Protein Crystallography

Mitchell Guss

University of Sydney

Australia

THE UNIVERSITY OF SYDNEY

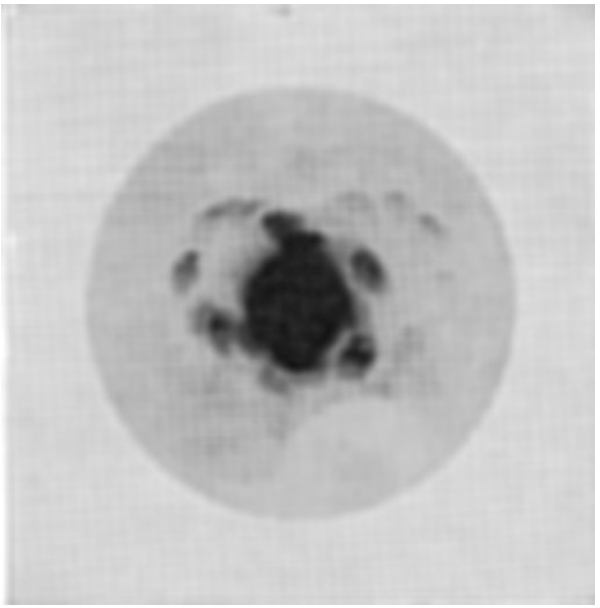
2014 international year of crystallography

2014 International Year of Crystallography

- Declared by the General Assembly of the United Nations in June 2013
  - Administered jointly by UNESCO and the IUCr
- See the website for list of activities and resources www.iycr.org

100+ years of X-ray crystallography

- 1912 Max von Laue and his colleagues demonstrate that a crystal (copper sulfate) can diffract X-rays
- Awarded the Nobel Prize 1914

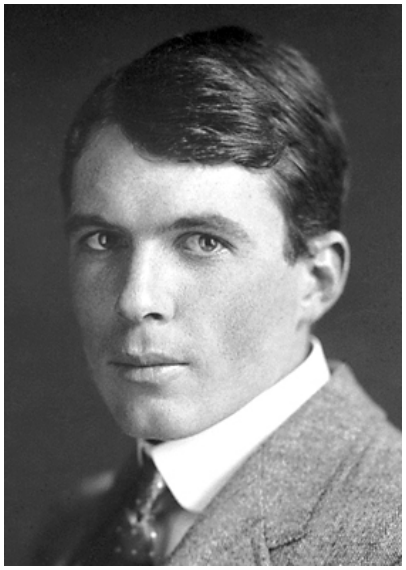


Friedrich, Knipping and von Laue, 1912

- Showed that X-rays behaved as waves
- Failed to correctly relate the scattering to the underlying structure

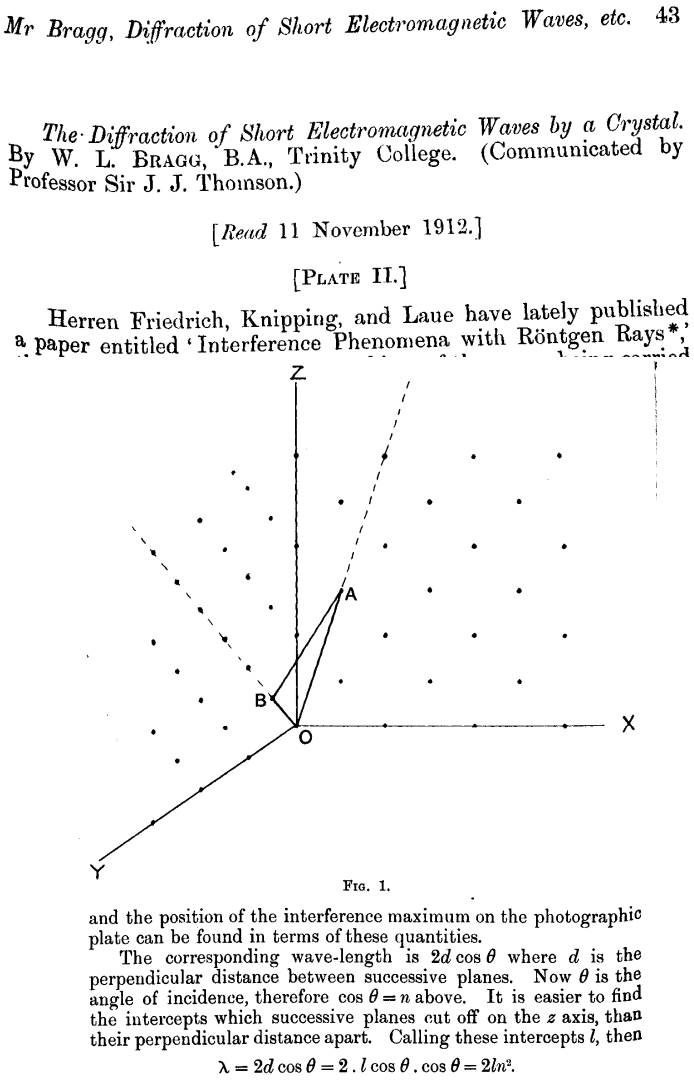
100+ years of X-ray crystallography

- 1912 William Lawrence Bragg interpreted the diffraction from sodium chloride
- Awarded the Nobel Prize with his father (W.H. Bragg) in 1915
  - W.L. Bragg still remains the youngest Nobel laureate at 25 years of age



Braggs, 1912

- Showed that the diffraction was equivalent to the reflection from planes drawn through the crystal
- Could predict the positions of the diffracted spots using Bragg's law: $\lambda = 2d\sin\theta$



Mr Bragg, Diffraction of Short Electromagnetic Waves, etc. 43

The Diffraction of Short Electromagnetic Waves by a Crystal.
By W. L. Bragg, B.A., Trinity College. (Communicated by Professor Sir J. J. Thomson.)

[*Read* 11 November 1912.]

[Plate II.]

Herren Friedrich, Knipping, and Laue have lately published a paper entitled 'Interference Phenomena with Röntgen Rays*',

Fig. 1.

and the position of the interference maximum on the photographic plate can be found in terms of these quantities.

The corresponding wave-length is $2d\cos\theta$ where d is the perpendicular distance between successive planes. Now θ is the angle of incidence, therefore $\cos\theta = n$ above. It is easier to find the intercepts which successive planes cut off on the z axis, than their perpendicular distance apart. Calling these intercepts l, then

$$\lambda = 2d\cos\theta = 2\,.\,l\cos\theta\,.\cos\theta = 2ln^2.$$

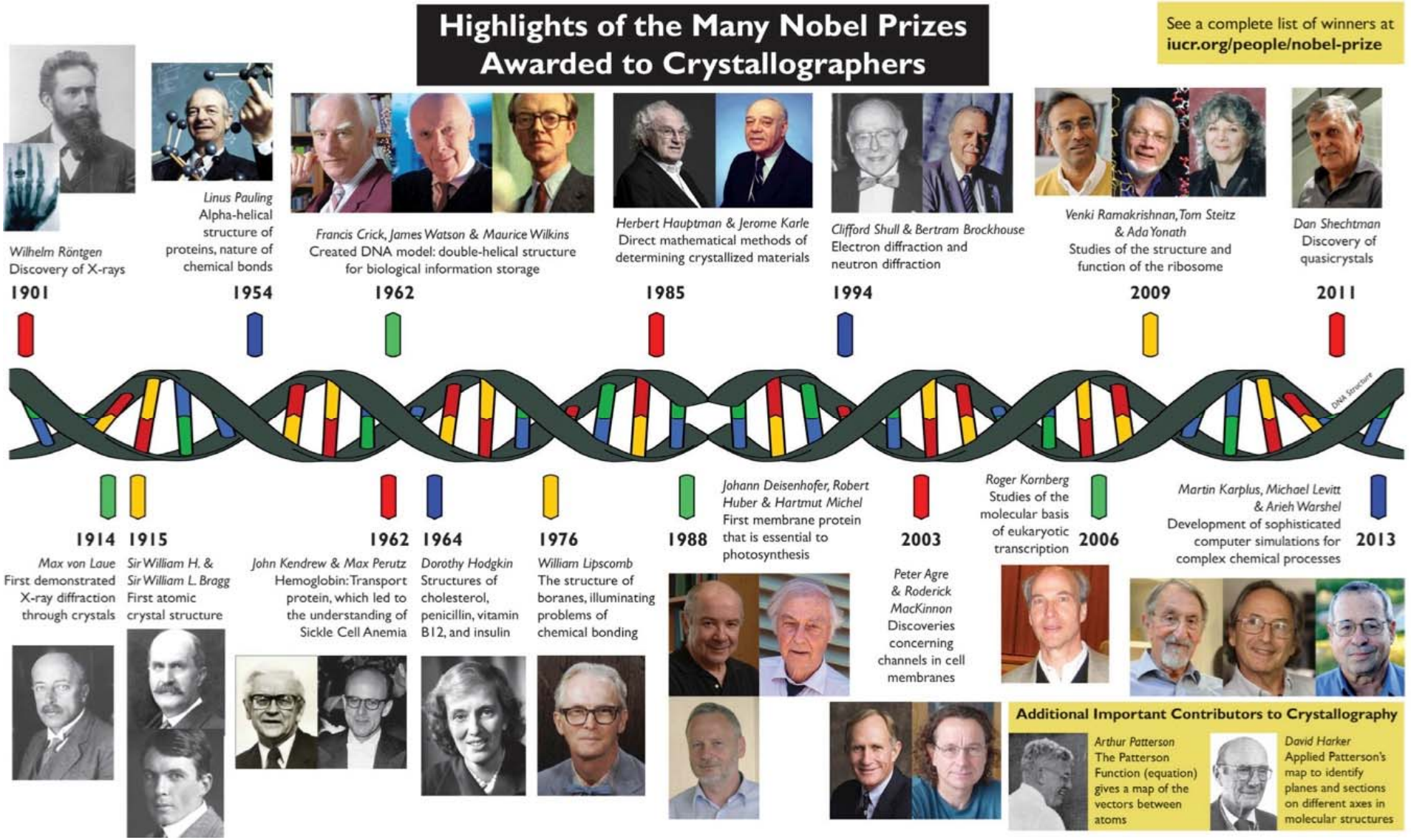
Timeline of biological crystallography - 1

- 1895 Röngten: discovers X-rays (Nobel 1901)
- 1912 von Laue: demonstrates diffraction (Nobel 1914)
- 1912 Braggs: first crystal structure (Nobel 1915)
- 1929 Sumner: crystallises an enzyme (Nobel 1946)
- 1951 Pauling: protein components α-helix and β-sheet (Nobel Chemistry 1954, Peace 1962)
- 1953 Watson, Crick & Wilkins: DNA structure (Nobel 1962)
- 1954 Crowfoot-Hodgkin: vitamin B12, penicillin (Nobel 1964)
- 1959 Kendrew & Perutz: myoglobin and haemoglobin (Nobel 1962)

Timeline of biological crystallography - 2

- 1970+ Klug crystallographic electron microscopy and nucleic acid-protein complexes (Nobel 1982)
- 1985 Michel, Deisenhofer & Huber: the photoreaction centre, first membrane protein (Nobel 1988)
- 1997 Boyer & Walker: ATPsynthase a rotating molecule (Nobel 1998)
- 1998 MacKinnon: K+ ion channel (Nobel 2003)
- 2000 Yonath, Ramakrishnan & Steitz: the ribosome (Nobel 2009)
- 2011 Lefkowitz & Kobilka: G-protein coupled receptors (Nobel 2012)

The Nobel Laureates in Crystallography



The beginnings of (X-ray) protein crystallography

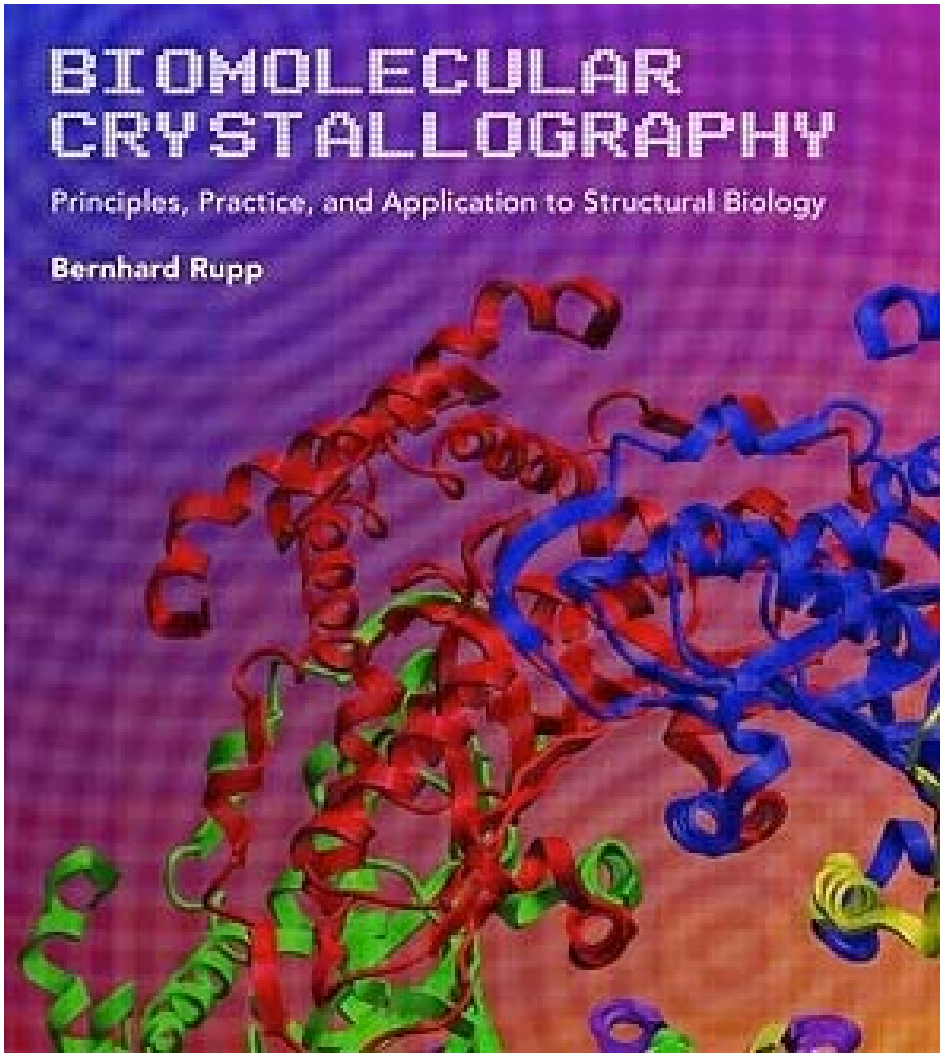
- 1930's Dorothy Crowfoot (later Hodgkin) and J.D. Bernal made the first X-ray diffraction pictures of an enzyme – they realised that the crystals had to be kept wet or they didn't diffract
- 1937 Max Perutz began the structure analysis of haemoglobin with W.L. Bragg as his supervisor – the structure was not solved until 1961!

What has changed from 1937 to today that enables the same structure to be solved in minutes?

Excellent text for biomolecular crystallography

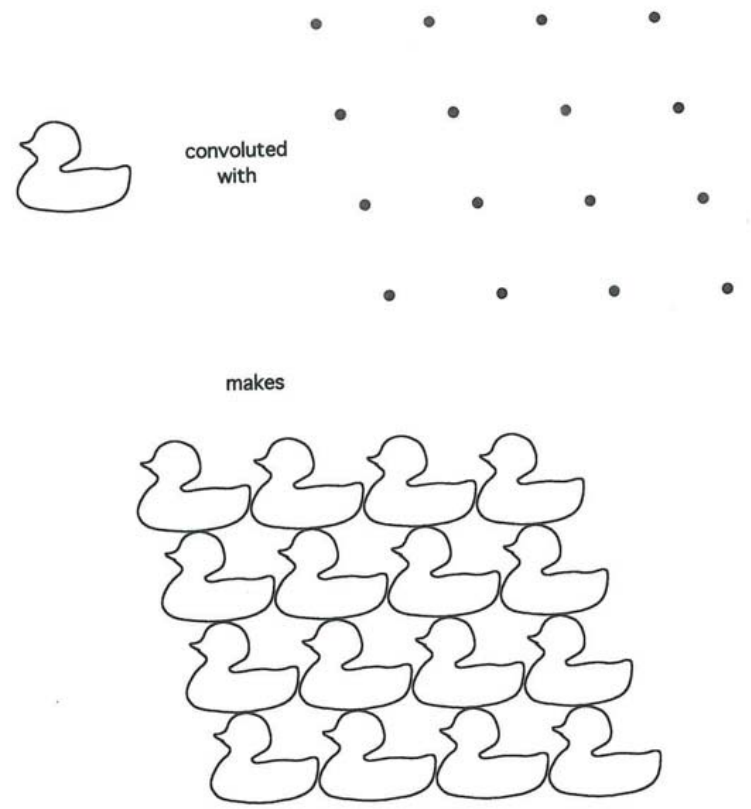
- Excellent coverage of both the theory and practice of the subject
- Many of my illustrations come from this text
- See Bernard Rupp's website for more detail

http://www.ruppweb.org/default.htm



What is a crystal?

- An solid containing of atoms or molecules that repeats in three dimensions
- *Or,* the convolution of a motif (molecule) with a lattice

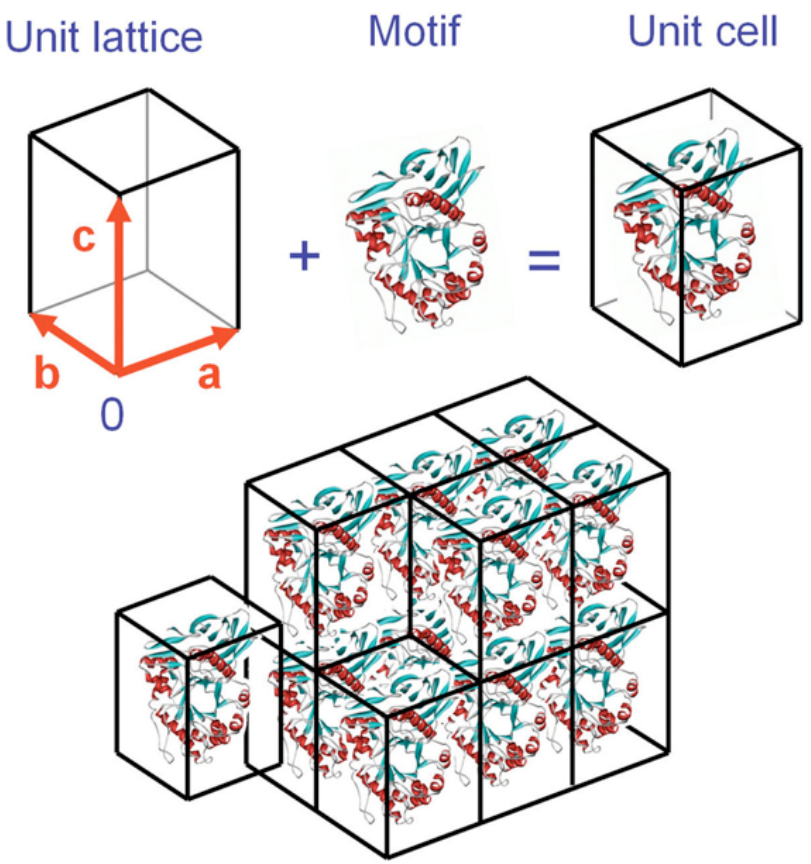


In this case a “duck” is convoluted with a 2-dimensional lattice to give a 2D “duck crystal”

Convolution
- Place an instance of the motif at each point of the lattice

3-dimensional protein crystals

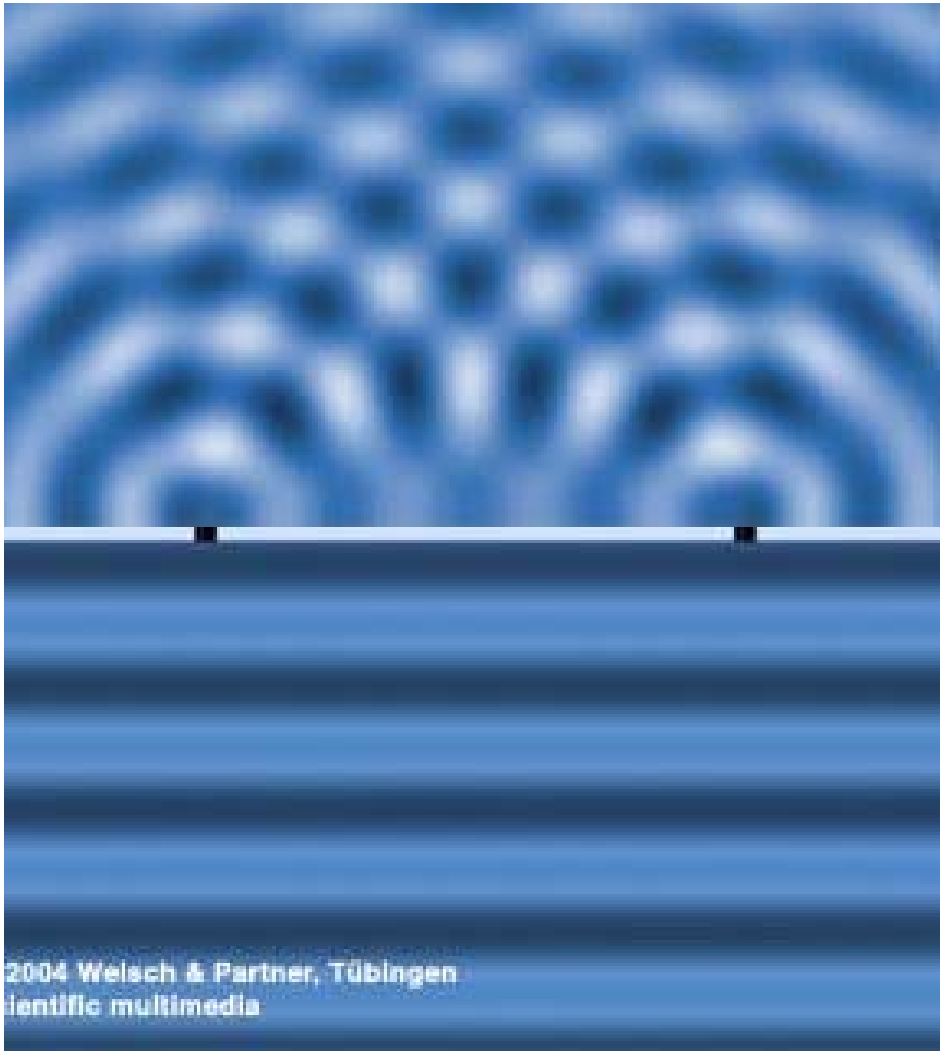
- The unit lattice is the 3D set of points described by the 3 non-colinear vectors that give the repeating units
- The motif is the content of the unit lattice (Note: it may be one or a number of actual molecules)
- The unit cell is the unit lattice with its contents
- The crystal is the 3D array of unit cells that fill space



What is diffraction?

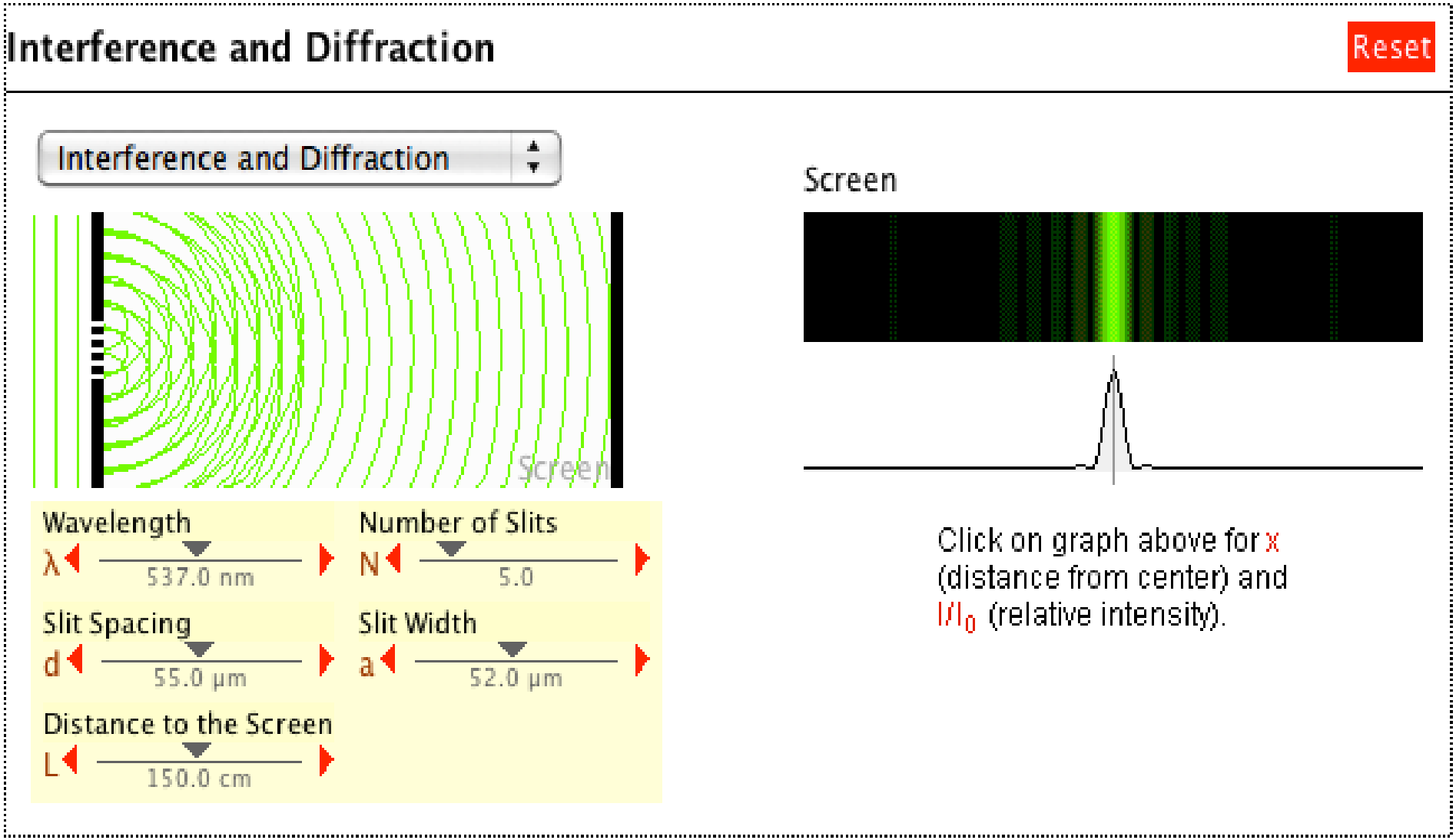
- Diffraction takes place with sound; with electromagnetic radiation, such as light, X-rays, and gamma rays; and with very small moving particles such as atoms, neutrons, and electrons, which show wavelike properties.
- The phenomenon is the result of interference (i.e., when waves are superposed, they may reinforce or cancel each other out) and is most pronounced when the wavelength of the radiation is comparable to the linear dimensions of the obstacle.

Since X-rays are of approximately the same wavelength as inter-atomic bonds they will be diffracted by molecules and crystals.

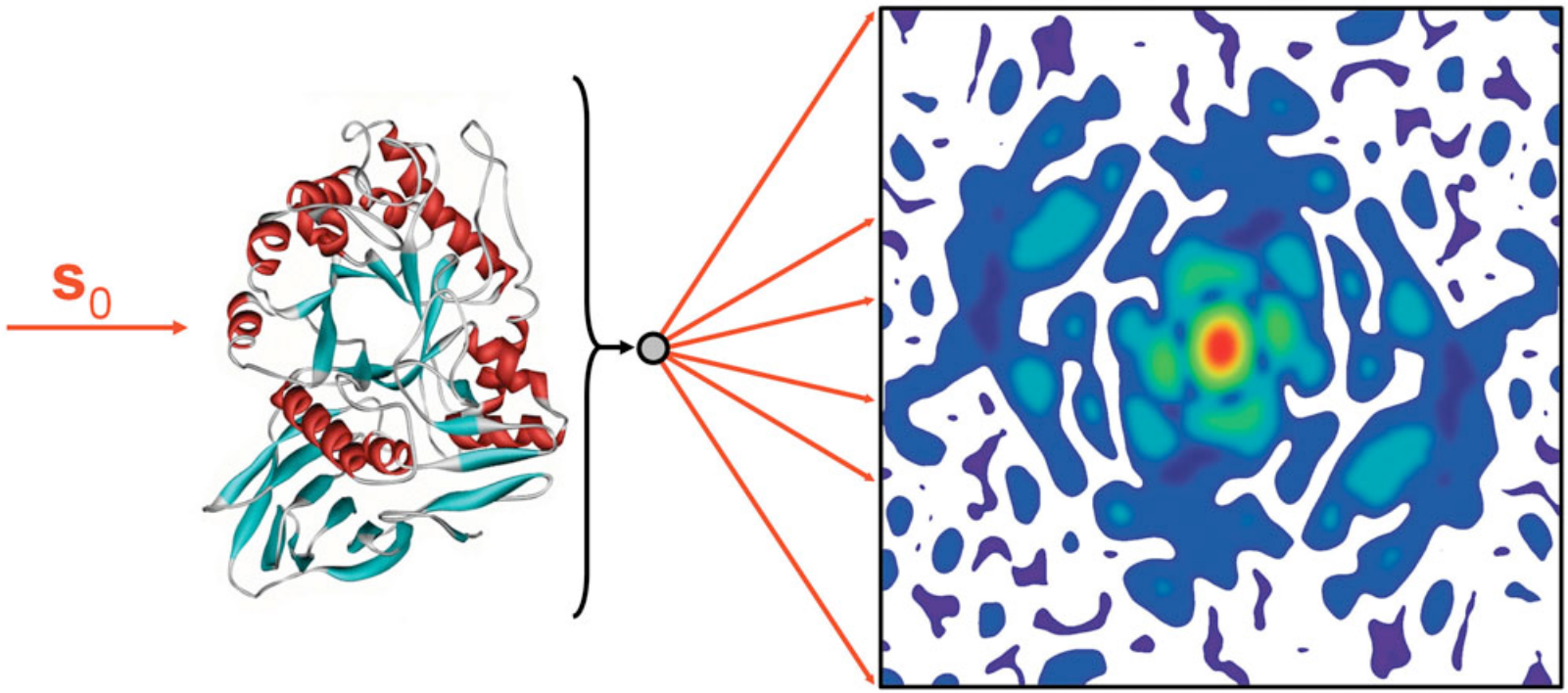


Diffraction simulation

http://www.pas.rochester.edu/~ksmcf/p100/java/Optics/Diffraction.html



Scattering from a molecule



The image is the scattering of the 2D projection protein molecule on the left. It is the sum of the scattering from each atom (*j* at position *S*). The diffraction (F_S) is continuous and can occur in any direction ***S***.

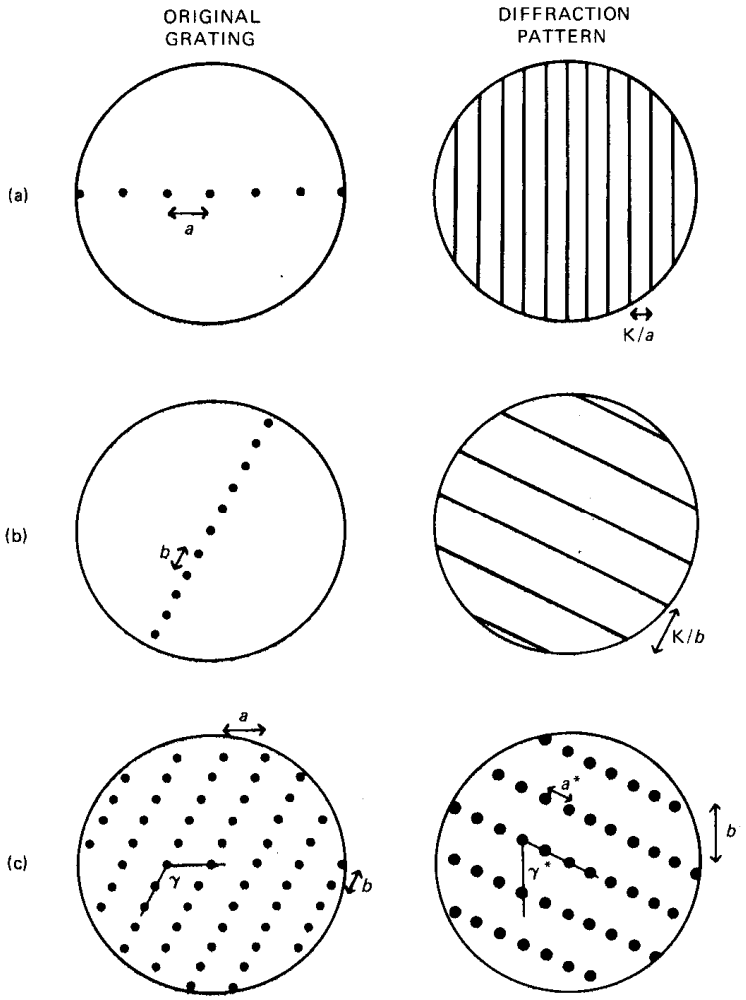
$$F_S = \sum_{j=1}^{atoms} f_{Sj}^0 . \exp(2\pi i S r_j)$$

Diffraction from a lattice

(a, b) Diffraction from a line of evenly spaced dots (or holes) is a series of lines perpendicular to the row of dots. The lines have inverse spacing to the dots.

(c) Diffraction from a 2D array of dots is an array of dots at right angles to the original rows.

The array of dots in (c) is the convolution of the dots in (a) and (b). The diffraction pattern in (c) is the product of the diffraction patterns in (a) and (b). What you see in the (c) is a real lattice and the associated reciprocal lattice

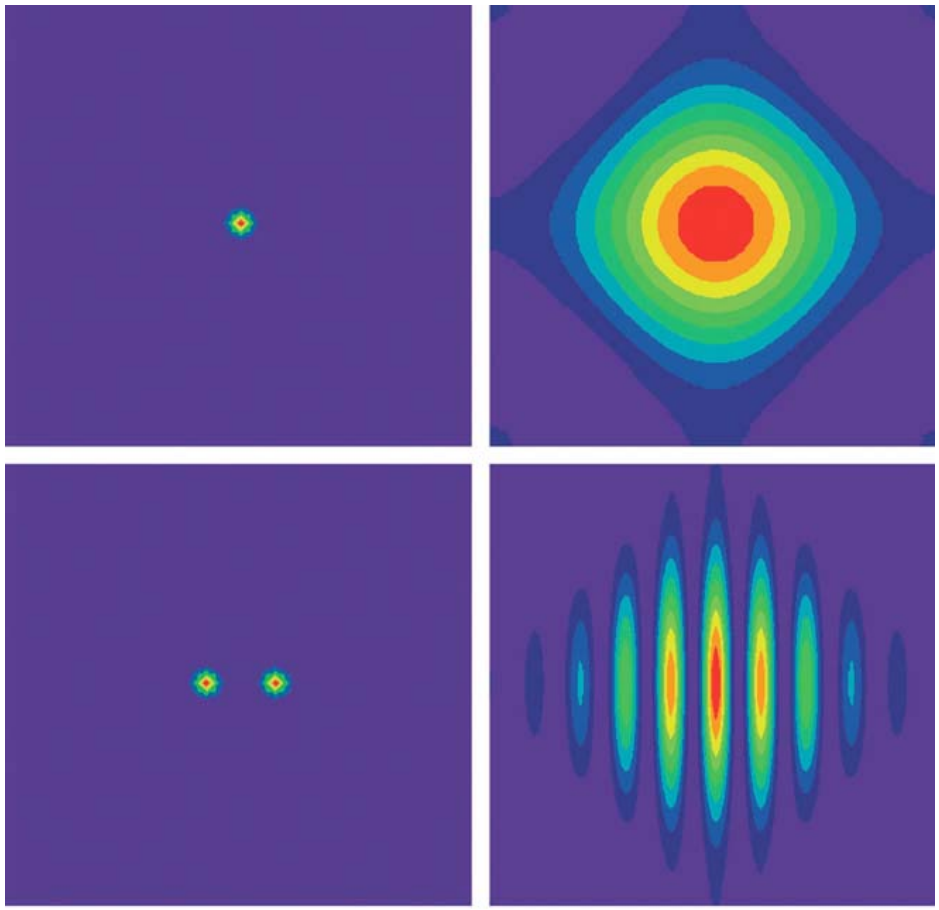


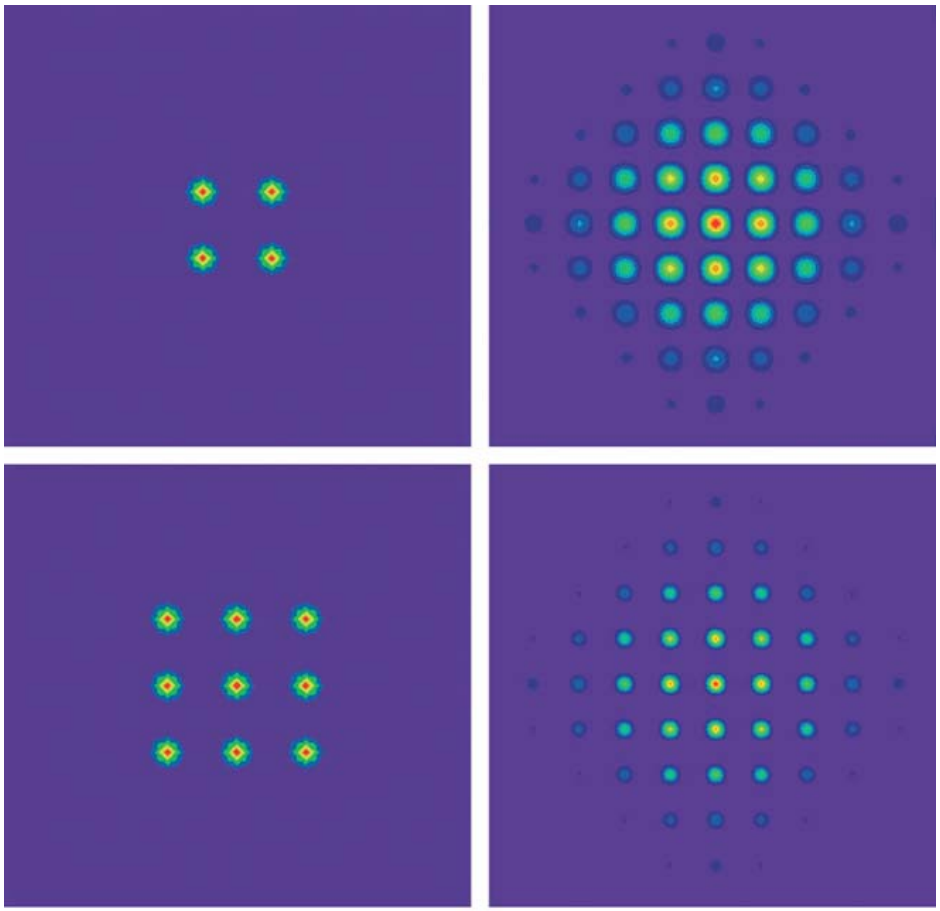
The convolution theorem and diffraction from a crystal

The convolution theorem states that under suitable conditions the Fourier transform of a convolution of two functions is the pointwise product of Fourier transforms of the individual functions.

Remember a “crystal” is the convolution of a lattice and a “molecule”. Therefore the diffraction from a crystal will be the product of the diffraction of the lattice and of one molecule.

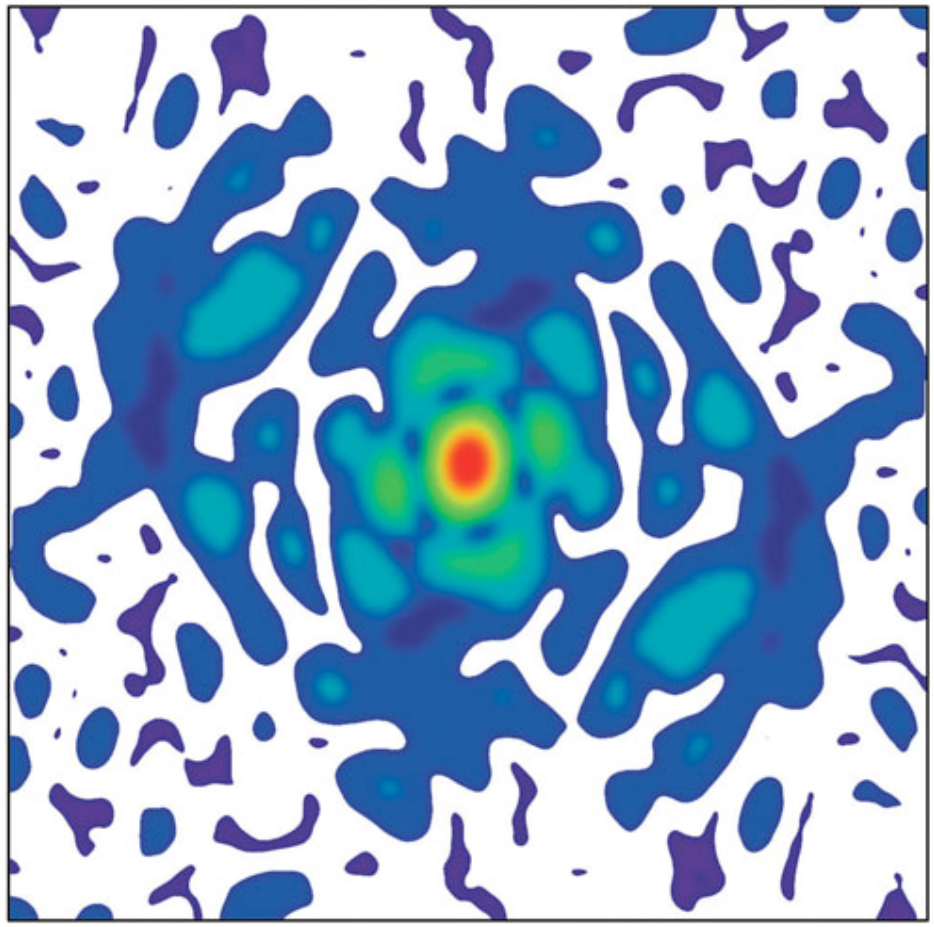
Scattering from a crystal

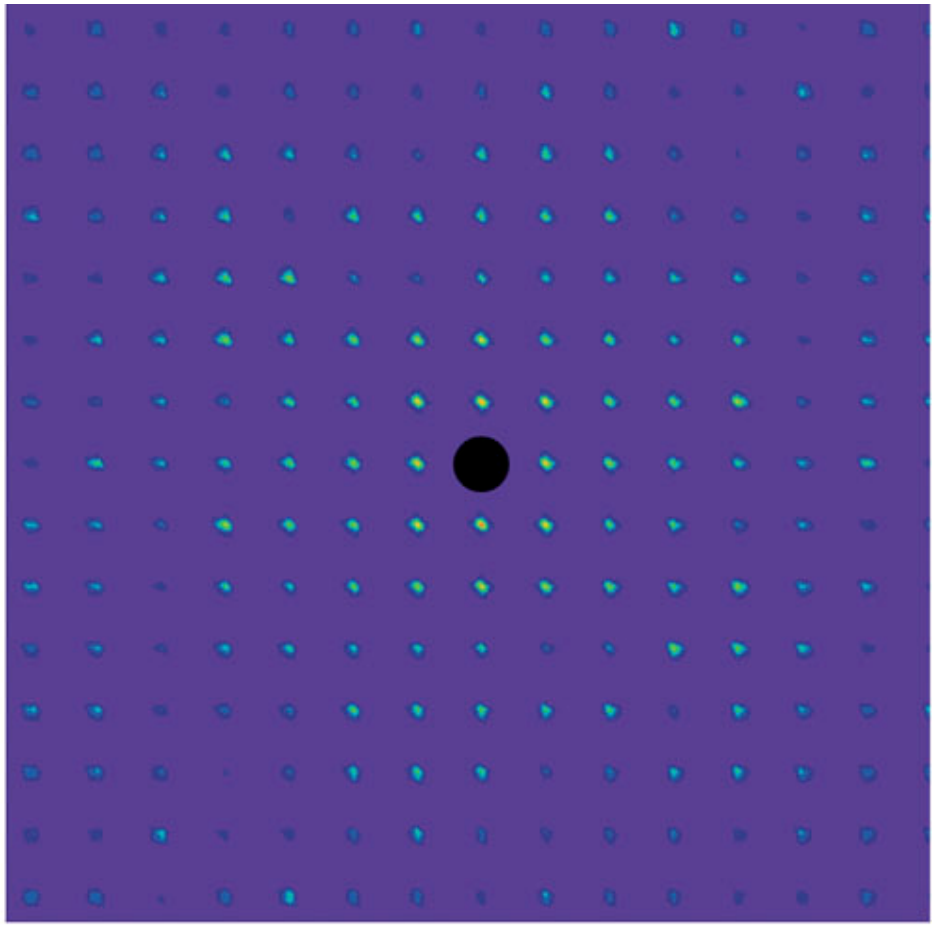




The underlying pattern remains the same. It is simply “sampled” at points allowed by the repeating lattice.

Scattering from a protein crystal





The effect of the crystal is to reinforce the diffraction signal at points where the waves from all the molecules are in phase. This makes the signal sufficiently large to be measured but all the information is theoretically in the scattering from one molecule.

Scattering from a crystal

The following formula shows how you can calculate the position and magnitude [F(hkl) is a vector] of the scattered X-rays if you know the electron-density [$\rho(x, y, z)$] at every position in space.

$$\mathbf{F}(hkl) = V \sum_{x} \sum_{y} \sum_{y} \rho(x, y, z) \bullet \exp[2\pi i(hx + ky + lz)]$$

Since the electrons are associated with atoms we can express **F**(*hkl*) in terms of the atomic positions and the scattering from a single atom of the appropriate type.

$$\mathbf{F}(hkl) = \sum_{j=1}^{atoms} f_{S,j}^{0} \bullet \exp[2\pi i(hx_j + ky_j + lz_j)]$$

Where f^0 represents the scattering from the jth atom at position (x_j,y_j,z_j). But that is the answer – we wish to calculate the electron density [ρ(x, y, z)] from the diffraction pattern.

Calculating the electron density

We can measure the intensities in the diffraction pattern [I(*hkl*)] that are proportional to the squares of the amplitudes of the structure factors.

$$I(hkl) \propto |\mathrm{F(hkl)}|^2$$

$$\rho(x,y,z) = 1/V \sum_{-h}^{h} \sum_{-k}^{-k} \sum_{-l}^{-l} |\mathbf{F}(hkl)| \exp[-2\pi i(hx+ky+lz-\alpha(hkl))]$$

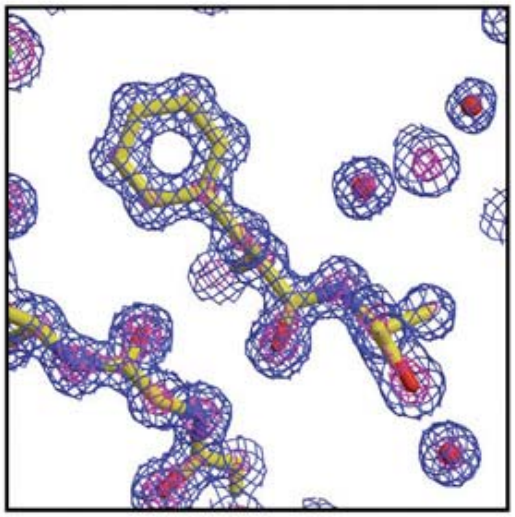
The electron density we are trying to find

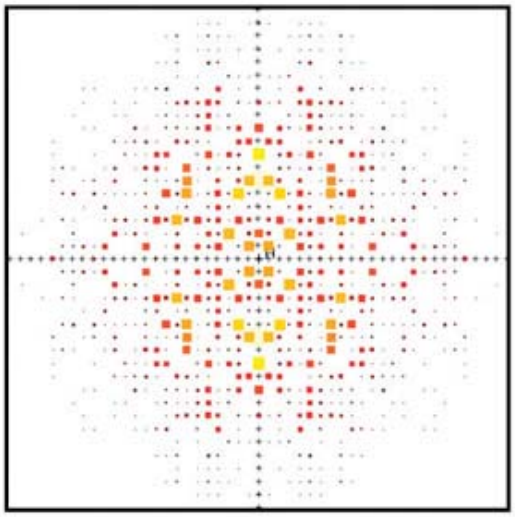
Structure factor amplitudes we can measure

We can't measure the relative phases – they need to be determined – "solving the structure"

The so-called ‘phase problem’

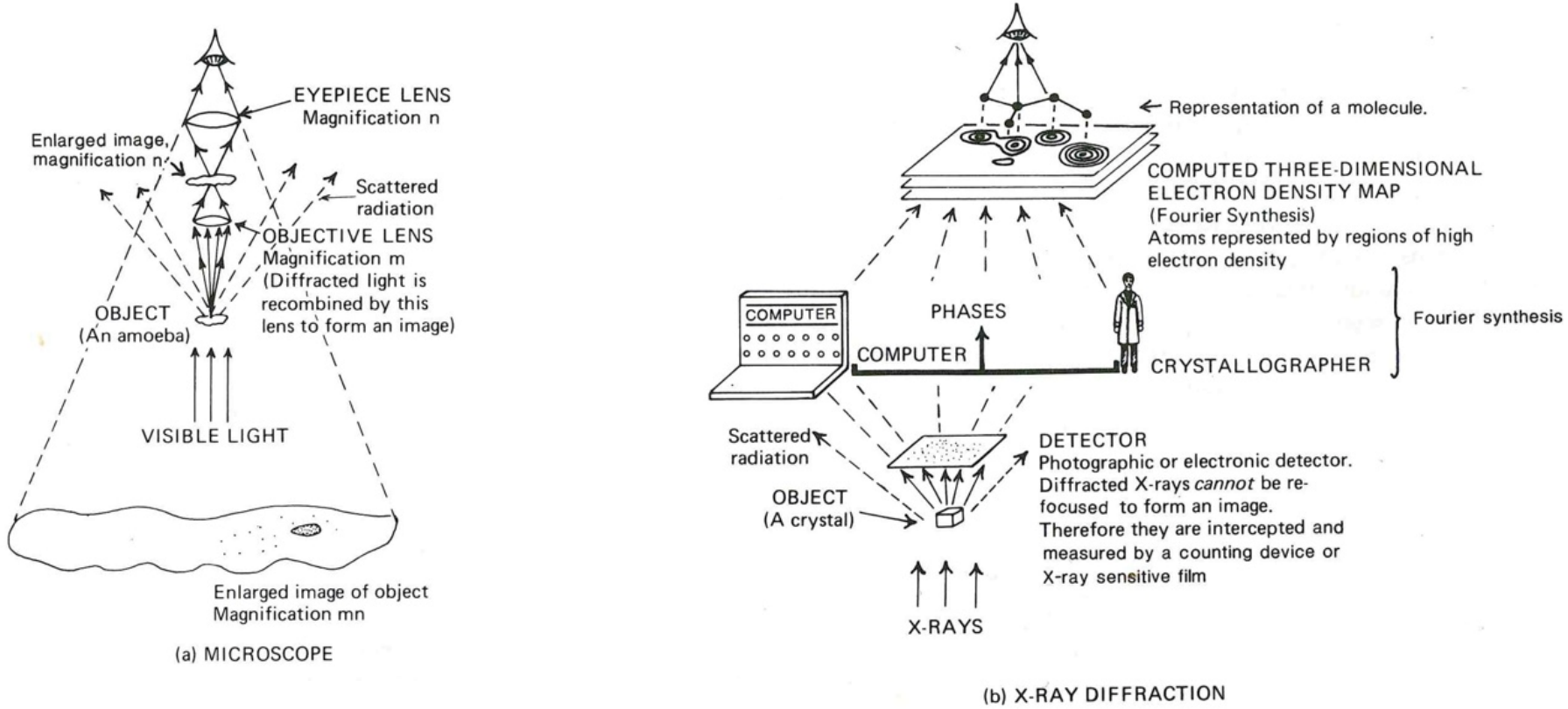
$$\rho(x,y,z) = \frac{1}{V}\sum_{-h}^{h}\sum_{-k}^{k}\sum_{-l}^{l} F_{hkl}\cdot\exp[-2\pi i(hx+ky+lz-\alpha_{hkl})]$$





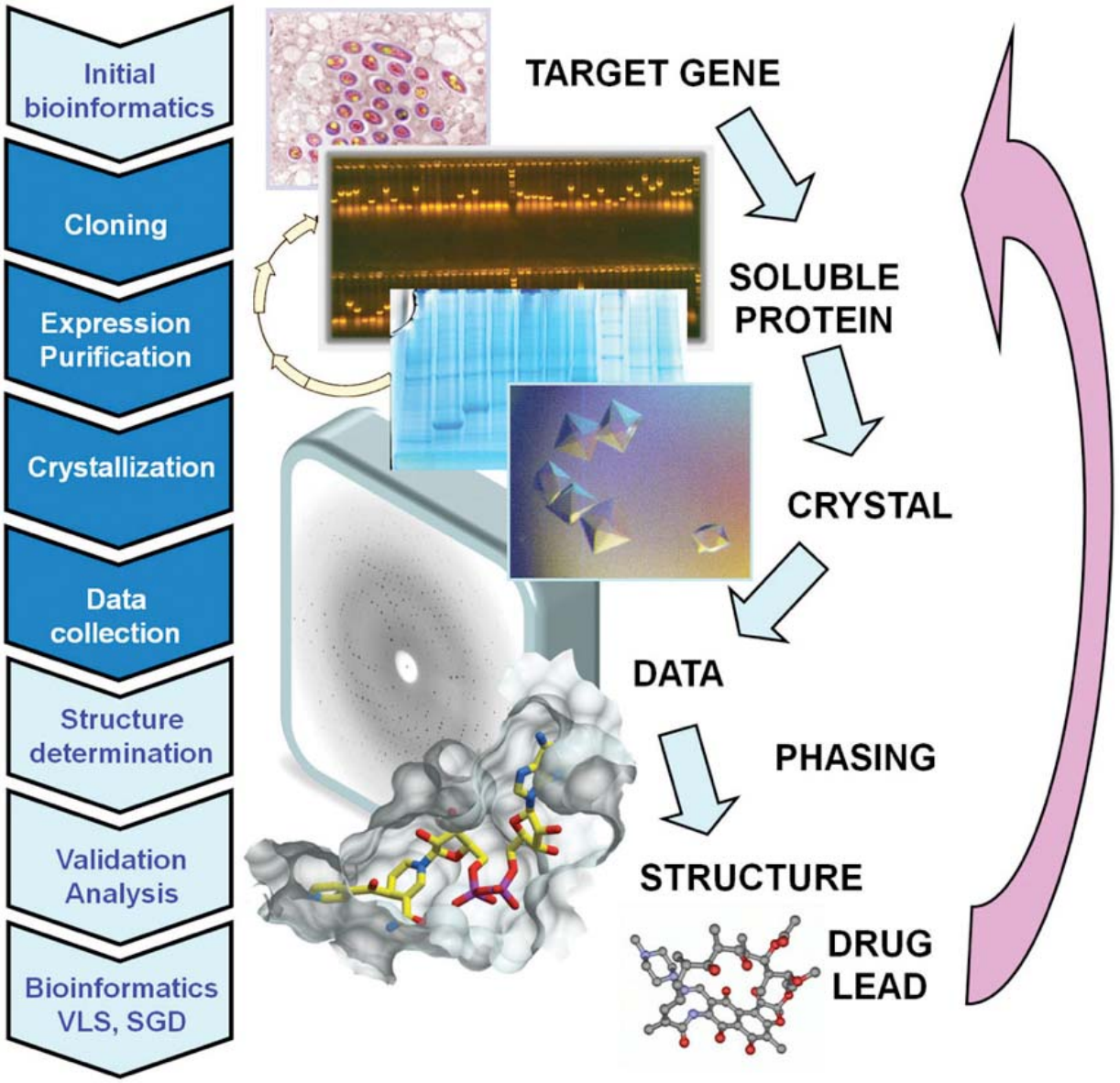
The crystallographic phase problem

The crystallography experiment

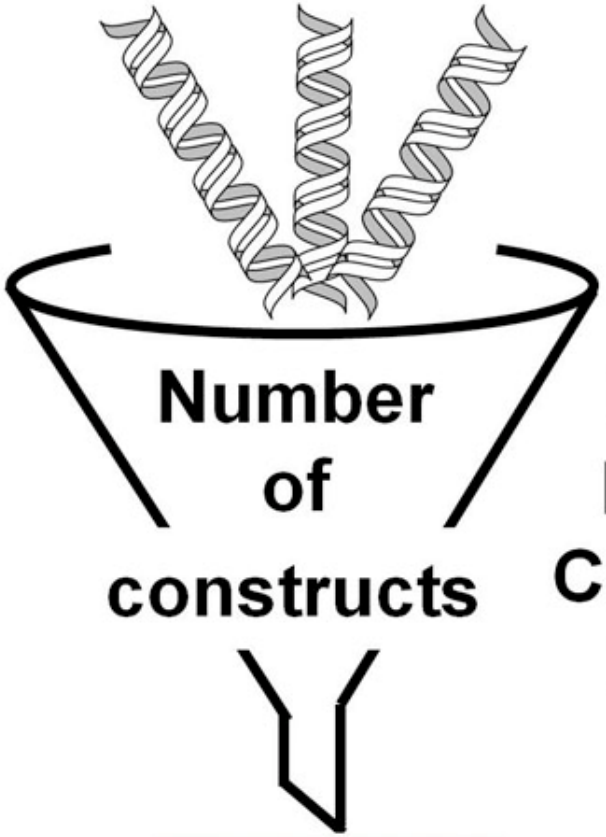


The X-ray crystallography experiment is like a light microscope except (1) use X-rays and not visible light and (2) we have no objective lens for X-rays so we cannot refocus to make the image. In other words while we can measure the amplitudes of the diffracted waves we cannot measure their relative phases. This is the "phase problem" of crystallography.

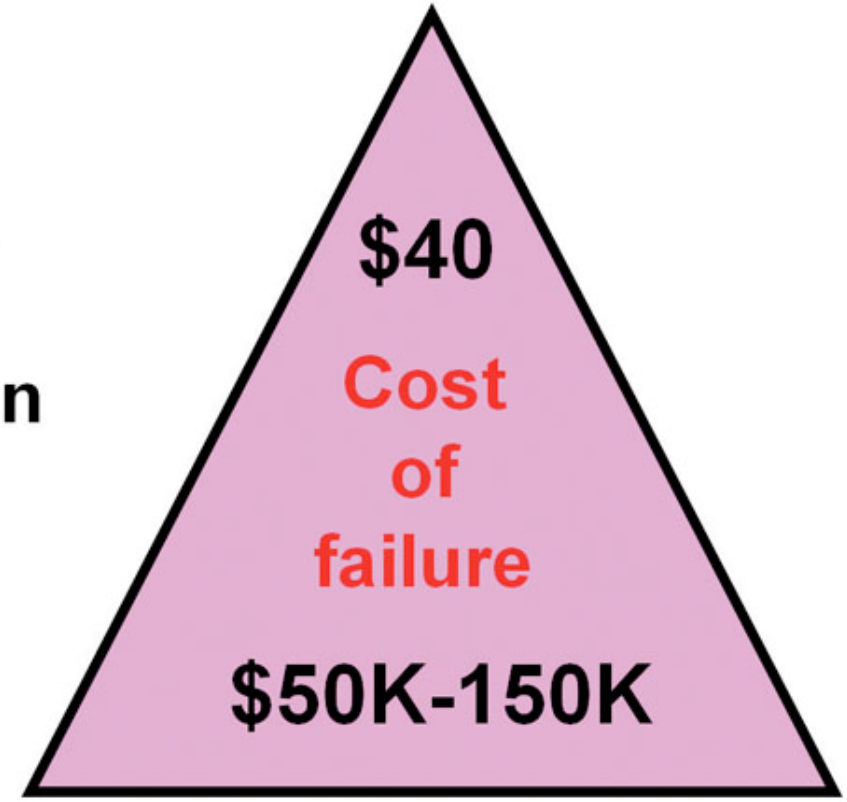
Overview of protein structure determination



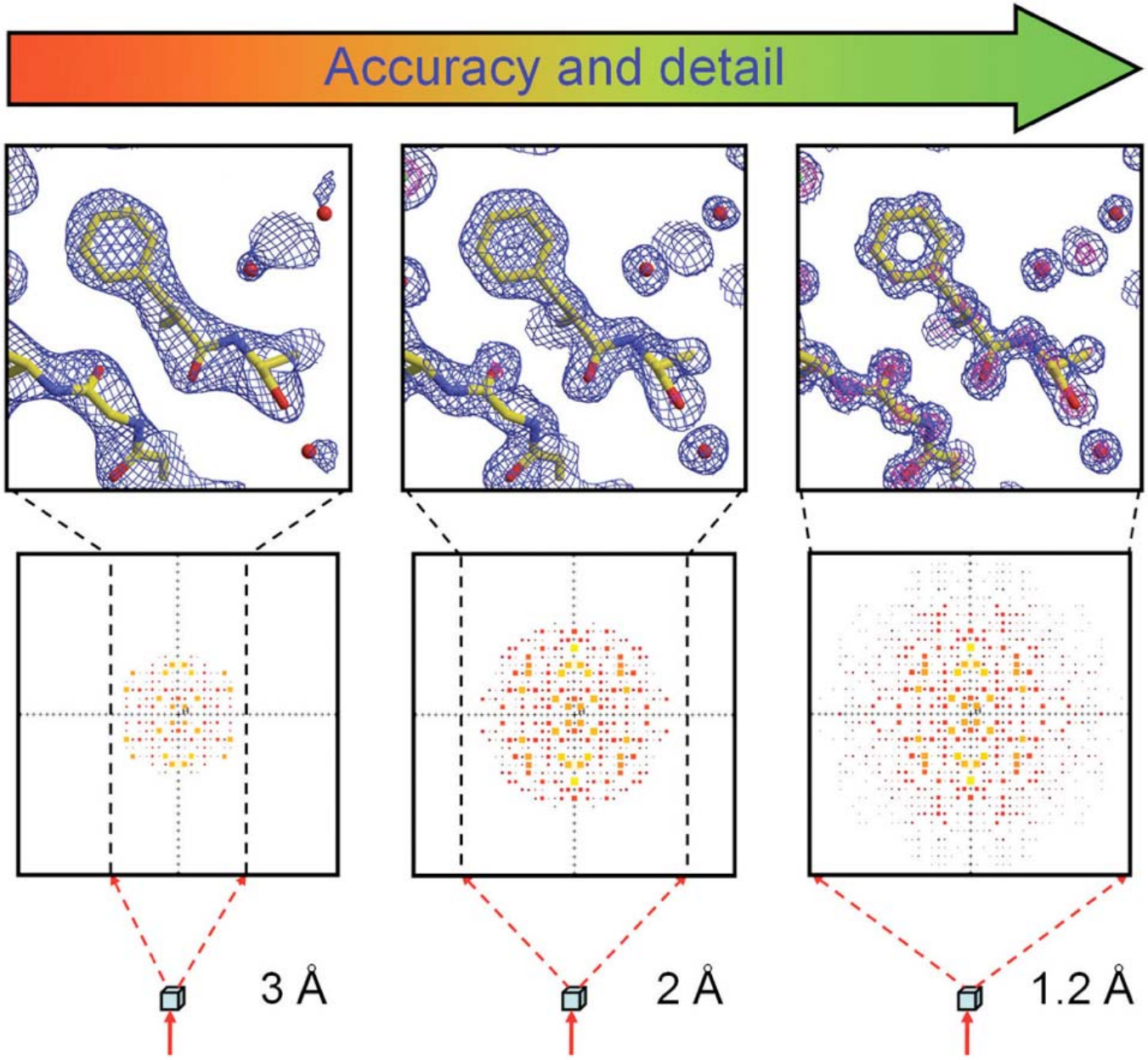
Life wasn't meant to be easy!



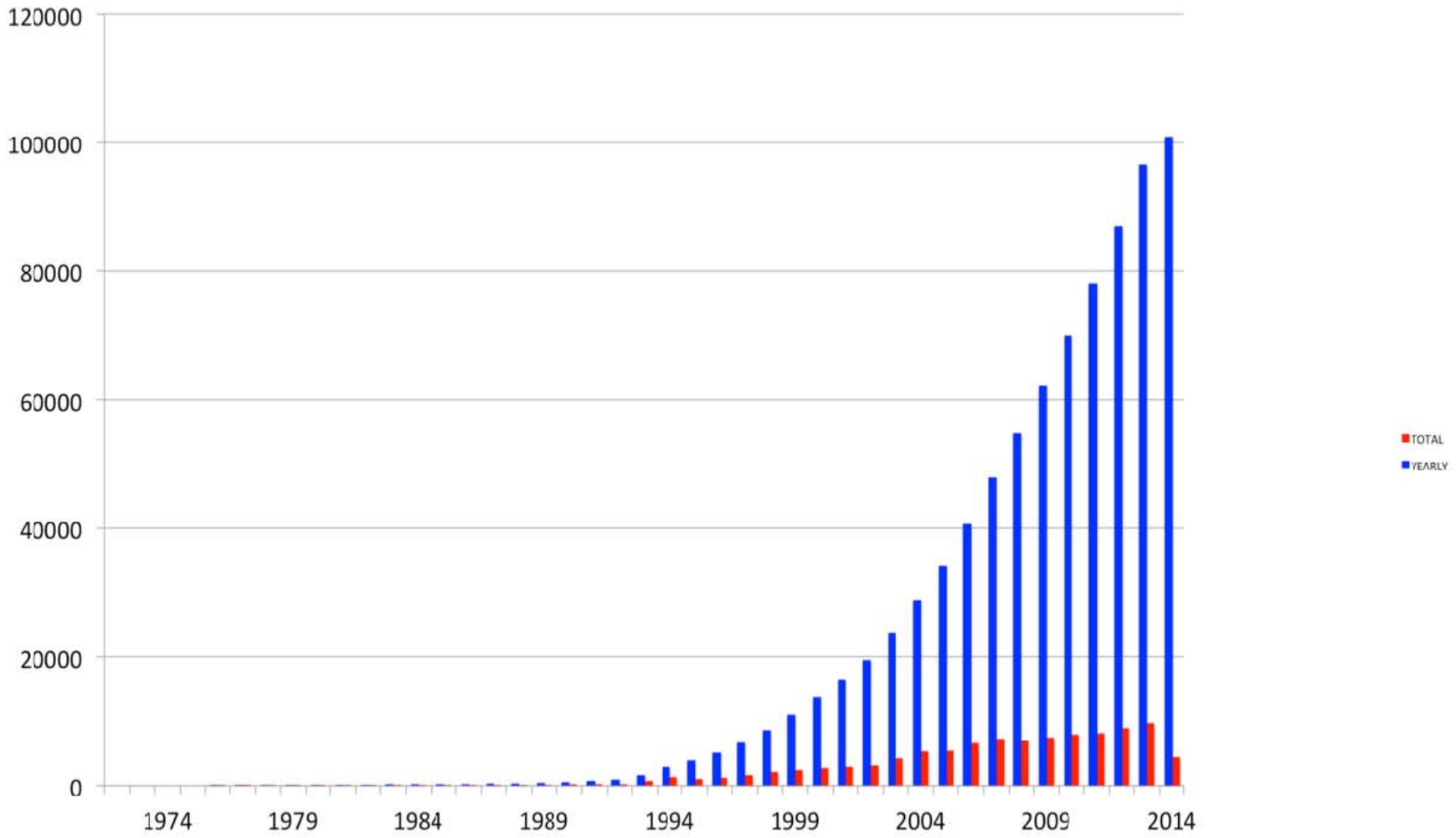
Cloning
Expression
Purification
Crystallization
Harvesting
Data
collection
Structure
determination



Importance of data quality (resolution)

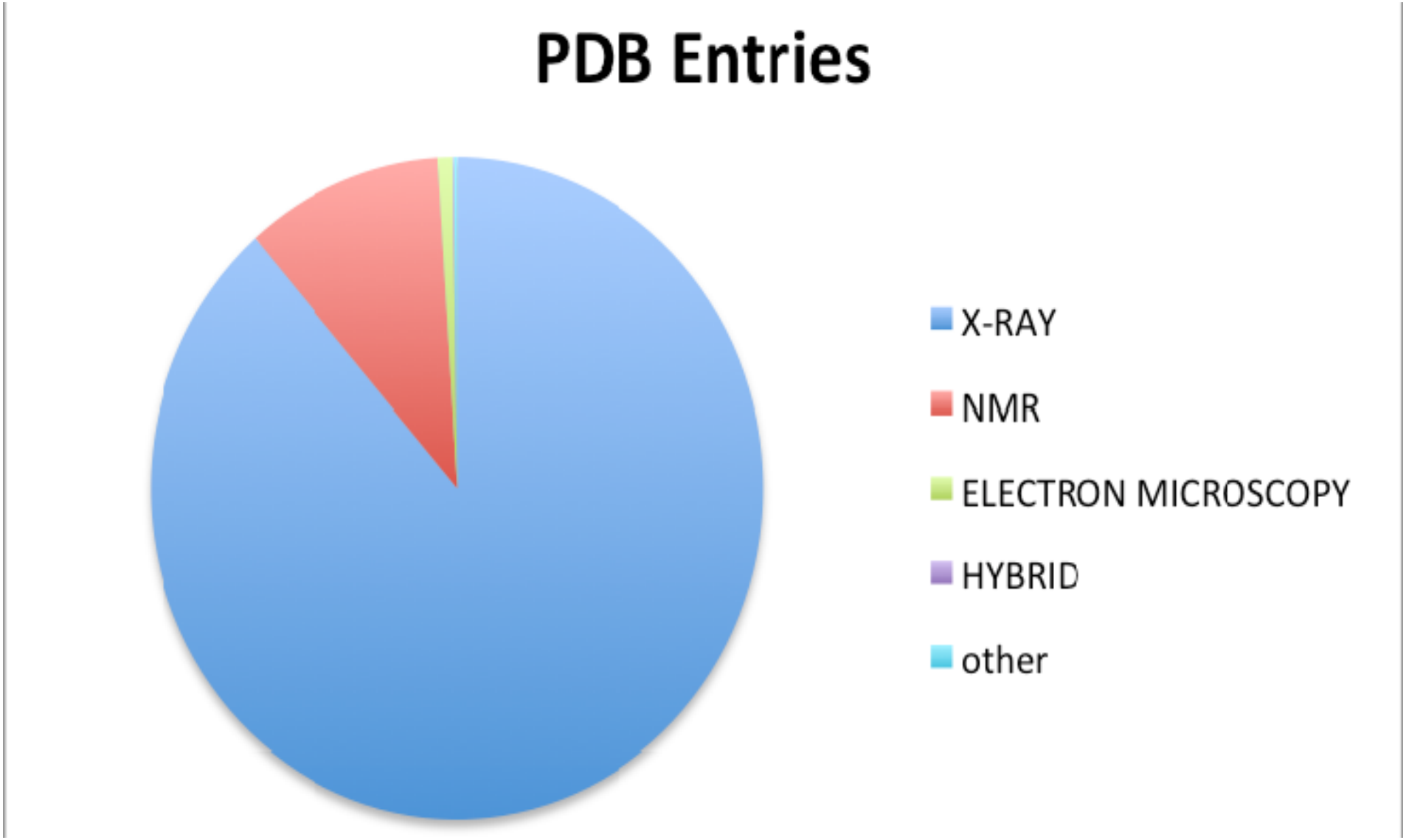


What has been achieved since 1962



Protein Data Bank (PDB) depositions by year

PDB depositions by technique



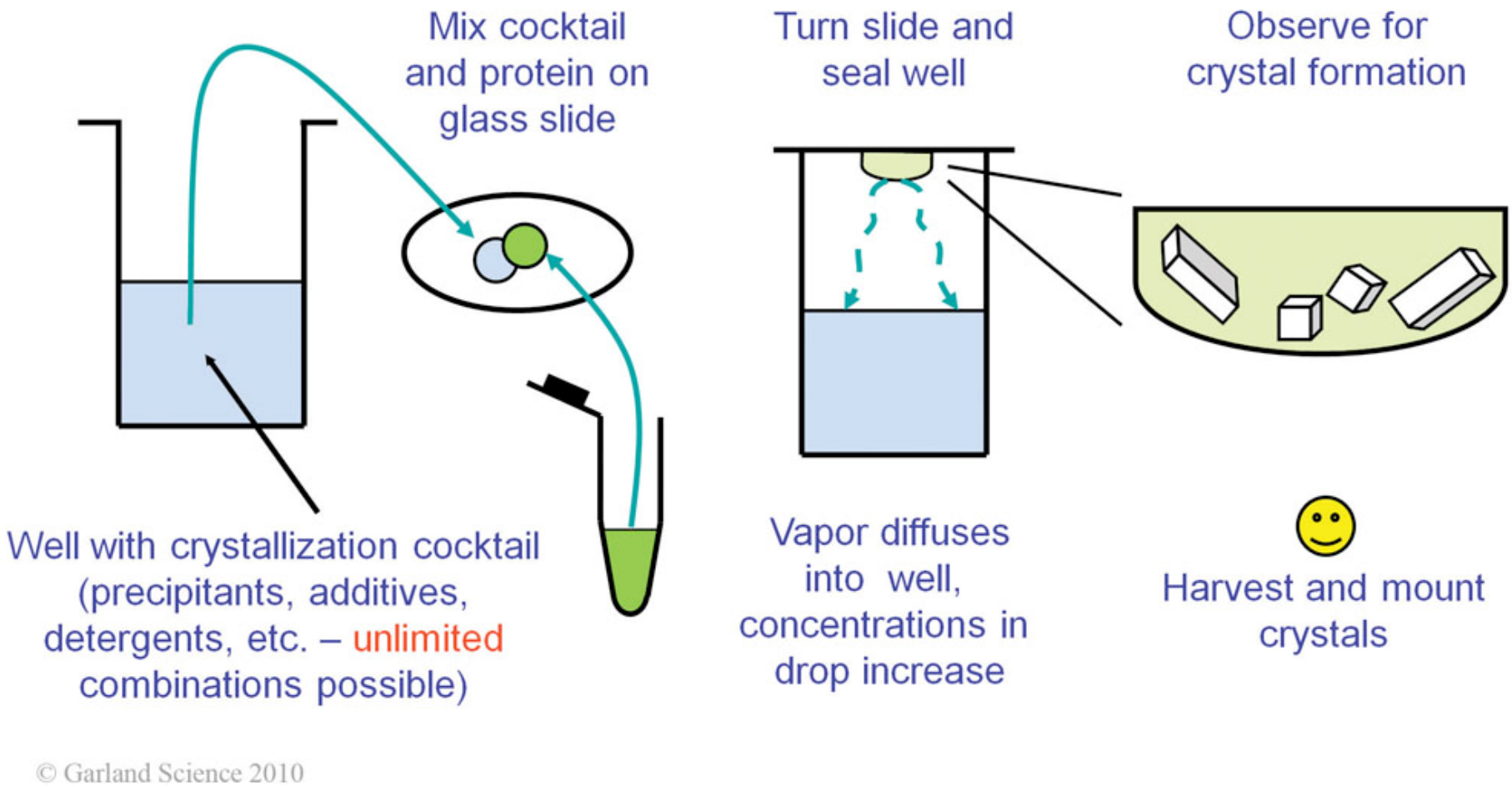
The experiment

1. Expression of proteins using molecular biology instead of isolation from natural sources – makes more and most importantly purer protein
2. Grow crystals – the most uncertain and often the most time consuming step – use of factorial screens and robotics
3. Record diffraction data – high speed detectors and intense synchrotron sources can reduce what once took months to minutes or even seconds
4. Solve the structure – determine the phases. New methods including anomalous phasing have significantly improved this step although it can still be difficult in some cases
5. Build a model – aided by highly intelligent computer graphics
6. Refine the structure – optimise the model to fit the data. Fast computers have reduced this step from hours to seconds
7. Deposit the structure in the Protein Data Bank, and;
8. Write a paper – often the slowest step

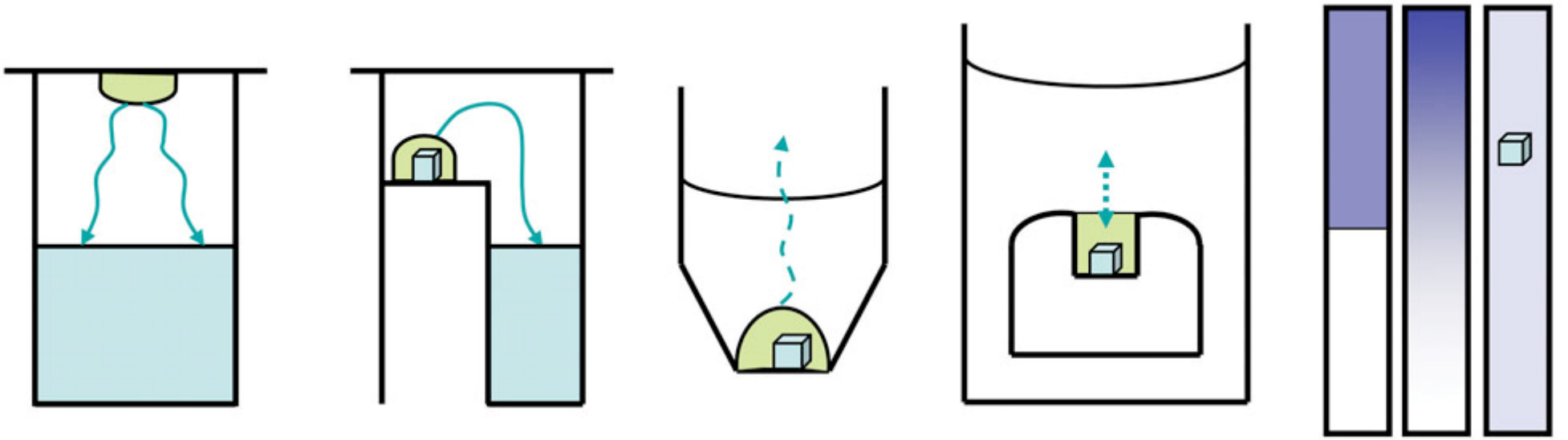
Introduction to:

CRYSTALLISATION

Vapour diffusion – in theory!



Different apparatus



The classic: hanging-drop vapor diffusion

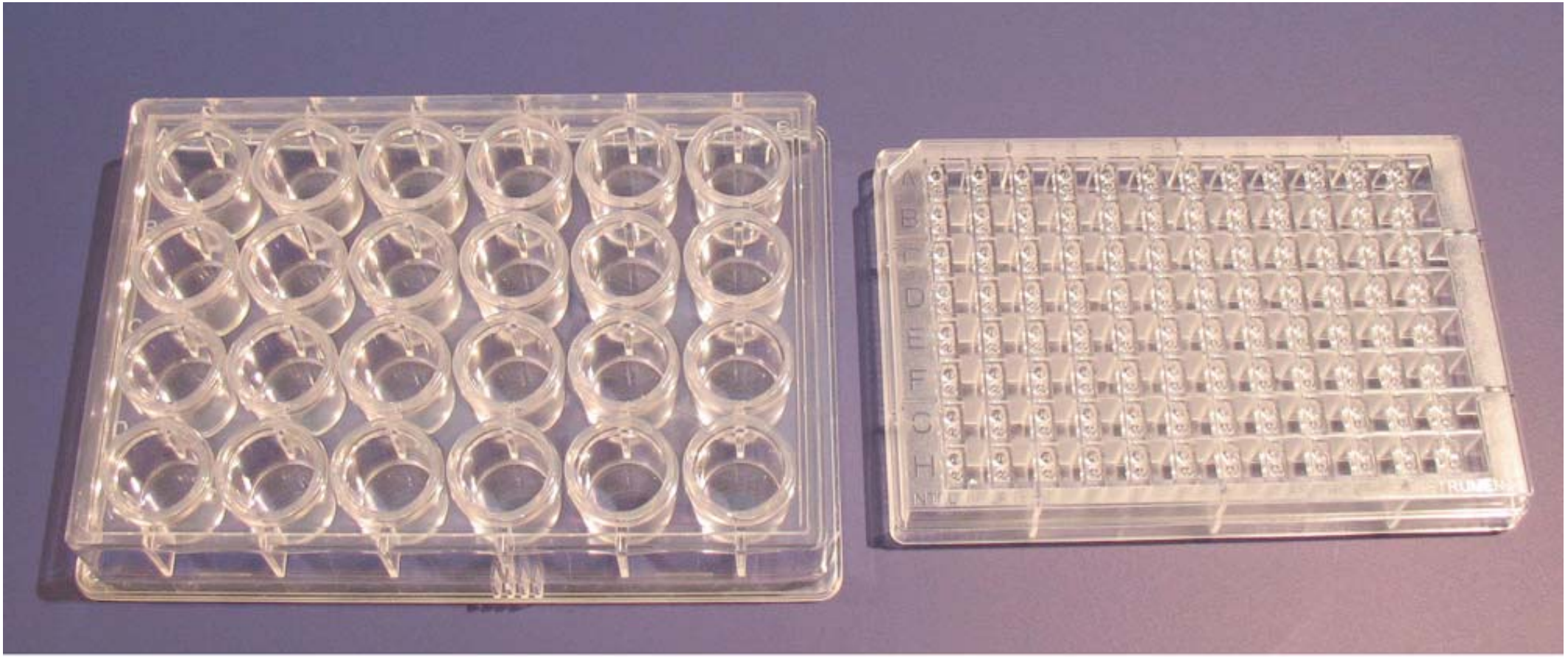
The variant: sitting-drop vapor diffusion

Micro-batch under oil

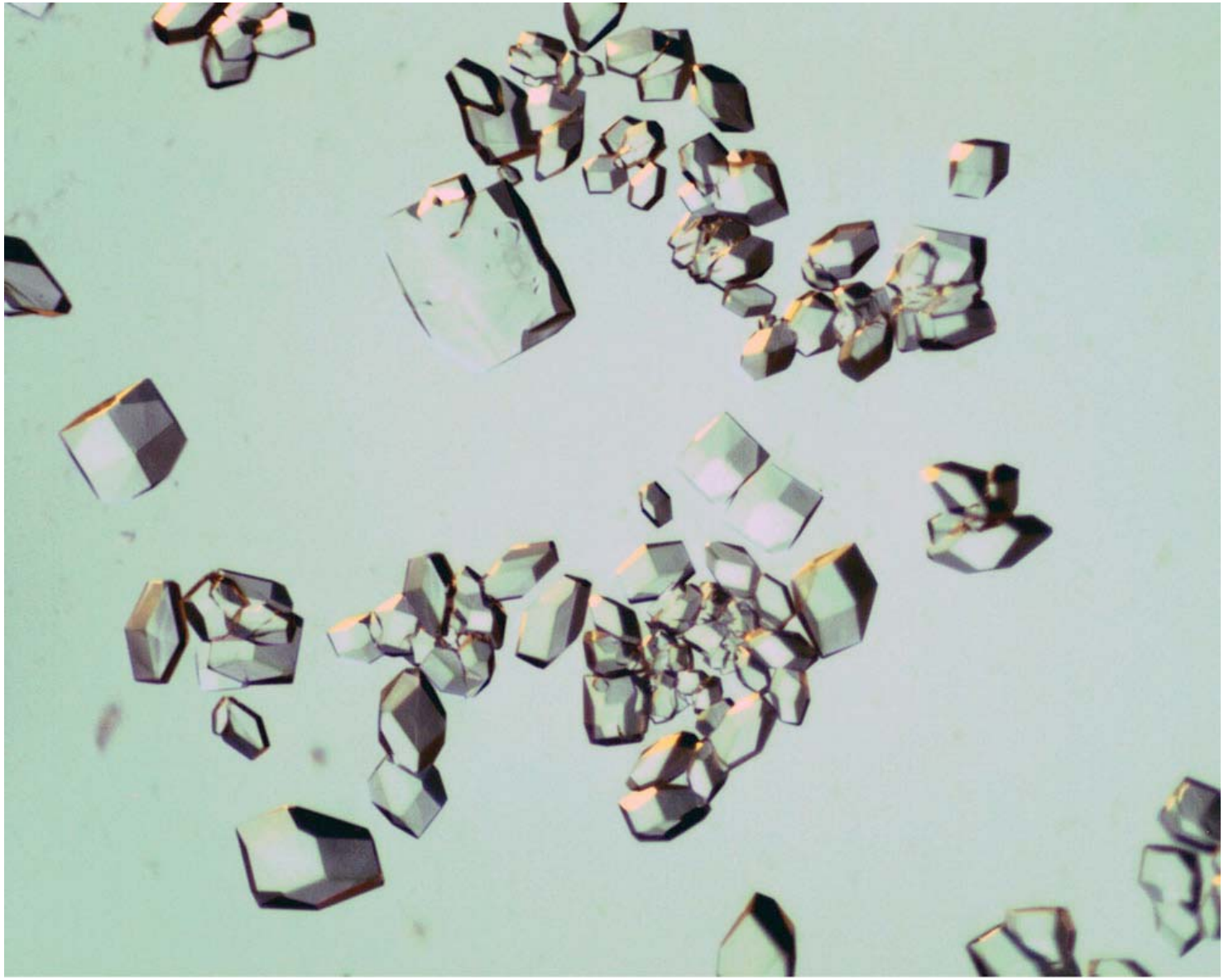
Micro-dialysis

Free-interface diffusion

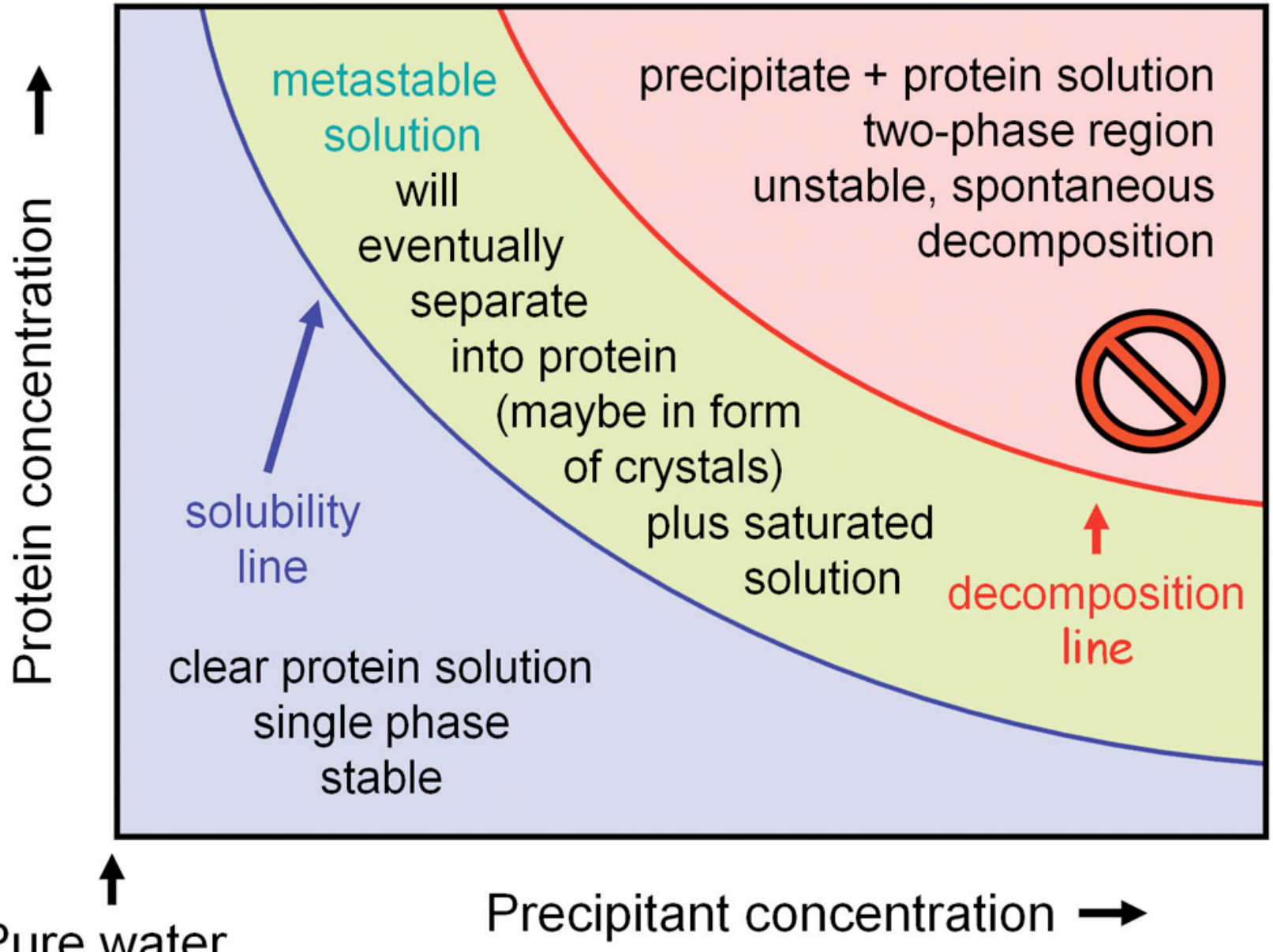
Crystallisation plates – for vapour diffusion



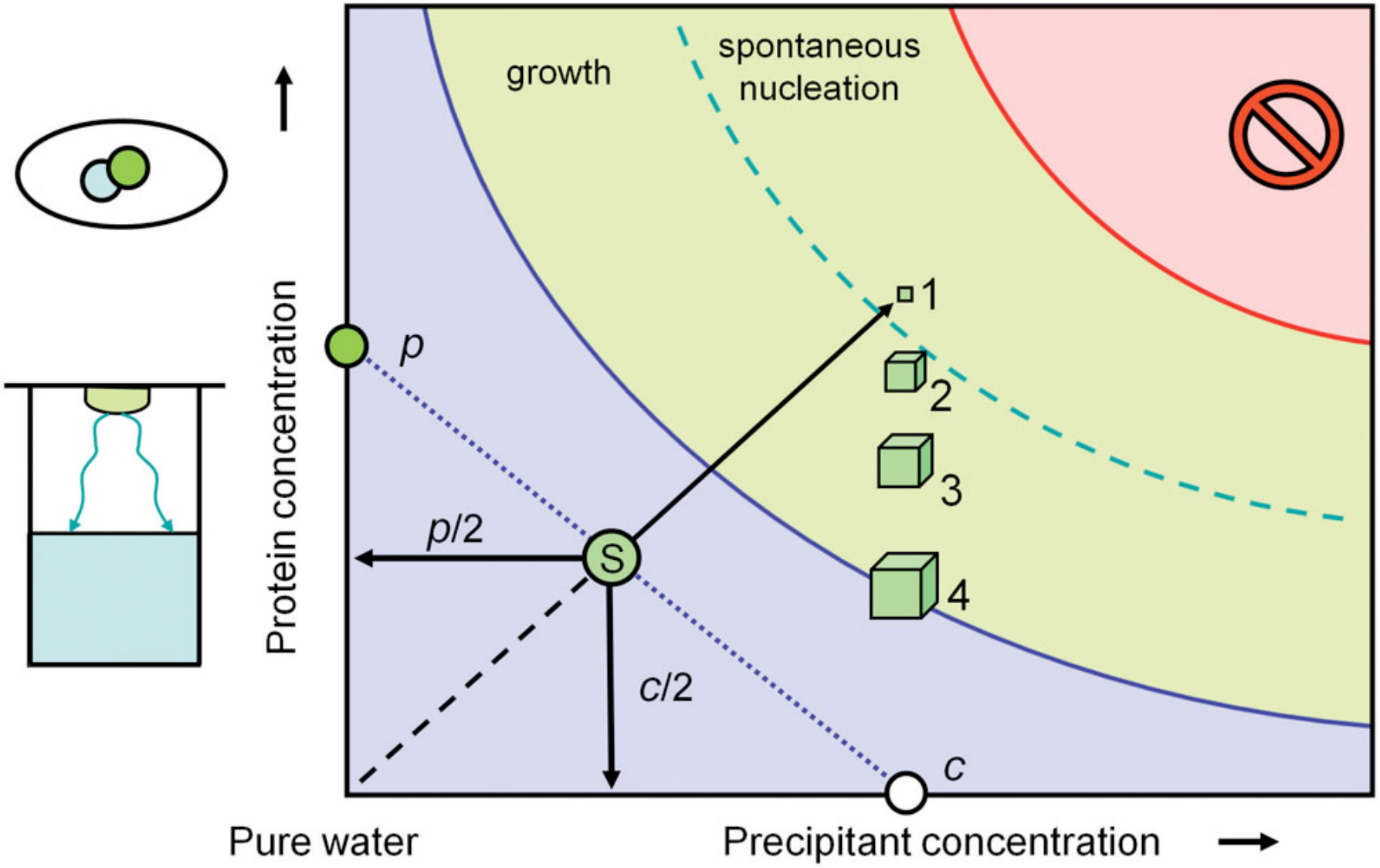
What you are looking for!



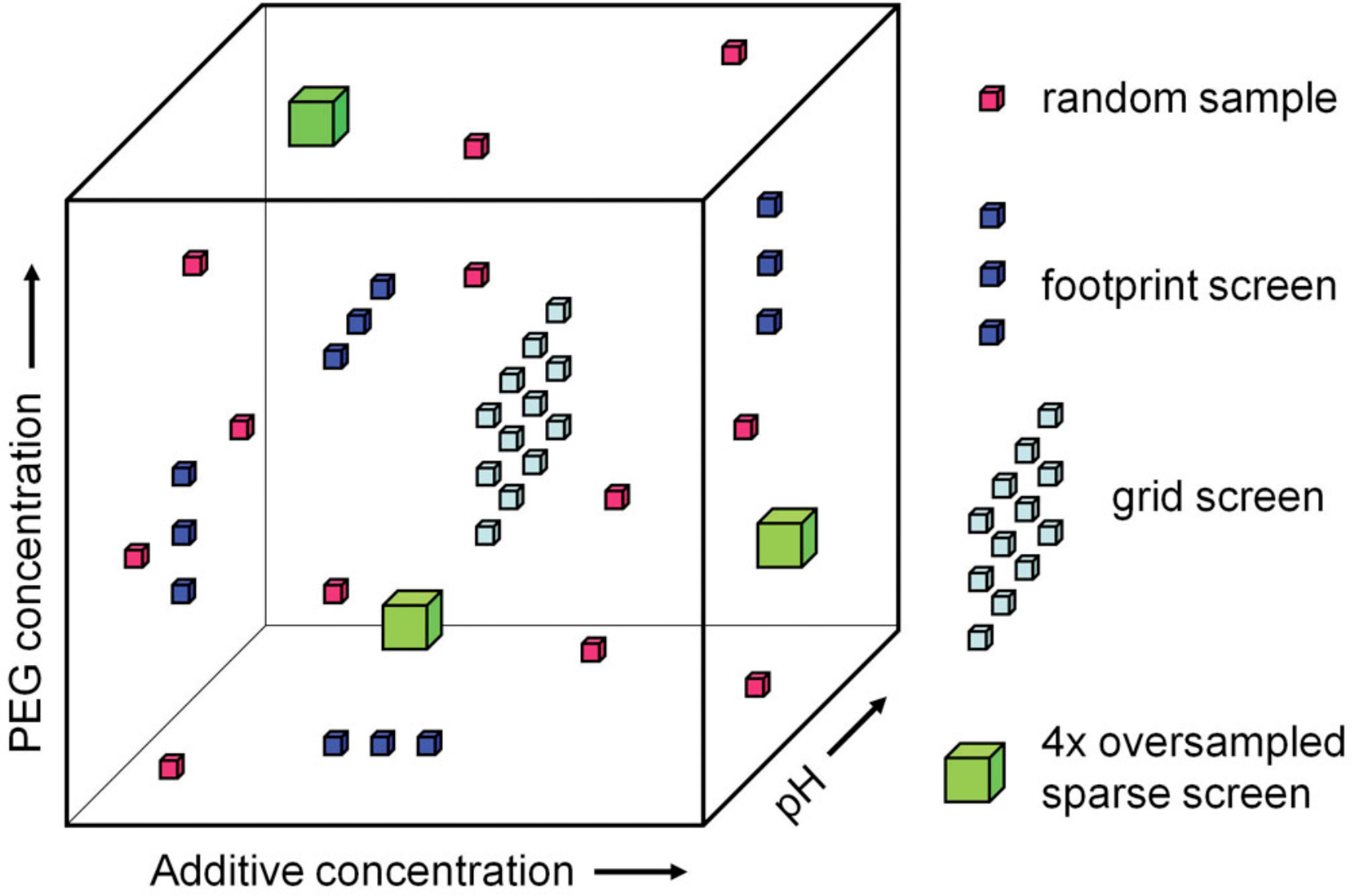
The basic crystallisation diagram



Successful vapour diffusion experiment

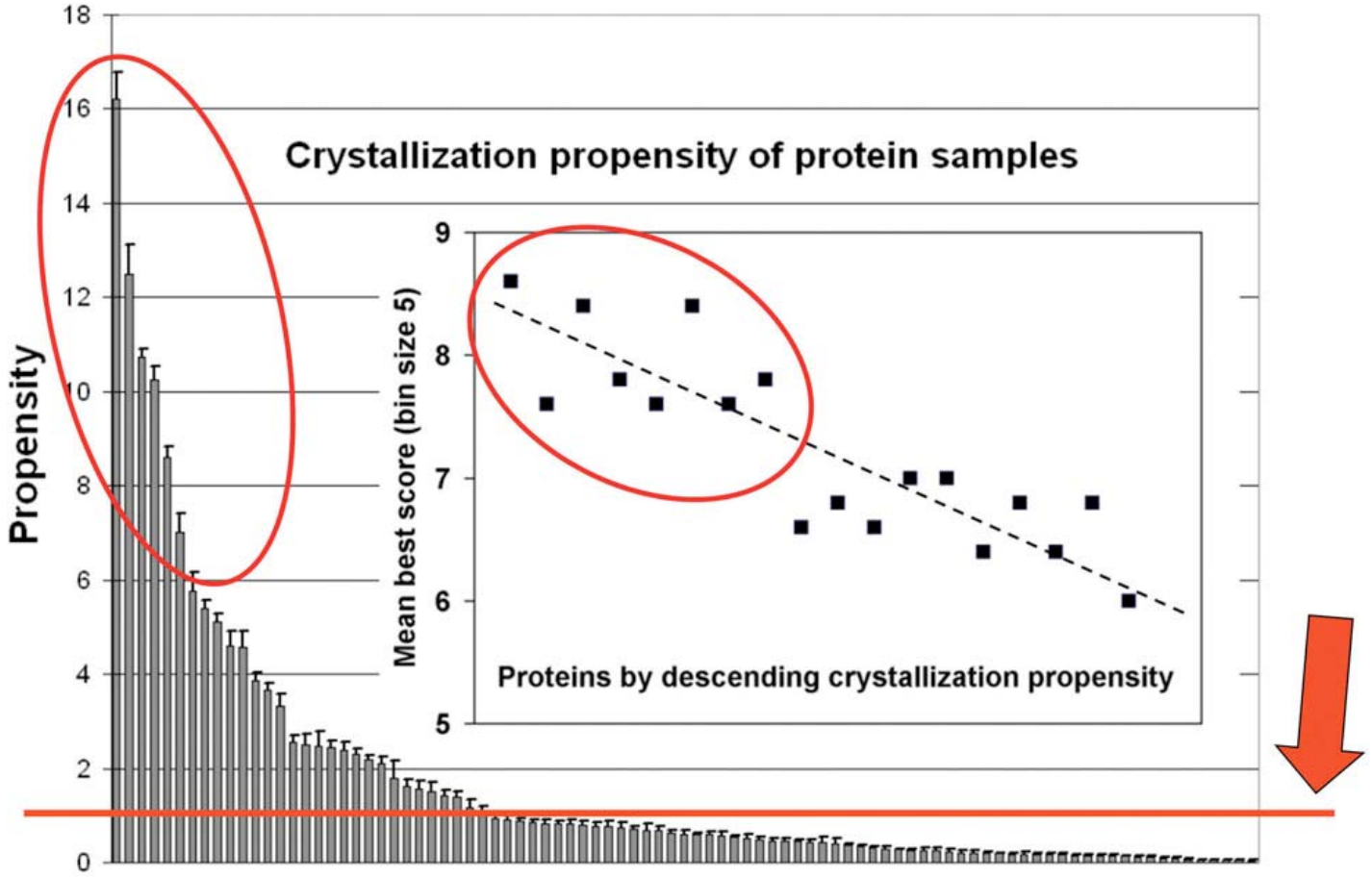


Screening – can't do everything



Can you predict whether a protein will crystallise

The xtalpred server: http://ffas.burnham.org/XtalPred-cgi/xtal.pl



"More efficient to improve the inherent crystallisability than to do further screening" Rupp, 2010

Introduction:

DATA COLLECTION (SOURCES)

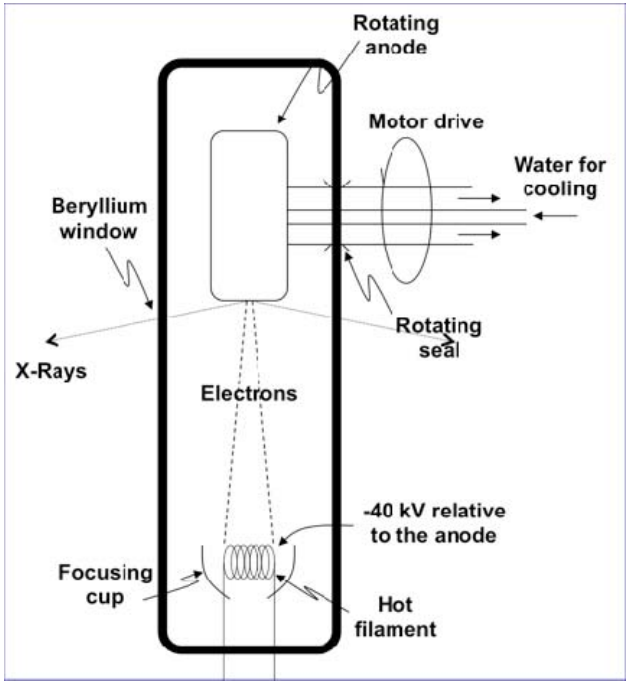
Laboratory X-ray sources

Table 1
Approximate X-ray beam brilliance for the main types of in-house sources with optics.

System	Power (W)	Actual spot on anode (μm)	Apparent spot on anode (μm)	Brilliance (photons s^{-1} mm^{-2} $mrad^{-1}$)
Standard sealed tube	2000	10000 × 1000	1000 × 1000	0.1×10^9
Standard rotating-anode generator	3000	3000 × 300	300 × 300	0.6×10^9
Microfocus sealed tube	50	150 × 30	30 × 30	2.0×10^9
Microfocus rotating-anode generator	1200	700 × 70	70 × 70	6.0×10^9
State-of-the-art microfocus rotating-anode generator	2500	800 × 80	80 × 80	12×10^9
Excillum JXS-D1-200	200	20 × 20	20 × 20	26×10^9

Beamsize at the sample 100 - 200μ depending on optics

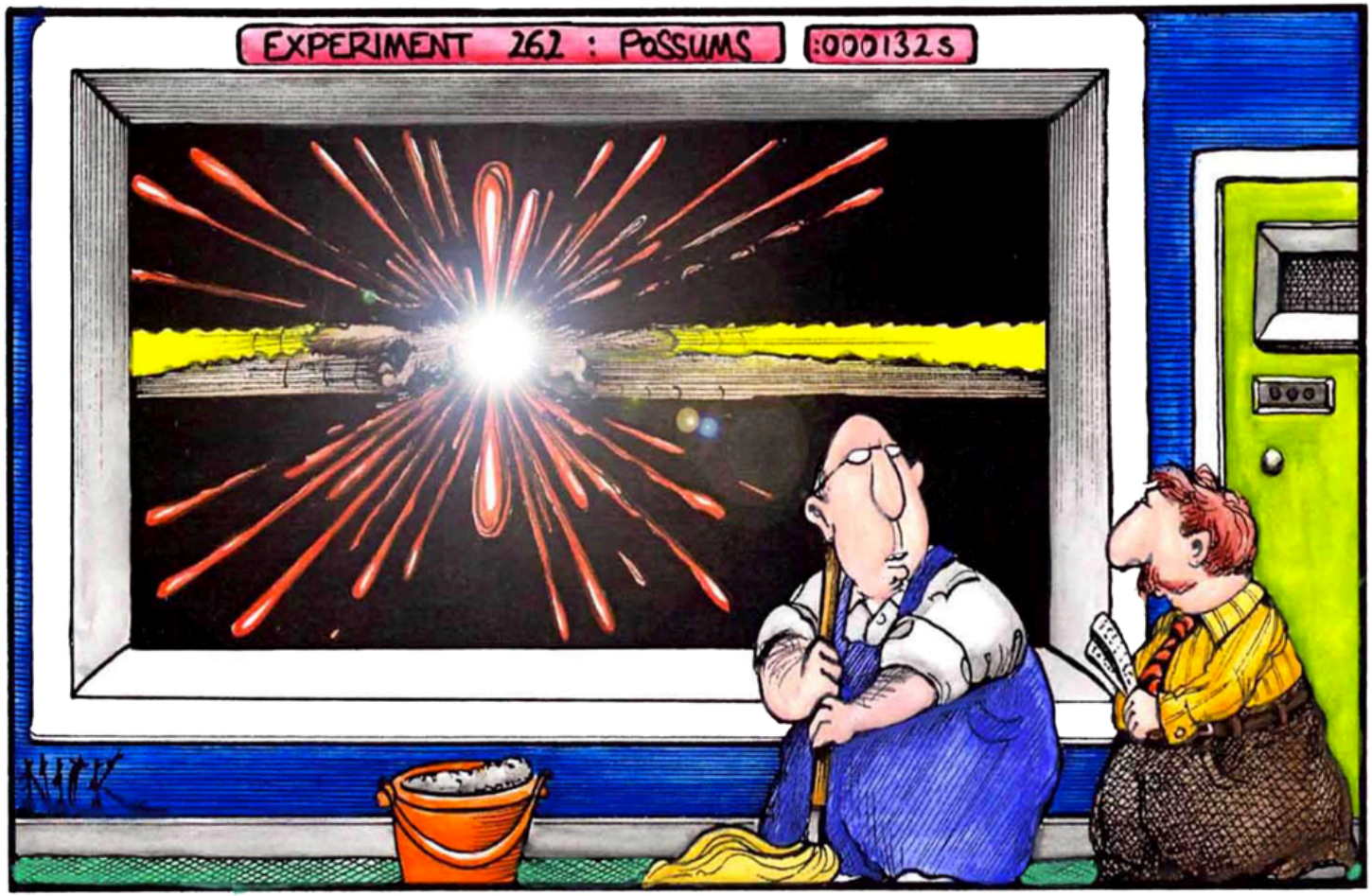
In the 1960s and 1970s typical problems were considered insoluble unless crystals at least 300μ could be grown. Even with modern optics most protein crystals suitable for a laboratory source are > 100μ.



Synchrotron radiation and protein crystallography

- Brilliance
  - Use of small crystals; rapid data collection
- Low Divergence
  - Resolve close reflections from large unit cells e.g. viruses
- Tunability
  - Enhance anomalous signal for phasing
- Pulsed beam
  - Perform time-resolved experiments on enzymes

Biologists were not always welcome at synchrotrons



"Sure been a heap more work for ME around here since those Biologists got granted research time on the ol' Synchrotron ."

Protein crystallography at the Australian Synchrotron

- Two beamlines
  1) Bending magnet, high throughput, MAD phasing, small molecule crystallography
  2) Undulator, high brilliance, small crystals
- Extensive use of remote access
  - large country, only one synchrotron

Protein crystallography endstations

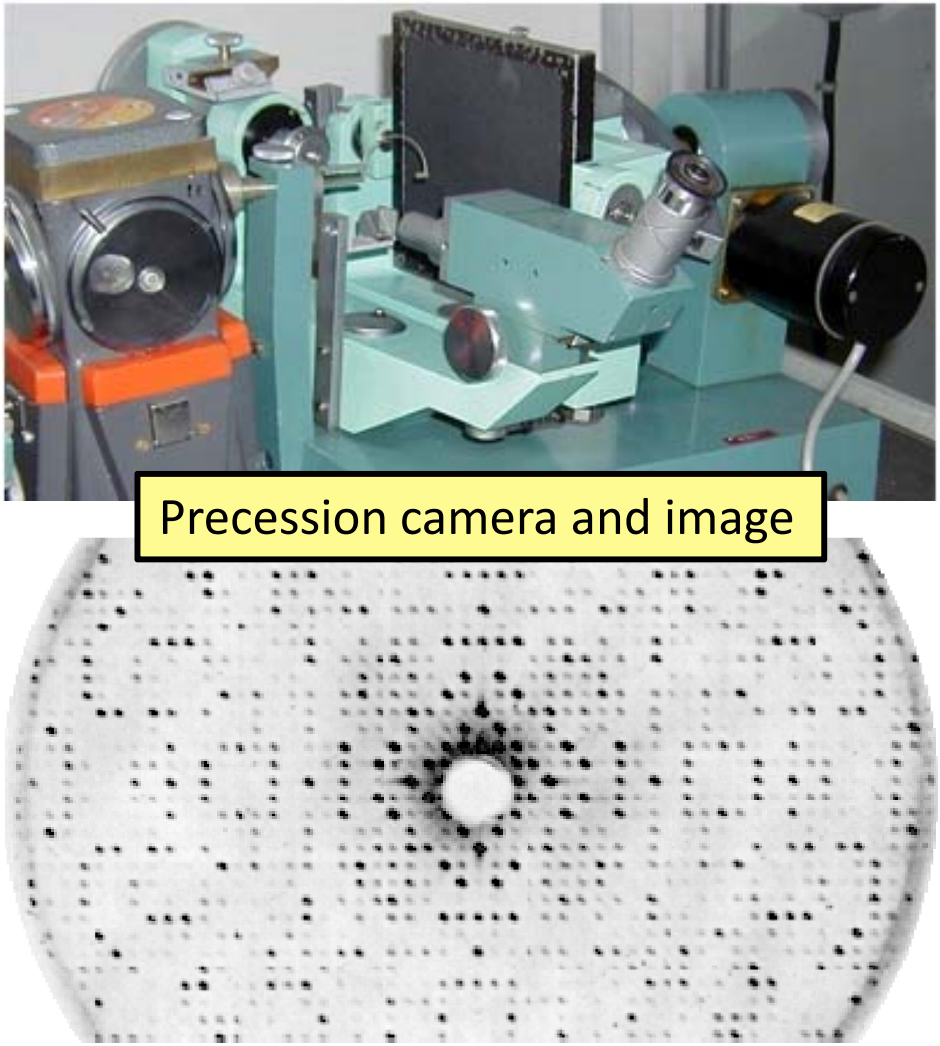
- Robot
- CCD detector
  - Would like Dectris!
- Overhead frame
  - Easy access
  - But less stable
- On-line software
  - Processing on-the-fly allows for rapid evaluation of crystals

Introduction:

DATA COLLECTION (DETECTORS)

Detectors - film

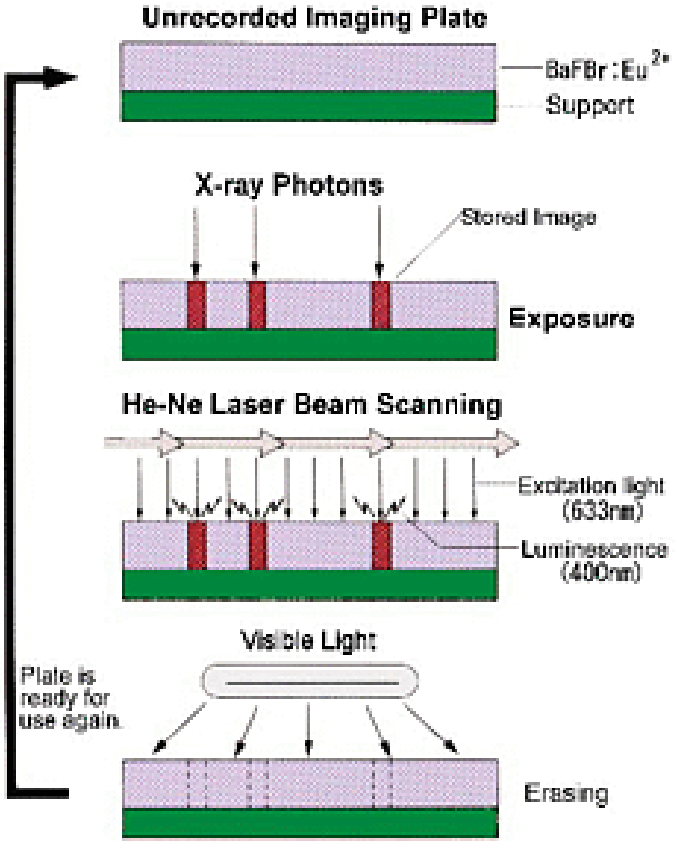
- Film
  - Cheap and linear response – no longer used except for teaching
  - Requires use of darkroom
  - Requires separate scanner to digitize the images
  - Can provide overall view illustrating the crystal symmetry



Precession camera and image

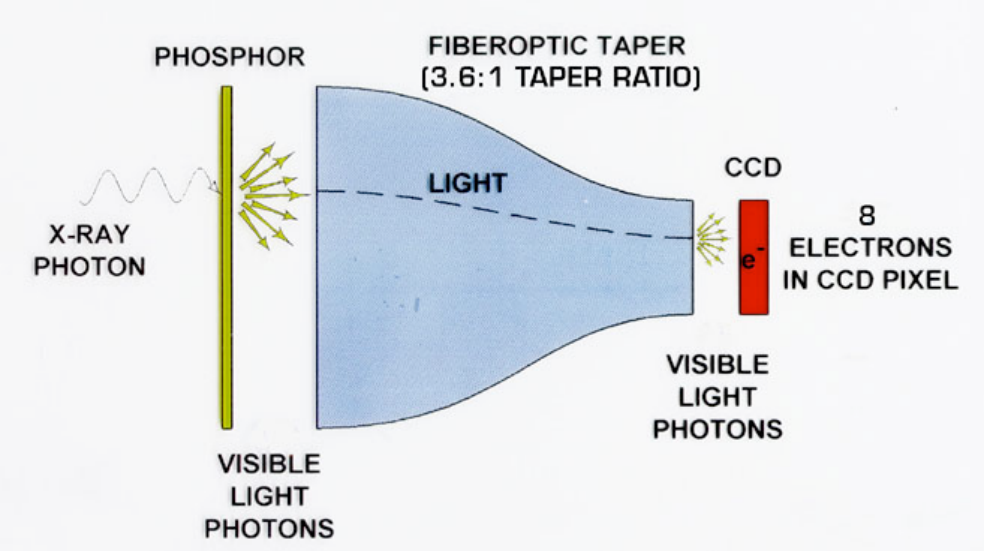
Detectors – image plates

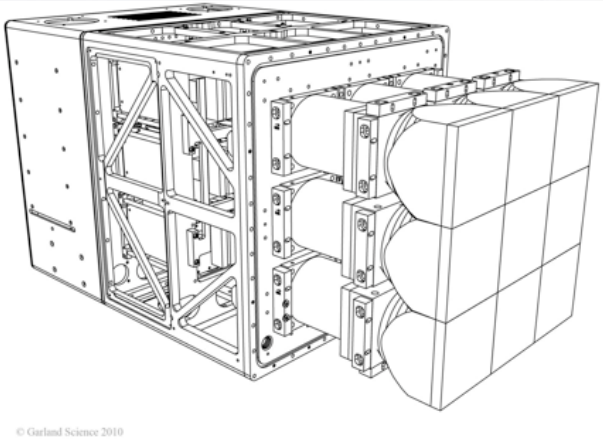
- Image plates
  - Essentially a reuseable plastic film
    - Read out by laser that detects excited centres created by X-rays
    - Erased by light
  - Relatively cheap for large area. Wide dynamic range. Low noise.
  - Relatively slow readout (about 120 s).
    - Restricts usefulness at a synchrotron where an exposure is about 1s.
    - Still used at home sources where exposures are 1-10 minutes



Detectors - CCD

- CCD
  - CCD. Need to be tiled for large area. Requiring cooling to reduce noise. Relatively fast readout (< 1s). Need to demagnify the image to match CCD size.
  - Relatively expensive. Need careful calibration.
  - Physically large

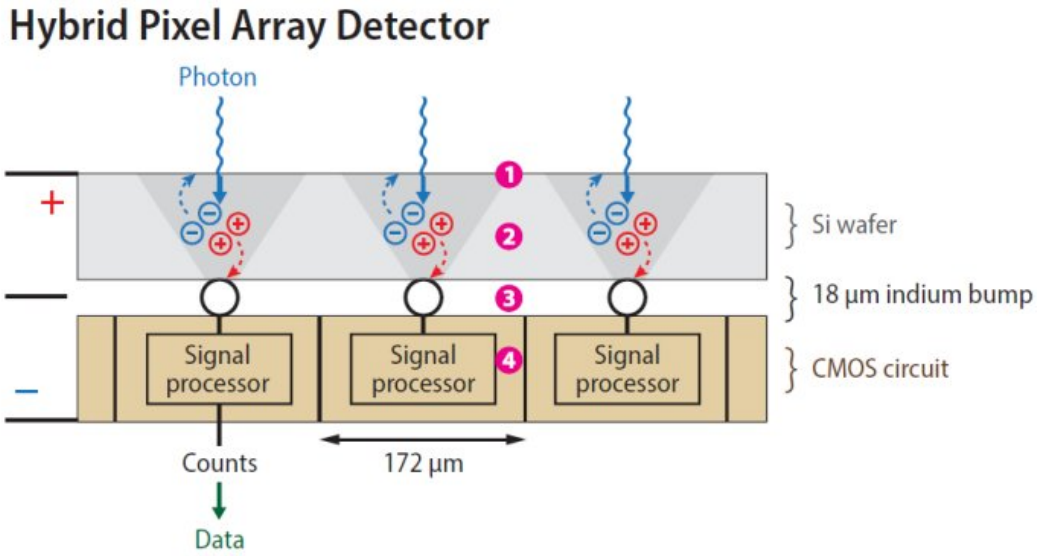


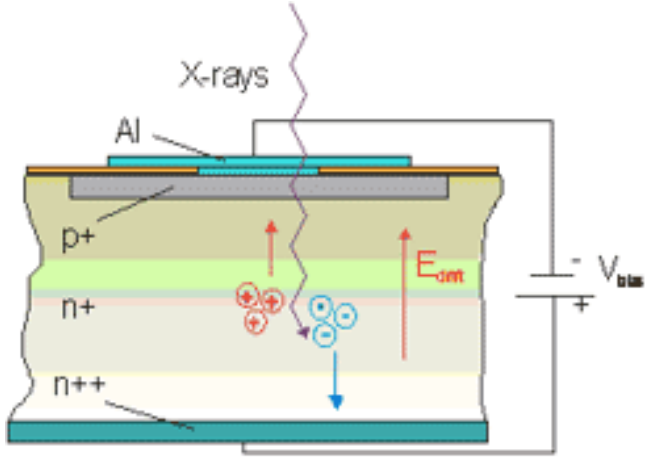


ADSC CCD detector

Detectors – pixel arrays

- Pixel arrays
  - Can tile many modules.
  - Virtually zero noise.
  - Operates at room temperature.
  - Ultrafast readout permits continuous exposure.
  - Very expensive – early adoption of new technology, not inherent.





Introduction:

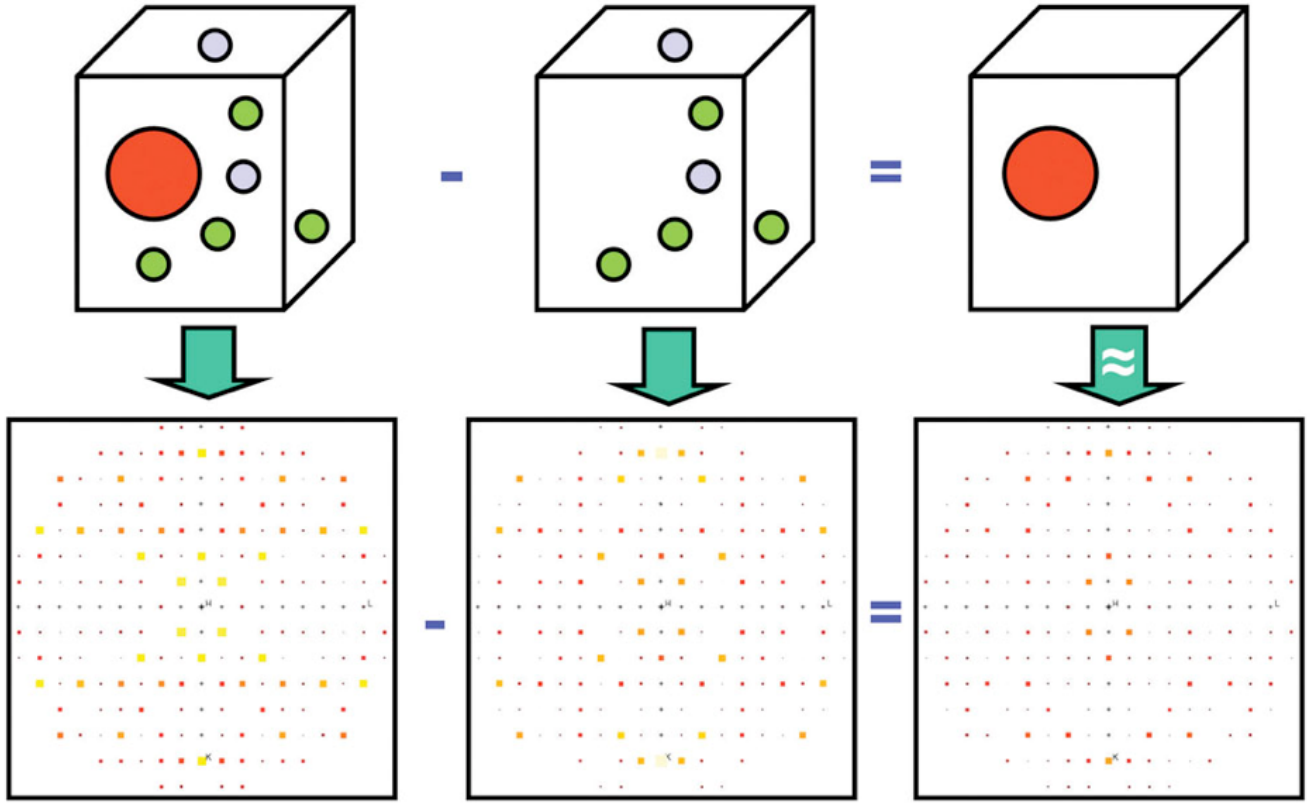
STRUCTURE SOLUTION AND REFINEMENT

Solving the phase problem

- **Marker atom substructure methods** do not depend on prior structural knowledge about the protein (other than the sequence and the structures of the amino acids). Once the marker atoms (heavy metals, selenium) have been located by Patterson or direct methods starting phases for the modelling can be calculated. These are *de novo* or experimental phasing methods.
- **Density modification** are powerful methods for *improving* the initial phases and are used in practically all *de novo* phasing experiments. These methods included solvent flattening, and non-crystallographic symmetry averaging. As they do not require any model information they may also be considered experimental techniques.
- **Molecular replacement** require a similar structure as a molecular search probe. *Replacement* is to be understood as the placing of the search model in the crystal structure not as "substitution". Phase bias can be a serious problem with this technique.
- **Direct methods** exploit the fact that relationships exist between certain sets of structure factors. They require very high resolution (1.2 Å or better) for *ab initio* determination of a protein structure and have been limited to very small proteins. They are commonly used for small molecule structures. They are, however, very important in sub-structure (marker atom) determination.

The concept of isomorphous difference

If we add an atom to a protein crystal and change nothing else (isomorphous) then the difference in the diffraction patterns of the “derivative” structure with added heavy atom and the “native” protein structure approximates the diffraction pattern from the heavy atom alone.



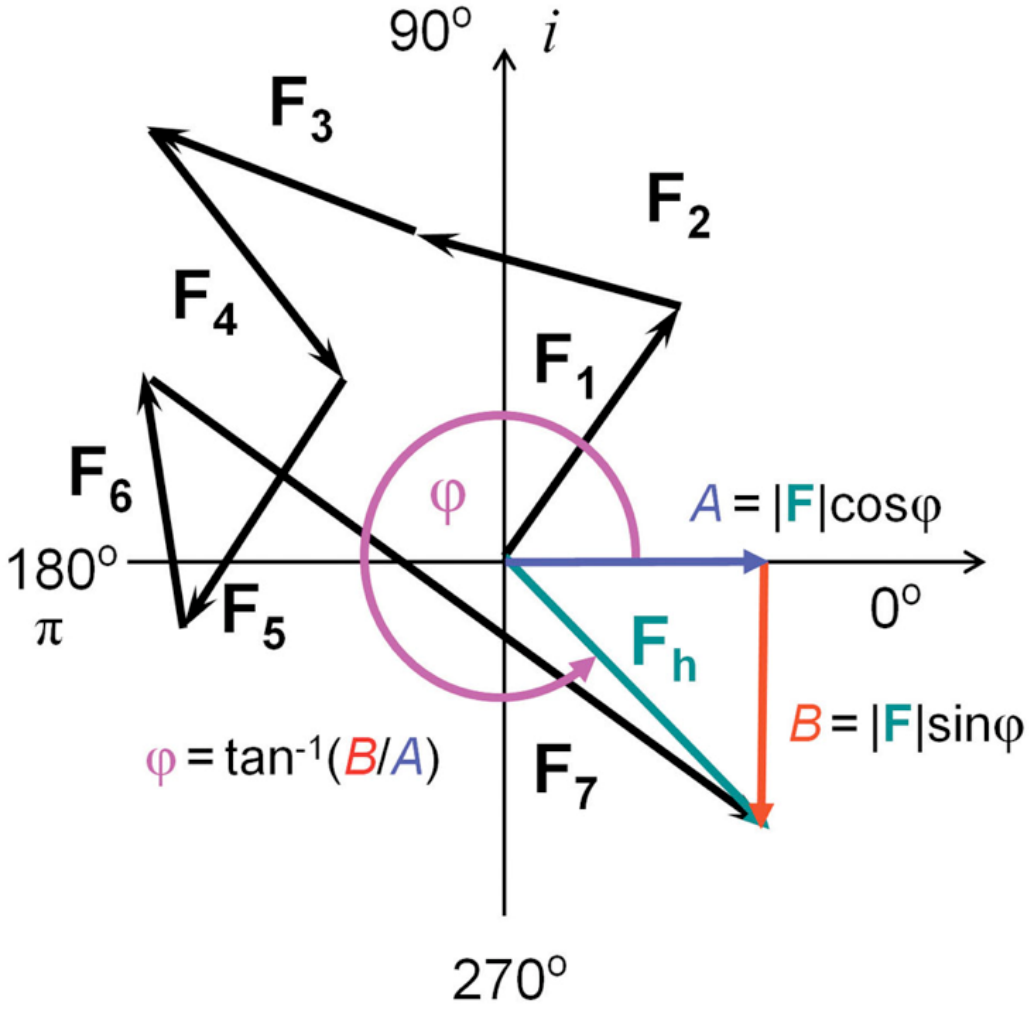
$\mathbf{F}_{PA} = \mathbf{F}_P + \mathbf{F}_A$ and thus

$\mathbf{F}_A = \mathbf{F}_P - \mathbf{F}_A$ but these are vectors and the amplitudes (things we can measure) don’t add up: $F_A \neq F_{PA} - F_P$ but we can still use the differences to solve the sub-structure.

Revision – graphical construction of the structure factor

Graphical illustration of the summation that yields one structure factor for a 7-atom structure – with 6 relatively light atoms and one heavier atom. The resultant structure factor is $\mathbf{F}_h$.

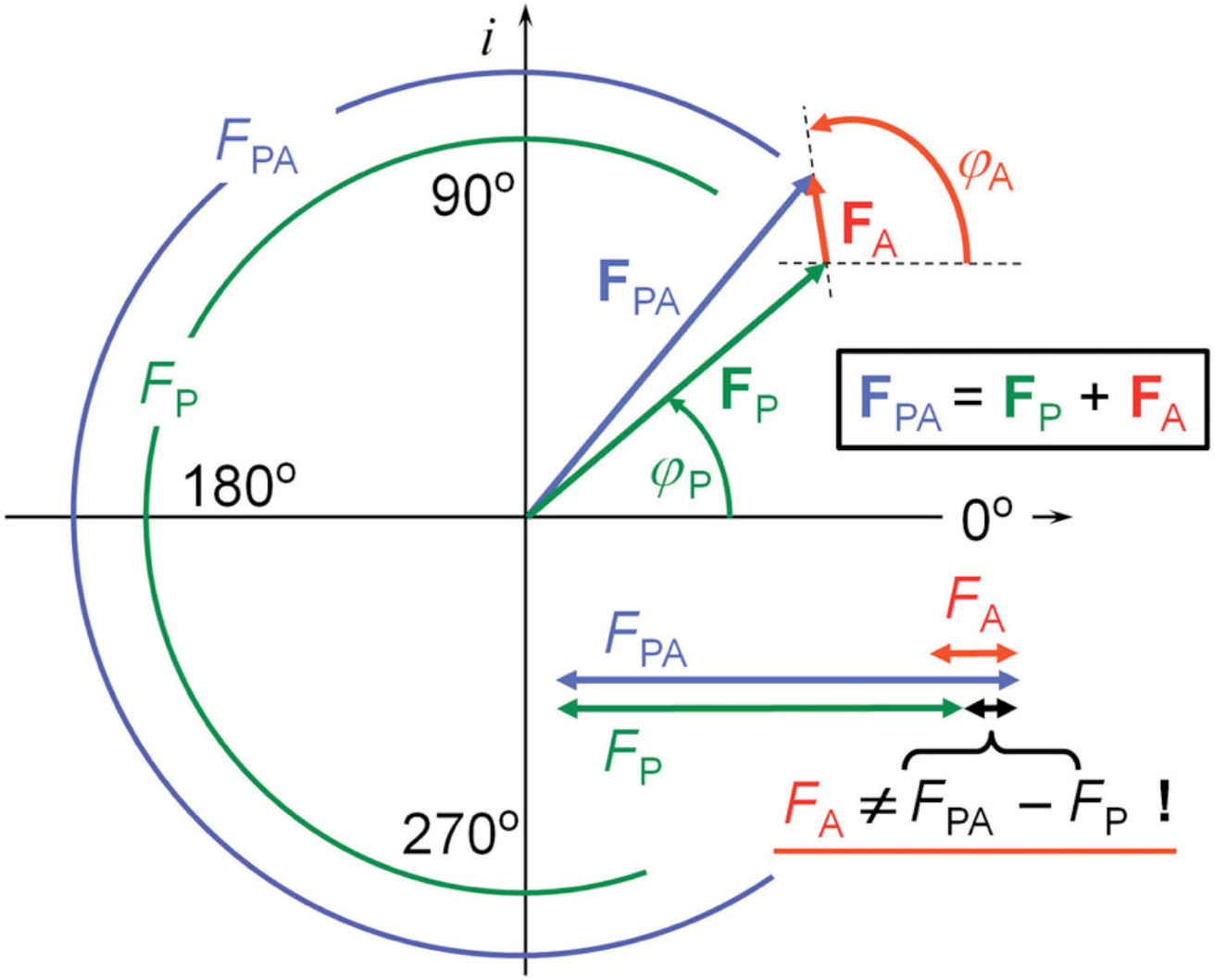
$$F_h = \sum_{j=1}^{atoms} f_{hj}^{0}.\exp(2\pi i h x_j)$$



The “true” isomorphous differences

When computing the total structure factor the contributions from the individual atoms are summed *vectorially.* Thus, if we add one more atom to a protein we simply add its structure factor to that of the total for the protein.

The radii of the circles are the amplitudes for the protein and heavy-atom derivative structure factors.

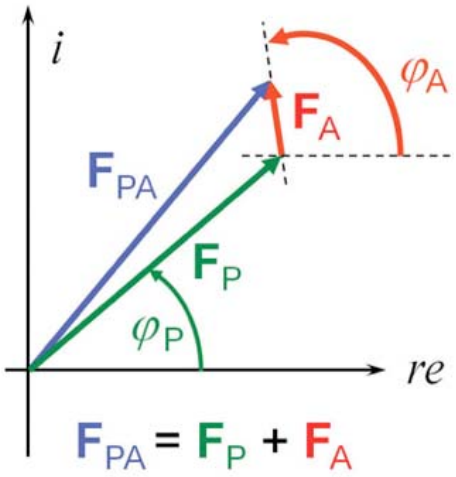


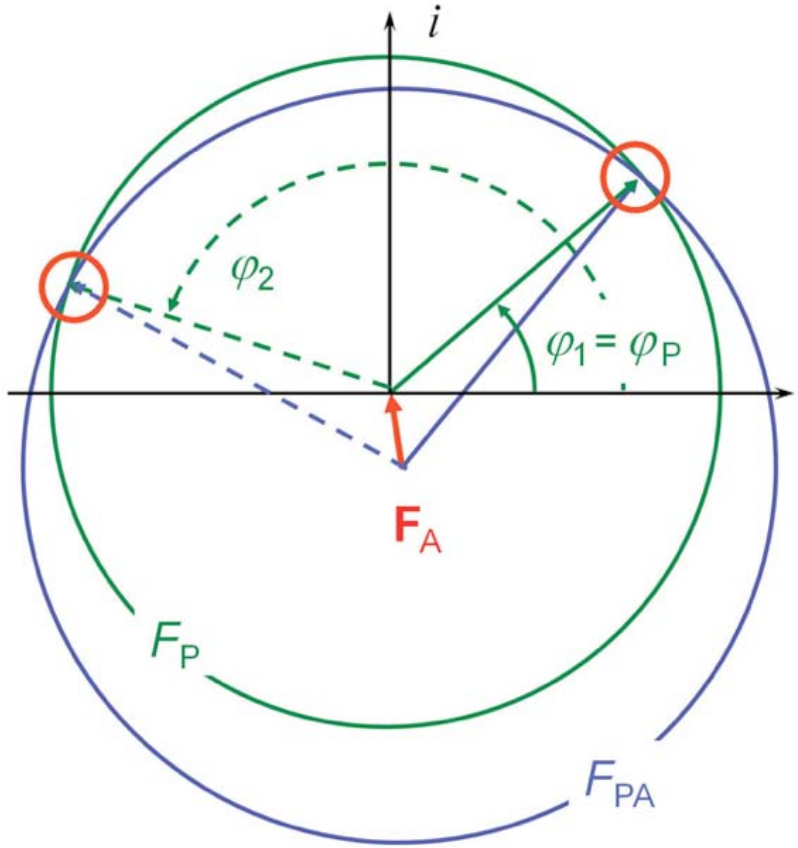
Graphical illustration of phase determination

Once we know where the heavy atom is we can calculate $\mathbf{F}_A$ – the amplitude AND phase of its contribution.

For each reflection we can make the construct closing the vector triangle. We know the amplitudes F_P and F_{PA} from the experimental measurements and the total contribution $\mathbf{F}_A$.

If we draw a circle radius, F_{PA}, at the origin and then move along $\mathbf{F}_A$ and draw a circle radius, F_P, there should be TWO points of intersection that satisfy the phase addition giving two solutions for ϕ_P.





Resolving the phase ambiguity

- We have a phase ambiguity if there is: a single isomorphous derivative (SIR) or anomalous scattering from the native crystal (e.g. from sulfur) (SAD).

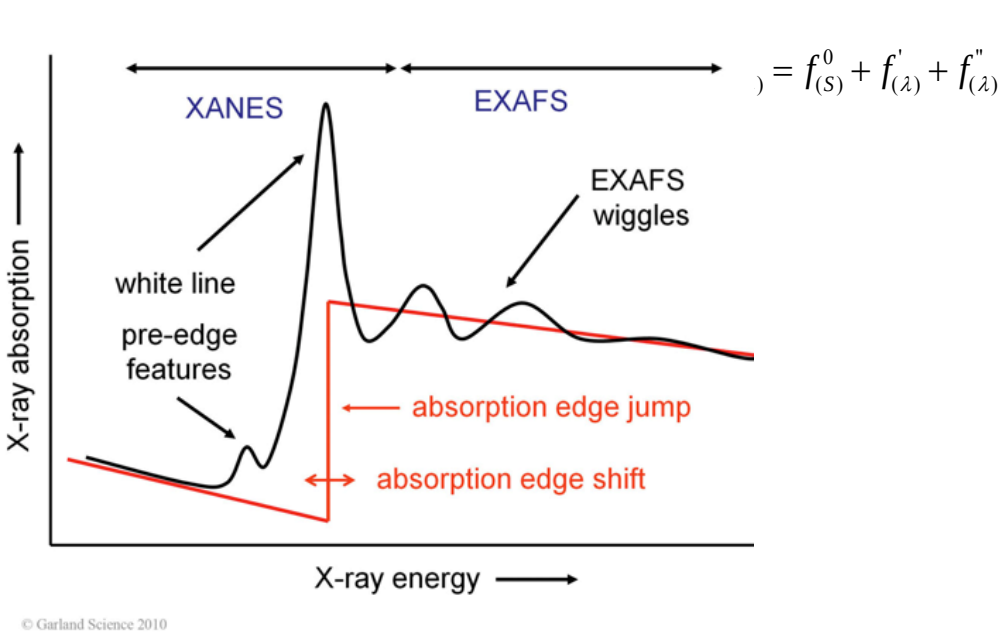
This problem can be resolved:

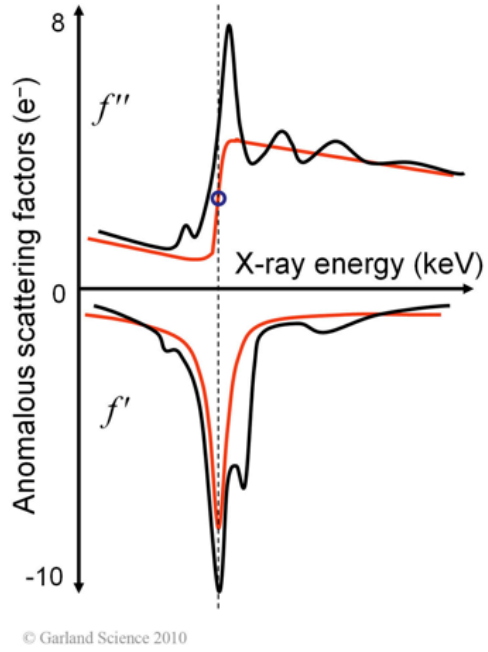
- By adding more derivatives – multiple isomorphous replacement (MIR)
- By using the anomalous signal – single isomorphous replacement with anomalous scattering (SIRAS) or multiple isomorphous replacement with anomalous scattering (MIRAS)
- By using the dispersive and anomalous signal together – multiple wavelength anomalous diffraction (MAD)
- With the help of density modification
- With dedicated direct methods programs

Anomalous scattering

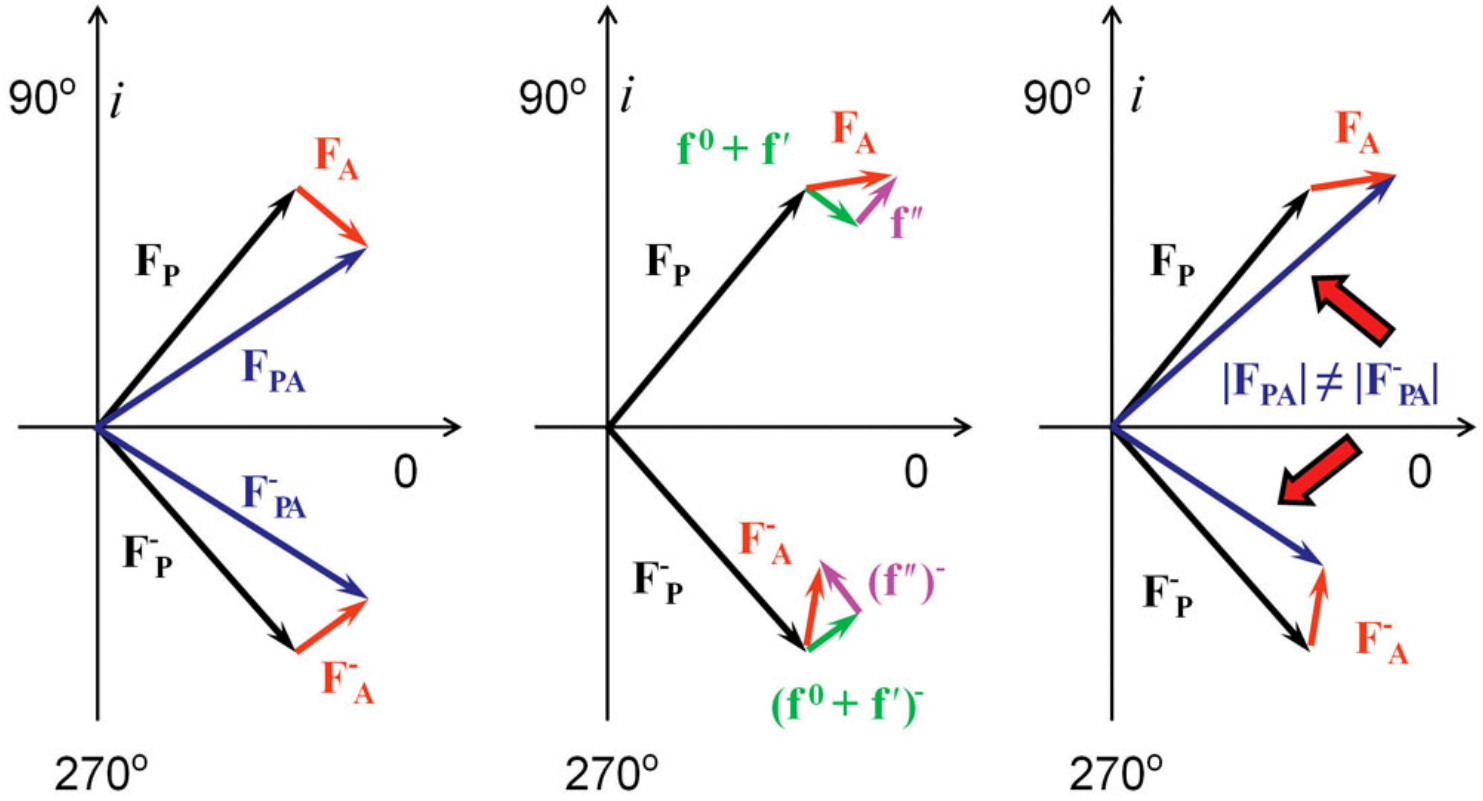
$$f_{(s,\lambda)} = f^{0}_{(s)} + f^{'}_{(\lambda)} + f^{''}_{(\lambda)}$$

The scattering from an atom actually contains two wavelength (energy) dependent terms. These vary rapidly near an absorption edge.



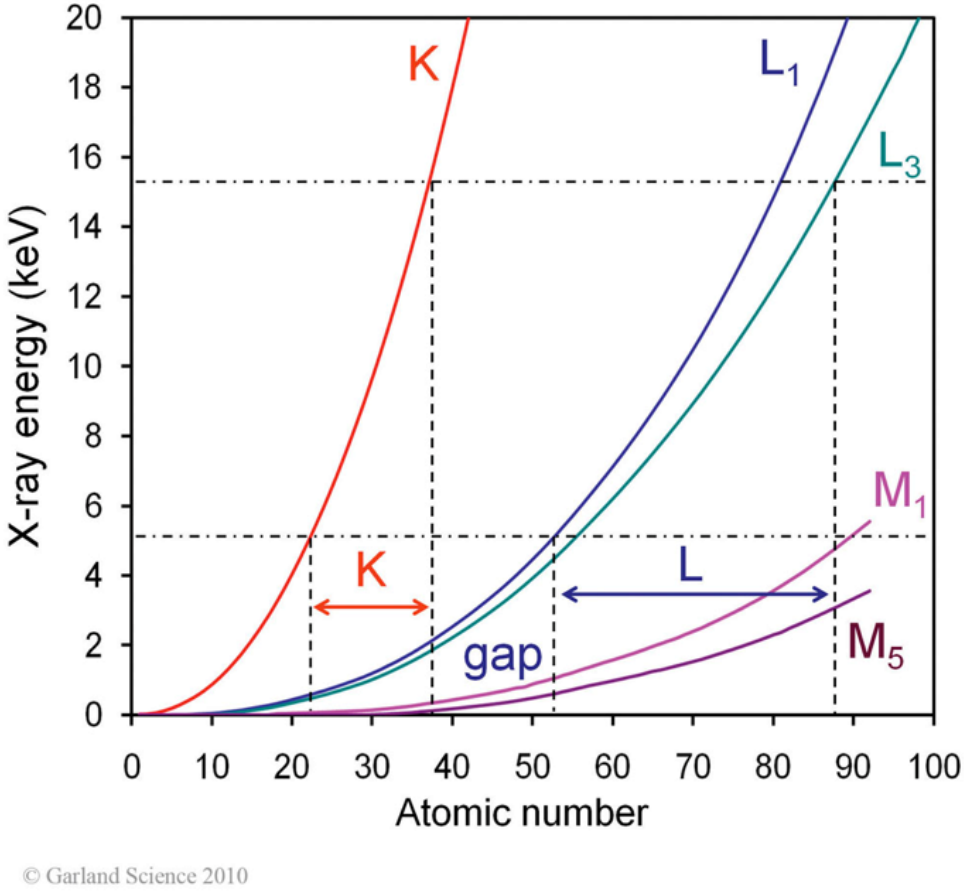


Anomalous scattering for phasing



Anomalous scattering only significant for heavier elements in a protein

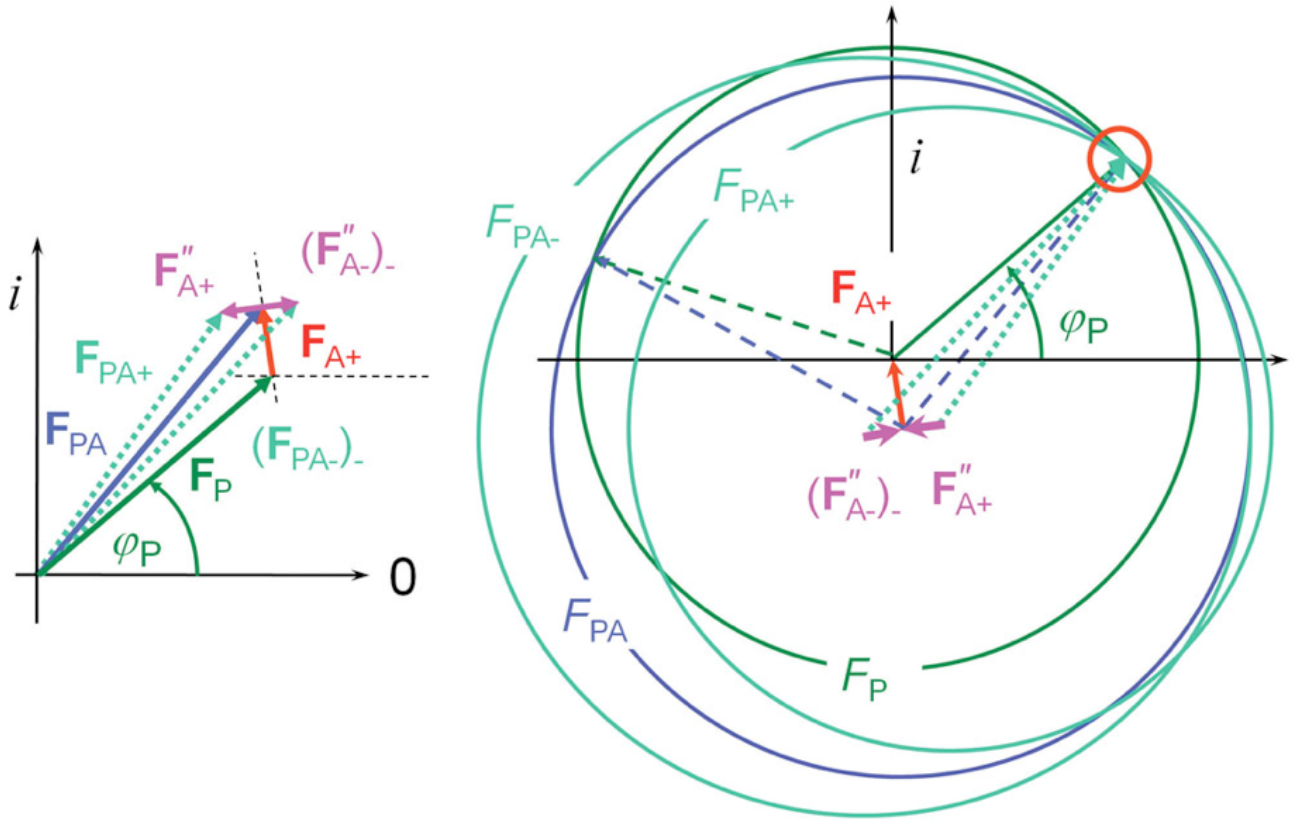
- At most synchrotrons PX beamlines one can access the L-edges for elements form iodine to uranium and the K-edges for transition metals
- Most importantly the edge for selenium is at 0.979 Å (1.27 Kev)
  - Se can be introduced into recombinantly expressed proteins replacing the sulfur in the amino acid methionine



Single Isomorphous Replacement with Anomalous Scattering (SIRAS)

In this case for each reflection we have 1 observation of the native amplitude, F_P, and two for the derivative, F_{PA-} and F_{PA-}.

In the absence of anomalous scattering (blue circle) there are two solutions with anomalous scattering (light green circles) there is only one.



Phasing method	Phasing marker	Remarks	Available data	Marker	Difference data	Data to be phased
SAD via sulfur atoms (S-SAD)	S in Met, Cys residues	Highly redundant data collection, must be combined with density modification	Anomalous pairs F_{PS}^{+}, F_{PS}^{-} also serve as native data	Sulfur positions, $\mathbf{F}_S$	ΔF_{ano} from $\lvert F_{PS}^{+} - F_{PS}^{-} \rvert$	Merged F_{PS}
SAD via naturally bound metals	Naturally bound metal ion, cofactor	Requires density modification for resolution of phase ambiguity	Anomalous pairs F_{PA}^{+}, F_{PA}^{-} also serve as native F_P	Anomalous scatterer positions, $\mathbf{F}_A$	ΔF_{ano} from $\lvert F_{PA}^{+} - F_{PA}^{-} \rvert$	Merged F_{PA}
SIR(AS) via isomorphous metals	Heavy atom ion, specifically bound anions Br^-, I^-, I^{3-}, also Xe	Isomorphous phasing power proportional to $z_{(H)}$, anomalous signal or density modification needed to break phase ambiguity	Native data F_P, isomorphous data F_{PA}, in pairs F_{PA}^{+}, F_{PA}^{-} for SIRAS	Isomorphous/ anomalous scatterer positions, $\mathbf{F}_A$	ΔF_{iso} from $\lvert F_{PA} - F_P \rvert$ and ΔF_{ano} from $\lvert F_{PA}^{+} - F_{PA}^{-} \rvert$	Native F_P
MIR(AS) via isomorphous metals	Heavy atom ions, clusters, specifically bound anions Br^-, I^-, I^{3-}	As above, except multiple derivatives or anomalous signal break phase ambiguity. Hg, Pt, Au, etc. phase several hundred residues, heavy atom clusters more.	Native data F_P, isomorphous data $n\cdot(F_{PA})$ in pairs $n\cdot(F_{PA}^{+}, F_{PA}^{-})$ for MIRAS	Isomorphous/ anomalous scatterer positions, $n\cdot(\mathbf{F}_A)$	$n\cdot\Delta F_{iso}$ from $n\cdot\lvert F_{PA} - F_P \rvert$, and ΔF_{ano} from $\lvert F_{PA}^{+} - F_{PA}^{-} \rvert$ pairs	Native F_P
MAD via Se	Se in Se-Met residues	1 Se phases 100–200 residues, introduced by expression host	Bijvoet pairs at n wavelengths, $n\cdot(F_{PA}^{+}, F_{PA}^{-})_{\lambda n}$, optional native data	Anomalous scatterer positions $\mathbf{F}_{Se}$	ΔF_{ano} from $\lvert F_{PA}^{+} - F_{PA}^{-} \rvert_{\lambda}$ pairs ΔF_{λ} from $\lvert F_{\lambda i} - F_{\lambda j} \rvert$ pairs	Best merged data F_{PA}, optional native F_P
MAD via isomorphous metals	Heavy atom, specifically bound	Strong signal on XAS "white lines," particularly at L-edges, can phase several hundred residues	Bijvoet pairs at n wavelengths $n\cdot(F_{PA}^{+}, F_{PA}^{-})_{\lambda n}$, native data F_P (not needed for phasing)	Anomalous scatterer positions $\mathbf{F}_A$	ΔF_{ano} from $\lvert F_{PA}^{+} - F_{PA}^{-} \rvert_{\lambda}$ pairs ΔF_{λ} from $\lvert F_{\lambda i} - F_{\lambda j} \rvert$ pairs	Native F_P, optional best merged F_{PA}
Direct methods	None	Near atomic resolution data (1.2 Å or better), relatively small proteins	F_P	All non-H atom positions *ab initio*	N/A	Native F_P
Density modification	None	Needs multiple copies of motif in asymmetric unit for *ab initio* phasing	F_P	Multiple copies of a subunit	N/A	Native F_P
MR via model structure	Positioned search model	Needs search model with structural similarity, subject to model bias, particularly at low resolution	Native data F_P and search model structure factors F_c	Entire model serves as search probe	N/A	Native F_P

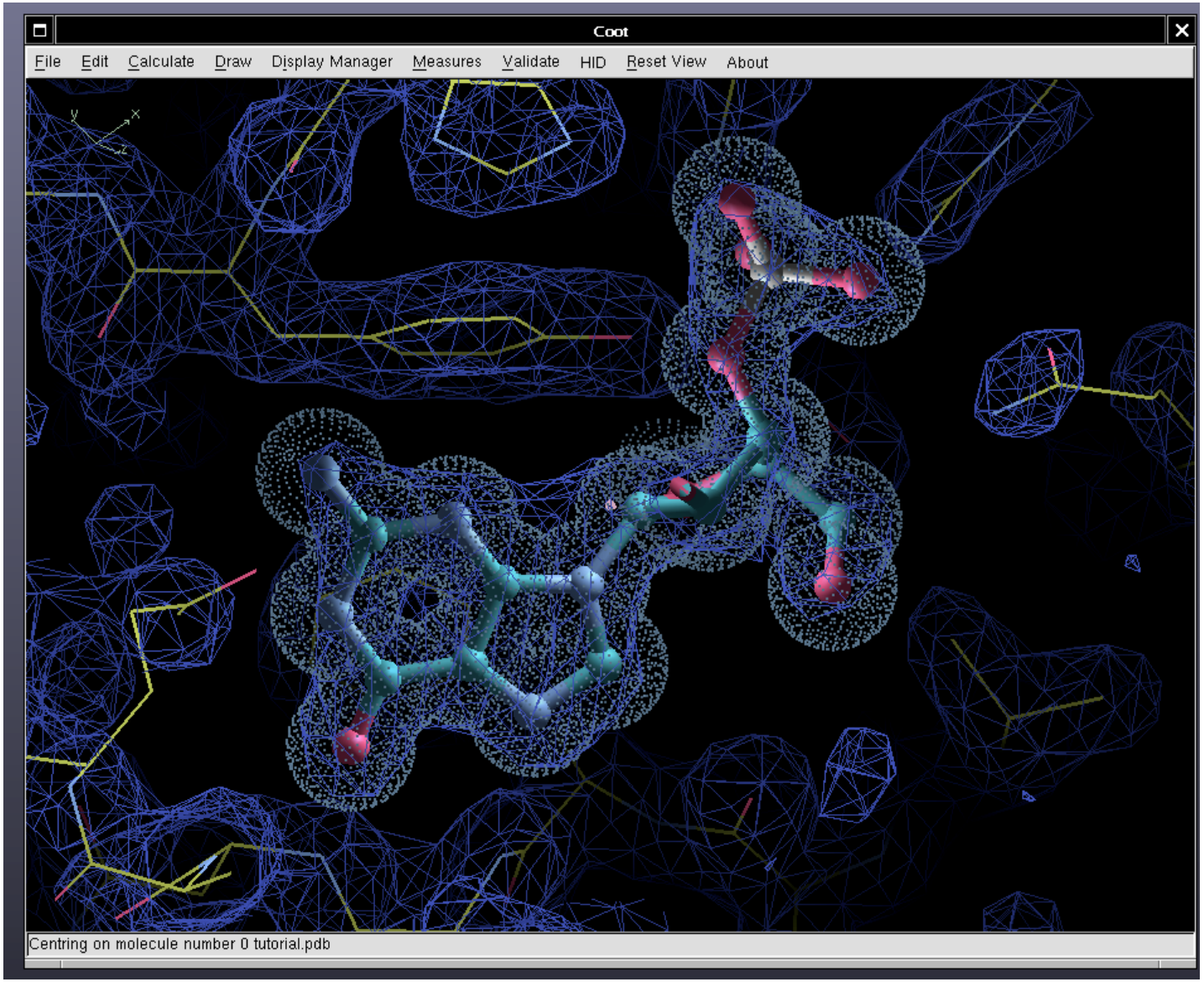
Early tools for model building

The “Richards” box – named after Fred Richards

Viewing the structure in a Richards' box

The reflection of the model in the half-silvered mirror is seen "superposed" on the electron-density contoured on transparent sheets behind the mirror.

Modern graphics - COOT



Refinement procedures

- **Local real-space refinement**: fragments of the model once situated in electron-density (using COOT or O). The fit (overlap) to electron-density can be automatically optimised for short segments of polypeptide. At the same time the peptide geometry can also be regularised by comparison with a dictionary.
- **Global reciprocal-space restrained refinement**: the parameters of the model atoms (x, y, z, and B) and overall parameters such as scale factor, bulk solvent correction, and anisotropy are refined against the experimental data. The most common target functions are either derived from maximum likelihood or least-squares.
- **Cycling in real- and in reciprocal-space**: fitting in real-space (especially manually) has a large radius of convergence and major errors and omissions in the model can be corrected. On the other hand reciprocal space refinement makes concerted movements of all the parameters to improve the overall agreement. Reciprocal space refinement will improve the electron-density maps that can then once again be subject to local real-space refinement.
- **Stereochemical restraints**: protein crystals rarely diffraction to atomic resolution so there are insufficient data to refine all the atomic parameters – that is the ratio of observations/variables is too small. This can be addressed either by increasing the number of observations (adding restraints) or reducing the number of variable parameters (constraints).

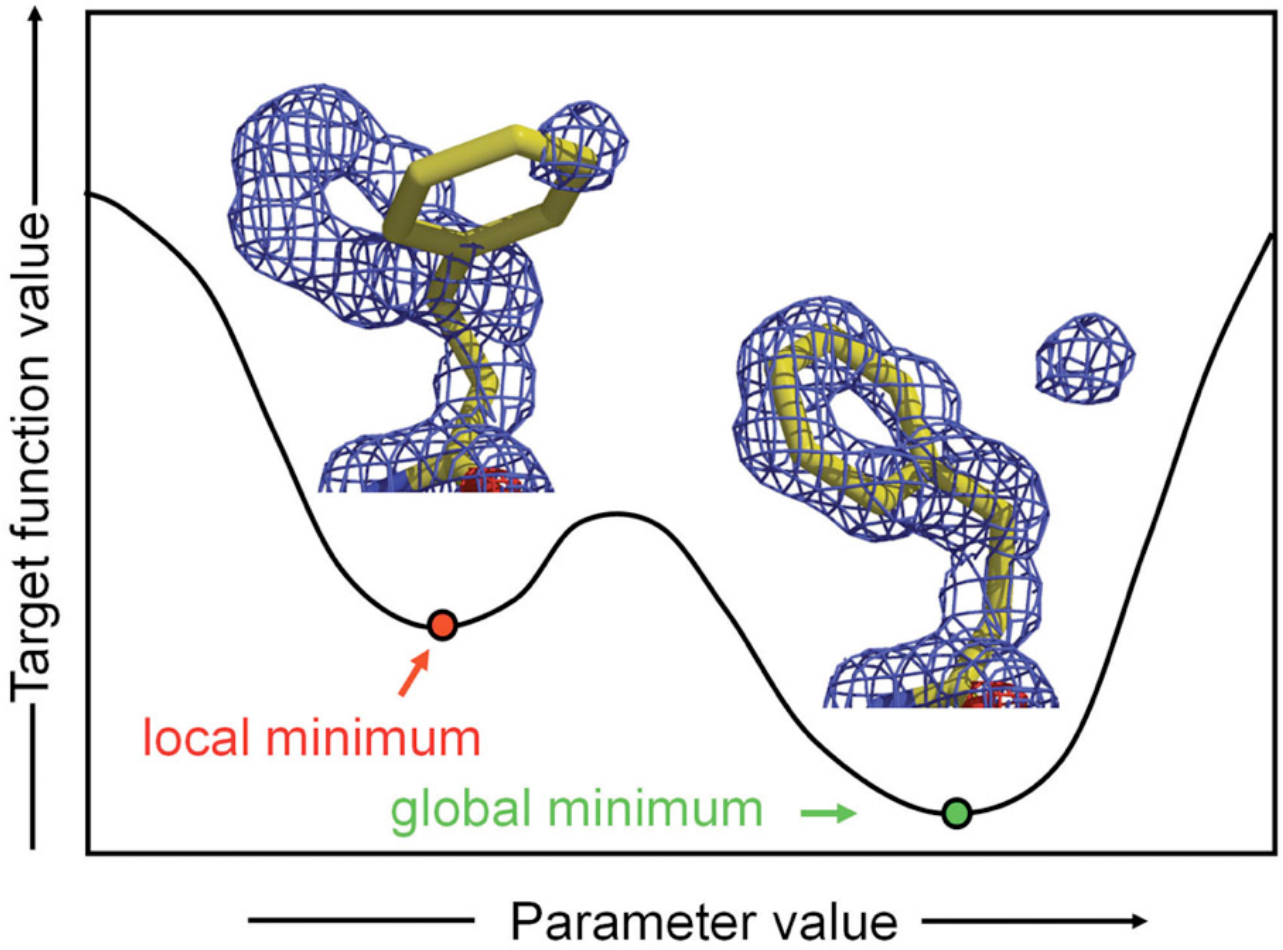
Monitoring refinement – linear *R*-value

$$R = \frac{\sum_{\mathbf{h}} | F_{obs} - F_{calc} |}{\sum_{\mathbf{h}} F_{obs}}$$

The *R*-value gives an estimate of the overall agreement between the observed and calculated structure factor amplitudes. It is relatively insensitive to local errors as the effect is spread over all the data. Hence the importance of the alternation of real- and reciprocal-space optimisation.

The *R*-value generally varies from 0.15 to 0.30 for most protein structures and is lower for higher resolution structures in general.

Local minima in a complex landscape



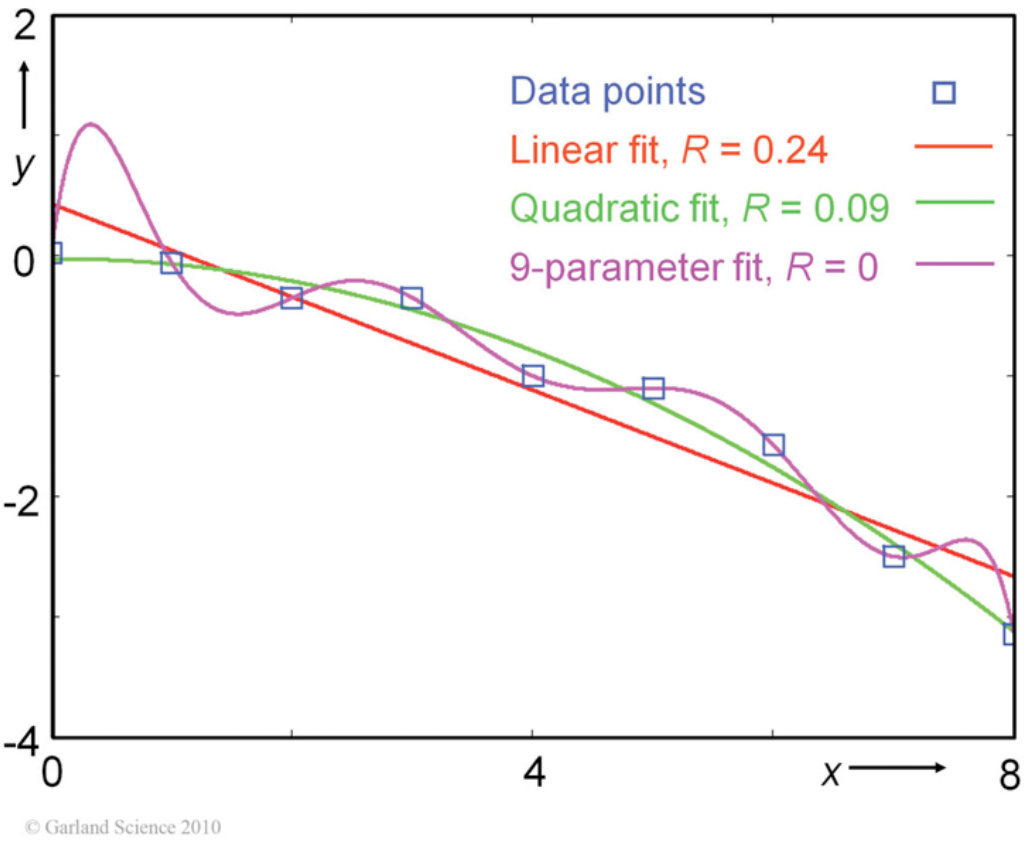
A refinement program may not be able to cross the barrier necessary to reach the global minimum but it would be obvious manually.

How many parameters?

"Numquam ponenda est pluralitas sine necessitate"
William of Ockham (1288-1347)

Shows in this example how with sufficient parameters a perfect fit can be achieved. The result, however, must be physically meaningful.

In protein crystallography we would normally restrict the number of parameters per atom to 4 or even 3 at low resolution. In small molecule crystallography we routinely describe the atomic motion by 6 parameters making 9 per atom in total.

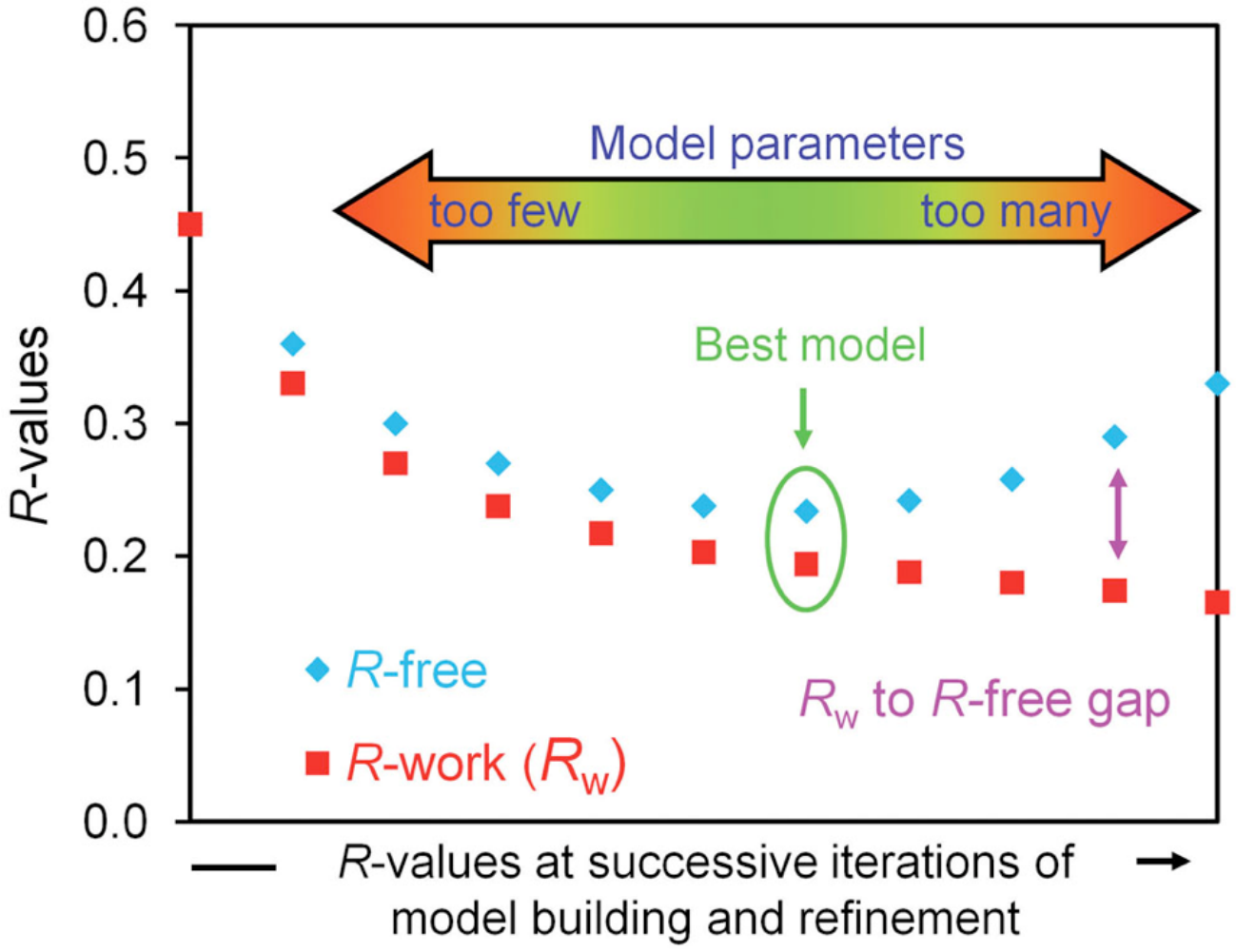


Cross-validation and R_{free}

R_{free} is simply the R-value calculated for a set of reserved reflections that have never been used in the refinement. The remaining reflections are termed the working set, R_{work}.

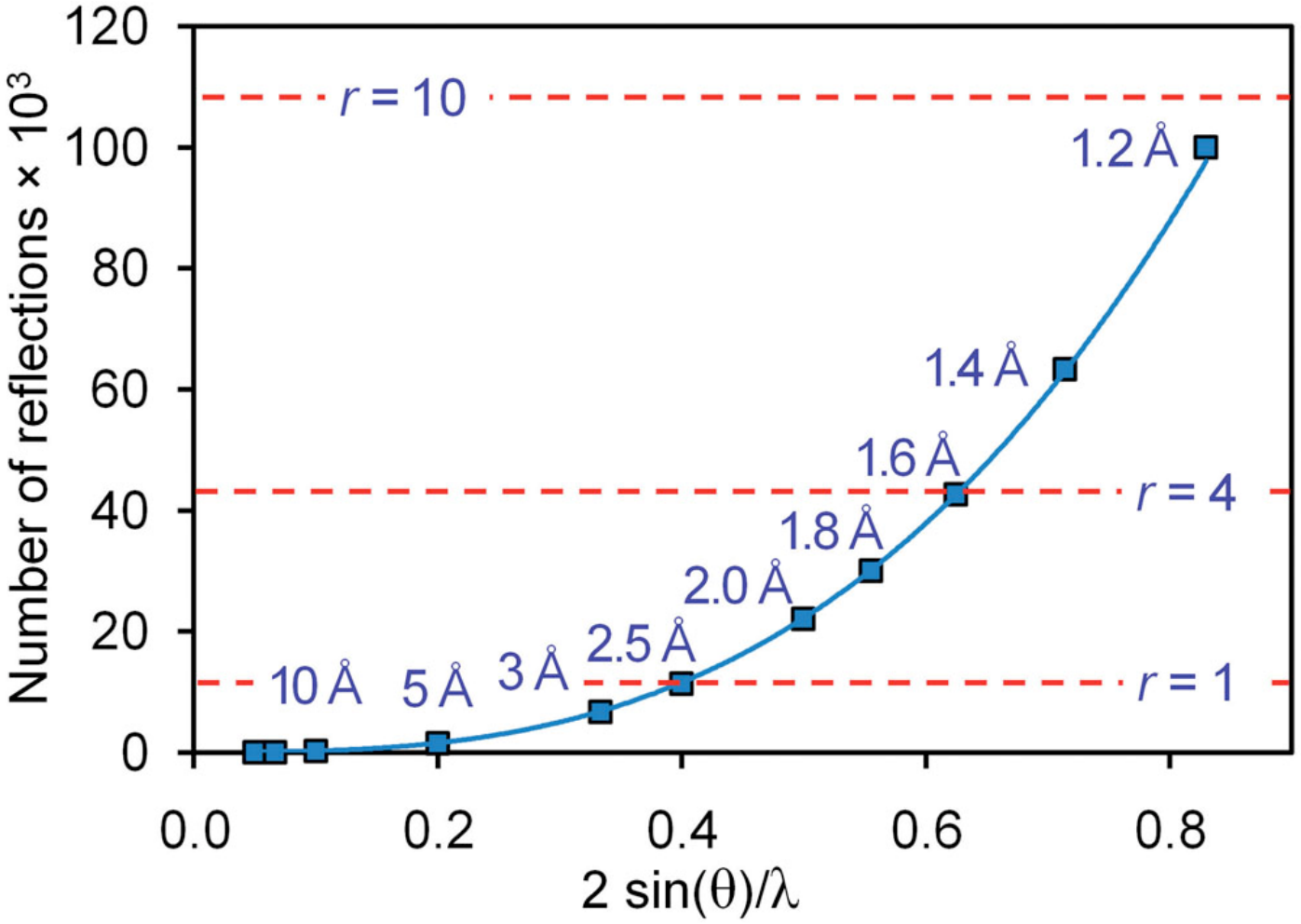
R_{free} is related to the mean phase error and is therefore a measure for phase accuracy and for model quality.

R_{free} is a very useful tool to monitor the refinement as long as it decreases along with R_{work} the refinement can continue. Beyond that point too many parameters may have been added.



Data-to-parameter ratio in protein crystallography

The number of reflections equals the number of parameters at a resolution of about 2.5 Å. The data only become sufficient for unrestrained refinement at 1.0 Å resolution.



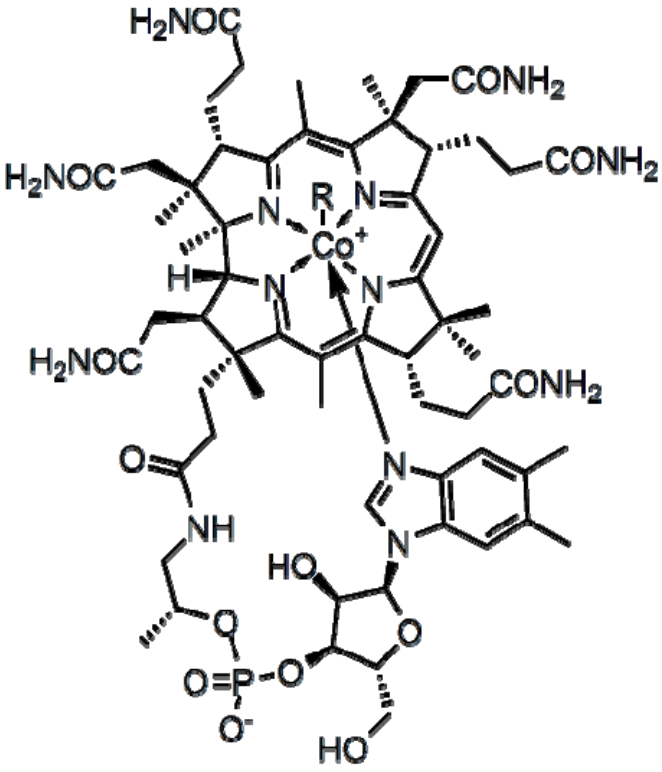
Other restraints

- Molecules related by non-crystallographic symmetry can be restrained to be similar.
  - The programs generally allow for different restraints for main-chain and side-chain atoms and weights for tight, medium or loose restraint. These can be useful in the early stages of refinement to ensure that subjective model-building doesn't make the structures deviate too much.
- Experimental phase restraints can be applied at early stages of refinement when the model phases are not very good.
- Displacement parameter (*B*-factor) restraints are applied so that atoms that are bonded to each other in the structure don't have wildly different *B*-factors, which would be physically unreasonable.

SOME EARLY ACHIEVEMENTS IN MACROMOLECULAR CRYSTALLOGRPAHY

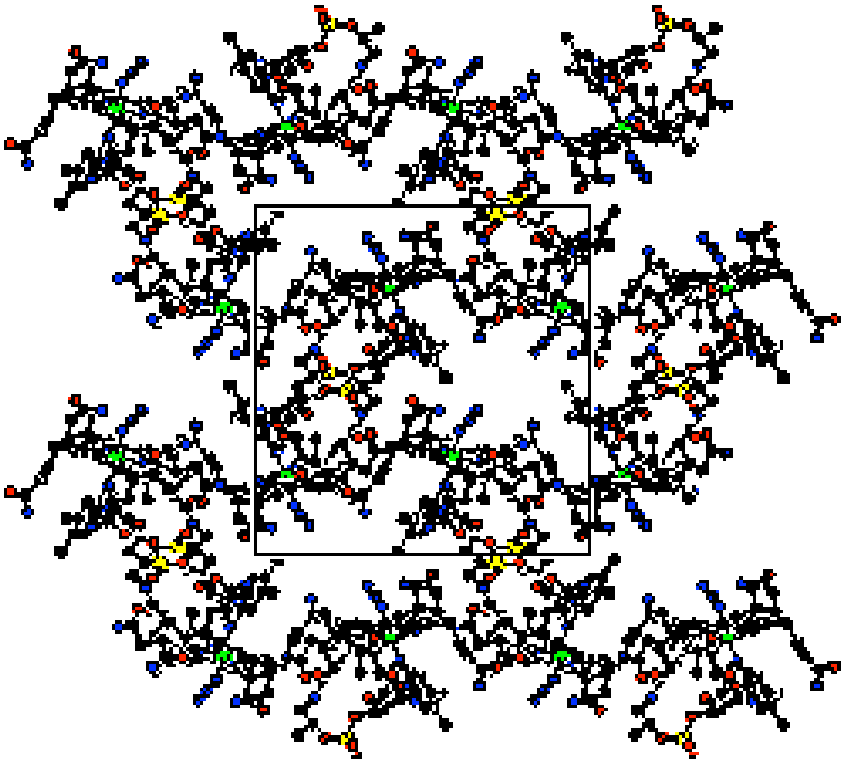
Vitamin B12

- A complex organic molecule – not able to be synthesised or characterised by other means in the 1950s
- Vitamin B12 is a water-soluble vitamin with a key role in the normal functioning of the brain and nervous system, and for the formation of blood.
- Neither fungi, plants, nor animals are capable of producing vitamin B12 only bacteria
- Humans get their normal intake of B12 from animal products
- It contains an essential cobalt atom at the centre of a pterin ring



R = 5'-deoxyadenosyl, Me, OH, CN

Crystal structure of vitamin B12



Provided the first complete structural AND chemical characterisation of vitamin B12

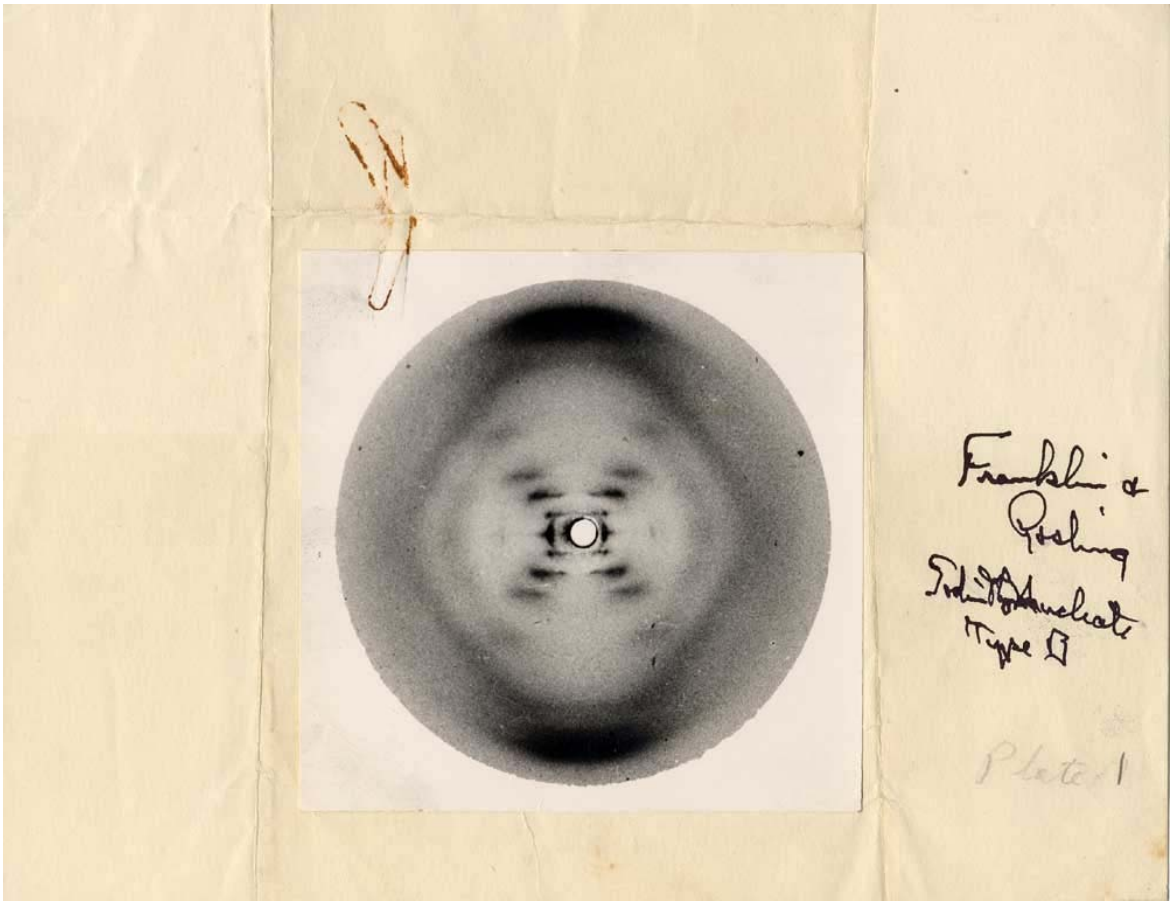
Royal Society stamp honouring Dorothy Crowfoot Hodgkin

10 stamps issued to celebrate the 350th anniversary of the Royal Society (of Great Britain) in 2010

Boyle (Chemistry) Newton (Optics) Franklin (Electricity) Jenner (Vaccination) Babbage (Computing)
Wallace (Evolution) Lister (Antisepsis) Rutherford (Atomic Structure) Hodgkin (Crystallography) Shackleton (Earth Sciences)

Chosen from more than 1400 members including 60 Nobel Laureates – crystallography 1 of 12 sciences, Dorothy Hodgkin the only woman

The most important discovery in molecular biology: the birth of the double helix

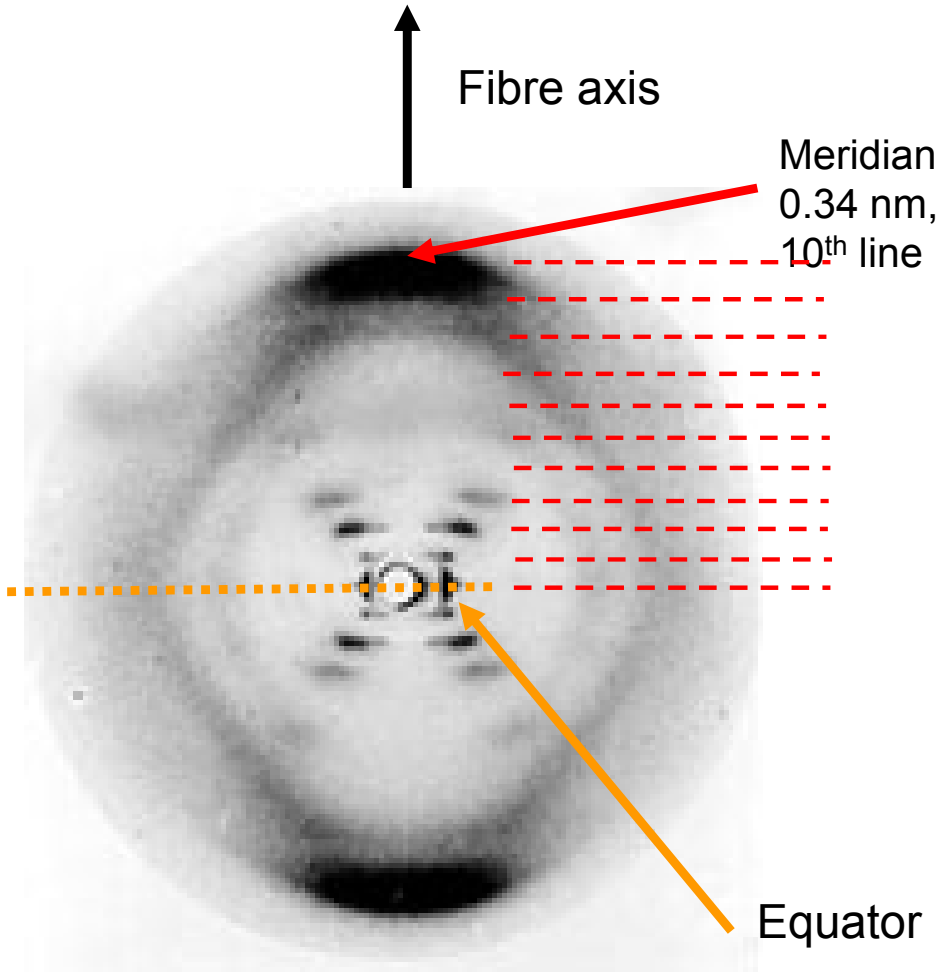


Sodium deoxyribose nucleate from calf thymus, Structure B, Photo 51, taken by Rosalind E. Franklin and R.G. Gosling. This is a diffraction photo from a **fibre** of DNA not a **crystal**!

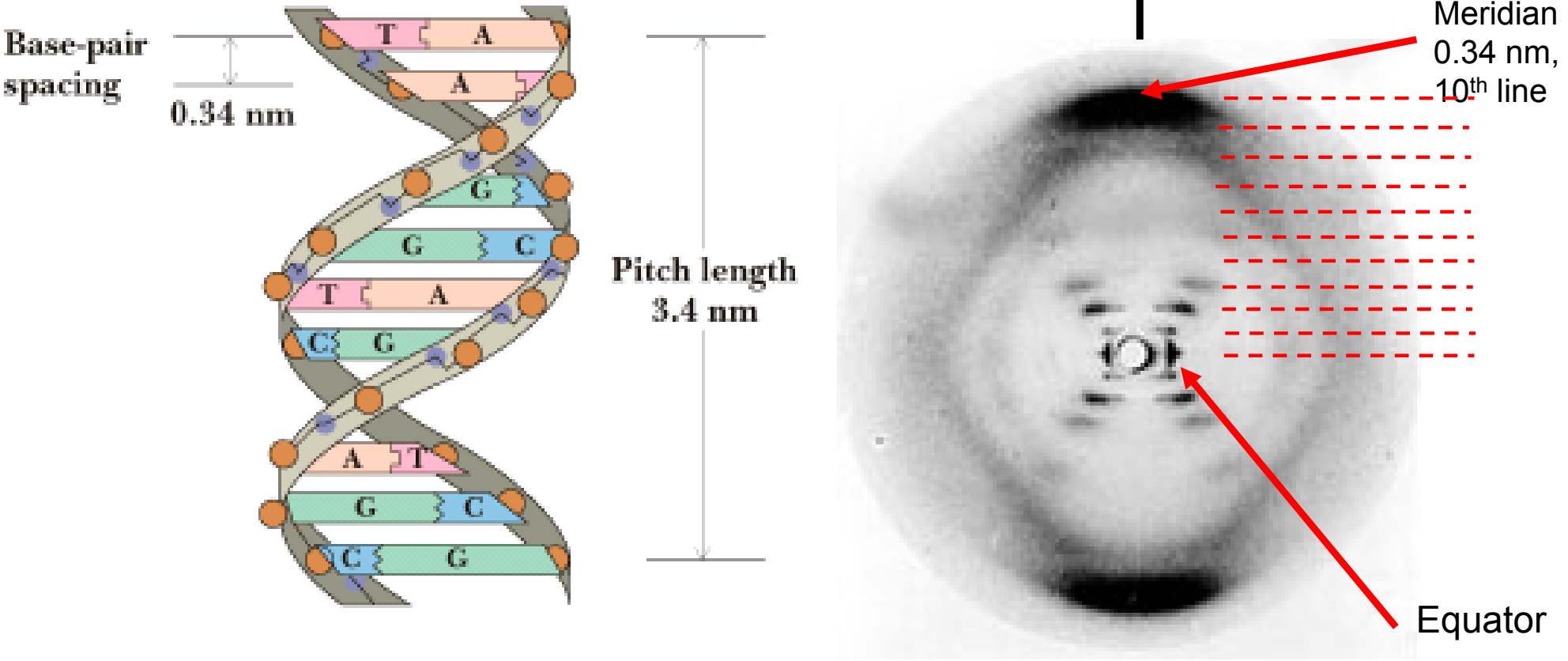
Rosalind Franklin - March 1956

X-ray diffraction and DNA

- Same experiment that Pauling and Corey used to define the α-helix & β-sheet protein structures
- Cross pattern indicates a helix
- Axial spacings define the dimensions of the helix:
  - 0.34 nm (3.4 Å) reflection gives projected height of repeating unit
  - 3.4 nm line spacing gives helix repeat
  - 10 lines = 10 repeating units/helix turn
- Equatorial spacings give radii of the helices and their separation
- Density and pattern of weak spots indicate a double helix



X-ray diffraction and DNA



0.1 nm = 1 Å = 10^{-9} m

Greatest understatement in history?

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three inter-twined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

Th... diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate di-ester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. WATSON
F. H. C. CRICK

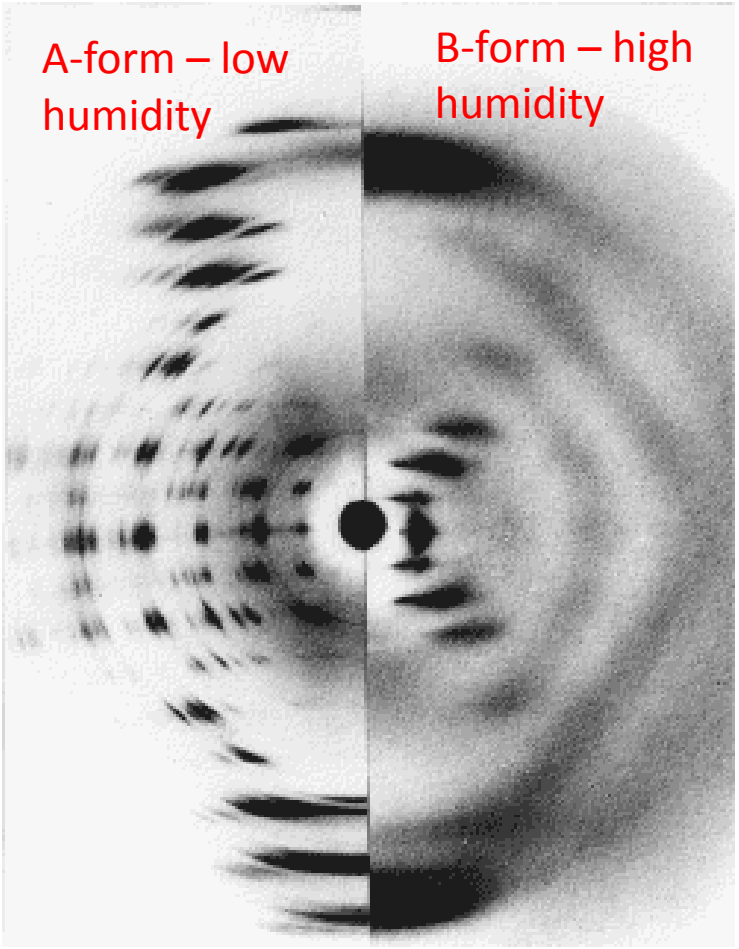
Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,
Cavendish Laboratory, Cambridge.
April 2.

[1] Pauling, L., and Corey, R. B., *Nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **39**, 84 (1953).
[2] Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952).
[3] Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, **9**, 402 (1952).
[4] Wyatt, G. R., *J. Gen. Physiol.*, **36**, 201 (1952).
[5] Astbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).
[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material"

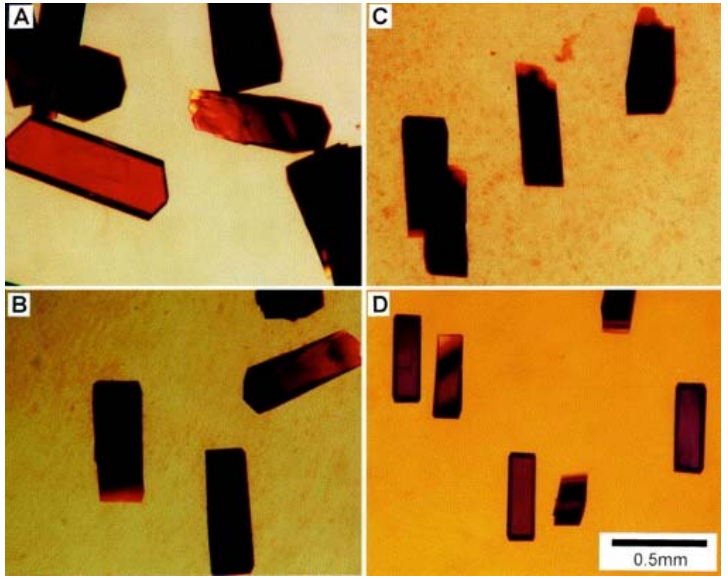
Did Rosalind Franklin examine the wrong pattern?

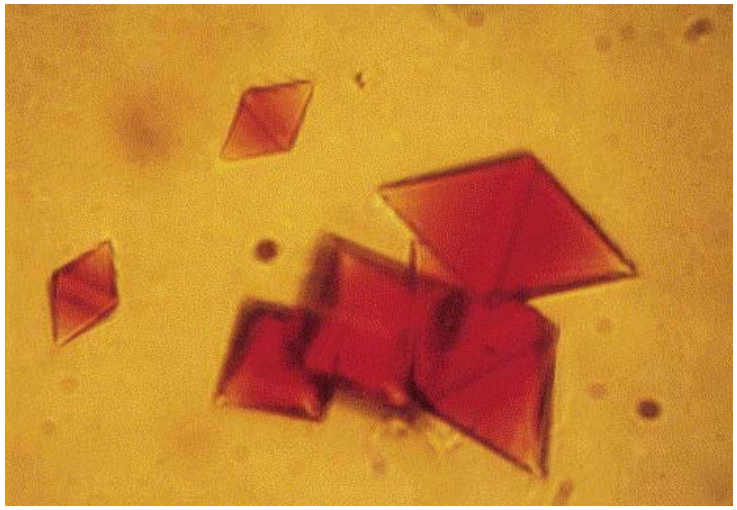
- 2 patterns
  - A-form semi-crystalline
  - B-form oriented fibre
- Rosalind Franklin worked on the A-form because as a crystallography it seemed to her to have more information
- The extra information gave greater detail about helix to helix packing
  - All the information is in the A-form it is just difficult to interpret
  - Ultimately A-form data were used for a full refinement of DNA by Arnott and Wilkins



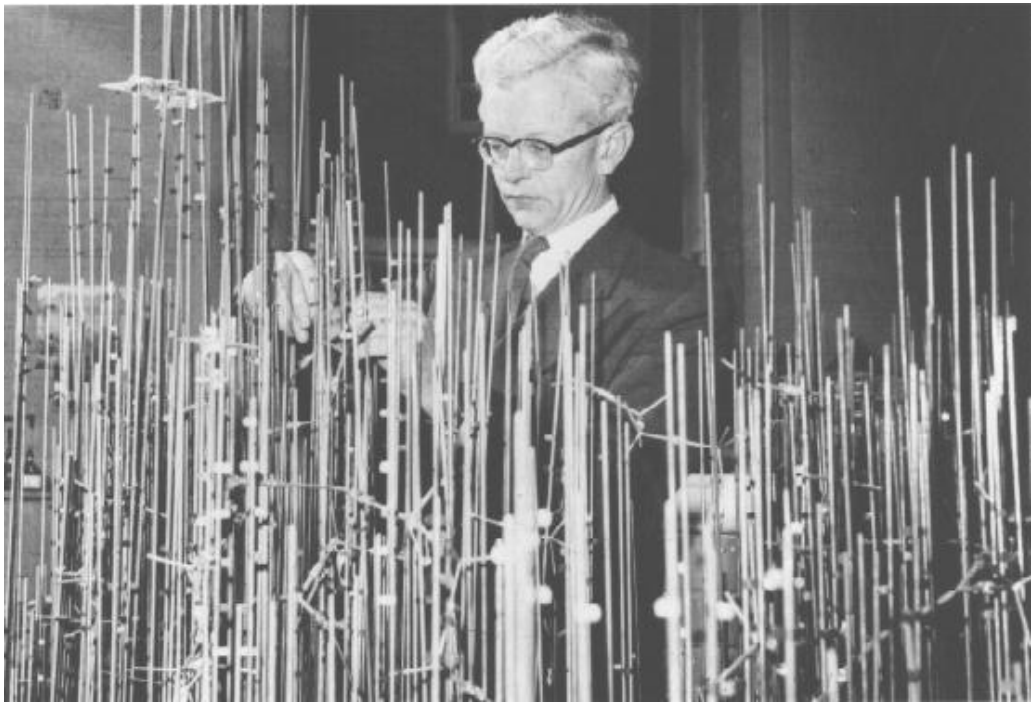
First protein structures: myoglobin & hemoglobin

- Crystals were first grown in 1930s
- Perutz began work on hemoglobin in 1938
- Dorothy Hodgkin had shown that the crystals needed to be kept wet or they fell apart
- At the time the intensities of diffraction spots were estimated ‘by eye’ – took teams of people months to get a partial data set

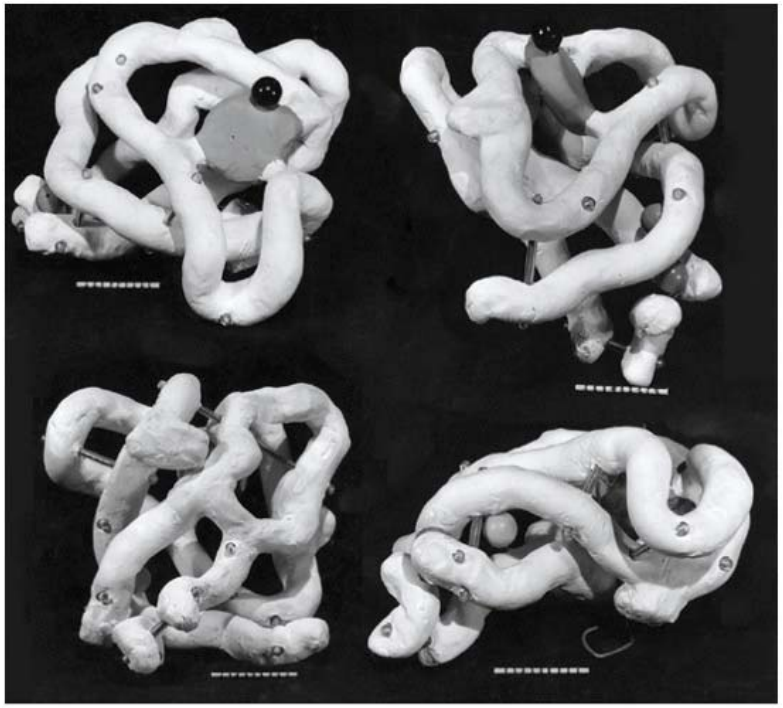




Structure of myoglobin (Mb)



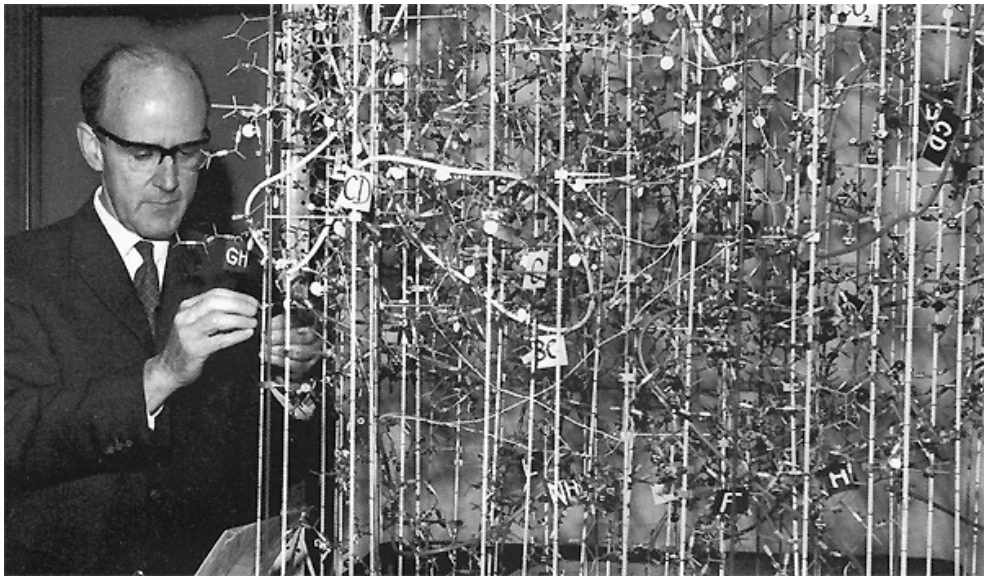
John Kendrew constructing the first atomic model of myoglobin



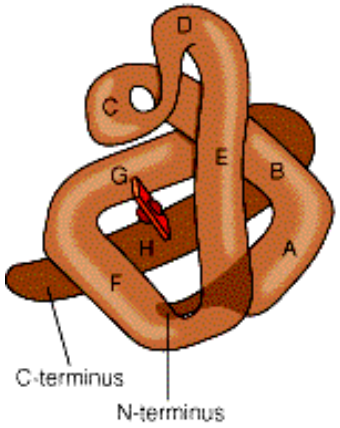
Nature Reviews | Molecular Cell Biology

Earlier low resolution models of myoglobin constructed from modeling clay. α-helices are seen as tubes at this resolution.

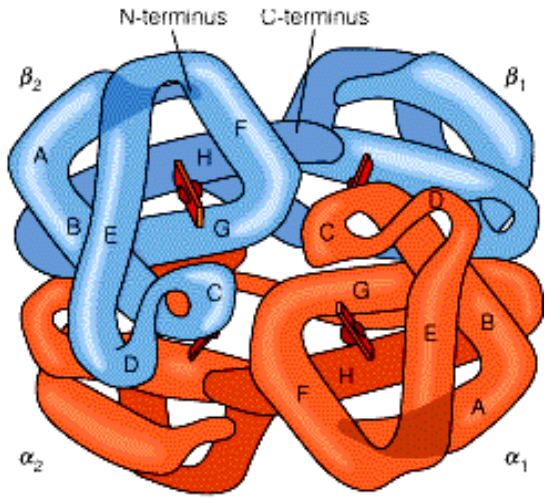
Structure of hemoglobin (Hb)



Max Perutz constructing the atomic model of hemoglobin (> 20 years after starting the project). Hb is like a complex of 4 myoglobin molecules pointing to an evolutionary relationship between the two proteins.

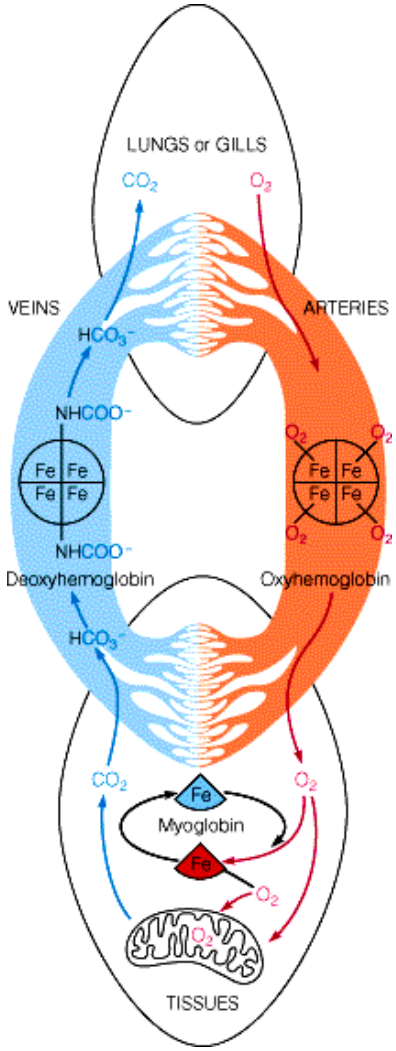


Myoglobin

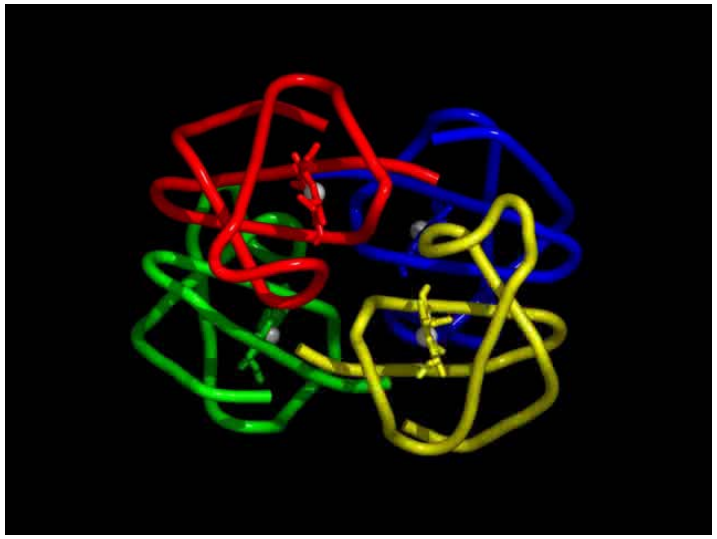


Hemoglobin

Relating structures of hemoglobin (Hb) and myoglobin (Mb) to their functions

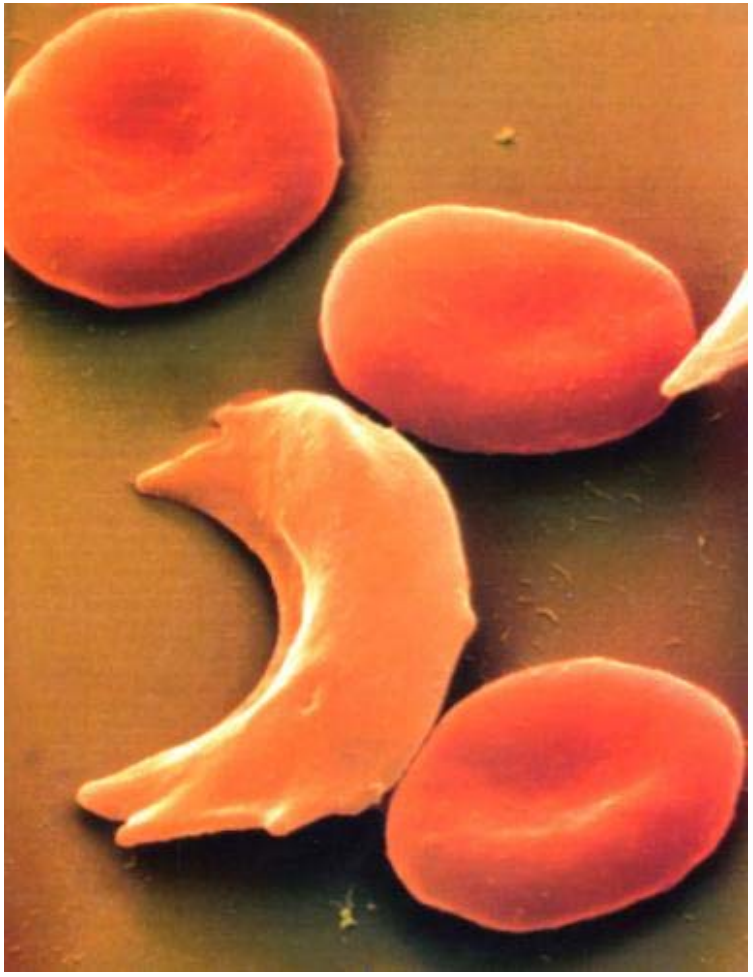


- Hb functions to transport O_2 from a region of high pO_2 (lungs) to one of relatively low pO_2 (resting muscle)
- Mb binds O_2 tightly at a point when Hb is binding less tightly but will eventually release the O_2 if the pO_2 falls to a low enough level (working muscle)
- When Hb binds O_2 its structure changes – this change can be related to the O_2 binding behaviour

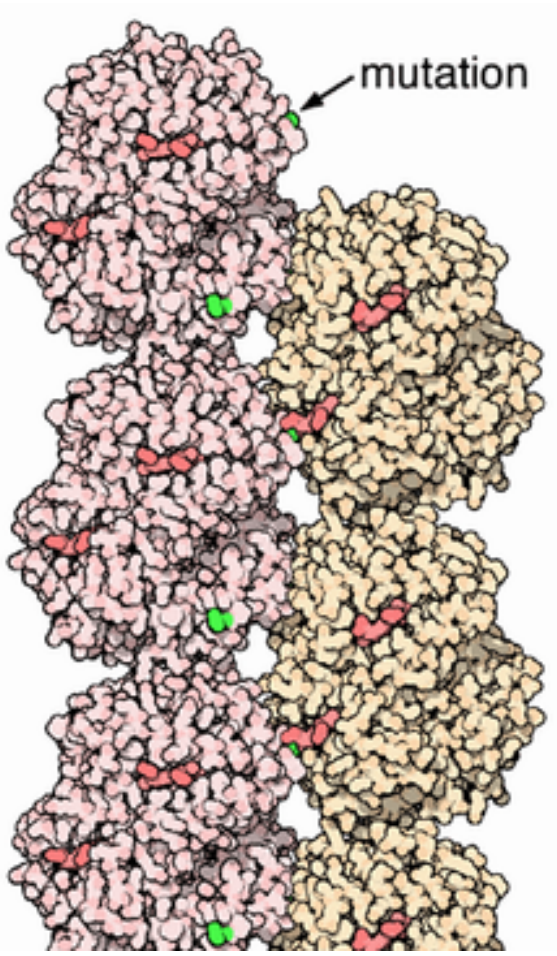


Use of Hb structure to understand disease

- Sickle cell anemia is caused by a single mutation (1 amino acid in 1 chain of Hb)
- The result is that the Hb molecules form fibers that distort the red blood cells



Normal & sickled red blood cells



Hb molecules form a fiber

Nobel laureates of 1962

Steinbeck

Wilkins ? Kendrew Perutz Crick Watson

Success:

AN EXAMPLE FROM MY LABORATORY

Something new about haemoglobin?

Structural basis of haemoglobin capture by *Staphylococcus aureus* – stealing haem from haemoglobin

THE UNIVERSITY OF SYDNEY

2014 international year of crystallography

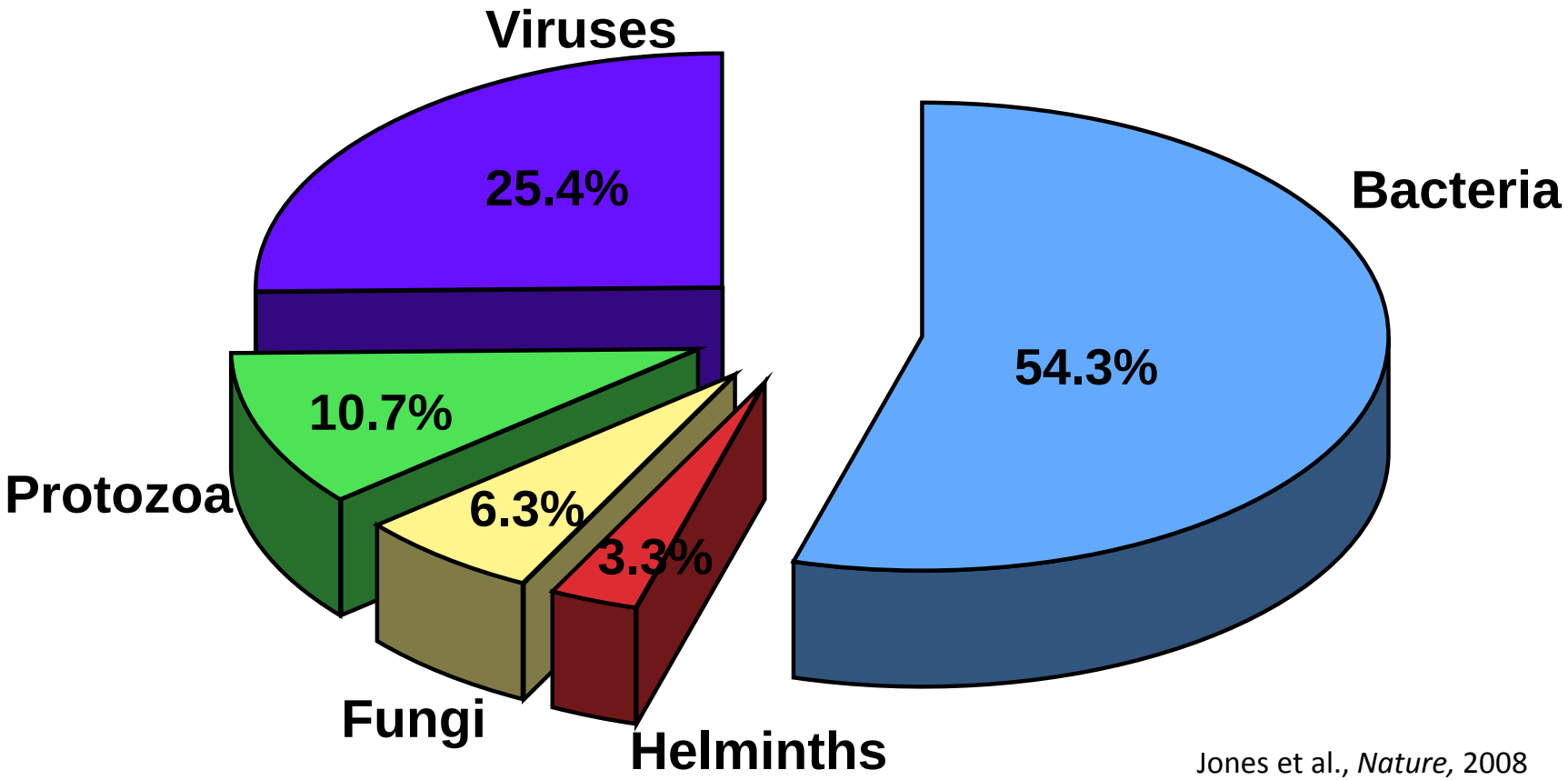
Max Perutz and haemoglobin

- 1937: Commenced work on his Ph.D. in Cambridge in the laboratory of J.D. Bernal (also supervisor of Dorothy Crowfoot Hodgkin) under the overall direction of Lawrence Bragg
- 1962: Nobel prize for Chemistry for the structure of haemoglobin
- Now > 600 structures in the PDB for haemoglobin – what can be new?

Max Perutz
with heamoglobin

John Kendrew
with myoglobin

Bacteria cause majority of emerging infectious disease



Jones et al., *Nature,* 2008

Staphylococcus aureus – 'Superbug'

- Pneumonia
- Meningitis
- Toxic shock syndrome
- Bacteremia
- Septic endocarditis

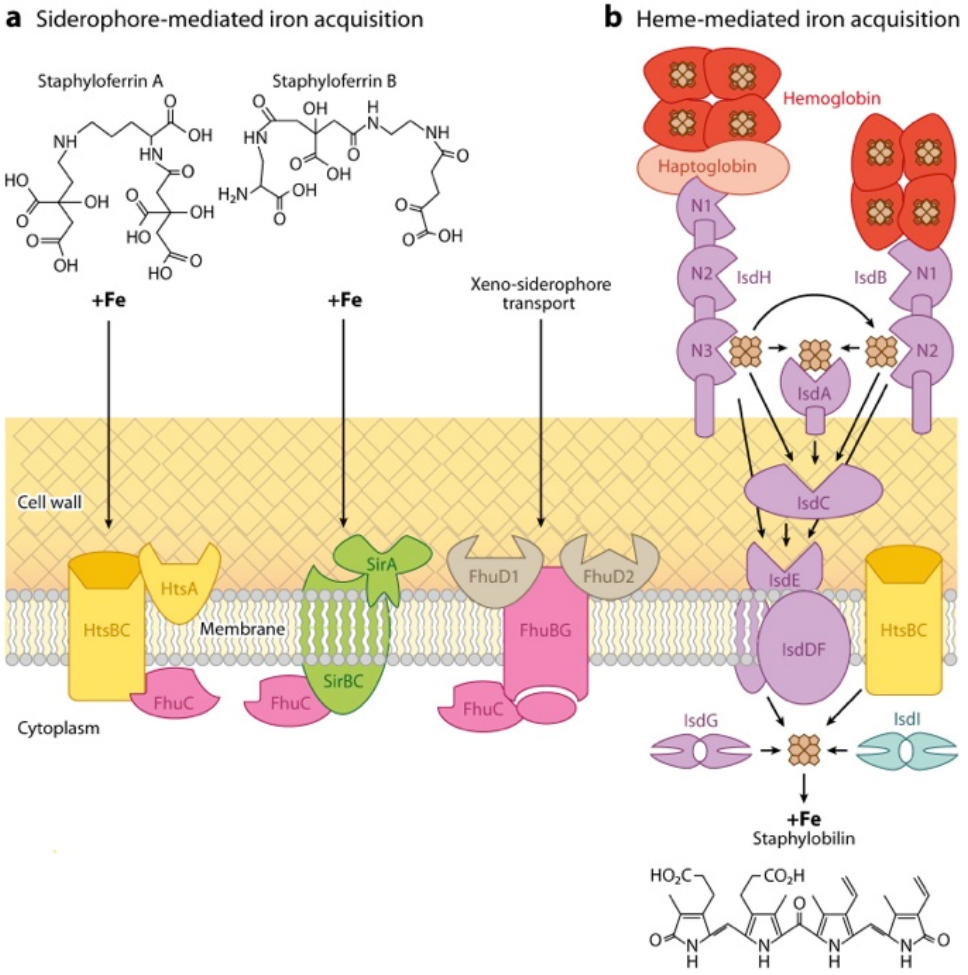
- Around 25 % of systemic infections in Australia are due to methicillin-resistant *S. aureus* (MRSA)
  - MRSA strains have until recently been restricted to hospital settings but now become evident in the community
- There is a pressing need to develop new therapeutics
  - Need to think outside the box!

Bacteria require iron

- Iron limitation is one of the most important innate defences against bacterial infection used by mammals
  - Low solubility of iron at physiological pH (exists as Fe^{3+} in free solution)
  - Intracellular location of iron
  - Sequestration of Fe within iron-binding proteins
    - Ferritin, transferin, lactoferin, haemoglobin, myoglobin
- In humans, 75% of total Fe is bound to haemoglobin!

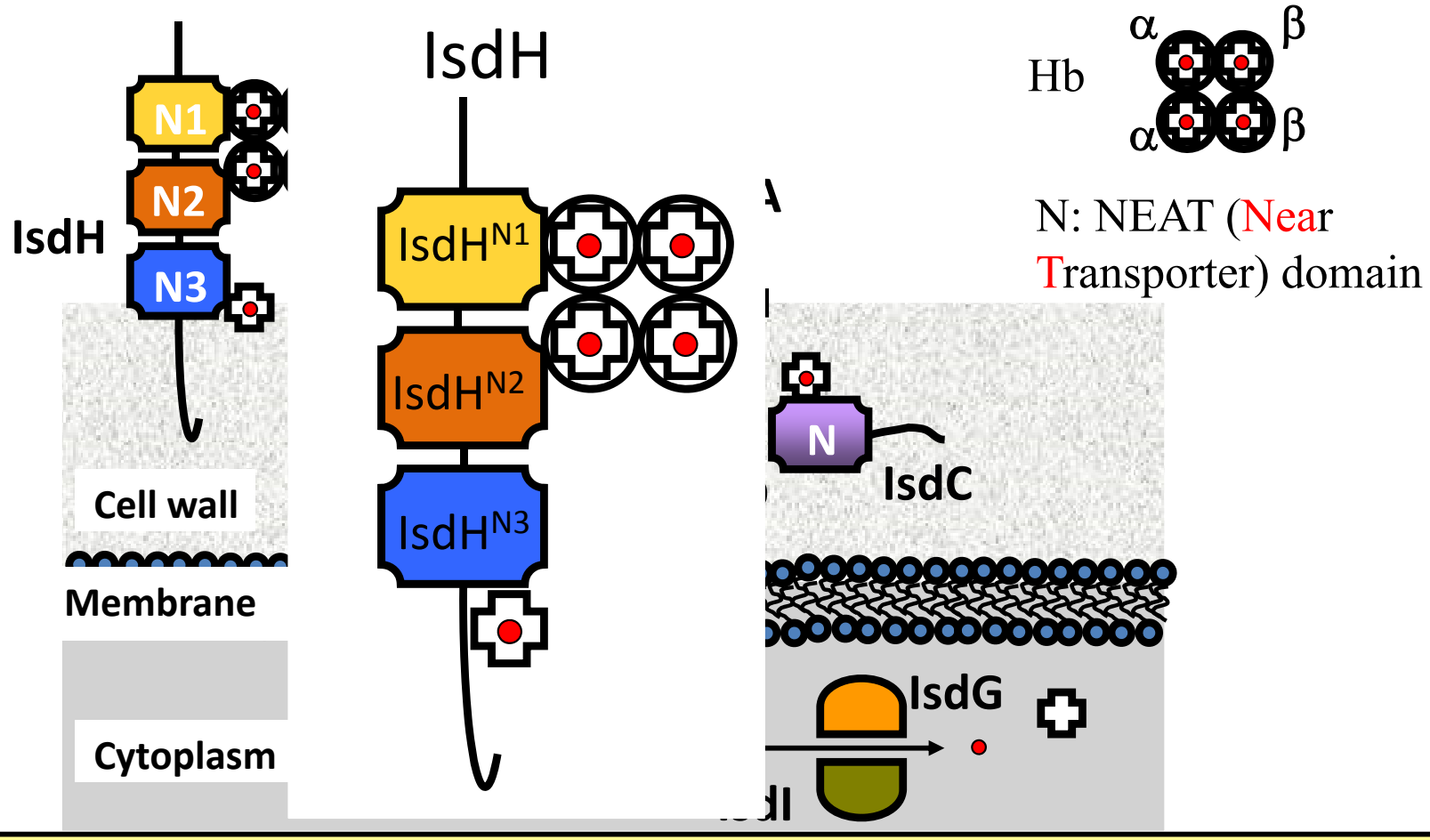
Iron acquisition by *Staphylococcus aureus*

- Siderophores bind Fe very tightly with affinities ranging over 30 orders to magnitude (enterochelin K_d 10^{52} M^{-1})
  - Bind free iron or rip iron from proteins such as transferrin
- Surface receptors, IsdB and IsdH can bind and extract haem from haemoglobin



Hammer & Skaar, Annu. Rev. Micro. 2011, 65:129.

Iron acquisition system
- Isd (Iron regulated surface determinant) system



In IsdH, N1 & N2 bind Hb but *not* haem; N3 binds haem but *not* Hb. N1 also binds haptoglobin but less tightly than Hb (35 vs 10 nM). N2 & N3 of IsdH are homologous with N1 and N2 of IsdB.

Questions to be addressed

- How does IsdH bind haemoglobin (Hb)?
- How is the haem extracted from Hb and transferred to IsdH^{N3}?

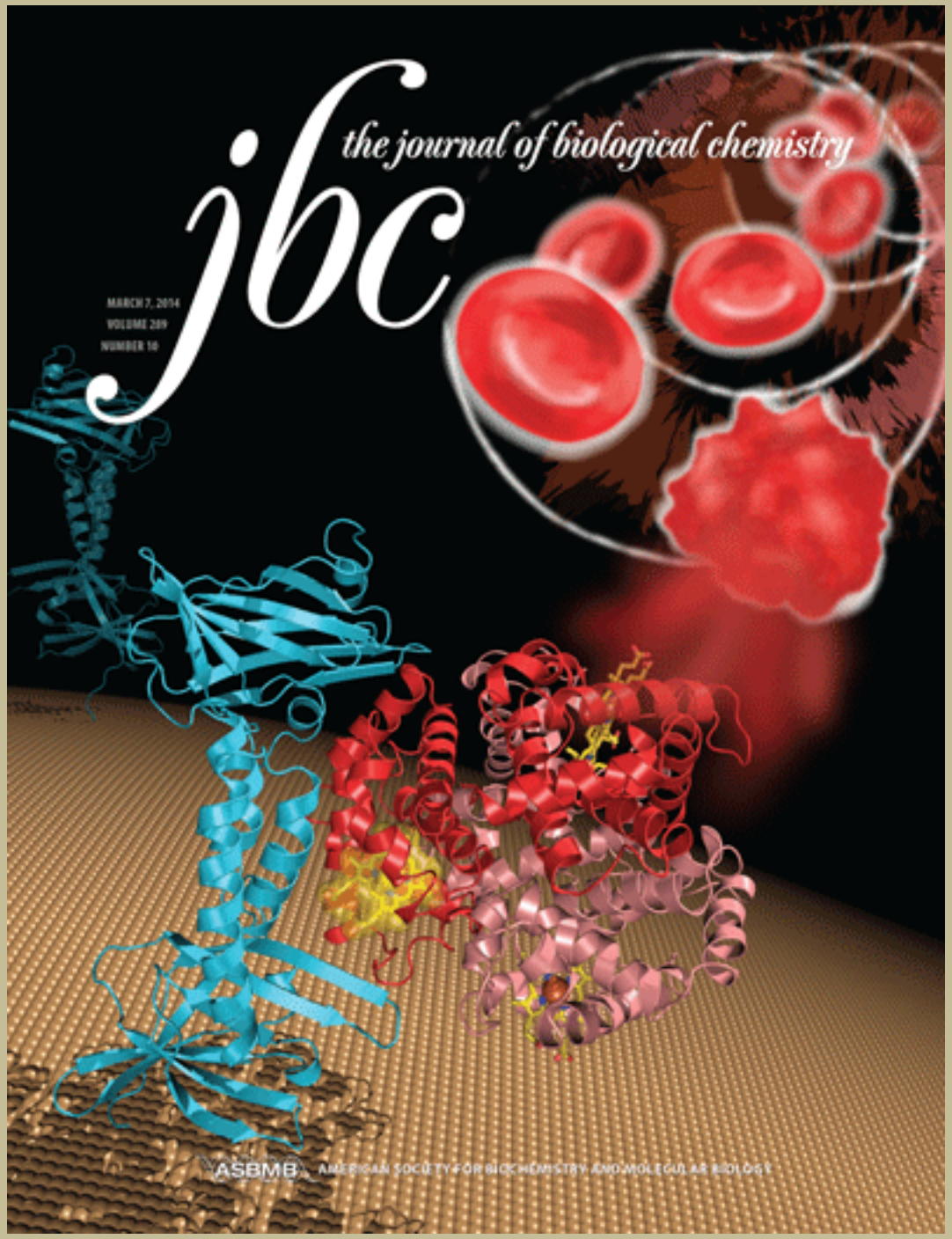
Pre-existing structural work addressed the issue of haem binding:

1. Structure of IsdA and IsdA-haem complex (Murphy and colleagues, X-ray; Clubb and colleagues, NMR)
2. IsdB^{N2}-haem complex (Murphy and colleagues; IsdB^{N2} is homologous to IsdH^{N3})
3. IsdC-haem complex (Paoli and colleagues)
4. IsdH^{N3}-haem complex (Tsumoto and colleagues, X-ray)

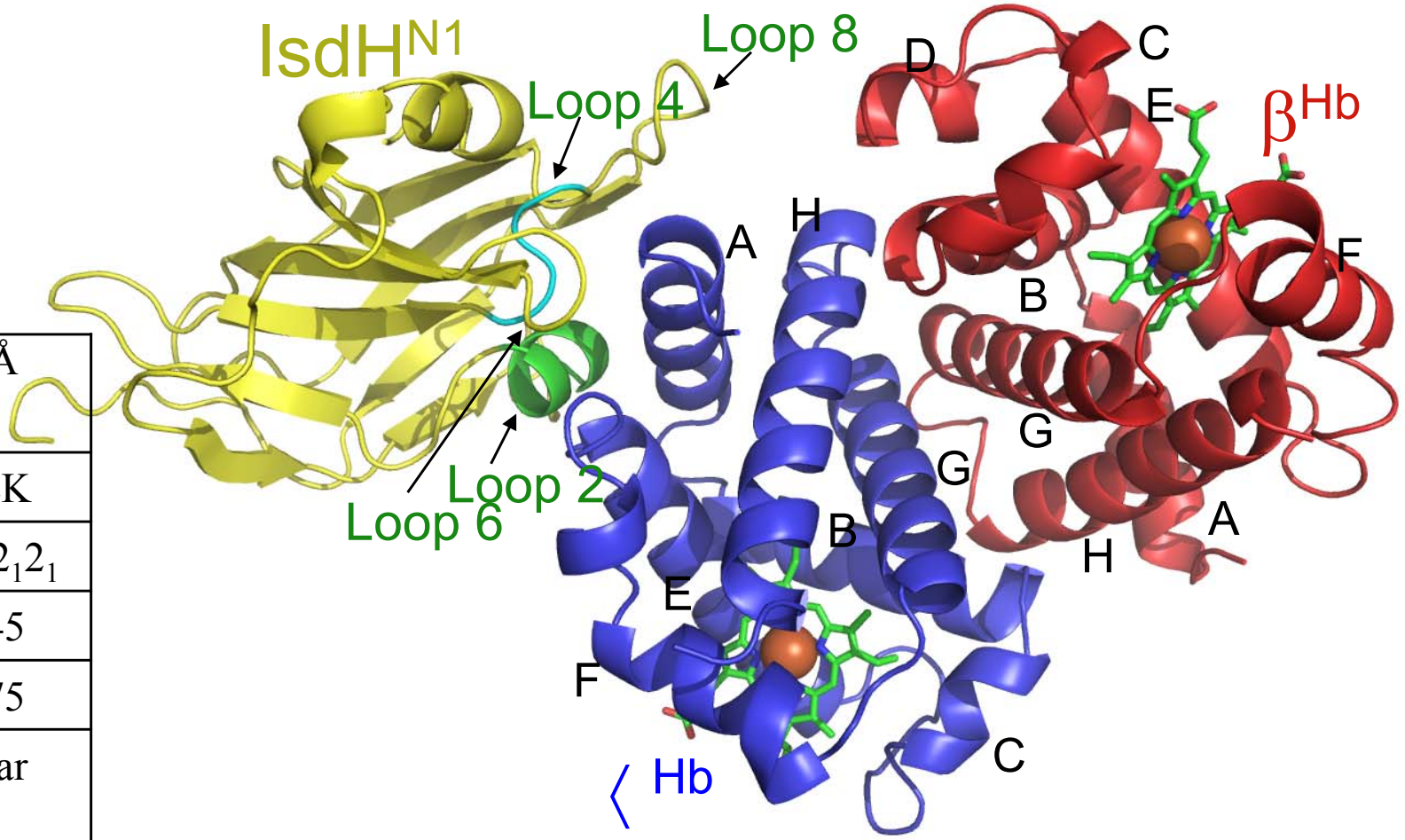
Our structural work:

1. IsdH^{N1} bound to Hb (Krishna Kumar *et al. J. Biol. Chem. 2011*)
2. IsdH^{N1} bound to Hb$^{\alpha}$ (Dickson *et al. Acta Cryst. 2014*)
3. IsdH^{N2} & IsdH^{N2N3} bound to Hb (Dickson *et al. J. Biol. Chem. 2014*)

Techniques used: X-ray crystallography, NMR spectroscopy, small angle X-ray scattering +


the journal of biological chemistry
jbc
MARCH 7, 2014
VOLUME 289
NUMBER 10
ASBMB
AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BIOLOGY

Structure of HbA(αβ):IsdH^{N1} complex



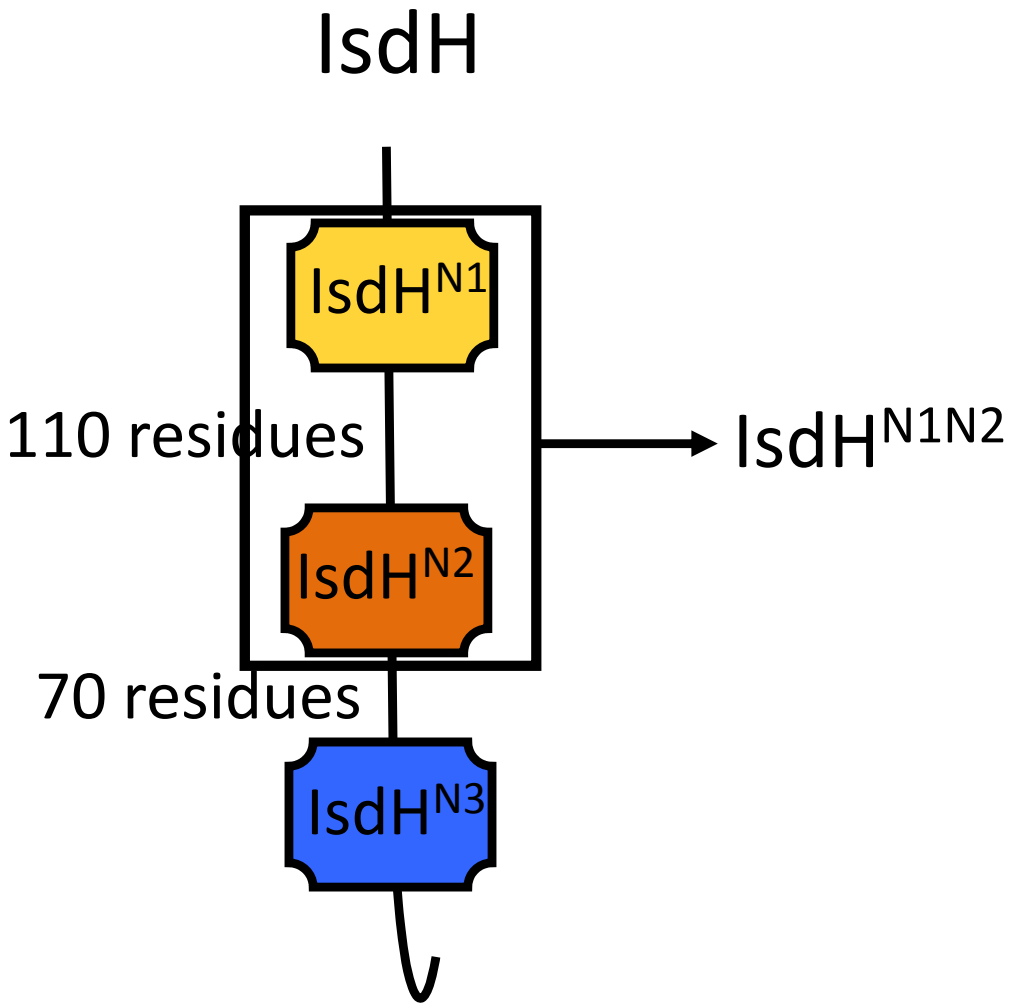
Resolution 3.0 Å	
PDB code	3SZK
Space group	$P2_12_12_1$
R_{cryst}	0.245
R_{free}	0.275
Solved by molecular replacement with either α^{Hb} or β^{Hb} as search model	

Krishna Kumar *et al.* J Biol Chem. (2011) 286, 38439-47.

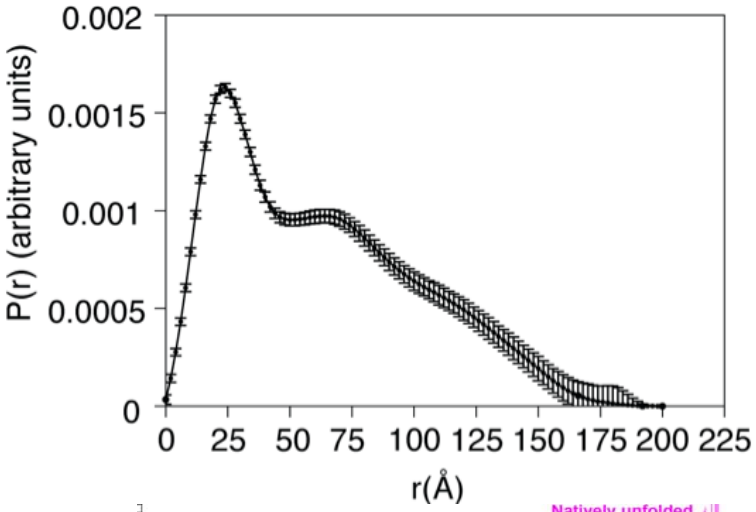
Key findings from IsdH^{N1} work

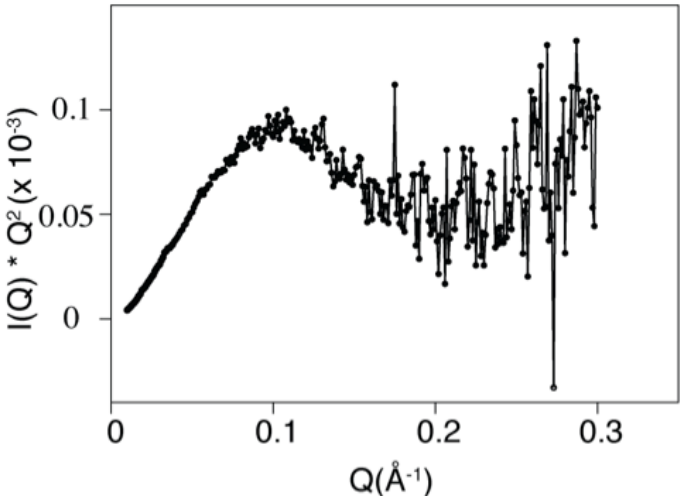
- IsdH^{N1} binds HbA through its α-subunit
- Sequence differences can explain the preference for α^{Hb} over β^{Hb}
- K11 of α^{Hb} is important for binding
  - Mutant αK11T does not bind IsdH^{N1}
- Binding is on opposite side to the haem pocket

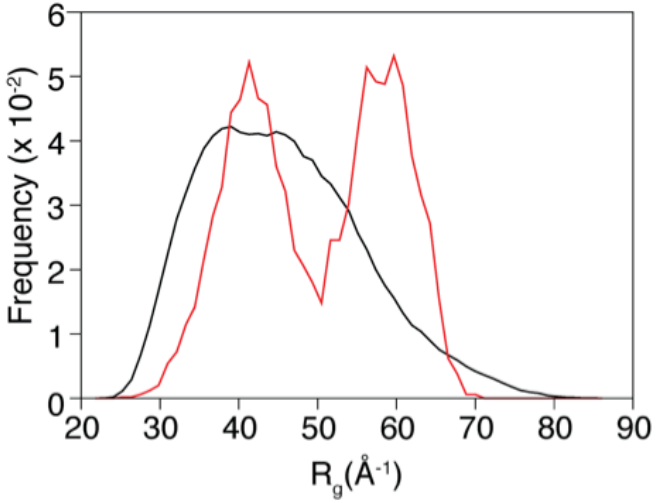
Haemoglobin receptors- IsdH

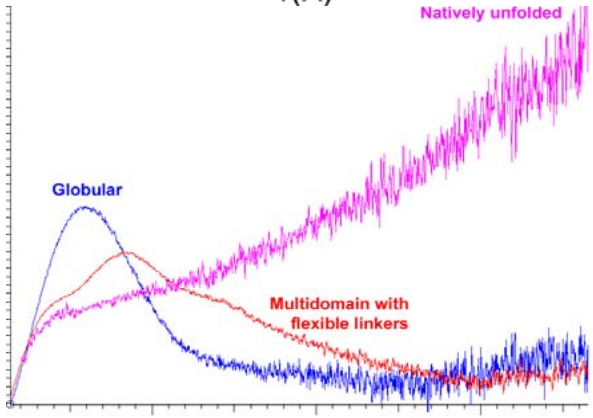


SAXS: IsdH^{N1} & IsdH^{N2} attached by a flexible linker

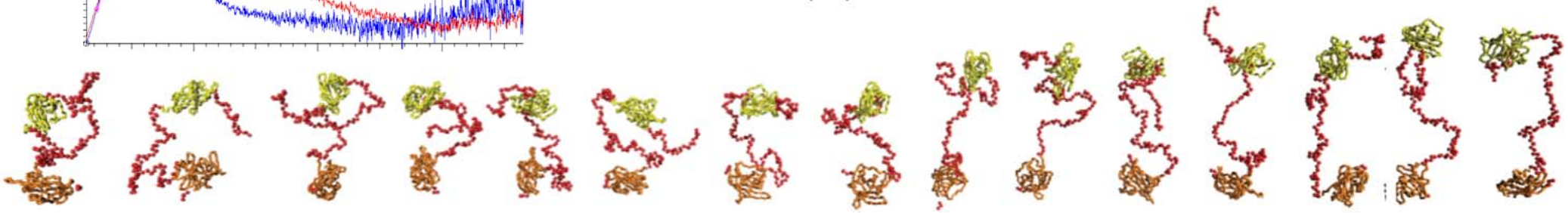






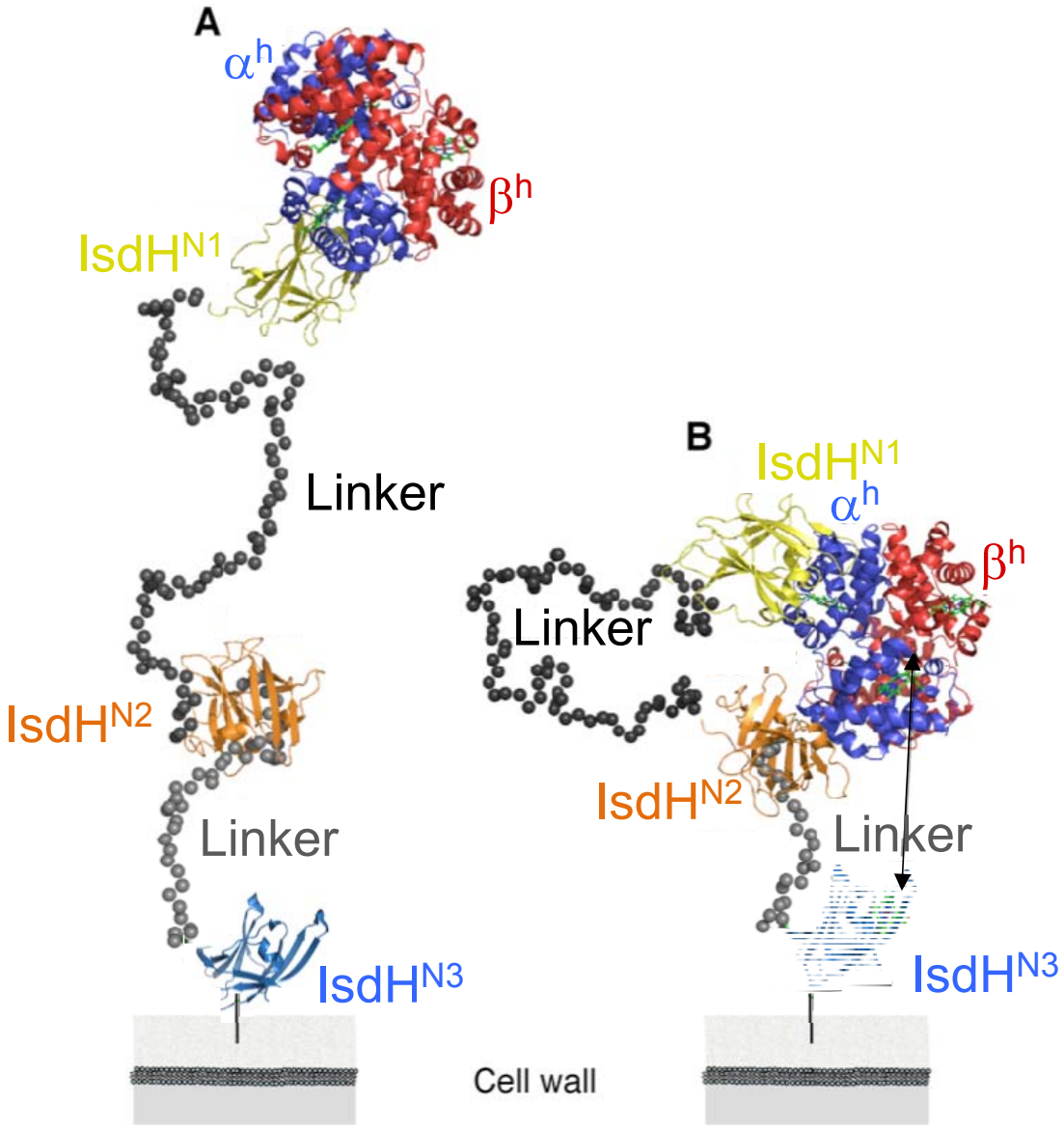


Can model the SAXS data with a flexible model with 2 populations – one with a compact form and the other with an extended form.

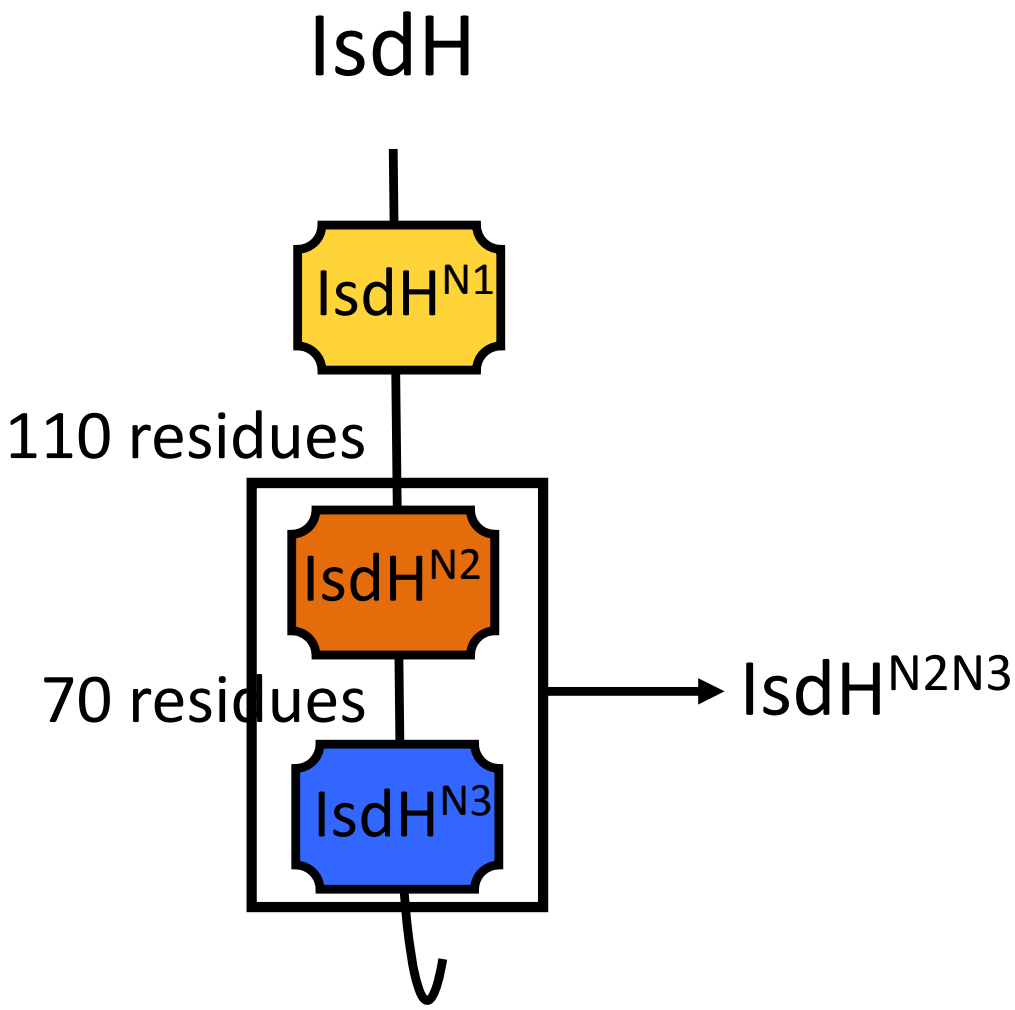


Manuscript in preparation

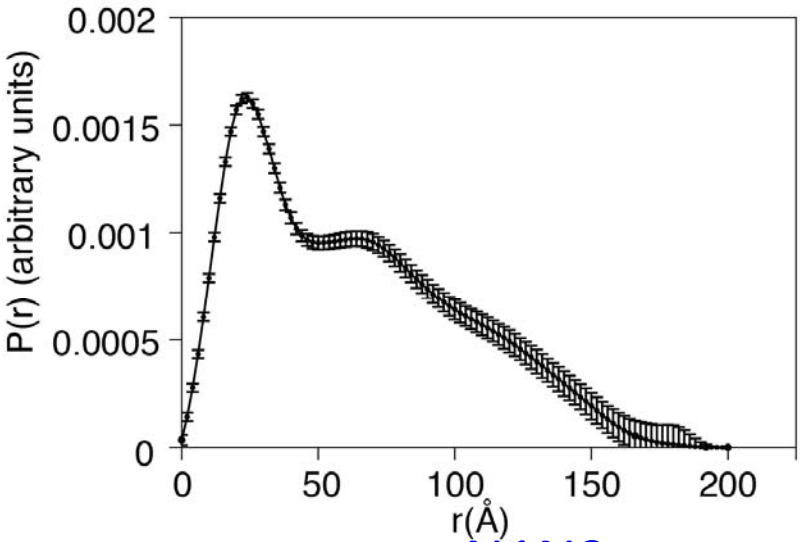
Model for HbA capture by IsdH

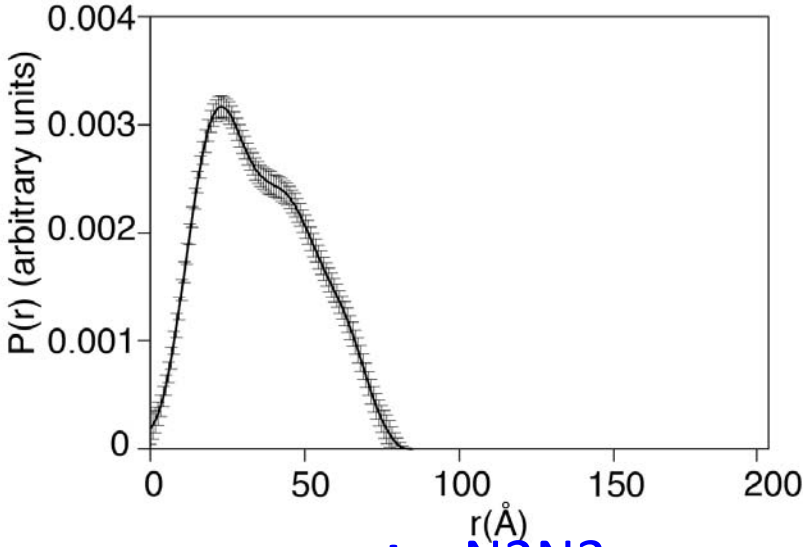


Haemoglobin receptors- IsdH



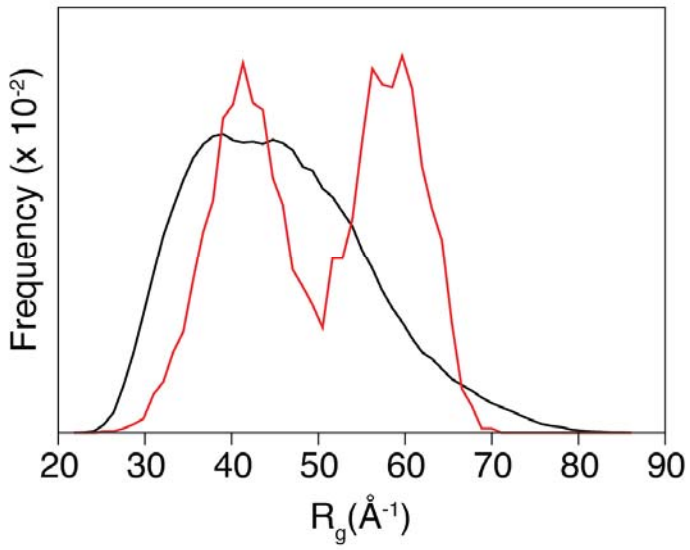
SAXS: IsdH^{N2N3} appears to be rigid

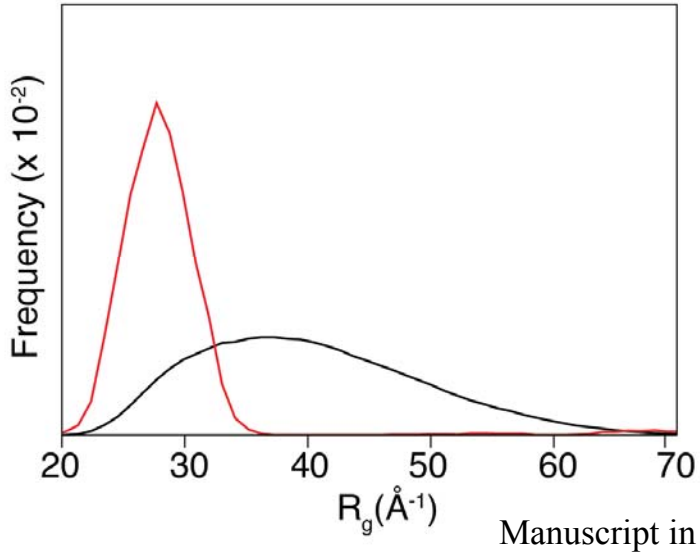




IsdH^{N1N2}

IsdH^{N2N3}

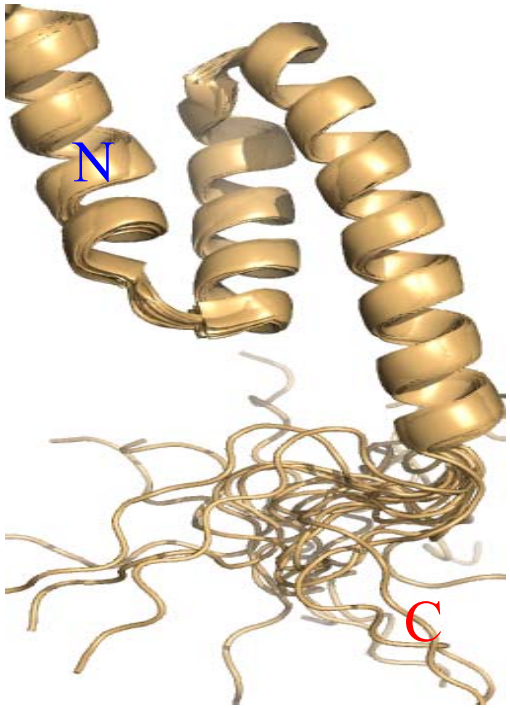


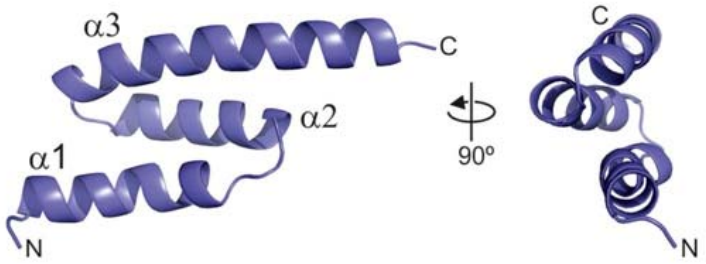


Manuscript in preparation

IsdH^{N2N3} linker is structured

PDB code: 2LHR





Structured region Val^{470}-Asp^{531}

20 lowest energy structures

Structural statistics for the solution structure of IsdH linker domain

The notation of the NMR structures is as follows. ⟨SA⟩ are the final 20 simulated annealing structures; (SA) is the average energy-minimized structure. The number of terms for each restraint is given in parentheses.

	⟨SA⟩[a]	(SA)
Root mean square. deviations		
NOE interproton distance restraints (Å) (1469)	0.046 ± 0.002	0.051
Dihedral angle restraints (degrees)[b] (118)	0.072 ± 0.099	0.306
$^{3}J_{HN}{}^{a}$ coupling constants (Hz) (54)	0.532 ± 0.018	0.543
Secondary ^{13}C shifts (ppm)		
$13C_{\alpha}$ (76)	1.212 ± 0.206	1.267
$^{13}C_{\beta}$ (76)	0.816 ± 0.206	0.791
Deviations from idealized covalent geometry		
Bonds (Å)	0.0044 ± 0.0002	0.0174
Angles (degrees)	0.623 ± 0.028	1.538
Impropers (degrees)	0.492 ± 0.031	1.181
PROCHECK results (%)[c]		
Most favorable region	96.7 ± 2.8	96.7
Additionally allowed region	3.3 ± 2.8	3.3
Generously allowed region	0.0 ± 0.0	0.0
Disallowed region	0.0 ± 0.0	0.0
Coordinate precision (Å)[d]		
Protein backbone	0.42 ± 0.10	
Protein heavy atoms	0.87 ± 0.07	

[a] None of the structures exhibits distance violations greater than 0.5 Å, dihedral angle violations greater than 5°, or coupling constant violations greater than 2 Hz.
[b] Experimental dihedral angle restraints comprised 48 φ, 48 ψ, and 16 χ_1 angles.
[c] PROCHECK-NMR data include residues Val^{470}–Val^{531} of the linker domain.
[d] The coordinate precision is defined as the average atomic root mean square deviation of the 20 individual simulated annealing structures and their mean coordinates. The reported values are for residues Val^{470}–Val^{531} of the linker domain.

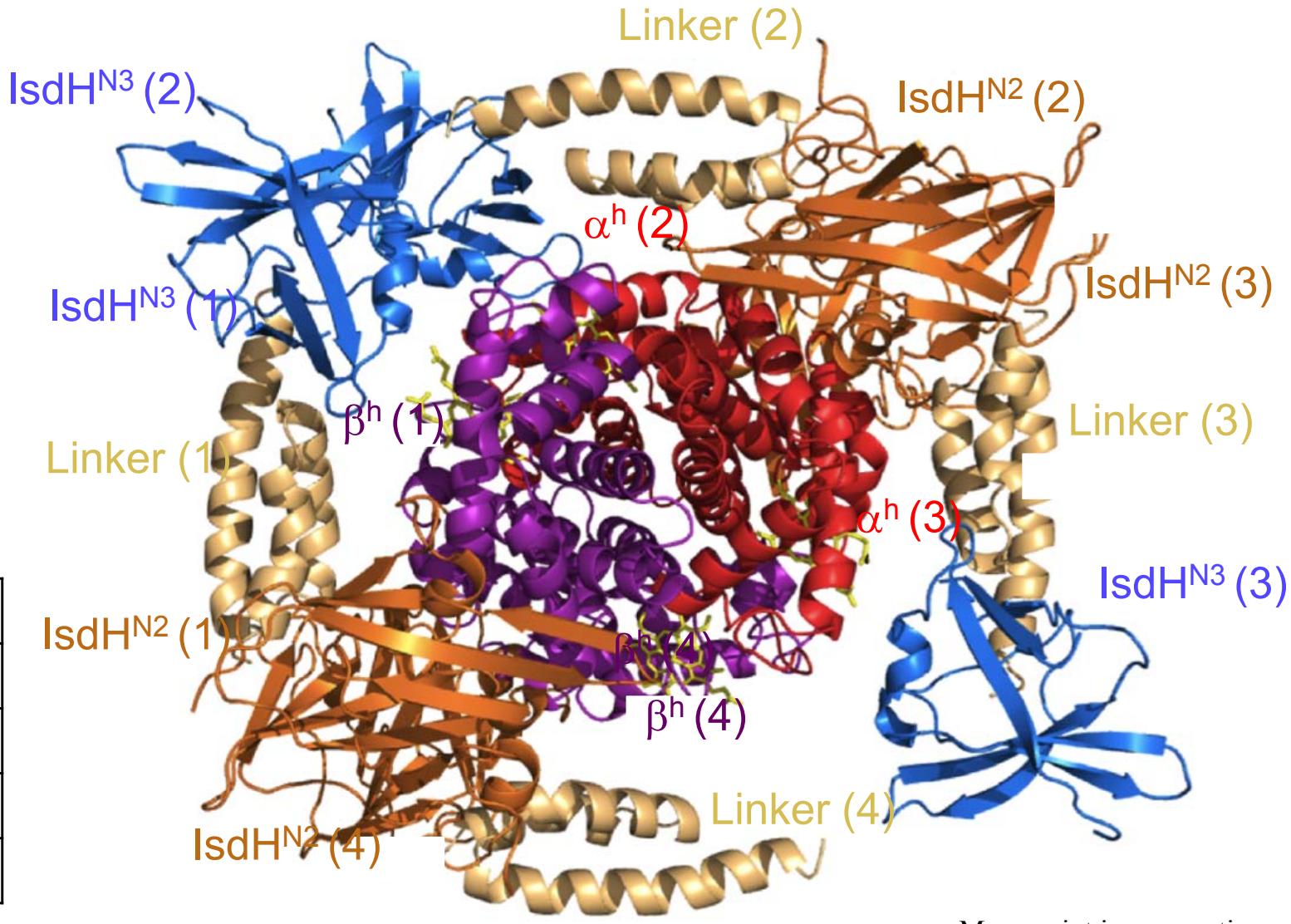
Spirig et al., J Biol Chem (2012) 288, 1065-1078.

Structure of IsdH^{N2N3} bound to HbA

Crystals:	HbA and Isd^{N2N3} (Y642A) purified complex
Diffraction data: synchrotron	MX2 Australian
Structure solution:	molecular replacement
Search models:	Stepwise searches with: Hb, IsdH^{N2} and IsdH^{N3}
Program:	PHASER – 4 IsdH^{N2N3} bound to HbA tetramer (1 N3 domain absent) N2-N3 linker was not included in search but clear electron-density was seen. NMR model was fitted to the density.
Refinement:	BUSTER with strong geometric and NCS restraints

	IsdH^{N2}:Hb	IsdH$^{N2\text{-}N3(Y642A)}$:Hb
Data collection		
Space group		P 2_1 2_1 2
Cell dimensions		
a, *b*, *c* (Å)		132.90, 185.30, 103.21
α, β, γ (°)		90, 90, 90
Resolution (Å)		49.7–4.23 (4.24–4.32)
R_{merge} (%)		9.2 (72.7)
$I / \sigma I$		12.5 (1.83)
Completeness (%)		99.9 (100)
Redundancy		3.8 (3.8)
Refinement		
Resolution (Å)		29.15–4.24
No. reflections		18514
R_{work} / R_{free}		0.310/0.299
No. atoms		
Protein		13375
Ligand/ion		172
Water		-
B-factors		
Protein		86.14
Ligand/ion		65.65
Water		-
R.m.s. deviations		
Bond lengths (Å)		0.008
Bond angles (°)		0.84

Structure of IsdH^{N2N3} bound to HbA

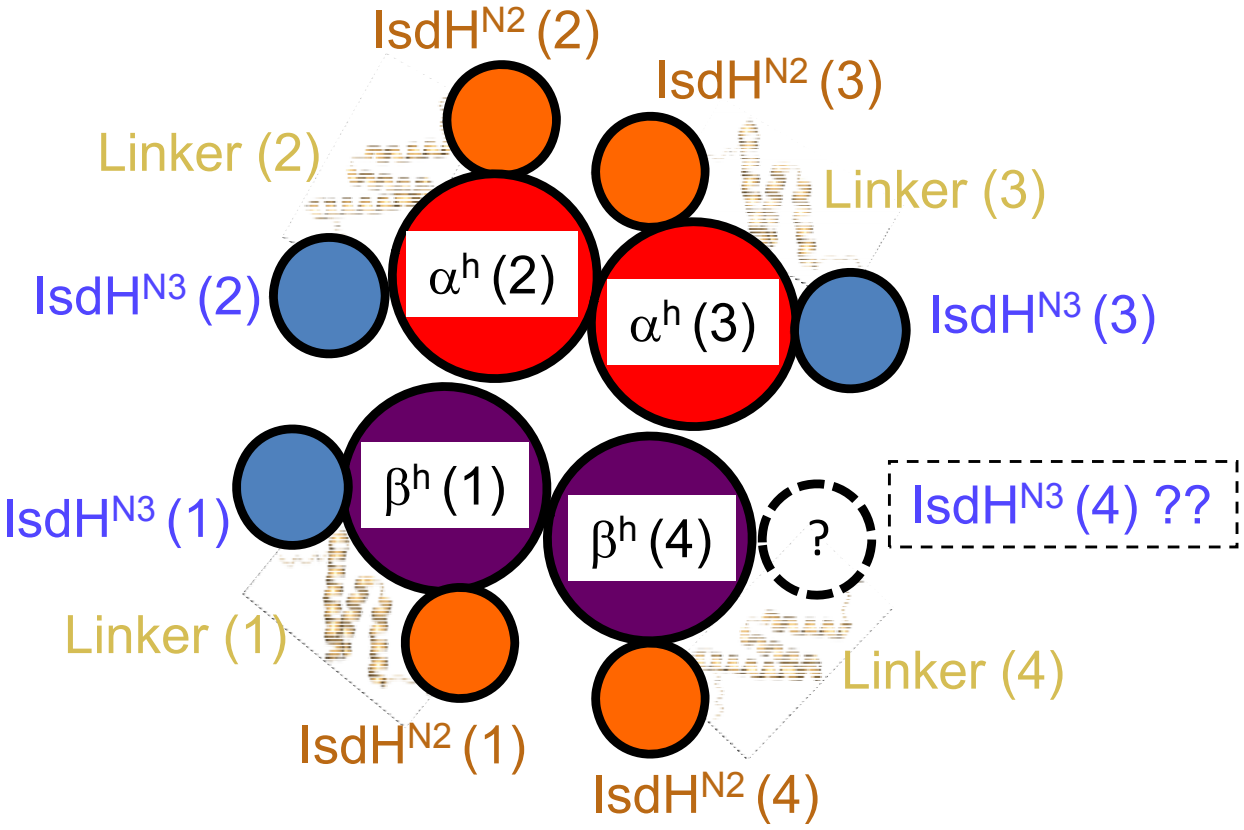


4.2 Å resolution	
PDB code	3FC4
Space group	$P2_12_12$
R_{cryst}	0.310
R_{free}	0.299

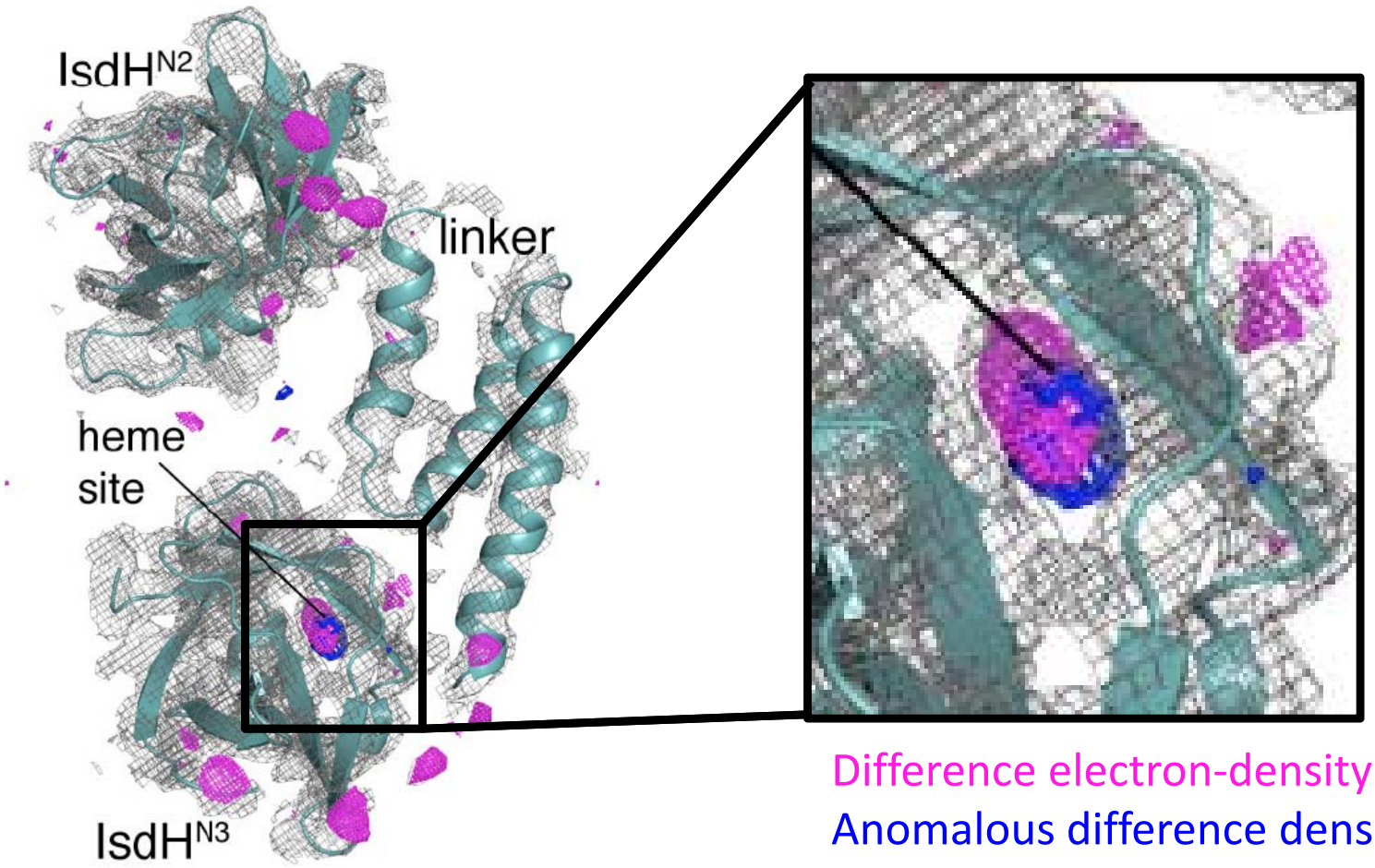
Manuscript in preparation

Structure of IsdH^{N2N3} bound to HbA

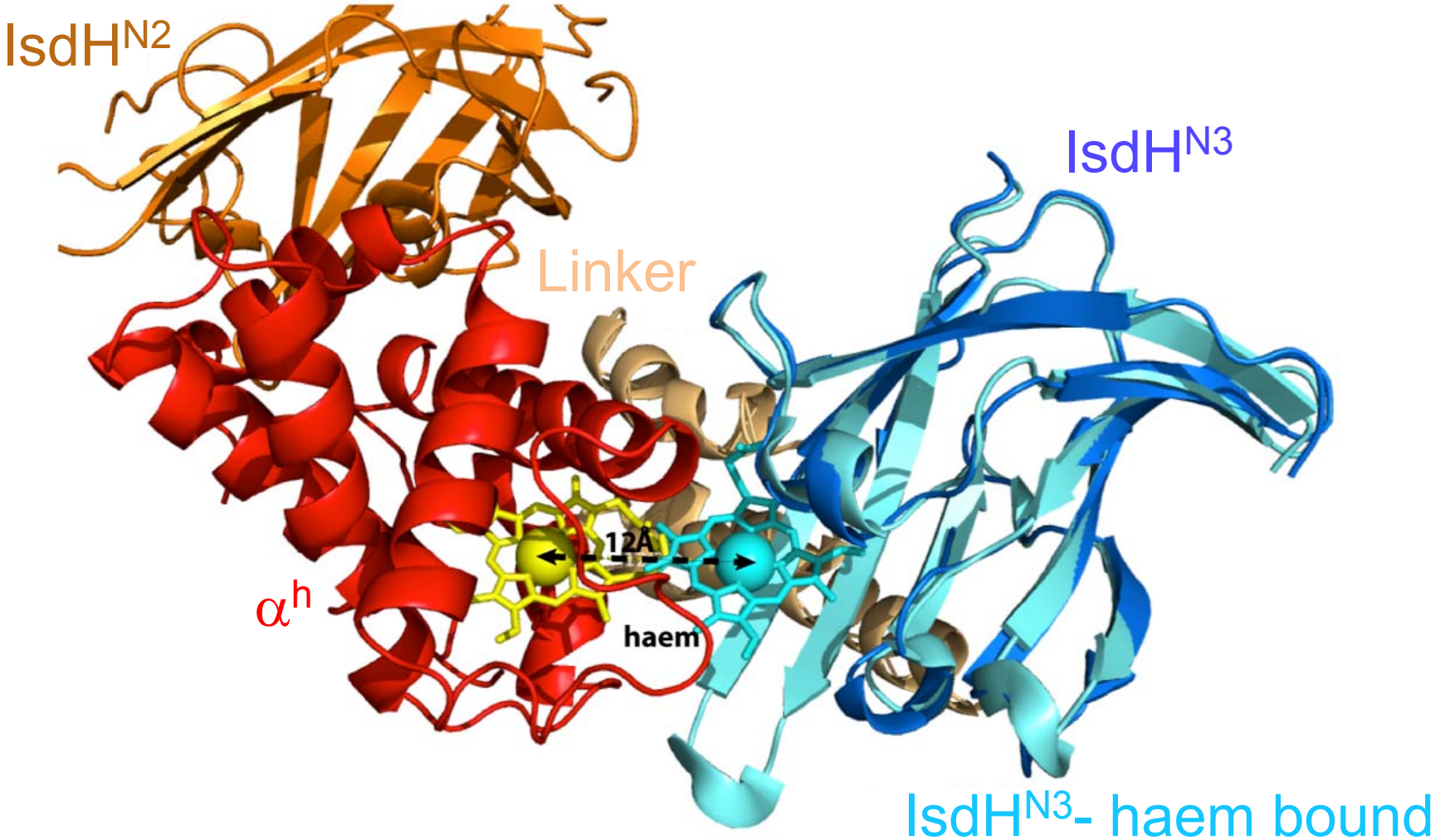
- IsdH^{N2N3} are bound to both α and β chains of Hb
- Density is seen for the links between the domains but not good enough to model
- The N3 domain of the 4th Isd molecule is not seen



Haem transfer to IsdH^{N3} from α^{Hb}

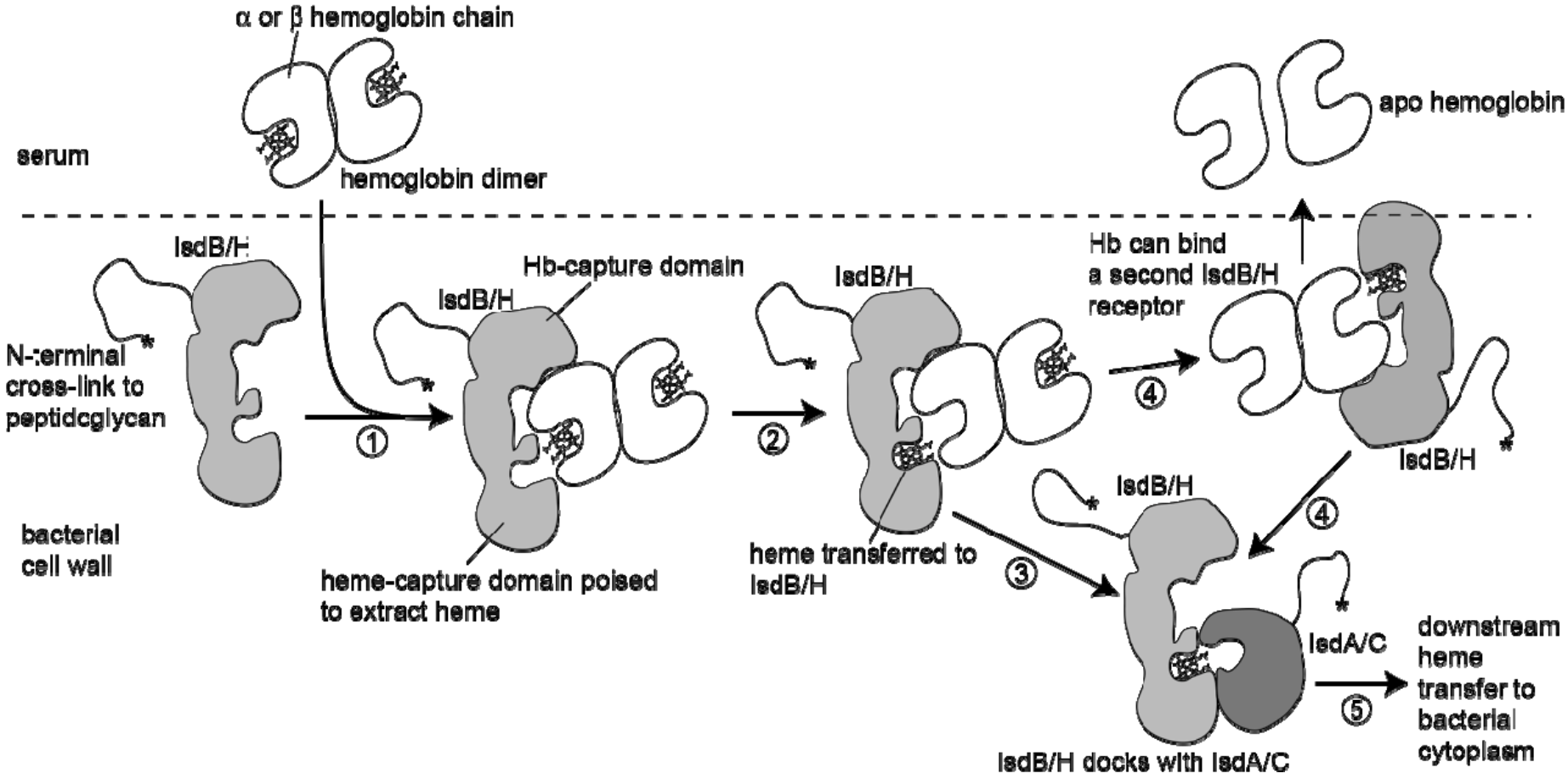


Haem transfer to IsdH^{N3} from α^{Hb}



PDB code: 2Z6F

Model for IsdH capture of HbA



Acknowledgement

David Gell

Kaavya Krishna Kumar

Claire Dickson

David Jacques

Joel Mackay

Tom Caradoc-Davies
Australian Synchrotron

Collaborators

Department of Chemistry and Biochemistry, UCLA, USA

Robert Clubb
Thomas Spirig
Reza Malmirchegini

Medical School, Vanderbilt University, USA

Eric P. Skaar
Gleb Pishchany

WHAT DOES THE FUTURE HOLD?

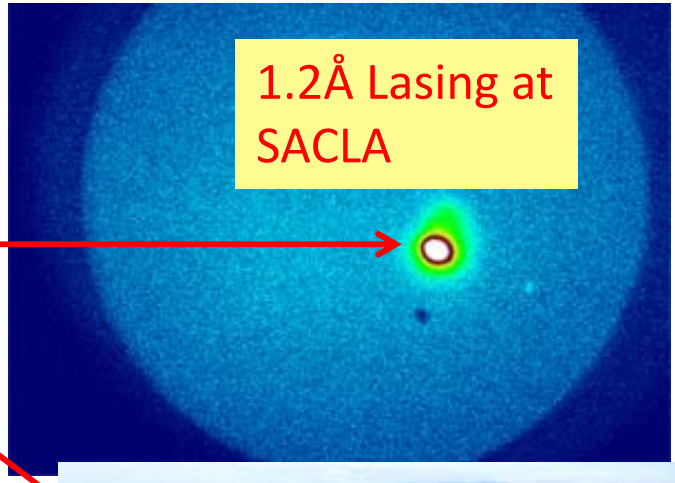
The future with and without crystals

- Modeling structures without any experimental data
  - By homology with reference to the increasingly large library of experimentally determined structures (currently >100,000 in the PDB)
  - Ab initio using evermore sophisticated energy algorithms – THIS IS THE HOLY GRAIL OF THEORETICAL STRUCTURAL BIOLOGY
- Free electron lasers providing many orders of magnitude more X-rays in very short pulses
  - Time-resolved studies of fast reactions
  - Structures from micro-crystals
  - Structures from single particles or – "crystallography without crystals" – THE DREAM OF FELs FOR BIOLOGY
- Increasing power of electron microscopy with new direct electron detectors
  - Single particle averaging with cryo electron microscopy has yielded structures to 3.5 Å resolution – MAY SUPPLANT THE NEED FOR FELs FOR LARGE STRUCTURAL COMPLEXES

FEL Facilities

Built/ In construction

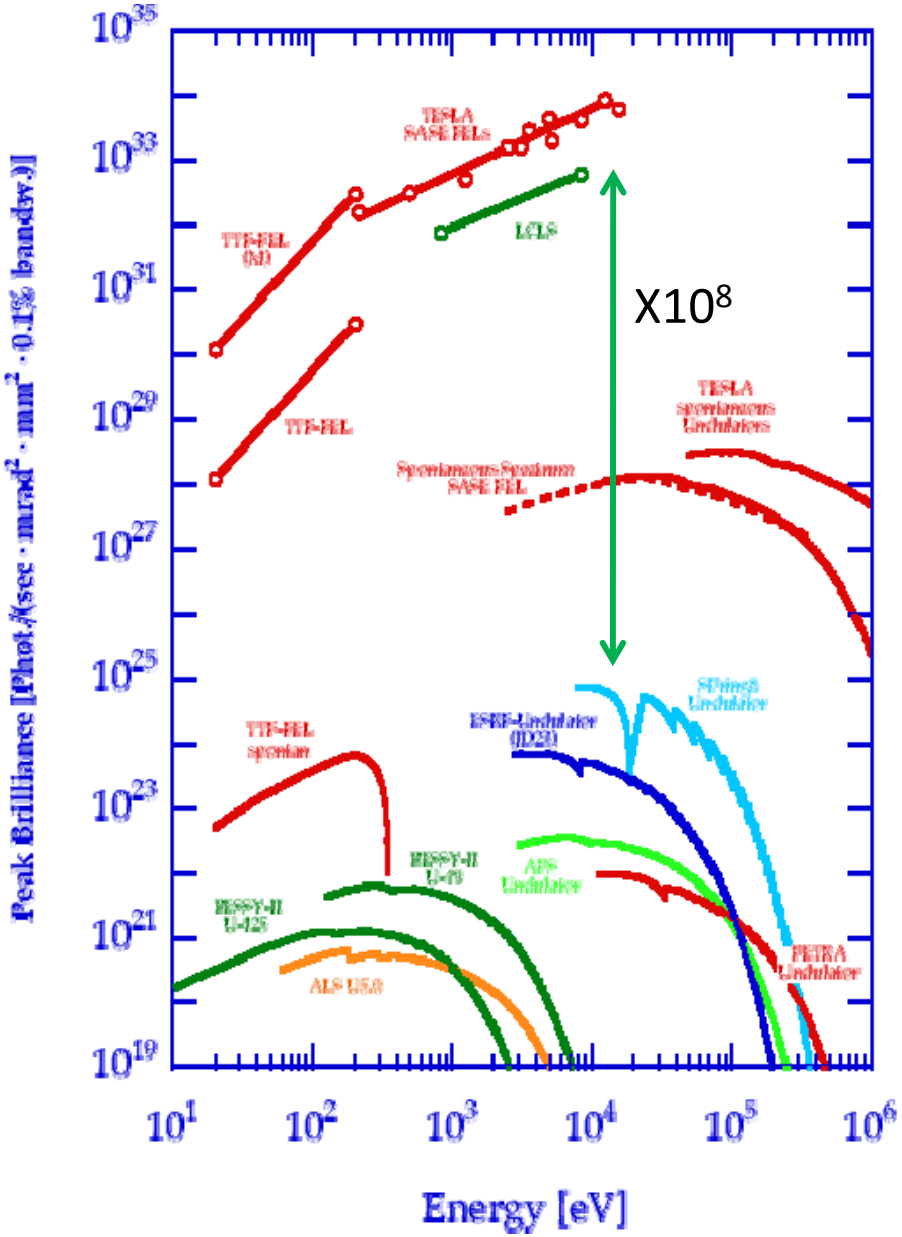
- DESY / TTF->FLASH: (2000)
  - Superconducting LINAC, ~1.2 GeV, 1000Å -> 41Å
- SLAC / LCLS (2007):
  - Room Temperature LINAC, 15 GeV 25Å -> 1.2Å
- Trieste / FERMI (2010):
  - Room Temperature LINAC, 1.2 GeV 1000Å - 100Å (planned)
- Spring-8 SACLA (2011)
  - Room Temperature LINAC, C-band, 8 GeV 1.2Å -> (0.8Å planned)
- DESY / European-XFEL (est 2015)
  - Superconducting LINAC, 17.5) GeV, 1Å – 60Å

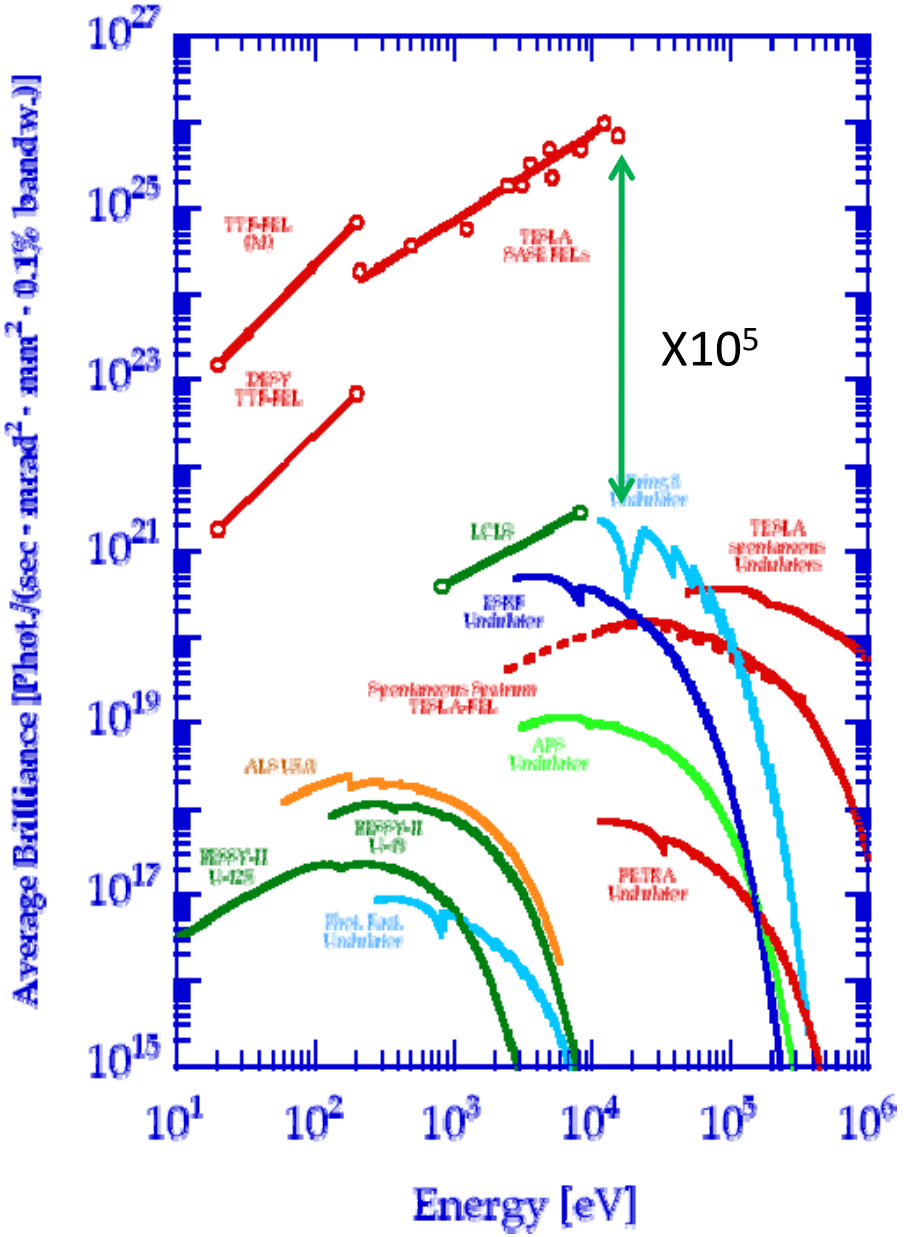


Proposed / Funded (Parameters and schedule subject to change)

- Pohang XFEL (~2015)
  - Room Temperature 10 GeV LINAC
  - 0.6Å to 50Å (planned)
- PSI Swiss-FEL (~2016)
  - Room Temperature 6 GeV LINAC (C-band)
  - 1Å - 70Å
- Shanghai XFEL (?)
  - Room temperature LINAC, 6 GeV ~1Å
- SLAC LCLS_II (2017)
  - Room Temperature LINAC 14 GeV 50Å- 1Å
- LBNL NGLS (2020s)
  - CW superconducting LINAC, 2 GeV, 50Å- 10Å
- LANL MaRIE (2020s)
  - Room Temperature LINAC, 0.25Å

Brilliance Comparison





Recent experiments - highlights

- Gedunken experiment – Hajdu
- Mimivirus – Chapman, Hajdu
- Photosystem I – Chapman, Fromme, Spence
- Cathepsin B - Chapman

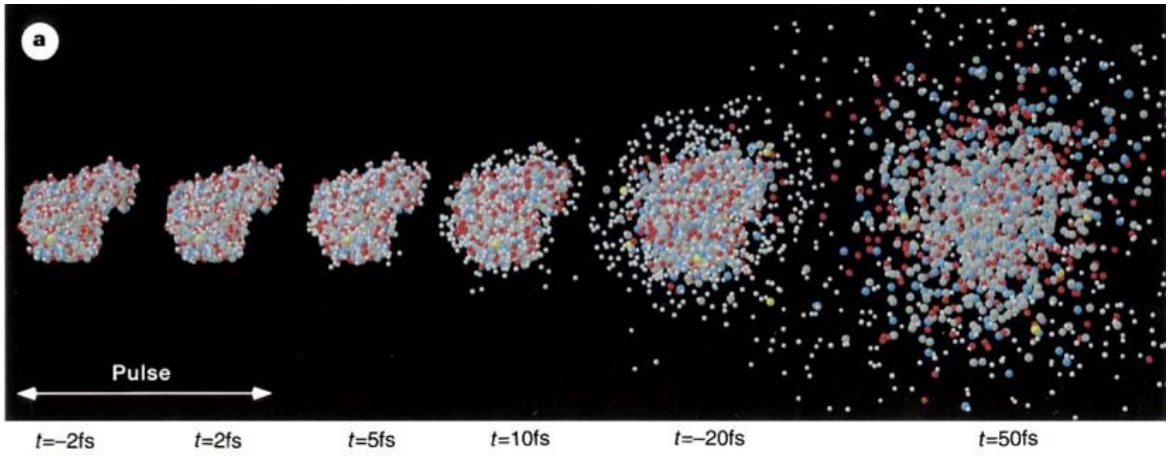
Hajdu's thought experiment – *Nature,* August 2000

Potential for biomolecular imaging with femtosecond X-ray pulses

Richard Neutze*, Remco Wouts*, David van der Spoel*, Edgar Weckert†‡ & Janos Hajdu*

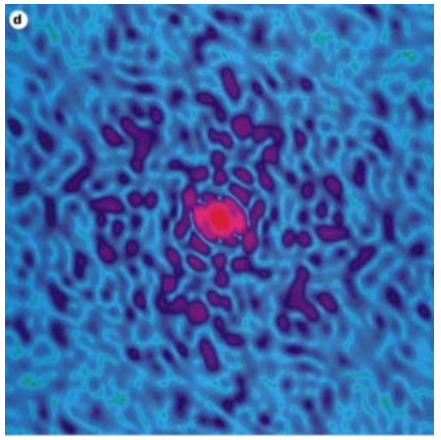
** Department of Biochemistry, Biomedical Centre, Box 576, Uppsala University, S-75123 Uppsala, Sweden*
† Institut für Kristallographie, Universität Karlsruhe, Kaiserstrasse 12, D-76128 , Germany

Sample damage by X-rays and other radiation limits the resolution of structural studies on non-repetitive and non-reproducible structures such as individual biomolecules or cells[1]. Cooling can slow sample deterioration, but cannot eliminate damage-induced sample movement during the time needed for conventional measurements[1,2]. Analyses of the dynamics of damage formation[3–5] suggest that the conventional damage barrier (about 200 X-ray photons per Å^2 with X-rays of 12 keV energy or 1 Å wavelength[2]) may be extended at very high dose rates and very short exposure times. Here we have used computer simulations to investigate the structural information that can be recovered from the scattering of intense femtosecond X-ray pulses by single protein molecules and small assemblies. Estimations of radiation damage as a function of photon energy, pulse length, integrated pulse intensity and sample size show that experiments using very high X-ray dose rates and ultrashort exposures may provide useful structural information before radiation damage destroys the sample. We predict that such ultrashort, high-intensity X-ray pulses from free-electron lasers[6,7] that are currently under development, in combination with container-free sample handling methods based on spraying techniques, will provide a new approach to structural determinations with X-rays.



(Above) Simulation of the destruction of a lysozyme molecule with an 2 fs (FWHM) X-ray pulse of 3 x 10^{12} (12 keV) photons per 100-nm diameter spot.

(Right) Simulated planar section through the continuous scattering image of a single T4 lysozyme molecule.



We predict that such ultrashort, high-intensity X-ray pulses from free-electron lasers[6,7] that are currently under development, in combination with container-free sample handling methods based on spraying techniques, will provide a new approach to structural determinations with X-rays.

The real thing – structure of photosystem I

LETTER *Nature* **470**, 73-77, 2011
doi:10.1038/nature09750

Femtosecond X-ray protein nanocrystallography

Henry N. Chapman[1,2], Petra Fromme[3], Anton Barty[1], Thomas A. White[1], Richard A. Kirian[4], Andrew Aquila[1], Mark S. Hunter[3], Joachim Schulz[1], Daniel P. DePonte[1], Uwe Weierstall[4], R. Bruce Doak[4], Filipe R. N. C. Maia[5], Andrew V. Martin[1], Ilme Schlichting[6,7], Lukas Lomb[7], Nicola Coppola[1]†, Robert L. Shoeman[7], Sascha W. Epp[6,8], Robert Hartmann[9], Daniel Rolles[6,7], Artem Rudenko[6,8], Lutz Foucar[6,7], Nils Kimmel[10], Georg Weidenspointner[11,10], Peter Holl[9], Mengning Liang[1], Miriam Barthelmess[12], Carl Caleman[1], Sébastien Boutet[13], Michael J. Bogan[14], Jacek Krzywinski[13], Christoph Bostedt[13], Saša Bajt[12], Lars Gumprecht[1], Benedikt Rudek[6,8], Benjamin Erk[6,8], Carlo Schmidt[6,8], André Hömke[6,8], Christian Reich[9], Daniel Pietschner[10], Lothar Strüder[6,10], Günter Hauser[10], Hubert Gorke[15], Joachim Ullrich[6,8], Sven Herrmann[10], Gerhard Schaller[10], Florian Schopper[10], Heike Soltau[9], Kai-Uwe Kühnel[8], Marc Messerschmidt[13], John D. Bozek[13], Stefan P. Hau-Riege[16], Matthias Frank[16], Christina Y. Hampton[14], Raymond G. Sierra[14], Dmitri Starodub[14], Garth J. Williams[13], Janos Hajdu[5], Nicusor Timneanu[5], M. Marvin Seibert[5]†, Jakob Andreasson[5], Andrea Rocker[5], Olof Jönsson[5], Martin Svenda[5], Stephan Stern[1], Karol Nass[2], Robert Andritschke[10], Claus-Dieter Schröter[8], Faton Krasniqi[6,7], Mario Bott[7], Kevin E. Schmidt[4], Xiaoyu Wang[4], Ingo Grotjohann[3], James M. Holton[17], Thomas R. M. Barends[7], Richard Neutze[18], Stefano Marchesini[17], Raimund Fromme[3], Sebastian Schorb[19], Daniela Rupp[19], Marcus Adolph[19], Tais Gorkhover[19], Inger Andersson[20], Helmut Hirsemann[12], Guillaume Potdevin[12], Heinz Graafsma[12], Björn Nilsson[12] & John C. H. Spence[4]

X-ray crystallography provides the vast majority of macromolecular structures, but the success of the method relies on growing crystals of sufficient size. In conventional measurements, the necessary increase in X-ray dose to record data from crystals that are too small leads to extensive damage before a diffraction signal can be recorded[1–3]. It is particularly challenging to obtain large, well-diffracting crystals of membrane proteins, for which fewer than 300 unique structures have been determined despite their importance in all living cells. Here we present a method for structure determination where single-crystal X-ray diffraction 'snapshots' are collected from a fully hydrated stream of nanocrystals using femtosecond pulses from a hard-X-ray free-electron laser, the Linac Coherent Light Source[4]. We prove this concept with nanocrystals of photosystem I, one of the largest membrane protein complexes[5]. More than 3,000,000 diffraction patterns were collected in this study, and a three-dimensional data set was assembled from individual photosystem I nanocrystals (~200 nm to 2 μm in size). We mitigate the problem of radiation damage in crystallography by using pulses briefer than the timescale of most damage processes[6]. This offers a new approach to structure determination of macromolecules that do not yield crystals of sufficient size for studies using conventional radiation sources or are particularly sensitive to radiation damage.

Radiation damage has always limited resolution in biological imaging using electrons or X-rays[2]. With the recent invention of the femtosecond X-ray laser, an opportunity has arisen to break the nexus between radiation dose and spatial resolution. It has been proposed that femtosecond X-ray pulses can be used to outrun even the fastest damage processes by using single pulses so brief that they terminate before the manifestation of damage to the sample[6]. Experiments at the FLASH free-electron laser (FEL), Germany, confirmed the feasibility of 'diffraction before destruction' at resolution lengths down to 60 Å on test samples fixed on silicon nitride membranes[7]. It was predicted that the irradiance (or power density) of focused pulses from a hard-X-ray FEL such as the Linac Coherent Light Source (LCLS), USA, would be sufficient to produce diffraction patterns at near-atomic resolution[6].

We demonstrate here that this notion of diffraction before destruction operates at subnanometre resolution, using the membrane protein photosystem I as a model system, and establish an approach to structure determination based on X-ray diffraction data from a stream of nanocrystals[6,8]. Membrane proteins have a central role in the functioning of cells and viruses, yet our knowledge of the structure and dynamics responsible for their functioning remains limited. Photosystem I is a large membrane protein complex (1-MDa molecular mass, 36 proteins, 381 cofactors) that acts as a biosolar energy converter in the process of oxygenic photosynthesis. Its crystals display the symmetry of space group $P6_3$, with unit-cell parameters $a = b = 281$ Å and $c = 165$ Å, and consist of 78% solvent by volume. We show that diffraction data can be recorded from these fragile protein nanocrystals before destruction occurs. Furthermore, we demonstrate that structure factors can be extracted from the 'partial' reflections of tens of thousands of single-crystal diffraction snapshots, showing that interpretable high-quality, three-dimensional (3D) structure factor data can be obtained from a suspension of submicrometre crystals.

Our experimental set-up (Fig. 1 and Methods) records single-crystal diffraction data from a stream of crystals carried in a 4-μm-diameter, continuous liquid water jet[9] that flows across the focused LCLS X-ray beam in vacuum at 10 μl min^{-1}. In contrast to cryo-electron microscopy[10,11] or standard crystallography on microcrystals[3], which require cryogenic cooling, these data were collected on fully hydrated, 3D nanocrystals. The crystal located in the interaction region when an X-ray pulse arrives gives rise to a diffraction pattern that is detected on a set of two low-noise, X-ray p–n junction charge-coupled device (pnCCD) modules[12] and read out before the arrival of the next pulse at the FEL repetition rate of 30 Hz, or 1,800 patterns per minute. The

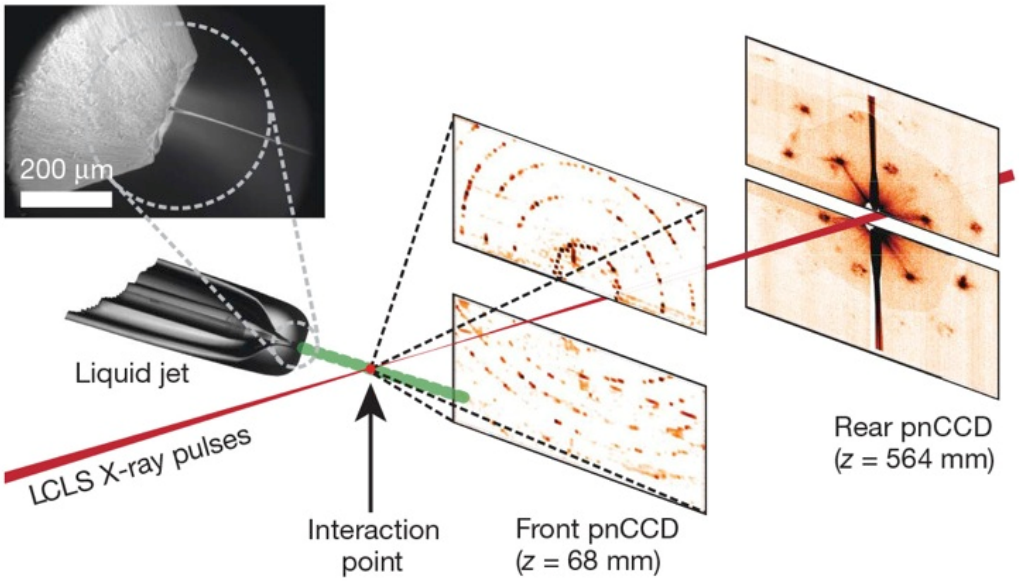
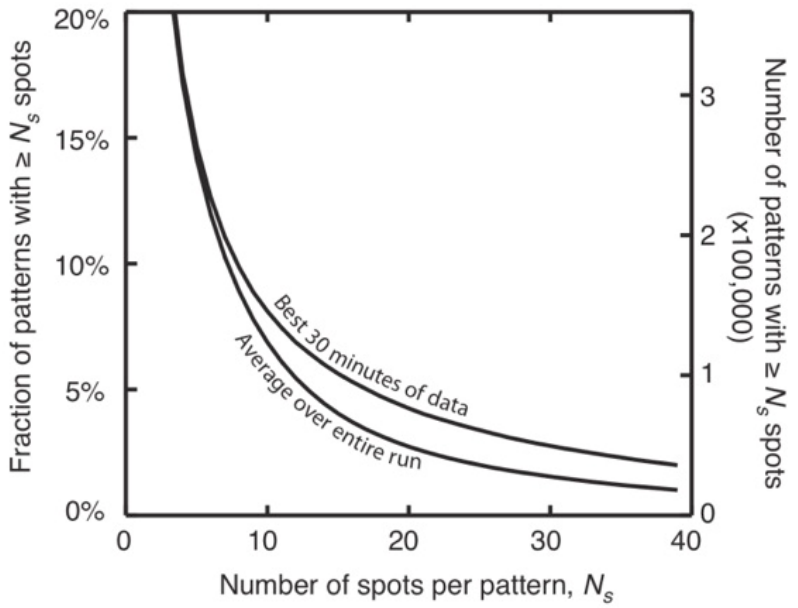


Figure 1 | Femtosecond nanocrystallography. Nanocrystals flow in their buffer solution in a gas-focused, 4-μm-diameter jet at a velocity of 10 m s^{-1} perpendicular to the pulsed X-ray FEL beam that is focused on the jet. Inset, environmental scanning electron micrograph of the nozzle, flowing jet and focusing gas[30]. Two pairs of high-frame-rate pnCCD detectors[12] record low- and high-angle diffraction from single X-ray FEL pulses, at the FEL repetition rate of 30 Hz. Crystals arrive at random times and orientations in the beam, and the probability of hitting one is proportional to the crystal concentration.

The real thing – structure of photosystem I



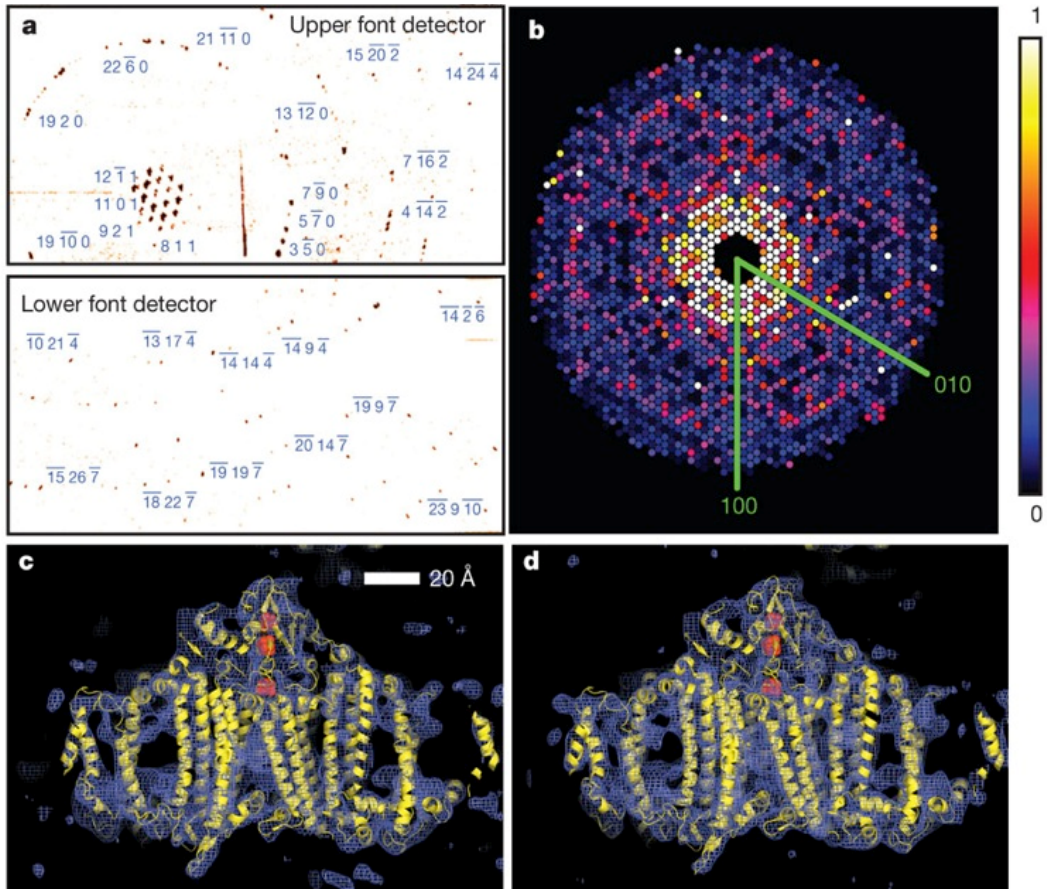


Figure S2 | Rate of hits. Complementary cumulative distribution of the number of single-pulse patterns with N_s or more measured Bragg reflections. From a run of 1.85 million pulses at 70 fs pulse duration, 6% of patterns (112,725) had 10 or more measured peaks, and 13% of those were indexed, for a total efficiency of 0.8%.

Rate of hits is very low since there is no coordination between the crystal stream and the X-ray pulses. The electron density maps at 8.5 Å resolution (left from the 70-fs data; right truncated conventional synchrotron data. Refined model is superposed.

Structure based on single particle diffraction

LETTER *Nature* **470**, 78-81, 2011

doi:10.1038/nature09748

Single mimivirus particles intercepted and imaged with an X-ray laser

M. Marvin Seibert[1]*, Tomas Ekeberg[1]*, Filipe R. N. C. Maia[1]*, Martin Svenda[1], Jakob Andreasson[1], Olof Jönsson[1], Duško Odić[1], Bianca Iwan[1], Andrea Rocker[1], Daniel Westphal[1], Max Hantke[1], Daniel P. DePonte[2], Anton Barty[2], Joachim Schulz[2], Lars Gumprecht[2], Nicola Coppola[2], Andrew Aquila[2], Mengning Liang[2], Thomas A. White[2], Andrew Martin[2], Carl Caleman[1,2], Stephan Stern[2,3], Chantal Abergel[4], Virginie Seltzer[4], Jean-Michel Claverie[4], Christoph Bostedt[5], John D. Bozek[5], Sébastien Boutet[5], A. Alan Miahnahri[5], Marc Messerschmidt[5], Jacek Krzywinski[5], Garth Williams[5], Keith O. Hodgson[6], Michael J. Bogan[6], Christina Y. Hampton[6], Raymond G. Sierra[6], Dmitri Starodub[6], Inger Andersson[7], Saša Bajt[8], Miriam Barthelmess[8], John C. H. Spence[9], Petra Fromme[10], Uwe Weierstall[9], Richard Kirian[9], Mark Hunter[10], R. Bruce Doak[9], Stefano Marchesini[11], Stefan P. Hau-Riege[12], Matthias Frank[12], Robert L. Shoeman[13], Lukas Lomb[13], Sascha W. Epp[14,15], Robert Hartmann[16], Daniel Rolles[13,14], Artem Rudenko[14,15], Carlo Schmidt[14,15], Lutz Foucar[13,14], Nils Kimmel[17,18], Peter Holl[16], Benedikt Rudek[14,15], Benjamin Erk[14,15], André Hömke[14,15], Christian Reich[16], Daniel Pietschner[17,18], Georg Weidenspointner[17,18], Lothar Strüder[14,17,18,19], Günter Hauser[17,18], Hubert Gorke[20], Joachim Ullrich[14,15], Ilme Schlichting[13,14], Sven Herrmann[17,18], Gerhard Schaller[17,18], Florian Schopper[17,18], Heike Soltau[16], Kai-Uwe Kühnel[15], Robert Andritschke[17,18], Claus-Dieter Schröter[15], Faton Krasniqi[13,14], Mario Bott[13], Sebastian Schorb[21], Daniela Rupp[21], Marcus Adolph[21], Tais Gorkhover[21], Helmut Hirsemann[8], Guillaume Potdevin[8], Heinz Graafsma[8], Björn Nilsson[8], Henry N. Chapman[2,3] & Janos Hajdu[1]

X-ray lasers offer new capabilities in understanding the structure of biological systems, complex materials and matter under extreme conditions[1–4]. Very short and extremely bright, coherent X-ray pulses can be used to outrun key damage processes and obtain a single diffraction pattern from a large macromolecule, a virus or a cell before the sample explodes and turns into plasma[1]. The continuous diffraction pattern of non-crystalline objects permits oversampling and direct phase retrieval[2]. Here we show that high-quality diffraction data can be obtained with a single X-ray pulse from a non-crystalline biological sample, a single mimivirus particle, which was injected into the pulsed beam of a hard-X-ray free-electron laser, the Linac Coherent Light Source[5]. Calculations indicate that the energy deposited into the virus by the pulse heated the particle to over 100,000 K after the pulse had left the sample. The reconstructed exit wavefront (image) yielded 32-nm full-period resolution in a single exposure and showed no measurable damage. The reconstruction indicates inhomogeneous arrangement of dense material inside the virion. We expect that significantly higher resolutions will be achieved in such experiments with shorter and brighter photon pulses focused to a smaller area. The resolution in such experiments can be further extended for samples available in multiple identical copies.

Diffraction studies of crystalline samples have led to spectacular breakthroughs in physics, chemistry and biology over the past hundred years. Many important targets are difficult or impossible to crystallize, and this creates systematic blank areas in the structural sciences. X-ray lasers offer the possibility of stepping beyond X-ray crystallography, to extend structural studies to single, non-crystalline particles or molecules[1]. In this Letter, we present results on biological imaging with an X-ray free-electron laser, and bring together all the elements required for structural studies of single, non-crystalline objects.

Mimivirus (*Acanthamoeba polyphaga* mimivirus) is the largest known virus[6]. Its size is comparable to the size of the smallest living cells (in fact, the name mimivirus stands for 'microbe-mimicking virus'). The viral capsid (0.45 μm in diameter) has a pseudo-icosahedral appearance and is covered by an outer layer of dense fibrils[7,8]. The total diameter of the particle, including fibrils, is about 0.75 μm. Mimivirus is too big for a full three-dimensional reconstruction by cryo-electron microscopy[7] and its fibrils prevent crystallization. The genome[9] has 1.2 million base pairs (comparable to a small bacterium) and contains several genes previously thought to be present only in cellular organisms, including components of the protein translation apparatus. Mimivirus can be infected by a smaller virus, named a 'virophage'[10], which seems to be the first example of a virus behaving as a parasite of another virus[8]. Studies of mimivirus are causing a paradigm shift in virology and have led to renewed debates about the origin and the definition of viral and cellular life[11].

Figure 1 shows the experimental arrangement for imaging single virus particles. The sample injector, which uses aerodynamic focusing, was mounted into the CFEL-ASG Multi-Purpose (CAMP) instrument[12] on the Atomic, Molecular and Optical Science (AMO) beamline[13] at the Linac Coherent Light Source[5] (LCLS). We recorded far-field diffraction patterns at a reduced pressure (10^{-6} mbar) to minimize background scattering. Mimivirus was aerosolized from a volatile buffer (250 mM ammonium acetate, pH 7.5) using a gas dynamic nebulizer[14] in a helium atmosphere. The beam of adiabatically cooled virus particles was guided through an aerodynamic lens stack (similar to the one described in ref. 15) and entered the interaction zone with an estimated velocity of 60–100 m s^{-1}. The particles were intercepted randomly by

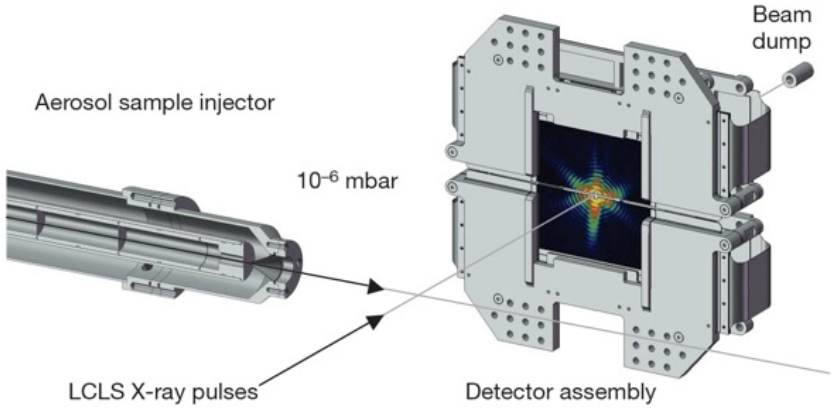
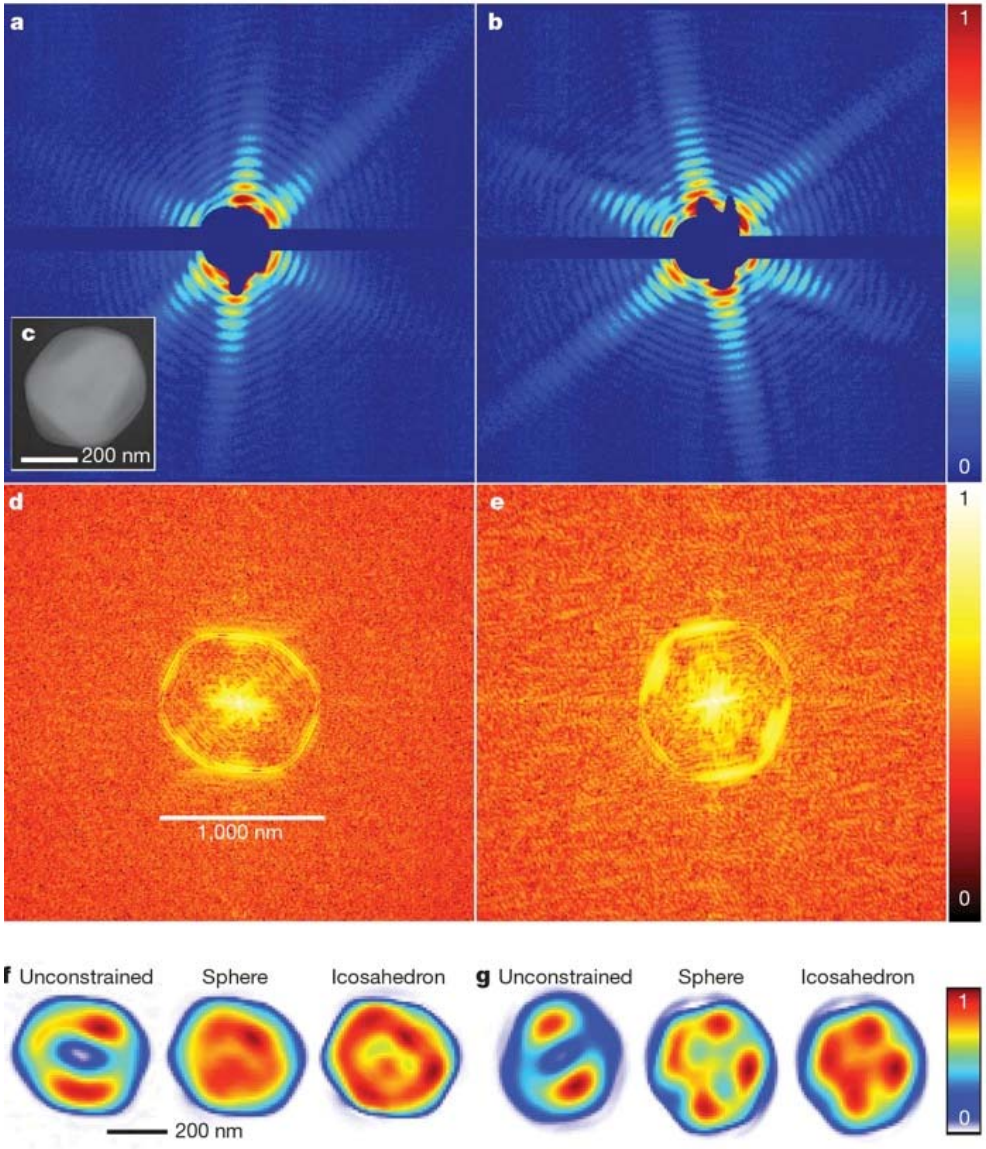


Figure 1 | The experimental arrangement. Mimivirus particles were injected into the pulse train of the LCLS at the AMO experimental station[13] with a sample injector built in Uppsala. The injector was mounted into the CAMP instrument[12]. The aerodynamic lens stack is visible in the centre of the injector body, on the left. Particles leaving the injector enter the vacuum chamber and are intercepted randomly by the LCLS pulses. The far-field diffraction pattern

Mimivirus (*Acanthamoeba polyphaga* mimivirus) is the largest known virus[6]. Its size is comparable to the size of the smallest living cells (in fact, the name mimivirus stands for 'microbe-mimicking virus'). The viral capsid (0.45 μm in diameter) has a pseudo-icosahedral appearance and is covered by an outer layer of dense fibrils[7,8]. The total diameter of the particle, including fibrils, is about 0.75 μm. Mimivirus is too big for a full three-dimensional reconstruction by cryo-electron microscopy[7] and its fibrils prevent crystallization. The genome[9] has 1.2 million base pairs (comparable to a small bacterium) and contains several genes previously thought to be present only in cellular organisms, including components of the protein translation apparatus. Mimivirus can be infected by a smaller virus, named a 'virophage'[10], which seems to be the first example of a virus behaving as a parasite of another virus[8]. Studies of mimivirus are causing a paradigm shift in virology and have led to renewed debates about the origin and the definition of viral and cellular life[11].

Structure based on single particle diffraction

- a,b. Experimental diffraction patterns from individual virus particles in two orientations
- c. Transmission electron micrograph of unstained mimivirus particle
- d,e. Autocorrelation functions for “a” and “b”
- f,g. Reconstructed images after iterative phase retrieval



The first *novel* structure

Science **339**, 227-230, 2013

Natively Inhibited *Trypanosoma brucei* Cathepsin B Structure Determined by Using an X-ray Laser

Lars Redecke,[1,2]† Karol Nass,[3,4]† Daniel P. DePonte,[3] Thomas A. White,[3] Dirk Rehders,[1] Anton Barty,[3] Francesco Stellato,[3] Mengning Liang,[3] Thomas R.M. Barends,[5,6] Sébastien Boutet,[7] Garth J. Williams,[7] Marc Messerschmidt,[7] M. Marvin Seibert,[7] Andrew Aquila,[3] David Arnlund,[8] Sasa Bajt,[9] Torsten Barth,[10] Michael J. Bogan,[11] Carl Caleman,[3] Tzu-Chiao Chao,[12] R. Bruce Doak,[13] Holger Fleckenstein,[3] Matthias Frank,[14] Raimund Fromme,[12] Lorenzo Galli,[3,4] Ingo Grotjohann,[12] Mark S. Hunter,[12,*] Linda C. Johansson,[8] Stephan Kassemeyer,[5,6] Gergely Katona,[8] Richard A. Kirian,[3,13] Rudolf Koopmann,[10] Chris Kupitz,[12] Lukas Lomb,[5,6] Andrew V. Martin,[3] Stefan Mogk,[10] Richard Neutze,[8] Robert L. Shoeman,[5,6] Jan Steinbrener,[5,6] Nicusor Timneanu,[15] Dingjie Wang,[13] Uwe Weierstall,[13] Nadia A. Zatsepin,[13] John C. H. Spence,[13] Petra Fromme,[12] Ilme Schlichting,[5,6] Michael Duszenko,[10] Christian Betzel,[16]‡ Henry N. Chapman[3,4]‡

The *Trypanosoma brucei* cysteine protease cathepsin B (TbCatB), which is involved in host protein degradation, is a promising target to develop new treatments against sleeping sickness, a fatal disease caused by this protozoan parasite. The structure of the mature, active form of TbCatB has so far not provided sufficient information for the design of a safe and specific drug against *T. brucei*. By combining two recent innovations, in vivo crystallization and serial femtosecond crystallography, we obtained the room-temperature 2.1 angstrom resolution structure of the fully glycosylated precursor complex of TbCatB. The structure reveals the mechanism of native TbCatB inhibition and demonstrates that new biomolecular information can be obtained by the "diffraction-before-destruction" approach of x-ray free-electron lasers from hundreds of thousands of individual microcrystals.

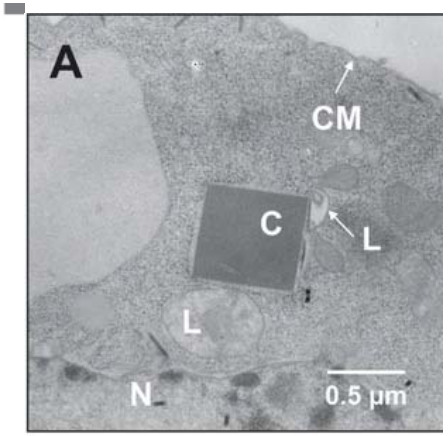
Over 60 million people are affected by human African trypanosomiasis (HAT), also known as sleeping sickness, which causes ~30,000 deaths per year (*1*). The protozoan parasite *Trypanosoma brucei*, transmitted by tsetse flies, infects the blood and the lymphatic system before invading the brain. Severe clinical manifestations occur within weeks or months. Current treatments of HAT rely on antiparasitic drugs developed during the last century, without knowledge of the biochemical pathways. These treatments are limited in their efficacy and safety, and drug resistance is increasing (*2*–*4*). Thus, new compounds that selectively inhibit vital pathways of the parasite without adverse affects to the host are urgently required. A promising strategy is to target lysosomal papainlike cysteine proteases that are involved in host-protein degradation, such as cathepsin B (*5*). The knockdown of this essential enzyme in *T. brucei* resulted in clearance of parasites from the blood of infected mice and cured the infection (*6*), which qualify cathepsin B as a suitable drug target. Cysteine proteases are synthesized as inactive precursors with N-terminal propeptides that act as potent and selective intrinsic inhibitors until the proteases enter the lysosome (*7*), where the propeptide is released and forms the mature active enzyme. Such native propeptide-inhibited structures have been used to develop species-specific protease inhibitors against proteases of other *Trypanosoma* species, e.g., cruzipain of *T. cruzi* (causing human Chagas disease in America) and congopain of *T. congolense* (causing nagana in cattle) (*8*, *9*). This approach could not be explored for *T. brucei* cathepsin B (TbCatB) because of the lack of structural information on the mode of propeptide inhibition and the large extent of structural conservation at the active site between mammalian and trypanosome cathepsin B (*10*–*12*). Previously solved mature *T. brucei* and human CatB structures show differences at the S2 and in part of the S1′ subsite of the substrate-binding cleft (Fig. 1C) and have been suggested as possible targets for the development of species-specific CatB inhibitors (*10*). Together with the natively inhibited human procathepsin B structure (*13*), our work fills the gap to understand the structural basis for species-specific inhibition.

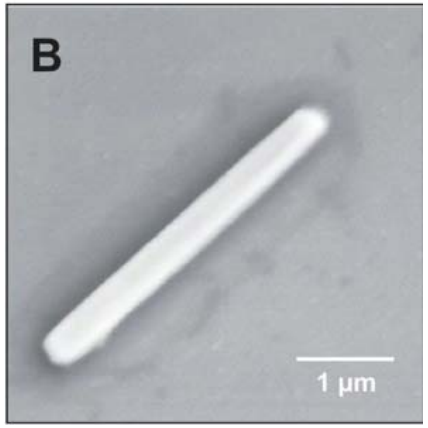
The growth of large well-ordered protein crystals is one of the major bottlenecks in structure determination by x-ray crystallography—with important biological targets, such as integral membrane proteins and posttranslationally modified proteins, proving particularly challenging to crystallize (*14*). Sizable crystals are required to obtain measurable high-resolution diffraction data within an exposure that is limited by the accumulation of radiation damage (*15*). Although microfocus beamlines enable the collection of diffraction data from micron-sized protein crystals (*16*), the tolerable dose limit of less than 30 MGy for cryogenically cooled protein crystals remains, which limits the achievable signal. The tolerable dose for room temperature measurements is about 1 MGy (*15*). We have previously shown that micron-sized crystals of glycosylated TbCatB spontaneously form in insect cells during protein overexpression (*11*). Such crystals are extremely well suited for the new method of serial femtosecond crystallography (SFX) (*17*). X-ray free-electron laser (FEL) pulses of less than 100-fs duration allow the dose to individual crystals to exceed the ~1 MGy limit by over a thousand times because of the "diffraction-before-destruction" principle (*17*, *18*). Diffraction data are recorded for each pulse as crystals are continually replenished by a microcrystal suspension in aqueous buffer flowing across the FEL beam in a vacuum in a fine liquid jet.

The Coherent X-ray Imaging (CXI) beamline (*19*) at the Linac Coherent Light Source (LCLS) enables high-resolution data collection using the SFX approach (*20*). We used this instrument to obtain diffraction data from in vivo grown crystals of TbCatB produced in the baculovirus-infected *Spodoptera frugiperda* (baculovirus-Sf9) insect cell system (*11*) (Fig. 1, A and B). Crystals with average dimensions of about 0.9 by 0.9 by 11 μm^3 (fig. S1) were sent in a 4-μm-diameter column of buffer fluid at room temperature, at a flow rate of 10 μl/minute, by using a liquid microjet (*21*). X-ray pulses from the FEL were focused onto this column to a spot 4 μm in diameter, before the breakup of the jet into drops (fig. S2). Single-pulse diffraction patterns of randomly oriented crystals that, by chance, were present in the interaction region, were recorded at a 120-Hz repetition rate by a Cornell-SLAC pixel array detector (CSPAD) (*19*, *20*) at 9.4-keV photon energy (1.3 Å wavelength). An average pulse energy of 0.6 mJ at the sample (4×10^{11} photons per pulse) with a duration of less than

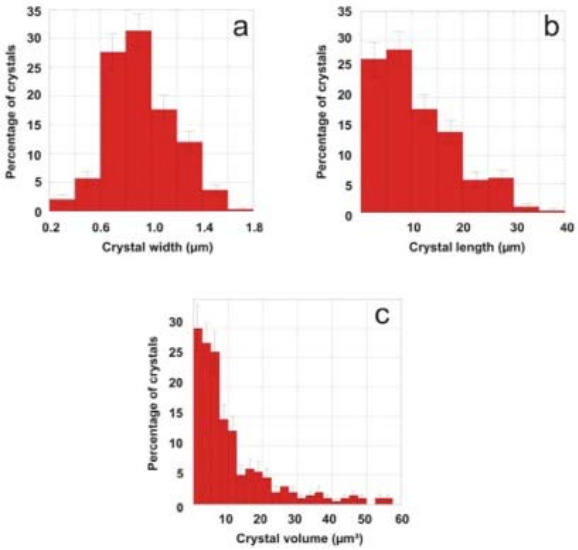
[1]Joint Laboratory for Structural Biology of Infection and Inflammation, Institute of Biochemistry and Molecular Biology, University of Hamburg, and Institute of Biochemistry, University of Lübeck, at Deutsches Elektronen-Synchrotron (DESY), Notkestrasse 85, 22607 Hamburg, Germany. [2]German Centre for Infection Research, University of Lübeck, 23538 Lübeck, Germany. [3]Center for Free-Electron Laser Science (CFEL), DESY, Notkestrasse 85, 22607 Hamburg, Germany. [4]Department of Physics, University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany. [5]Max-Planck-Institut für medizinische Forschung, Jahnstrasse 29, 69120 Heidelberg, Germany. [6]Max Planck Advanced Study Group, Center for Free-Electron Laser Science (CFEL), DESY, Notkestrasse 85, 22607 Hamburg, Germany. [7]Linac Coherent Light Source, Stanford Linear Accelerator Center (SLAC) National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. [8]Department of Chemistry and Molecular Biology, University of Gothenburg, SE-405 30 Gothenburg, Sweden. [9]Photon Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany. [10]Interfaculty Institute of Biochemistry, University of Tübingen, Hoppe-Seyler-Strasse 4, 72076 Tübingen, Germany. [11]Photon Ultrafast Laser Science and Engineering (PULSE) Institute, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. [12]Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287, USA. [13]Department of Physics, Arizona State University, Tempe, AZ 85287, USA. [14]Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA. [15]Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, SE-75124 Uppsala, Sweden. [16]Institute of Biochemistry and Molecular Biology, University of Hamburg, at DESY, Notkestrasse 85, 22607 Hamburg, Germany.

*Present address: Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA.
†These authors contributed equally to this study.
‡To whom correspondence should be addressed. E-mail: henry.chapman@desy.de (H.N.C.) or christian.betzel@uni-hamburg.de (C.B.)





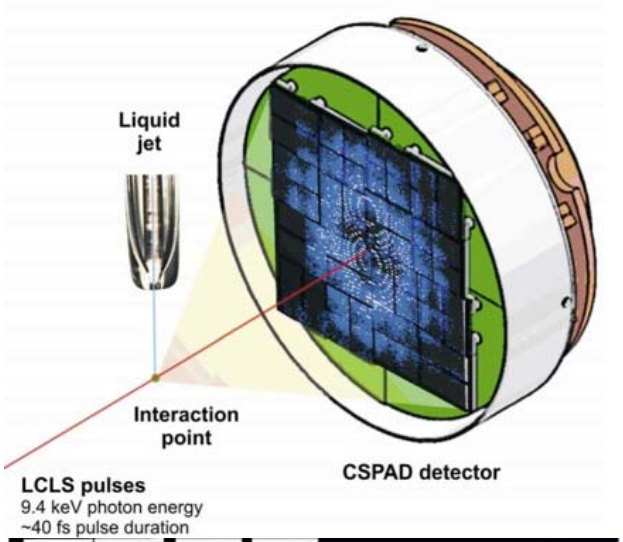
Crystals are grown *in vivo* in insect cells

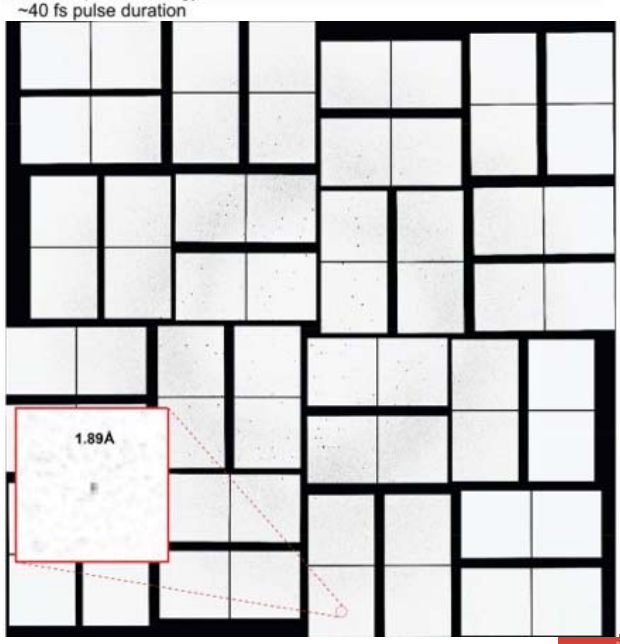


The first *novel* structure

Table S3.
Quality indicators for the individual resolution shells of the dataset

Resolution shell [Å]	Number of unique reflections	Redundancy	Merged I/sigma(I)	R_{split} [%]
20.000 – 4.509	2,793	7,541	32.91	3.0
4.509 – 3.585	2,648	8,094	27.57	3.3
3.585 – 3.134	2,609	8,353	19.12	3.5
3.134 – 2.848	2,588	7,656	12.04	5.8
2.848 – 2.645	2,588	7,656	8.22	11.5
2.645 – 2.489	2,568	7,968	5.93	16.2
2.489 – 2.365	2,552	8,212	4.55	19.6
2.365 – 2.262	2,560	7,899	3.59	24.4
2.262 – 2.175	2,536	7,505	2.93	28.5
2.175 – 2.100	2,540	7,060	2.37	35.3





Note: These are extremely good statistics for compared with traditional synchrotron data. The method results in extremely high values of redundancy.

The first *novel* structure

The quality of the structure is evident in the unbiased electron difference density revealing the unmodelled carbohydrate and bound peptide.

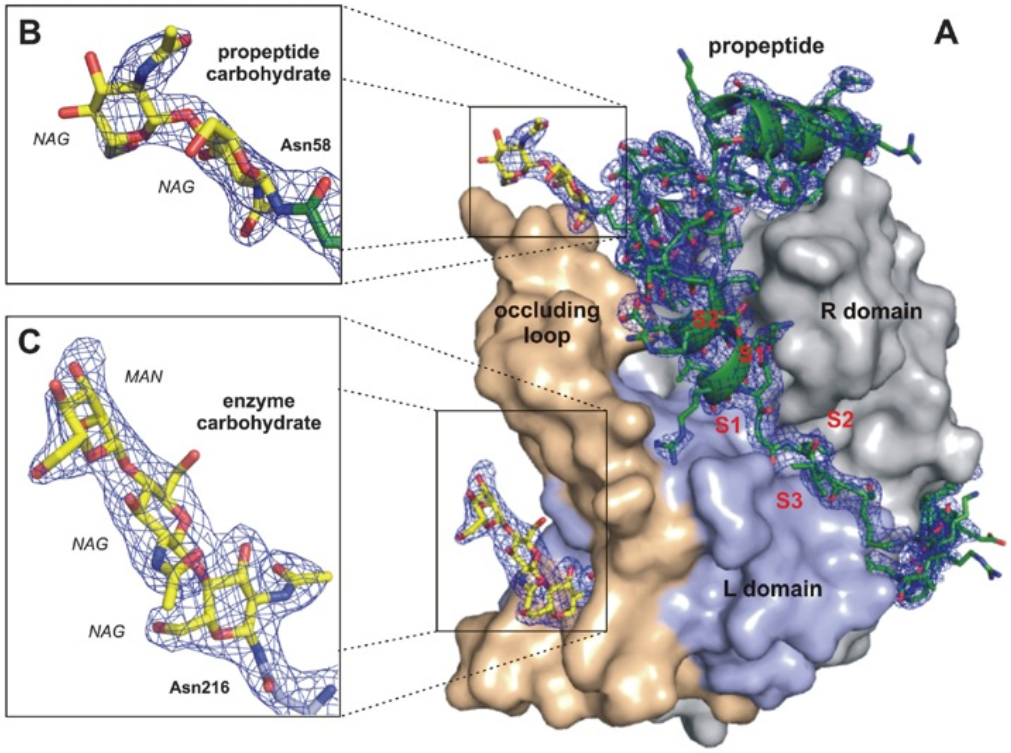
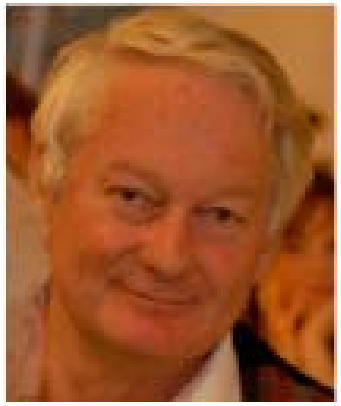


Fig. 2. Quality of the calculated electron density. (**A**) Surface representation of the TbCatB-propeptide complex solved by molecular replacement using the mature TbCatB structure (*11*) as a search model. The solution revealed additional electron density ($2F_{obs} - F_{calc}$, 1σ, blue) of the propeptide (green) that is bound to the V-shaped substrate-binding cleft and of two carbohydrate structures (yellow) N-linked to the propeptide (**B**) and to the mature enzyme (**C**). The propeptide, as well as both carbohydrates, are well-defined within the electron density map (blue), which confirms that the phases are not biased by the search model. Color codes correspond to Fig. 1C.

Names to remember – Nobel prize?

Janos Hajdu
Laboratory of Molecular Biophysics
Uppsala University, Sweden
(Ph.D. Biology, Hungarian Academy of Sciences, 1980)
Will his vision of single molecule diffraction be realised? If so, when?



John H. Spence
Department of Physics
Arizona State University, U.S.A.
(Ph.D. Physics, University of Melbourne, 1973)

Henry Chapman
Centre for Free Electron Science
DESY, Germany
(Ph.D. Physics, University of Melbourne, 1992)