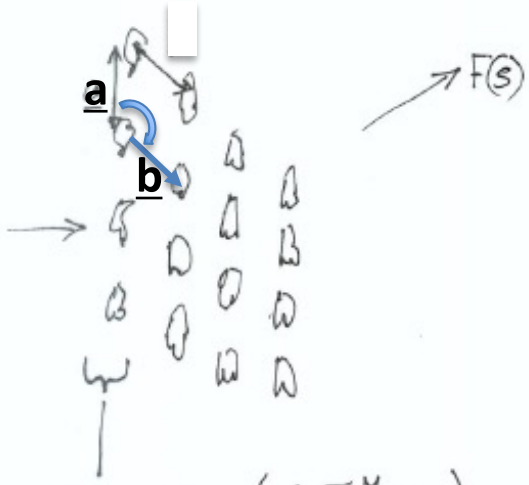
add a second repeat axis b



$F(s)_{tot}$: $F(s)\left(\frac{\sin \pi M \underline{a}.\underline{s}}{\sin \pi \underline{a}.\underline{s}}\right)$

Next column, is just $\underline{b}$ away

$= F(s)\left(\frac{\sin \pi M \underline{a}.\underline{s}}{\sin \pi \underline{a}.\underline{s}}\right) e^{2\pi i \underline{b}.\underline{s}}$

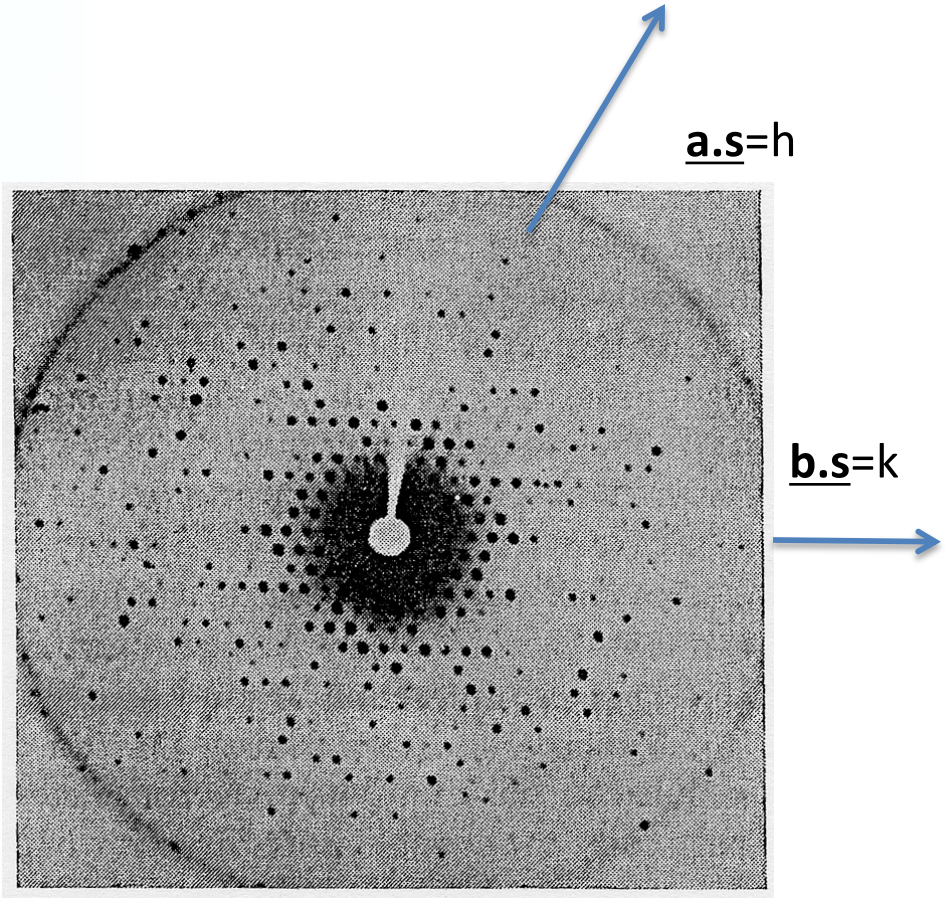
etc.

So for the 2D array:-

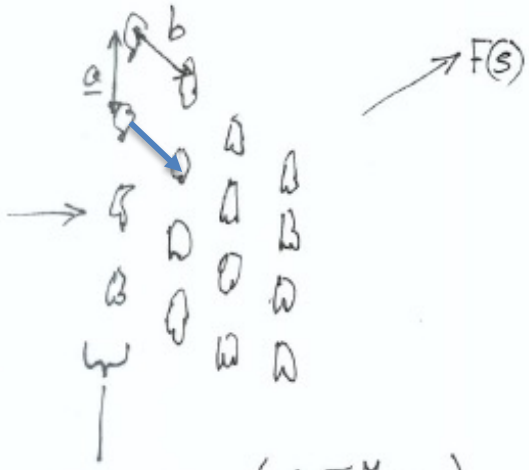
$= F(s)\left(\frac{\sin \pi M a.s}{\sin \pi a.s}\right)\left(\frac{\sin \pi N \underline{b}.\underline{s}}{\sin \pi \underline{b}.\underline{s}}\right)$

for 3D array

$= F(s) \frac{\sin \pi M \underline{a}.\underline{s}}{\sin \pi \underline{a}.\underline{s}} \frac{\sin \pi N \underline{b}.\underline{s}}{\sin \pi \underline{b}.\underline{s}} \frac{\sin \pi P \underline{s}.\underline{s}}{\sin \pi \underline{c}.\underline{s}}$



add a second repeat axis b



$F(s)_{tot}$: $F(s)\left(\frac{\sin \pi M \underline{a}.\underline{s}}{\sin \pi \underline{a}.\underline{s}}\right)$

Next column, is just $\underline{b}$ away

$= F(s)\left(\frac{\sin \pi M \underline{a}.\underline{s}}{\sin \pi \underline{a}.\underline{s}}\right) e^{2\pi i \underline{b}.\underline{s}}$ etc.

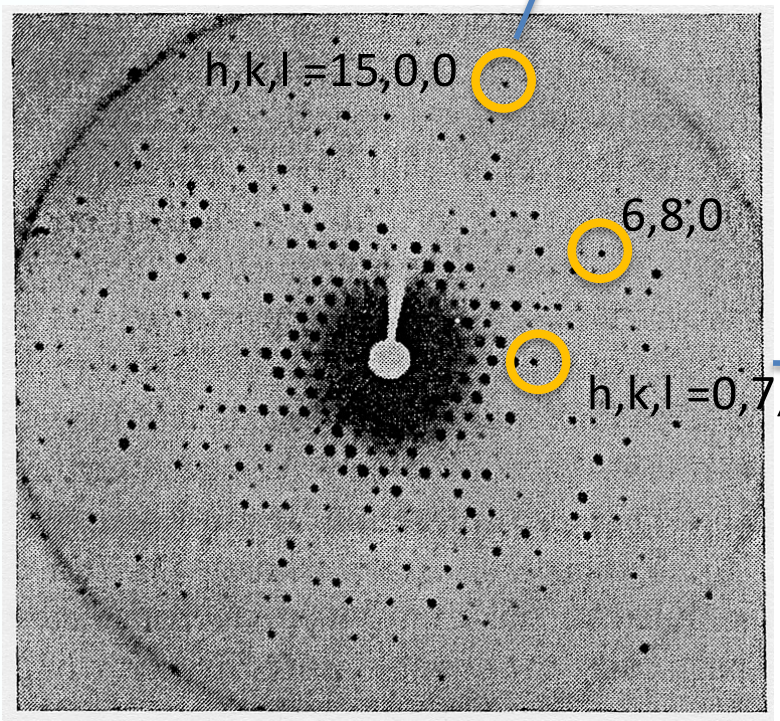
So for the 2D array:-

$= F(s)\left(\frac{\sin \pi M a.s}{\sin \pi a.s}\right)\left(\frac{\sin \pi N \underline{b}.\underline{s}}{\sin \pi \underline{b}.\underline{s}}\right)$

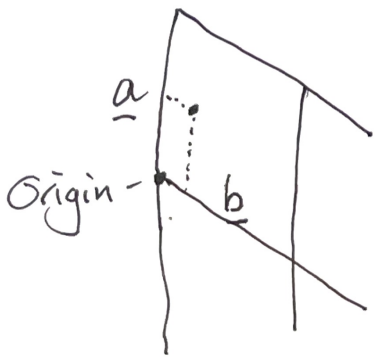
for 3D array

$= F(s) \frac{\sin \pi M \underline{a}.\underline{s}}{\sin \pi \underline{a}.\underline{s}} \frac{\sin \pi N \underline{b}.\underline{s}}{\sin \pi \underline{b}.\underline{s}} \frac{\sin \pi P \underline{s}.\underline{s}}{\sin \pi \underline{c}.\underline{s}}$

a.s=2
a.s=1
a.s=0



Positions can be described in a "unit cell"



atom j

$$\underline{r} = x_j \underline{a} + y_j \underline{b} + z_j \underline{c}$$

X-ray "reflections" only when

$\underline{a} \cdot \underline{S} = h$ ≡≡≡ (perpendicular to $\underline{a}$)

$\underline{b} \cdot \underline{S} = k$

$\underline{c} \cdot \underline{S} = l$

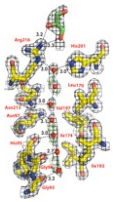
Thus

$$\underline{r}_j \cdot \underline{S} = (h x_j + k y_j + l z_j)$$

so $$\underline{F}(\underline{S}) = \sum_j f_j \, e^{2\pi i (h x_j + k y_j + l z_j)}$$

and $$\rho(\underline{r}) = \rho(x,y,z) = \sum_{hkl} |F| e^{2\pi i \varphi_{hkl}} \cdot e^{2\pi i (hx + ky + lz)}$$

This is all there is? YES!!



FT

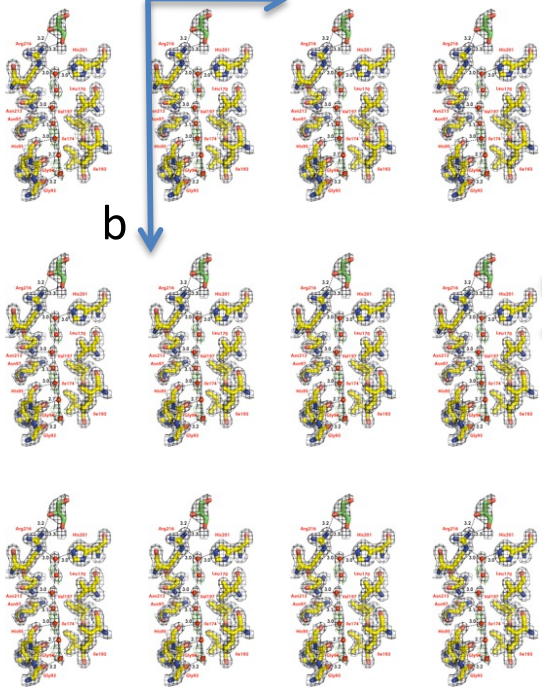
FT^{-1}

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

$$\underline{F}(\underline{s}) = \Sigma_j f_j e^{(2\pi i \underline{r}_j . \underline{s})}$$

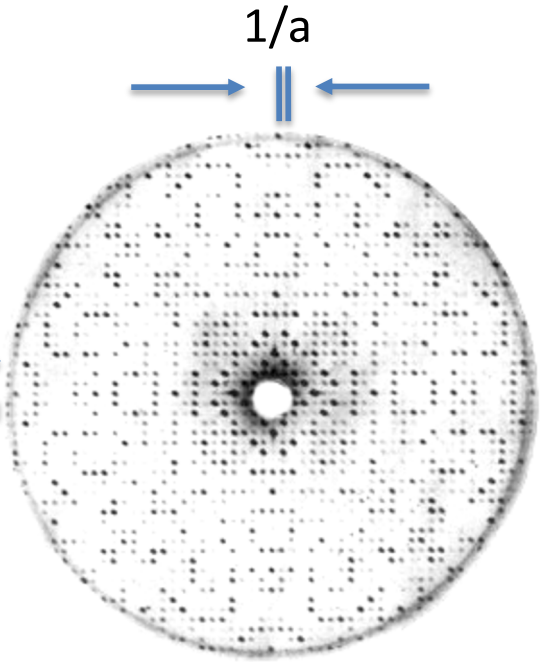
Observe I($\underline{S}$) = $\underline{F}(\underline{s}).\underline{F}^*(\underline{s})$

Structure is the 'inverse' Fourier transform of the Scattering pattern $\underline{F}(\underline{s})$



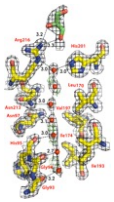
FT

FT^{-1}



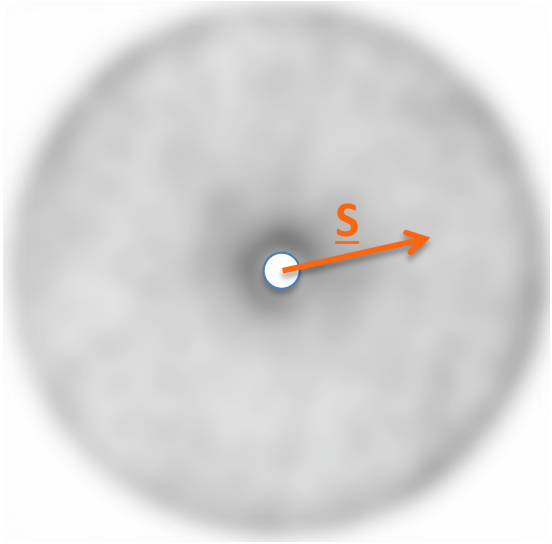
$$\rho(\underline{r}) = \Sigma_s \underline{F}(\underline{s}) e^{(-2\pi i \underline{r}.\underline{s})}$$

This is all there is?



FT

FT^{-1}

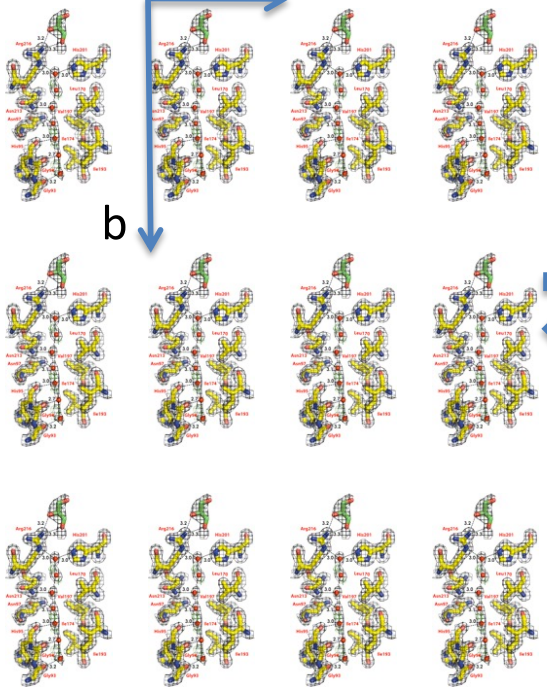


Scattering pattern is the Fourier transform of the structure

$$\underline{F}(\underline{S}) = \Sigma_j f_j e^{(2\pi i \underline{r}_j.\underline{s})}$$

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

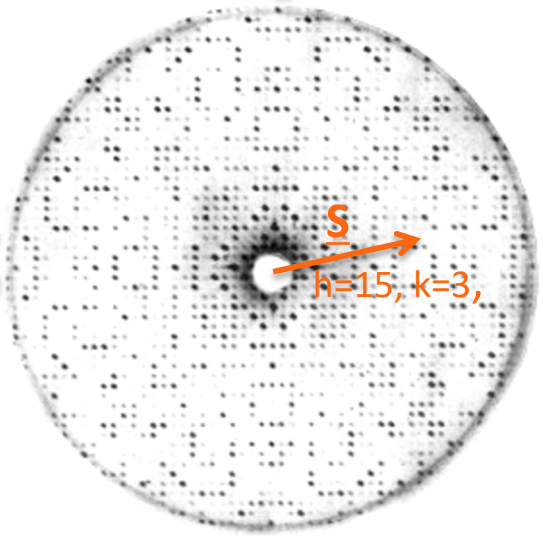
$$\rho(\underline{r}) = \Sigma_s \underline{F}(\underline{s}) e^{(-2\pi i \underline{r}.\underline{s})}$$



FT

FT^{-1}

1/a



1/b

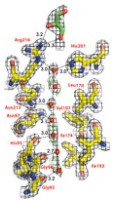
$$\underline{F}(h,k,l) = \Sigma_j f_j e^{(2\pi i (hx_j+ky_j+lz_j))}$$

$$\rho(x,y,z) = \Sigma_{h,k,l} \underline{F}_{(\underline{h,k,l})} e^{(-2\pi i (hx+ky+lz))}$$

This is all there is?

PHASES-as colors !

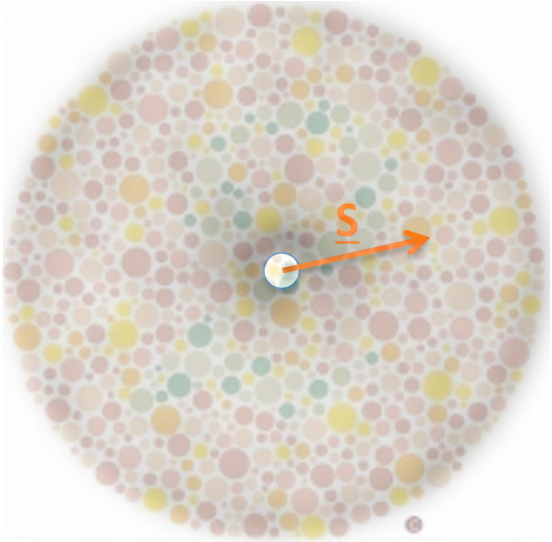
$\rho(\underline{r})$



FT

$\Leftrightarrow$

FT^{-1}

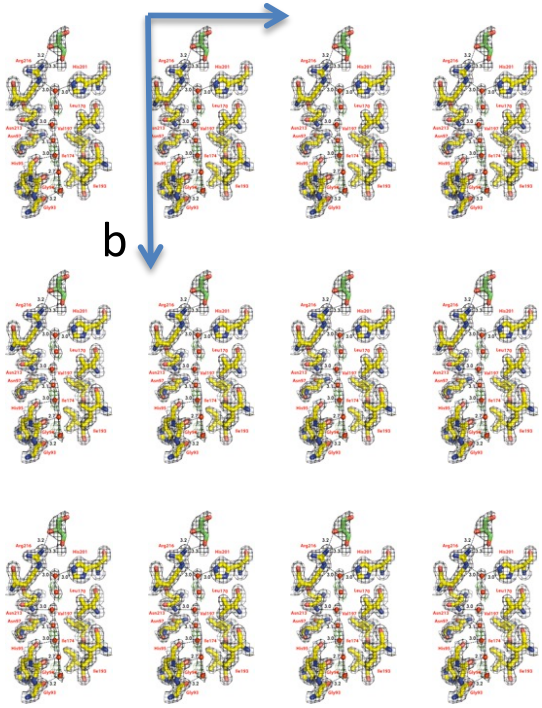


Scattering pattern is the Fourier transform of the structure

$$\underline{F}(\underline{S}) = \Sigma_j f_j e^{(2\pi i \underline{r}_j . \underline{s})}$$

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

$\rho(\underline{r})$

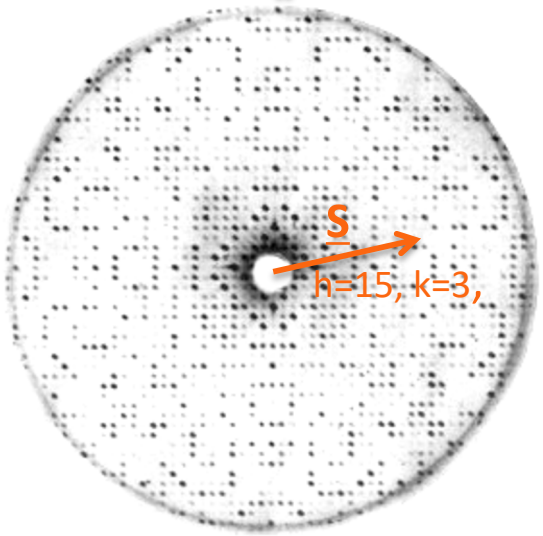


FT

$\Leftrightarrow$

FT^{-1}

1/a



1/b

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

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

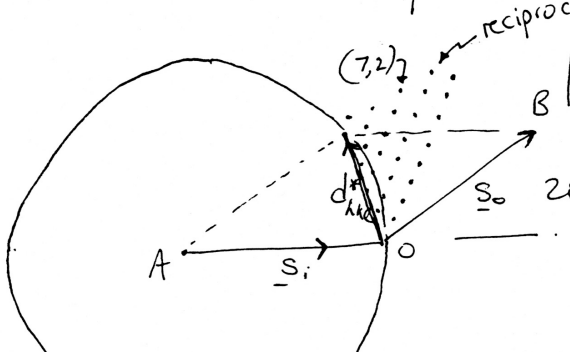
$$\rho(x,y,z) = \Sigma_{h,k,l} \underline{F}(h,k,l) e^{(-2\pi i \mathbf{r}.\mathbf{S})}$$

©Robert M. Stroud 2022

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~~ ends in the Ewald sphere.

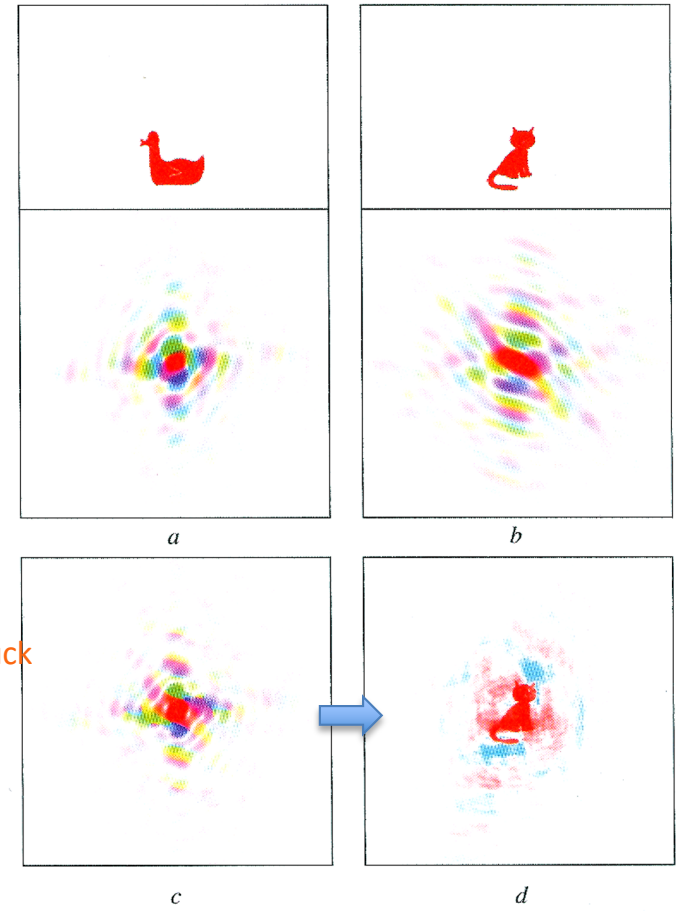


$$I_{hkl} \propto |F_{hkl}|^2 = G(\underline{S})\,G^*(\underline{S})$$

Relative Information in Intensities versus phases

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

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



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

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

$\underline{F}_{(\underline{S})} = \Sigma_r \rho_{(r)} e^{(2\pi i \underline{r}.\underline{S})}$

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

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

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

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.

©Robert M. Stroud 2022

Relative Information in Intensities versus phases

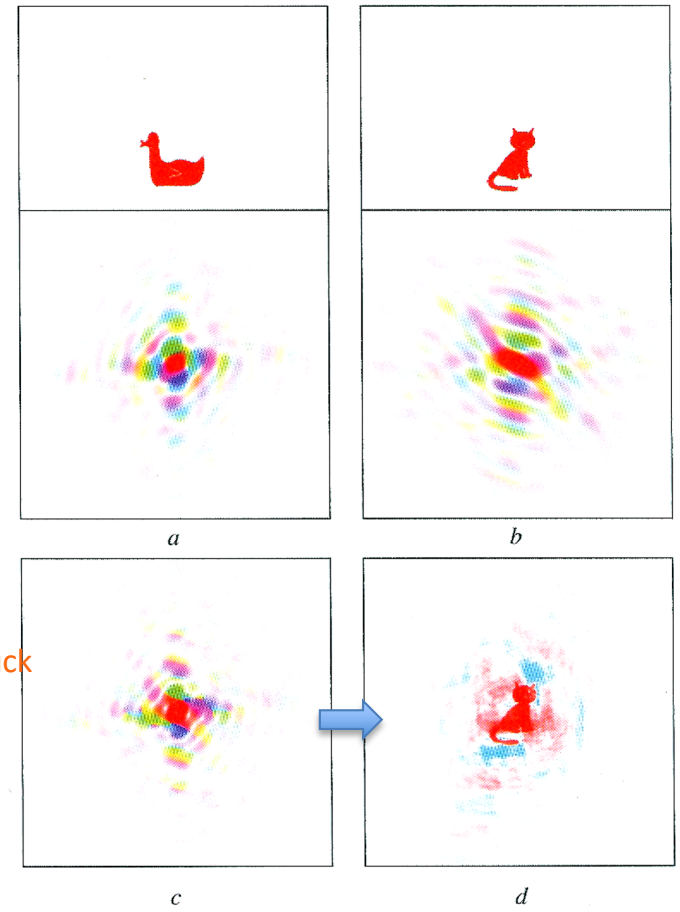
$\rho(\underline{r})$ duck

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

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

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

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



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

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

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

©Robert M. Stroud 2022

2. Molecular Replacement

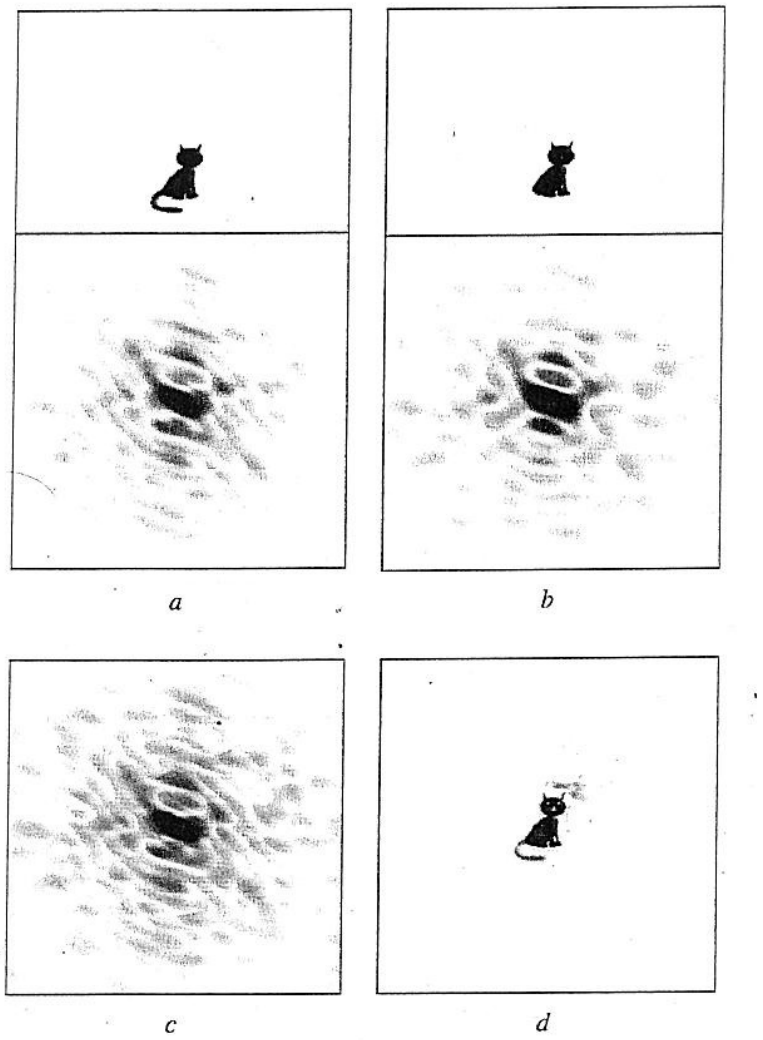


Figure 6.17 ► Structure determination by molecular replacement. (*a*) Unknown structure, cat, and its diffraction pattern (not colored, because phases are unknown). (*b*) Known structure and phasing model, Manx cat, and transform computed from the model (colored, because calculation of transform from a model tells us phases). (*c*) Manx-cat phases combined with unknown-cat intensities. (*d*) Back-transform of (*c*). Intensities contain enough information to reveal differences (the tail) between phasing model and unknown structure.

Molecular Replacement

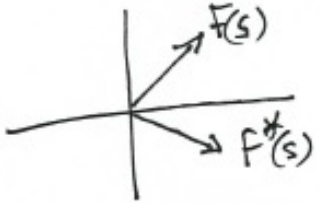
$$\underline{F}(\underline{s}) = \sum f_i \, e^{2\pi i \, \underline{r}_i \cdot \underline{s}}$$

What happens if we transform the observed intensities $\underline{I}(s)$?

ie $$P(\underline{r}) = \sum_{\underline{s}} \underline{I}(\underline{s}) \, e^{2\pi i \, \underline{r} \cdot \underline{s}} = \sum_{hkl} I_{hkl} \, e^{2\pi i (hx + ky + lz)}$$

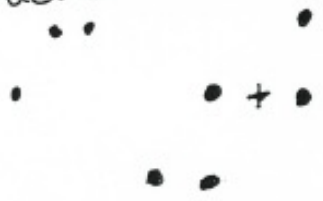
$$P(\underline{r}) = \left(\underline{F}(\underline{s}) \times \underline{F}^*(\underline{s}) \right) e^{2\pi i \, \underline{r} \cdot \underline{s}}$$

$$= \underline{I}(\underline{s}) \, e^{2\pi i \, \underline{r} \cdot \underline{s}}$$

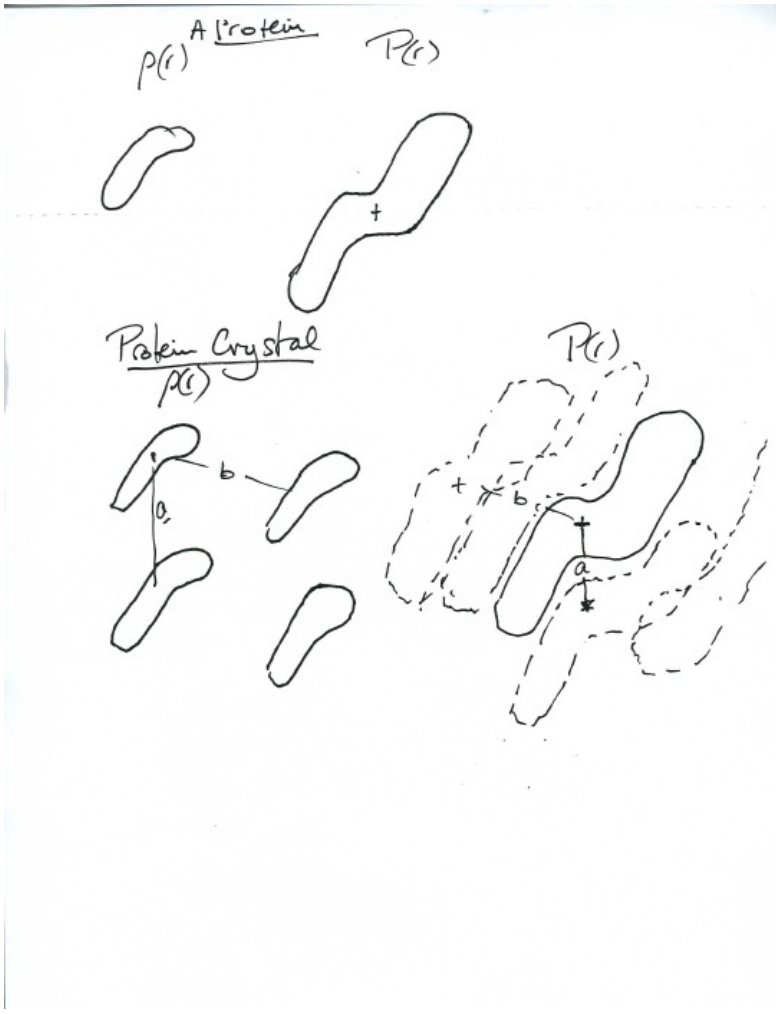


$P(r)$ = Patterson function
= "All vectors in the crystal, weighted by electron, brought to a common origin"

$\rho(r)$ Molecule



$$P(r) = \rho(r) \otimes \rho(-r)$$

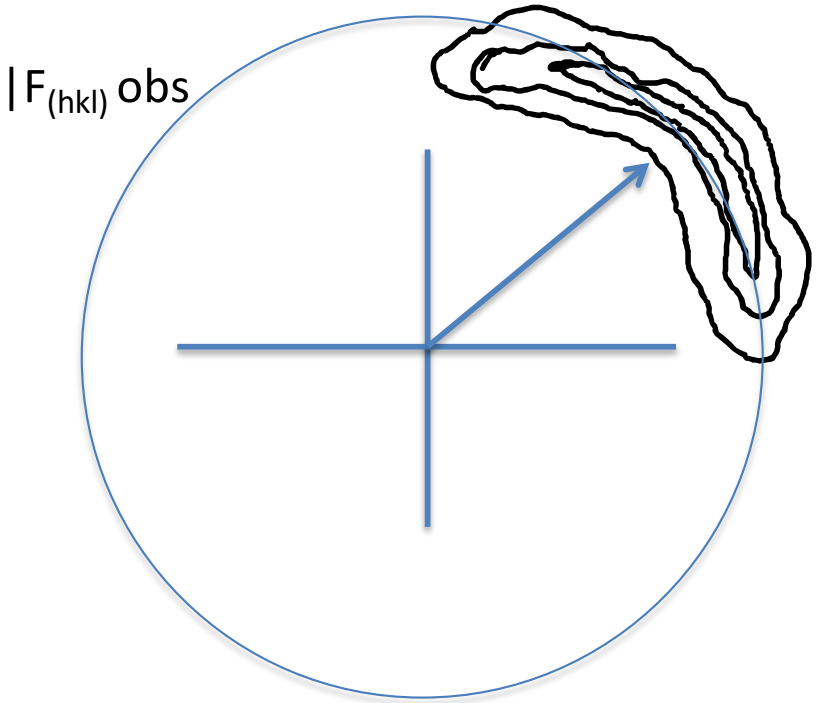


A protein Molecule

Protein Crystal

Molecular Replacement:
Search P(r) (calculated from observed I(s) only) with a P(r) calculated for a similar molecule
⇒ find orientation
⇒ search all translations.

Phase determination by any means, ends up as a probabilty distribution. So $F_{h,k,l}$, $\phi_{(h,k,l)}$



Then what to use for the best map?

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

the signal towards some F true will be

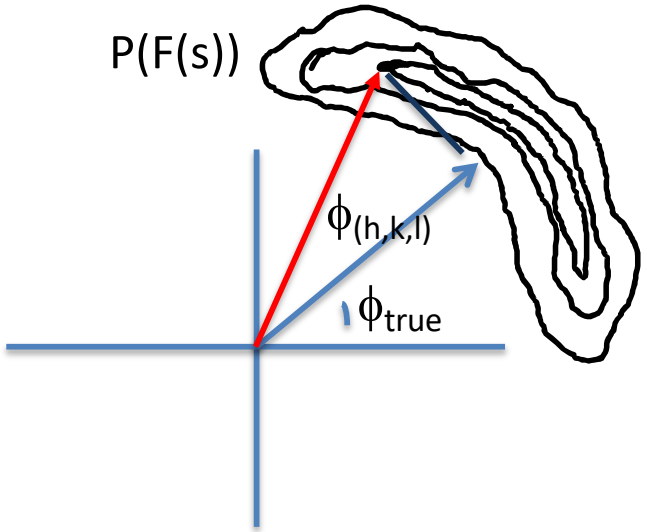
$\int_{\phi} P(F(S)) \cos(\phi_{(h,k,l)} - \phi_{true})$

and the 'noise ' will be

$\int_{\phi} P(F(S)) \sin(\phi_{(h,k,l)} - \phi_{true})$

The map with the least noise will have F(s) = center of mass of P(F(S))

Phase determination by any means, ends up as a probabilty distribution. So $F_{h,k,l}$, $\phi_{(h,k,l)}$



Then what to use for the best map?

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

the signal towards some F true will be

$\int_{\phi} P(|F(\underline{s})|\, \cos\, (\phi_{(h,k,l)} - \phi_{true})$

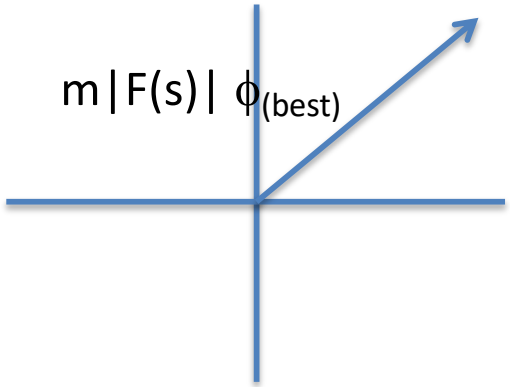
and the 'noise ' will be

$\int_{\phi} P(|F(\underline{s}))\, \sin\, (\phi_{(h,k,l)} - \phi_{true})$

The map with the least noise will have F(s) = center of mass of P(F(S))

Figure of merit weights to 'minimum error'

Phase determination by any means, ends up as a probabilty distribution. So $F_{h,k,l}$, $\phi_{(h,k,l)}$



Then what to use for the best map?

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

The map with the least noise will have

$F(s)$ = center of mass of $P(F(S)) = \int_\phi P_\phi F \sin(\phi - \phi_{(best)})$ is a minimum. Then $m = \int_\phi P_\phi F \cos(\phi - \phi_{(best)})/F$

Signal = $m|F(s)|\ \phi_{(best)} = \int_\phi P.F$

where m = figure of merit = $\int_\phi P_\phi(F)\, F(s)$

$m = \langle \cos \Delta\phi \rangle$

noise = $\int_\phi F(s) \sin \Delta\phi$

If a map is produced with some $\phi_{(hkl)}$
The probability of it being correct is $\Pi_{(hkl)} P_{(hkl)}(\phi_{(hkl)})$

Maximum value of $P_{(hkl)}(\phi_{(hkl)})$ gives the 'Most probable' map

Map with the least mean square error, is when noise is minimum,
Int find $\phi_{(best)}$ such that
$Q = \int_{\phi} [|F| P_{(hkl)}(\phi_{(hkl)}) \exp(i\phi_{(hkl)}) - \mathbf{F}_{best}\phi_{(best)})]^2 d\phi$ is minimum.

is minimum when $dQ/d\mathbf{F}_{best} = 0$
so $\mathbf{F}_{best}\phi_{(best)} = \int_{\phi} |F| P_{(hkl)}(\phi_{(hkl)}) \exp(i\phi_{(hkl)})\, d\phi$

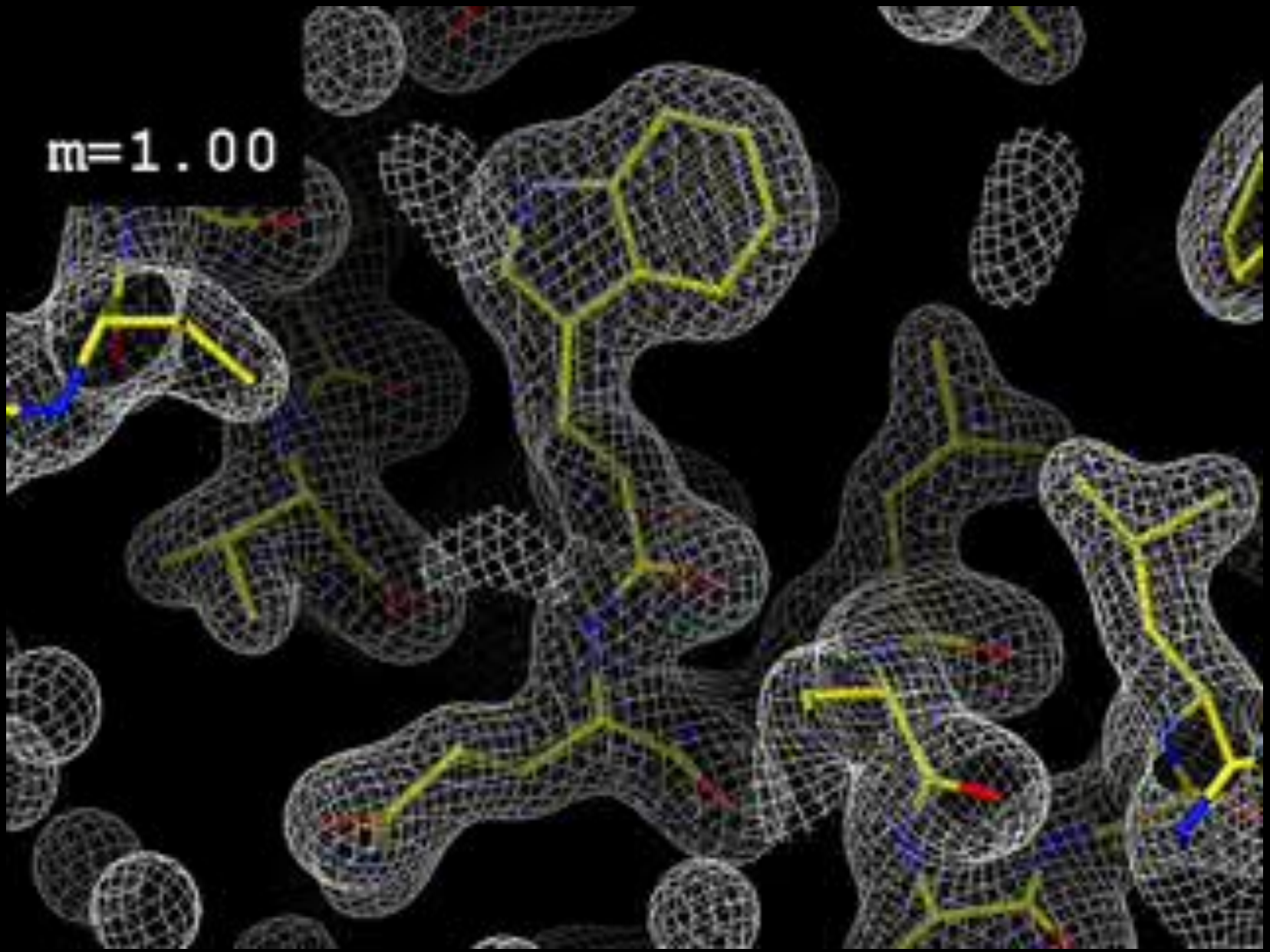
$\mathbf{F}_{best}\phi_{(best)}$ = m|F| center of 'mass' of the Probability distribution

where $m = \int_{\phi} P_{(hkl)}(\phi_{(hkl)}) \cos(\phi - \phi_{(best)})$
consider rms errors from one reflection, and its complex conjugate
$\langle(\Delta\rho)^2\rangle = 2/V^2 \int_{\phi} P_{(hkl)}(\phi_{(hkl)}) (\sin(\phi - \phi_{(best)}))^2$

Then $|F|_{best} = \int_{\phi} F \cos(\phi - \phi_{(best)}) / F$
Noise $\langle(\Delta\rho)^2\rangle = 2/V^2 \int_{\phi} F (\sin(\phi - \phi_{(best)}))^2/F = F(1-m^2)$

mF where $m = \int_{\phi} F \cos(\phi - \phi_{(best)})$

Figure of Merit



http://bl831.als.lbl.gov/~jamesh/movies/dephase.mpeg

‘Difference maps’

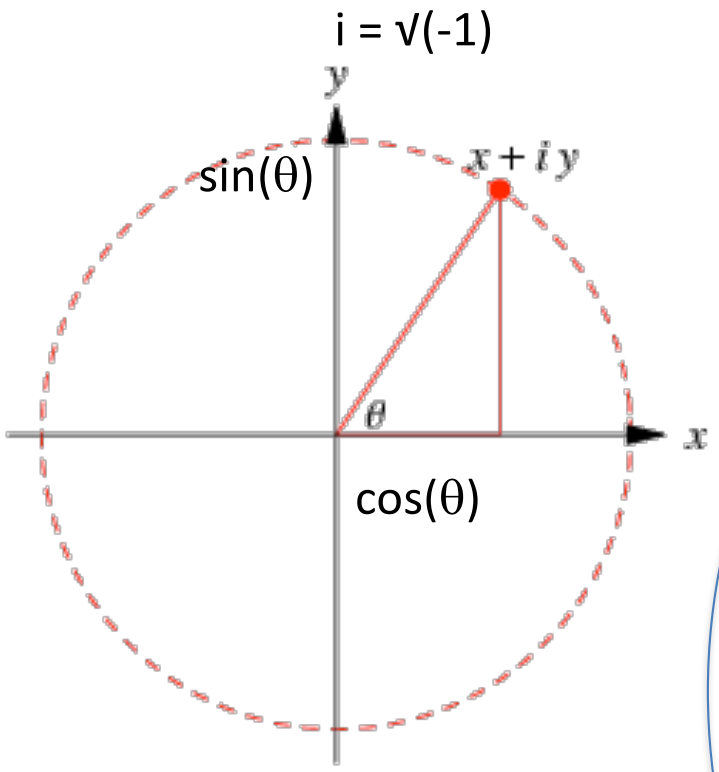
-Define bound ligands
-to find any missing atoms during refinement,
-to find ligands
-define movements of protein or water
-determine ion positions
-determine changes in dynamic motion

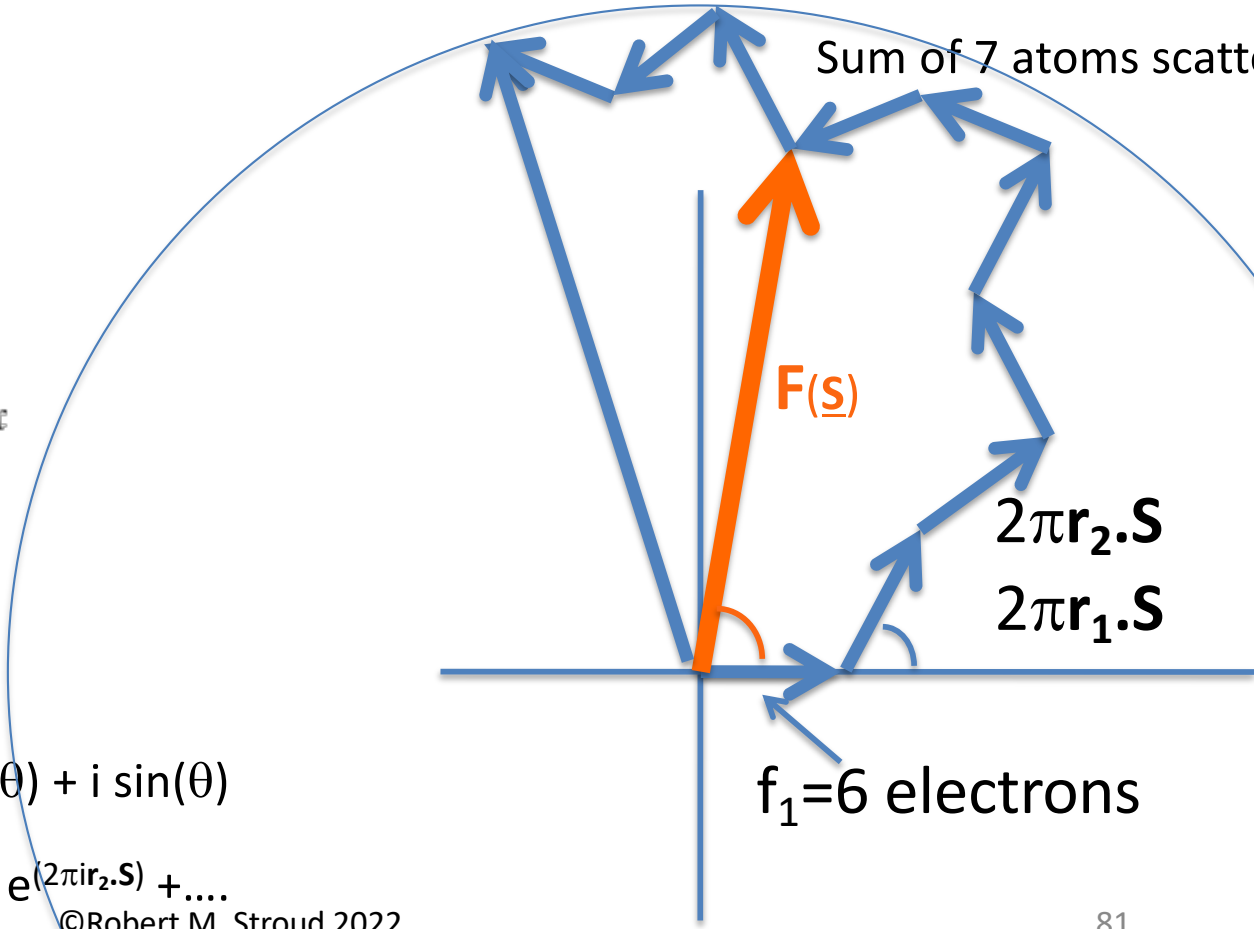
©Robert M. Stroud 2022

Suppose we interpret 7 atoms; but 3 remain to be found in density

Result is a wave of
amplitude $|\mathbf{F}(\underline{\mathbf{S}})|$
phase $\Phi(\underline{\mathbf{S}})$

In reality, maybe 3 atoms are missing.
How to see what is missing?





$$e^{(i\theta)} = \cos(\theta) + i \sin(\theta)$$

$$\mathbf{F}(\underline{\mathbf{S}}) = f_1 e^{(2\pi i \mathbf{r}_1.\mathbf{S})} + f_2 e^{(2\pi i \mathbf{r}_2.\mathbf{S})} + \ldots$$

©Robert M. Stroud 2022

USES: 1. Determining missing regions

$|\mathbf{F}(\underline{\mathbf{S}})|_{obs}$ Compare with $|\mathbf{F}(\underline{\mathbf{S}})|_{calc}$

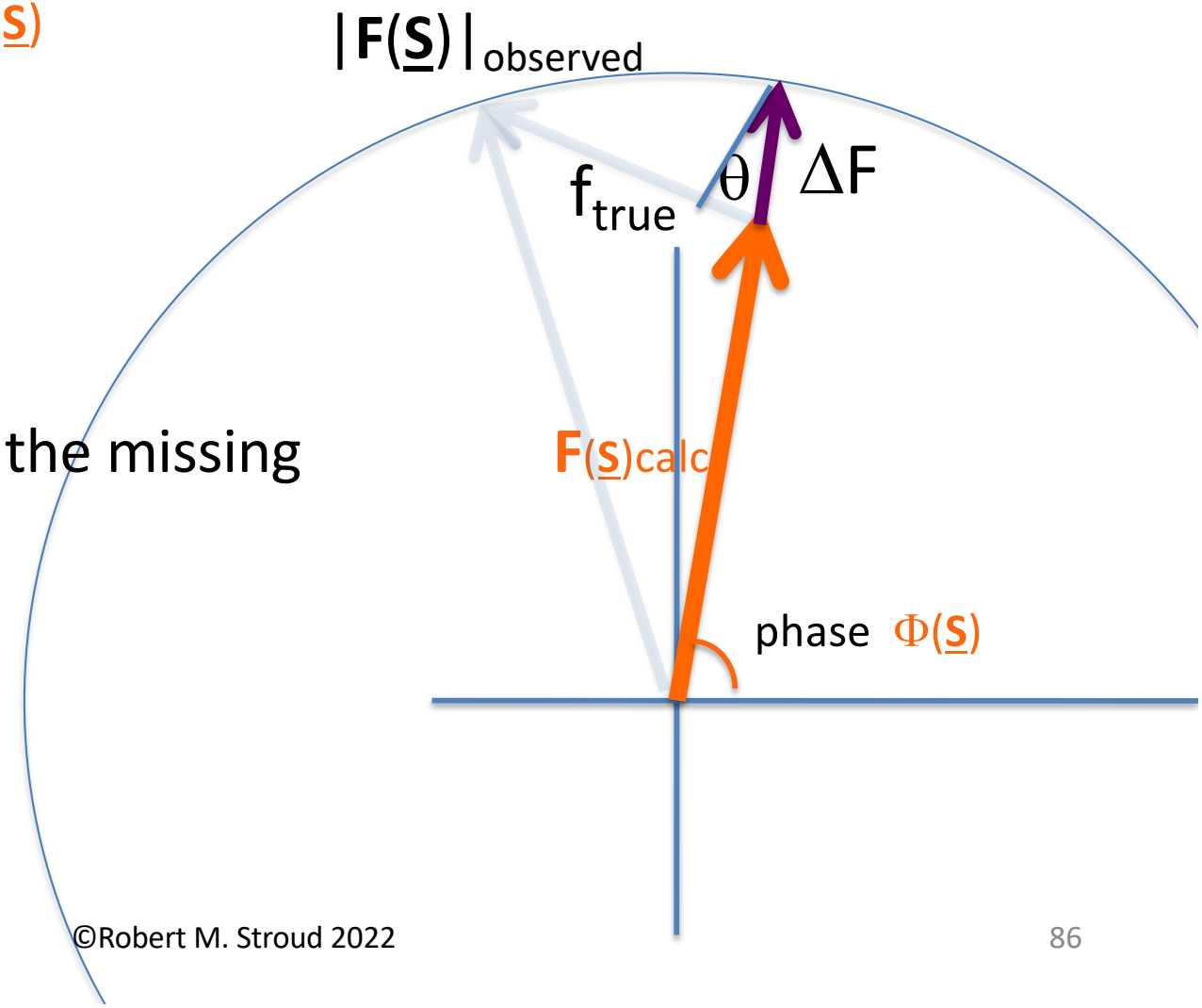
Transform $\Delta F = \big|\, |\mathbf{F}(\underline{\mathbf{S}})|_{obs} - |\mathbf{F}(\underline{\mathbf{S}})|_{calc} \,\big|\ \Phi(\underline{\mathbf{S}})$

or

$[2|\mathbf{F}(\underline{\mathbf{S}})|_{obs+substrate} - |\mathbf{F}(\underline{\mathbf{S}})|_{obs}]\ \Phi(\underline{\mathbf{S}})$

= a '$2F_0$-F_o map'

It is **unbiased** as to where the missing Atoms are.



©Robert M. Stroud 2022

USES: 2. Add a substrate, Grow a new crystal

Measure New $|\mathbf{F}(\underline{\mathbf{S}})|_{obs+substrate}$ Compare with the apo-protein.

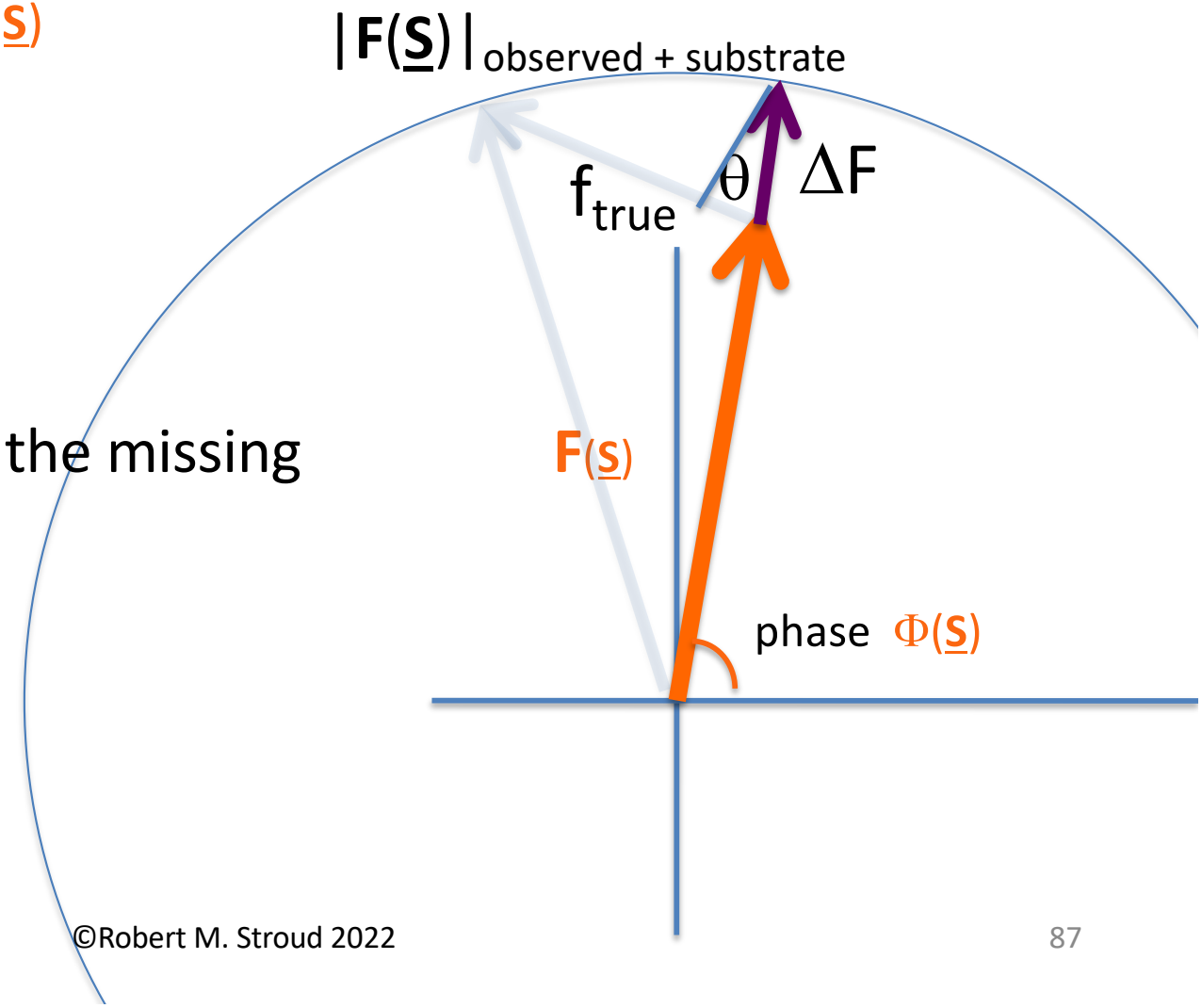
Transform $\Delta F = \left| |\mathbf{F}(\underline{\mathbf{S}})|_{obs+substrate} - |\mathbf{F}(\underline{\mathbf{S}})|_{obs} \right| \Phi(\underline{\mathbf{S}})$

or

$[2|\mathbf{F}(\underline{\mathbf{S}})|_{obs+substrate} - |\mathbf{F}(\underline{\mathbf{S}})|_{obs}] \Phi(\underline{\mathbf{S}})$

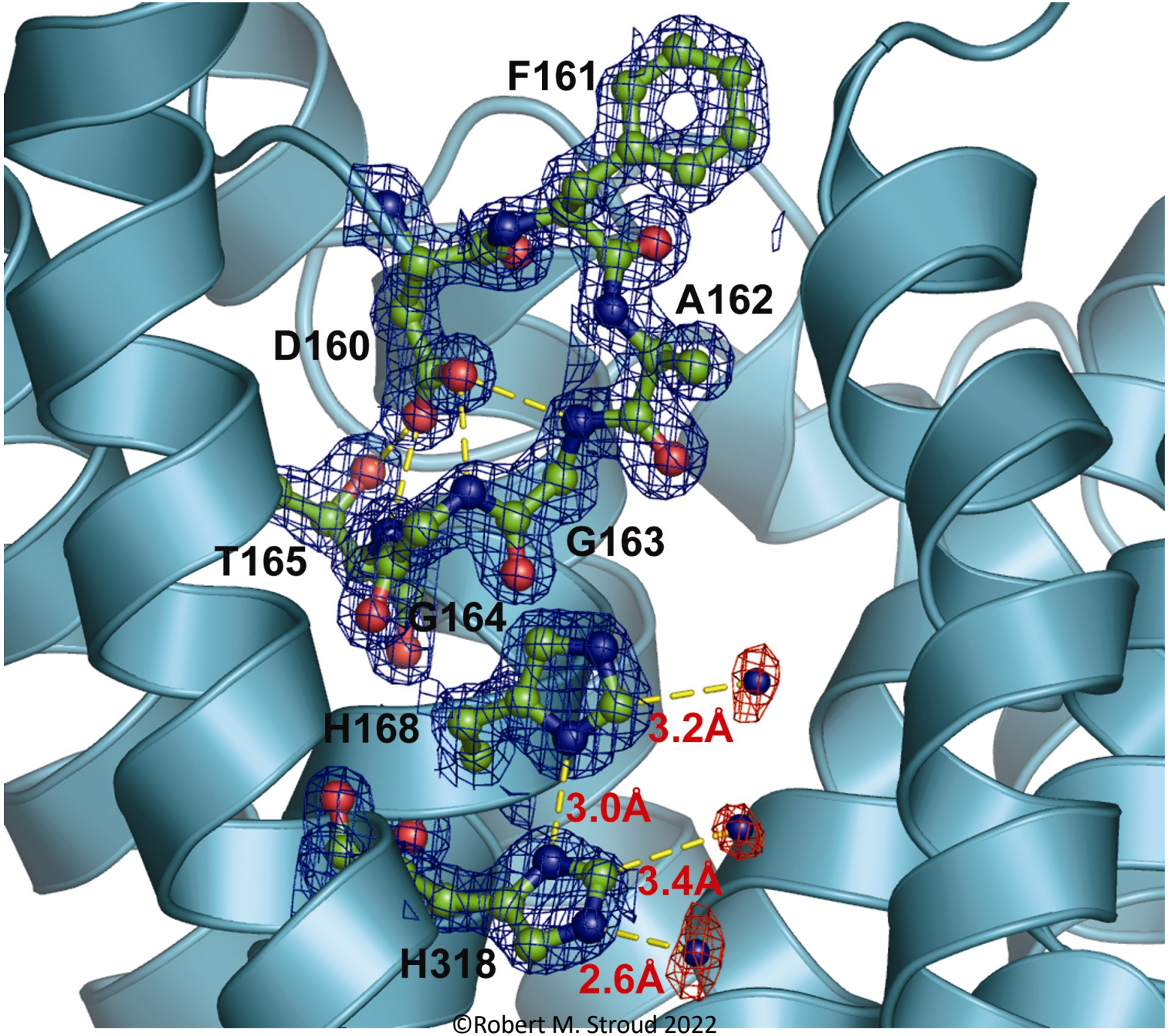
= a '$2F_0$-F_o map'

It is **unbiased** as to where the missing substrate is.



©Robert M. Stroud 2022

A Dfference map shows 1/3 occupied NH3 sites and the role of D160 at 1.35Å Resolution. Here are 0.3 NH_3 peaks!



©Robert M. Stroud 2022

Khademi..Stoud 2003

Fo-Fc maps identify everything ordered that is 'missing'

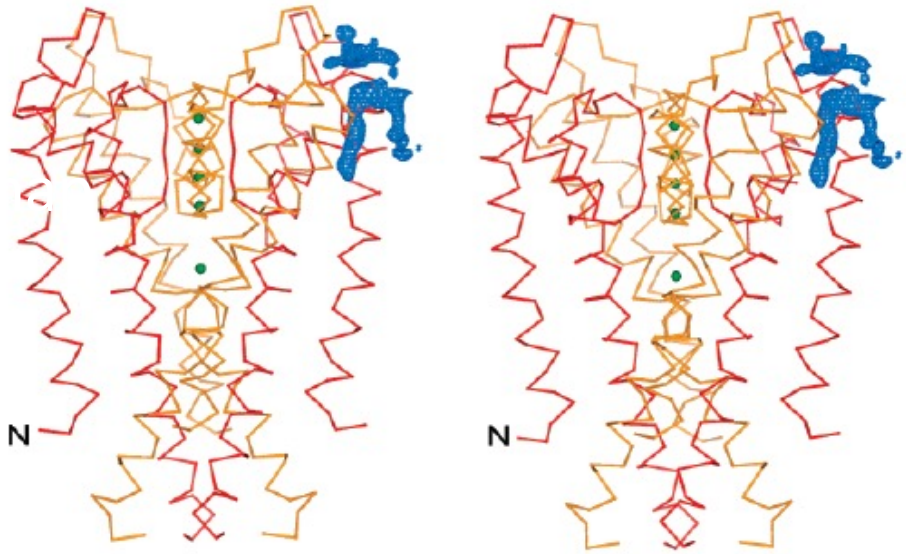


FIGURE 1: Lipid molecules in KcsA crystals. A stereoview of the KcsA structure with electron density corresponding to the lipid molecule. The backbone of KcsA is shown as a red and yellow trace. Green spheres represent potassium ion binding sites. The $F_o - F_c$ map (contoured at 3σ) was calculated using a model that does not contain lipid molecules. For clarity, density corresponding to only one of the lipid molecules is shown. The KcsA monomer consists of an N-terminal outer helix, a central pore helix, and a C-terminal inner helix. This figure was prepared with MOLSCRIPT (*31*) and Raster3D (*32*).

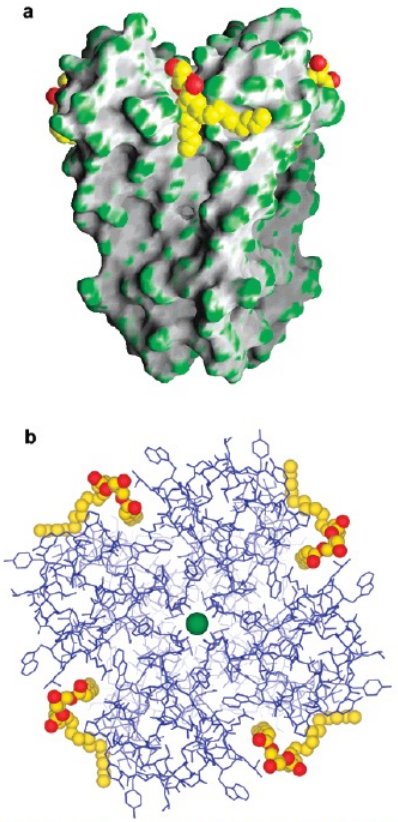


FIGURE 3: Structural analysis of lipid binding to KcsA. (a) Binding surface of the lipid molecule. The surface of KcsA is colored according to curvature (green, convex; gray, concave). The lipid molecule, built as 1,2-diacylglycerol, is shown in CPK representation with oxygen atoms colored red and carbon atoms colored yellow. (b) Lipid-binding site viewed from the extracellular side along the 4-fold axis of KcsA. The channel is colored blue. The green sphere represents the potassium ion. The lipid molecule is in CPK representation colored as in panel a. Panel a was prepared with GRASP (*33*). Panel b was prepared with MOLSCRIPT (*31*) and Raster3D (*32*).

-Eliminate Bias
-Half electron content
-See electrons

©Robert M. Stroud 2022

The closer you get –the lower the noise.
Can see single electrons.

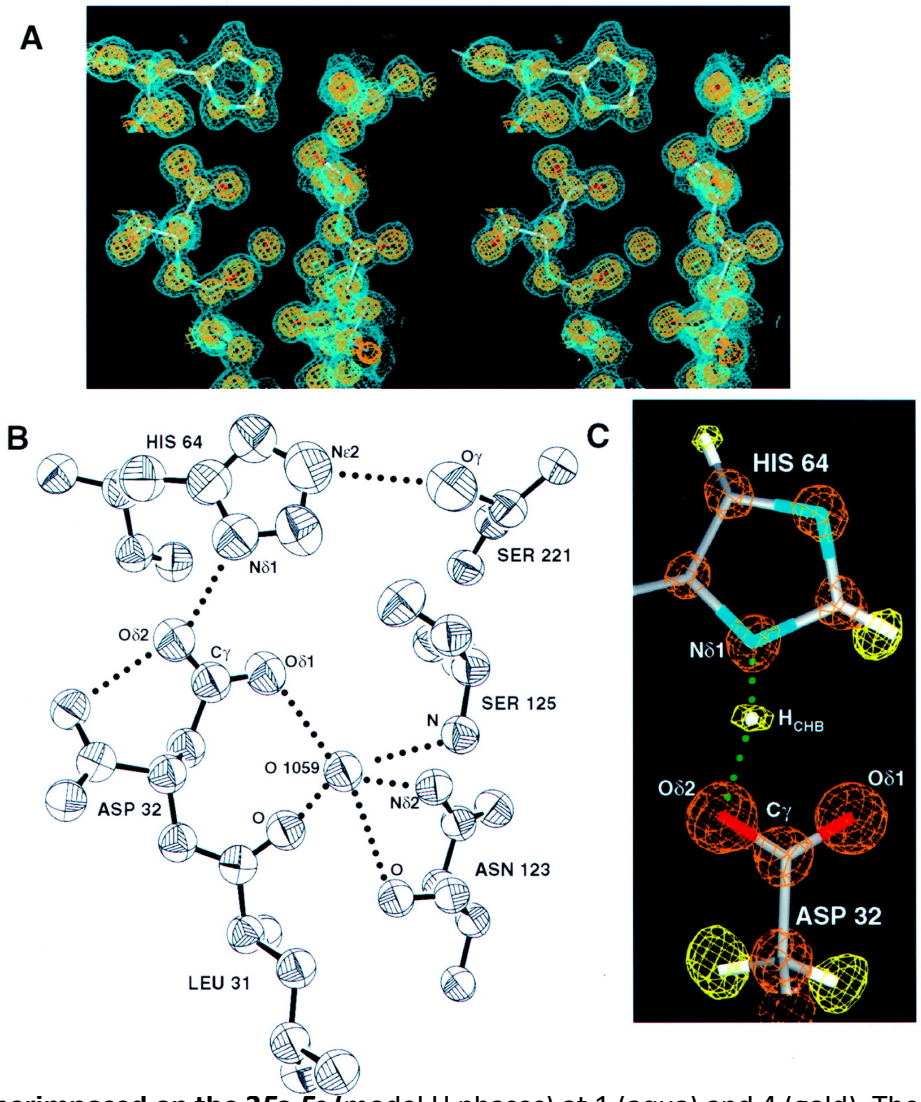
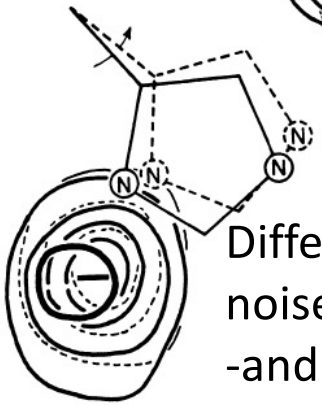


Figure 3 The catalytic triad. **(A) Stereoview displaying Model H superimposed on the 2*Fo Fc* (**model H phases) at 1 (aqua) and 4 (gold). The densities for C and N in His 64 are weaker than in Asp 32. The Asp 32 CO2 bond at 4 is continuous, while the density for the C and O1 are resolved. (B) Schematic of the catalytic residues and hydrogen bonded neighbors with thermal ellipsoid representation countered at 50% probability (*29*). Catalytic triad residues Ser 221 and His 64 show larger thermal motion than the Asp 32. Solvent O1059 appears to be a relatively rigid and integral part of the enzyme structure. (C) Catalytic hydrogen bond (CHB)**. A *Fo Fc* (model H phases) difference map** contoured at +2.5 (yellow) and 2.5 (red) and a 2*Fo Fc* (model H phases) electron density map contoured at 4 (gold). The position of the short hydrogen atom (labeled HCHB) in the CHB is positioned in the positive electron density present between His 64 N1 and Asp 32 O2.

Published in: Peter Kuhn; Mark Knapp; S. Michael Soltis; Grant Ganshaw; Michael Thoene; Richard Bott; *Biochemistry* **1998,** 37, 13446-13452.
DOI: 10.1021/bi9813983
Copyright © 1998 American Chemical Society
©Robert M. Stroud 2022

Seeing small shifts
-down to ±0.1Å



Difference Map
noise ~20% of noise in the parent protein
-and only two peaks!

FIG. 7. The peaks associated with His57 on the difference map. The lower peak is negative density (−) while the other one is positive (+). The latter peak is a composite with a solvent molecule density (see text).

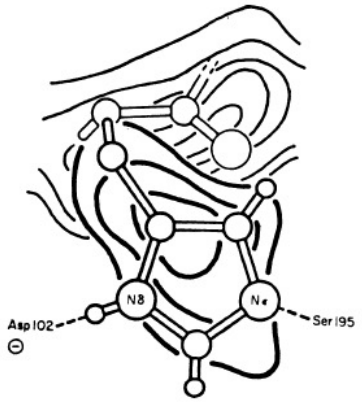


FIG. 8. Electron density for His57 in the DIP-trypsin Fourier map, computed for the plane parallel to the imidazole ring.

TABLE 1

Analysis of Fourier maps

Map	$\langle\lvert F_{obs}\rvert\rangle$ (e)	σ (e)	Calculated† $\langle\Delta\rho^2\rangle^{\frac{1}{2}}$ (e Å^{-3})	Observed‡ r.m.s. error (e Å^{-3})	Observed highest noise (e Å^{-3})	s.d.§	Observed highest peak (e Å^{-3})	s.d.
BA-trypsin—DIP-trypsin	84·7	2·3	0·069	0·059	0·17	2·5	0·75	11
DIP-trypsin	573·0	21·0	0·38	—	—	—	—	—

† ΔF: $\langle\Delta\rho^2\rangle = \frac{1}{2V^2}\sum_{hkl}\Delta F^2\,(2-m^2)$,

F_{DIPT}: $\langle\Delta\rho^2\rangle = \frac{1}{V^2}\sum_{hkl}F^2_{DIPT}\,(1-m^2)$,

(after Henderson & Moffat, 1971).

‡ The observed root mean-square density error is based on a relatively featureless region of the map.
§ s.d., the electron density given as a multiple of the observed r.m.s. error.

©Robert M. Stroud 2022

Difference maps; to the last electron! Why?
Supplementary Proof of the ‘Random Walk’ calculation

©Robert M. Stroud 2022

The 'Random Walk' problem? (p33.1-33.3)

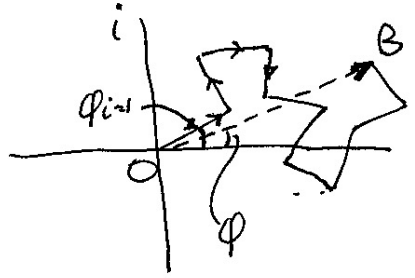
What is the average sum of
n steps in random directions?

(What is the average amplitude
<|F(s)|> from an n atom structure?)

-AND why do we care?!.....

How much difference in <|F(s)|> from adding a
4 carbon atom substrate?
mercury atom (f=80)?

"Random Walk", in 2d - -
(What is scattering amplitude from
n atoms, each with f electrons)



n steps, of length f
relatively randomly
different angles.
ie. Probability
$P(\varphi)$

Easiest to consider as the expected value
maximum for $|OB|^2$ (rather than OB)

$$|OB|^2 = \underline{OB}.\underline{OB}^*$$

[Review: $\underline{OB} = |OB|(\cos(\varphi) + i\sin\varphi)$
$\underline{OB}^* = |OB|(\cos(\varphi) - i\sin\varphi)$

so $\underline{OB}.\underline{OB}^* = |OB|^2(\cos^2\varphi + \sin^2\varphi) = |OB|^2$]

$$\text{So } \underline{OB} = \sum_{i=1}^{n} f\, e^{2\pi i \varphi_i}$$

$$\underline{OB}^* = \sum_{i=1}^{n} f\, e^{-2\pi i \varphi_i}$$

©Robert M. Stroud 2022

The average intensity for an n atom structure, each of f electrons is $\langle I \rangle = nf^2$

The average amplitude is Square root of n, times f

$$|OB|^2 = \left(\sum_{i=1}^{n} e^{2\pi i \varphi_i}\right)\left(\sum_{j=1}^{n} e^{-2\pi i \varphi_j}\right) \cdot f^2$$

$$= \sum_i \sum_j e^{2\pi i (\varphi_i - \varphi_j)} \cdot f^2$$

since $P(\varphi_i) = P(\varphi_j)$ = constant (all equally probable)

the sums..

- for $i \neq j$ $P(\varphi_i)$ all equally probable

so $\sum_i \sum_j \Rightarrow \int_0^{2\pi} \cos 2\pi(\varphi_i - \varphi_j) + i \sin 2\pi(\varphi_i - \varphi_j)$

$\cos \to 0$ $\quad$ $\sin \to 0$

- for $i = j$ $e^{2\pi i(\varphi_i - \varphi_j)} = e^{2\pi i \cdot 0} = 1$

$$\langle |OB|^2 \rangle = nf^2$$

$$OB = \sqrt{n}\, f \qquad (1)$$

©Robert M. Stroud 2022

Can we "see" an added 4 atoms?

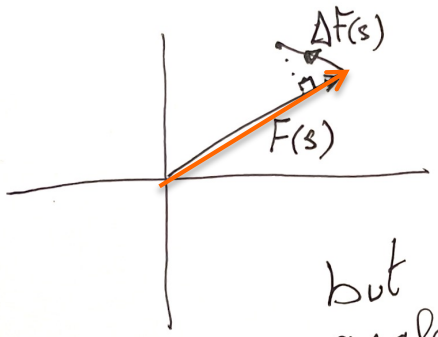
25 kDa Protein ~ 2500 atoms of $f \sim 7$ electrons each.

We measure $I = |F(s)|^2$

where $\langle |F(s)| \rangle \sim \sqrt{2500} \cdot 7 = 350$ electrons

Change in $\langle |F(s)| \rangle$ from adding 4 atoms?

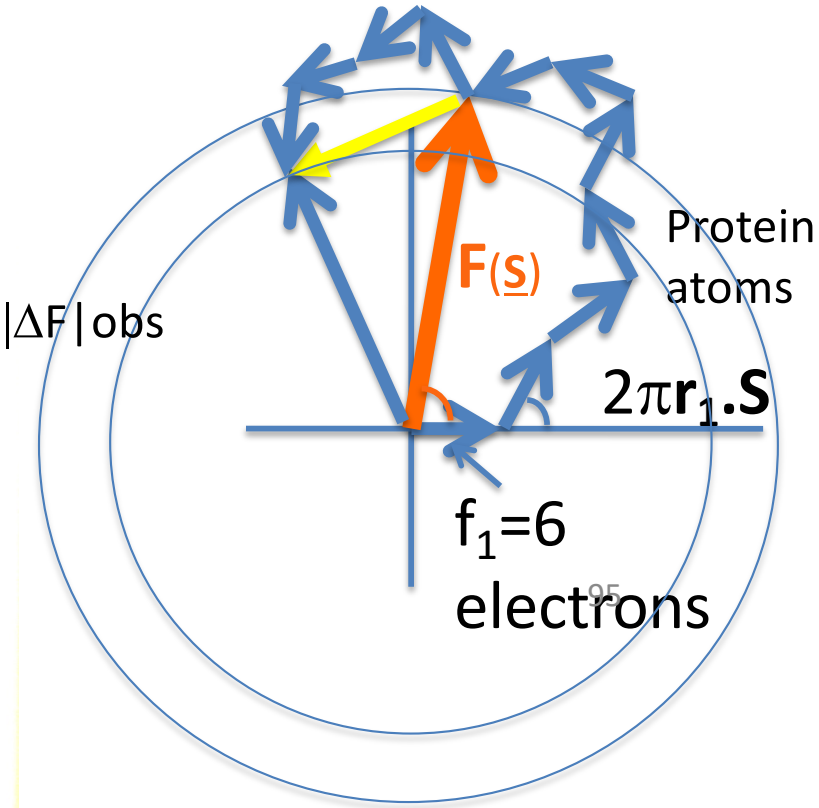
$\langle |\Delta F(s)| \rangle \sim \sqrt{4} \cdot 7 = 14$ electrons



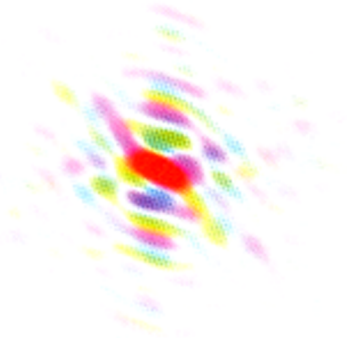
but $\Delta F(s)$ is at a 'random' angle to $F(s)$ so the difference in $|F(s)|$ will only be $\frac{2}{\pi} \Delta F(s) \sim 9$ electrons

so $\langle |\Delta F|_{obs} \rangle / \langle |F_{obs}| \rangle = \frac{9}{350} \sim 2.5\%$

$\langle \Delta I \rangle / \langle I \rangle = 5\%$



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



Crystal

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

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

|Fo-Fc| Difference Maps

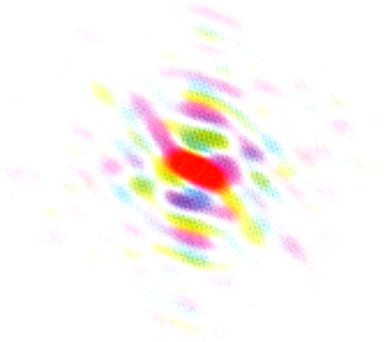
Phases $\phi_{(h,k,l)}$

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

Alphafold Molecular Replacement
Similar molecule Replacement
Experimental
heavy atom labels
selenium for sulfur
Trial & error similar structure

Atom positions (x,y,z)

Change one side ⟺ Change the other



Crystal

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

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

|Fo-Fc| Difference Maps

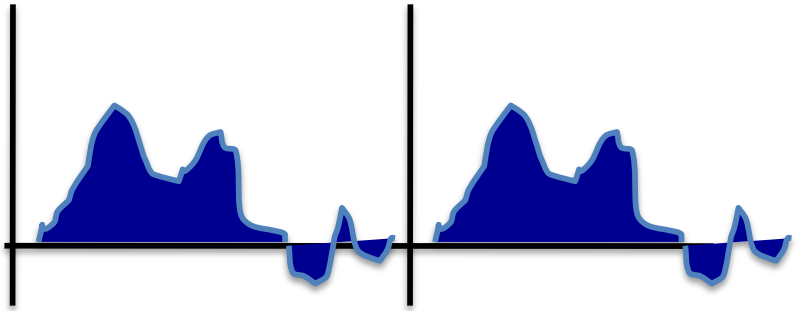
Phases $\phi_{(h,k,l)}$

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

Alphafold Molecular Replacement
Similar molecule Replacement
Experimental
heavy atom labels
selenium for sulfur
Trial & error similar structure

Atom positions (x,y,z)

Density Modification: Solvent Flattening



Crystal

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

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

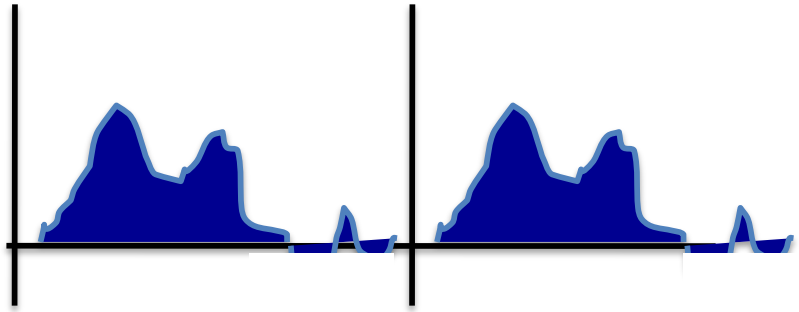
Phases $\phi_{(h,k,l)}$

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

Experimental
heavy atom labels
selenium for sulfur
Trial & error similar structure

Atom positions (x,y,z)

Density Modification: Solvent Flattening



Crystal

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

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

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

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

Experimental
heavy atom labels
selenium for sulfur
Trial & error similar structure

Atom positions (x,y,z)

Refinement

- Least Squares Refinement is common when errors in observations are presumed to be random errors that obey Gaussian statistics.
- Refine x_i, y_i, z_i, B_i with respect to the F_o

Minimize $E = \Sigma_{hkl} 1/\sigma^2 (k|F_{obs}| - |F_{calc}|)^2$ with respect to $(xyzB)_i$ of all atoms.

To include an energy term, that constrains the structure toward acceptable geometry

Minimize E = (1-w) Energy + w $\Sigma_{hkl} 1/\sigma^2 (k|F_{obs}| - |F_{calc}|)^2$ where w is the fractional weighting on geometry versus X-ray terms. Energy has vdW, torsional restraints, bond length and dihedral angles.

Maximum Likelihood refinement seeks the most probable solution most consistent with all observations. ie Least squares refinement alone minimizes the difference between |Fo| and |Fc|.

Validation? 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

©Robert M. Stroud 2022

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

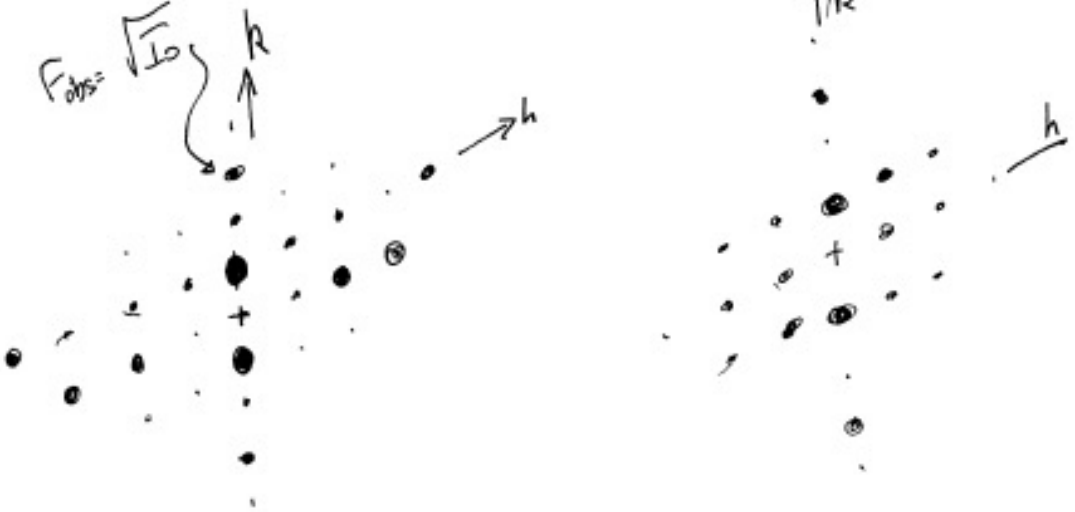
Intensity = Amplitude²

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

$\sqrt{I_{obs}}_{hkl} = F_{obs}{}_{hkl}$

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

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



How do we judge the Quality of structure?

2. Overall quality criteria: agreement of observations 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 Fo are ~ 3%. inadequate models of solvent, atom motion, anharmonicisity

6 Accuracy ~ 0.5*res*R

©Robert M. Stroud 2022

Residual "R" factors

R_{cryst} (or just "R") How well does structure match the observations?

observed vs calculated data (F(s))

R_{free} Remove bias; leave some observations out of determination

cross-check with "random" subset of data

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

R_{sym} = R_{merge} How self consistent are observations that should be identical? = measuring errors.

(self-consistency of data: Is)

“R” factors

$$R = \frac{\sum \left| F_{obs} - F_{calc} \right|}{\sum F_{obs}}$$

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” factors

R_{cryst} (or just “R”)

observed vs calculated data (Fs)

R_{free}

cross-check with “random” subset of data

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

R_{merge}

$$R_{merge} = \frac{\sum \left| I_{obs} - \langle I \rangle \right|}{\sum I_{obs}}$$

← blows up as $I_{obs} \rightarrow 0$

completely random: **0.59**

weak data (high angle): **0.7- ∞**

wrong symmetry choice?: **~0.2-0.55**

small or disordered crystal: **~0.1-0.2**

typical: **~ 0.05**