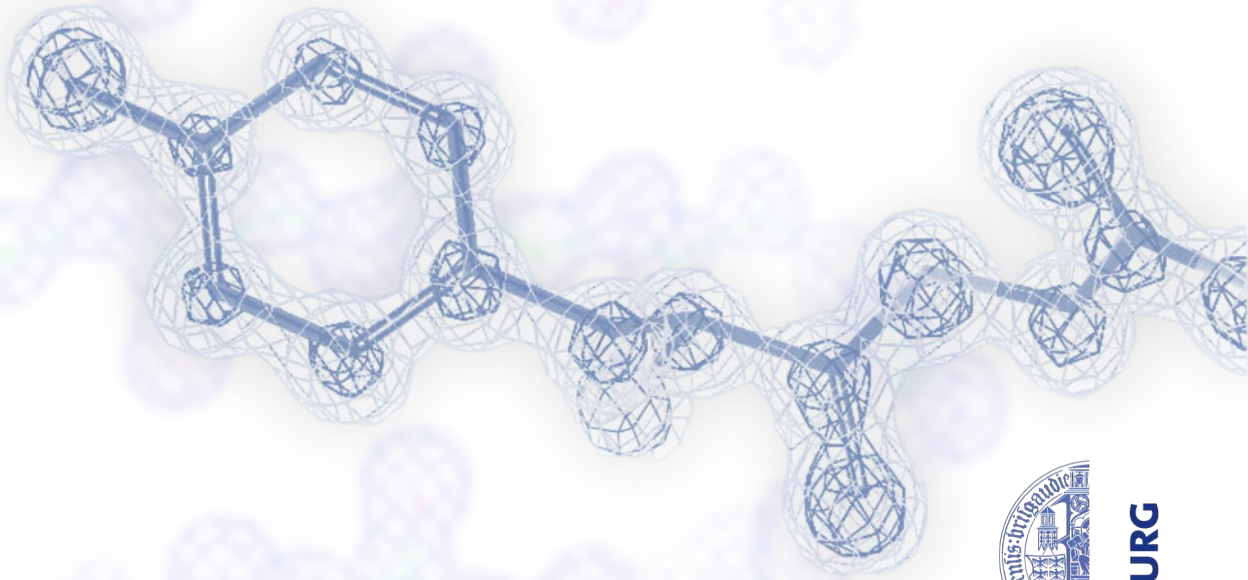
Protein X-ray Crystallography

Oliver Einsle

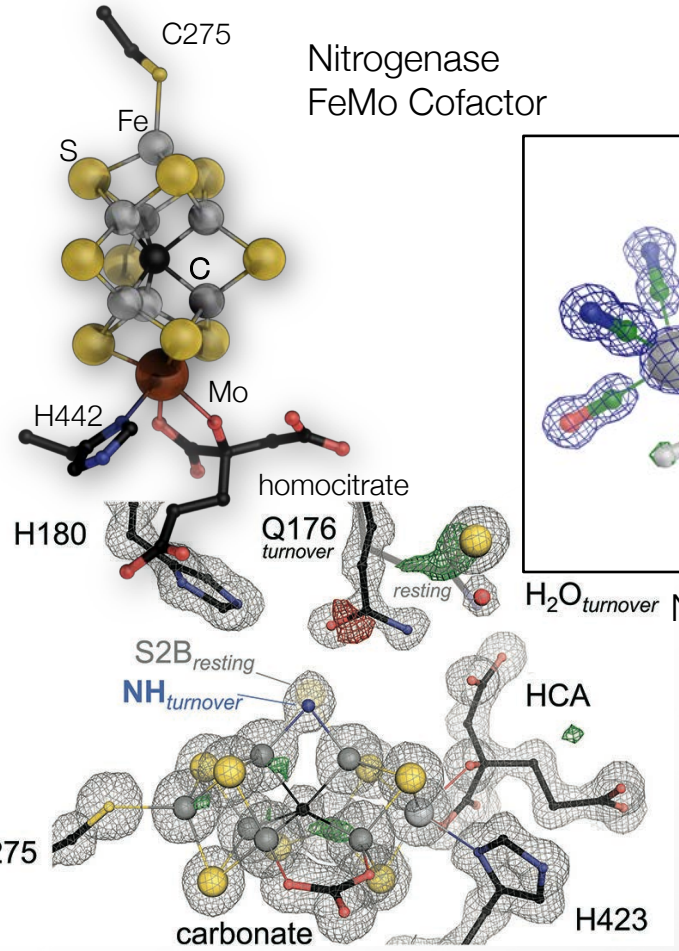


5th Penn State Bioinorganic Workshop 2018

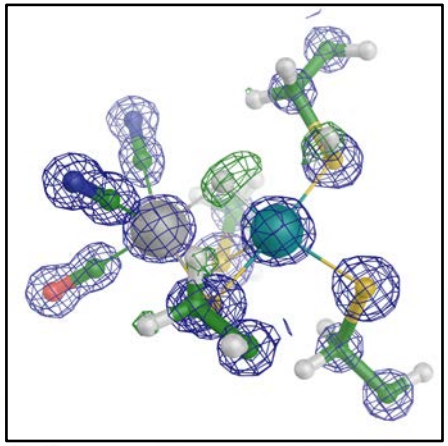
UNI FREIBURG

The Role of Structure in Bioinorganic Chemistry

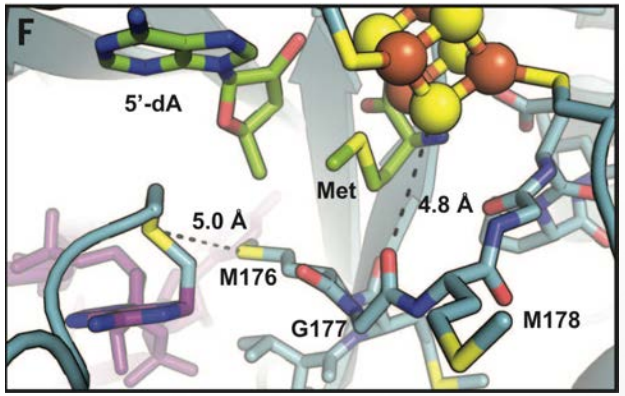
⇒ Many (most) bioinorganic active sites are non-standard (contrary to polypeptides of the 20 amino acids).



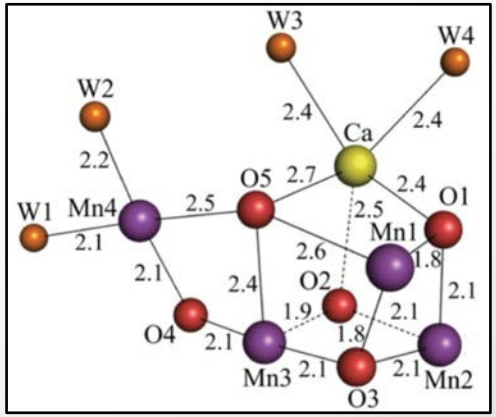
Nitrogenase
FeMo Cofactor



$H_2O_{turnover}$ NiFe hydrogenase
Ni-R state



RlmN, caught in the act …



PSII – Oxygen-Evolving Complex

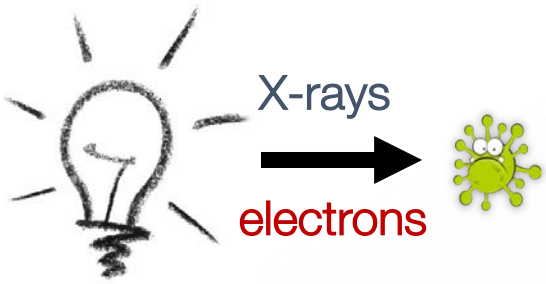
Schwalm, Grove, Booker & Boal (2016) *Science* 352:309
Ogata, Nishikawa & Lubitz (2016) *Nature* 520:571
Burger, Andrade & Einsle (2015) *Curr. Op. Struct. Biol.* 35:32
Sippel *et al.* (2018) *Science* 356:1434
Spatzal *et al.* (2011) *Science* 334:940
Umena *et al.* (2011) *Nature* 473:55

Structural Biology: Seeing Molecules at Work

⇒ Microscopy on an atomic scale

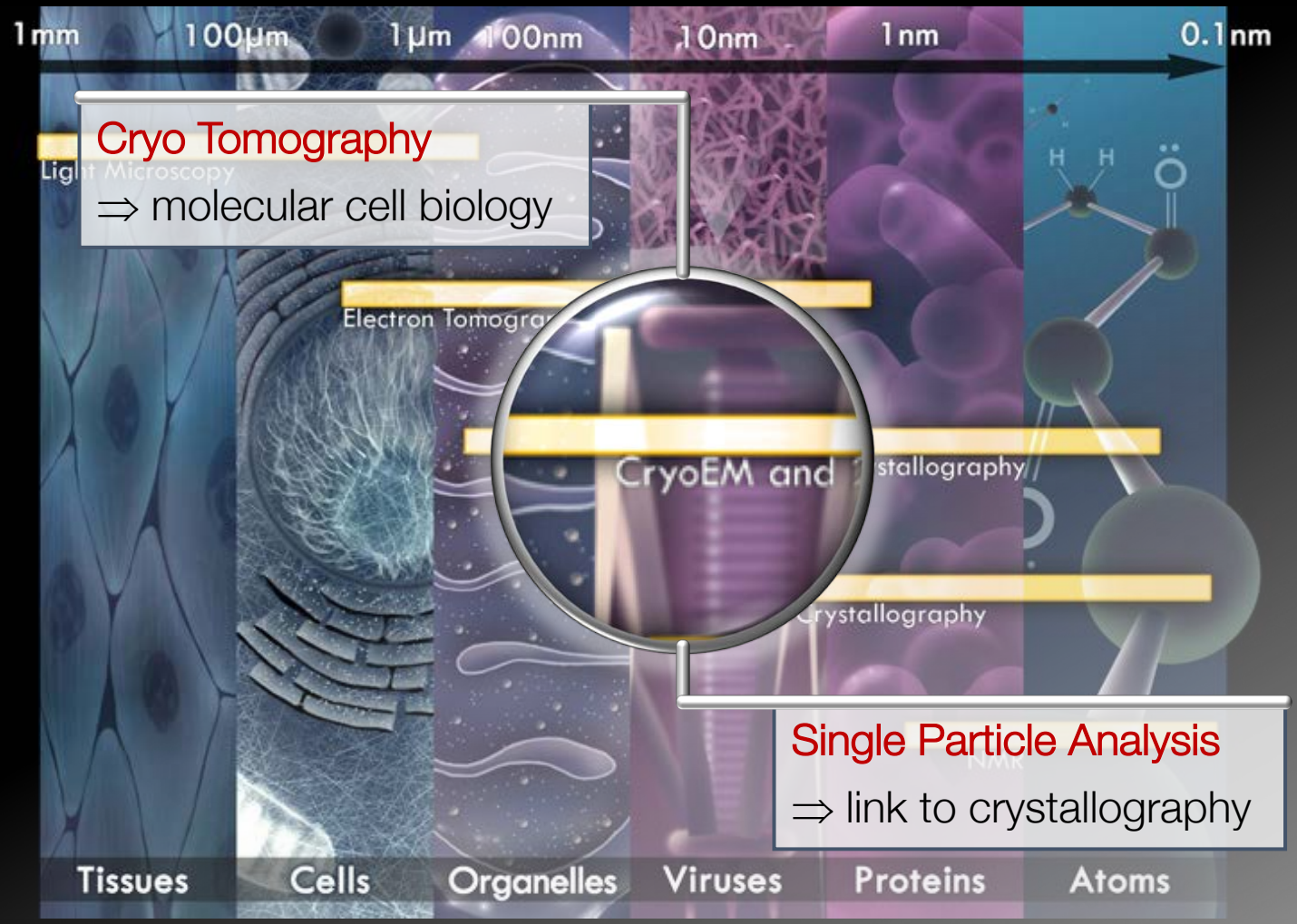
greek: *μικρός*: small; *σκοπειν*: to observe

visible light



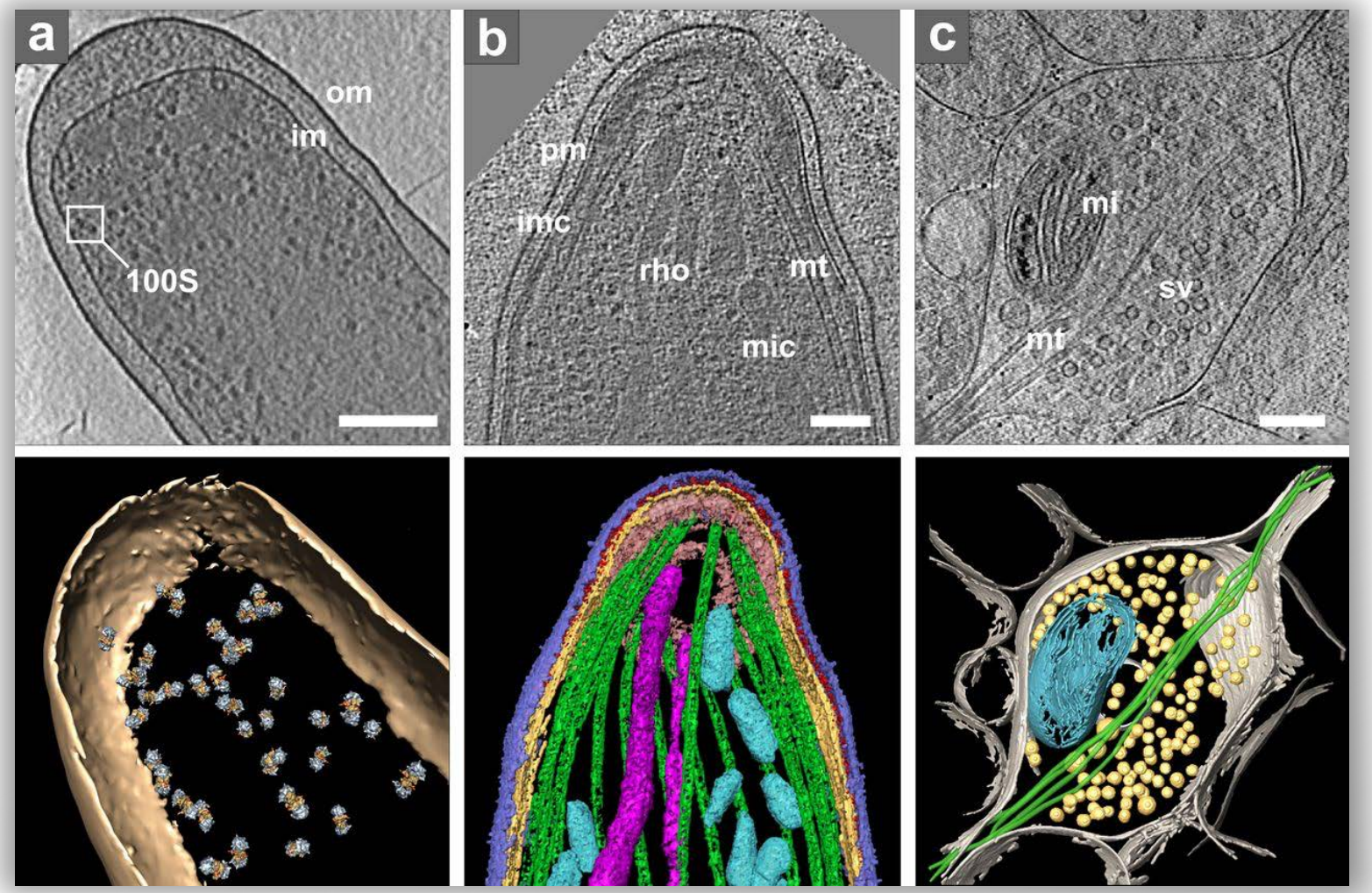
light source object optics image

The Resolving Power of Microscopes

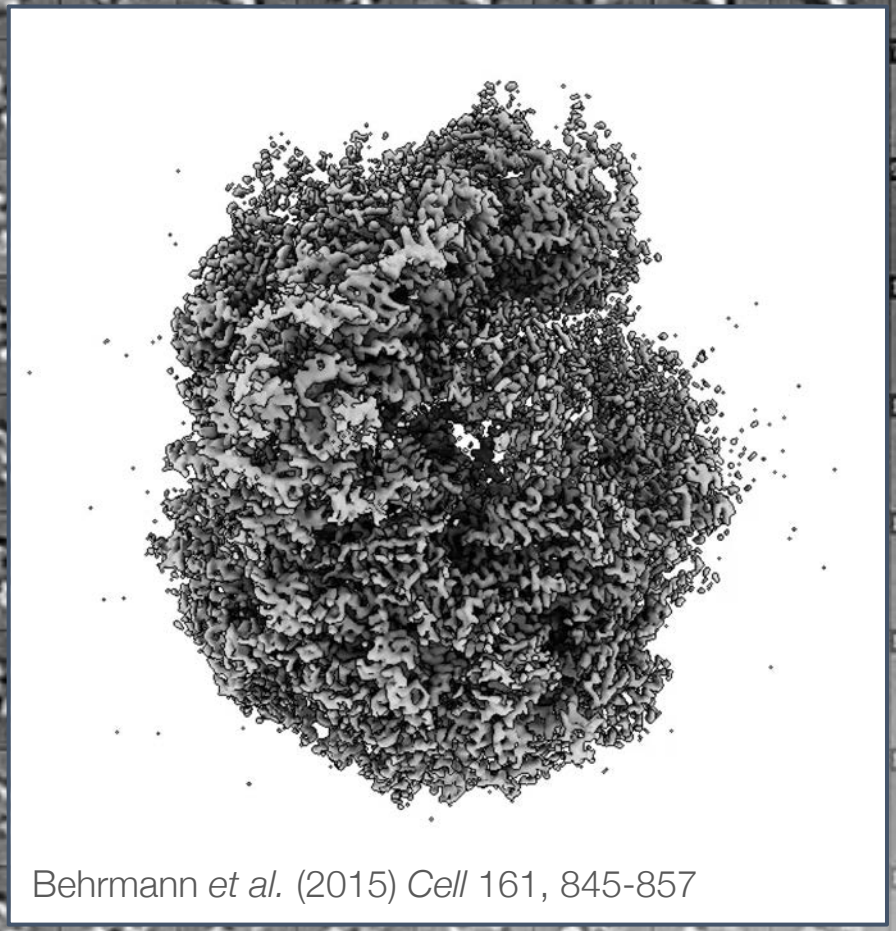


UNI FREIBURG

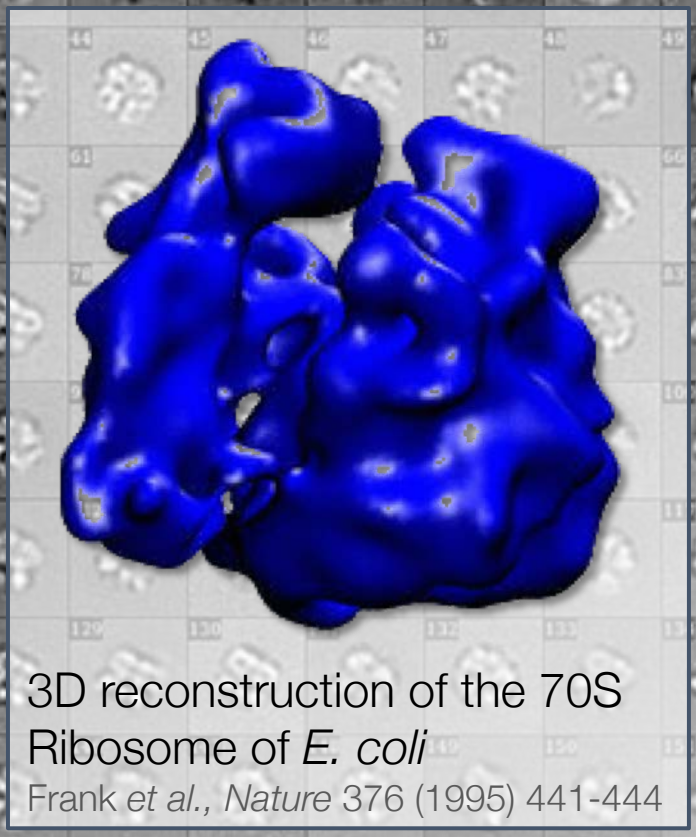
Electron Tomography



Single-Particle Analysis



Behrmann *et al.* (2015) *Cell* 161, 845-857



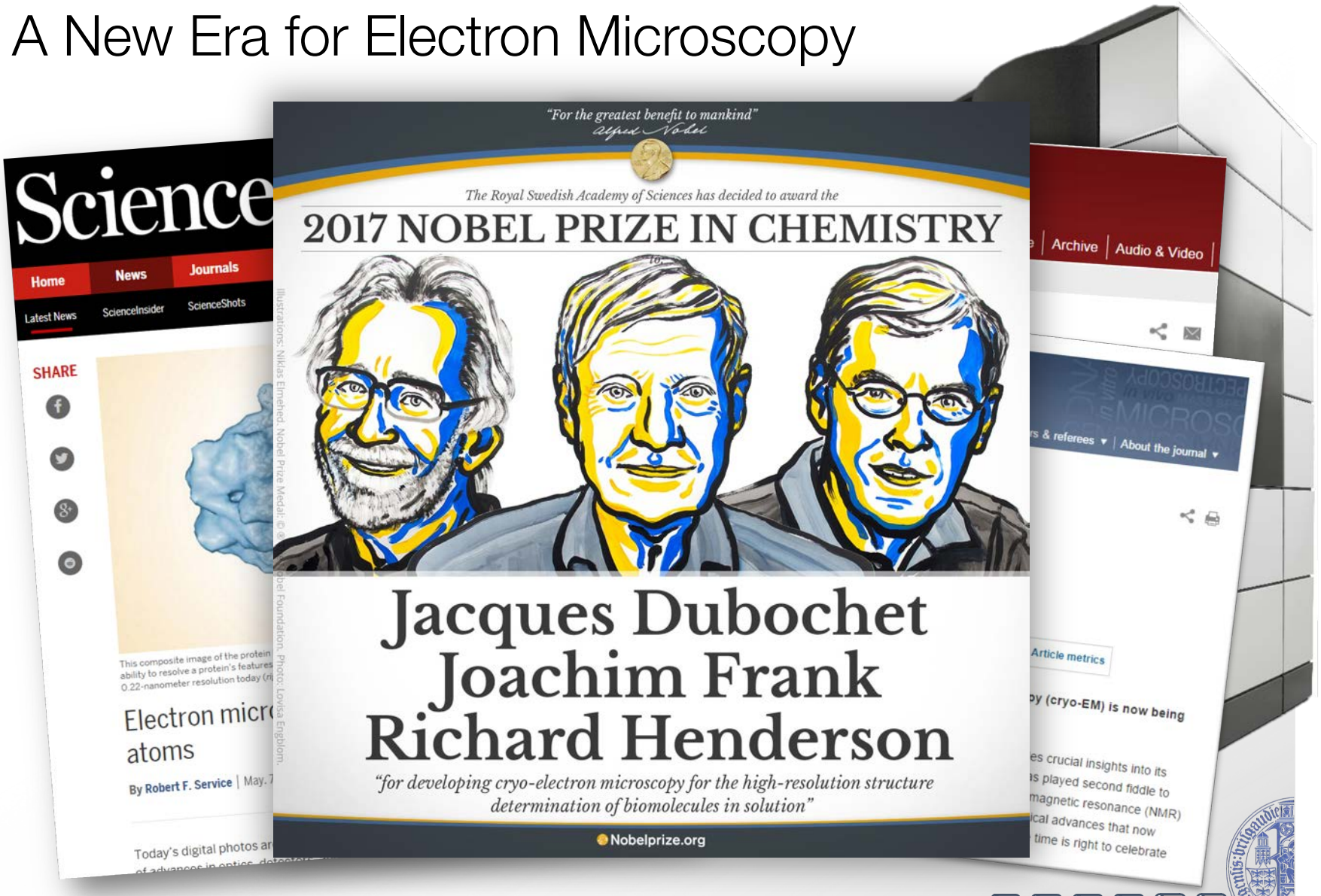
3D reconstruction of the 70S Ribosome of *E. coli*

Frank *et al.*, *Nature* 376 (1995) 441-444

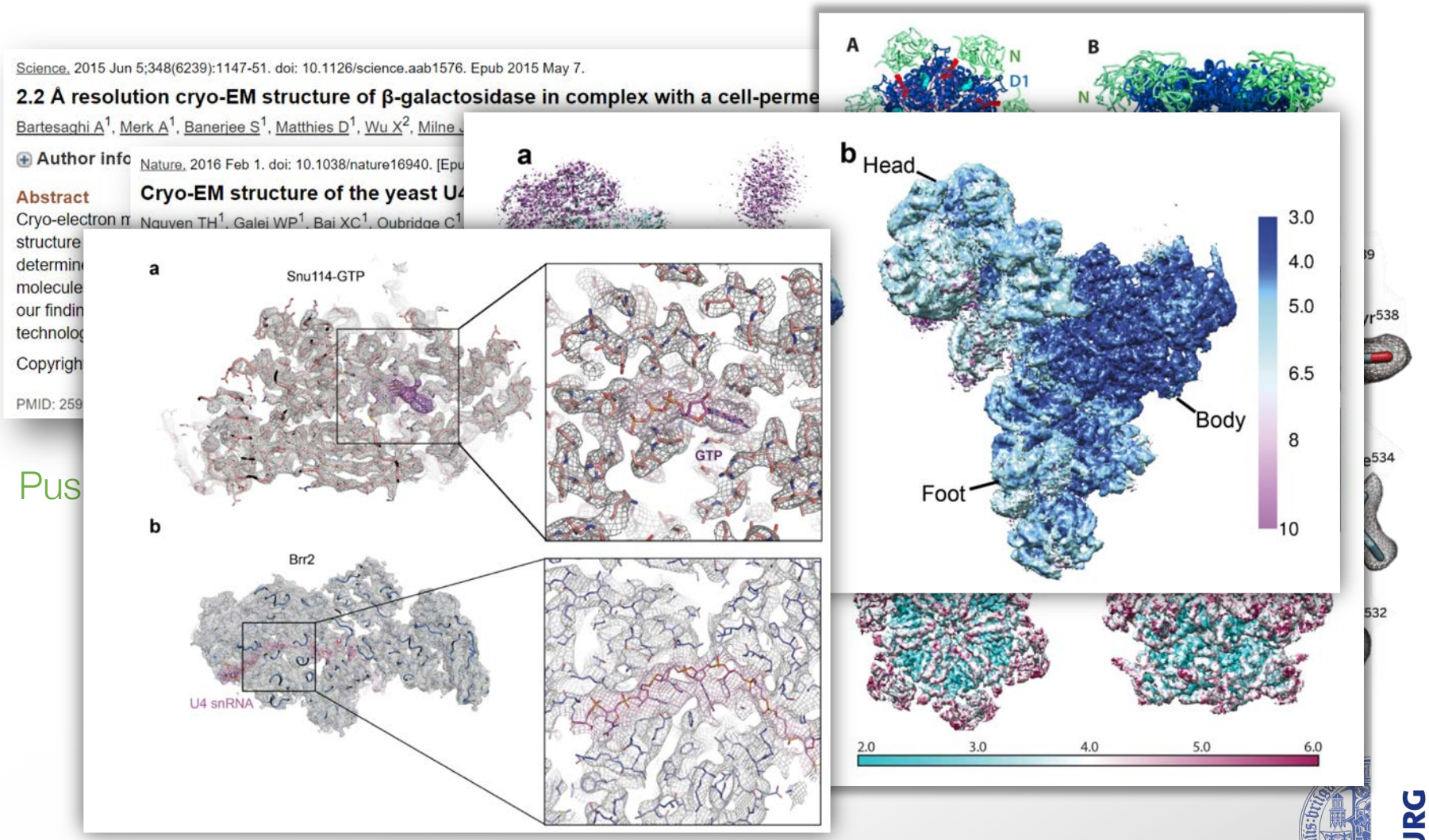
Averaging a large number of observations *in silico in order* to enhance signal-to-noise.

UNI FREIBURG

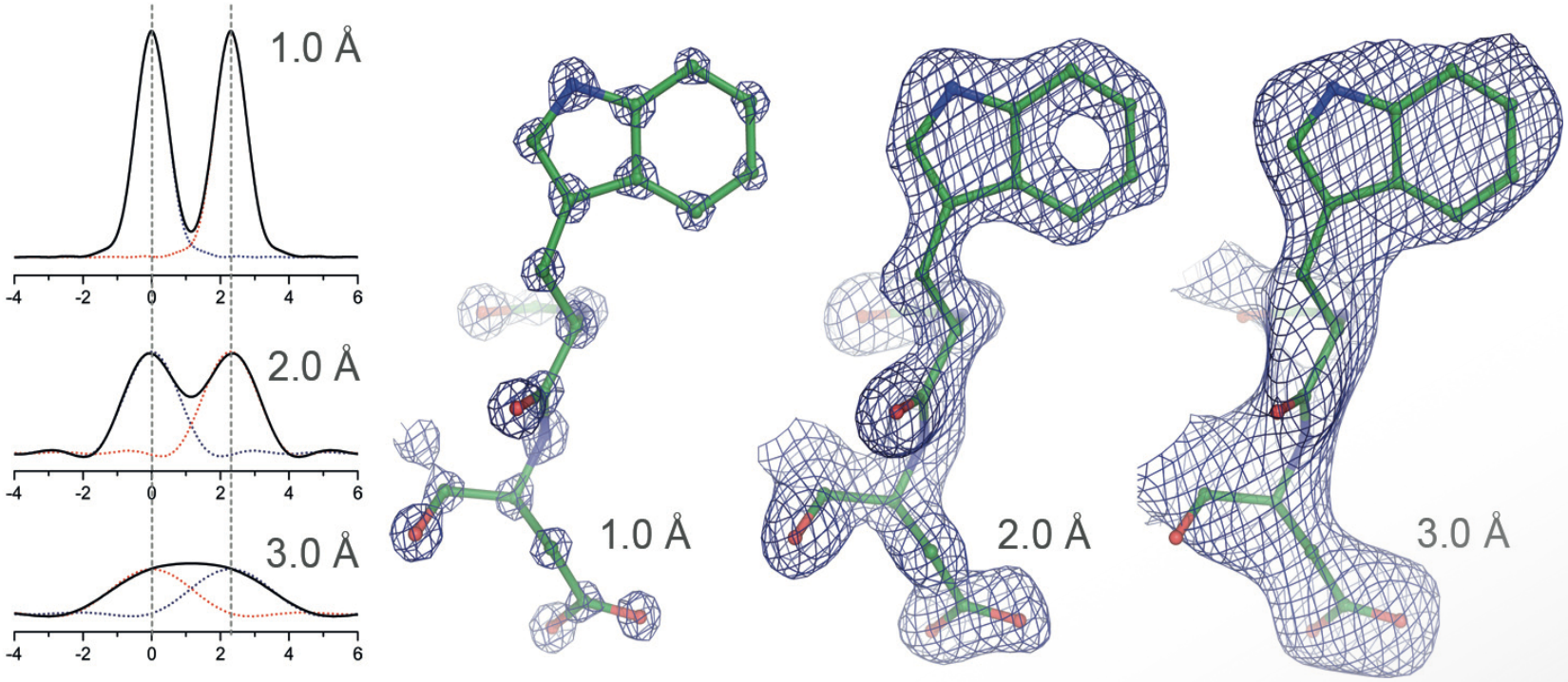
A New Era for Electron Microscopy



Pushing the Limits



The Power of Crystallography

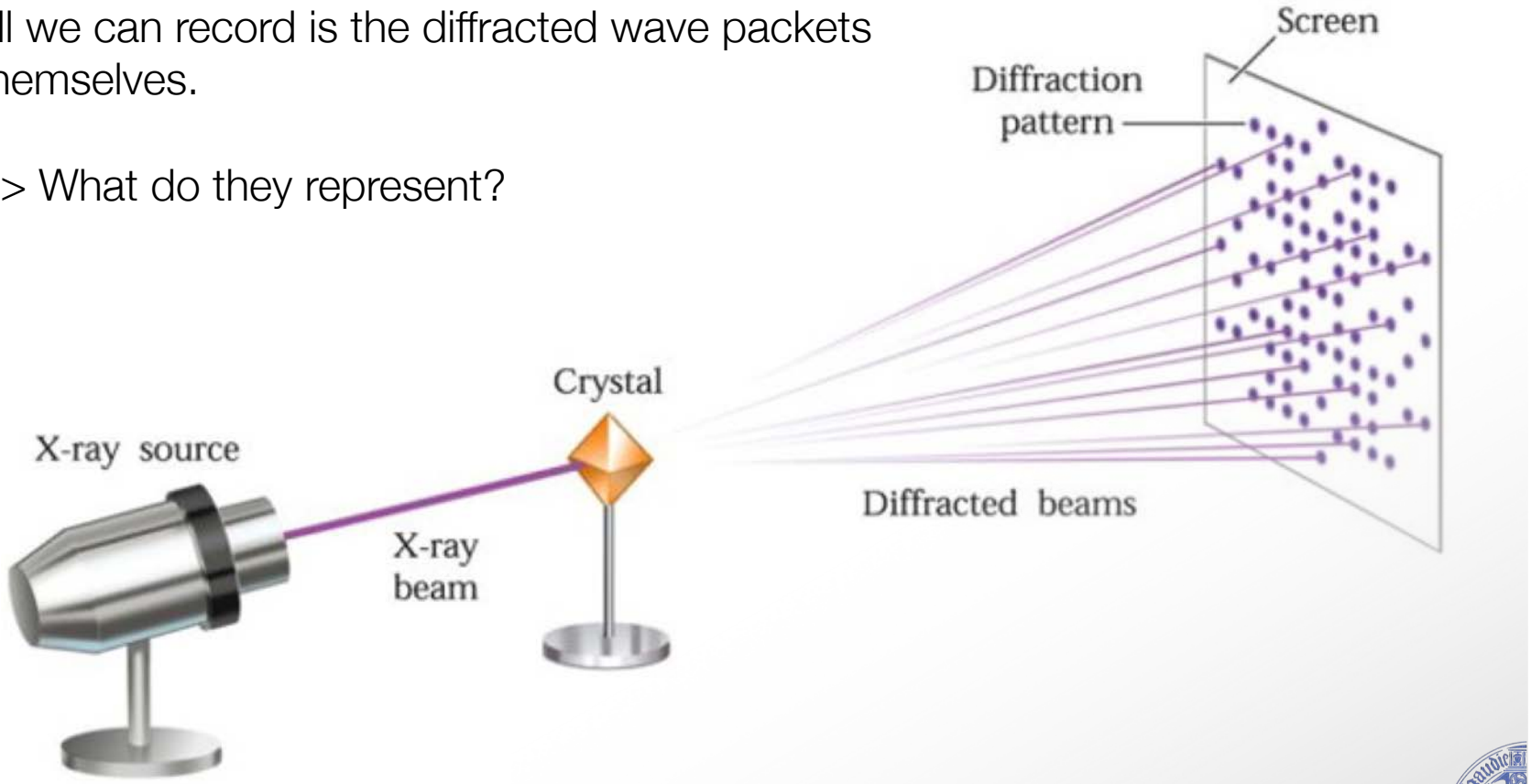


- **established method**: Structure solution is mostly straightforward (and fast)
- **real atomic resolution**: Precision of structure is the basis for mechanistic understanding!

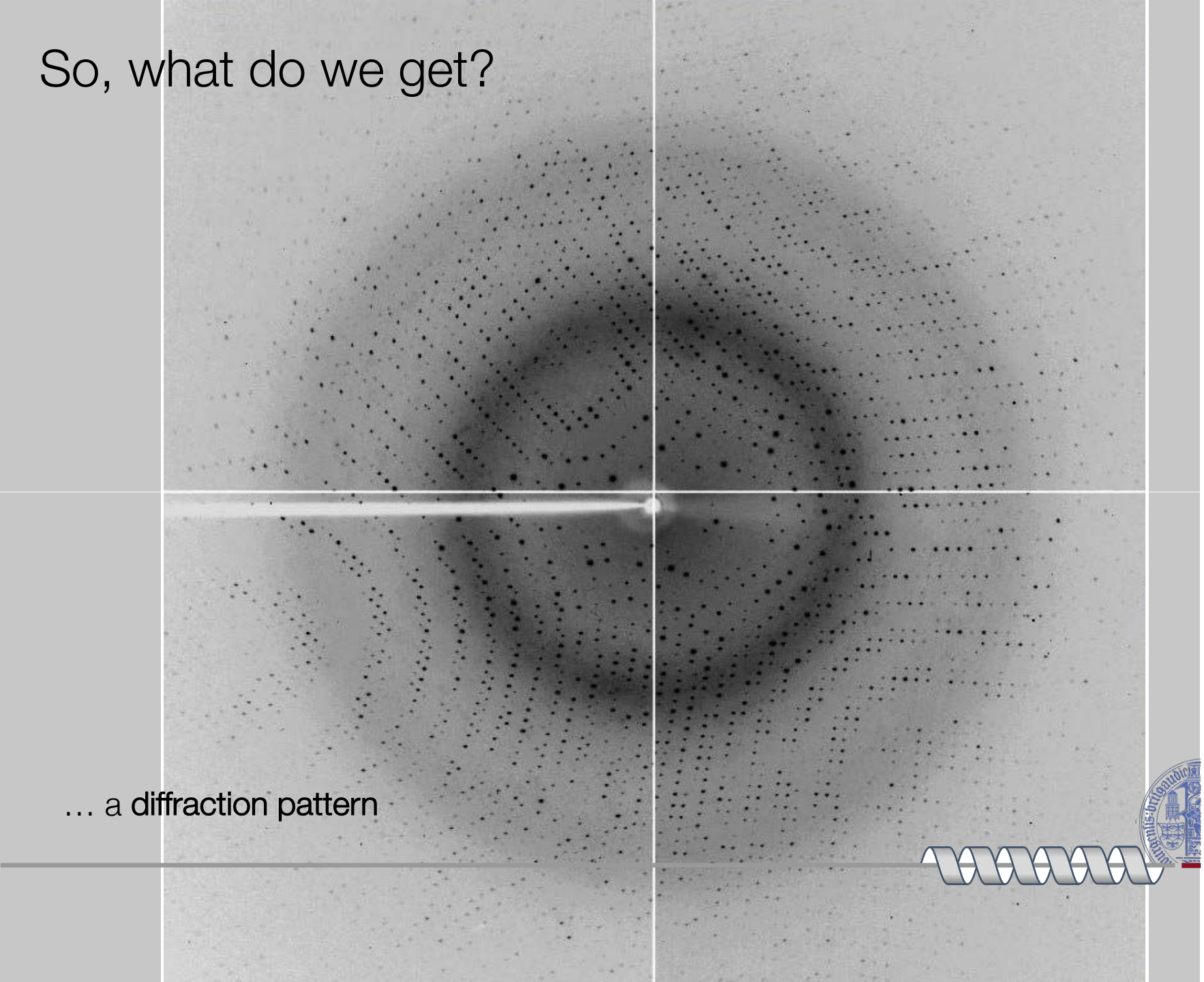
Caveat: There are no Lenses for X-rays

An X-ray beam is diffracted by a sample just like an electron beam.
However, as there are no focusing lenses for X-rays,
all we can record is the diffracted wave packets
themselves.

=> What do they represent?



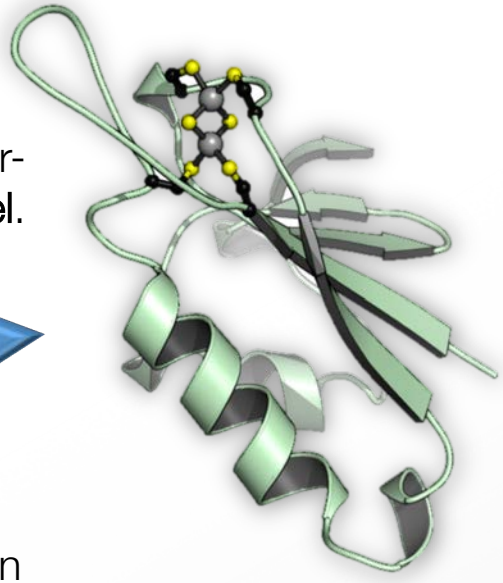
So, what do we get?

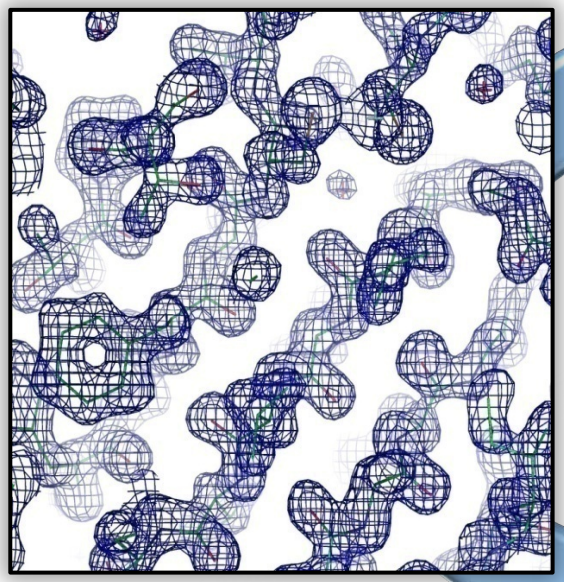


… a **diffraction pattern**

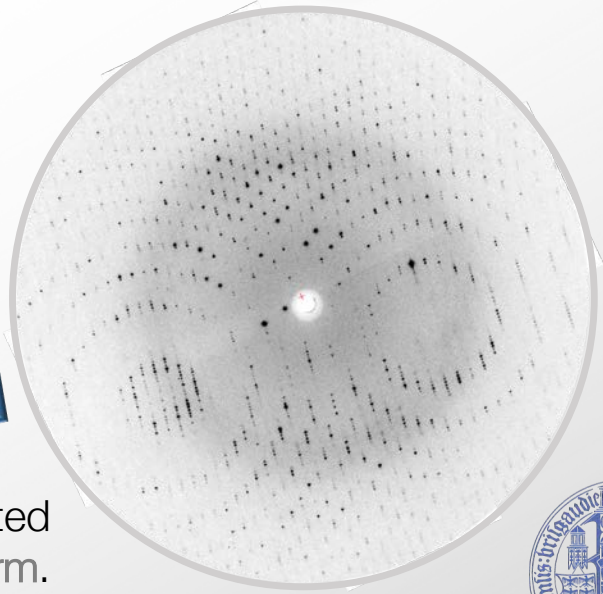
Solving a Crystal Structure

The final goal of a crystallographic structure determination is a **three-dimensional structural model**.



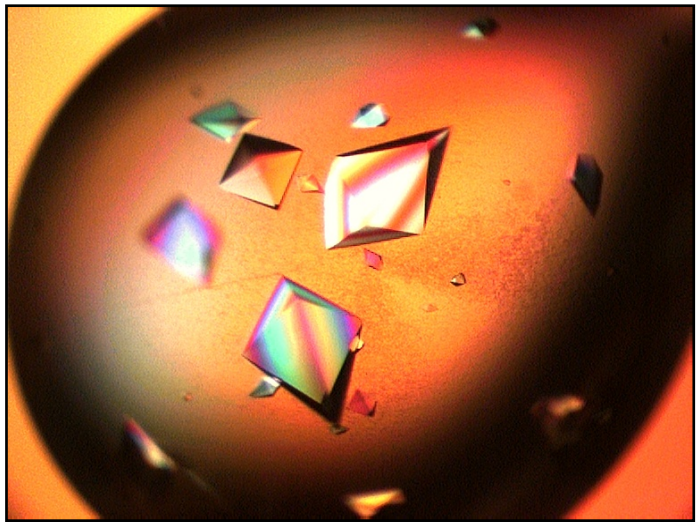


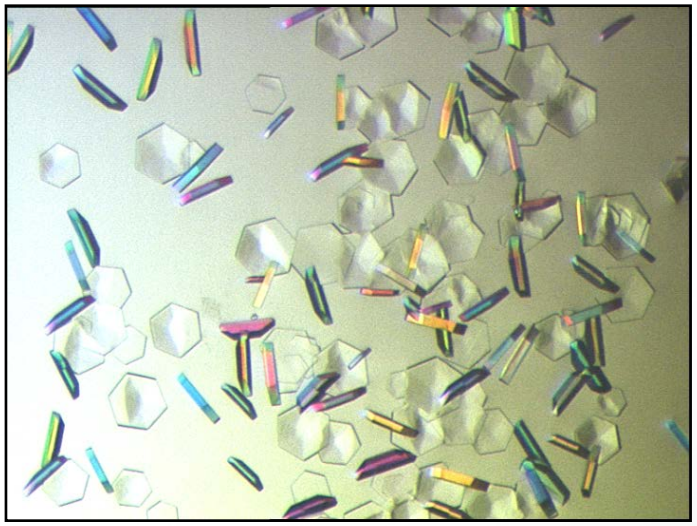
It is obtained by fitting an atomic model into an **electron density map** that represents the actual experimental result.

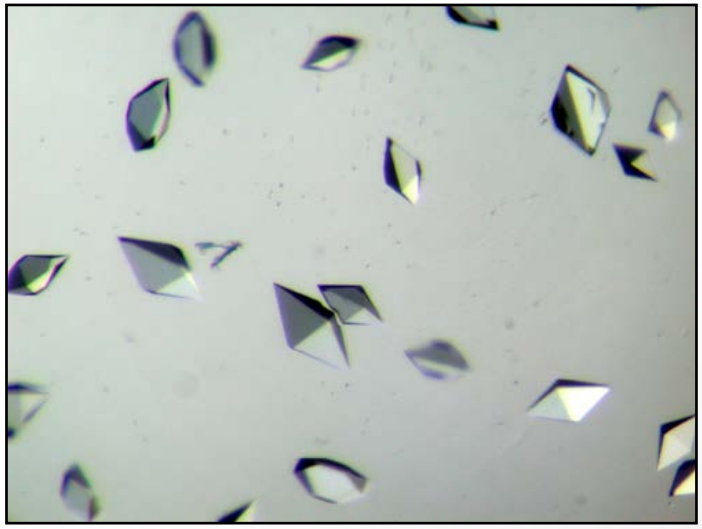


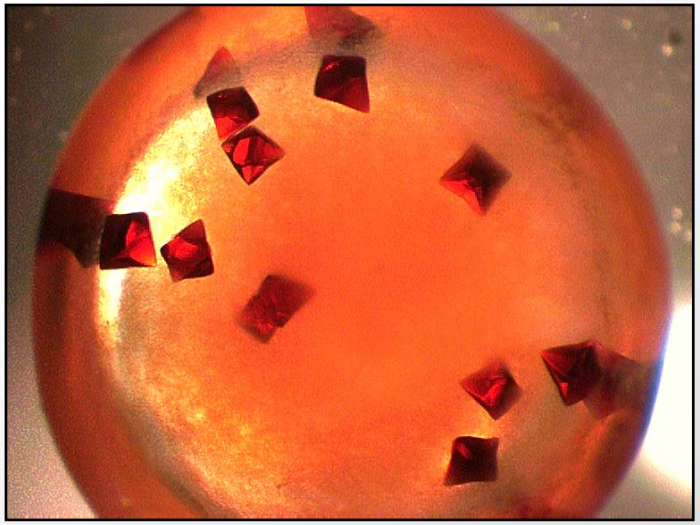
The electron density map is calculated from the collected **diffraction data** of a crystal via a Fourier transform.

Protein Crystals

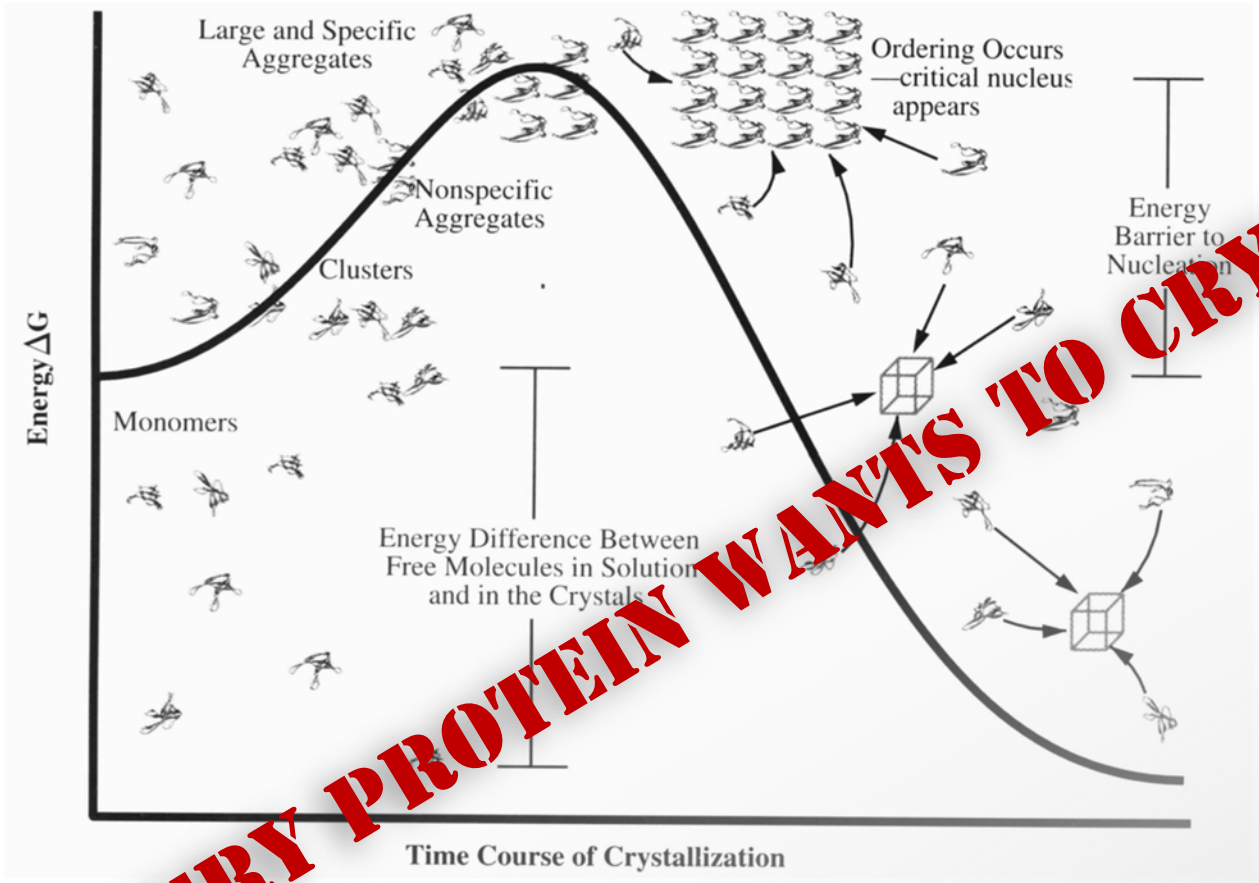








Crystallization is a Two-Step Process



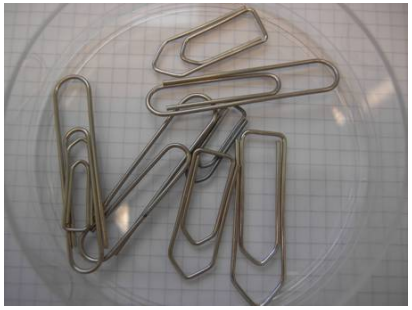
EVERY PROTEIN WAYS TO CRYSTALLIZE !!!

Protein solubility is a function of:

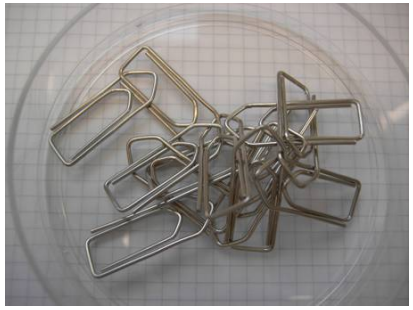
- ⇒ concentration
- ⇒ pH
- ⇒ Ionic strength
- ⇒ temperature
- ⇒ dielectricity
- ⇒ (...)

crystallization lowers the free energy of proteins by ~ 3-6 kcal·mol^{-1}

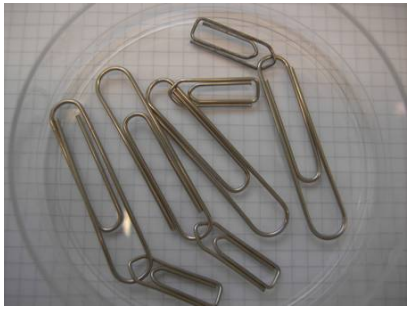
Obstacles in Forming Protein Crystals



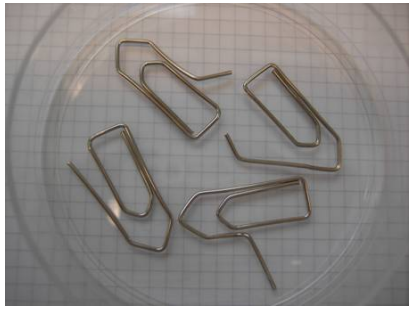
Impurities



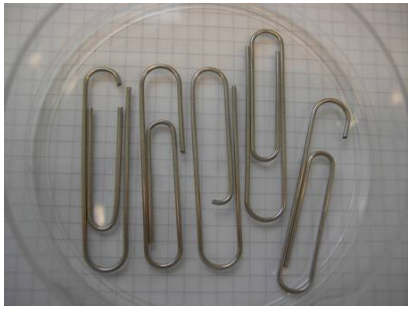
Aggregation



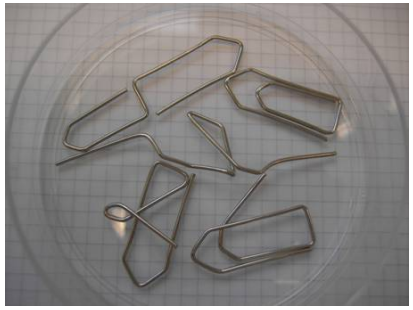
Flexible domains



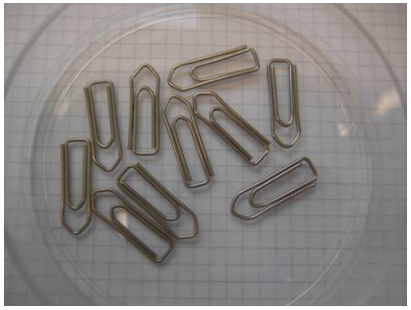
Flexible tail



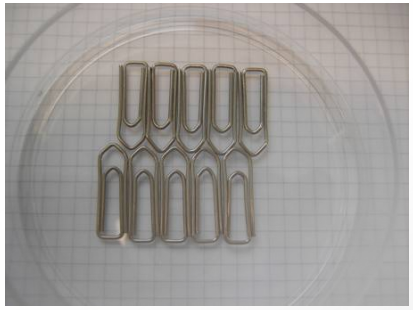
Proteolysis



Unfolding



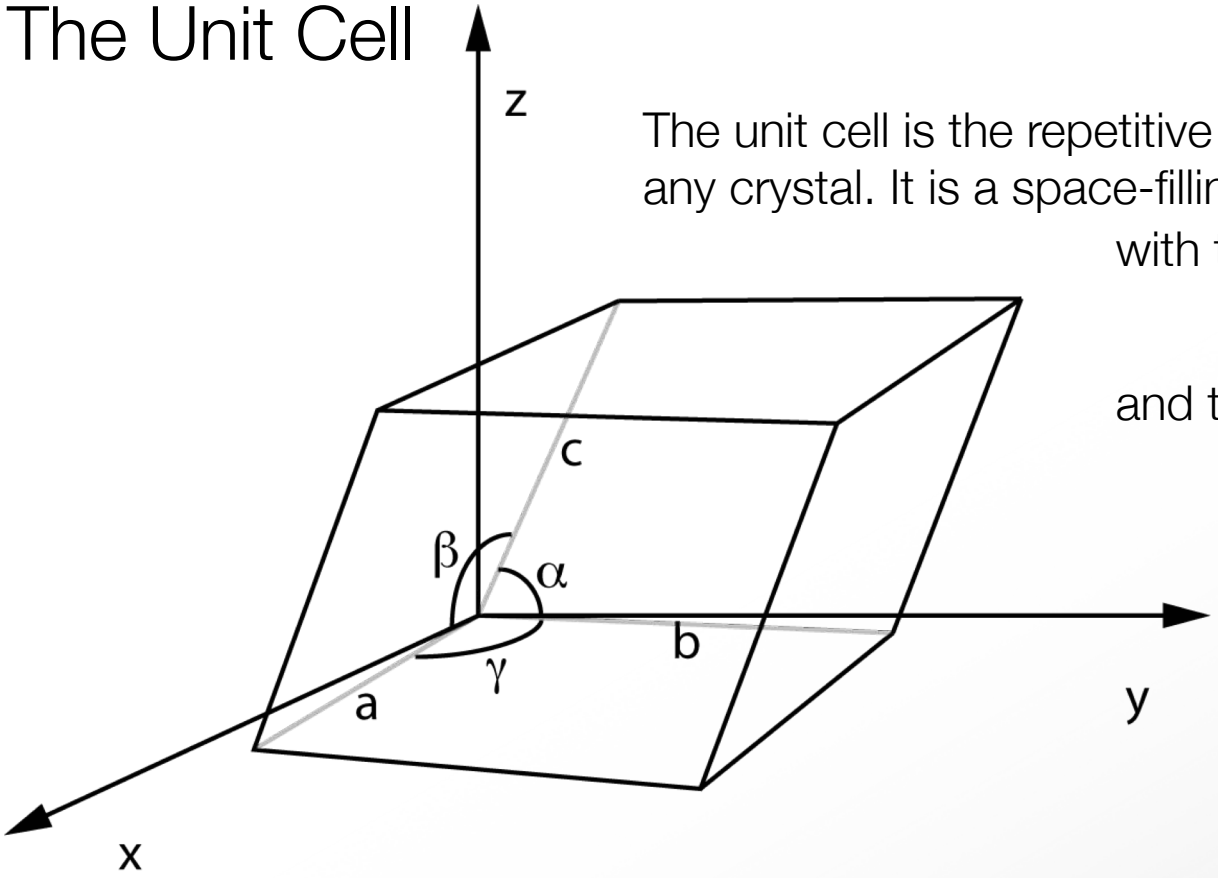
„Pure“ protein



Crystals

Many things can go wrong – good protein biochemistry is the key to success!

The Unit Cell



The unit cell is the repetitive building block of any crystal. It is a space-filling **parallelepiped** with the unit cell axes

$\mathbf{a}, \mathbf{b}, \mathbf{c}$

and the angles

α, β, γ

The general volume of the unit cell is

$$V = abc\sqrt{1 - \cos^2\alpha - \cos^2\beta - \cos^2\gamma + 2\cos\alpha\cos\beta\cos\gamma}$$

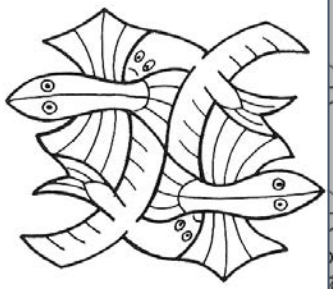
Basic Crystal Architecture

Only packing arrangements of objects that fill space completely generate single crystals: **7 crystal systems** lead to **14 Bravais lattices**, from which **230 space groups** arise through additional sy

As proteins 65 *enantiom*

crystal syster		
triclinic		
monoclinic		
orthorhombic		
trigonal		
tetragonal	*P*	$a = b \neq c$, 90, 90, 90
hexagonal	*P*	$a = b \neq c$, 90, 90, 120
cubic	*P, F, I*	$a = b = c$, 90, 90, 90

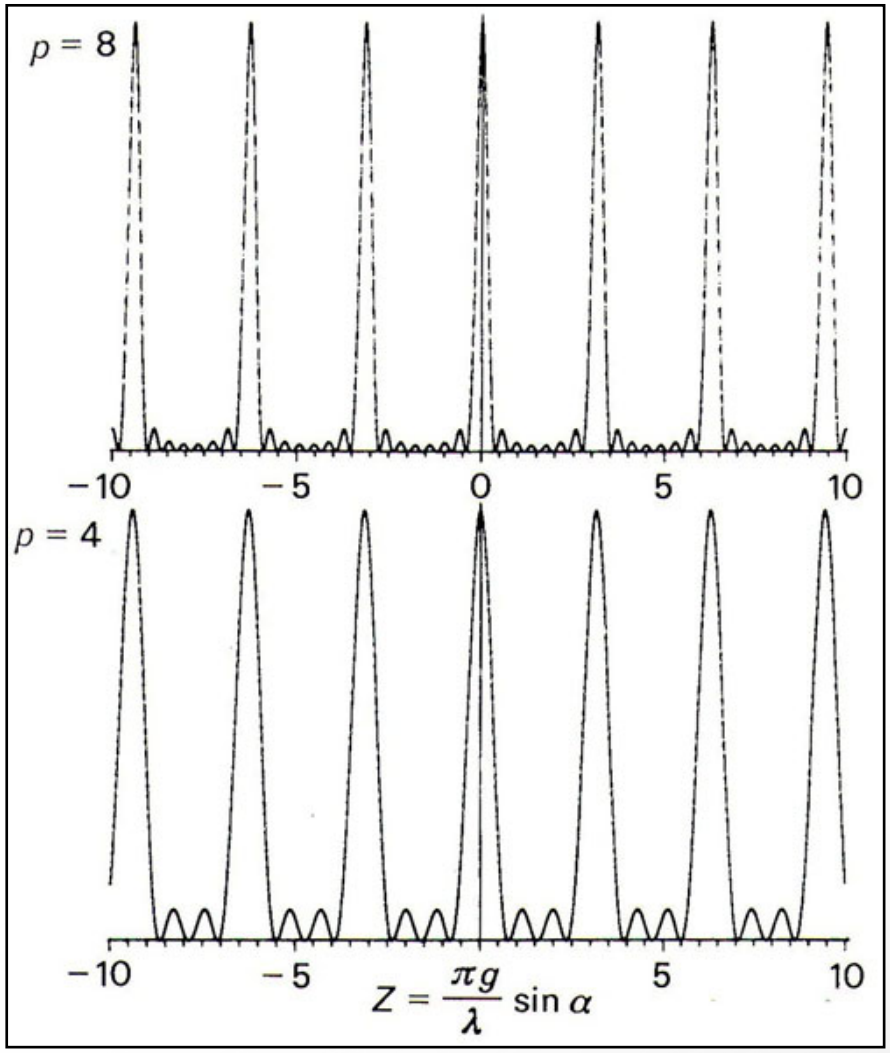
Unit cells (can) contain internal symmetry elements.



The basic building block to be modelled is the asymmetric unit.

the 14 Bravais lattices

Lattice Interference



With increasing number of slits/ grid points, the maxima, caused by **constructive interference**, become more focused and more intense.

Between the maxima, destructive interference cancels out any signal.

Diffraction on a crystal lattice results in a

- **discrete** and
- **amplified**

diffraction pattern. This is pre-requisite for recording a signal.

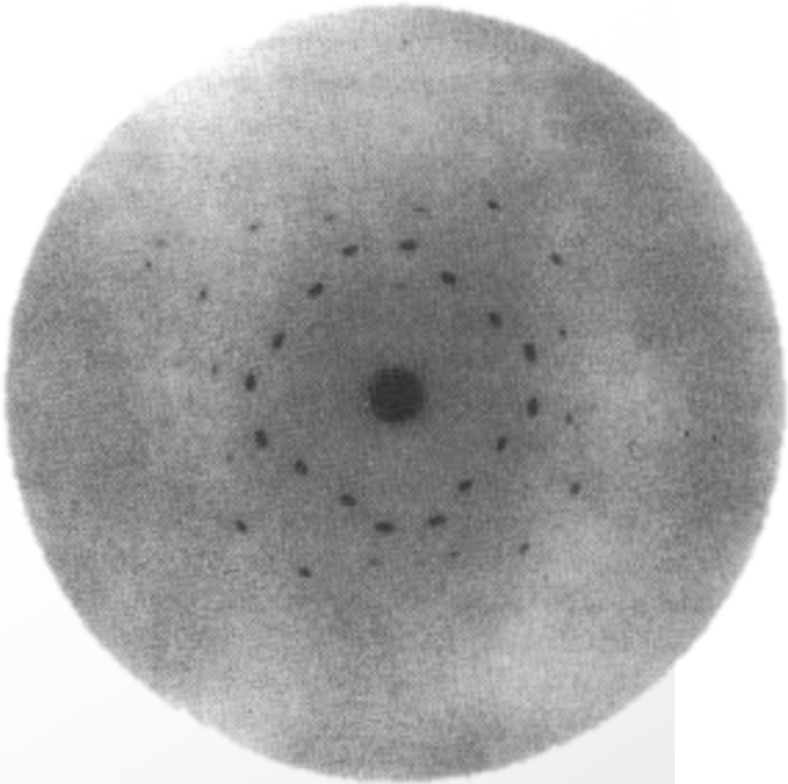
The Birth of Crystallography

Max von Laue
(*1879 †1960)

Diffraction of X-rays by crystals.

"Als Herr v. LAUE nach der theoretischen Einleitung die erste Aufnahme zeigte, die den Durchgang eines Strahlenbündels durch ein ziemlich willkürlich orientiertes Stück von triklinem Kupfervitriol darstellte - man sah auf der photographischen Platte neben der zentralen Durchstoßungsstelle der Primärstrahlen ein paar kleine sonderbare Flecken -, da schauten die Zuhörer gespannt und erwartungsvoll, aber doch wohl nicht ganz überzeugt auf das Lichtbild an der Tafel. Aber als nun jene Figur 5 sichtbar wurde, das erste typische LAUEdiagramm, welches die Strahlung durch einen genau zur Richtung der Primärstrahlung orientierten Kristall regulärer Zinkblende wiedergab mit ihren regelmäßig und sauber in verschiedenen Abständen vom Zentrum angeordneten Interferenzpunkten, da ging ein allgemeines "ah" durch die Versammlung. Ein jeder von uns fühlte, daß hier eine große Tat vollbracht war".

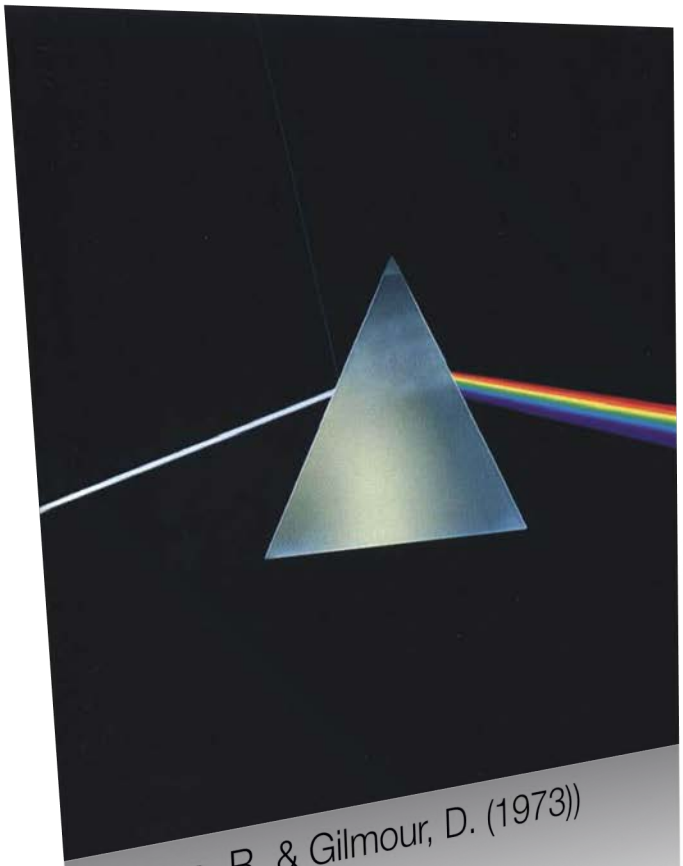
Max Planck, 1937



June 14, 1912:
diffraction image of
of zinc blende (ZnS)

UNI FREIBURG

Diffraction of X-rays



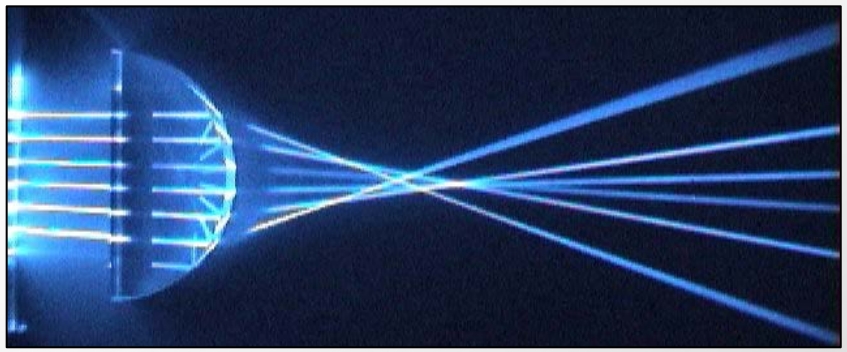
(Waters, R. & Gilmour, D. (1973))

Incoming Radiation is split into discrete wave packets by the object:

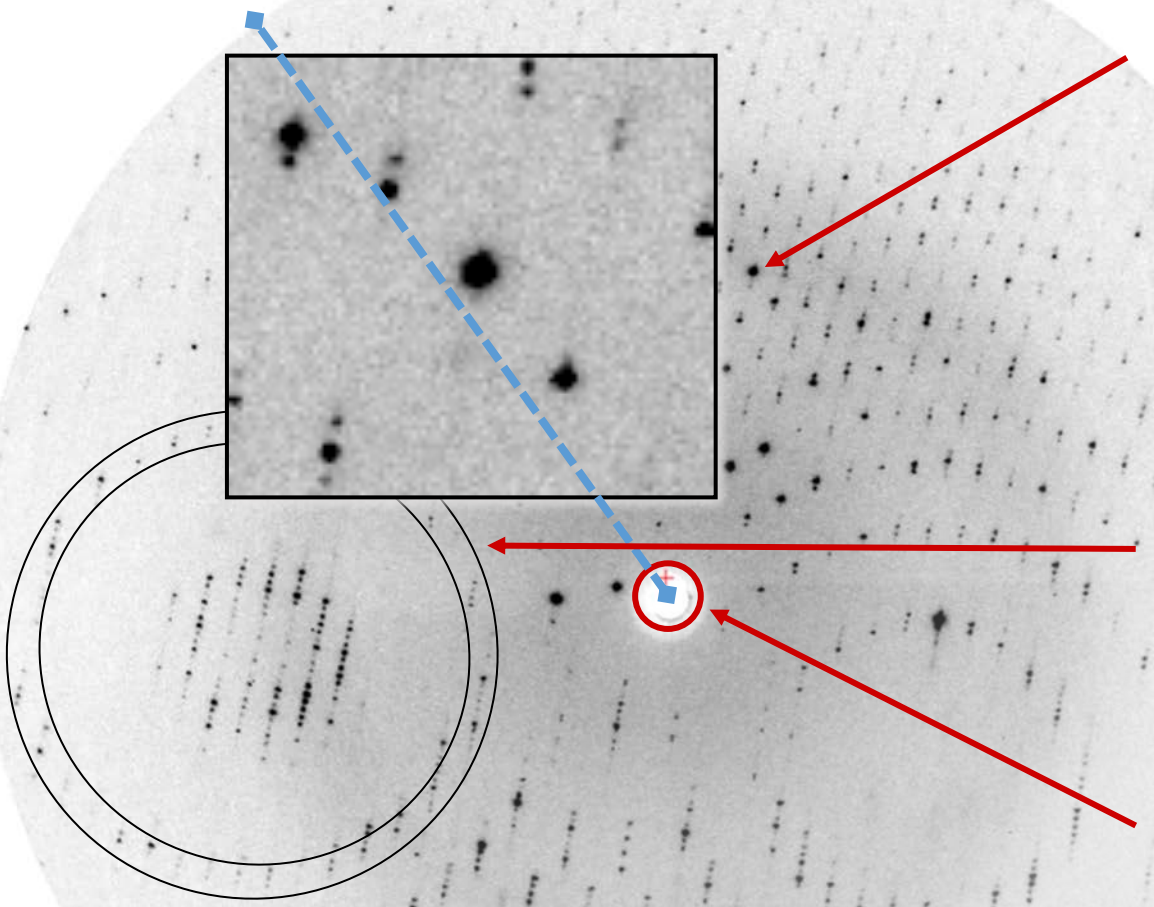
The structure of the object becomes encoded in the distribution of diffracted waves.

In mathematical terms, diffraction yields the **Fourier Transform** of the object.
Through its reversion, the structure of the object can be recalculated.

This reverse transform is what a lens does:



Understanding a Diffraction Image



Reflections:

Distinct intensity maxima appear in a regular arrangement on the image.

Resolution:

The further away from the center a reflection appears, the more detailed is the structural information encoded in it.

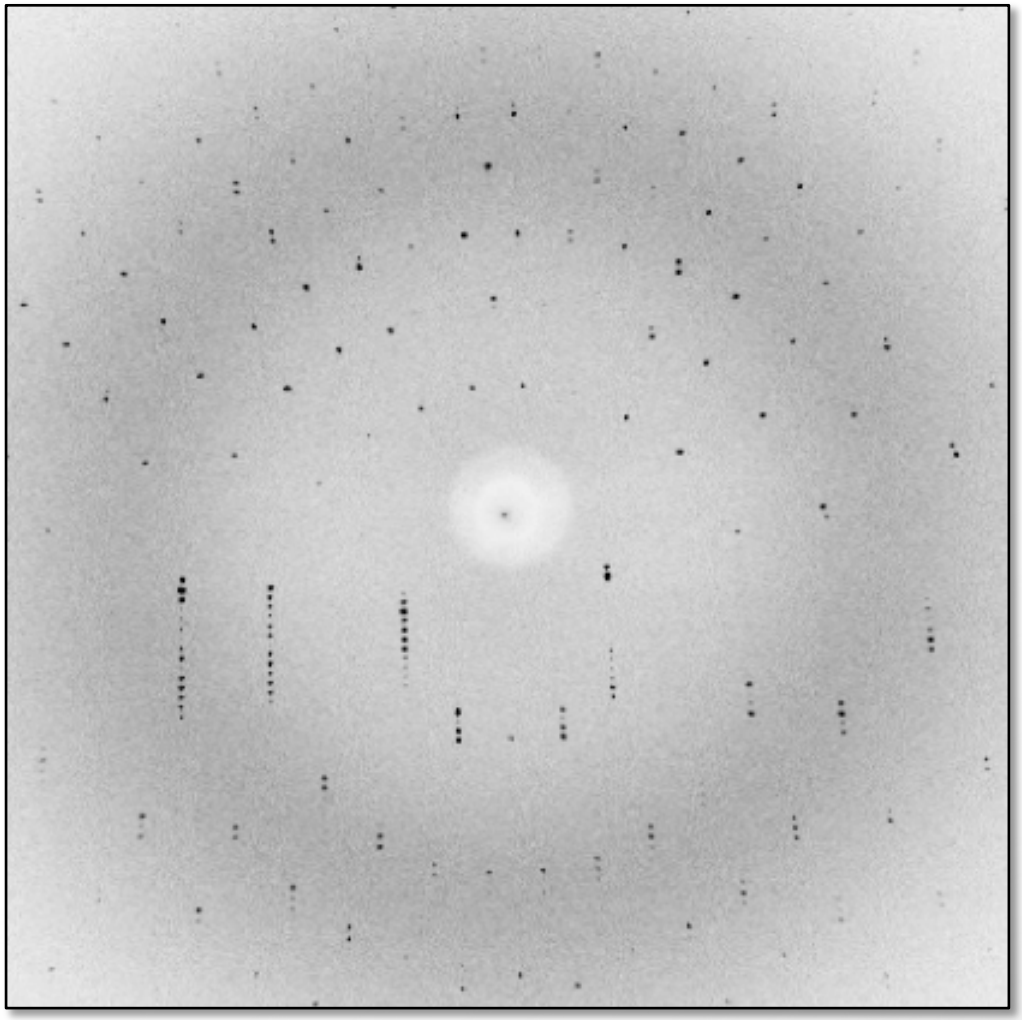
Lunes:

Reflections appear in rings or circles in the diffraction image. Within these, they are arranged in lines.

Beam Stop Shadow:

Most X-rays pass the crystals without interference and must be stopped by a lead beam stop.

Indices for Diffraction Spots



In order to process diffraction data, each reflection has to be assigned an unambiguous identifier, a position in the reciprocal lattice in Fourier Space.

During indexing, reflections are assigned **Miller Indices** (h,k,l) that mark their position in the reciprocal lattice.

In the following step, integration, the intensity of the reflection is measured, resulting in a table:

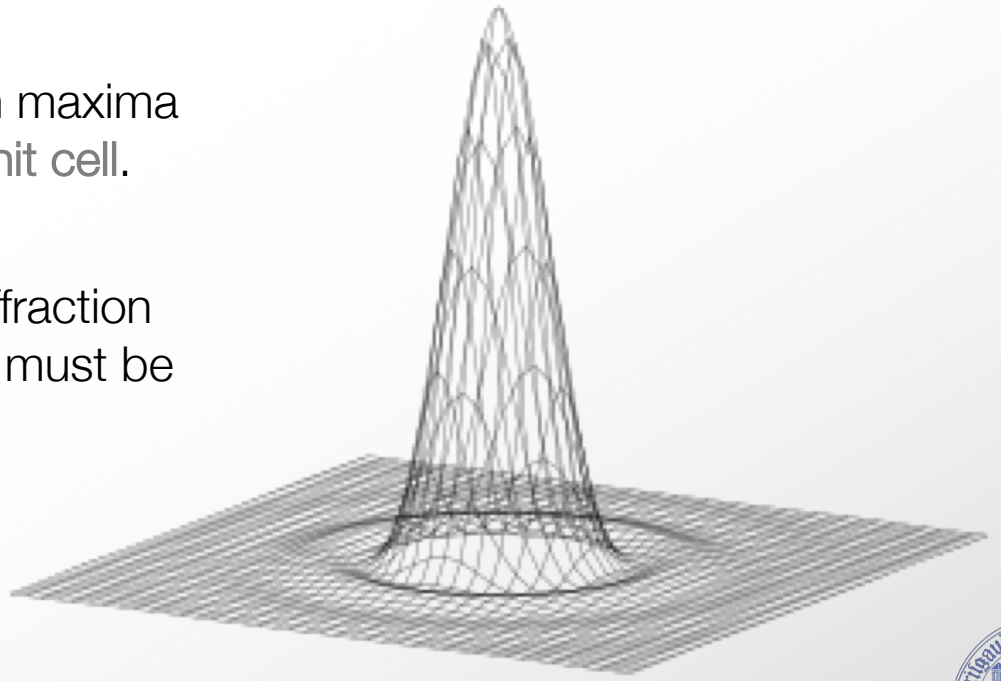
h	k	l	$\lvert\mathbf{F}\rvert$	$\sigma(\lvert\mathbf{F}\rvert)$

Information in a Diffraction Image

two levels of information:

- The **position** of diffraction maxima is solely dependent on the dimensions of the unit cell.
- The **intensity** of the diffraction maxima reflects the contents of the unit cell.

In order to obtain a complete diffraction data set, every single maximum must be **indexed** and **integrated**.

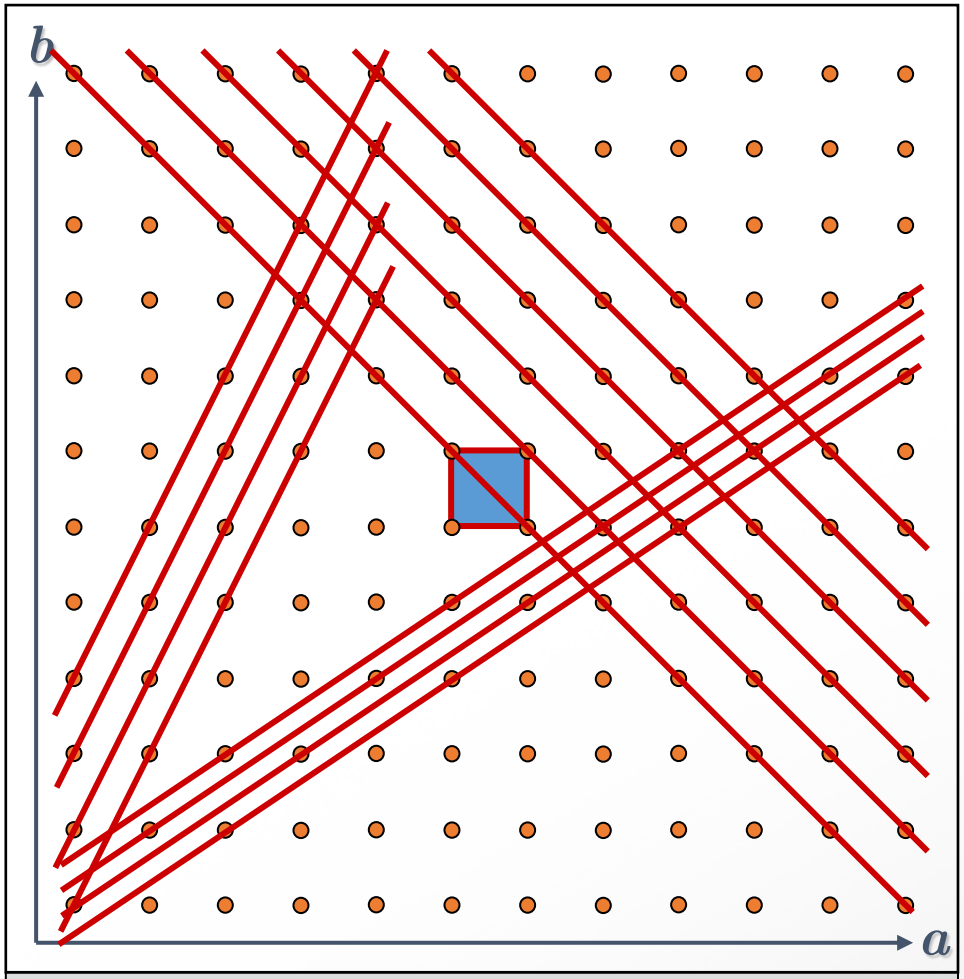


The Concept of Lattice Planes

How does the reciprocal lattice relate to the crystal lattice?

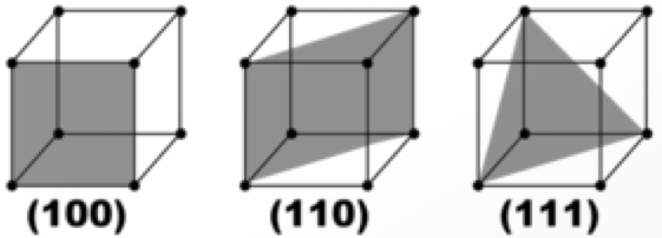
The diffraction pattern can be geometrically constructed if individual diffraction spots are considered as actual **reflections** of the X-ray beam on sets of imaginary (!) **lattice planes** in the crystal lattice.

Miller Indices (h,k,l)



the unit cell

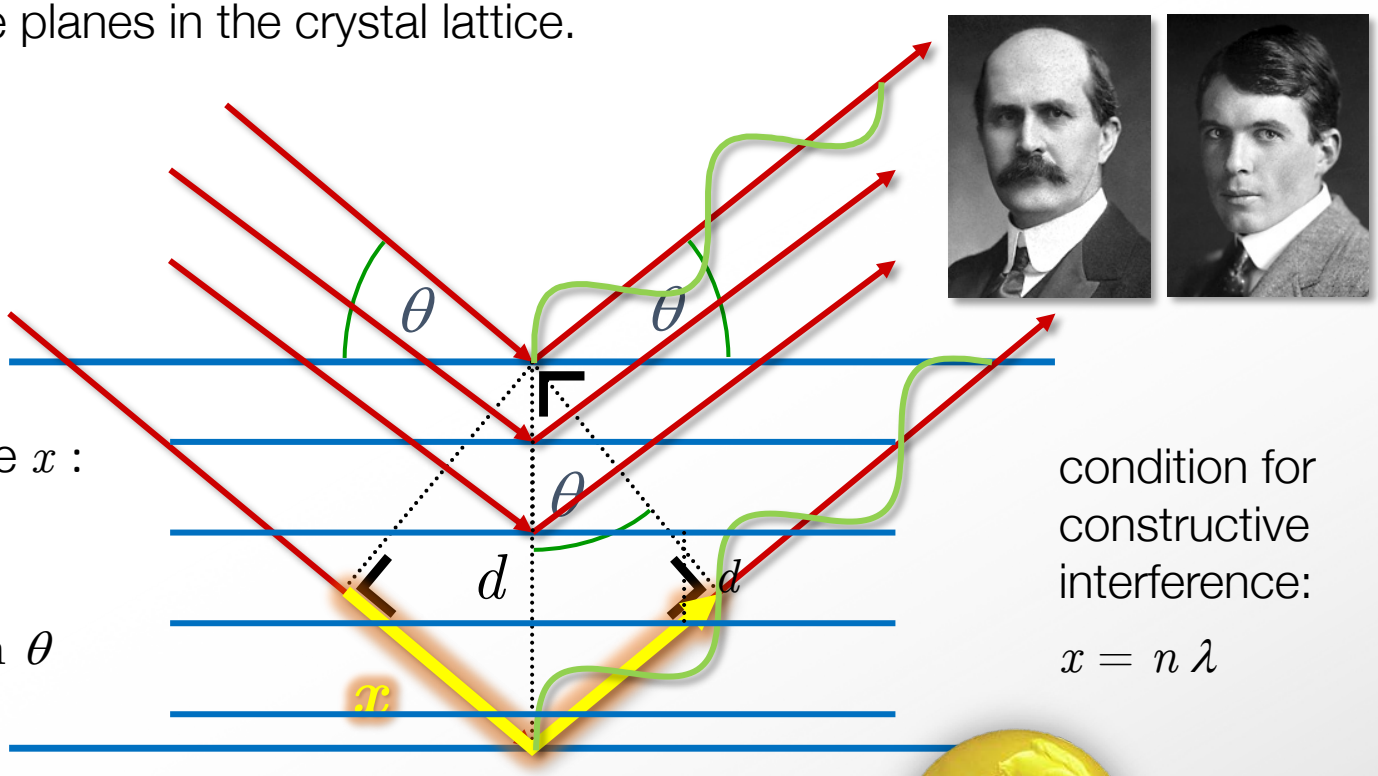
(2, 3, 1)



The Miller Index states in how many equal parts the unit cell axis is divided by the respective lattice plane.

Bragg's Law

Geometrical interpretation: "Reflections" are actual reflections of the X-ray beam by *virtual* lattice planes in the crystal lattice.



path difference x :

$$\sin\ \theta = \frac{x}{2d}$$

$$x = 2d \sin\ \theta$$

condition for constructive interference:

$$x = n\,\lambda$$

Bragg's Law:

$$n\ \lambda = 2d \sin\ \theta$$

W.H. Bragg, W.L. Bragg, 1913

The Ewald Sphere

Bragg's Law allows to calculate which points of the reciprocal lattice lead to an observable reflection in a given orientation of the crystal. Paul Peter Ewald visualized this as follows...

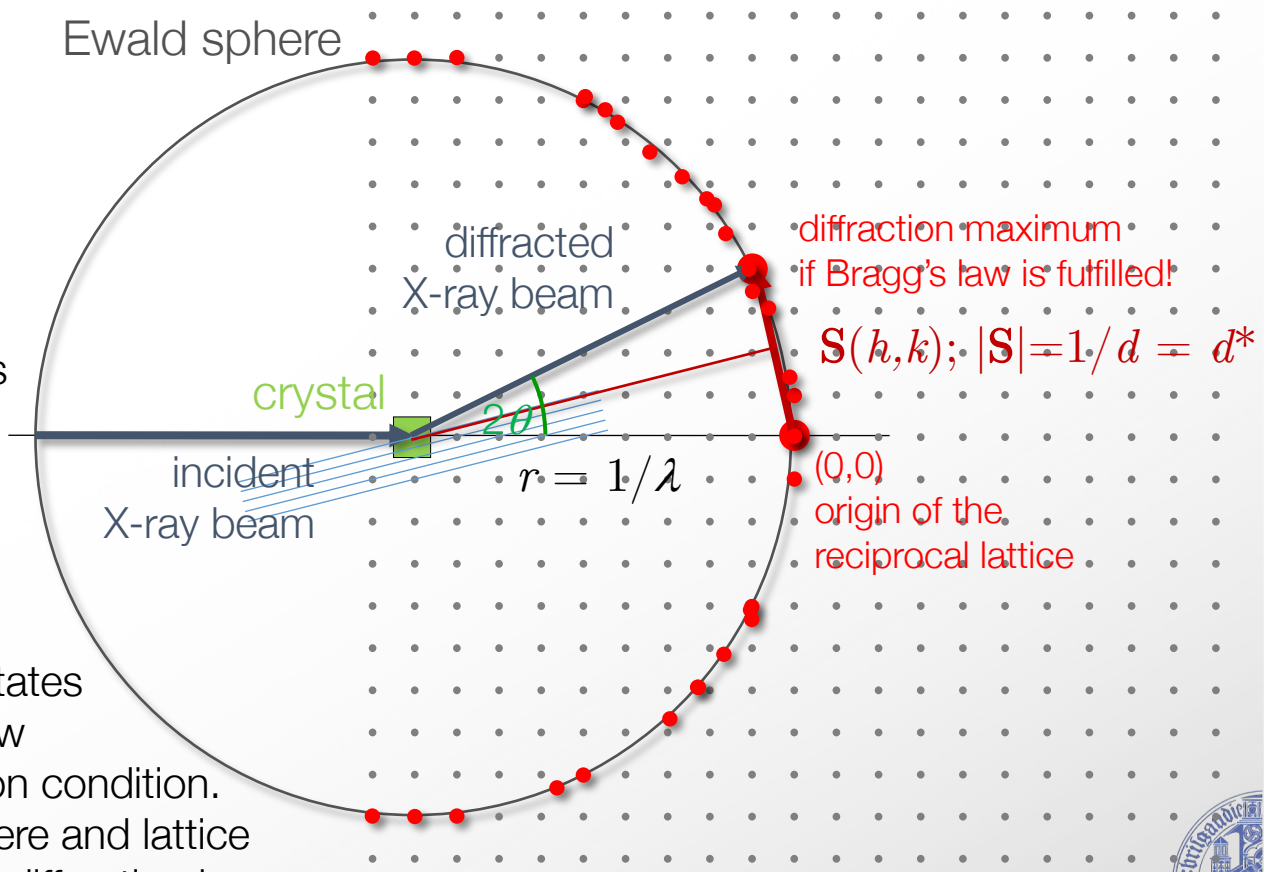
$$n\,\lambda = 2d\,\sin\,\theta$$

becomes:

$$\sin\,\theta = \frac{\lambda}{2d} = \frac{d^*}{2/\lambda}$$

Only those lattice points that lie on the Ewald sphere give rise to observable diffraction maxima.

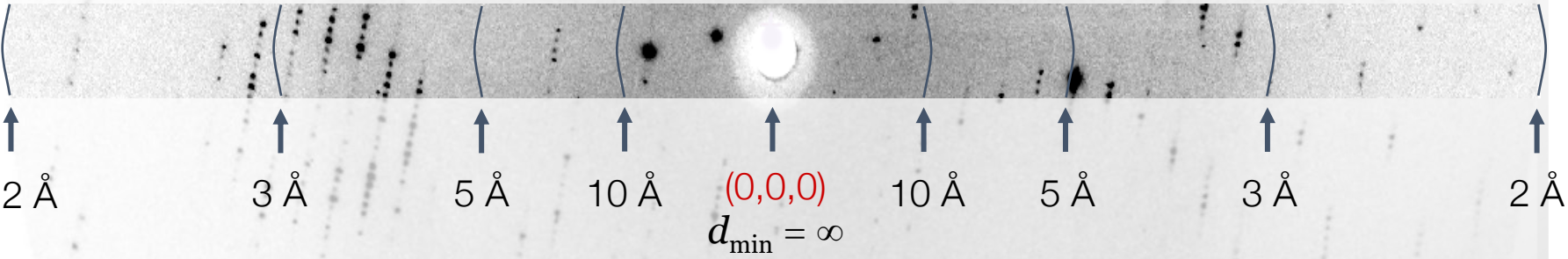
The reciprocal lattice rotates with the crystal, and new points fulfill the diffraction condition. This intersection of sphere and lattice created the lunes in the diffraction image.



The Resolution of a Crystal Structure

According to Bragg's Law: $d_{min} = \frac{\lambda}{2 \sin \theta_{max}}$

This means that reflections that correspond to high resolutions (low d_{min}) have large Miller indices ***(h,k,l)*** and are far from the center of the diffraction image:

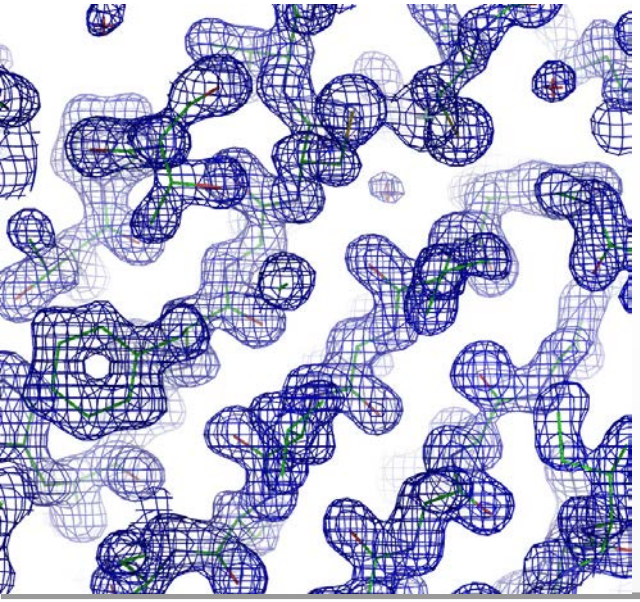


In practice this definition coincides well with the definition known from microscopy, where the limiting resolution is the distance at which two points can be observed as separate.

The Electron Density Equation

X-rays are diffracted by electrons. The distribution of electrons in space, the **electron density function** $\rho(\mathbf{r})$ is obtained as a discrete Fourier transform of the structure factors.

$$\rho(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{S}} |\mathbf{F}(\mathbf{S})| \exp(-2\pi i \mathbf{r} \cdot \mathbf{S} - i\varphi(\mathbf{S}))$$



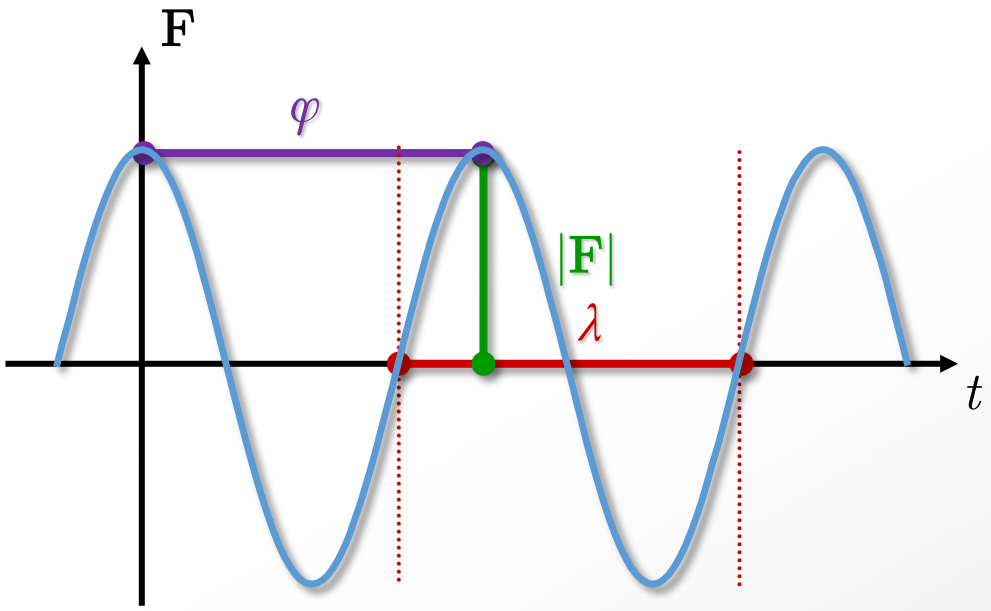
At each point $\mathbf{r} = (x,y,z)$ in real space, the electron density $\rho(\mathbf{r})$ is calculated from **all** structure factors $\mathbf{F}(\mathbf{S})$.

Note: This works, although $\rho(\mathbf{r})$ is a continuous function, while $\mathbf{F}(\mathbf{S})$ is a discrete function!

Wave functions

A simple, periodic wave function is described by three parameters:

- wavelength λ
- amplitude $|\mathbf{F}|$
- phase angle φ



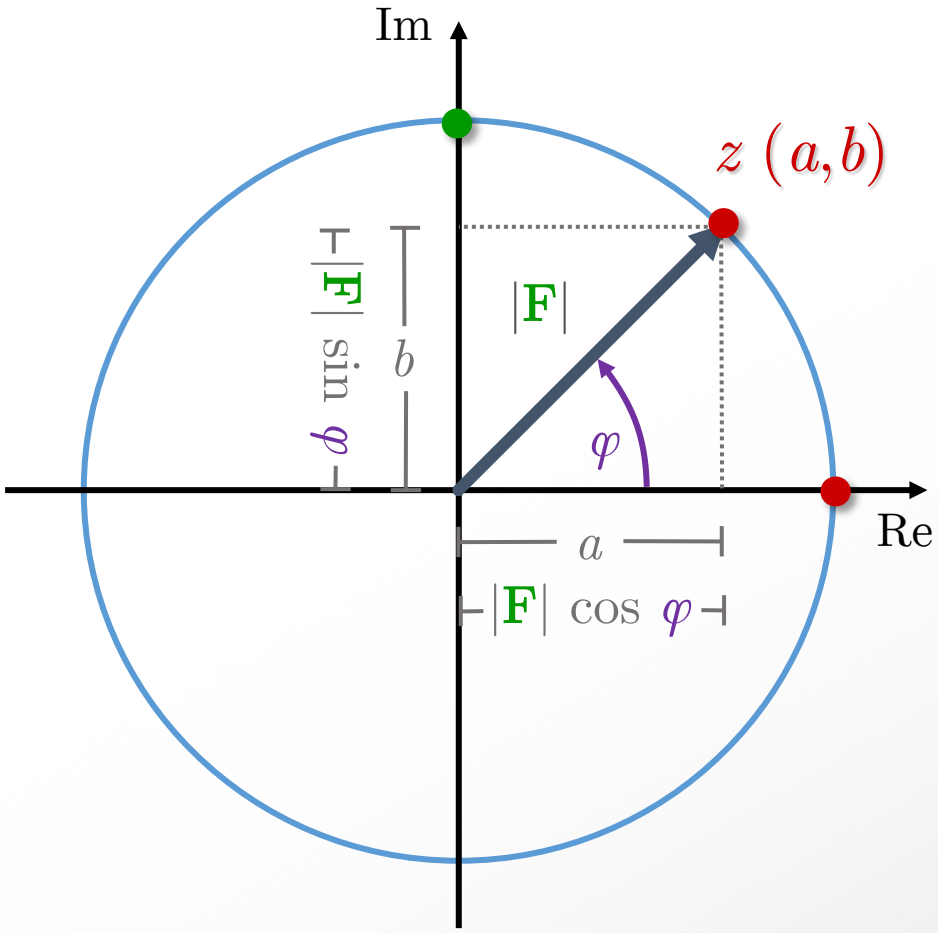
In a diffraction experiment we use monochromatic X-ray radiation, so that λ is constant and known. Every single wave function is therefore fully described by its amplitude and its phase angle:

$$\mathbf{F} = |\mathbf{F}| \cos(\omega t - \varphi)$$

angular velocity $\omega = 2\pi\upsilon = 2\pi\frac{c}{\lambda}$

Describing a Wave Function

Plot of a complex number in the Gaussian plane:
Argand Diagram



A complex number:

$$z = a + i \cdot b$$

with:

a: real part
b: imaginary part

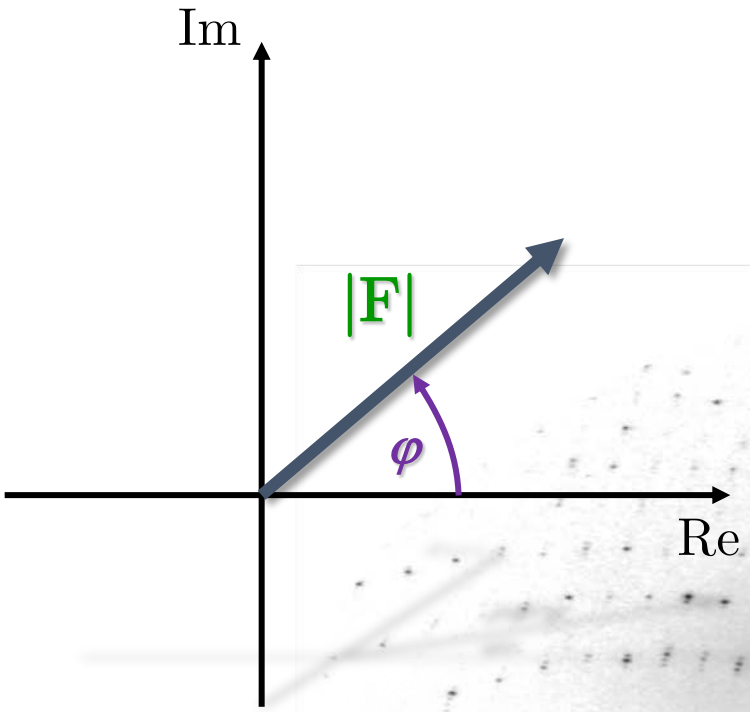
or, accordingly:

$$z = |\mathbf{F}|\ \cos\ \varphi + i\ |\mathbf{F}|\ \sin\ \varphi$$
$$= |\mathbf{F}|\ \exp(i\ \varphi)$$

Leonhard Euler
(1707-1783)

UNI FREIBURG

Structure Factors



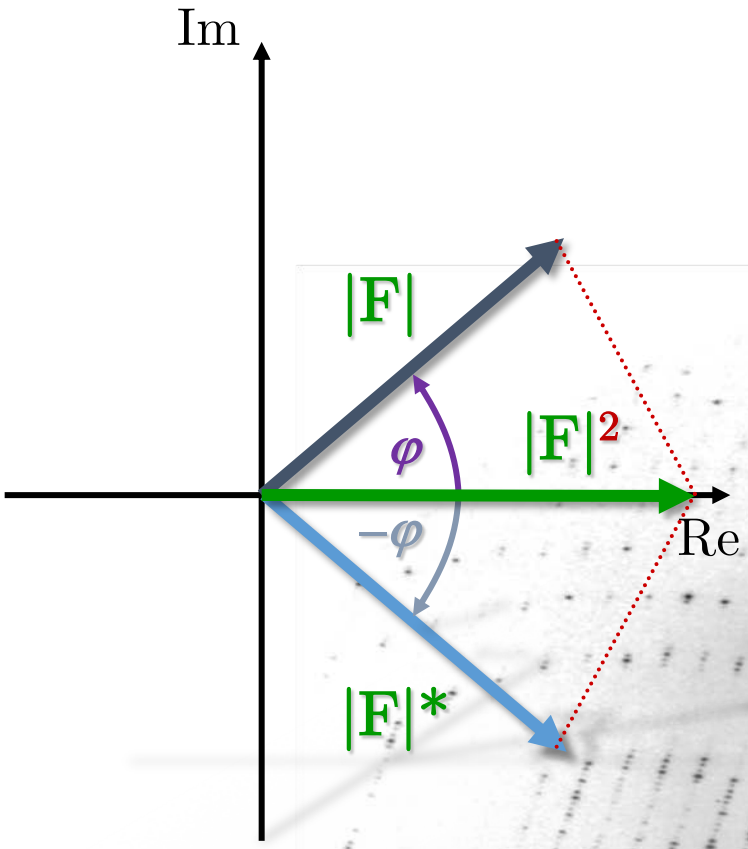
Each individual wave is fully defined an amplitude $|\mathbf{F}|$ and a phase angle φ and can be represented by a vector $\mathbf{F}$ in an Argand diagram.

This vector is called the **structure factor**.

In a diffraction image, the **intensity I** of *each spot* represents the squared value of a structure factor amplitude:

$$\mathbf{I} \approx |\mathbf{F}|^2$$

Not quite all we need



The square of a complex number:

$$|z|^2 = z \cdot z^*$$

is the product of z with its complex conjugate.

$|\mathbf{F}|^2$ is a real number ($\varphi = 0$).

The observed intensity $\mathbf{I}$ contains information on the structure factor amplitude $|\mathbf{F}|$, but not on the phase angle φ. φ can not be determined in a diffraction experiment.

This is known as the phase problem of crystallography.

$$\mathbf{I} \approx |\mathbf{F}|^2$$

The Phase Problem

Electron density is calculated as a Fourier sum of structure factors:

$$\rho(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{S}} |\mathbf{F}(\mathbf{S})| \exp(-2\pi i \mathbf{r} \cdot \mathbf{S} - i\varphi(\mathbf{S}))$$

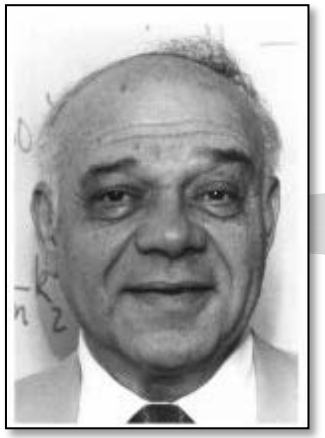
We measure this: $\mathrm{I} \approx |\mathbf{F}|^2$

φ can not be measured.

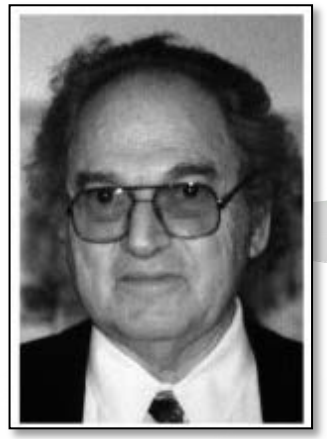
Obviously, crystallographers solve structures:

How is phase information obtained?

The Significance of Phase Information



Jerome Karle
(1918-2013)



Herbert Hauptman
(1917-2011)

Nobel Prize in Chemistry 1985

"for their outstanding achievements in the development of direct methods for the determination of crystal structures"

amplitudes: $|F_{JK}(S)|$

e angles: $\varphi_{JK}(S)$

amplitudes: $|F_{HH}(S)|$

phase angles: $\varphi_{HH}(S)$

?

?

Structural information is strongly determined by the phase angles.

How to Solve the Phase Problem

A. Based on a known structure

=> **Molecular replacement**

Phase information is 'borrowed' from a structurally similar molecule. This model has to be oriented within the unit cell of the crystal to match the position of the unknown protein.

B. *De novo* structure solution

=> **Substructure Analysis**

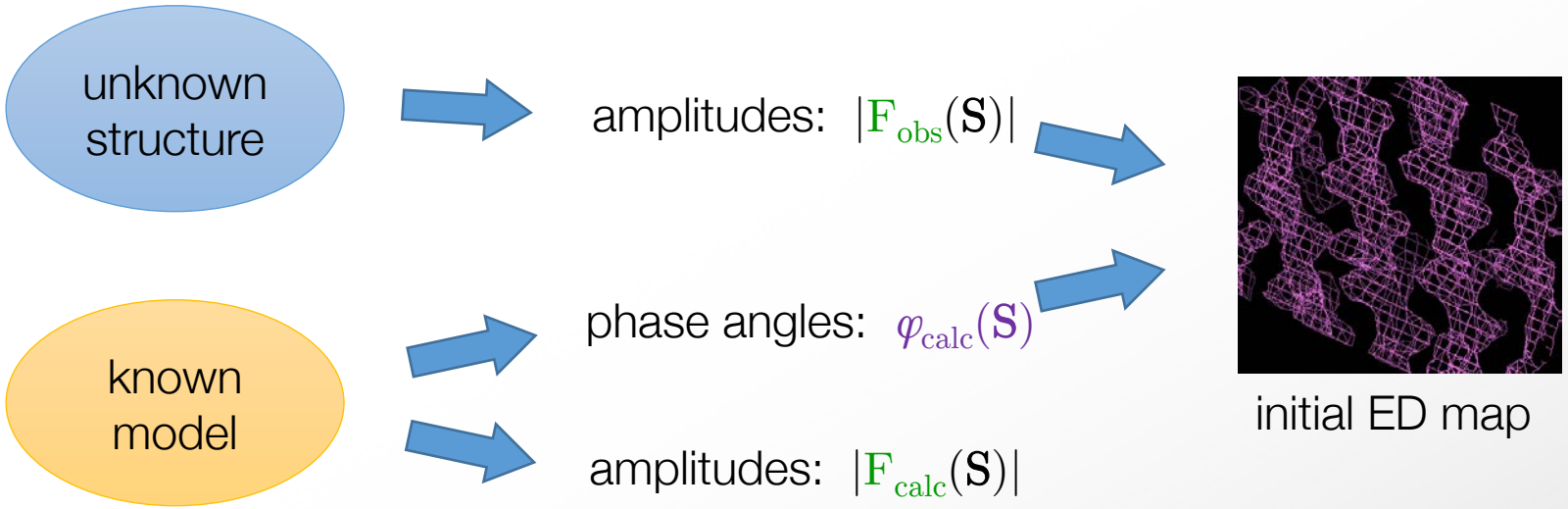
Reduction of the size of the problem: Identify a small subset of atoms to identify their real space positions. Yes: this actually helps to solve the entire structure.

=> **Direct methods**

The actual solution of the phase problem (sic!) – for small molecules (100 atoms).
Jerome Karle & Herbert Hauptman, NP 1985.

Molecular Replacement

The idea: If a sufficiently similar structure is known, its structure factor phases can be calculated and used as a first approximation for the new molecule.



But: Amplitudes and phase angles not only depend on the structure of the molecule, but also on its orientation within the unit cell.

Rossmann & Blow (1962); Crowther (1972)

The Patterson Function

Without phases, the electron density can not be computed. But what can we do?

$$\rho(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{S}} |\mathbf{F}(\mathbf{S})| \exp(-2\pi i \mathbf{r} \cdot \mathbf{S} - i\varphi(\mathbf{S}))$$

If we just use the intensities we measure, we technically square the equation and we lose all complex terms:

$$P(\mathbf{u}) = \frac{1}{V} \sum_{\mathbf{S}} |\mathbf{F}(\mathbf{S})|^2 \cos(2\pi \mathbf{u} \cdot \mathbf{S})$$

This is the Patterson Function. It does not describe electron density in real space and it is no Fourier Transform. To emphasize this we do not use (x,y,z), but instead we use (u,v,w) and call it **Patterson Space**.

- Peaks in an electron density map correspond to **atom positions.**
- Peaks in a Patterson map correspond to **interatomic distance vectors.**

Structure Solution by Substructure Solution

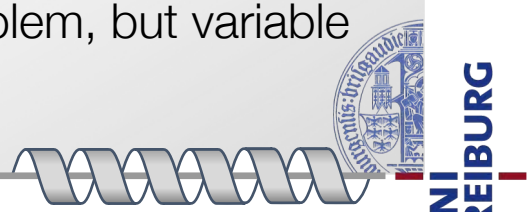
To identify a subset of atoms, they must be discernible from the others.

1. Extrinsic scatterers: Isomorphous Replacement

- can be added to the native protein and might bind specifically
- due to the phase ambiguity, two different derivatives are needed to solve a structure.
- The heavy atoms must not alter the structure and packing, or the difference map will be too noisy to identify the heavy atoms.

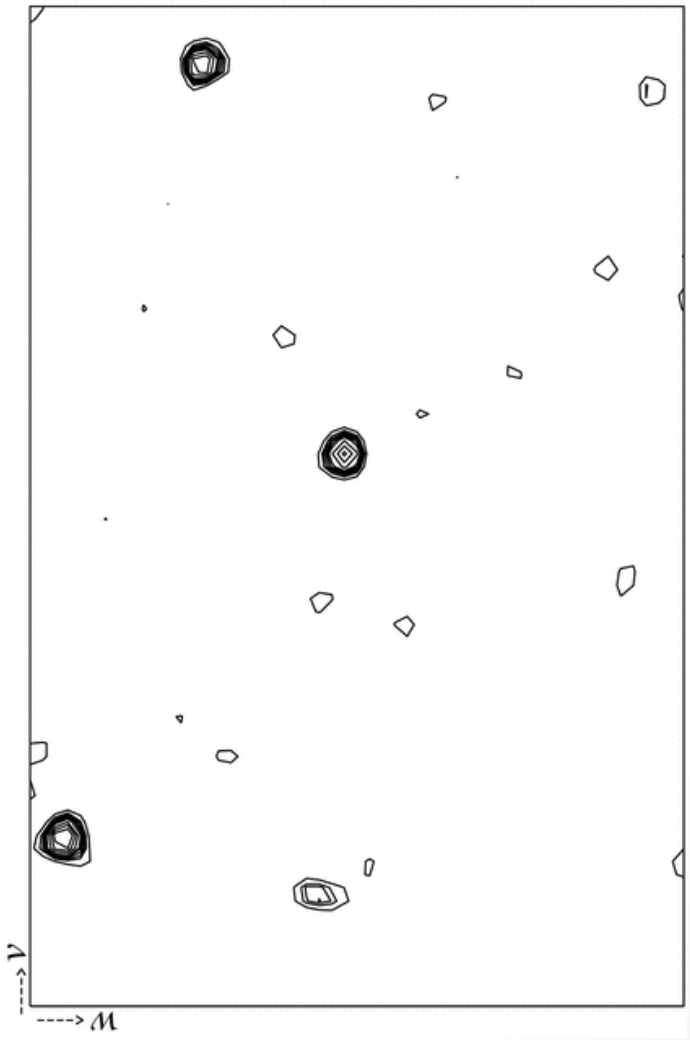
2. Intrinsic scatterers: Anomalous Dispersion

- Data collection at energies close to an absorption edge leads to breakdown of **Friedel's law**: $\Delta_{ano} = |\mathbf{F}(\mathrm{S})| - |\mathbf{F}(-\mathrm{S})|$
- This difference can be measured on a single crystal within a single data set.
- Due to the use of intrinsic atoms, isomorphy is not a problem, but variable X-ray energies, i.e. a synchrotron source, are required.



UNI FREIBURG

An Experimental Patterson Map



A simple substructure is calculated from 'amplitudes' that originate only from the respective subset of atoms.

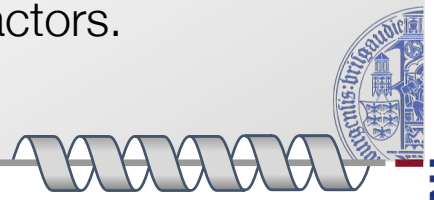
They can be obtained from

dispersive differences: $|\mathbf{F}_{\mathrm{H}}| = |\mathbf{F}_{\mathrm{PH}}| - |\mathbf{F}_{\mathrm{P}}|$

anomalous differences: $\Delta_{ano} = |\mathbf{F}(\mathrm{S})| - |\mathbf{F}(-\mathrm{S})|$

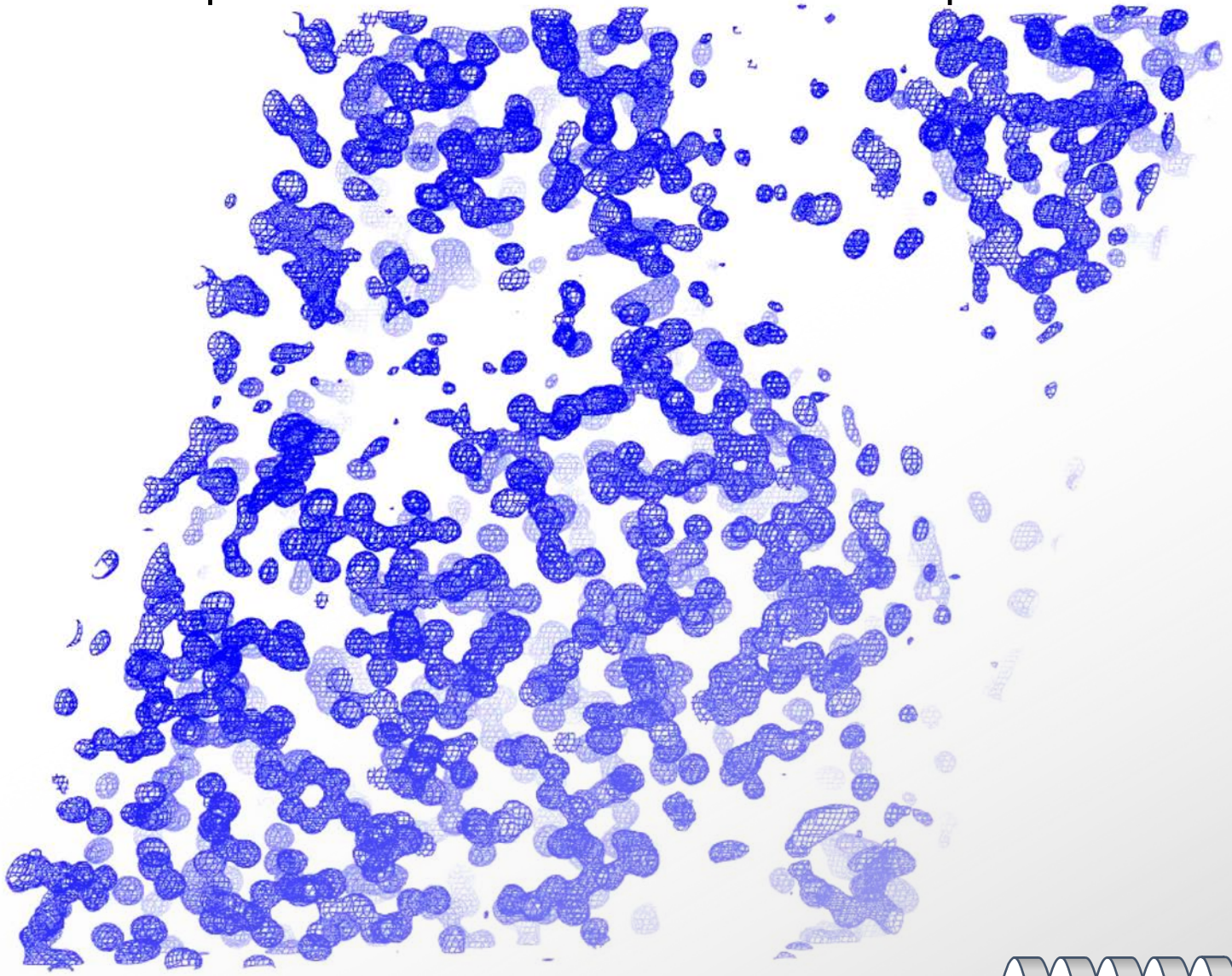
The position of the peaks in Patterson space can be used to derive the actual position in $\mathbf{r}(x,y,z)$ space.

From the position of an atom in the unit cell, we can calculate its structure factors.

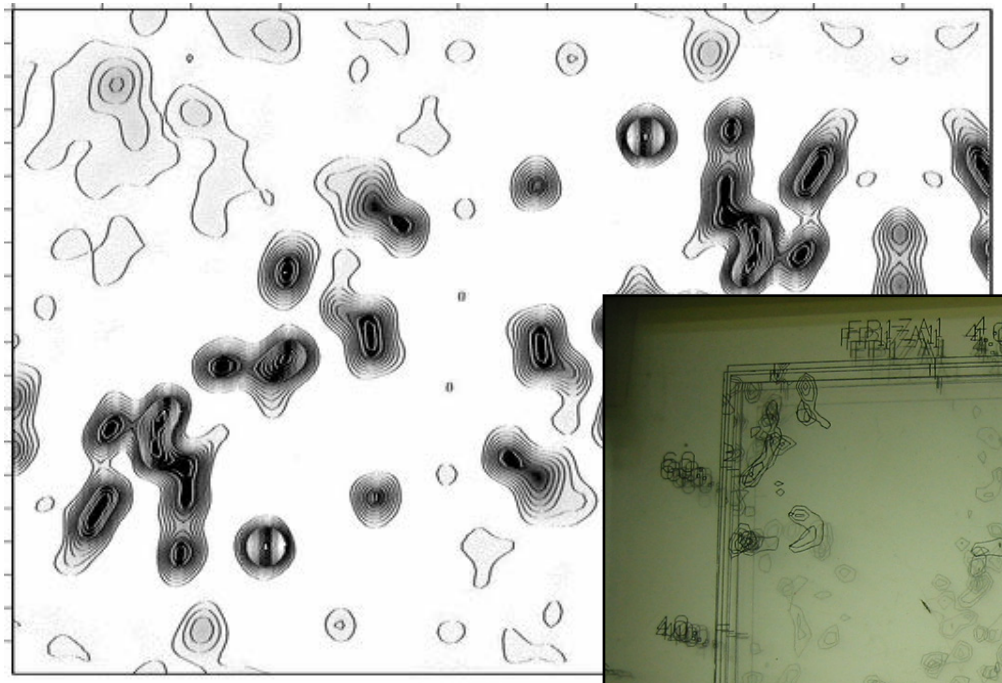


UNI FREIBURG

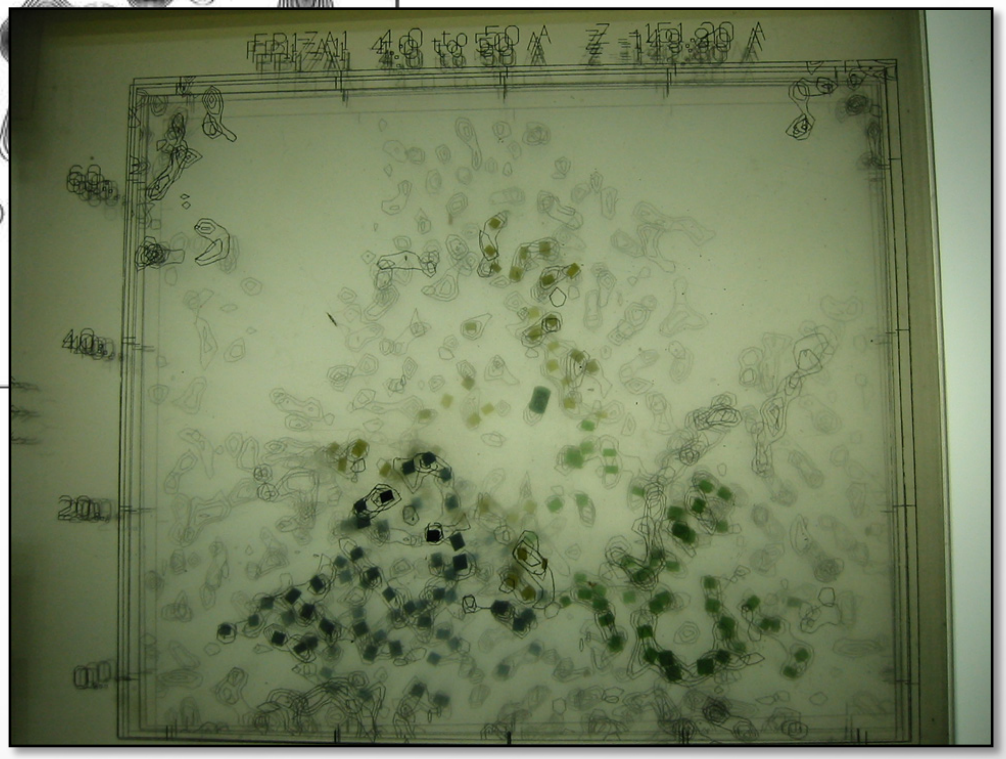
The ED Map: Result of a Diffraction Experiment



Display and Interpretation of Electron Density



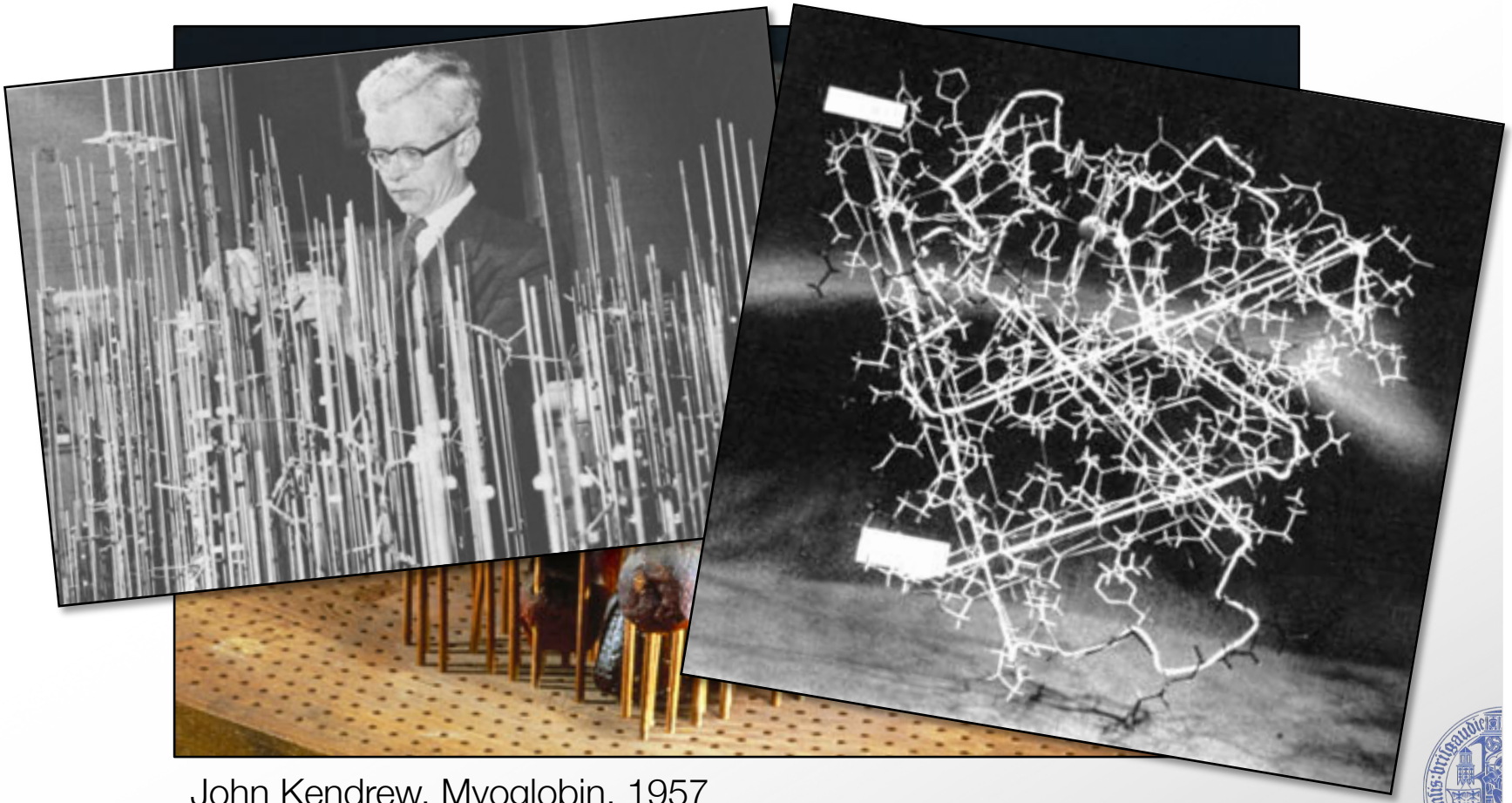
A 2D slice through an electron density map is similar to a topographic map.



In pre-PC times their interpretation used to be somewhat tedious.

The First Structural Model of a Protein

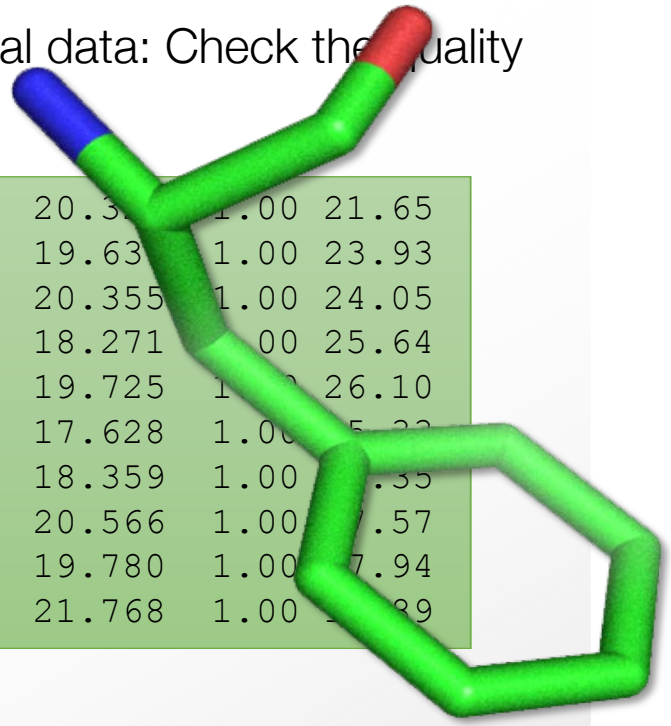
Interpretation of the electron density map involves building a model of the protein. Initially these were physical models.



John Kendrew, Myoglobin, 1957

A Word of Caution: Assessment of Quality

Structure models are **interpretations** of the experimental data: Check the quality of deposited structures!



```
ATOM      1  CB  PHE A   3      17.872  -1.006  20.3   1.00 21.65
ATOM      2  CG  PHE A   3      17.897   0.332  19.63  1.00 23.93
ATOM      3  CD1 PHE A   3      18.147   1.495  20.355 1.00 24.05
ATOM      4  CD2 PHE A   3      17.646   0.429  18.271  00 25.64
ATOM      5  CE1 PHE A   3      18.146   2.739  19.725  1  26.10
ATOM      6  CE2 PHE A   3      17.642   1.671  17.628  1.00
ATOM      7  CZ  PHE A   3      17.893   2.828  18.359  1.00
ATOM      8  C   PHE A   3      15.410  -0.790  20.566  1.00  .57
ATOM      9  O   PHE A   3      14.860  -1.565  19.780  1.00  .94
ATOM     10  N   PHE A   3      16.616  -2.596  21.768  1.00  .89
```

Quality indicators

R_{cryst}	crystallographic residual	< 0.2
R_{free}	cross-validation residual	$R_{\text{free}} - R_{\text{cryst}} < 0.05$
$R_{\text{p.i.m.}}$	precision-independent merging R	0.02–0.05
$I/\sigma(I)$	signal/noise ratio	> 1.0
$\text{CC}_{1/2}$	half-width correlation coefficient	> 0.33

Selected Further Reading

Rupp, B. (2009) Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology.
1st ed., Taylor & Francis, Oxford, UK.

Rhodes, G.(2008) Crystallography Made Crystal Clear.
2nd ed., Academic Press, New York, USA.

Messerschmidt, A. (2006) X-Ray Crystallography of Biomacromolecules.
1st ed., Wiley-VCH, Weinheim, Germany.

Glusker, J. P., Lewis, M. & Rossi, M. (1994) Crystal Structure Analysis for Chemists and Biologists.
1st ed., Wiley-VCH, Weinheim, Germany.

Drenth, J. (1994) Principles of Protein X-Ray Crystallography.
1st ed., Springer, New York, USA.