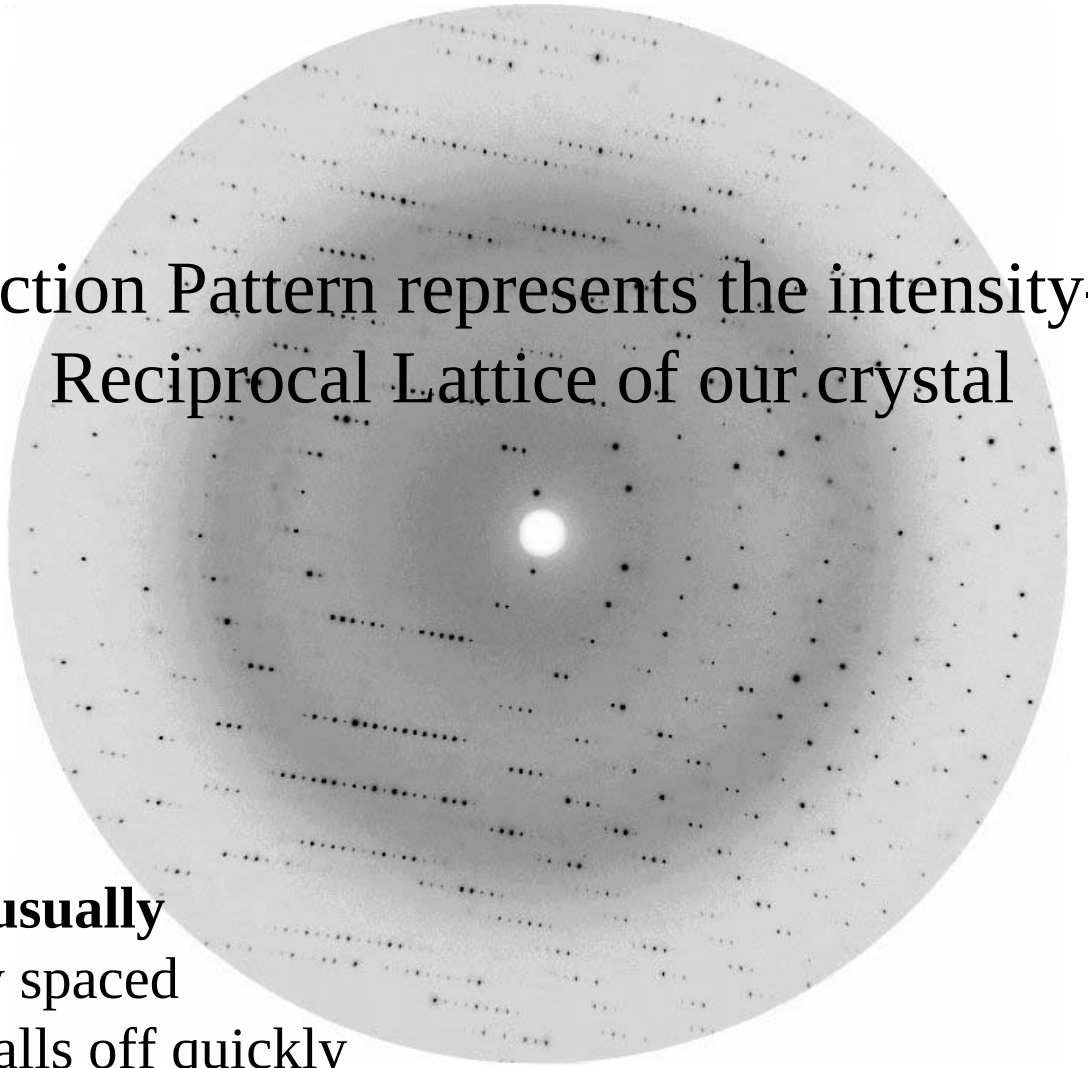
Crystallographic studies of biological macromolecules: from atomic resolution to molecular giants

Mariusz Jaskolski

Department of Crystallography, Faculty of Chemistry, UAM
Center for Biocrystallographic Research, IBCh PAN
Poznan

Protein crystal X-ray diffraction pattern



The Diffraction Pattern represents the intensity-weighted Reciprocal Lattice of our crystal

The reflections usually

- are very closely spaced
- their intensity falls off quickly
- are on strong background visible as a dark ring

The Reciprocal Lattice

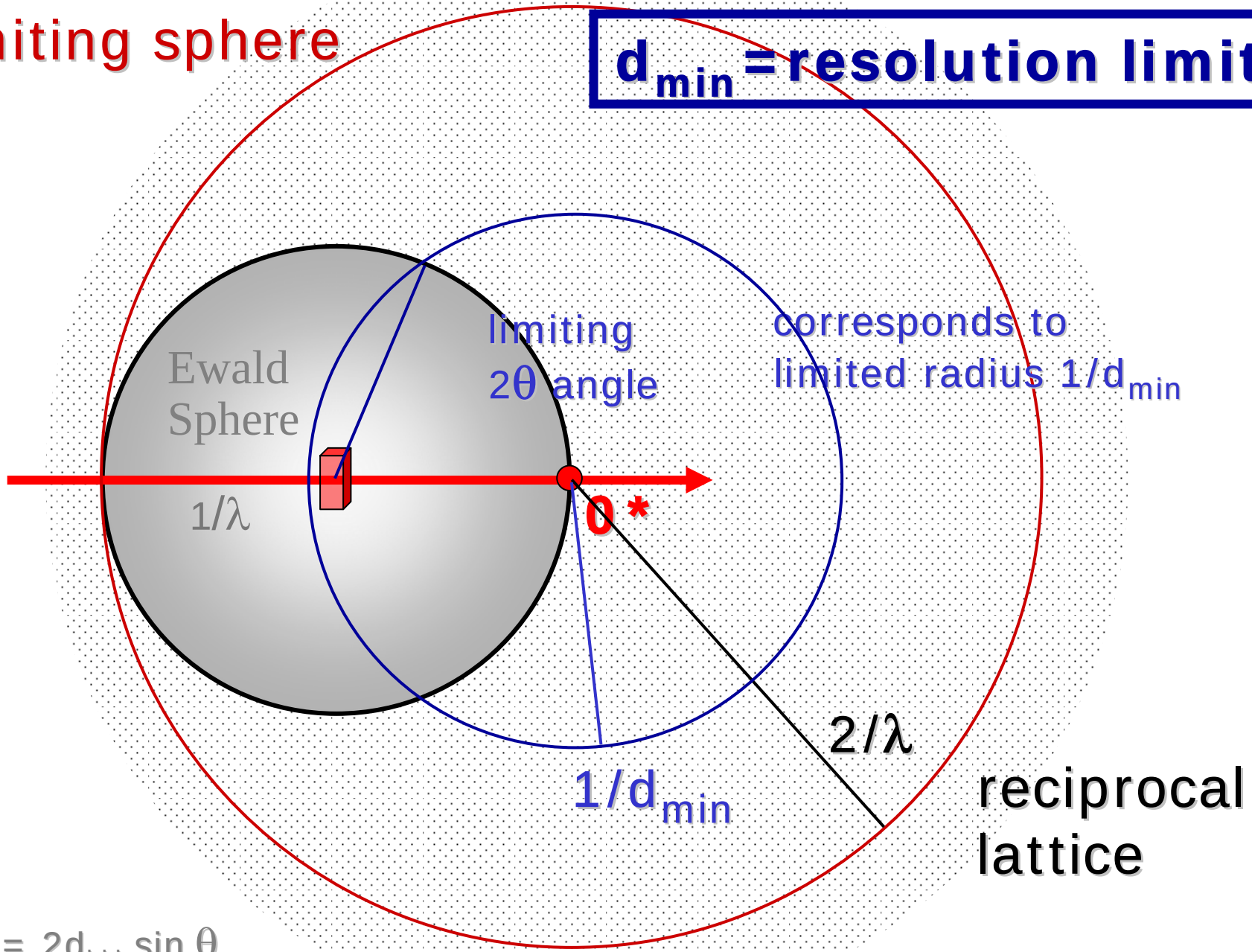
collection of points *hkl* associated with planes (hkl) in direct lattice

collection of nodes generated by basis vectors **a* b* c*** defined by:

- direct-lattice planes (100) (010) (001)
- vector products in direct lattice, e.g. $\mathbf{c^*} = (\mathbf{a} \times \mathbf{b})/V$
- mixed scalar products: $\mathbf{a}_i \cdot \mathbf{a}_j^* = \delta_{ij}$

limiting sphere

d_{min} = resolution limit

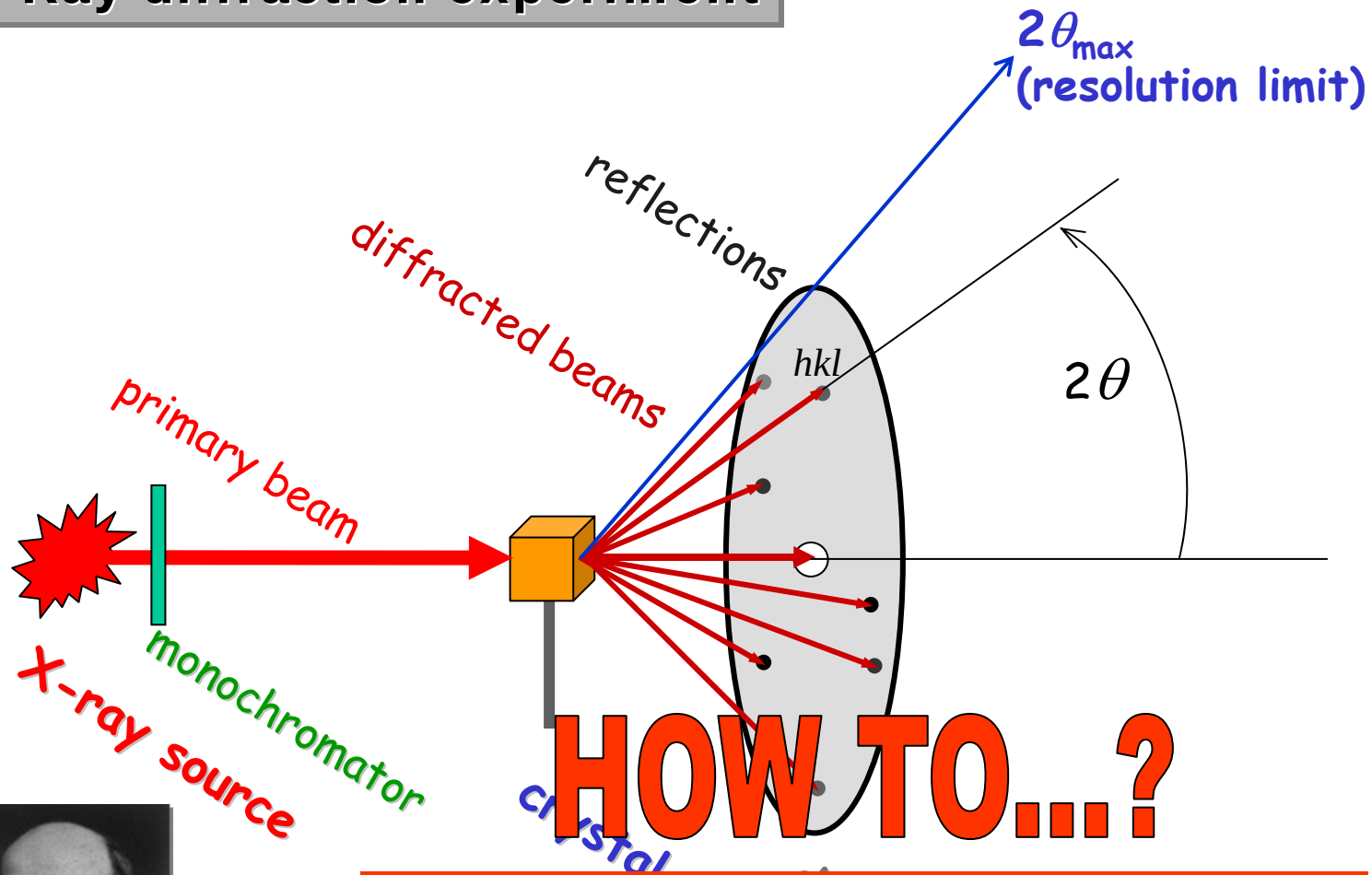


$\lambda = 2d_{hkl} \sin \theta$

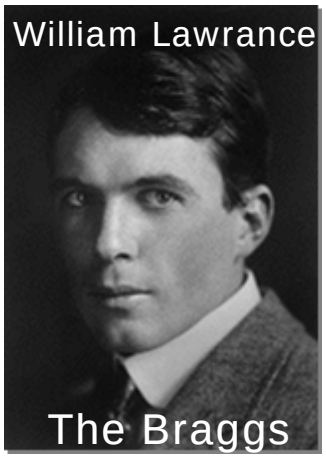
X-Ray diffraction experiment

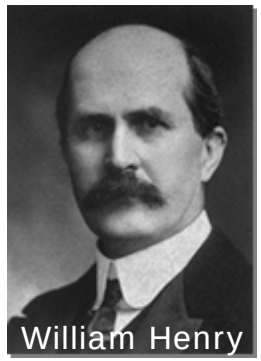
1912

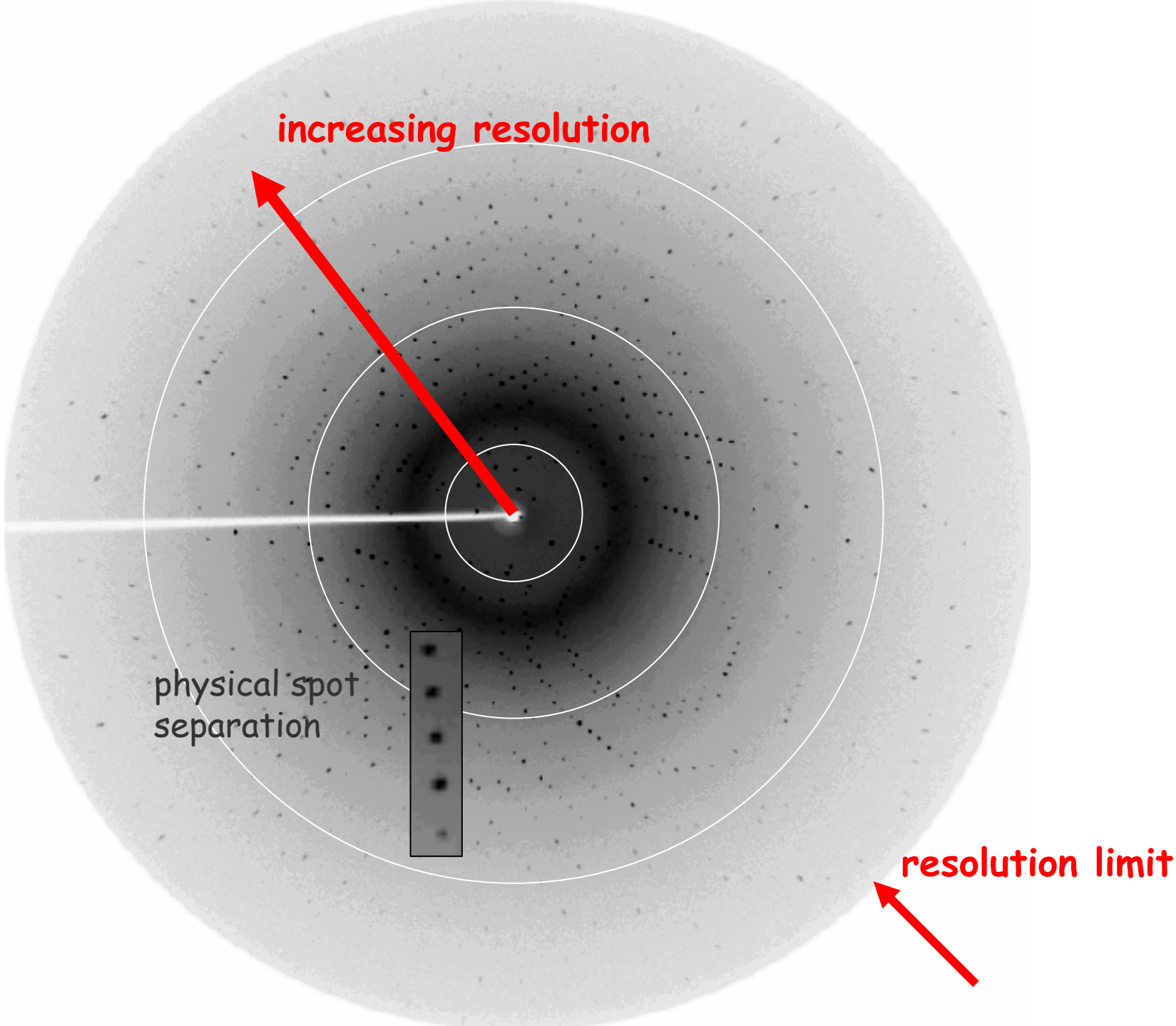
$\lambda = 2d_{hkl} \sin \theta$



- capture more reflections on the detector
- change separation between reflections






increasing resolution
physical spot separation
resolution limit

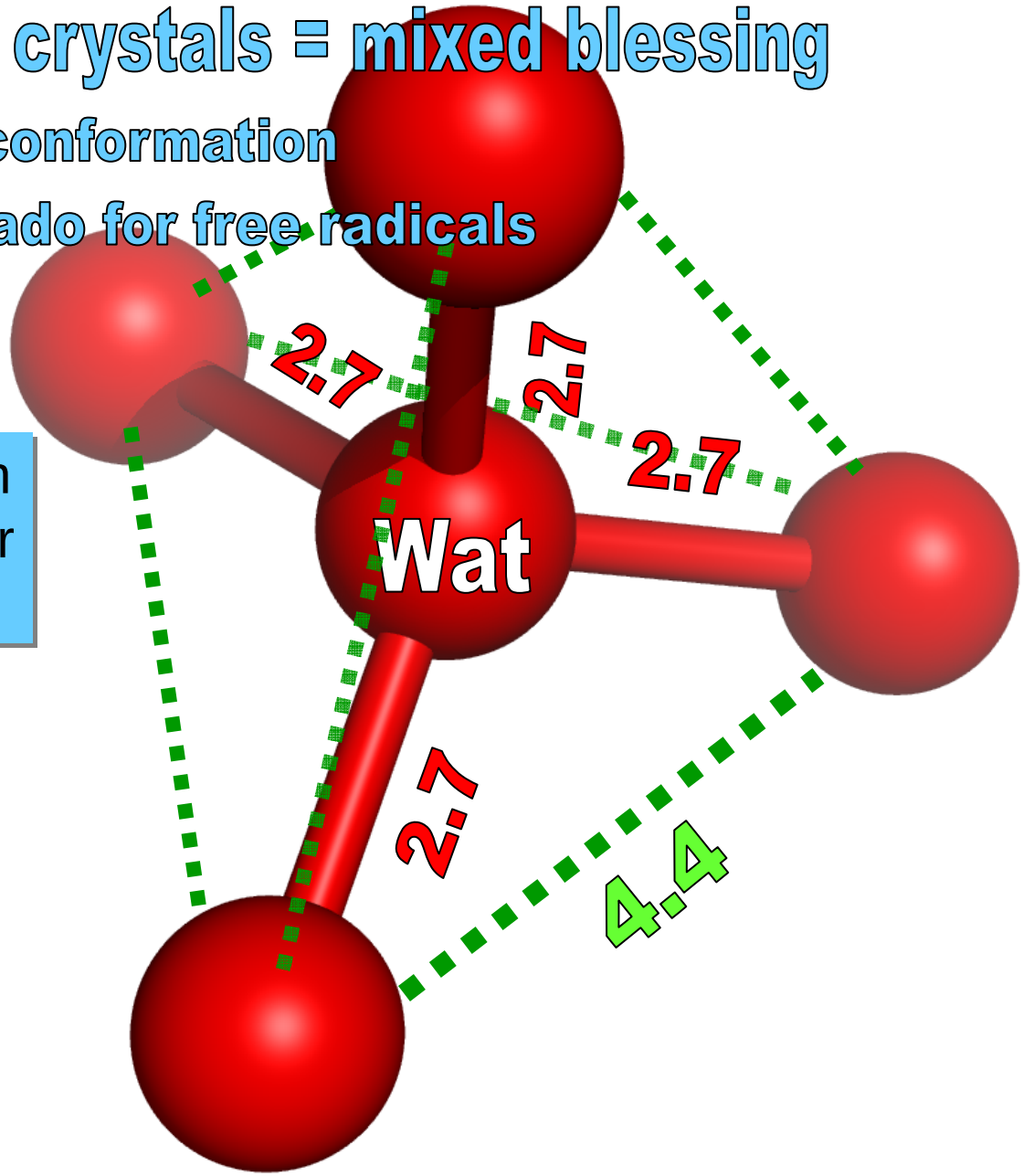
Water in protein crystals = mixed blessing

+ native protein conformation

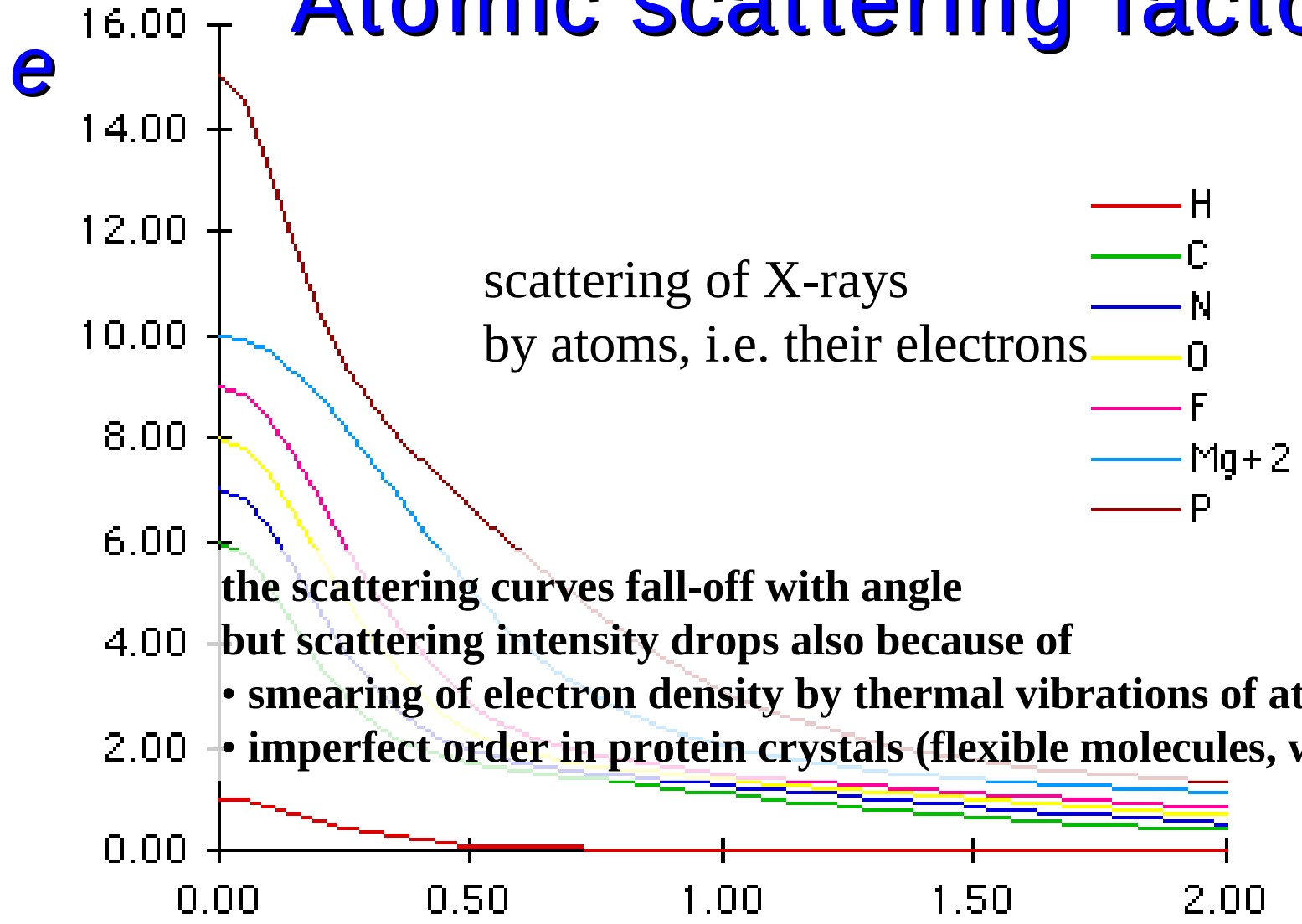
- disorder, El Dorado for free radicals

Protein crystals contain a large volume of water (about 50%)

Water molecule with four neighbors H-bonded in a tetrahedral fashion. Variously oriented 1-3 O...O distances 4.4 Å long will abound.



Atomic scattering factors



scattering of X-rays
by atoms, i.e. their electrons

the scattering curves fall-off with angle
but scattering intensity drops also because of
- **smearing of electron density by thermal vibrations of atoms**
- **imperfect order in protein crystals (flexible molecules, water)**

Jean Baptiste Joseph Fourier
1768-1830

The relation between

electron density distribution

(= crystal structure)

and

the X-ray diffraction pattern

is governed by

Fourier Transformation

To be able to determine this electron distribution

- or electron density -

(in the form of electron density maps)

is of great importance to us:

in chemistry – everything is explained by electrons:

the character of different atoms

and the bonds between them in molecules

Scattering of X-rays by matter

- X-rays are scattered by electrons
- the Amplitude of scattering is proportional to the number of electrons

$$\mathbf{F}(\mathbf{r}^*) = \int_V \rho(\mathbf{r}) \exp[2\pi i(\mathbf{r}^*.\mathbf{r})]\, d\mathbf{r}$$

ρ is the density of electrons at point **r** within a volume *V*

r* is a point in the reciprocal space definig the direction of scattering

F(r*) is called the Structure Factor

- the Intensity of scattering is proportional to squared Amplitude

$$\mathbf{I}(\mathbf{r}^*) \sim \mathbf{F}^2(\mathbf{r}^*)$$

The Fourier Transform

the expression

$$\mathbf{F}(\mathbf{r}^*) = \int_v \rho(\mathbf{r}) \exp[2\pi i(\mathbf{r}^*.\mathbf{r})]\, d\mathbf{r}$$

is at the heart of a mathematical theory formulated by **Jean Baptiste Joseph Fourier** *(1768-1830)*

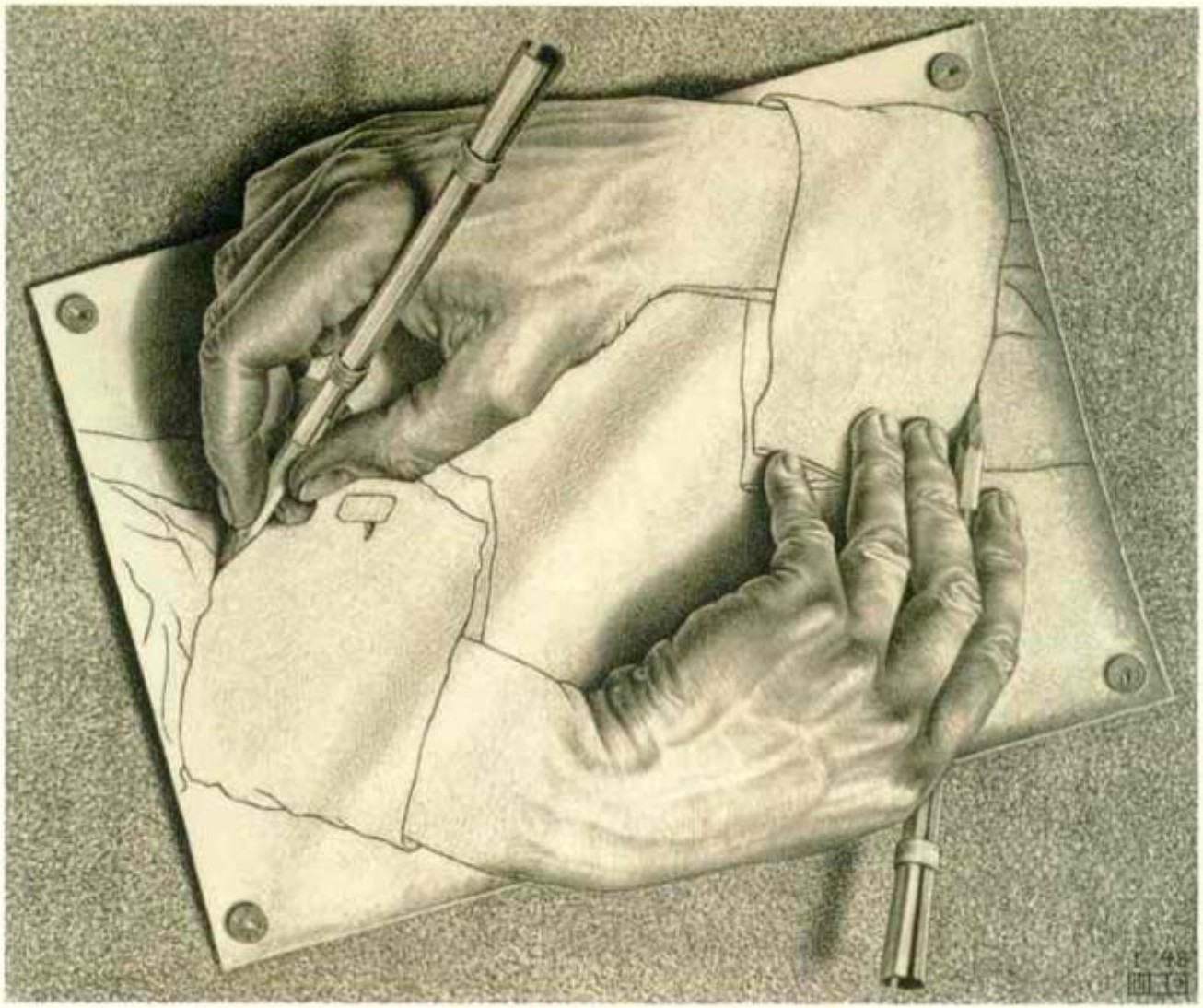
It defines, that the (complex) function

$\mathbf{F}(\mathbf{r}^*)$ ***is the*** **Fourier Transform** ***of the function*** **$\rho(\mathbf{r})$**

Simultaneously, an almost identical expression defines $\rho(\mathbf{r})$ *in terms of* $\mathbf{F}(\mathbf{r}^*)$

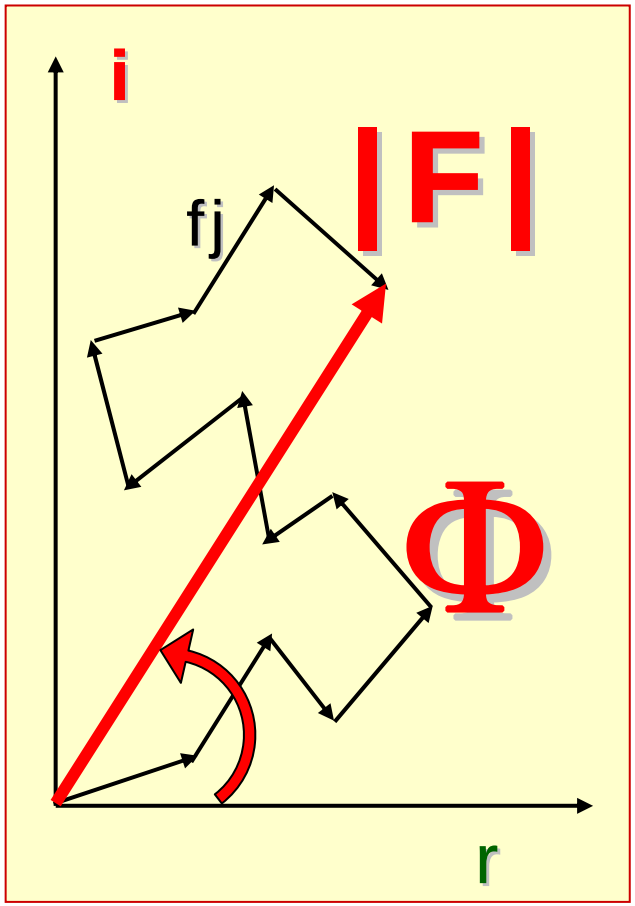
$$\rho(\mathbf{r}) = \int_{v^*} \mathbf{F}(\mathbf{r}^*) \exp[-2\pi i(\mathbf{r}^*.\mathbf{r})]\, d\mathbf{r}^*$$

Fourier Transform



Maurits Cornelis Escher (1898-1972) *Drawing Hands* *1948*

Structure Factor $F^2(hkl)$ = Intensity and electron density $\rho(xyz)$



$$F(hkl) = \sum fj \exp[2\pi i(hxj + kyj + lzj)]$$

$$\rho(xyz) = V^* \sum F(hkl) \exp[-2\pi i(hx + ky + lz)]$$

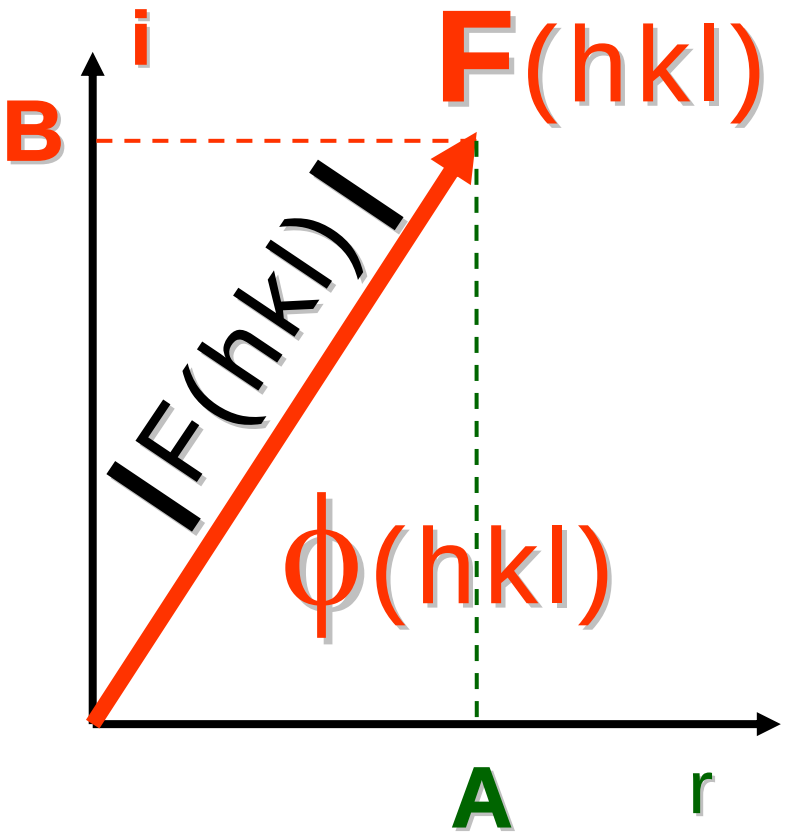
Fourier Transforms

$$F(hkl) = |F| . \exp(i\phi)$$

$$|F(hkl)| = \sqrt{I(hkl)}$$

The Φase Problem

The Structure Factor is a complex entity



$$\mathbf{F}(hkl) = A + iB$$

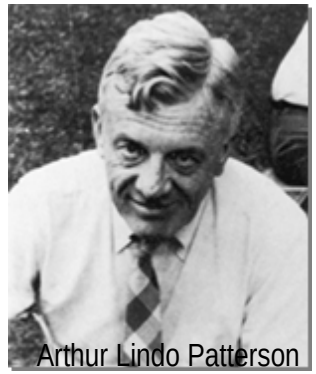
$$= |F(hkl)|.(\cos\phi + i\sin\phi)$$

$$= |F(hkl)|.\exp(i\phi)$$

Euler notation

magnitude *phase*

The Patterson Function


Arthur Lindo Patterson

- Fourier transform of the squares of the structure factors, i.e. of reflection intensities:

$$P(\mathbf{u}) = \Sigma|F(\mathbf{h})|^2 \exp[-2\pi i(\mathbf{h}.\mathbf{u})]$$

- the distribution of atoms convoluted with its centrosymmetric image (autocorrelation function):

$$P(u) = \rho(x) \otimes \rho(-x) = \int \rho(x)\, \rho(x-u)\, dx$$

- the peaks represent interatomic vectors
- each pair of atoms has a unique Patterson peak
- the peak height is the product of atomic numbers
- very complex: N^2 Patterson peaks for an N-atom structure

Protein crystallographic methods for solving the **Phase Problem**

all rely to some extent on the Patterson function

- Isomorphous Replacement (MIR/SIR)
  (Max Perutz)
- **Molecular Replacement (MR)**
  (Michael Rossmann)
- Multiwavelength Anomalous Diffraction (MAD)
  (Wayne Hendrickson)

An electron density map is the primary product of an X-ray diffraction study of a protein crystal. Everything else is its interpretation.

Electron density maps are generated via Fourier Transformation. This is only possible when suitable Phases can be attributed to the experimental Structure Factor Amplitudes

There are two fundamental ways to obtain the Phases:

Through additional
experimental measurements
in the methods of
Isomorphous Replacement (MIR) or
Multiwavelength Anomalous Diffraction (MAD)

Experimental Phases

Through
calculations
using an **assumed model**
in the method of
Molecular Replacement (MR)

Model-derived Phases can lead to model bias

Molecular Replacement (Michael Rossmann)

MR

- Atomic model of a similar protein

  becomes a probe with which

  we can solve the unknown crystal structure

  of a new protein

Main steps of the method of Molecular Replacement

we must have a sufficiently similar molecular model for the unknown crystal structure

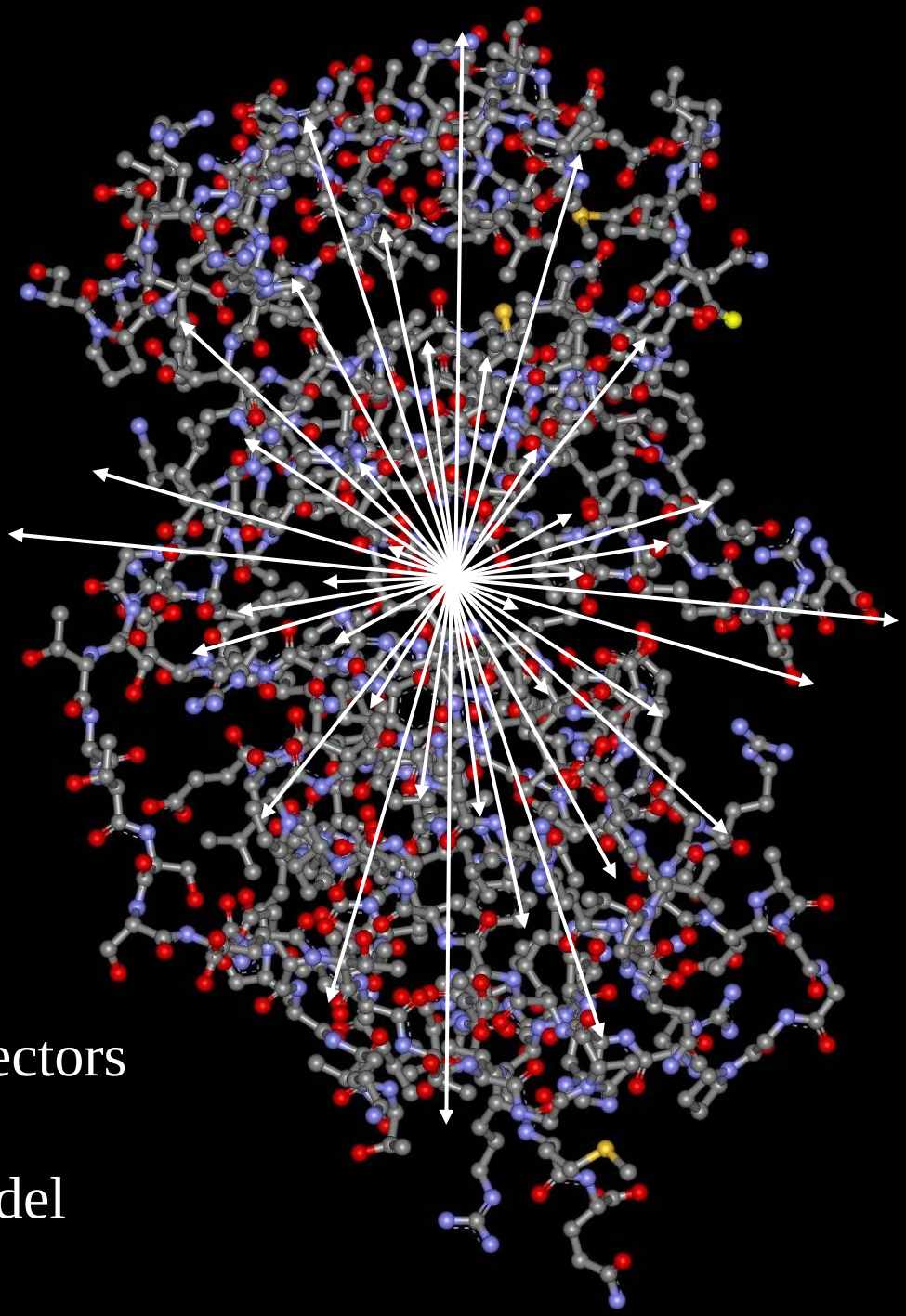
we do not know the crystal architecture, i.e. the orientation and placement of the model in the unit cell are unknown

calculate the experimental Pattersona map, Po, based on Fo^2, for the unknown structure

generate a set of interatomic vectors, Pm, from the model

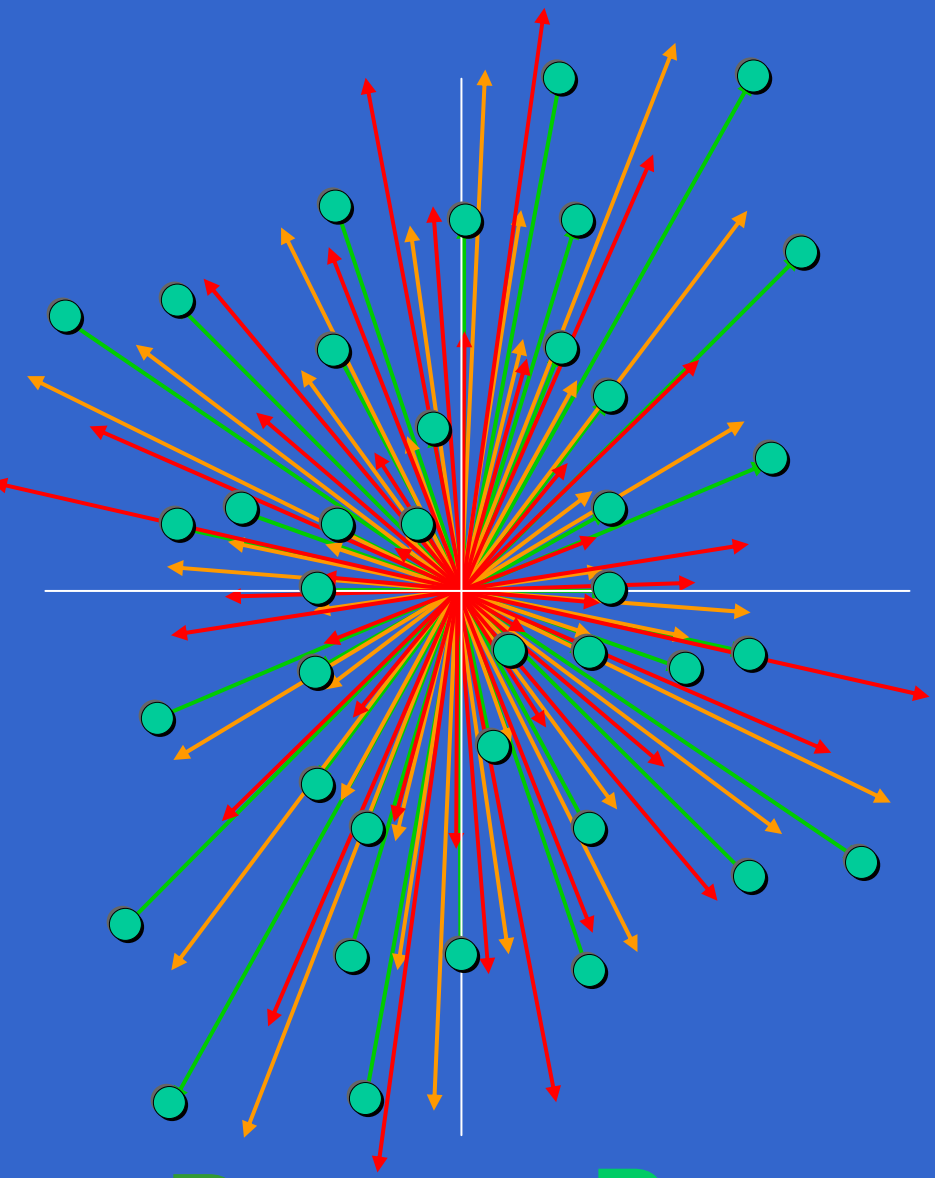
find the orientation of Pm when superposed on the Po map

find the position of the model relative to the unit cell origin in the unknown crystal structure



interatomic vectors
derived from
an atomic model

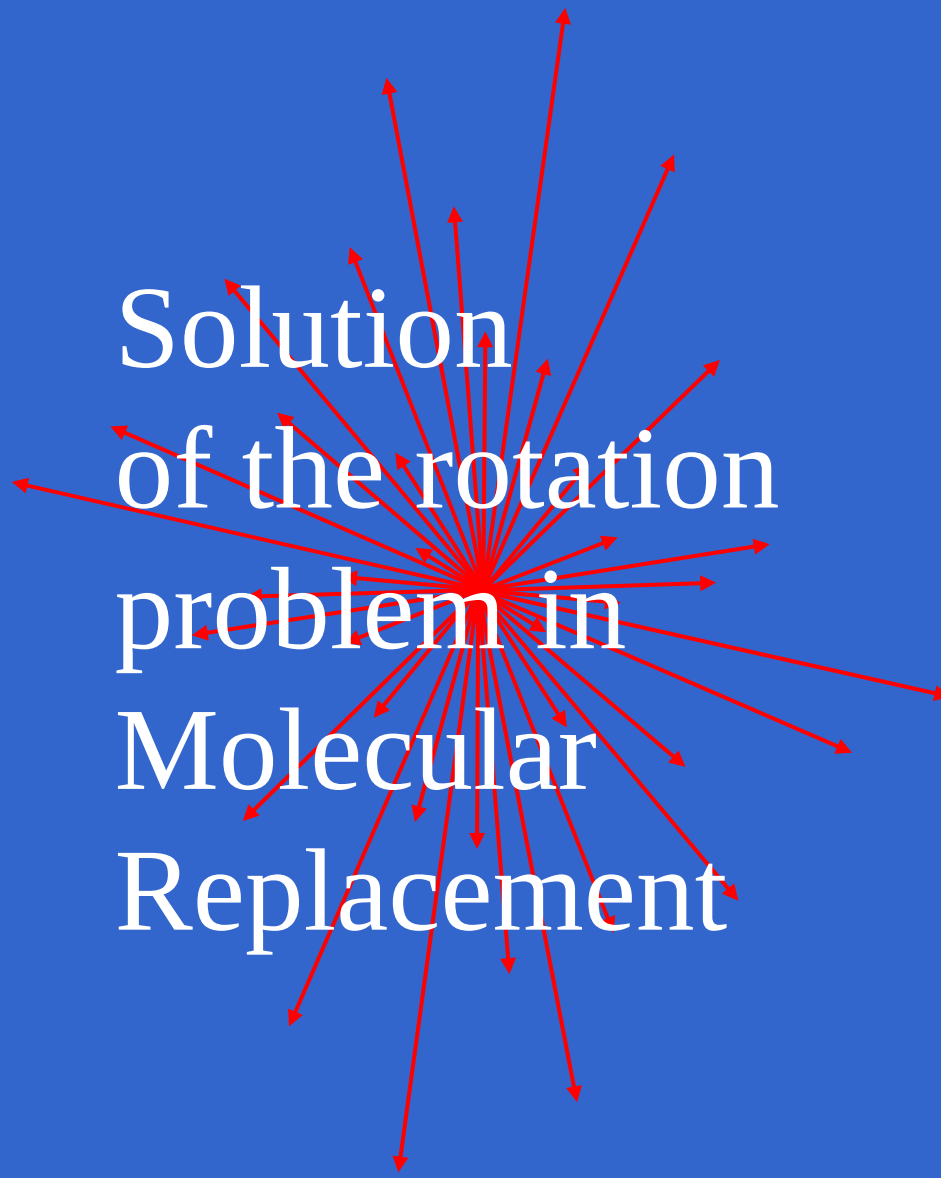
Pm



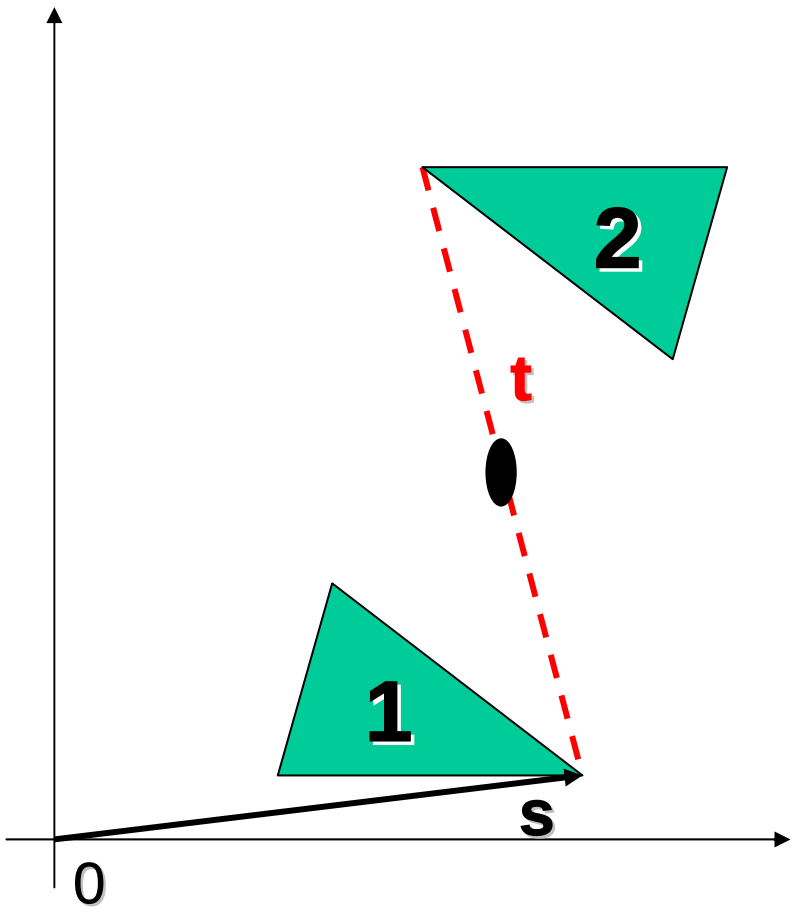
Po

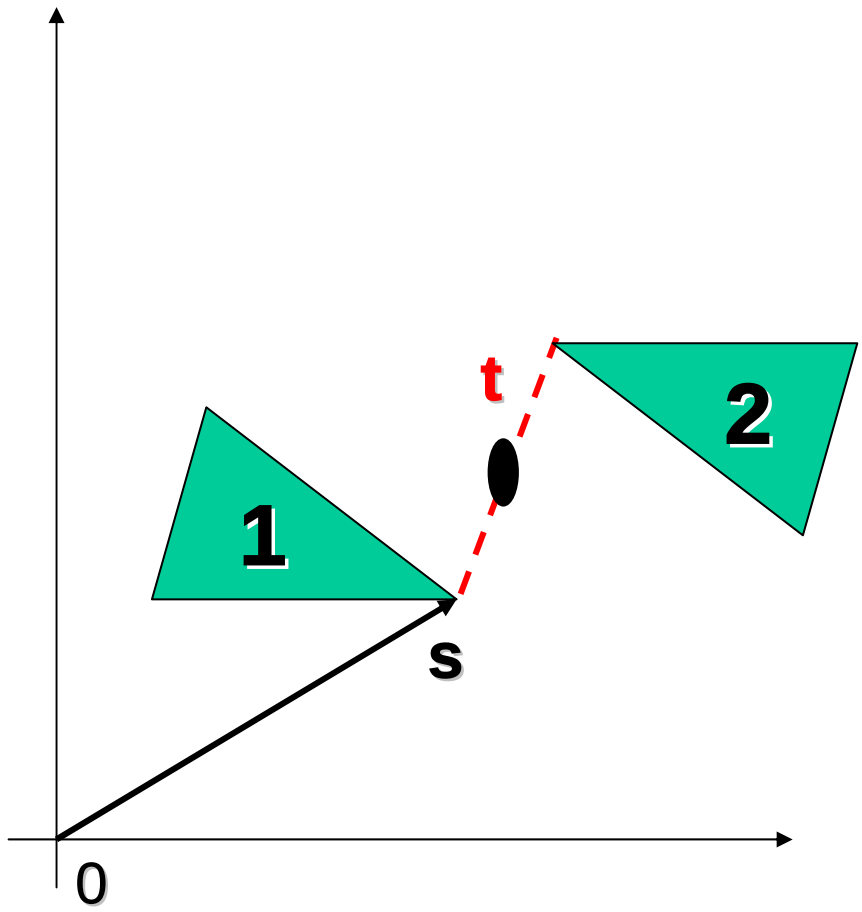
Pm

Solution of the rotation problem in Molecular Replacement



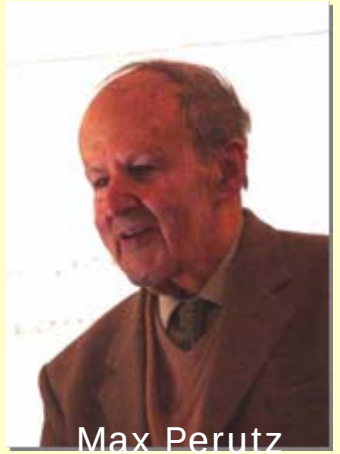
Pm





Translation **s** of a correctly oriented model changes the intermolecular vectors **t** generated by symmetry

Isomorphous Replacement (Max Perutz)



SIR/MIR

- A few heavy atoms in an isomorphously derivatized protein crystal become markers of electron density in the structure – they are the starting point for deciphering of the entire structure

Main steps of the Method of Isomorphous Replacement

introduce to the crystal lattice, without distorting its structure, a few electron-rich atoms

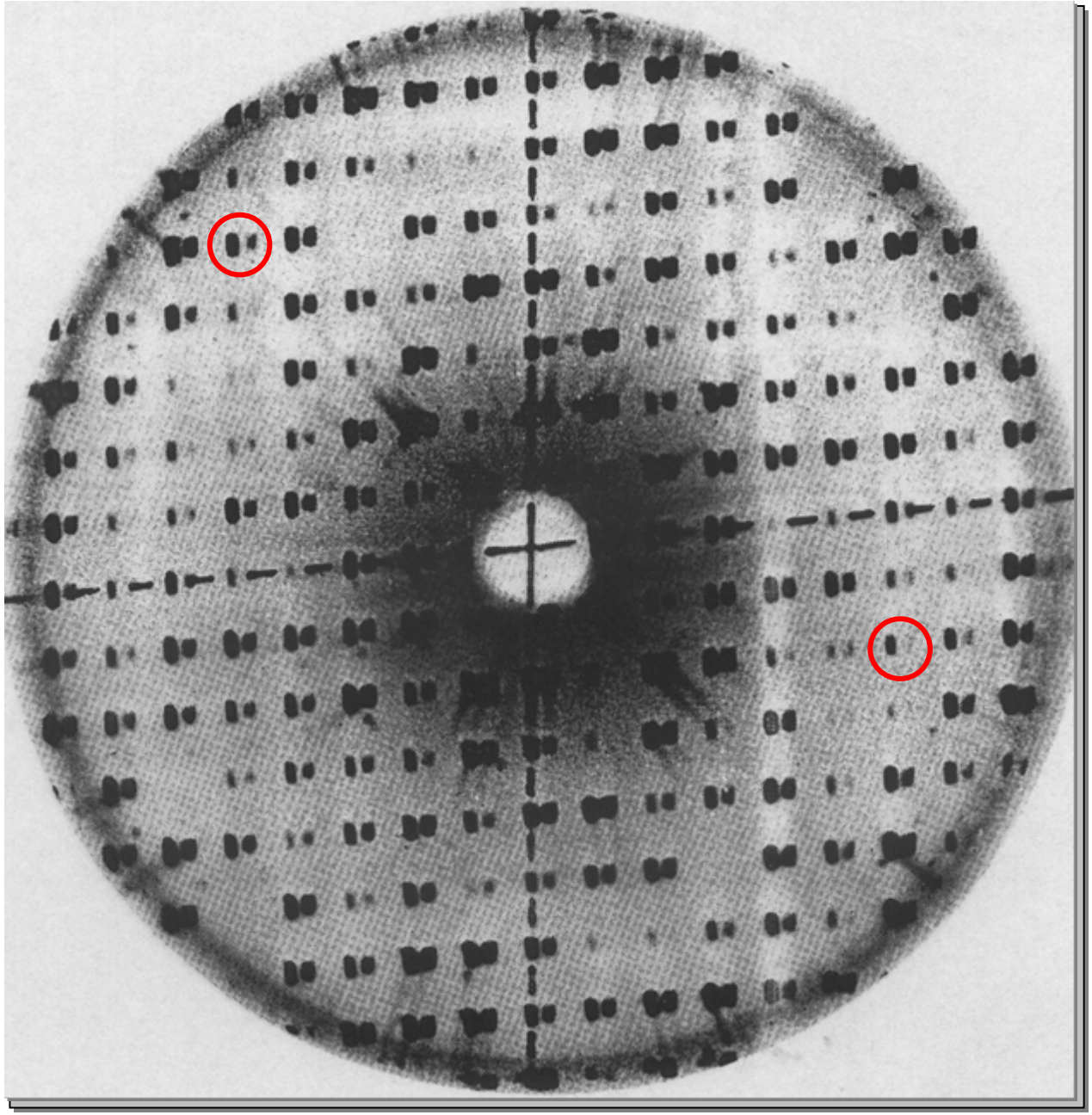
compare the diffraction of the native crystal and of the isomorphous derivative

determine the location of the heavy atoms in the structure

use the information about the heavy atoms to estimate the phases of the reflections

F_{PH}

Heavy atom derivative

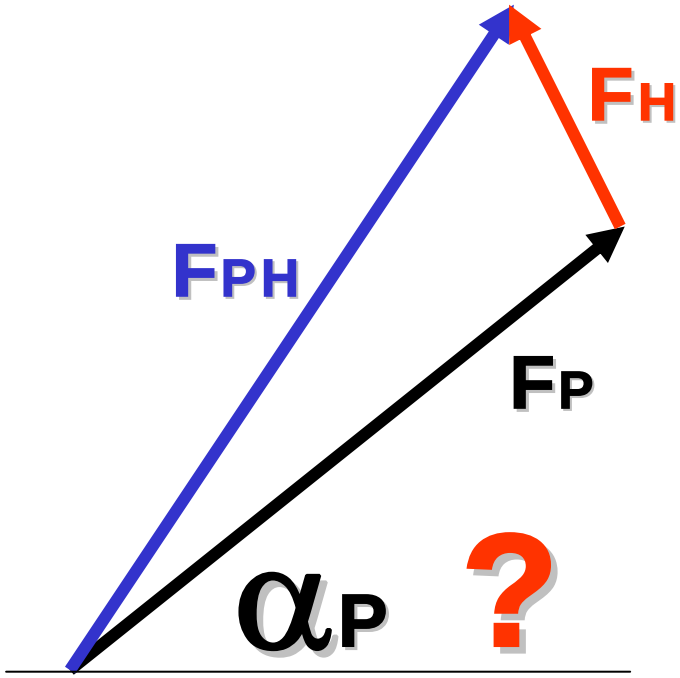


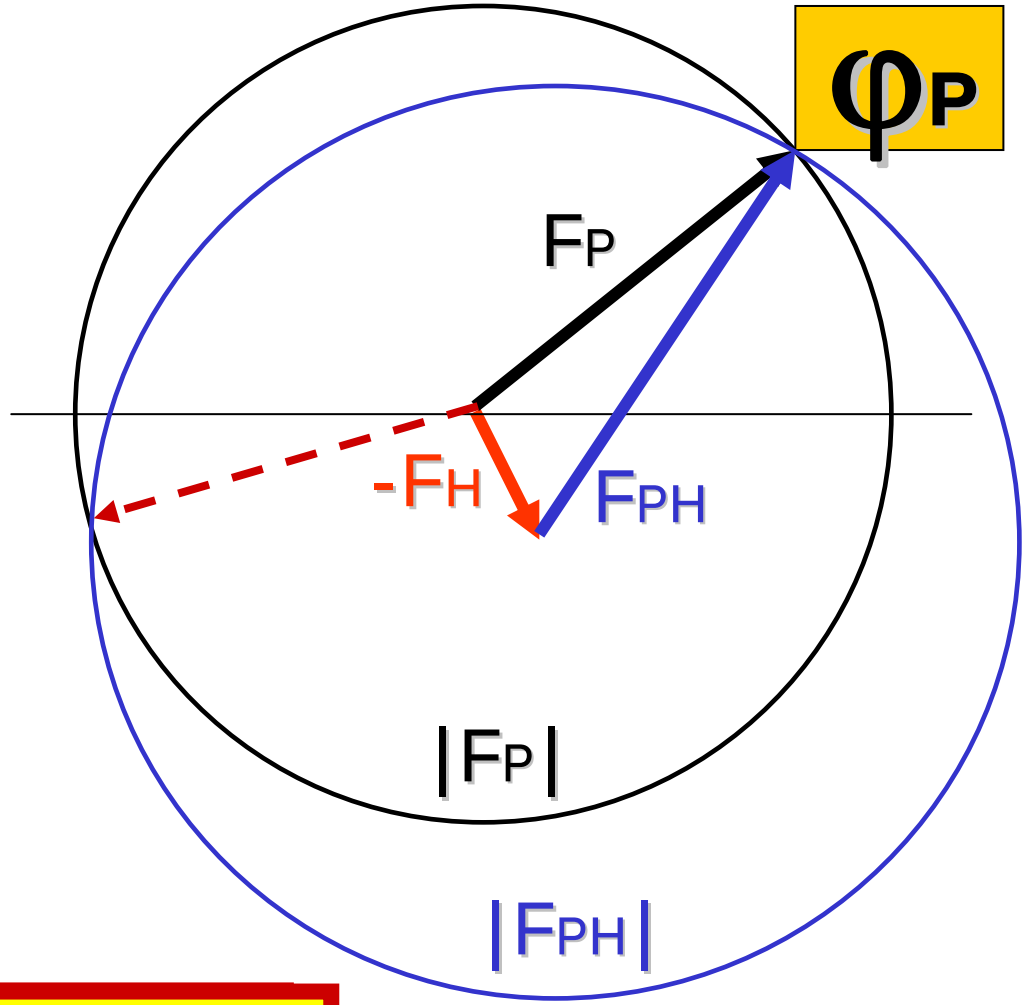
F_P

Native crystal

Two X-ray photographs superposed

Isomorphous replacement
SIR



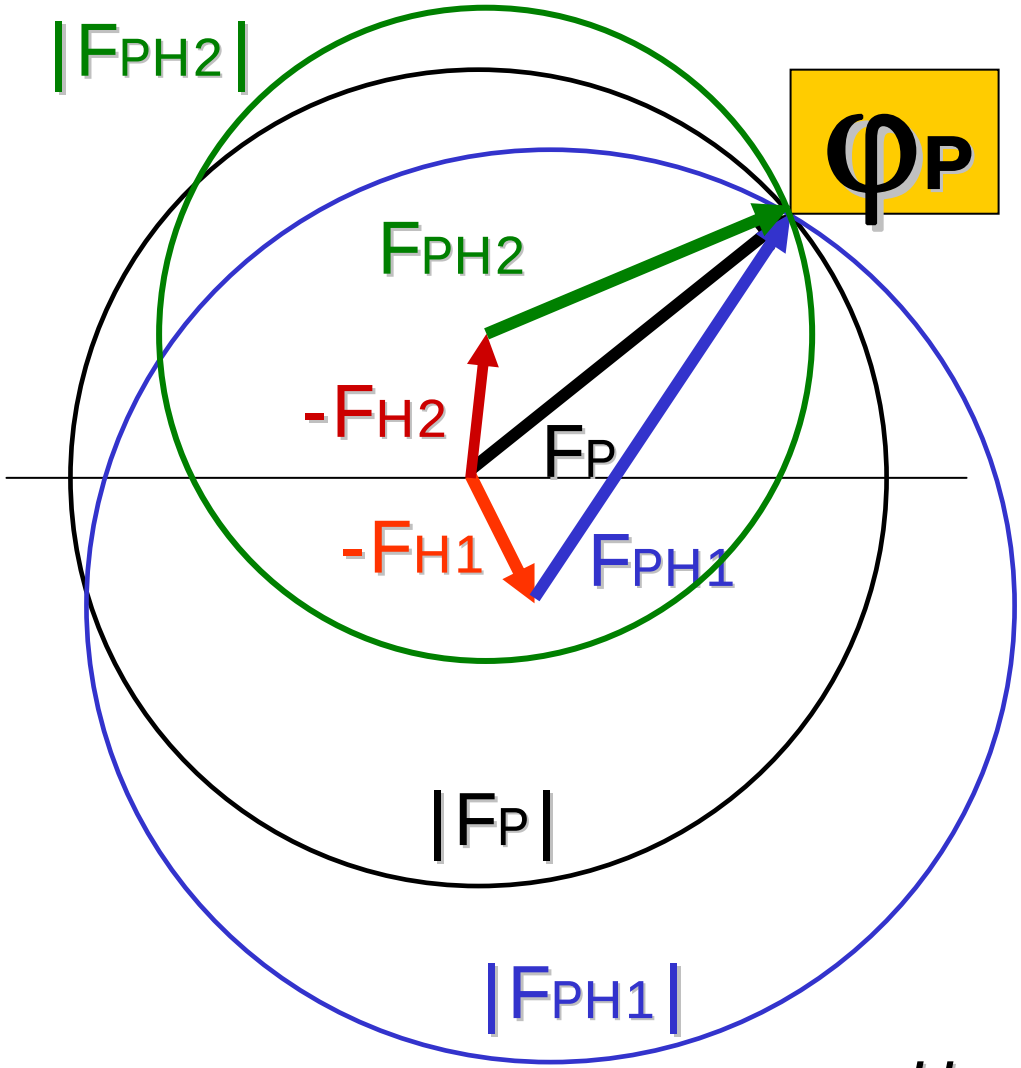


$$-F_H + F_{PH} = F_P$$

Harker diagram

Isomorphous replacement

MIR



Harker diagram

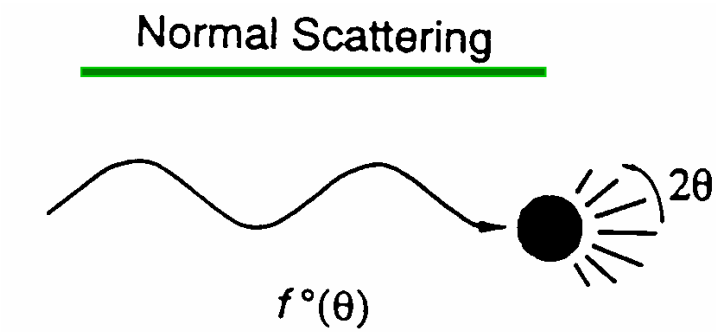
MAD (Wayne Hendrickson)

- Our protein is 'labeled' with an atom (usually Se introduced as Se-Met) that can scatter X-rays anomalously
- The energy of the X-rays is tuned (synchrotron) to maximize the anomalous effect
- An analysis of the anomalous effects reveals the phases, and the phases reveal the crystal structure

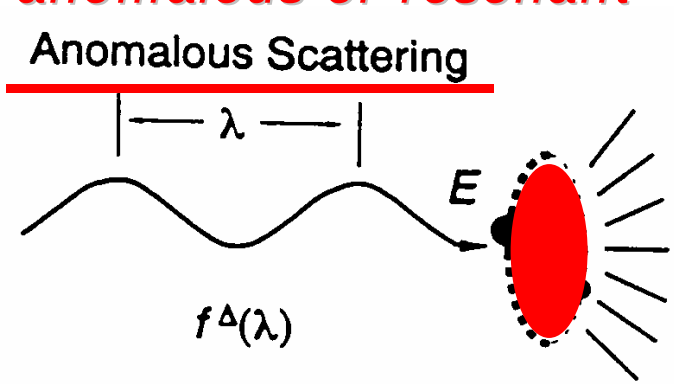
MAD: *Multiwavelength Anomalous Diffraction*

Scattering of X-rays by electrons in atoms

normal



anomalous or resonant



resonance

1. depends on the scattering angle 2θ
2. does not depend on wavelength (except resonance conditions)

1. does not depend on the scattering angle 2θ
2. critically depends on wavelength

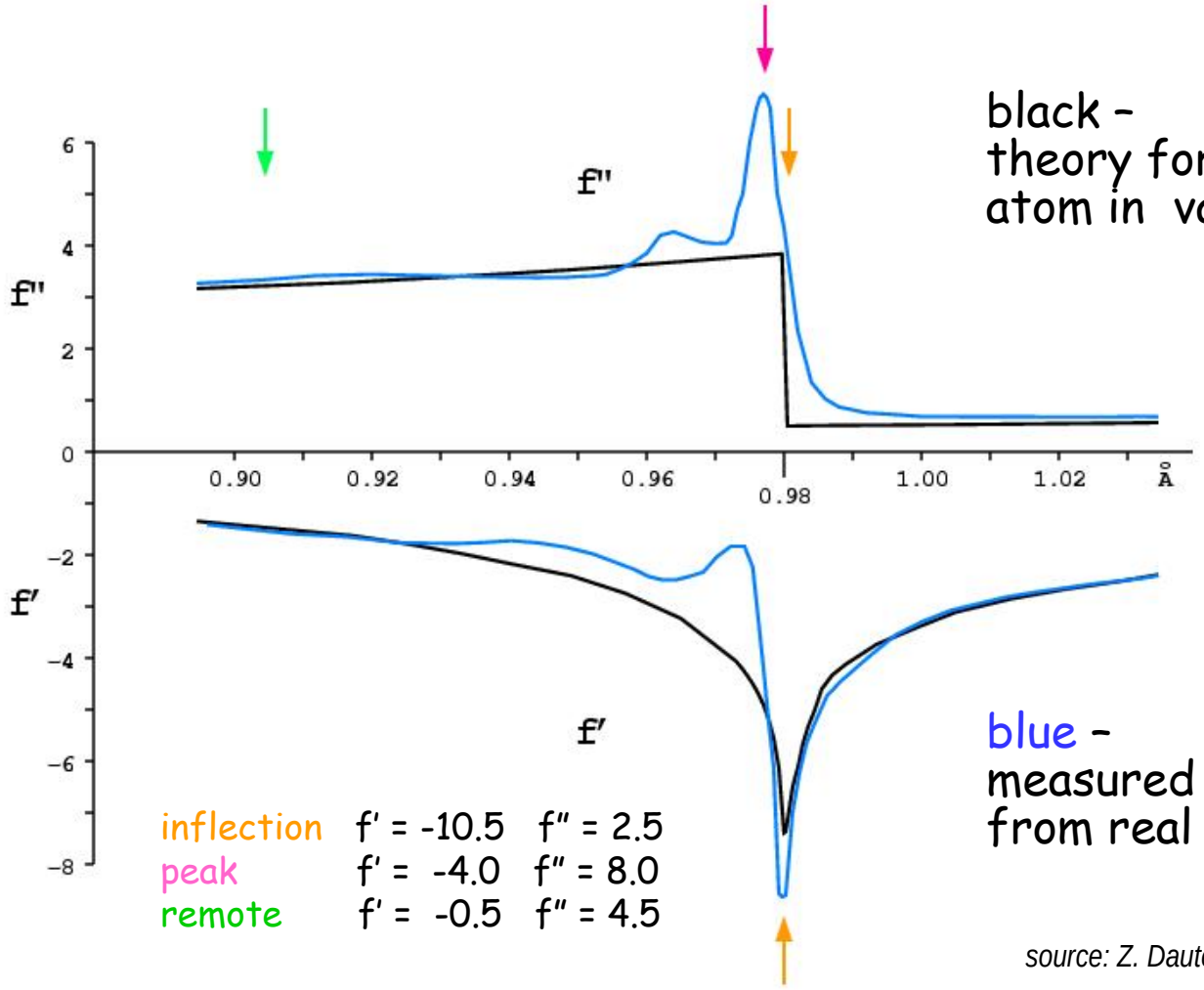
$$f_a = f_o + f' + if''$$

dispersive correction (real)

Bijvoet absorptive correction (imaginary)

The atomic scattering factor in the case of anomalous scattering

Anomalous corrections f' and f" for Se



black –
theory for single atom in vacuum

blue –
measured curve from real sample

inflection	f′ = -10.5	f″ = 2.5
peak	f′ = -4.0	f″ = 8.0
remote	f′ = -0.5	f″ = 4.5

source: Z. Dauter

Algebraic solution of the phase problem in MAD

(Jerome Karle, Wayne Hendrickson)

$$|{}^{\lambda}F(h)|^2 = |{}^{o}F_T|^2 + a(\lambda)\,|{}^{o}F_A|^2 + b(\lambda)\,|{}^{o}F_A|\,|{}^{o}F_T|\cos({}^{o}\phi_T - {}^{o}\phi_A) + c(\lambda)\,|{}^{o}F_A|\,|{}^{o}F_T|\sin({}^{o}\phi_T - {}^{o}\phi_A)$$

separation of λ-independent:

$$(|{}^{o}F_T|,\ |{}^{o}F_A|,\ {}^{o}\phi_T - {}^{o}\phi_A)$$

and λ-dependent variables:

$$a(\lambda) = (f'^2 + f''^2)/f_o^2$$
$$b(\lambda) = 2(f'/f_o)$$
$$c(\lambda) = 2(f''/f_o)$$

Main steps of MAD

1. Elaborate an expression system of your protein in *Escherichia coli*

2. Express, purify and crystallize an Se-Met variant of your protein

3. Tune the wavelength of synchrotron radiation to resonance with Se atoms

4. Collect X-ray diffraction data at λ_{peak}, λ_{edge} and λ_{remote}

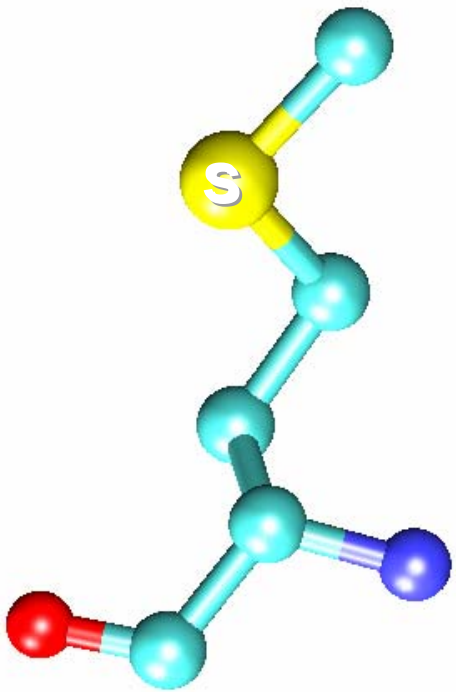
5. Solve the phase problem using the MAD algebra

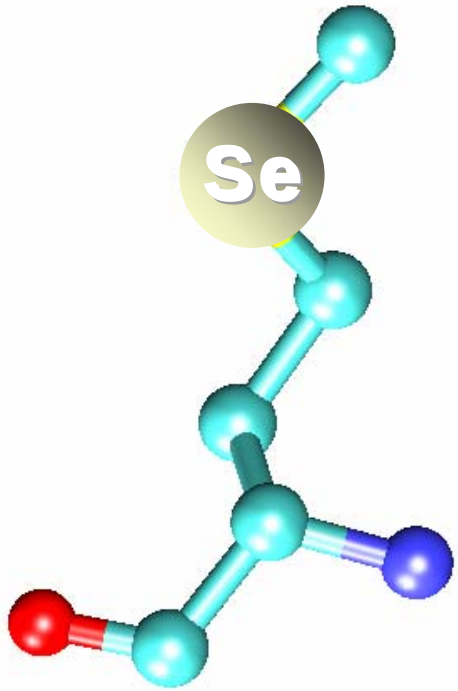
Selenomethionine

in the method of

MAD

(**M**ultiwavelength **A**nomalous **D**iffraction)





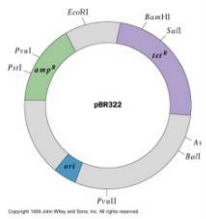
Met

Se-Met

foreign DNA

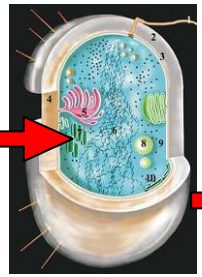
transformation

Met auxotroph



plasmid

cloning vector

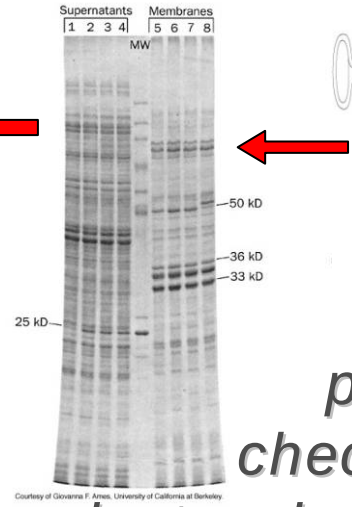


E.coli

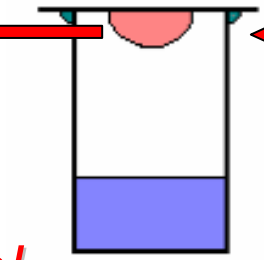
bacterial culture

purification by chromatography

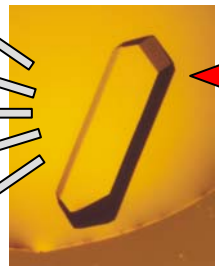
Recombinant proteins for protein crystallography



purity check by electrophoresis



crystallization



single crystal

X-ray diffraction

When the phase problem has been solved,
by any of the three methods,
electron density maps can be calculated
and an atomic model can be built in them.
But even if a preliminary model of our protein
has been built,
it still needs to be **refined**.

Structure Refinement

- Minimization of:

$$S1 = \Sigma w(|F_o| - |F_c|)^2$$

- Adjustable parameters:

  for each atom, its coordinates x y z (position) and
  Atomic Displacement Parameter B (amplitude of vibration)

- Gigantic number of model parameters! *e.g.* a 50 kDa protein – *ca* 450 residues – *ca* 3500 non-H atoms – *ca* 14 000 parametrs !
  At 3 Å resolution only *ca* 8500 reflections!

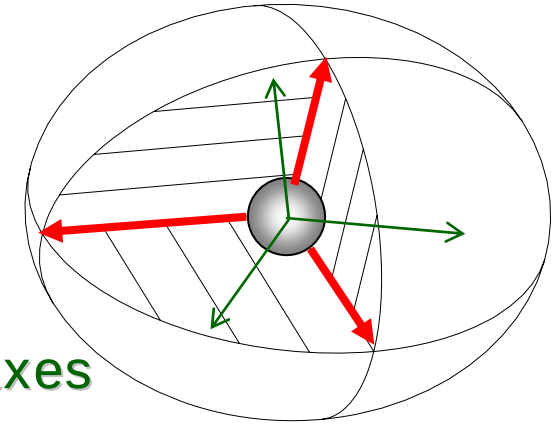
Solution: get more experimental data (always the best) or
reduce the number of parameters (constraints) or
better increase the number of "observations" (restraints)

- Stereochemical restraints as additional equations:

$$S2 = \Sigma w[p(ideal) - p(calc)]^2$$

(analogy to "energy" minimization)

The Temperature Factor or **ADP**
Atomic Displacement Parameter



anisotropic thermal ellipsoid
expressed in crystallographic axes

Atomic vibrations *(**displacements** from equilibrium positions) and **disorder** smear the electron density.*

*The vibrations are assumed to be **harmonic** and generally are **anisotropic** (described by a symmetric second-rank tensor, or **six Bij parameters**).*

*If a simple **isotropic model** is assumed, the magnitude of displacement is the same in all directions, the thermal ellipsoid becomes a sphere, and only **one parameter**, the **isotropic Displacement Parameter**, **Biso**, is necessary for atom k.*

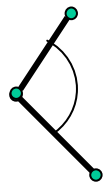
$$B_k = 8\pi^2 \langle u_k^2 \rangle$$

"A crystal should not be considered a chemical cemetery" (J. Dunitz)

Stereochemical restraints

Bond lengths, angles, etc. are parameters that must be reproduced by the model.

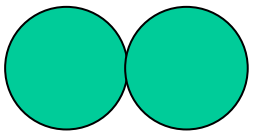
bond lengths



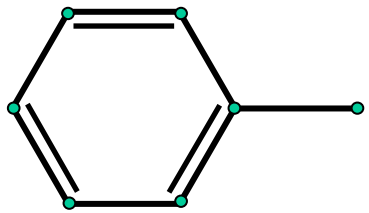
bond angles

$$S = \Sigma_{hkl}\, w[|F_{obs}| - |F_{calc}|]^2 + \Sigma_{par}\, (\sigma)^{-2}[d_{obs} - d_{std}]^2$$

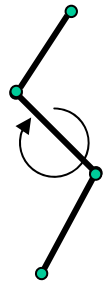
…



van der Waals contacts



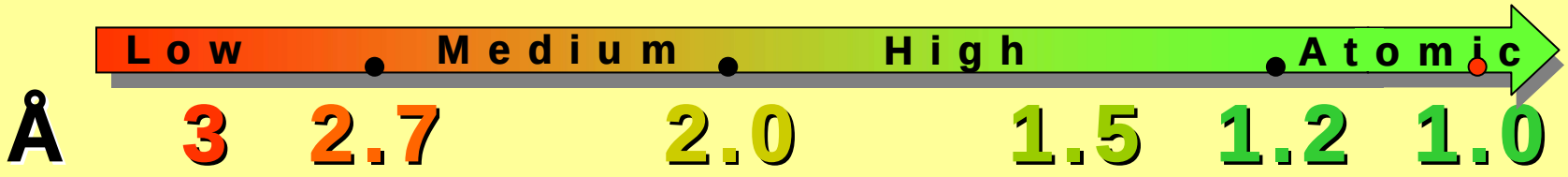
planar groups



torsion angles

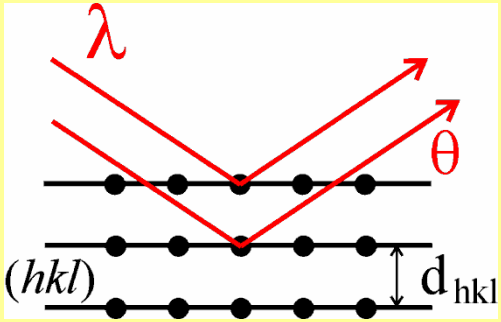
Resolution in crystallography

Resolution $d_{min} = \lambda / (2\sin\theta_{max})$



Bragg equation:

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



Sheldrick criterion

it can be shown that the **technically defined** resolution $\mathbf{d_{min}}$ is almost exactly equivalent to **optical resolution**, i.e. the minimum separation of two points that are still distinguishable in the Fourier Transform of the diffraction data, i.e. in electron density map

In **small-molecule crystallography**, **atomic resolution** is nearly always achieved and **atomic coordinates** are determined directly as peak coordinates in Fourier or difference-Fourier maps

In **macromolecular crystallography** this is rarely the case; the **primary product** is a (lower resolution) **electron density map**, and an **atomic model** of a macromolecule is (at least to some degree) a result of its **subjective interpretation** by a crystallographer

small-molecule crystallography

(full Cu sphere, λ=1.54 Å)
d_{min} = 0.77 Å

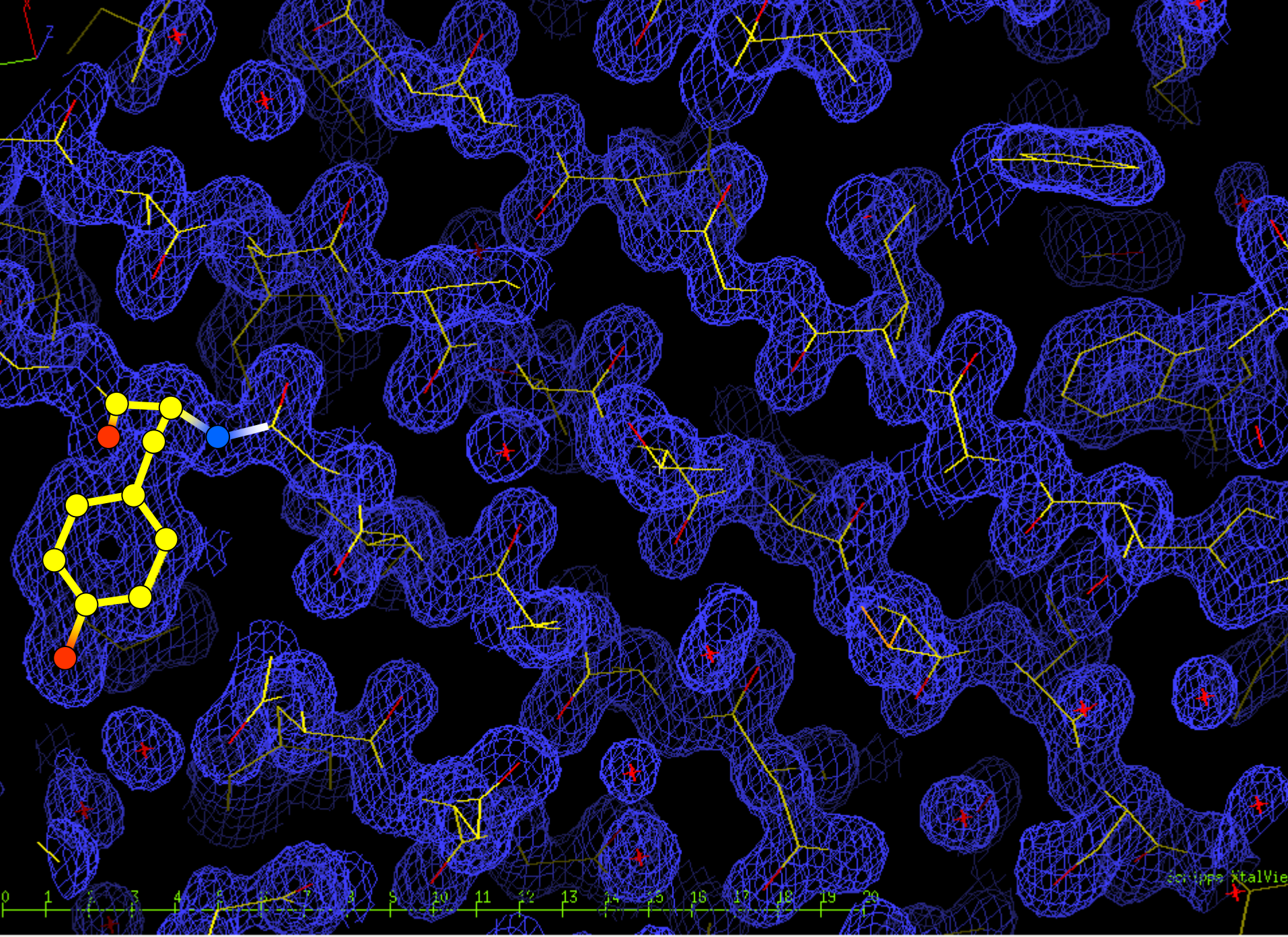
Mo radiation, λ=0.71 Å
2θ=60º
d_{min} = 0.71 Å

macromolecular crystallography

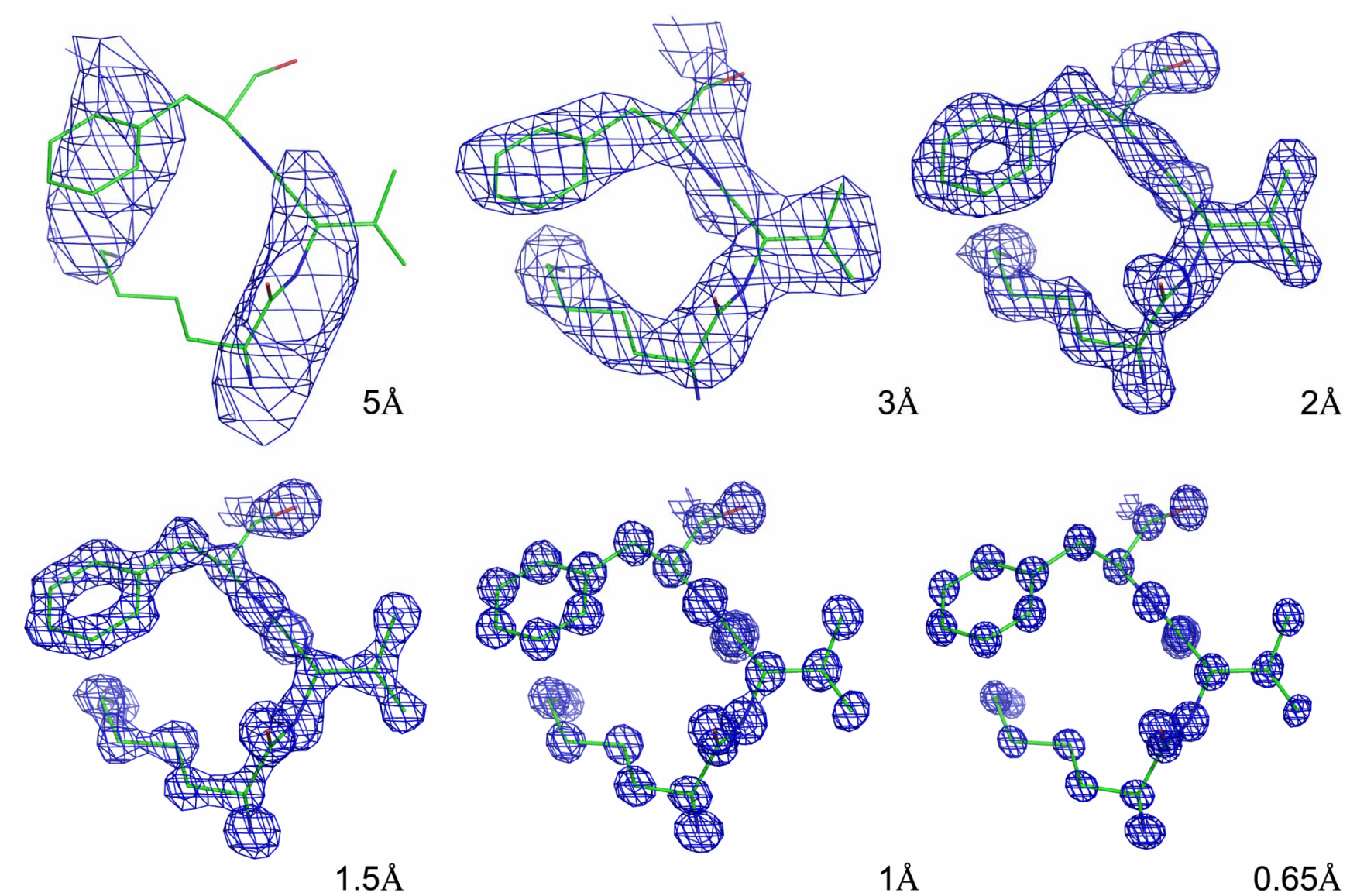
d_{min} = 1.2 Å (Sheldrick criterion)
(atomic resolution)

already a big success

but even if atomic resolution is not achieved – atomic models can be built


0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
XtalVie

Model building at low and high resolution



figures courtesy of Z. Dauter

Resolution of electron-density maps and the corresponding model detail

Resolution	Features Observed
5.0 Å	Overall shape of the molecule
3.5 Å	Cα trace
3.0 Å	Side chains
2.7 Å	Carbonyl O atoms (bulges) First chance to see water molecules
2.5 Å	Side chains well resolved, Peptide bond plane resolved
1.4 Å	Holes in aromatic rings Anisotropic B-factors possible
1.2 Å	True atomic resolution
1.0 Å	First chance to see H atoms
0.6 Å	Current limit for best protein crystals

Research Collaboratory for Structural Bioinformatics

1971 - Protein Data Bank created in Brookhaven Natl Lab (USA)
1998 - moved to RCSB (USA)
2003 - International organization (wwPDB)
Current Director – Dr. Helen Berman

http://www.rcsb.org/pdb/

The Worldwide Protein Data Bank (**wwPDB**)
is an international organization.
It consists of organizations that act as:
deposition, data processing and distribution centers for PDB data.
The founding members are:
RCSB PDB (USA), **PDBe** (Europe) and **PDBj** (Japan).
The BMRB (USA) group joined the wwPDB in 2006.
The mission of the wwPDB is to
maintain a single Protein Data Bank Archive
of macromolecular structural data
that is freely and publicly available
to the global community

How to assess crystallographic structures?

Resolution $d_{min} = \lambda/(2\sin\theta_{max})$

Å 3 2.7 2.0 1.5 1.2 1.0

R (resid

R_{free}: how well does the model predict data it has never 'seen'?

$||F_o|-|F_c||/\Sigma|F_o|$

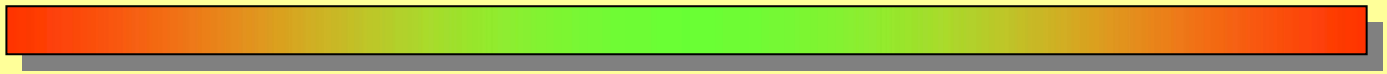
% 30 20 15

R_{free} - R

% 8 5 2

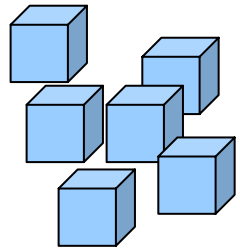
Deviations from „ideal” geometry

rms(bond-length – standard)

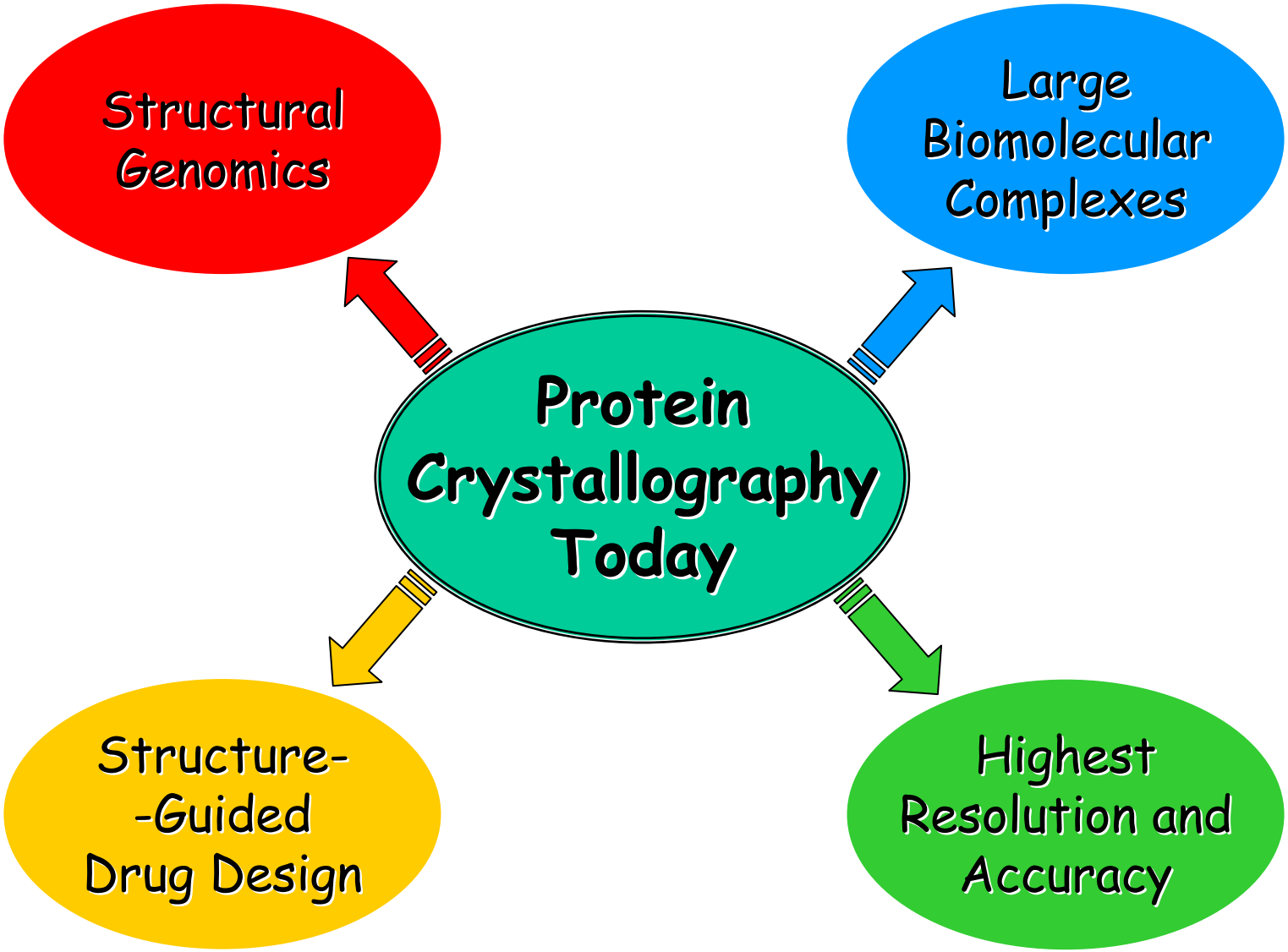


Å 0.03 0.02 0.01

Protein crystallography
areas of advancement



- **advancements in crystallization methods**
- **development of cryocrystallographic techniques**
- **powerful synchrotron X-ray sources**
- **faster and more sensitive detectors**
- **faster computers**
- **better algorithms**


Structural Genomics
Large Biomolecular Complexes
Protein Crystallography Today
Structure--Guided Drug Design
Highest Resolution and Accuracy

Structural Genomics

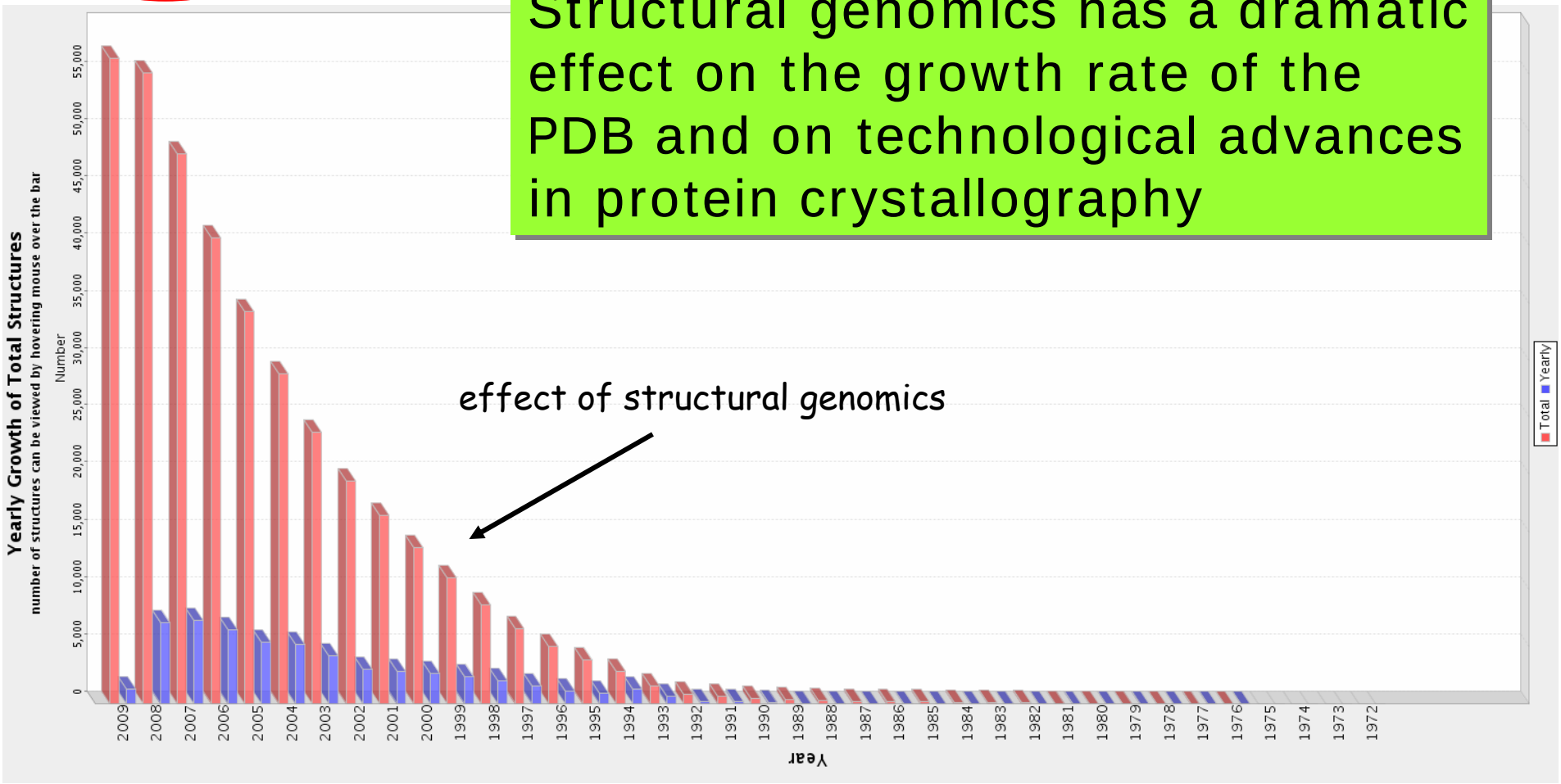
The aim of **structural genomics** is to determine in a **high-throughput** automated approach the 3D structure of **all the proteins** encoded in the genome of a given organism in order to understand their function. Contrary to classical biochemistry, the structure is studied before biochemical characterization of the target.

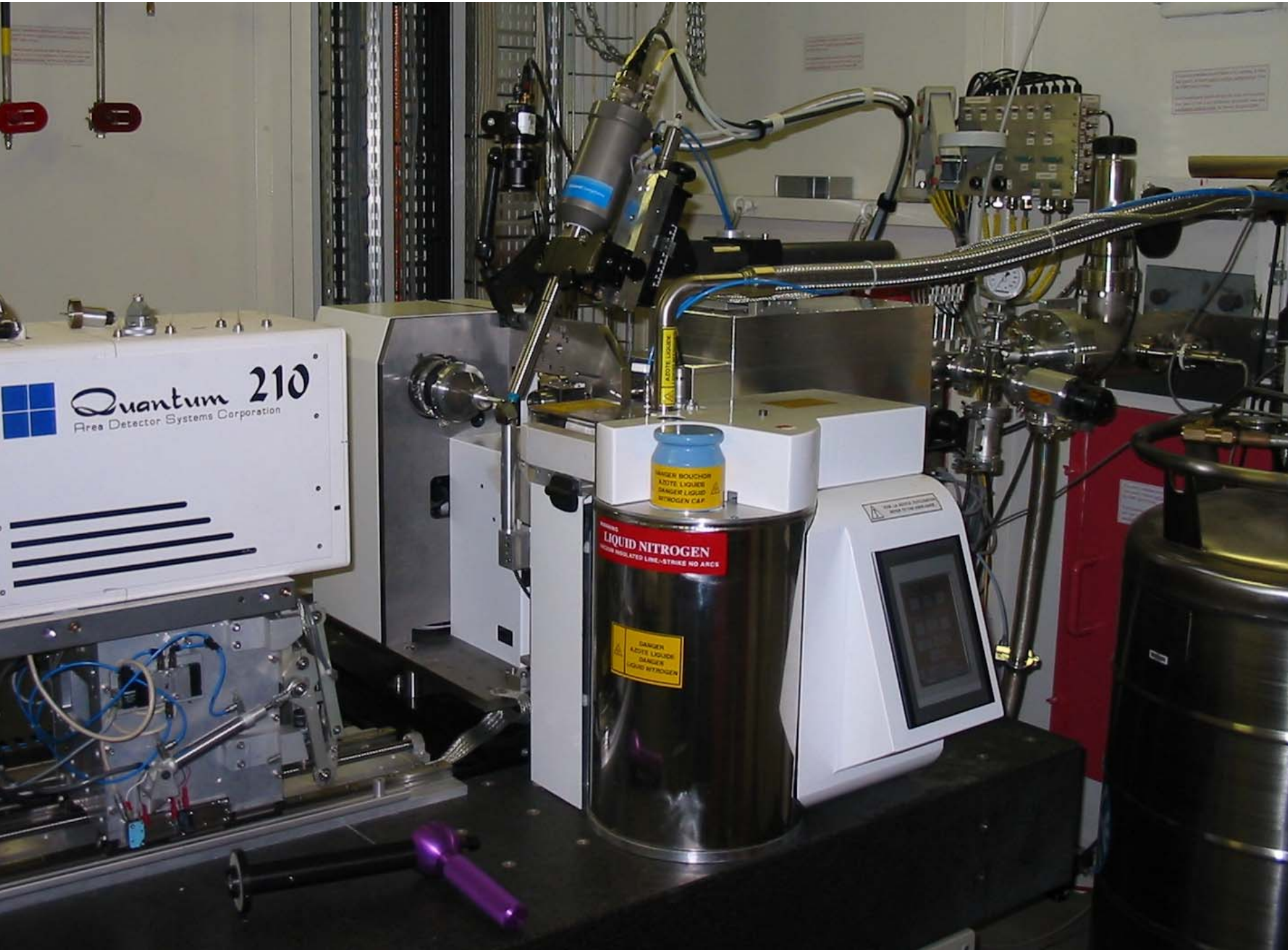
A number of human pathogens are being studied in this way, in search of new therapeutic targets.

Structural Genomics

Growth of the PDB

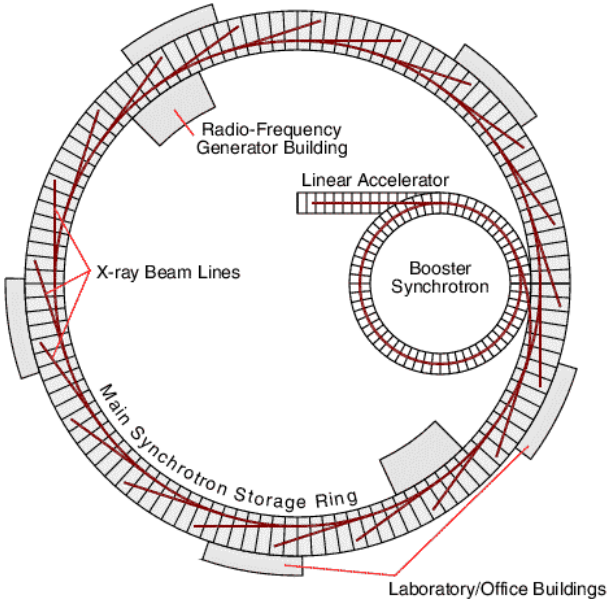
Structural genomics has a dramatic effect on the growth rate of the PDB and on technological advances in protein crystallography

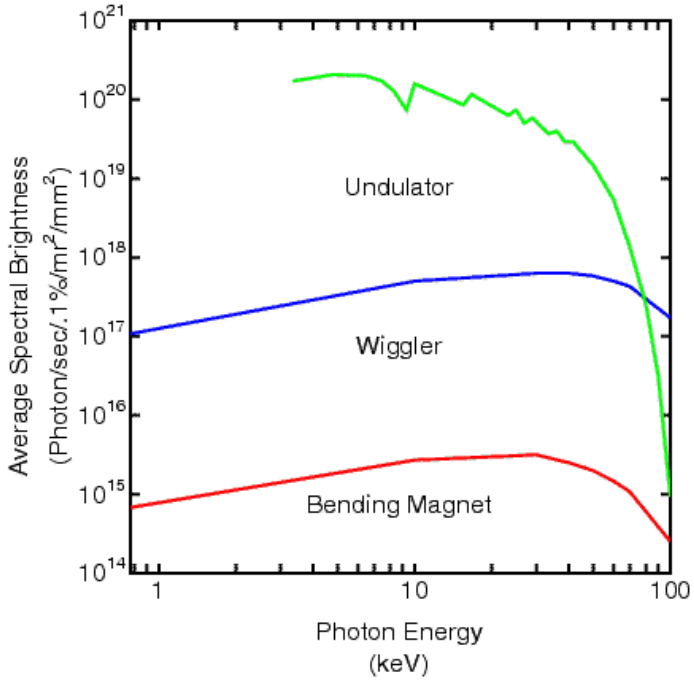


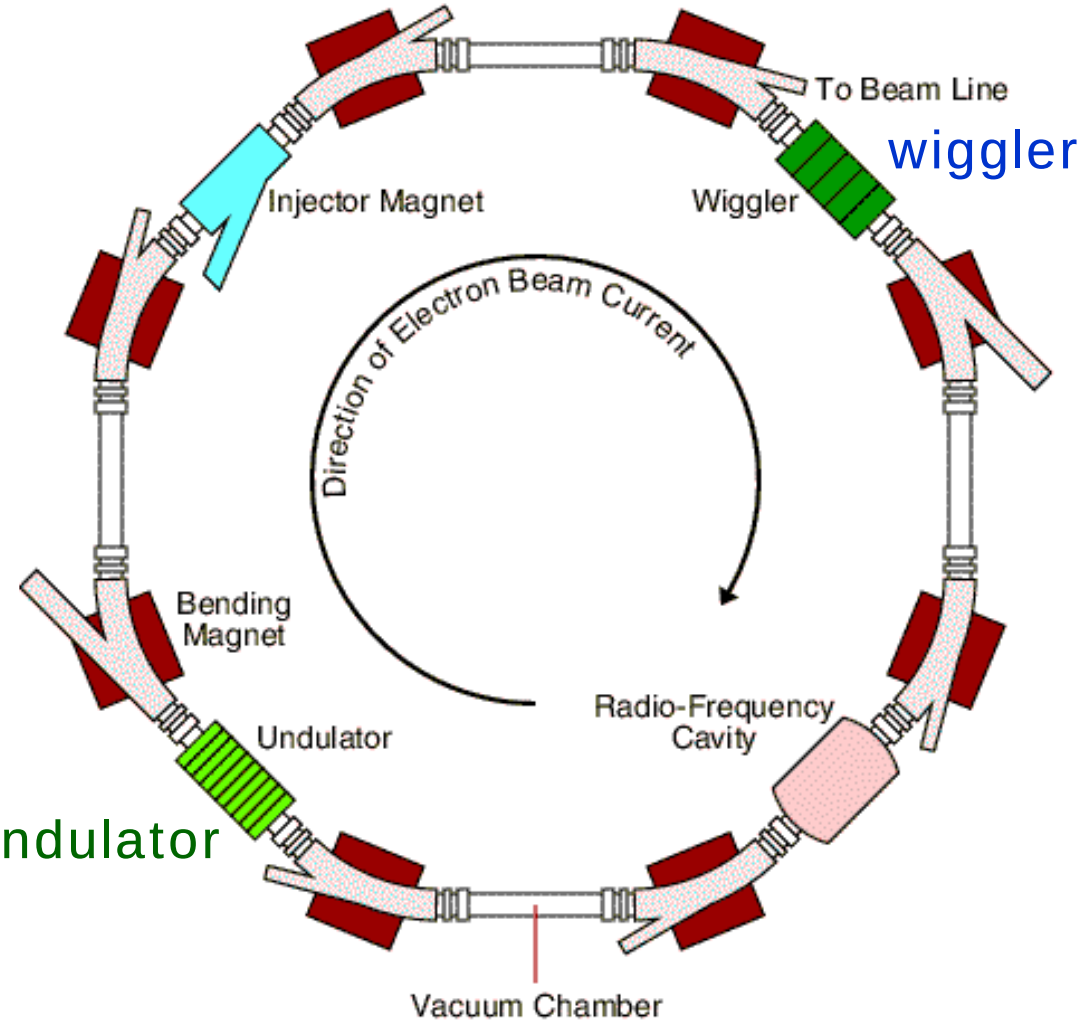

Quantum 210
Area Detector Systems Corporation
LIQUID NITROGEN

European Synchrotron Radiation Facility, Grenoble

Synchrotron



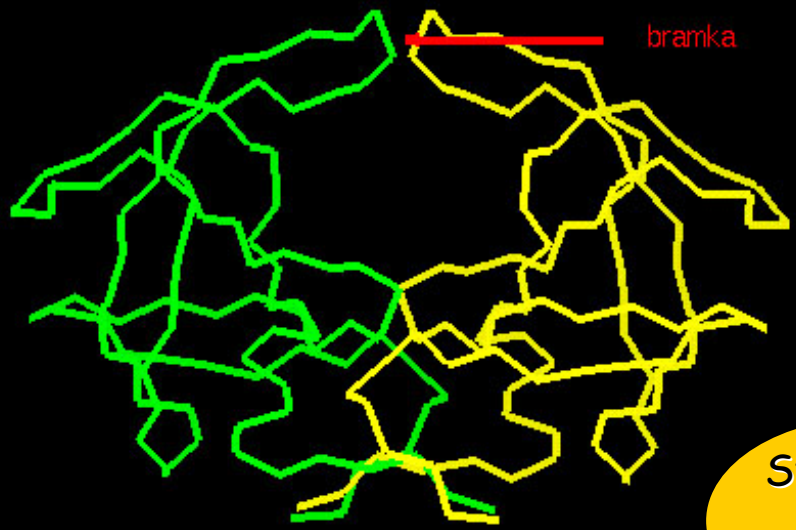


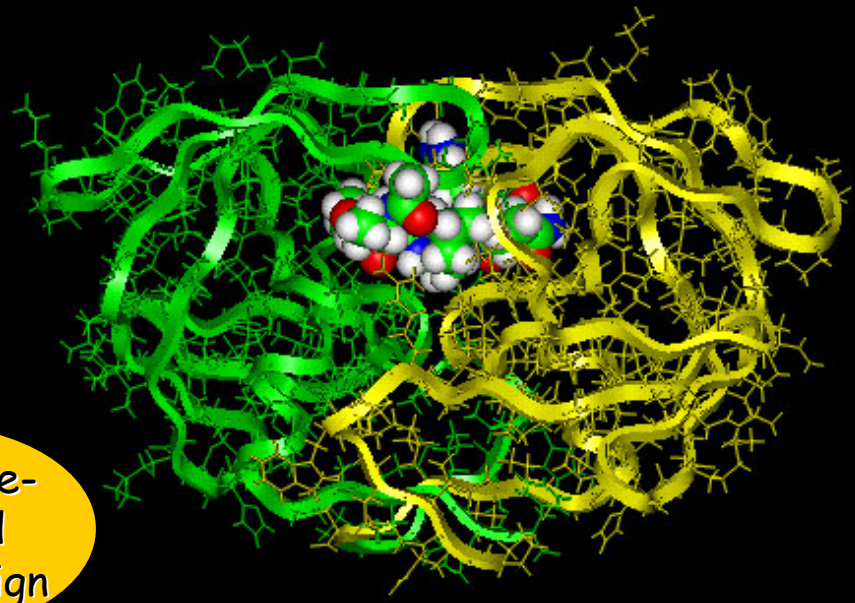


Characteristics of synchrotron radiation

- Wide spectral range (λ)
- Tuneability
- Very high intensity
- Polarization (*E* in equatorial plane)
- Has a time structure

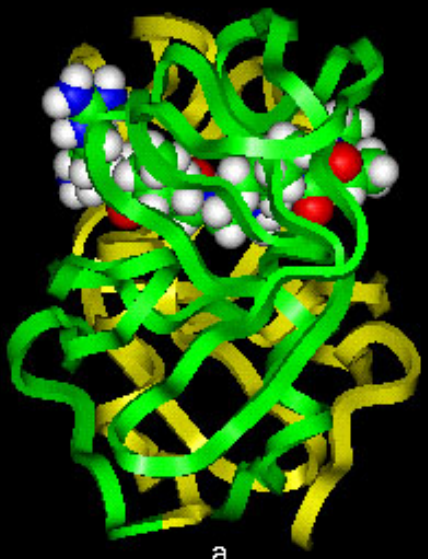
Poznan-ESRF remote access session

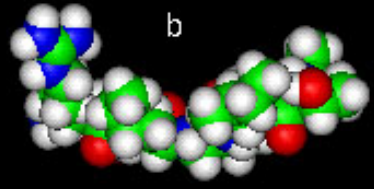




Structure-
-Guided
Drug Design

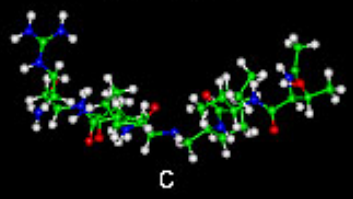
HIV-1 Protease

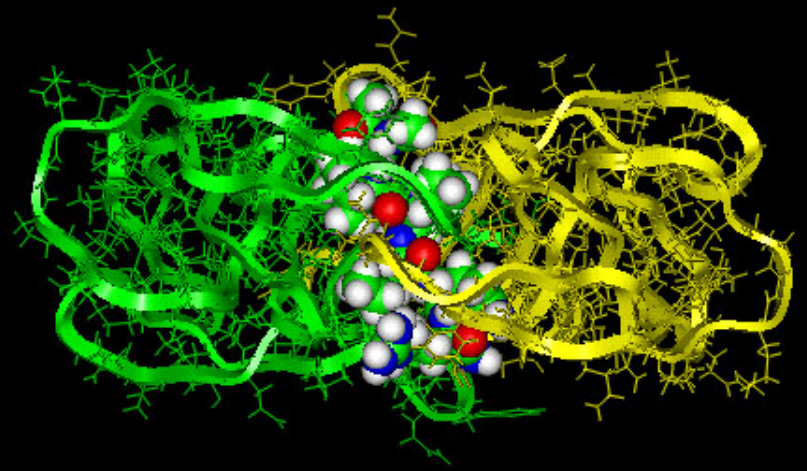




INHIBITOR

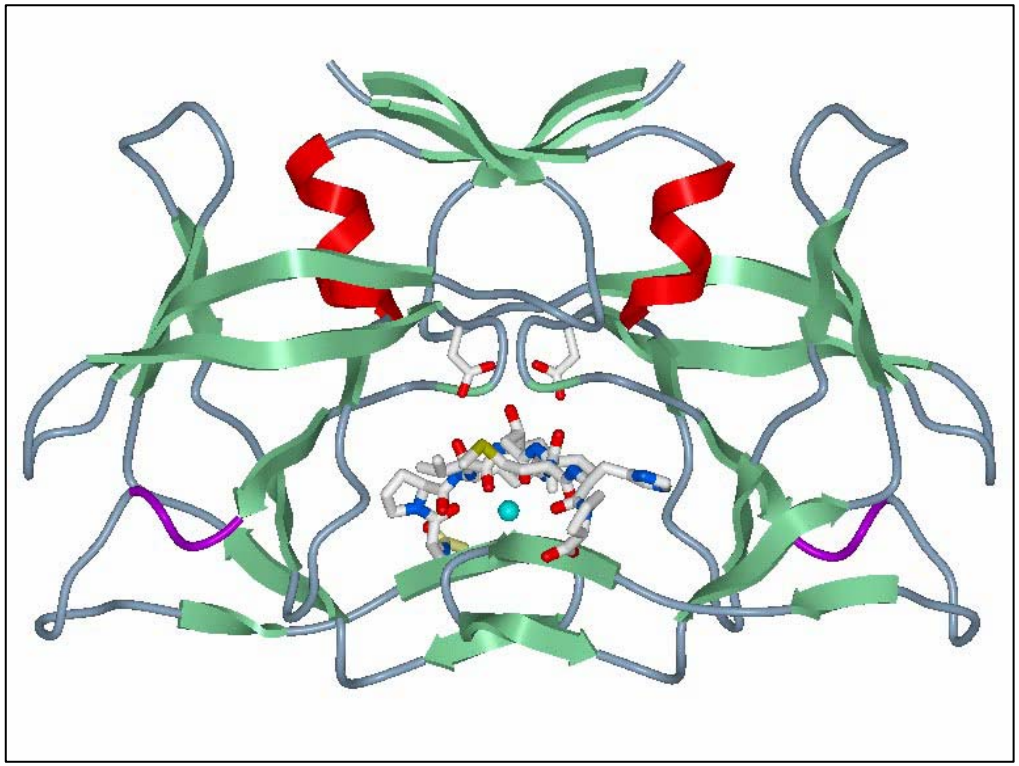
MVT – 101

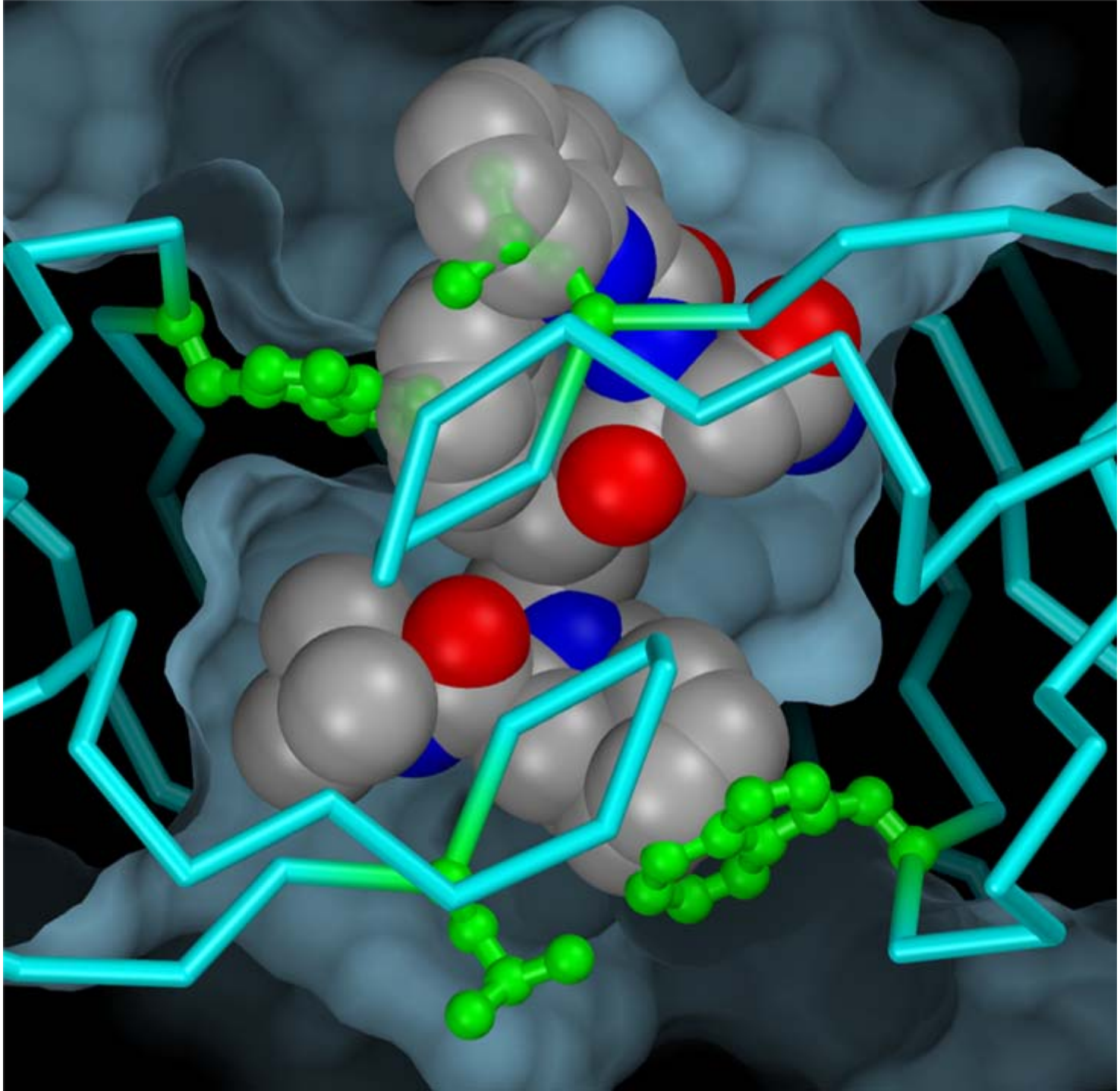




source: Dr M.Miller
Prof. M.Geller

Overall structure of HTLV-1 PR





Large Biomolecular Complexes

Examples:

- viruses
- the ribosome

Virus structure by X-ray crystallography

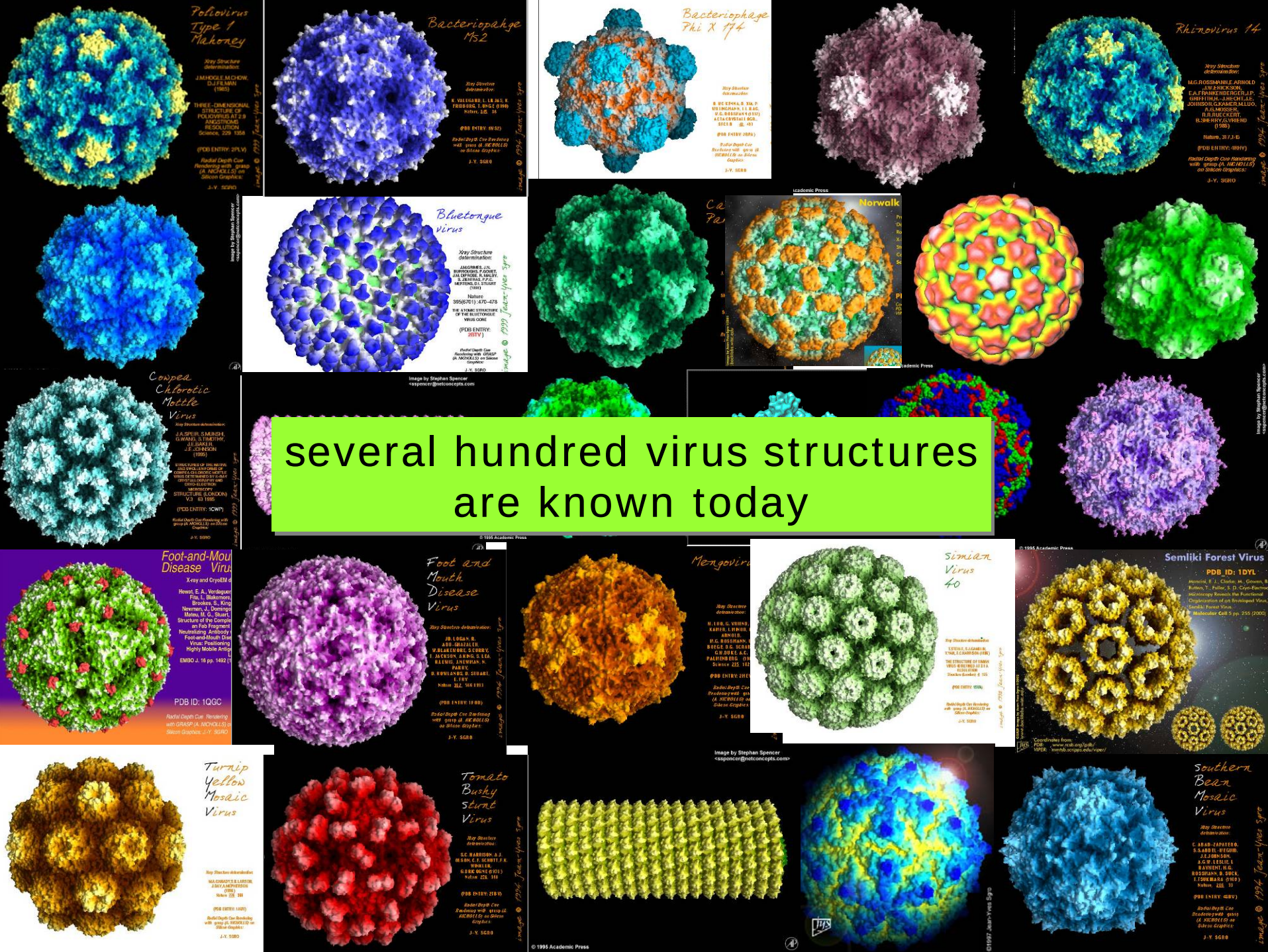
helical viruses

TMV

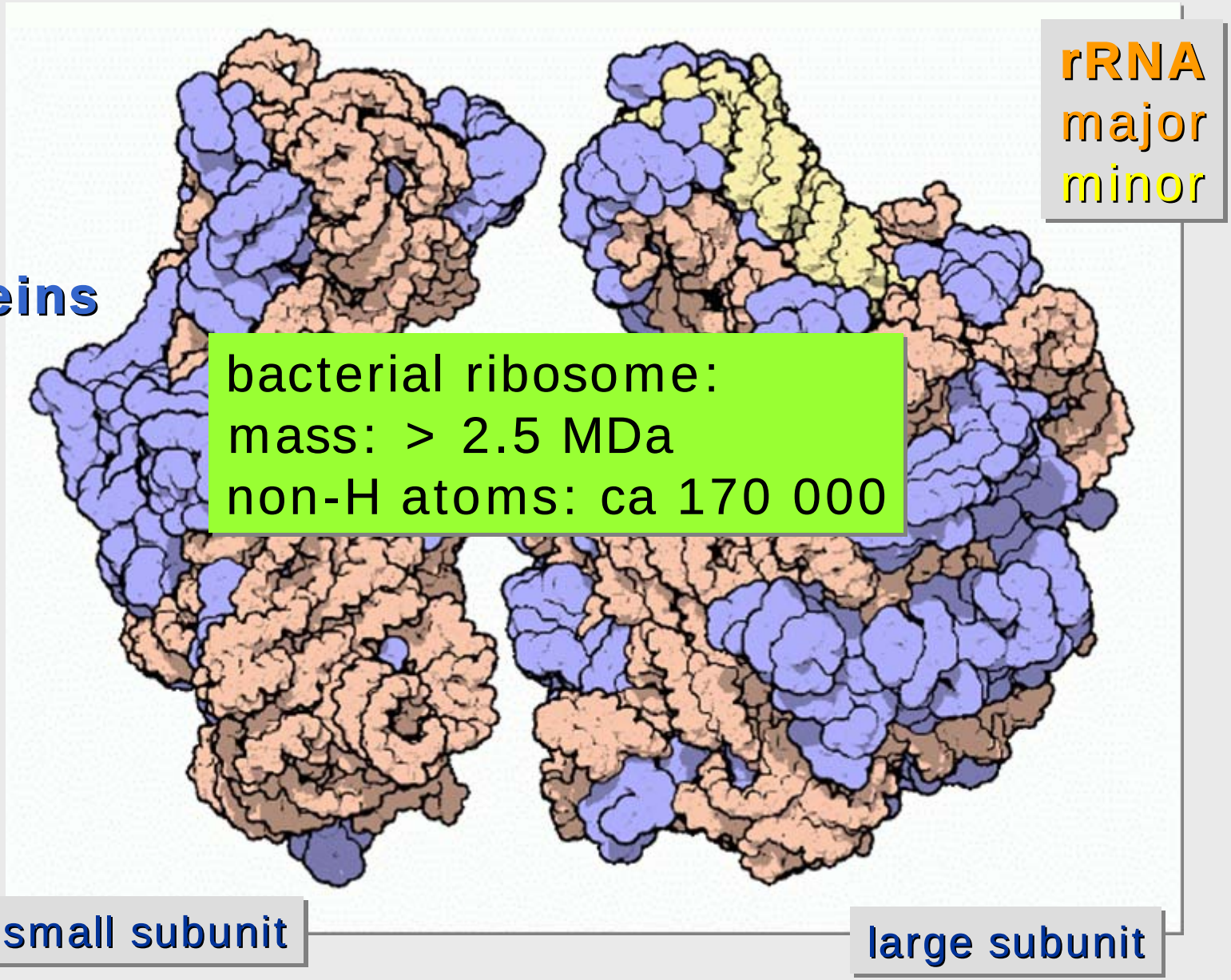
- Aaron Klug (1978) single-crystal X-ray diffraction, 2.8 Å
- Kenneth Holmes, Gerald Stubbs (1986) fiber X-ray diffraction, 3.6 Å

icosahedral (spherical) viruses

TBSV	- Stephen Harrison (1978) 2.9 Å
Bean mosaic virus	- Michael Rossmann (1980) 2.8 Å
STNV	- Lars Liljas (1984) 2.5 Å
Polio virus	- James Hoggle (1985) 2.9 Å
Common cold virus	- Michael Rossmann (1985) 2.8 Å


Poliovirus Type 1 Mahoney
Bacteriophage Ms2
Bacteriophage Phi X 174
Rhinovirus 14
Bluetongue Virus
Norwalk
Cowpea Chlorotic Mottle Virus
several hundred virus structures are known today
Foot-and-Mouth Disease Virus
Simian Virus 40
Semliki Forest Virus
PDB ID: 1DYL
PDB ID: 1QGC
Turnip Yellow Mosaic Virus
Tomato Bushy Stunt Virus
Southern Bean Mosaic Virus

Ribosome – the protein factory of the cell

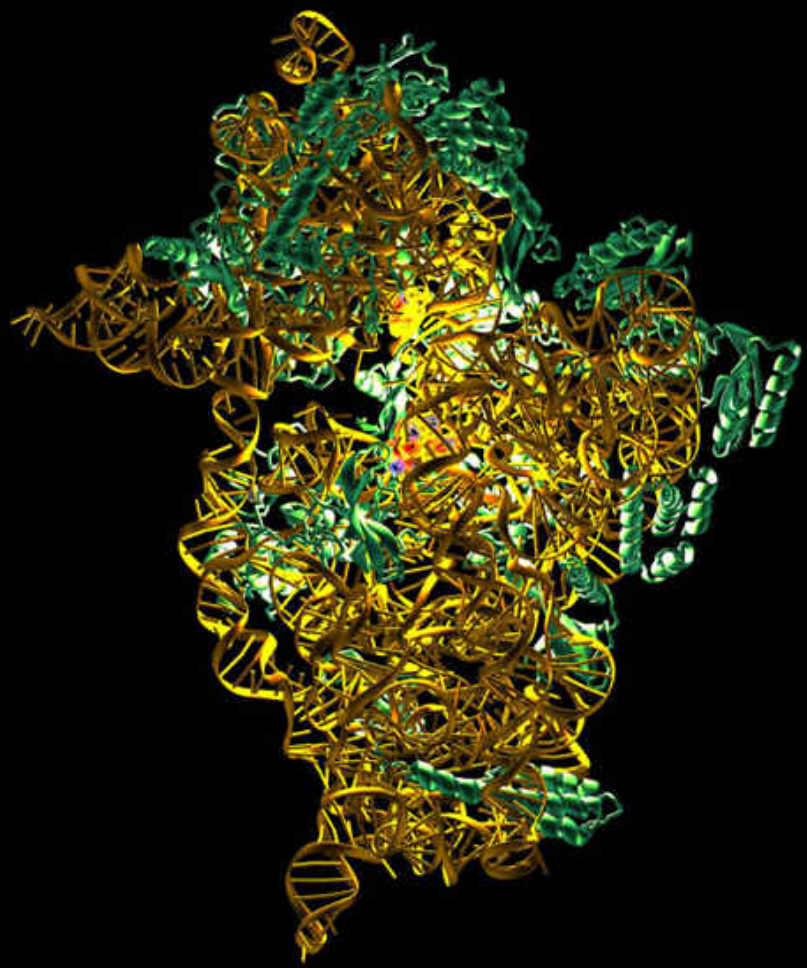


Ribosome structure

- *small subunit, 3.0 Å, Venki Ramakrishnan et al., 2000; 3.3 Å, Ada Yonath et al., 2000*
- *large subunit, 2.4 Å, Thomas Steitz et al., 2000; 3.1 Å, Ada Yonath et al., 2003*
- ***complete ribosome with tRNA & mRNA,*** *5.5 Å, Harry Noller et al., 2001;* ***3.2 Å, Jamie Cate et al., 2006; 2.8 Å, Venki Ramakrishnan et al., 2006***

Thermus thermophilus, Haloarcula marismortui, Deinococcus radiodurans, Escherichia coli

Ribosome structure

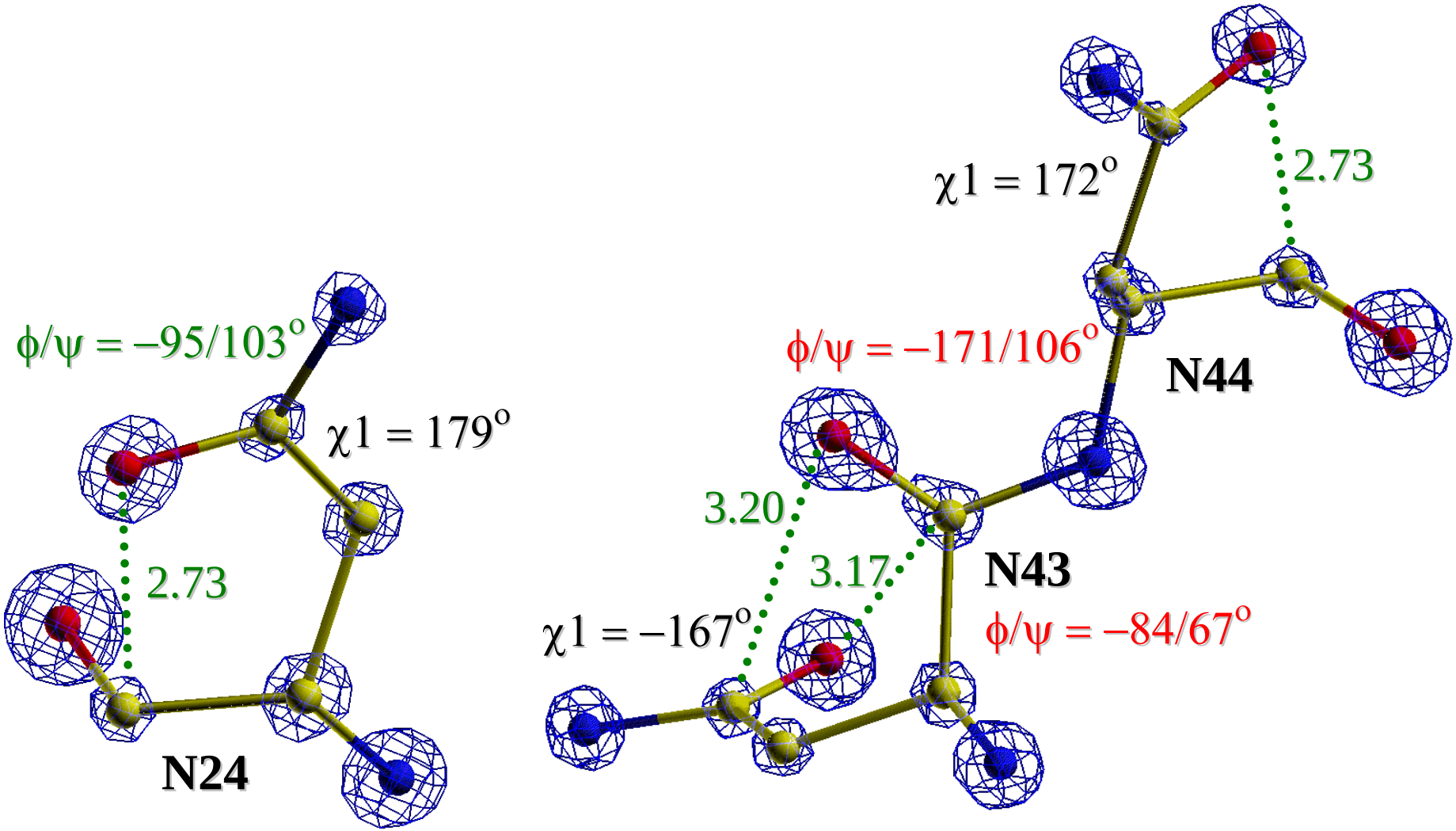


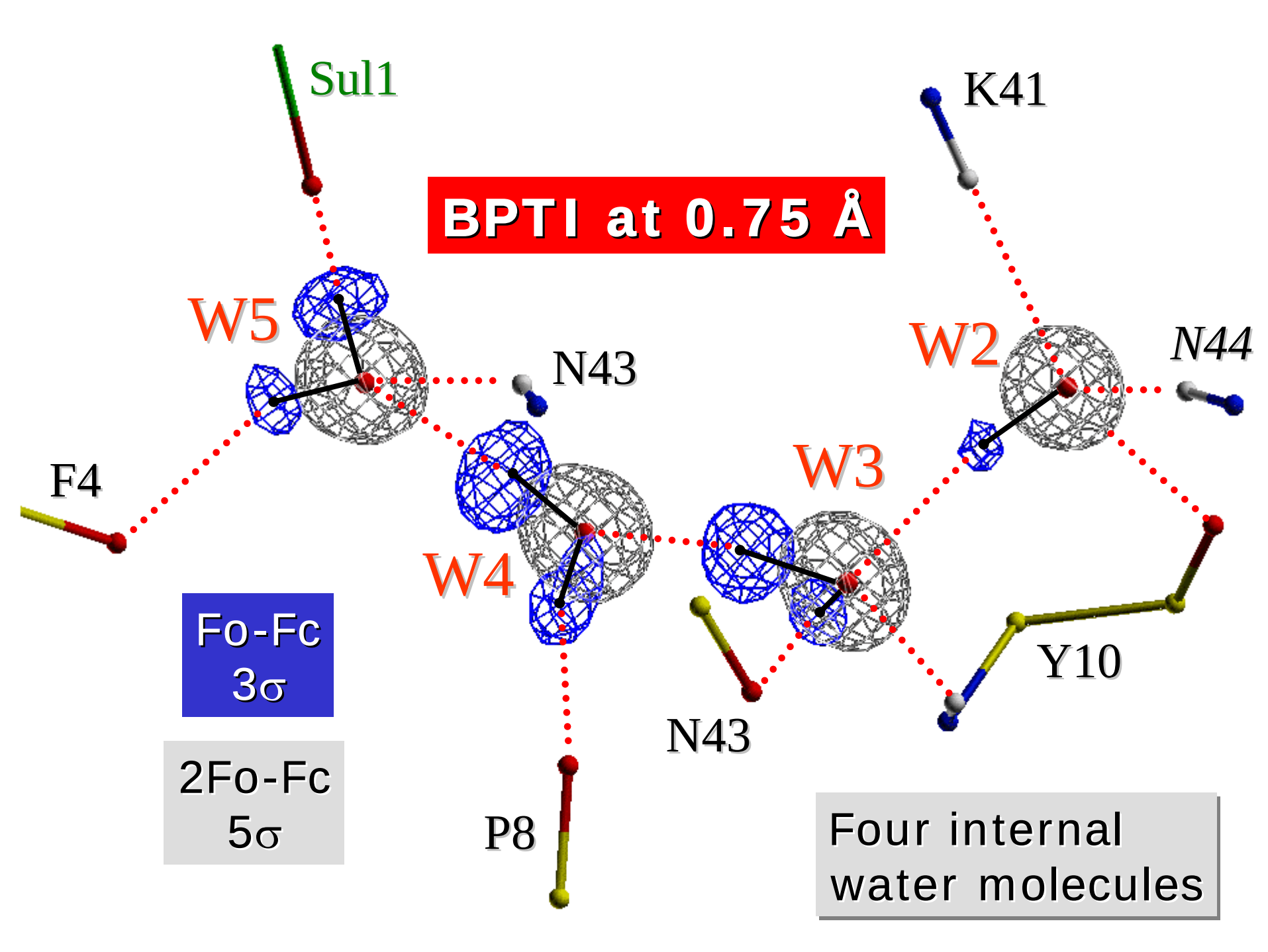
Highest Resolution and Accuracy

Record breaker: crambin at 0.54 Å

Examples from the structure of BPTI (Bovine Pancreatic Trypsin Inhibitor) at 0.86 and 0.75 Å

BPTI at 0.86 Å resolution

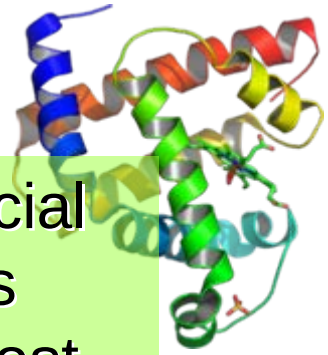



Sul1
K41
BPTI at 0.75 Å
W5
N43
W2
N44
F4
W3
W4
Fo-Fc
3σ
2Fo-Fc
5σ
N43
Y10
P8
Four internal
water molecules

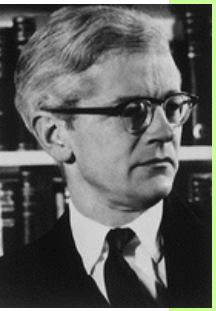
Protein Data Bank – growth of atomic-resolution entries

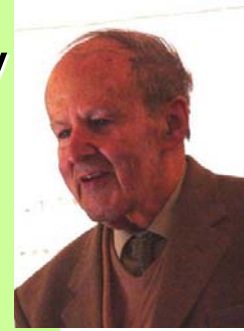
Resol (Å) at least	All holdings	1.4	1.2	1.0	0.8
1981	98	7	0	0	0
1986	224	13	3	**1**	0
1991	1103	40	15	7	0
1996	5929	116	44	16	0
2001	17694	590	284	102	6
July 09	59000	**2800**	**950**	**360**	**20**

5PTI - BPTI, Wlodawer & Huber (1984): N- and X-ray refinement



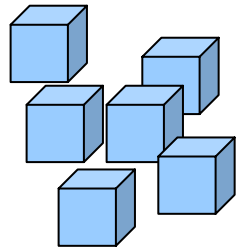
Crystallography – through the study of this special solid state of matter – crystalline protein, was historically the first method to reveal to us, almost exactly 50 years ago,
the secret of protein structure. This was done for myoglobin by John Kendrew and for hemoglobin by Max Perutz.





Today, we have accumulated in the PDB an enormous amount of structural information about proteins.
Crystallography is now assisted in this effort by other methods: NMR, electron microscopy, bioinformatics. But crystallography still remains the main source of information about the structure of proteins, especially in the context of Structural Genomics.

Protein crystals
high resolution – high goal

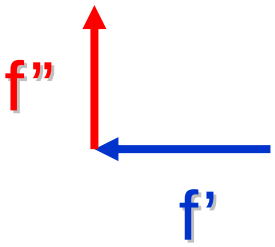


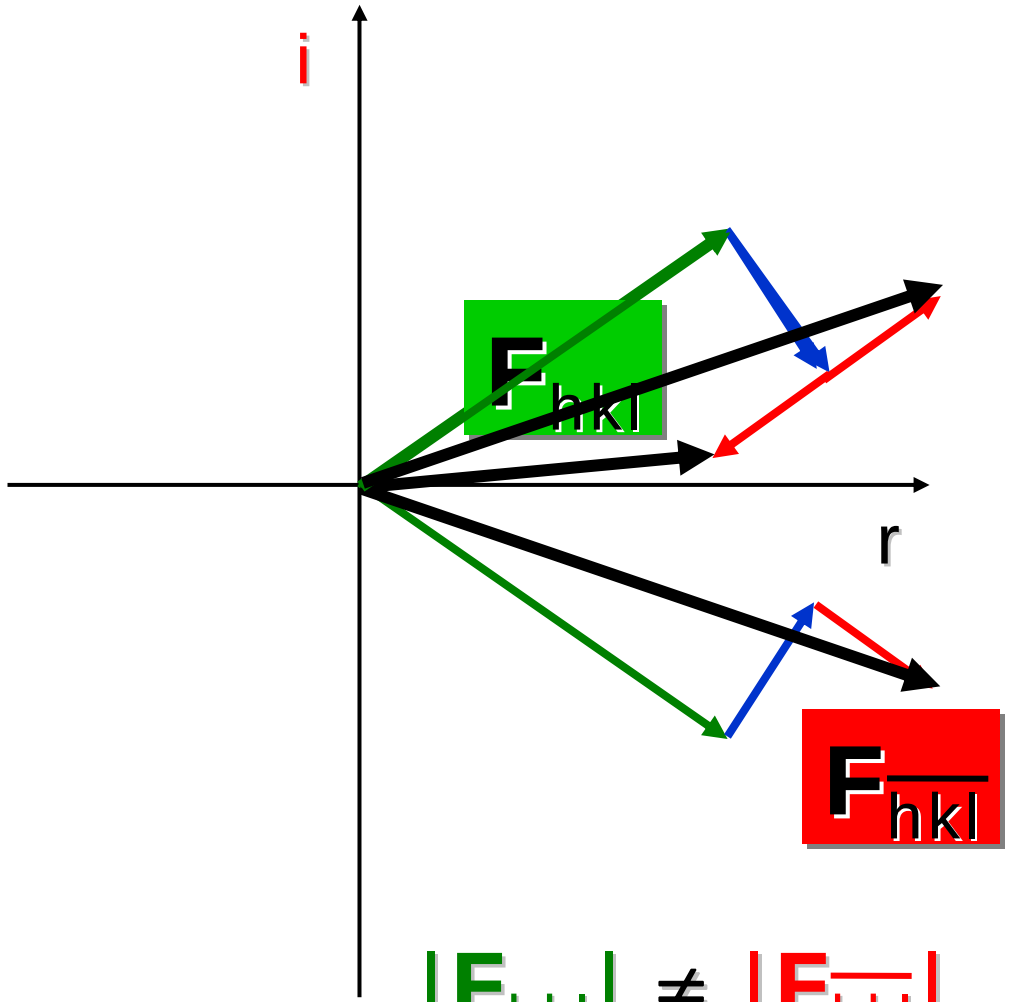
- only proper symmetry possible
- huge unit cells (*ca* 100 Å) – very weak diffraction ($I \sim V^{-2}$)
- huge numbers of reflections (e.g. 10^6 collected, 10^5 unique)
- gigantic molecules, often multiple, > 10,000 non-H atoms not rare
- molecules with a degree of flexibility - disorder
- high solvent content in crystal volume (50% or more)
- disorder in bulk water region
- crystals often sensitive and unstable

The Structure Factor in the case of anomalous scattering

normal scattering

anomalous scattering

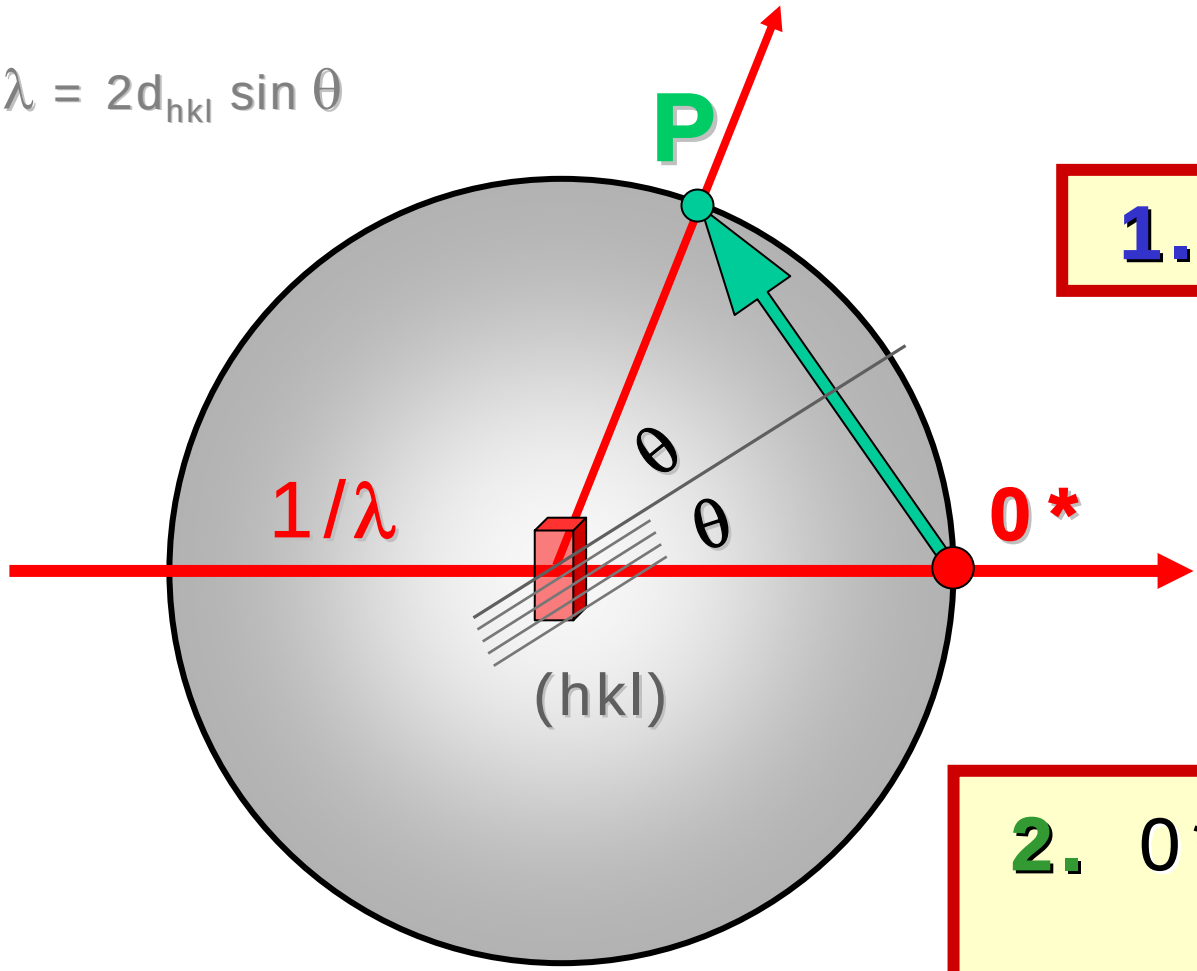




$$|F_{hkl}| \neq |F_{\overline{hkl}}|$$

The Ewald construction

$\lambda = 2d_{hkl} \sin\theta$



1. $0^*P \perp (hkl)$

2. $0^*P = (2/\lambda)\sin\theta$

$= 1/d_{hkl}$