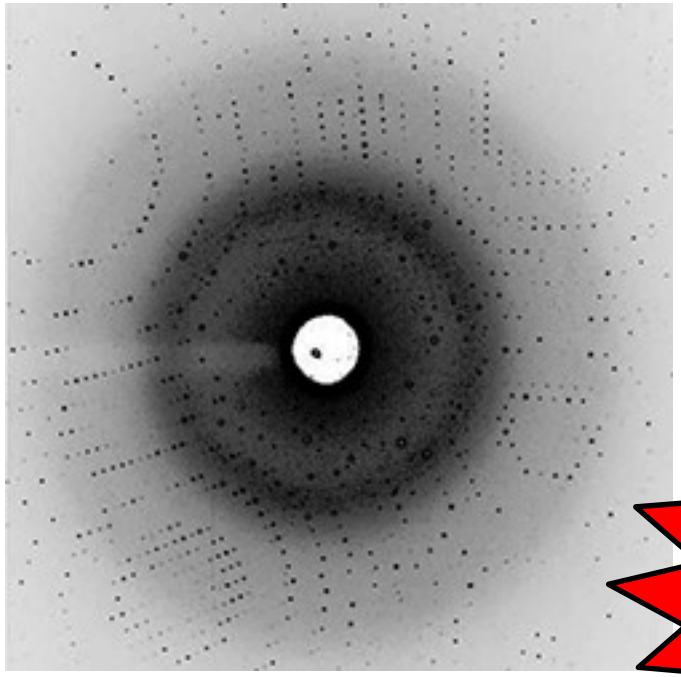
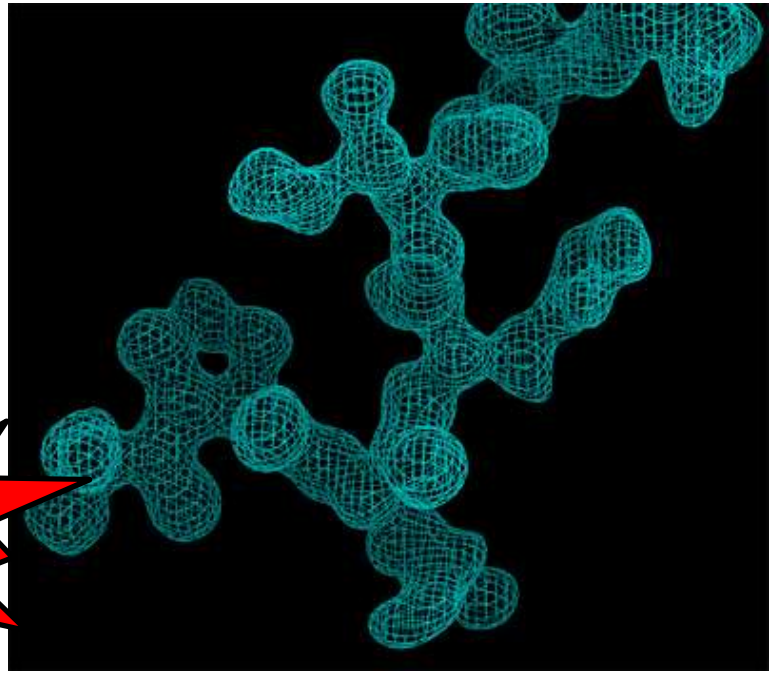
The phase problem



FT^{-1}

ϕ_{hkl}?



$I_{hkl} \propto F_{hkl}^2$

ρ_{xyz}

Structure factors

In the diffraction experiment, X-ray radiation interacts with electrons present in the sample. The incident beam is scattered around the atom.

For a single atom:

$f_{2\theta} = f_{\boldsymbol{S}} = \int_{\boldsymbol{r}}^{V} \rho(\boldsymbol{r}) \exp(2\pi i\, \boldsymbol{S} \cdot \boldsymbol{r})\, d\boldsymbol{r}$ with $\rho(\boldsymbol{r})$ electron density of the atom

To describe the diffracted beam, interference between diffracted beams has to be considered. In a crystal, due to periodicity conditions of the lattice, interference of the diffracted beams is constructive only in discrete directions (Laue conditions).

For the unit cell of the crystal:

$$\boldsymbol{F_h} = \sum_j^N n_j f_{\boldsymbol{h},j}^0 \exp\left(-B_j (\sin\theta/\lambda)^2\right) \exp\left(2\pi i \boldsymbol{h} \cdot \boldsymbol{x}_j\right)$$

with $f_{\boldsymbol{h},j}^0$ scattering factor of the specific atom,

B_j atomic displacement of the atom, n_j occupancy of the atom,

$\boldsymbol{x}_j$ position of the atom, and

$\boldsymbol{h}$ vector defining a node in the reciprocal lattice (and associated to the family of planes that generated the reflection according to Bragg's law).

Therefore, considering the electron density of the whole unit cell:

$$\boldsymbol{F_h} = \int_V \rho(\boldsymbol{r}) \exp(2\pi i \boldsymbol{h} \cdot \boldsymbol{r})\, d\boldsymbol{r} = FT[\rho(\boldsymbol{r})]$$

Fourier Transform

The Fourier Transform (FT) is a mathematical transformation between a space (real space) and another space (of frequencies, reciprocal space).

Given a function $f(x)$ defined in the real space *R*, the Fourier Transform $g(x^*)$ is defined in the reciprocal space *R**:

$$FT[f(x)] = g(x^*) \quad \text{and} \quad FT^{-1}[g(x^*)] = f(x)$$

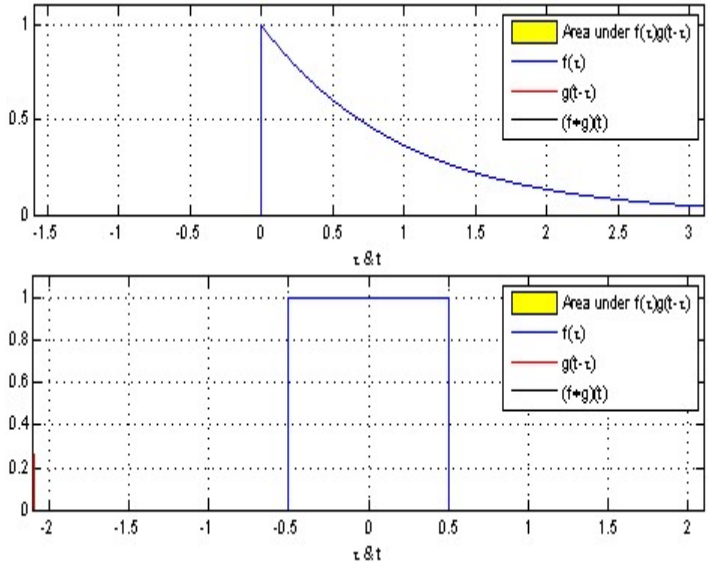
$$g(x^*) = \int_R f(x) \exp(2\pi i\, x^* \cdot x)\, dx$$

$$f(x) = \int_{R^*} g(x^*) \exp(-2\pi i\, x^* \cdot x)\, dx^*$$

Convolution

The convolution is a mathematical operation between two functions that generates a third function defined in the same space:

$$Conv(u) = f(x) \otimes g(x) = \int_x f(x) g(u - x) dx$$



Properties of the Fourier Transform

1) Fourier convolution theorem: the Fourier Transform of the convolution between two functions is the product of the Fourier Transform of each function:

$$FT[f(x) \otimes g(x)] = FT[f(x)] \cdot FT[g(x)]$$

2) The Fourier Transform of a Gaussian function is a Gaussian function with inverse variance.

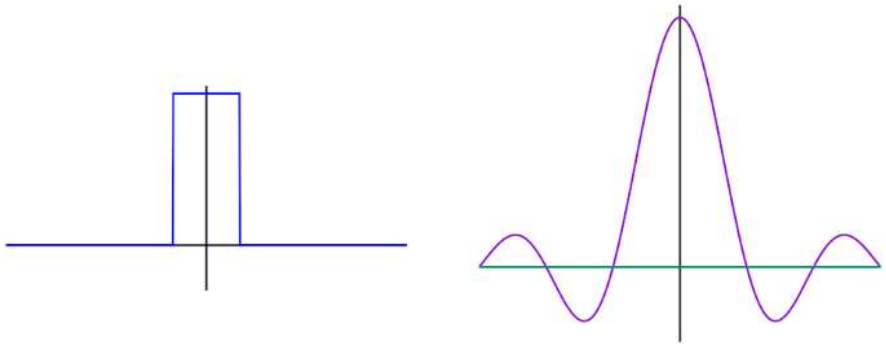
   i.e. A broad Gaussian is transformed in a sharp Gaussian.

3) The Fourier Transform of a constant function is a delta function centered in the origin.

   i.e. The Fourier Transform of a constant is zero everywhere except in the origin.

4) The Fourier Transform of a periodic delta function is a periodic delta function with inverse periodicity.

5) The Fourier Transform of a step function generates a peak with ripples:



Lattice function

The crystal can be expressed as a convolution of the electron density of the unit cell, $\rho_{uc}(\boldsymbol{r})$, and a lattice function:

$$\rho(\boldsymbol{r}) = \rho_{uc}(\boldsymbol{r}) \otimes L(\boldsymbol{r})$$

The lattice function $L(\boldsymbol{r})$ has the form of a periodic delta function with the maxima located in the nodes of the lattice.

According to the Fourier Transform properties:

$$\boldsymbol{F_S} = FT[\rho(\boldsymbol{r})] = FT[\rho_{uc}(\boldsymbol{r}) \otimes L(\boldsymbol{r})\,] = FT[\rho_{uc}(\boldsymbol{r})] \cdot FT[L(\boldsymbol{r})] = FT[\rho_{uc}(\boldsymbol{r})] \cdot L^*(\boldsymbol{S})$$

The lattice function $L^*(\boldsymbol{S})$ describes the reciprocal lattice.

- *As previsouly discussed, the presence of the lattice repetition induces a discretization of the reciprocal space: the function $\boldsymbol{F_S}$ has a non-zero value only in points of the reciprocal lattice.*
- *The reciprocal lattice has an inverse periodicity compared to the periodicity of the real lattice: a larger unit cell in the real lattice corresponds to a smaller unit cell in the reciprocal lattice.*

The phase problem

According to the definition of the Fourier Transform:

$$\boldsymbol{F_S} = FT[\rho(\boldsymbol{r})] \quad \text{and} \quad \rho(\boldsymbol{r}) = FT^{-1}[\boldsymbol{F_S}]$$

Therefore, to determine the atomic structure of the protein in the real space, we can apply the inverse Fourier Transform to the structure factors:

$$\rho(\boldsymbol{r}) = \int_{R^*} \boldsymbol{F_S} \exp(-2\pi i\, \boldsymbol{S} \cdot \boldsymbol{r})$$

Using the crystallography fractional coordinates and considering that the reciprocal space is discrete:

$$\rho(\boldsymbol{x}) = \frac{1}{V} \int_{\boldsymbol{h}=-\infty}^{+\infty} \boldsymbol{F_h} \exp(-2\pi i\, \boldsymbol{h} \cdot \boldsymbol{x}) = \frac{1}{V} \sum_{\boldsymbol{h}=-\infty}^{+\infty} \boldsymbol{F_h} \exp(-2\pi i\, \boldsymbol{h} \cdot \boldsymbol{x})$$

The structure factor is a complex number, with a real component (*A*) and an imaginary component (*B*). It can also be expressed by the modulus *F* and the phase α:

$$\boldsymbol{F} = F\, e^{i\alpha}$$

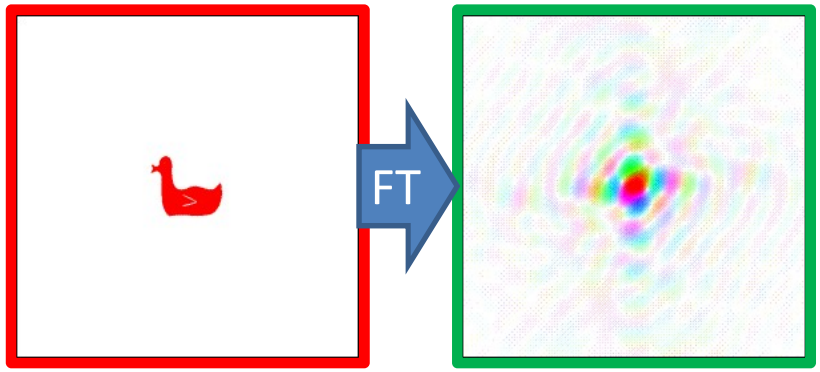
$$\rho(\boldsymbol{x}) = \frac{1}{V} \sum_{\boldsymbol{h}=-\infty}^{+\infty} F_{\boldsymbol{h}} \exp(i\alpha_{\boldsymbol{h}}) \exp(-2\pi i\, \boldsymbol{h} \cdot \boldsymbol{x}) = \frac{2}{V} \sum_{h=0}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} F_{\boldsymbol{h}} \cos(\boldsymbol{h} \cdot \boldsymbol{x} - \alpha_{\boldsymbol{h}})$$

While intensities of the diffracted spots yield moduli of the structure factors, phases cannot be retrieved from the diffraction experiment.

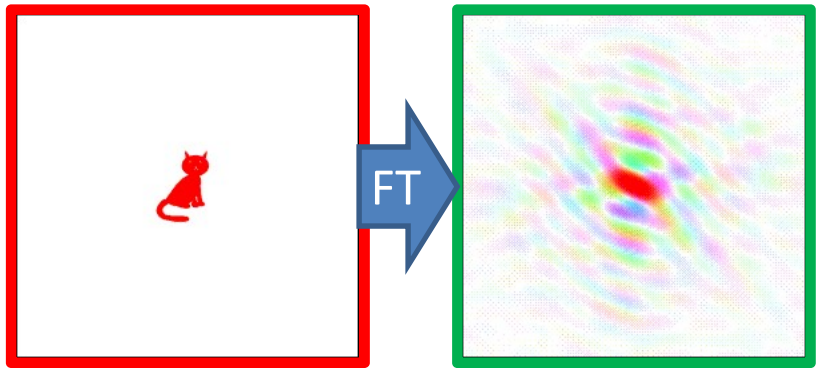
Phase and amplitude

What is the contribution of each of these terms to the electron density?

Consider the representation of a duck and its Fourier Transform:

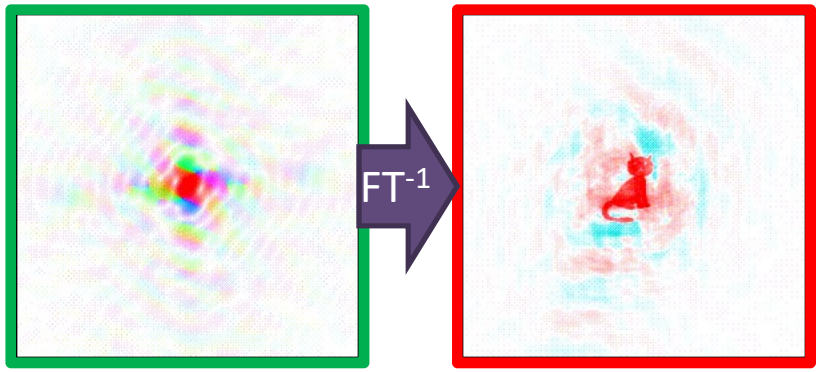


Consider also the representation of a cat and its Fourier Transform:



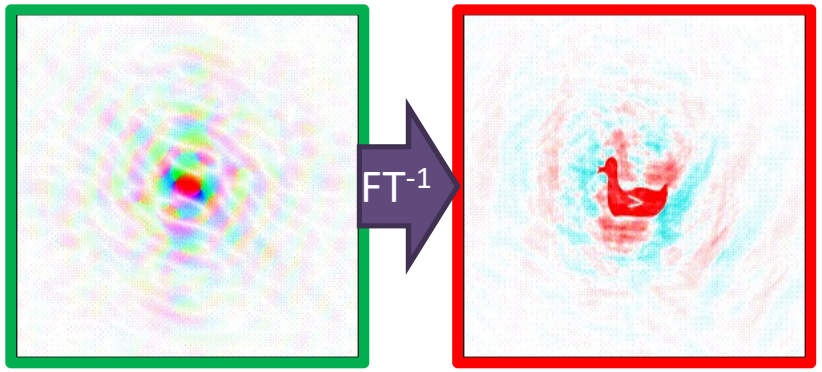
Let's mix!

Amplitudes (shape of the function) of the duck, with the *phases* (color) of the *cat*



Real space

Amplitudes of the cat, with the *phases* of the *duck*



Reciprocal space

Resolution

The summation that yields the electron density covers the entire reciprocal space (from $-\infty$ to $+\infty$). However, experimental intensities of the diffracted beams are limited by attenuation effects (at high resolution) and experimental conditions.

What effect has the truncation of the summation on the electron density?

Truncation is equivalent to multipling the structure factor function with a square function, that is 1 where the structure factors are present and 0 for reflections not collected or with zero intensity.

If the truncation is due to attenuation effects of the signal (e.g. atomic displacement):

Broadening of the peaks in the electron density.

If the truncation is due to experimental conditions that did not allow the collection of high resolution reflections:

Ripples in the electron densities, particularly strong around heavy atoms.

Data collection issues

- Random missing reflections in the reciprocal space due to the presence of ice rings (excluded during integration), missing detector edges, randomly removed data for validation...

  No significant changes in the electron density.

- Sistematically missing strong reflections (particularly at low resolution) due to overload of detector and insuffcient dynamic range

  FATAL ERROR! Phasing strongly hampered and noisy electron density map.

- Missing data wedge due to wrong identification of crystal symmetry or unsuitable geometry of the data collection

  Serious error! Noisy electron density map and streaking in density.

- Structural disorder (isotropic) due to flexible groups, imperfect crystal, high thermal motion of atoms...

  Reduction of resolution in the diffraction space. High resolution reflections become too weak.

- Anisotropic structural disorder due to anisotropy in packing interactions

  Anisotropic resolution limit.

Patterson function

The Patterson function is defined as the autoconvolution of the electron density in the real space:

$$P(\boldsymbol{u}) = \rho(\boldsymbol{x}) \otimes \rho(-\boldsymbol{x}) = \int_R \rho(\boldsymbol{x})\rho(\boldsymbol{u}-\boldsymbol{x})d\boldsymbol{x}$$

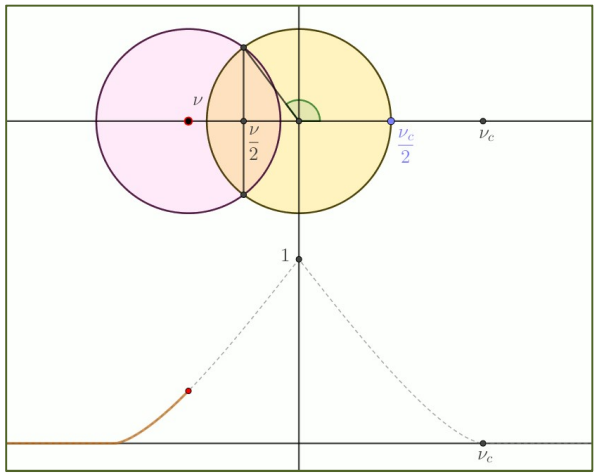
From the diffraction intensities:

$$I_{\boldsymbol{h}} \propto \boldsymbol{F}_{\boldsymbol{h}} \cdot \boldsymbol{F}_{\boldsymbol{h}}^{*} = FT[\rho(\boldsymbol{x})] \cdot FT[\rho(-\boldsymbol{x})] = FT[\rho(\boldsymbol{x}) \otimes \rho(-\boldsymbol{x})] = FT[P(\boldsymbol{u})]$$

The Patterson function, therefore, can be computed as the inverse Fourier Transform of the diffraction intensities:

$$P(\boldsymbol{u}) = FT^{-1}[I_{\boldsymbol{h}}]$$

The autoconvolution can be visualized as a window containing the function sliding over the same function. For each sliding vector ($\boldsymbol{u}$), functions are multiplied.



The maxima of the autoconvolution appear when the sliding vector brings a peak of the first function onto a peak of the second. This happens in the origin (high peak due to the product of each peak with itself) and when the vector $\boldsymbol{u}$ is an interatomic vector.

The peaks of the Patterson function are interatomic vectors.

Matthews coefficient

How many copies of the protein are present in the unit cell?

1) How many copies of the asymmetric unit are present in the unit cell?
   From symmetry elements of the space group: z

2) Volume of the asymmetric unit:
   $V_a = V/z$

3) In protein structures, the volume occupied by solvent molecules is between 20% and 80% of the cell volume
   Usually, protein crystals diffracting at higher resolution contain a smaller amount of solvent...

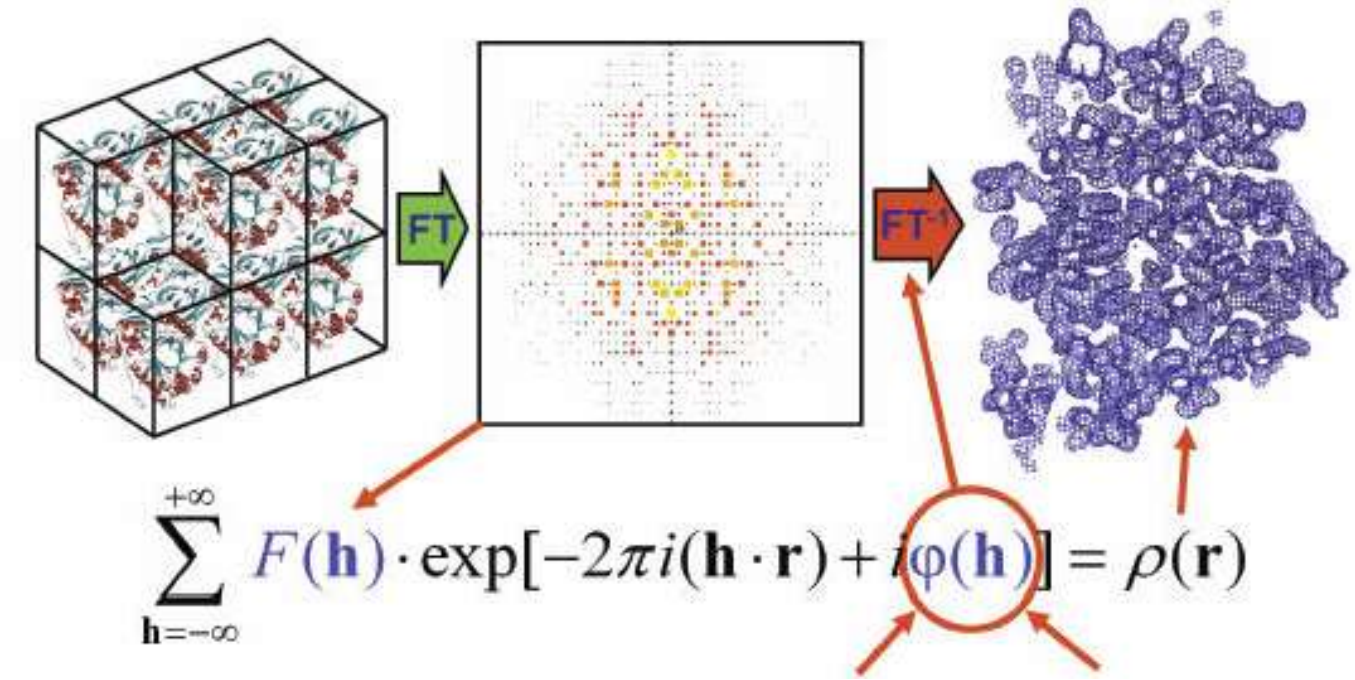
4) The **Matthews volume** can be computed as: $\boldsymbol{V_M = V_a / MW_{asymmetric\ unit}}$
   The Matthews volume is expressed in $\text{Å}^3/Da$ and should have a value between $\boldsymbol{1.5\ \text{Å}^3/Da < V_M < 6\ \text{Å}^3/Da}$.

5) Considering the Matthews coefficient, the number of protein molecules in the asymmetric unit can be obtained.

6) More sofisticated analysis express the **Matthews probability**, or the conditional probability distribution of Matthews coefficients in PDB, given a specific resolution.

Phasing

To calculate electron density in real space, both structure factor moduli and phases are required.



Moduli are obtained directly from the diffraction experiment ($F_{hkl} \propto \sqrt{I_{hkl}}$).

Phases need to be provided by different methods (phasing): ϕ_{calc}

If moduli and phases are known, **electron density maps** can be calculated and the atomic structure can be determined in the electron density map:

$$\rho(\boldsymbol{r}) = FT^{-1}(F_{obs}, \phi_{calc})$$

Phasing methods

For small molecules:

- Heavy atom methods, based on Patterson function
- Direct methods

For proteins (lower resolution, more atoms!):

- **Molecular replacement (MR)** methods: use a molecular search probe in the unit cell
- Marker atom substructure methods:
  - Methods based on heavy atom isomorphous derivatives:

    Single Isomorphous Replacement (SIR)

    Multiple Isomorphous Replacement (MIR)
  - Methods based on anomalous dispersion:

    Single-wavelength Anomalous Dispersion (SAD)

    Multiple-wavelength Anomalous Dispersion (MAD)
- Density modification methods
- Mixed methods