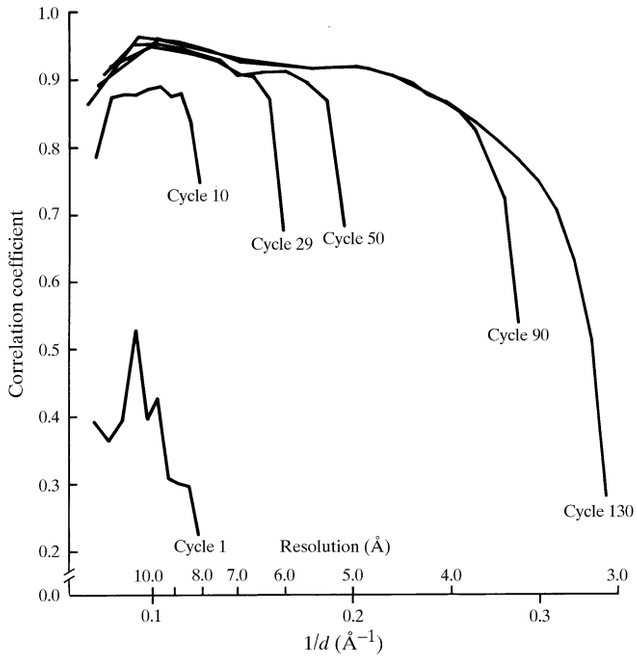

Fig. 13.4.11.1. Plot of a correlation coefficient as the phases were extended from 8 to 3 Å resolution in the structure determination of Mengo virus. [Reproduced with permission from Luo *et al.* (1989). Copyright (1989) International Union of Crystallography.]

(1986, 1988) discussed phase error as a function of error in the NCS definition and applied rigid-body least-squares refinement for refining particle position and orientation of human rhinovirus 14. The 'climb' procedure has been found especially useful (Muckelbauer *et al.*, 1995). This depends upon searching one at a time for the parameters (rotational and translational) that minimize the near r.m.s. deviation of the individual densities to the resultant averaged densities.

Improvement of the NCS parameters is dependent upon an accurate knowledge of the cell dimensions. In the absence of such knowledge, the rotational NCS relationship cannot be accurate, since elastic distortion will result, leading to very poor averaged density. This was the case in the early determination of southern bean mosaic virus (Abad-Zapatero *et al.*, 1980), where the structure solution was probably delayed at least one year due to a lack of accurate cell dimensions.

Another aspect to phase extension is the progressive decrease in or quality of observed structure amplitudes. The observed amplitudes can be augmented with the calculated values obtained by Fourier back-transformation of the averaged map. However, clearly, as the number of calculated values increases in proportion to the number of observed values, the rate of convergence decreases. In the limit, when there are no available $F_{\rm obs}$ values, averaging a map based on $F_{\rm calc}$ values will not alter it, and, thus, convergence stops entirely.

13.4.11. Convergence

Iterations consist of averaging, Fourier inversion of the average map, recombination of observed structure-factor amplitudes with calculated phases, and recalculation of a new electron-density map. Presumably, each new map is an improvement of the previous map as a consequence of using the improved phases resulting from the map-averaging procedure. However, after five or ten cycles, the procedure has usually converged so that each new map is essentially the same as the previous map. Convergence can be usefully measured by computing the correlation coefficient (*CC*) and *R* factor (*R*) between calculated ($F_{\rm calc}$) and observed ($F_{\rm obs}$) structure-factor amplitudes as a function of resolution (Fig. 13.4.11.1). These factors are defined as

$$CC = \frac{\sum_h (\langle F_{\rm obs}\rangle - F_{\rm obs})(\langle F_{\rm calc}\rangle - F_{\rm calc})}{\left[\sum_h (\langle F_{\rm obs}\rangle - F_{\rm obs})^2 (\langle F_{\rm calc}\rangle - F_{\rm calc})^2\right]^{1/2}},$$

$$R = 100 \times \sum |(|F_{\rm obs}| - |F_{\rm calc}|)| \Big/ \sum |F_{\rm obs}|.$$

Because of the lack of information immediately outside the resolution limit, these factors must necessarily be poor in the outermost resolution shell. Nevertheless, the outermost resolution shell will be the most sensitive to phase improvement as these structure factors will be the furthest from their correct values at the start of a set of iterations after a resolution extension.

Convergence of *CC* and *R* does not, however, necessarily mean that phases are no longer changing from cycle to cycle. Usually, the small-amplitude structure factors keep changing long after convergence appears to have been reached (unpublished results). However, the small-amplitude structure factors make very little difference to the electron-density maps.

The rate of convergence can be improved by suitably weighting coefficients in the computation of the next electron-density map. It can be useful to reduce the weight of those structure factors where the difference between observed and calculated amplitudes is larger than the average difference, as, presumably, error in amplitude can also imply error in phase. Various weighting schemes are generally used (Sim, 1959; Rayment, 1983; Arnold *et al.*, 1987; Arnold & Rossmann, 1988).

As mentioned above, the rate of convergence can also be improved by inclusion of $F_{\rm calc}$ values when no $F_{\rm obs}$ values have been measured. However, care must be taken to use suitable weights to ensure that the $F_{\rm calc}$'s are not systematically larger or smaller than the $F_{\rm obs}$ values in the same resolution range.

Monitoring the *CC* or *R* factor for different classes of reflections (*e.g.* $h+k+l=2n$ and $h+k+l=2n+1$) can be a good indicator of problems (Muckelbauer *et al.*, 1995), particularly in the presence of pseudo-symmetries. All classes of reflections should behave similarly.

The power (*P*) of the phase determination and, hence, the rate of convergence and error in the final phasing has been shown to be (Arnold & Rossmann, 1986) proportional to

$$P \propto (Nf)^{1/2}/[R(U/V)],$$

where *N* is the NCS redundancy, *f* is the fraction of observed reflections to those theoretically possible, *R* is a measure of error on the measured amplitudes (*e.g.* $R_{\rm merge}$) and U/V is the ratio of the volume of the density being averaged to the volume of the unit cell. Important implications of this relationship include that the phasing power is proportional to the square root of the NCS redundancy and that it is also dependent upon solvent content and diffraction-data quality and completeness.

13.4.12. *Ab initio* phasing starts

Some initial low-resolution model is required to initiate phasing at very low resolution. The use of cryo-EM reconstructions or available homologous structures is now quite usual. However, a phase determination using a sphere or hollow shell is also possible. In the case of a spherical virus, such an approximation is often very reasonable, as is evident when plotting the mean intensities at low

references