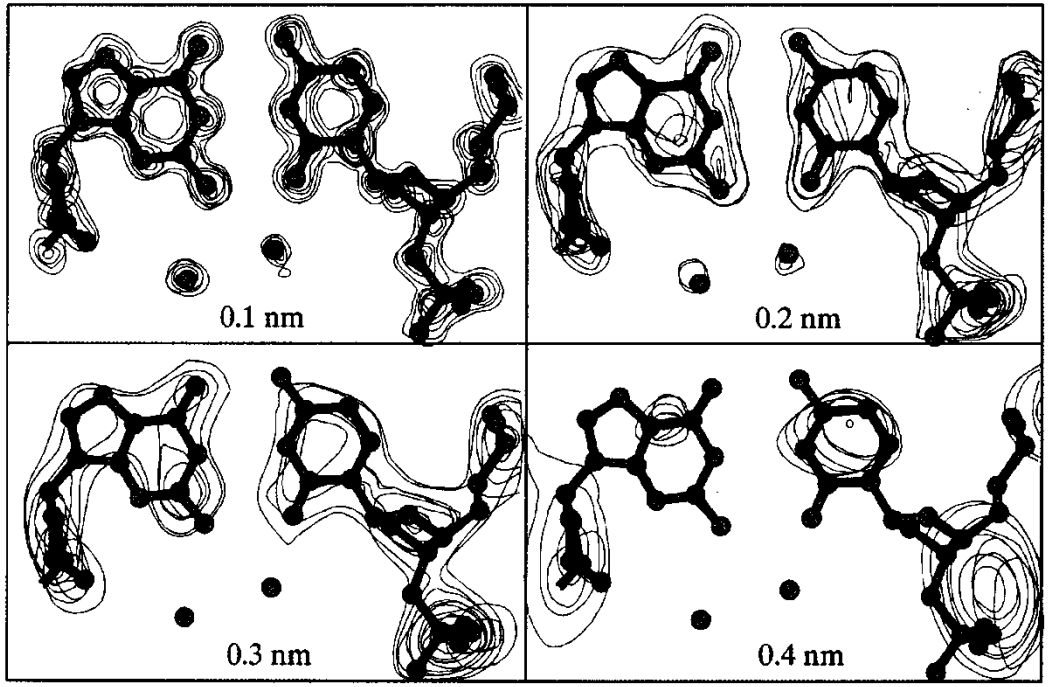
Starting Protein Model

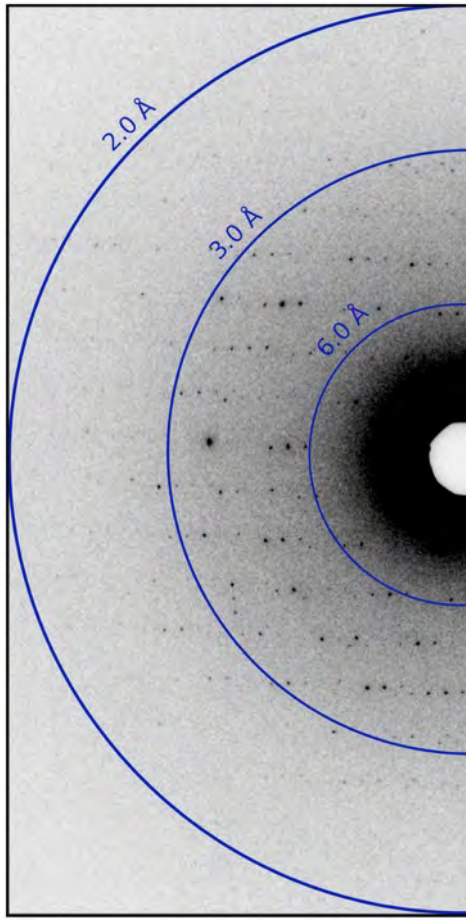
The next step is to fit the amino acid sequence of the protein into the electron density map. You do this “by hand” *in silico*, i.e. using a computer.

Once you have a reasonable fit – i.e. a starting model - you then use computer programs to rotate bond angles, adjust bond lengths, etc. to improve fit between the amino acid sequence and the electron density to arrive at the final protein model.

A structural model

From electron density to an atomic model – Parameters
i) Resolution

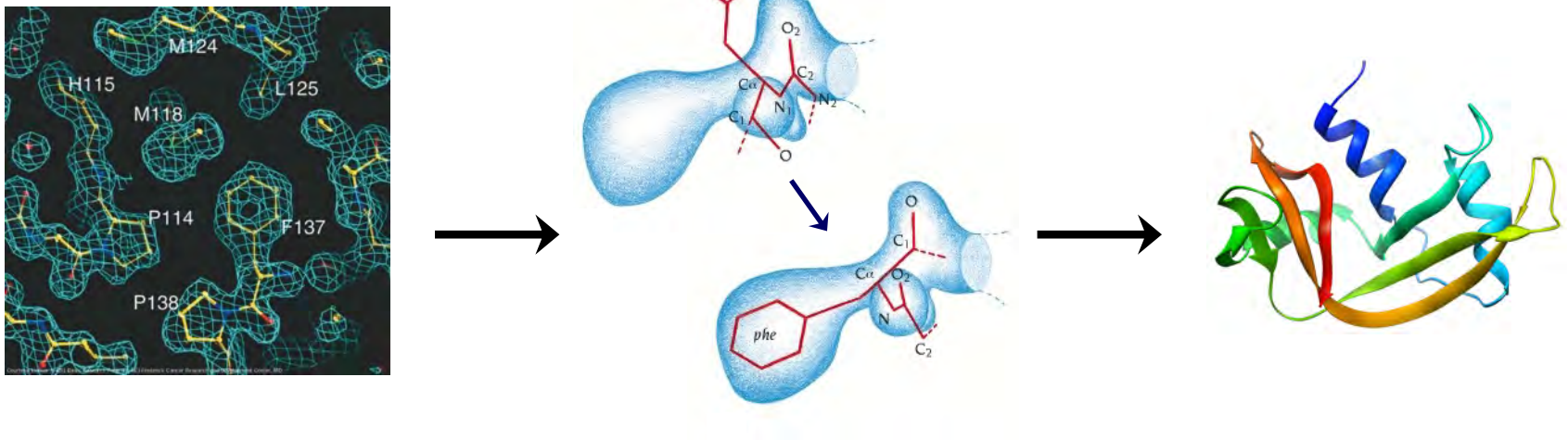




The farther from the center that you can detect spots in the diffraction pattern, the higher the resolution of the resulting electron density map, and thus the more accurate the protein model.

A structural model

From electron density to an atomic model – Parameters
ii) R-factor (accuracy of "fit")



Electron Density Map

Fitting the aa sequence into the map

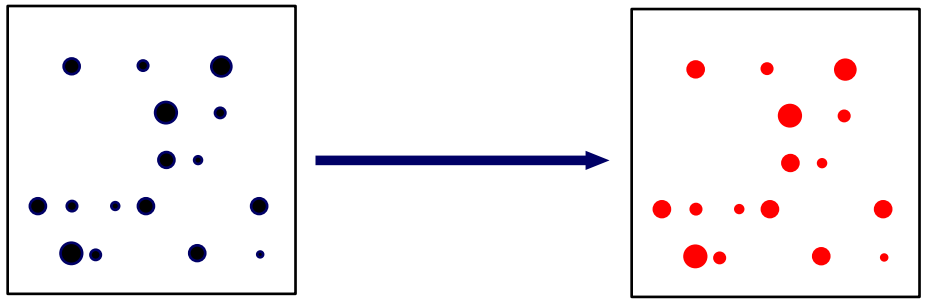
Starting Protein Model

It is not easy to fit a polypeptide chain into an electron density map. **How do we know that we have done a good job?**

The **R-factor** allows us to assess the accuracy of our model – how well the protein has been fit into the electron density map.

A structural model

From electron density to an atomic model – Parameters
ii) R-factor (accuracy of “fit”)



Experimental (observed) Diffraction Pattern

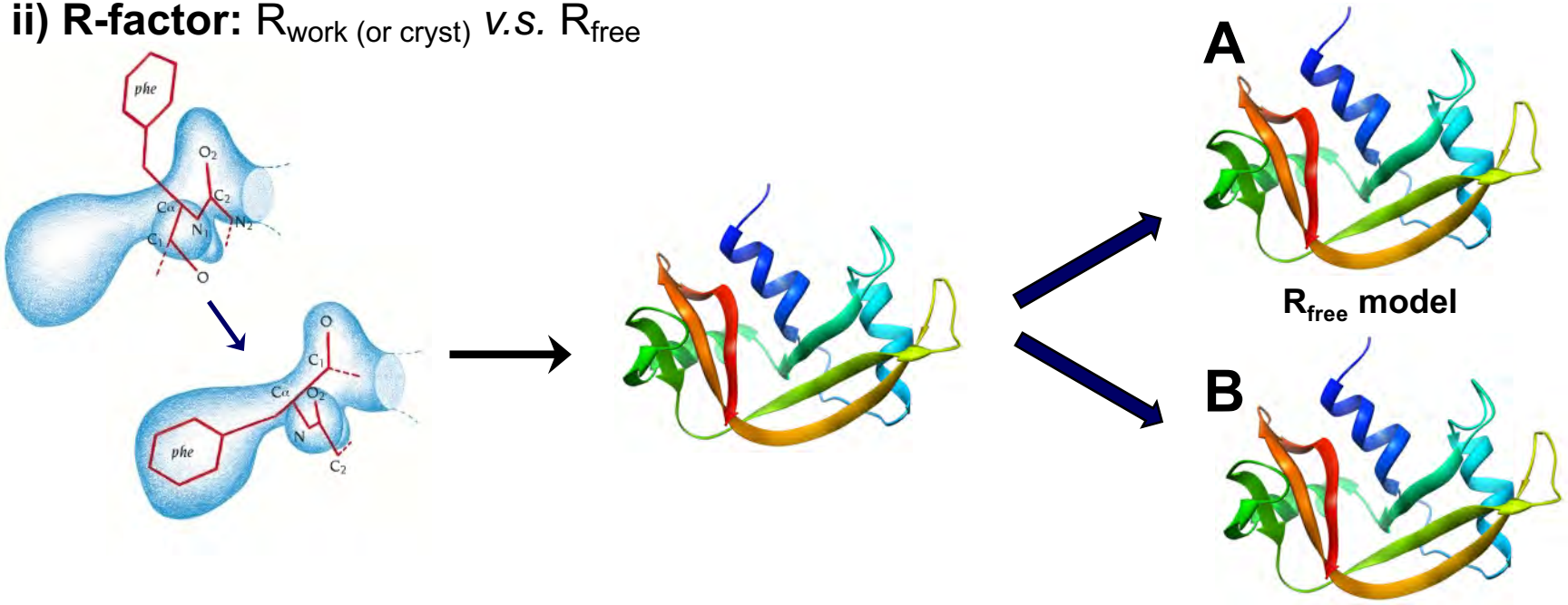
Calculated Diffraction Pattern

$$R = \frac{\sum \left| |F_{obs}| - |F_{calc}| \right|}{\sum |Fobs|}$$

A structural model

From electron density to an atomic model – Parameters

ii) R-factor: $R_{work\ (or\ cryst)}$ *v.s.* R_{free}



Fitting the amino acid sequence into the map

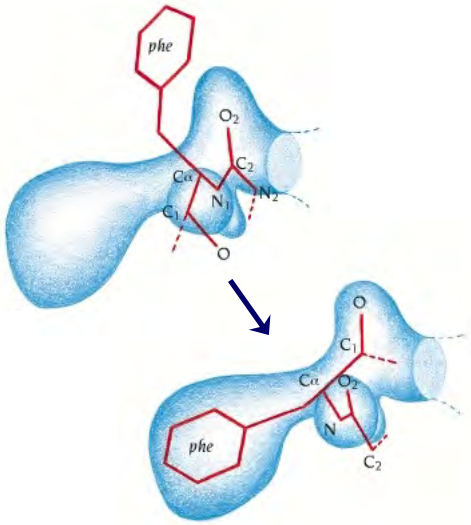
Starting Protein Model

To avoid over-fitting, we assess by setting aside a small set of data and measuring free R-factor (**R_{free}**), how well the model predicts experimental observations that are not themselves used to fit the model. Then compare with the R-factor calculated for the working set (**R_{work}**).

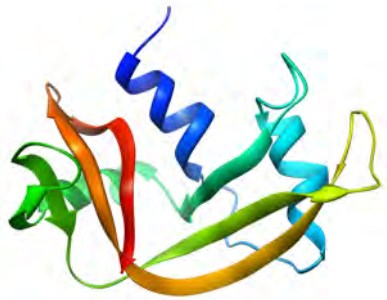
In a good model (<2.5Å), R_{free} is close in value to R_{work} (i.e. within ~10% difference).

A structural model

From electron density to an atomic model – Parameters
iii) Cross Correlation ($CC_{1/2}$)



Fitting the amino acid sequence into the map



Starting Protein Model

Model (1st half data)

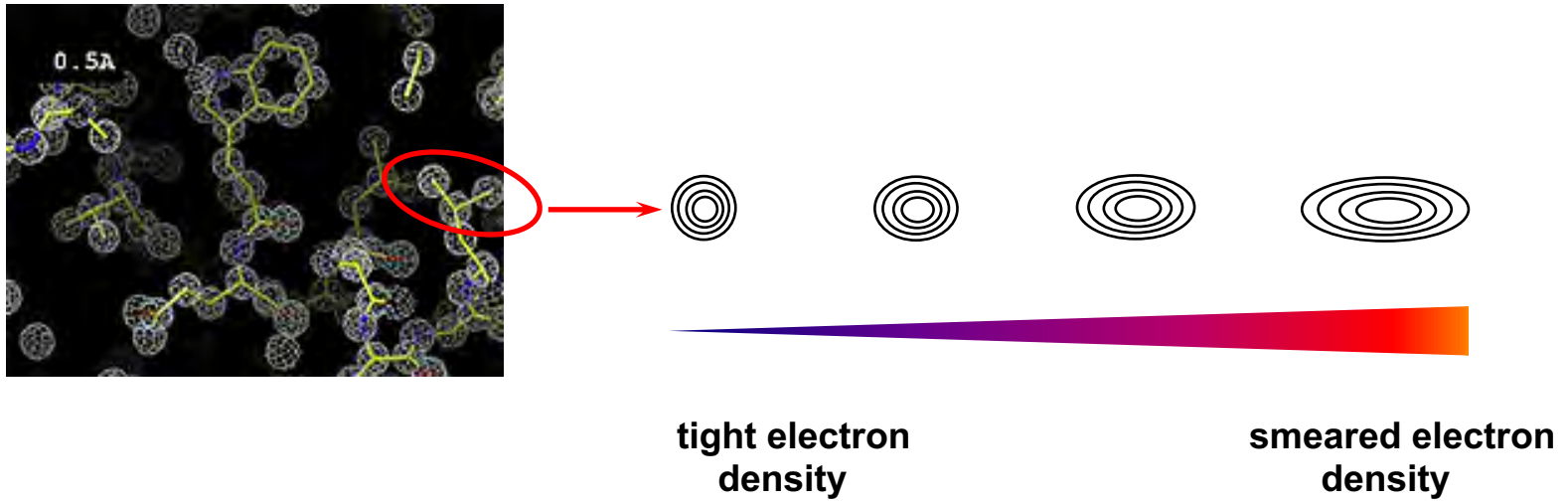
Model (2nd half data)

$$CC_{1/2} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_\varepsilon^2} = \frac{<I^2> - <I>^2}{<I^2> - <I>^2 + \sigma_\varepsilon^2}$$

$CC_{1/2}$ between intensity estimates from half data sets. Primary indicator for use for selecting high resolution cutoff for data processing. Is related to the effective signal to noise of the data

A structural model

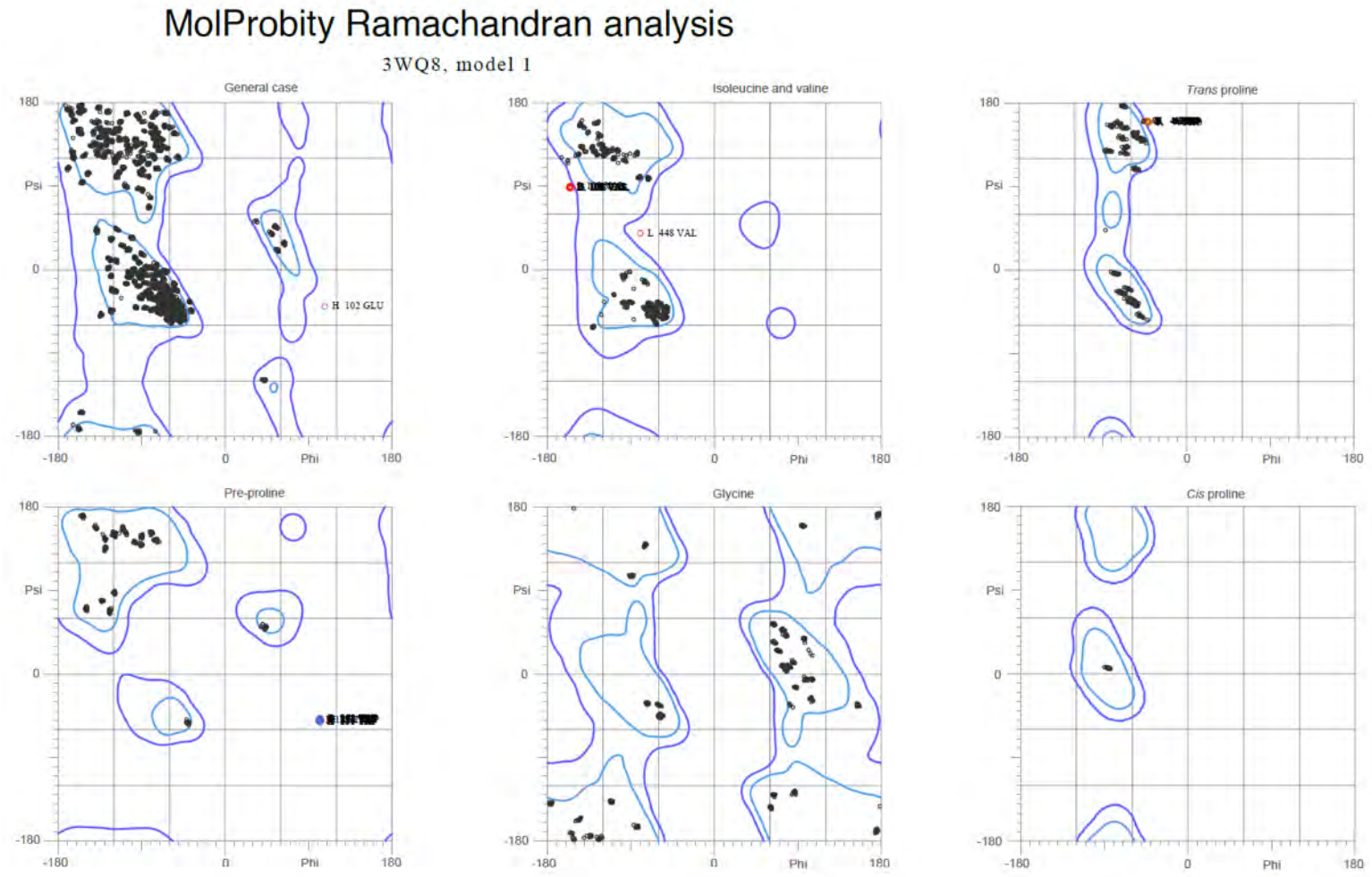
From electron density to an atomic model – Parameters
iv) B-factor (mobility of atoms)



Even in a high resolution electron density map, some atoms have tight density around each atom (often in the protein core), while others have “smeared” electron density (often on the protein surface). The smearing of density can arise due to different factors, including movement of the side chain in the crystal. The B-factor or “temperature” fact defines the amount of smearing and thus how accurately one can define the position of a side chain.

A structural model

From electron density to an atomic model – Parameters
v) Ramachandran Plots



When solving a structure, one always checks the Ramachandran Plots to ensure that residues are found within the “allowed regions”.

"Table 1" – data collection and refinement statistics

	S1V2-83-H3-MSK-1999 Fab-HA head complex	S8V2-18-H1-X181-2009 Fab-HA head complex	S8V2-37-H3-TX-2012 Fab-HA head complex	S8V2-47-H1-BJ-1995 Fab-HA head complex
PDB ID	6XPZ	6XQ0	6XQ2	6XQ4
Data Collection	APS 24-ID-C	APS 24-ID-E	APS 24-ID-C	APS 24-ID-C
Number of datasets	1	1	1	1
Resolution, Å	48.59 - 3.45 (3.573 - 3.45)	43.95 - 2.3 (2.38 - 2.3)	48.8 - 3.00 (3.11 - 3.00)	48.22 - 3.35 (3.47 - 3.35)
Wavelength (Å)	0.9791	0.9792	0.9791	0.9791
Space Group	P 21 2 21	C 1 2 1	C 1 2 1	P 1 21 1
Unit cell dimensions (a, b, c), Å	85.79, 103.71, 220	160.2, 55.26, 168	294.8 42.38 155.76	96.09 66.47 135.18
Unit cell angles (α, β, γ) °	90, 90, 90	90, 95.02, 90	90 111.50 90	90 101.96 90
I/σ	9.0 (9.2)	8.03 (1.34)	9.71 (1.28)	9.38 (2.10)
Rmeas	0.22 (1.22)	0.13 (0.83)	0.17 (1.14)	0.07 (0.43)
Rpim	0.07 (0.40)	0.08 (0.47)	0.08 (0.54)	0.05 (0.30)
Rmerge, %	21.20 (114.90)	10.86 (67.38)	15.30 (99.94)	5.19 (30.33)
CC*	1.00 (0.96)	1.00 (0.90)	1.00 (0.94)	1.00 (0.93)
CC½	1.00 (0.84)	0.99 (0.68)	0.99 (0.79)	1.00 (0.77)
Completeness, %	99.49 (99.08)	95.24 (96.14)	98.32 (98.09)	96.60 (97.92)
Number of reflections	239288 (23792)	177025 (17199)	149811 (14847)	46537 (4652)
Redundancy	9.0 (9.2)	2.8 (2.8)	4.1 (4.2)	2.0 (2.0)
Refinement				
Number of reflections:				
Working	26450 (2579)	62570 (6230)	36261 (3545)	23634 (2357)
Free	1325 (117)	3038 (306)	1881 (180)	1142 (110)
Rwork, %	25.11 (34.42)	22.30 (31.35)	24.72 (34.08)	23.79 (33.65)
Rfree, %	30.84 (39.73)	26.71 (37.27)	29.07 (36.02)	28.04 (43.05)
Ramachandran plot, % (favored, disallowed)	94.27 (0.21)	97.25 (0.08)	94.6 (0.07)	94.18 (0.16)
Rmsd bond lengths, Å	0.002	0.005	0.002	0.002
Rmsd bond angles, °	0.52	1.17	0.56	0.6
Average B-factor	140.07	43.35	125.44	102.4

(McCarthy et al, mBio, 2021)