**Electron Density at various resolutions, and fitting a model as accurately as possible.**

$$\varrho_{xyz} = (\mathrm{Vol})^{-1} \sum_h \sum_k \sum_l m_{hkl} \bullet | F_{hkl} | \bullet e^{i\phi_{hkl}} \bullet e^{-i2\pi( hx + ky + lz )}$$

Amplitude (Amplitude-factors) • (Phase-factors)
(electron density has no phase, just use the real part of computed result.)

The x,y,z coordinates actually used in the calculaton are fractions of the unit cell edges a,b,c. This is in accord with the *h,k,l* actually representing spacings defined in terms of Bragg planes that cut the edges of the unit cell into integral fractions. Just as *h,k,l* defines directions $h \perp$ y,z ; $k \perp$ z,x ; $l \perp$ x,y perpendicular to the planes of the unit cell parallelepiped, the x,y,z directions are aligned with the edges of the unit cell. Thus the natural coordinate system of the model is (often) non-orthogonal and non-normalized. However, most graphics programs (and most people) work with ortho-normal Cartesian coordinates in standard units (usually Ångstroms). Since a,b,c can be expressed in Å at the known angles of the unit cell, Cartesian model atom coordinates are calculated from the size and shape of the unit cell.

**Experimental Phases, Model Phases**

When some sort of starting values of the phases are available, a model can be built into the electron density. This electron density is dependent on both the amplitudes and the phases of the *h,k,l* data points. The appearance of the image turns out to be most dependent on the phases, i.e. on the part not known directly from experiment, and thus quite susceptible to errors and misconceptions. Initially errors in deriving the starting phases, and later as the model itself is used to calculate phases, from errors and misconceptions about molecule.

When we know (some) of the coordinates of a model, we can use a Fourier transform of these to get calculated phases for each reflection in order to make a (hopefully) better electron-density image.

**A diffracted wave is the sum of contributions from all atoms.**

$$\mathbf{F}_{hkl} = | F_{hkl} | \bullet e^{i\phi_{hkl}} = \sum_n \bullet O_n \bullet f_{n,\theta} \bullet e^{-B_n(\sin\theta_n/\lambda)^2} \bullet e^{i2\pi(hx + ky + lz)}$$

**Residuals, R-values, assess agreement between datasets.**
(amplitudes F calculated from the model and F observed from the experiment)

$$R_{cryst} = \Sigma| F_{obs} - F_{calc} | / \Sigma| F_{obs} |$$

But since $R_{cryst}$ can be forced to appear good by warping the model -- the model must also be evaluated by other criteria. Also, a small subset (e.g. 5%) of the data is now standardly withheld from the refinement process. These data points can be used to calculate an $R_{free}$ which should get better as $R_{cryst}$ gets better as long as the model changes are really an improvement toward matching what the molecule really is.

**$R_{free}$ is a very valuable control against over-fitting.**
**$R_{free}$ is usually just a few percentage points greater than $R_{cryst}$**

---

# Refinement

$$\mathbf{F}_{hkl} = | F_{hkl} | \bullet e^{i\phi_{hkl}} = \sum_n \bullet O_n \bullet f_{n,\theta} \bullet e^{-B_n(\sin\theta_n/\lambda)^2} \bullet e^{i2\pi(hx + ky + lz)}$$

(Note: equations at the bottom of section "3.2 Molecular Scatter" where symbols are defined.)

Model improvement involves (re)building a model into the electron density (real space "refinement"), and shifting parameters to improve the fit of the calculated "structure factors" (data) to the observed data (reciprocal space refinement -- what is commonly called "Refinement Cycles" since the relationship of parameters to data is non-linear and is matched through a series of successive approximations).

The main target is to match the calculated |F| with that observed.

The parameters are $x_n$ $y_n$ $z_n$ $B_n$
For macromolecules there are essentially only these four parameters.
$B_n$ is the place where all errors are lumped, including, of course, uncertainty in location e.g. from motion of the atom.
(At the best obtainable resolutions, sometimes rather than the single isotropic B factor, it is possible to use a 6 parameter aniostropic expression.
$O_n$ (occupancy) is usually presumed to be 100% except where there is evidence of alternate conformations or the genuine partial occupancy of ions or ligands).

The molecule has a certain number of atoms, which sets the number of parameters, but the number of data points increases by the volume of reciprocal space that is measured. So at poorer resolutions there is a problem of numbers of parameters with respect to the number of data points.

Geometrical target functions can be defined based on “previous knowledge” like bond lengths and bond angles. Now many of the criteria used to validate final structural models are checked as refinement is done. There is a compromise between making the best possible model and being able to fairly validate the final model. However, meeting all known criteria is so complicated and difficult that the meeting all validation criteria is still a very strict accounting.

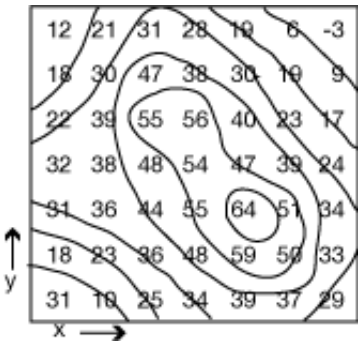
Even at the best (highest) (smallest-value) resolution, there can be regions where the electron density is weak and geometrical target functions are needed.

Refinement balances fit to data and fit to stereo-chemistry.

# Electron Density Maps

Now that we know the $\varphi_P$ for each reflection, we can use the Fourier transform formula to get a picture of the electron density in the molecule:

$$\varrho_{xyz} = (\mathrm{Vol})^{-1} \sum_h \sum_k \sum_l m_{hkl} \bullet | F_{hkl} | \bullet e^{i\phi_{hkl}} \bullet e^{-i2\pi( hx + ky + lz )}$$

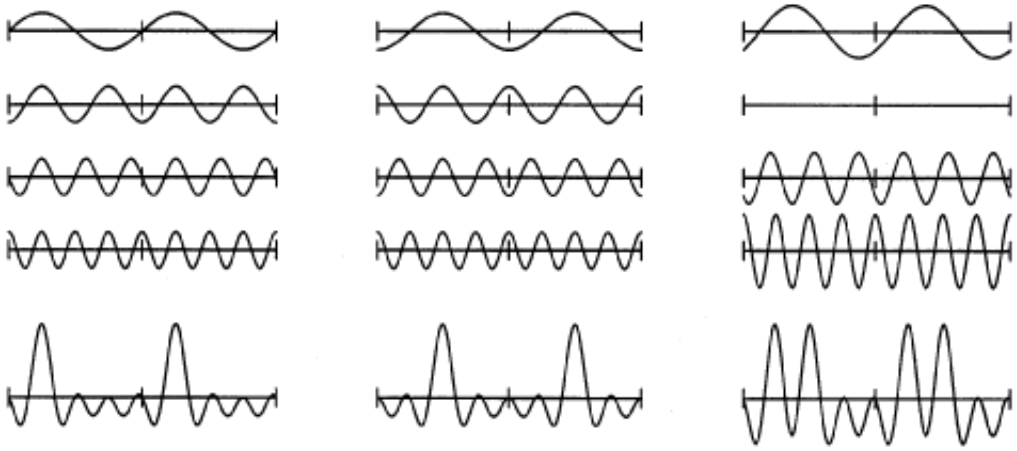



For each chosen grid point x,y,z in the repeating unit of the crystal, the above expression must be summed over all the measured reflections *h, k, l*. The top illustration shows part of one layer of such a map, where each number is the value of the electron density (on an arbitrary scale) at that grid point. Contours have been drawn at 10, 20, 30, etc.

The middle illustration shows 7 superimposed layers of a part of the 2.5 Å resolution map of staphylococcal nuclease .

In the bottom illustration this same piece of map is shown interpreted as two residues of backbone in extended chain conformation, with arginine and phenylalanine side groups.

Example of summing waves of 1 , 2 , 3 , 4 wavelengths across a unit cell, of different amplitudes and different relative phases to get different number and positions of reconstructed “atoms”.



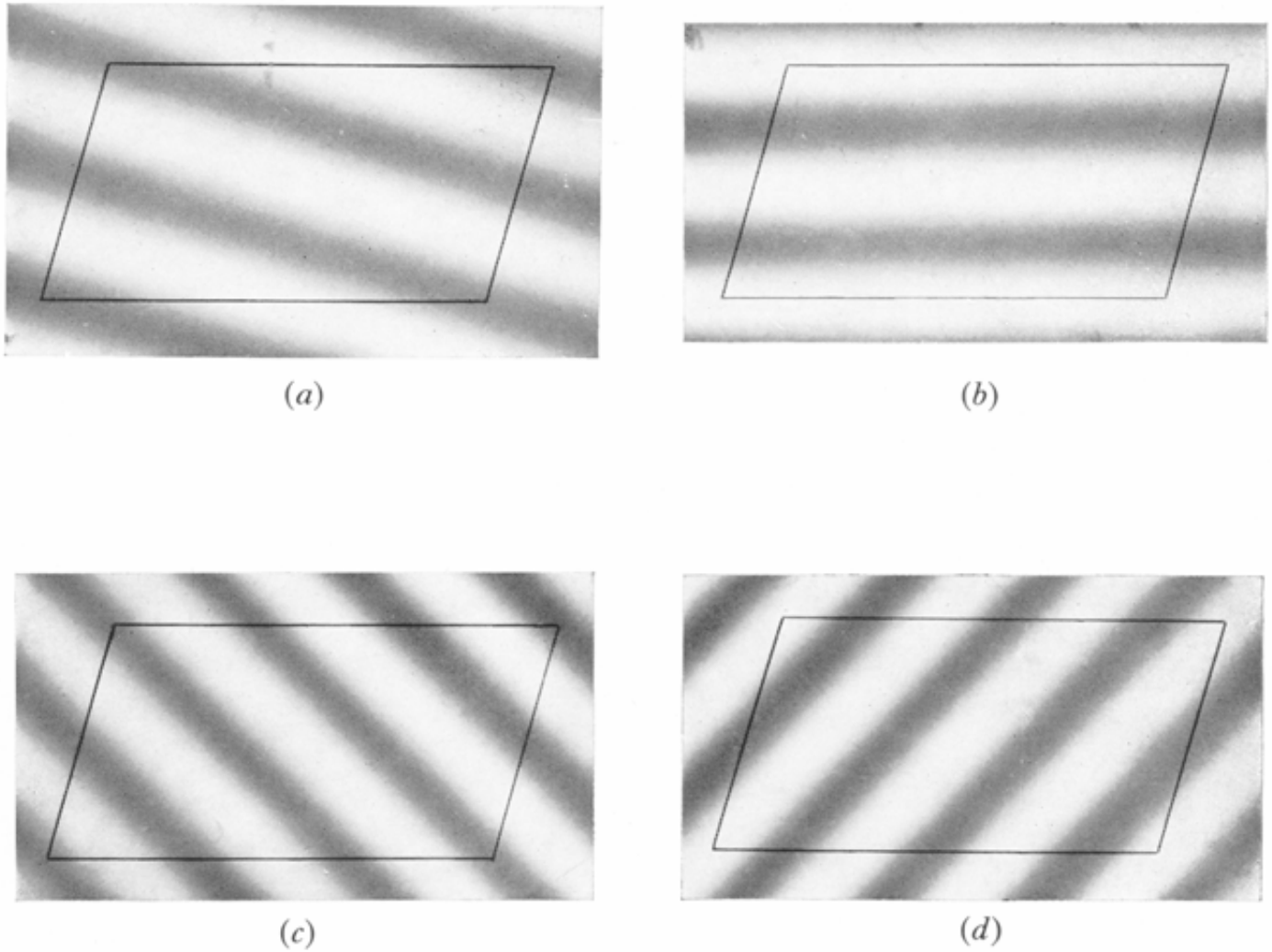
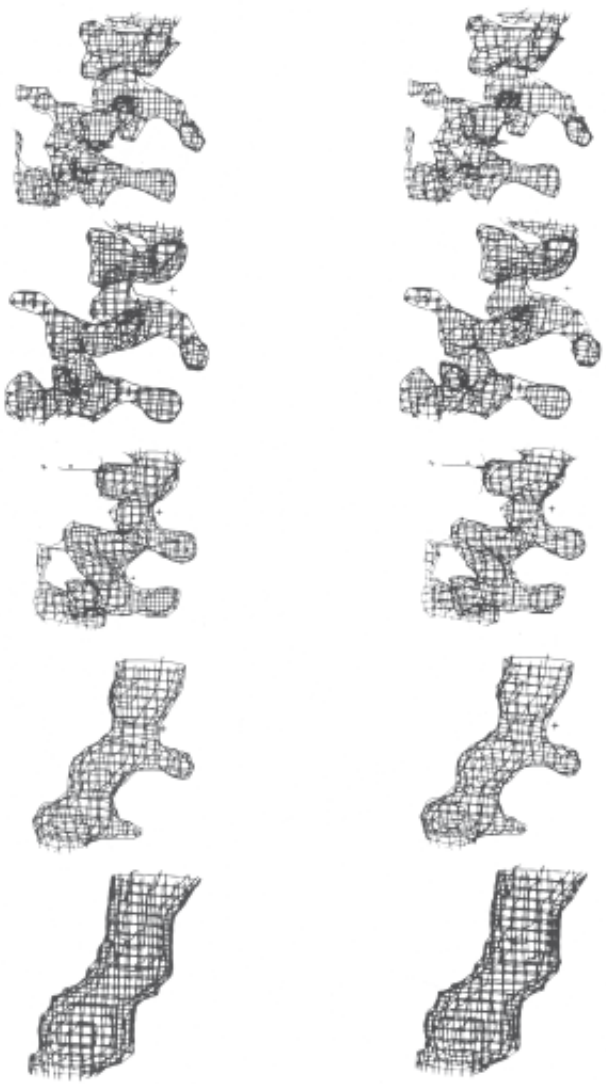



FIG. 138.—Sinusoidal alternations of light and shade. The bands in the figure represent the contributions to the image due to the following spectra:

(*a*) F(102), phase negative
(*b*) F(002), phase positive
(*c*) F(302), phase negative
(*d*) F(30$\bar{1}$), phase positive

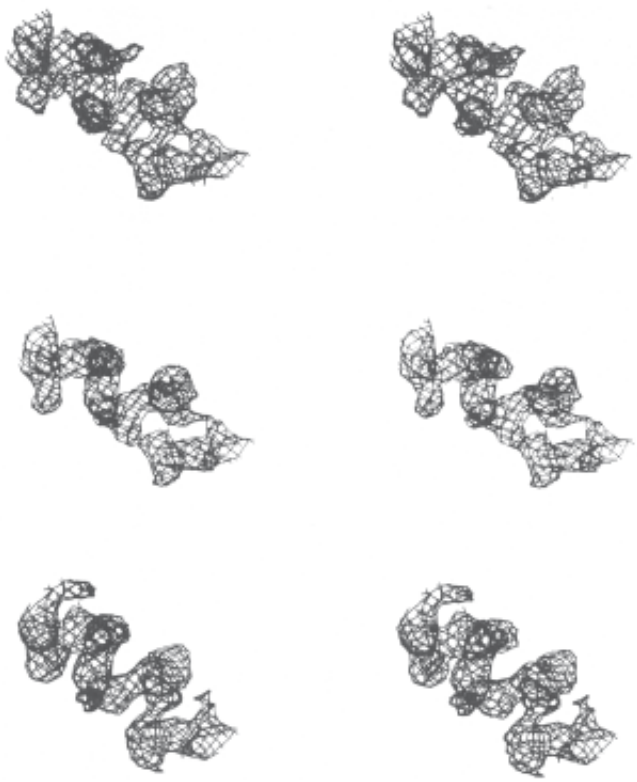
(*Zeit. f. Krist.*, **70**, 483, 1929)



Stereo views of the electron density for an α-helix in staphylococcal nuclease at (from top to bottom) 2, 3, 4, 5, and 6 Å resolution. All maps were made using $F_{obs}$, the original MIR phases, and the same grid spacing. Viewpoint is the same, and contour levels were adjusted to be approximately equivalent.

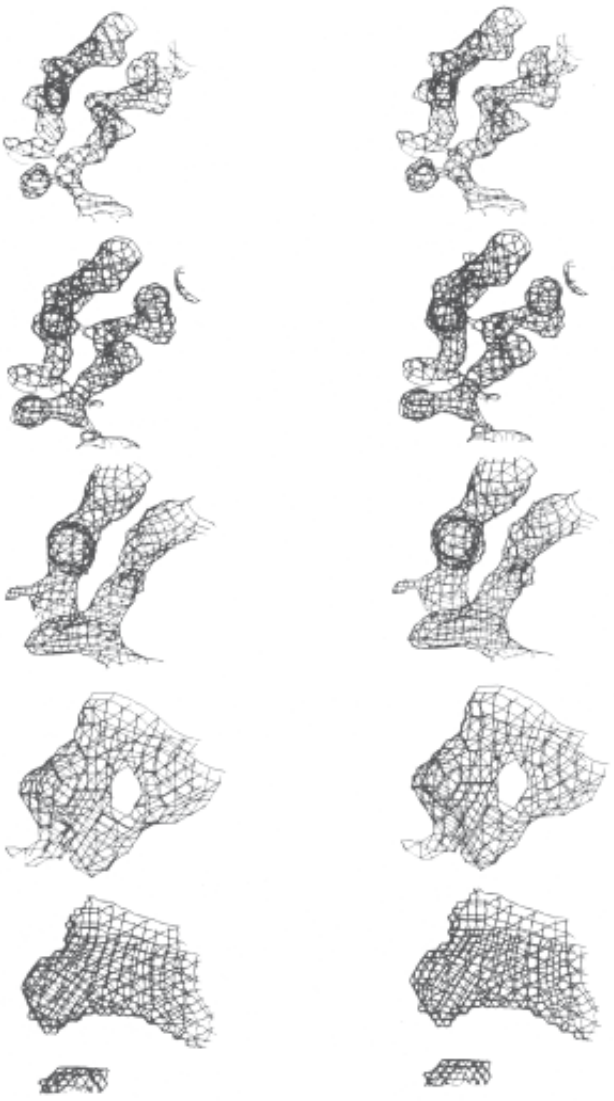
All carbonyl oxygens are clear at 2 Å, but almost all of them are absent at 3 Å, although side chains can still be judged. At 4 Å, density has begun to coalesce along the helix axis, and there is a false connection between side chains at the lower left.

This and the next two figures are from JS and DC Richardson (1985) "Interpretation of Electron Density Maps" in Methods in Enzymology; HW Wyckoff, CHW Hirs, SN Timasheff, eds.; 115: 189-206.



Side views of the same helix as on previous page, at 2, 3, and 3.5 Å resolution.

At intermediate resolution the density connects through a hydrogen bond (lower right) more strongly than through the nearby helical main chain, although the connectivity is correct at both higher and lower resolutions.



Stereo views of the electron density for two strands of antiparallel β-sheet in staphylococcal nuclease at 2, 3, 4, 5, and 6 Å resolution.

In this case the strands separate correctly at 4 Å but that would not always be true. At 5 and 6 Å the density is sheetlike, but with holes in variable locations. At 6 Å the right-hand side extends further out because it is no longer separated from a third strand.