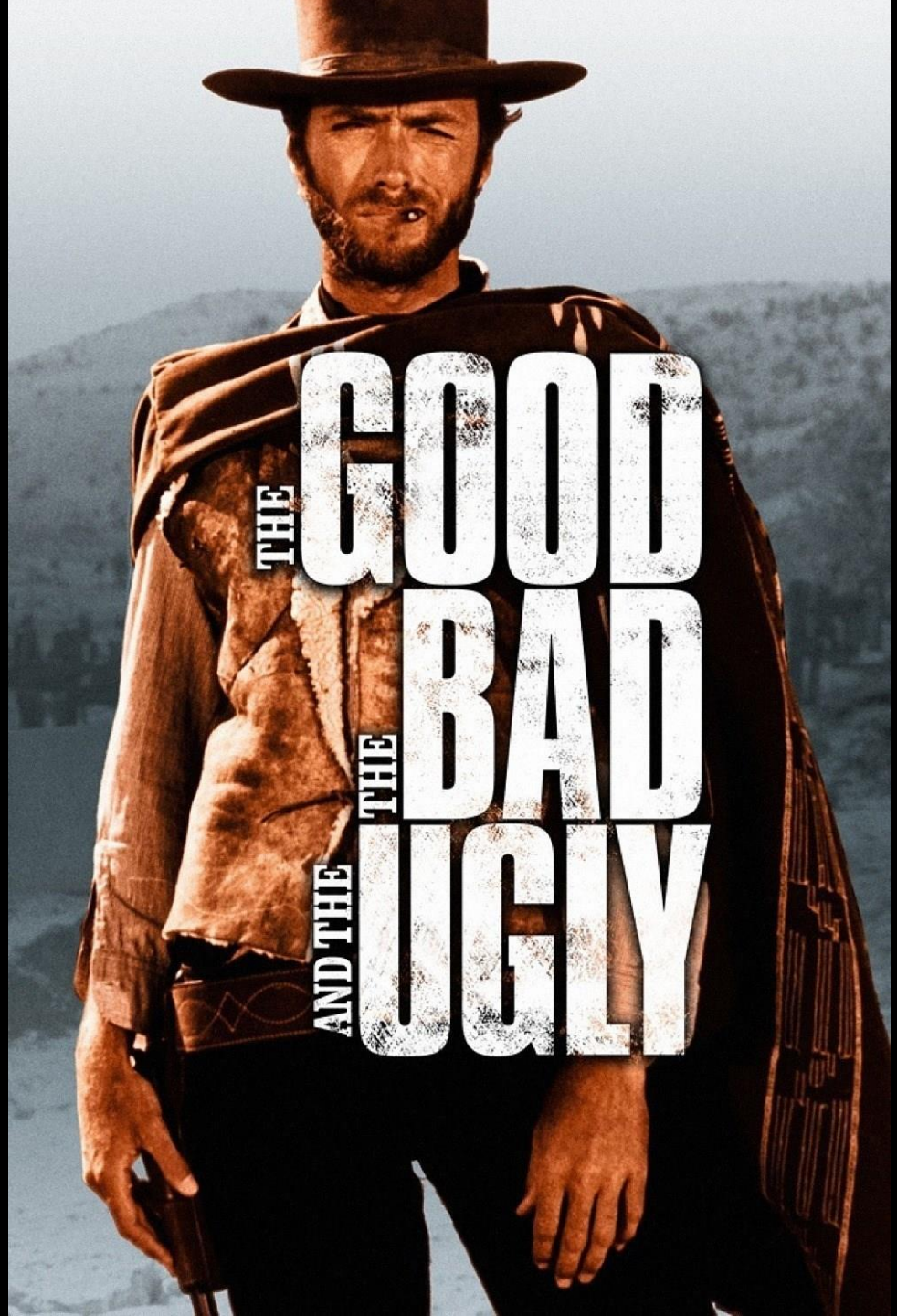
All that glitters is not necessarily gold –

The accurate identification of metals in metalloproteins and post X-ray diffraction structural remediation

Edward H. Snell


THE GOOD
THE BAD
AND THE UGLY

Structure

It is the pervading law of all things organic and inorganic,

Of all things physical and metaphysical,

Of all things human and all things super-human,

Of all true manifestations of the head,

Of the heart, of the soul,

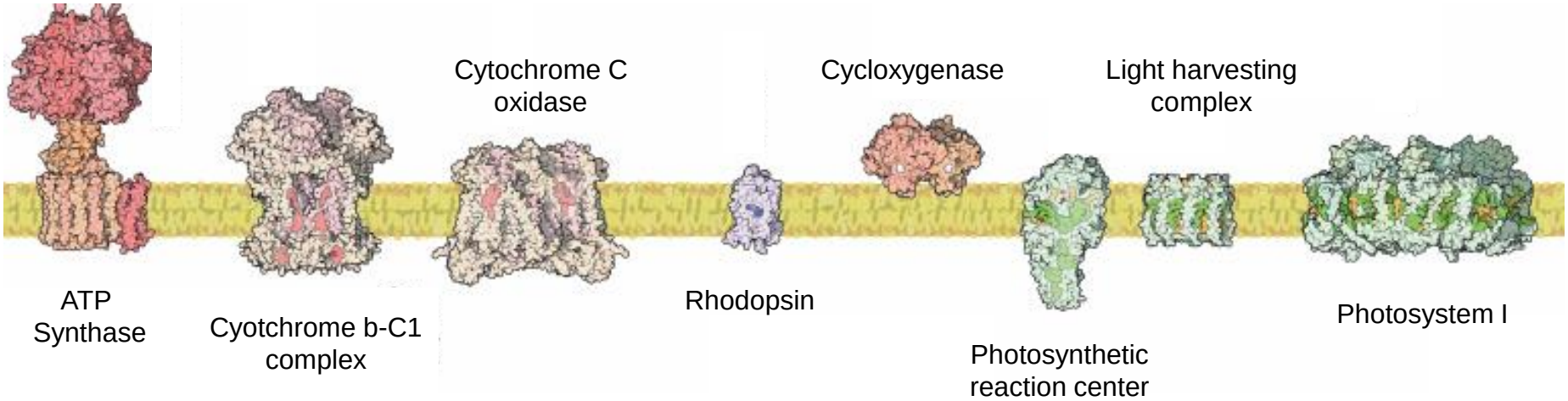
That the life is recognizable in its expression,

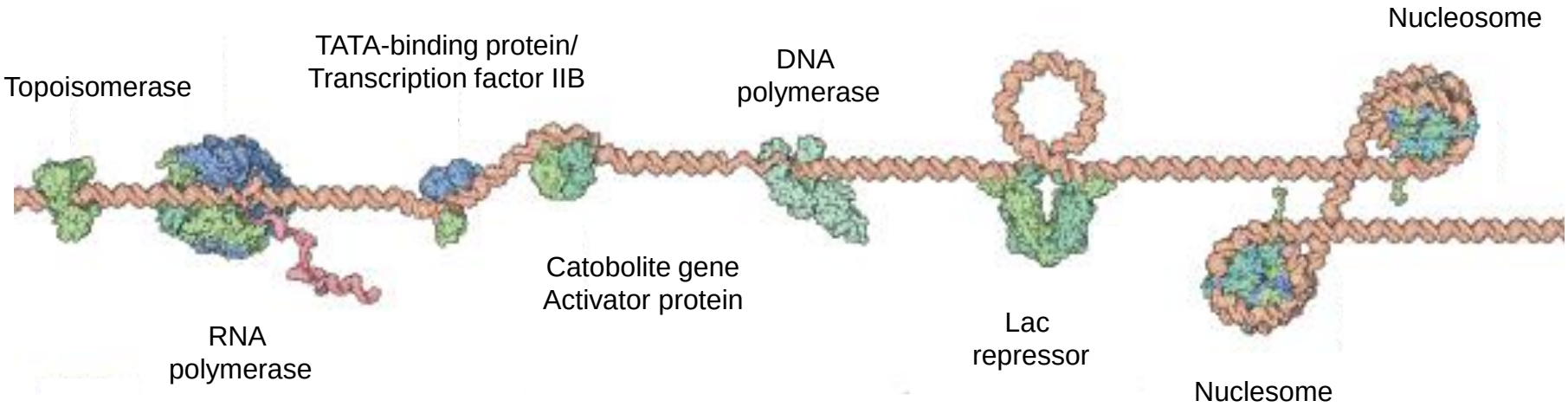
That ***form ever follows function.***

This is the law

American sculptor, Horation Greenhough but made famous by architect, Louis Sullivan, father of the skyscraper.

The Guaranty Building, 28 Church Street, Buffalo, completed 1895.

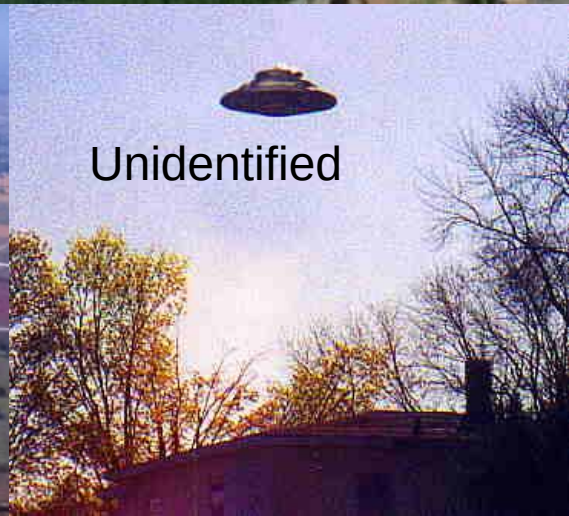




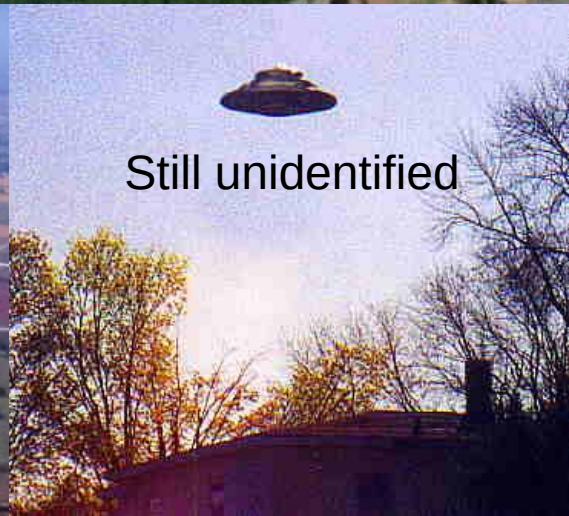
Form (or structure) gives a **clue** to the function.

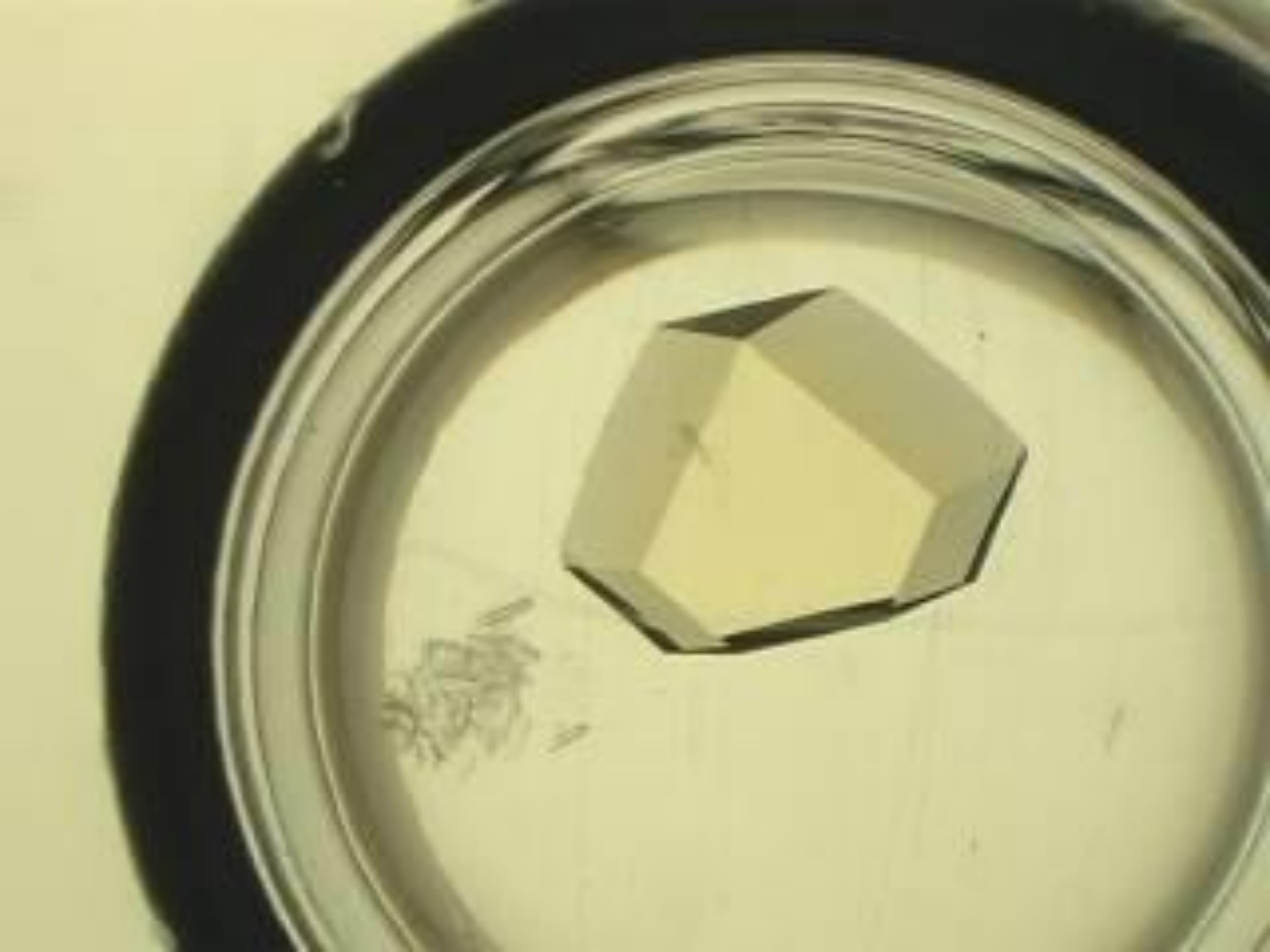
Adapted from Molecular Machinery: A tour of the Protein Data Bank, http://www/rcsb/org

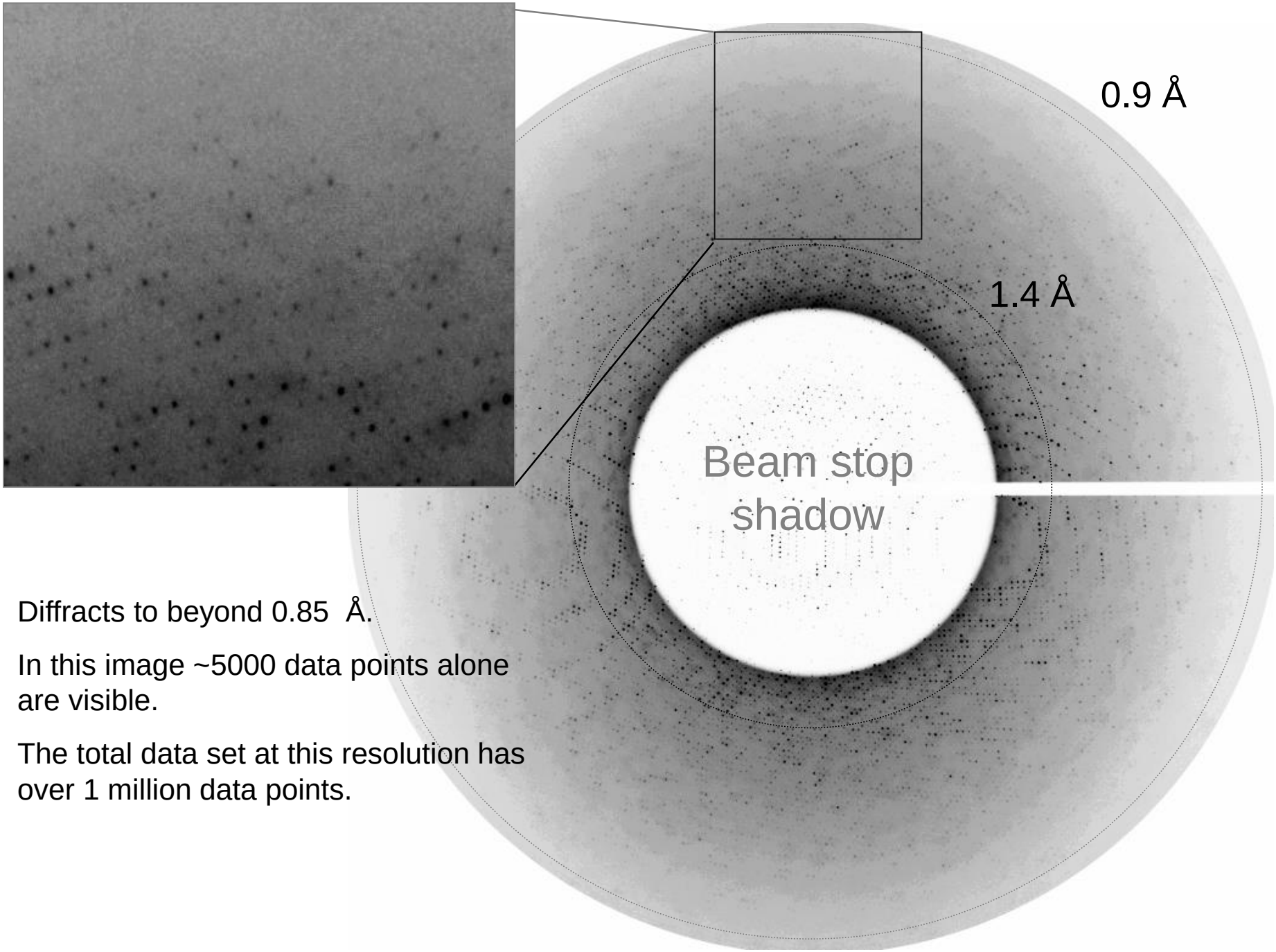
Stealth fighter



Nothing there,
false observation?



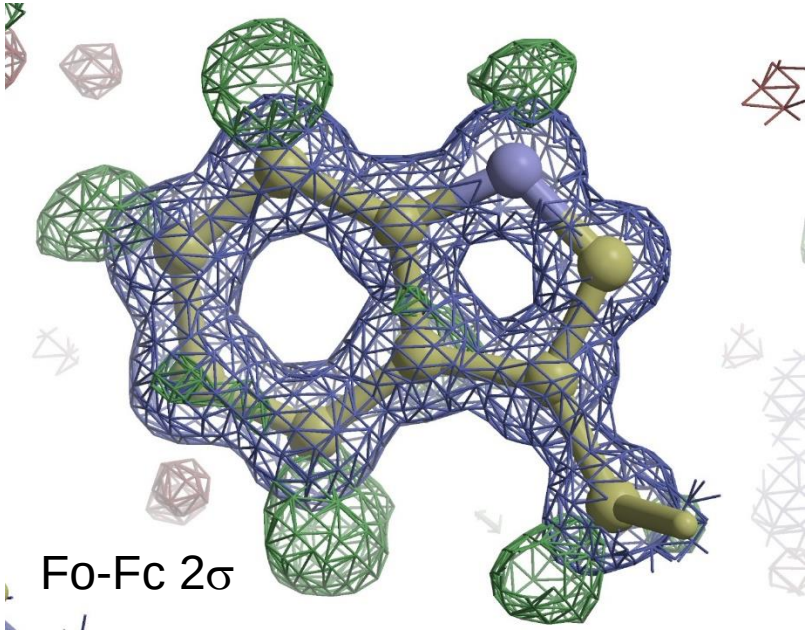


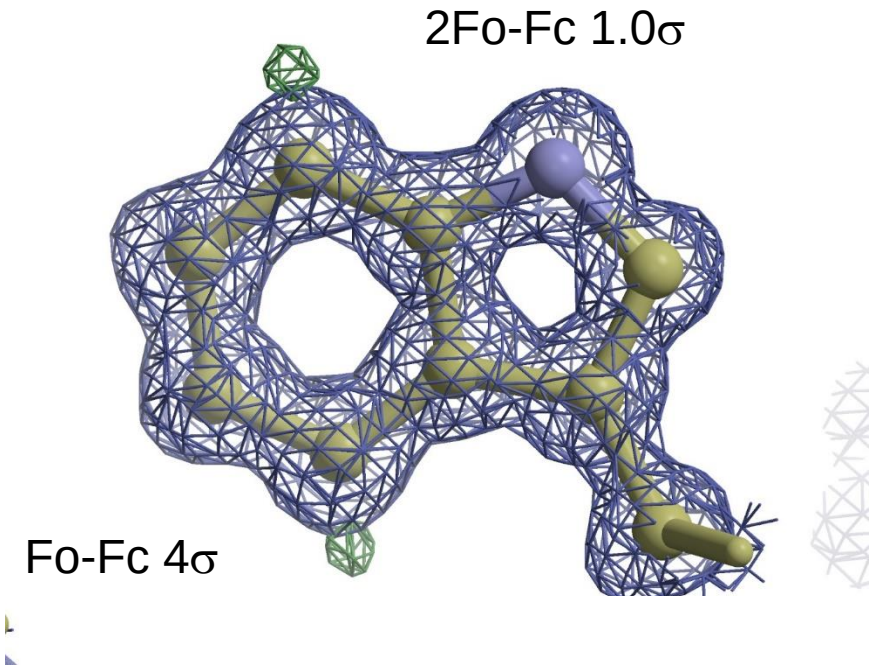


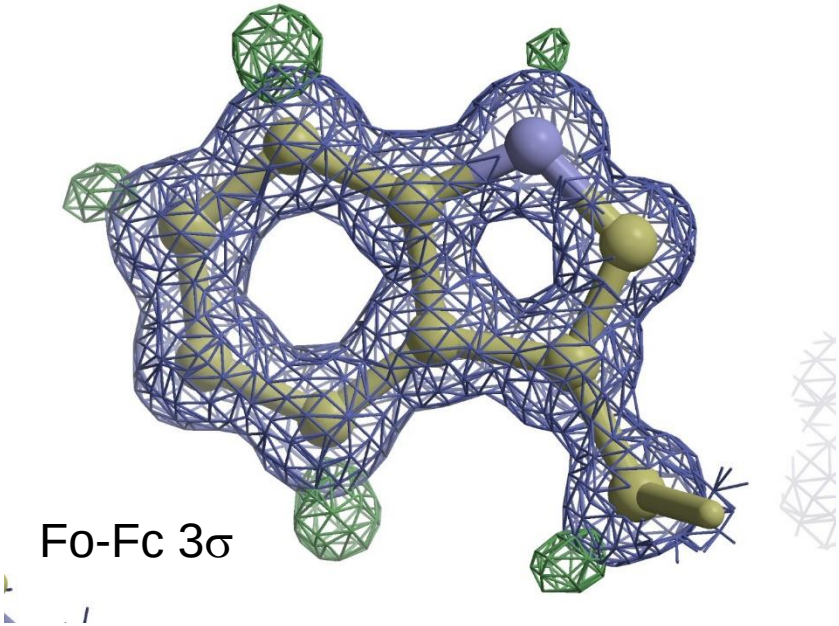
Diffracts to beyond 0.85 Å.

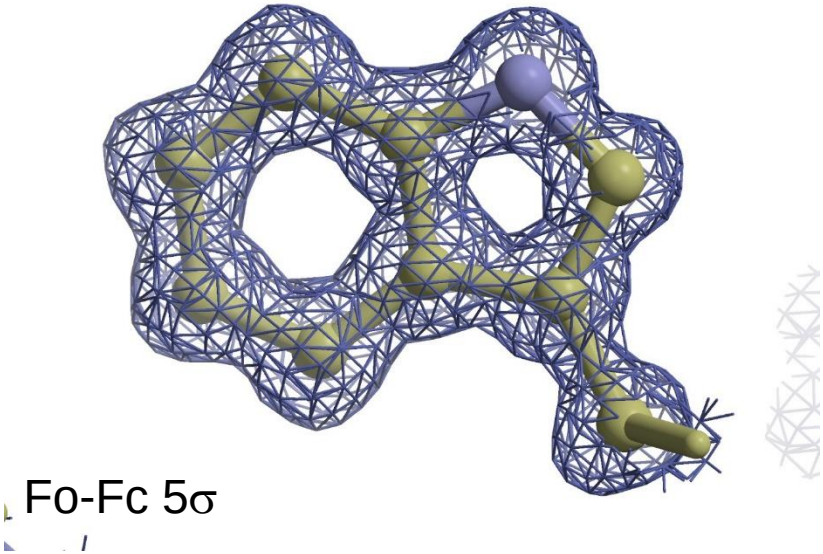
In this image ~5000 data points alone are visible.

The total data set at this resolution has over 1 million data points.


Fo-Fc 2σ


2Fo-Fc 1.0σ
Fo-Fc 4σ


Fo-Fc 3σ


Fo-Fc 5σ

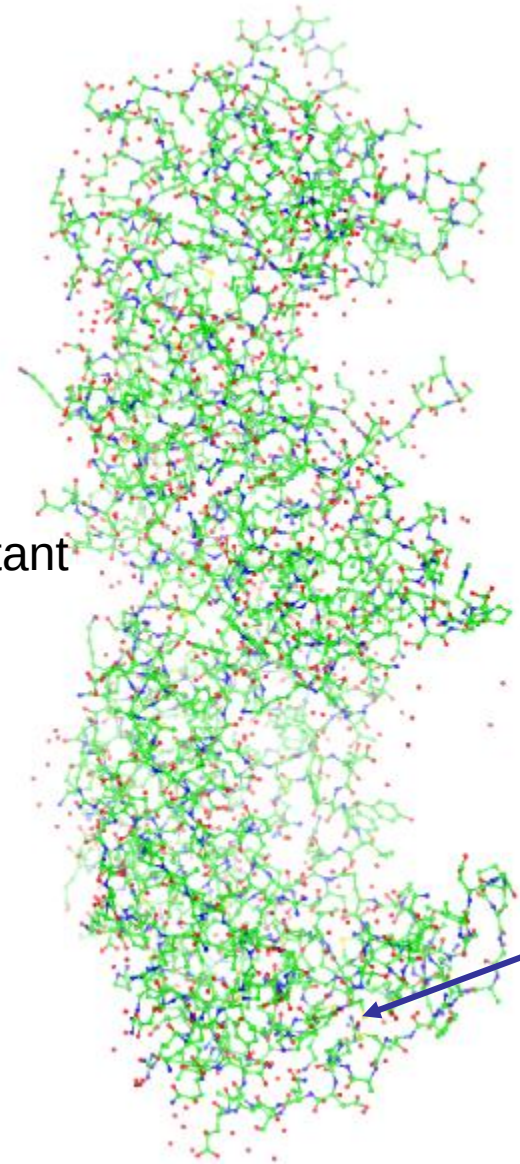
Crystals!

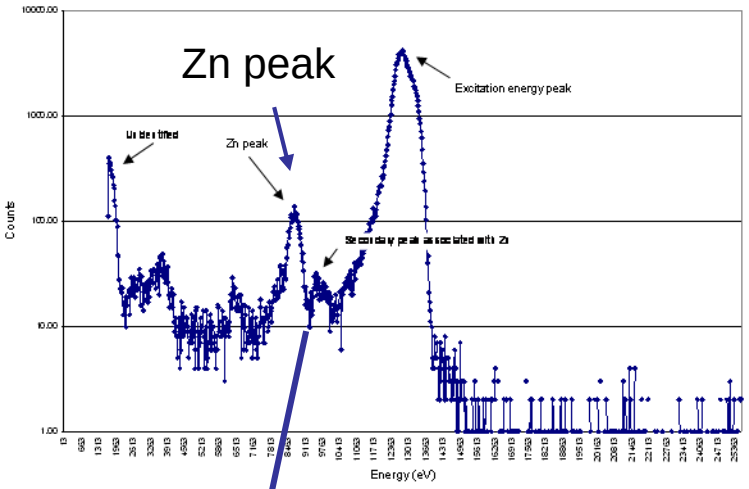
- But macromolecules are not found as crystals in plants, animals, or us (for the most part)?
  - The structures from X-ray diffraction are the same when the macromolecule crystallized from very different conditions.
  - The structures are the same (for the most part) when determined by NMR (NMR needs no crystals but is limited in macromolecuar size).
  - The solvent content of the crystal (30-70%) is comparable to that found in the cell.
  - When co-crystallized with ligands interactions seen in the structure seem to explain the known biochemistry.
  - Many macromolecules are active in the crystalline form.

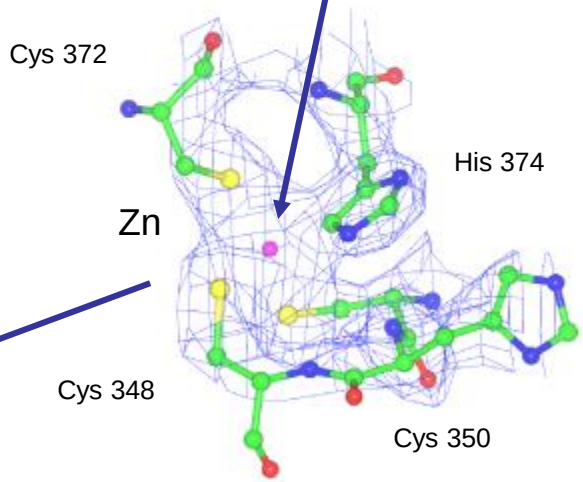
Macromolecules and Macromolecular Crystals are complex systems

Multiple components:

- Target
- Precipitant
- Buffer
- Cryoprotectant
- Solvent

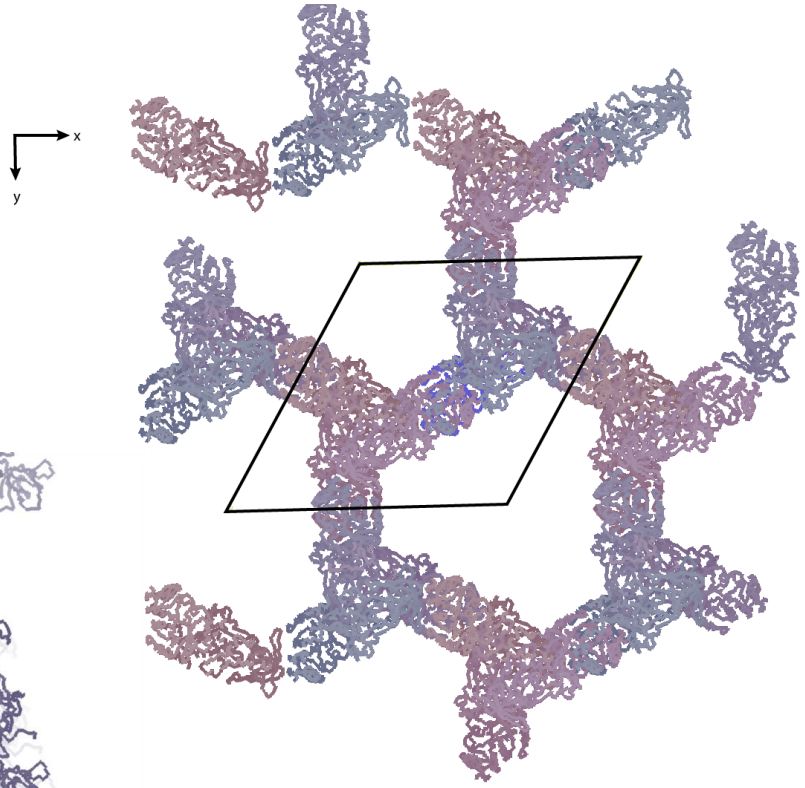


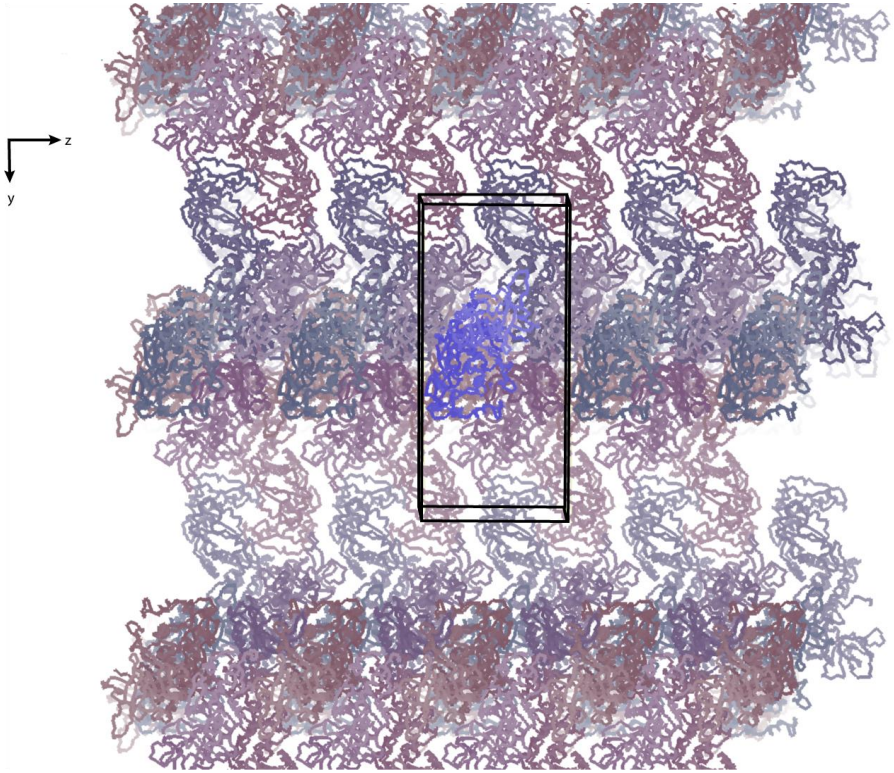




Current R=17.4%, R_{free}=21.6%

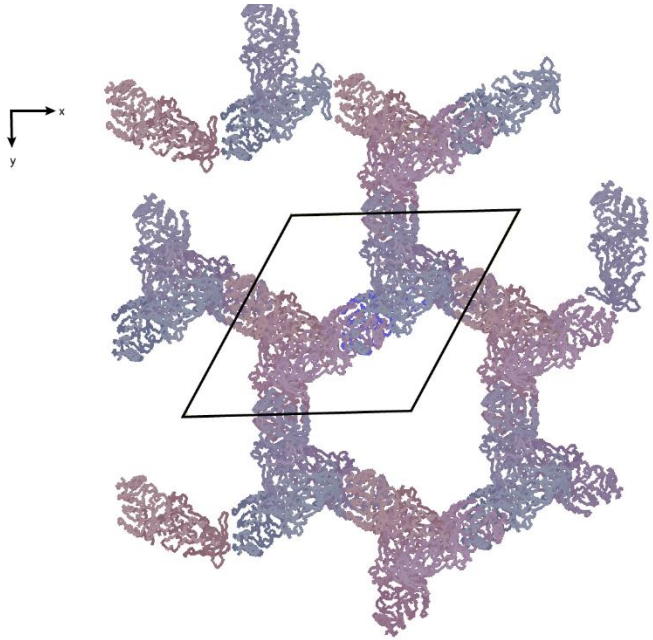
How are the molecules packed in the crystal?





Large water channels!

How many atoms are there in a <u>single</u> molecule in the crystal?

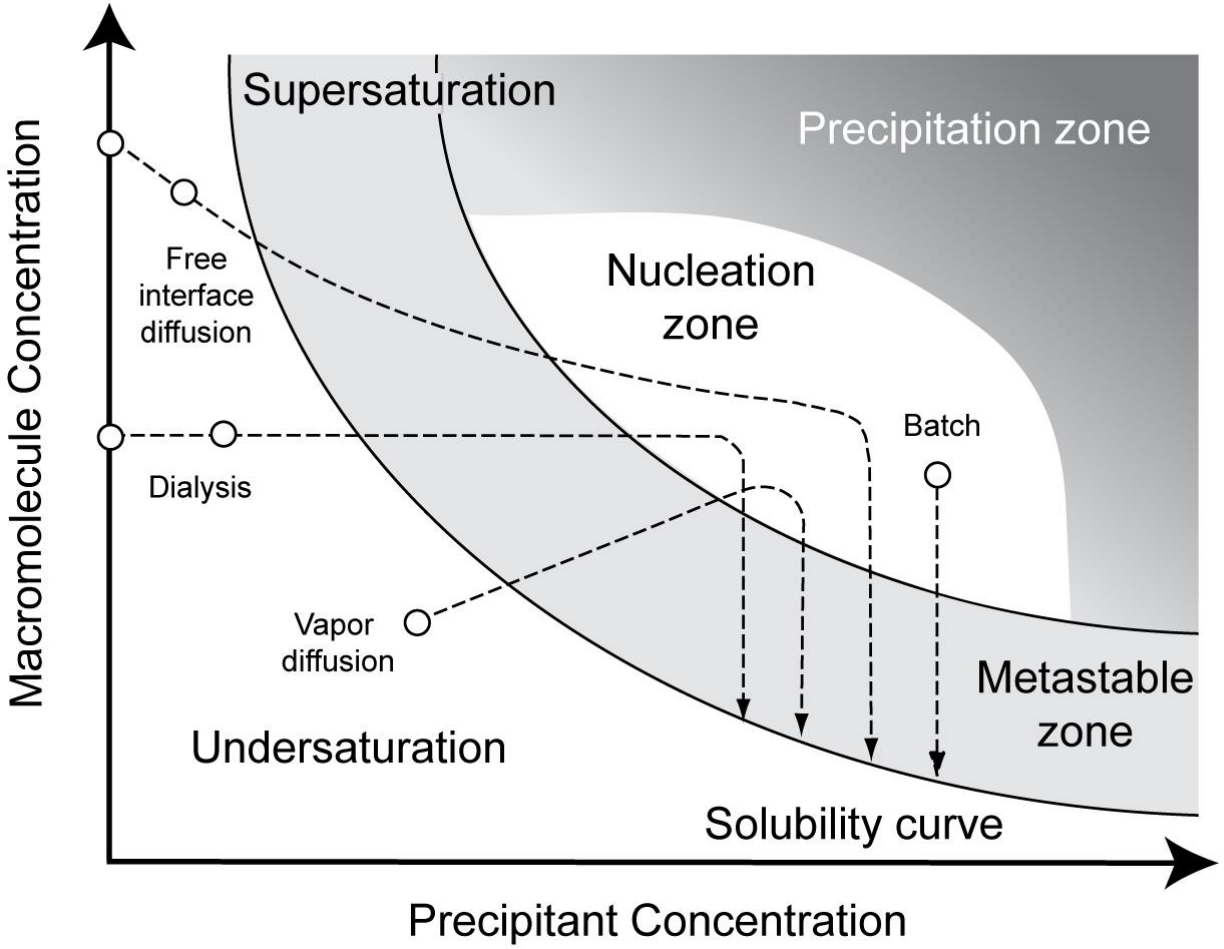


Atomic composition

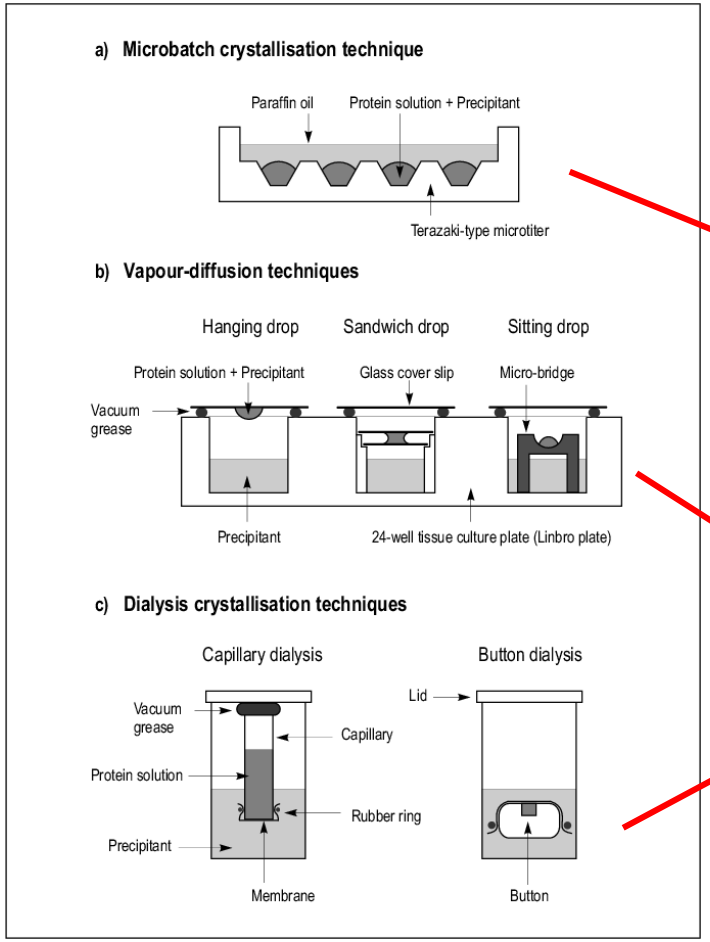
Carbon	4166
Hydrogen	6525
Nitrogen	1129
Oxygen	1248
Sulfur	23
Total number of atoms	13091
Formula = $C_{4166}H_{6525}N_{1129}O_{1248}S_{23}$	

On the order of 500,000,000 molecules in the crystal.

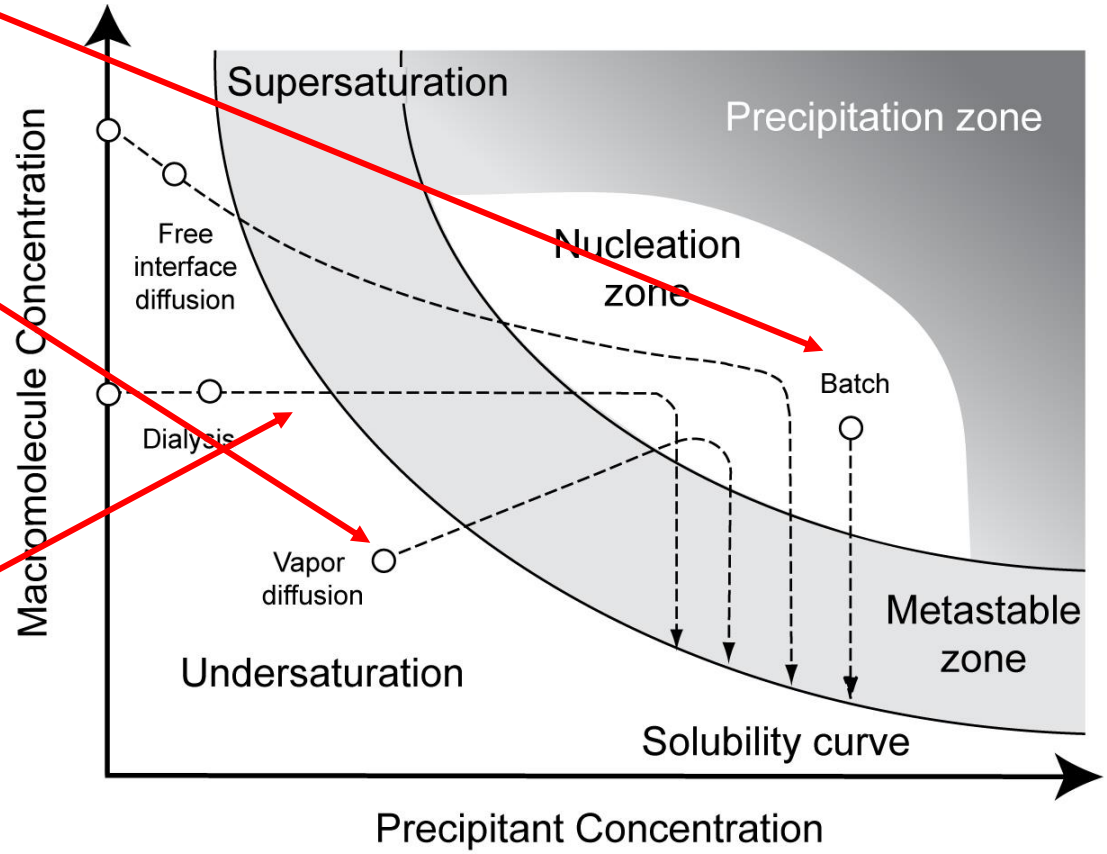
Simplified phase diagram for crystallization



Simplified phase diagram for crystallization



Different techniques traverse the crystallization space differently.



The Crystallization Screening Center at the Hauptman-Woodward Medical Research Institute

Since February of 2000 the High Throughput Crystallization Center has been screening potential crystallization conditions as a high-throughput service

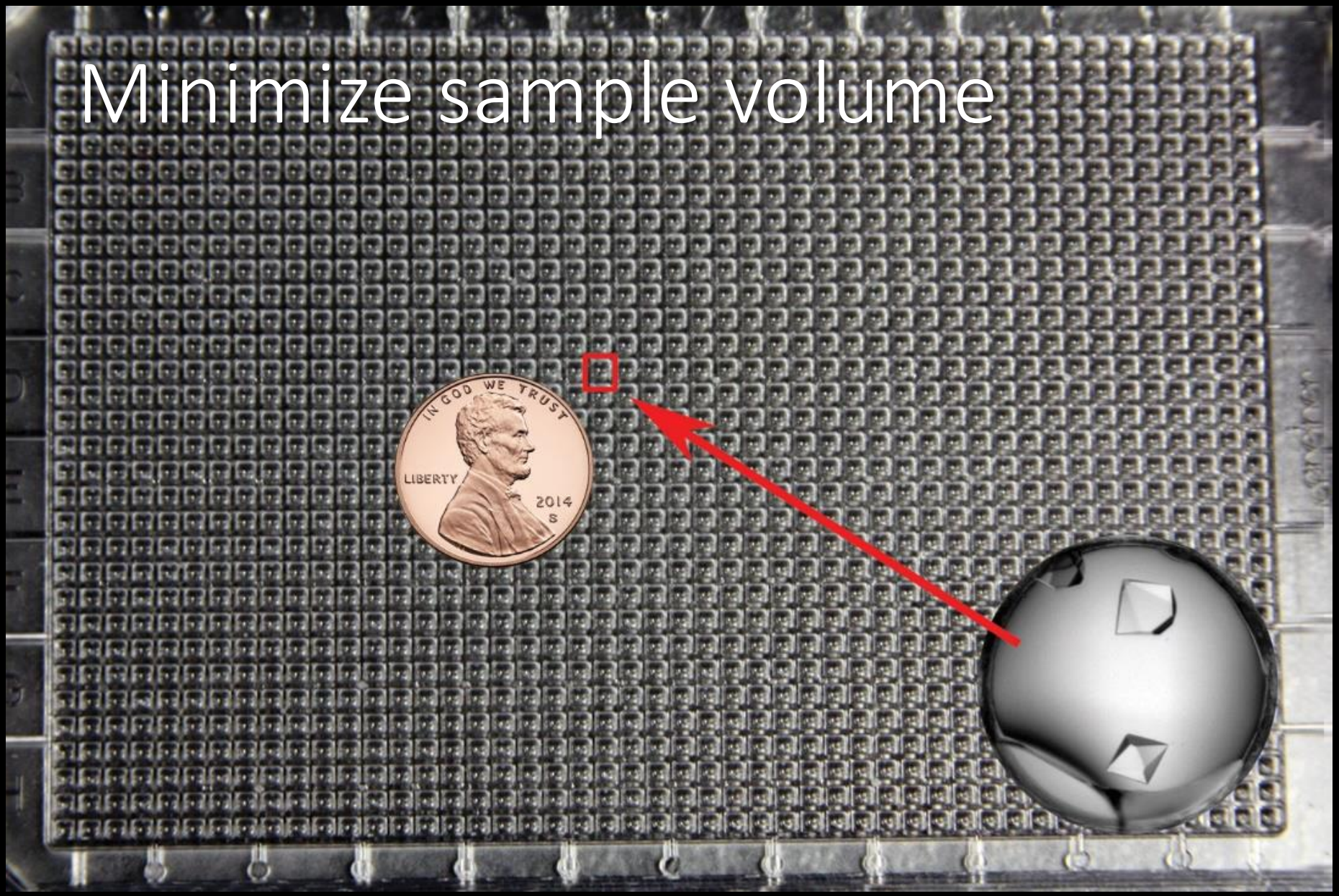
The HTS lab screens samples against three types of cocktails:

1. Buffered salt solutions varying pH, anion and cation and salt concentrations
2. Buffered PEG and salt, varying pH, PEG molecular weight and concentration and anion and cation type
3. Almost the entire Hampton Research Screening catalog.

The HTSlab has investigated the crystallization properties of over 16,000 individual proteins archiving approximately 160 million images of crystallization experiments.

All data and in many cases, dead volume recovered samples are available

Minimize sample volume



http://Getacrystal.org

High-Throughput Crysta

hwi.buffalo.edu/science/high-throughput-crystallization-center/

HAUPTMAN–WOODWARD
MEDICAL RESEARCH INSTITUTE

High-Throughput Crystallization Center

Our expertise – your results

The High-throughput Crystallization Screening Center performs crystal-growth screening experiments in 1,536-well microassay plates for both soluble and membrane biological macromolecules. You send the sample, we do the rest.

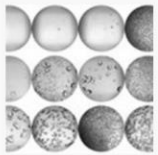
- Some 1,536 different chemical conditions are sampled representing many of the commercially available crystallization screens. Each condition is imaged visually over a period of six weeks. Second Order Non-linear Imaging of Chiral Crystals (SONICC) and UV Two Photon Emitted Fluorescence (UV-TPEF) imaging of crystallization screening also take place to identify biological crystals not picked up visually or those obscured by precipitate.
- Color images, and SONICC and UV-TPEF images are integrated with the analysis software provided as part of the service. Lead crystallization conditions can be identified and the complete set of results used to analyze the samples solubility phase diagram.

From February 2000, the laboratory has been used to set up over 24 million crystallization experiments on >16,000 biological macromolecules. Over 1,000 laboratories worldwide have used and benefited from the service.

Click "Get a crystal" for details.

This service is for academic and other not-for-profit users only. Our commercial users go through '*the structure people*', our friends at HarkerBIO (www.harkerbio.com).

Get a Crystal (send us your sample)



Analyze your Results to Optimize

Frequently Asked Questions

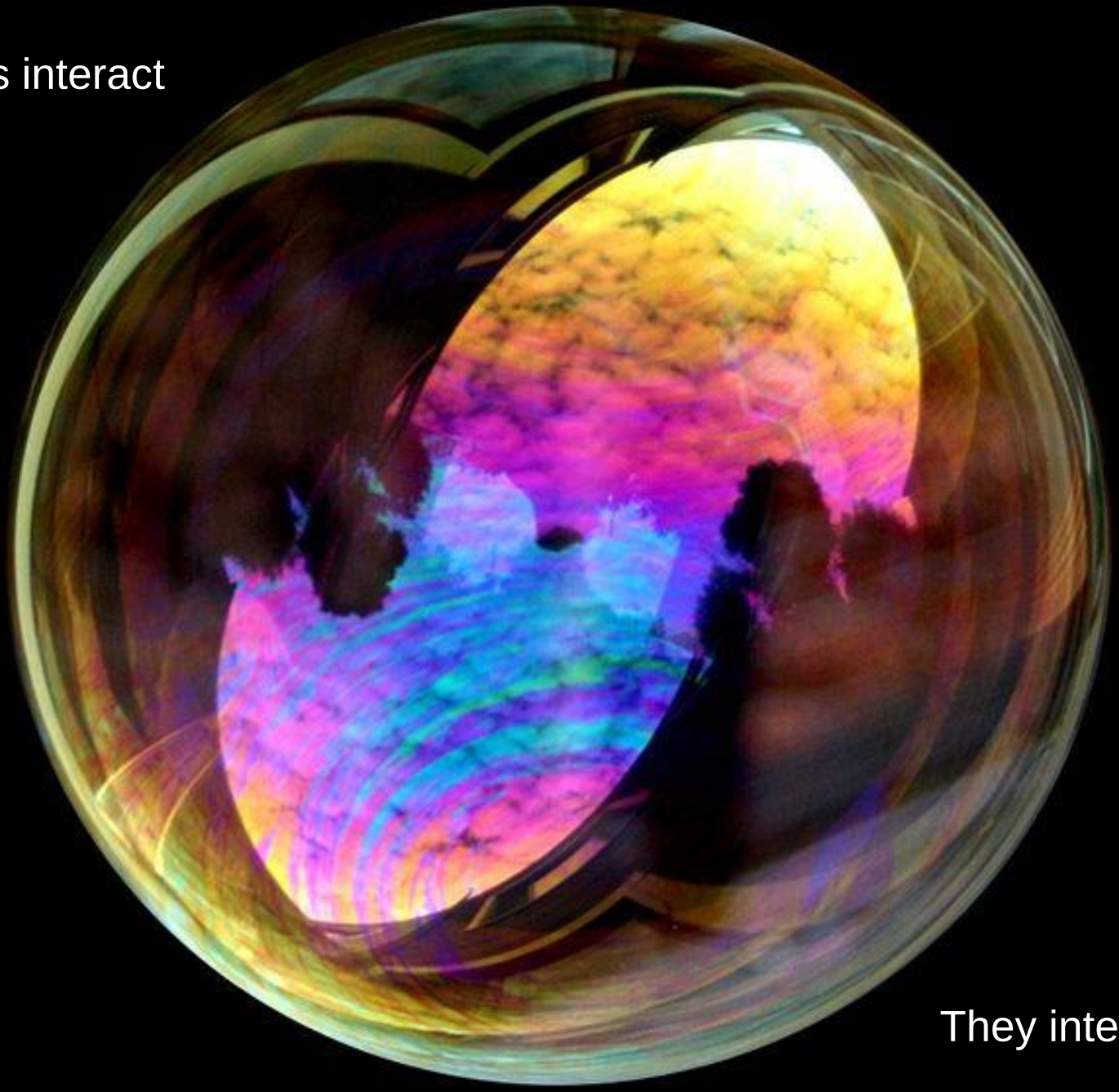
Crystallization Research

Crystallization screening, 1536 conditions including most of Hampton Research. SONICC and UV imaging.

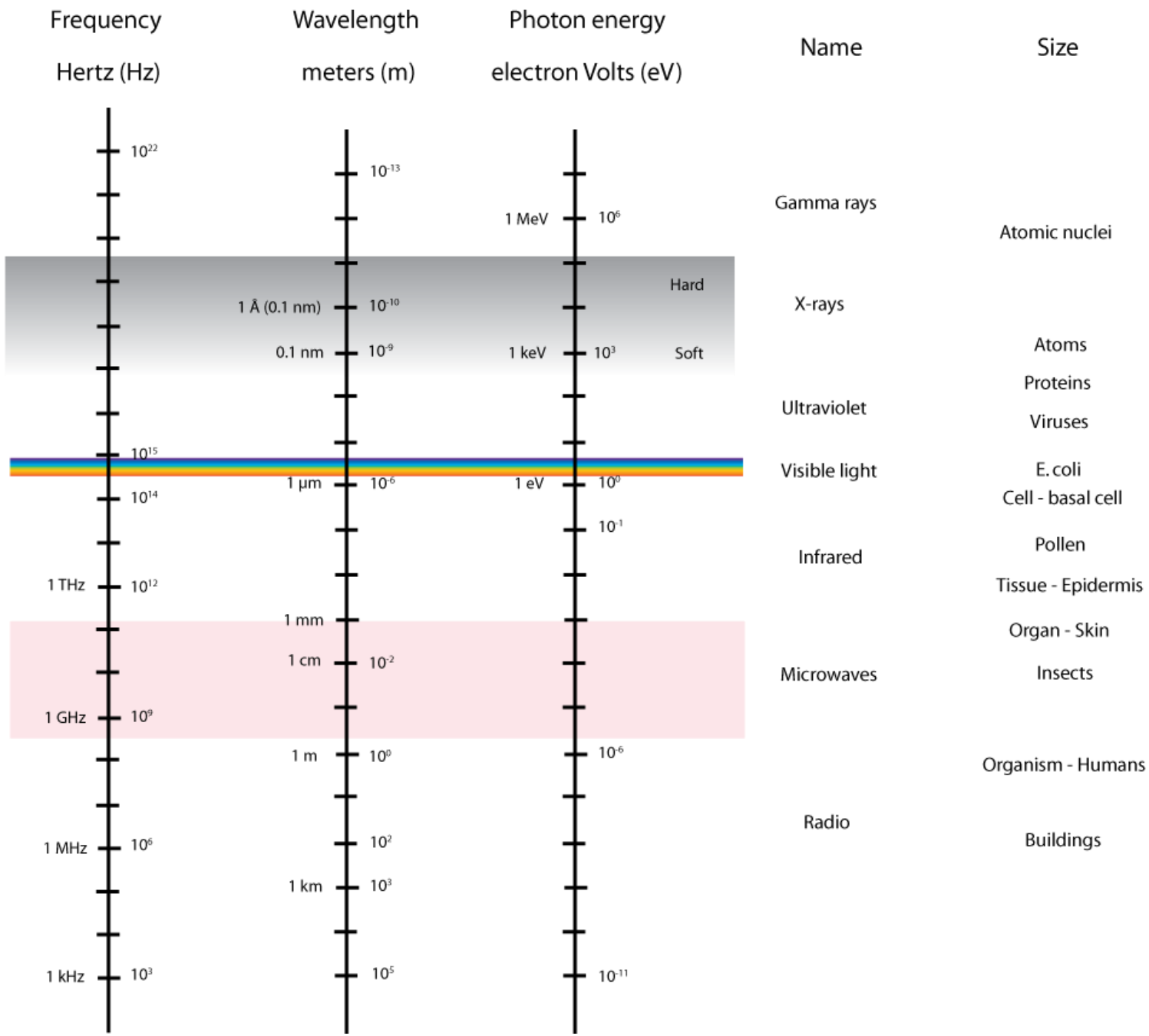
$375 per sample

Waves

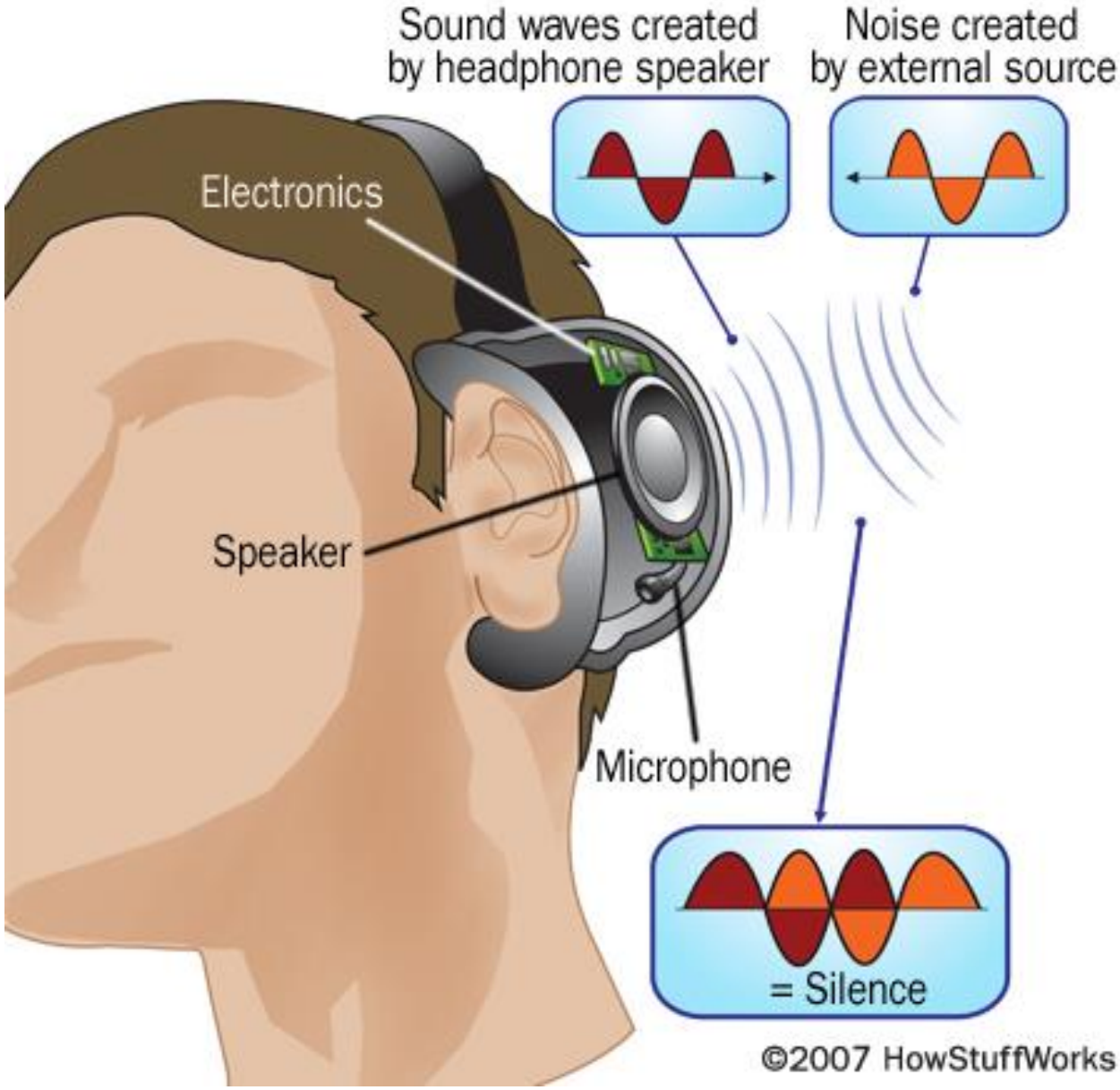
Waves interact



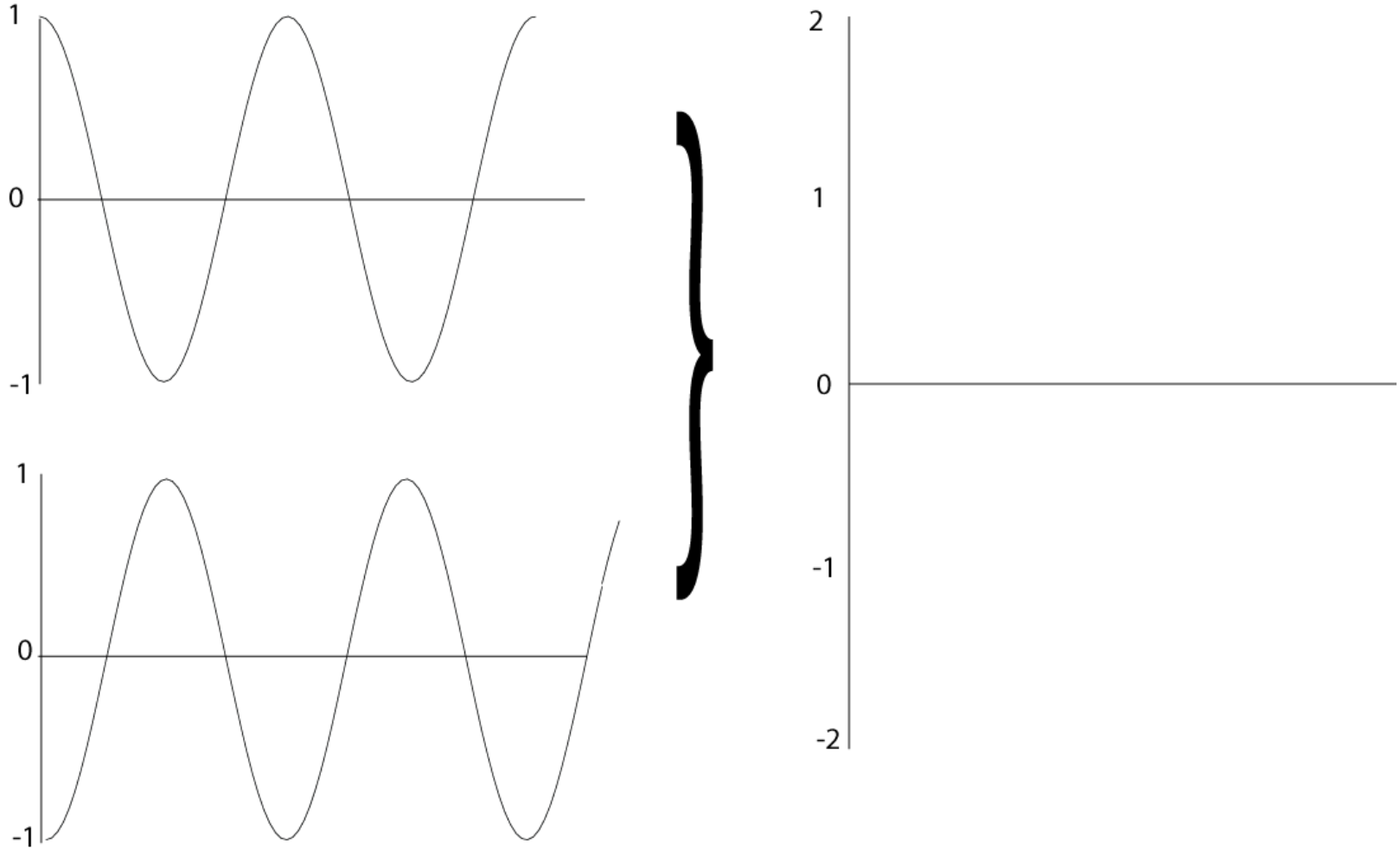
They interfere


Frequency
Hertz (Hz)
Wavelength
meters (m)
Photon energy
electron Volts (eV)
Name
Size
10^{22}
10^{-13}
1 MeV
10^{6}
Gamma rays
Atomic nuclei
Hard
1 Å (0.1 nm)
10^{-10}
X-rays
0.1 nm
10^{-9}
1 keV
10^{3}
Soft
Atoms
Proteins
Ultraviolet
Viruses
10^{15}
Visible light
E. coli
1 µm
10^{-6}
1 eV
10^{0}
10^{14}
Cell - basal cell
10^{-1}
Infrared
Pollen
1 THz
10^{12}
Tissue - Epidermis
1 mm
Organ - Skin
1 cm
10^{-2}
Microwaves
Insects
1 GHz
10^{9}
1 m
10^{0}
10^{-6}
Organism - Humans
Radio
1 MHz
10^{6}
10^{2}
Buildings
1 km
10^{3}
1 kHz
10^{3}
10^{5}
10^{-11}

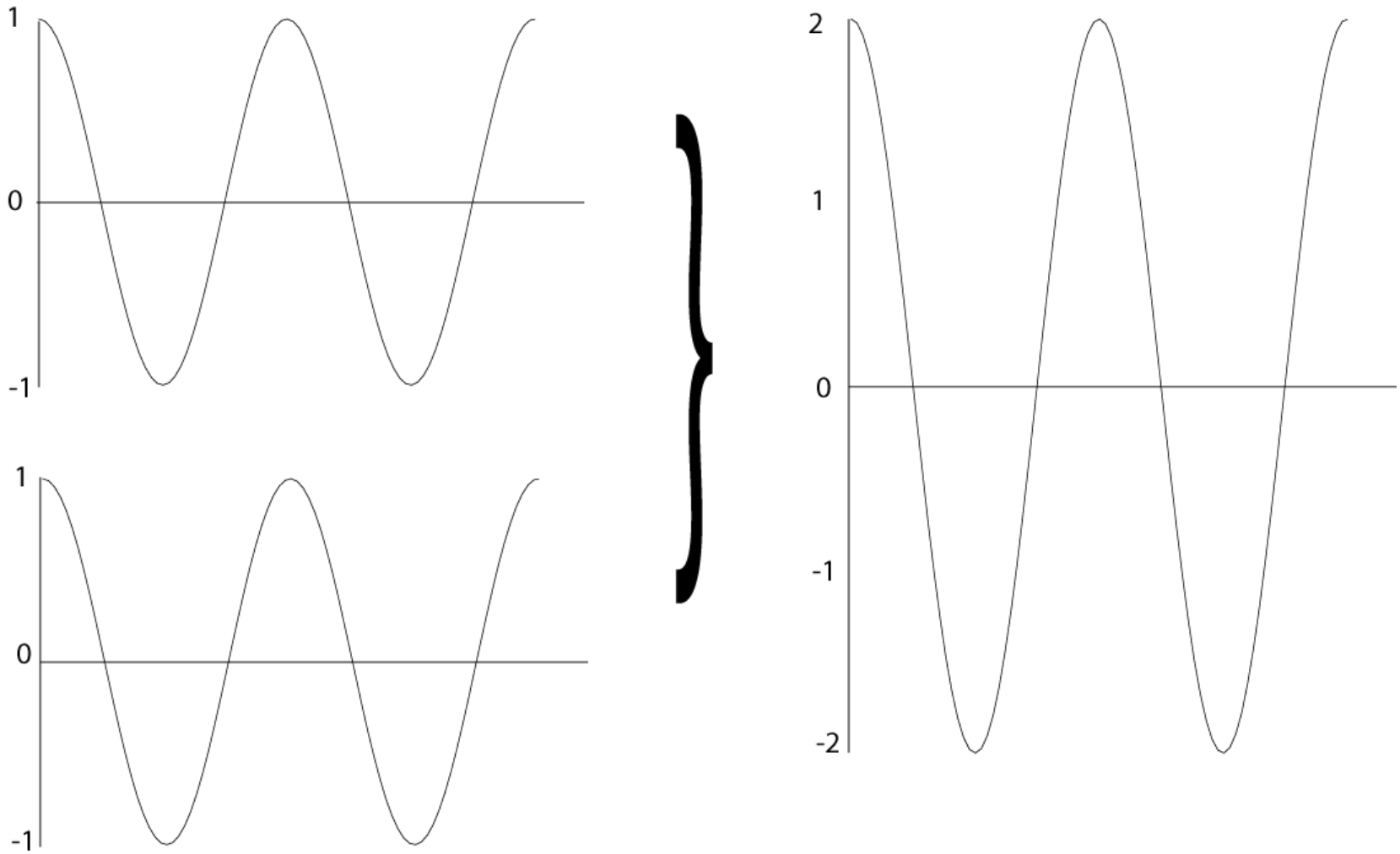
Inside noise-canceling headphones



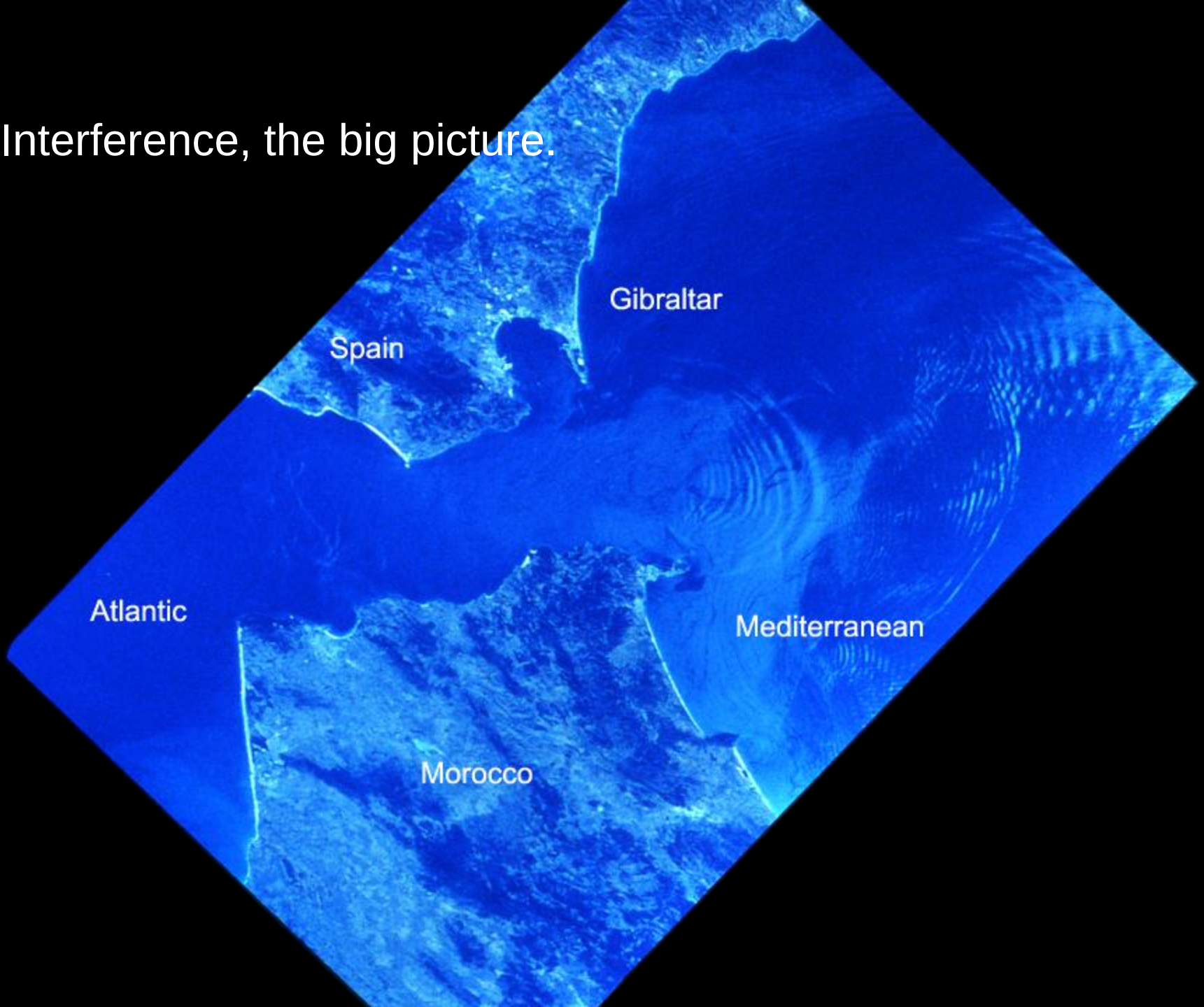
Destructive interference.

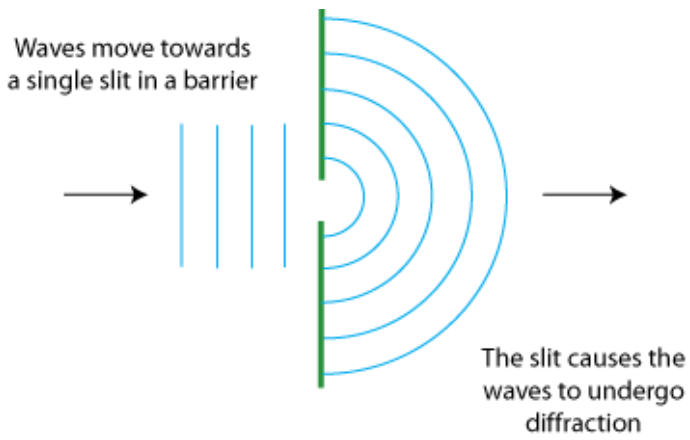


Constructive interference.



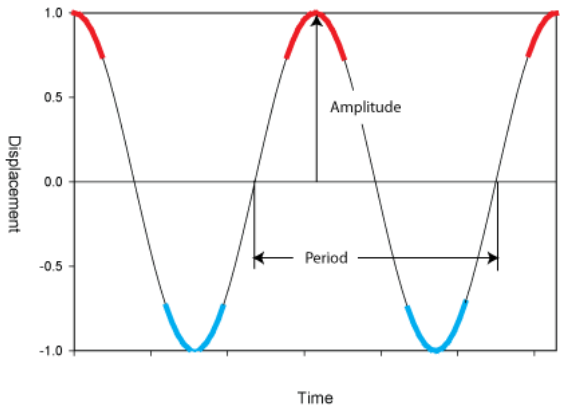
Interference, the big picture.

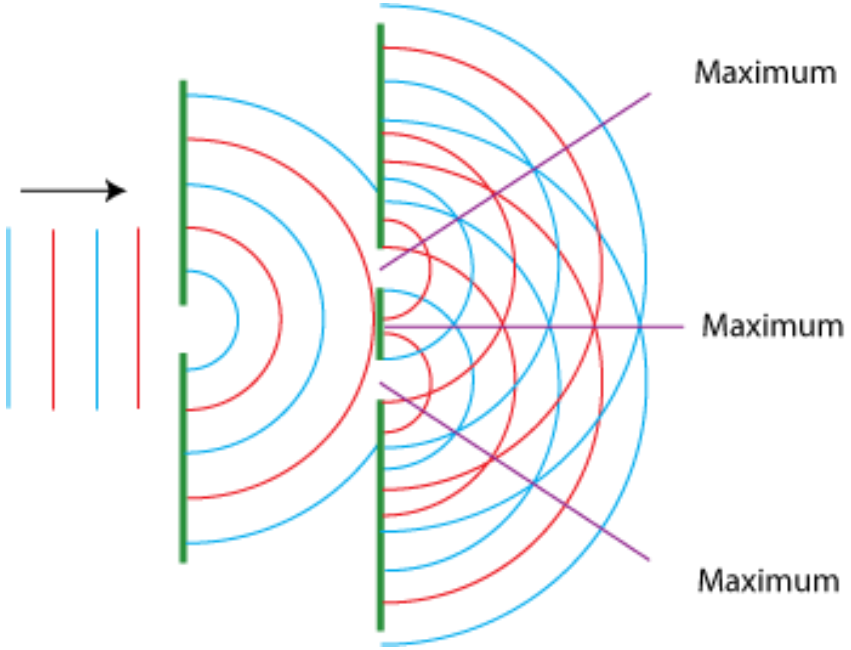


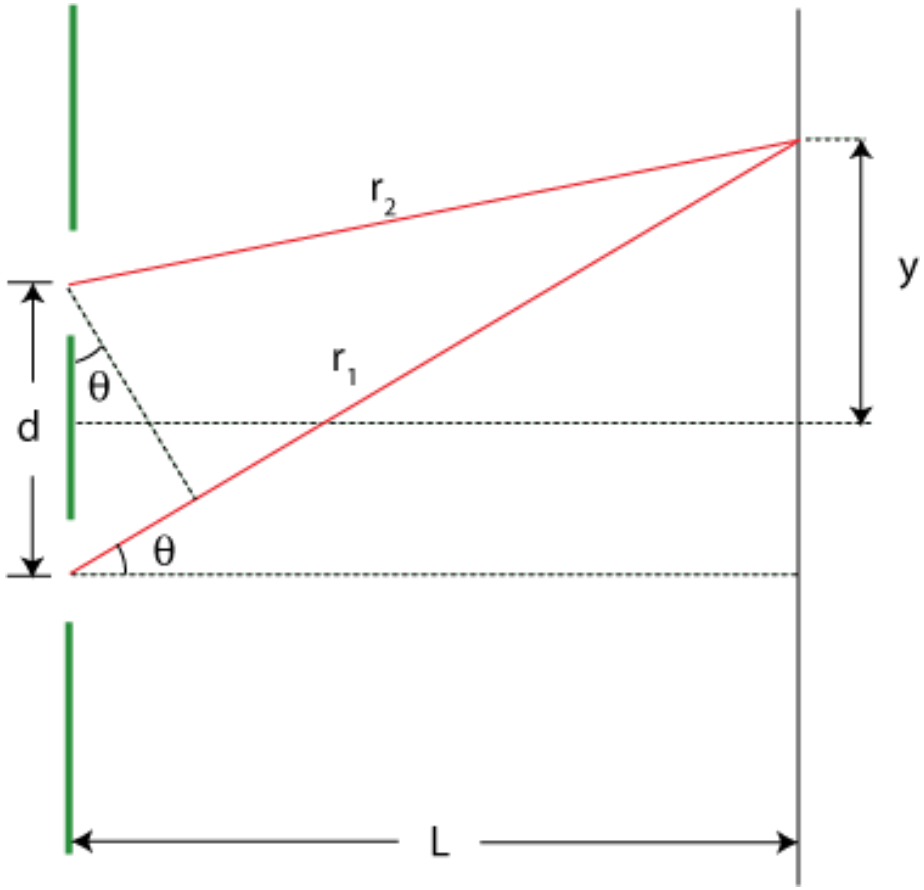


Huygen's principle – waves spread out as they pass through slits.

This spreading out is called diffraction. In general this occurs when the waves pass through small openings, around obstacles or by sharp edges







For constructive interference the path difference between r_1 and r_2 has to be a whole wavelength λ

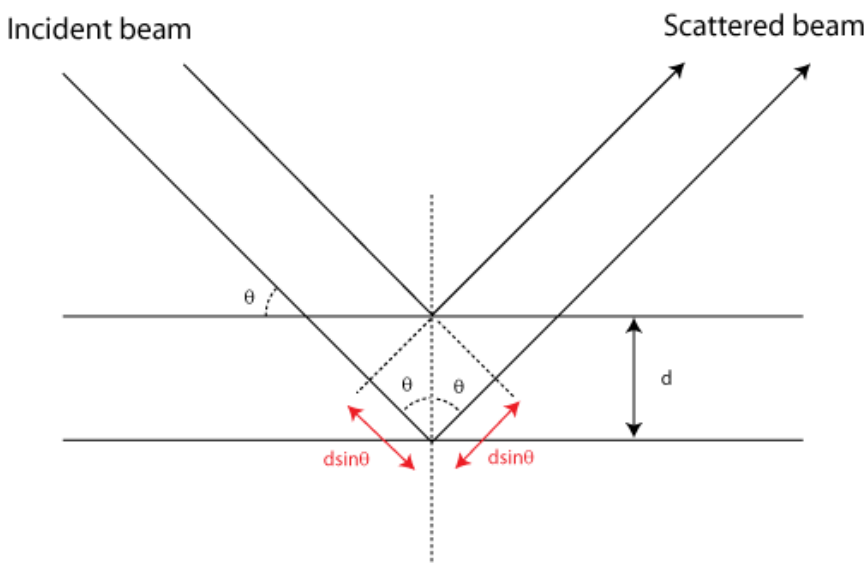
$r_1 - r_2 = \lambda = d\sin\theta$

remembering that sine is the ratio of the opposite over hypotenuse.

For interference at successive λ path differences the equation becomes

$n\lambda = d\sin\theta$

Where n is +/- 1, +/-2 etc.



Path difference for constructive interference is when the wavelength is a multiple of 2dsinθ or Braggs law:

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

Braggs law

Jean Baptiste Fourier (1763-1830)

In 1807 came up with an idea ….

Any periodic function could be rewritten as a weighted sum of sines and cosines of different frequencies.

This was not regarded as possible by other mathematicians of the time and it was not until 1878 that the idea was published in English.

The name of the idea is the Fourier series

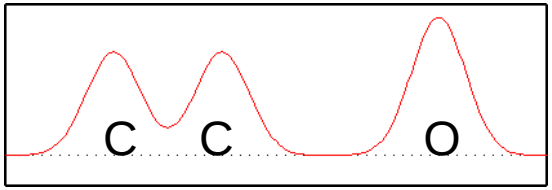
To change a signal to a Fourier series we perform a Fourier transform. To change the Fourier transform to a signal we perform an inverse Fourier transform.

Fourier's theorem is not only one of the most beautiful results of modern analysis, but it may be said to furnish an indispensable instrument in the treatment of nearly every recondite question in modern physics.

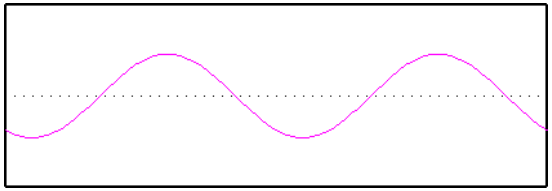
Lord Kelvin (1824-1907)

Other Lord Kelvin quotes

- Heavier-than-air flying machines are impossible.
- Radio has no future.
- In science there is only physics; all the rest is stamp collecting.

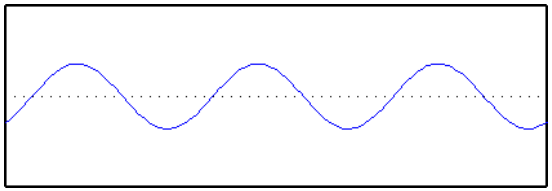


= A unit cell of a crystal with 2 carbons and an oxygen



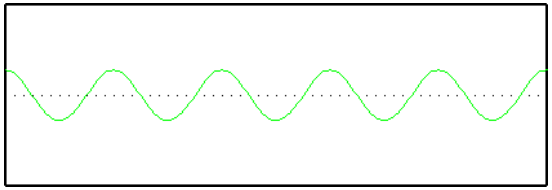
A cosine wave with frequency of 2, one peak represents the oxygen and the other the two carbons

+



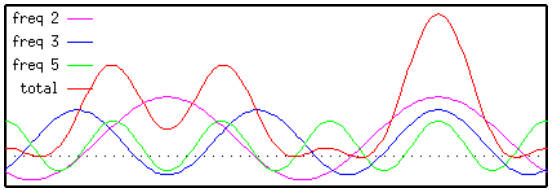
Add a cosine wave with frequency of 3, three repeats across the crystal. Note the phase is different, it starts in a different place.

+

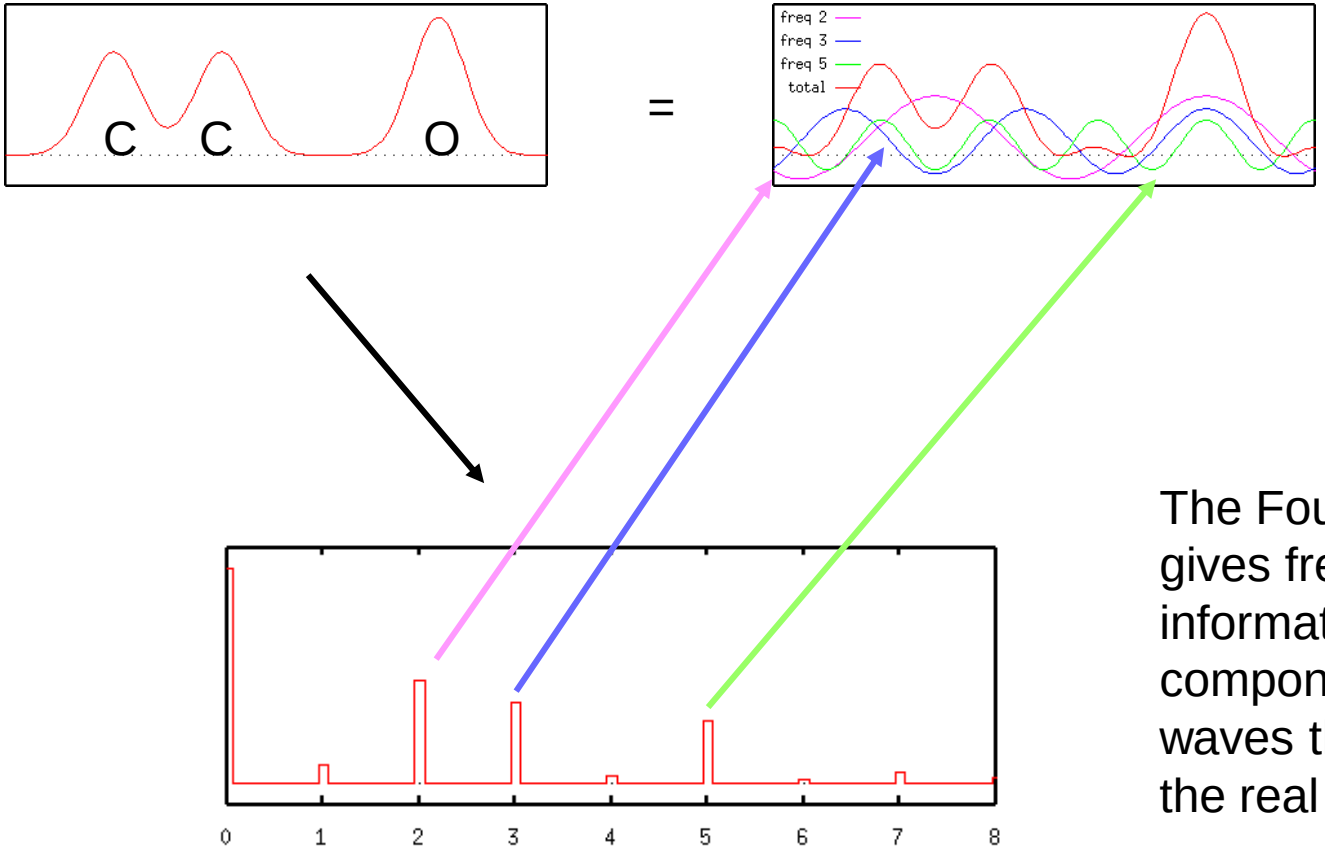


Add a third cosine wave with frequency of 5, with the peaks lined up on the carbons

=



Add all the waves and the result is the original unit cell of the crystal

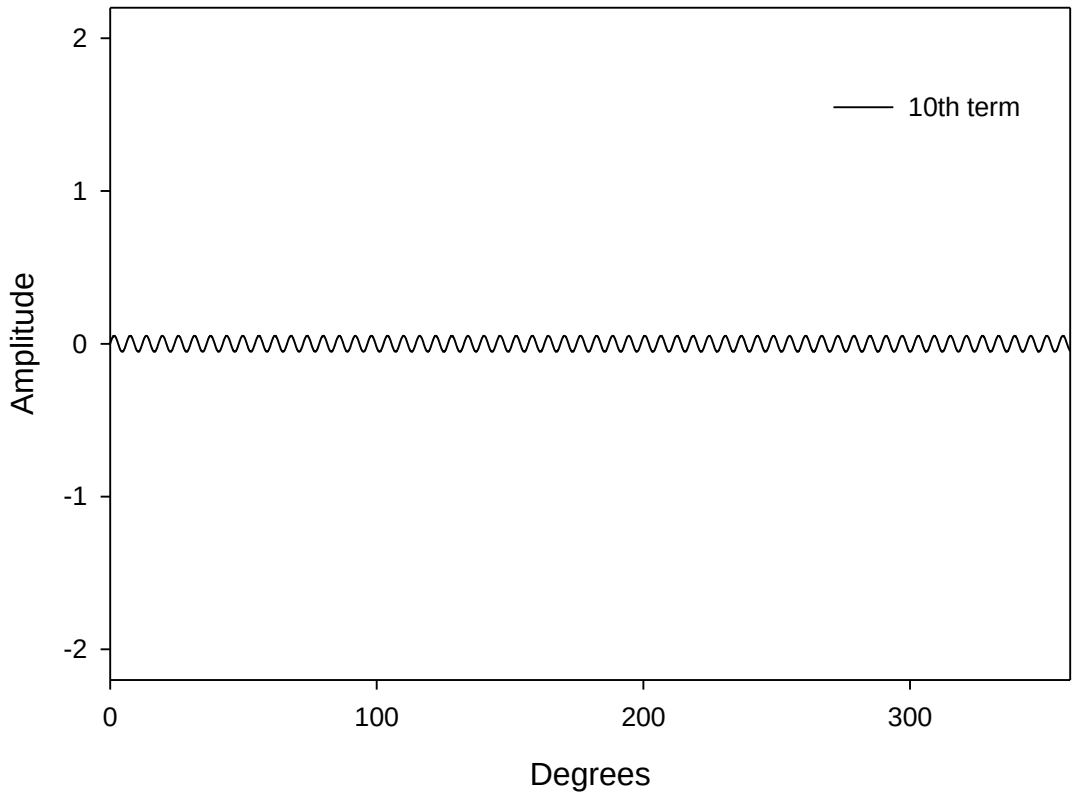


The Fourier transform gives frequency information on the components of the waves that describe the real space.

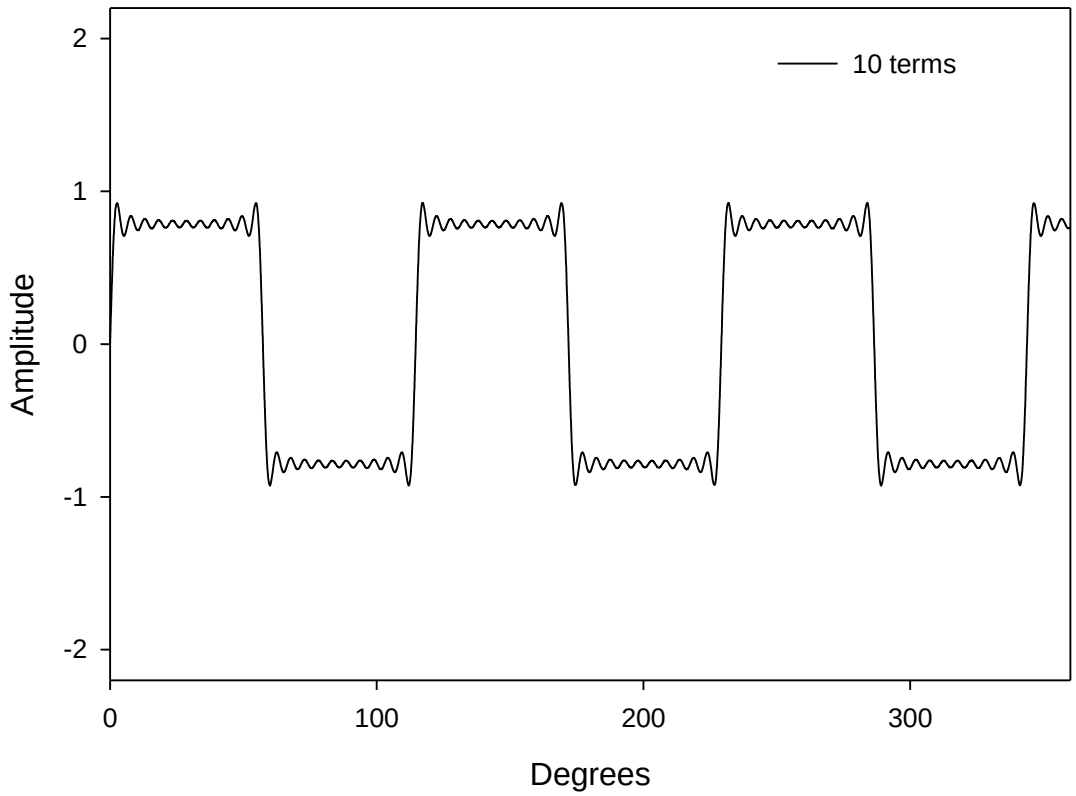
The Fourier transform of the unit cell

Check out http://www.ysbl.york.ac.uk/~cowtan/fourier/ftheory.html

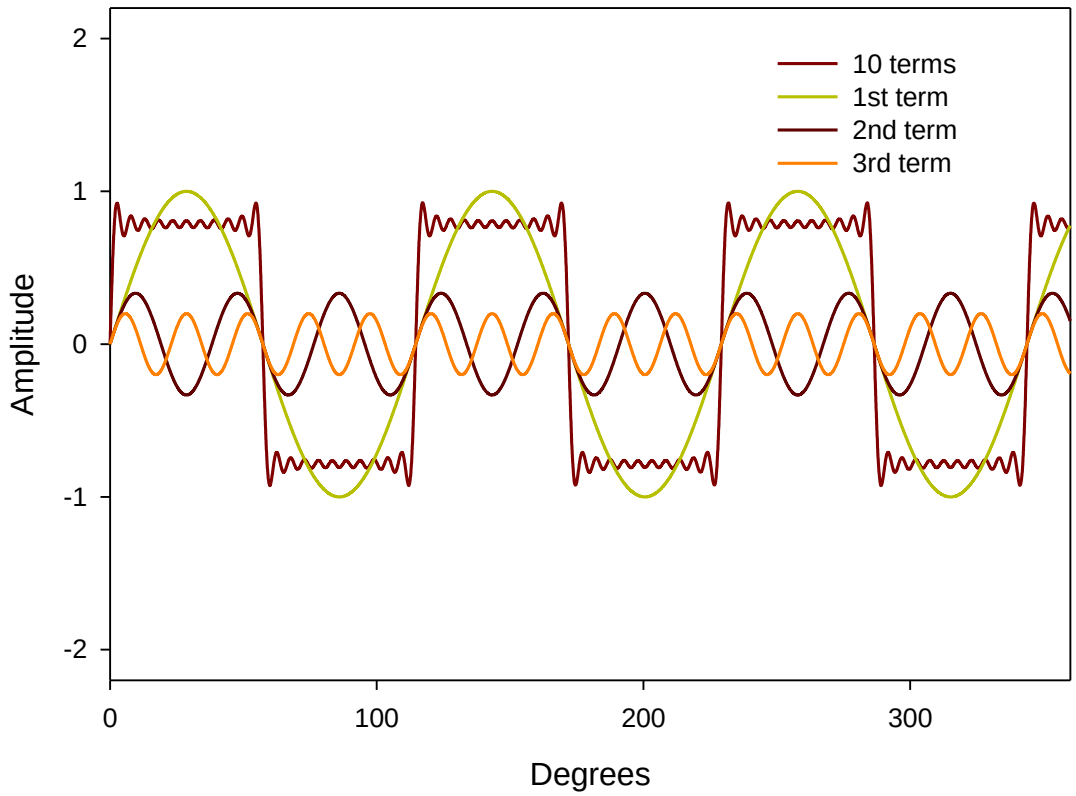
$$f(x) = \frac{4}{\pi}\left(\frac{\sin \pi x}{1} + \frac{\sin 3\pi x}{3} + \frac{\sin 5\pi x}{5} + \frac{\sin 7\pi x}{7} + \ldots\right)$$

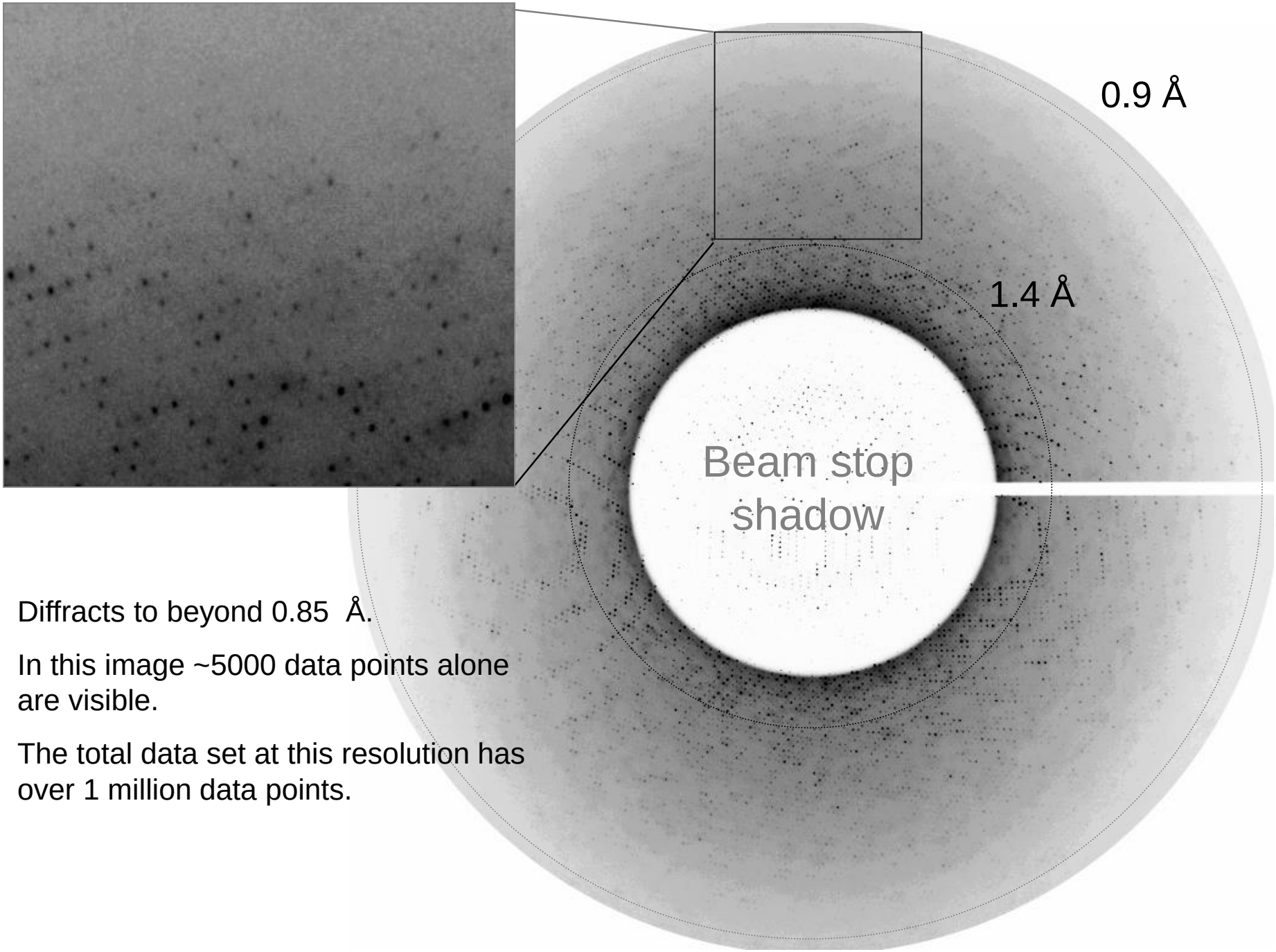


$$f(x) = \frac{4}{\pi}\left(\frac{\sin \pi x}{1} + \frac{\sin 3\pi x}{3} + \frac{\sin 5\pi x}{5} + \frac{\sin 7\pi x}{7} + \ldots\right)$$



$$f(x) = \frac{4}{\pi}\left(\frac{\sin \pi x}{1} + \frac{\sin 3\pi x}{3} + \frac{\sin 5\pi x}{5} + \frac{\sin 7\pi x}{7} + \ldots\right)$$

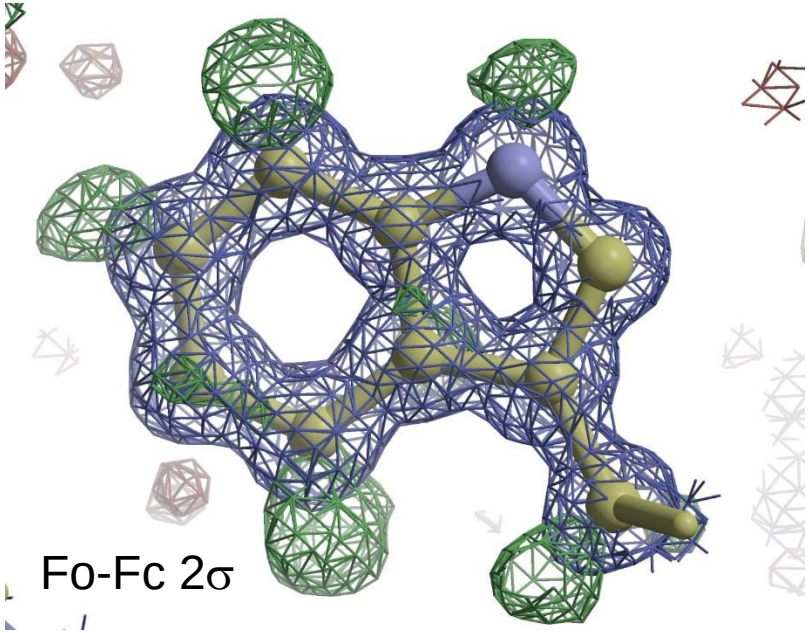


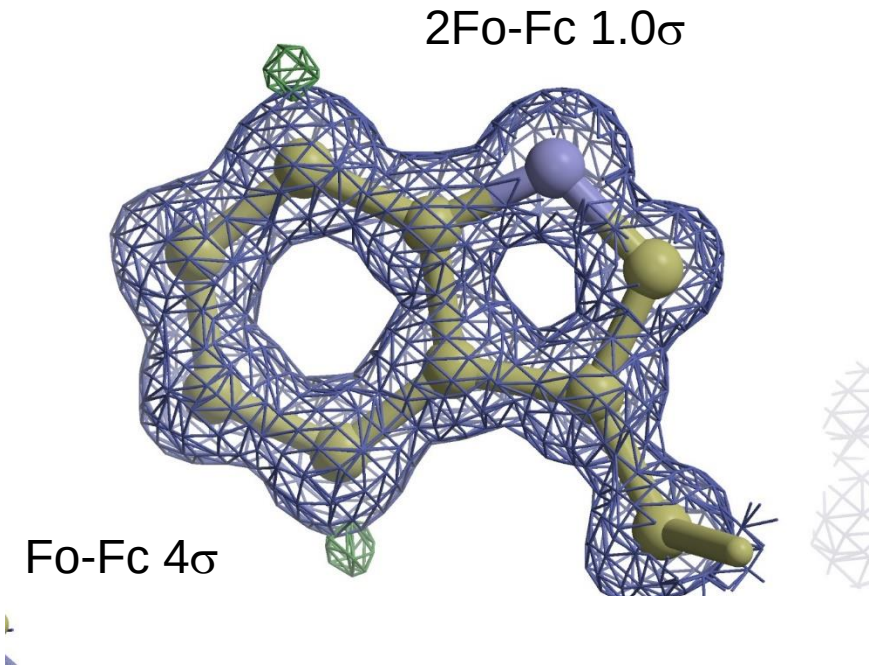


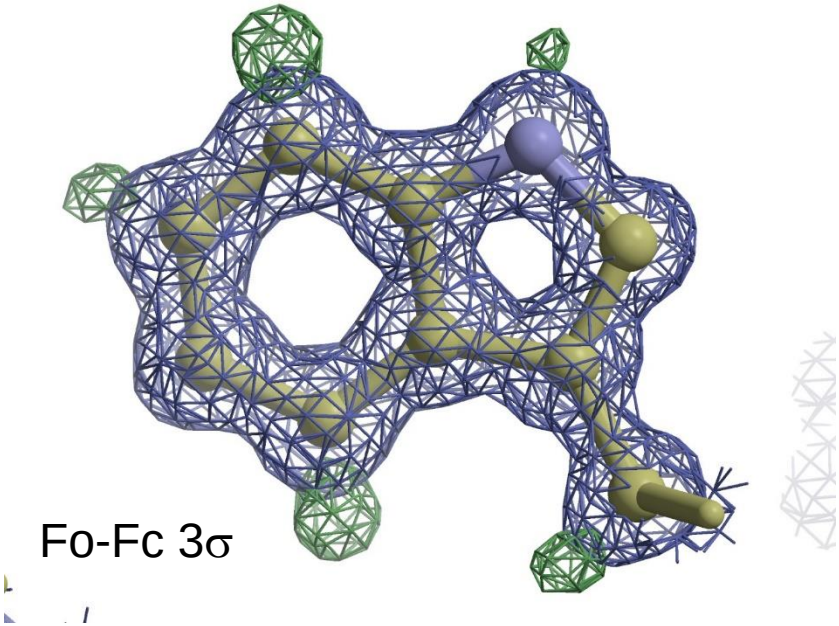
Diffracts to beyond 0.85 Å.

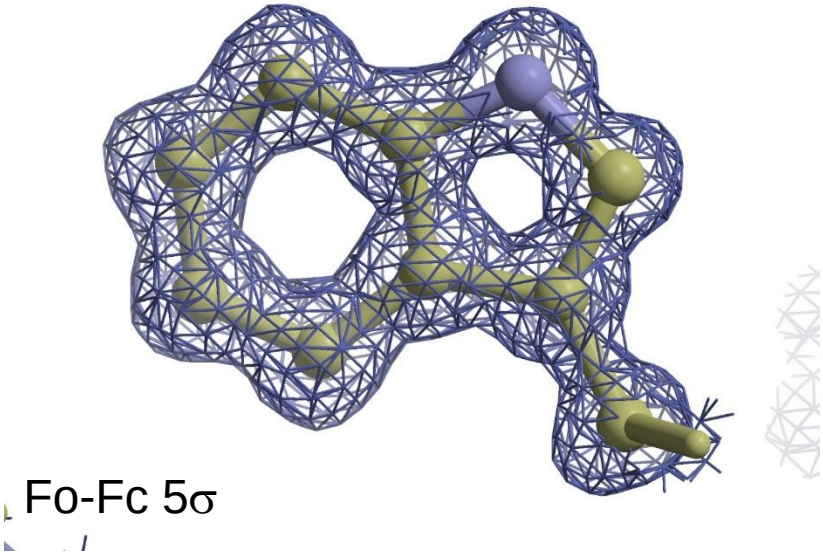
In this image ~5000 data points alone are visible.

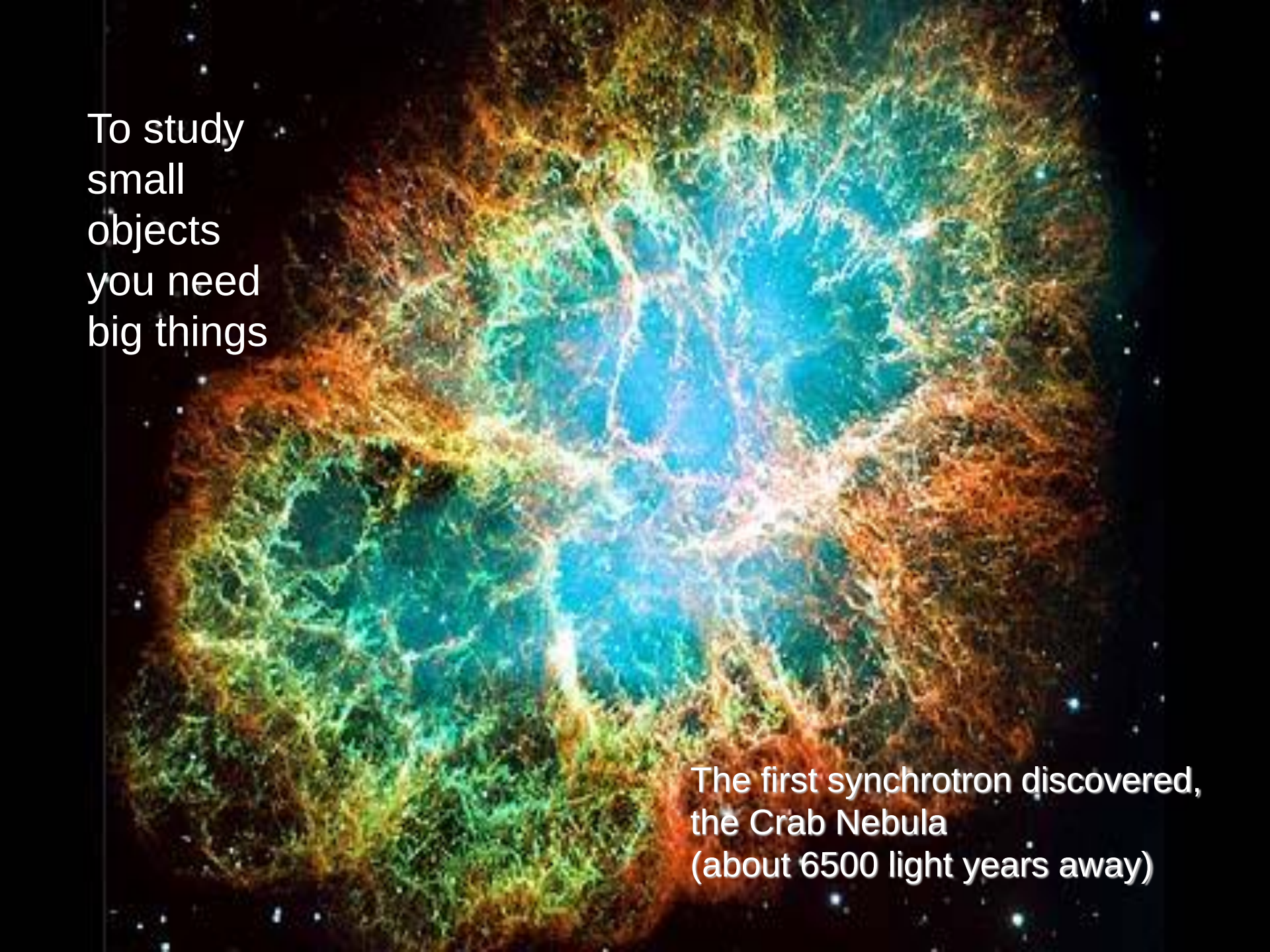
The total data set at this resolution has over 1 million data points.


Fo-Fc 2σ


2Fo-Fc 1.0σ
Fo-Fc 4σ


Fo-Fc 3σ


Fo-Fc 5σ


To study small objects you need big things
The first synchrotron discovered, the Crab Nebula
(about 6500 light years away)

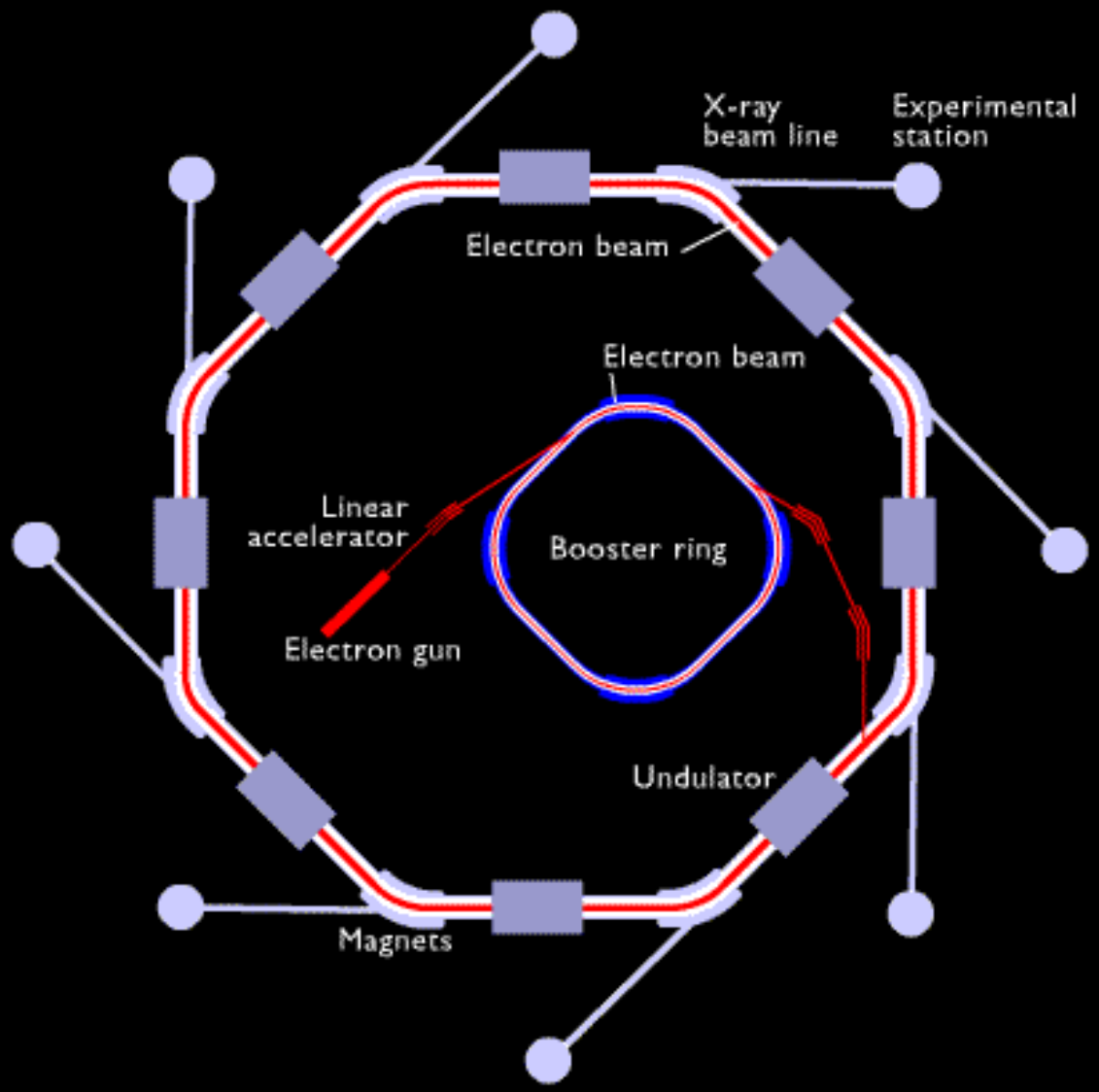
A synchrotron accelerates and stores particles (electrons or protons) moving at speeds close to that of light.

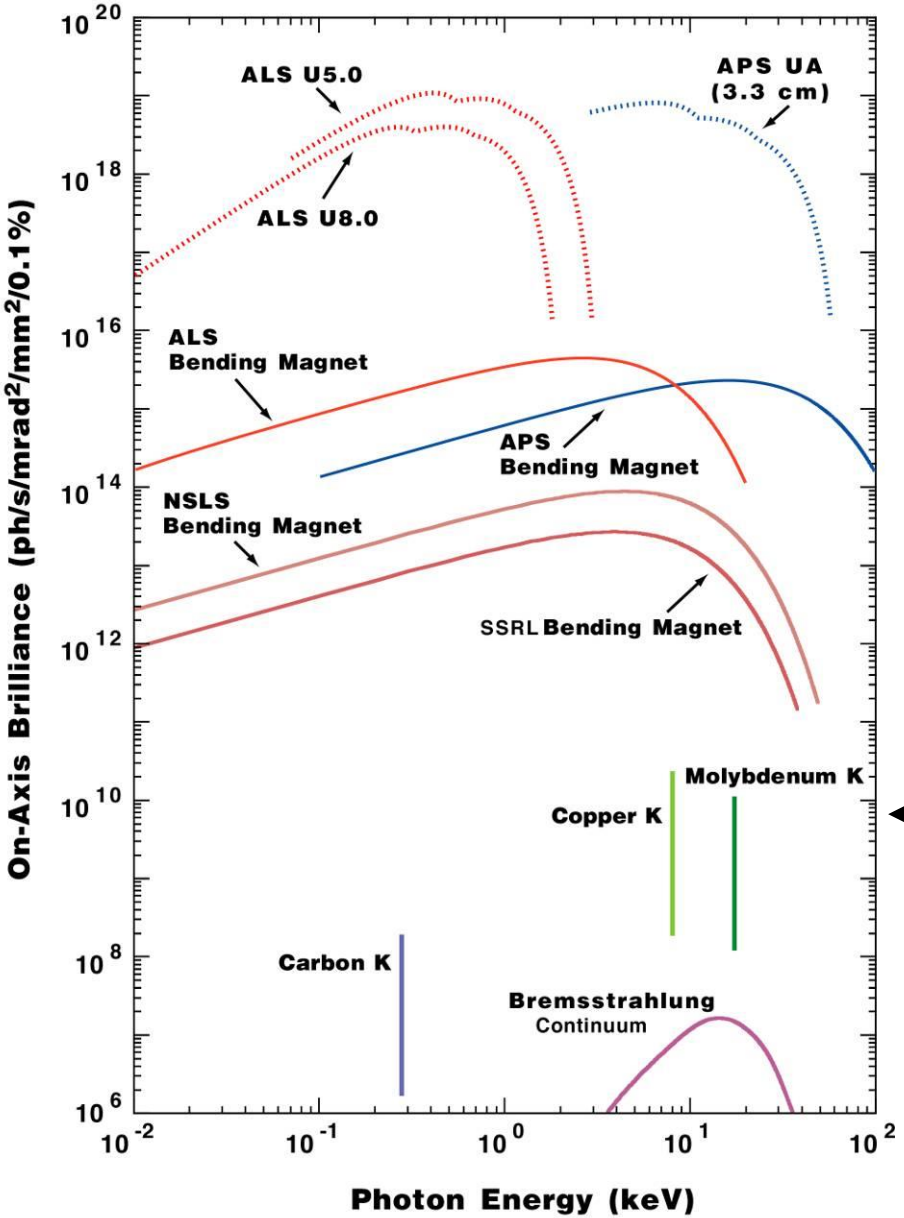
As the particles loose energy they give of electromagnetic radiation.

The particles are steered by magnetic fields.

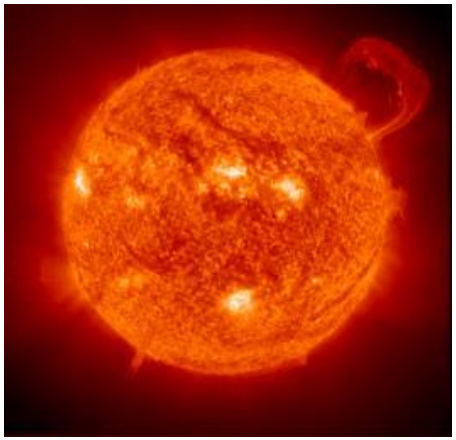
Electromagnetic radiation (photons) is not affected by these fields and is emitted at the tangent to the change in direction.

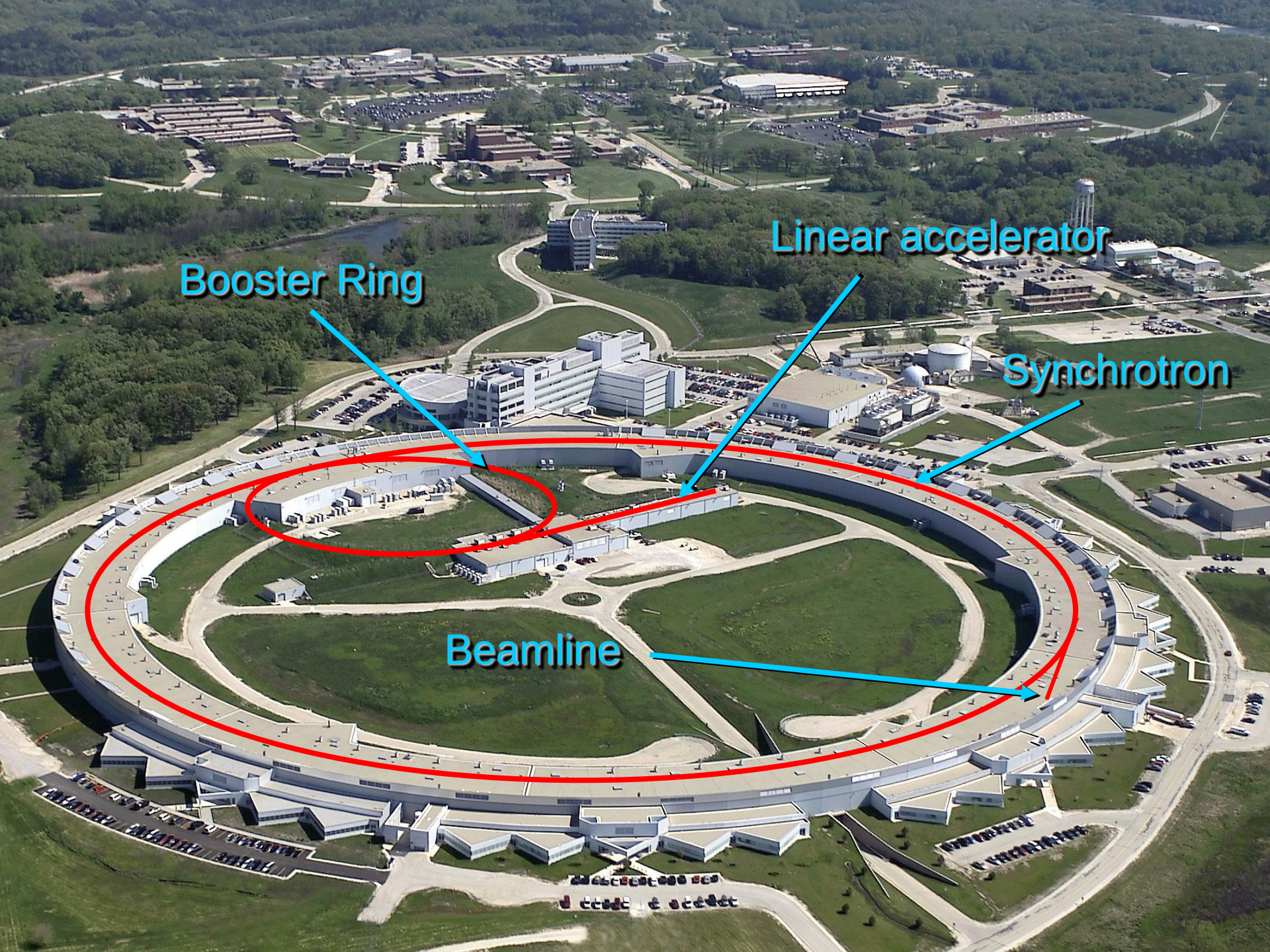
Insertion devices (undulators and wigglers) ‘amplify’ this radiation





Synchrotron radiation is 10^9 times
More brilliant than the sun
and about 100 million miles closer




Linear accelerator
Booster Ring
Synchrotron
Beamline

The PDB file

The PDB stores coordinates, experimental detail and comments and more recently the processed diffraction image data used to generate the coordinates.

```
CRYST1   77.453   77.453   37.183  90.00  90.00  90.00 P 43 21 2     8
ORIGX1      1.000000  0.000000  0.000000        0.00000
ORIGX2      0.000000  1.000000  0.000000        0.00000
ORIGX3      0.000000  0.000000  1.000000        0.00000
SCALE1      0.012911  0.000000  0.000000        0.00000
SCALE2      0.000000  0.012911  0.000000        0.00000
SCALE3      0.000000  0.000000  0.026894        0.00000
ATOM      1  N   LYS A   1      29.393  40.729   0.892  1.00 15.98           N
ATOM      2  CA  LYS A   1      28.951  39.778  -0.166  1.00 15.98           C
ATOM      3  C   LYS A   1      27.443  39.827  -0.378  1.00 15.87           C
ATOM      4  O   LYS A   1      26.684  39.780   0.587  1.00 14.49           O
ATOM      5  CB  LYS A   1      29.349  38.350   0.233  1.00 17.08           C
ATOM      6  CG  LYS A   1      28.843  37.250  -0.704  1.00 20.49           C
ATOM      7  CD  LYS A   1      29.418  35.906  -0.283  1.00 22.21           C
ATOM      8  CE  LYS A   1      28.758  34.735  -0.990  1.00 22.43           C
ATOM      9  NZ  LYS A   1      29.208  34.569  -2.386  1.00 25.22           N
ATOM     10  N   VAL A   2      27.014  39.943  -1.635  1.00 15.27           N
ATOM     11  CA  VAL A   2      25.584  39.924  -1.941  1.00 14.91           C
ATOM     12  C   VAL A   2      25.281  38.513  -2.439  1.00 14.86           C
ATOM     13  O   VAL A   2      25.698  38.133  -3.542  1.00 15.04           O
ATOM     14  CB  VAL A   2      25.192  40.921  -3.057  1.00 15.23           C
ATOM     15  CG1 VAL A   2      23.695  40.820  -3.334  1.00 16.00           C
ATOM     16  CG2 VAL A   2      25.552  42.337  -2.640  1.00 15.40           C
ATOM     17  N   PHE A   3      24.589  37.736  -1.603  1.00 14.21           N
ATOM     18  CA  PHE A   3      24.203  36.353  -1.915  1.00 12.95           C
ATOM     19  C   PHE A   3      23.063  36.267  -2.923  1.00 13.95           C
ATOM     20  O   PHE A   3      22.212  37.154  -2.989  1.00 12.93           O
ATOM     21  CB  PHE A   3      23.688  35.632  -0.659  1.00 13.15           C
ATOM     22  CG  PHE A   3      24.750  34.999   0.195  1.00 11.18           C
ATOM     23  CD1 PHE A   3      25.551  35.764   1.046  1.00 12.22           C
ATOM     24  CD2 PHE A   3      24.905  33.620   0.193  1.00 10.60           C
ATOM     25  CE1 PHE A   3      26.480  35.161   1.885  1.00 12.38           C
ATOM     26  CE2 PHE A   3      25.831  33.000   1.026  1.00 11.09           C
ATOM     27  CZ  PHE A   3      26.626  33.760   1.879  1.00 12.19           C
```

Atom ID

Residue name

Residue number

B factor

Comment

```
ATOM     58  N   LEU A   8      21.906  31.721  -0.789  1.00 11.49          N
ATOM     59  CA  LEU A   8      21.726  31.982   0.631  1.00 10.44          C
ATOM     60  C   LEU A   8      20.912  30.860   1.284  1.00 10.51          C
ATOM     61  O   LEU A   8      21.191  30.474   2.422  1.00 10.37          O
ATOM     62  CB  LEU A   8      21.066  33.358   0.859  1.00 11.15          C
ATOM     63  CG  LEU A   8      20.854  33.696   2.336  1.00 10.23          C
ATOM     64  CD1 LEU A   8      22.195  33.727   3.068  1.00 10.73          C
ATOM     65  CD2 LEU A   8      20.183  35.053   2.459  1.00 12.22          C
```

Atom type

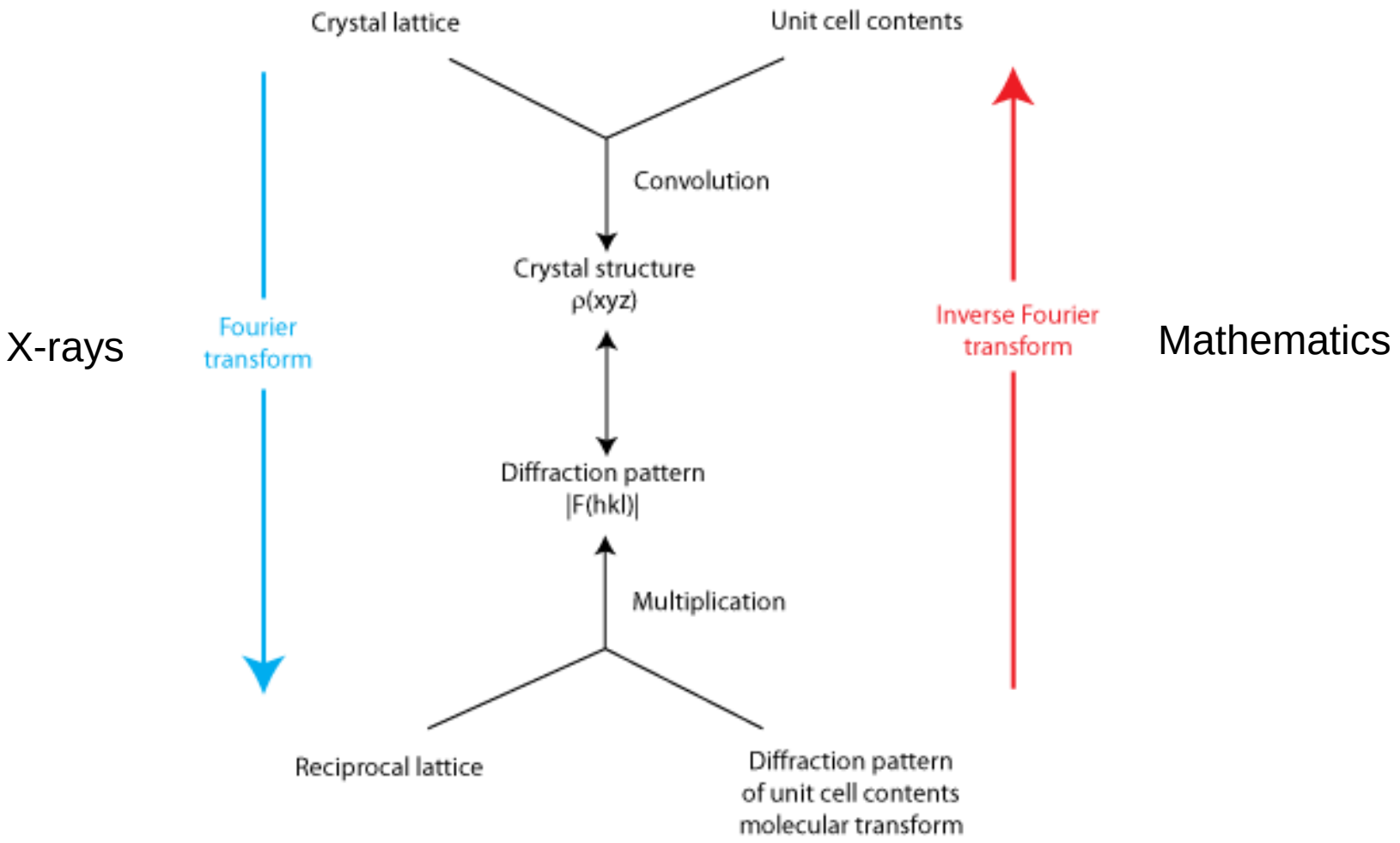
Chain

Occupancy

Numerical sequence of atom

x,y and z coordinates in Angstroms

Crystal – X-ray data - Model


Crystal lattice
Unit cell contents
Convolution
Crystal structure
ρ(xyz)
X-rays
Fourier
transform
Inverse Fourier
transform
Mathematics
Diffraction pattern
|F(hkl)|
Multiplication
Reciprocal lattice
Diffraction pattern
of unit cell contents
molecular transform

The Protein Data Bank

- The Protein Data Bank contains depositions for 133,589 biological macromolecules.
- Some 119,652 of those are from data derived by X-ray crystallography.
- Simple validation tests are available but a deposition can still be accepted even if a test is failed.
- How accurate are the ‘structures’ in the PDB?

More importantly ….

- The 133,589 biological macromolecules have been built up since 1976.
- About 10,000 structures are now being deposited per year.
- In each year, the PDB is accessed by over 1,000,000 unique users.
- The majority of those users will be those who make use of the models but are not trained to look at the data.
- How accurate are the ‘structures’ in the PDB?

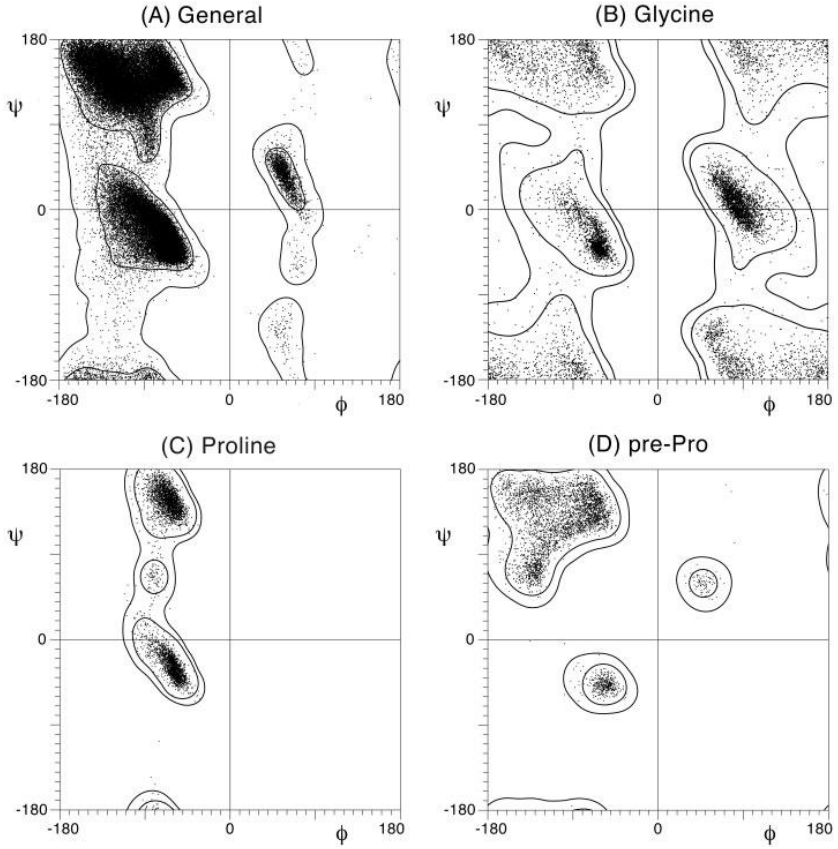
Can we identify errors in the models, if any?

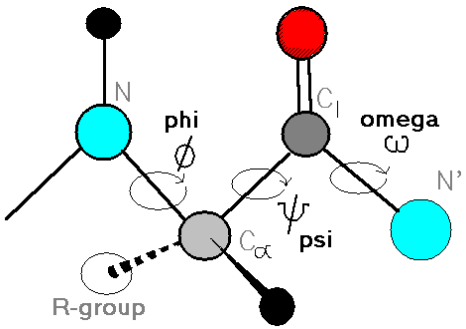
- Residues have well defined geometries.
- Sequence information is well known.
- Potential problems are:
  - Structural perturbation due to radiation damage
  - Incorrect ligand identification
  - Missing ligands
  - Just generally bad refinement
  - Crystallographic oligomer

A ‘structure’

- A structure is a model that best represents the measured data.
- Think about what you are measuring:
  - The data is an average taken over many macromolecules. For example, a 100 μm^3 crystal produced from a macromolecule that has a typical size of 200 Å on edge will consist of ~ 5,000 molecules on edge or 125,000,000,000 molecules in total.
  - The data is not static, it represents an average of those molecules over time.
  - The data is dynamic. X-rays cause chemical changes which can also be captured over time.

Known knowns – we know what to expect

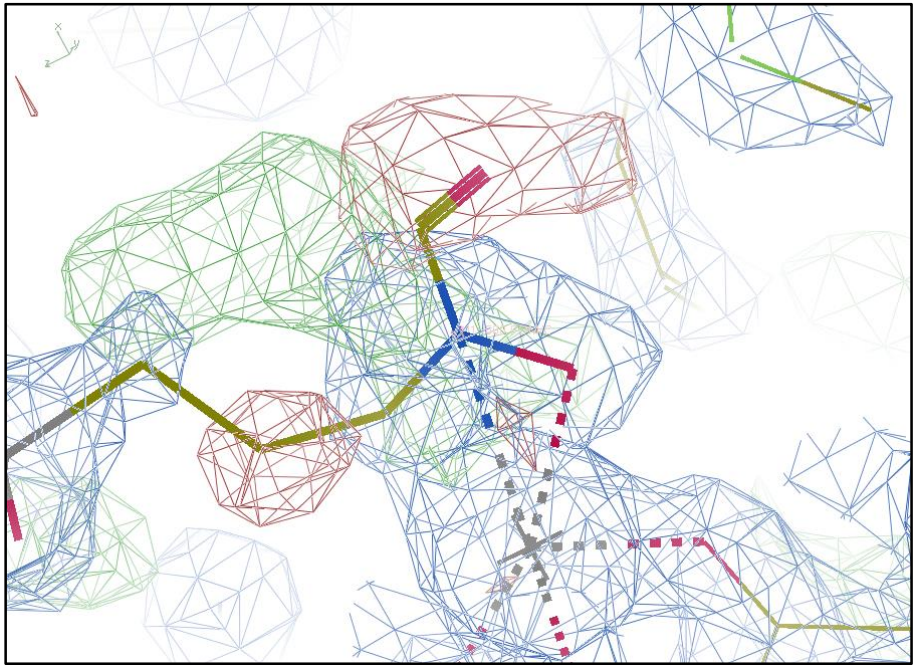


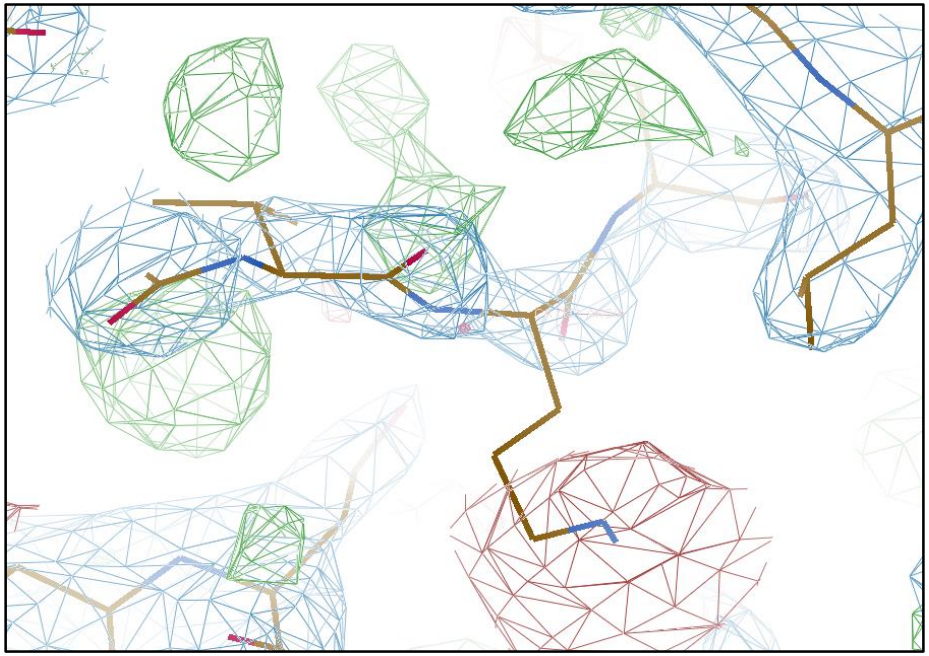


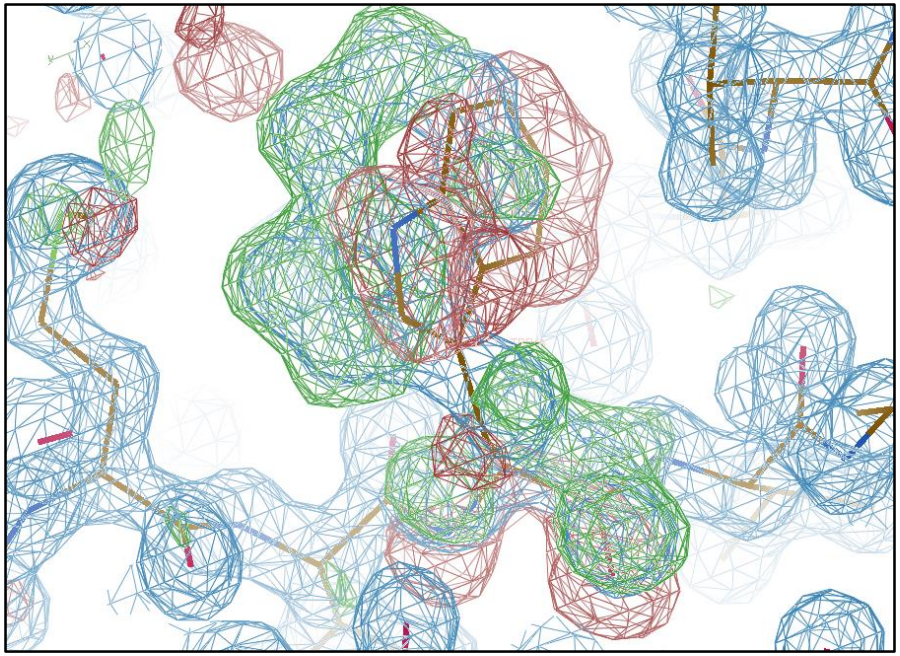
Structure Validation by Cα Geometry: φ, ψ and Cβ Deviation PROTEINS: Structure, Function, and Genetics 50:437–450 (2003)

Simon C. Lovell, Ian W. Davis, W. Bryan Arendall III, Paul I. W. de Bakker, J. Michael Word, Michael G. Prisant, Jane S. Richardson, and David C. Richardson

The dihedral angles in the main chain have allowed and disallowed regions that are well known – developed by Gopalasamudram Narayana Ramachandran and called the Ramachandren plot. Available as part of several software packages.







Error propagation …

- More common than you may think
- The examples presented are in the PDB and all come from well respected structural biologists
- Despite care and diligence, errors still get through
- There are serious problems in many models yet the non-crystallographic community for the most part use these models as 'structures' on the assumption that the model accurately represents the structure

How can we over come these problems?

- Structural perturbation due to radiation damage
  - Radiation damage studies, knowledge of the chemical processes and signatures
- Incorrect ligand/Metal identification
  - Better ligand treatment during refinement
  - Careful analysis of the crystallization conditions
  - Analysis of the sample pre or post crystallization
- Missing ligands
  - Similar approaches to the above
- Just generally bad refinement
  - To paraphrase Bernard Rupp, sometimes is worthwhile to look at the map!
- Crystallographic oligomer
  - Solution scattering

How can we over come these problems?

- Structural perturbation due to radiation damage
  - Radiation damage studies, knowledge of the chemical processes and signatures
- Incorrect ligand identification
  - Better ligand treatment during refinement
  - Careful analysis of the crystallization conditions
  - Analysis of the sample pre or post crystallization
- Missing ligands
  - Similar approaches to the above
- Just generally bad refinement
  - To paraphrase Bernard Rupp, sometimes is worthwhile to look at the map!
- Crystallographic oligomer
  - Solution scattering

Metals in proteins

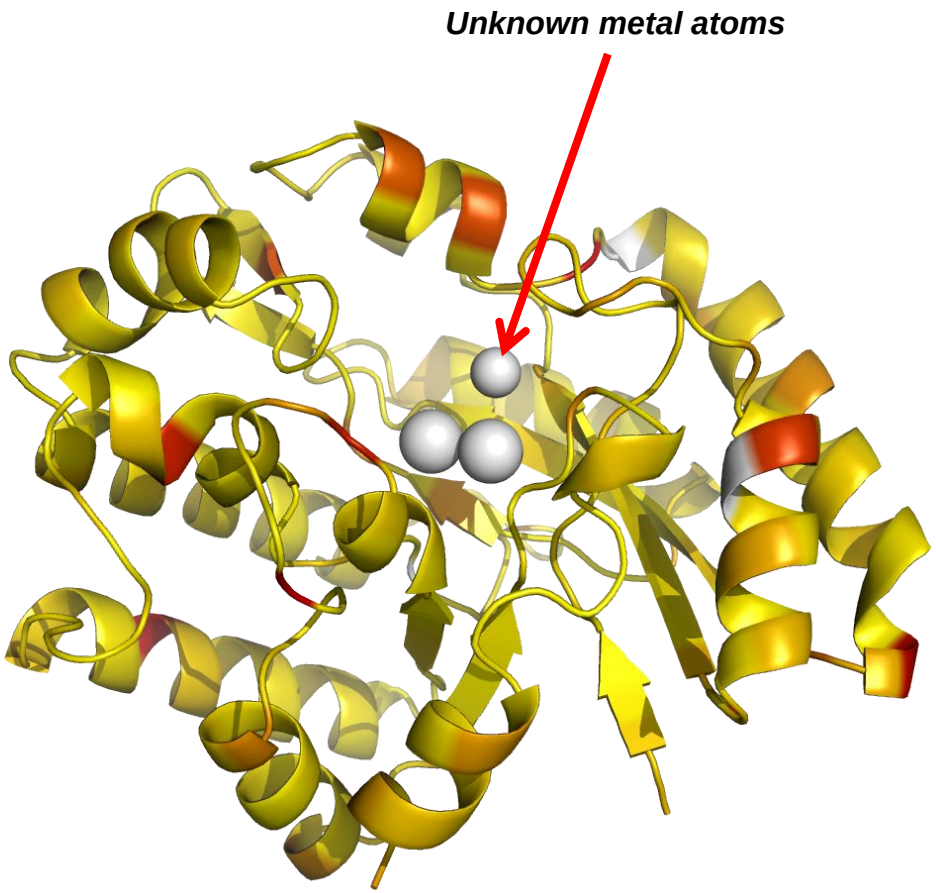
- Many proteins contain small numbers of metal atoms (estimated to be at least 30%)
  - Binding and transport of metals
  - A single metal atom helps to determine the folded shape of the molecule

X-ray crystallography measures electron density

Cannot determine Z of metal atoms

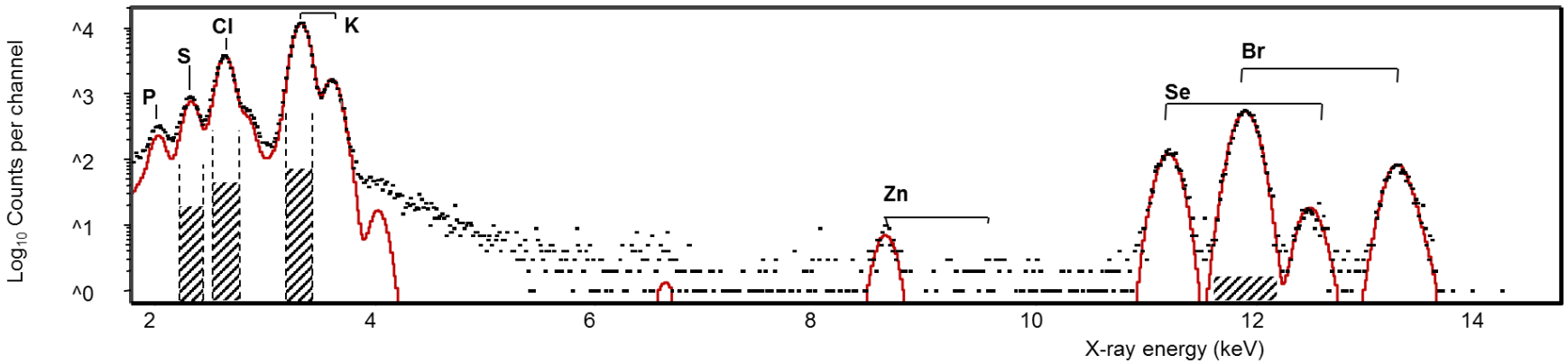
Z is often inferred indirectly from molecular modelling

Excitation scans should be used



Particle induced X-ray emission

The energy of an X-ray emitted when an atomic electron undergoes an energy transition between its shell and a vacant electron site in a lower energy shell (e.g. for an M to L shell transition, sulphur gives a 2.3 keV X-ray) gives an unambiguous identification of atoms.



Emission of the characteristic X-rays from a sample can be induced by an incident beam of high energy protons (Particle Induced X-ray Emission: PIXE).

PIXE analysis of proteins

- Concentrations are typically 1 atom per molecule of 10 – 100 kDa (10s to 100s ppm)
- Available sample size is small (microlitres of solution)
- MicroPIXE is ideal for identifying and quantifying unknown metal atoms in proteins
- Well developed technique.

Structure 1999

Ways & Means R291

Leaving no element of doubt: analysis of proteins using microPIXE

Elspeth Garman

Address: Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK.

E-mail: elspeth@biop.ox.ac.uk

Structure December 1999, 7:R291–R299

number. In contrast, for ionisation of atoms with electrons (EPMA) and protons (PIXE), the cross-sections for interactions which subsequently cause X-ray production are high for elements with low atomic numbers but decrease with increasing atomic number. Thus for identifying

PIXE conference 2004

UniS

University of Surrey

PIXE analysis of single atoms: measuring metal atom concentrations in protein molecules

G.W. Grime[(a)], I. Gomez-Morilla[(a)], D. Yates[(b)], E.F. Garman[(b)]

(a)University of Surrey, Department of Physics, Guildford U.K.
(b)University of Oxford, Department of Biochemistry, South Parks Road, Oxford, U.K.

PBMB 2005

ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Progress in Biophysics and Molecular Biology 89 (2005) 173–205

Progress in Biophysics & Molecular Biology

www.elsevier.com/locate/pbiomolbio

Review

Elemental analysis of proteins by microPIXE

Elspeth F. Garman[a,*], Geoffrey W. Grime[b]

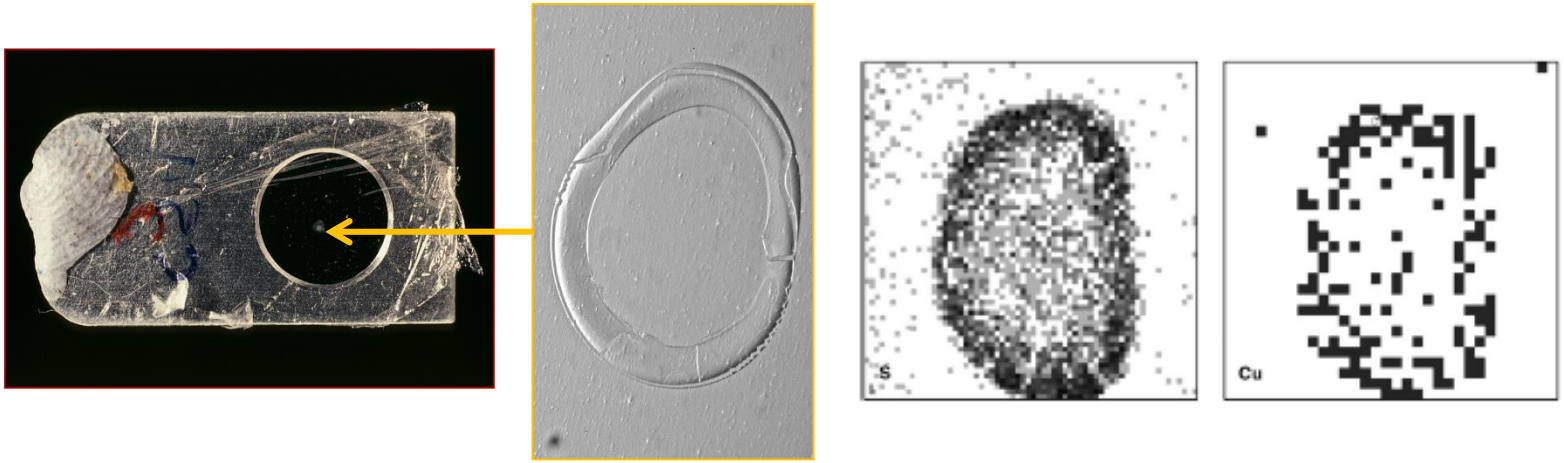
[a]*Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK*
[b]*Department of Physics, University of Surrey, Guildford GU2 7XH, Surrey, UK*

Available online 5 November 2004

However there are issues

- Samples are prepared by manual pipetting onto foils
- Samples are analysed by manual positioning on the PIXE maps to locate the precipitated protein. Differential precipitation of buffer may require accurate positioning using elemental maps of sulphur.
- Spectra are processed manually

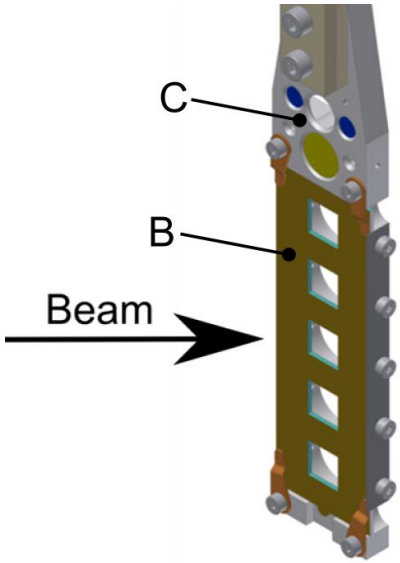
It is difficult to analyse more than 10 samples in a run day



An atomic technique – can be applied to samples that are biologically ‘past their sell by date’

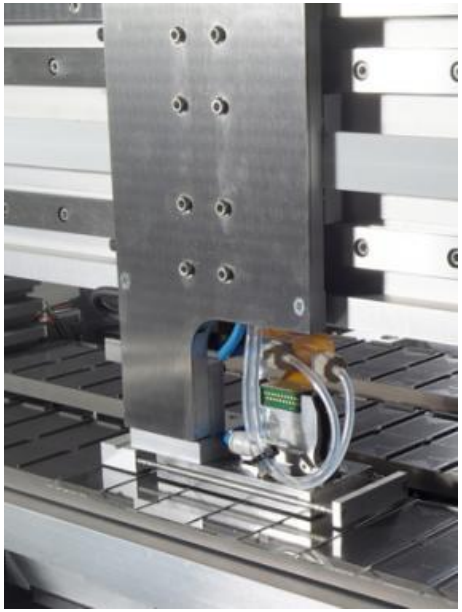
Can we apply it in high-throughput?

Sample preparation: support film and printing



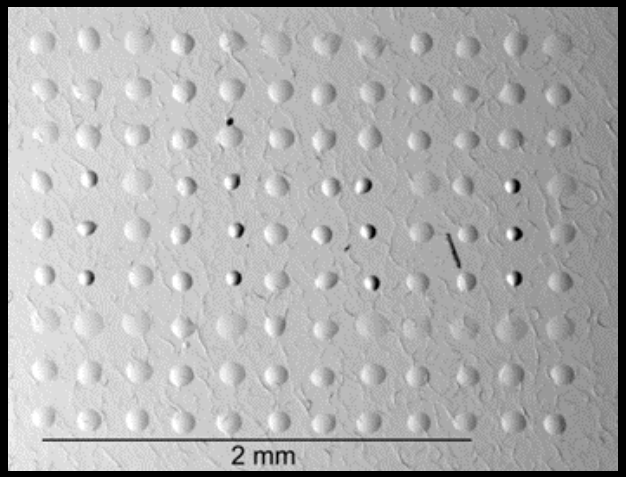
- Sample holders have the same dimensions as a standard microcope slide and are adapted for compatibility with both printer and sample stage
- Five 8 × 8 mm sample windows per slide covered in polypropylene film using a specially developed coating machine and non-instant contact adhesive

- Samples supplied as solutions in well plates
- Printed by a non-contact ArrayJet microarrayer
- Up to 144 samples per 12 × 12 array, 5 arrays per slide



High-throughput Sample Preparation

Dispense samples with a non-contact microarray printer

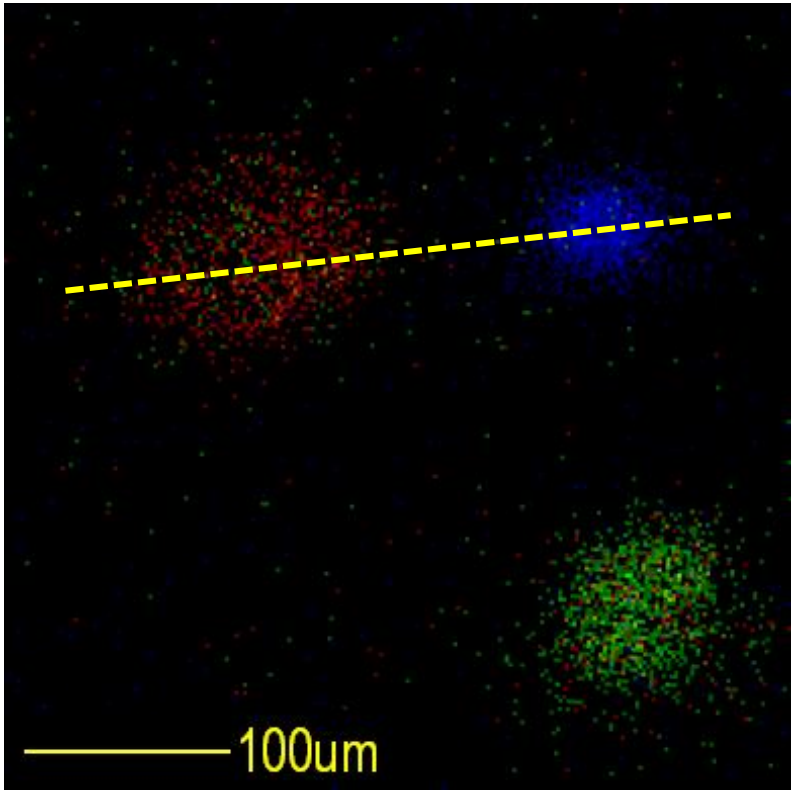


Up to 144 samples dispensed into a 384 well plate and printed into a 12x12 array of 60 um drops with 200 um spacing.

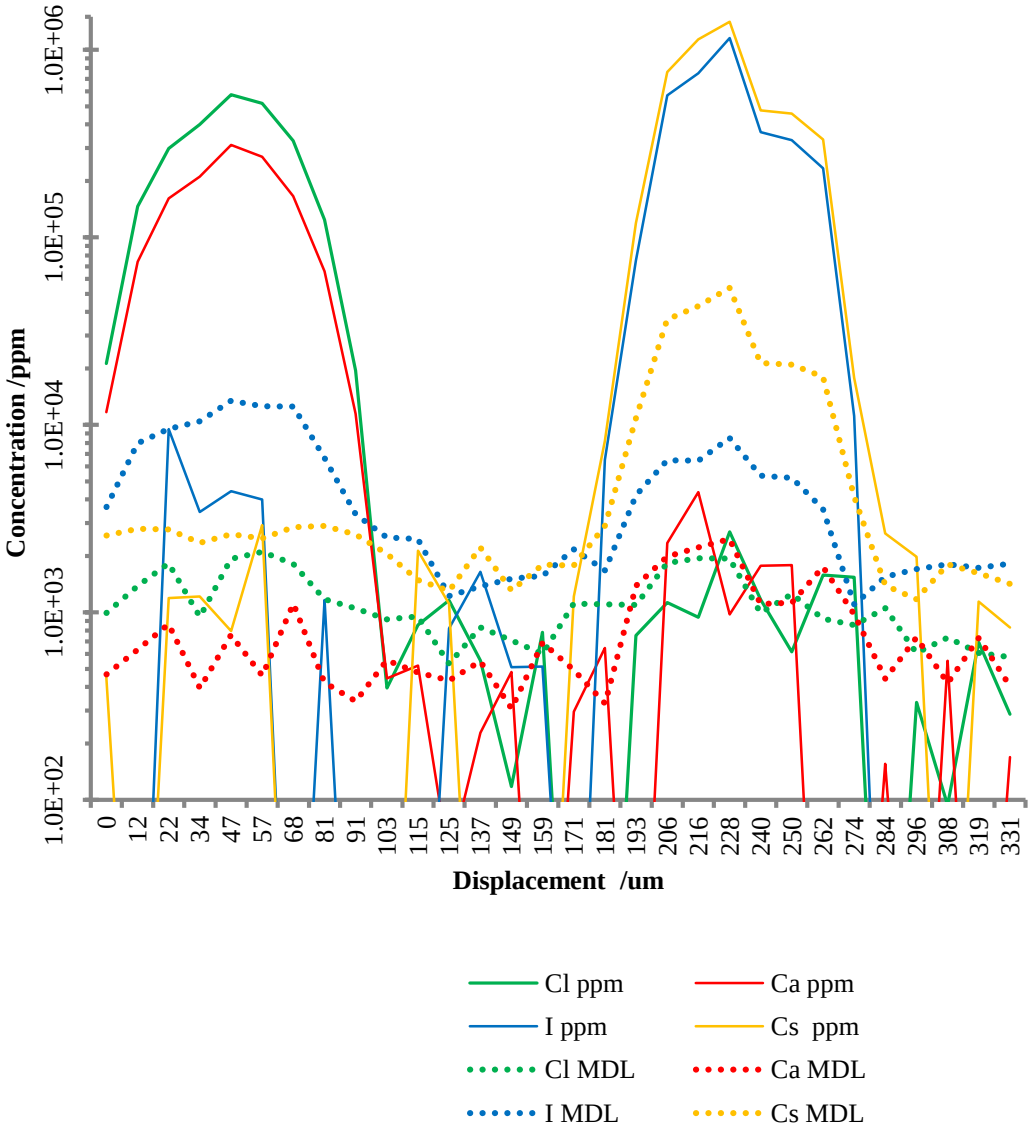
Up to five arrays can be mounted into a single sample holder giving a total of 720 samples per slide.

Cross-talk?

- Array of standard salt compounds.
- Verify that there is no cross talk between adjacent spots.

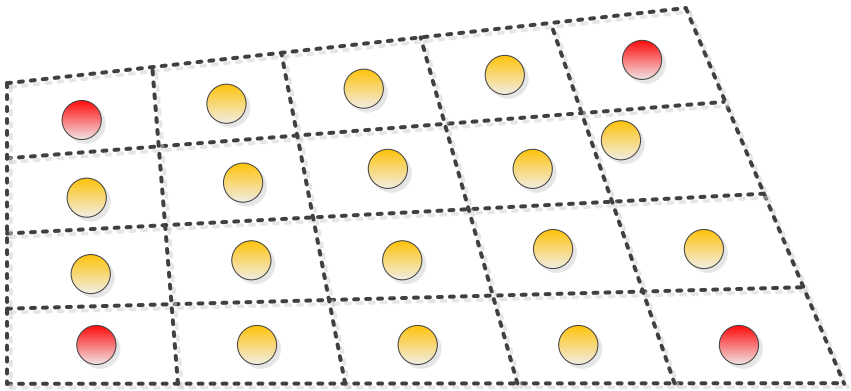


$CaCl_2$ CsI

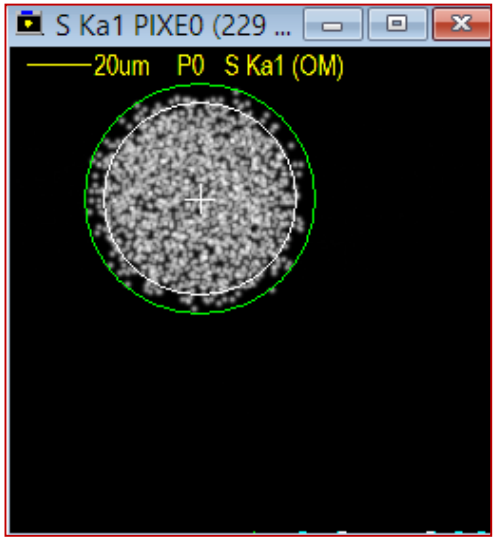


Finding the spots

1. Print '**Landing Lights**' at the corners of the array. *Spots of metal salt (e.g. KBr) which are easy to find with PIXE.*

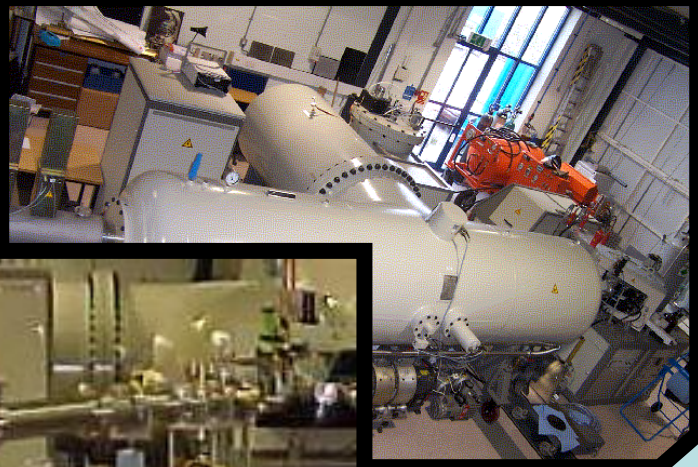


2. Move the stage to each corner (operator control) and use a least-squares fitting routine to find the centre of the spot from the PIXE map. **(This is the only manual operation for each array)**



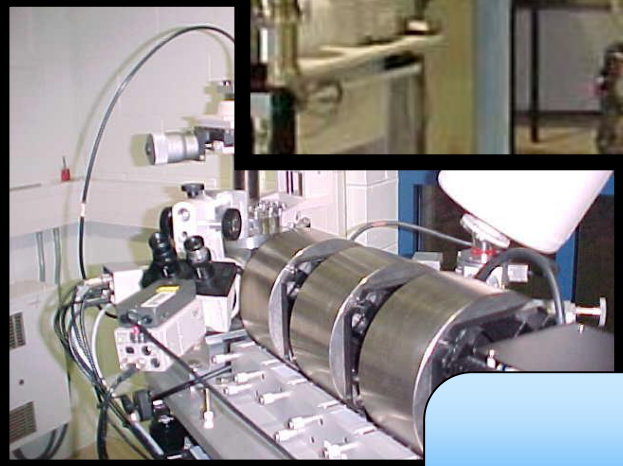
3. Store the stage coordinates of the corners

4. Interpolate the stage coordinates of each cell in the array. This corrects for linear geometric distortions,

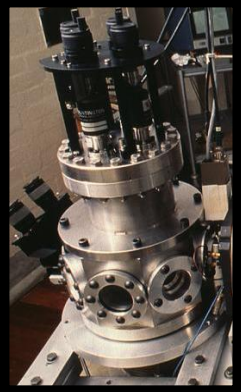


Source & Accelerator

Scanning System

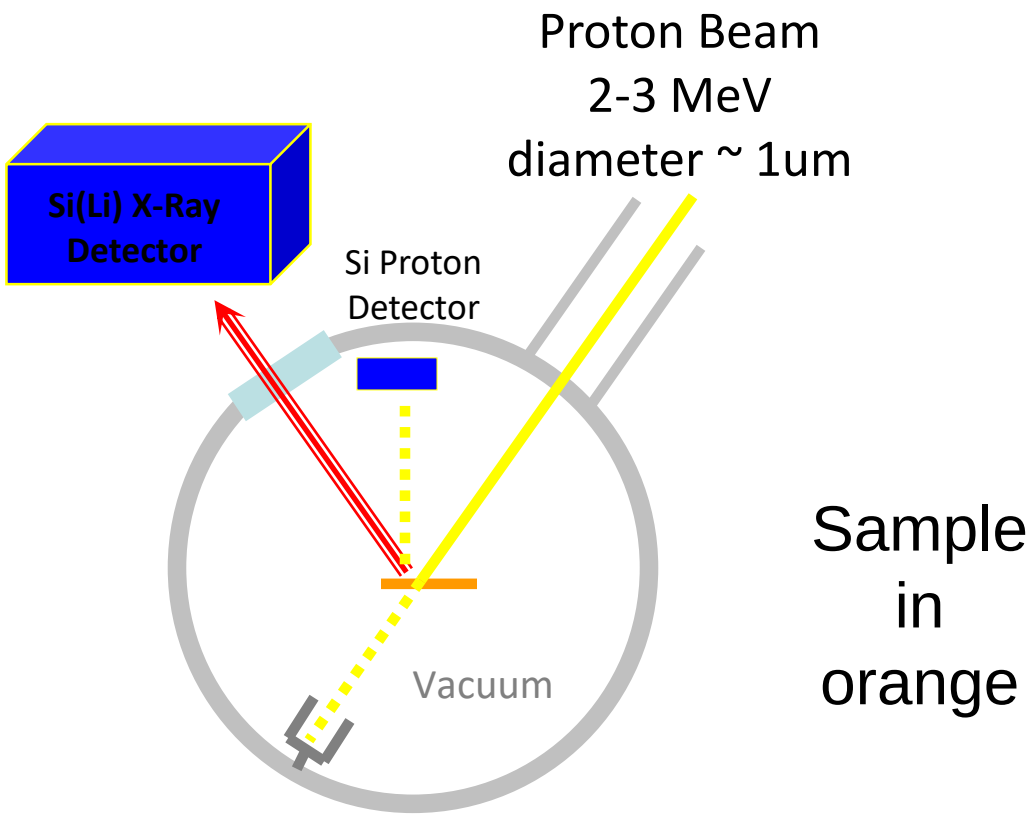


Focusing System



~1 µm diameter beam on target

Scanning Proton Microprobe for PIXE analysis. 2-3 MeV protons emerge from the van de Graaff accelerator and are focussed by high precision magnets onto the sample. The whole beamline is kept under vacuum.



Why would we want to do this?

Measuring the metal content

- MicroPIXE can be used to determine the proportion of methionine substitution where no sulfur is present in the buffer.
- The concentration of an element is determined by fitting the area of the X-ray peak corresponding to the element.
- If the total number of Se atoms per protein molecule is α_{Se}, the total number of S atoms left per protein molecule is α_S, and the original number of S atoms (cysteine + methionine) in the sequence was α then $\alpha = \alpha_S + \alpha_{Se}$ and we can write:

$$\frac{\alpha_S}{\alpha_{Se}} = \frac{c_S}{c_{Se}} \frac{A_{Se}}{A_S} \frac{(\alpha - \alpha_{Se})}{\alpha_{Se}}$$

- Where A_S and A_{Se} are the atomic masses of S and Se respectively and c_S and c_{Se} are the mass concentrations.

High-Throughput PIXE

- In our case the buffer for all the samples contained Sulfur so we could not use Cys and Met as a calibration standard.
- All the proteins studied were expressed with SeMet for phasing purposes.
- The number of atoms of element Z per protein can be determined from this by

$$\alpha_Z = \frac{c_Z}{c_{Se}} \frac{A_{Se}}{A_Z} \alpha_{Se}$$

- Where A_Z and A_{Se} are the atomic masses of element Z and Se respectively and c_Z and c_{Se} are the mass concentrations determined from the PIXE spectrum.

- We do make the assumption of full Se incorporation but because we already have structural data, we can confirm this assumption.

The initial experiment

- 34 metalloprotein samples chosen from a set of samples successfully crystallized in the High-Throughput Crystallization Screening Center.
- All were SeMet samples.
- All produced crystals and a structural model.
- PIXE analysis was carried out on each sample.

- The samples used were split into four groups based on PIXE analysis
  - Those where the PDB was inconsistent with the PIXE data
  - Those where extra metals were seen in the PIXE data (but not present in the PDB)
  - Those that were consistent with the PIXE data.
  - Those that produced no signal.

Table of results

	PDB ID	Gene	Residues	Metal in PDB	Metals in PIXE (>3xLOD)	Potential metals in PIXE (1-3xLOD)
PIXE data consistent with PDB						
1		BfR258E	168	Ca	Ca (1.7)	Fe
2		BcR147A	103	Ca	Ca (0.8)	
3		HR4604D	100	Zn	Zn (2.5), Fe (0..3)	Ca, Co, Cu
4		OR3	114	Zn	Ca, Zn*	Fe, Ni*
5		LkR105	290	-	Fe (0.04)	Ca, Cu
6		MjR117B	80	-	Ca (0.2)	Fe
7		EwR179	129	-	-	Ca, Fe
8		MjR118E	105	Na◆	-	-
9		VfR176	149	Na◆	-	Co
10		SyR86	212	Na◆	-	Fe
11		SyR101A	100	Na◆	-	Ca, Fe, Cu
12		VcR193	255	Mg◆	-	-
13		DhR1A	147	Mg◆	-	Ca, Fe*
14		BuR114	223	Mg◆	-	Fe, Ni
Sample too dilute for PIXE (no S signal)						
1		LpR108	284	Ca	-	K, Mn
2		LpR109	232	Mg/Na◆	-	-

*S signal was below 3 times the limit of detection, so accurate stoichiometries could not be established.
◆Presence of sodium and magnesium could not be confirmed at the proton energies used in these experiments.

Table of results

	PDB ID	Gene	Residues	Metal in PDB	Metals in PIXE (>3xLOD)	Potential metals in PIXE (1-3xLOD)
PDB inconsistent with PIXE						
1		BiR14	456	Ca	-	Ca, Mn
2		NsR437I	106	Mn	-	-
3		SnR135D	161	Ca	-	Ca
4		LmR141	283	Fe/Zn	Ca (3.3), Mn (0.5), Fe (1.2), Co (1.2)	Zn
5		NsR236	119	K	-	Ca
6		NsR437H	141	Mn	-	Fe, Co
7		SoR237	137	Na	Co (0.7), Zn (0.7)	Fe, Ni
8		BtR324A	169	Zn	-	Ca, Mn, Fe*
9		GR157	262	Zn	-	Co

	PDB ID	Gene	Residues	Metal in PDB	Metals in PIXE (>3xLOD)	Potential metals in PIXE (1-3xLOD)
Extra metals present in PIXE						
1		MuR16	210	Fe/Zn	Fe (0.6), Co (0.9), Ni (0.4), Zn (0.7)	-
2		MqR88	420	Na◆	Ca (7.1)	Fe
3		SR677	222	Mg◆	Ca (0.7), Fe (0.05)	K/Br
4		DrR130	296	Mg◆	Ca*	-
5		BtR319D	172	Mg◆	Ca (1.74)	-
6		ShR87	320	Mg◆	Mn (0.8), Fe (0.7)	-
7		SmR83	218	Mg◆	Ca (0.5), Fe (0.1)	Ti, Co, Cu
8		NsR141	225	Mg◆	Mn (0.2), Fe (0.4), Ni (0.4)	Co
9		ZR319	289	Mg◆	-	Ca, Fe, Cu

- More than half of the proteins analysed were inconsistent with their entry in the PDB!
- This highlights a deep problem in identifying metal constituents of proteins.

- Of the 34 samples analyzed, 9 were inconsistent with the PDB results, 9 had extra metals present, 18 were consistent, and 2 were unsuitable for analysis due to low protein concentration on the sample.
- In total, 18 of the 32 analyzable samples (56%) were not correctly or fully described in the PDB deposition.

	PDB ID	Gene	Residues	Metal in PDB	Metals in PIXE (>3xLOD)	Potential metals in PIXE (1-3xLOD)	Crystallization conditions
PDB inconsistent with PIXE							
1		BiR14	456	Ca	-	Ca, Mn	18% PEG 3350, 0.2M Ca acetate, 0.1M MES, pH 6.15
2		NsR437I	106	Mn	-	-	20% PEG 4000, 0.1M Mn chloride, 0.1M MES, pH 6.0
3		SnR135 D	161	Ca	-	Ca	20% PEG 8000, 0.1M Ca acetate, 0.1M MES, pH 6.0
4		Protein A	283	Fe/Zn	Ca (3.3), Mn (0.5), Fe (1.2), Co (1.2)	Zn	15% PEG 8000, 0.17 M sodium acetate, 0.01 M L-cysteine, 0.1 M MES pH 6.2
5		NsR236	119	K	-	Ca	8.64 M K acetate, 0.1 M TAPS, pH 9.0
6		NsR437 H	141	Mn	-	Fe, Co	20% PEG 1000, 0.1M Mn chloride, 0.1M MES, pH 6.0
7		SoR237	137	Na	Co (0.7), Zn (0.7)	Fe, Ni	NaCl 200 mM, MES PH6, PEG 3350 20%, pH 6.15
8		BtR324A	169	Zn	-	Ca, Mn, Fe*	0.75M Mg Formate, 0.1M Bis-Tris, pH 7.0
9		GR157	262	Zn	-	Co	100 mM Na Acetate (pH 4.6), 30% MPD, and 200 mM NaCl.

Model in the PDB containing a metal from the crystallization cocktail and not protein

Model in the PDB containing an incorrect metal

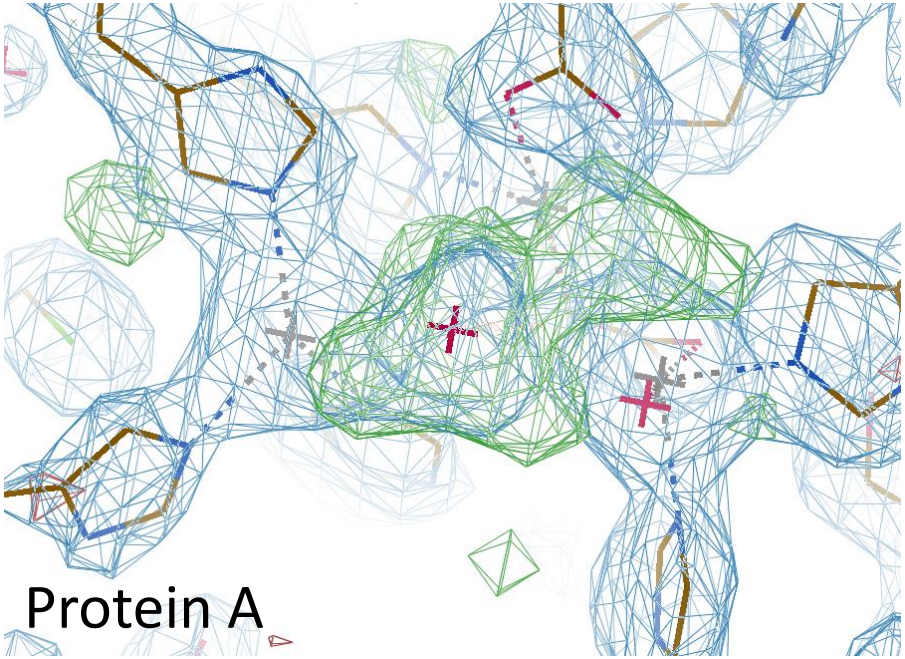
	PDB ID	Gene	Residues	Metal in PDB	Metals in PIXE (>3xLOD)	Potential metals in PIXE (1-3xLOD)	Crystallization conditions
PDB inconsistent with PIXE							
1		BiR14	456	**Ca**	-	Ca, Mn	18% PEG 3350, 0.2M **Ca** acetate, 0.1M MES, pH 6.15
2		NsR437I	106	**Mn**	-	-	20% PEG 4000, 0.1M **Mn** chloride, 0.1M MES, pH 6.0
3		SnR135 D	161	**Ca**	-	Ca	20% PEG 8000, 0.1M **Ca** acetate, 0.1M MES, pH 6.0
4		Protein A	283	Fe/Zn	Ca (3.3), Mn (0.5), Fe (1.2), Co (1.2)	Zn	15% PEG 8000, 0.17 M sodium acetate, 0.01 M L-cysteine, 0.1 M MES pH 6.2
5		NsR236	119	**K**	-	Ca	8.64 M **K** acetate, 0.1 M TAPS, pH 9.0
6		NsR437 H	141	**Mn**	-	Fe, Co	20% PEG 1000, 0.1M **Mn** chloride, 0.1M MES, pH 6.0
7		SoR237	137	**Na**	Co (0.7), Zn (0.7)	Fe, Ni	**Na**Cl 200 mM, MES PH6, PEG 3350 20%, pH 6.15
8		BtR324A	169	Zn	-	Ca, Mn, Fe*	0.75M Mg Formate, 0.1M Bis-Tris, pH 7.0
9		GR157	262	Zn	-	Co	100 mM Na Acetate (pH 4.6), 30% MPD, and 200 mM NaCl.

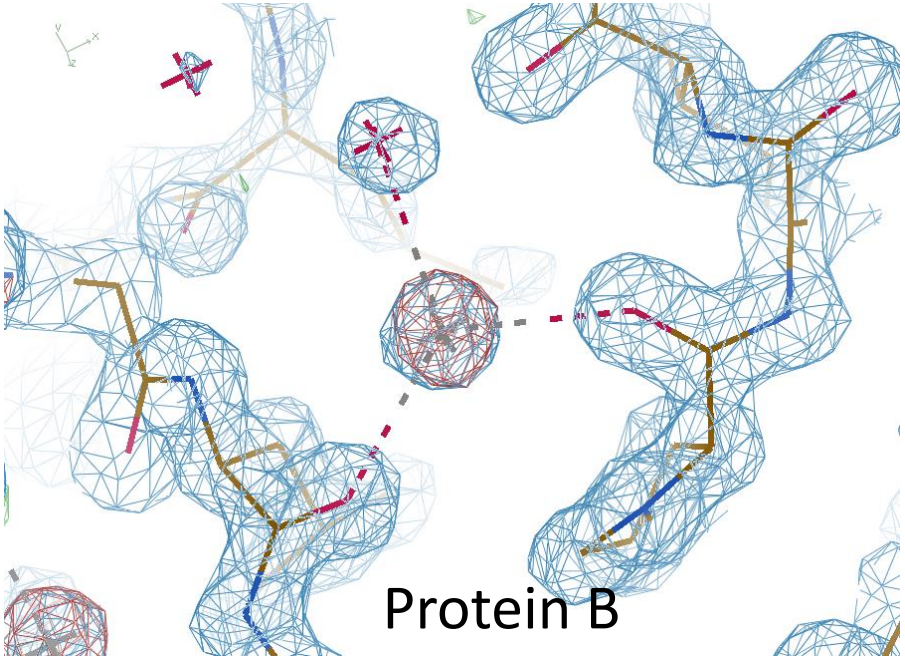
Model in the PDB containing a metal from the crystallization cocktail and not protein

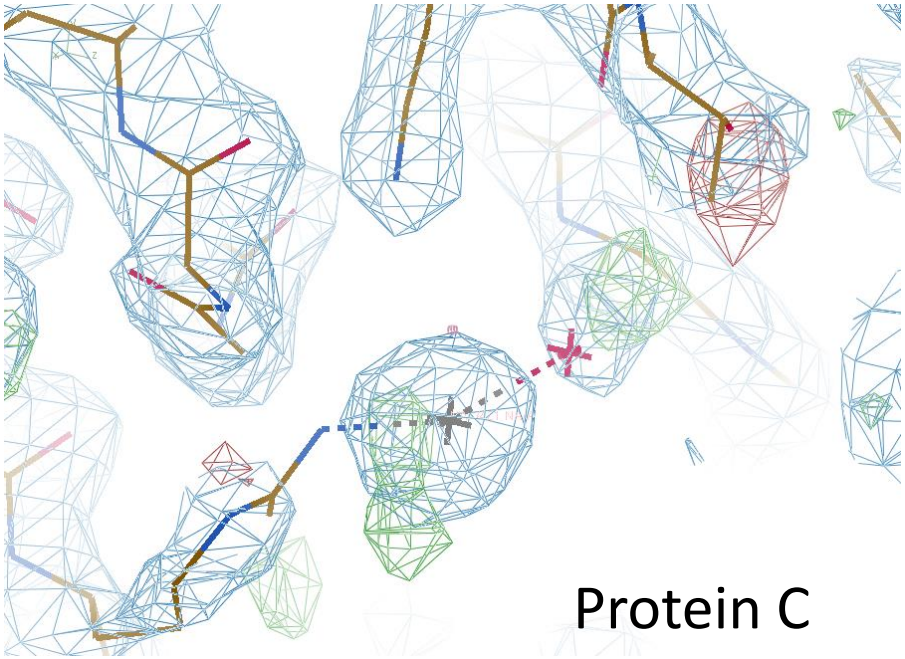
Model in the PDB containing an incorrect metal

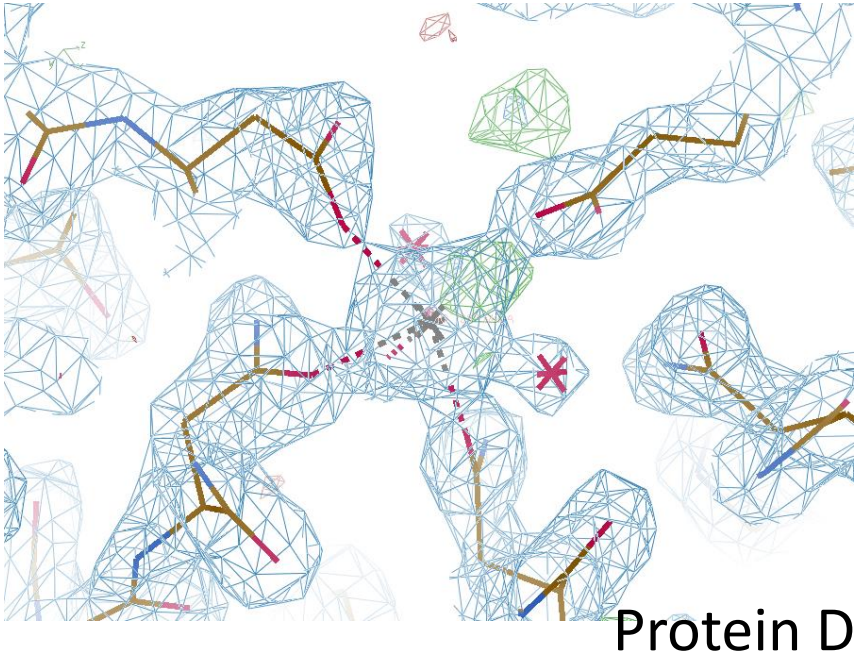
	PDB ID	Gene	Residues	Metal in PDB	Metals in PIXE (>3xLOD)	Potential metals in PIXE (1-3xLOD)	Crystallization conditions
Extra metals present in PIXE							
1	[illegible]	MuR16	210	Fe/Zn	Fe (0.6), Co (0.9), Ni (0.4), Zn (0.7)	-	0.1 M Na_2MoO_4*2H2O, 0.1 M Bis-Tris propane, 12% PEG 20000
2	[illegible]	MqR88	420	Na◆	Ca (7.1)	Fe	0.1 M Na_2MoO_4, 0.1 M Tris, pH 8.0, 20% PEG 8000
3	[illegible]	SR677	222	Mg◆	Ca (0.7), Fe (0.05)	K/Br	0.1 M KH_2PO_4, 0.1 M $NaC_2H_3O_2$, pH 5.0, 12% PEG 20000
4	[illegible]	DrR130	296	Mg◆	Ca*	-	0.1 M NaCl, 0.1 M TAPS, pH 9.0, 18% PEG 3350, $MgCl_2$
5	[illegible]	BtR319D	172	Mg◆	Ca (1.74)	-	None given
6	[illegible]	ShR87	320	Mg◆	Mn (0.8), Fe (0.7)	-	0.1 M Na citrate, pH 5.2, 1.25 M Li_2SO_4, 0.5 M $(NH_4)_2SO_4$
7	[illegible]	SmR83	218	Mg◆	Ca (0.5), Fe (0.1)	Ti, Co, Cu	0.1 M $LiCl_2$, 0.1 M Bis-Tris, pH 5.5, 18% PEG 3350
8	[illegible]	NsR141	225	Mg◆	Mn (0.2), Fe (0.4), Ni (0.4)	Co	0.1 M citric acid, pH 5.0, 1.6 M $(NH_4)_2SO_4$
9	[illegible]	ZR319	289	Mg◆	-	Ca, Fe, Cu	0.1 M Tris-HCl, pH 9.1, 18% PEG 3350, 0.1 M $MgSO_4$

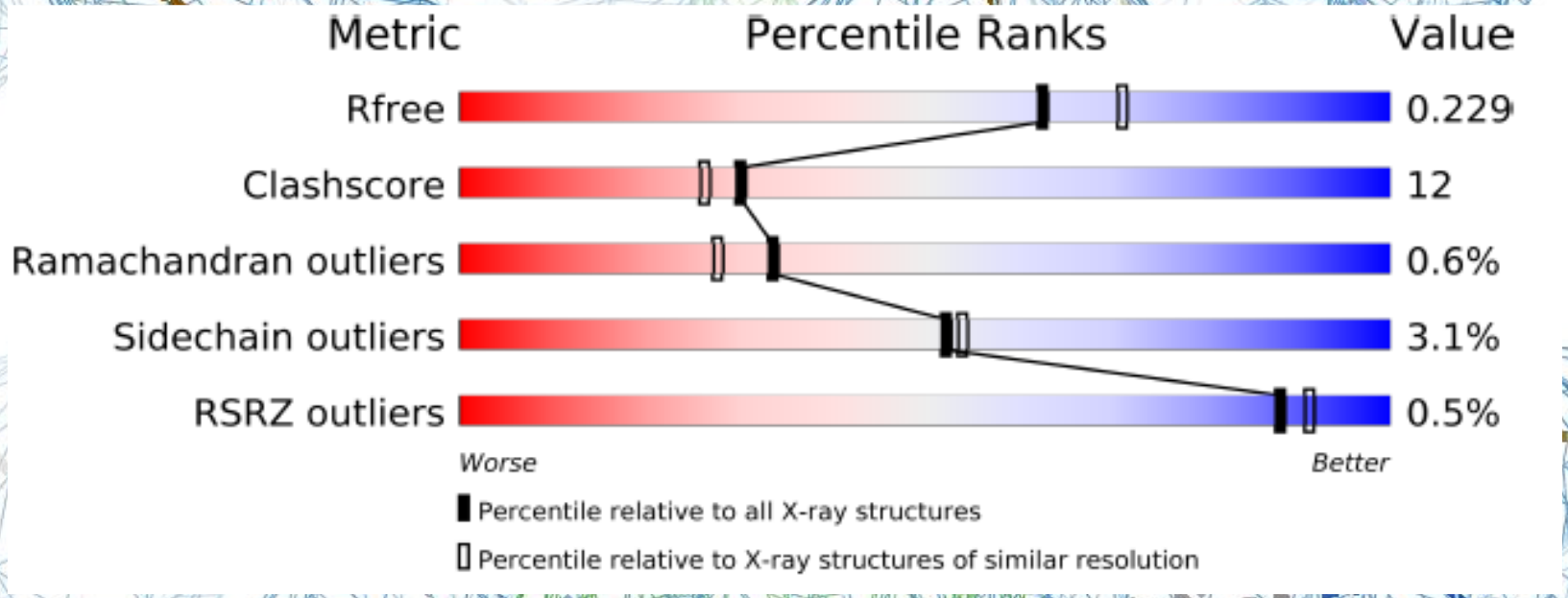
Model in the PDB containing an extra misidentified metal


Protein A


Protein B


Protein C


Protein D



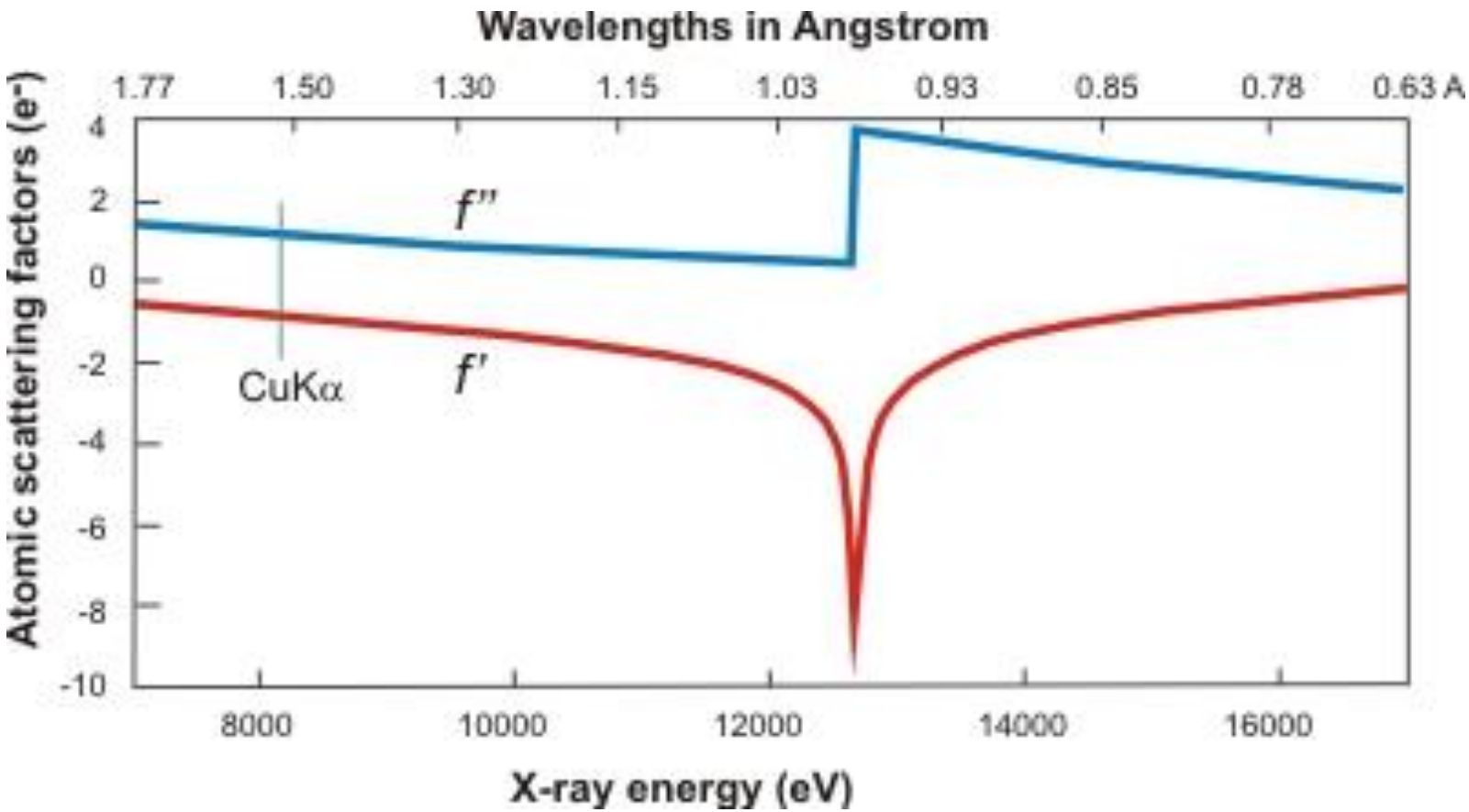
Protein A is a putative histidinol phosphatase from Listeria monocytogenes

	PDB ID	Gene	Residues	Metal in PDB	Metals in PIXE (>3xLOD)	Potential metals in PIXE (1-3xLOD)	Crystallization conditions
PDB inconsistent with PIXE							
1		BiR14	456	Ca	-	Ca, Mn	18% PEG 3350, 0.2M Ca acetate, 0.1M MES, pH 6.15
2		NsR437I	106	Mn	-	-	20% PEG 4000, 0.1M Mn chloride, 0.1M MES, pH 6.0
3		SnR135 D	161	Ca	-	Ca	20% PEG 8000, 0.1M Ca acetate, 0.1M MES, pH 6.0
4		Protein A	283	Fe/Zn	Ca (3.3), Mn (0.5), Fe (1.2), Co (1.2)	Zn	15% PEG 8000, 0.17 M sodium acetate, 0.01 M L-cysteine, 0.1 M MES pH 6.2
5		NsR236	119	K	-	Ca	8.64 M K acetate, 0.1 M TAPS, pH 9.0
6		NsR437 H	141	Mn	-	Fe, Co	20% PEG 1000, 0.1M Mn chloride, 0.1M MES, pH 6.0
7		SoR237	137	Na	Co (0.7), Zn (0.7)	Fe, Ni	NaCl 200 mM, MES PH6, PEG 3350 20%, pH 6.15
8		BtR324A	169	Zn	-	Ca, Mn, Fe*	0.75M Mg Formate, 0.1M Bis-Tris, pH 7.0
9		GR157	262	Zn	-	Co	100 mM Na Acetate (pH 4.6), 30% MPD, and 200 mM NaCl.

Model in the PDB containing a metal from the crystallization cocktail and not protein

Model in the PDB containing an incorrect metal

Focus on one example

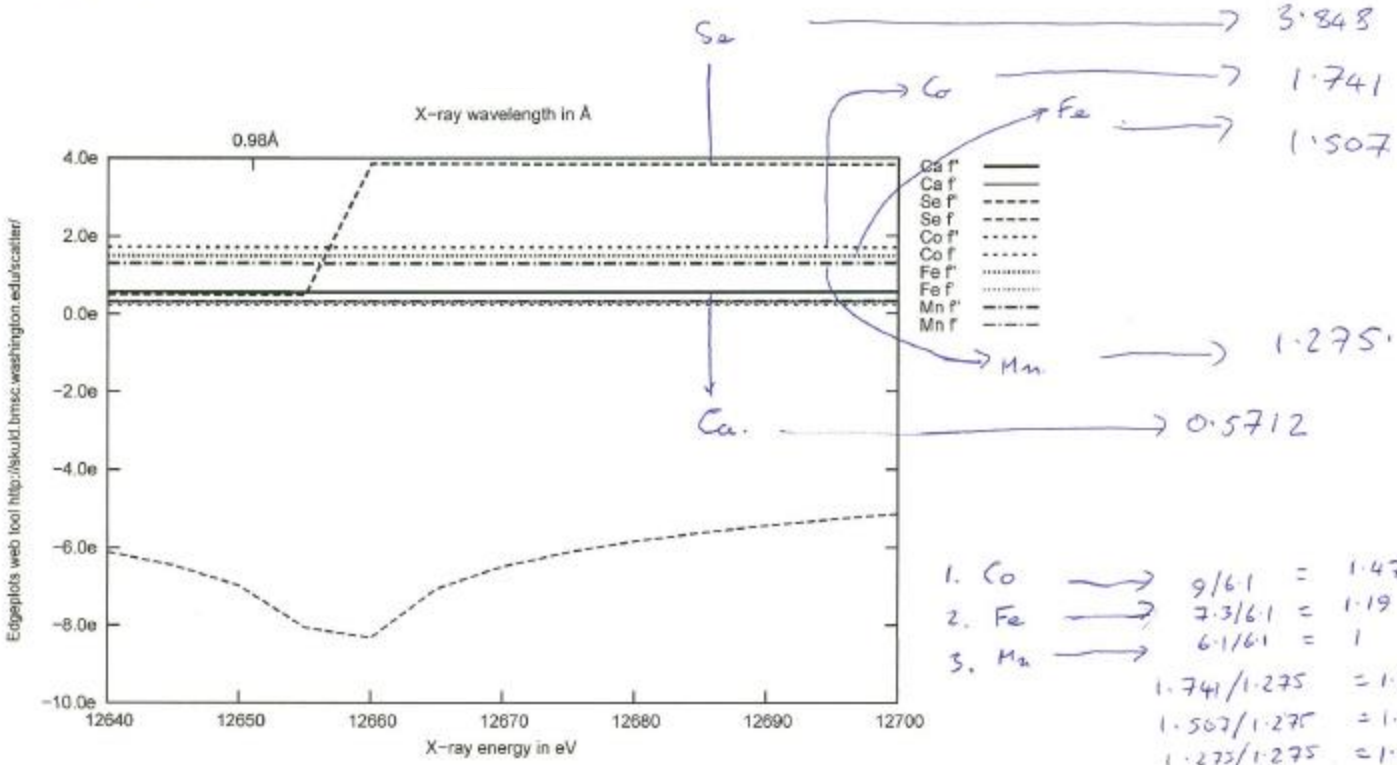

Wavelengths in Angstrom
1.77
1.50
1.30
1.15
1.03
0.93
0.85
0.78
0.63 A
Atomic scattering factors (e⁻)
4
2
0
-2
-4
-6
-8
-10
f''
f'
CuKα
8000
10000
12000
14000
16000
X-ray energy (eV)

Se 3.848
Co 1.741
Fe 1.507
Mn 1.275
Cu 0.5712.

A Chm

ED.

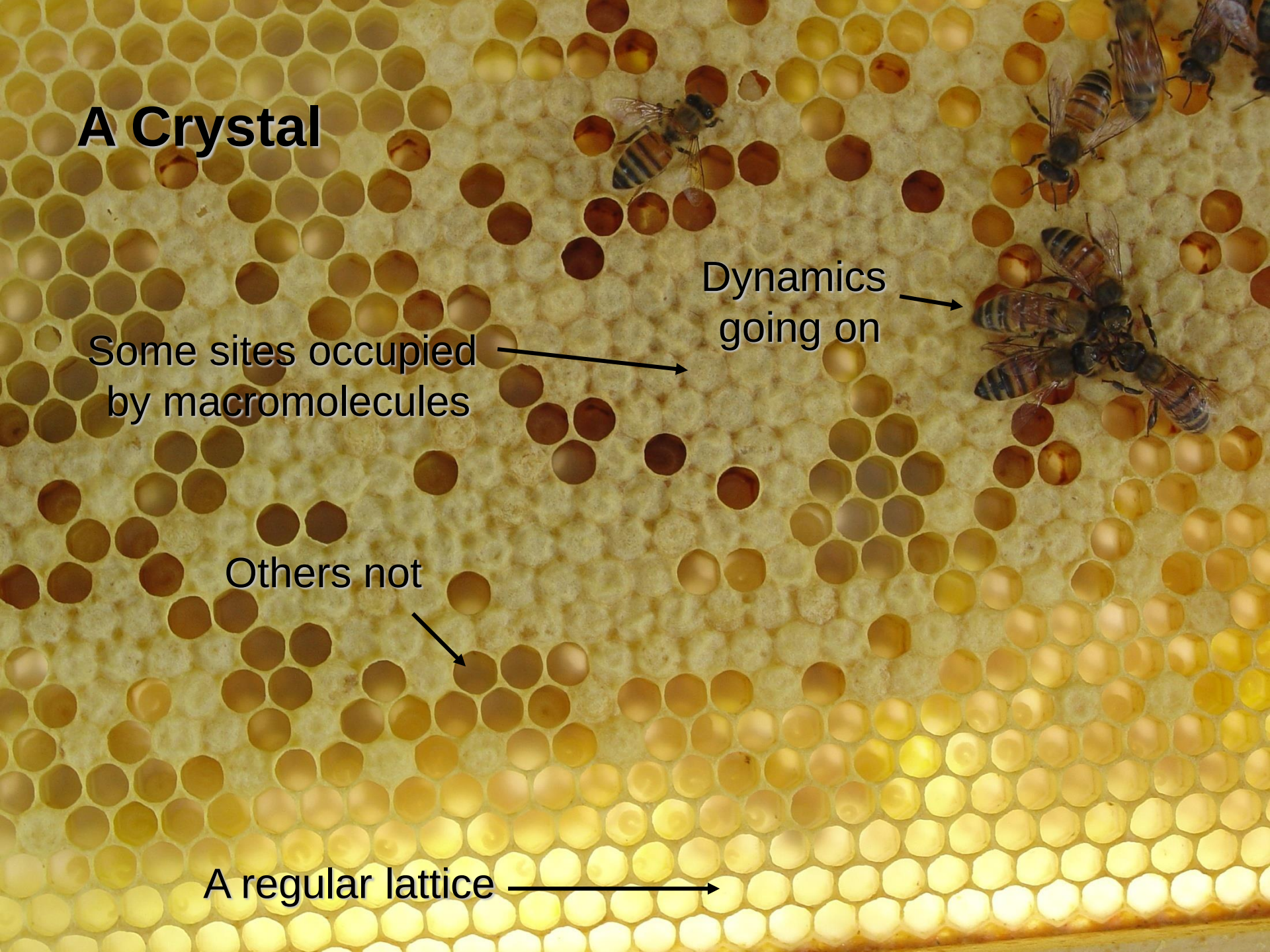
→ (A) 1. 25.79 → Next lowest (2) 7.3 mm
2. 17.19 → Lowest ED (3) 6.1 mm
3. 13.95 → Highest (1) 9.0 mm



1. Co → 9/6.1 = 1.47
2. Fe → 7.3/6.1 = 1.19
3. Mn → 6.1/6.1 = 1

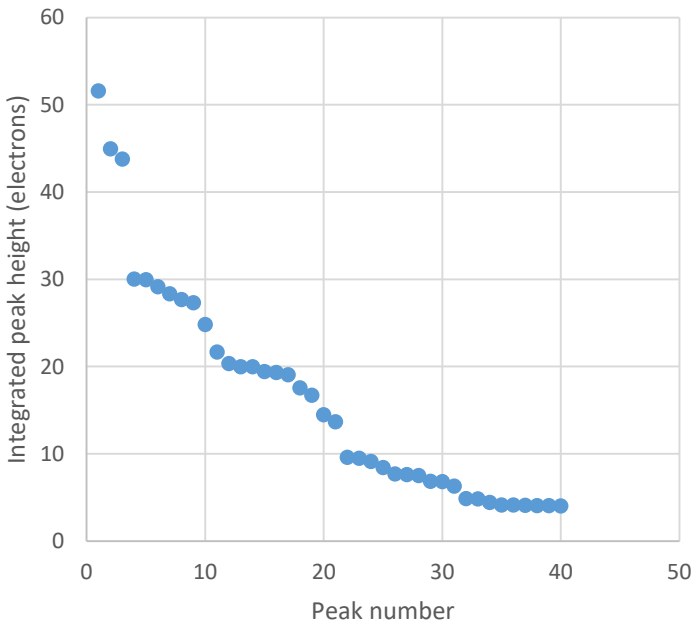
1.741/1.275 = 1.365
1.507/1.275 = 1.18
1.275/1.275 = 1.00

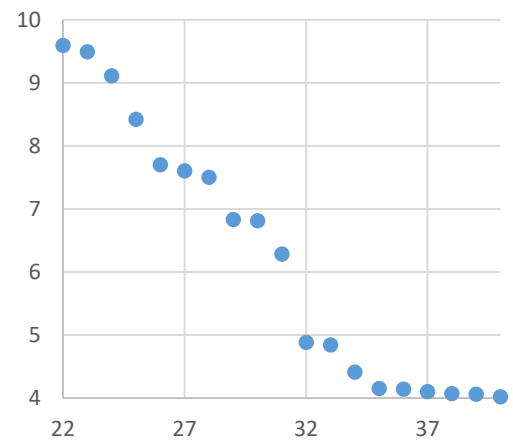
A Crystal


Dynamics going on
Some sites occupied by macromolecules
Others not
A regular lattice

Wavelength 0.97931

	f'	f''	f'' n_Se	f'' n_Zn	f'' n_Co	f'' n_Fe
Se	-8.6571	3.843	1.000			
Zn	-0.3843	2.477	0.645	1.000		
Co	0.1697	1.715	0.446	0.692	1.000	
Fe	0.2421	1.500	0.390	0.606	0.875	1.000
Mn	0.2905	1.303	0.339	0.526	0.760	0.869
Ca	0.2938	0.565	0.147	0.228	0.329	0.376
O	0.0163	0.012	0.003	0.005	0.007	0.008





21 Se atoms, 7 in each chain A,B and C

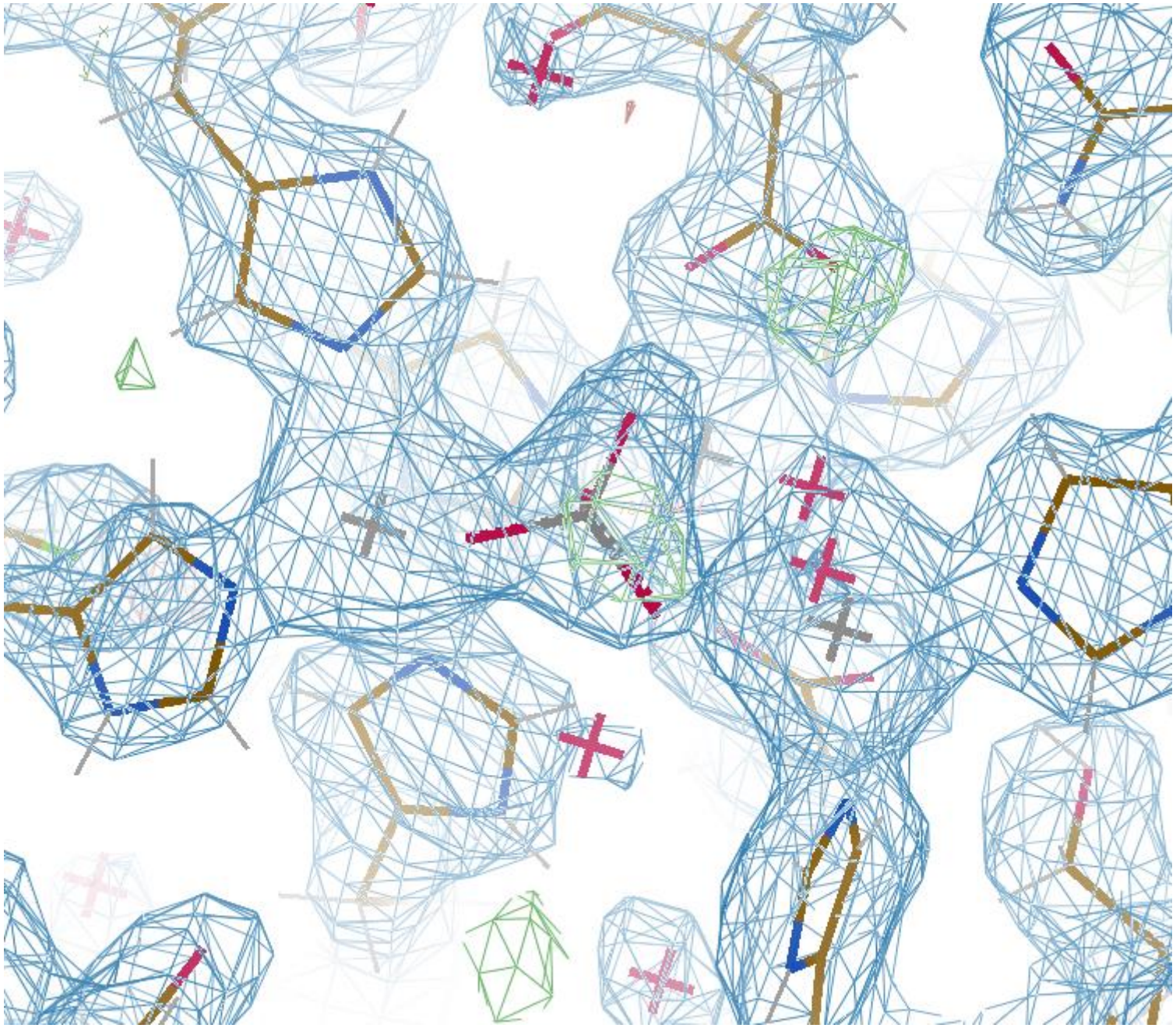
Wavelength 0.97931

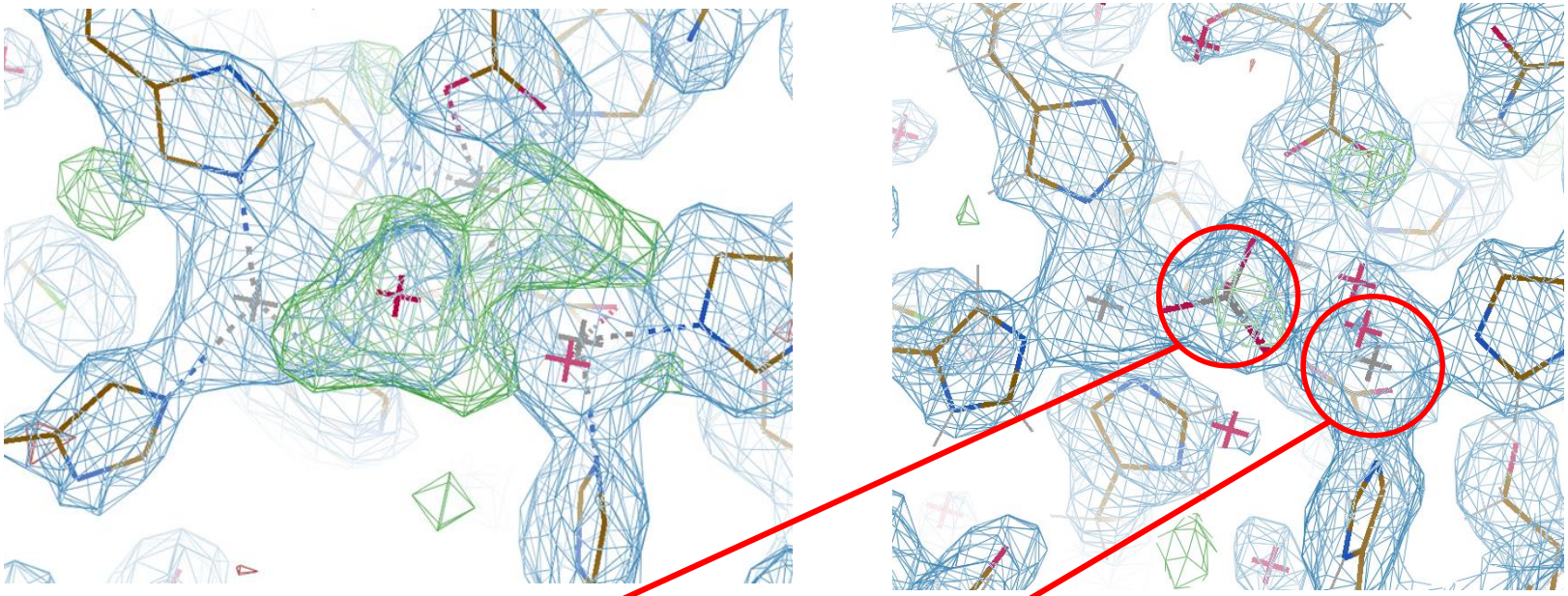
	f'	f''	f'' n_Se	f'' n_Zn	f'' n_Co	f'' n_Fe
Se	-8.6571	3.843	1.000			
Zn	-0.3843	2.477	0.645	1.000		
Co	0.1697	1.715	0.446	0.692	1.000	
Fe	0.2421	1.500	0.390	0.606	0.875	1.000
Mn	0.2905	1.303	0.339	0.526	0.760	0.869
Ca	0.2938	0.565	0.147	0.228	0.329	0.376
O	0.0163	0.012	0.003	0.005	0.007	0.008

Chain A	9.59	1.00
	6.83	0.71
	6.81	0.71
Chain B	8.42	1.00
	7.60	0.90
	7.50	0.89
Chain C	9.11	1.00
	7.70	0.85
	6.28	0.69

Re-Refining Protein A

	R_{work}	R_{free}	RMS(bonds)	RMS(angles)	Clash	Ram-fav	Ram-out	Rot-out
PDB								
	0.193	0.212	0.008	1.2	11.97	96.07	0.61	
Re-refined								
	0.1847	0.2143	0.0031	0.744	1.9	96.81	0.61	2.82
Metal	Metals replaced with Co, Fe and Mn, PO_4 added in active site. Ca added in places							
	15.60	18.50						





PO_4

Fe

Metals in new structure, Fe, Mn, Co cluster

BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

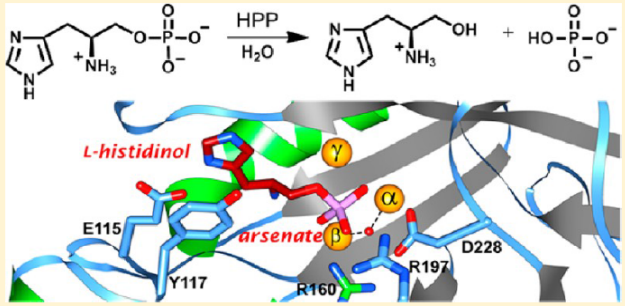
pubs.acs.org/biochemistry

A closely related protein

site and had a catalytic efficiency of ~10^3 M^{-1} s^{-1}. Expression of the protein under iron-free conditions resulted in the production of an enzyme with a 2 order of magnitude improvement in catalytic efficiency and a mixture of zinc and manganese in the active site. Solvent isotope and viscosity effects demonstrated that proton transfer steps and product dissociation steps are not rate-limiting. X-ray structures of HPP were determined with sulfate, L-histidinol phosphate, and a complex of L-histidinol and arsenate bound in the active site. These crystal structures and the catalytic properties of variants were used to identify the structural elements required for catalysis and substrate recognition by the HPP family of enzymes within the amidohydrolase superfamily.

Supporting Information

ABSTRACT: L-Histidinol phosphate phosphatase (HPP) catalyzes the hydrolysis of L-histidinol phosphate to L-histidinol and inorganic phosphate, the penultimate step in the biosynthesis of L-histidine. HPP from the polymerase and histidinol phosphatase (PHP) family of proteins possesses a trinuclear active site and a distorted $(\beta/\alpha)_7$-barrel protein fold. This group of enzymes is closely related to the amidohydrolase superfamily of enzymes. The mechanism of phosphomonoester bond hydrolysis by the PHP family of HPP enzymes was addressed. Recombinant HPP from *Lactococcus lactis* subsp. *lactis* that was expressed in *Escherichia coli* contained a mixture of iron and zinc in the active site and had a catalytic efficiency of ~10^3 M^{-1} s^{-1}. Expression of the protein under iron-free conditions resulted in the production of an enzyme with a 2 order of magnitude improvement in catalytic efficiency and a mixture of zinc and manganese in the active site. Solvent isotope and viscosity effects demonstrated that proton transfer steps and product dissociation steps are not rate-limiting. X-ray structures of HPP were determined with sulfate, L-histidinol phosphate, and a complex of L-histidinol and arsenate bound in the active site. These crystal structures and the catalytic properties of variants were used to identify the structural elements required for catalysis and substrate recognition by the HPP family of enzymes within the amidohydrolase superfamily.



Metal content measured with an inductively coupled mass spectrometer

Accurate Metal identification is important

- The original structure contained Fe and Zn.
- The revised structure shows the phosphate and Co.
- The phosphate and tri-nuclear metal center are important for mechanism.

- Where they from crystallization? In some cases we don’t know due to incomplete crystallization information.

Important notes about the technique

- Because PIXE is an elemental analysis the sample does not have to be in any preserved state.
- Samples from years ago can be used to collect experimental data.
- The number and ratio of different metals (or other atoms) per protein molecule can be determined.
- Not discussed today, but the data reveals signatures in protein models coupled with the deposited X-ray data that identify suspect metals.

Summary

- Crystallization analysis and elemental analysis have great potential in improving structural models.
- This improvement is needed as our limited study shows a an error rate of greater than 50%.
- Experimentally identifying errors defines signatures of those same errors in other structural models.
- The work leads to a potential quality control mechanism to identify suspect structural models.
- It also allows native metals (at least from expression) to be distinguished from opportune ones.

The Team

Elspeth Garman, Geoffrey Grime,
Elizabeth Snell, and Oliver Zeldin

Special thanks to the 'Pixie'

Thank you and questions?

esnell@hwi.buffalo.edu