CHAPTER 6

REFINEMENT RESULTS

6.1 Low Resolution Refinement

The results of the preliminary R-factor search are discussed at the end of chapter 4. It was hoped that the best 16 Ångstrom model, with an R-factor of 0.517 when compared with observed structure factors, would have phases sufficiently close to the true ones that they would converge to the latter by symmetry-averaging phase refinement.

If the starting model were incorrect in any place by a displacement of half a wavelength at the limiting resolution, then the contribution of that part of the electron density to those structure factors would be completely out of phase with the expected contribution for the structure. From this starting point, the phase refinement might just as easily modify the structure in the wrong direction as the right one; it is not clear that the lower resolution terms, fewer in number, would have the strength to outweigh the noise introduced by the wrongly phased terms at limiting resolution. On the other hand, if the starting model is within a quarter of a wavelength, all contributions to the calculated structure factors are either positive or neutral. We conclude from this argument that the expected radius of convergence under phase refinement for the position of any

portion of the structure in direct space is about a quarter of a limiting wavelength. For this reason, cycles of refinement should be performed at low resolution and gradually enlarged to include higher and higher resolution shells of observed data.

A separate question to pose is: if part of the structure is missing form the model, what tendency is there for this to appear during phase refinement? To place confidence in such unprecedented features that might appear in the final high resolution map, it is worthwhile to perform some of these control experiments at low resolution, where the calculations are much faster.

Figure 6.1 shows the course of the low resolution refinement. Three cycles were done with 3662 terms from 16 to 35 Ångstroms starting with the best 3 parameter model of $\theta = 97^{o}$, $x_S = 18\text{Å}$ and $x_P = 20\text{Å}$. These were followed by four cycles with 9408 terms to 12 Ångstroms, again deriving new starting phases from the model. Both calculations were done on the standard 12 Ångstrom grids of table 5.1. Convergence of some kind is clearly taking place, as evidenced by the decrease in conventional R-factor, increase of correlation coefficient, and decrease of the mean absolute phase shift per cycle. The convergence is rapid, as is expected with such a small redundancy ratio, U/V. Also shown in figure 6.1 is the convergence behaviour of the phase refinement of a structure based on a different starting model. The model

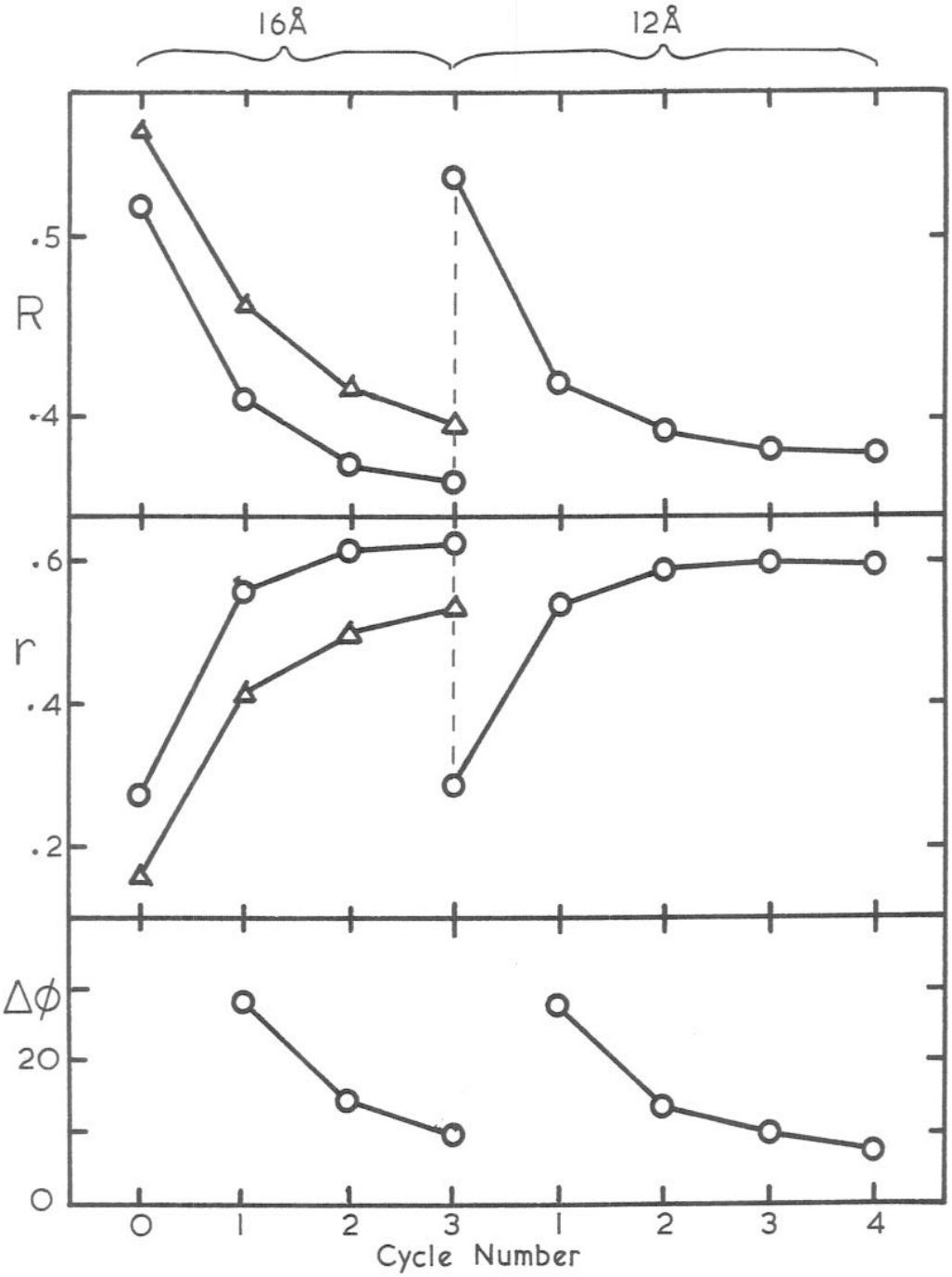


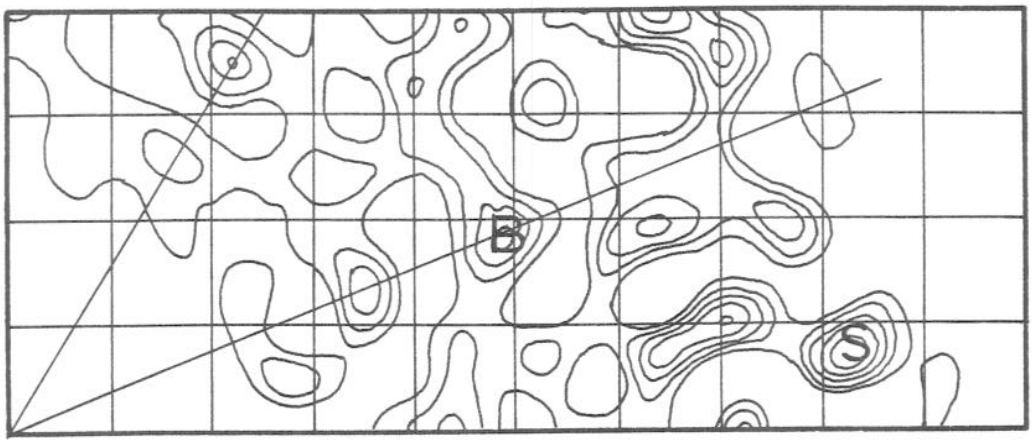
Figure 6.1. Progress of the low resolution phase refinement with 16Å and 12Å data. R-factor (top), correlation coefficient (middle) and mean phase change from previous cycle (bottom) are shown.
△ starting phases derived from $\theta = 97^{\circ}$, x = 14Å model.
○ starting phases derived from $\theta = 97^{\circ}$, x = 18Å model.

has $\theta = 97^{\circ}$ and $x_S = x_P = 14\text{Å}$ and also has the β-annulus (residues 1-37 of the C chain) deleted. The convergence is very similar; proceeding at about the same rate, but does not achieve as good an agreement with the observed structure factors in the asymptotic limit.

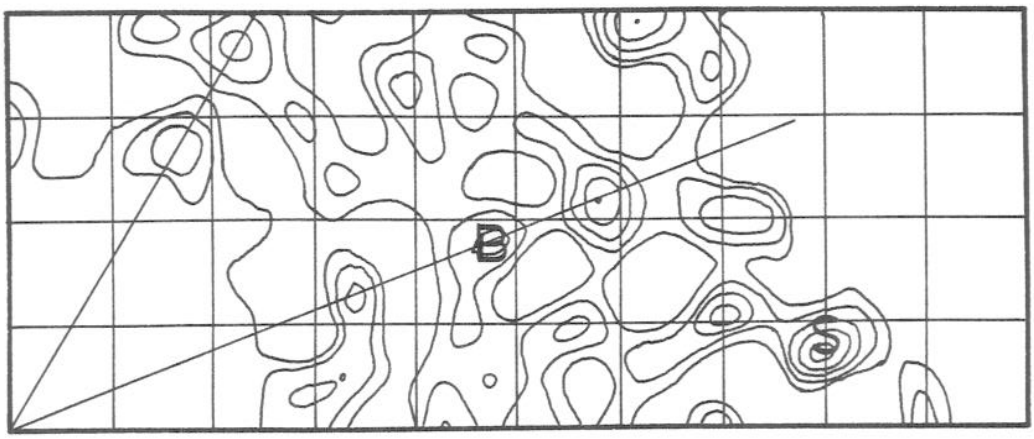
Examination of the 16 Ångstrom electron density maps for the two cases, examples of which are shown in figure 6.2, shows that the density of the map derived from the correct model is largely unaltered by the refinement after the first cycle, while the one derived from the wrong model changes by a radial displacement of the positive peaks corresponding to the S-domains into something resembling the first map. After two cycles, most of the redistribution has taken place, accompanied by the large part of the R-factor improvement. The shift in the main peaks is about 7 Ångstroms in magnitude, as would be expected for a model artificially shrunken by 6 Ångstroms. Furthermore, the β-annulus, which was omitted from the model, has reappeared after three cycles. Comparison with the map of the compact structure at the same resolution shows that the position of the β-annulus, as well as the other main peaks, is meaningful for a virus particle expanded radially by 18 Ångstroms.

Several conclusions can be drawn from these results:

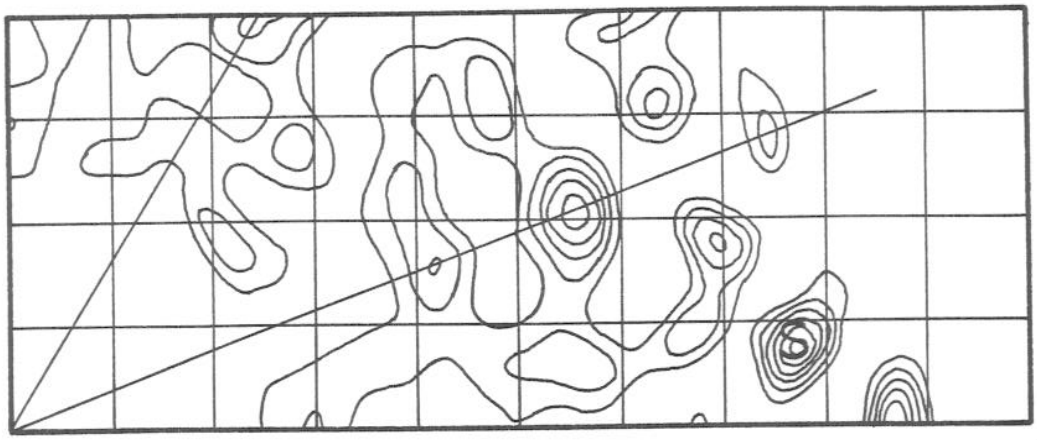
i) The radius of convergence in direct space is fairly large. 6 Ångstroms is more than a quarter of the 16



(a) After three cycles starting from the best model with x = 18Å.



(b) After three cycles starting from the x = 14Å model.



(c) After one cycle starting from the x = 14Å model.

Figure 6.2. Continued on next page.

Figure 6.2 (continued). Part of the z = 0 section of some 16Å phase refined structures. Equal contour spacings above zero are shown. Each grid is 20Å square and the virus origin is in the bottom left hand corner. The 3-fold and 5-fold local symmetry axes of the particle are drawn in. The dense shell of protein between radii of 130 and 190Å is clearly visible. (a) shows the structure fully refined from the best three parameter model, while (b) and (c) show the 3 cycle and 1 cycle stages of the refinement starting from the wrong model with x = 14Å and the beta annulus omitted. The S-domain feature marked 'S' is seen to migrate radially outwards during the refinement from the wrong start position. The beta annulus, marked 'B', does not appear in the first cycle, but is a strong feature of the 3 cycle map.

Ångstrom wavelength, but this amount of error is not too great to allow convergence to a meaningful structure.

ii) The method has the ability to bring to full density in a few cycles, features that are not present in the starting phases. We can be confident that any features that are present in the crystals of the expanded virus will become visible as peaks in the refined electron density, even when they are not anticipated by the model used for starting phases.

iii) Although a wide range of models might be expected to converge towards something close to the right structure, the agreement with observation of an attempt is still better when the starting phases are derived from a good model than from a bad one.

Particularly in view of this third conclusion, it was decided to interpret fully the 12 Ångstrom refined results before moving on to high resolution. Methods of real space refinement, that might give an optimal fit of model into density, were considered, but no satisfactory way was found. Instead, it was decided to examine the heavy atom derivative data.

6.2 Heavy Atom Derivative Interpretation.

Nowhere in the phase refinement scheme described is there an independent test of the validity of the phases.

The fact that the agreement with observation improves when the known non-crystallographic symmetry of the virus is imposed as a constraint is a strong indication. To satisfy all doubt, a test of the phases, which is completely objective and independent of the method by which they were obtained, is required. The calculation of a heavy atom difference map is such a test.

6.2.1 The Difference Fourier Method.

The protein phase can be applied to the difference in structure factor amplitudes between the native and an isomorphous heavy atom derivative, to form a set of structure factors closely related to the set formed by the constellation of heavy atoms in the crystal alone. The map calculated from these difference structure factors should show an image of each of the heavy atoms upon a smooth background, being an approximation to the difference in electron density point by point between the native and derivative crystals. The method was first used by Stryer et. al. (1964) and has been widely used since; it is highly reliable for locating isomorphous differences from a known structure. As a test of phases, the method is frequently used in multiple isomorphous replacement (MIR) phasing: existing phases are used to locate heavy atom positions in a new derivative, which can then be refined to produce better phases (Dickerson et. al., 1967); if the existing phases are poor, it is impossible to interpret the difference map.

It has been found that only a small fraction of the difference Fourier coefficients are required in the synthesis in order to obtain clear, unambiguous heavy atom peaks. This is particularly true in situations with a large amount of non-crystallographic symmetry, where the difference map can be averaged to improve the signal to noise ratio (Hogle and Harrison, 1981). This is fortunate from the point of view of data collection for the expanded TBSV derivatives, where a single crystal was used for each heavy atom data set.

6.2.2 Heavy Atom Difference Maps.

Symmetry-averaged difference maps were calculated for each of the derivative datasets listed in table 3.1, using the phases from the final cycle at 12 Ångstroms resolution. The intensity data were first scaled carefully together by fitting a straight line through the graph of average intensity ratio as a function of $\sin^2\theta/\lambda^2$ and calculating the relative scale and temperature factors. These all appear to be very clean with the main peaks five times or more greater than the r.m.s. noise level. An example is shown in figure 6.3. Although the subsets of the data in the two platinum derivative datasets come from different parts of reciprocal space, the peak positions are very close indeed; the same is true for the two mercury datasets, where it is also clear that the relative heavy atom occupancies of the various sites are not all the same. There are three

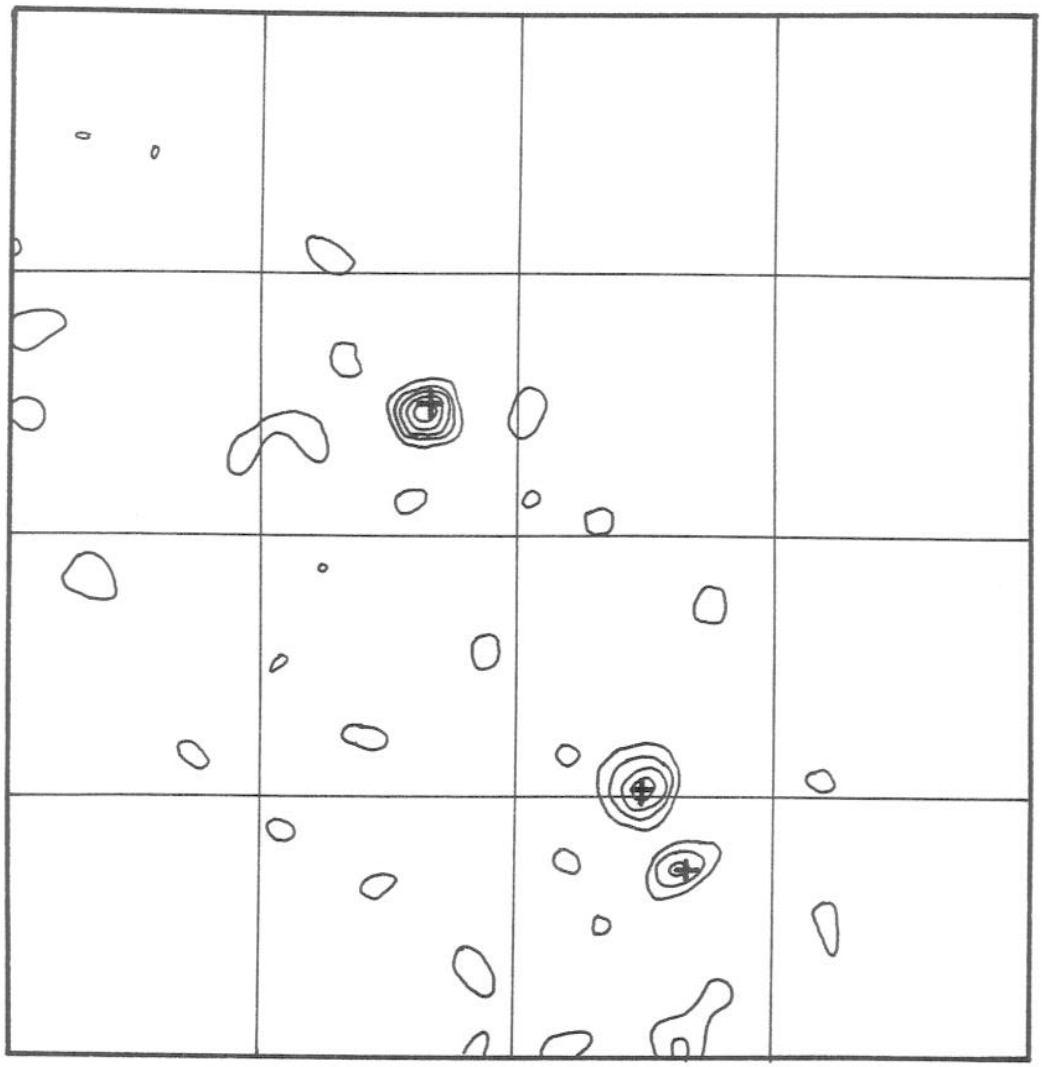


Figure 6.3. The z = 16Å section of a $PtCl_4$ difference map calculated with data from crystal X131 and the native, taking phases from the refined 12 Ångstrom structure, and averaged over the icosahedral symmetry of the virus. The grids are 50Å squares and the virus origin is bottom left. The platinum positions are clustered around the particle 5- and 3-fold axes at a radius of 135Å. The single site labelled 'A' in the upper half of the map is bound to the A position S-domain, directly above the 5-fold; the others are the B and C sites above the particle 3-fold. The atomic positions, after refinement by PHARE, are marked '+' and are quoted in table 6.1.

platinum sites in the icosahedral asymmetric unit, all on the same z section (standard orientation). These show the quasi symmetry of the virus very clearly. Furthermore, the positions are related to three of the four platinum sites of the compact structure (Winkler et. al., 1977), which are also quasi symmetry related because they bind to the same residue in the three chains. The platinum positions are in a plane 18Å higher in the expanded structure than in the compact one; this is precisely the position expected by the simple model of the expansion. The same is true for the four mercury positions which show about the same radial displacement from four of the six compact site positions.

The clean appearance of the difference maps is strong evidence for the high quality of the phases. The concern that the phases might be completely artefactual can be safely dismissed. It is particularly encouraging that, while the starting model and phase refinement method are completely independent of the heavy atom information of the compact structure, the same sites appear in the difference maps. Moreover, there is the additional possibly that the heavy atom sites will refine by the conventional methods so the exact positions of the metal sites can be determined. Such information would be extremely valuable as it would allow a more accurate expansion model to be constructed based on the displacement vectors from the compact site positions to the expanded ones. The three platinum atoms are attached to the A, B and C S-domains at methionine 210

while the mercuries are attached to cysteine 223 in the A and B S-domains and to an unidentified ligand in the A and B P-domains. Thus five of the six domains are labelled by at least one site, which will in principle allow 10 of the 13 parameters of the most general expansion model to be fixed. Self-consistency checks are also possible in the form of site to site separation distances across supposedly conserved contacts in the expanded virus; the accuracy with which such distances are conserved is a measure of the error in the new model that is derived.

6.3 Heavy Atom Refinement.

The approximate heavy atom positions can be measured by locating the point of maximum density in the averaged difference map. However, these positions are erroneous because of the large fraction of missing data and also because of the approximations intrinsic to the difference Fourier method. To find the correct positions, it is necessary to perform the structure factor calculation for the heavy atoms alone, add these vectorially to the known protein structure factors to obtain an estimate for the derivative amplitudes, and calculate the r.m.s. 'lack of closure' (of the vector triangle) by summing the square differences from the observations for the derivative. The lack of closure can be minimised with respect to the parameters defining the heavy atoms by the conventional least squares

algorithm (Dickerson et. al., 1961, 1968; Lipscomb et. al., 1966). This method is independent, in principle, of the total number of observations, provided, of course, there are enough to determine all the parameters uniquely.

The program that was used to do the minimisation is 'PHARE'. This required certain modifications to allow up to 120 equivalent positions of each site in the unit cell, and to allow it to read the phase input directly. The parameters refined for each heavy atom were its three positional coordinates and its occupancy. The latter was refined independently for each crystal to allow for the variations in the soaking conditions. The overall scale factor and temperature factor relating the datasets were also refined. The constraint of non-crystallographic symmetry was imposed by supplying the program with a list of transformations relating all of the 120 equivalent positions in the unit cell (the program regards all problems as having space group P1). Since these transformations were not trivial to write down, being defined in the skew coordinate frame of the monoclinic unit cell, a separate program was written to generate them, exactly analogous to that used for the symmetry averaging programs. The starting positions of the heavy atoms were the positions of the peaks in the difference maps; the occupancies were started at unity. No special tricks were required to force convergence of the algorithm for the 12 Ångstrom data: all parameters were refined simultaneously by full-matrix least squares within a

few cycles; no damping of the parameter shifts was needed either.

6.3.1 Platinum Sites.

The refined positions and relative occupancies of the platinum sites are quoted in table 6.1. The coordinates are given in the orthogonal viral frame in Ångstroms, because the crystallographic coordinates are not easily compared with the compact positions. The standard deviations of the positions are derived from the residual lack of closure after refinement. The corresponding compact site positions are also quoted for reference and the sites are named after the subunits they are attached to.

For reasons detailed in chapter 4, the motion of the B and C sites from the compact to the expanded positions should be considered to be a single rigid body motion. The component of the motion parallel to the 3-fold axis is 19.1Å for B and 18.8Å for C, showing good agreement within the bounds of the positional error. The perpendicular distance from the B site to the 3-fold is 23.2Å compact and 23.1Å expanded; that of the C site is 20.4Å compact and 20.4Å expanded (see table 6.3). Both these distances should be conserved if the model is to be supported, and they do so within the experimental limits. There is also a net clockwise rotation (looking down on the particle) of 2.7° for B and 2.2° for C, but this is within the error limits of zero.

(a) Expanded sites

Site	x	y	z	Occupancy (x131)	Occupancy (x132)
A	-7.63 ±.46	67.47 ±.37	134.10 ±.48	7.1 ±1.0	9.1 ±1.0
B	36.67 ±.60	18.96 ±.46	132.99 ±.63	6.9 ±1.1	5.9 ±1.1
C	-30.21 ±.57	7.05 ±.42	132.63 ±.57	6.1 ±1.1	7.2 ±1.1

(b) Compact sites

Site	x	y	z
A	-6.02	57.21	117.26
B	28.97	18.43	115.50
C	-23.72	7.78	115.00

Table 6.1. Heavy atom parameters of the 12Å refined platinum sites and the corresponding sites in compact TBSV. Nomenclature is after the compact structure (Winkler et. al., 1977). All positions are in Ångstroms; the occupancies are on an arbitrary scale.

The corresponding description of the motion of the platinum A site during expansion is a 19.7Å shift parallel to the 5-fold and a clockwise rotation of 5.4°. The perpendicular distance to the 5-fold changes from 14.3Å compact to 15.2Å expanded, again consistent with the model.

6.3.2 Mercury Sites.

The four sites obtained from the difference map all refined well with PHARE. Their positions and occupancies are given in table 6.2. There was a marginal fifth site in the C_1 position quasi symmetry related to the refined A_1 and B_1 S-domain sites, but this had much lower density in the difference map. Interestingly, this also refined well and had comparable occupancy to the A_1 and B_1 sites. An additional D site, unrelated to any of the compact sites refined well too. This position is in close proximity to cysteine 168. There is no apparent chemical reason why a P-domain C_2 site, quasi symmetry related to A_2 and B_2, and present in the compact, should not also exist. A starting position for this was estimated and refined. Although the refinement was not as good as before, taking many cycles to converge and leaving a large residual standard deviation in the coordinates (see table 6.2), the occupancy did not diminish, so the site is real. Tests of PHARE with false sites showed that these vanished rapidly. It is satisfying that all six of the compact sites are now accounted for and that the quasi symmetry is upheld.

(a) Expanded sites

Site	x	y	z	Occupancy (x140)	Occupancy (x144)
A_1	-16.84 ±.63	43.80 ±.47	145.81 ±.64	1.9 ±.7	6.2 ±.8
B_1	18.45 ±.61	40.04 ±.45	144.70 ±.62	3.0 ±.7	5.9 ±.8
C_1	-5.24 ±.90	8.53 ±.62	145.64 ±.85	2.3 ±.7	4.0 ±.8
A_2	-29.40 ±1.15	48.21 ±.85	177.61 ±1.13	2.8 ±.7	1.6 ±.8
B_2	28.52 ±.72	45.53 ±.53	178.95 ±.75	2.6 ±.7	5.0 ±.8
C_2	9.47 ±1.35	4.69 ±.93	178.64 ±1.34	1.4 ±.7	2.7 ±.8
D	-42.59 ±.74	1.03 ±.65	150.55 ±.74	1.9 ±.7	4.5 ±.8

(b) Compact sites

Site	x	y	z
A_1	-14.26	34.68	127.95
B_1	13.72	35.37	128.14
C_1	0.38	10.38	128.76
A_2	-26.06	46.18	162.25
B_2	31.50	37.40	161.98
C_2	-5.59	5.56	164.81

Table 6.2. Heavy atom parameters of the 12Å refined mercury sites and the corresponding sites in compact TBSV. Nomenclature is after the compact structure (Winkler et. al., 1977). All positions are in Ångstroms; the occupancies are on an arbitrary scale.

6.3.3 Interpretation of the Heavy Atom Positions.

The refined positions for the S-domains were analysed in the way described above. The results are displayed in table 6.3. An average value of the expansion displacements is deduced, assuming that the motion of the heavy atom label describes that of the domain to which it is attached. There is a noticeable difference in the amount of expansion along the 5-fold direction (20.0Å) and along the 3-fold (18.3Å). The P-domain positions tell an even more interesting story, which is summarised in table 6.4. The C subunit moves 13.8Å up the particle 2-fold, a considerably shorter distance than that of the S-domains; this P-domain therefore drops down 5Å relative to the S-domains. There is in addition a clockwise rotation of 110° (equivalent to a rotation of -70°) of this domain. This conclusion is based entirely on the position of the poorest mercury site and therefore should be verified by an R-factor search (see below). The dimer consisting of the A and B position P-domains also rotates, but through a smaller angle. The magnitude of the translational part of the AB dimer motion was obtained by comparing the centroid position of the expanded sites with the compact. This gives both the expansion displacement of 16.8Å and its direction, quoted in table 6.4.

(a) Expansion displacements parallel to the symmetry axis.

	S-domain subunit		
	A	B	C
Pt	19.7±.5	19.1±.6	18.8±.6
Hg	20.2±.6	17.3±.6	18.2±.9
Average	20.0±.4	18.3±.3	

(b) Perpendicular distances to symmetry axis.

	S-domain subunit		
	A	B	C
Pt Exp.	15.2±.5	23.1±.6	20.4±.6
Pt Comp.	14.3	23.2	20.4
Hg Exp.	42.8±.6	52.8±.6	47.8±.9
Hg Comp.	40.7	48.5	47.6

(c) Clockwise rotations about symmetry axis from Compact to Expanded.

	S-domain subunit		
	A	B	C
Pt	$5.4^{o}\pm1.8^{o}$	$2.7^{o}\pm1.6^{o}$	$2.2^{o}\pm1.6^{o}$
Hg	$2.3^{o}\pm0.8^{o}$	$2.0^{o}\pm0.7^{o}$	$2.3^{o}\pm1.0^{o}$

Table 6.3. Modification parameters for the S-domain heavy atom sites for the refined positions using 12Å phases. All distances are in Ångstroms. The symmetry axis refers to the 5-fold for A position sites and to the nearest 3-fold for B and C subunits.

(a) Expansion displacements parallel to the symmetry axis.

	P-domain dimer	
	AB	CC
Hg	16.8±.7	13.8±1.3

Direction cosines of AB displacement
= (-.186, .398, .898)

(b) Site to site separation across the dimer.

	P-domain dimer	
	AB	CC
Expanded	16.6±1.4	21.2±3.3
Compact	15.8	16.3

(c) Clockwise rotations about symmetry axis from Compact to Expanded.

	P-domain dimer	
	AB	CC
Hg	$32.3^{\circ}\pm4.7^{\circ}$	$108.5^{\circ}\pm8.8^{\circ}$

Table 6.4. Modification parameters for the P-domain heavy atom sites for the refined positions using 12Å phases. All distances are in Ångstroms. The symmetry axis here refers to the particle 2-fold for the CC dimer, and to the direction of the vector joining the centroid of the A and B sites in the compact structure to that of the expanded sites for the AB dimer, whose cosines are given in (a).

6.4 Improved Model.

The heavy atom information obtained with the 12 Ångstrom phases can now be used to produce a much better model than the original 3-parameter one described in chapter 4. The new model can then be tested be R-factor searches for small variations in each of the new parameters.

The one critical parameter that is not directly fixed by the heavy atom analysis is the crystal packing angle, Θ. The original value of $\Theta = 97^{\circ}$ was implicitly used in the heavy atom refinement in the form of the matrices describing symmetry equivalent positions. An attempt was made to determine the value empirically by calculating difference maps with symmetry matrices corresponding to 96.7° and 97.3° and comparing peak heights and positions. These difference maps were virtually identical to the 97° one and so the method proved to be not as sensitive as the R-factor search method, which was therefore trusted instead. The result does however remove the possible question of the heavy atom parameters being biased by the choice of Θ angle.

The full 13 parameter model is not completely determined by the heavy atom positions. Only 5 of the 6 degrees of freedom in the rigid body transformation of the AB P-domain dimer motion are fixed by two sites; the dimer still has the freedom to rotate about the axis joining the two sites. For this reason, it was considered preferable to use the intuitively sensible local 2-fold axis of the dimer

as the axis of rotation. This removes <u>two</u> degrees of freedom from the motion, but in a way that is reasonably consistent with the observed site positions. The translational part of the motion is completely described by the heavy atom positions. The direction cosines and magnitude are derived from the motion of the mid-point of the line joining the A_2 site to the B_2 site (see table 6.4). This direction is quite different from that of the local 2-fold of the dimer, which is obtained by the optimal superposition of the two domains; these axes are 16° apart.

6.4.1 9-Parameter Model Specification.

S-domain rotations were left out of the model as being negligible, which leaves a total of 9 parameters in the model: the complete specification is

$x_{SA} = 20.0\text{Å}$

$x_{SBC} = 18.3\text{Å}$

$x_{PC} = 13.8\text{Å}$ $\quad \alpha_{PC} = 108.5^{\circ}$

$x_{PAB} = 16.8\text{Å}$ $\quad \alpha_{PAB} = 32.3^{\circ}$

AB displacement along (-.186, .398, .898)

$\Theta = 97.0^{\circ}$

The AB P-domain rotation was fixed as before to be the local diad of the dimer in the compact structure. The direction cosines were (-.264, .140, .954). This model was subjected to verification by R-factor searches before the 8 Ångstrom phase refinement was started.

6.4.2 R-factor Searches.

There are three purposes of performing a search at this stage:

i) To determine whether the improved model really is better than the previous one.

ii) To verify that a quarter turn rotation of the C position P-domains makes sense.

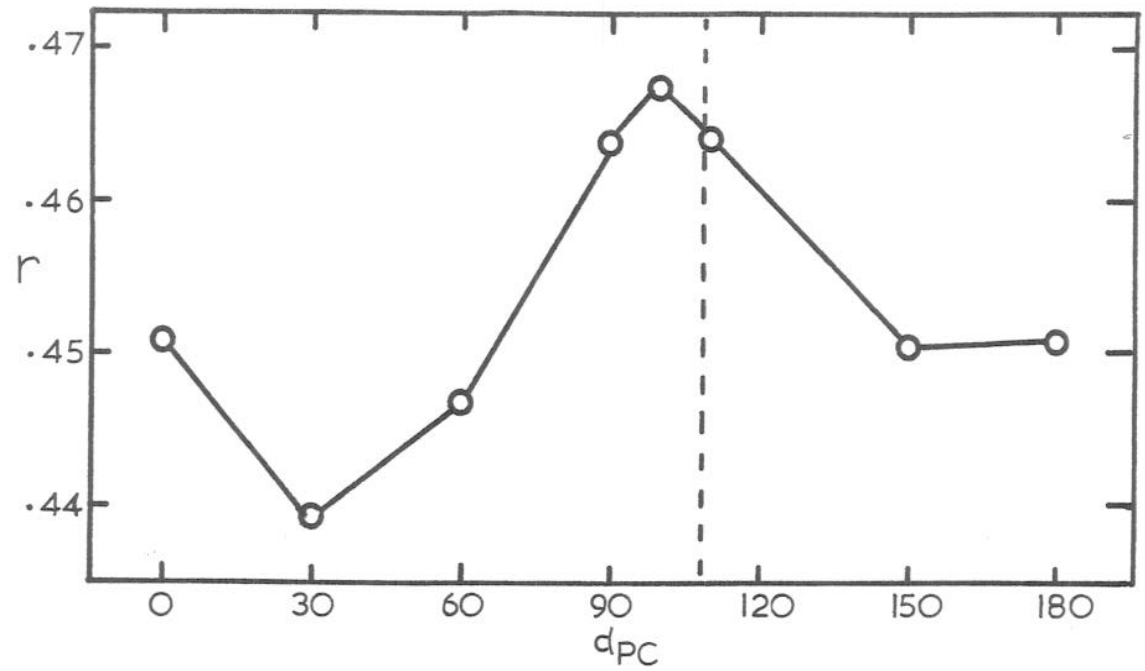
iii) To fine-tune the value of Θ that was not derivable from the heavy atom positions.

R-factors and correlation coefficients (for definitions see section 4.3) were calculated for 29449 terms between 35 and 8 Ångstroms. They were scaled independently in six equal sized resolution shells as described in chapter 4.

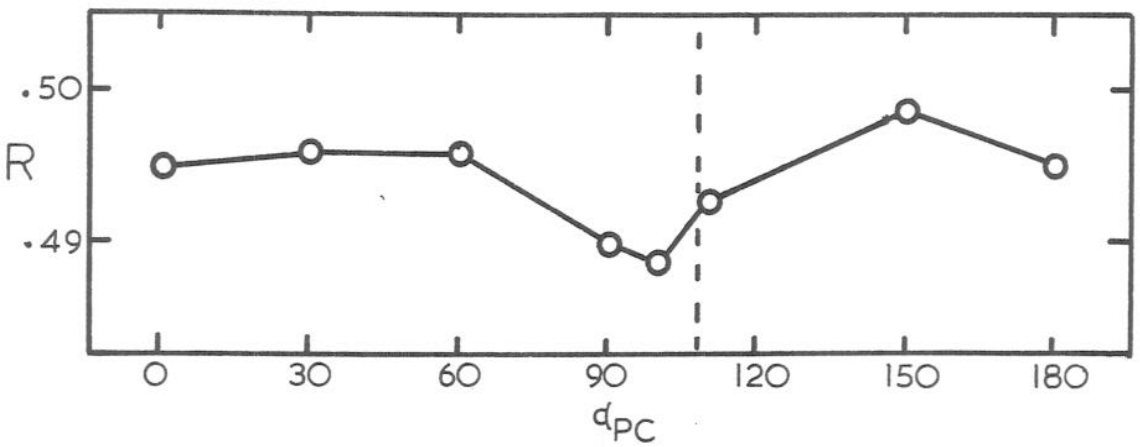
Figure 6.4 shows the R-factor search for the rotation of the P-domain CC dimer. Three conclusions are immediately apparent from this result:

i) The new model is a dramatic improvement over the three parameter one. The absolute R-factor and correlation coefficient are <u>twice</u> as far from their values for random structure factors, when correctly scaled (see chapter 4).

ii) The fit of the model now shows a significant response to small perturbations. The C position P-domains constitute about one eighth of the virus mass included in the model, but their orientation affects the correlation coefficient by 0.03. Such sensitivity was not



(a) Correlation coefficient.



(b) R-factor.

Figure 6.4. R-factor search for the orientation angle of the CC P-domain dimer. All data to 8 Angstroms are included. Because of the 2-fold symmetry of the dimer, all possible positions are covered by the range $0 < \alpha_{PC} < 180^{\circ}$.

seen with the earlier model.

iii) The value of d_{PC} with the best agreement is very close to the value predicted by the heavy atom positions, the difference being about one standard deviation of the latter.

To make the predicted position of the C position P-domains absolutely certain, a second search was made for its radial position, x_{PC}. It was found that a perturbation to either side of the heavy atom value worsens the agreement. The packing angle Θ was determined by the final search to be 97.1^{o}, showing little variation from its previous value.

6.5 High Resolution Phase Refinement.

Because significant changes in the model had taken place, a new envelope was calculated by the method previously described in section 5.4. The progress of the 7.4 Ångstrom phase refinement was similar to that at low resolution as is shown in figure 6.5. Two cycles were performed first using 9408 terms from 35 to 12 Ångstroms resolution which brought the R-factor down to 0.336 and the correlation coefficient to 0.689; these values already represent a distinct improvement from the previous refinement. Two cycles were then calculated using 32172 non-zero terms to 7.4 Ångstroms resolution. The response was not as great as expected, the final R-factor being 0.452 and the correlation

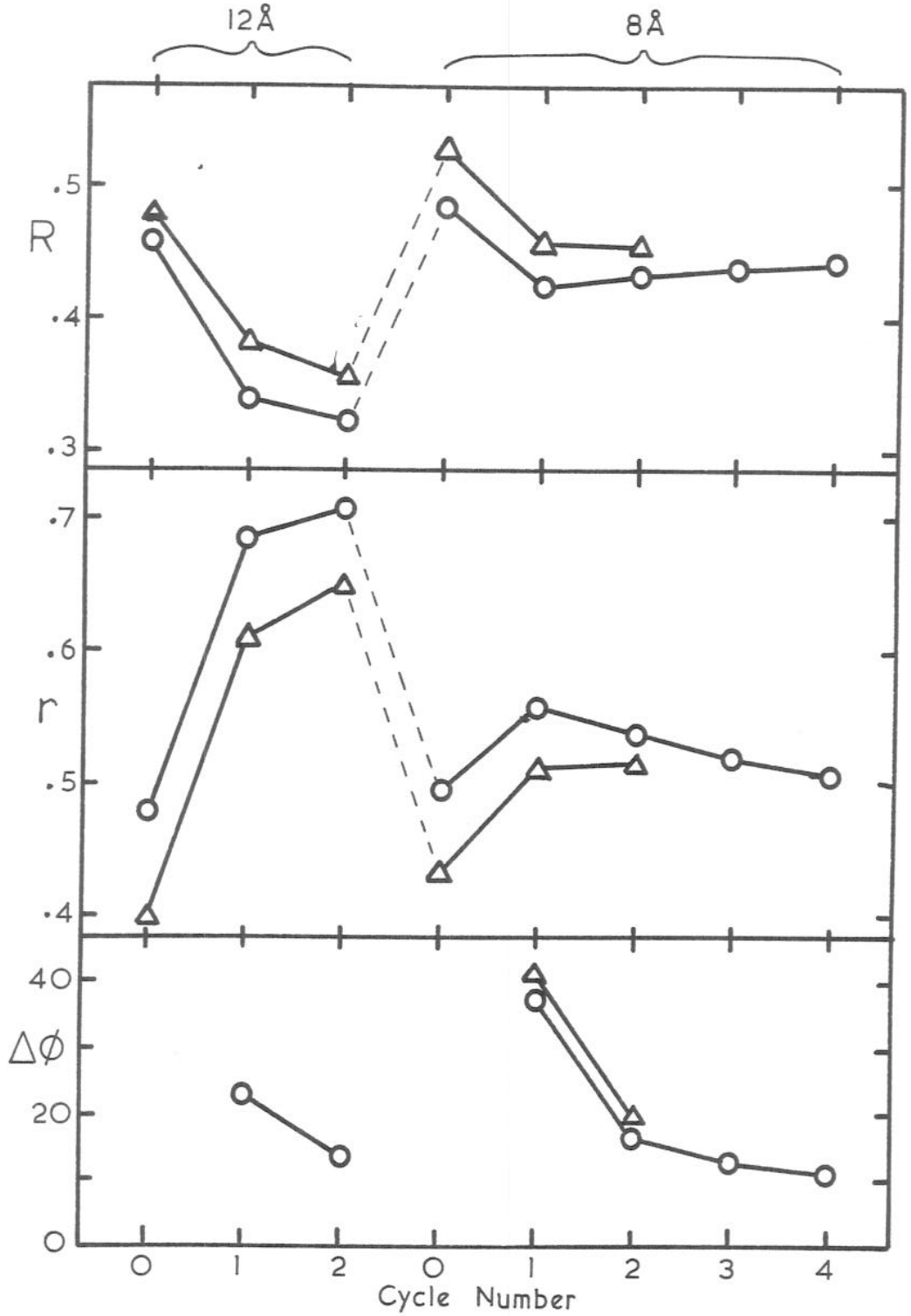


Figure 6.5. Progress of the low resolution phase refinement with 12Å and 8Å data. R-factor (top), correlation coefficient (middle) and mean phase change from previous cycle (bottom) are shown.
△ starting phases derived from 9 parameter model.
○ starting phases derived from 13 parameter model after optimisation.

coefficient 0.521, probably reflecting the lower quality of the data in the outermost shells.

The final map was sectioned and plotted with a superimposed image of the model α carbon coordinates so that the fit of the model into density could be examined. The fit is extremely good; a detailed analysis is presented below in chapter 7. The shape of the density of the domains is not so much what would be expected from a fake map of the model, but is much closer to their image in the 8 Ångstrom maps of the compact structure. This probably indicates the convolution effect of the imposition of a resolution cutoff in reciprocal space. One inadequacy of the model was apparent from close examination. Figure 6.6 shows a section of the map just inside the inner surface of the S-domains. The model would fit the density somewhat better if all of the S-domains were rotated slightly clockwise. This was the prediction of the 12 Ångstrom refined heavy atom positions, but had been neglected. This suggests that the next model should include the S-domain rotational parameters in addition.

6.6 High Resolution Heavy Atom Refinement.

Once a 7.4 Ångstrom phase set was available, the refinement of the heavy atom positions could be repeated so that the new parameters (hopefully more accurate than before) could be used to make a new model.

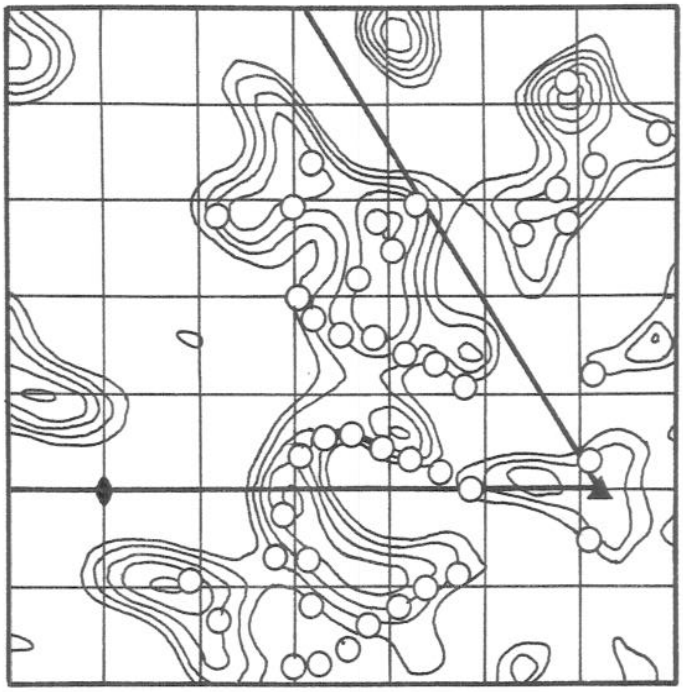


Figure 6.6. Part of a map calculated with phases refined at 7.4Å from the 9 parameter model, with α carbon positions of that model superimposed. The section is at z = 136Å and contains -10 < x < 60Å, -20 < y < 50Å; each grid is 10Å square. The density is of the B and C position S-domains (labelled) near to the inner protein surface. The particle 2- and 3-fold axes are marked, as is the boundary of the icosahedral asymmetric unit. Examination of the density suggests that a small clockwise rotation of the S-domains about the 3-fold axis would improve the fit of the model.

The high resolution refinement was not quite as straightforward as that at lower resolution. It was necessary to restrict PHARE to refine positions and occupancies on alternate cycles. The overall scale and temperature factors were refined on every cycle. The convergence was considerably slower, requiring about 10 cycles for the platinum occupancies to settle to within 5% of their standard deviation, and 16 cycles for the mercury, both starting with the best 12 Ångstrom positions.

The results of the refinement are displayed in the same form as before in tables 6.5, 6.6 and 6.7. Several of the sites have moved by more than 1Å from their low resolution positions, an amount greater than the expected error in the site positions. Some of this variation may be due to the improvement of the phases between the 12 Ångstrom calculations and the 7.4 Ångstrom ones. Furthermore, the spread among the derived structural modification parameters (tables 6.6 and 6.7) is also greater than the experimental errors in the individual parameters. For instance, the four independent estimates of the expansion displacement of the B and C position S-domains along the 3-fold axis have a spread of 1.1Å, but an experimental error of less than 0.6Å each. To investigate the source of the discrepancy, the platinum data were divided into five equal groups in an arbitrary way and refined separately until convergence. The spread of the refined values of each of the heavy atom parameters was about 1.5 times the quoted experimental error, which

(a) Platinum sites

Site	x	y	z	Occupancy (x131)	Occupancy (x132)
A	-6.93 ±.30	67.66 ±.21	133.34 ±.30	8.1 ±.9	7.8 ±.8
B	36.26 ±.34	19.95 ±.24	133.94 ±.33	6.5 ±.9	7.8 ±.8
C	-30.91 ±.35	6.53 ±.25	133.32 ±.35	8.8 ±.9	5.1 ±.8

(b) Mercury sites

Site	x	y	z	Occupancy (x140)	Occupancy (x144)
A_1	-17.48 ±.37	43.79 ±.27	145.57 ±.37	1.2 ±.6	7.2 ±.7
B_1	21.51 ±.52	40.08 ±.37	145.71 ±.51	.4 ±.6	5.3 ±.7
C_1	-5.52 ±.62	8.79 ±.45	144.66 ±.61	.9 ±.6	4.4 ±.7
A_2	-28.62 ±.39	48.88 ±.28	179.24 ±.39	3.5 ±.6	5.5 ±.7
B_2	30.59 ±.37	45.80 ±.27	179.68 ±.38	3.2 ±.6	6.2 ±.7
C_2	13.38 ±1.14	3.30 ±.83	178.76 ±1.14	1.9 ±.9	2.4 ±1.0
D	-42.94 ±.52	1.97 ±.38	150.56 ±.54	1.6 ±.6	4.6 ±.7

Table 6.5. Heavy atom parameters of the 7.4Å refined sites. All positions are in Ångstroms; the occupancies are on an arbitrary scale.

(a) Expansion displacements parallel to the symmetry axis.

	S-domain subunit		
	A	B	C
Pt	19.2±.3	19.8±.3	19.6±.4
Hg	20.0±.4	19.3±.5	17.4±.6
Average	19.6±.2	19.0±.2	

(b) Perpendicular distances to symmetry axis.

	S-domain subunit		
	A	B	C
Pt Exp.	14.3±.3	24.3±.3	19.8±.4
Pt Comp.	14.3	23.2	20.4
Hg Exp.	43.0±.4	51.2±.5	47.3±.6
Hg Comp.	40.7	48.5	47.6

(c) Clockwise rotations about symmetry axis from Compact to Expanded.

	S-domain subunit		
	A	B	C
Pt	$4.1^{\circ}\pm1.2^{\circ}$	$2.6^{\circ}\pm0.8^{\circ}$	$3.2^{\circ}\pm1.0^{\circ}$
Hg	$3.2^{\circ}\pm0.5^{\circ}$	$4.2^{\circ}\pm0.6^{\circ}$	$1.9^{\circ}\pm0.8^{\circ}$
Average	$3.6^{\circ}\pm0.7^{\circ}$	$2.9^{\circ}\pm0.4^{\circ}$	

Table 6.6. Modification parameters for the S-domain heavy atom sites for the refined positions using 7.4Å phases. All distances are in Ångstroms. The symmetry axis refers to the 5-fold for A position sites and to the nearest 3-fold for B and C subunits.

(a) Expansion displacements parallel to the symmetry axis.

	P-domain dimer	
	AB	CC
Hg	18.1±.3	13.7±1.1

Direction cosines of AB displacement
= (-.137, .352, .926)

(b) Site to site separation across the dimer.

	P-domain dimer	
	AB	CC
Expanded	16.1±0.5	27.5±2.8
Compact	15.8	16.3

(c) Clockwise rotations about symmetry axis from Compact to Expanded.

	P-domain dimer	
	AB	CC
Hg	$27.4^{o}\pm1.9^{o}$	$121.0^{o}\pm5.8^{o}$

Table 6.7. Modification parameters for the P-domain heavy atom sites for the refined positions using 7.4Å phases. All distances are in Ångstroms. The symmetry axis here refers to the particle 2-fold for the CC dimer, and to the direction of the vector joining the centroid of the A and B sites in the compact structure to that of the expanded sites for the AB dimer, whose cosines are given in (a).

accounts for most of the apparent underestimation of errors.

By comparing the various multiple estimates of the model parameters in parts (a) and (c) of tables 6.6 and 6.7, bearing in mind the possible underestimation of the experimental errors, we conclude that the model of expansion is consistent with the heavy atom data. The model is also supported by the constancy of those parameters which should not vary during the expansion. These are in parts (b) of tables 6.6 and 6.7. Of these distances, half are conserved to better than one standard deviation between the compact and expanded forms, but the rest are several standard deviations in disagreement. Part of this must be attributable to imperfect phases and part to incomplete convergence of PHARE, but there remains the possibility of small (of the order of 1-2Å) changes <u>within</u> the domains that would be reflected in the position of the heavy atoms.

The site that refined with the most difficulty was again the mercury C_2 site, which migrated considerably before settling down. A different starting position was tried and, after similarly slow convergence, refined to a rather <u>different</u> final position, much too close to the 2-fold axis, but with just as large an occupancy. The one constant feature of all the intermediate positions of this site, and both final positions, was its radius from the centre of the virus. In fact all of these positions lay approximately along a straight line in space, intersecting

the 2-fold axis. This behaviour is what might be expected if the C position P-domains were partially disordered by the existence of multiple hinge conformations, so that the averaged shape of the mercury atom is highly prolate.

6.7 Final Optimisation of the Model.

Having made use of all the collected data in the high resolution heavy atom refinement, the model parameters obtained are probably as accurate as is possible. For a completely independent confirmation, one more R-factor search was performed, searching for the values of all of the model parameters, one at a time. Conclusion ii) of the previous search (section 6.4) suggests that the method is sensitive enough to determine all of the parameters reliably.

As in all multidimensional extremal value problems, possible strong correlations between the variables must be anticipated. If two variables are correlated, then the minimum R value state of one will be a function of the other variable. For this reason care had been taken to parameterise the model in a way that maintains orthogonality between the variables as much as possible, within the limits of convenience. The positional parameters of different domains are unlikely to be highly correlated, but care should be taken with different parameters for the same domain. The order in which the searches are done is important; if any parameter turns out to be seriously in error,

then correcting it will endanger the accuracy of all the parameters examined before it, and these should all be retested. The least well determined parameters should therefore be looked at first. When a complete pass of searches in all the parameters shows each to be at an R-factor minimum simultaneously, then the solution is found.

Calculations were done, as described above, with 29449 structure factors between 35 and 8 Å Angstroms resolution, scaled independently in six equal shells. The best model beforehand had R = .4949 and r = .4868. Both R and r were used as measures of the quality of the fit, though the latter is preferred because:

i) It is much less sensitive to the scaling of the data.

ii) It is found to be more sensitive than R to small perturbations in the structure. The reverse would be true in situations where the fit is very close and the observed and calculated structure factors are almost completely correlated (e.g. $R < 0.2$ or $r > 0.9$).

iii) The first derivative of r with respect to the parameters is continuous, which means that the shape of the search profile is quadratic around the minimum. This is not true of R which has a discontinuous first derivative.

In fact there was never any disagreement between the point of maximum r and that of minimum R, so the distinction was not important. The course of the detailed searches went as follows.

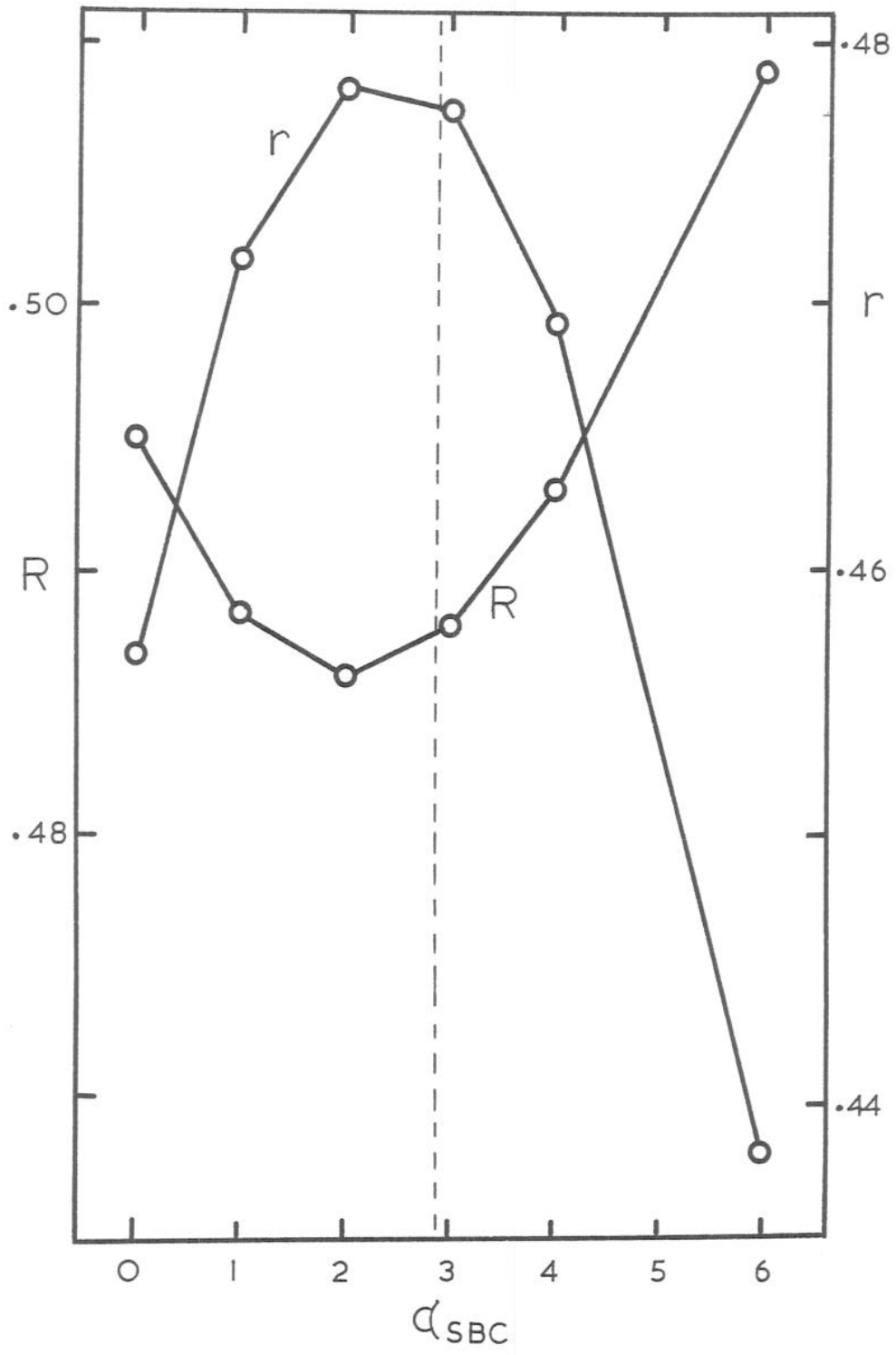
1. Search for α_{SBC} (figure 6.7(a)). As expected from the previous high resolution maps and the analysis of the heavy atom sites, a small clockwise rotation around the 3-fold did lead to a better model. Interpolating the profile of figure 6.7(a) between the search points gave $\alpha_{SBC} = 2.1^{\circ}$ as the best value with R = .4860 and r = .4787 at the nearest point. This parameter value is 2 standard deviations from the heavy atom prediction of 2.9°.

2. Search for α_{SA} (figure 6.7(b)). Similarly here a best value of $\alpha_{SA} = 4.5^{\circ}$ gave a significant improvement. This a little more than one standard deviation from the heavy atom prediction of 3.6°. With α_{SBC} and α_{SA} at their best values the measure of fit was R = .4728 and r = .5047.

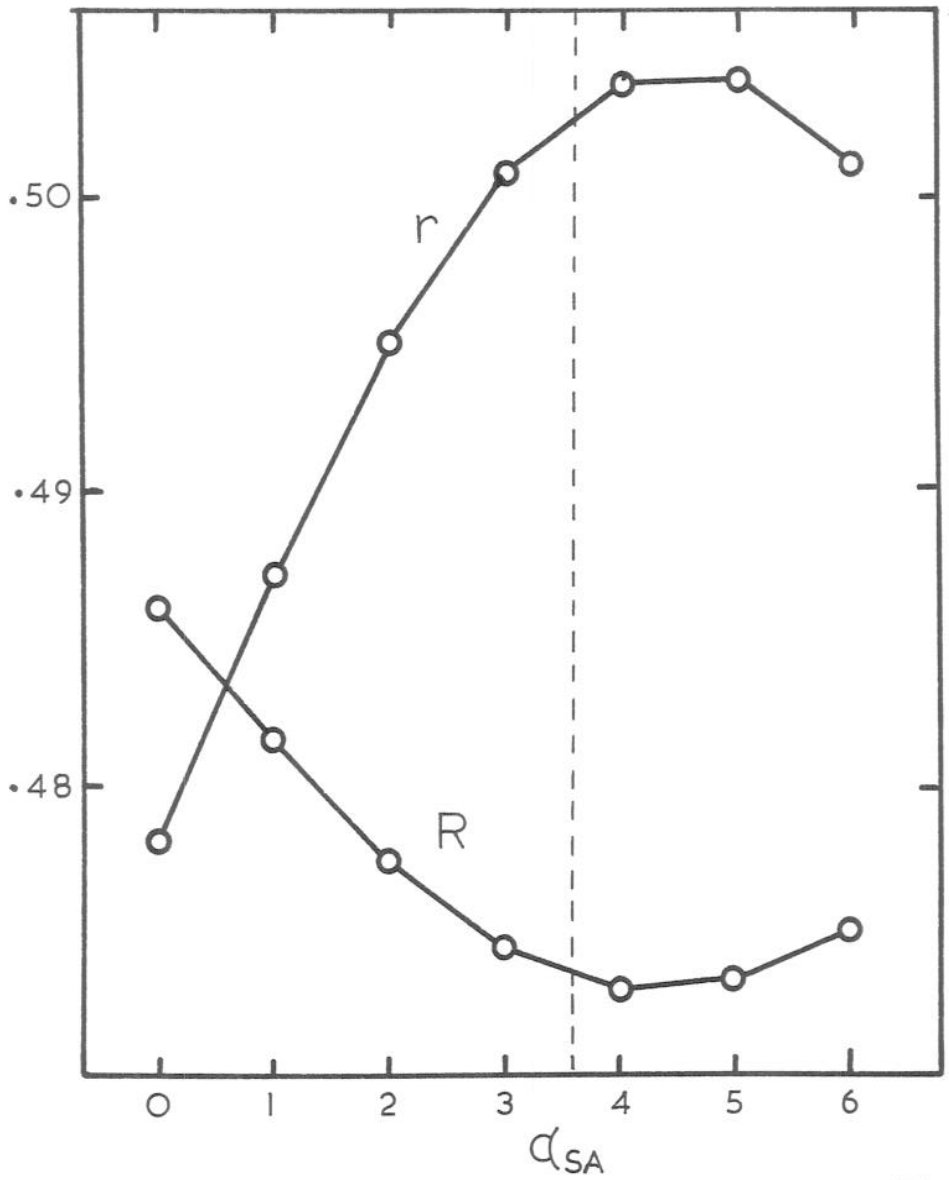
3. Search for α_{PC}. This was looked at next because it was so poorly determined by the heavy atom positions. The previous value of 100° was better than 90° or 110° and the profile was symmetric so the value was left unaltered.

4 Search for α_{PAB}. A change from 27.4° (heavy atom) to 30°, by about 2 standard deviations, improved the fit to R = .4717 and r = .5064.

5. Search for x_{PAB}. The heavy atom value of 18.1Å was better than 17.1Å or 19.1Å and the profile was also symmetric.



(a) R-factor search for the S-domain hexamer orientation angle about the 3-fold axis, α_{SBC}.



(b) R-factor search for the S-domain pentamer orientation angle about the 5-fold axis, α_{SA}.

Figure 6.7. R-factor and correlation coefficient searches (on the same scale) for the S-domain rotational parameters. The vertical dashed lines indicate the predicted values from the heavy atom positions. Between the two changes, the R-factor drops from 0.495 to 0.473, which is the greatest part of the improvement from the 9 to the 13 parameter model.

6. Search for x_{SBC}. A serious discrepancy was discovered here. A value of 17.8Å was considerably better than the heavy atom prediction of 19.0Å. The change of 5 standard deviations may indicate some inadequacy of the model or some difference in the binding of the heavy atoms to the protein between the compact and expanded structures. The fit after this modification was R = .4614 and r = .5275.

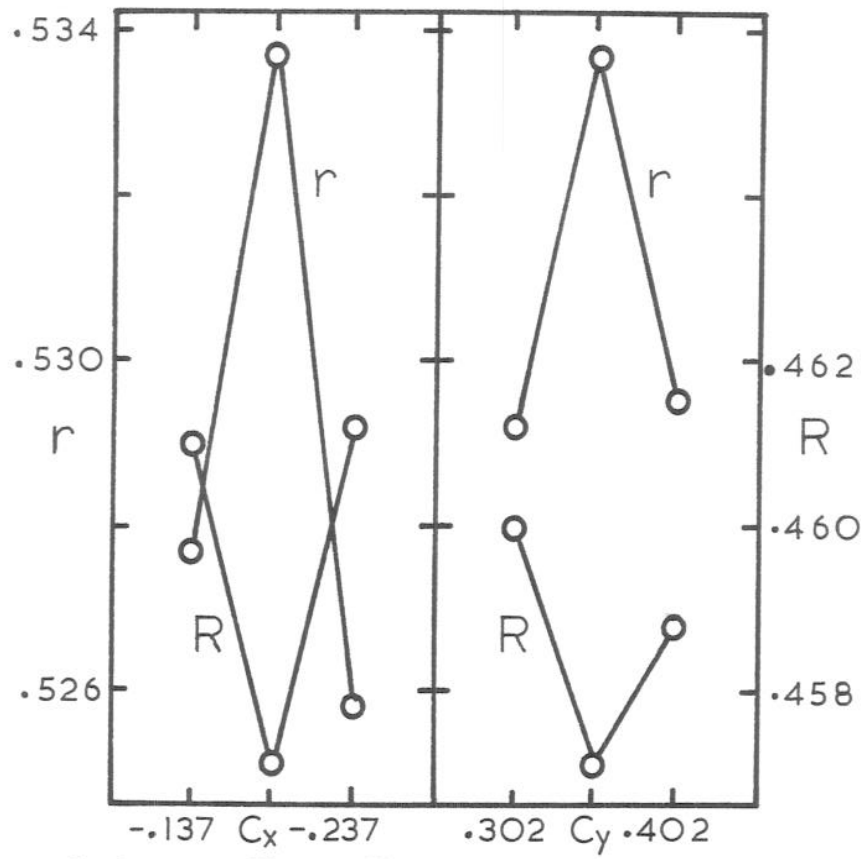
7. Search for x_{PC}. Both R and r were very insensitive to this parameter. However, both indicated that the heavy atom value of 13.7Å was better than 13.2Å or 14.7Å.

8. Search for α_{SBC}. To make sure that subsequent changes (especially to x_{SBC}) had not shifted the minimum of this parameter, its search was repeated. $\alpha_{SBC} = 2.1^{o}$ remained better than 1.6^{o} or 2.6^{o} and the profile was still symmetric.

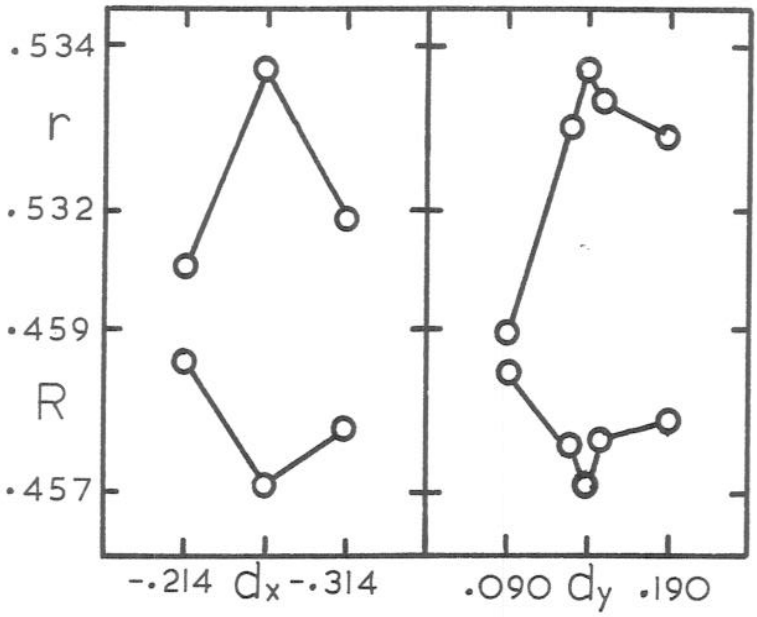
9. Search for Θ. Perturbations in this parameter of 0.1^{o} showed that it remained at the optimal value.

10. Search for the direction of displacement of AB dimer (figure 6.8(a)). Changes of 0.05 in the first two direction cosines correspond to less than 3^{o} in the axis direction or less than 1Å in the domain position. A single such change in the x-direction led to an improvement to R = .4571 and r = .5337.

11. Search for the direction of the rotation axis of the AB dimer (figure 6.8(b)). Both changes in the x- and y-components of the direction cosines showed symmetric



(a) Search for AB dimer displacement direction, c.



(b) Search for AB dimer rotation axis, d.

Figure 6.8. R-factor and correlation coefficient searches for the four P-domain AB dimer directional parameters. See table 4.1 for definitions.

worsening of the measure of fit, so the direction was left unchanged as being the local diad of the dimer.

12. Repeat searches for all eight positional parameters (figure 6.9). These were all retested for optimality. Three very slight modifications were found to be necessary:

    x_{SA} from 19.6Å to 19.4Å

    x_{PC} from 13.7Å to 14.2Å

    α_{PC} from 100° to 103°

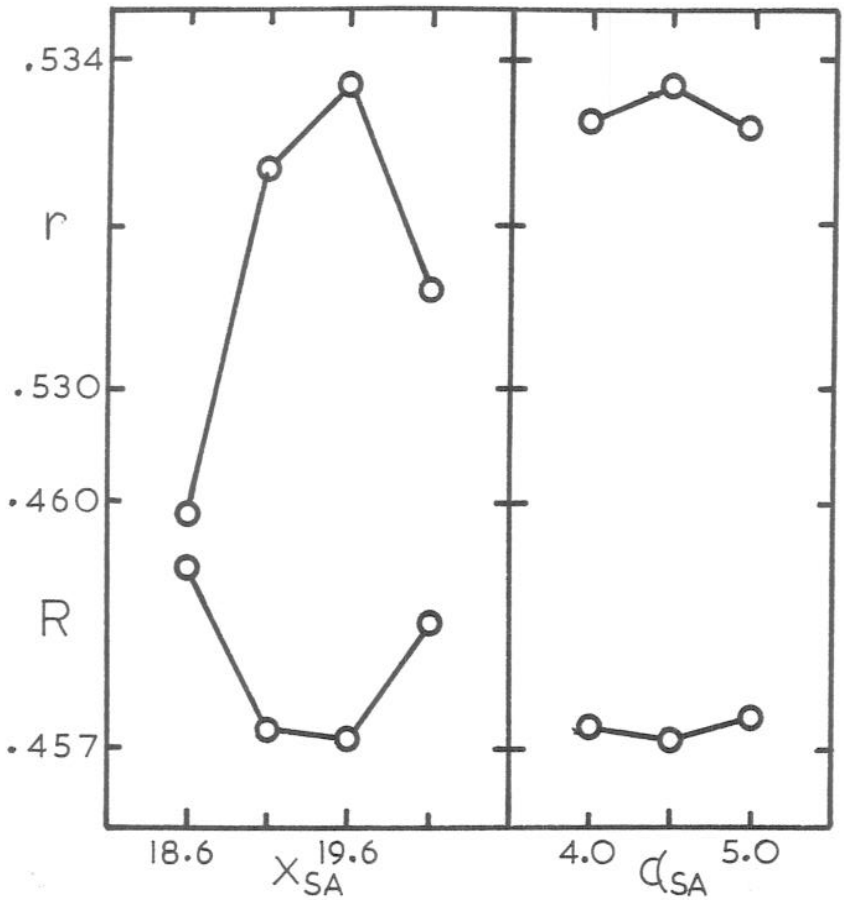
    With these changes, the R-factor dropped to .4566 and the correlation coefficient reached .5340.

13. Final search for Θ (figure 6.10). This confirmed the value of $\Theta = 97.1°$.

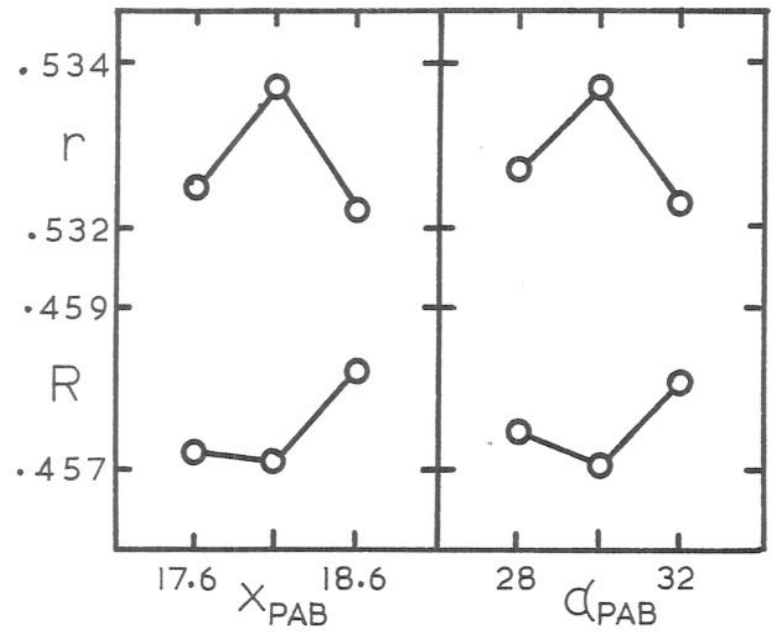
Altogether the profiles for all 13 parameters in figures 6.8 through 6.10 show that they are all optimised simultaneously and that the model is therefore completed. A comparison of the parameters of the various stages of the evolution of this model are shown in table 6.8.

6.8 Calculation of Final Map.

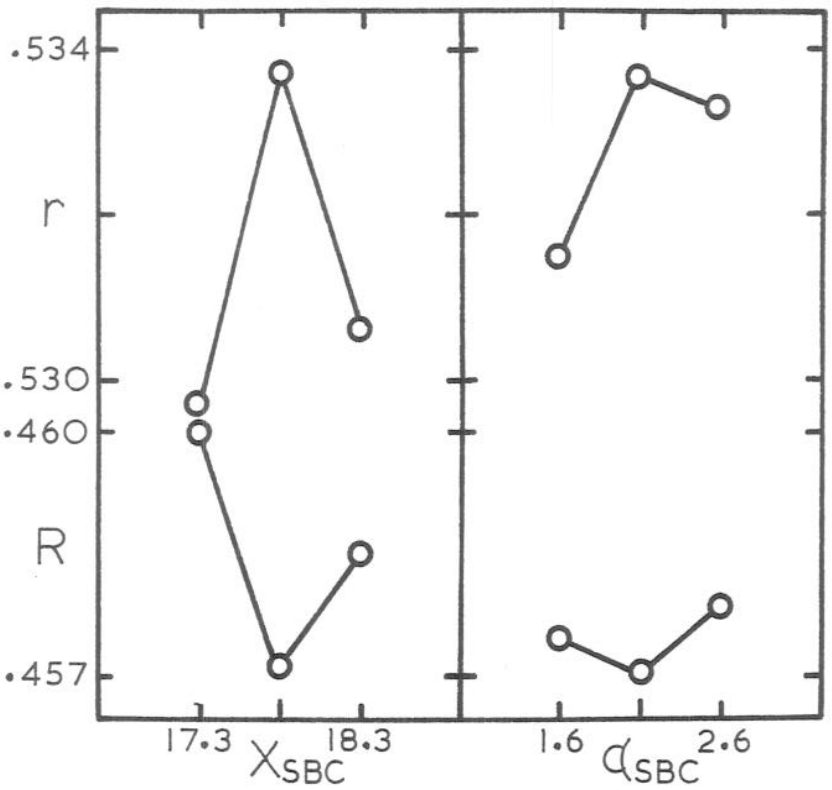
The model had changed sufficiently that an additional round of non-crystallographic symmetry averaging refinement was performed, 2 cycles at 12 Ångstroms resolution followed by 4 cycles at 8 Ångstroms. A total of 29484 phases for non-zero reflections were refined in the latter. The refinement statistics are shown in figure 6.5. There is



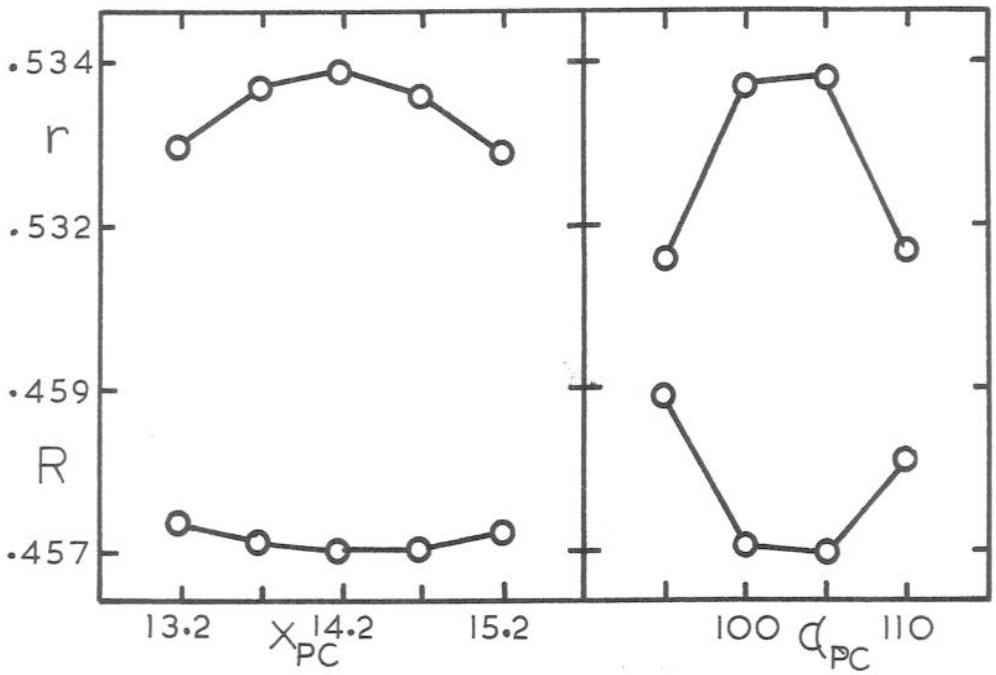
(a) Search for the A position S-domain parameters.



(b) Search for the P-domain AB dimer displacement and rotational parameters.



(c) Search for the B and C position S-domain parameters.



(d) Search for the P-domain CC dimer displacement and rotational parameters.

Figure 6.9. Fine tuning of the displacement and rotational parameters of all the domains by R-factor and correlation coefficient search.

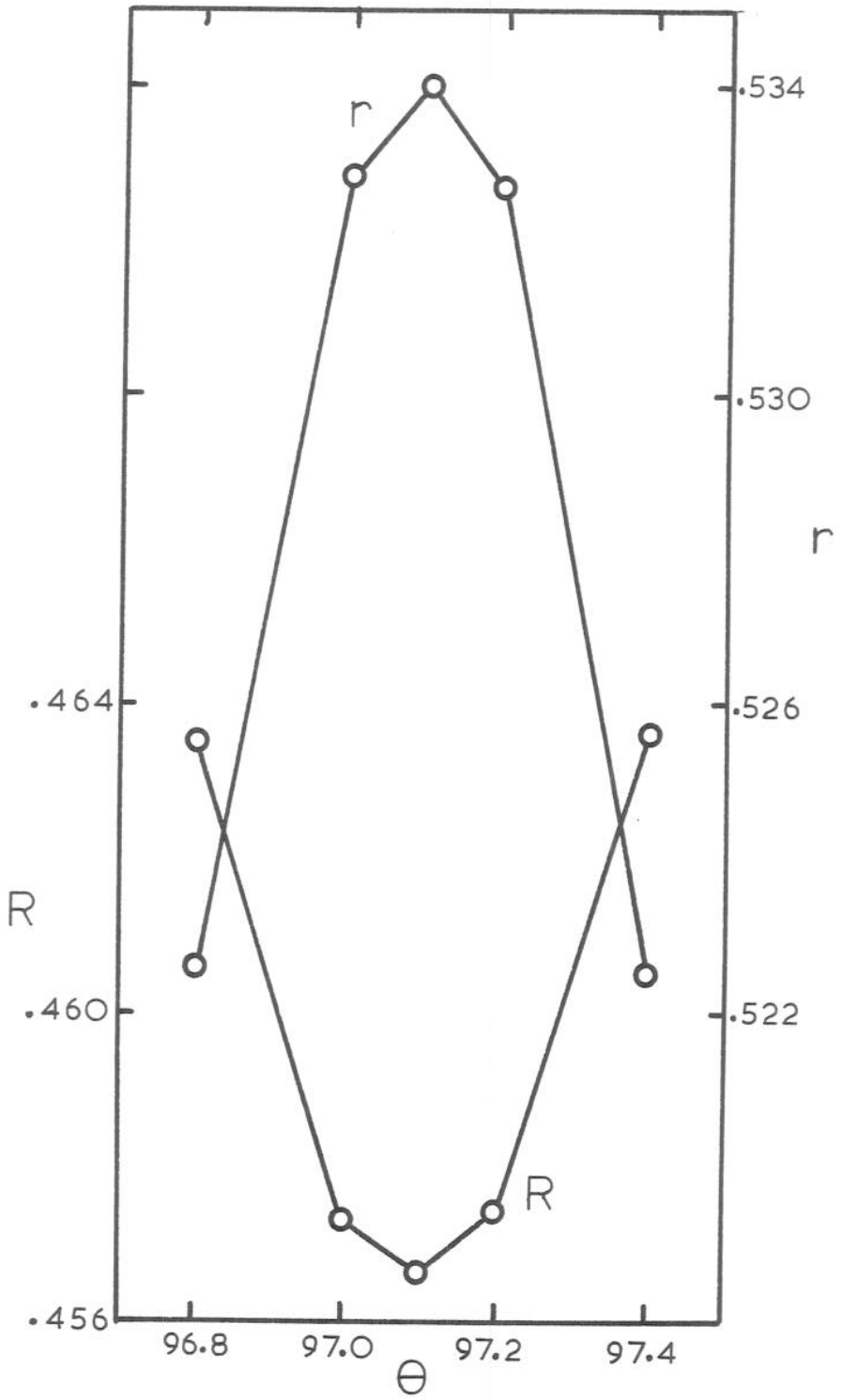


Figure 6.10. Final search for the value of θ, the crystallographic packing angle.

	- - - - - - M O D E L - - - - - -			
	3 param.	9 param.	13 param. (before)	13 param. (after)
x_{SA}	18.	20.0	19.6	19.4
σ_{SA}			3.6	4.5
x_{SBC}	18.	18.3	19.0	17.8
σ_{SBC}			2.9	2.1
x_{PC}	20.	13.8	13.7	14.2
σ_{PC}		108.5	121.0	103.
x_{PAB}	20.	16.8	18.1	18.1
σ_{PAB}		32.3	27.4	30.
AB disp. direction	-.188 .277 .942	-.186 .398 .898	-.137 .352 .962	-.187 .352 .917
AB rotn. direction		-.264 .140 .954	-.264 .140 .954	-.264 .140 .954
θ	97.	97.1	97.1	97.1
R-factor	0.52	.492	.486	0.457

Table 6.8. Comparison of model parameters from different stages of refinement. The 3 parameter model was derived from the first R-factor search alone. The 9 parameter model was after the 12 Ångstrom heavy atom refinement, and the 13 parameter ones after the 7.4 Ångstrom refinement, both before and after optimisation by R-factor search. The R-factor applies to comparisons of model calculated structure factors with observations to 16Å (3 parameter model) or 8Å (remaining three).

a slight improvement in the final agreement of R = .438* and r = .550 over the previous round. The agreement after the two 12 Ångstrom cycles was R = 0.325. The refinement shows the worrying behaviour of the R-factor increasing slightly from cycle to cycle. This indicates either that there is a considerable amount of noise in the data or that the basic icosahedral assumption is not completely upheld. The fact that only 50% of all possible data to 8 Ångstroms were included in the calculation certainly contributes a lot of additional noise over the R = .20 error level intrinsic to the data collection.

This final map was plotted along with the final model psuedoatom positions on acetate sheets and stacked for detailed examination.

---

* These values cannot be compared directly with those from the model calculations because the method of scaling is different. Model calculation R-factors tend to be artificially lower.