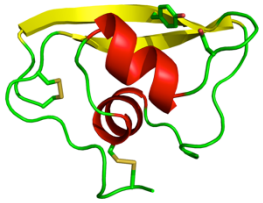


Central concepts in (bio-)crystallography

4. Electron Density Maps

Jeroen Mesters
UNIVERSITÄT ZU LÜBECK
STIFTUNGSUNIVERSITÄT
SEIT 2015

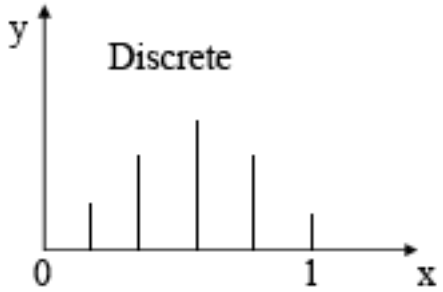
Electron density

Fourier pair of equations:

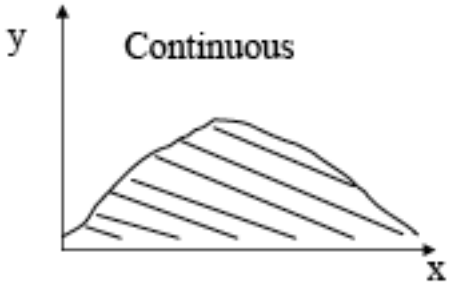
$$\underline{\mathbf{F}}_{hkl} = V \int_{xyz} \rho(xyz)\, \exp^{[2\pi i\cdot(hx + ky + lz)]}\, dxdydz$$

$$\rho(xyz) = 1/V \sum_{hkl} |\underline{\mathbf{F}}_{hkl}|\, \exp^{[-2\pi i\cdot(hx + ky + lz) + i\alpha(hkl)]}$$

$$\rho(xyz) = {}^{1}/_{V}|\underline{\mathbf{F}}_{000}| + {}^{2}/_{V} \sum_{+hkl} |\underline{\mathbf{F}}_{hkl}| \cos[2\pi\cdot(hx + ky + lz) - \alpha_{hkl}]$$



$f(x_{0\to1}) = x_1 + x_2 + x_3 + x_4 + x_5$
$= \sum_x f(x_n)$

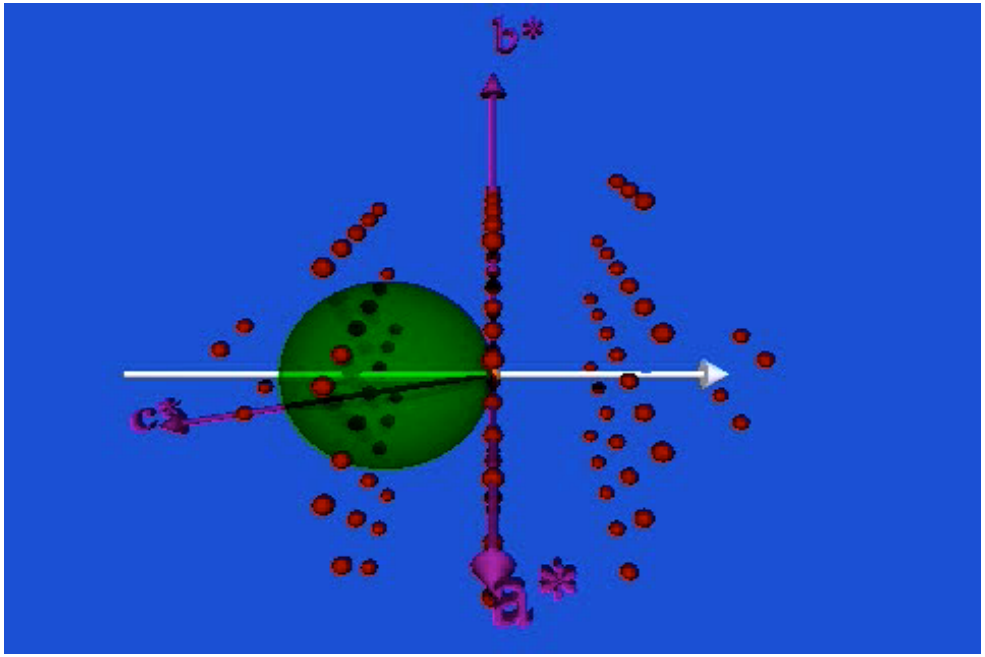


$f(x_{0\to1}) = \int_{0\to1} f(x)dx$
{the area under the function curve}

* Note that $e^{ix} = \cos(x) + i\sin(x)$.

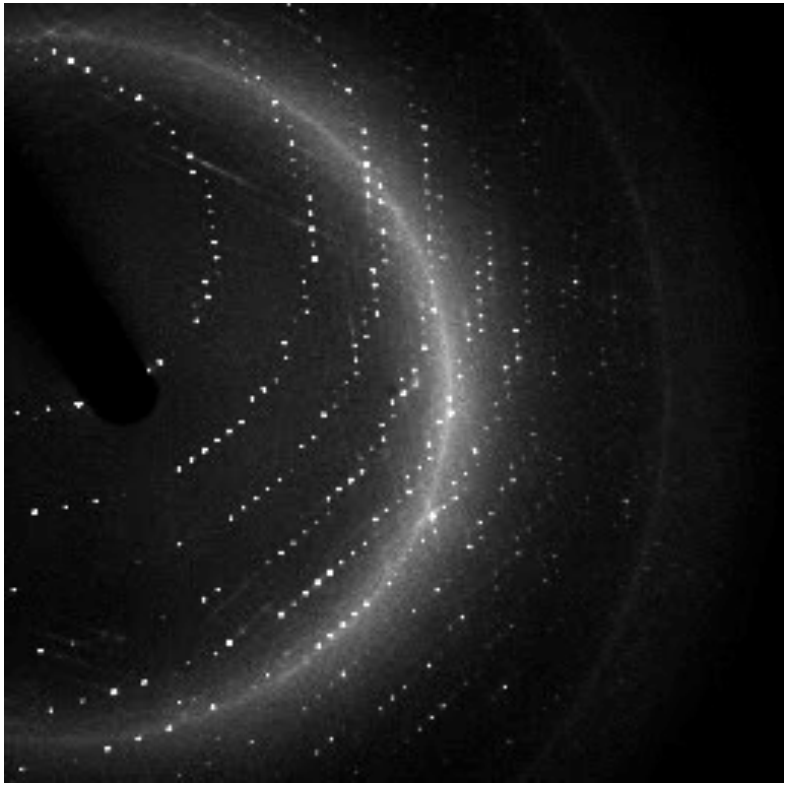
Bragg‘s Law & Ewald Movies

Ewald sphere



http://escher.epfl.ch/x-ray/diff.mpeg

X-ray detector



$$n / d = (2 \cdot \sin\theta) / \lambda$$

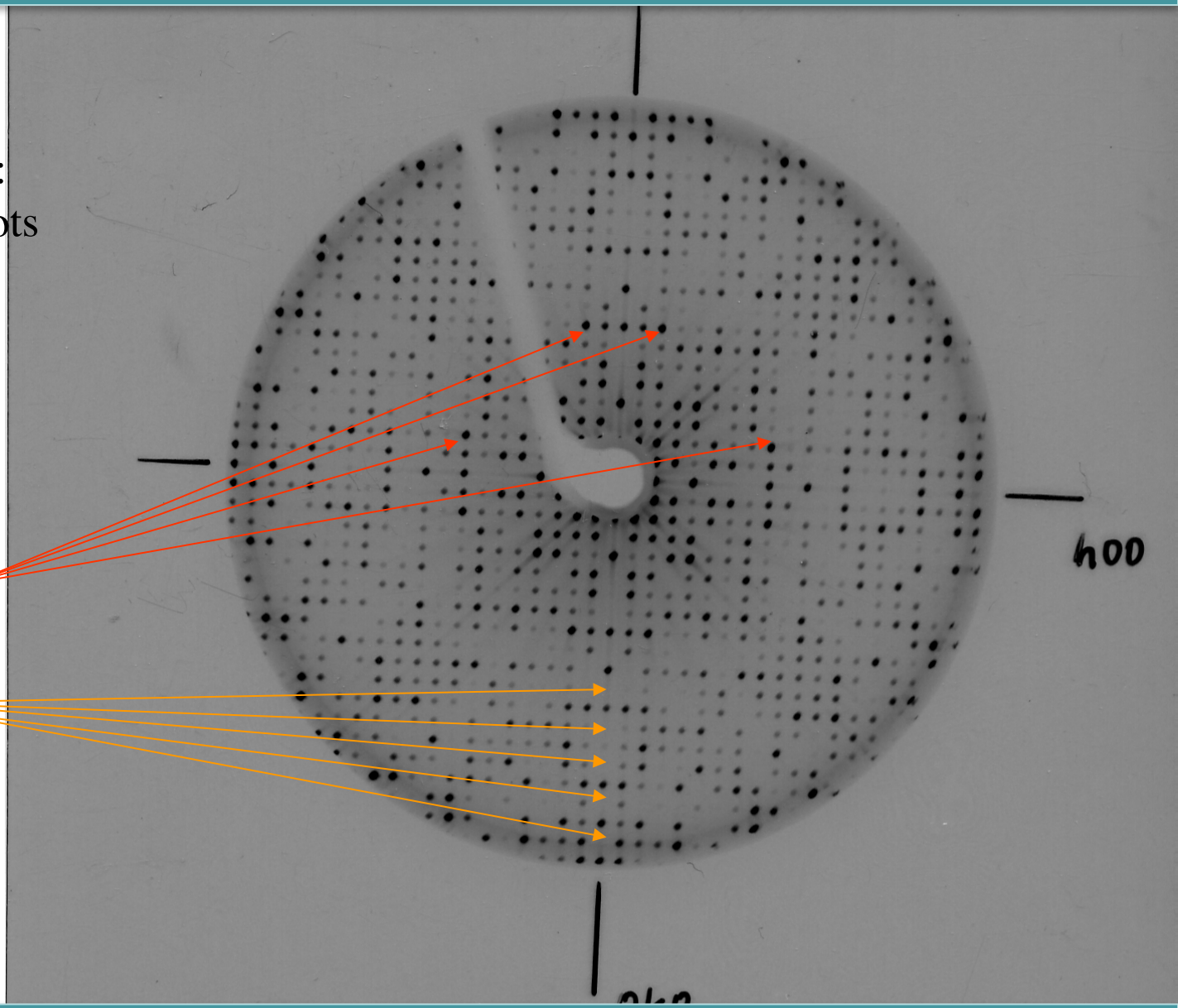
Convolution of Lattice and Motif

Lattice determines position and arrangement of spots: Distance between spots reveals unit cell parameters

Motif influences intensities of spots

4-fold Symmetry

Screw-axis: systematic absences



B-factor

The size of the electron cloud is temperature independent under biological conditions. However, thermal vibration smears out the atom position thereby increasing the "size" of the atom. Therefore, atomic scattering factors must be multiplied by a "temperature dependent factor"!

Correction along $\underline{\mathbf{S}}$ equal to

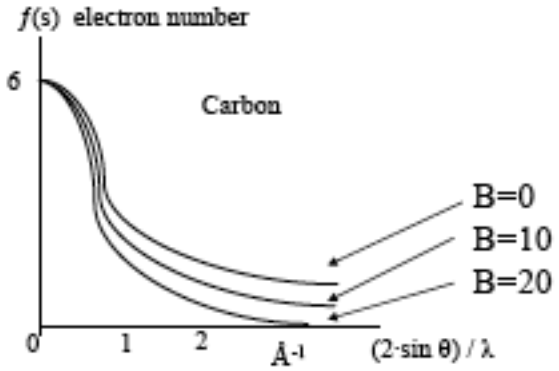
$$T(\text{isotropic}) = \exp\left[(-B/4)(2\sin(\theta)/\lambda)^2\right]$$

The thermal parameter B is related to the mean square displacement $\underline{\mathbf{u}}^2$ of the vibration: $B = 8\pi^2 \cdot \underline{\mathbf{u}}^2$

For B = 30 Å^2, the thermal displacement (r.m.s.) = 0.62 Å

B factor effect

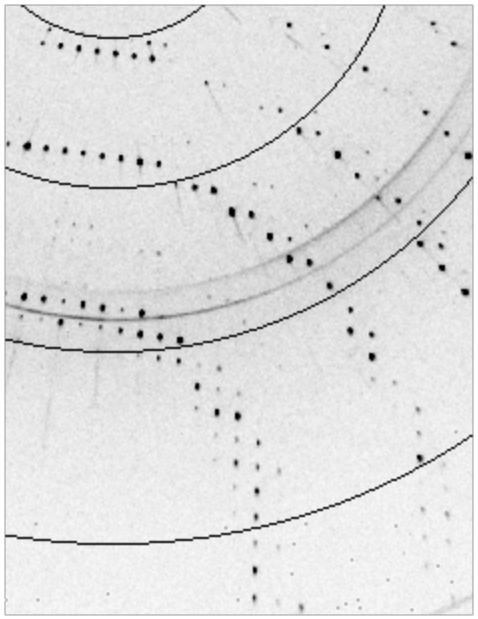
Thermal vibration leads to diminished scattering at higher scattering angles



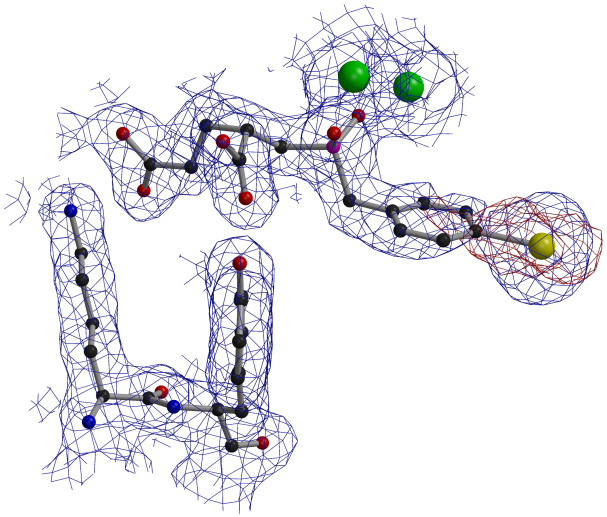
Example from PDB database (B factor displayed in red):

CRYST1 58.452 85.756 46.746 90.00 90.00 90.00 P 21 21 2 8
ATOM 1 N PRO A 1 28.721 40.079 5.613 1.00 36.25 N
ATOM 2 CA PRO A 1 29.610 38.971 5.185 1.00 35.47 C
ATOM 3 C PRO A 1 29.034 38.272 3.988 1.00 33.71 C
ATOM 4 O PRO A 1 27.890 38.531 3.647 1.00 27.37 O
ATOM 5 CB PRO A 1 29.610 38.012 6.382 1.00 36.82 C
ATOM 6 CG PRO A 1 28.502 38.398 7.164 1.00 37.19 C
ATOM 7 CD PRO A 1 28.240 39.862 6.980 1.00 37.01 C

$|F_{hkl}|$ and 3D structure



1 0 0 F_{100} ?
2 0 0 F_{200} ?
3 0 0 F_{300} ?
4 0 0 F_{400} ?
5 0 0 F_{500} ?
6 0 0 F_{600} ?
7 0 0 F_{700} ?
etc.



|F(hkl)| → ρ(xyz)

MR, MIR, MAD, etc..

Crystallographic Phase Problem

$$\rho(xyz) = {}^{1}/_{V}|\underline{\mathbf{F}}_{000}| + {}^{2}/_{V}\Sigma_{+h}\Sigma_{k}\Sigma_{l}\,|\underline{\mathbf{F}}_{hkl}|\cos[2\pi\cdot(hx+ky+lz)-\alpha_{hkl}]$$

How to obtain information about α?

1) Isomorphous crystal structure (difference-Fourier synthesis)
2) Molecular replacement technique (MR)
3) Multiple isomorphous replacement (MIR)
4) Multi-wavelength anomalous diffraction (MAD)

) Patterson

5) Ab initio

ED calculation

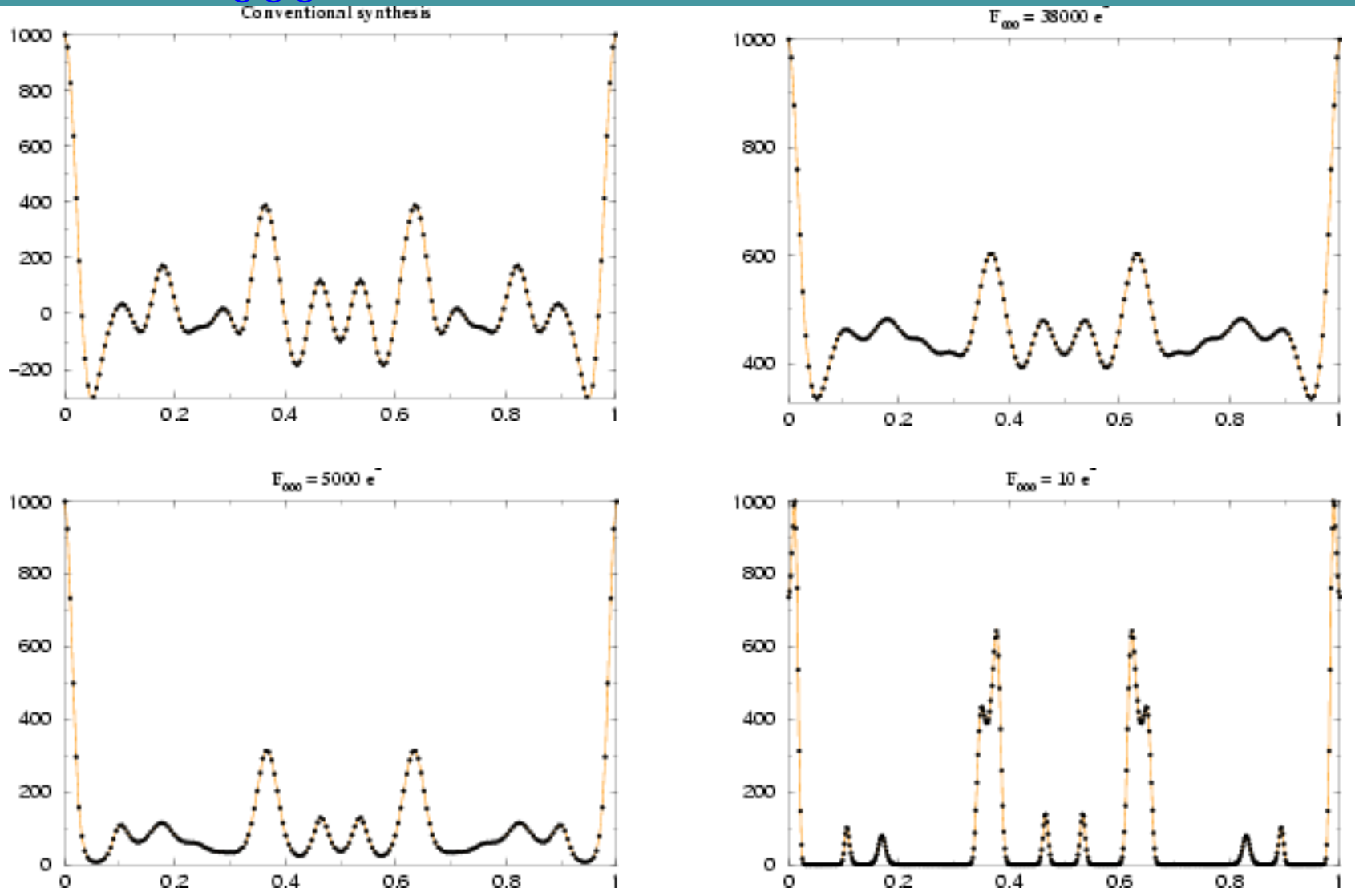
$$\rho(xyz) = 1/V \sum_{hkl} |\underline{F}_{hkl}| \exp^{[-2\pi i \cdot (hx + ky + lz) + i\alpha(hkl)]}$$

$$\rho(xyz) = {}^{1}/_{V}|\underline{F}_{000}| + {}^{2}/_{V} \sum_{+hkl} |\underline{F}_{hkl}| \cos[2\pi \cdot (hx + ky + lz) - \alpha_{hkl}]$$

1 0 0 F_{100} α_{100}
2 0 0 F_{200} α_{200}
3 0 0 F_{300} α_{300}
4 0 0 F_{400} α_{400}
5 0 0 F_{500} α_{500}
6 0 0 F_{600} α_{600}
7 0 0 F_{700} α_{700}
etc.

The electron density at a position xyz can be calculated by summing the contributions of as many F_{hkl} as possible: the more F_{hkl} the better the resolution of the single atoms *i.e.* the better the quality and the higher information content.

F_{000} and effect on e.d.-map



https://utopia.duth.gr/glykos/GraphEnt-html/node92.html

Mean e.d. 0.40 e^{-}Å^{-3}

A crystal with 50% 2M $(NH_4^+)_2SO_4$ and 50% protein has a mean e.d. of $\pm$ 0.40 e^{-} Å^{-3} >>> $F_{000} = 0.40 V_{cell}\ e^{-}$.

For the majority of macromolecular problems this is an overestimate which is no harm as the maps will not be too sharp and it will hard to misinterpret them.

Average specific protein density is 1.35 g cm^{-3}, and average protein electron density therefore is 0.44 e^{-} Å^{-3}.

(difference) electron-density maps

$$\rho(xyz) = {}^{1}/_{V} |\underline{\mathbf{F}}_{000}| + {}^{2}/_{V} \sum_{+hkl} |\underline{\mathbf{F}}_{obs}| \cos [2\pi\cdot(hx+ky+lz) - \alpha_{mod}(hkl)]$$

Unweighted maps of type $(n|\underline{\mathbf{F}}_{\mathbf{obs}}| - (n-1)|\underline{\mathbf{F}}_{\mathbf{calc}}|)$*:

1) $\rho(xyz) = {}^{1}/_{V} |\underline{\mathbf{F}}_{000}| + {}^{2}/_{V} \sum_{hkl} (|\underline{\mathbf{F}}_{\mathbf{obs}}| - |\underline{\mathbf{F}}_{\mathbf{calc}}|) \cos [.....]$

2) $\rho(xyz) = {}^{1}/_{V} |\underline{\mathbf{F}}_{000}| + {}^{2}/_{V} \sum_{+hkl} (|\underline{\mathbf{F}}_{\mathbf{obs}}| + |\underline{\mathbf{F}}_{\mathbf{obs}}| - |\underline{\mathbf{F}}_{\mathbf{calc}}|) \cos [.....]$

3) $\rho(xyz) = {}^{1}/_{V} |\underline{\mathbf{F}}_{000}| + {}^{2}/_{V} \sum_{hkl} (2|\underline{\mathbf{F}}_{\mathbf{obs}}| - |\underline{\mathbf{F}}_{\mathbf{calc}}| + |\underline{\mathbf{F}}_{\mathbf{obs}}| - |\underline{\mathbf{F}}_{\mathbf{calc}}|) \cos [..]$

F_{obs} and F_{calc}

F_{obs} is calculated from the diffraction experiment:

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

F_{calc} is derived from the structure model build ($\rho_{structure}$) i.e. amino acid sequence threaded into the electron density maps:

$$\underline{\mathbf{F}}_{calc} = V \int_x \int_y \int_z \rho_{structure}(xyz) \exp^{[2\pi i \cdot (hx + ky + lz)]} dxdydz$$

A comparison of $|F_{obs}|$ and $|F_{calc}|$ is useful in refining (i.e. model rebuilding) the correct 3D structure.

F_{obs} - F_{calc} electron density map

$$\rho(xyz) = {}^{1}/_{V} \sum_{hkl} (|\underline{\mathbf{F}}_{o}(hkl)| - |\underline{\mathbf{F}}_{c}(hkl)|) \exp[-2\pi i \cdot (hx+ky+lz)+i\alpha(hkl)]$$

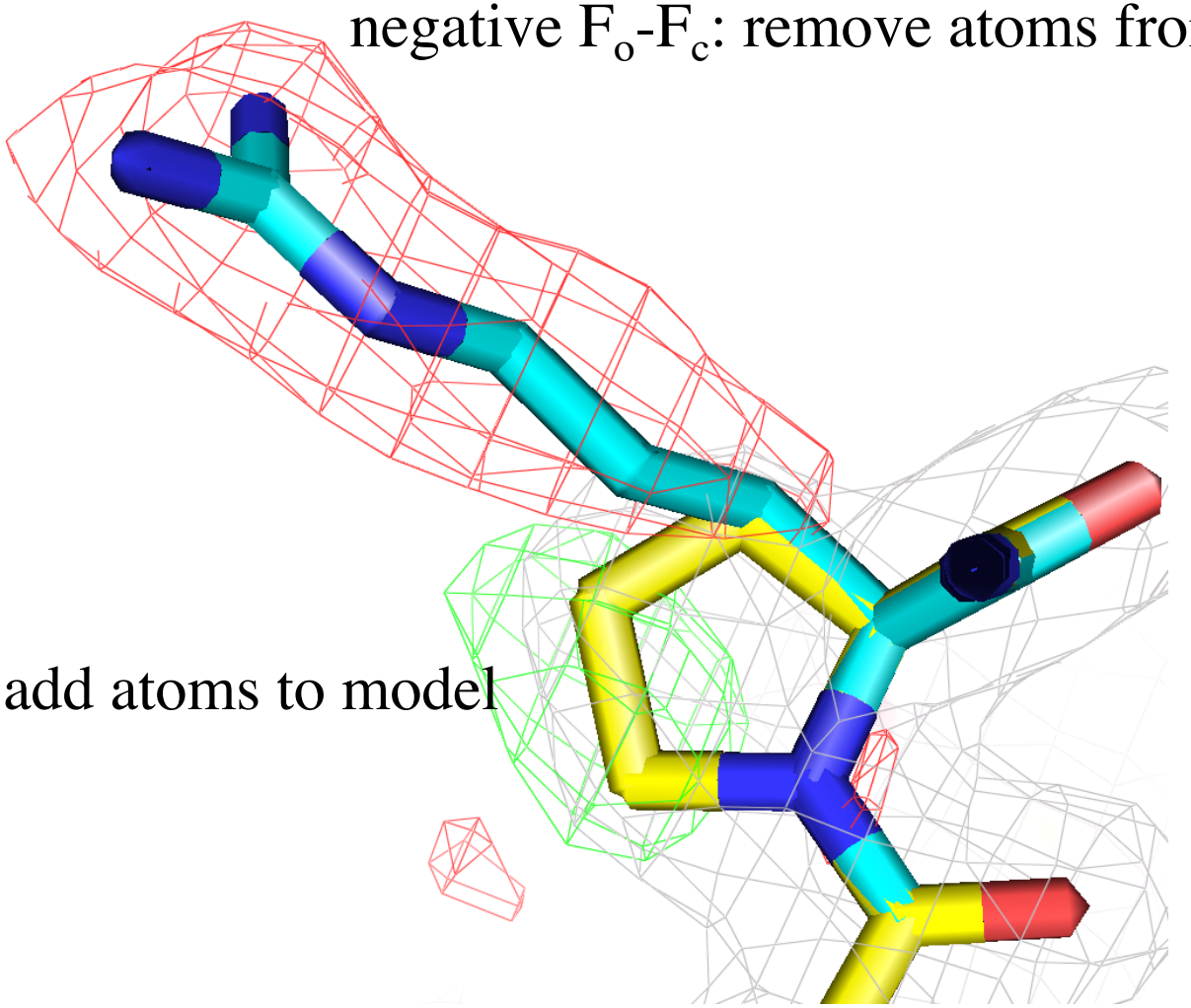
Case 1: if the model, *e.g.* $|\underline{\mathbf{F}}_{calc}(hkl)|$, is equal to what we measure/observe, *e.g.* $|\underline{\mathbf{F}}_{obs}(hkl)|$, then $\rho(xyz) = 0$

Case 2: if the model is incomplete with respect to what we measure/observe, then $\rho(xyz) > 0$ (positive)

Case 3: if the model is over-complete what we measure/observe then $\rho(xyz) < 0$ (negative)

Difference densities

negative F_o-F_c: remove atoms from model



positive F_o-F_c: add atoms to model

$2F_{obs} - F_{calc}$ electron density map

$$\rho(xyz) = {}^{1}/_{V} \sum_{hkl} (2 \cdot |\underline{\mathbf{F}}_o(hkl)| - |\underline{\mathbf{F}}_c(hkl)|) \exp[-2\pi i \cdot (hx+ky+lz) + i\alpha(hkl)]$$

This map can be regarded as the sum of the electron density of a model plus a difference electron density map.

$3F_{obs} - 2F_{calc}$ electron density map

$$\rho(xyz) = {}^{1}/_{V} \sum_{hkl} (3 \cdot |\underline{\mathbf{F}}_o(hkl)| - (2 \cdot |\underline{\mathbf{F}}_c(hkl)|) \exp[\ \ldots\ldots\ ]$$

$$3F_{obs} - 2F_{calc} = (2F_{obs} - F_{calc}) + (F_{obs} - F_{calc})$$

In other words, this shows a summation of the map covering the model and the map showing the differences.

M. Vijayan: „as the amount of missing structure increases, the relative height of peaks for missing atoms drops, so a higher n is needed to equalise peaks from modeled and missing atoms“

ED maps

$$\rho(xyz) = {}^{1}/_{V} \sum_h \sum_k \sum_l |\underline{\mathbf{F}}(hkl)| \exp^{[-2\pi i \cdot (hx + ky + lz) + i\alpha(hkl)]}$$
$$= {}^{1}/_{V} F_{000} + {}^{2}/_{V} \sum_{+hkl} |\underline{\mathbf{F}}(hkl)| \cos[2\pi \cdot (hx + ky + lz) - \alpha(hkl)]$$

$$\rho(xyz) = {}^{1}/_{V} F_{000} + {}^{2}/_{V} \sum_{+hkl} (2|\underline{\mathbf{F}}_{\mathbf{obs}}| - |\underline{\mathbf{F}}_{\mathbf{calc}}|) \cos [\ldots..]$$

if $\underline{\mathbf{F}}_{\mathbf{obs}} = \underline{\mathbf{F}}_{\mathbf{calc}}$ then map covers whole model

$$\rho(xyz) = {}^{1}/_{V} F_{000} + {}^{2}/_{V} \sum_{+hkl} (|\underline{\mathbf{F}}_{\mathbf{obs}}| - |\underline{\mathbf{F}}_{\mathbf{calc}}|) \cos [\ldots..]$$

if $\underline{\mathbf{F}}_{\mathbf{obs}} = \underline{\mathbf{F}}_{\mathbf{calc}}$ then there is no map (clean/empty)

$$\rho(xyz) = {}^{1}/_{V} F_{000} + {}^{2}/_{V} \sum_{+hkl} (3|\underline{\mathbf{F}}_{\mathbf{obs}}| - 2|\underline{\mathbf{F}}_{\mathbf{calc}}|) \cos [\ldots..]$$

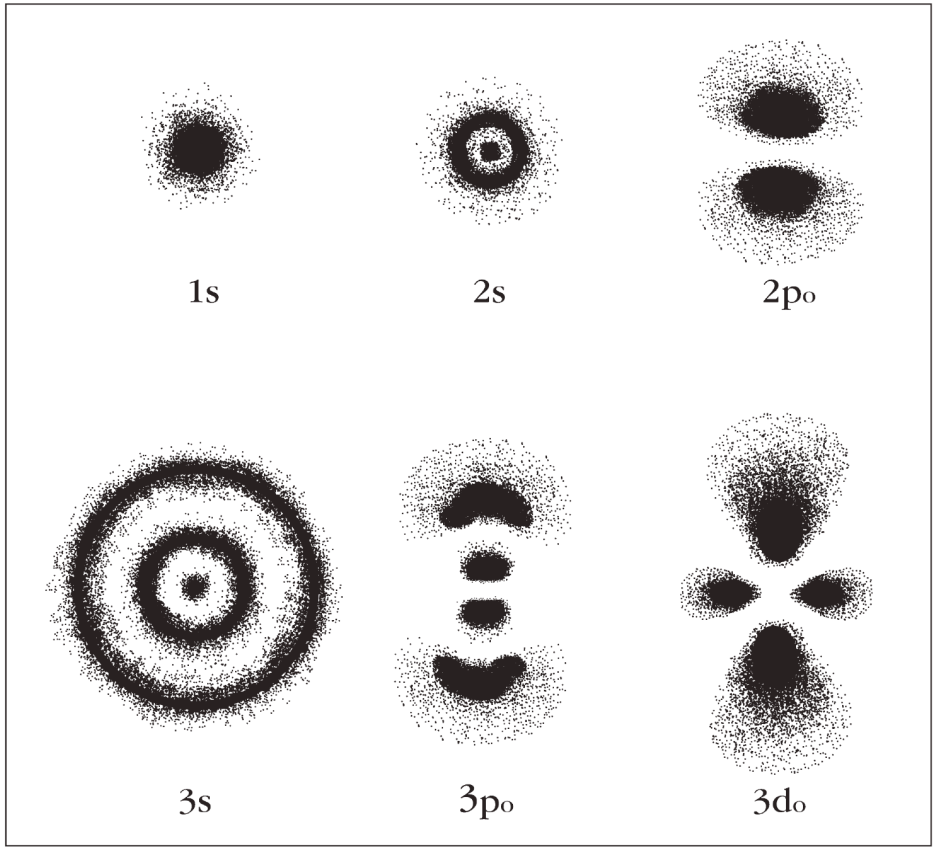
summation of $2F_o$-F_c and F_o-F_c and if $\underline{\mathbf{F}}_{\mathbf{obs}} = \underline{\mathbf{F}}_{\mathbf{calc}}$ then map equal to F_o

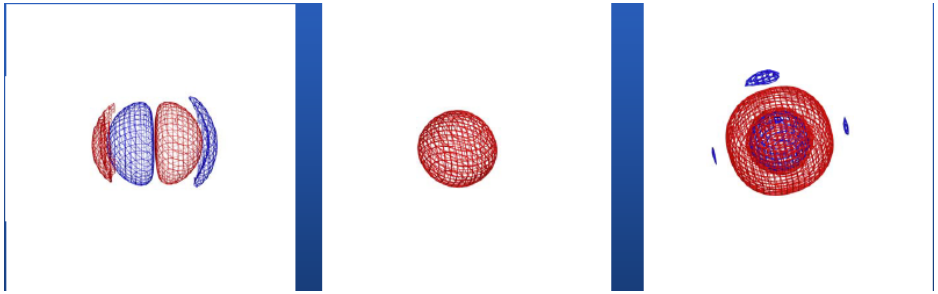
Simple scaling vs. Babinet scaling

$F_{model} = k_{overall}(F_{calc} + k_{mask}F_{mask})$ *vs.* $F_{total} = -k_{sol}F_{prot}\exp(-B_{sol}\sin^2\theta/\lambda^2)$

Bulk solvent mask artifacts can only occur at narrow channels, where the mask radius is too big > *negative* difference density
Remedy: changing from simple scaling to Babinet scaling (Refmac but „uncheck“ calculate contribution from the solvent region) and/or you can optimize the solvent mask parameters by running Refmac with the keyword "solvent optimize“ which determines optimum parameters

Errors in the e.d. maps





- Occupancy errors look like 1s orbitals
- B errors look like 2s orbitals
- Positional errors look like 2p orbitals
- Anisotropic B errors look like 3d orbitals

But only if you ignore series termination…

Negative difference density in a big hydrophobic cavity > wrongly determined low-resolution |Fobs|
Remedy: cut very low resolution data

Lang *et al.*, PNAS **111**, 237-242

Two problems with normal maps:

- maps are not on an absolute scale ($e^-/Å^3$)
- calculating noise at every point in map

Consequently:

- σ-scaled maps cannot be compared correctly
- at 1σ noise is overestimated (6-8 fold)
- information is ignored

SigmaA weighting (mFo-DFc)*

R. Read:
„the weighted difference map (mFo-DFc)* is much less affected by considerations of relative heights of peaks for included and missing atoms .... but look at 2mFo-DFc and 3mFo-2DFc maps to see which shows regions of error more clearly“

*Accounting for incomplete and/or erroneous models
m: Sim weights as a structure factor (later on, maximum likelihood by French/Bricogne/Read) probability distribution correction and further modified to account for errors in the model (sigma-A by Srinivasan and Ramachandran)

D: Luzzati factor as weighted average over the complex error distribution of Fcalc

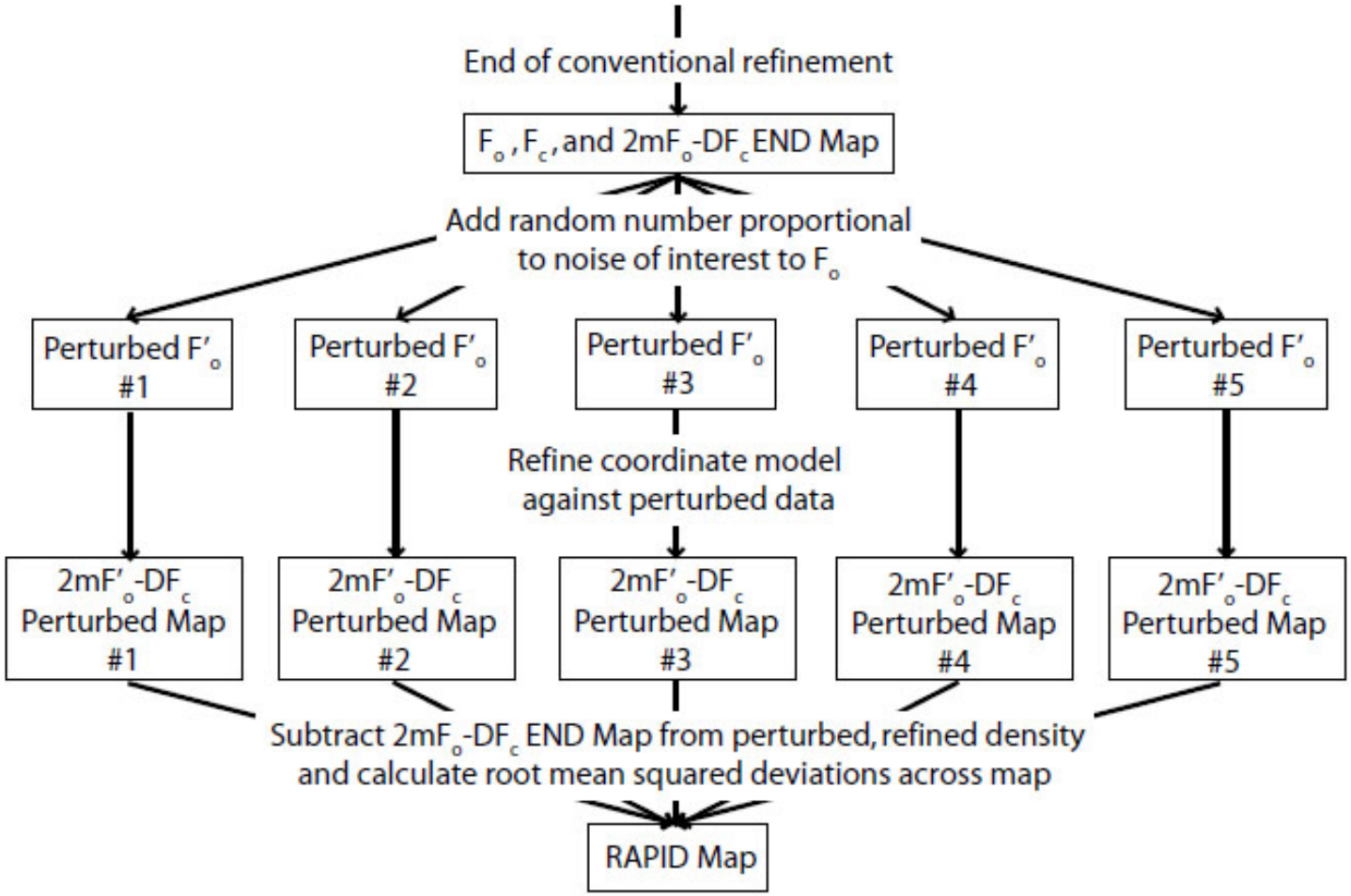
SigmaA weighting (mFo-DFc)

Dale Tronrud:
„the common practice of the field is to contour maps at 1 sigma and difference maps at ± 3 sigma“ but this habit throws away some significant information:
– Once you have built a model, you can calculate maps on an absolute scale (*i.e.* electrons/Å^3) - the absence of an atom should result in a peak of a particular height when expressed in e$^-$/Å^3, but will not be consistent when viewed in sigma's
– The sigma of a difference map will drop as refinement progresses, but that does not mean that peaks are becoming more significant
– The sigma of a density map will depend on the solvent fraction of the map calculated.

Simulated-annealing OMIT maps

Any atom that is located within the OMIT or border regions (2 Å) is given zero occupancy in all calculations, is however included in geometric restraint calculations and model building but do not contribute to the structure-factor calculation!

RAPID map



Refinement against perturbed input data (RAPID): introduce errors / simulated noise to Fobs in order to analyse spatial distribution of errors in map

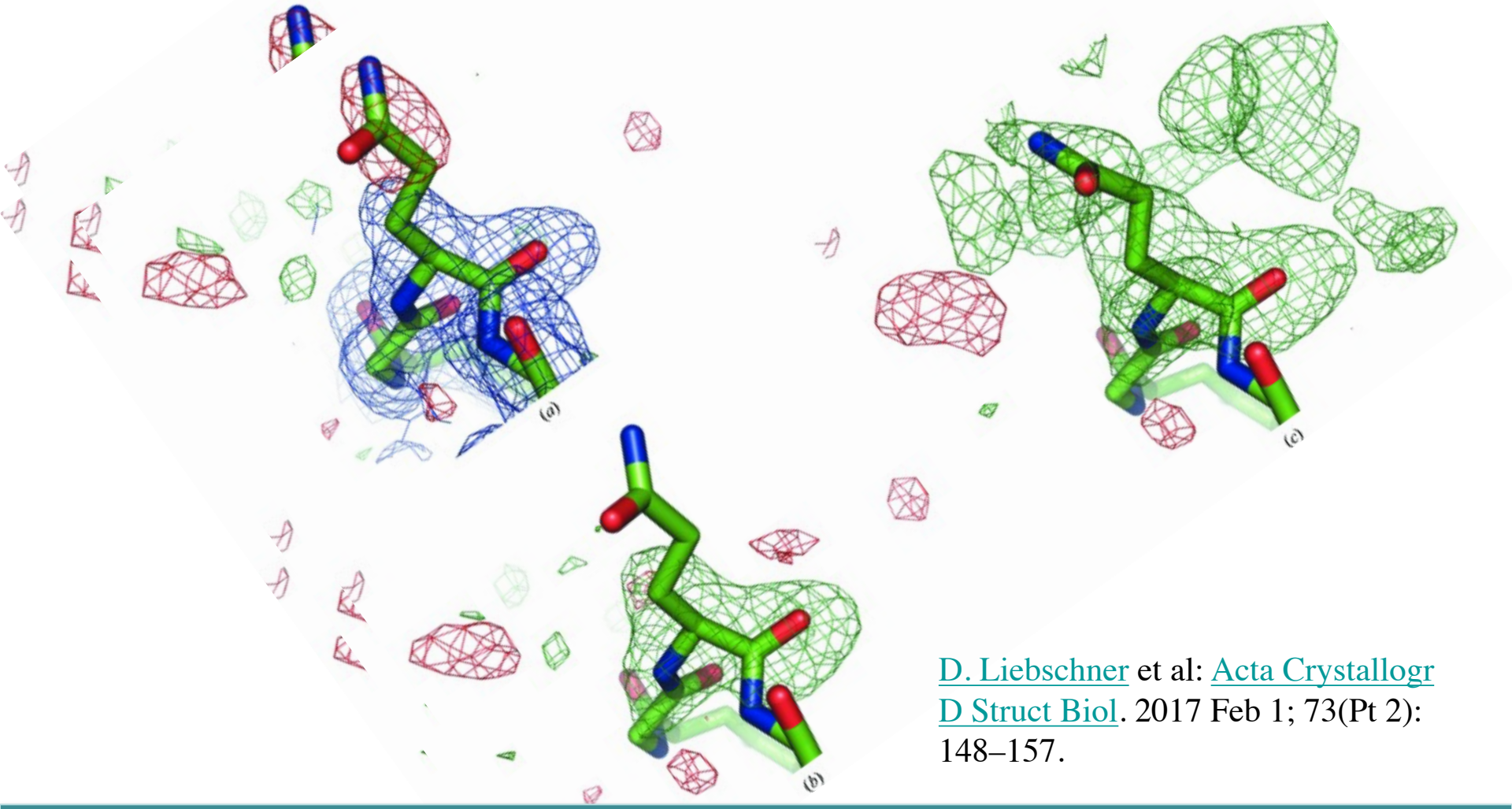
Praznikar et al., Acta **D**65, 921-931

AK maps (reduced model bias):

average of series of kicked (random modified atom positions) maps

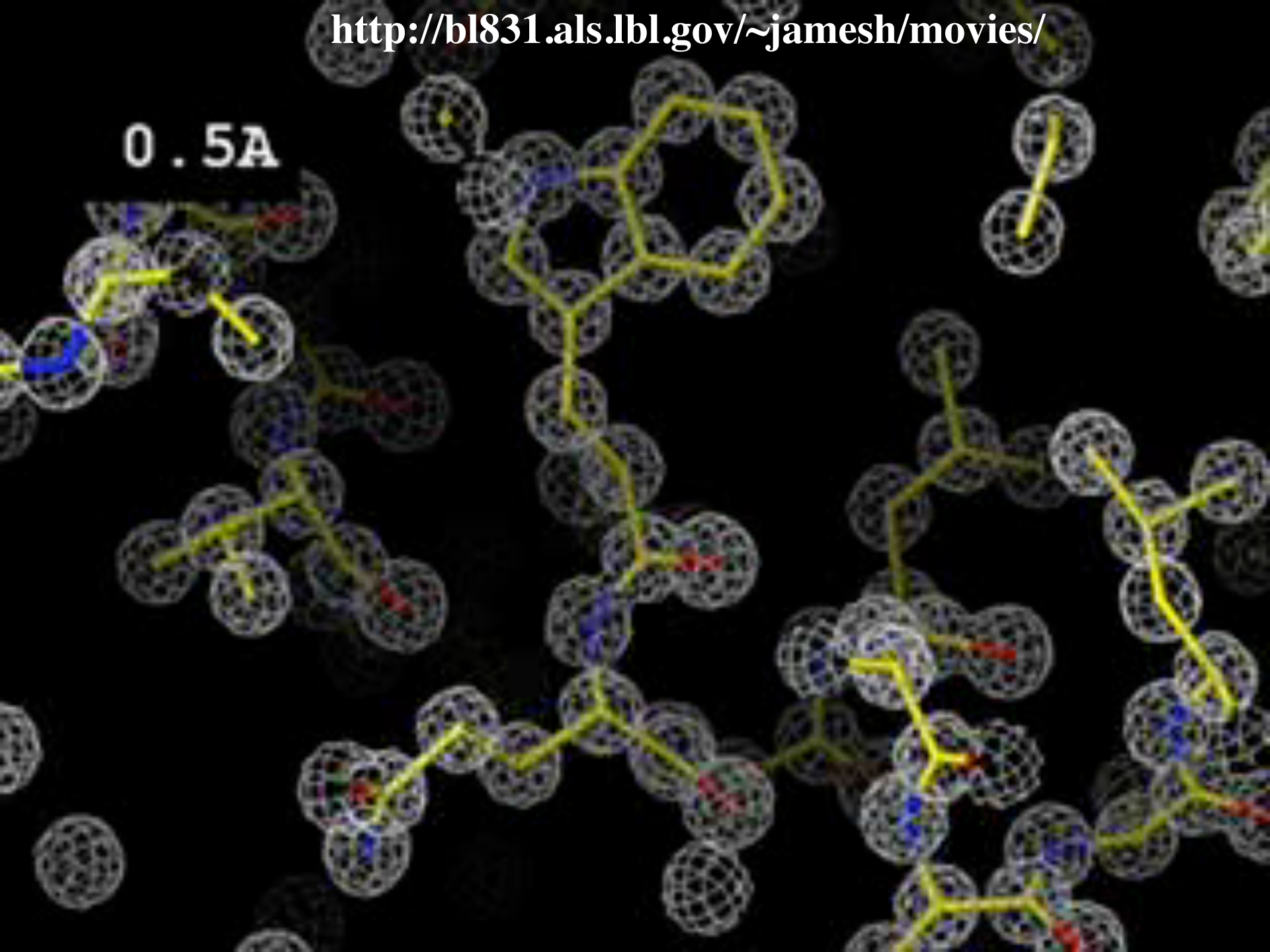
ML theory proposes that a current model can be corrected by introducing random errors and suggests structure-factor corrections after theoretical averaging of such models

Polder maps: improving OMIT maps by excluding bulk solvent

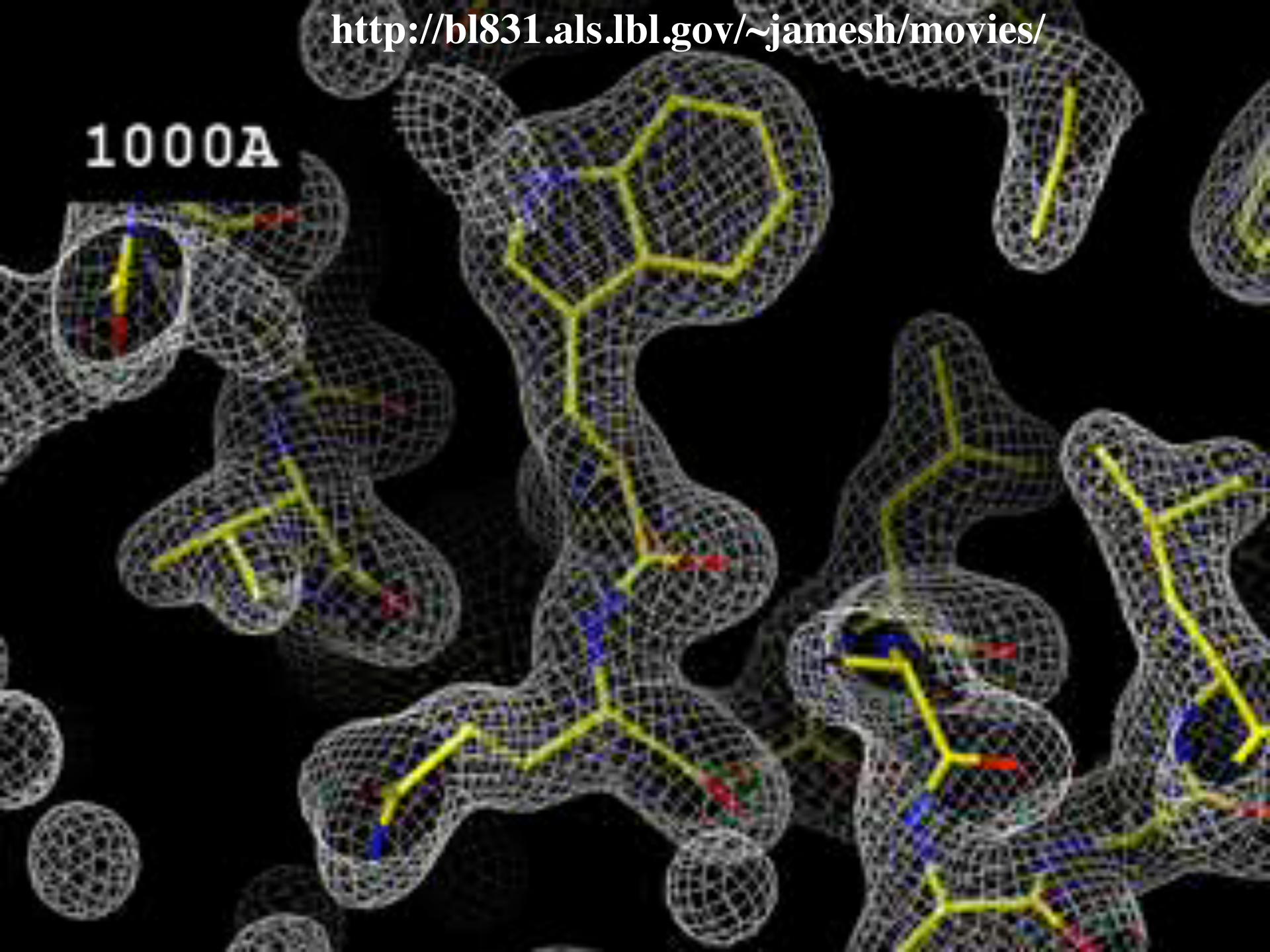


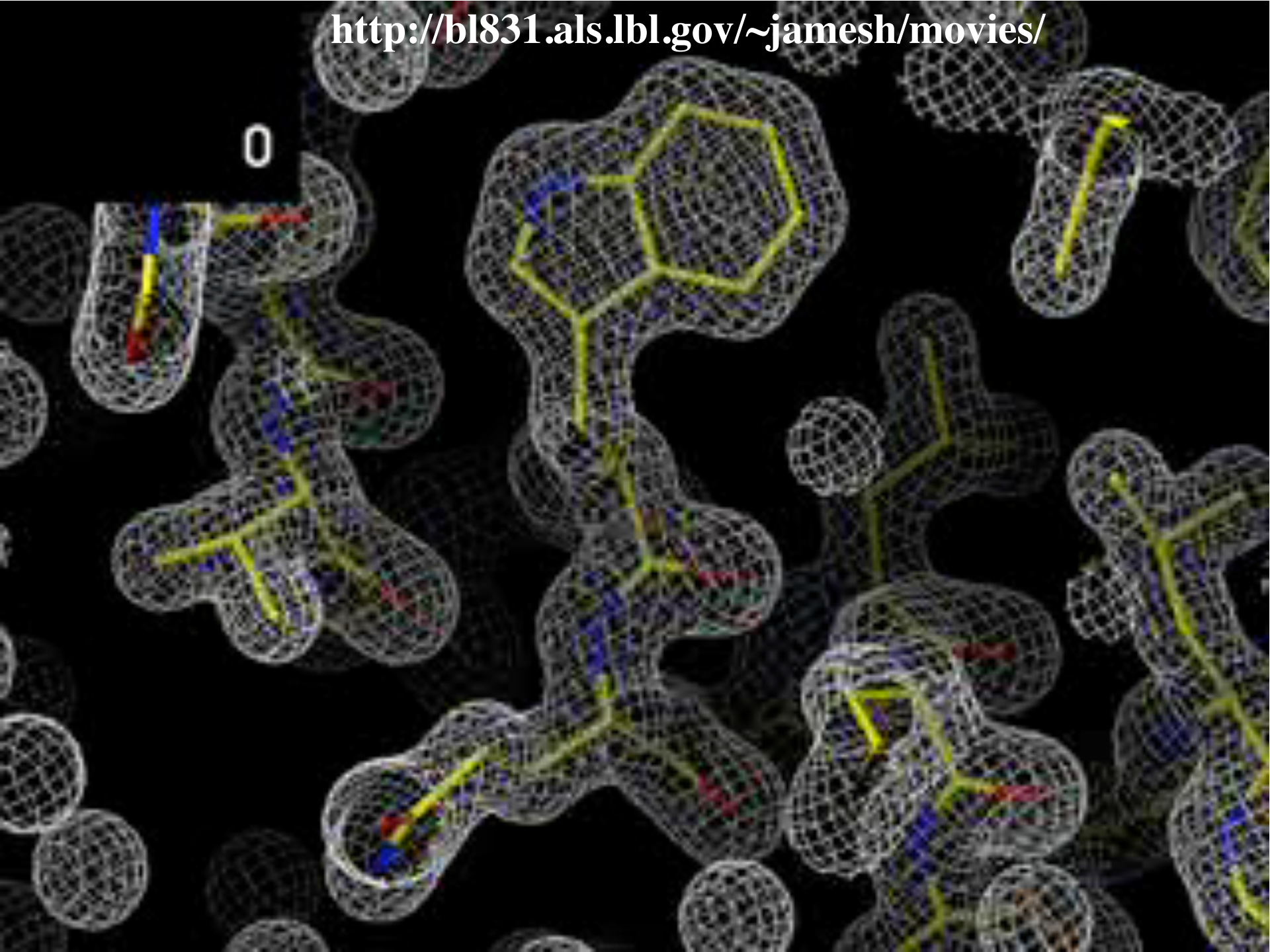
D. Liebschner et al: Acta Crystallogr D Struct Biol. 2017 Feb 1; 73(Pt 2): 148–157.

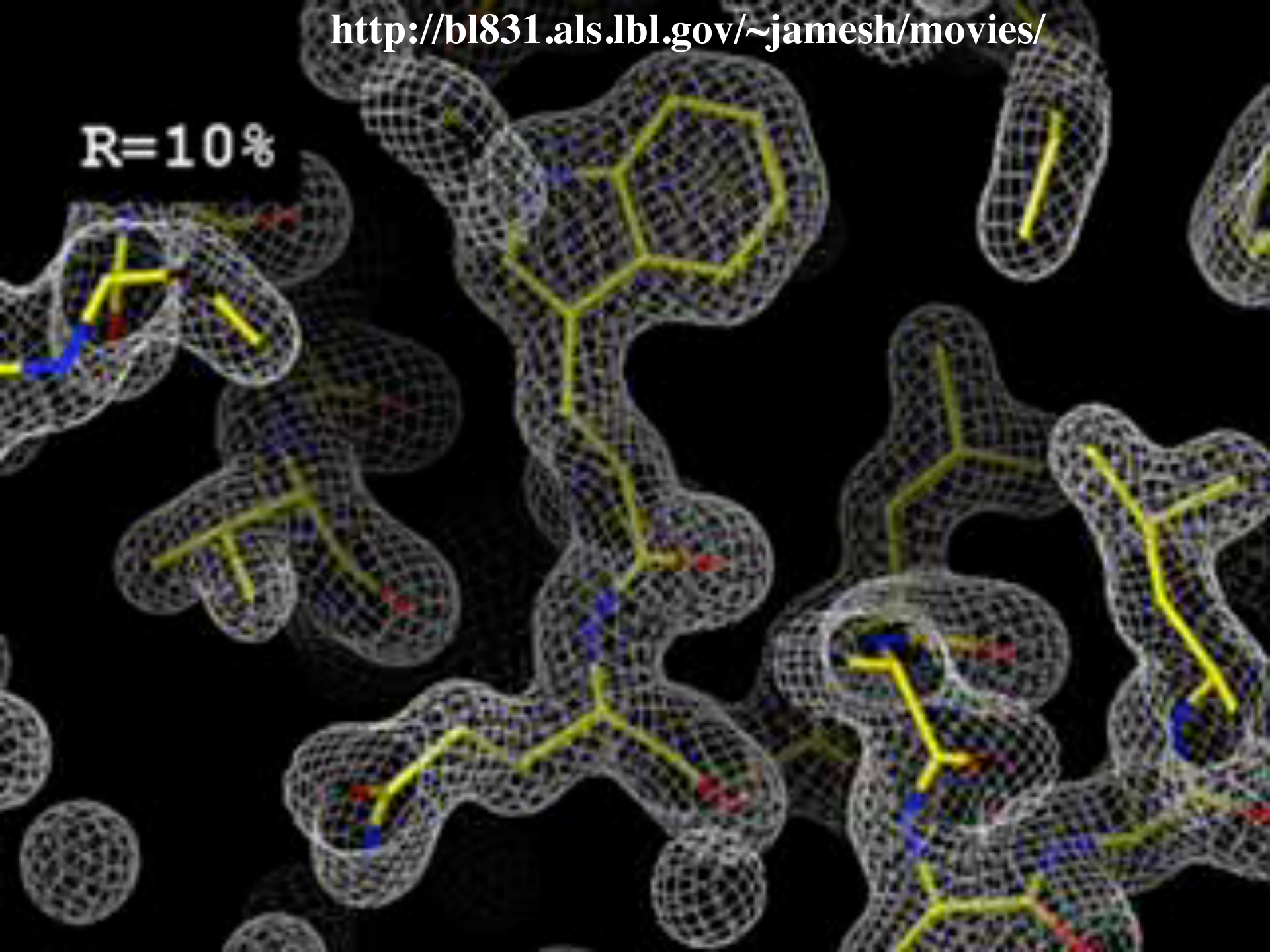
http://bl831.als.lbl.gov/~jamesh/movies/



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




http://bl831.als.lbl.gov/~jamesh/movies/
0


http://bl831.als.lbl.gov/~jamesh/movies/
R=10%