

G. Rudenko,<sup>a</sup> L. Henry,<sup>b</sup> C. Vonrhein,<sup>c</sup> G. Bricogne<sup>c</sup> and J. Deisenhofer<sup>b\*</sup>

<sup>a</sup>Department of Biochemistry, UT Southwestern Medical Center, 5323 Harry Hines Boulevard Y 4-206, Dallas, Texas 75390, USA, <sup>b</sup>Howard Hughes Medical Institute, UT Southwestern Medical Center, 5323 Harry Hines Boulevard Y 4-206, Dallas, Texas 75390, USA, and <sup>c</sup>Global Phasing Limited, Sheraton House, Castle Park, Cambridge CB3 0AX, England

Correspondence e-mail:  
johann.deisenhofer@utsouthwestern.edu

# 'MAD'ly phasing the extracellular domain of the LDL receptor: a medium-sized protein, large tungsten clusters and multiple non-isomorphous crystals

The crystal structure of the extracellular domain of the LDL receptor (LDL-R) was determined in a MAD experiment using 12-tungstophosphate clusters as anomalous scatterers. While useful for phasing, the tungsten clusters rendered the crystals radiation-sensitive and non-isomorphous and profoundly altered the diffraction data, causing complications. The work is presented as a case study for phasing a medium-sized protein (700 residues) at low resolution (4 Å) with multiple non-isomorphous crystals containing 31 W atoms in the asymmetric unit.

Received 17 March 2003  
Accepted 26 September 2003

**PDB Reference:** human LDL receptor extracellular domain, 1n7d.

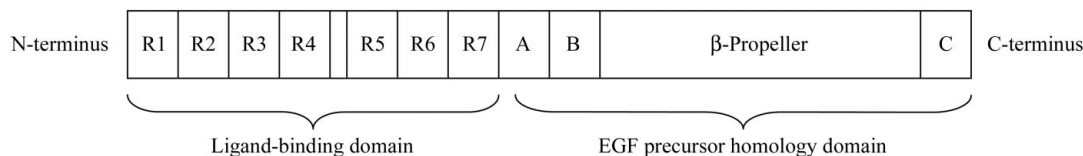
## 1. Abbreviations

*d*, Bragg spacing. FOM, figure of merit. MAD, multiple-wavelength anomalous dispersion. MIR, multiple isomorphous replacement. PhP, phasing power.  $R_{\text{factor}}(\text{SCALEIT}) = \sum |F_{PH} - F_P| / \sum F_P$ .  $R_{\text{weighted}}(\text{SCALEIT}) = \{ \sum [ |F_P^2 - F_{\text{Fmean}}| / \text{Var}(F_P^2) + |F_{PH}^2 - F_{\text{Fmean}}| / \text{Var}(F_{PH}^2) ] \} / \{ \sum [ F_P^2 / \text{Var}(F_P^2) + F_{PH}^2 / \text{Var}(F_{PH}^2) ] \}$ , where  $\text{Var}(F_P^2) = \text{Var}(F_P) \times 4F_P^2$ ,  $\text{Var}(F_P) = \sigma(F_P)^2$  and  $F_{\text{Fmean}} = [F_P^2 / \text{Var}(F_P^2) + F_{PH}^2 / \text{Var}(F_{PH}^2)] / [1 / \text{Var}(F_P^2) + 1 / \text{Var}(F_{PH}^2)]$ . R.m.s., root mean square.

## 2. Introduction

The low-density lipoprotein receptor (LDL-R), a cell-surface protein, plays an important role in cholesterol metabolism (reviewed by Goldstein *et al.*, 2001). LDL-R clears lipoprotein particles carrying cholesterol and cholesterol derivatives from the blood *via* receptor-mediated endocytosis. The first 695 residues of the LDL-R extracellular domain are divided into a 'ligand-binding domain' containing seven cysteine-rich repeats and an 'EGF precursor homology domain' containing in sequence two EGF-like repeats, a six-bladed  $\beta$ -propeller domain and a third EGF-like repeat (Yamamoto *et al.*, 1984; Jeon *et al.*, 2001; Fig. 1). Over the years, the structures of different fragments from the LDL-R extracellular domain have been elucidated by NMR and X-ray crystallography (Daly, Djordjevic *et al.*, 1995; Daly, Scanlon *et al.*, 1995; Fass *et al.*, 1997; Bieri *et al.*, 1998; Beglova *et al.*, 2001; North & Blacklow, 1999*a,b*; Clayton *et al.*, 2000; Jeon *et al.*, 2001; Saha *et al.*, 2001; Kurniawan *et al.*, 2001). However, it was not possible to deduce the precise arrangement of the modules with respect to each other, an absolute prerequisite to understand important functional aspects of LDL-R.

We have determined the structure of the extracellular domain of LDL-R (residues 1–699) by soaking large 12 W-atom clusters into crystals of LDL-R and carrying out a MAD experiment at the  $L_{III}$  tungsten edge. While providing many anomalous scatterers yielding a very large anomalous

**Figure 1**

Modular organization of the extracellular region of LDL-R. The 'ligand-binding domain' and the 'EGF precursor homology domain' are indicated. The cysteine-rich repeats R1–R7 are separated by short linkers of approximately four residues, with the exception of R4 and R5 which are separated by a 12-residue linker. The EGF-like repeats A, B and C flank the  $\beta$ -propeller domain.

signal and routinely improving the diffraction limits of the crystals from 6 Å to better than 4 Å, the tungsten clusters also complicated the structure determination by rendering the crystals non-isomorphous (both with respect to native crystals as well as other tungsten-cluster soaked crystals). The combination of weak low-resolution diffraction data and the presence of the clusters (profoundly altering the diffracted intensities) caused difficulties in scaling and merging the data as well as in assessing the magnitude of the anomalous signal. With a total of 31 W atoms in the asymmetric unit (two in special positions on a crystallographic twofold axis) and approximately 700 amino acids, the ratio of heavy-atom electrons to protein electrons is much higher than typically used to solve a protein structure. Our experiences are presented here as a case study.

### 3. Methods and results

#### 3.1. Preparation of polytungstate-derivatized crystals of LDL-R

Crystals of the extracellular domain of LDL-R (amino acids 1–699 overexpressed in insect cells, with mutations N494Q, N636Q and a hexahistidine affinity tag) were grown in 3% (v/v) 1,2-hexanediol, 50 mM sodium acetate pH 5.3, 0.5 mM calcium chloride as reported by Rudenko *et al.* (2002). The crystals, long hexagonally shaped rods with dimensions 800 × 100 × 100 μm, were soaked overnight at 294 K in 3% (v/v) 1,2-hexanediol, 50 mM sodium acetate pH 5.3, 5 mM calcium chloride supplemented with nanomolar to micromolar concentrations of sodium 12-tungstophosphate. While native crystals of the mutant N494Q, N636Q diffracted routinely to 6 Å at the synchrotron, incorporating clusters in the crystal lattice shrunk the unit cell and improved the diffraction limit of the crystals to better than 3.7 Å. The concentration of the cluster and the pH of the crystal-soaking solution had to be carefully controlled in order to limit crystal cracking and non-isomorphism between soaked crystals. The tungstophosphate-derivatized crystals were cryoprotected in crystal-soaking solution containing 30% (v/v) glycerol and cooled in liquid propane prior to data collection under cryogenic conditions. The crystals exhibit the symmetry of space group  $P3_121$ , average unit-cell parameters  $a = 185$ ,  $b = 185$ ,  $c = 85$  Å and contain one molecule in the asymmetric unit (74% solvent). Because the crystals diffracted weakly, long exposure times were needed to collect data with sufficient signal-to-noise, thereby inflicting significant radiation damage, particularly at

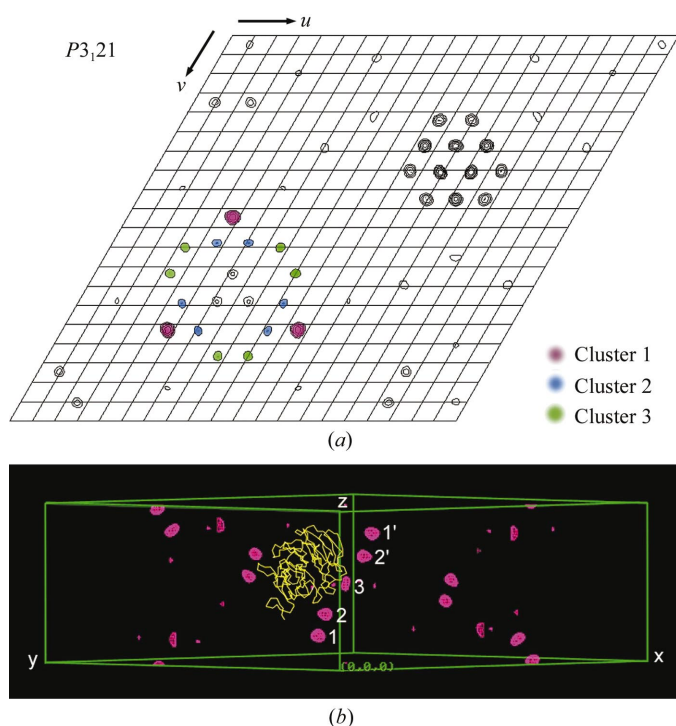
the peak and inflection-point wavelengths. For this reason, the crystals were translated every 10–20° during data collection to expose fresh pieces of crystal. More than 700 crystals were screened (first in-house for useable diffraction patterns and then at the synchrotrons APS and ALS for isomorphous diffraction) in the process of gathering sufficient data to complete a MAD experiment, which were ultimately obtained from 11 crystals and revealed the architecture of the extracellular domain of LDL-R.

#### 3.2. Overview of the structure determination of LDL-R

The structure determination of the extracellular domain of LDL-R was carried out in three cycles of MAD phasing. Each MAD experiment was the result of a different data-collection strategy, summarized here to first give an overview. Cycle 1, our initial attempt, contained a peak data set to 4.2 Å and a low-energy remote data set to 4.7 Å collected on different crystals because of radiation damage. We were able to identify the sites for the tungsten clusters, but the resulting electron-density map was of very low quality. Thorough comparison of the two data sets subsequently revealed non-isomorphism between the crystals. Cycle 2 contained a peak data set to 4.4 Å, an inflection-point data set and a high-energy remote data set both to 4 Å, all collected from a single crystal by shortening the exposure times. Although higher resolution data was sacrificed by shortening the exposure times, the phases derived from this single-crystal experiment were sufficient to generate a MAD electron-density map visualizing all but one module of the extracellular domain of LDL-R. While the modules were clearly visible, their backbone trace and side chains were generally still not well resolved. Lastly, in cycle 3 we shifted our emphasis to collecting higher resolution data. We added small segments of data collected from multiple crystals that suffered little radiation damage to the existing data from cycle 2 and were able to increase the resolution and the quality of the electron-density map. At this stage, the MAD experiment now contained a low-energy remote data set to 3.7 Å and a peak data set to 3.8 Å as well as inflection-point and high-energy remote data sets to 4.0 and 3.8 Å, respectively. Data-collection and phasing statistics for the three MAD experiments are summarized in Tables 1 and 2. We will now discuss the three MAD experiments in greater detail to illustrate the incremental progress towards the structure determination of LDL-R as the hurdles posed by non-isomorphism, low-resolution diffraction, radiation damage and large anomalous signal were overcome.

## 3.3. Cycle 1: the first experiment

Cycle 1 was carried out at the Advanced Photon Source (APS) at beamline 19-ID. The peak-wavelength and low-energy remote wavelength data sets were collected from different crystals. Data were processed with *HKL2000* (Otwinowski & Minor, 1997). The scaling procedure indicated that diffraction data at the peak wavelength rapidly became weaker owing to radiation damage (indicated by the scale factors decreasing from 1.0 to 0.45, under restrained *B* factors, within collecting 20° of data, *i.e.* 20 min accumulated exposure time). The presence of a strong anomalous signal was evident. Merging the data in the peak data set resulted in an  $R_{\text{sym}}$  of 18% compared with an  $R_{\text{sym}}$  of 11% for the low-energy remote data (Table 1). The two data sets were scaled together, with the low-energy remote as a reference set, using *FHSCAL*



**Figure 2**

Deriving positions for the tungsten clusters in the crystal. (a) Anomalous difference Patterson map calculated for the peak data collected in cycle 1 [data between 10 and 5 Å with  $F_{\text{obs}}/\sigma(F_{\text{obs}}) > 5$ ]. The Harker section  $w = 1/3$  is shown and reveals strong Harker peaks as well as some cross-peaks. The positions for cluster 1 and 2 were found by hand using Harker peaks. Harker peaks for cluster 1 had a height of 26% of the origin (fortuitously high, as the position of cluster 1 resulted in the expected six Harker peaks overlapping to show only three peaks). Harker peaks for cluster 2 had a height of 17% of the origin. A number of strong cross-peaks are also seen in the Harker section  $w = 1/3$ . The position of cluster 3, first identified in an anomalous difference Fourier map, was confirmed with the anomalous difference Patterson map. (b) Anomalous difference Fourier map calculated with the peak-wavelength data set (cycle 1) from 15 to 6 Å and molecular-replacement phases (search model shown in yellow). Cluster 1 and 2 are found as electron-density peaks with the height of  $6.5\sigma$  and  $6.0\sigma$ , respectively; cluster 3 is weaker. The anomalous difference density was contoured at  $4\sigma$  (shown in magenta for part of the unit cell). Cluster 3 is located on a crystallographic twofold axis lying along a diagonal between the *a* and *b* axes. The clusters 1 and 2 and half of cluster 3 fall within one asymmetric unit, while clusters 1', 2' and the other half of cluster 3 are located in another (indicated by white numbers).

(Collaborative Computational Project, Number 4, 1994). *SCALEIT* (Collaborative Computational Project, Number 4, 1994) was used to analyze the anomalous contributions, the extent of the isomorphism between the two data sets and the phasing potential of the data. The normal probability analysis plot of the dispersive differences, as calculated by *SCALEIT*, showed a gradient of 2.4 for acentric data between 27 and 4.7 Å, with no resolution bin having a value greater than 3.1. The  $R_{\text{weighted}}$  exceeded the  $R_{\text{factor}}$  only in the resolution range 12–7.4 Å ( $R_{\text{weighted}} \simeq 19\%$  and  $R_{\text{factor}} \simeq 17\%$ ). Typically, in *SCALEIT*, a ‘good derivative’ is identified as having a gradient greater than 5, with significant isomorphous differences in the range of data where the  $R_{\text{weighted}}$  is larger than the  $R_{\text{factor}}$ . At this point in our case we now questioned whether our tungstophosphate-derivatized crystals would prove amenable to a MAD experiment at the tungsten edge.

Molecular-replacement calculations using the largest known fragment (PDB code 1ijq; 308 residues containing the  $\beta$ -propeller and third EGF-like repeat) were carried out with *CNS* (Brünger *et al.*, 1998) using the peak data in the resolution range 8–4 Å. The rotation function gave a clear solution  $6\sigma$  above the mean and the translation function gave a clear solution with a correlation coefficient of 0.342 for space group  $P3_121$  (but not for  $P3_221$  or  $P321$ ). Although a clear molecular-replacement solution was found, the electron-density map calculated with the molecular-replacement phases did not retrieve density for the missing protein modules (~400 residues). Nevertheless, molecular replacement was useful to assign the symmetry of the crystals to space group  $P3_121$  and later, as discussed below, the molecular-replacement phases ( $\alpha_{\text{molrep}}$ ) proved invaluable to confirm the tungsten-cluster sites.

The sites of the tungsten clusters were identified with surprising ease. Clusters 1 and 2 were located in a Bijvoet difference anomalous Patterson map calculated with the peak data set (Fig. 2a). The best Patterson map was obtained using data between 10 and 5 Å with  $F_{\text{obs}}/\sigma(F_{\text{obs}}) > 5.0$ . The anomalous difference Fourier map, calculated between 15 and 6 Å with Bijvoet differences ( $|F_{\text{peak}+}| - |F_{\text{peak}-}|$ ) and phases  $\alpha_{\text{molrep}} - 90^\circ$ , confirmed the two cluster sites and revealed a third weaker site ( $<3\sigma$ ) on a crystallographic twofold axis (Fig. 2b). The shape of the anomalous difference Fourier densities (using data between 15 and 4.3 Å) suggested an orientation for the clusters with the central holes corresponding to the phosphate core of the clusters (Figs. 3a and 3b). The two and a half clusters found in the asymmetric unit generated a heavy-atom constellation containing 31 W atoms (two W atoms on special positions were given a starting occupancy of 0.5) that was used for phasing in *SHARP* (version 1.3; de La Fortelle & Bricogne, 1997). Phasing statistics for cycle 1 are given in Table 2. The resulting MAD map had a figure of merit of 0.43 to 4.2 Å (for acentric reflections) and revealed density for most of the modules, although not well enough to place them unambiguously (Fig. 4b).

The phasing was repeated in *SHARP* (version 1.3) using only the peak data in a SAD experiment. The SAD phases

**Table 1**

Diffraction data statistics from three MAD experiments, crystals grown with symmetry space group  $P3_121$ .

Values in parentheses are for the outer shell. HER, high-energy remote; LER, low-energy remote.

(a) Cycle 1: APS 19-ID, SBC2 detector unbinned mode, filters Al 0.254 + 0.127, exposure time 30 s per 0.5°, processed with *HKL2000*, peak 10.20821 keV, LER 10.00 keV.

Data set	$\lambda$ (Å)	Unit-cell parameters (Å)	Resolution (Å)	Total/unique reflections	Completeness (%)	$R_{\text{sym}}(I)$ (%)	$\langle(I)/\sigma(I)\rangle$
Peak	1.21455	185.85, 185.85, 84.94	32.2–4.2 (4.39–4.20)	33441/12304	97.6 (98.5)	18.0 (41.7)	8.2 (2.7)
LER	1.23984	186.92, 186.92, 85.36	45.6–4.7 (4.87–4.70)	43325/8757	96.3 (94.7)	10.7 (42.6)	10.6 (3.9)

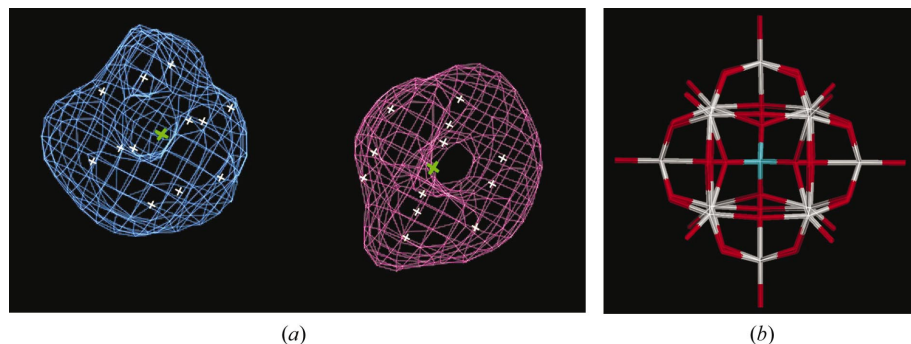
(b) Cycle 2: ALS 5.0.2, ADSC Q4 detector, unbinned mode, no filter, exposure time 50 s (25 s in dezing mode) per 0.5°, processed with *HKL2000*, peak 10.216 keV, inflection point 10.211 keV, high-energy remote 11.200 keV.

Data set	$\lambda$ (Å)	Unit-cell parameters (Å)	Resolution (Å)	Total/unique reflections	Completeness (%)	$R_{\text{sym}}(I)$ (%)	$\langle(I)/\sigma(I)\rangle$
Peak	1.21363	184.26, 184.26, 84.76	41.0–4.45 (4.63–4.45)	34576/9406	89.1 (75.5)	17.1 (53.1)	10.0 (2.0)
Inflection	1.21423	184.78, 184.78, 84.83	36.2–4.0 (4.16–4.00)	50003/14149	98.2 (97.1)	14.9 (74.3)	11.7 (1.7)
HER	1.10696	184.42, 184.42, 84.82	34.7–4.0 (4.16–4.00)	47315/14058	98.0 (98.8)	9.7 (48.6)	12.7 (2.3)

(c) Cycle 3: ALS 8.2.1, ADSC Q210 detector, unbinned mode, no filter, exposure time 120 s (60 s in dezing mode) per 0.5°, processed with *MOSFLM/SCALA*, LER 10.0 keV; for other wavelengths see cycle 2. Wilson  $B$  factors (*TRUNCATE*; Collaborative Computational Project, Number 4, 1994) calculated from the lowest to highest resolution shell exemplify the linear fall-off of the data (peak 133 Å<sup>2</sup>, inflection point 122 Å<sup>2</sup>, HER 123 Å<sup>2</sup> and LER 125 Å<sup>2</sup>). See also Rudenko *et al.* (2002).

Data set	$\lambda$ (Å)	Unit-cell parameters (Å)	Resolution (Å)	Total/unique reflections	Completeness (%)	$R_{\text{merge}}(I+, I-)^{\dagger}$ (%)	$R_{\text{merge}}(I)^{\ddagger}$ (%)	$\langle(I)/\sigma(I)\rangle$
Peak	1.21363	184.80, 184.80, 84.96	39.2–3.8 (3.93–3.80)	101571/14057	84.6 (20.7)	11.4 (39.6)	19.3 (42.5)	10.1 (1.6)
Inflection	1.21423	184.80, 184.80, 84.88	49.4–4.0 (4.14–4.00)	43739/13002	91.4 (51.2)	12.0 (50.1)	17.5 (49.0)	7.7 (2.4)
HER	1.10696	184.21, 184.21, 84.72	45.2–3.8 (3.93–3.80)	47496/14789	89.6 (32.0)	9.5 (59.6)	11.4 (55.6)	8.1 (2.4)
LER	1.23985	185.29, 185.29, 85.19	33.7–3.7 (3.83–3.70)	64801/17303	95.6 (64.5)	—	9.0 (44.9)	9.8 (2.7)

<sup>†</sup> Multiplicity-weighted  $R_{\text{merge}}$  keeping  $I+$  and  $I-$  separate, calculated by *SCALA* (Collaborative Computational Project, Number 4, 1994). Bijvoet and symmetry-related pairs were kept together for scaling, but separated for merging, in peak, inflection point and HER, but not LER data. <sup>‡</sup> Multiplicity-weighted  $R_{\text{merge}}$  with respect to  $I_{\text{mean}}$  (*i.e.*  $I+$ ,  $I-$  and their symmetry-related mates), calculated by *SCALA* (Collaborative Computational Project, Number 4, 1994).

**Figure 3**

Distinct orientation of the tungsten clusters in the crystal. (a) Anomalous difference Fourier density calculated for the peak data set (cycle 1) using data between 15 and 4.3 Å with the coefficients  $(|F_{\text{peak}+}| - |F_{\text{peak}-}|)$  and phases  $\alpha_{\text{molrep}} - 90^\circ$ . The electron density is contoured at  $1.5\sigma$  in maroon for cluster 1 and slate blue for cluster 2. The W atoms are indicated as white crosses and the central phosphate atoms are shown as green crosses. (b) Atomic model for the 12-tungstophosphate cluster based on a small-molecule structure determined by Gabriel *et al.* (1997). W atoms are shown in white, O atoms in red and the phosphate atom in cyan.

gave a figure of merit of 0.46 (for acentric reflections) to 4.2 Å; the electron density was surprisingly similar and in some places actually better than the electron density phased with two wavelengths (Fig. 4a versus Fig. 4b). We concluded that as predicted by *SCALEIT*, significant non-isomorphism existed between crystals which led to the quality of our MAD map being degraded (compared with our SAD map) even at relatively low resolution.

### 3.4. Cycle 2: adjusting the data-collection strategy for phasing

Cycle 2 was carried out at the Advanced Light Source (ALS) beamline 5.0.2 using a strategy focused on collecting non-isomorphous data in order to provide better phases. Data at three wavelengths [peak, inflection point and high-energy remote (HER)] were collected from a single crystal by shortening the exposure time per frame. To accomplish this task, the decay of the crystals in the X-ray beam was investigated beforehand to roughly determine how many degrees of data (accumulated exposure time) could be collected from a piece of crystal before radiation damage reduced the scale factors to 0.5 as determined by *HKL2000*. By trans-

lating the crystal (a long rod) in the beam after each wavelength, the inflection point and HER could be collected from a single fresh piece of crystal, while the peak-wavelength data set was constructed using two such pieces. The data were processed with *HKL2000* (Table 1) and scaled using *FHSCAL* with the high-energy remote as a reference data set. A *SCALEIT* analysis indicated dispersive differences for the peak and inflection-point data sets (both with respect to the

**Table 2**Phasing statistics from three MAD experiments, 31 W atoms (including 2 W atoms on special positions) in  $P3_121$ .

	Wavelength	Resolution (Å)	$R_{\text{cutlis}}^{\dagger\ddagger}$ (%)		Phasing power $^{\dagger\ddagger}$ (%)	
			$\Delta_{\text{iso}}$	$\Delta_{\text{ano}}$	$\Delta_{\text{iso}}$	$\Delta_{\text{ano}}$
Cycle 1: single-wavelength SAD	Peak	50–4.2	—/—	0.62/—	0.00/—	3.81/—
	Figure of merit $^{\ddagger}$ : 0.46/0.32 to 4.2 Å (0.19/0.21 in the range 4.48–4.20 Å)					
Cycle 1: two-wavelength MAD	Peak	50–4.2	0.84/0.84	0.58/—	1.13/0.85	1.86/—
	LER	50–4.7	—/—	0.92/—	0.00/0.00	1.05/—
	Figure of merit $^{\ddagger}$ : 0.43/0.31 to 4.2 Å (0.18/0.16 in the range 4.49–4.20 Å)					
Cycle 2: three-wavelength MAD	Peak	50–4.4	0.76/0.78	0.55/—	1.25/0.77	3.04/—
	Inflection	50–4.0	0.54/0.48	0.71/—	2.22/1.81	1.89/—
	HER	50–4.0	—/—	0.77/—	0.00/0.00	1.64/—
	Figure of merit $^{\ddagger}$ : 0.52/0.41 to 4.0 Å (0.39/0.23 in the range 4.27–4.20 Å)					
Cycle 3: four-wavelength MAD	Peak	50–3.8	0.75/0.82	0.53/—	1.03/0.75	5.24/—
	Inflection	50–4.0	0.80/0.78	0.69/—	2.39/1.42	3.67/—
	HER	50–3.8	0.63/0.55	0.77/—	0.60/1.28	2.78/—
	LER	50–3.7	—/—	0.71/—	0.00/0.00	1.81/—
	Figure of merit $^{\ddagger}$ : 0.64/0.53 to 3.7 Å (0.34/0.25 in the range 3.95–3.70 Å)					
Global phasing: four-wavelength MAD	Peak	42–3.9	0.51/0.57	0.41/—	1.91/1.69	4.84/—
	Inflection	42–3.9	—/—	0.51/—	—/—	3.25/—
	HER	42–3.9	0.32/0.32	0.73/—	3.64/3.55	2.00/—
	LER	52–3.5	0.46/0.80	0.92/—	2.65/1.41	1.12/—
	Figure of merit $^{\ddagger}$ : 0.44/0.42 to 3.5 Å					

$^{\dagger}$   $R_{\text{cutlis}}$  phasing power as given by *SHARP* (de La Fortelle & Bricogne, 1997).  $^{\ddagger}$  Given for acentric/centric.

high-energy remote) extending to approximately 6 Å. We concluded this because in the range 35–6 Å,  $R_{\text{weighted}}$  exceeded or was roughly equal to the  $R_{\text{factor}}$  ( $R_{\text{weighted}} \simeq 10.4\%$  and  $R_{\text{factor}} \simeq 10.9\%$  for inflection point *versus* HER;  $R_{\text{weighted}} \simeq 7.8\%$  and  $R_{\text{factor}} \simeq 6.8\%$  for peak *versus* HER). In cycle 1,  $R_{\text{weighted}}$  was never more than approximately 2% larger than  $R_{\text{factor}}$  for the best resolution bin. However, in cycle 2 the differences between  $R_{\text{weighted}}$  and  $R_{\text{factor}}$  were much larger, up to a difference of 10% for inflection point *versus* HER for data between 35 and 10 Å, indicating improved isomorphism, more significant dispersive differences and the possibility of deriving better phases.

The tungsten-cluster positions for cycle 2 were manually optimized in the anomalous difference Fourier density ( $|F_{\text{peak}^+}| - |F_{\text{peak}^-}|$ ) with phases  $\alpha_{\text{molrep}} - 90^\circ$ .

The molecular-replacement phases used were calculated from a slightly repositioned search fragment (rotated less than  $0.5^\circ$  and shifted less than 1 Å). The locations of the clusters found in cycle 1 and 2 differed by less than 1 Å. The optimized tungsten positions were used to calculate and refine phases with *SHARP* (version 1.3). Phasing statistics for cycle 2 are given in Table 2. The resulting MAD electron-density map had an improved figure of merit of 0.52 to 4.0 Å (for acentric reflections) and revealed all the modules in the extracellular domain with the exception of R1 (Fig. 4c). The density clearly showed the shape and orientation of the modules, but the details of the backbone trace and the side chains (especially in the cysteine-rich repeats) remained poorly resolved.

### 3.5. Cycle 3: completing the MAD experiment with higher resolution data

In contrast to the data-collection strategy used in cycle 2, the strategy in cycle 3 was optimized to extend the resolution of our data past 4 Å. First, to improve our experimental MAD electron-density map, we collected a new low-energy remote

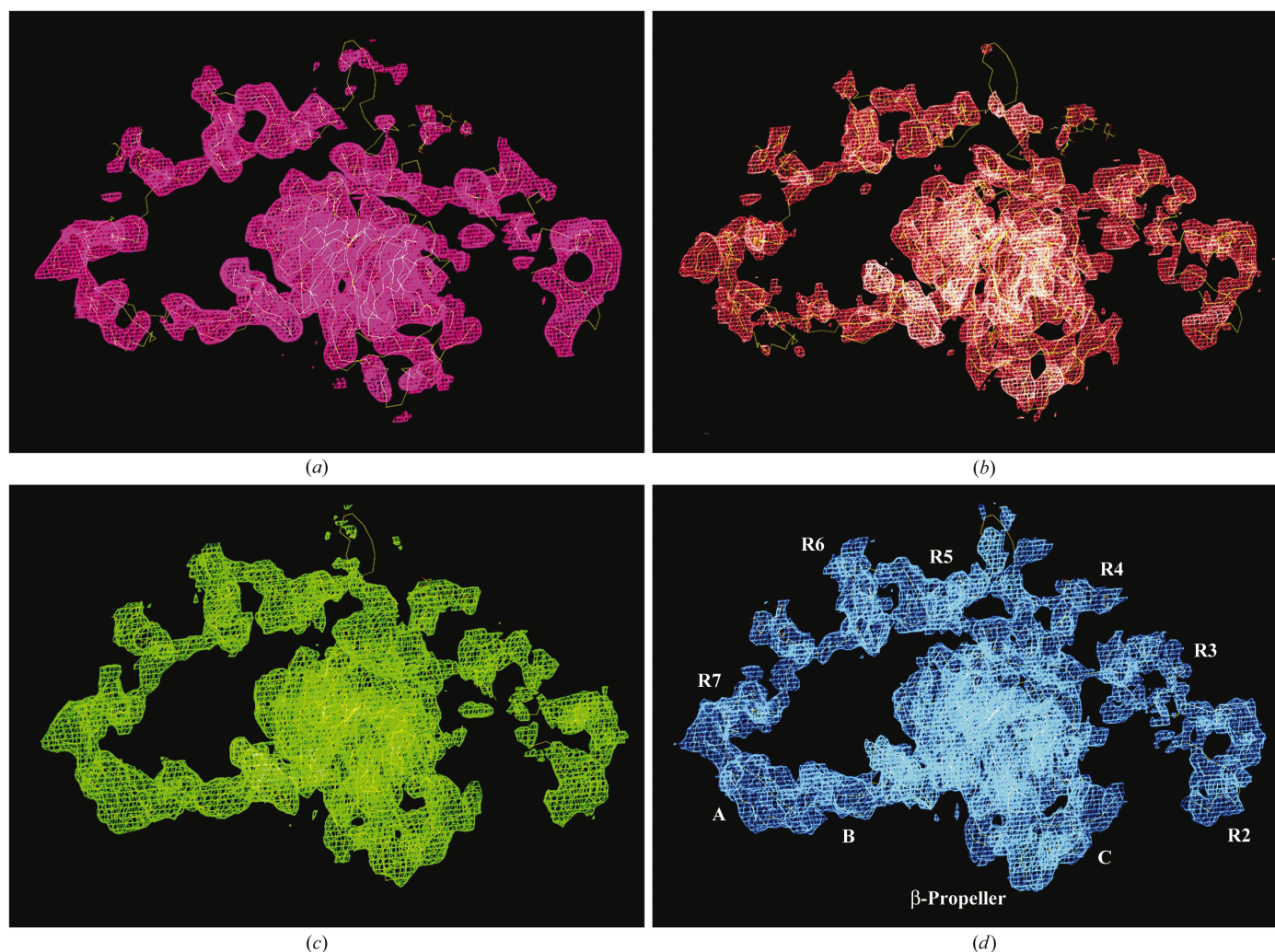
(LER) data set at the Advanced Light Source (ALS) beamline 8.2.1. The exposure time to X-rays per  $0.5^\circ$  of data was increased and anisotropic diffraction to 3.5 Å was observed. However, when a piece of crystal decayed to the extent that the scale factor per frame was reduced to 0.75 (roughly  $10^\circ$  of data), the crystal was translated to place a fresh piece in the X-ray beam. A complete data set was constructed after screening 128 regions of 53 crystals to yield 23 data segments diffracting to better than 4 Å. Of these, 11 segments (collected from eight different crystals) could be combined with an  $R_{\text{merge}} < 10\%$  to 3.7 Å. We purposely sought higher resolution data at the low-energy remote wavelength in order to avoid large Bijvoet differences. This was important as a large Bijvoet anomalous signal complicated the assignment of a consistent indexing scheme and obscured non-isomorphism between the different segments. Secondly, we added two incomplete data sets (collected as tests during cycle 2) to the peak data to increase multiplicity and extend the resolution. Finally, all data were integrated with *MOSFLM* (Leslie, 1992) using anisotropic resolution limits, with each wavelength scaled separately in *SCALA* (Evans, 1997) (Table 1). Thus, as a result of carefully characterizing different batches of diffraction data and assembling them into complete data sets, the data in cycle 3 featured a similar low degree of non-isomorphism as obtained in cycle 2, but with significantly improved resolution. The 31 W-atom positions derived from cycle 2 were used to phase cycle 3 with *SHARP* (version 1.4) (Table 2). The resulting MAD map with a figure of merit of 0.64 to 3.7 Å (for acentric reflections) showed improved density for all the modules in the extracellular domain, with the exception of R1 (Fig. 4d).

### 3.6. Model building

Fragments of LDL-R (PDB codes 1ajj, 1d2j, 1i0u, 1ijq) were used as a starting point to build a model for the extracellular

domain of LDL-R. Atomic coordinates from PDB entry 1fx5 were used to construct the N-linked carbohydrate cores on Asn135 and Asn251 (see also Rudenko *et al.*, 2002). While the density from cycle 3 allowed modules to be placed unambiguously (residues 44–693), the density for R1 remains poor and is not included in the model. Loops in the cysteine-rich repeats (which are known from NMR studies to be highly flexible in solution) had to be adjusted to fit the electron density. The density for a large number of side chains was further resolved after including the model in the phase calculation and sharpening the observed amplitudes with a resolution-dependent  $B$  factor [ $\exp(-B/d^2)$ , where  $d$  is the resolution and  $B = -120 \text{ \AA}^2$  derived from the Wilson plot of the data] (Fig. 5). At this stage, approximately 10% of the side chains were left modeled as alanines. Electron density for loops close to or contacting the clusters was still poorly visible, with no density for residues 57–58, 75, 99–102, 388, 441 and 453–454 (Fig. 6a). Crystallographic refinement at this

resolution is problematic because of the low ratio of observations to parameters (17 303 unique reflections and more than 5000 atoms), the lack of non-crystallographic symmetry providing additional restraints and/or lack of perfectly fitting high-resolution models (Rudenko *et al.*, 2002). A cautious positional refinement protocol using the experimentally determined phases as a cross-reference (MLHL target) was carried out with *CNS* to regularize the model, while placing strong restraints on the core residues in the modules. At present, the  $R_{\text{work}}$  is 38.2% using all reflections with  $F > 0$  between 33.1 and 3.7  $\text{\AA}$ ; the  $R_{\text{free}}$  is 38.7% (using 6% of the total reflections as a test set). The model has r.m.s. bond-length deviations of 0.028  $\text{\AA}$  and 4.3° for bond angles. Of the 547 non-glycine and non-proline residues, 484 residues are in the most favored/allowed regions, while one residue is currently in the unfavorable region of the Ramachandran plot (*PROCHECK*; Collaborative Computational Project, Number 4, 1994).

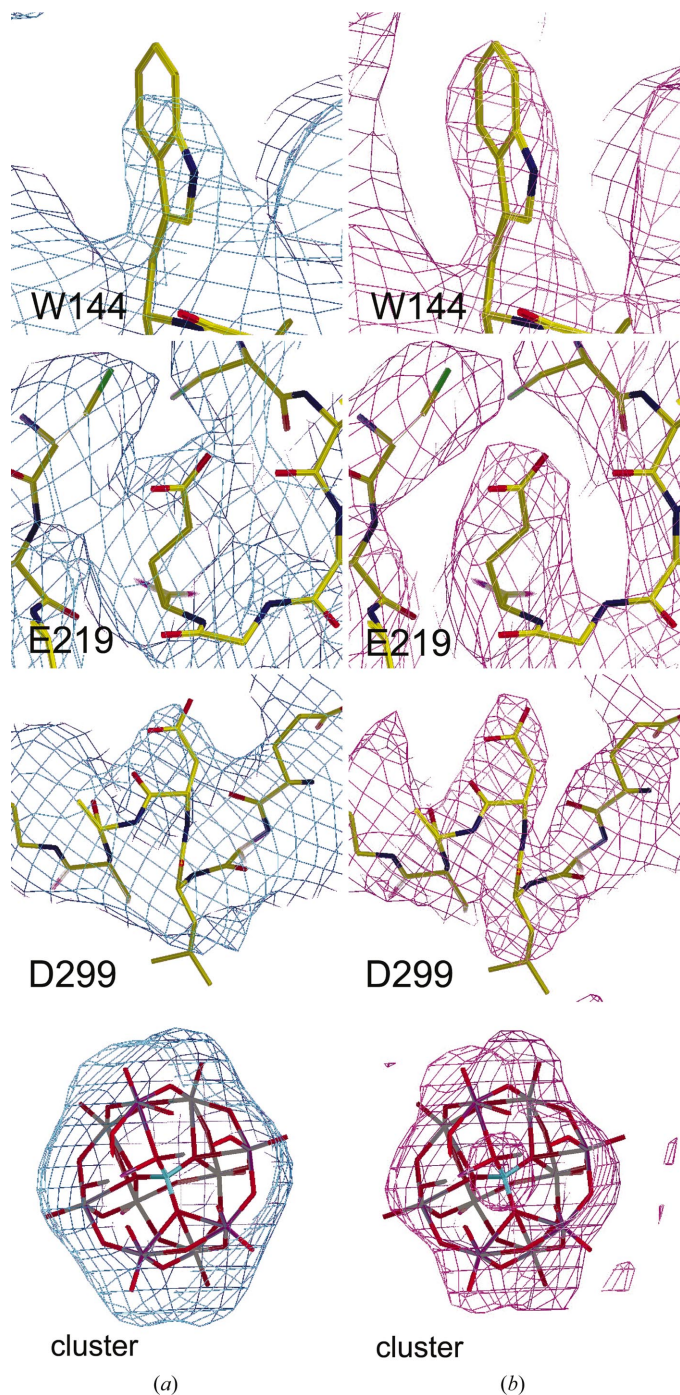


**Figure 4**

Experimental SAD and MAD electron-density maps calculated using one-, two-, three- or four-wavelength data sets. The phases were calculated using *SHARP* and improved through density modification. The density for the monomer is shown within 1.8  $\text{\AA}$  from the protein atoms. The clusters and their densities have been removed for the sake of clarity. (a) SAD map calculated with peak data from cycle 1; (b) MAD map derived from peak and low-energy remote data from cycle 1; (c) MAD map derived from peak, inflection-point and high-energy remote data from cycle 2; (d) MAD map derived from peak, inflection-point, high-energy remote and low-energy remote data from cycle 3. The modules are labeled according to the convention in Fig. 1.

3.7. New phasing procedure with *SHARP* version 2.0

To address the severe distortion of the electron-density map located near the tungsten clusters, the data were reprocessed and phased with a new version of *SHARP* 2.0 using Global Phasing. Recent developments in *SHARP* 2.0 have led to an enriched representation of two-dimensional structure-factor distributions (Bricogne *et al.*, 2003). This new eight-parameter



**Figure 5**  
Improvement of the electron-density map before (a) and after (b) sharpening the data, as discussed in the text. Examples are shown for residues Trp144, Glu219 and Asp299 (contoured at  $1\sigma$ ) and the 12-tungstophosphate cluster (contoured at  $3.5\sigma$ ).

model offers the possibility of transferring phase information from the phasing step to the density-modification step in a more faithful manner, especially in cases where heavy atoms are systematically present in the macromolecule. The data was scaled in *SCALA* (Evans, 1997) using an absorption correction as implemented in version 3.1.15. The LER data set was used to  $3.5 \text{ \AA}$ , although the highest resolution shells have poor statistics. The inflection-point data set was taken as the reference data set. To arrive at more accurate phases, an overall anisotropic scale was applied, a new feature available in *SHARP* 2.0, and relative anisotropic scale factors were refined. The phasing statistics are given in Table 2 (note: statistics in *SHARP* 2.0 are calculated slightly differently than in *SHARP* 1.4 and earlier versions). This procedure retrieved almost all of the electron density that had been obliterated next to the clusters, as shown for the loop 431–447 binding cluster 3 (Fig. 6b).

## 4. Discussion

## 4.1. Using polyoxotungstates in protein X-ray crystallography

12-Tungstophosphate ( $\text{PW}_{12}\text{O}_{40}^{3-}$ ; a polyoxometallate) is a water-soluble compound ideally suited to derivatize crystals for MIR or MAD experiments. The cluster can be seen as a central phosphate tetrahedron surrounded by 12  $\text{WO}_6$  octahedra, with a total net charge of  $-3$  and a span of  $10 \text{ \AA}$  in the largest dimension. Related polyoxotungstates with 17 or 18 W atoms have been used successfully to phase ribosomal subunits (Tocij *et al.*, 1999; Ban *et al.*, 1998; Clemons *et al.*, 1999). In our structure, cluster 1 is inserted between the modules R2 and R3, cluster 2 is found stabilizing module R3 against two  $\beta$ -propeller domains and cluster 3 is located on a twofold crystallographic axis which is also between two  $\beta$ -propeller domains (Rudenko *et al.*, 2002). The clusters each contact two or more arginines to form ionic interactions.

The presence of such huge tungsten clusters in our crystal was a mixed blessing. The tungsten clusters induced profound changes to the native crystals, shrinking all three cell dimensions by approximately  $5 \text{ \AA}$  (excluding the possibility of an MIR experiment). However, the 12-tungstophosphate-soaked crystals also had drastically improved diffractive qualities (better than  $4 \text{ \AA}$ , compared with  $6 \text{ \AA}$  routinely seen for native crystals of LDL-R N494Q, N636Q). Though the diffraction limits were improved, the crystals nevertheless still diffracted very weakly, so that the exposure times needed to obtain diffraction data with sufficient signal-to-noise induced significant radiation damage, severely limiting the amount of useable data that could be collected from a single crystal. Despite the low resolution of our crystals and the inherent non-isomorphism, the tungsten clusters gave rise to such a large anomalous signal that these hurdles were overcome and the structure of LDL-R was determined.

A high ratio of heavy atoms to protein atoms can pose unique problems to a structure determination. In our case, the presence of 2.5 tungsten clusters in the asymmetric unit and only one LDL-R molecule (*i.e.* one W atom per 25 residues)

**Table 3**

Effects of 2.5 tungsten clusters per asymmetric unit on the MAD data (31 W atoms including two W atoms on special positions).

Comparison of the predicted and observed effects of the tungsten clusters in our crystals. The theoretical scattering factors ( $f'_{\text{theor}}$  and  $f''_{\text{theor}}$ ) were calculated with *CROSSEC* (Collaborative Computational Project, Number 4, 1994) and the experimental values ( $f'_{\text{exp}}$  and  $f''_{\text{exp}}$ ) were measured at ALS from EXAFS scans. The anomalous (Bijvoet) and dispersive signals of 2.5 clusters per asymmetric unit were theoretically estimated using formulas given by Hendrickson & Ogata (1997). The anomalous and dispersive signals actually present in our experiment were calculated from the diffraction data (cycle 3) using *CNS*.

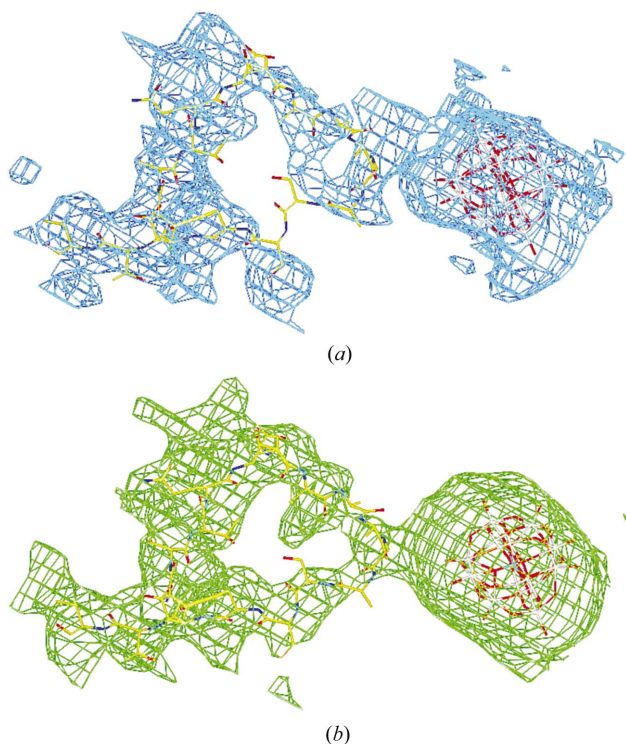
	W scattering factor ( $e^-$ )				Estimated signal				Experimental signal					
	$f'_{\text{theor}}$	$f''_{\text{theor}}$	$f'_{\text{exp}}$	$f''_{\text{exp}}$	Bijvoet $\Delta\ddagger$ (%)	Dispersive $\Delta\ddagger$ (%)			Resolution ( $\text{\AA}$ )	Bijvoet $\Delta\ddagger$ (%)	Dispersive $\Delta\ddagger$ (%)			
						Infl.	HER	LER			Infl.	HER	LER	
Cycle 3														
Peak	-17.3	10.5	-14.0	22.8	35.7	10.7	4.7	2.5	50–3.8	16.6	12.0	11.7	15.3	
Inflection	-19.8	10.5	-27.7	13.1	20.5		15.4	13.2	50–4.0	14.2		13.5	17.3	
HER	-7.9	8.6	-8.0	8.5	13.3			2.2	50–3.8	10.6			12.4	
LER	-10.3	3.9	-10.0	4.0	6.3				50–3.7	–				

† The Bijvoet diffraction ratio,  $\text{r.m.s.}(\Delta F_{\pm hkl})/\text{r.m.s.}|F|$ , is estimated to be roughly equal to  $(N_A/2N_T)^{1/2}(2f''_{\text{theor}}/Z_{\text{eff}})$ , and the dispersive difference ratio,  $\text{r.m.s.}(\Delta F_{\Delta\lambda})/\text{r.m.s.}|F|$ , is estimated to be roughly equal to  $(N_A/2N_T)^{1/2}[(f'_A(\lambda_1) - f'_A(\lambda_2))/Z_{\text{eff}}]$ , as given by Hendrickson & Ogata (1997), with  $Z_{\text{eff}} = 6.7 e^-$ ,  $N_A$  being the number of anomalous scatterers and  $N_T = 7.8$  times the number of residues. ‡ Bijvoet  $\Delta$ ,  $(|F_{\lambda_+}| - |F_{\lambda_-}|)/(\frac{1}{2}(|F_{\lambda_+}| + |F_{\lambda_-}|)/2)$ ; dispersive  $\Delta$ ,  $(|F_{\lambda_1}| - |F_{\lambda_2}|)/(|F_{\lambda_2}| + |F_{\lambda_1}|)/2$  as given by *CNS* (Brünger *et al.*, 1998).

profoundly altered the diffraction data (Table 3). In total, 8% of the X-ray scattering electrons came from the clusters (considering 3060 electrons from 2.5 clusters and  $\sim 36\,600$  electrons from non-H protein atoms). The Crick & Magdoff formula (Crick & Magdoff, 1956) predicted that the clusters

would induce an average fractional change in the intensity of more than 160% for centric reflections and more than 110% change for acentric reflections compared with native crystals. The Bijvoet diffraction ratio was estimated to be 36% for our peak data if 2.5 clusters were bound. The largest dispersive diffraction ratio was estimated to be 15% between our inflection-point and high-energy remote data sets (Table 3). Even the low-energy remote data was predicted to show significant anomalous signal, with an estimated Bijvoet diffraction ratio of 6%. In practice, however, the anomalous contributions were lower judging from the dispersive and anomalous diffraction ratios calculated directly from the data (Table 3). While *SHARP* was able to handle these very large anomalous differences at the phasing stage, complications occurred during data processing when we tried to combine data from multiple non-isomorphous crystals.

A number of factors contributed to our success. Careful incremental optimization of the data-collection strategy proved to be absolutely critical. Poor-quality low-resolution data from just the single peak wavelength in cycle 1 were quite sufficient to locate the cluster positions. Even though the data were of low resolution, we could obtain clear anomalous Patterson maps by adjusting the  $F_{\text{obs}}/\sigma(F_{\text{obs}})$  cutoff to remove the weakest data. Exploring different resolution ranges in the anomalous difference Fourier analysis enabled us not only to clearly locate the clusters (data to 6  $\text{\AA}$ ), but also to determine their likely orientations (data to 4.3  $\text{\AA}$ ). Low-resolution data were in fact extremely useful, as even with only 14–7  $\text{\AA}$  data it was possible to place the fragment  $\beta$ -propeller + C correctly by molecular replacement. The molecular-replacement calculations were likely to have been aided by the asymmetry of the search model (broken by the EGF-like repeat C). To improve the phases, we needed multiple wavelengths. We re-collected the MAD experiment using only one crystal to reduce non-isomorphism and sacrificed higher resolution data by shortening the exposure times to accomplish this. Only as the final step did we seek higher resolution data. On the one hand, higher resolution data were needed to improve the electron density; on the other hand, we were concerned that if the

**Figure 6**

Four-wavelength MAD electron density (*a*) before and (*b*) after removal of Fourier ripples at the tungsten-cluster sites as discussed in the text. The phases were calculated using *SHARP* version 1.4 (*a*) and 2.0 (*b*), respectively, and subsequently optimized by density modification within *SHARP*. The model was not included in the phase calculations and the data was not sharpened. The density is contoured at  $1\sigma$ . The loop 431–447 binding cluster 3 is shown (C atoms in yellow, O atoms in red, N atoms in blue, W atoms in white). The electron density is shown within 2.5  $\text{\AA}$  of the atoms.



clusters were not ordered to the full extent of the data, we would not be able to provide an accurate heavy-atom model to phase higher resolution reflections. As the figure of merit is still 0.34 in our highest resolution shell 3.9–3.7 Å (see cycle 3, Table 2), it seems that the clusters are indeed still largely ordered to this resolution. We constructed a higher resolution data set to 3.7 Å by combining segments of data from different crystals; it was important to collect these data at the low-energy remote wavelength in order to minimize the Bijvoet anomalous signal that interfered with selecting small wedges of isomorphous data. Extending the resolution of the data to 3.7 Å improved the electron density, especially after sharpening the data and enabled us to further improve our model.

While a number of very large macromolecular assemblies have been solved with very large heavy-atom clusters, we show here that these clusters can be equally useful for much smaller proteins. The complications introduced by derivatizing our crystals with such larger clusters were outweighed by the benefits in our case.

We thank the staff at the synchrotrons Advanced Photon Source (APS) 19-ID and Advanced Light Source (ALS) 5.0.2, 8.2.1 for excellent support, in particular Dr Keith Henderson at ALS for screening potentially derivatized crystals with EXAFS. We gratefully acknowledge Dr Mischa Machius for reading the manuscript. Dr Elspeth Garman and Robin Owen are thanked for insightful comments and theoretical calculations regarding radiation damage in our crystals.

## References

- Ban, N., Freeborn, B., Nissen, P., Penczek, P., Grassucci, R. A., Sweet, R., Frank, J., Moore, P. B. & Steitz, T. A. (1998). *Cell*, **93**, 1105–1115.
- Beglova, N., North, C. L. & Blacklow, S. C. (2001). *Biochemistry*, **40**, 2808–2815.
- Bieri, S., Atkins, A. R., Lee, H. T., Winzor, D. J., Smith, R. & Kroon, P. (1998). *Biochemistry*, **37**, 10994–11002.
- Bricogne, G., Vonnrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). *Acta Cryst. D* **59**, 2023–2030.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst. D* **54**, 905–921.
- Clayton, D., Brereton, I. M., Kroon, P. A. & Smith, R. (2000). *FEBS Lett.* **479**, 118–122.
- Clemons, W. M., May, J. L. C., Wimberly, B. T., McCutcheon, J. P., Capel, M. & Ramakrishnan, V. (1999). *Nature (London)*, **400**, 833–840.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst. D* **50**, 760–763.
- Crick, F. H. C. & Magdoff, B. S. (1956). *Acta Cryst.* **9**, 901–908.
- Daly, N. L., Djordjevic, T. L., Kroon, P. A. & Smith, R. (1995). *Biochemistry*, **34**, 14474–14481.
- Daly, N. L., Scanlon, M. J., Djordjevic, J. T., Kroon, P. A. & Smith, R. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 6334–6338.
- Evans, P. R. (1997). *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **33**, 22–24.
- Fass, D., Blacklow, S., Kim, P. S. & Berger, J. M. (1997). *Nature (London)*, **388**, 691–693.
- Gabriel, J.-C. P., Nagarajan, R., Natarajan, S., Cheetham, A. K. & Rao, C. N. R. (1997). *J. Solid State Chem.* **129**, 257.
- Goldstein, J. L., Hobbs, H. H. & Brown, M. S. (2001). *The Metabolic and Molecular Bases of Inherited Disease*, edited by C. R. Scriver, A. L. Beaudet, W. S. Sly & D. Valle, Vol. II, ch. 120, pp. 2863–2913. New York: McGraw-Hill.
- Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
- Jeon, H., Meng, W., Takagi, J., Eck, M. J., Springer, T. A. & Blacklow, S. C. (2001). *Nature Struct. Biol.* **8**, 499–504.
- Kurniawan, N. D., Aliabadizadeh, K., Brereton, I. M., Kroon, P. A. & Smith, R. (2001). *J. Mol. Biol.* **311**, 341–356.
- La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
- Leslie, A. G. W. (1992). *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **26**.
- North, C. L. & Blacklow, S. C. (1999a). *Biochemistry*, **38**, 3926–3935.
- North, C. L. & Blacklow, S. C. (1999b). *Biochemistry*, **39**, 2564–2571.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Rudenko, G., Henry, L., Henderson, K., Ichtchenko, K., Brown, M. S., Goldstein, J. L. & Deisenhofer, J. (2002). *Science*, **298**, 2353–2358.
- Saha, S., Boyd, J., Werner, J. M., Knott, V., Handford, P. A., Campbell, I. D. & Downing, A. K. (2001). *Structure*, **9**, 451–456.
- Tocilj, A., Schlunzen, F., Janell, D., Gluhmann, M., Hansen, H. A., Harms, J., Bashan, A., Bartels, H., Agmon, I., Franceschi, F. & Yonath, A. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 14252–14257.
- Yamamoto, T., Davis, C. G., Brown, M. S., Schneider, W. J., Casey, M. L., Goldstein, J. L. & Russell, D. W. (1984). *Cell*, **39**, 27–38.