

## Accelerated X-ray Structure Elucidation of a 36 kDa Muramidase/Transglycosylase Using *wARP*

ERIK J. VAN ASSELT,<sup>a</sup> ANASTASSIS PERRAKIS,<sup>b†</sup> KOR H. KALK,<sup>a</sup> VICTOR S. LAMZIN<sup>c</sup> AND BAUKE W. DIJKSTRA<sup>a\*</sup>

<sup>a</sup>BIOSON Research Institute and Laboratory of Biophysical Chemistry, Groningen University, Nijenborgh 4, 9747 AG Groningen, The Netherlands, <sup>b</sup>Netherlands Cancer Institute (NKI), Department of Molecular Carcinogenesis, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands, and <sup>c</sup>European Molecular Biology Laboratory (EMBL) Hamburg, c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany. E-mail: bauke@chem.rug.nl

(Received 27 May 1997; accepted 18 July 1997)

### Abstract

The X-ray structure of the 36 kDa soluble lytic transglycosylase from *Escherichia coli* has been determined starting with the multiple isomorphous replacement method with inclusion of anomalous scattering at 2.7 Å resolution. Subsequently, before any model building was carried out, phases were extended to 1.7 Å resolution with the weighted automated refinement procedure *wARP*, which gave a dramatic improvement in the phases. The electron-density maps from *wARP* were of outstanding quality for both the main chain and the side chains of the protein, which allowed the time spent on the tracing, interpretation and building of the X-ray structure to be substantially shortened. The structure of the soluble lytic transglycosylase was refined at 1.7 Å resolution with *X-PLOR* to a final crystallographic *R* factor of 18.9%. Analysis of the *wARP* procedure revealed that the use of the maximum-likelihood refinement in *wARP* gave much better phases than least-squares refinement, provided that the ratio of reflections to protein atom parameters was approximately 1.8 or higher. Furthermore, setting aside 5% of the data for an *R<sub>free</sub>* test set had a negative effect on the phase improvement. The mean *W<sub>wARP</sub>*, a weight determined at the end of the *wARP* procedure and based on the variance of structure factors from six individually refined *wARP* models, proved to be a better indicator than the *R<sub>free</sub>* factor to judge different phase improvement protocols. The elongated Slt35 structure has three domains named the alpha, beta and core domains. The alpha domain contains mainly  $\alpha$ -helices, while the beta domain consists of a five-stranded antiparallel  $\beta$ -sheet flanked by a short  $\alpha$ -helix. Sandwiched between the alpha and beta domains is the core domain, which bears some resemblance to the fold of the catalytic domain of the previously elucidated 70 kDa soluble lytic transglycosylase from *E. coli*. The putative active site is at the bottom of a large deep groove in the core domain.

### 1. Introduction

Almost all bacteria possess a cell wall consisting of peptidoglycan, a biopolymer built up from linear polysaccharide strands of alternating *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) residues, which are cross-linked by short oligopeptide chains (reviewed by Dijkstra & Keck, 1996). The cell wall gives the organism its shape and protects it from lysis. Maintaining the integrity of the cell wall is of vital importance to the bacterium, and compounds that interfere with the correct action of peptidoglycan metabolizing enzymes are often lethal. Prime examples of such antibacterial compounds are penicillins and related  $\beta$ -lactams, which inhibit the enzymes involved in the cross-linking of peptidoglycan. However, the widespread use of such antibiotics has led to the emergence of resistance and novel compounds to inhibit bacterial growth and multiplication are of vital importance (Gold & Moellering, 1996).

Our research is focused on enzymes that are important for the metabolism of the glycan strands, with the objective to characterize their role in peptidoglycan metabolism and to make them amenable for the design of novel antibiotics. Previously, we described the X-ray structures of the 70 kDa soluble lytic transglycosylase (Slt70) from *Escherichia coli* (Thunnissen *et al.*, 1994) and its complex with bulgecin (Thunnissen, Rozeboom, Kalk & Dijkstra, 1995), a naturally occurring Slt70 inhibitor produced by *Pseudomonas acidophila* and *Pseudomonas mesoacidophila*. Here we describe the X-ray structure elucidation of a 36 kDa soluble lytic transglycosylase from *E. coli*, which has been named Slt35 (Engel, Smink, van Wijngaarden & Keck, 1992) and which has no significant sequence similarity to Slt70. Both Slt35 and Slt70 cleave the  $\beta(1,4)$ -glycosidic bond between a MurNAc and GlcNAc residue of peptidoglycan. The cleavage is subsequently followed by an intramolecular transglycosylation reaction, in which the C6-hydroxyl group of MurNAc attacks the C1 atom of the same residue forming an 1,6-anhydromuropeptide product (Höltje, Mirelman, Sharon & Schwarz, 1975; Dijkstra, Hermann & Keck, 1995).

† Present address: EMBL, c/o ILL, BP 156, F-38042 Grenoble CEDEX, France.

The X-ray structure of Slt35 was solved by multiple isomorphous replacement including anomalous differences (MIRAS) with subsequent application of the weighted automated refinement procedure *wARP* (Perrakis, Sixma, Wilson & Lamzin, 1997). The recently developed *wARP* procedure makes use of models consisting of free atoms placed in significant electron density in an initial map. Several so-called pseudo protein models are generated and subjected to unrestrained least-squares or maximum-likelihood refinement and iterative model updating with the automated refinement procedure, *ARP* (Lamzin & Wilson, 1993). To minimize noise and to enhance the overall accuracy of the phases, the calculated structure factors from the refined pseudo protein models are subsequently averaged to obtain the final phase set. In the case of Slt35, *wARP* proved to be very effective in improvement and extension of MIRAS phases to high resolution, prior to any model building. The *wARP* electron-density map was of outstanding quality, considerably facilitating map interpretation and model building. In this article we give details of the structure determination of Slt35 and discuss the influence of various parameters on the effectiveness of *wARP*.

## 2. Materials and methods

### 2.1. Crystallization

The Slt35 protein, encoded by the *mltB* gene, was cloned, overexpressed and purified as previously reported (Dijkstra *et al.*, 1995). As Slt35 is a lipoprotein with the lipid anchor and active domain separated by a proteolysis susceptible spacer, a 36 kDa soluble fragment lacking the membrane anchor and spacer, was constructed starting at residue 41 with two additional residues (Met39 and Val40) at the N terminus. In the presence of 150 mM NaCl, 1% isopropanol, 1% polyethylene glycol 20 000 (PEG 20K) and 20 mM Hepes (pH 7.5), this protein is stable and can be stored at 253 K. For crystallization experiments, the protein was first diluted to a concentration of 7.0 mg ml<sup>-1</sup> and subsequently dialysed for 6 h against a solution containing 10 mM Tris-HCl, pH 7.2, 1% isopropanol and 1 mM phenylmethylsulfonyl fluoride (PMSF).

Rod-shaped crystals with typical sizes of 0.25 × 0.25 × 1.0 mm were grown at 295 K in 1 d to 1 week by equilibrating a hanging drop, that consisted of 3 μl of dialysed protein solution and 3 μl of reservoir solution, against 1 ml of reservoir solution of 100 mM Bicine-NaOH buffer, pH 7.8–8.5 and 0–6% PEG 20K. After growth, crystals were stored in a stabilizing solution containing 100 mM Bicine-NaOH (pH 8.0–8.5), 1% isopropanol, and 5–10% PEG 20K. Slt35 crystallizes in the primitive orthorhombic space group *P*2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> with cell dimensions *a* = 58.3, *b* = 67.9 and

*c* = 98.9 Å. With one molecule per asymmetric unit, the value of the Matthews coefficient, *V<sub>m</sub>* (Matthews, 1968), is 2.72 Å<sup>3</sup> Da<sup>-1</sup>. Assuming a partial specific volume for globular proteins of 0.74 cm<sup>3</sup> g<sup>-1</sup>, the solvent content of the crystals is estimated to be about 55%.

### 2.2. Data collection and processing

For diffraction experiments, crystals were flash frozen in a stream of evaporating nitrogen of 120 K after being soaked for 1 min in a cryo-protectant consisting of 25–35% glycerol added to the stabilizing solution. Data to 1.7 Å resolution were collected from a native Slt35 crystal (stored at pH 8.5) at a wavelength of 0.920 Å at the A1 station at CHESS (Cornell High Energy Synchrotron Source) in Ithaca (USA), using a CCD camera as detector. A second data set to 2.0 Å resolution was collected from the same crystal with the beam attenuated by aluminium foil to measure strong low-angle reflections saturated in the first set. Data were processed with *DENZO* and *SCALEPACK* (Otwinowski, 1993). Intensities were converted to structure-factor amplitudes with programs from the Groningen *BIOMOL* software package (Groningen protein crystallography group). Fig. 1 shows the completeness and the *R<sub>merge</sub>* of the diffraction data as a function of resolution.

In the course of an extensive heavy-atom derivative search, diffraction data were collected on a FAST area detector (Enraf-Nonius, Delft, The Netherlands) using Cu *Kα* radiation from an Enraf-Nonius FR571 rotating-anode generator with a graphite monochromator. Data reduction was performed with *MADNES* (Messerschmidt & Pflugrath, 1987), *XDS* (Kabsch, 1988) and *BIOMOL* programs. Data for a mercury derivative (Hg-III) were measured at beamline BW7A of the EMBL Outstation at DESY, Hamburg (Germany), at a wavelength of 0.970 Å with a MAR Research (Hamburg) imaging plate as detector. These data were processed with *DENZO*. Data-collection statistics are summarized in Table 1.

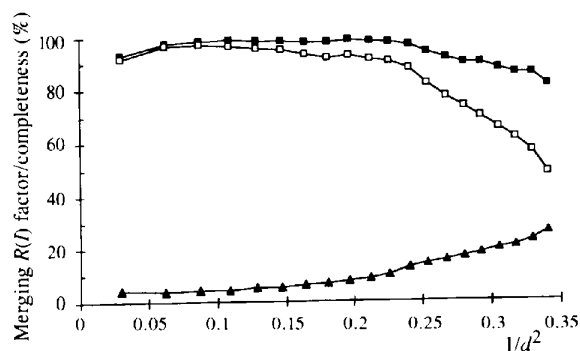


Fig. 1. Completeness and *R<sub>merge</sub>* of the data as a function of resolution. Filled squares indicate the completeness for all reflections with *I* ≥ 0, open squares for those with *I* ≥ 3σ(*I*). The filled triangles show the merging *R* factor ( $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$ ) for symmetry-related and multiply recorded reflections.

Table 1. Data-collection statistics for *Slt35* data sets

	Native	Os-I†	Pb-I†	Hg-I†	Hg-II†	Hg-III†
X-ray source	CHESS A1 station	Rotating anode	Rotating anode	Rotating anode	Rotating anode	EMBL/DESY BW7A
Wavelength (Å)	0.920	1.542	1.542	1.542	1.542	0.970
Detector	CCD	FAST	FAST	FAST	FAST	MAR
Soaking						
Concentration (mM)	—	3.0	12	3.0	2.0	3.0
Time	—	4 d	24 h	6 d	8 d	2 h
Space group $P2_12_12_1$						
Cell dimensions (Å)						
<i>a</i>	58.3	58.3	58.3	58.3	58.4	58.3
<i>b</i>	67.9	67.8	67.5	67.6	67.9	67.7
<i>c</i>	98.9	98.1	98.9	98.7	98.8	98.5
Resolution (Å)	29.6–1.70	33.9–3.00	33.8–3.44	33.8–2.69	33.9–2.70	17.5–2.02
No. of reflections						
Total	206582	23113	21880	20201	23732	92150
Unique	41653	7954	5096	9063	9339	24767
Overall						
Completeness (%)	94.4	96.6	91.8	80.0	82.6	94.0
( $I/\sigma(I)$ )	24.4	22.5	20.4	27.2	26.6	30.4
$R_{\text{merge}} \ddagger$ (%)	5.0	5.6	7.4	3.7	4.2	3.8
$R_{\text{nat}} \S$ (%)	—	9.3	6.0	6.2	7.1	6.9

† Os-I =  $\text{K}_2\text{OsO}_4 \cdot 2\text{H}_2\text{O}$ ; Pb-I = trimethyllead acetate; Hg-I = *para*-chloromercurypheylsulfonic acid; Hg-II = *para*-chloromercury-benzoic acid and Hg-III = 2-chloromercury-4-nitrophenol.  $\ddagger R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$ .  $\S R_{\text{nat}} = \sum_{hkl} |(F_P - F_{PH})| / \sum_{hkl} 0.5(F_P + F_{PH})$ .

Table 2. Phasing statistics for *Slt35*

Heavy-atom compound	Os-I	Pb-I	Hg-I	Hg-II	Hg-III
High-resolution limit (Å)	3.03	3.44	2.70	2.70	2.70
Number of sites	4	2	7	5	6
Phasing power†					
Iso	1.29	0.97	1.56	1.39	0.88
Ano	0.51	0.37	0.69	0.69	1.20
$R_{\text{Cullis}} \ddagger$	0.578	0.657	0.552	0.608	0.657
$R_{\text{Kraut}} \S$					
Iso¶	0.121	0.083	0.074	0.090	0.097
Ano¶¶	0.125	0.083	0.075	0.089	0.099

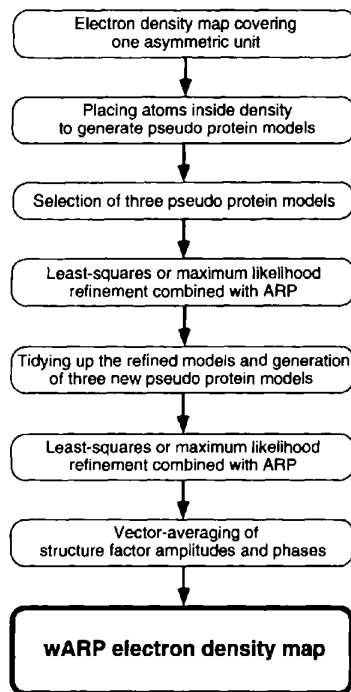
Figure of merit (overall) 0.604 (for 10418 reflections from 24.0 to 2.70 Å resolution)

† Phasing power =  $\sum_{hkl} |F_H(\text{calc})| / \sum_{hkl} (|F_{PH}| - |F_{PH}(\text{calc})|)$ .  
 $\ddagger R_{\text{Cullis}} = \sum_{hkl} (|F_{PH} \pm F_P| - |F_H(\text{calc})|) / \sum_{hkl} |F_{PH} \pm F_P|$ .  
 $\S R_{\text{Kraut}}(\text{iso}) = \sum_{hkl} (|F_{PH}| - |F_{PH}(\text{calc})|) / \sum_{hkl} |F_{PH}|$ .  $\S R_{\text{Kraut}}(\text{ano}) = \sum_{hkl} (|F_{PH^+} - F_{PH^-(\text{calc})}| + |F_{PH^-} - F_{PH^-(\text{calc})}|) / \sum_{hkl} (|F_{PH^+}| + |F_{PH^-}|)$ .

### 2.3. Multiple isomorphous replacement including anomalous scattering

The structure of *Slt35* was solved by multiple isomorphous replacement including anomalous scattering differences using five heavy-atom derivatives:  $\text{K}_2\text{OsO}_4 \cdot 2\text{H}_2\text{O}$  (data set Os-I), trimethyl-lead acetate (Pb-I), parachloromercury-phenylsulfonic acid (Hg-I), parachloromercury-benzoic acid (Hg-II) and 2-chloromercury-4-nitrophenol (Hg-III). The derivative data were scaled to the native data set with programs from the

*PHASES* package (Furey & Swaminathan, 1998). The first two heavy-atom sites for derivatives Os-I, Hg-I and Hg-II were determined from difference Patterson maps. All other heavy-atom positions were located in cross-difference Fourier maps with phases calculated from

Fig. 2. Flow diagram of the *wARP* procedure.

other derivatives. Anomalous scattering information was used to establish the correct hand of the structure and to improve phases. Heavy-atom parameters were refined with *PHASES*, keeping individual atomic displacement parameters constant at  $20.0 \text{ \AA}^2$ . The MIRAS map was solvent flattened using *PHASES* assuming 50% solvent content. Heavy-atom parameters, now including the atomic displacement parameters, were further improved by refinement against the phases resulting from solvent flattening. Although the data for Hg-III nominally extend to  $2.0 \text{ \AA}$  resolution, the phasing power was low beyond  $2.7 \text{ \AA}$ , and, therefore, this derivative was only used to  $2.7 \text{ \AA}$  resolution (Table 2). The final figure of merit after completion of the heavy-atom-parameter refinement was 0.60 at  $2.7 \text{ \AA}$  resolution.

#### 2.4. Phase extension and improvement using *wARP*

The recently developed *wARP* procedure (Perrakis *et al.*, 1997) was used to extend and improve the phases for Slt35. The entire method can be divided in seven major steps as can be seen in the flow diagram (Fig. 2). In the following we discuss the use of these steps in more detail.

**2.4.1. Input electron-density map.** To apply the *wARP* procedure for phase extension and improvement, an input electron-density map with phases extending to  $2.5 \text{ \AA}$  resolution is recommended (Perrakis *et al.*, 1997). Therefore, MIRAS phases to  $2.7 \text{ \AA}$  resolution were extended to  $2.5 \text{ \AA}$  resolution using the program *DM* (Cowtan & Main, 1993) with solvent flattening and histogram matching. The phases to  $2.5 \text{ \AA}$  resolution were

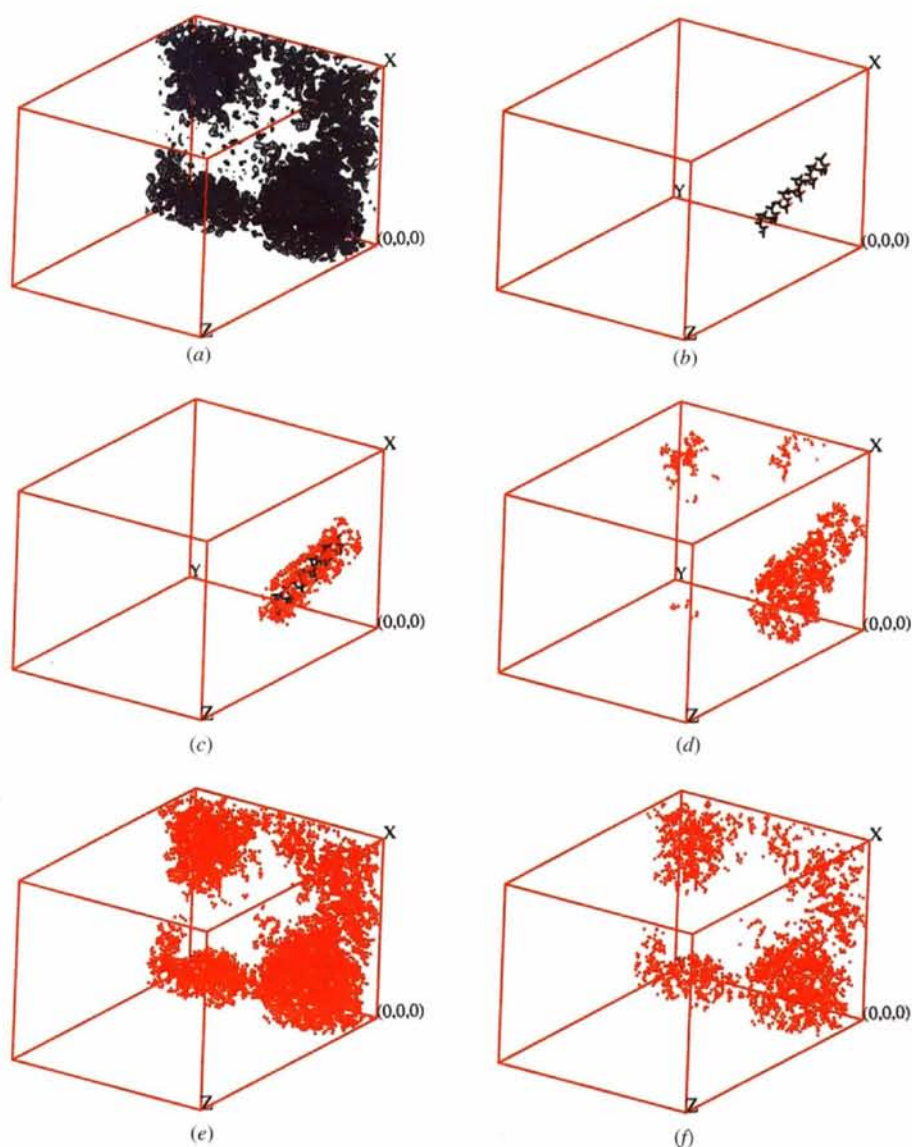


Fig. 3. *wARP* procedure of placing atoms in electron density. (a)  $2.5 \text{ \AA}$  resolution electron-density map after phase extension with *DM* contoured at  $1.0\sigma$  and covering one asymmetric unit of the cell. For clarity, the map is not shown in (b)–(e). (b) A polyaniline  $\alpha$ -helix of 20 residues placed in the protein part of the electron density. (c) 600 atoms placed near the  $\alpha$ -helical seed. (d) Atoms placed in  $2.0\sigma$  density after removal of the  $\alpha$ -helical seed. (e) Pseudo protein model containing 8187 atoms after finishing the placement of atoms in electron density. (f) Pseudo protein model consisting of 2750 atoms after truncation. See §2 for details.

subsequently used to calculate a figure-of-merit weighted  $F_o$  electron-density map covering the asymmetric unit with a very fine grid spacing of 0.25 Å.

**2.4.2. Placing atoms inside density to generate pseudo protein models.** *wARP* makes use of models consisting of free atoms placed in significant electron density. As these atoms are not bonded to each other we call these sets of atoms pseudo protein models. The generation of pseudo protein models requires a seed of approximately 100 atoms placed somewhere in the protein region of the input electron-density map. For that purpose, we took a polyalanine helix of 20 residues which was placed in an arbitrary orientation inside density representing the protein molecule. The building of the pseudo protein models was initiated by placing O atoms with the program *ARP* (Lamzin & Wilson, 1993) at grid points with a density of at least  $2.0\sigma$  above the mean, and within 5.0 Å from an atom in the polyalanine helix. After the first cycle, in which 600 atoms were placed, the initial polyalanine helix was removed and the pseudo protein model was slowly expanded in the next cycles by adding gradually atoms in significant electron density at distances 1.1 to 1.8 Å from existing atoms. When no more atoms could be found that obeyed these conditions, the density threshold was lowered in steps of  $0.1\sigma$  from 2.0 to  $0.9\sigma$  to place more atoms in weaker sections of the electron density. Occasionally, breaks in the electron density are present in the protein main chain or between a side chain and the main chain. Therefore, by placing atoms in weaker density, good density regions become connected to each other. Fig. 3 illustrates the procedure of placing atoms inside electron density. The intermediate sets of atoms generated at the various levels were kept for the next step in the *wARP* procedure.

**2.4.3. Selection of three pseudo protein models.** In the Slt35 crystal, there is one protein molecule per asymmetric unit containing approximately 2 500 atoms. In the *wARP* procedure, three starting models were selected from the atom sets generated in the previous step that contained approximately 1.5, 2.0 and 3.0 times the expected number of protein atoms in the asymmetric unit of the cell. Subsequently, the number of atoms in each of these models was reduced to 110% of the expected number of atoms (2 750) to allow for strongly bound solvent molecules, by rejecting atoms in the weakest density. The three atom sets are called pseudo protein models 1, 2 and 3. The individual displacement parameters for all atoms were set to the  $B$  factor derived from a Wilson plot (Wilson, 1949),  $18.4 \text{ \AA}^2$ .

**2.4.4. Refinement of the first three pseudo protein models.** The three pseudo protein models were refined in a cyclic procedure of unrestrained reciprocal-space refinement either with *SFALL* (least-squares refinement; Agarwal, 1978; Collaborative Computational Project, Number 4, 1994) or with *REFMAC* (maximum-likelihood refinement; Murshudov, Vagin & Dodson, 1997), followed by real-space refinement and iterative model

Table 3. *Refinement statistics of Slt35*

Resolution range (Å)	20.0–1.70
Number of reflections	
Total	41498
Work set	37284
Test set	4214
Final $R$ factor (%)†	18.9
$R_{\text{work}}$ factor (%)‡	18.7
$R_{\text{free}}$ factor (%)§	22.1
R.m.s.d. bond lengths (Å)¶	0.009
R.m.s.d. bond angles (°)¶	1.4
Non-H protein atoms	2456
Residues	42–97 and 109–361
Alternate side chains	5
Solvent molecules	312
Average $B$ factor (Å <sup>2</sup> )	
All atoms	23.4

† Final  $R$  factor =  $\frac{\sum_{hkl} [|F_o(hkl, \text{all})| - |F_c(hkl, \text{all})|]}{\sum_{hkl} |F_o(hkl, \text{all})|}$ .  
 ‡  $R_{\text{work}}$  factor =  $\frac{\sum_{hkl} [|F_o(hkl, \text{work set})| - |F_c(hkl, \text{work set})|]}{\sum_{hkl} |F_o(hkl, \text{work set})|}$ .  
 §  $R_{\text{free}}$  factor =  $\frac{\sum_{hkl} [|F_o(hkl, \text{test set})| - |F_c(hkl, \text{test set})|]}{\sum_{hkl} |F_o(hkl, \text{test set})|}$ .  
 ¶ With respect to the set of Engh & Huber parameters (Engh & Huber, 1991).

updating by *ARP*. Both atomic positional and displacement parameters were refined. Depending on the high-resolution limit (Res) and the number of atoms ( $N$ ), a maximum of  $(N \times 0.08/\text{Res}^3)$  atoms were located in  $F_o - F_c$  difference density above the  $3.5\sigma$  level and added to the model in each cycle. Likewise, the same number of atoms was allowed to be rejected from the model for one of the following reasons. (1) If the value of the  $2mF_o - DF_c$  electron density at the atomic centre was less than  $1.5\sigma$  above the mean or (2) if any pair of atoms was closer than 0.6 Å to each other, one atom was rejected and the other was assigned to the mean position of the two atoms. In total, 20 cycles of refinement with *SFALL* or *REFMAC* and *ARP* were performed for each of these three starting models. The progress of the refinement was followed with an  $R_{\text{free}}$  factor using a test set containing 5% of randomly selected observations (Brünger, 1992a). For reasons of comparison, a low-resolution cutoff of 8.0 Å was chosen for both least-squares and maximum-likelihood refinement as no bulk solvent correction could be applied within *SFALL*. Several different high-resolution limits were taken to test the convergence of *wARP* (see §3).

**2.4.5. Tidying up the refined models and generation of three new pseudo protein models.** The three refined pseudo protein models were slightly changed in two steps. First, atoms with an atomic displacement parameter larger than two times the Wilson  $B$  factor were discarded. After that, the displacement parameters of all remaining atoms were reset to the  $B$  factor derived from the Wilson plot. To create three new different atomic models, a random positional shift of 0.5 Å was applied to

Table 4. Location of the heavy-atom sites in the Slt35 structure

Derivative	Site	x	y	z	B	Occupancy	Protein ligands/remarks
Os-I	1	0.72	0.10	0.81	36	1.00	Tyr234, His263
	2	0.60	0.08	0.82	37	0.70	His340
	3	0.52	0.07	0.83	52	0.37	His340
	4	0.57	0.06	0.79	53	0.37	His340
Pb-I	1	0.70	0.10	0.81	14	1.00	His263
	2	0.23	0.08	0.05	18	0.94	Tyr223
Hg-I	1	0.35	0.24	1.00	19	1.00	Asp208, Asn245
	2	0.94	0.22	1.00	18	0.96	Asp205, Glu206, His243
	3	0.70	0.10	0.81	20	0.23	Tyr234, His263
	4	0.55	0.09	0.82	18	0.26	His340
	5	0.25	0.21	0.97	29	0.19	Probably sulfonic acid moiety of the Hg-complex in site 1
	6	0.42	0.25	0.01	14	0.16	Glu206, His243; probably sulfonic acid moiety of the Hg-complex in site 2
	7	0.03	0.18	0.01	29	0.21	Val168, Gly170, Lys171, Gln207 of symmetry-related molecule.
Hg-II	1	0.94	0.22	1.00	19	1.00	Asp205, Glu206, His243
	2	0.35	0.24	0.99	19	0.92	Asp208, Asn245
	3	0.55	0.09	0.82	21	0.62	His340
	4	0.70	0.11	0.81	22	0.37	Tyr234, Tyr259, His263, Arg337
Hg-III	5	0.62	0.10	0.81	23	0.28	His340, Arg337
	1	0.72	0.10	0.81	27	1.00	Tyr234, Tyr259, His263
	2	0.61	0.09	0.82	32	0.51	His340
	3	0.53	0.09	0.81	33	0.48	His340
	4	0.86	0.21	0.99	28	0.51	Glu206
	5	0.94	0.21	0.99	34	0.26	Asp208, Ile244
6	0.35	0.24	1.00	22	0.16	Glu206, His243	

all atom coordinates of the modified models using the *MOLEMAN* program (Kleywegt, 1995). The application of random shifts and resetting of *B* factors, together with the removal of atoms in weak density gives a considerable improvement of the refinement at the cost of negligible computing time (see *e.g.* Table 5*d* and 5*e*). Apparently, shaking the models is very effective in escaping local refinement minima. The newly generated atom sets were called pseudo protein models 4, 5 and 6, respectively.

**2.4.6. Refinement of all six pseudo protein models.** Each of the six pseudo protein models was subsequently refined with *SFALL* or *REFMAC* and updated with *ARP* in 20 cycles, as described in §2.4.4. At the end of these refinement cycles, atoms located in electron density below  $2.0\sigma$  in a  $2mF_o - DF_c$  map were removed. Subsequently, a random shift of 0.5 Å was applied to all remaining atoms in the six models and the atomic displacement parameters were truncated to a minimum of half the Wilson *B* factor and to a maximum of the value of the Wilson *B* factor. The modified models were subsequently refined in 20 additional cycles.

**2.4.7. Vector-averaging of structure factors.** Structure-factor amplitudes and phases were calculated with *SFALL* for each of the six refined models, and scaled to the observed amplitudes with the program *RSTATS* (Collaborative Computational Project, Number 4, 1994). These

calculated structure factors were vector-averaged using the program *REFLAVER* (Perrakis *et al.*, 1997) to minimize noise and to enhance the overall accuracy of the phases. To each structure factor a weight,  $W_{wARP}$ , was assigned based on the variance of the two-dimensional distribution of the individual structure factors around the average. The value of  $W_{wARP}$  is 1.0, if the structure factors from all six models are the same. Large differences between structure factors will lead to a smaller  $W_{wARP}$  value with zero as a minimum. With the final vector-

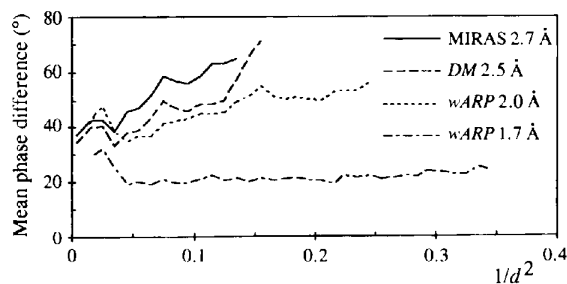


Fig. 4. Mean phase differences with the phases calculated from the final refined Slt35 structure as a function of resolution with (—) MIRAS phases to 2.7 Å resolution, (---) DM-modified phases to 2.5 Å resolution, (· · · ·) vector-averaged phases to 2.0 Å resolution using maximum-likelihood refinement in *wARP* (used to build the Slt35 structure), and (- · - ·) vector-averaged phases to 1.7 Å resolution using maximum-likelihood refinement.



averaged structure-factor amplitudes, phases and  $W_{wARP}$  weights,  $F_o$  and  $2F_o - F_c$  maps were computed.

### 2.5. Model building and refinement

Model building was performed with the interactive graphics program *O* (Jones, Zou, Cowan & Kjeldgaard,

1991) on Silicon Graphics workstations. The quality of the *wARP* map allowed to take complete advantage of the model-building procedure included in the *O* package and especially the real-space refinement capabilities. The model was refined with the program *X-PLOR* version 3.843 (Brünger, 1992*b*) starting with residues 51–99 and 109–360. All atomic displacement parameters were set to  $20.0 \text{ \AA}^2$ . For a test set, 10% of randomly selected

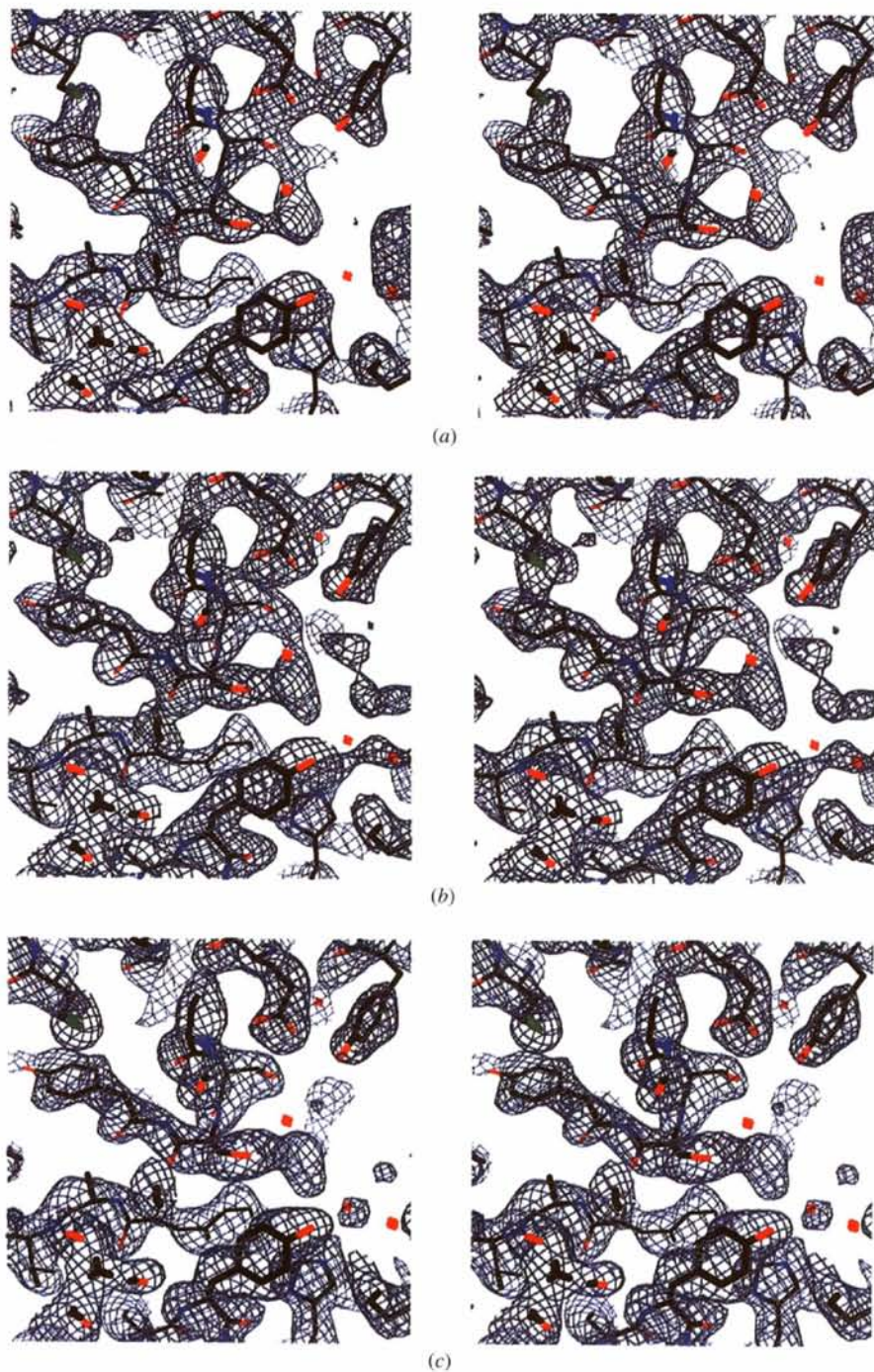


Fig. 5.  $F_o$  electron-density maps contoured at  $1.0\sigma$  at various stages of the structure determination with the final refined SlT35 model superimposed. Maps were calculated with (a) MIRAS phases and FOM weights to 2.7 Å resolution, (b) DM phases and FOM weights to 2.5 Å resolution, (c) vector-averaged phases and  $W_{wARP}$  weights to 2.3 Å using least-squares refinement in *wARP*.



reflections were set aside (Brünger, 1992*a*). After two cycles of refinement including simulated annealing (3000–300 K), Powell energy minimization, overall and individual  $B$ -factor refinement and model rebuilding, the crystallographic  $R_{\text{work}}$  and  $R_{\text{free}}$  factors were respectively 28.3 and 31.4% for data between 8.0 and 1.7 Å resolution and with residues 48–99 and 109–361 included in the model.

In six subsequent refinement and model-rebuilding cycles, the low-resolution limit for the diffraction data was extended to 20.0 Å and a bulk solvent correction was applied. A part of the N terminus (residues 42–47) showed up in a  $\sigma_A$ -weighted  $2F_o - F_c$  omit map (Read, 1986; Bhat, 1988; Vellieux & Dijkstra, 1997) and these residues were added to the model. Residues 98 and 99 had poor electron density and were removed. The final

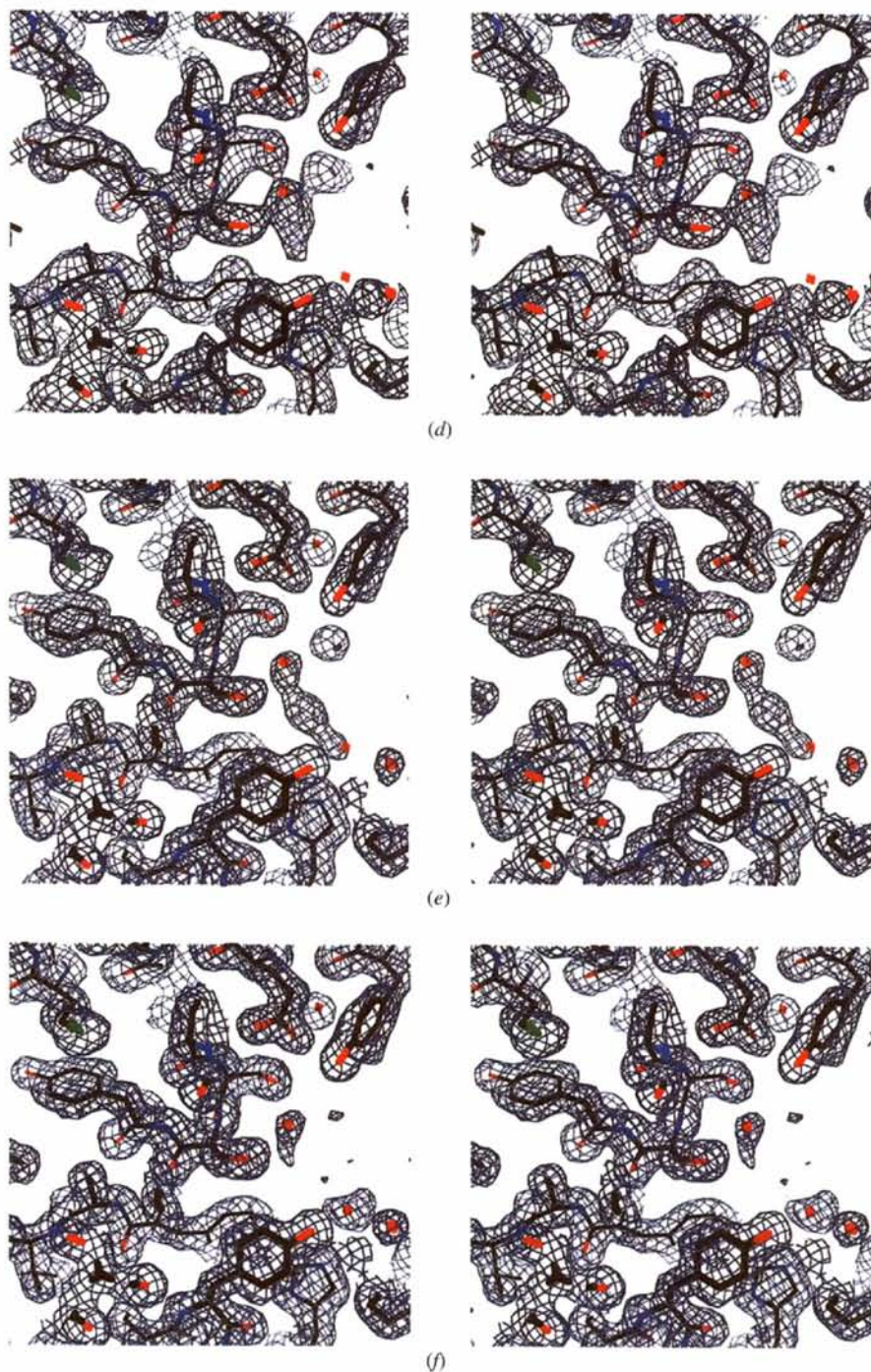


Fig. 5. (cont.) (d) Vector-averaged phases and  $W_{\text{wARP}}$  weights to 2.0 Å resolution using maximum-likelihood refinement in *wARP* (used to build the Slit35 structure), (e) vector-averaged phases and  $W_{\text{wARP}}$  weights to 1.7 Å resolution using maximum-likelihood refinement in *wARP* and (f)  $\sigma_A$ -weighted  $2F_o - F_c$  map using phases of the final Slit35 model. The map correlation coefficients are, respectively, (a) 60.2, (b) 74.3, (c) 78.2, (d) 80.8 and (e) 94.5%, with respect to a electron-density map calculated with phases from the final Slit35 structure.



protein model contains residues 42–97 and 109–361. Alternate side-chain conformations were observed and built for Asp74, Asp139, Trp145, Glu153 and Arg204. Occupancies for double conformations were refined with the total occupancy of each pair constrained to 1.0 in *X-PLOR* using a locally written script (E. van Asselt, unpublished work). In addition, 312 solvent molecules were gradually placed in electron density of at least  $3\sigma$  above the mean in a difference Fourier map, within a distance of 2.4–3.5 Å from a neighbouring O or N atom, and provided that good density in a  $\sigma_A$ -weighted  $2F_o - F_c$  omit map was present. The final refinement statistics are summarized in Table 3. The last refinement cycle was repeated with all data including the test set and resulted in a structure with an *R* factor of 18.9%. Stereochemistry was inspected with the programs *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993) and *WHAT CHECK* (Vriend & Sander, 1993).

### 2.6. Comparison of phase sets

Phase differences between phase sets were calculated with the programs *PH\_RMS* (V. Lamzin, unpublished work) and *SFTOOLS* (*BIOMOL* software). Map correlation coefficients were calculated with the program *MAPMAN* (Kleywegt & Jones, 1996) and *OVERLAP-MAP* (Collaborative Computational Project, Number 4, 1994) by comparing  $F_o$  electron-density maps computed with the observed structure-factor amplitudes and selected phase sets. The different pseudo protein models were analysed for common atoms with the program *MCFMAN* (*BIOMOL* software).

## 3. Results and discussion

### 3.1. *MIRAS* phasing

The X-ray structure of Slt35 was elucidated using five heavy-atom derivatives with inclusion of anomalous scattering differences (Tables 1 and 2). Although derivatives Hg-I, Hg-II and Hg-III share similar binding sites, all sets were useful for the phasing, as the occupancies of the sites differed significantly. Furthermore, Hg-I and Hg-II contributed in particular to the isomorphous phasing power, while Hg-III was collected with synchrotron radiation at a wavelength optimized for anomalous scattering differences. Another reason to include both Hg-I and Hg-II in the phasing was that these derivatives had a poor completeness of 43.2 and 57.1%, respectively, in the highest resolution shell (2.91–2.70 Å), but a more reasonable combined completeness of 72.0%.

The heavy-atom binding sites were examined in the final Slt35 structure (Table 4). In all derivatives, one common heavy-atom binding site was present near residue His263 with additional binding contributions from Tyr234, Tyr259 or Arg337. A second common

binding site for the mercury and osmium derivatives was observed close to another histidine residue (His340). The side chain of this histidine is exposed to the solvent explaining why the mercury and osmium compounds bind at several slightly different positions to this residue. As all heavy-atom soaking experiments have been carried out at pH 8.0 (only Pb-I was soaked at pH 8.5), the unprotonated histidine imidazole ring appeared to be a good nucleophile to displace chloride from the mercury compounds (Petsko, 1985). Also His243 provided a mercury binding site. In total, three out of the seven histidine residues in Slt35 are involved in the binding of the mercury compounds. The remaining histidine residues are involved in a hydrogen-bonding network and are not accessible to the heavy-atom compounds.

### 3.2. The initial map and phase improvement by standard methods

An electron-density map calculated with the *MIRAS* phases to 2.7 Å resolution did indicate elements of the secondary structure. However, several regions were not readily interpretable. Solvent flattening, histogram matching and phase extension (Cowtan & Main, 1993) to 2.5 Å resolution decreased the mean phase difference with the phases derived from the final refined model from 54.3 to 45.4° (24.0–2.7 Å, Fig. 4). In the 2.7–2.5 Å resolution range, the inclusion of histogram matching in the solvent flattening and phase extension procedure had a very favourable effect on the extended phases, which had their mean phase error decreased from 79.5 to 68.1°. The overall mean phase error was 50.0° (24.0–2.5 Å). Nevertheless, the resulting 2.5 Å map (Fig. 5*b*), although improved, still had regions that were very difficult to interpret.

### 3.3. Phase improvement and extension with *wARP*

To further improve the phases we applied the *wARP* procedure. Initially, this was carried out with data to 2.0 Å resolution, which resulted in a map that was readily interpretable (Fig. 5*d*). Main chain and side chains could be easily recognised and, therefore, the building of the Slt35 model was very easy. Subsequently, even better results were obtained using data extending to 1.7 Å resolution (Fig. 4 and 5*e*).

In the following sections we analyse the capabilities and the limitations of *wARP*. The influence of parameters such as the resolution of the diffraction data and the reciprocal-space refinement method is discussed. The use of the  $R_{\text{free}}$  factor, as opposed to the standard *R* factor combined with *wARP* weights, to judge the effectiveness of *wARP* is also discussed.

**3.3.1. Refinement of the pseudo protein models.** The central concept of *wARP* is that independent refinement of several pseudo protein models, consisting of free atoms only, at sufficiently high resolution will lead to sets

Table 5. Comparison of the various pseudo protein models

(a) Comparison of the three starting models

Pseudo protein model	$\sigma$ cutoff	No. of atoms placed	No. of atoms selected	Common atom positions to model†	Mean phase differences (8.0–1.7 Å) with model			
					2	3	2	3
1	1.6	4160	2750	2518	2676	26.1	24.6	67.4
2	1.3	5946	2750	—	2674	—	25.7	67.6
3	1.0	8187	2750	2674	—	25.7	—	67.1

(b) Final mean phase differences (°) between *SFALL/ARP*-refined pseudo protein models with each other, their averaged phase set and the final Slt35 model. Data from 8.0 to 1.7 Å resolution was used during the refinement, with 5% of the data set aside for a test set.

Model	1	2	3	4	5	6	Averaged	Final Slt35 model
1	—	55.9	54.9	48.6	57.7	57.0	35.6	53.4
2		—	55.8	58.1	49.1	58.1	36.6	54.7
3			—	57.1	57.5	48.4	35.8	55.2
4				—	58.7	59.1	37.9	56.1
5					—	58.8	38.3	55.8
6						—	38.0	56.8
Averaged							—	44.9

(c) Final mean phase differences (°) between *SFALL/ARP*-refined pseudo protein models with each other, their averaged phase set and the final Slt35 model. All data from 8.0 to 1.7 Å resolution was used during the refinement, including the data in the test set.

Model	1	2	3	4	5	6	Averaged	Final Slt35 model
1	—	53.7	56.1	46.9	57.1	56.5	35.7	53.6
2		—	53.9	55.7	46.1	55.2	33.9	51.9
3			—	57.0	56.2	46.7	35.7	54.0
4				—	57.8	57.4	37.1	54.4
5					—	57.7	37.1	53.5
6						—	36.6	54.3
Averaged							—	43.2

(d) Final mean phase differences (°) between *REFMAC/ARP* refined pseudo protein models with each other, their averaged phase set and the final Slt35 model. Data from 8.0 to 1.7 Å resolution was used during the refinement, with 5% of the data set aside for a test set.

Model	1	2	3	4	5	6	Averaged	Final Slt35 model
1	—	31.6	34.3	26.9	29.2	31.8	19.3	30.2
2		—	35.0	31.1	29.6	34.1	21.4	32.8
3			—	35.0	35.3	32.0	24.0	36.1
4				—	27.8	31.8	19.0	28.1
5					—	31.4	19.1	29.0
6						—	21.3	32.8
Averaged							—	24.3

(e) Final mean phase differences (°) between *REFMAC/ARP*-refined pseudo protein models with each other, their averaged phase set and the final Slt35 model. All data from 8.0 to 1.7 Å resolution was used during the refinement, including the data in the test set.

Model	1	2	3	4	5	6	Averaged	Final Slt35 model
1	—	25.5	26.2	23.4	24.7	24.5	15.9	27.3
2		—	26.9	25.3	23.6	25.5	16.7	27.6
3			—	26.4	26.4	24.7	17.7	28.7
4				—	23.6	24.3	15.6	26.1
5					—	24.5	15.5	25.8
6						—	15.8	25.9
Averaged							—	21.7

† 2450 atoms were placed at common positions (same grid points) in all three starting pseudo protein models.

of phases, which can be combined into a new set of phases with lower noise and less systematic error than those from each of the individual models. In the three starting models about 90% of the atoms (2 450 out of 2 750) occupy the same positions (Table 5a), leading to modest mean phase differences between the models of about  $25^\circ$ . In contrast, the phases from each of the individual models differ by about  $67^\circ$  from the final Slt35 structure (Table 5a).

At the end of a *REFMAC/ARP* refinement to 1.7 Å resolution, the six pseudo protein models still had mean phase differences of around  $25^\circ$  with each other, but only about  $27^\circ$  with the final refined Slt35 structure (Table 5e). Although at this resolution stand-alone *REFMAC/ARP* refinement of a single pseudo protein model is thus sufficient to eliminate random noise to a large extent and to dramatically improve the phases, vector averaging of structure factors from the six refined models results in phases that are even closer to those from the final Slt35 model (mean phase difference of  $21.7^\circ$ ). This shows clearly that vector averaging is a very effective method in *wARP* to further improve phases without the need for any prior map interpretation or model building. Fig. 5 compares representative parts of the electron-density map at the various stages of the structure determination. Our results are in agreement with earlier observations that averaging of structure factors from a family of refined structures, reduces the effect of local errors (Rice & Brünger, 1994; Brünger, Adams & Rice, 1997).

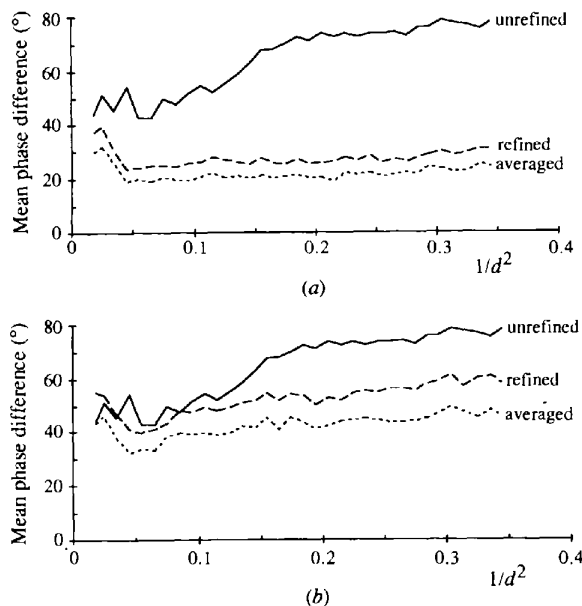


Fig. 6. Mean phase differences as a function of resolution between a pseudo protein model and the final refined Slt35 structure at the beginning of the refinement (—), at the end of the refinement (---), and after averaging of all six refined pseudo protein models (----) using *REFMAC/ARP* (a) or *SFALL/ARP* (b) with diffraction data between 8.0 and 1.7 Å resolution. No test set was used during the refinement.

3.3.2. *Comparison of different refinement protocols.* *wARP* allows the use of different refinement programs. In our analysis we compared *REFMAC* (maximum-likelihood refinement) and the least-squares refinement program *SFALL*. Both refinement protocols worked well in combination with *ARP* in our hands when using data extending to 1.7 Å resolution. However, Fig. 6 shows clearly that *REFMAC* produces much better phases than *SFALL*. Not only is the overall phase error lower, but it also increases less with resolution.

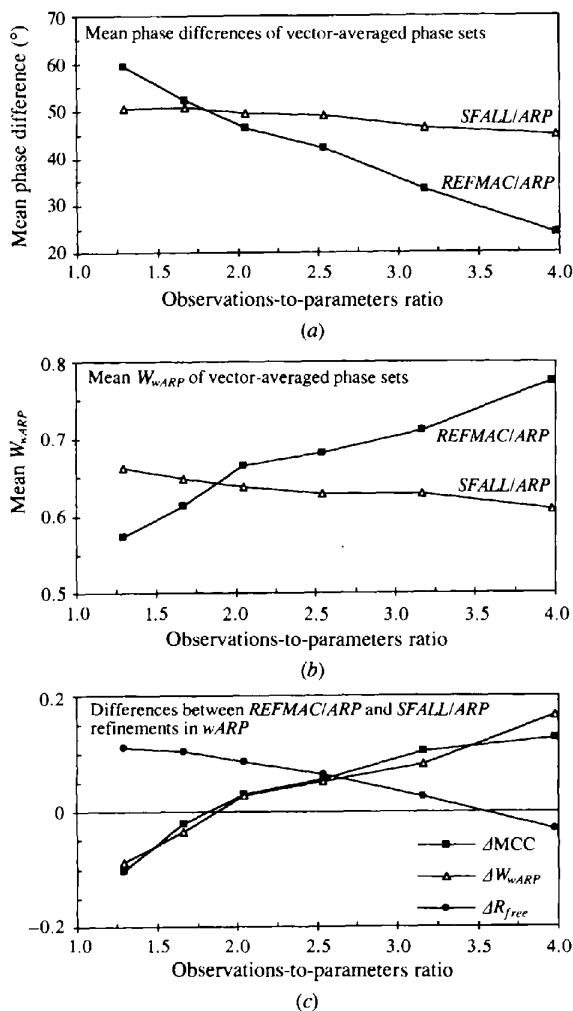


Fig. 7. Analysis of the effectiveness of *wARP* as a function of the number of observations in the work set divided by the number of protein atom parameters ( $x$ ,  $y$ ,  $z$  and  $B$  for 2456 atoms). (a) The mean phase difference to the phases from the final refined Slt35 structure and (b) the mean  $W_{wARP}$  from vector averaging of pseudo protein models refined with *REFMAC/ARP* (filled squares) or *SFALL/ARP* (open triangles). (c)  $\Delta MCC$ , the map correlation coefficient (MCC) to the final map for *REFMAC/ARP* refinement minus that for the *SFALL/ARP* refinement.  $\Delta W_{wARP}$ , the difference in mean  $W_{wARP}$  of *REFMAC/ARP* and *SFALL/ARP* refinement, which is fully correlated with  $\Delta MCC$ , while  $\Delta R_{free}$ , the difference in  $R_{free}$  of *REFMAC/ARP* and *SFALL/ARP* refinement, shows completely opposite behaviour. 5% of the reflections was set aside for a test set during the refinement. See Table 6 for more details.



Table 6. Statistics of vector-averaged structure factors from *REFMAC/ARP* or *SFALL/ARP* refinements at various resolutions

Resolution range (Å)	8.00–2.50	8.00–2.30	8.00–2.15	8.00–2.00	8.00–1.85	8.00–1.70	8.00–1.70†
Observations-to-parameters ratio‡	1.30	1.67	2.04	2.54	3.17	3.98	4.18
<i>(a) REFMAC/ARP refinement</i>							
Mean phase difference (°)§	59.3	52.2	46.4	42.2	33.4	24.3	21.7
$R_{\text{work}}$ factor (%)	18.0	18.0	16.1	16.0	16.0	15.0	15.2
$R_{\text{free}}$ factor (%)	46.0	42.0	39.0	35.6	31.5	25.2	—
Mean $W_{\text{wARP}}$	0.574	0.613	0.665	0.681	0.710	0.774	0.817
Map correlation coefficient§							
Using no $W_{\text{wARP}}$ weights	0.653	0.740	0.802	0.837	0.891	0.942	0.954
Using $W_{\text{wARP}}$ weights	0.685	0.764	0.820	0.852	0.900	0.945	0.956
<i>(b) SFALL/ARP refinement</i>							
Mean phase difference (°)§	50.4	50.8	49.6	49.0	46.4	44.9	43.2
$R_{\text{work}}$ factor (%)	13.7	13.1	14.0	15.1	16.1	17.3	17.3
$R_{\text{free}}$ factor (%)	34.8	31.6	30.4	29.2	29.0	28.3	—
Mean $W_{\text{wARP}}$	0.662	0.648	0.637	0.628	0.628	0.608	0.618
Map correlation coefficient¶							
Using no $W_{\text{wARP}}$ weights	0.755	0.760	0.772	0.780	0.787	0.814	0.825
Using $W_{\text{wARP}}$ weights	0.776	0.782	0.794	0.801	0.816	0.830	0.841

† All diffraction data were used in the refinement, including the data in the test set. ‡ The observations-to-parameters ratio is here defined as the number of reflections in the work set divided by the number of protein atom parameters ( $x$ ,  $y$ ,  $z$  and  $B$  times 2456 atoms). § Mean phase differences between vector-averaged phases from the final Slt35 structure. ¶ Using the vector-averaged phases, non- $W_{\text{wARP}}$  weighted and  $W_{\text{wARP}}$  weighted  $F_o$  electron-density maps were calculated and compared with a  $F_o$  electron-density map calculated with phases of the final refined Slt35 structure. The map correlation coefficient for the  $F_o$  electron-density map calculated with the *DM*-modified phases to 2.50 Å resolution was 0.732 and 0.743 if FOM weights were used. The mean phase error for the *DM*-modified phases was 50.0°.

To further analyse the efficacy of *wARP*, we applied the two refinement protocols at different resolutions to the same starting pseudo protein models. As the number of observations in a data set does not only depend on the resolution but also on the solvent content of the crystal and the completeness of the data, we prefer to use the observations-to-parameters ratio. Fig. 7*a* shows that from an observations-to-parameters ratio higher than about 1.8, *REFMAC/ARP* refinement gives lower phase errors than *SFALL/ARP* refinement. From this we suggest that also for other crystals the *REFMAC/ARP* protocol is preferable provided that the observations-to-parameters ratio is sufficient. It is also evident from the figure that the more reflections are included in the refinement, the larger the phase improvement is.

3.3.3. *The  $R_{\text{free}}$  factor.* Initially we thought that the  $R_{\text{free}}$  factor was a good indicator to compare the different phase-improvement protocols in the *wARP* procedure. However, Table 6 shows that there is no clear correlation between  $R_{\text{free}}$  factor and phase error. *REFMAC* produces lower phase errors and higher map correlation coefficients than *SFALL* at resolution limits of 2.15, 2.00 and 1.85 Å, while the  $R_{\text{free}}$  factor is higher. This indicates that a low  $R_{\text{free}}$  value does not necessarily mean lower phase errors, and therefore the  $R_{\text{free}}$  factor is not a useful parameter to compare least-squares and maximum-likelihood refinement in *wARP*.

A more important point to be considered is that the use of the  $R_{\text{free}}$  factor decreases the number of reflections used in the refinement, reducing the observations-to-parameters ratio. The influence of omitting 5% of the

data during the refinement can be directly seen in Table 6. Comparison of the last two columns shows that inclusion of all data to 1.7 Å resolution leads to significantly lower phase errors. As the atoms in the pseudo protein models are refined without restraints on atomic positions and displacement parameters, the inclusion of all diffraction data means more crystallographic restraints on the free atoms during the refinement and this results in lower phase errors. Therefore, omitting reflections in *wARP* reduces the efficacy of the procedure and the use of an  $R_{\text{free}}$  factor is not recommended.

3.3.4. *The mean  $W_{\text{wARP}}$*  As shown in the previous paragraph, the  $R_{\text{free}}$  factor is not the best choice to evaluate the progress of the phase improvement. A useful alternative indicator seems to be the weight  $W_{\text{wARP}}$ . The parameter  $W_{\text{wARP}}$  is a weight with a value between zero and one that is assigned to each structure factor on the basis of the vector averaging of the structure factors from the different pseudo protein models. If the models resemble each other very closely, this means that there is a high correlation between the calculated structure factors, which results in a mean  $W_{\text{wARP}}$  close to 1.0. Large deviations between models lead to a lower mean  $W_{\text{wARP}}$ . As can be seen from Fig. 7 and Table 6 the highest mean  $W_{\text{wARP}}$  between *SFALL/ARP* and *REFMAC/ARP* refinement always corresponds to the set with the best phases in contrast to the behaviour of the  $R_{\text{free}}$  factor. From this we conclude that the mean  $W_{\text{wARP}}$  is a better indicator than the  $R_{\text{free}}$  factor to monitor the convergence of the procedure. Moreover, the mean  $W_{\text{wARP}}$  correlates reasonably well with the cosine of the mean phase error, and

thus it may serve as a figure of merit analogous to that in *e.g.* multiple isomorphous replacement. Indeed, the use of the individual  $W_{wARP}$  of each reflection as a weight in the electron-density map calculations improves the map quality (see Table 6). However, it should be stressed that the mean  $W_{wARP}$  is a measure of the self-consistency of the different phase sets, rather than a true figure of merit.

**3.3.5. Phase combination in case of low observations-to-parameters ratios.** At low resolution when the observations-to-parameters ratio is not high enough for *ARP* to give phase improvement on its own (Lamzin & Wilson, 1993), *wARP* may help to reduce the phase errors. We could slightly improve the 2.5 Å phases from solvent flattening and histogram matching by 2.1° (data not shown), by combining them with the phases derived from the three initial unrefined pseudo protein models using the program *SIGMAA* (Read, 1986). Attempts to improve the phases by combining them with the vector-averaged phase set from the six pseudo protein models refined at 2.5 Å with *SFALL* or *REFMAC* were not successful, however. Probably the refinements at 2.5 Å resolution did not give sufficient meaningful improvements to the pseudo protein models.

#### 3.4. Structure of Slt35

The 2.0 Å resolution electron-density maps from *wARP* were of outstanding quality and were used to build the Slt35 structure. The protein main chain could be easily followed and side chains could be readily recognised. Even solvent molecules could be identified in the density maps. After the model building, the X-ray structure of Slt35 was refined against diffraction data between 20.0 and 1.70 Å resolution with *X-PLOR* to an  $R_{work}$  and  $R_{free}$  of 18.7 and 22.1%, respectively. The final crystallographic *R* factor for all diffraction data is 18.9%. The current X-ray structure contains residues 42–97 and 109–361, five alternate side-chain conformations and 312 solvent molecules. The first three N-terminal residues (Met39, Val40 and Leu41) and residues 98–108 are not visible in the electron-density maps. Continuous electron density is observed for the rest of main chain, except for two small breaks in the main chain at residues 46 and 48. According to *PROCHECK* (Laskowski *et al.*, 1993), 91.2% of the  $\phi/\psi$  angles fall inside the most favoured regions of the Ramachandran plot and the rest in the additionally allowed regions (Fig. 8).

The X-ray structure of Slt35 consists of three domains named the alpha, beta and core domains. The alpha domain contains mainly  $\alpha$ -helices as secondary-structure elements and it is found at the bottom of the model in the orientation as given in Fig. 9. A  $\beta$ -sheet domain of five antiparallel  $\beta$ -strands flanked by one  $\alpha$ -helix is observed on top of the protein and is called the beta domain. Sandwiched between the alpha and beta domains is the core domain containing  $\alpha$ -helices and a large groove. The latter domain bears some resemblance to the fold of the

catalytic domain of Slt70 (Thunnissen *et al.*, 1994). A superposition of these two domains revealed that the catalytic Glu478 of Slt70 coincides with Glu162 of Slt35 (not shown). A more detailed analysis of the architecture of Slt35 is in preparation (van Asselt, Dijkstra, Kalk, Keck & Dijkstra, 1998).

Slt70 is a doughnut-shaped exo-muramidase, which cleaves 1,6-anhydromuropeptides from the 1,6-anhydromuramate end of glycan strands processing towards the *N*-acetylglucosamine end. It has been proposed that the doughnut domain of Slt70 is responsible for this exo-activity (Thunnissen *et al.*, 1994). In the case of Slt35, there is no such doughnut domain. Nevertheless, enzyme activity studies with radioactive double-labelled peptidoglycan material have shown that Slt35 is an exo-muramidase as well (Romeis, Vollmer & Höltje, 1993). It might be possible that residues 98–108, unfortunately not seen in the electron density, form a mobile loop from the alpha to the core domain that might fulfil a similar role to the Slt70 doughnut domain and prevent the endo-activity of the enzyme. As the active site of Slt70 superimposes on the core domain of Slt35, it seems likely that Slt35 starts degradation of peptidoglycan at the 1,6-anhydromuramate end of the glycan strand as well. To more clearly understand how Slt35 interacts with its substrate, further crystallographic studies are under way.

#### 4. Conclusions

The first application of *wARP* in the structure determination of a protein at high resolution starting from medium-resolution MIRAS phases is described. Phases could be extended from 2.7 to 1.7 Å in an automatic way without

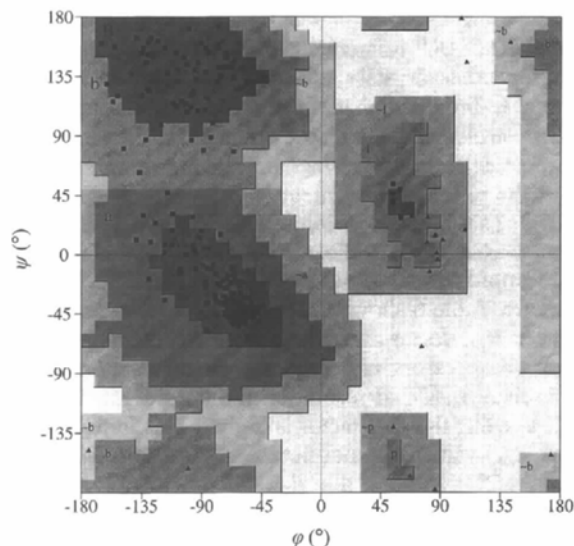


Fig. 8. Ramachandran plot for the final Slt35 structure, as calculated with the program *PROCHECK* (Laskowski *et al.*, 1993). 91.2% of the residues are in the most favoured regions and 8.8% in the additionally allowed regions. Glycine residues are shown as filled triangles and the other residues as filled squares.

the requirement of an atomic model. Our first approach to apply *wARP* using data to 2.0 Å resolution resulted already in electron density of such excellent quality that the complete protein model could be built in only a few days. Later studies showed that even better results could be obtained by using all diffraction data to 1.7 Å resolution. This latter approach resulted in a mean phase difference of only 21.7° compared with the phases from the final Slt35 structure. The structure of Slt35 consists of three domains and it has been refined to 1.7 Å resolution with *X-PLOR* to a final crystallographic *R* factor of

18.9%. The core domain bears some resemblance to the catalytic domain of Slt70, while the alpha and beta domains are completely different.

To monitor the progress and the reliability of the different refinement protocols in *wARP*, we initially thought that the  $R_{\text{free}}$  factor might be a good indicator. However, it appeared that setting aside 5% of the diffraction data for an  $R_{\text{free}}$  test set, negatively affects the phase improvement. Instead, the mean  $W_{\text{wARP}}$  is a better parameter to judge different phase improvement protocols. Hence, it is recommended not to use an  $R_{\text{free}}$

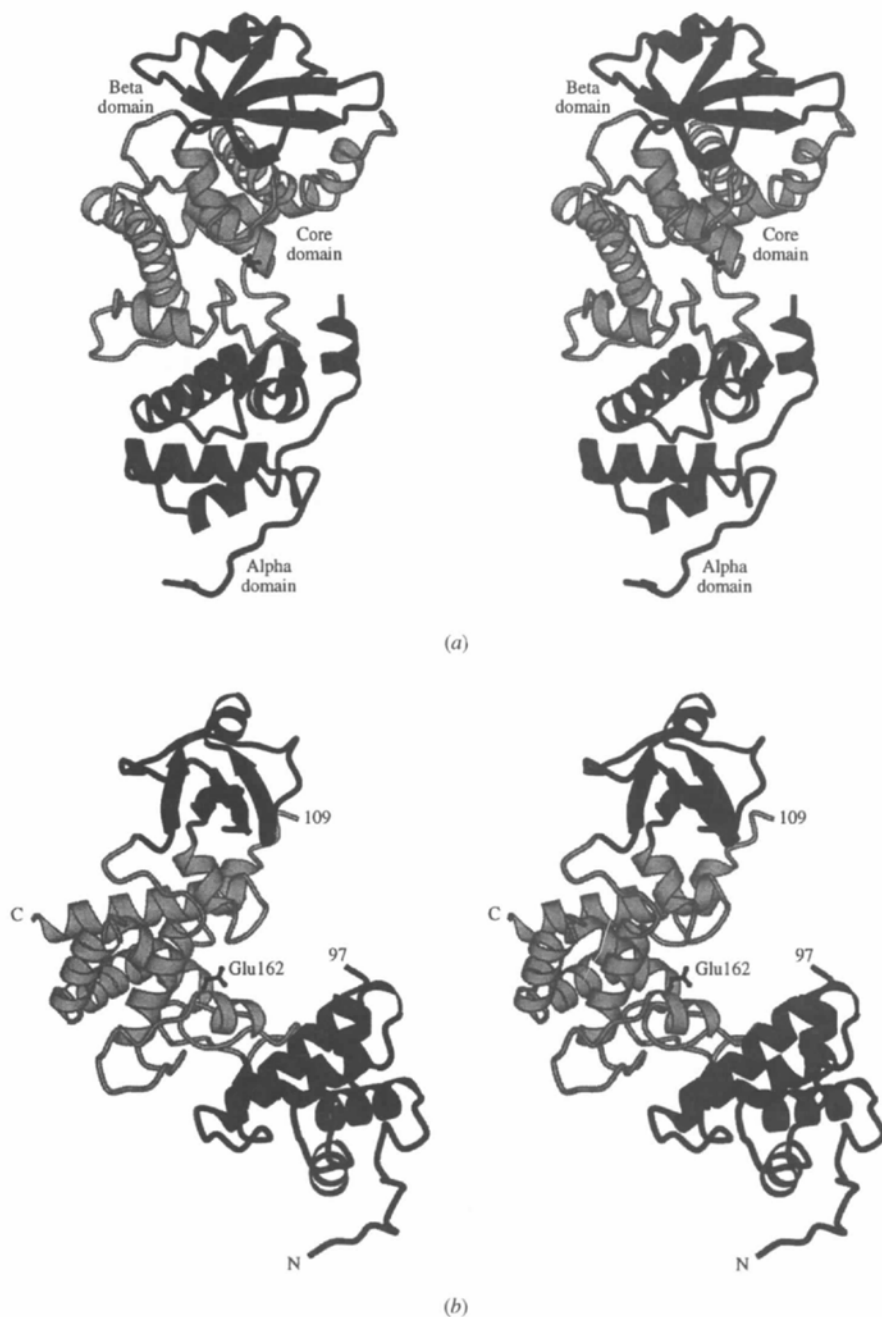


Fig. 9. (a) Stereoview of the Slt35 structure with the alpha domain at the bottom, the beta domain at the top of the molecule and the core domain in the centre. (b) Stereo figure of Slt35 viewed along the peptidoglycan binding cleft. The proposed catalytic acid Glu162 is indicated in ball and stick representation in the core domain. The picture was produced with *MOLSCRIPT* (Kraulis, 1991).



test set, but to include all diffraction data to the highest resolution available in the *wARP* procedure. The better observations-to-parameters ratio will improve the convergence rate of the refinement and the chance to succeed with *wARP*. In the case of Slt35, maximum-likelihood refinement in *wARP* gave much better phases than least-squares refinement, provided that the observations-to-parameters ratio was approximately 1.8 or higher.

In our hands, *wARP* appeared to be a very powerful tool for improving experimental phases and extending them to higher resolution. The electron-density maps calculated with the final *wARP* phases facilitated the interpretation of the map enormously and accelerated the building of the Slt35 X-ray structure.†

The result of this study helped to completely automate the entire *wARP* procedure, starting from the placement of O atoms in a fine-grid electron-density map to the calculation of the final *wARP* electron-density map. The user has to provide the best available electron-density map, covering one asymmetric unit, together with diffraction data and an arbitrarily placed piece of atomic model (e.g. an  $\alpha$ -helix) inside the core of the protein electron density. Additional parameters that have to be provided by the user are the space group, cell dimensions, resolution limits, grid spacing for calculating electron-density maps, asymmetric unit limits, the *B* factor derived from a Wilson plot and the estimated number of protein atoms. A choice should be made between least-squares and maximum-likelihood refinement. Six pseudo protein models are generated, refined and finally vector-averaged in one go and the high-resolution output map is calculated. All log files are summarized to check the progress during refinement of the different models. The refinement of the models can be parallelized by automatically distributing the calculations to up to six different computers, which share common disk storage space, reducing the total calculation time. Another requirement is that the user has the CCP4 program suite, *ARP* and *MOLEMAN*. All other programs and C-shell scripts are distributed together within the *wARP* package, which is available through the WWW at <http://den.nki.nl/~perrakis/arp.html>.

We would like to thank Arnoud Dijkstra and Wolfgang Keck for a stimulating collaboration on soluble lytic transglycosylases. Andy-Mark Thunnissen, Daniel Thiel and Marian Szebenyi are acknowledged for their help during data collection at CHESS. We are grateful for access to the EMBL outstation, and we thank the European Union for support of the work at EMBL,

† Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: ILTM, RILMSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: JN0033). At the request of the authors, the atomic coordinates will remain privileged until 1 January 1999 and the structure factors will remain privileged until 1 January 2002.

Hamburg, through the HCMP Access to Large Installations Project, Contract Number CHGE-CT93-0040. The investigations were supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO). AP is supported by a long-term EMBO fellowship (ALTF-215 1995) and thanks Thomas Schneider and Peter Metcalf for some Unix magic tricks they shared with him while developing the *wARP* scripts. VSL and AP thank Zbygniew Dauter for invaluable discussions in the EMBL library on the real-space averaging of density maps, which provided stimulating ideas for initiating development of *wARP*.

## References

- Agarwal, R. C. (1978). *Acta Cryst.* **A34**, 791–809.
- van Asselt, E. J., Dijkstra, A. J., Kalk, K. H., Keck, W. & Dijkstra, B. W. (1998). In preparation.
- Bhat, T. N. (1988). *J. Appl. Cryst.* **21**, 279–281.
- Brünger, A. T. (1992a). *Nature (London)*, **355**, 472–475.
- Brünger, A. T. (1992b). *X-PLOR. A System for X-ray Crystallography and NMR*, Yale University, New Haven, CT, USA.
- Brünger, A. T., Adams, P. D. & Rice, L. M. (1997). *Structure*, **5**, 325–336.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* **D49**, 148–157.
- Dijkstra, A. J., Hermann, F. & Keck, W. (1995). *FEBS Lett.* **366**, 115–118.
- Dijkstra, A. J. & Keck, W. (1996). *J. Bacteriol.* **178**, 5555–5562.
- Engel, H., Smink, A. J., van Wijngaarden, L. & Keck, W. (1992). *J. Bacteriol.* **174**, 6394–6403.
- Engh, R. A. & Huber, R. (1991). *Acta Cryst.* **A47**, 392–400.
- Furey, W. & Swaminathan, S. (1998). *Methods Enzymol.* In the press.
- Gold, H. S. & Moellering, R. C. Jr (1996). *N. Engl. J. Med.* **335**, 1445–1453.
- Höltje, J.-V., Mirelman, D., Sharon, N. & Schwarz, U. (1975). *J. Bacteriol.* **124**, 1067–1076.
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110–119.
- Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 916–924.
- Kleywegt, G. J. (1995). *ESF/CCP4 Newslett.* **31**, 45–50.
- Kleywegt, G. J. & Jones, T. A. (1996). *Acta Cryst.* **D52**, 826–828.
- Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
- Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* **D49**, 129–147.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Messerschmidt, A. & Pflugrath, J. W. (1987). *J. Appl. Cryst.* **20**, 306–315.
- Murshudov, G., Vagin, A. & Dodson, E. (1997). *Acta Cryst.* **D53**, 240–255.
- Otwinowski, Z. (1993). *Oscillation Data Reduction Program. In Proceedings of the CCP4 Study Weekend: Data Collection and Processing*. Warrington: Daresbury Laboratory.
- Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *Acta Cryst.* **D53**, 448–455.

- Petsko, G. A. (1985). *Meth. Enzymol.* **114**, 147–156.
- Read, R. J. (1986). *Acta Cryst.* **A42**, 140–149.
- Rice, L. M. & Brünger, A. T. (1994). *Proteins*, **19**, 277–290.
- Romeis, T., Vollmer, W. & Höltje, J.-V. (1993). *FEMS Microbiol. Lett.* **111**, 141–146.
- Thunnissen, A.-M. W. H., Dijkstra, A. J., Kalk, K. H., Rozeboom, H. J., Engel, H., Keck, W. & Dijkstra, B. W. (1994). *Nature (London)*, **367**, 750–753.
- Thunnissen, A.-M. W. H., Rozeboom, H. J., Kalk, K. H. & Dijkstra, B. W. (1995). *Biochemistry*, **34**, 12729–12737.
- Vellieux, F. M. D. & Dijkstra, B. W. (1997). *J. Appl. Cryst.* **30**, 396–399.
- Vriend, G. & Sander, C. (1993). *J. Appl. Cryst.* **26**, 47–60.
- Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.