

9 Phase Improvement by Non-Crystallographic Symmetry Averaging with Phase Extension

9.1 The Method of Non-Crystallographic Symmetry Averaging and Phase Extension

Proteins sometimes crystallise with several (almost) identical subunits in the asymmetric unit. These subunits are related by non-crystallographic symmetry (NCS). Averaging over NCS has played an essential role in the determination of spherical virus structures such as the tomato bushy stunt virus (Winkler *et al.*, 1977; Harrison *et al.*, 1978), the Mengo virus (Luo *et al.*, 1987) and the foot-and-mouth disease virus (Acharya *et al.*, 1989). Some virus crystals contain as many as sixty copies of the viral-coat protein in the asymmetric unit. The subunits related by non-crystallographic symmetry may have, however, slight differences due to different crystal environments as there are no crystallographic restraints which force identity. This redundancy can be used to improve the map by averaging the copies of the equivalent subunit densities. Analogously, intensity data collected from such structures exhibit redundancy. This redundancy imposes a constraint on the protein structure amplitudes and phase angles (Rossmann & Blow, 1962).

Thus, it is possible to improve the electron density by averaging and solvent flattening outside the molecular boundary. Phases obtained from Fourier transforming the resulting electron density are generally more accurate than the original phases. A new improved map can be obtained by using the observed structure factors with the improved phases. The procedure is repeated

until convergence is achieved. After a cycle of averaging, the new map contains higher resolution information than the unaveraged map. This procedure can therefore also be used to extend phasing beyond the resolution of the original map (Rossmann, 1992). Rossmann & Blow (1963), Main (1967) and Crowther (1969) proposed a reciprocal space procedure to derive NCS phase relationships; the amounts of computation was proportional to the square of the number of reflections. At the time, computing resources imposed limits of a few thousand reflections. Bricogne (1974) developed an equivalent approach in real space which involved the determination of the molecular envelope in the asymmetric unit and the following iteration procedure: calculation of an electron density map and of figures of merit from the current observed structure factors; averaging of the electron densities of all crystallographically independent subunits and setting the electron density outside the molecular boundaries to an average solvent value; rebuilding of the asymmetric unit from this averaged aggregate; obtaining the phase angles for the new model by back-transforming the electron density map; combination with any heavy-atom phase information using Sim's formula to estimate figures of merit (Bricogne, 1976). The amount of computation involved is only proportional to the number of grid points. The averaged structure is regarded as the "known" part of the structure in the Sim formula (Equation 14). The Bricogne algorithm has been shown to converge rapidly if more than three subunits are present in the asymmetric unit. One needs to know precisely the relationship between the subunits in the asymmetric unit. This could be a perfectly symmetric arrangement or there might be imperfect rotational and translational components. A rotation

function analysis allows the orientation of the NCS axes to be determined (see Chapter 6).

The position of the molecular symmetry axes may often be derived from the positions of a heavy-atom derivative, as long as the heavy-atom binds at equivalent sites on each protein subunit and does not lie on a special position (e.g. on the NCS symmetry axes). Here, the self-rotation indicated 222 symmetry and the intersection of these local axes could be determined from the PIP heavy-atom positions. If the operations between subunits are known, the molecular envelope can be identified from an electron density map with the initially available phases by globally averaging using the local symmetry operators. Beyond the molecular boundaries, the electron density is not expected to be correlated and an initial envelope can be estimated (Bricogne, 1976).

In the procedure of phase extension, phases are generated for a few unphased reflections in a slightly higher resolution shell usually only one or two reciprocal lattice units beyond the resolution used to calculate the present map. The averaged electron density map contains higher-resolution components because it is calculated on an over-sampled grid and the electron density at each grid point has been replaced by the average of others or is set to zero if it lies outside the molecular envelope (Lawrence, 1991).

The averaging iterations are repeated until convergence before extending the resolution. The procedure can be enhanced by weighting the structure factors used in the calculation of the electron density map using the weighting

function (Rayment, 1983)

$$w = \exp\left(-\frac{||F_{calc}| - |F_{obs}||}{|F_{obs}|}\right) \quad (15)$$

where $|F_{calc}|$ and $|F_{obs}|$ are the calculated and observed amplitudes. This function attaches a high weight to reflections with calculated amplitudes close to their observed amplitudes.

9.2 NCS Averaging with Phase Extension Protocol Application to Neu5Ac Lyase

The programs used here for symmetry averaging and phase extension were Bricogne's original averaging and phase extension programs (Bricogne, 1976) modified by Lawrence (1991) to employ a ten-point (fully quadratic) interpolation algorithm and with the double-sorting algorithm replaced by entirely in-core procedures. The original calculated SIRAS map was packed into two byte integers by the program `SHRINK`, which also closes the edges of the map by unit cell translations. The resulting map was averaged using the program `QIKAVERAGE`, which is Lawrence's modified version of Bricogne's programs `GENERATE`, `SORT1`, `INTERPOL`, `SORT2` and `RECONST`. The averaging program `QIKAVERAGE` first generates the mapping between grid coordinates in the initial and final maps (corresponding to Bricogne's program `GENERATE`). For interpolating accurately, the grid sampling needs to be of the order of $\frac{1}{3}$ to $\frac{1}{5}$ of the final resolution (2.5 Å here) and was set to 0.5 Å. The program `QIKAVERAGE` requires here an envelope [determined by Leslie's (1987b) modified reciprocal space procedure as discussed in the previous chapter] and the transformation matrices Q , relating the intermediate orthonormal frame to the crystallographic frame, which for the trigonal space group is given by

$$Q = \begin{pmatrix} 1 & \tan 30^\circ & 0 \\ 0 & \frac{1}{\cos 30^\circ} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (16)$$

and P describing the orientation of the skew grid relative to the intermediate orthonormal frame; P is a 3 x 3 matrix obtained from the direction cosines (see

equation 6, page 54) of the three orthogonal two-fold axes describing the non-crystallographic symmetry. The conversion from the crystallographic frame into the skewed frame can be performed using

$$R = QPU \quad (17)$$

and

$$t = QPV + X \quad (18)$$

where (U, V) is the rotation matrix and translation vector of the averaging operation and X the coordinates in the crystallographic frame. For lyase, the electron density of the four independent subunits in the asymmetric unit was averaged and the solvent density outside the envelope was set to the average solvent density. The asymmetric unit of electron density was rebuilt from this averaged aggregate and the program `SHRINK` was again applied to the new map. The program `QIKGENERATE` generates the unit cell from the asymmetric unit (where the information about the envelope is not needed here) and this is then back transformed to give calculated structure factors by Bricogne's program `BFFTP1`. This step was necessary due to the absence of a space group specific `FFT` program. The grid sampling was set to 1 Å. The phases were not combined with the `SIRAS` phases as suggested by Bricogne (1976) since the phases calculated by the extension and averaging process were anticipated to be more accurate than the ones obtained from isomorphous replacement. The structure factors were weighted by Bricogne's program `NEWCOMB` applying Rayment's (1983) weighting function. This procedure was iterated for eight cycles before the resolution was increased by one reciprocal lattice unit of the

longest cell dimension. Unmeasured structure factors were included in the electron density calculation by substituting these with appropriately scaled calculated structure factor values (Rayment, 1983; Lawrence, 1991). Phases were extended from the initial 6 Å SIRAS set in 46 shells of 0.005 \AA^{-1} out to 2.5 Å resolution, with eight cycles of averaging about the local 222 axes after each resolution increment.

9.3 Results

The origin and orientation of the local molecular axes were improved before the NCS averaging and phase extension procedure by an R -factor ($R = \frac{\sum ||F_{calc}| - |F_{obs}||}{\sum |F_{obs}|}$) grid search (in steps of 0.3 Å and 1°, respectively) at 6 Å resolution which led to interpolated changes of 0.27 Å and 1.16° respectively. The R -factor decreased from 0.216 to 0.157. The refined non-crystallographic symmetry operator [the P matrix which corresponds to the Eulerian angles (see Figure 11, page 53, illustrating the Eulerian conventions) $\theta_1 = 218.20^\circ, \theta_2 = 71.31^\circ, \theta_3 = 352.11^\circ$ centered at \vec{t}] was

$$\begin{pmatrix} -0.80559 & -0.57801 & -0.13007 \\ 0.08844 & -0.33440 & 0.93827 \\ -0.58583 & 0.74436 & 0.32051 \end{pmatrix}$$

$$\vec{t} = (0.53340, 0.75556, 0.14620)$$

for space group $P3_121$.

The averaging and phase extension was performed in the space group $P3_121$. Subsequently, it became obvious from features in the averaged electron density map at 2.5 Å resolution that helical segments were left-handed coils and that the correct space group is therefore $P3_221$. As noted in chapter 7 (section 7.4), the contribution of anomalous scattering to the phasing is low, and the SIRAS maps in the + and - hand are approximately mirror images of each other. In this case inclusion of anomalous scattering with the wrong hand has been rescued by the solvent flattening and NCS averaging procedures. A 10 Å thick slab of the final map obtained from the averaging and phase ex-

R -factor^a, Correlation coefficient^b, Relative lack of closure^c

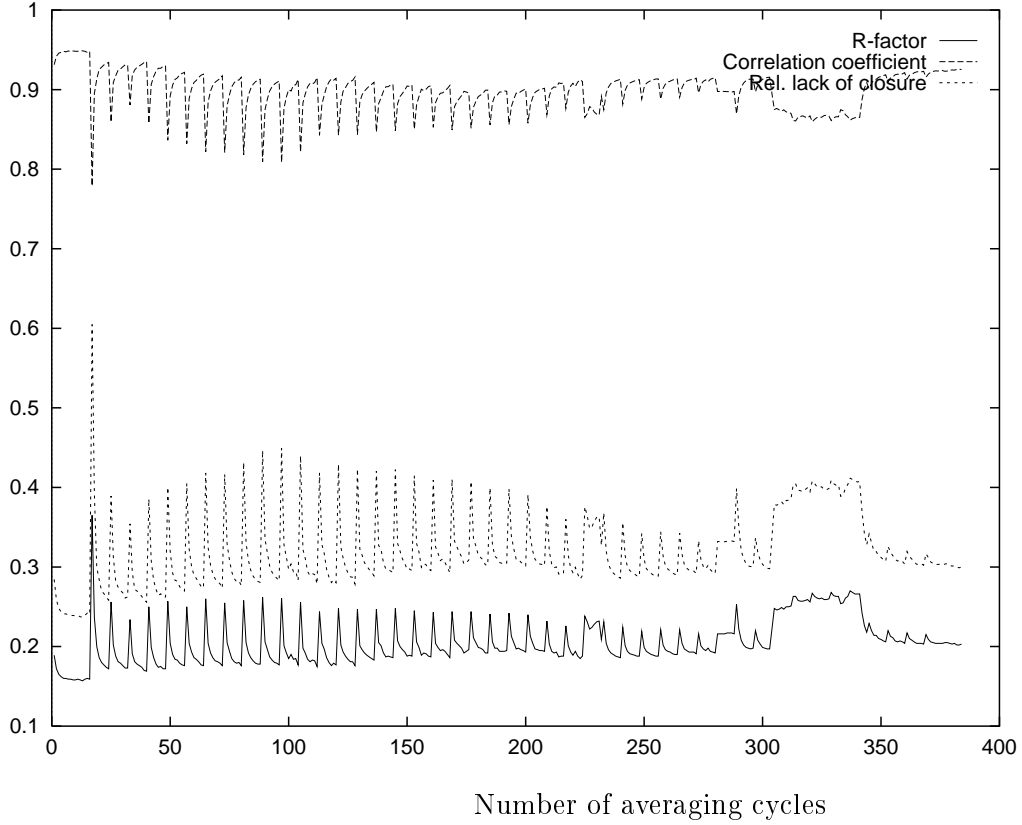


Figure 19: Variation of the R -factor^a, correlation coefficient^b and relative lack of closure^c as a function of averaging and phase extension cycles. After the first 16 averaging cycles about the local 222 axes at 6 Å resolution phases were extended in 46 shells of 0.005 Å⁻¹ out to 2.5 Å resolution, with eight cycles of averaging about the local 222 axes after each resolution increment. The unexpected increase in the R -factor^a and in the relative lack of closure^b and the decrease in the correlation coefficient^c between cycles 313 and 341 is unexplained and because of the great amount of computer resources needed during the NCS averaging and phase extension has not been investigated further.

$$^a R\text{-factor} = \frac{\sum_h ||F_{h,obs}| - |F_{h,calc}||}{\sum_h |F_{h,obs}|}$$

$$^b \text{ correlation coefficient} = \frac{\langle F_{obs} F_{calc} \rangle - \langle F_{obs} \rangle \langle F_{calc} \rangle}{\sqrt{(\langle F_{obs}^2 \rangle - \langle F_{obs} \rangle^2)(\langle F_{calc}^2 \rangle - \langle F_{calc} \rangle^2)}}$$

$$^c \frac{\sum ||F_{obs}^2| - |F_{calc}^2||}{\sum |F_{obs}|^2}$$

R -factor^a, Correlation coefficient^a

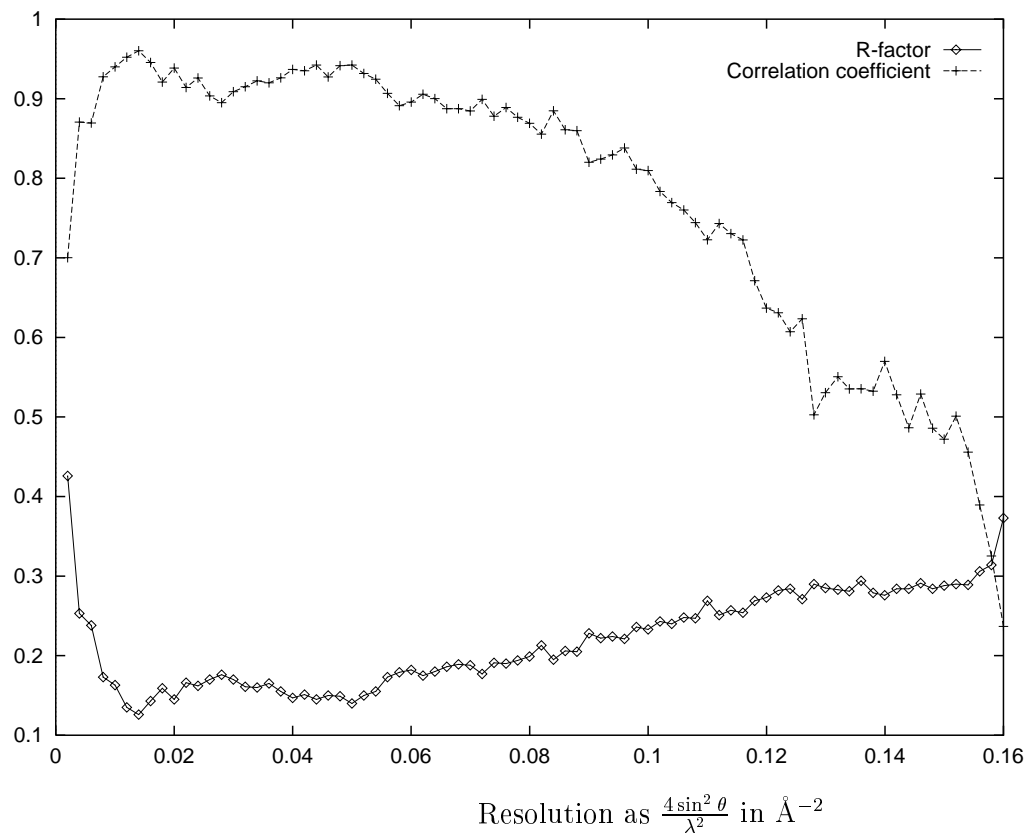


Figure 20: Final phase extension statistics as a function of resolution. The Final R -factor^a and correlation coefficient^a at 2.5 \AA resolution were 0.21 and 0.93 respectively.

^a see Figure 1.1 (page 91) for definitions

Table 8: Phasing statistics after each round of averaging (eight cycles) before extending the phases to higher resolution.

<i>Number of reflections</i>	<i>Resolution</i>	<i>R-factor^a</i>	<i>Correlation coefficient^a</i>	<i>Number of reflections</i>	<i>Resolution</i>	<i>R-factor^a</i>	<i>Correlation coefficient^a</i>
24631	5.976	0.159	0.9482	123419	3.450	0.193	0.9006
29383	5.590	0.172	0.9341	129599	3.410	0.194	0.9017
32047	5.423	0.173	0.9310	135995	3.333	0.188	0.9082
34705	5.270	0.169	0.9354	142391	3.297	0.188	0.9101
37672	5.130	0.172	0.9297	148682	3.227	0.209	0.8917
40679	5.000	0.176	0.9217	155243	3.162	0.186	0.9130
43802	4.880	0.176	0.9194	162059	3.131	0.188	0.9125
47306	4.767	0.176	0.9174	168551	3.071	0.188	0.9133
50760	4.663	0.176	0.9155	175436	3.043	0.190	0.9133
54537	4.564	0.178	0.9130	182385	2.988	0.191	0.9143
58401	4.472	0.177	0.9108	188982	2.936	0.192	0.9151
62479	4.385	0.177	0.9129	196050	2.911	0.216	0.8976
66709	4.303	0.175	0.9152	203214	2.863	0.198	0.9129
70921	4.152	0.177	0.9142	210642	2.817	0.223	0.8940
75622	4.082	0.176	0.9159	217611	2.774	0.251	0.8701
80158	4.016	0.183	0.9093	224919	2.732	0.256	0.8671
84961	3.953	0.186	0.9068	232233	2.712	0.260	0.8667
90080	3.835	0.188	0.9058	239799	2.673	0.263	0.8659
95285	3.780	0.187	0.9049	247362	2.635	0.218	0.9084
100586	3.727	0.186	0.9052	255006	2.599	0.206	0.9186
106118	3.627	0.196	0.8956	262515	2.565	0.204	0.9213
111758	3.581	0.192	0.8994	270285	2.532	0.203	0.9231
117548	3.536	0.195	0.8978	277992	2.500	0.204	0.9240

^a see Figure 1.1 (page 91) for definitions

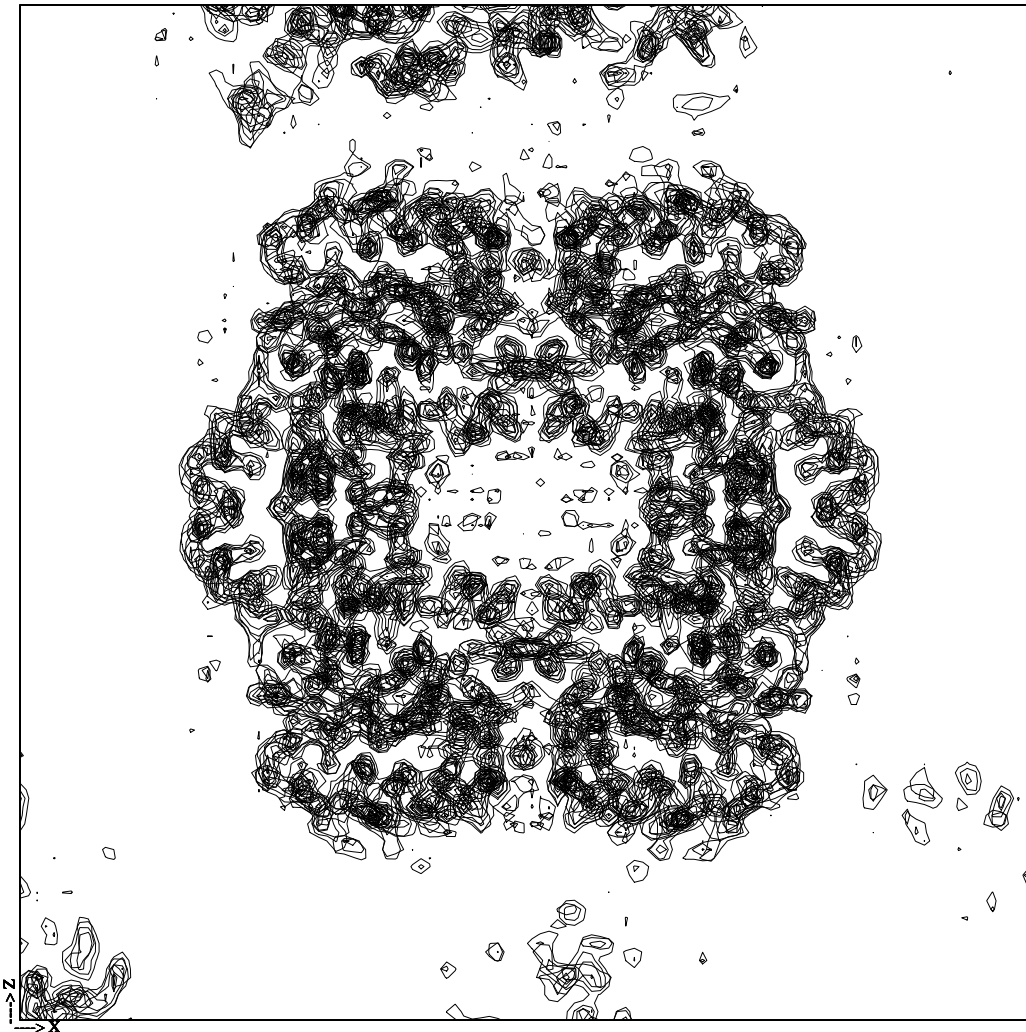


Figure 21: Projection of a 10 Å thick slab of the map calculated from the final extended and averaged phases obtained looking down a non-crystallographic symmetry axis (the other two two-fold axes are horizontal and vertical intersecting at the center of this projection) and using data between 2.5 Å and 20 Å resolution contoured at 1σ with 1σ increments.

tension procedure at 2.5 Å resolution skewed along a local two-fold axis is shown in Figure 21. The other two-fold axes are horizontal and vertical and intersect at the center of this projection.

The final R -factor and correlation coefficient for the phase extension at 2.5 Å were 0.21 and 0.93 respectively (Table 8 and Figures 1.1 and 20). The average phase shift in the last cycle of averaging was 1.0°.