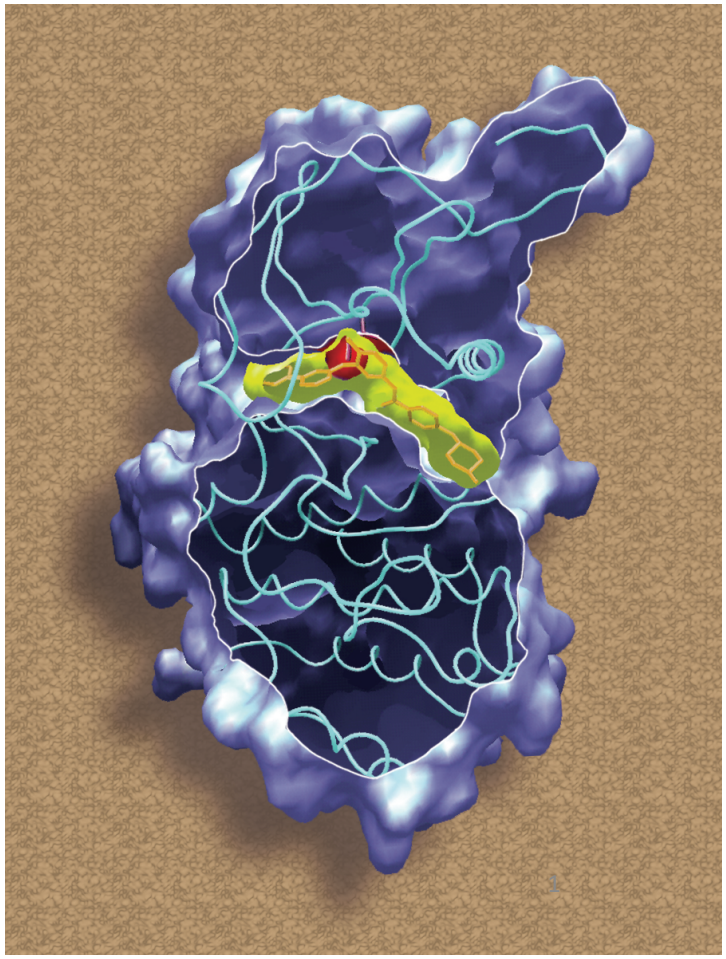
Bi 204 Methods:

# Seeing atomic Structure: Calibrating Molecular Interactions

Bob Stroud
2021

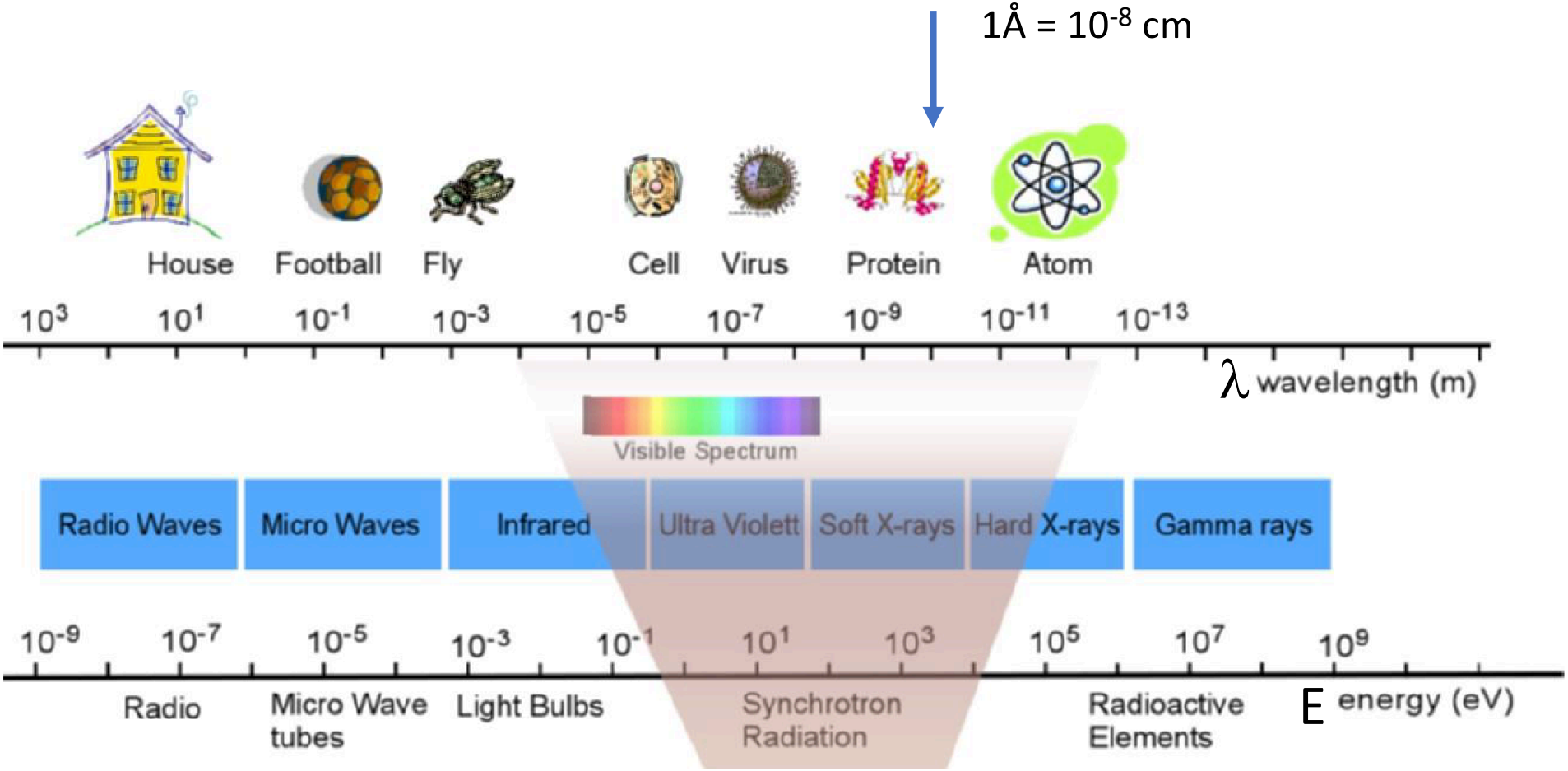
stroud@msg.ucsf.edu



A ‘Ligand’ the cancer drug imatinib (Gleevec) bound to the tyrosine kinase Abl.

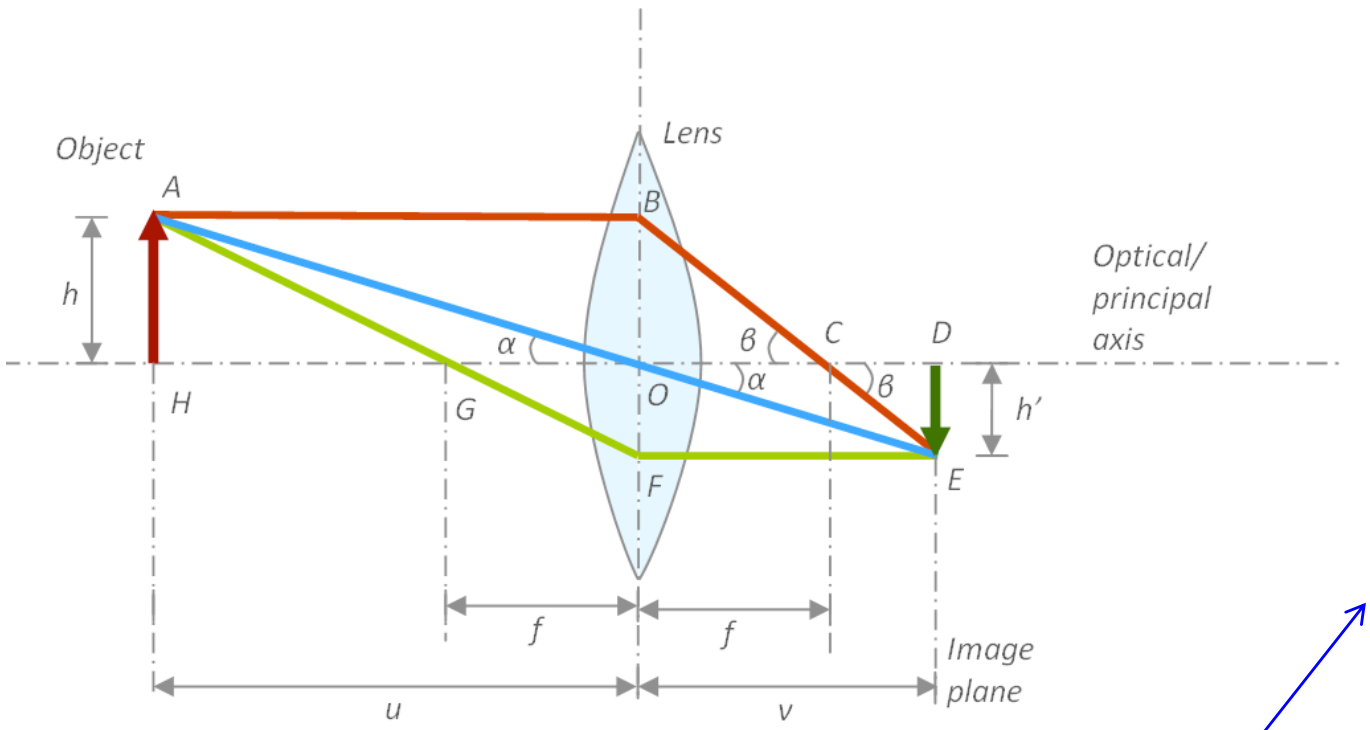
# To ‘see’ anything,  need Wavelength < detail in object

## The Electromagnetic Spectrum



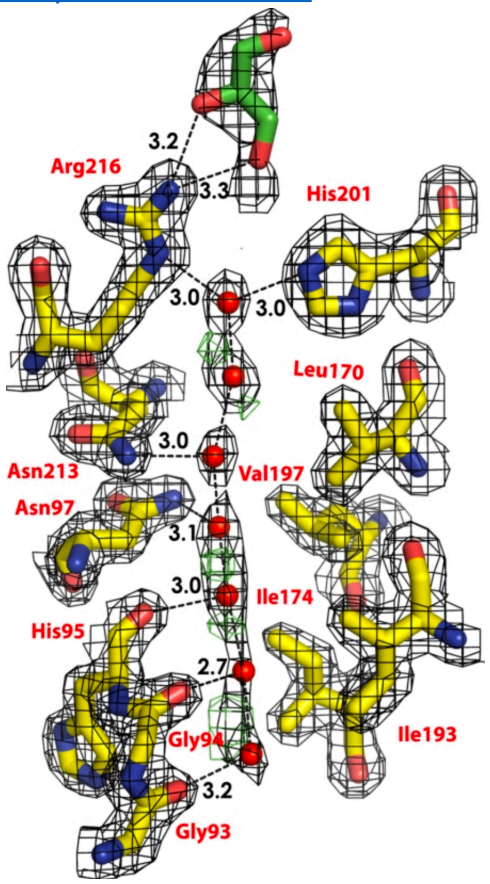

$E = h\nu = hc/\lambda$

# Optical image formation




Human Aquaporin 1.8Å resolution

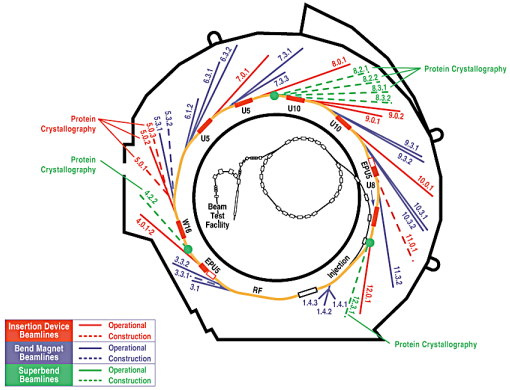
DOI: 10.1073/pnas.0902725106



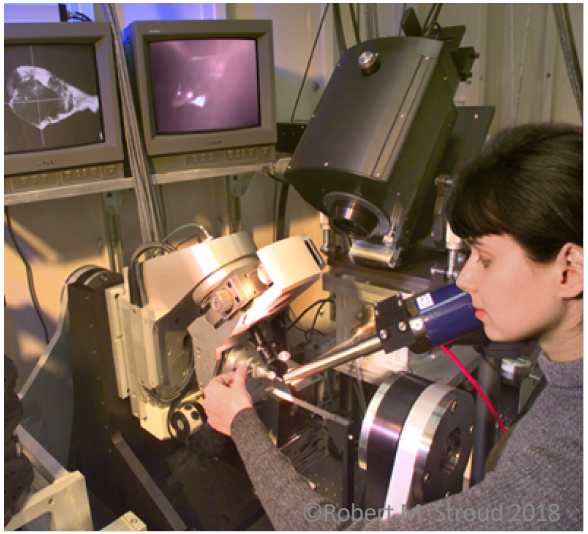

| Type of light | wavelength | what we see? | character | speed |
|---|---|---|---|---|
| Light | 5000Å | dielectric | em. waves-good lenses++ | speed of light |
| X-rays | 1 to 3 Å | electron density; f ~ $n_e$ | em. waves-NO lenses | speed of light |
| neutrons | 1 to 5 Å | nuclei | particles NO lenses | slow speed thermal neutrons |
| electrons | 0.01 - 0.1 Å | electric fields | particles Poor lenses. | eV~0.5mv$^2$. |



Demonstration.


$\lambda \sim 1$ Å



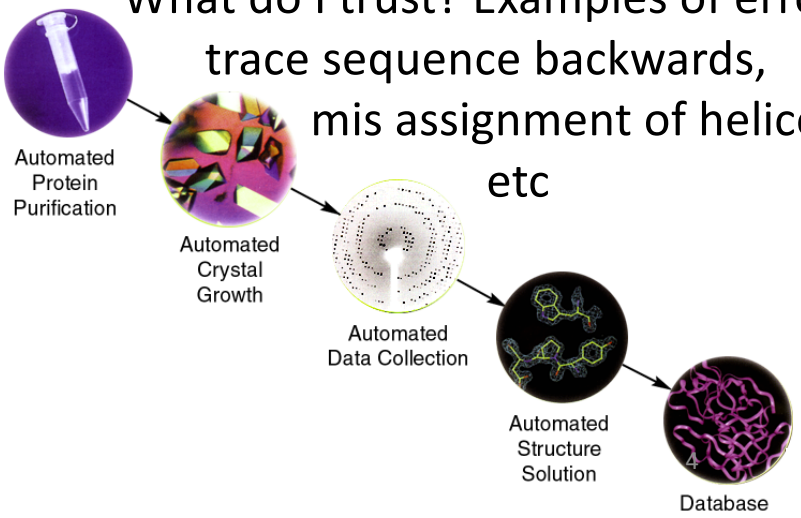

If automated- why are there errors?

What do I trust? Examples of errors

trace sequence backwards,

mis assignment of helices

etc



# Particle wave length. (electrons/neutrons..)

- $\lambda = h/p = h/mv$ Louis de Broglie. (v=velocity)
- $eV = 1/2\ mv^2$ (V= accelerating Voltage)

  $m = 9.11 \times 10^{-31}$ kg (m= electron mass

  $h = 6.6 \times 10^{-34}$ J.s (Planck's const

  $e = 1.6 \times 10^{-19}$ C (charge on electron)
- $\lambda = h/(2meV)^{½} = 12.25/(V)^{½}$ Å

- $\lambda_e$ = 0.04 Å at 100 keV,

  0.027 Å at 200 keV,

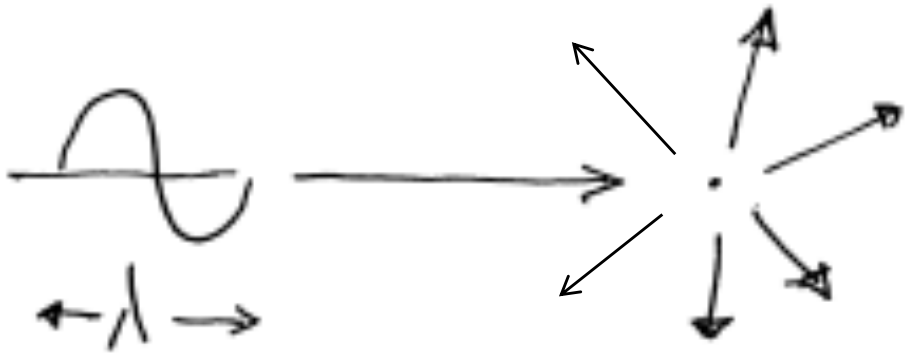
  0.022 Å at 300 keV

Why crystals for X-ray?

Why need to rotate the crystal?

OK its technology—It works-- How do we judge results??

# Elastic Scattering from a point is equal in every direction



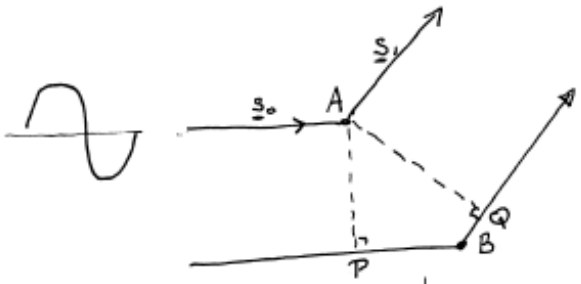

Scattering from a point is equal in all directions.

# Scattering from multiple points? Add wave amplitudes with phase change

Scattering by matter – (interference) of a single wavelength X-ray


Scattering from a point is equal in all directions.

add a second point, scattering in some direction $\underline{s}_1$



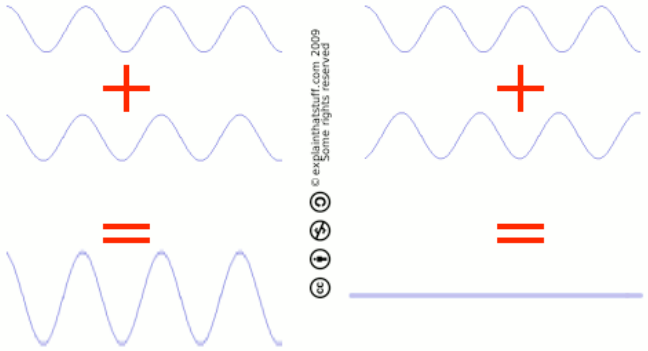

The second wave, scattered by B travels further by the distance PB + BQ. Its scattered wave lags in phase by

$$= \frac{2\pi}{\lambda}\left([P-B] + [B-Q]\right)$$

$$= \frac{2\pi}{\lambda}\left(\Phi\right)$$

where $\Phi$ = path length extra for B versus the reference A.

Adding up the scattering of Atoms: Amplitudes, 'interference' of waves



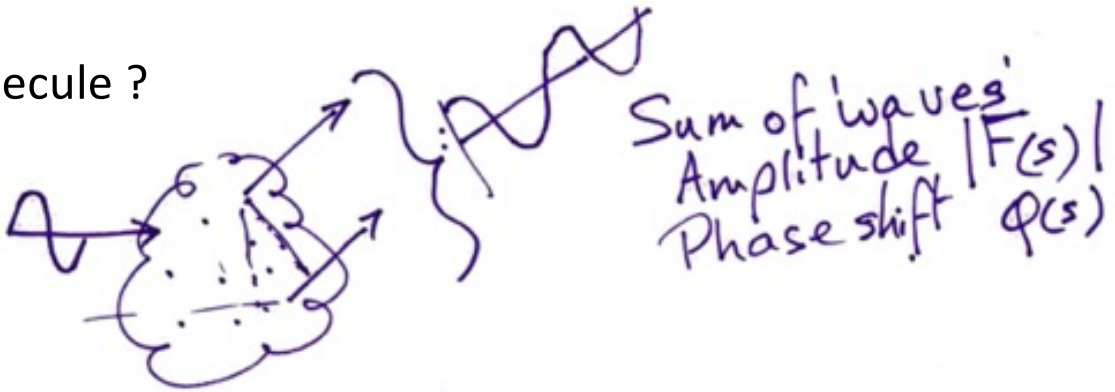

Constructive Interference

Destructive Interference

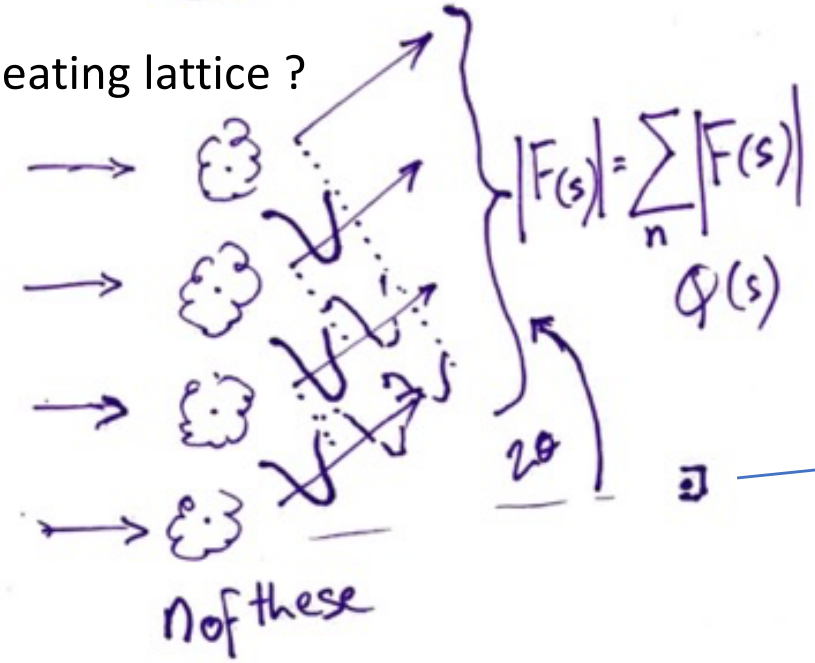
Waves add out of phase by $2\pi$[extra path/$\lambda$]

**Scattering in each direction is different.**

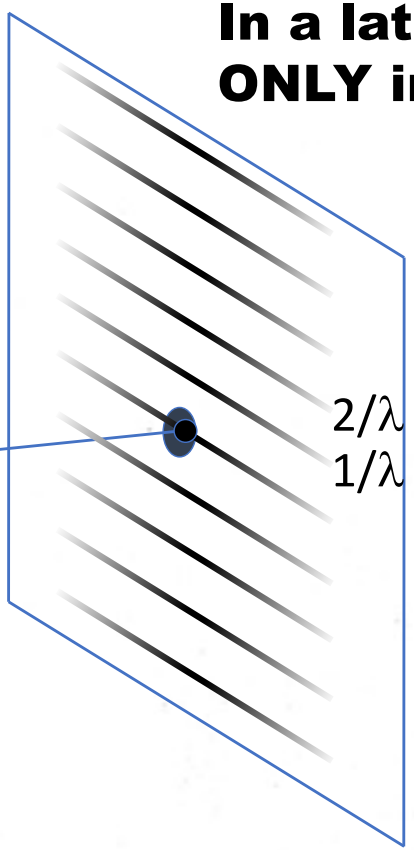
A molecule ?



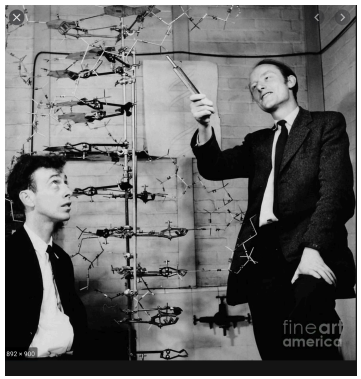

A repeating lattice ?



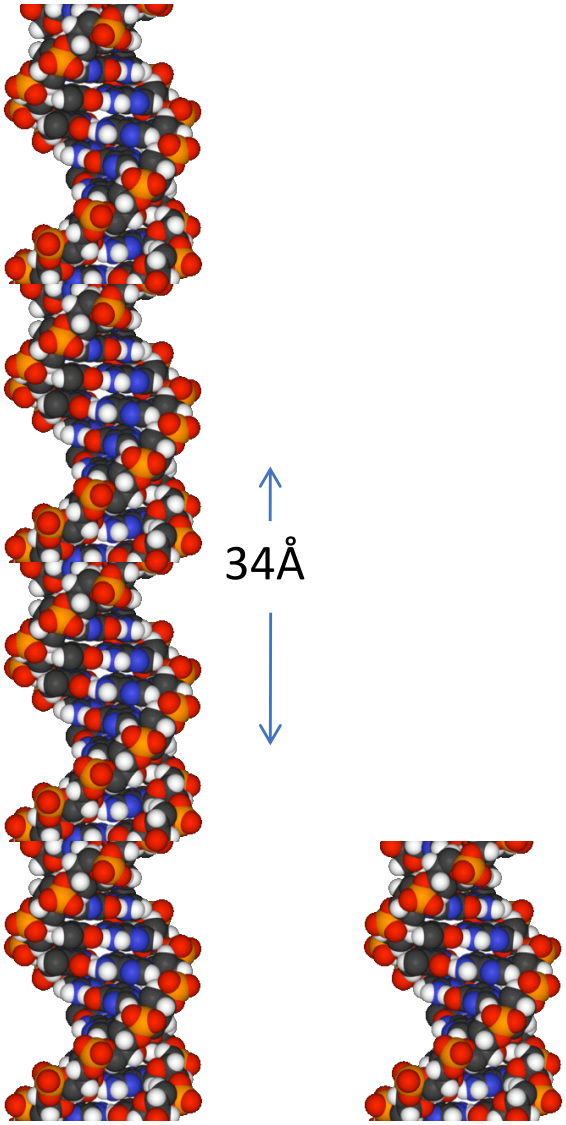

**In a lattice Waves add up ONLY in certain directions**



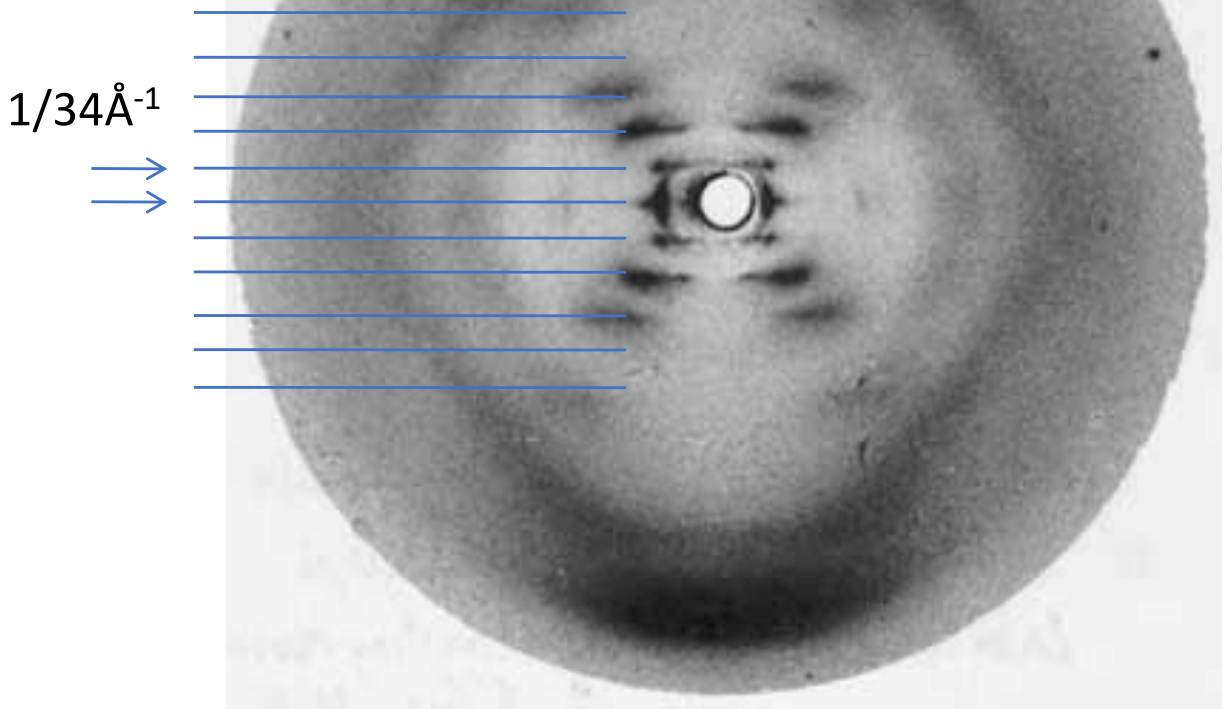
# Consequences of being a 1 D repeat?



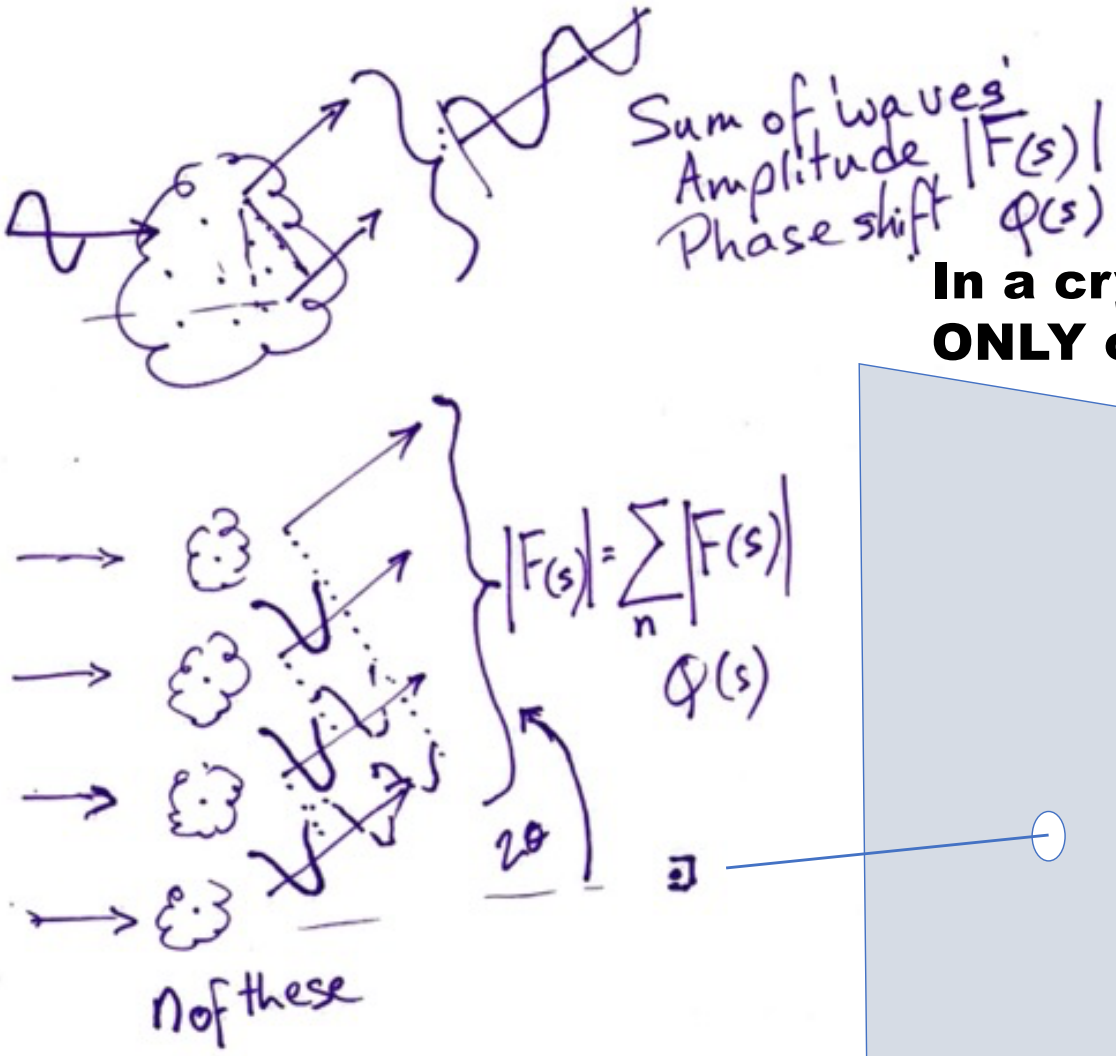
- Repetition = sampling of $F_{(S)}$



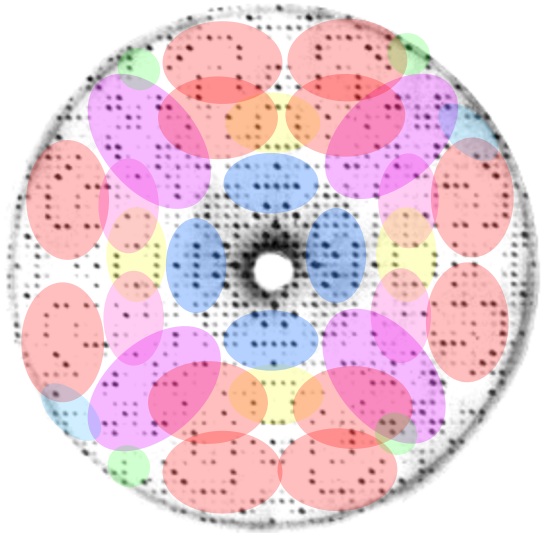

Rosalind Franklin's 'picture 53'

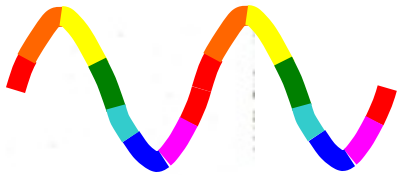


# Crystal lattice; Scattering in each direction is different.



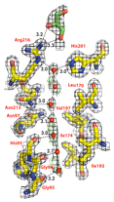

**In a crystal Waves add up ONLY on 'inverse' lattice lines**





Color by phase

# This is all there is? YES!!



FT

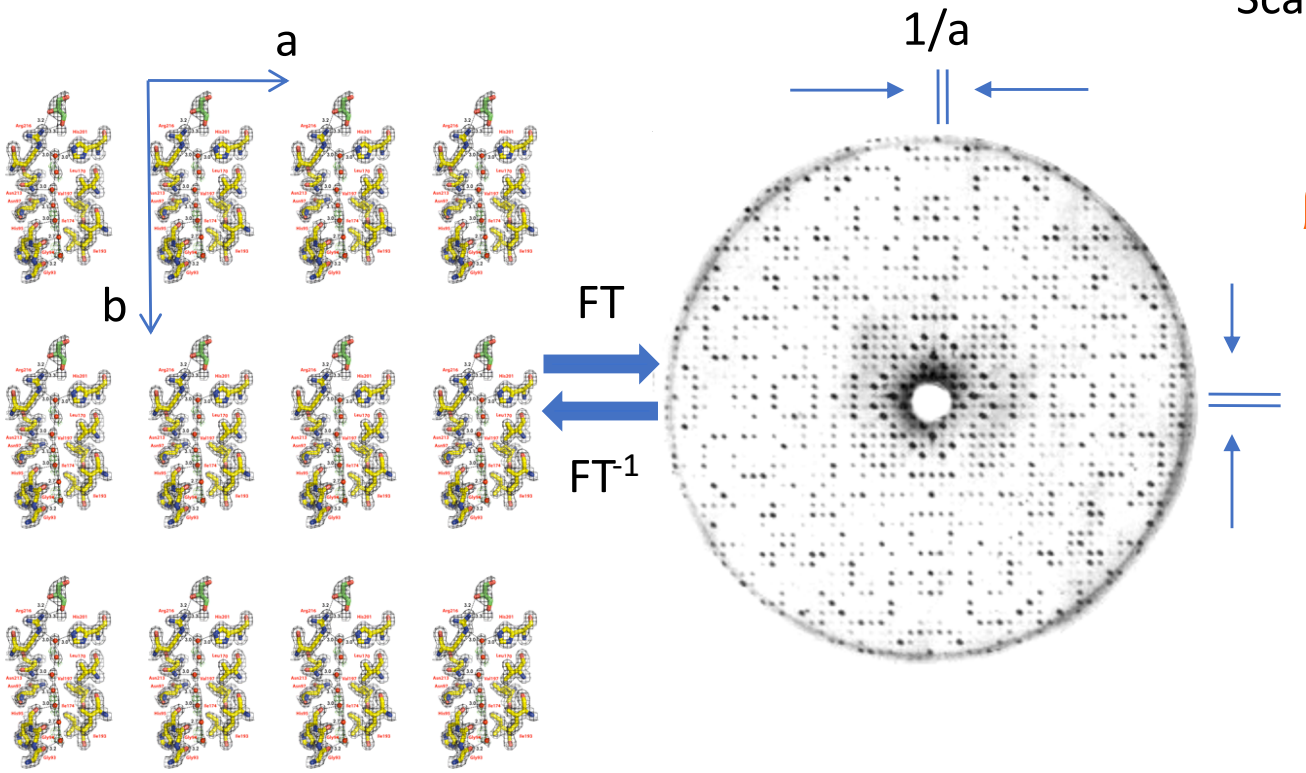
$FT^{-1}$

Scattering pattern is the Fourier transform (FT) of the structure: Amplitude and phase of waves is a sum of waves from each atom

$$\mathbf{F(\underline{S})} = \Sigma_j f_j e^{(2\pi i \mathbf{r_j.S})}$$

**Observe I(S) = F(S).F*(S)**

Structure is the 'inverse' Fourier transform of the Scattering pattern **F(S)**




$$\rho(\underline{r}) = \Sigma\, \mathbf{F(\underline{S})}\, e^{(-2\pi i \mathbf{r.S})}$$

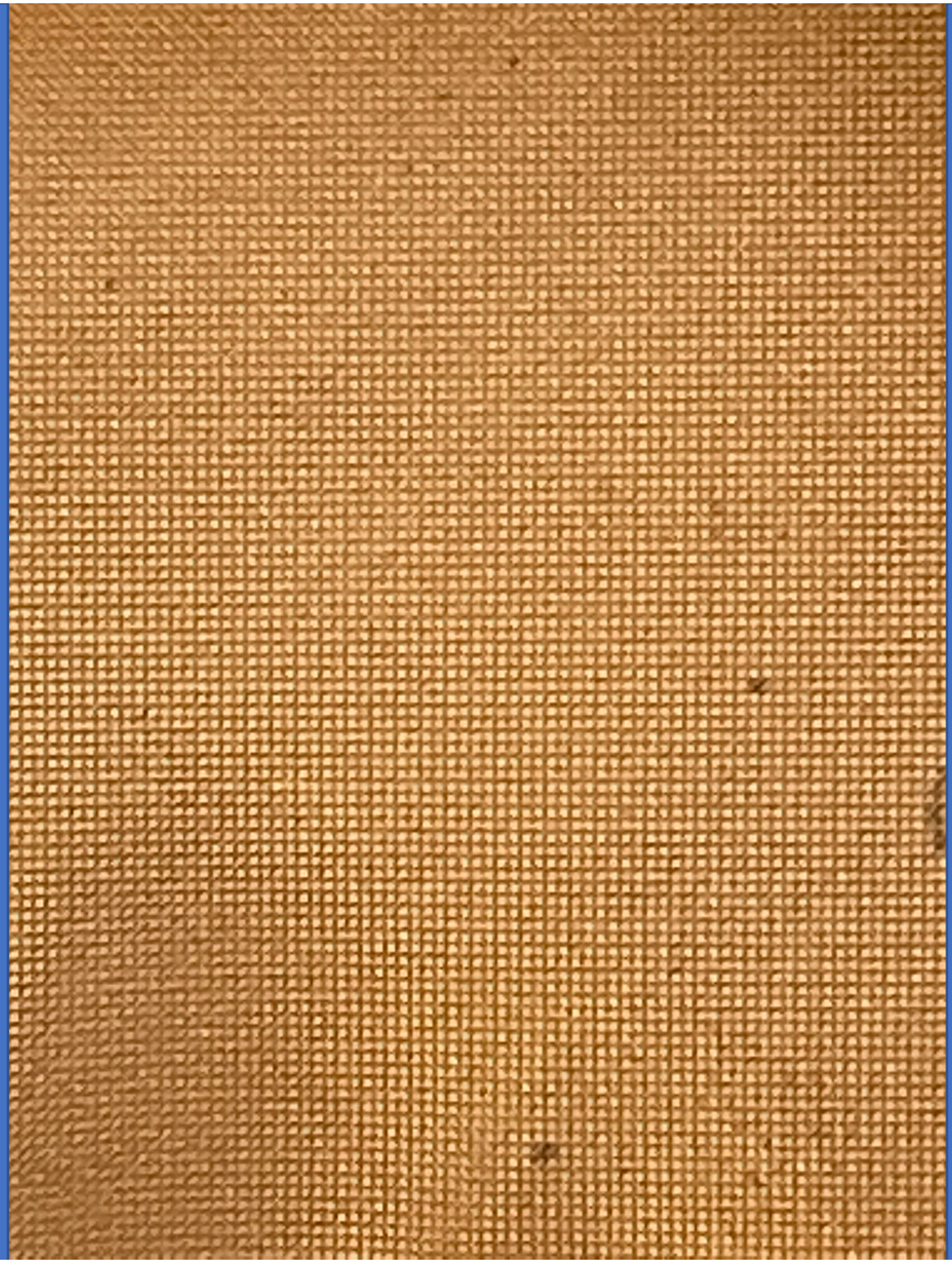
Demonstration.

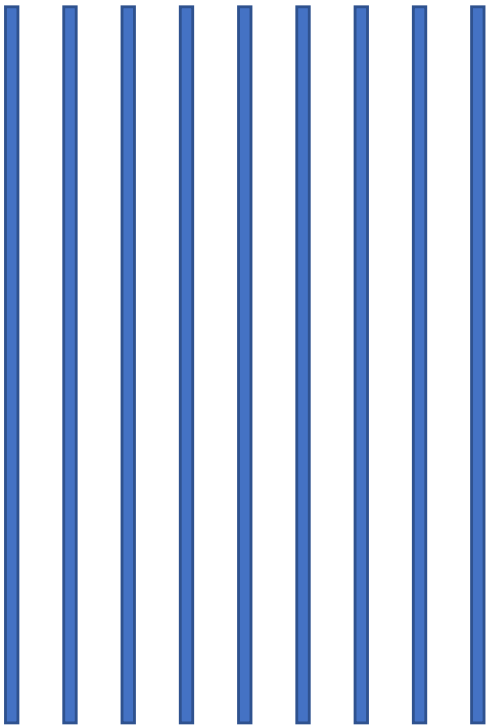
Demonstration of diffraction from an array of ‘little squares’

I’ll use a few sieves of different inter-wire spacings with a weave of wires used for chemicals etc). I show that the diffraction pattern from the repeated ‘little squares’ Is identical to the diffraction pattern from a single square, but only sampled at positions allowed by the repetitions in the ‘lattice’.

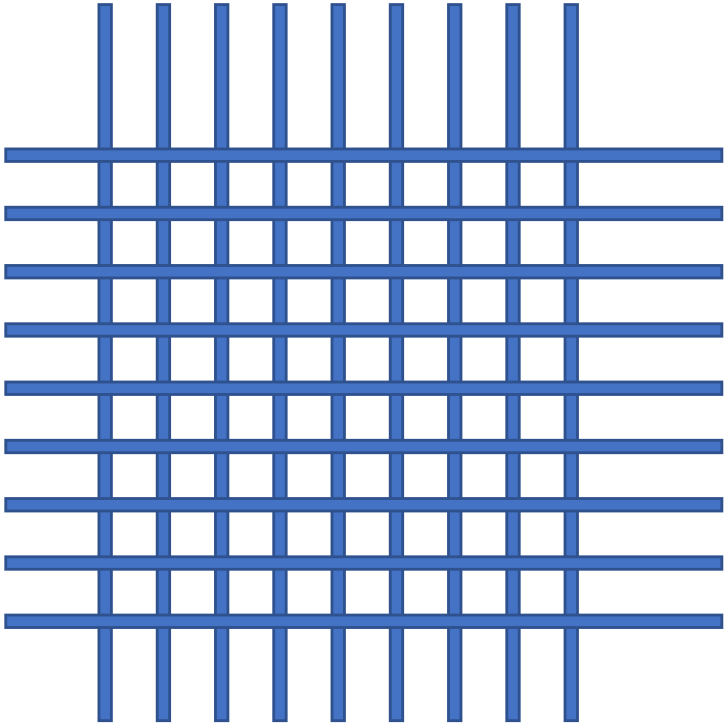
Then we change wavelength of the light, see that the distance between The lattice points in diffraction space indeed change with wavelength Proportional to $2\sin(\Theta) = \lambda/d$. On the screen we see distances between spots that Increase as the spacing between the wires decrease.

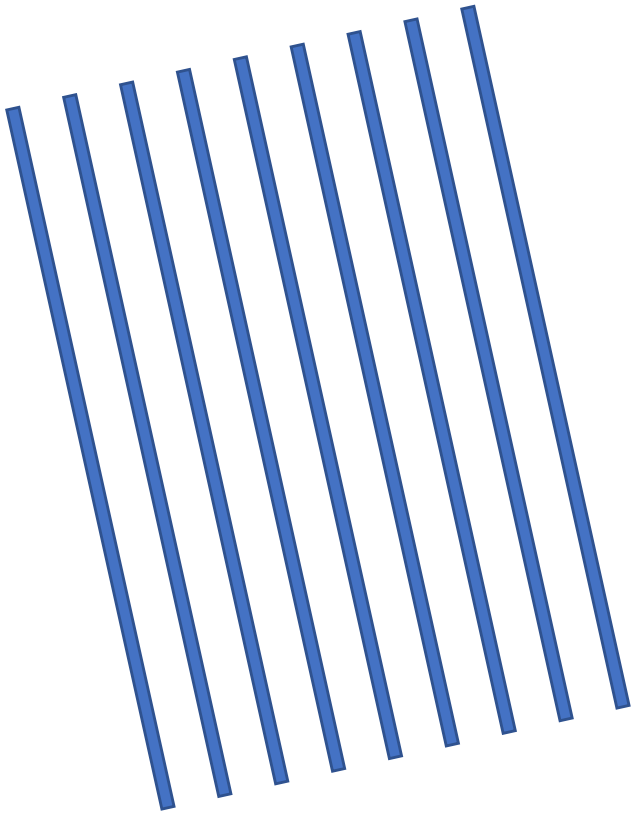
Using rulings, multiplication in ‘real space’ give ‘convolution’ in ‘reciprocal space’.





Multiplication





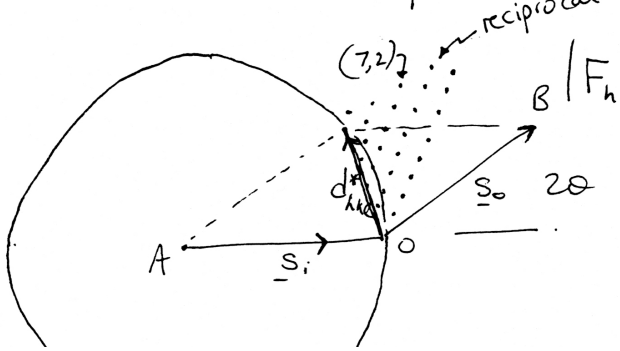

J.D.Bernal



describe any reciprocal lattice point as

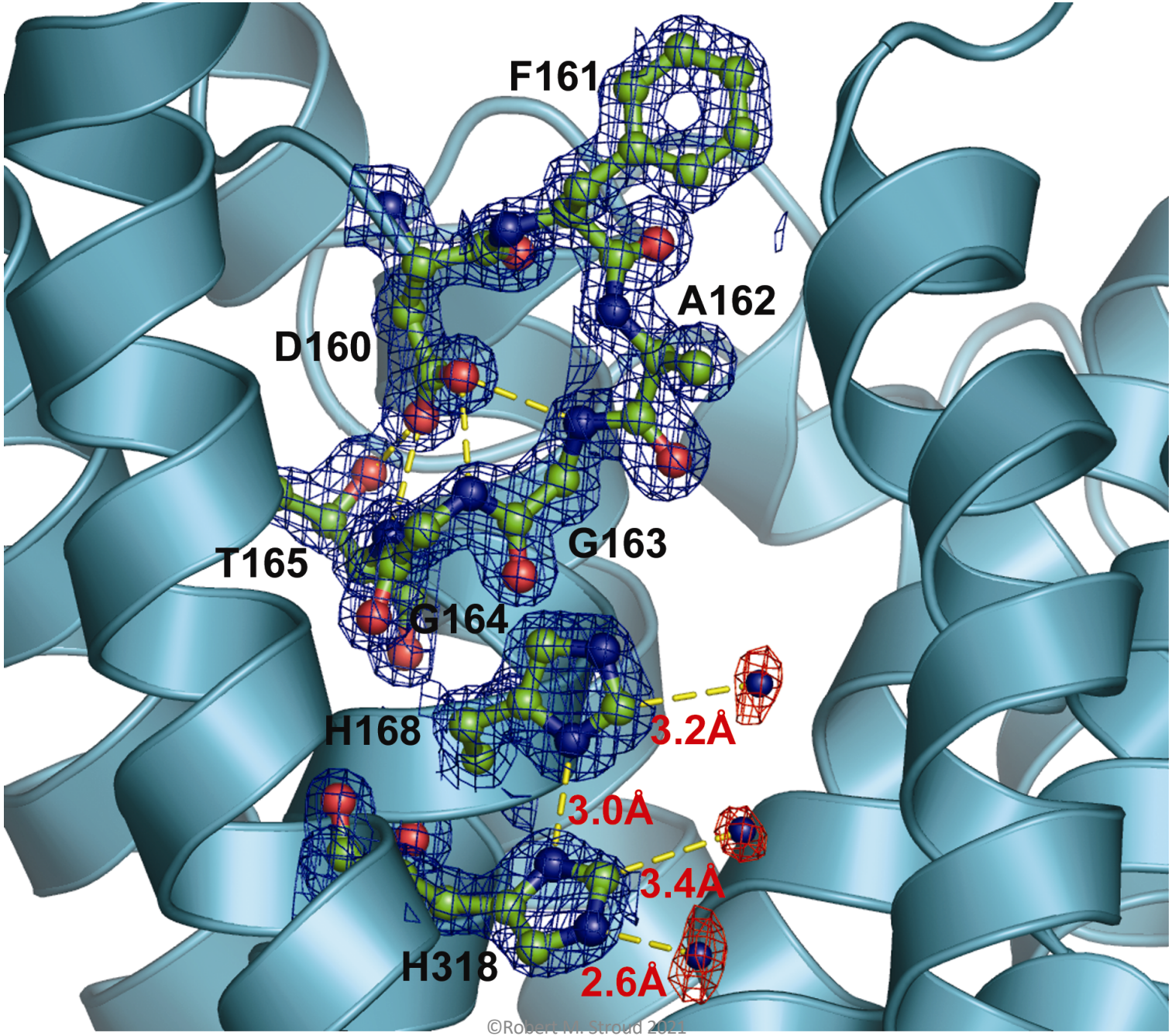
$$\underline{S} = \underline{d}^* = h\underline{a}^* + k\underline{b}^* + l\underline{c}^*$$

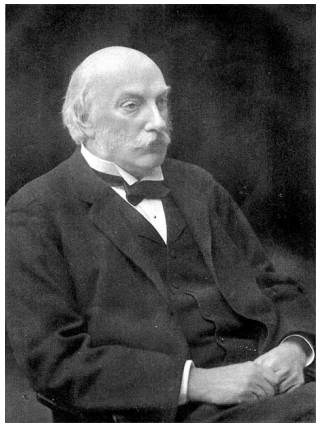
This (hkl) lattice point will only be observed when the crystal is turned so that the (hkl) point $\underline{S} = \underline{d}^*_{hkl}$ ~~lies~~ ~~lands~~ in the Ewald sphere.




$$I_{hkl} = |F_{hkl}|^2 = G(\underline{S})\,G^*(\underline{S})$$

# RESOLUTION ? NH3 sites and role of D160 at 1.35Å Resolution

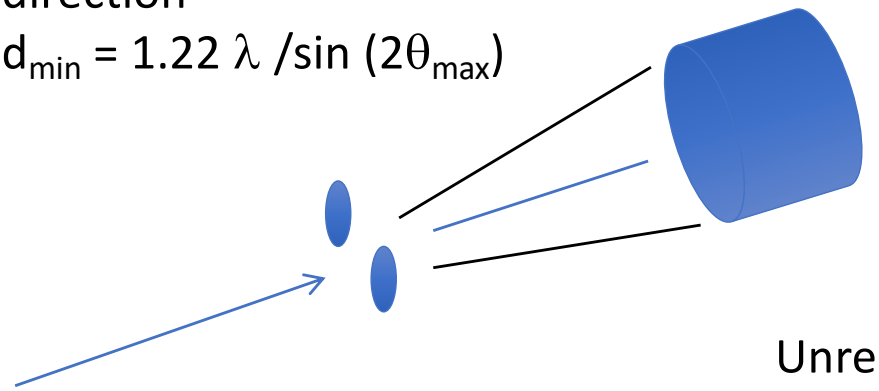




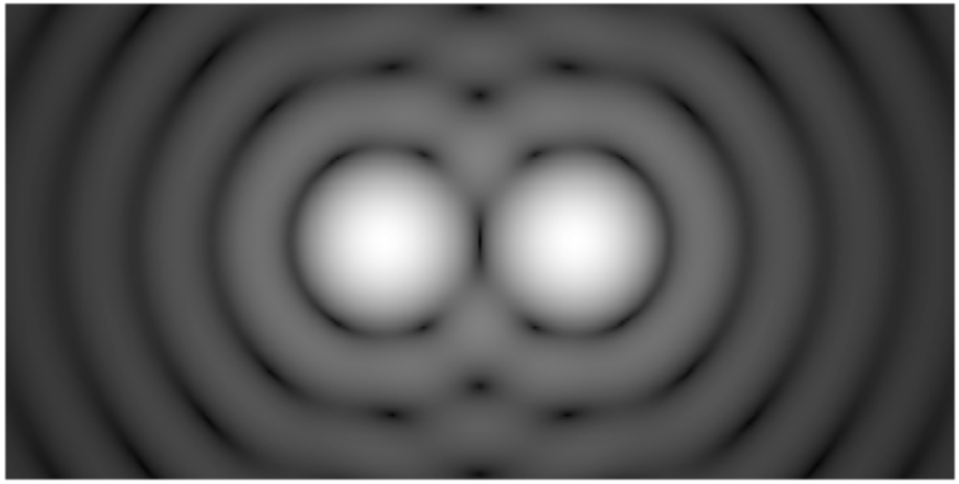
Lord Rayleigh
U.Cambridge
Nobel 1904

Diffraction from Circular hole, forward direction

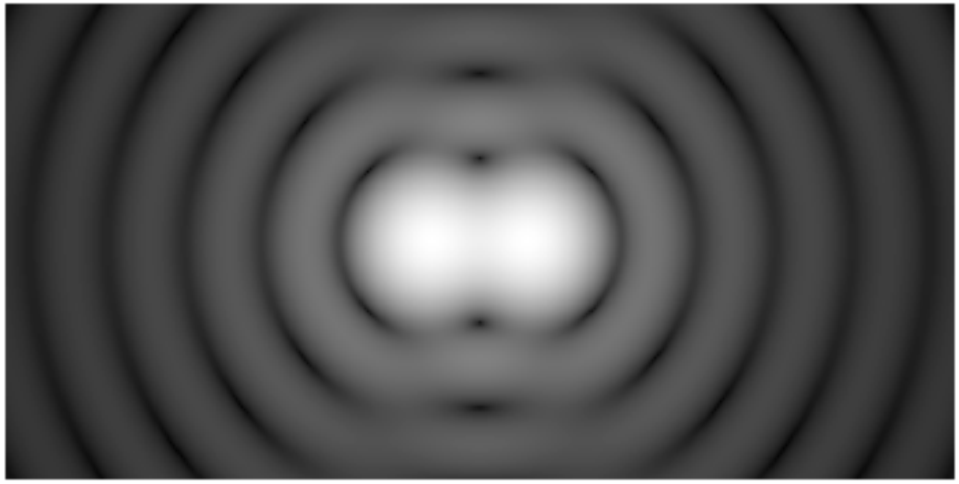
$d_{min} = 1.22\ \lambda / \sin(2\theta_{max})$



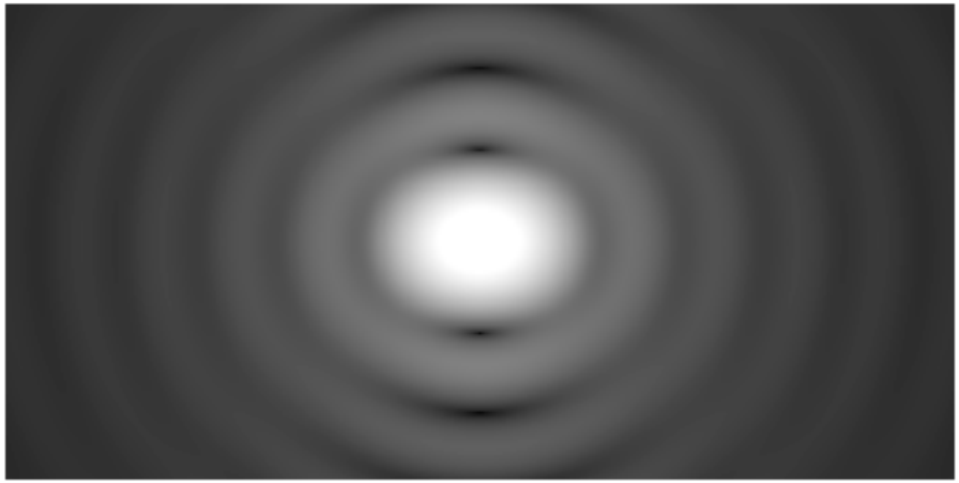
Crystal



Rayleigh

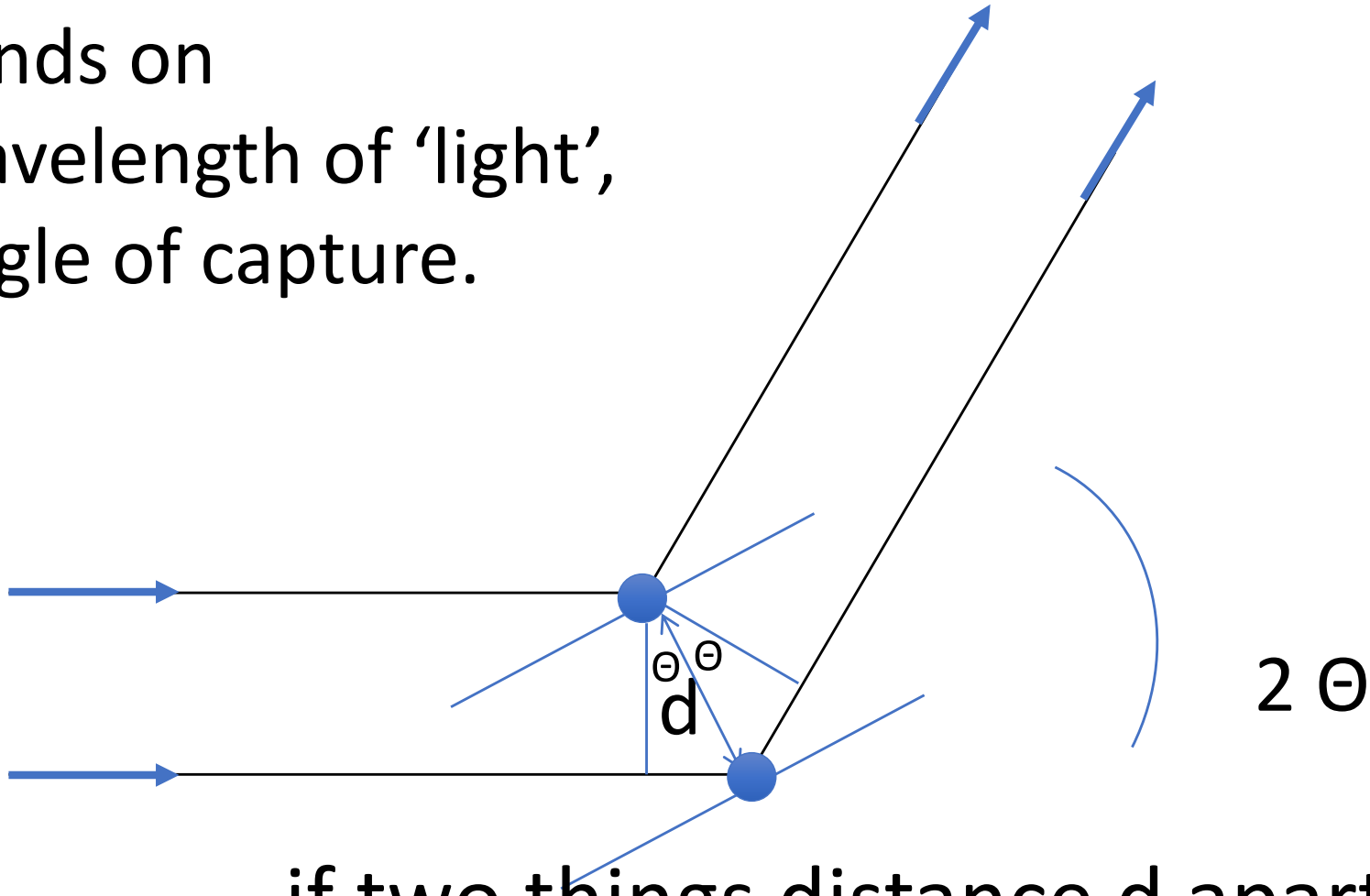


Unresolved



**Resolution** ‘d’ –smallest distance resolved depends on

1. Wavelength of ‘light’,
2. Angle of capture.

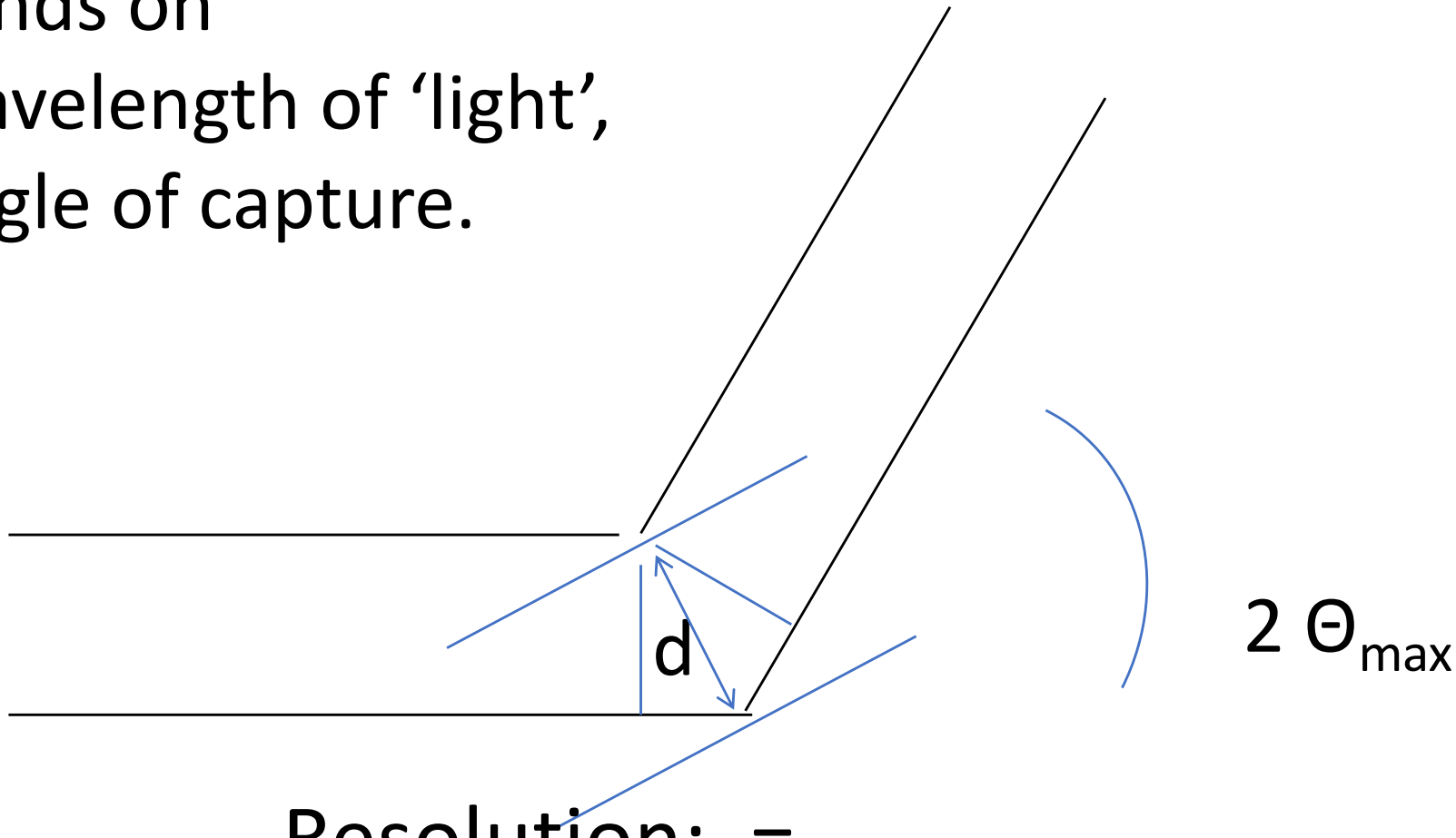



if two things distance d apart. scattered waves reinforce when

$2\ d \sin(\Theta) = \lambda$

**Resolution** ‘d’ –smallest distance resolved depends on

1. Wavelength of ‘light’,
2. Angle of capture.



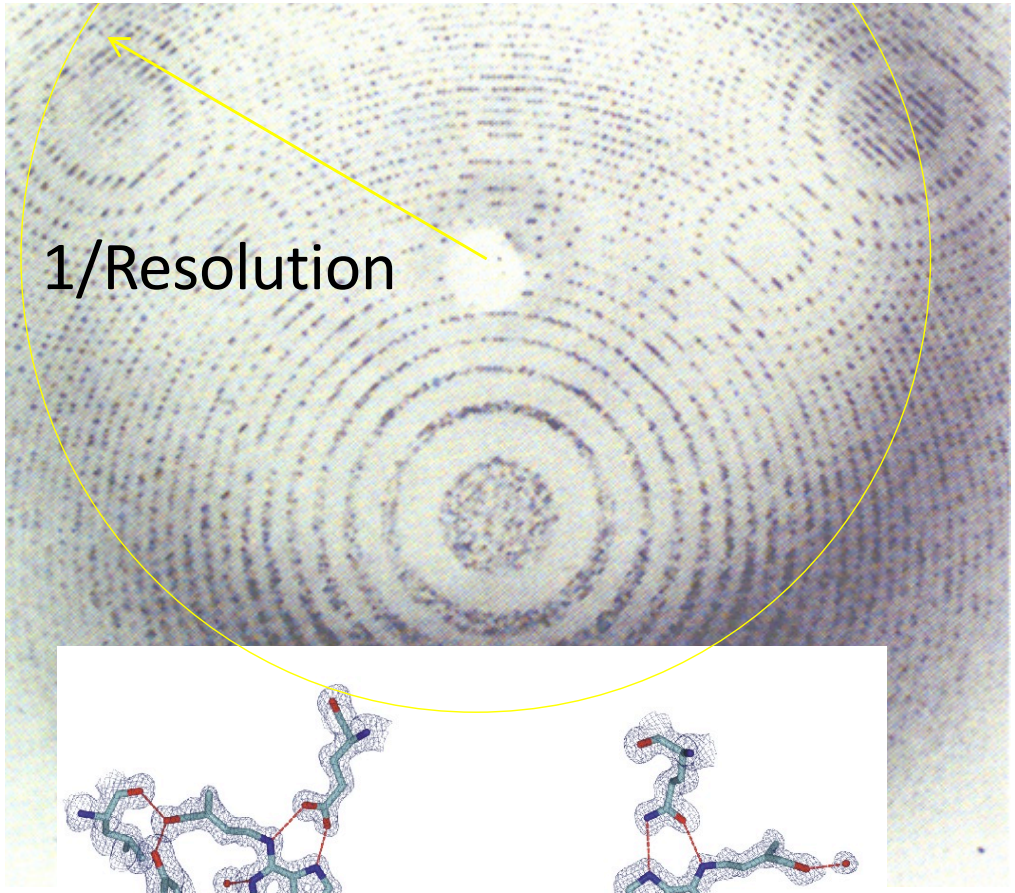

Resolution: =

$d = \lambda/2 \sin(\Theta_{max})$

At best!

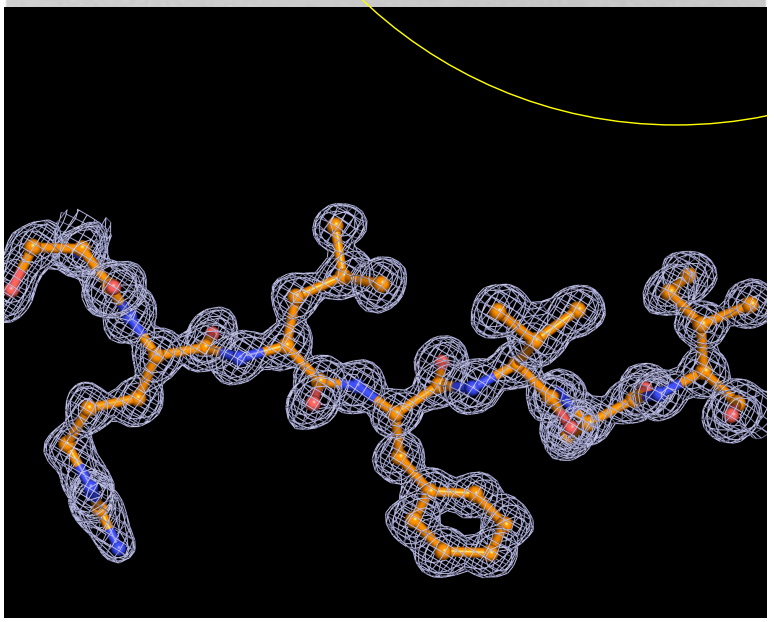
Data/Parameter is the same for all molecular sizes at the same resolution $d_{min}$

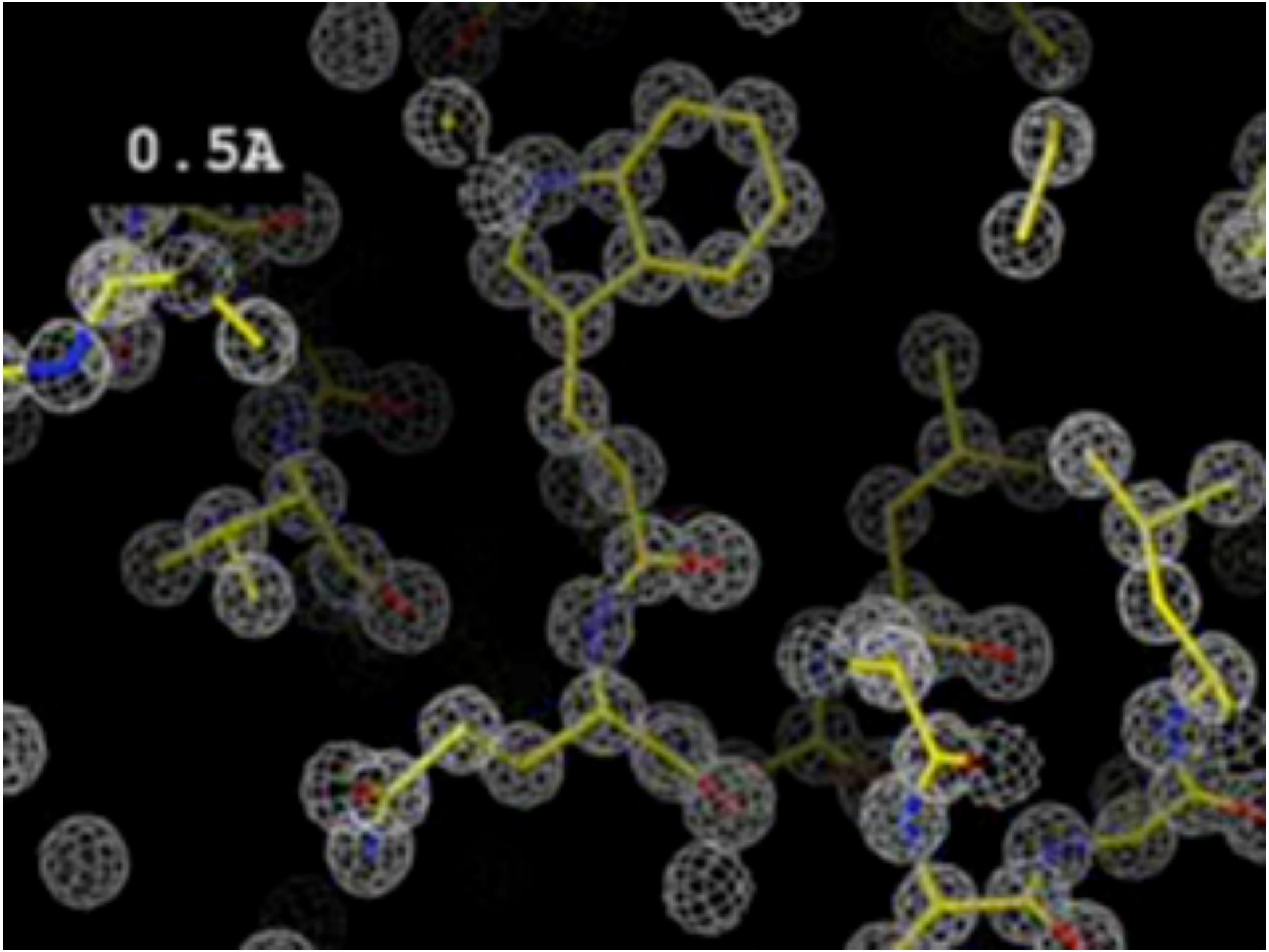
**ie. quality is the same!**





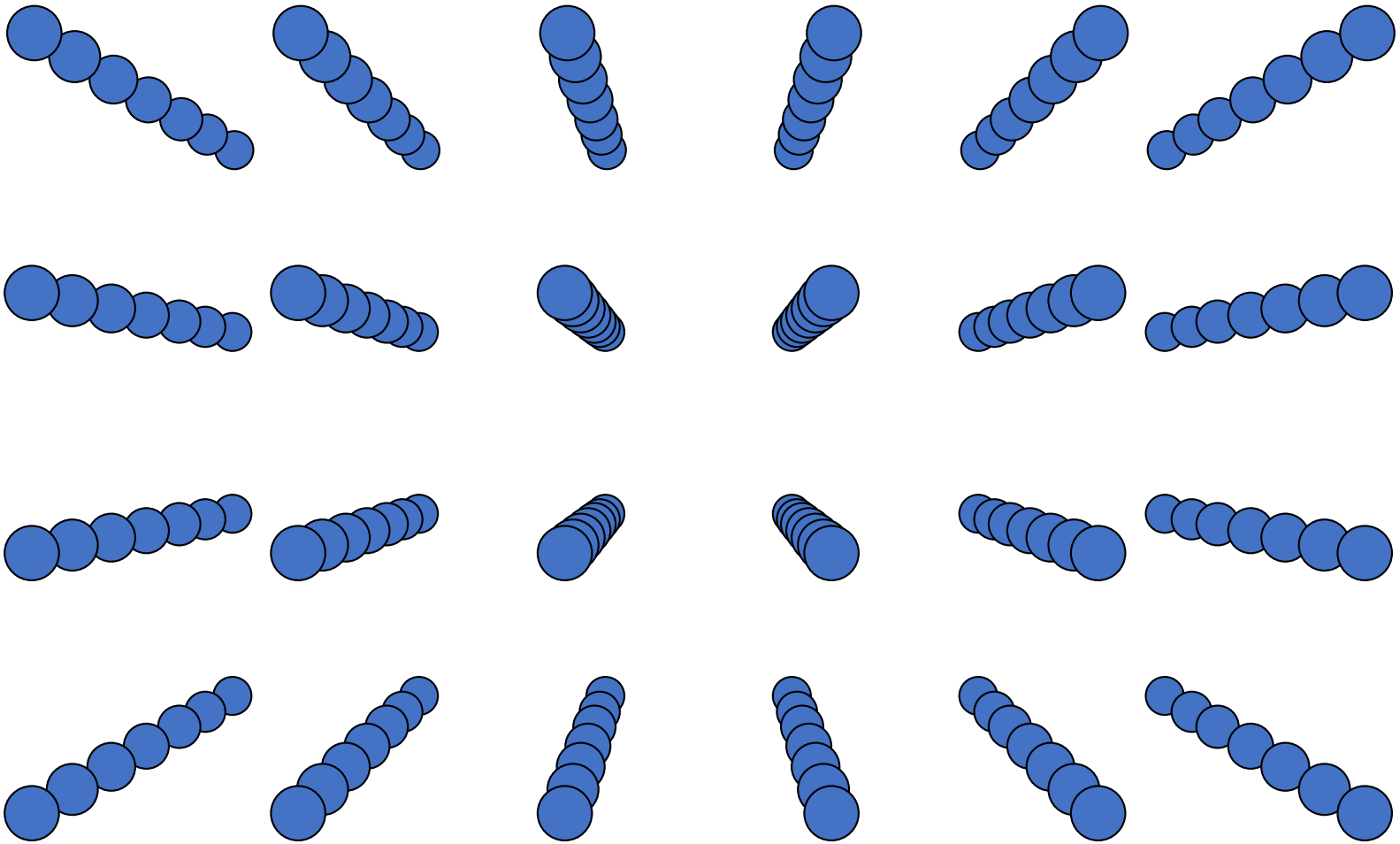


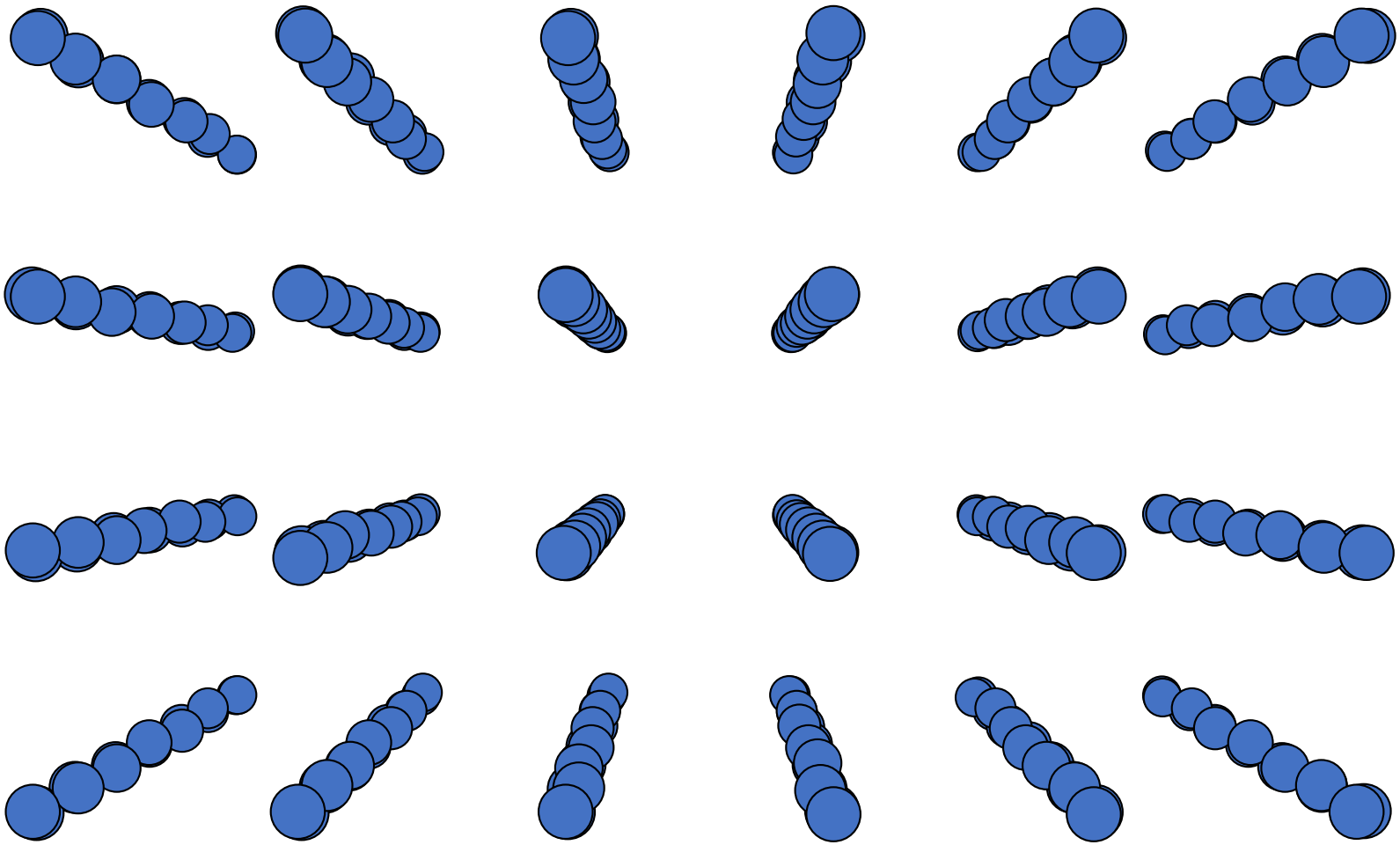

0.5A

# What limits Resolution?

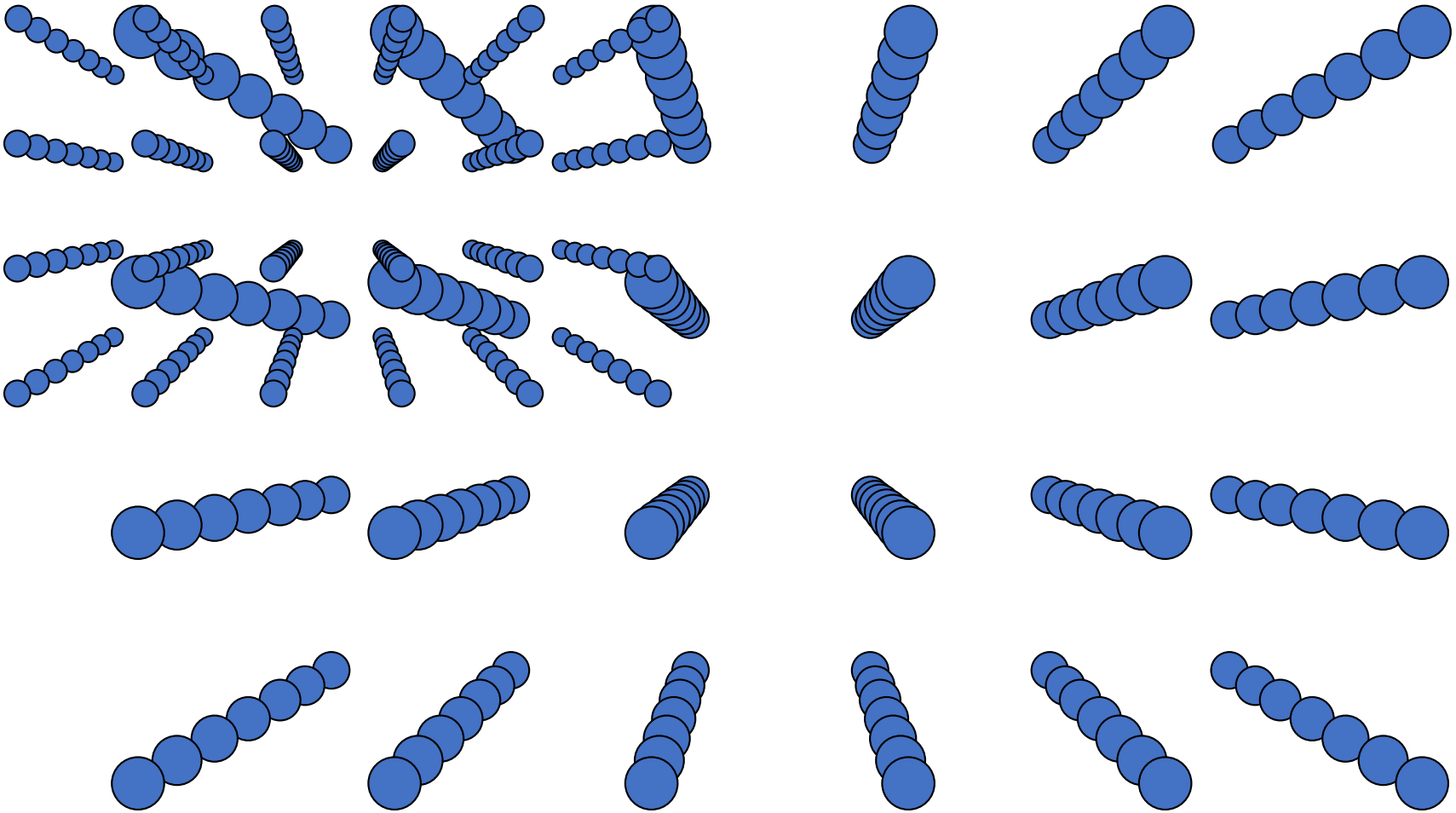
# What is “disorder”?



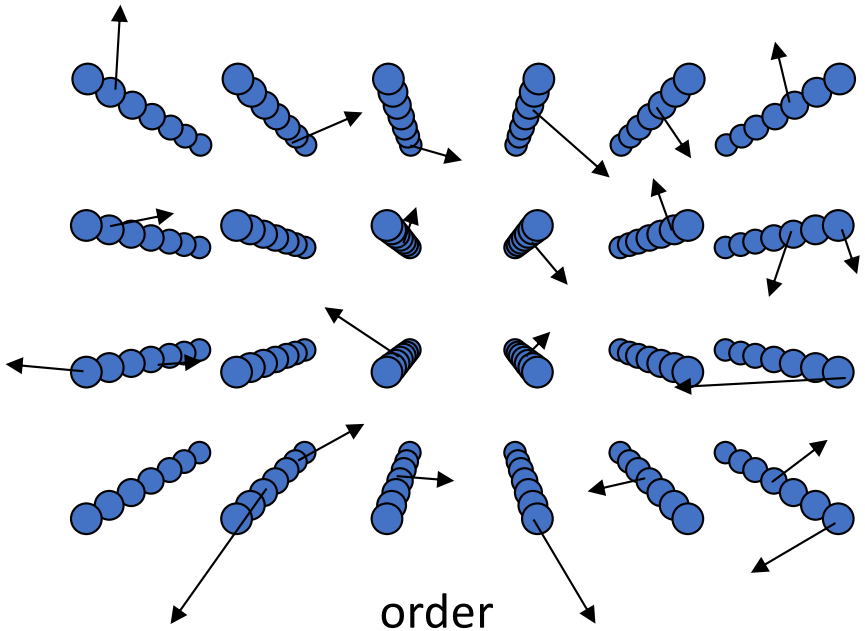
# What is “disorder”?



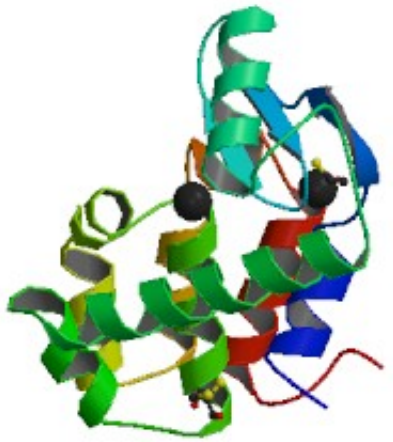
# What is “disorder”?



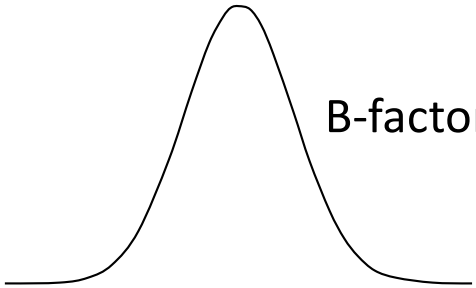
# What is “disorder”?



order



order



B-factor

$B_i = 8\pi^2 U_i^2$

disorder

# “B” factors

$$B_i = 8\pi^2 U_i^2$$

```
ATOM    122  N    LEU A  13      -3.244  25.808  19.998  1.00 16.96
ATOM    123  CA   LEU A  13      -2.877  25.448  21.355  1.00 15.29
ATOM    124  C    LEU A  13      -2.792  23.966  21.561  1.00 17.54
ATOM    125  O    LEU A  13      -1.814  23.493  22.143  1.00 16.35
ATOM    126  CB   LEU A  13      -3.907  26.164  22.268  1.00 18.72
ATOM    127  CG   LEU A  13      -3.577  25.982  23.738  1.00 21.19
ATOM    128  CD1  LEU A  13      -2.283  26.820  24.019  1.00 19.43
ATOM    129  CD2  LEU A  13      -4.702  26.474  24.639  1.00 24.65
ATOM    130  N    SER A  14      -3.677  23.149  20.979  1.00 15.96
ATOM    131  CA   SER A  14      -3.646  21.711  21.061  1.00 18.26
ATOM    132  C    SER A  14      -2.373  21.203  20.360  1.00 18.71
ATOM    133  O    SER A  14      -1.747  20.315  20.930  1.00 17.47
ATOM    134  CB   SER A  14      -4.875  21.077  20.419  1.00 17.62
ATOM    135  OG ASER A  14      -4.825  19.665  20.388  0.50 20.89
ATOM    136  OG BSER A  14      -6.027  21.408  21.164  0.50 18.67
ATOM    137  N    LYS A  15      -2.045  21.772  19.215  1.00 18.03
ATOM    138  CA   LYS A  15      -0.799  21.361  18.555  1.00 18.12
ATOM    139  C    LYS A  15       0.446  21.707  19.351  1.00 18.81
ATOM    140  O    LYS A  15       1.400  20.948  19.411  1.00 17.77
ATOM    141  CB   LYS A  15      -0.700  22.034  17.177  1.00 14.49
ATOM    142  CG   LYS A  15      -1.727  21.368  16.256  1.00 16.12
ATOM    143  CD   LYS A  15      -1.663  22.147  14.936  1.00 19.40
ATOM    144  CE ALYS A  15      -2.725  21.614  13.986  0.50 17.42
```

# 2. Phase Determination.

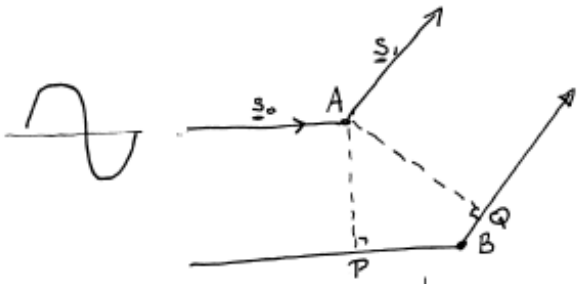
Adding waves together.

# Scattering from multiple points? Add wave amplitudes with phase change

Scattering by matter – (interference) of a single wavelength X-ray


Scattering from a point is equal in all directions.

add a second point, scattering in some direction $\underline{s}_1$



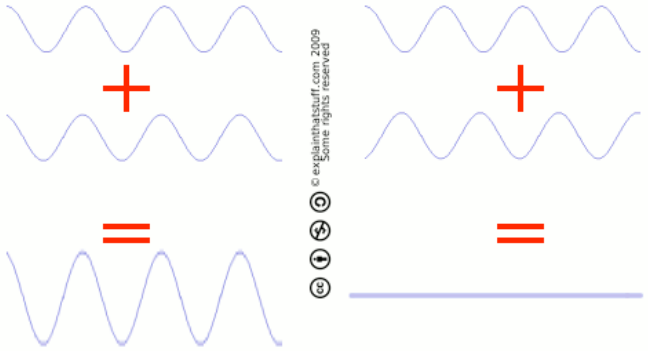

The second wave, scattered by B travels further by the distance PB + BQ. Its scattered wave lags in phase by

$$= \frac{2\pi}{\lambda}\left([P-B] + [B-Q]\right)$$

$$= \frac{2\pi}{\lambda}\left(\Phi\right)$$

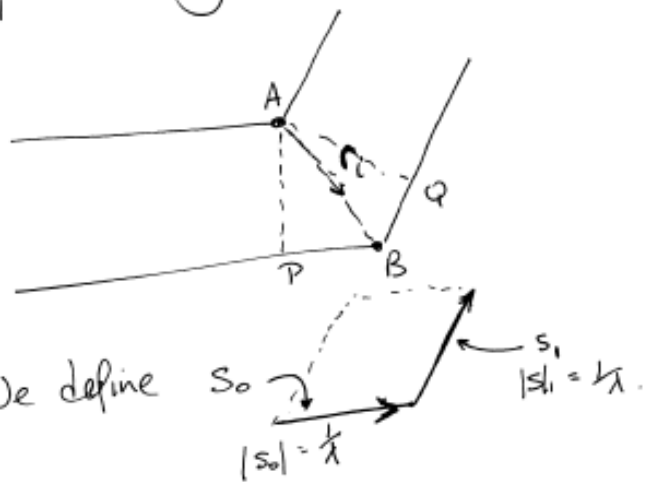
where $\Phi$ = path length extra for B versus the reference A.

Adding up the scattering of Atoms: Amplitudes, 'interference' of waves




Waves add out of phase by $2\pi$[extra path/$\lambda$]

The phase lag can also be simplified

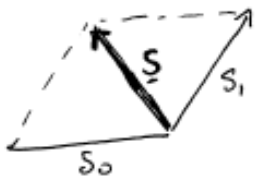



the PB = projection of $r_i$ on $S_0$
BQ = projection of $r_i$ on $S_1$

$$\text{So } \frac{\Phi}{\lambda} = \frac{PB + BQ}{\lambda} = -\underline{r_i} \cdot \underline{S_0} + \underline{r_i} \cdot \underline{S_1}$$

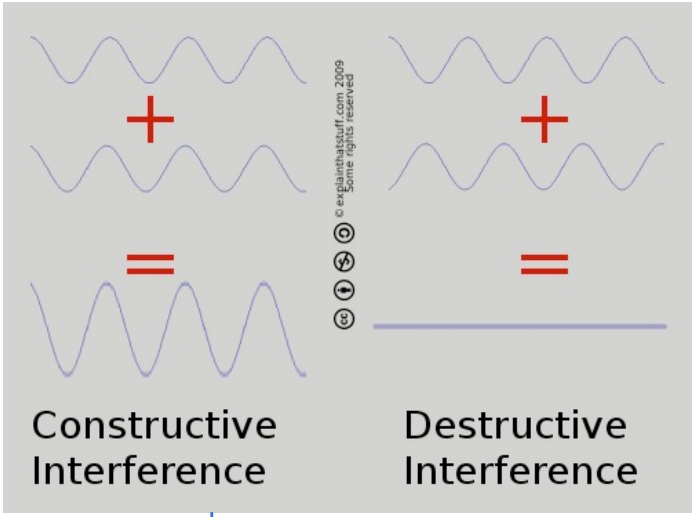
$$= \underline{r_i} \cdot (\underline{S_1} - \underline{S_0})$$

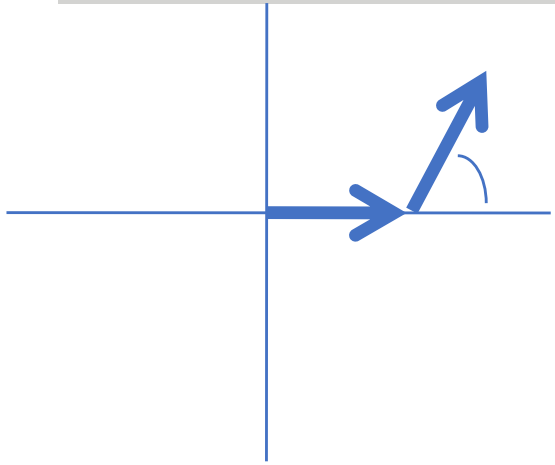
So we define $\underline{S} = S_1 - S_0$



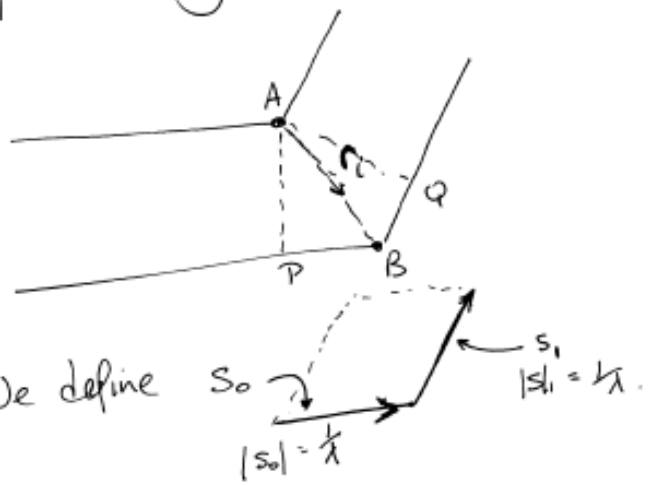

Then $\frac{\Phi}{\lambda} = \underline{r} \cdot \underline{S}$

Adding up the scattering of Atoms: 'interference' of waves



The phase lag can also be simplified

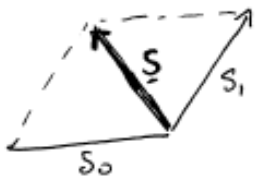



the PB = projection of $r_i$ on $S_0$
BQ = projection of $r_i$ on $S_1$

$$\text{So } \frac{\Phi}{\lambda} = \frac{PB + BQ}{\lambda} = -\underline{r_i} \cdot \underline{S_0} + \underline{r_i} \cdot \underline{S_1}$$

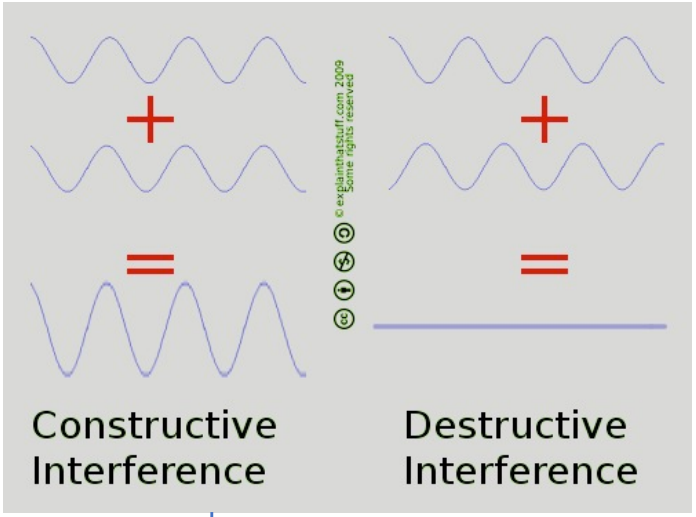
$$= \underline{r_i} \cdot (\underline{S_1} - \underline{S_0})$$

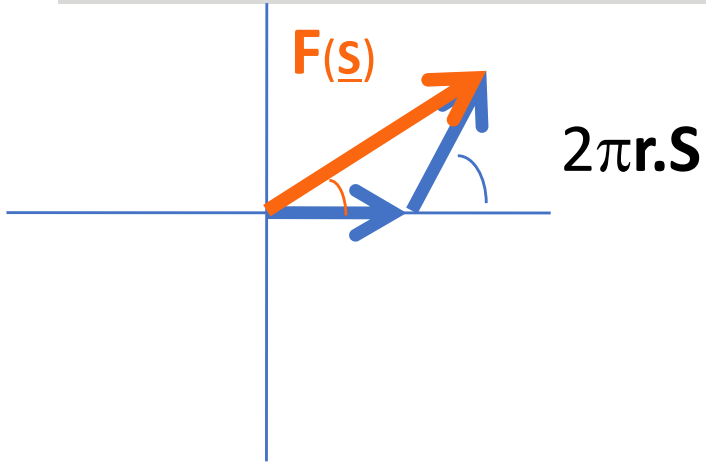
So we define $\underline{S} = S_1 - S_0$



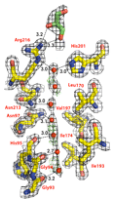

$$\text{Then } \frac{\Phi}{\lambda} = \underline{r} \cdot \underline{S}$$

Adding up the scattering of Atoms: 'interference' of waves



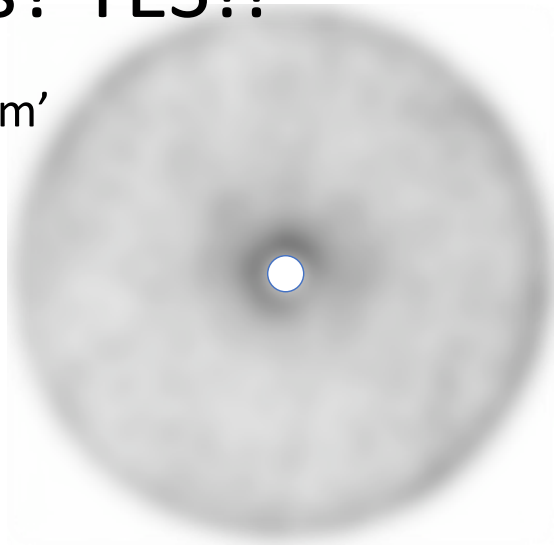
# This is all there is? YES!!



'Fourier Transform'

FT

$FT^{-1}$

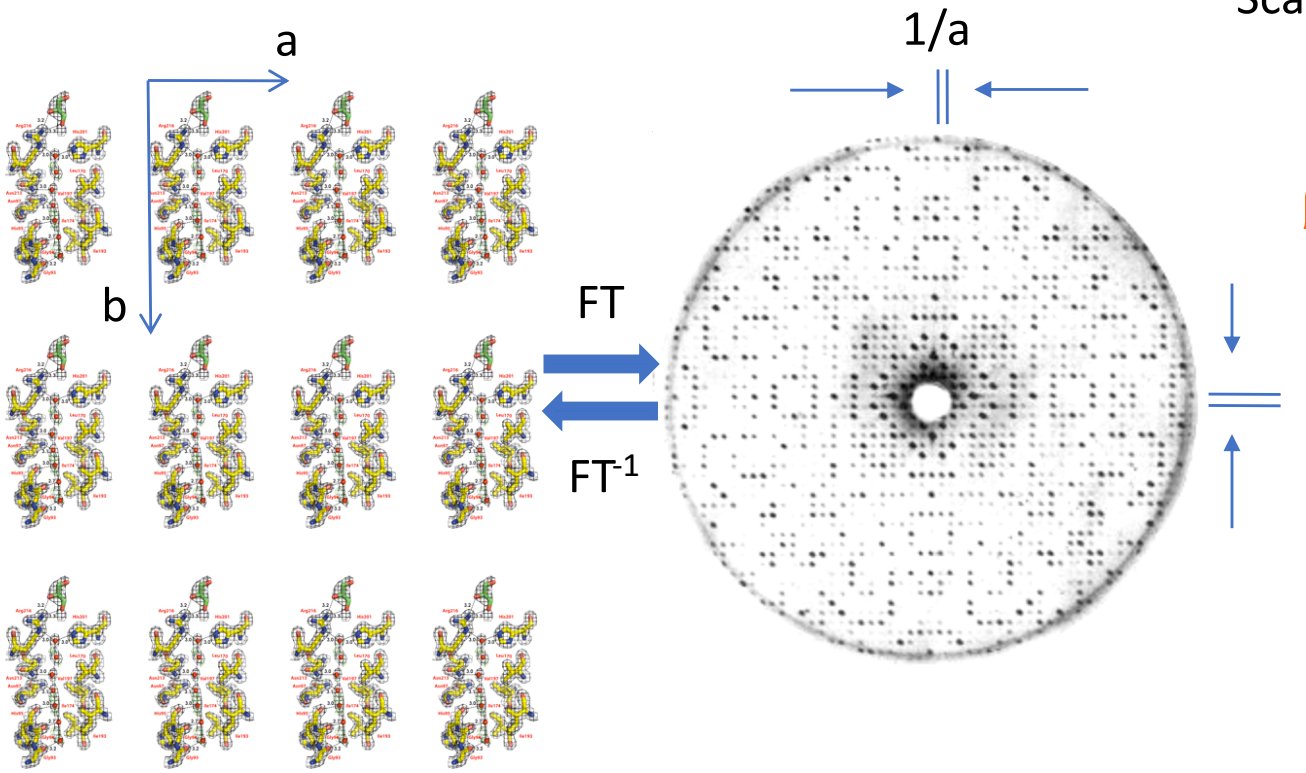


Scattering pattern is the Fourier transform (FT) of the structure: Amplitude and phase of waves is a sum of waves from each atom

$$\mathbf{F(\underline{S})} = \Sigma_j f_j e^{(2\pi i \mathbf{r_j.S})}$$

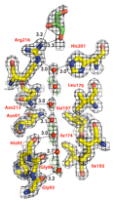
**Observe I(S) = F(S).F*(S)**

Structure is the 'inverse' Fourier transform of the Scattering pattern **F(S)**



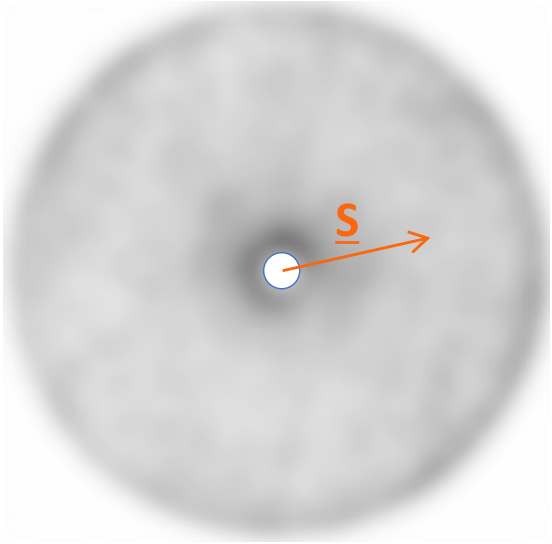

$$\rho(\underline{r}) = \Sigma\, \mathbf{F(\underline{S})}\, e^{(-2\pi i \mathbf{r.S})}$$

# This is all there is?



FT

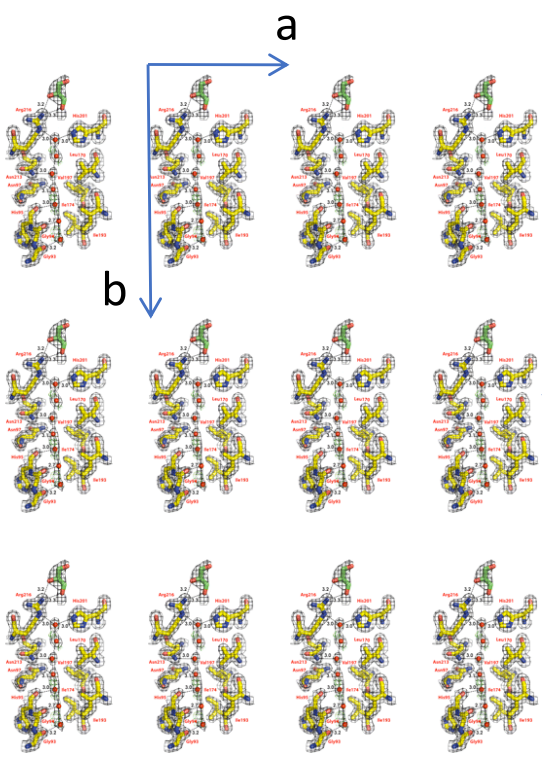
$FT^{-1}$




Scattering pattern is the Fourier transform of the structure

$$\underline{F}(\underline{S}) = \Sigma_j f_j e^{(2\pi i \mathbf{rj.S})}$$

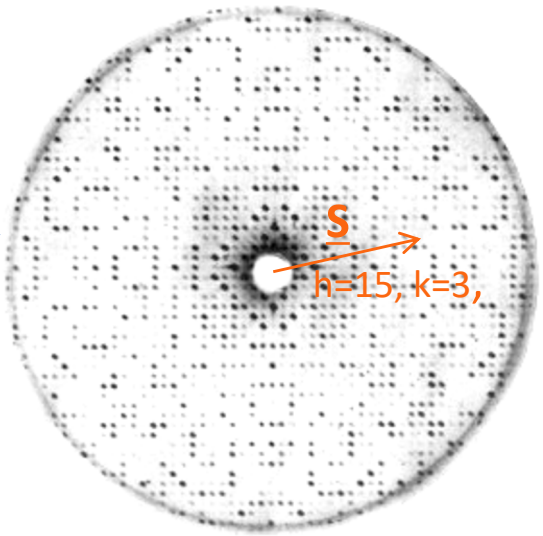
Structure is the 'inverse' Fourier transform of the Scattering pattern




FT

$FT^{-1}$

1/a




1/b

$$\rho(\underline{\mathbf{r}}) = \Sigma \underline{F}(\underline{S}) e^{(-2\pi i \mathbf{r.S})}$$

$$\underline{F}(h,k,l) = \Sigma_j f_j e^{(2\pi i\ (hx+ky+lz))}$$

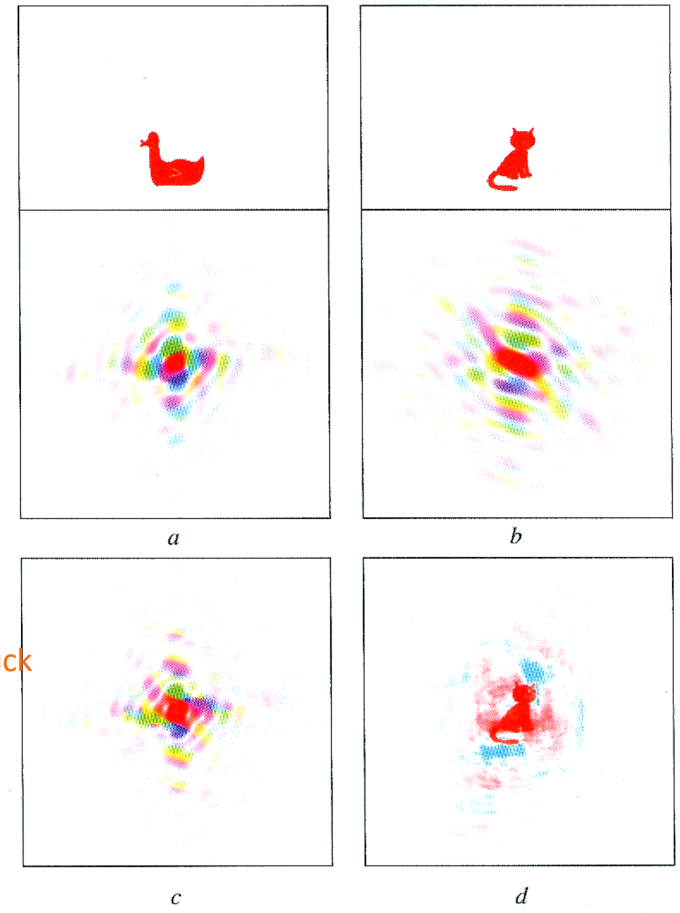
$$\rho(x,y,z) = \Sigma \underline{F}(h,k,l) e^{(-2\pi i \mathbf{r.S})}$$

# Relative Information in Intensities versus phases??

$\rho(\underline{r})$ duck

$\rho(\underline{r})$ cat

$\underline{F}_{(\underline{S})}$ duck → $|F_{(\underline{S})}|$ $\phi_{(\underline{S})}$




$\underline{F}_{(\underline{S})}$ cat

$\underline{F}_{(\underline{S})} = \Sigma_j f_j e^{(2\pi i \mathbf{rj.S})}$

$|F_{(\underline{S})}|$ duck

$\phi_{(\underline{S})}$ cat

$\rho(\underline{r}) = \Sigma \underline{F}_{(\underline{S})} e^{(-2\pi i \mathbf{r.S})}$

Looks like a .....

**Figure 6.1** ▶ Relative amounts of information contained in reflection intensities and phases. (*a*) and (*b*) Duck and cat, along with their Fourier transforms. (*c*) Intensity (shading) of the duck transform, combined with the phases (colors) of the cat transform. (*d*) Back-transform of (*c*) produces recognizable image of cat, but not duck. Phases contain more information than intensities. Figure generously provided by Dr. Kevin Cowtan.

# Relative Information in Intensities versus phases

(bold letters= vectors)

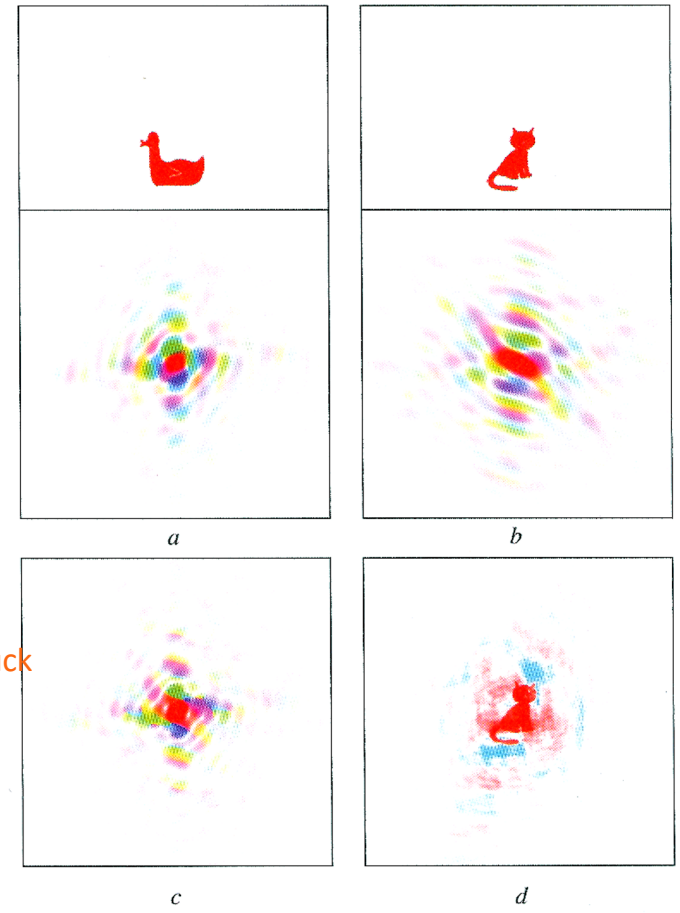
$\rho(\underline{\mathbf{r}})$ duck

$\rho(\underline{\mathbf{r}})$ cat

$\underline{\mathbf{F}}_{(\underline{\mathbf{S}})}$ duck → $|F_{(\underline{\mathbf{S}})}|$, $\phi_{(\underline{\mathbf{S}})}$

$\underline{\mathbf{F}}_{(\underline{\mathbf{S}})}$ cat

$\underline{\mathbf{F}}_{(\underline{\mathbf{S}})} = \Sigma_j f_j e^{(2\pi i \mathbf{rj.S})}$




$|F_{(\underline{\mathbf{S}})}|$ duck, $\phi_{(\underline{\mathbf{S}})}$ cat

$\rho(\underline{\mathbf{r}}) = \Sigma\, \underline{\mathbf{F}}_{(\underline{\mathbf{S}})} e^{(-2\pi i \mathbf{r.S})}$

**Figure 6.1** ▶ Relative amounts of information contained in reflection intensities and phases. (*a*) and (*b*) Duck and cat, along with their Fourier transforms. (*c*) Intensity (shading) of the duck transform, combined with the phases (colors) of the cat transform. (*d*) Back-transform of (*c*) produces recognizable image of cat, but not duck. Phases contain more information than intensities. Figure generously provided by Dr. Kevin Cowtan.

Looks like a CAT

PHASES DOMINATE:

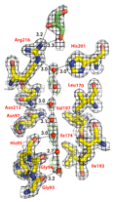
-Incorrect phases = incorrect structure

-incorrect model = incorrect structure

-incorrect assumption = incorrect structure

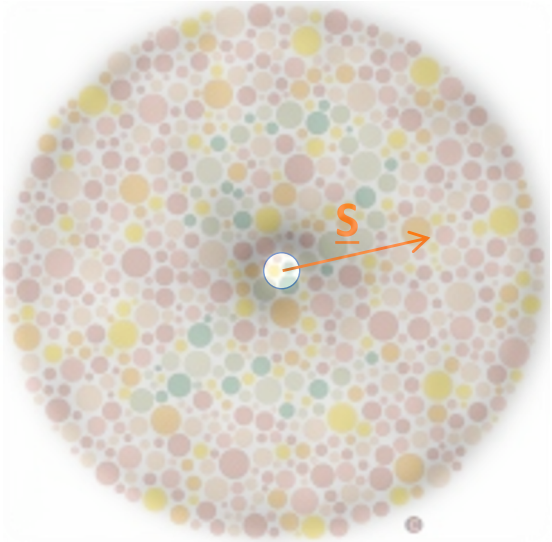
This is all there is?

# PHASES-as colors !



FT

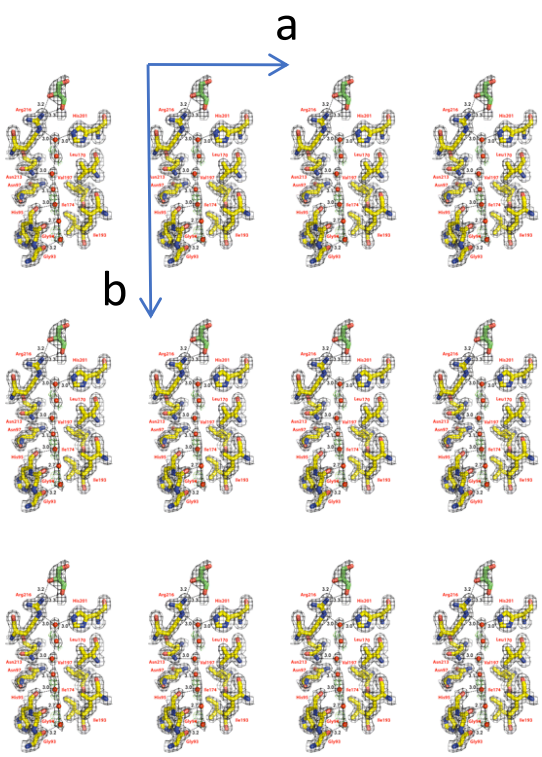
FT$^{-1}$




Scattering pattern is the Fourier transform of the structure

$$\underline{F}(\underline{S}) = \Sigma_j f_j e^{(2\pi i \mathbf{rj.S})}$$

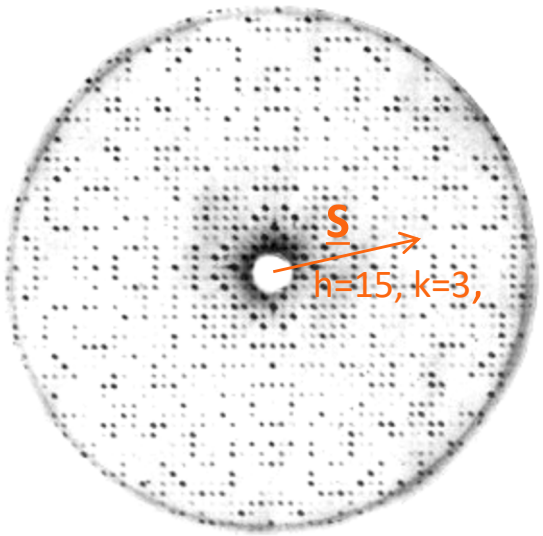
Structure is the 'inverse' Fourier transform of the Scattering pattern




FT

FT$^{-1}$

1/a




1/b

$$\rho(\underline{r}) = \Sigma \underline{F}(\underline{S}) e^{(-2\pi i \mathbf{r.S})}$$

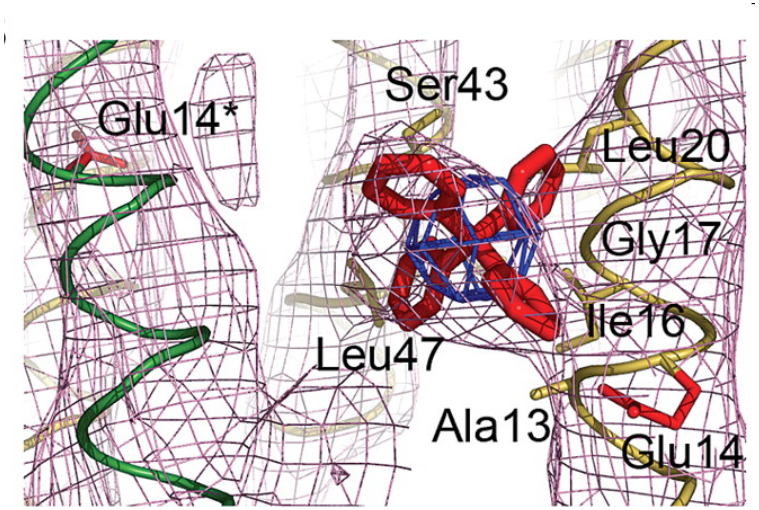
$$\underline{F}(h,k,l) = \Sigma_j f_j e^{(2\pi i\ (hx+ky+lz))}$$

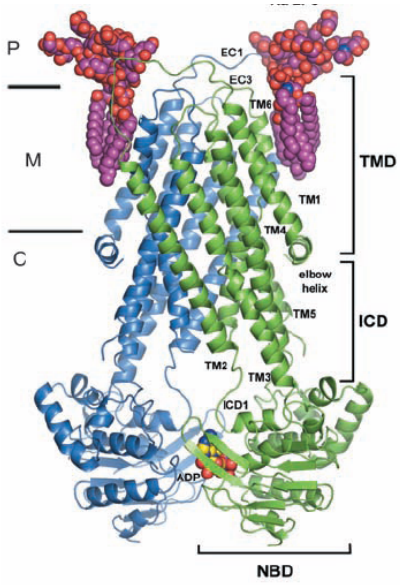
$$\rho(x,y,z) = \Sigma \underline{F}(h,k,l) e^{(-2\pi i \mathbf{r.S})}$$

OK OK. = I get it,- computers.....why should I care???

Why should I care??

# Unintentionally Incorrect Results



1. Chang, G., Roth, C. B. (2001) Structure of MsbA from E. coli: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science* 293(5536):1793-800.
2. Chang, G. (2003). Structure of MsbA from Vibrio cholera: a multidrug resistance ABC transporter homolog in a closed conformation. *J Mol Biol* 330(2):419-30.
3. Ma, C, Chang, G. (2004). Structure of the multidrug resistance efflux transporter EmrE from Escherichia coli. *Proc Natl Acad Sci USA* 101(9):2852-7.
4. Pornillos, O., Chen, Y. J., Chen, A. P., Chang, G. (2005) X-ray structure of the EmrE multidrug transporter in complex with a substrate. *Science* 310(5756):1950-3.
5. Reyes, C. L., Chang, G. (2005) Structure of the ABC transporter MsbA in complex with ADP.vanadate and lipopolysaccharide. *Science* 308(5724):1028-31.

# Intentionally Incorrect Results

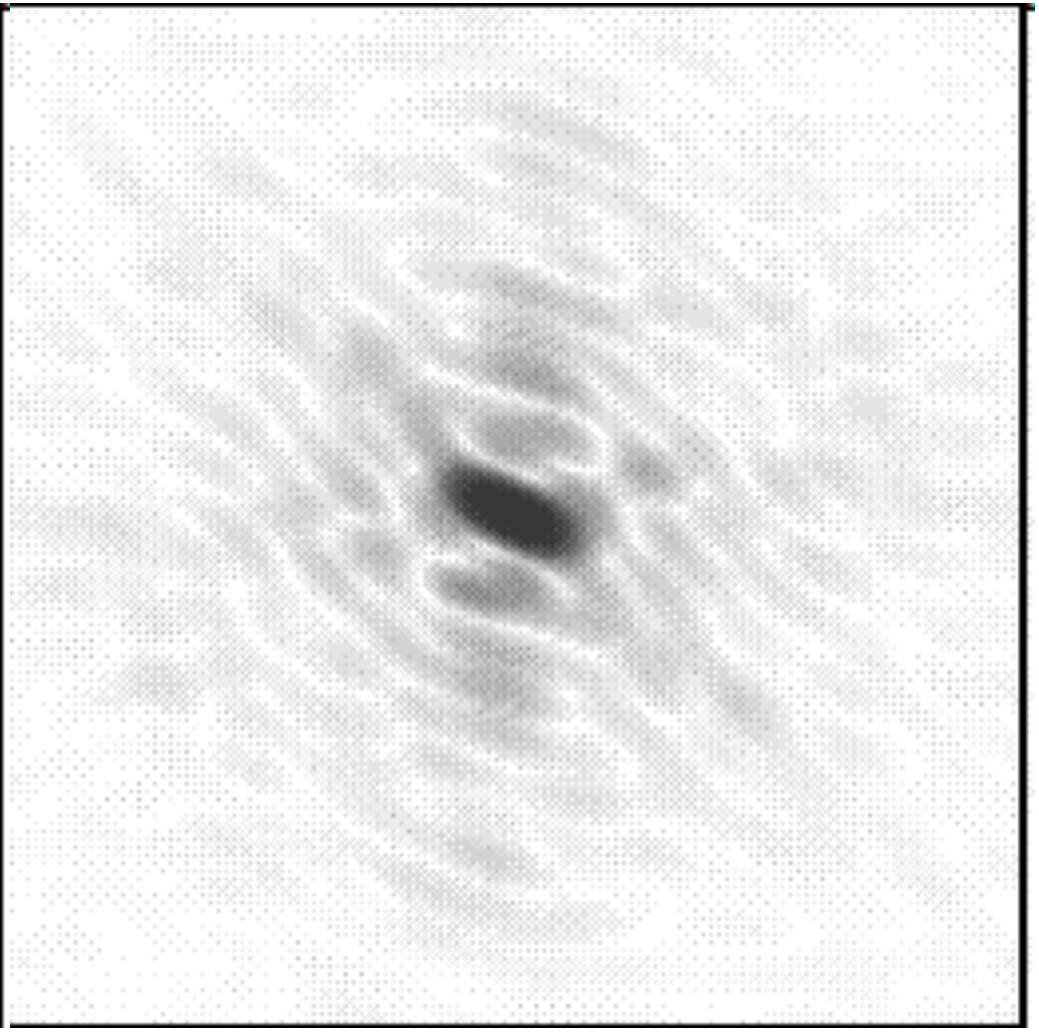

1. Murali, R., Sharkey, D. J., Daiss, J. L. & Murthy, H. M. (1998)."Crystal structure of Taq DNA polymerase in complex with an inhibitory Fab: The Fab is directed against an intermediate in the helix-coil dynamics of the enzyme", *PNAS USA* **95**, 12562.
2. Murthy, H. M., Clum, S. & Padmanabhan, R. (1999)."Dengue virus NS3 serine protease. Crystal structure and insights into interaction of the active site with substrates by molecular modeling and structural analysis of mutational effects", *J. Biol. Chem.* **274**, 5573.
3. Urs, U. K., Murali, R. & Murthy, K. (1999)."Structure of Taq DNA polymerase shows a new orientation for the structure-specific nuclease domain", *Acta Cryst. D* **55**, 1971-1977.
4. Murthy, H. M. K., Judge, K., DeLucas, L. & Padmanabhan, R. (2000)."Crystal structure of Dengue virus NS3 protease in complex with a Bowman-Birk inhibitor: implications for flaviviral polyprotein processing and drug design", *J. Mol. Biol.* **301**, 759-767.
5. Murthy, K. H. M., Smith, S. A., Ganesh, V. K., Judge, K. W., Mullin, N., Barlow, P. N., Ogata, C. M. & Kotwal, G. J. (2001)."Crystal structure of a complement control protein that regulates both pathways of complement activation and binds heparan sulfate proteoglycans", *Cell* **104**, 301-311.
6. Kumar, M. S., Carson, M., Hussain, M. M. & Murthy, H. M. K. (2002)."Structures of Apolipoprotein A-II and a Lipid- Surrogate Complex Provide Insights into Apolipoprotein- Lipid Interactions", *Biochemistry* **41**, 11681-11691.
7. Ganesh, V. K., Smith, S. A., Kotwal, G. J. & Murthy, K. H. M. (2004)."Structure of vaccinia complement protein in complex with heparin and potential implications for complement regulation", *PNAS USA* **101**, 8924.
8. Ganesh, V. K., Muthuvel, S. K., Smith, S. A., Kotwal, G. J. & Murthy, K. H. M. (2005)."Structural Basis for Antagonism by Suramin of Heparin Binding to Vaccinia Complement Protein", *Biochemistry* **44**, 10757-10765.
9. Ajees, A. A., Anantharamaiah, G. M., Mishra, V. K., Hussain, M. M. & Murthy, H. M. (2006)."Crystal structure of human apolipoprotein AI: insights into its protective effect against cardiovascular diseases", *PNAS USA* **103**, 2126.
10. Ajees, A. A., Gunasekaran, K., Volanakis, J. E., Narayana, S. V. L., Kotwal, G. J. & Murthy, H. M. K. (2006)."The structure of complement C3b provides insights into complement activation and regulation", *Nature* **444**, 221-225.

# Major Phasing techniques

- Molecular Replacement
- Multiple Isomorphous Replacement
- Multiwavelength Anomalous Diffraction
- Single-wavelength Anomalous Diffraction

# Molecular Replacement

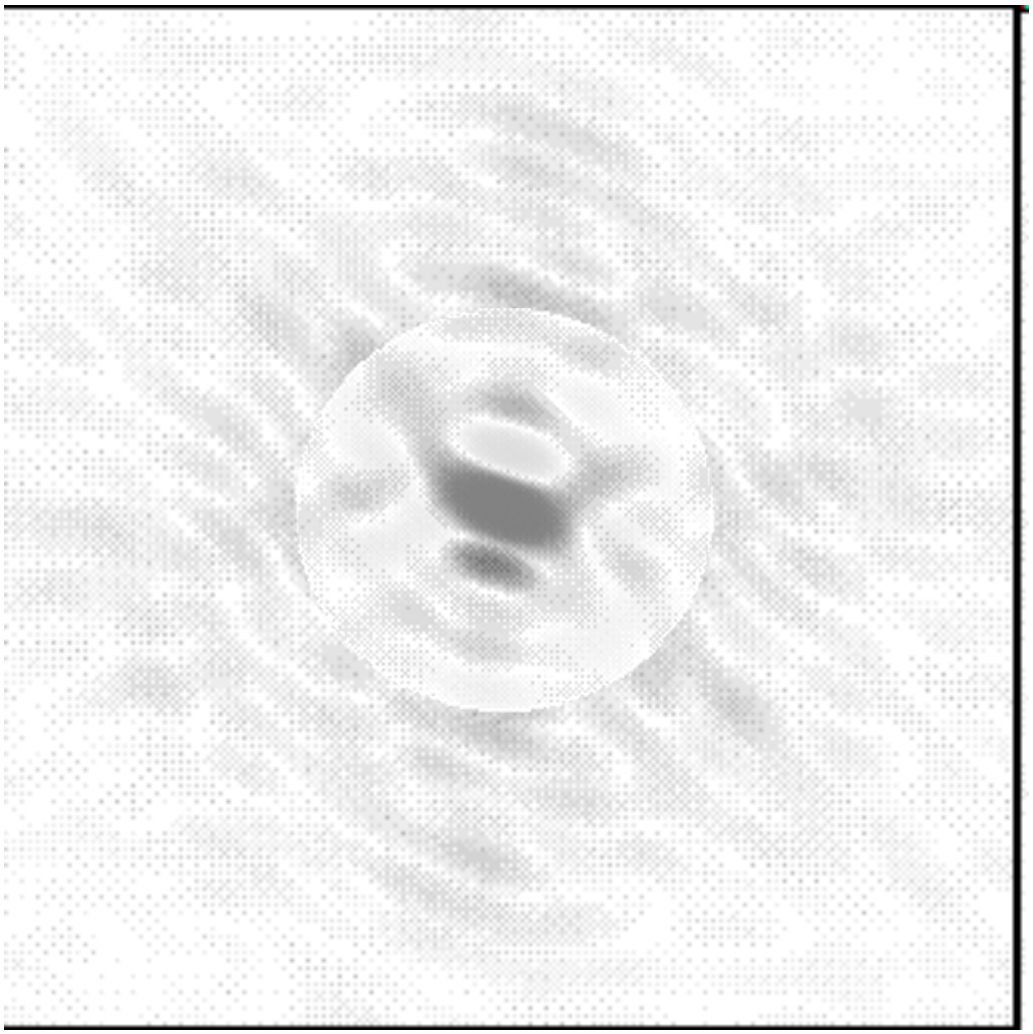
correct structure and intensities



http://www.ysbl.york.ac.uk/~cowtan/fourier/coeff.html

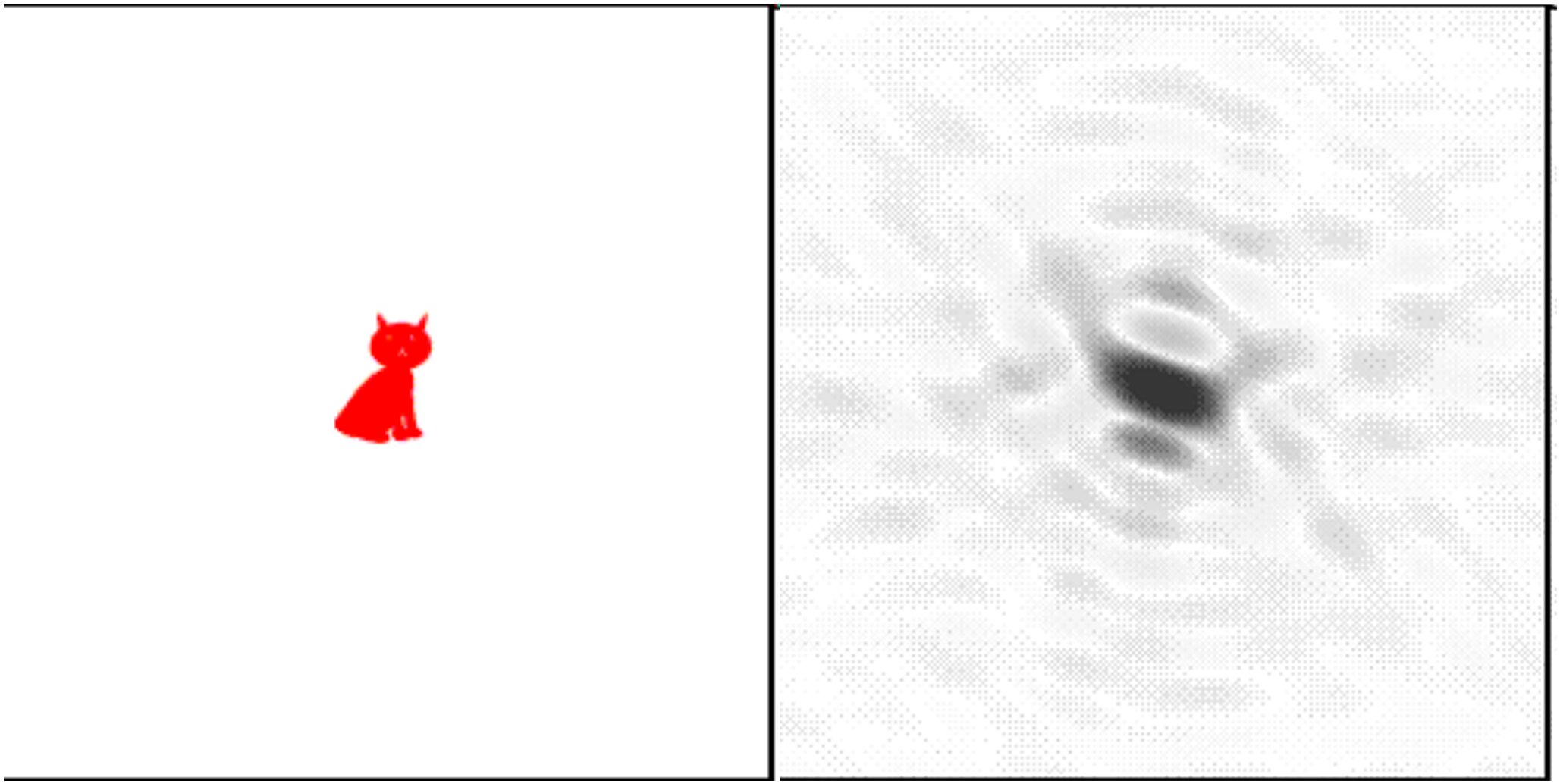
# Molecular Replacement

use something similar as a starting model



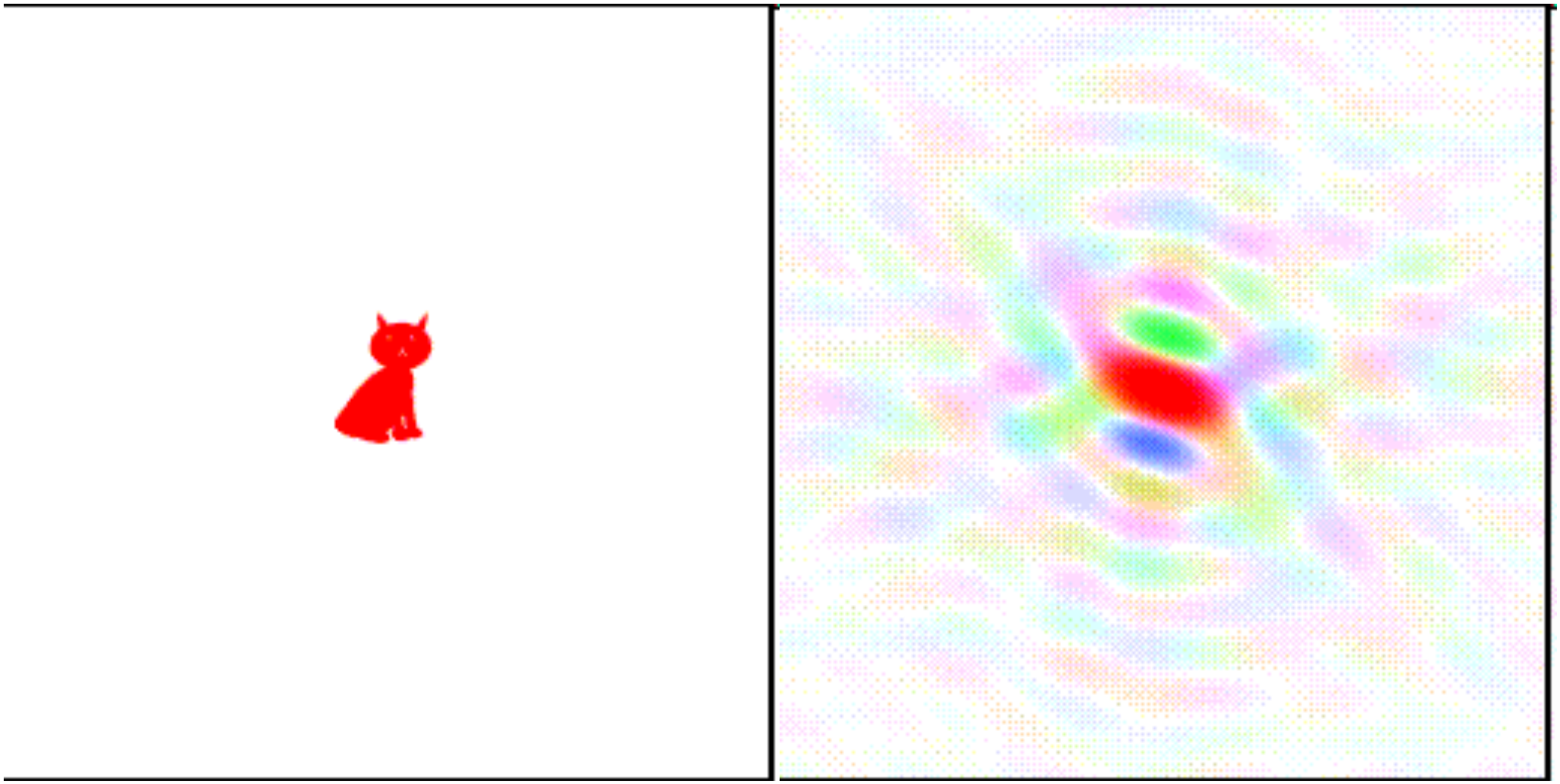
# Model Building

current model is missing something



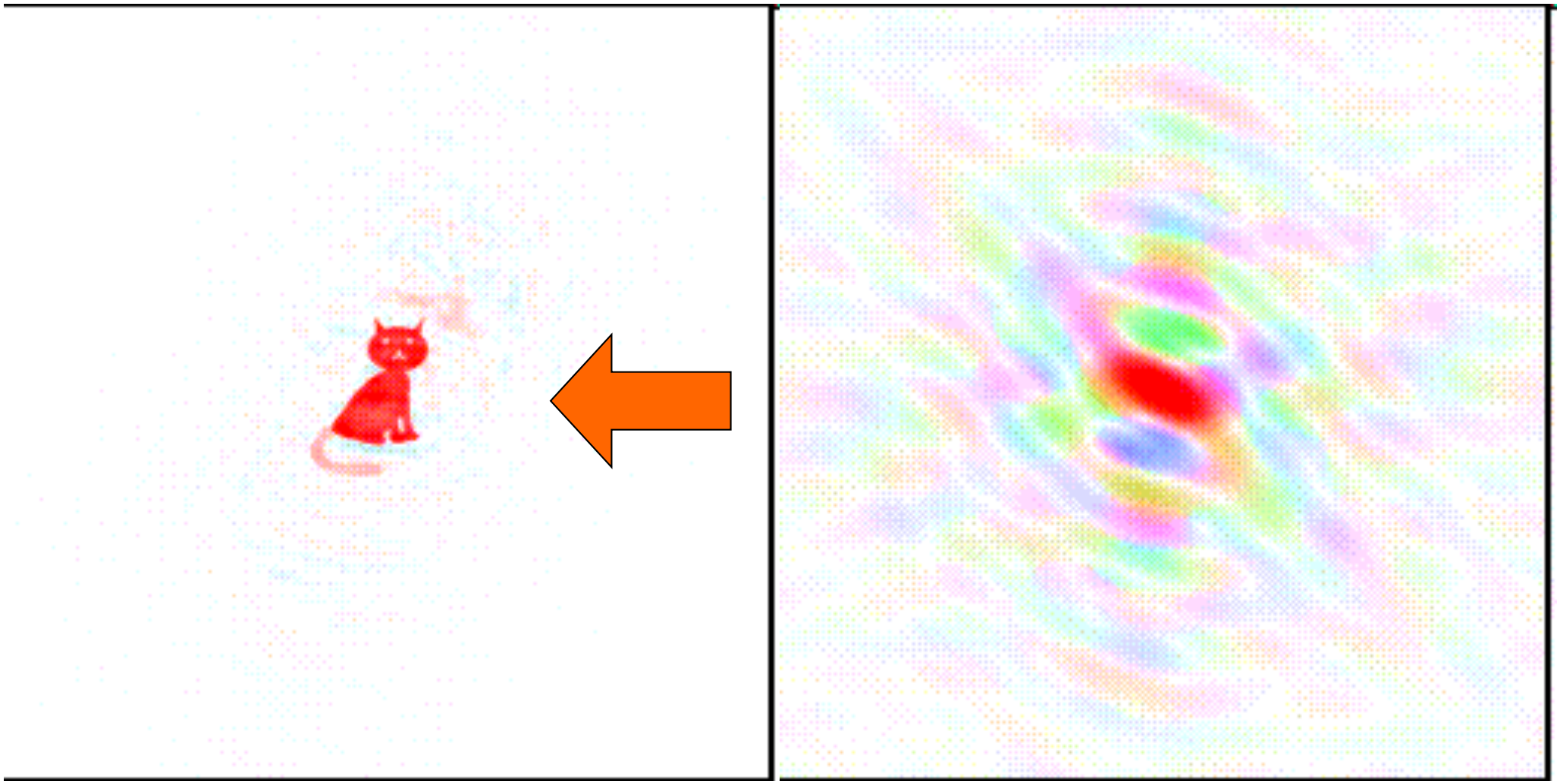
# Model Building

phases from model



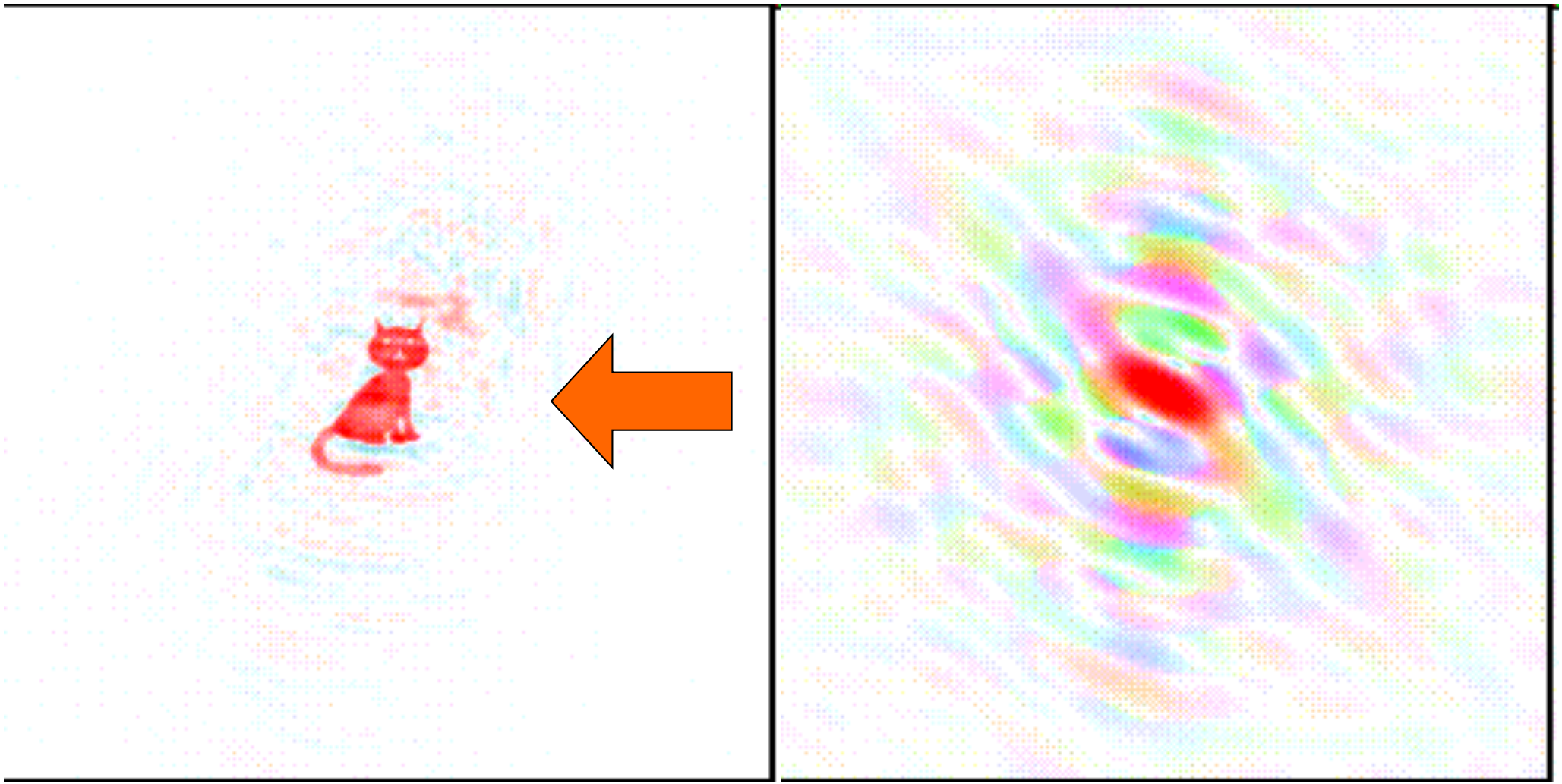
# Model Building

missing bits show up in “difference map”

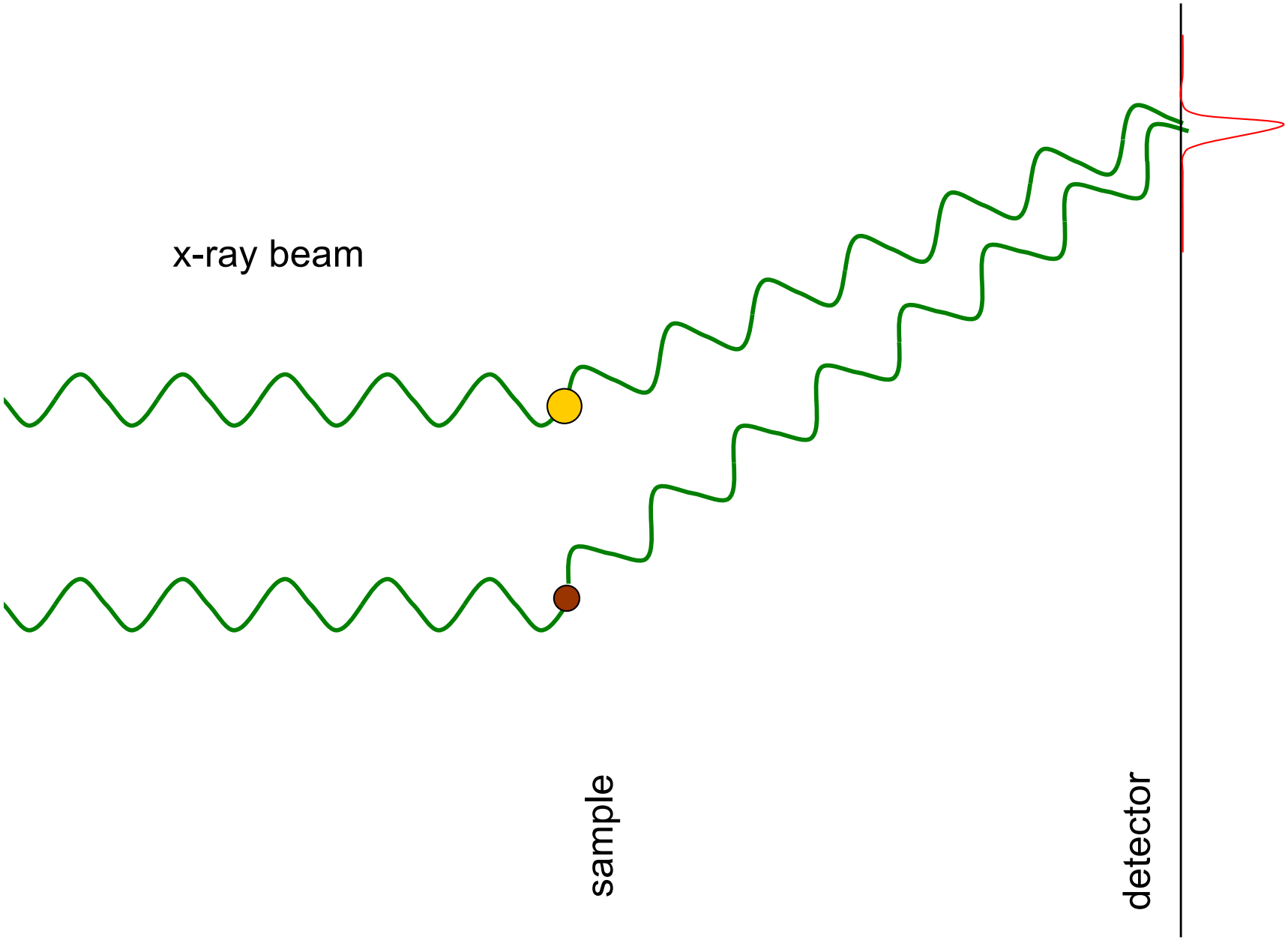


# Model Building

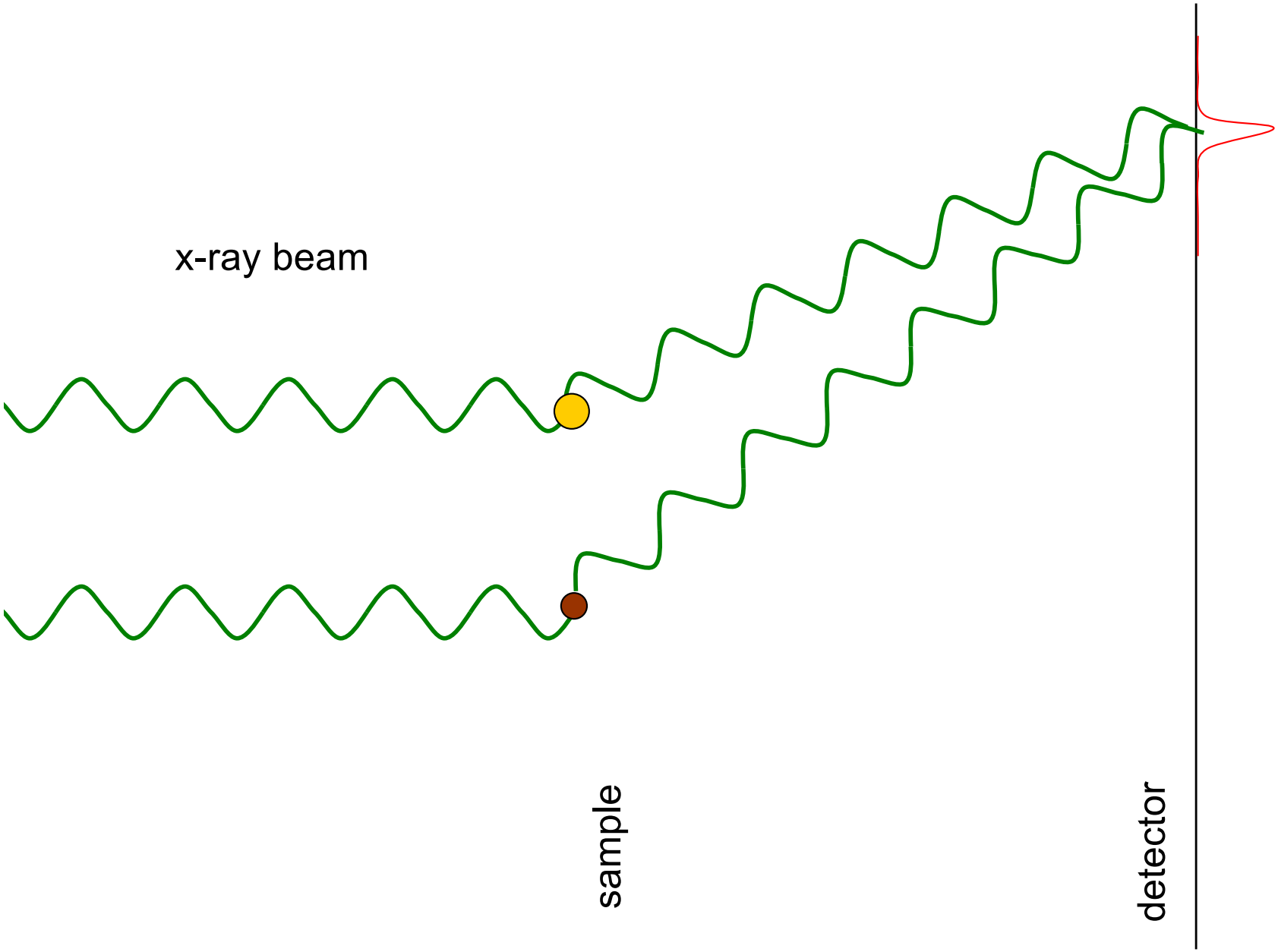
missing bits show up better in $F_O + (F_O - F_C)$ map



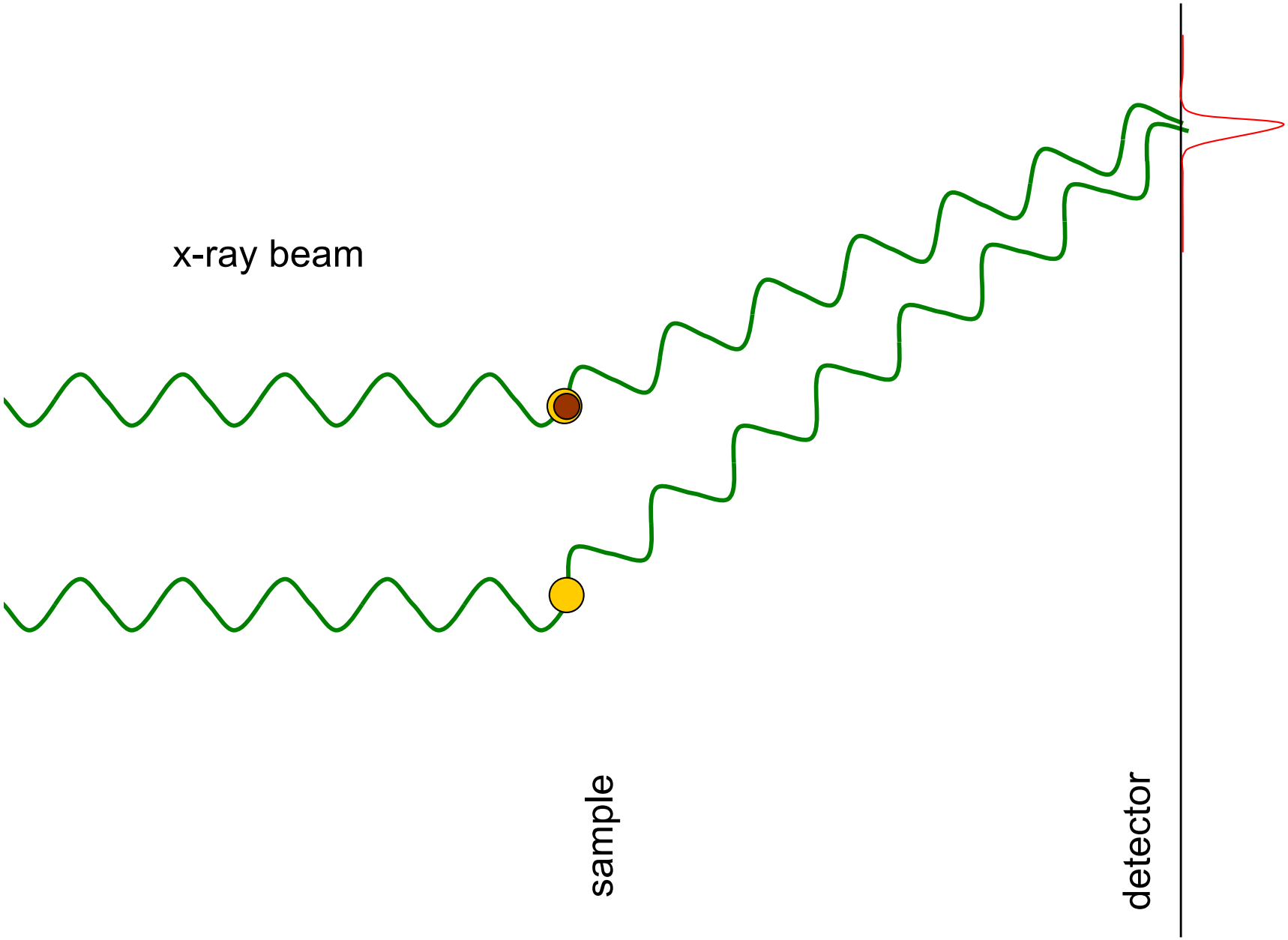
# anomalous scattering



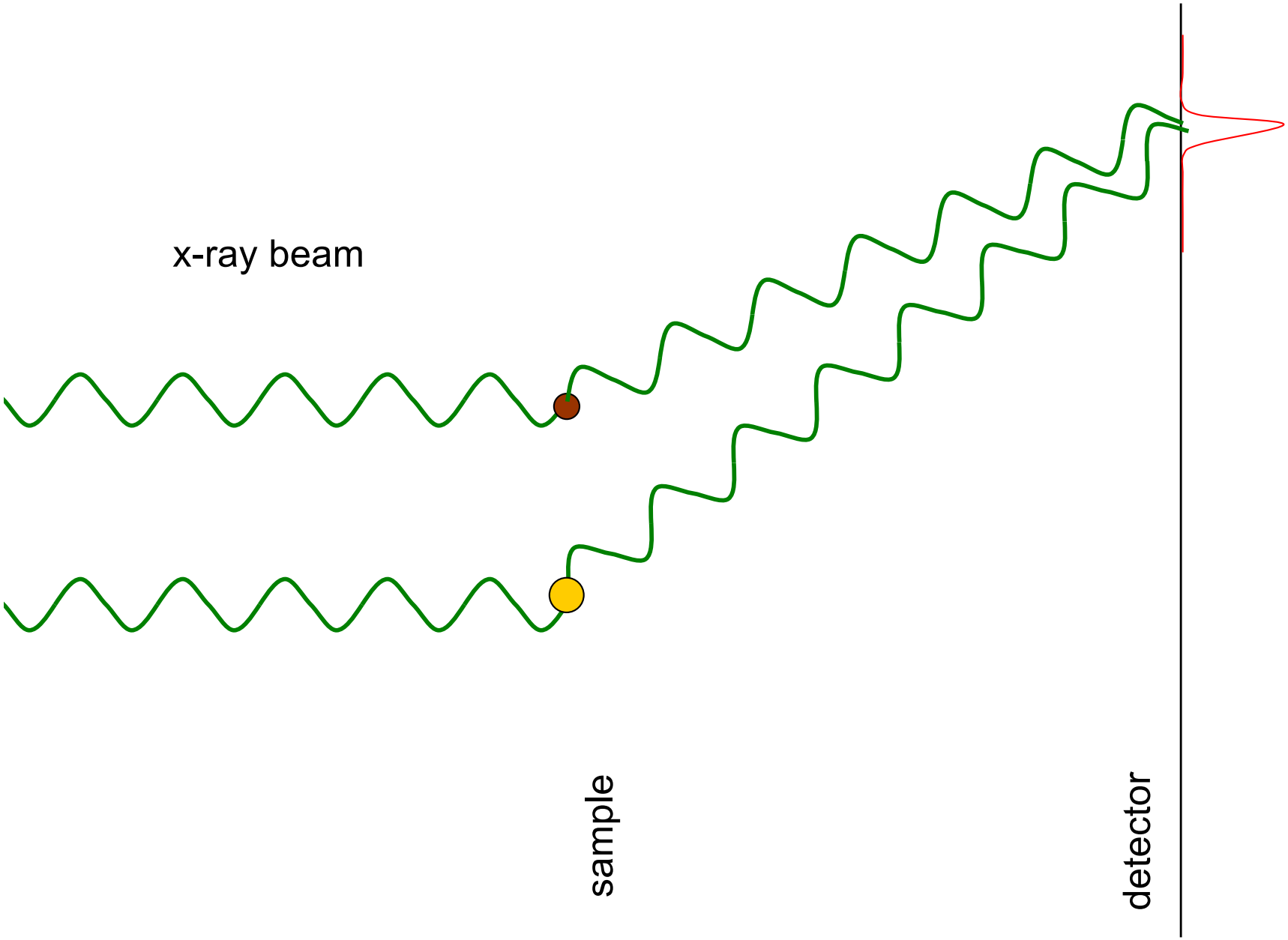
# anomalous scattering



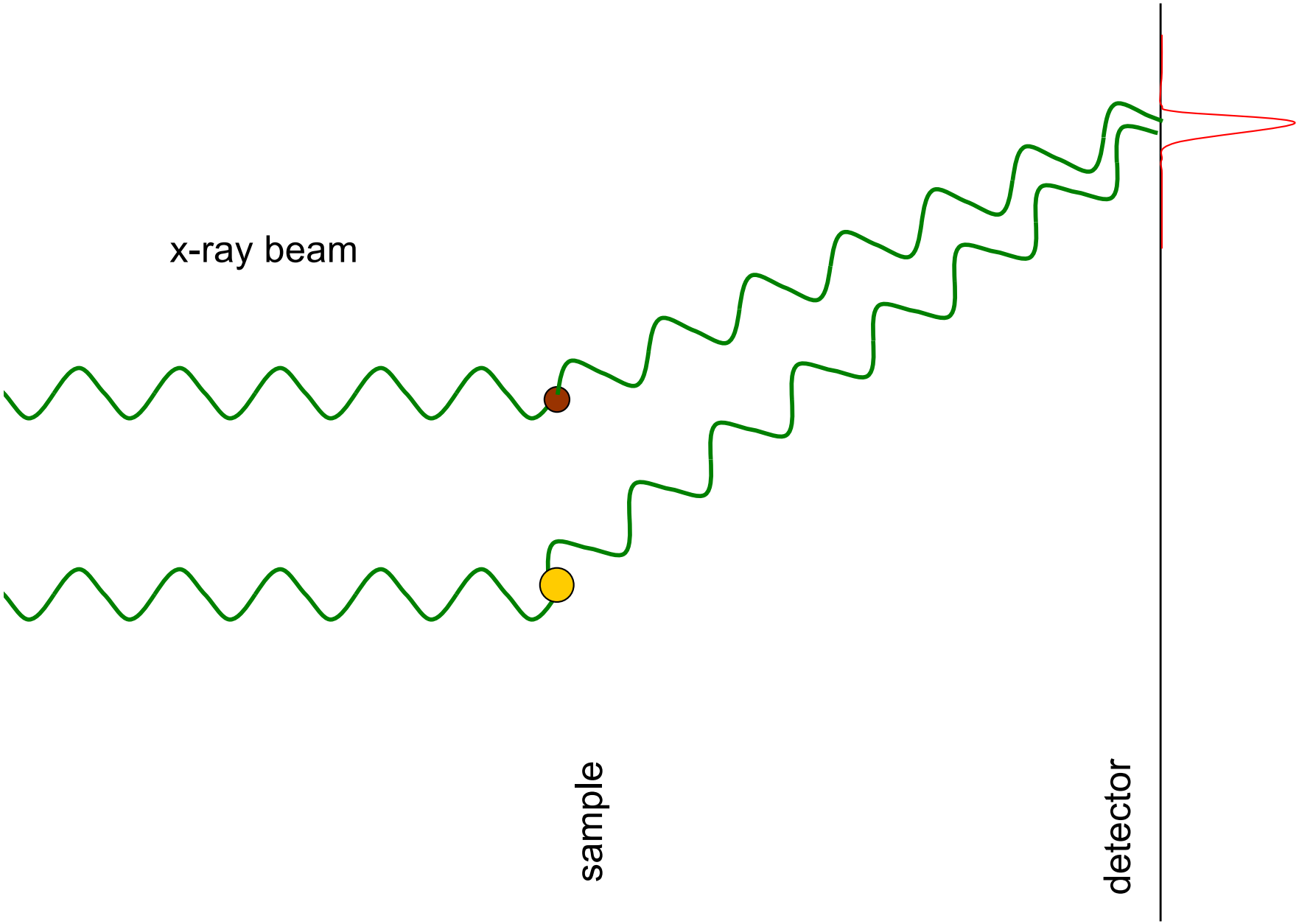
# anomalous scattering



# anomalous scattering



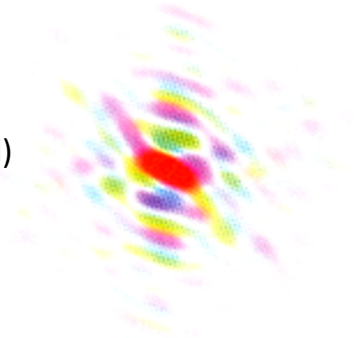
# anomalous scattering



# Lecture 3.

Data to Parameter Ratio
Refinement
Constraints/Restraints
Validation:  R factors..

# AXIOM: Forward FT ⟺ Back FT-1 are Truly Inverse



Crystal $I_{(h,k,l)}$

Amplitudes $F_{(h,k,l)}$

Electron density $\rho(x,y,z)$

Phases $\phi_{(h,k,l)}$

Known:
Amino acid sequence
Ligands
Bond lengths angles
Constraints on geometry

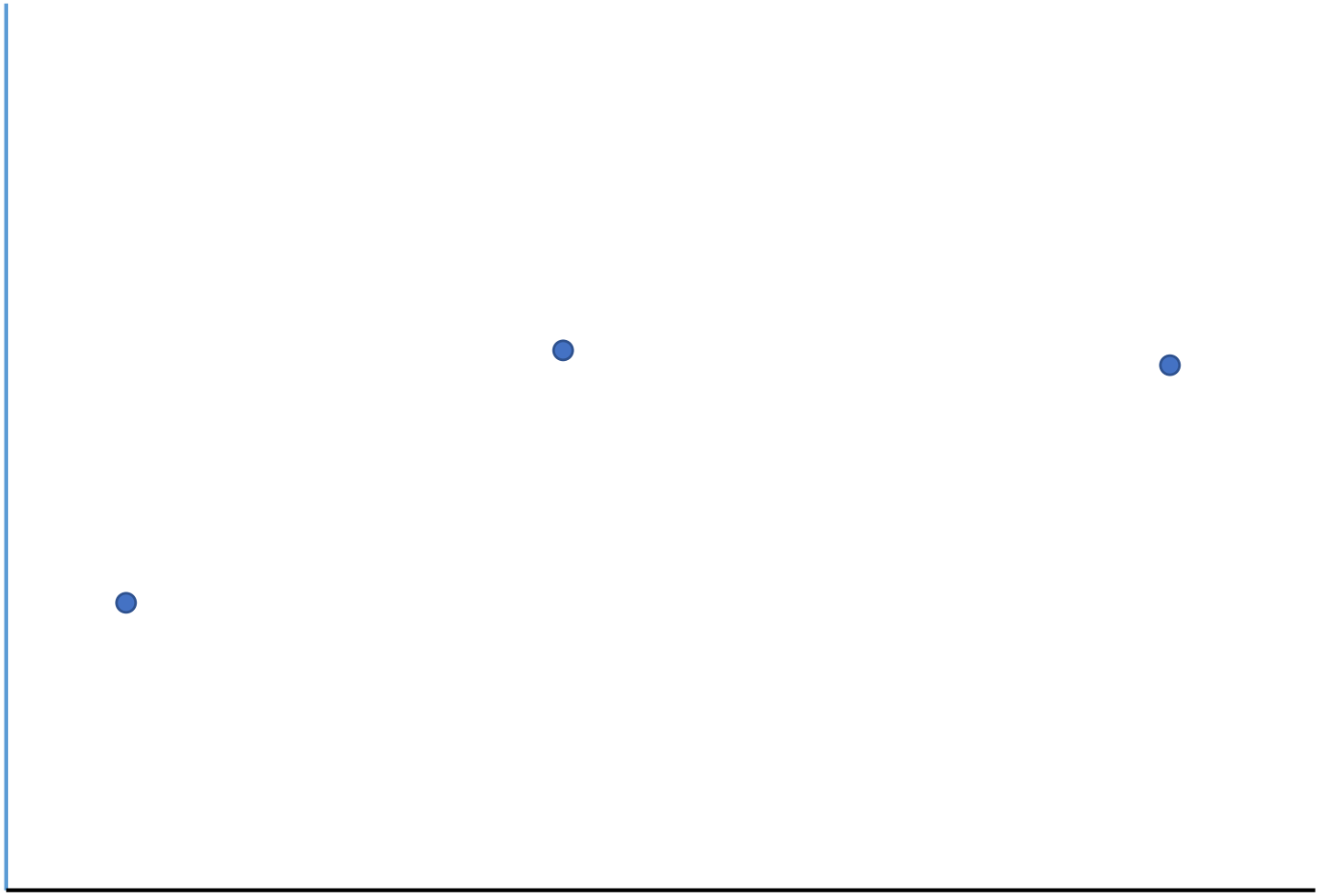
1. Experimental:
heavy atom labels
selenium for sulfur

2. Trial & error similar structure

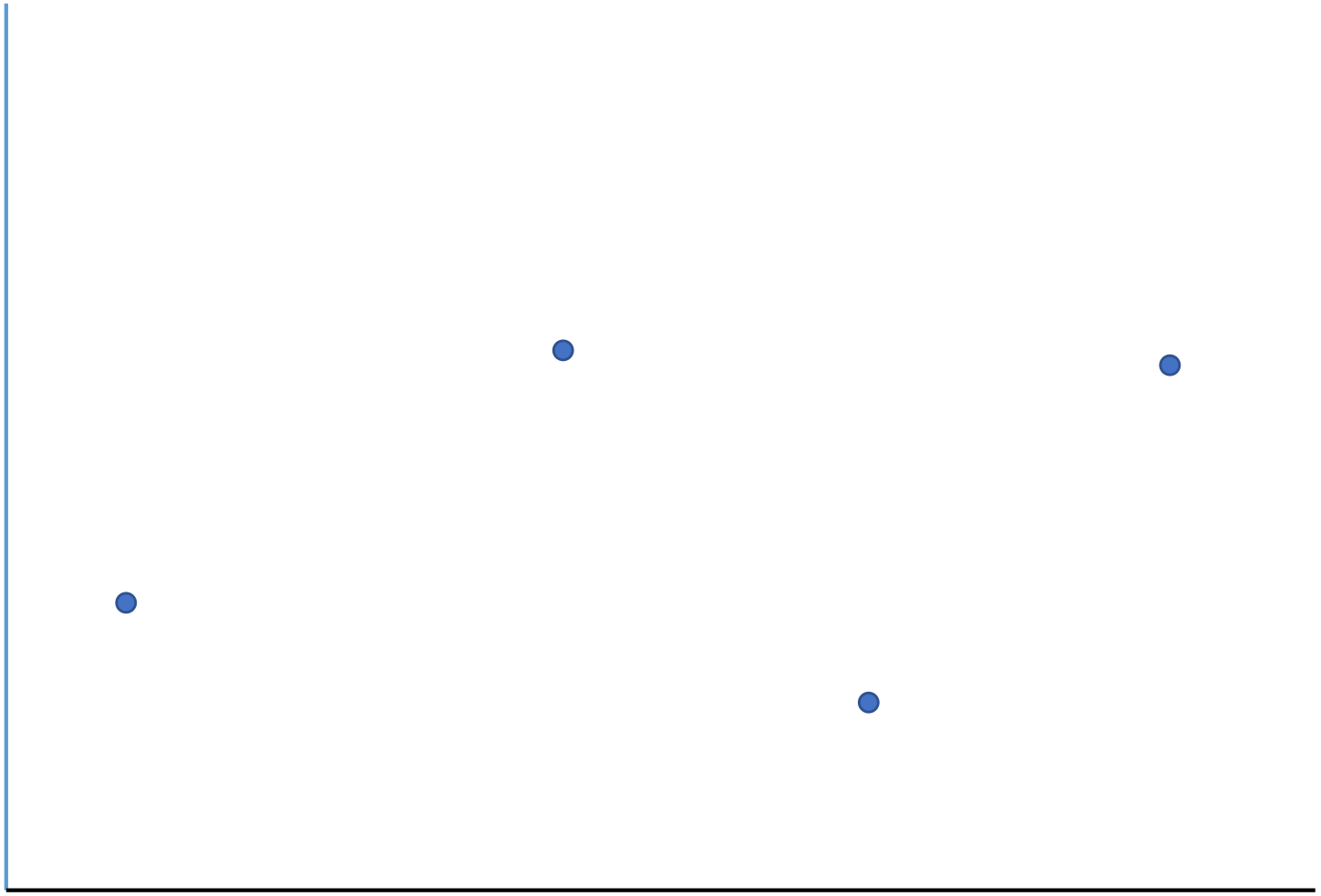
Atom positions (x,y,z)

# Data to parameter ratio

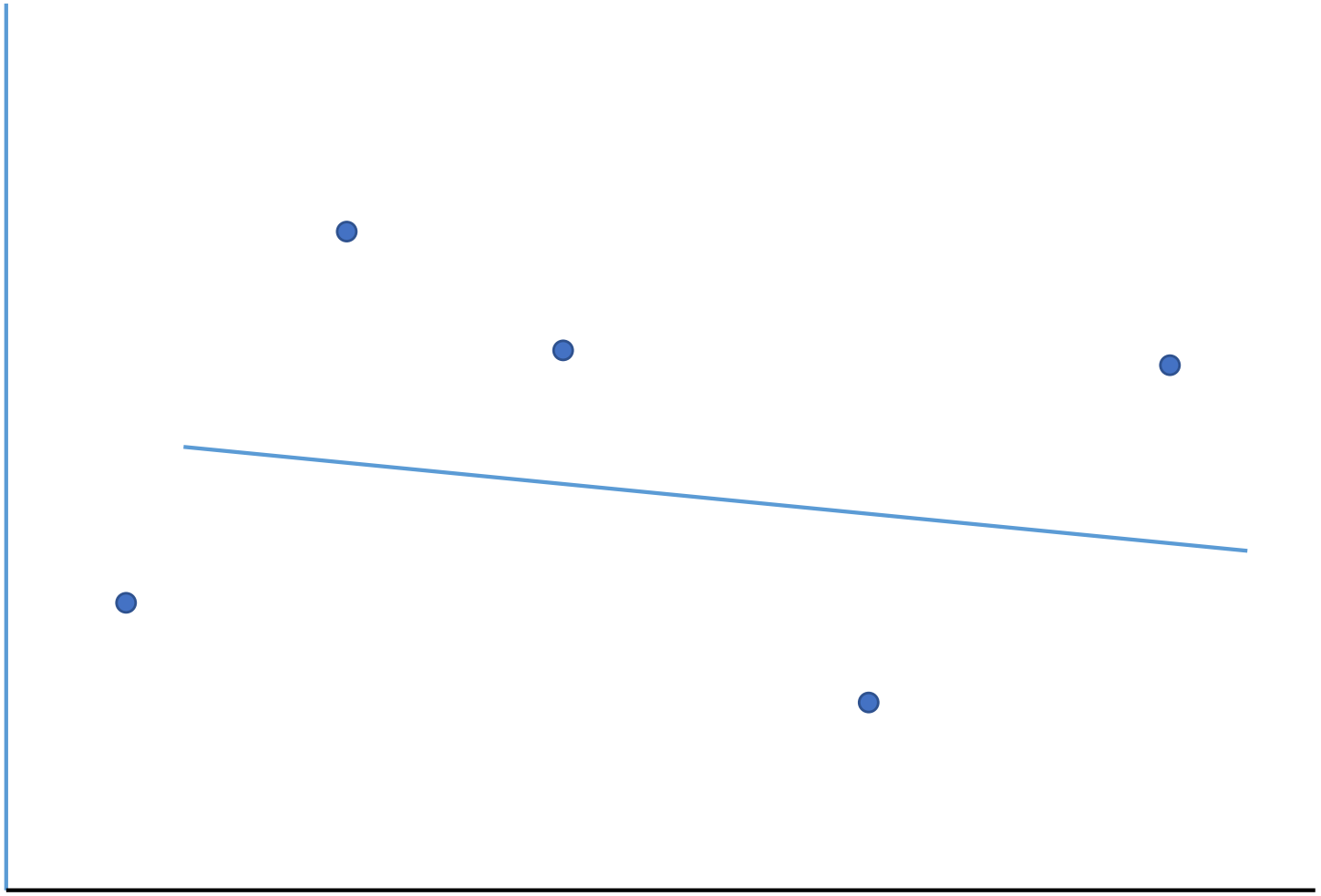
# Data to parameter ratio



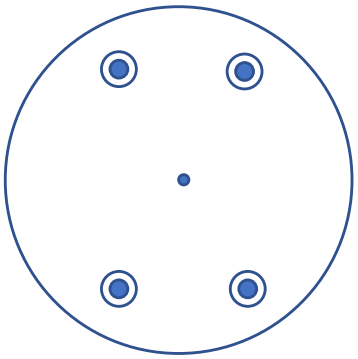
# Data to parameter ratio



# Data to parameter ratio



# Data to parameter ratio?? 5PTP



Bovine trypsin Unit cell V = 54.8Å 58.7Å 67.6Å

$V=217{,}453\ Å^3$

$V^*= 1/V$

**To 5Å Resolution**

hkl max 10.96 x 11.6 x 13.5

Volume Vs $=4/3\ \pi\ r^3 = 4.1888\ (1/5)^3$

No spots $=Vs/V$ = 7286 less the number in hk0 etc planes = 1350 hkls =5936

Multiplicity is 8 (all except planes, so ) ~ < 742 hkls

The planes have multiplicity of 4 so n of hk0 $\pi\ r^2$ = 338

Planes have multiplicity of 4 ~< 84

**Independent $I_{hkl}$ = 826**

number goes up as $(1/\text{resolution})^3$

**5Å 826 Ihkl**

**3.75 1,958 Ihkl**

**2.5Å 6,608 Ihkl**

**2Å 12,906 Ihkl**

**1.25Å 52,864Ihkl**

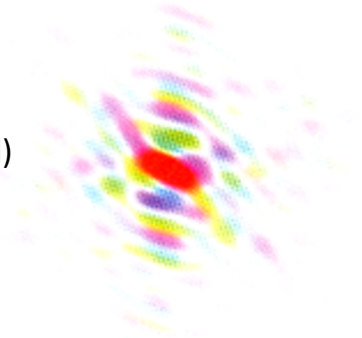
**5 PTP. Not enough Data??????**

**Parameters 223 amino acids 1655 atoms in protein 212 waters**

**=1867 atoms x, y, z, B x4 parameters = 7468 parameters**

| Resln. | Independent ($F_{hkl}$) obs | Data to parameter ratio Biso (4 para) |
|---|---|---|
| 5Å | 826 Ihkl | |
| 3.75 | 1,958 Ihkl | |
| 2.5Å | 6,608 Ihkl | 0.88 |
| 2Å | 12,906 Ihkl | 1.7 |
| 1.25Å | 52,864Ihkl | 7.1 |

# AXIOM: Forward FT ⟺ Back FT-1 are Truly Inverse



Crystal $I_{(h,k,l)}$

Amplitudes $F_{(h,k,l)}$

Electron density $\rho(x,y,z)$

Phases $\phi_{(h,k,l)}$

Known:
Amino acid sequence
Ligands
Bond lengths angles
Constraints on geometry

1. Experimental:
heavy atom labels
selenium for sulfur

2. Trial & error similar structure

Atom positions (x,y,z)

# Reduce parameters?

**1. 223 amino acids , phi, psi, xi 1 xi2 = 892 parameters + Bfactors protein only 1655=2,547**

Or

**2. 12 helices 6 parameters =72 parameters**

Match restraints in planarity, bond lengths in <1 Å resolution structures.
'Molprobity' (inc in Phenix refine)

Solvent Flattening

Density modification, Nothing should be negative electron density, All positive
Chemical reason, Hydrogen bonds distance, angles.
Proline isomerizations, oxidation?

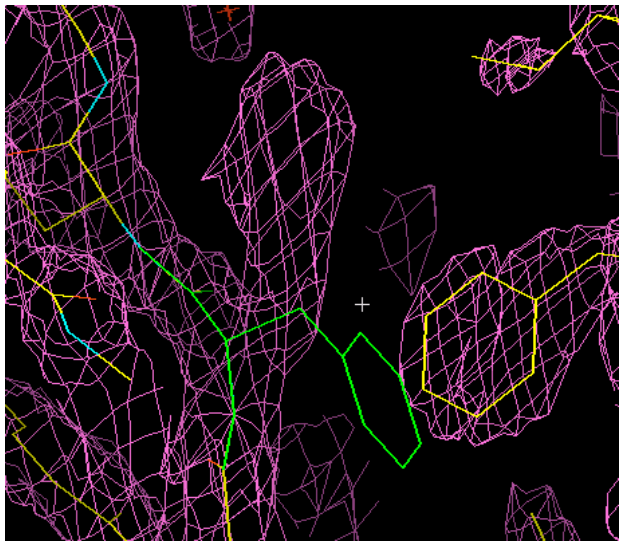
**5 PTP**

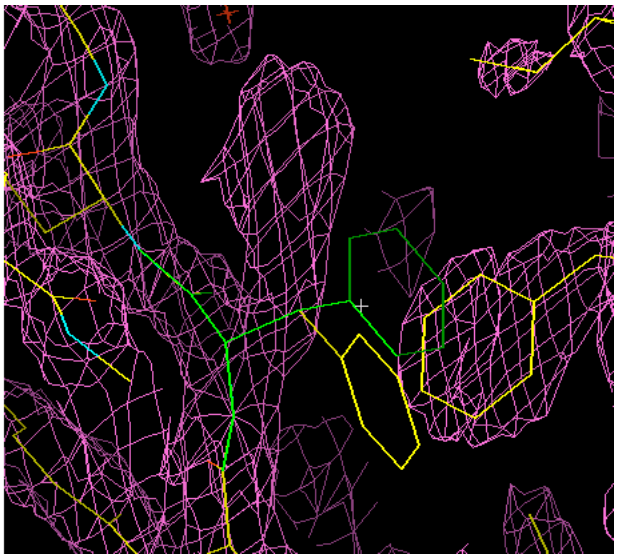
**Parameters 223 amino acids 1655 atoms in protein 212 waters**

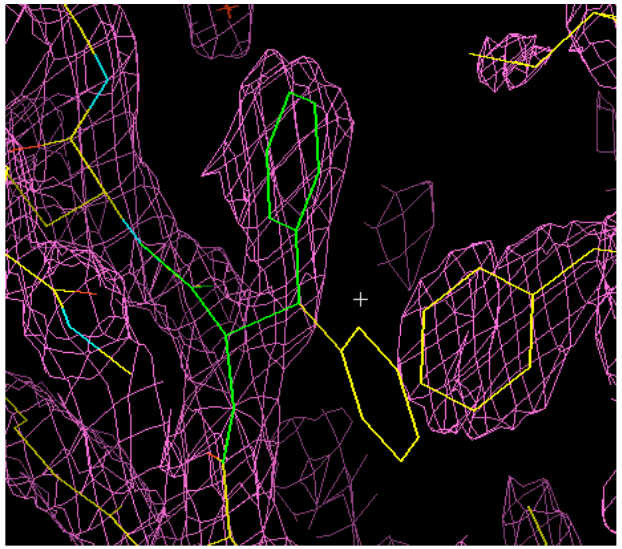
**=1867 atoms x, y, z, B x4 parameters = 7468 parameters**

| Resln. | Observations. ($F_{hkl}$) obs | Biso (4 para) | Bij (3+6 para) | Data to parameter ratio. (>10??) phi/psi B | 12 helices |
|---|---|---|---|---|---|
| 5Å | 826 Ihkl | | | 0.32 | 11.5 |
| 3.75 | 1,958 Ihkl | | | 0.76 | 27 |
| 2.5Å | 6,608 Ihkl | 0.88 | 0.4 | 2.6 | 92 |
| 2Å | 12,906 Ihkl | 1.7 | 0.76 | 5 | |
| 1.25Å | 52,864Ihkl | 7.1 | 3.2 | 21 | |

# Density fitting –with reduced parameters





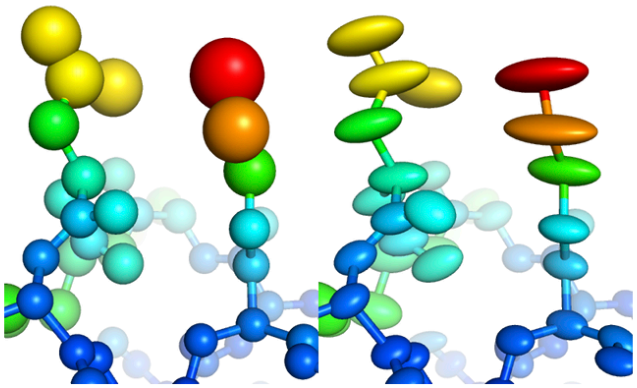


5 PTP

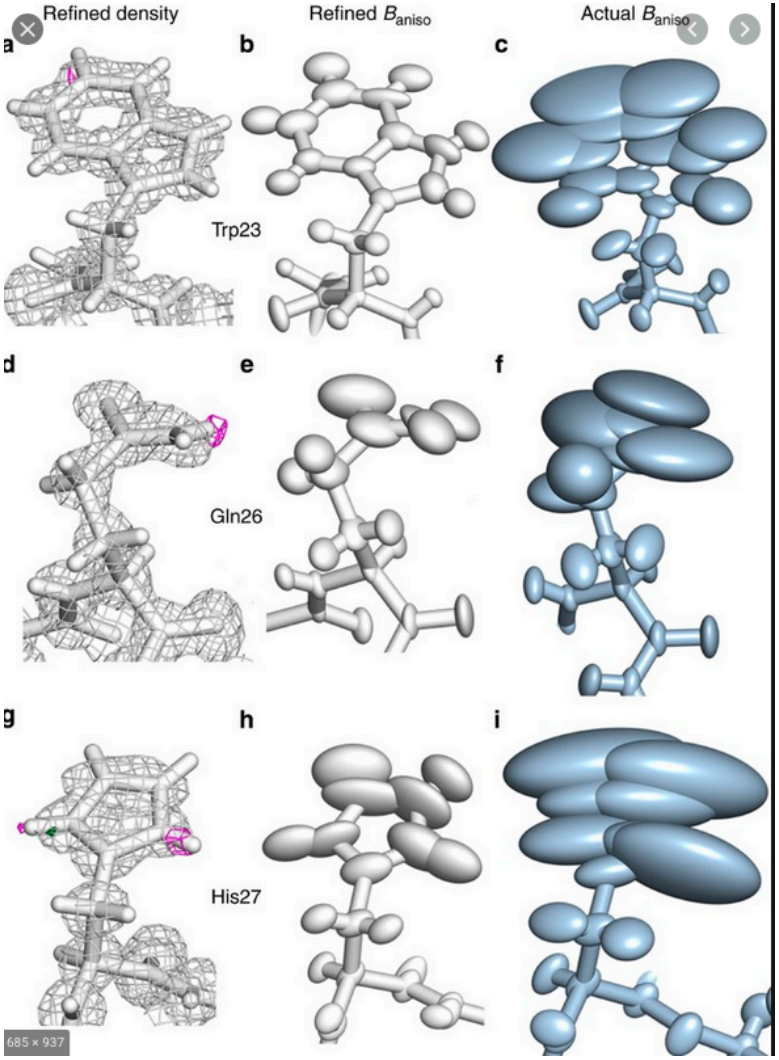
**Parameters** 223 amino acids 1655 atoms in protein 212 waters

=1867 atoms x, y, z, B x4 parameters = **7468 parameters**

| Resln. | Observations ($F_{hkl}$) obs | Data to parameter ratio Biso (4 para) | Bij (3+6 para) |
|---|---|---|---|
| 5Å | 826 Ihkl | | |
| 3.75 | 1,958 Ihkl | | |
| 2.5Å | 6,608 Ihkl | 0.88 | 0.4 |
| 2Å | 12,906 Ihkl | 1.7 | 0.76 |
| 1.25Å | 52,864Ihkl | 7.1 | 3.2 |



*Visualization of individually refined isotropic (left) versus anisotropic (right) B-factors for a high-resolution structure*



# Validation? Residual 'R' factors

- Use Current structure to calculate Amplitudes
- $F_{(h,k,l)}$calc and Phase $_{(h,k,l)}$calc
- Compare differences between Observed and Calculated Amplitudes – as a percentage

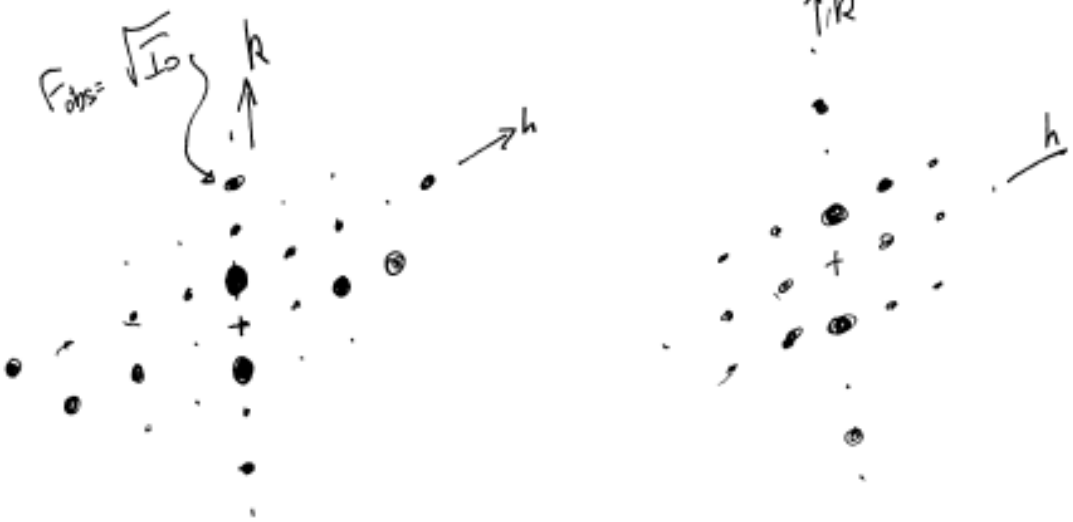
Bottom Lines: $I_{hkl} = |F_{hkl}|^2$

**Intensity = Amplitude²**

- "R factor" = Agreement between Amplitudes calculated, $F_{calc}^{hkl} = \sum_j f_j e^{2\pi i(hx+ky+lz)}$ and Amplitudes observed $\sqrt{I_{obs}}_{hkl} = F_{obs}{}_{hkl}$

$$R = \frac{\sum_{hkl} ||F_o| - |F_c||}{\sum_{hkl} |F_o|}$$

$F_{calc} = \sum_{atoms} f_i R$




judge the Quality of structure?

2. Overall agreement of observation with diffraction calculated from the interpreted structure.

3. Since we refine the structure To match the $I_{hkl}$ overfitting ?

Define $R_{free}$ for a 'hold-out ' set of observations.

4. OK? R < 20%, R free< 25%

5. But the experimental errors in measuring $F_o$ are ~ 3%. inadequate models of solvent, atom motion, anharmonicisity

6 Accuracy ~ 0.5*res*R

"R" factors

$$R = \frac{\sum_{hkl} ||Fobs| - |Fcalc||}{\sum_{hkl} |Fobs|}$$

| | |
|---|---|
| **completely random:** | **0.59** |
| **starting MR solution:** | **0.4-0.55** |
| **something still wrong?:** | **> 0.3** |
| **correct chain trace:** | **< 0.2** |
| **small molecule:** | **~ 0.05** |

# “R” (Residual) factors

**$R_{cryst}$**

**observed vs calculated data (|Fs|)**

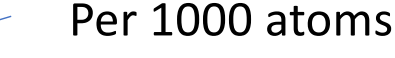
**$R_{free}$**

**cross-check with “random” subset of data**

**should be < 0.3 and < $R_{cryst}$ + 0.1**

**$R_{sym}$ = $R_{merge}$ (self-consistency of data: Intensities)**

| | EcTS |
|---|---|
| **Table 1. Data collection and refinement statistics.** Statistics for the highest-resolution shell are shown in parentheses. | |
| **Wavelength** | 0.8855 |
| **Resolution range** | 56.89 - 1.05 (1.088 - 1.05) |
| **Space group** | P 63 |
| **Unit cell** | 125.53 125.53 66.757 90 90 120 |
| **Total reflections** | 4948169 (204831) |
| **Unique reflections** | 277294 (27104) |
| **Multiplicity** | 17.8 (7.6) |
| **Completeness (%)** | 99.72 (97.52) |
| **Mean I/sigma(I)** | 24.22 (0.73) |
| **Wilson B-factor** | 10.31 |
| **R-merge** | 0.0648 (2.414) |
| **R-meas** | 0.06666 (2.592) |
| **R-pim** | 0.0154 (0.9278) |
| **CC1/2** | 1 (0.257) |
| **CC*** | 1 (0.639) |
| **Reflections used in refinement** | 277202 (27007) |
| **Reflections used for R-free** | 2007 (195) |
| **R-work** | 0.1224 (0.3145) |
| **R-free** | 0.1351 (0.2927) |
| **CC(work)** | 0.982 (0.612) |
| **CC(free)** | 0.976 (0.601) |
| **Number of non-hydrogen atoms** | 5503 |
| **macromolecules** | 4588 |
| **ligands** | 141 |
| **solvent** | 774 |
| **Protein residues** | 532 |
| **RMS(bonds)** | 0.009 |
| **RMS(angles)** | 1.09 |
| **Ramachandran favored (%)** | 98.66 |
| **Ramachandran allowed (%)** | 0.96 |
| **Ramachandran outliers (%)** | 0.38 |
| **Rotamer outliers (%)** | 0.00 |
| **Clashscore** | 2.37 |
| **Average B-factor** | 14.95 |
| **macromolecules** | 12.85 |
| **ligands** | 20.21 |
| **solvent** | 26.42 |



| EcTS | **Table 1. Data collection and refinement statistics.** Statistics for the highest-resolution shell are shown in parentheses. |
|---|---|
| **Wavelength** | 0.8855 |
| **Resolution range** | 56.89 - 1.05 (1.088 - 1.05) |
| **Space group** | P 63 |
| **Unit cell** | 125.53 125.53 66.757 90 90 120 |
| **Total reflections** | 4948169 (204831) |
| **Unique reflections** | 277294 (27104) |
| **Multiplicity** | 17.8 (7.6) |
| **Completeness (%)** | 99.72 (97.52) |
| **Mean I/sigma(I)** | 24.22 (0.73) |
| **Wilson B-factor** | 10.31 |
| **R-merge** | 0.0648 (2.414) |
| **R-meas** | 0.06666 (2.592) |
| **R-pim** | 0.0154 (0.9278) |
| **CC1/2** | 1 (0.257) |
| **CC*** | 1 (0.639) |
| **Reflections used in refinement** | 277202 (27007) |
| **Reflections used for R-free** | 2007 (195) |
| **R-work** | 0.1224 (0.3145) |
| **R-free** | 0.1351 (0.2927) |
| **CC(work)** | 0.982 (0.612) |
| **CC(free)** | 0.976 (0.601) |
| **Number of non-hydrogen atoms** | 5503 |
| **macromolecules** | 4588 |
| **ligands** | 141 |
| **solvent** | 774 |
| **Protein residues** | 532 |
| **RMS(bonds)** | 0.009 |
| **RMS(angles)** | 1.09 |
| **Ramachandran favored (%)** | 98.66 |
| **Ramachandran allowed (%)** | 0.96 |
| **Ramachandran outliers (%)** | 0.38 |
| **Rotamer outliers (%)** | 0.00 |
| **Clashscore** | 2.37 |
| **Average B-factor** | 14.95 |
| **macromolecules** | 12.85 |
| **ligands** | 20.21 |
| **solvent** | 26.42 |

Overlaps>0.4Å
Per 1000 atoms

# Table S1. Cryo-EM data collection, refinement and validation statistics

| | rVGLUT2/(EMDB-21040)/(PDB 6V4D) |
|---|---|
| **Data collection and processing** | |
| Microscope | Talos Arctica |
| Camera | Gatan K3 direct detector |
| Magnification | 36,000 on scope/ 50,000 calibrated at detector |
| Voltage (kV) | 200 |
| Electron exposure ($e^{-}$/Å$^2$) | 50 to 60 |
| Exposure rate | ~0.5 $e^{-}$/frame |
| Number of frames per micrograph | 120 |
| Automation software | SerialEM |
| Underfocus range (μm) | 1.0 to 2.5 |
| Pixel size (Å) | 1.14 |
| Symmetry imposed | C1 |
| Number of micrographs used | 5704 |
| Initial particle images (no.) | 1,919,729 |
| Refined particle images (no.) | 431,655 |
| Final particle images (no.) | 243,615 |
| Map resolution (Å) (masked/unmasked)<br>FSC threshold | 3.8/3.9<br>0.143 |
| Map resolution range (Å) | 3.8-50 |
| Estimated accuracy of translations | 1.11 Å |
| Estimated accuracy of rotations | 3.41º |
| | |
| **Refinement** | |
| Refinement package | Phenix real space refinement |
| Resolution cut-off | 4.0 |
| Initial model used (PDB code) | None |
| Model resolution (Å)<br>FSC threshold | 4.0<br>0.5 |
| Model resolution range (Å) | 3.1-4.0 |
| Map sharpening B factor (Å$^2$) | -146 |
| Model vs map CC | 0.76 |
| Model composition<br>Non-hydrogen atoms<br>Protein residues<br>Ligands | <br>3062<br>417 |
| B factors (Å$^2$)<br>Protein<br>Ligand | <br>175 |
| R.m.s. deviations<br>Bond lengths (Å)<br>Bond angles (°) | <br>0.007 (0)<br>0.781 (0) |
| Validation<br>MolProbity score<br>Clashscore<br>Poor rotamers (%) | <br>2.22<br>18.51<br>0.34 |
| CaBLAM outliers (%) | 3.21 |
| Cβ deviation | 0 |
| EMRinger score | 1.90 |
| Ramachandran plot<br>Favored (%)<br>Allowed (%)<br>Disallowed (%) | <br>92.94<br>7.06<br>0.00 |

## Extras: Optical diffraction shows

Scattering from a spherical ‘atom’

Scattering from a molecule. = pattern

Scattering from a lattice of points = Diffraction

Scattering from a lattice of molecules = scattering from a molecule, sampled so visible only at the diffraction

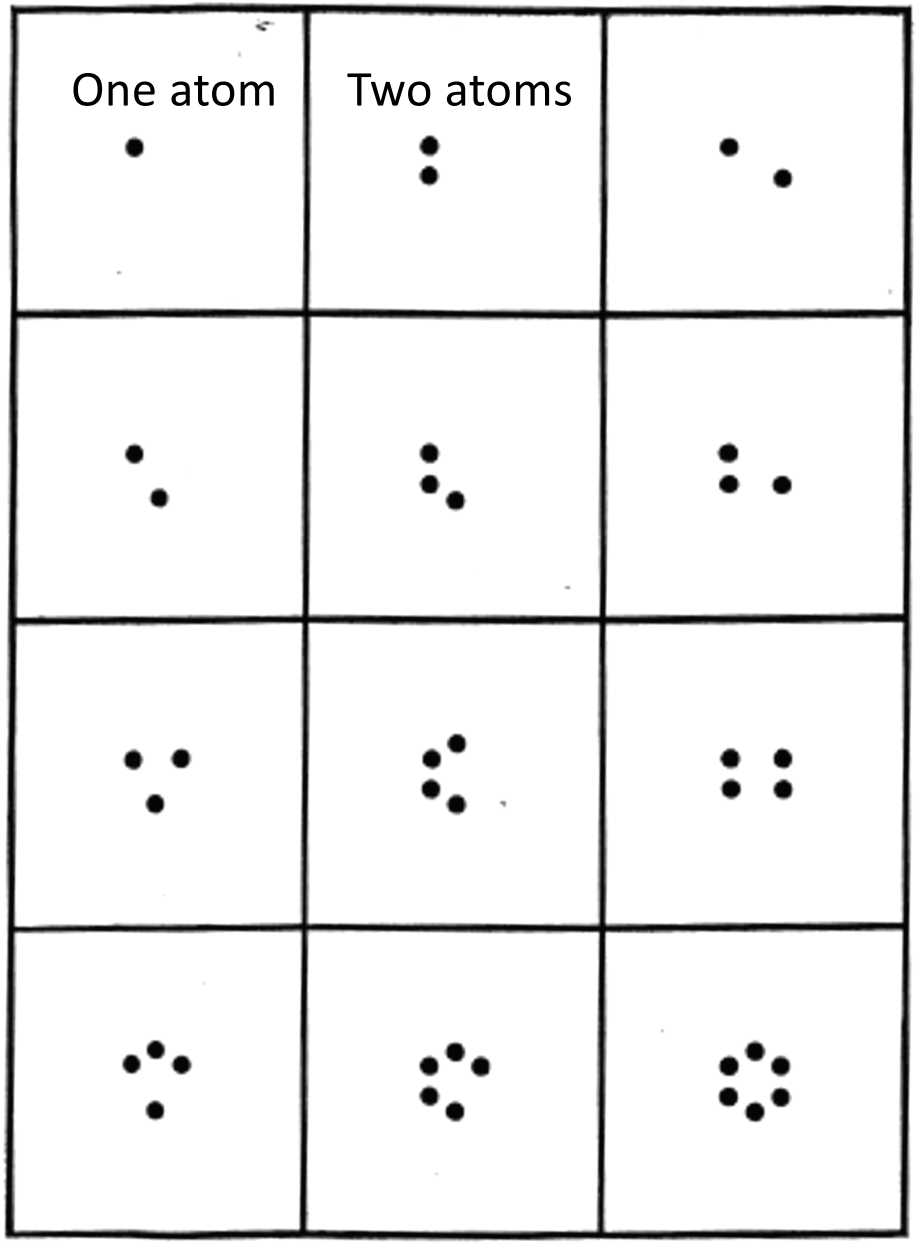
Plate 2 Superimposed fringes

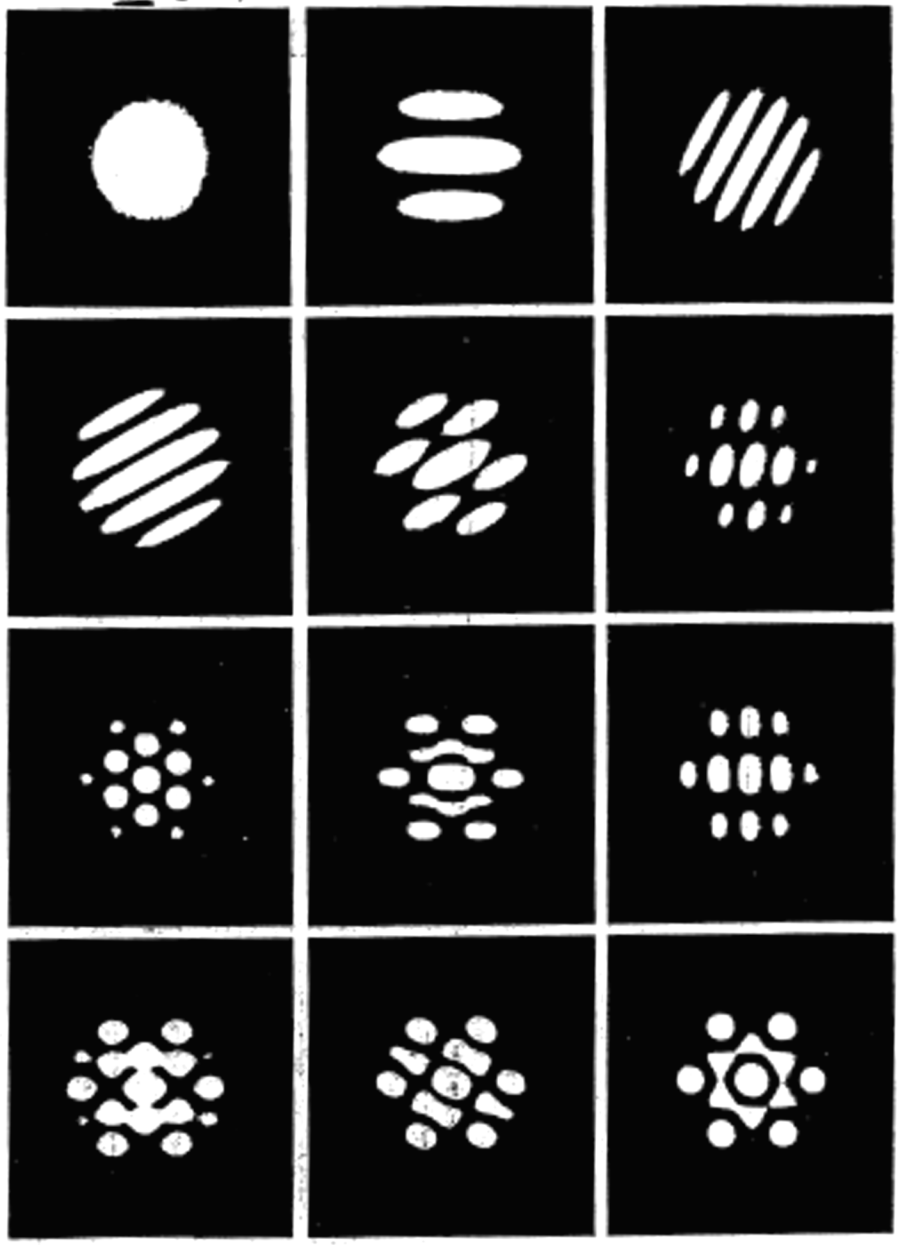
The sequence shows how the diffraction pattern of a simple object is built up by superposition of sets of fringes.

Plate 2

$\rho(r)$

$F(s)$



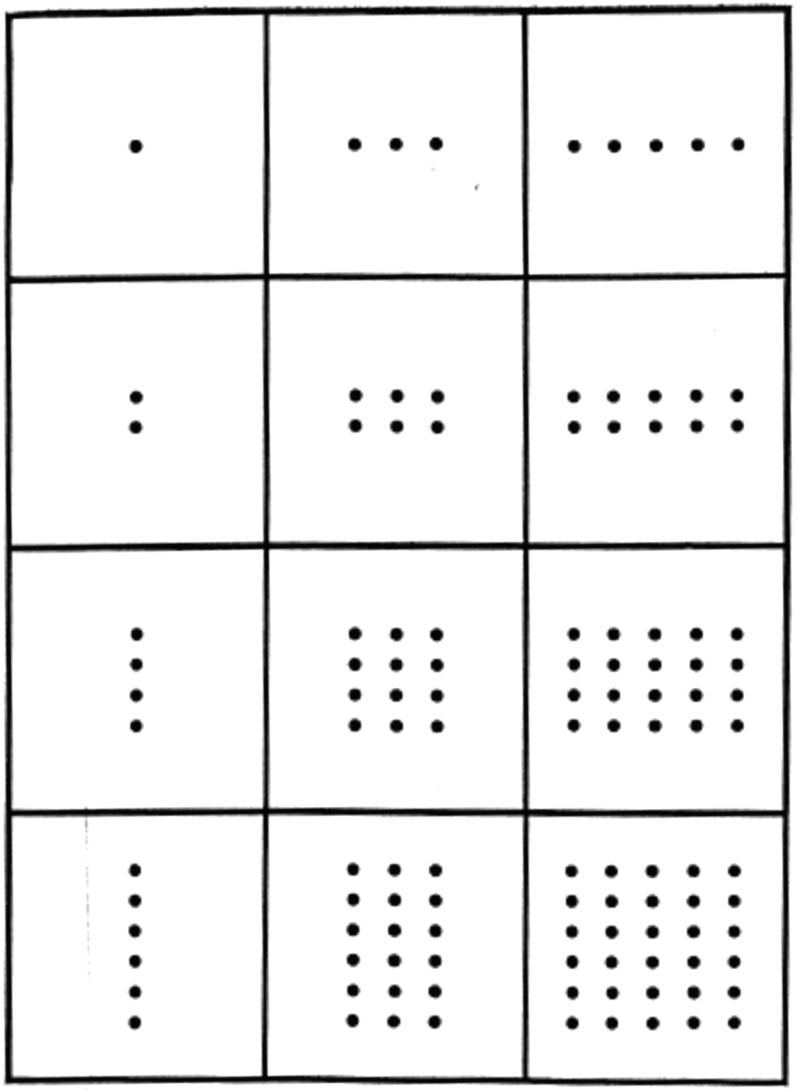



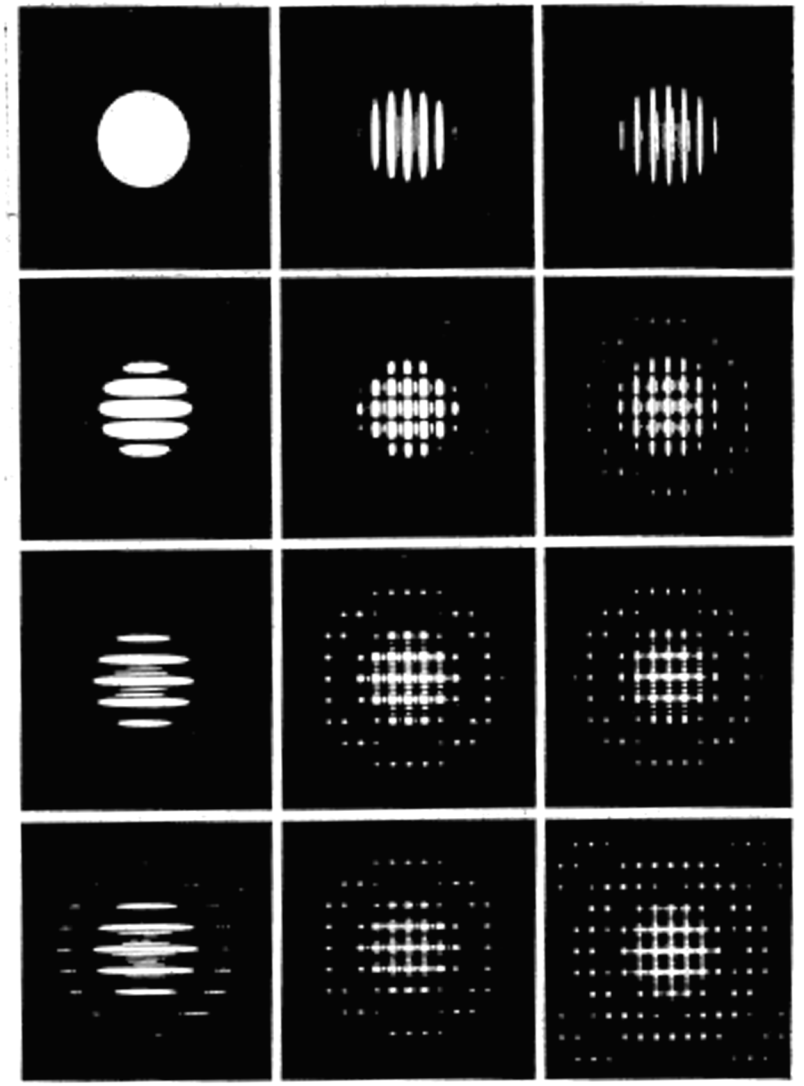
# Diffraction due to repeats, is the same as the object, sampled by (1/repeat)

Plate 10 Development of a lattice

A single aperture is repeated to build up rows and columns which are combined to produce the lattice of 12. Subsidiary diffraction maxima due to the small number of apertures can be seen clearly.

Plate 10

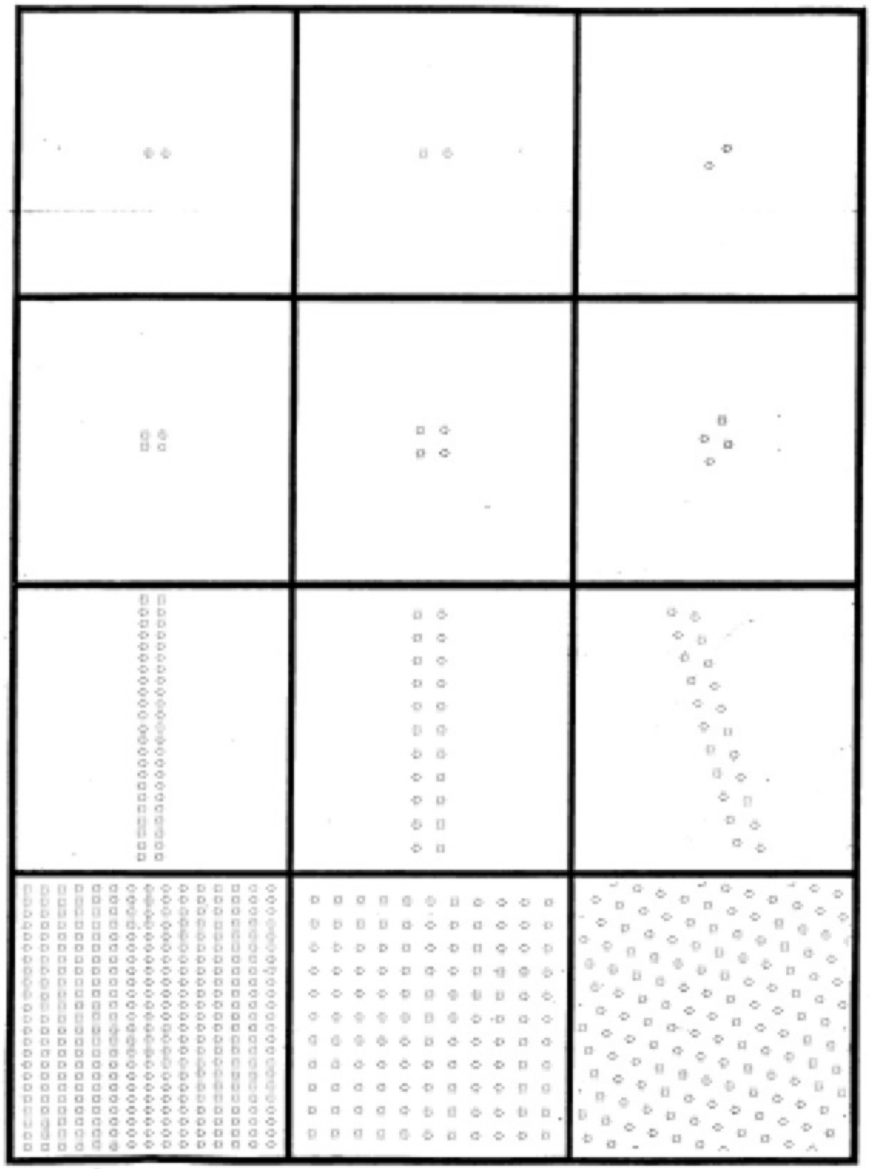


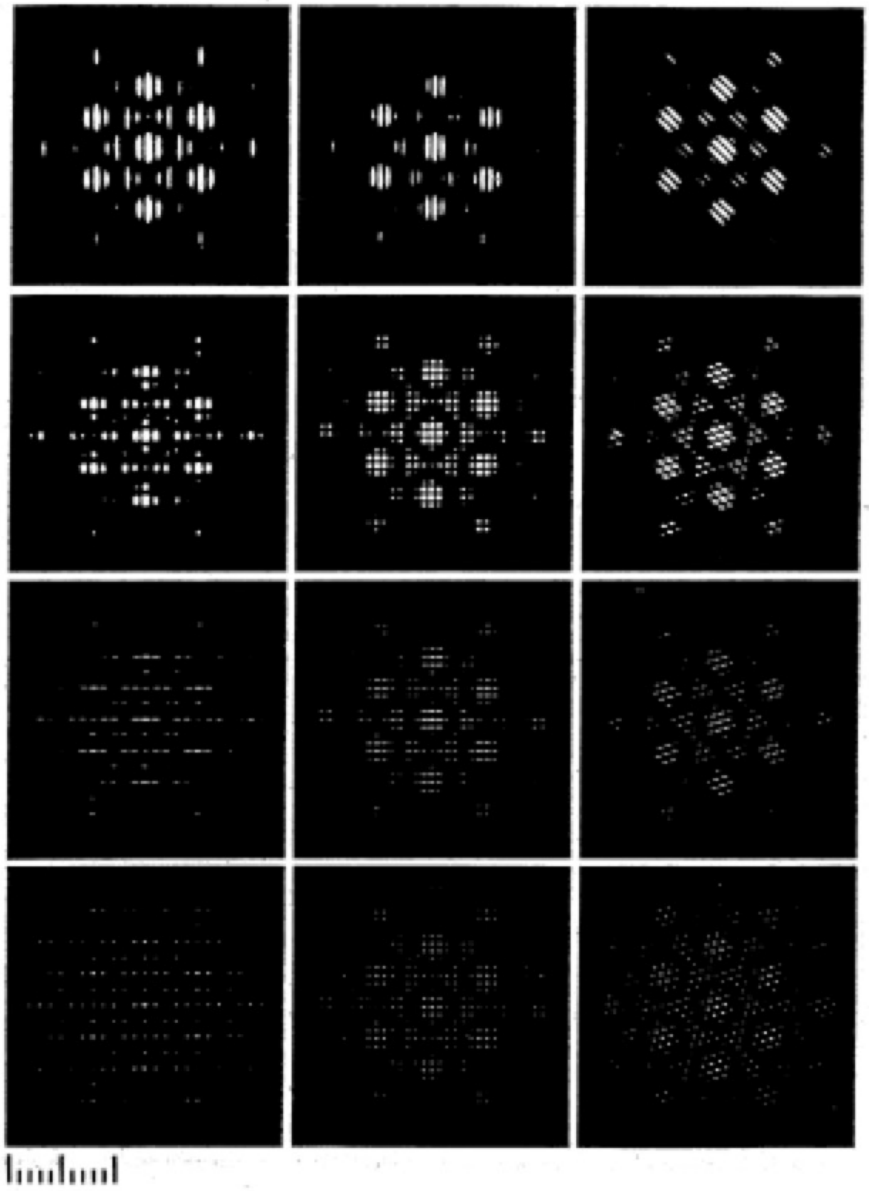


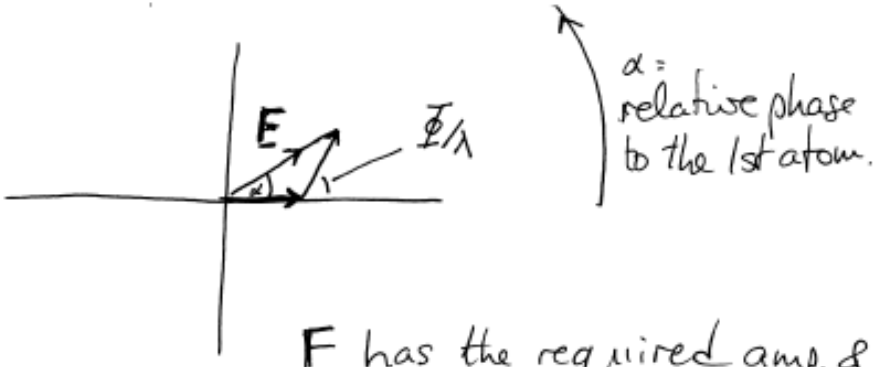
Build a crystal

Object

Scattering

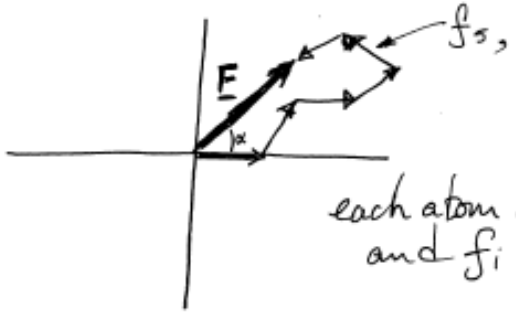







$\underline{F}$ has the required amp. & phase.

If we use this method we can add $i = 1$ to $n$ different atoms; each amplitude $f_i$




each atom has $f_i \cos\alpha_i$ along $x$ and $f_i \sin\alpha_i$ along $y$

If we put 'units' on the axes, we can add up the 'x' and 'y' components to write the sum over "x", the sum over "y", — hence calculate $\underline{F}$ as a wave of amplitude

$$|\underline{F}| = \sqrt{\left(\sum_x f_i \cos\alpha_i\right)^2 + \left(\sum_y f_i \sin\alpha_i\right)^2}$$

and $$\alpha = \tan^{-1}\left(\frac{\sum_y f_i \sin\alpha_i}{\sum_x f_i \cos\alpha_i}\right)$$

Many atoms add by the same rules.

Different in every direction.