Haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10 refined at 1.15 Å resolution

Ivo S. Ridder, Henriëtte J. Rozeboom and Bauke W. Dijkstra*

Laboratory of Biophysical Chemistry, Department of Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Correspondence e-mail: bauke@chem.rug.nl

Crystals of the 35 kDa protein haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10 diffract to 1.15 Å resolution at cryogenic temperature using synchrotron radiation. Blocked anisotropic least-squares refinement with *SHELXL* gave a final conventional R factor of 10.51% for all reflections in the 15–1.15 Å resolution range. The estimated r.m.s. errors of the model are 0.026 and 0.038 Å for protein atoms and all atoms, respectively. The structure comprises all 310 amino acids, with 28 side chains and two peptide bonds in multiple conformations, two covalently linked Pb atoms, 601 water molecules, seven glycerol molecules, one sulfate ion and two chloride ions. Water molecules accounting for alternative solvent structure are modelled with a fixed occupancy of 0.5. The structure is described in detail and compared with previously reported dehalogenase structures refined at 1.9–2.3 Å resolution. An analysis of the protein's geometry and stereochemistry reveals eight mean values of bond lengths and angles which deviate significantly from the Engh & Huber parameters, a wide spread in the main-chain ω torsion angle around its ideal value of 180 (6)° and a role for C—H···O interactions in satisfying the hydrogen-bond acceptor capacity of main-chain carbonyl O atoms in the central β-sheet.

1. Abbreviations

DhlA, haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10; DCE, 1,2-dichloroethane; AR, atomic resolution; PDB, Protein Data Bank; MES, 2-(*N*-morpholino)-ethanesulfonic acid; PEG, polyethylene glycol; SEM, scanning electron microscopy; AAS, atomic absorption spectrometry; 1be0, haloalkane dehalogenase structure at pH 5.0 and cryo-temperature; *ARP*, automated refinement procedure; ADP, anisotropic displacement parameter; B_{eq}, B factor equivalent to a given set of ADPs; U_{eq}, isotropic mean-square displacement equivalent to a set of ADPs; $B_{eq} = 8\pi^2 U_{eq}$; F_o, F_c, observed and calculated structure-factor amplitudes; r.m.s., root-mean-square; CGLS, conjugate-gradient least-squares; LS, least-squares; CSD, Cambridge Structural Database; 2had, haloalkane dehalogenase structure at pH 6.2 and room temperature; p.p.m., parts per million.

2. Introduction

Haloalkane dehalogenase (DhlA) is the key enzyme in the degradation of 1,2-dichloroethane (DCE) by the bacterium *X. autotrophicus* GJ10 (Janssen *et al.*, 1985). It converts a broad range of haloalkanes to their corresponding alcohols by hydrolytic cleavage of the carbon–halogen bond (Keuning *et*

al., 1985). As several of its substrates are toxic xenobiotics, the enzyme is of fundamental interest in environmental biotechnology (Stucki & Thüer, 1994, 1995). Detailed biochemical and crystallographic studies have supplied insight into the structure and catalytic mechanism of DhlA (Franken *et al.*, 1991; Krooshof *et al.*, 1998; Pries *et al.*, 1994; Schanstra, Kingma *et al.*, 1996; Schanstra, Ridder *et al.*, 1996; Verschueren, Franken *et al.*, 1993; Verschueren, Seljée *et al.*, 1993). The protein consists of two domains. The main domain is composed of a central eight-stranded mainly parallel β-sheet flanked on either side by α-helices, typical of the α/β-hydrolase fold family of enzymes (Ollis *et al.*, 1992). This domain provides a scaffold for the catalytic triad residues Asp124, His289 and Asp260, which are located at the bottom of a hydrophobic cavity. The cavity is closed off from the solvent by the second or cap domain which is made up of five α-helices. The four-step reaction mechanism starts with the formation of a Michaelis complex, with the halogen moiety of the substrate bound between Trp125 and Trp175. Upon nucleophilic attack of Asp124 OD1 on the substrate's C1 atom, the carbon–halogen bond is cleaved and a halide ion is produced. The resulting covalent alkyl-enzyme intermediate is hydrolysed in the next step by a water molecule which is activated by the general base His289, followed by fast release of the alcohol product. Finally, the halide ion is removed from the active site.

These insights into the structure and mechanism of DhlA have been obtained from crystal structures at medium to high resolution, ranging from 2.7 to 1.9 Å. Until recently, this was a common situation in macromolecular crystal structure determination. In general, the large size and flexibility of protein molecules in combination with a high crystal solvent content results in weak diffraction, making it impossible to obtain atomic resolution (AR) data. This situation is changing now crystallographic techniques such as cryogenic data collection are routinely applied and as high-intensity synchrotron sources become more widely available, equipped with more efficient two-dimensional detectors. An increasing number of data sets are being collected at atomic resolution, *i.e.* extending to at least 1.2 Å, with over 50% of the theoretically measurable reflections in the outer shell having $I > 2\sigma(I)$ (Dauter *et al.*, 1997; Sheldrick, 1990). A major advantage of a complete data set at AR is a far greater ratio of experimental observations to parameters, which in principle allows the refinement of the model without any stereochemical restraints, although in practice the more flexible regions of the protein need to be restrained even at very high resolution. Other advantages include a better modelling of side-chain disorder and anisotropy, a solvent model which includes partially occupied sites, accurate calculation or even refinement of H-atom positions and a rigorous estimation of the errors for each individual atomic parameter.

From a methodological point of view, atomic resolution structures are highly interesting as they provide a more accurate protein model. This allows one to obtain more insight into geometric and conformational properties and might help in the creation of better chemical and stereochemical target libraries. The most commonly used library is based on X-ray structures of amino acids and small peptide fragments (Engh & Huber, 1991), assuming that protein stereochemistry is very similar. High-resolution protein structure data can be useful in validating this assumption. Furthermore, AR structures supply valuable data for the construction of validation tools such as *PROCHECK* (Laskowski *et al.*, 1993) and *WHATCHECK* (Hooft *et al.*, 1996), which are used in the quality assessment of protein models (Wilson *et al.*, 1998). The study by Wilson and co-workers is based on only eight structures, but since then the number of AR structures available in the Protein Data Bank (PDB; Abola *et al.*, 1997) has increased to well over 30. However, most of the reported structures are relatively small, and only five are of a size comparable with that of DhlA. These larger proteins provide large sampling and thus the AR DhlA structure presented here might contribute significantly to the advance of library development.

Detailed structural information on haloalkane dehalogenase is already available from 2.0 Å resolution structures of the wild-type enzyme obtained from both room-temperature and cryo-temperature diffraction experiments carried out in-house (Krooshof *et al.*, 1998; Verschueren, Franken *et al.*, 1993). The use of synchrotron radiation enables us to report here the structure of DhlA with a chloride ion bound refined at atomic resolution. Apart from the general advantages of an AR model mentioned above, halide binding in the active site and anisotropy in the nucleophilic Asp124 can now be assessed very accurately in DhlA. Furthermore, side-chain disorder is now better defined, as well as the position of one flexible loop and the N- and C-termini of the protein. Moreover, the structure allows us to re-examine the hydrogen-bond pattern in the β-sheet, providing evidence for the existence of attractive C—H···O interactions.

3. Materials and methods

3.1. Crystallization and product complex formation

Haloalkane dehalogenase was purified and crystallized as described previously (Rozeboom *et al.*, 1988). Lozenge-shaped crystals (typical dimensions $0.3 \times 0.3 \times 0.15$ mm) were obtained within a few weeks at room temperature by vapour diffusion in hanging drops. The drop contained 8 µl 2.5 mg ml^{-1} protein, 24–26% ammonium sulfate (saturated at 273 K) and 100 mM 2-[*N*-morpholino]ethanesulfonic acid (MES) buffer pH 5.6–5.8. The well contained 1 ml 49–53% ammonium sulfate and 100 mM MES buffer pH 5.6–5.8. The crystal used in the experiment was recovered from a three year-old dried-in drop by placing the cover slip with the drop over fresh storage solution containing 70% ammonium sulfate and 100 mM MES buffer pH 5.8.

Subsequently, the crystal was equilibrated for 0.5 h in a solution containing 70% ammonium sulfate and 100 mM MES buffer pH 5.0 and was then soaked for 3 h at room temperature in 10 mM 1-chloropentane, 70% ammonium sulfate and 100 mM citrate buffer pH 5.0. A solution of 30%(*w/v*) PEG

Table 1
Data collection and processing statistics.

Values in parentheses are for the highest resolution shell.

Beamline	ID14-EH3
Temperature (K)	100
Wavelength (Å)	0.9475
Space group	$P2_12_12$
Unit-cell dimensions (Å)	$a = 92.18$, $b = 72.03$, $c = 40.91$
Resolution (Å)	30–1.15 (1.17–1.15)
$R_{merge}(I)$† (%)	3.9 (31.6)
$\langle I/\sigma(I)\rangle$	30.6 (2.7)
Completeness (%)	97.6 (92.5)
Number of observations	347849
Number of observations discarded‡	3566
Number of unique reflections	94837

† R_{merge} (%) $= 100 \times \sum_{hkl}\sum_{i=1}^{N}|\langle I^{hkl}\rangle - \langle I_i^{hkl}\rangle|/\sum_{hkl}\sum_{i=1}^{N}\langle I_i^{hkl}\rangle$. ‡ For example, owing to oversaturation.

6000, 20%(v/v) glycerol and 100 mM citrate pH 5.0 was applied as cryoprotectant during data collection.

The soaking solution was analysed because the electron density indicated the presence of heavy metals in the experiment. The solution was evaporated and calcinated and a scanning electron microscopy (SEM) elemental analysis showed the presence of C, N, O, S, Cl and Pb and/or Bi. After microwave digestion of the powder in a closed system, atomic absorption spectroscopy using a Perkin–Elmer 1100B spectrophotometer demonstrated traces of Pb at a concentration of about 3 p.p.m. and no detectable amounts of Bi (detection limit 2 p.p.m.) or Hg (detection limit 0.5 p.p.m.).

3.2. X-ray data collection and processing

Diffraction data were collected at 100 K from one single frozen crystal using synchrotron radiation on experimental station ID14-3 ($\lambda = 0.9475$ Å) at ESRF, Grenoble, equipped with a MAR Research CCD detector system. One data set was collected in the resolution range 15–1.15 Å, applying an exposure time of 12 s in three passes. A second data set was taken from the same crystal over the lower resolution range 30–2.32 Å using a shorter exposure time of 1 s in order to measure accurately the high-intensity reflections. The data were integrated and merged with *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Oversaturated reflections in the long-exposure data set were discarded and those from the short-exposure set were included in the merging. If the two data sets were scaled separately, the merging R factors were 4.0 (30.7) and 3.1 (7.1)% for the high-resolution and low-resolution data, respectively. Combined merging of the two sets results in an overall R_{merge} of 3.9%. The percentage completeness and R_{merge} of the data as a function of resolution are shown in Fig. 1. The drop in R_{merge} and completeness at $1/d^2 = 0.19$ Å^{-2} corresponds to the maximum resolution of the low-resolution data set. Data-collection parameters and processing statistics are given in Table 1. The mosaicity of the crystal refined to 0.6°. Intensities were reduced to structure factors and brought to an absolute scale with programs from the *CCP*4 suite (Collaborative Computational Project,

Number 4, 1994). From a Wilson plot (Wilson, 1942) an overall B factor of 10.0 Å^2 was estimated.

3.3. Refinement protocol

The 2.0 Å structure of haloalkane dehalogenase at pH 5.0 and cryo-temperature was used as the starting model for the refinement (PDB entry 1be0; Krooshof *et al.*, 1998). This structure consisted of the complete DhlA molecule with an acetic acid molecule bound in the active site. The acetic acid was removed from the model, but all 334 water molecules were kept. The random set of 5% of the unique reflections set apart to calculate a free R factor (Brünger, 1992a) as an independent cross-validation of the refinement progress in the structure determination of 1be0 was extended to also cover 5% of the additional data, resulting in a test set of 5193 reflections.

An initial round of rigid-body refinement was performed with *REFMAC* (Murshudov *et al.*, 1997) to position the molecule more accurately in the cell. Ten further cycles of *REFMAC* maximum-likelihood positional and B-factor refinement were carried out in combination with the automated refinement procedure *ARP* (Lamzin & Wilson, 1993, 1997), which was used to automatically model and update the solvent structure. For the rigid-body refinement, X-ray data in the 15–2.5 Å resolution range were used with no σ cutoff on the structure-factor amplitudes, and in subsequent *REFMAC* stages this range was changed to include reflections from 5 Å to the maximum resolution of 1.15 Å. Standard geometrical restraints and a 1:1 weight ratio of the X-ray *versus* geometry terms were applied. The only manual intervention at this stage was the replacement of one water molecule with a B value of 2.0 Å^2 by a chloride ion.

Refinement was continued with *SHELXL*, employing a strategy analogous to that suggested by the authors (Sheldrick & Schneider, 1997). Calculated values for the anomalous scattering factors f' and f'' at the wavelength of the experi-
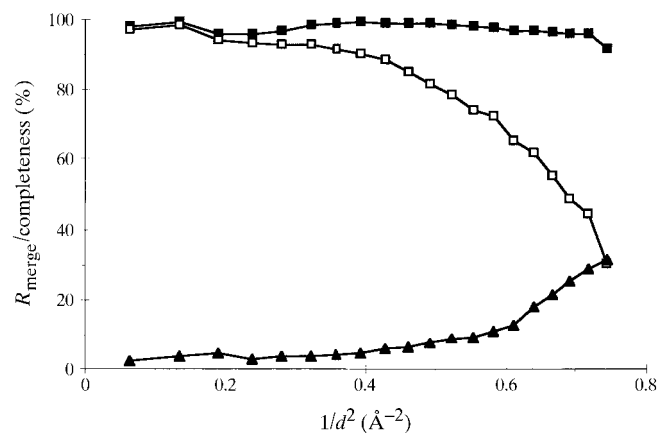
Figure 1
Completeness and R_{merge} of the data as a function of resolution. Filled squares indicate the completeness for all reflections, open squares for reflections with $I \geq 3\sigma(I)$. The filled triangles show the merging R factor for symmetry-related and multiply recorded reflections. The reflection bins are chosen such that they contain an equal number of reflections and the data are displayed at the mid-point of the bin.

ment were included for chloride and lead. In all *SHELXL* refinement cycles, restraints were applied to bond lengths and angles, chiral volume, planarity of appropriate torsion angles, atomic displacement parameters (ADPs) of neighbouring atoms, summed occupancy of alternative side-chain conformations and van der Waals contacts ('anti-bumping restraints'). At all stages, diffuse solvent was modelled using Babinet's principle (Moews & Kretsinger, 1975). An initial round of isotropic refinement was carried out against data in the resolution range 5–1.15 Å, after which, badly placed water molecules were removed. In subsequent rounds of refinement, 15–1.15 Å reflections were used and anisotropic ADPs were introduced, first for the heavier S atoms and the chloride ion and subsequently for all atoms. Alternative main-chain and side-chain conformations were modelled and the atom occupancies were refined using one common occupancy value for the atoms of one conformer. Protein H atoms were introduced at calculated positions and not refined. In the early stages of the refinement process, each round consisted of five (later ten) cycles of conjugate-gradient least-squares minimization against structure-factor amplitudes, including bulk-solvent correction. In the later stages, the model was refined in rounds which included five cycles of conjugate-gradient least-squares (CGLS) minimization plus four 'supercycles' of blocked-matrix least-squares (LS) minimization. Each 'supercycle' contained as many cycles as were needed to refine all parameters in blocks of about equal size (*ca* 1200–1300 parameters), allowing for some overlap between successive blocks. In the last two cycles, the 'anti-bumping' restraints were relaxed to 0.05 Å in order to decrease the bias that this restraint would introduce in the analysis of the interatomic contacts (see below).

After every refinement round, *PROCHECK* (Laskowski *et al.*, 1993) and *WHATCHECK* (Hooft *et al.*, 1996) were used to identify unusual stereochemical features of the model and σ_A-weighted $2F_o - F_c$ and $F_o - F_c$ Fourier electron-density maps (Read, 1986) were calculated. Visual inspection of the maps and manual rebuilding of the model were performed with *O* (Jones *et al.*, 1991). Water molecules were placed in spherical electron density of at least 4σ in the $F_o - F_c$ map, provided that the $2mF_o - DF_c$ map also showed reasonable electron density above the 1σ level. Moreover, the solvent molecule had to be within a distance of 2.2–3.5 Å from a neighbouring hydrogen-bond donor or acceptor. If the density criteria were fulfilled but the water molecule was positioned too close to a neighbouring atom, the solvent molecule was given a fixed occupancy of 0.5 to account for alternative solvent structure. Water molecules were removed if they lacked electron density above the 1.25σ level in the $2mF_o - DF_c$ map or if they had $B_{eq} \geq 80$ Å^2 (or $B_{eq} \geq 45$ Å^2 for molecules with half occupancy). When the refinement gave no further decrease in free R nor any improvement in stereochemistry, it was considered complete and one final round was performed including all X-ray diffraction data.

The structure was analysed using the program *VOIDOO* (Kleywegt & Jones, 1994) and programs from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) and

the *BIOMOL* package (Protein Crystallography Group, University of Groningen). C···C, C···O and N···O interatomic distances in the range 2.5–4.6289 Å were analysed as described by Derewenda *et al.* (1995), applying a radial distribution of 35 bins of an equal volume of 10 Å^3. The contacts were expressed binwise as a percentage of the total number of contacts found in the analysed range, in order to assess the portion of electrostatic contacts in the C···O contacts by subtracting the van der Waals contacts, estimated from the C···C contacts distribution, from the total number of C···O contacts. Only fully occupied water molecules and, for residues with multiple conformations, the main conformer were used. Short C—H···O contacts were analysed as in Derewenda *et al.* (1995). Calculated H-atom positions from the *SHELXL* refinement were used and all sp^2 hybridized O atoms were checked for the presence of an H atom from a C—H group within 2.7 Å distance. This value was chosen as a cutoff since it is the sum of the accepted van der Waals radii for a C—H H atom (1.2 Å) and an O atom (1.5 Å) (Taylor & Kennard, 1982). Geometrical parameters for the donor–acceptor distance C···O (d) and for the actual hydrogen-bonding distance H···O (d_H) were calculated and grouped in shells of an equal volume of 10 Å^3 (Derewenda *et al.*, 1995). Angular parameters ζ for the C—H···O and ξ for the H···O=C angles were determined as well as θ (the 'elevation angle'), which is the angle between the H···O=C hydrogen bond and the plane of the carbonyl C-atom sp^2 electron orbitals (Taylor & Kennard, 1982). Only contacts between amino acids which are at least three residues apart in sequence were taken into account and then only if $\zeta > 90°$ and $|\theta| < 45°$. All geometrical parameters were calculated with a modified version of *BIOMOL* software.

4. Results

4.1. Details of the refinement procedure

The first stage of the refinement (cycles 1 and 2), in which *REFMAC/ARP* was used, led to a decrease in the R value from an initial 35.7 to 28.0% after rigid-body refinement, and a further decrease to 18.6% after maximum-likelihood refinement of atomic positions and B values. The corresponding free R factors dropped from 38.1 to 27.7% and subsequently to 21.3%. A brief visual inspection of the model and corresponding $3F_o - 2F_c$ and $F_o - F_c$ maps identified one water molecule with an extremely low B value and additional positive difference Fourier electron density located between Trp125 and Trp175. It was replaced by a chloride ion. No continuous electron density for 1-chloropentane was found.

After the first round of isotropic refinement with *SHELXL* the model was inspected more carefully. The N- and C-terminal residues Met1 and Glu310 were removed because they lacked reliable $2mF_o - DF_c$ electron density. About 15 amino-acid side chains clearly showed more than one side-chain conformation, but only for Ser44 and Ser204 was the map sufficiently clear to model the alternative OG atoms. Furthermore, a difference density peak >20σ was observed at

Table 2
Course of the crystallographic refinement procedure.

Rounds 1–2: no rebuilding. Round 3: introduction of chloride ion in active site. Round 4: removed Met1, Glu310. Added covalently linked O atom to Cys150 SG. Double conformation assigned for Ser44 and Ser204. S atoms and chloride refined anisotropically. Round 5: added Met1, one sulfate ion. Double conformations built for the peptide bond between Asn14 and Leu15 and between Glu47 and Asp48 and for side chains of Pro29, Val49, Glu72, Glu95, Glu111, Met152, Thr181, Glu200, Thr213 and Gln228. All atoms refined anisotropically. Round 6: removed Met1, sulfate. Added Glu310, one glycerol. Replaced Cys150 OD with S atom. Double conformation assigned for Asp9, Asp137, Arg143, Lys192, Met225, Gln245 and Glu296. Introduced water molecules with half occupancy. Calculated 'riding' H atoms added on main-chain N, CA. Round 7: removed alternative conformation Gln228. Added Met1, one sulfate, two glycerol. Double conformation assigned for Pro138, Thr153 and Glu239. Calculated 'riding' H atoms added to all protein atoms. Round 8: removed alternative conformation Thr153, Lys192 and Glu296. Added two glycerols. Double conformation (re)introduced for Gln228, Lys261, Arg300 and triple conformation for Ser204 and Thr213. Round 9: double conformation assigned for Arg112, Lys192 and Gln283. Changed weighting scheme, CGLS + LS refinement against F_o^2s. Round 10: removed one glycerol and alternative conformation Asp137. Mutated Met1Ala. Added covalently linked O atom to Cys233 SG. Round 11: removed one sulfate and alternative conformation Lys192. Added one chloride and one glycerol. Replaced Cys150 SD with Hg atom. Double conformation (re)assigned for Asp137, Lys176, Arg186 and Ile241. Round 12: added one sulfate and one glycerol. Replaced Cys150 HgD and Cys233 OD with Pb atom. Double conformation assigned for Asp184. Round 13: Removed one sulfate. Added one glycerol. Double conformation reintroduced for Lys192. Refined against all data.

Round	1	2	3	4	5	6	7
Number of cycles	5	30	5	5	5	5	10
Number of atoms							
All (non-H)	2759	3055	3055	2911	3058	3113	3160
Protein†	2478	2478	2478	2465 (4)	2551 (164)	2587 (232)	2597 (240)
Water‡	281	577	576	445	501	519 (17)	539 (38)
Other	—	—	1 chloride	1 chloride	1 chloride, 1 sulfate	1 chloride, 1 glycerol	1 chloride, 1 sulfate, 3 glycerol
Data range (Å)	15–2.5	5.0–1.15	5.0–1.15	15–1.15	15–1.15	15–1.15	15–1.15
wR_2 at end§ (%)	—	—	42.0	43.2	33.0	30.9	28.0
R factor at end (%)	28.0	18.6	16.4	17.1	13.4	12.6	11.3
R_{free} at end (%)	27.7	21.3	19.6	19.9	17.3	16.8	15.4
Refinement type¶	rb-REF/ARP	iso-REF/ARP	iso-SH	(an)iso-SH	ani-SH	ani-SH	ani-SH

Round	8	9	10	11	12	13
Number of cycles††	10	5 + 3 (37)	5 + 3 (37)	5 + 4 (37)	5 + 4 (28)	5 + 4 (28)
Number of atoms						
All (non-H)	3296	3232	3199	3270	3300	3289
Protein†	2700 (249)	2623 (283)	2617 (277)	2629 (297)	2636 (309)	2639 (313)
Water‡	560 (46)	573 (52)	547 (62)	604 (69)	618 (89)	601 (113)
Other‡‡	1 chloride, 1 sulfate (1), 5 glycerol(1)	1 chloride, 1 sulfate (1), 5 glycerol(1)	1 chloride, 2 sulfate (1), 4 glycerol (1)	2 chloride (1), 1 sulfate (1), 5 glycerol (1)	2 chloride (1), 2 sulfate (1), 6 glycerol (1)	2 chloride (1), 1 sulfate (1), 7 glycerol (1)
Data range (Å)	15–1.15	15–1.15	15–1.15	15–1.15	15–1.15	15–1.15
wR_2 at end§ (%)	28.1	25.7	25.7	25.2	24.9	24.8
R factor at end (%)	11.2	10.4	10.9	10.7	10.6	10.5
R_{free} at end (%)	15.0	14.4	14.9	14.7	14.5	—
Refinement type¶	ani-SH	ani-SH	ani-SH	ani-SH	ani-SH	ani-SH

† Values in brackets are the number of protein atoms with occupancy less than unity. ‡ Values in brackets are the number of water molecules with half occupancy. § $wR_2 = 100\{\sum_{hkl}[w(F_o^2 - F_c^2)^2]/\sum_{hkl}[w(F_o^2)^2]\}^{1/2}$. ¶ rb-REF/ARP, rigid-body refinement with *REFMAC* in combination with *ARP*; iso-REF/ARP, isotropic refinement with *REFMAC* in combination with *ARP*; iso-SH, isotropic refinement with *SHELXL*; (an)iso-SH, isotropic refinement with *SHELXL*, including anisotropy for heavier atoms; ani-SH, anisotropic refinement with *SHELXL*. †† Values given are for the number of conjugate-gradient least-squares cycles (CGLS) plus the number of blocked-matrix least-squares 'supercycles' (LS); values in brackets are the number of blocked-matrix least-squares cycles that each 'supercycle' consisted of. ‡‡ Values in brackets are the number of molecules with occupancy less than unity.

a distance of 1.7 Å from Cys150 SG, which was preliminarily interpreted as a covalently linked O atom. Therefore, residue 150 was included in the model as an S-hydroxycystine, as has been observed previously in rhodanese (Gliubich *et al.*, 1996). Solvent sites which had no $2mF_o - DF_c$ electron density above 2σ were removed.

These adjustments resulted in an improvement of the density after round 4, which allowed modelling of alternative side-chain conformations of ten amino-acid side chains with confidence. Additionally, alternative orientations of the carbonyl group of the peptide bonds between residues 14–15 and 47–48 could be included in the model. Full anisotropic refinement then commenced. This led to a dramatic drop of 10.3% in wR_2, the R value on F^2, which can be used to assess the progress of the refinement in *SHELXL*, since the program

refines against F^2s and not Fs. The corresponding R and free R values decreased by 3.7 and 2.6%, respectively. Refinement proceeded with alternate cycles of model building and refinement (Table 2). After round 6, 'riding' H atoms were added to protein atoms at calculated positions. This lowered both the R and free R values by about 1.3%. From round 9 onward, the model was refined directly against measured F_o^2s (*SHELXL* option HKLF 4), instead of the structure-factor amplitudes F_o derived from them. This allows the direct inclusion of reflections for which $F_o^2 \le 0$ with good error estimates. Even when F_os are input, *SHELXL* refines against F_o^2s, since for small molecules this was shown to be more correct (Hirshfeld & Rabinovich, 1973; Watkin, 1994; Wilson, 1973, 1976). Furthermore, from this stage on, three or four 'supercycles' of blocked-matrix LS refinement were carried

Table 3
Refinement statistics and stereochemistry of the final model.

	REFMAC/ARP	SHELXL†
Resolution (Å)	5.0–1.15	15.0–1.15
R factor‡ (%)	18.6	10.5
Free R factor§ (%)	21.3	14.5
Number of non-H atoms		
Protein	2478	2480 (154)¶
Water	575	601 (113)††
Other‡‡	—	49 (12)††

R.m.s. deviations from ideal geometry

	Target	REFMAC/ARP	SHELXL†
Bond length (1–2) (Å)	0.02	0.008	0.014
Angle distance (1–3) (Å)	0.04	0.021	0.029
Planar distance (1–4) (Å)	0.05	0.027	0.044
Chiral volume (Å³)	0.15	0.10	0.56
Deviations from planes (Å)	0.02	0.02	0.03
Planar torsion angles (°)	7.0	4.0	5.9

	Target	REFMAC/ARP	SHELXL†
R.m.s. ΔB bond, main chain (Å²)	2.5	0.82	2.3
R.m.s. ΔB bond, side chain (Å²)	3.0	1.71	3.5
Mean B_{eq} (Å²)			
All atoms		16.3	18.9
Protein atoms		12.8	15.1
Other atoms		31.8	33.8

† Calculated geometry values output from *REFMAC*, using the final *SHELXL* model. The R factor results from a final round, including all reflections; the free R factor is taken from the penultimate round. ‡ $R = 100\sum_{hkl} ||F_o| - |F_c|| / \sum_{hkl} |F_o|$. § Free R factor calculated for 5% of the data omitted in the refinement. ¶ Value in brackets is the number of atoms for which an alternative conformation is modelled. †† Value in brackets is the number of atoms with occupancy less than unity. ‡‡ Glycerol molecules, chloride and sulfate ions.

out after the CGLS refinement. This reduced the R and free R values by a further 0.5%, at the cost of considerable computing time. The last cycles were used to make some final corrections to the geometry of the model, especially to the protein side chains with alternative conformations and to the modelling of the solvent structure. To analyse the nature of the atoms covalently linked to Cys150 and Cys233 SG, an anomalous difference Fourier electron-density map was calculated with coefficients $\Delta(F_{hkl}^{+} - F_{hkl}^{-})$ and phases derived from the model. The map showed clear peaks at the $+10\sigma$ and -10σ level around the Cys150 D atom and peaks at $+3\sigma$ and -3σ level around the Cys233 D atom (Figs. 5c and 5d). No other peaks were found above this level. Intuitively the atom was refined as Hg, as this element is commonly known to bind to cysteine S atoms (Blundell & Johnson, 1976). As SEM and AAS analyses indicated the sole presence of lead in the soaking solution, the heavy atom was modelled as such. The occupancy of the Cys150 PbD atom was related to that of the B conformer of Met152, as it was clear that the Pb atom was too close to the A conformer's methyl group. In the last stage of the refinement all data were used, yielding a final R factor of 9.72% for 80648 reflections with $F_o > 4\sigma(F_o)$ and 10.51% for all 94752 reflections in the resolution range 15–1.15 Å. The course of refinement is presented in Table 2.

4.2. Final model

The final model for haloalkane dehalogenase consists of 2639 protein-atom sites, representing 2478 protein atoms and two covalently bound Pb atoms, 601 water molecules, seven glycerol molecules, one sulfate ion and two chloride ions. Two peptide bonds and 26 side chains are modelled in two alternative conformations; two side chains have three alternative conformations. 113 water molecules are included with half occupancy and one glycerol molecule, one chloride ion and the sulfate ion are partially occupied. In general, the polypeptide chain fits the electron density very well and only a small number of side chains are poorly defined. Glu72 is modelled in two alternative conformations, but the CD, OE1 and OE2 atoms still lack electron density at the 1.0σ level. The same holds for the side-chain atoms of Lys192 and Arg300. All three residues point into the solvent and can easily adopt multiple side-chain orientations. Furthermore, the side chain of the N-terminal Met1 and the complete C-terminal Glu310 residue are not clearly visible in the 1.0σ-contoured $2mF_o - DF_c$ map, most likely because of flexibility, which is also represented by their high B_{eq} values of 51.6 and 71.6 Å², respectively. At a lower contouring level of 0.8σ, both terminal residues can be identified. The quality of the DhlA final model is summarized in Table 3. The excellent quality of the electron density can be judged from Fig. 2(a), where a well defined part of the model is shown.

The coordinate errors were determined from blocked-matrix least-squares refinement of positional parameters and ADPs (Fig. 3a). The overall r.m.s. coordinate errors for protein main-chain atoms and all atoms are 0.019 and 0.026 Å, respectively. For all atoms in the model the error is 0.038 Å, a value which is close to the mean value for the coordinate error of 0.045 Å estimated from a Luzzati plot (data not shown) (Luzzati, 1952). The coordinate error is highly correlated to the B_{eq} value (Fig. 3b); the correlation coefficients are 0.954 and 0.885 for main-chain atoms and all atoms, respectively.

Most of the mean values for main-chain bond lengths are in agreement with those from Engh & Huber (1991) which were used to restrain the protein's geometry in the refinement process. Nevertheless, the mean CA—C and C=O bond lengths of 1.520 (10) and 1.235 (11) Å, respectively, appear to deviate significantly from the small-molecule data [1.525 (21) and 1.231 (20) Å, respectively]. From the standard deviations, it is seen that the atomic resolution DhlA values are clustered more tightly around their mean.

For the main-chain bond angles a similar observation is made. The mean values for the AR structure have somewhat smaller standard deviations, and significant deviations from the Engh & Huber data are found for CA—C—N (except Gly, Pro), 117.4 (1.8) in DhlA *versus* 116.2 (2.0)° as found by Engh & Huber (1991); O=C—N (except Pro), 122.1 (1.4) *versus* 123.0 (1.6)°; C—N—CA (Gly), 122.1 (1.4) *versus* 120.6 (1.7)°; CA—C=O (except Gly), 120.4 (1.6) *versus* 120.8 (1.7)°; CB—CA—C (Ile, Thr, Val), 111.2 (1.8) *versus* 109.1 (2.2)° and CB—CA—C (except Ala, Ile, Thr, Val), 110.7 (1.9) *versus* 110.1 (1.9)°.
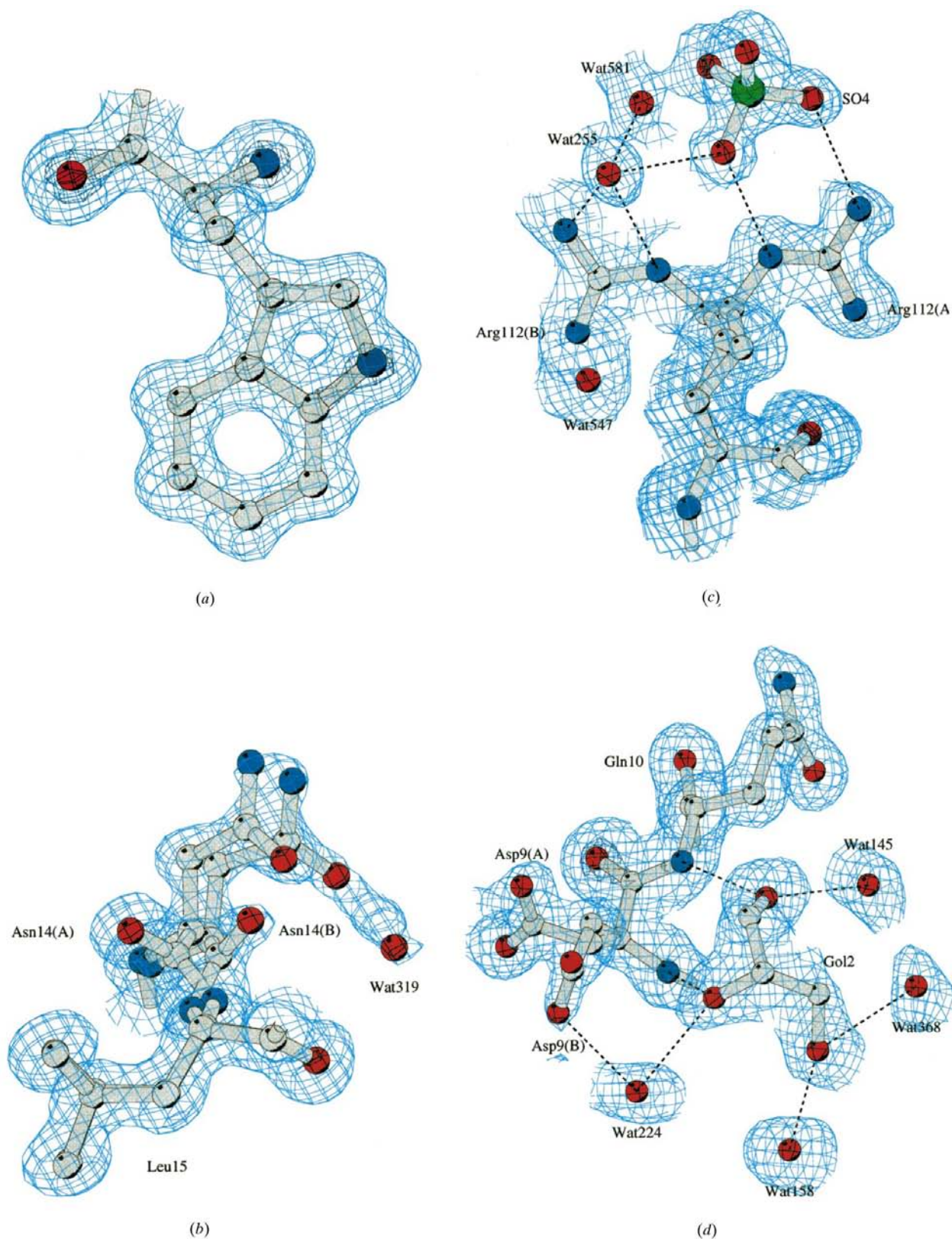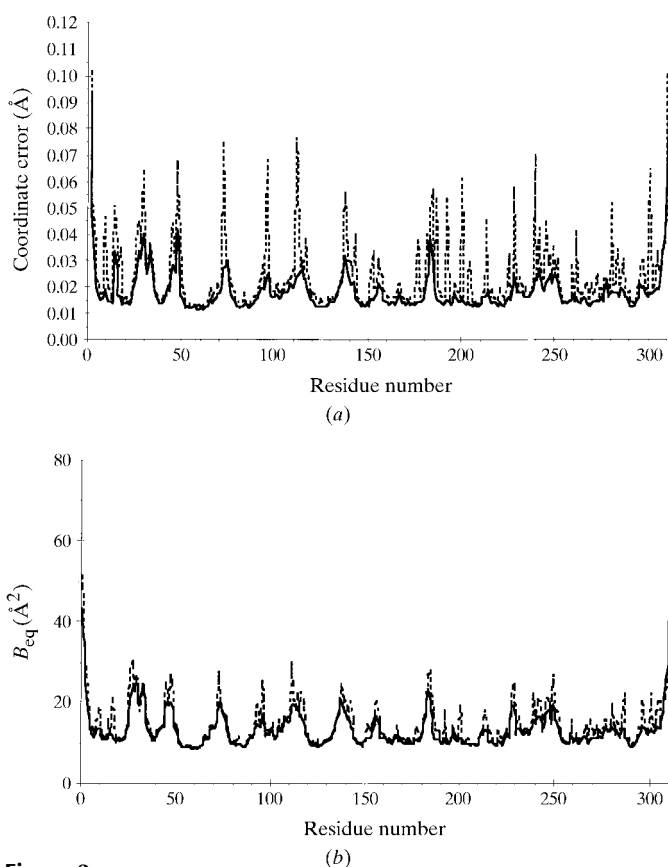
Figure 2
$2mF_o - DF_c$ electron density contoured at 0.82 e Å^{-3} (1.5σ, cyan) and 2.45 e Å^{-3} (4.5σ, dark blue) levels for (a) Trp194, (b) the peptide bond between Asn14 and Leu15, showing the two alternative main-chain conformations, (c) residue Arg112 as an example of a double side-chain conformation, also showing the partially occupied sulfate ion and two partially occupied water molecules, Wat547 and Wat581, and (d) a glycerol molecule. In (c) and (d), the low contouring levels are 0.55 e Å^{-3} (1.0σ) and 0.68 e Å^{-3} (1.25σ), respectively. Hydrogen bonds are depicted as dashed lines. This figure and Fig. 5 were generated with *BOBSCRIPT* (Esnouf, 1997).

Although the absolute differences from 'ideal values' are small, typically of the order of 0.005 Å and 1° for bond lengths and angles, respectively, they are significant at the 99% confidence level in a t-test. Deviations have not been found in other DhlA structures, suggesting that they can only be observed if data of sufficiently high resolution are available. As similar results can be obtained for other AR structures (data not shown), these results prove the need for a statistical analysis of protein geometry in atomic resolution structures.

In the Ramachandran plot of the main-chain φ,ψ angles (Ramachandran & Sasisekharan, 1968), 88.9% of the residues are found in the most-favoured regions, 10.7% in the additionally allowed regions and one residue, Asp124, is located in the generously allowed part of the plot (Fig. 4a). The main-chain ω angle was restrained by imposing the CA_n, C_n, O_n, N_{n+1} and CA_{n+1} atoms to lie in an unweighted least-squares plane with a standard deviation of 0.5 Å^3. For DhlA, the angle has an average uncertainty of 1.1°. The distribution of the ω values is approximately symmetric around its average value of 180 (6)° (Fig. 4b). A maximum deviation of 20° is observed for the peptide bond between residues Ile118 and Thr119. Large deviations from the ideal value of 180° are not uncommon and have been found throughout the PDB (MacArthur & Thornton, 1996) and especially in recent high-resolution structures (Sevcik $et\ al.$, 1996; Walsh $et\ al.$, 1998; Wilson $et\ al.$,

Figure 3
(a) Coordinate error estimated from matrix inversion (r.m.s. values per residue) for main-chain atoms (continuous line) and all protein atoms (dashed line); (b) B_{eq} values for main-chain (continuous line) and all protein atoms (dashed line), respectively.
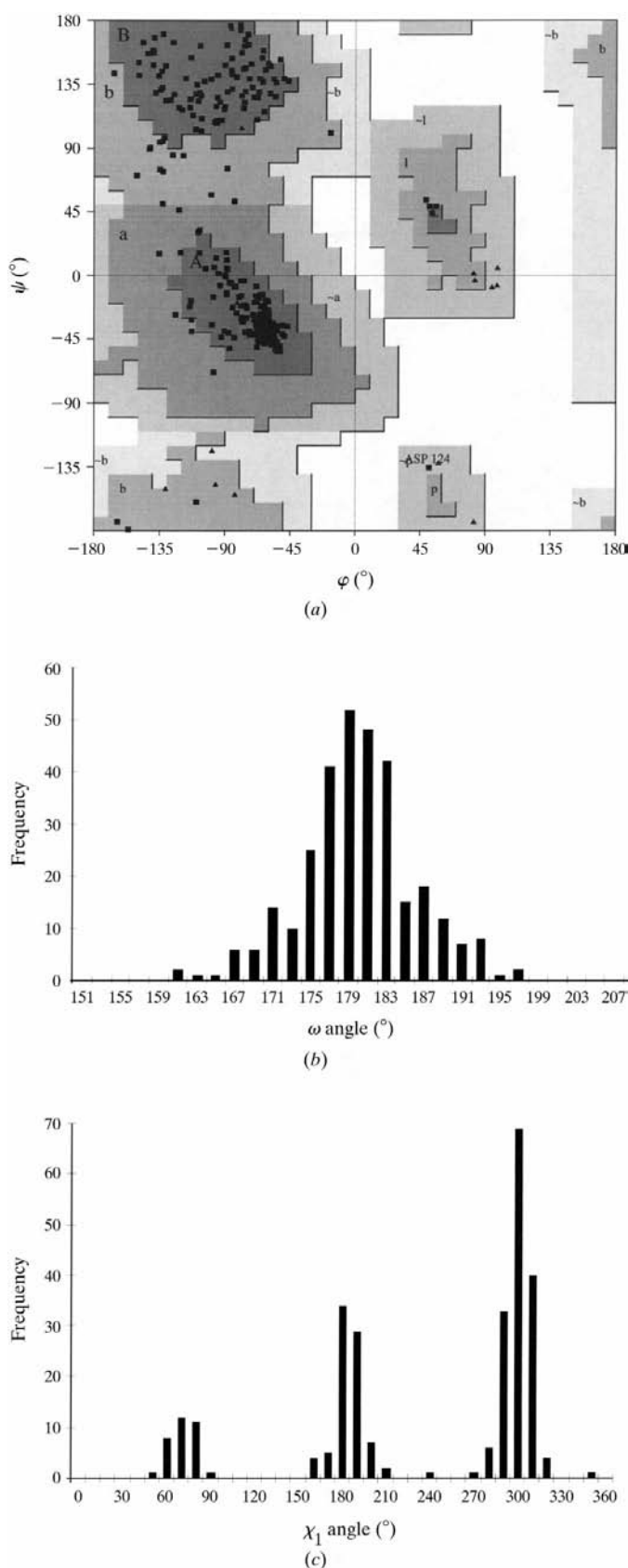
Figure 4
Distribution of the DhlA torsion angles (a) Ramachandran plot with glycine residues shown as triangles; (b) histogram of ω main-chain angles (°); (c) histogram of side-chain χ_1 angles (°).

1998). The two *cis*-peptides which are found, Glu56–Pro57 ($\omega = -20.3°$) and Gln167–Pro168 ($\omega = -5.4°$), were not included in the calculations. The side-chain rotamer χ_1 torsion angles were not restrained during the refinement since they can provide a useful validation tool. A histogram of the χ_1 distribution shows a relatively tight clustering around the three possible rotamers g^-, t and g^+ (Fig. 4*c*). The mean values and standard deviations of the three populations are 66 (8), 179 (8) and 295 (9)°, respectively, if alternative conformations are not included, and 66 (9), 181 (12) and 295 (10)° if they are taken into account. The two outliers in the histogram belong to alternative conformations of Glu72 ($\chi_1 = 233°$) and Asp137 ($\chi_1 = 265°$). Proline-residue χ_1 angles were omitted from both distributions. The mean values for the torsion angles are in agreement with those from Morris *et al.* (1992), but the smaller standard deviations indicate a tighter clustering around the mean. This was also observed by Wilson *et al.* (1998) in their profound analysis of eight atomic resolution structures.

5. Description of the structure and comparison with DhlA refined at 1.9 Å resolution

5.1. Overall structure and active site

The atomic resolution structure of DhlA has been determined at cryo-temperature. The high resolution of the data has allowed interpretation of the electron density to a high level of detail, including the modelling of anisotropy and disorder. The overall structure is essentially identical to the room-temperature apo-DhlA structures determined previously at pH 6.2 and pH 8.2 (PDB entries 2had and 1ede; Verschueren, Franken *et al.*, 1993) and also to the pH 6.2 structure complexed with a chloride ion (PDB entry 1edb; Verschueren, Seljée *et al.*, 1993).

The r.m.s. differences on CA atoms with respect to the room-temperature structures are in the range 0.31–0.34 Å. For all protein atoms, the values lie between 0.71 and 0.74 Å. If a residue had multiple conformations, only the major conformation was considered in the latter calculations. Differences in CA atom positions larger than 1 Å occur at the N- and C-termini and in the loop which connects strands β1 and β2. In the room-temperature structures, this loop (residues 26–33) was not very well defined and had relatively high *B* factors. The cryo-conditions might have decreased the mobility, enabling correct remodelling of the loop. The N- and C-terminal residues Met1 and Glu310 convincingly fit the electron density at a lower contour level of 0.8σ. When these regions are left out of the comparison, r.m.s. differences of 0.25 and 0.65 Å are calculated for CA atoms and all protein atoms, respectively. These remaining differences cannot be attributed to a specific region of the protein and probably result from movements in the protein induced by freezing. Compared with the starting model of DhlA at pH 5 and 120 K, the atomic resolution structure has r.m.s. differences on CA and all protein atoms of 0.24 and 0.54 Å, respectively. These differences can be attributed almost entirely to different modelling of the N- and C-terminal residues.

Table 4
Amino acids with more than one conformation.

Residue number	Accessibility of side-chain†	Occupancy	Atoms used to refine occupancy
Asp9	0.79	0.61	CB CG OD1 OD2
Asn14	—	0.59	N CA CB CG OD1 ND2 C O
Leu15	—	0.59	N CA CB CG CD1 CD2
Pro29	0.80	0.53	N CD CG CB CA
Ser44	0.70	0.59	CB CG
Glu47	0.90	0.63	N CA CB CG CD OE1 OE2 C O
Asp48	—	0.63	N CA CB CG OD1 OD2
Val49	0.04	0.66	CB CG1 CG2
Glu72	0.86	0.58	CB CG CD OE1 OE2
Glu95	0.77	0.64	CB CG CD OE1 OE2
Glu111	0.63	0.63	CB CG CD OE1 OE2
Arg112	0.58	0.58	CB CG CD NE CZ NH1 NH2
Asp137	0.31	0.63	CB CG OD1 OD2
Arg143	0.06	0.52	CB CG CD NE CZ NH1 NH2
Met152	0.08	0.86	CB CG SD CE
Lys176	0.29	0.65	CB CG CD CE NZ
Thr181	0.51	0.69	CB OG1 CG2
Asp184	0.57	0.56	CB CG OD1 OD2
Arg186	0.56	0.56	CG CD NE CZ NH1 NH2
Lys192	0.33	0.61	CB CG CD CE NZ
Glu200	0.75	0.68	CB CG CD OE1 OE2
Ser204‡	0.64	0.45, 0.41	CB OG
Thr213‡	0.46	0.51, 0.33	CB OG1 CG2
Met225	0.09	0.53	CB CG SD CE
Gln228	0.92	0.50	CB CG CD OE1 NE2
Glu239	0.60	0.64	CB CG CD OE1 OE2
Gln245	0.48	0.57	CB CG CD OE1 NE2
Lys261	0.56	0.60	CB CG CD CE NZ
Glu280	0.65	0.50	CB CG CD OE1 OE2
Glu283	0.38	0.56	CB CG CD OE1 OE2
Arg300	0.53	0.62	CG CD NE CZ NH1 NH2

† Fractional solvent accessibility of the main side-chain conformer. ‡ Residue modelled with three alternative conformations.

In the Ramachandran plot (Fig. 4*a*) only one residue (Asp124) is found outside the most-favoured or additionally allowed regions, whereas in all previously described structures Asn148 was also observed to have unfavourable φ,ψ torsion angles (Krooshof *et al.*, 1998; Verschueren, Franken *et al.*, 1993). The main-chain torsion angles in this area of the atomic resolution structure are $\omega_{147} = -176°$, $\varphi_{148} = -17°$, $\psi_{148} = 101°$, $\omega_{148} = 172°$ and in the structure of DhlA at pH 6.2 and room temperature (Verschueren, Franken *et al.*, 1993) these values differ by up to 10° ($\omega_{147} = 179°$, $\varphi_{148} = -10°$, $\psi_{148} = 94°$, $\omega_{148} = -177°$), yet the r.m.s. difference between all atoms of residues 147–149 is small (0.12 Å). These angular values for 2had are very similar to those of 1be0, the starting model in the refinement. It is clear that not only do the φ,ψ torsion angles in the AR structure differ, but also the ω dihedral angle. In the crystallographic refinement of both 2had and 1be0, the ω dihedral angle was restrained. In 2had, which was refined with *TNT* (Tronrud, 1997), a restraint was set on the plane formed by CA_n, C_n, O_n, N_{n+1} and CA_{n+1} atoms, with a standard deviation equal to that of other planes, *e.g.* of aromatic side chains. In *X-PLOR* (Brünger, 1992*b*), which was applied in the refinement of 1be0, an appreciable force constant was put on the ω dihedral angle to keep it in a *trans* configuration, as well on the two improper angles formed by the CA_n, C_n, O_n, N_{n+1} and $H(N_{n+1})$ atoms to keep them in

one plane. The *SHELXL* parameters chosen in the refinement of the AR model impose flatness on the five atoms which define the peptide plane. However, the standard deviation was set to a value five times higher than on the planes of the aromatic side chains, allowing much more flexibility in the ω dihedral angle (see above). Although the conclusion still holds that Asn148 has unusual main-chain torsion angles implying strain in this region of the structure, the AR structure provides a nice example of how restraints in crystallographic refinement might overemphasize such features if these restraints are incorrectly parameterized.

The active site of DhlA is located in a central cavity between the main and core domains of the protein. The shape of the cavity in the AR structure is very similar to that described previously for 2had. The r.m.s. differences for the positions of all atoms of the residues which line the cavity are in the range 0.15–0.19 Å. The volume of the active-site cavity is 75 Å^3 for the atomic resolution structure; the cavity volume of other DhlA structures does not deviate by more than 15%. The largest differences are found for atoms of Pro223, which has a flat ring – the average of the two possible rotameric states – in most of the other structures, and the Asp124 carboxylate O atoms. In the electron-density map, clear anisotropy is present around the Asp124 OD1 atom (Fig. 5a). It can be seen as static disorder in which, in some of the molecules, Asp124 OD1 is rather far apart from His289 NE2. From the anisotropy, the distance can be estimated to range from 2.6 to 3.2 Å. These values are close to the distances observed in the pH 8.2 apo-structure ($d_{O-N} = 3.3$ Å) and the pH 6.2 apo-structure ($d_{O-N} = 2.7$ Å). In previously determined DhlA structures with chloride bound at low pH this distance is even shorter, with a minimum of 2.3 Å. This is a consequence of the protonation state of His289; if the NE2 atom is deprotonated, no strong attractive interaction with Asp124 exists, whereas if the histidine is protonated, a hydrogen bond can be formed. Hence, from the anisotropy in this high-resolution structure it is apparent that both the protonated and deprotonated state are present in the enzyme, even at a pH as low as pH 5.

Apart from a chloride ion, discussed in §5.3, two water molecules are also bound inside the cavity. One of them, Wat197, is hydrogen bonded to the chloride ion, the second water molecule and Asp124 OD1 (Fig. 5b). This water molecule has been found in previous structures, unlike the second water molecule, Wat518. This newly observed water molecule was refined at half occupancy to match the height of the electron-density peak. The molecule is hydrogen bonded to Wat197 only, but it makes van der Waals interactions with the apolar side chains of Leu262 and Leu263 (Fig. 5b). When the structure is superimposed on the DhlA structure with DCE bound (PDB entry 2dhc; Verschueren, Seljée *et al.*, 1993; not shown), the second water molecule is at a distance of 1.4 Å from the Cl2 atom of the substrate. It is known that 1-chlorobutane is the chlorinated substrate with the longest alkyl tail that is efficiently degraded by DhlA (Janssen *et al.*, 1991). Thus, Wat518 might indicate the position of the C4 atom of 1-chlorobutane. Clearly, the environment of the water
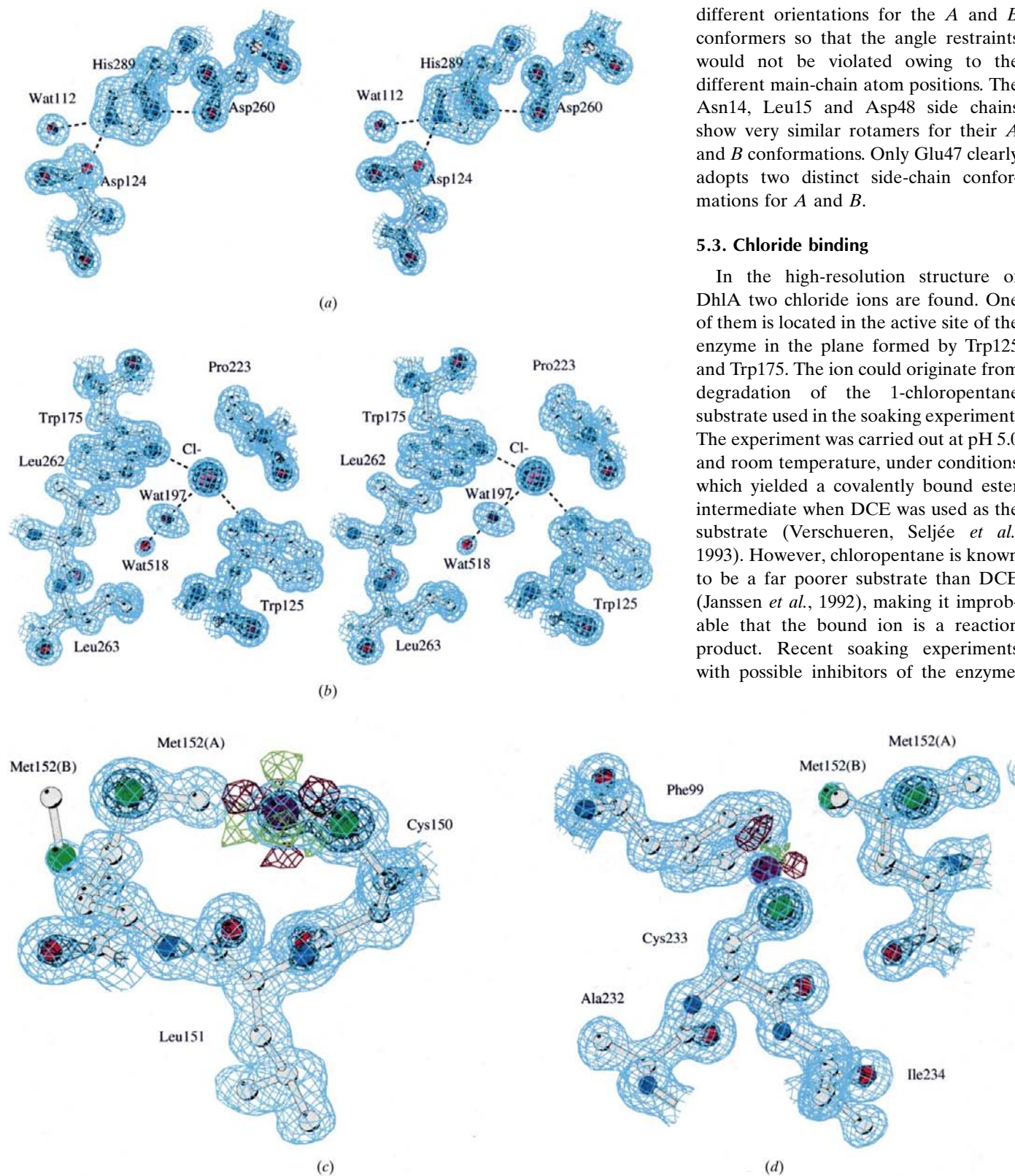
molecule in this structure is more suitable for an apolar functional group and the position of the water molecule can therefore facilitate modelling of 1-chlorobutane binding.

5.2. Regions of the protein with alternate conformations

In haloalkane dehalogenase, two parts of the main chain and 28 side chains were modelled with multiple conformations, two of which consist of three conformers and the other 26 having double conformations. Most of these side chains are charged, with a high percentage of the total number of glutamate and arginine residues in the protein involved: 44 and 27%, respectively. Two out of every three discreetly disordered residues are located at the surface of the protein, but a few completely buried side chains also display disorder. Residues with alternate conformations are listed in Table 4 and two examples of how they fit the electron density are shown in Figs. 2(b) and 2(c).

Nearly all amino acids with alternative conformations are isolated from one another; only two pairs of residues are found which are close enough to form a small hydrogen-bonding network. The negatively charged Glu111 and Asp137 are close in space and the *A* conformers of their side chains are oriented such that they form a hydrogen bond from Glu111 OE2 through Wat486 to Asp137 OD2. This is similar to the conformation which was observed in the 1.9 Å structure (Verschueren, Franken *et al.*, 1993). The *B* conformers use Wat250 as an intermediate in an indirect contact between Glu111 OE2 and Asp137 OD1. The other pair of residues with alternate conformations are Asp184 and Arg186. The *A* conformers make a direct hydrogen bond from Asp184 OD2 to Arg186 NH2, whereas the *B* conformers form a direct contact from Asp184 OD1 to Arg186 NE and NH2. The previously determined DhlA structures present a kind of mixture between the *A* and *B* state, with the guanidinium group of Arg186 in a *B*-like orientation and the carboxylate group of Asp184 similar to the *A* conformer. However, the residues are located further apart from each other. Walsh *et al.* (1998) observed in the 0.925 Å resolution structure of lysozyme that a large part of the disordered residues interacted with each other, either directly or *via* a water molecule. That this is not found with DhlA is largely a consequence of the different maximum resolution which allowed the assignment of alternative conformations for about 10% of the side chains, whereas in the lysozyme structure 24 out of 129 residues are modelled with disorder.

Apart from side-chain disorder, two parts of the main chain also occur in more than one conformation. The peptide bond between Asn14 and Leu15 can be described by the main-chain angles $\psi_{13} = 103$, $\omega_{14} = 179$, $\varphi_{14} = -140°$ for the *A* conformation and $\psi_{13} = 30$, $\omega_{14} = 176$, $\varphi_{14} = -58°$ for the *B* conformation, which is partially flipped with respect to *A* (Fig. 2b). A similar observation can be made for the peptide bond between Glu47 and Asp48, with main-chain angles $\psi_{46} = -44$, $\omega_{47} = -175$, $\varphi_{47} = -87°$ and $\psi_{46} = -3$, $\omega_{47} = 177$, $\varphi_{47} = 149°$ for the *A* and *B* conformation, respectively. For these residues, the side chains were also allowed to take up

different orientations for the *A* and *B* conformers so that the angle restraints would not be violated owing to the different main-chain atom positions. The Asn14, Leu15 and Asp48 side chains show very similar rotamers for their *A* and *B* conformations. Only Glu47 clearly adopts two distinct side-chain conformations for *A* and *B*.

5.3. Chloride binding

In the high-resolution structure of DhlA two chloride ions are found. One of them is located in the active site of the enzyme in the plane formed by Trp125 and Trp175. The ion could originate from degradation of the 1-chloropentane substrate used in the soaking experiment. The experiment was carried out at pH 5.0 and room temperature, under conditions which yielded a covalently bound ester intermediate when DCE was used as the substrate (Verschueren, Seljée *et al.*, 1993). However, chloropentane is known to be a far poorer substrate than DCE (Janssen *et al.*, 1992), making it improbable that the bound ion is a reaction product. Recent soaking experiments with possible inhibitors of the enzyme,

Figure 5
$2mF_o - DF_c$ electron density contoured at 0.82 e Å^{-3} (1.5σ, cyan) and 2.45 e Å^{-3} (4.5σ, dark blue) levels for (*a*) the catalytic triad Asp124–His289–Asp260, illustrating the anisotropy around the Asp124 OD1 atom (stereoview); (*b*) the chloride ion, Wat197 and Wat518 in the active site plus surrounding residues (stereoview); (*c*) the Pb atom (purple) bound to Cys150 SG, also showing the neighbouring Met152 in two conformations, (*d*) the same for Cys233 SG with the neighbouring Phe99 residue. In the latter two panels the anomalous difference electron density is also displayed, contoured at +0.15 e Å^{-3} (+8.0σ, green) and −0.15 e Å^{-3} (−8.0σ, brown) and at +0.06 e Å^{-3} (+3.0σ, green) and −0.06 e Å^{-3} (−3.0σ, brown) levels for (*c*) and (*d*), respectively. In (*c*) especially, the orbitals of the electrons in the *L* shell can clearly be discriminated in the shape of the anomalous difference density.

taking a crystal from the same batch, also showed electron density for a chloride ion in this position (Nardini, 1998). Thus, the presence of the ion is more likely to arise from the crystallization medium or the PEG 6000 in the cryoprotectant solution, which might have contained traces of chloride.

The position of the chloride ion has been very accurately determined with a coordinate error of 0.004 Å, which is the lowest in the structure. The B_{eq} of the ion is 9.64 Å^2, which is close to 10.28, the B_{eq} of the surrounding atoms. The ion is hydrogen bonded to Trp125 NE1 (d_{Cl-N} = 3.28 Å), Trp175 NE1 (d_{Cl-N} = 3.25 Å) and a water molecule, Wat197, which is bound in the active site (d_{Cl-O} = 3.07 Å). These hydrogen-bond donor–acceptor distances are close to the values reported for organic molecules of d_{Cl-N} = 3.181 (6) Å and d_{Cl-O} = 3.190 (3) Å, respectively (Steiner, 1998). Other residues within 4 Å include Glu56 (CB, CG), Phe222 (C, O), Pro223 (N, CD, CA, CB) and Val226 (CG2) (Fig. 5b). The chloride-ion environment is essentially identical to that observed in the chloride-bound structure at pH 8.2 and room temperature (Verschueren, Kingma et al., 1993) and the chloride-bound structure which was observed after soaking with DCE substrate at pH 6.2 and room temperature (PDB entry 2dhe; Verschueren, Seljée et al., 1993). In the previously determined structures, a tendency for the ion to bind closer to Trp175 was observed. In view of the more accurate atomic resolution structure described here, where the chloride ion is equidistant from the tryptophan rings, this tendency should not be attributed any significance. However, it can be argued that binding to Trp175 is stronger, as the chemically more relevant H···Cl$^-$ distance to Trp175 is shorter by 0.20 Å as the N—H···Cl$^-$ angle is more linear in the case of Trp175 (165 versus 137°).

A second chloride ion is bound on the surface of the protein close to Thr197. It was identified during the refinement process, when the water molecule which was originally modelled at that position had a B_{eq} lower than its environment and still showed over 4σ positive electron density in the difference Fourier map. This did not allow for the incorporation of a fully occupied chloride ion. In the structure, a chloride ion and a water molecule are now modelled at this position with an occupancy and coordinate error of 0.38 (2) and 0.02 Å, respectively, for the chloride ion and 0.62 (2) and 0.04 Å, respectively, for the water molecule. The B_{eq} values of the ion and the water molecule are 11.66 and 13.32 Å^2, respectively, which are both close to 13.80, the B_{eq} of the surrounding atoms. The ion is hydrogen bonded to Thr197 OG1 (d_{Cl-O} = 3.01 Å) and two water molecules, Wat12 (d_{Cl-O} = 3.01 Å) and Wat150 (d_{Cl-O} = 2.96 Å). Other residues within 4 Å include Ala195 (CA, CB, C), Pro196 (N, CD) and one further water molecule, Wat335.

This secondary halide-binding site in DhlA was identified recently in a crystal structure which resulted from the soaking of a DhlA crystal in 0.5 M sodium bromide (PDB entry 1cij; Pikkemaat et al., 1999). The 2.3 Å resolution DhlA structure with a fully occupied bromide ion bound at this position is similar to the high-resolution structure described in this article, except for the two water molecules Wat12 and Wat150,

which were not previously observed owing to the medium resolution. In the same 2.3 Å resolution DhlA structure, a second surface bromide-binding site was found in the cap domain, close to Gln216. This position is taken up by a water molecule in the atomic resolution structure. Our findings that at low chloride concentration only the first halide-binding site is occupied supports the conclusion from the site-directed mutagenesis and kinetic experiments of Pikkemaat et al. (1999) that only the first site is important. It is the location of the 'collision complex' which was identified by kinetic measurements as an intermediate in the import of halide ions to the active site (Schanstra & Janssen, 1996). Although only the reverse process, halide export, is of direct relevance as the rate-limiting step in the dehalogenation reaction, both structures provide valuable information on the halide-transport route, since the 'collision complex' is located close to the exit of a tunnel which leads out of the active site to the surface of the enzyme. The existence of this tunnel was proposed earlier by Franken et al. (1991).

5.4. Bound ions and solvent structure

5.4.1. Sulfate ion. In the atomic resolution structure of DhlA, a previously undetected sulfate ion is bound at the surface of the protein, close to Asp26 and partially overlapping the A conformer of the Gln228 OE1 atom. Therefore, the ion was constrained to have the same occupancy as the Gln228 B conformer, which refined to 0.50. Its $\langle B_{eq}\rangle$ is 16.89 Å^2, similar to that of the protein atoms to which it is hydrogen bonded. Each of the four sulfate O atoms is hydrogen bonded to two proton donors: Asp26 N, the NE and NH2 atoms of the A conformer of Arg112 and four water molecules with full occupancy and two water molecules with half occupancy (Fig. 2c). Three of the water molecules belong to a symmetry-related molecule. Although the protein is crystallized from ammonium sulfate, the bound sulfate ion is not necessarily crucial for crystallization, since it is present in only 50% of the intermolecular contacts. In those molecules where the sulfate is absent, intermolecular contacts are formed from Gln228A OE1 to Arg112A NE and/or Arg112B NE.

5.4.2. Heavy-atom binding. In the initial Fourier maps, two of the four cysteine residues in the structure showed spherical electron density beyond the SG atom at a distance indicative of covalent binding. At first, it was believed that the sulfur was oxidized by molecular oxygen, but a clear peak in an anomalous difference Fourier map indicated the binding of a heavy atom. Chemical analysis of the soaking solution showed the presence of lead, which must have been a contaminant of one of the reagents. As the lead L_{III} edge (λ = 0.9507 Å) is close to the wavelength of the diffraction experiment, a considerable anomalous signal can be expected. Lead binding to DhlA has not been observed previously, although it has not been tested extensively. In the heavy-atom search during the DhlA structure determination (Franken et al., 1991) one lead compound was tested, PbSO$_4$, but no differences were observed on a precession photograph. This

does not exclude the possibility of lead binding to DhlA, as it was also impossible to localize the heavy atoms in our current structure in an isomorphous difference Patterson map (data not shown). We assume that the two amino acids have been modified to a Cys–Pb$^+$ entity. The S—Pb distances and CB—SG—Pb angles were not restrained and they refined to values of 1.64 Å, 138° and 1.62 Å, 98° for residues 150 and 233, respectively. No positively charged C—S—X (X = any element) entity is present in the PDB or the Cambridge Structural Database (CSD; Allen *et al.*, 1979), preventing a comparison with known data.

The Cys150 Pb atom has an occupancy of 0.15, constrained in the refinement to be equal to that of the B conformer of Met152, as the A conformer is too close to the Pb atom (Fig. 5c). Its B_{eq} is 10.61 Å^2, a little lower than that of the S atom to which it is bound and it has a small coordinate error of 0.006 Å. The atom does not have any interactions within 3.5 Å, but it is located about 3.5 Å above the aromatic plane of the Phe128 side chain, possibly interacting with the partially negative cloud of π electrons. A similar interaction is observed for the Pb atom attached to Cys233 SG, which is close to the aromatic ring of Phe99 (Fig. 5d). The only other atom within 3.5 Å is Met152B SD. The Cys233 Pb atom has a low occupancy of 0.05, *i.e.* accounting for four electrons, and a B_{eq} of 14.71 Å^2, which is close to the value for the covalently linked S atom. Its coordinate error is 0.02 Å, comparable with that of other light atoms.

5.4.3. Glycerol molecules. Furthermore, seven glycerol molecules were found in the structure at the surface of the protein: for an example, see Fig. 2(d). Glycerol molecules were not observed in any of the room-temperature structures – they originate from the cryoprotectant solution – nor in any of the other DhlA structures determined at cryo-temperature. This might be caused by the inherent rotational freedom of the molecule. In DhlA, the glycerol molecules have a high $\langle B_{eq}\rangle$ of 39.9 Å^2 and an average coordinate error of 0.08 Å. The molecules are fully occupied, except for GOL4 which partially overlaps the A conformation of Glu95 and is therefore refined with the same occupancy as the Glu95 B conformer (0.36). All glycerol O atoms form at least one hydrogen bond, with an average of 7.2 per molecule, one-third of which are interactions with the protein and two-thirds with solvent.

5.4.4. Water structure. At the start of the refinement, all 334 water molecules from the model were retained, since this 2.0 Å resolution structure resulted from a similar cryo-experiment. Initially, *ARP* was used to update the solvent model and in later stages manual rebuilding based on geometrical and electron-density based criteria was applied. These criteria first led to a rigorous removal of over 100 water molecules, but in subsequent building cycles the solvent structure was slowly extended. In all stages, a two-parameter bulk-solvent model was used. The final DhlA model contains 601 water molecules, 113 of which were included with half occupancy to account for an alternative solvent structure. The occupancies were not refined. The molecules with full occupancy have a $\langle B_{eq}\rangle$ of 34.3 Å^2 and an average coordinate error of 0.09 Å, and the half-occupied molecules have a $\langle B_{eq}\rangle$ of

25.6 Å^2 and an average coordinate error of 0.09 Å. Somewhat less than half the water molecules have equivalents in the 2.0 Å resolution cryo-structure (1be0), *i.e.* located within 1.1 Å when the two structures are superimposed (Fig. 6). In total, 270 molecules are close to one of the 334 solvent sites which were identified in 1be0. This is slightly lower for the 2.0 Å room-temperature structure (2had), which has 170 out of 248 sites in common with the final DhlA model. It should be noted that nearly all internal water molecules have an equivalent in both 1be0 and 2had, except for those in the active site, as discussed above. Moreover, many of the water molecules were retained from 1be0, the starting model. Differences to previously determined DhlA structures are in part a consequence of the introduction of multiple side-chain conformations. Other differences are caused by the different temperature of the experiment (with respect to 2had) and obviously by the higher resolution of the data, which allowed for the modelling of more solvent at the protein surface.

5.5. C—H$\cdots$O hydrogen bonds

In a recent analysis of several medium and high-resolution structures, Derewenda *et al.* (1995) concluded that C—H$\cdots$O bonds are ubiquitous in macromolecular structures and might contribute significantly to the free-energy difference between the folded and unfolded state of proteins. An analysis of C$\cdots$C, C$\cdots$O and N$\cdots$O interatomic distances can be used as a first indication of the presence of such cohesive interactions (Fig. 7). The strong cohesive N$\cdots$O electrostatic interaction is immediately obvious from the distance distribution, with a maximum around 2.9 Å, clearly below the sum of the van der Waals radii of the two atoms. In contrast, the C$\cdots$C distance curve shows no apparent maximum, and less than 1.5% of the contacts are shorter than 3.5 Å, the sum of the van der Waals radii of two C atoms. This confirms that carbon–carbon interactions are essentially of the van der Waals type. The C$\cdots$O distance distribution has mixed character, typical of both electrostatic and van der Waals contacts. This can be
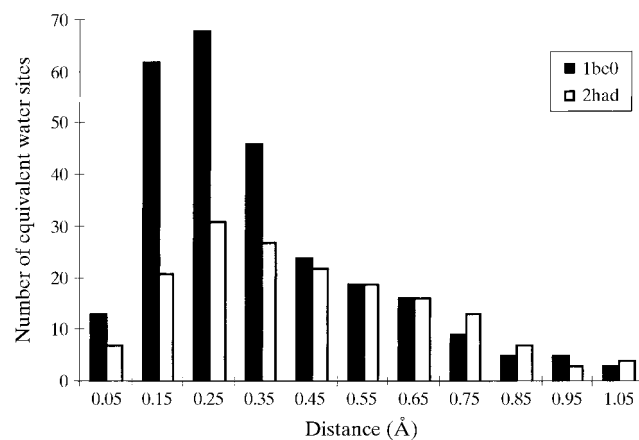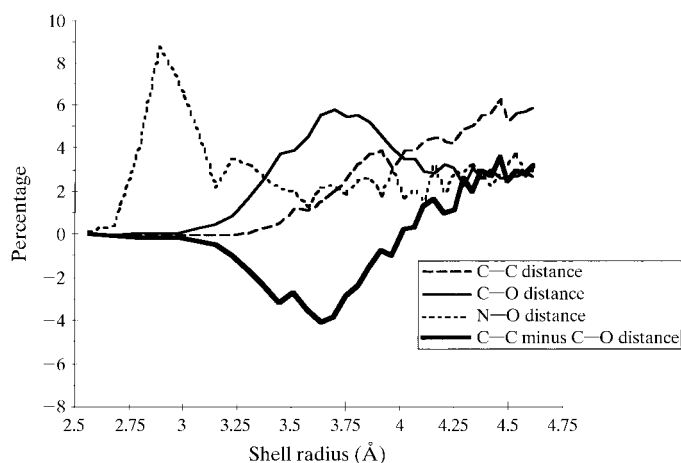

Figure 6
Histogram of number of equivalent water molecules in DhlA at atomic resolution *versus* the 2.0 Å cryo-temperature model (1be0, filled bars) and *versus* the 2.0 Å room-temperature model (2had, open bars).

illustrated by the subtraction of the van der Waals portion of the C$\cdots$O contacts, represented by the typical van der Waals C$\cdots$C contact-distance distribution, from all C$\cdots$O contacts. The resulting difference curve shows a broad minimum around 3.65 Å, signifying that a major portion of these C$\cdots$O contacts is shorter than 3.7 Å – the sum of the van der Waals radii for carbon and oxygen. Hence, these contacts can be attributed to cohesive electrostatic interactions, namely hydrogen bonds (Derewenda *et al.*, 1995).

The use of the sum of the van der Waals radii of the donor and acceptor atom as a cutoff criterion for hydrogen bonding is inaccurate (Jeffrey & Sänger, 1991). The hydrogen–acceptor atom distance and directionality of the electrostatic interaction should also be included in any analysis. The high resolution of our data enabled the reliable inclusion of calculated H atoms in haloalkane dehalogenase during refinement. Based on these positions, the short C$\cdots$O contacts were analysed for C—H$\cdots$O hydrogen-bond formation. The analysis was restricted to carbonyl and carboxyl sp^2 hybridized O atoms, since they are the most abundant C—H$\cdots$O hydrogen-bond acceptors (Taylor & Kennard, 1982), and the plane in which the lone-pair electrons are located can reliably be derived from the adjacent non-H atoms. In DhlA, a total number of 141 C—H$\cdots$O interactions are found for which $d_H < 2.7$ Å. Other restrictions applied to the analysis were a spacing of at least three residues in sequence between donor and acceptor atom, an interatomic C—H$\cdots$O angle ζ of at least 90° and an absolute value of the 'elevation angle' θ of less than 45°.

The average d_H is 2.55 (13) Å, and if only the 23 C—H$\cdots$O interactions of main-chain atoms in the central β-sheet are counted, this value is 2.46 (11) Å (Fig. 8a). The CA—H$\cdots$O interaction in β-sheets can be considered a special class of this type of interactions, as they are relatively common (Derewenda *et al.*, 1995). The directionality of the hydrogen bond is

not very pronounced, as the average ζ is 142 (15) and 146 (9)° for all interactions and β-sheet interactions, respectively. It should be noted that, from pure geometrical considerations, the number of contacts found in the range from ζ to $\zeta + \delta\zeta$ is proportional to $\sin\zeta \; \delta\zeta$ (Kroon & Kanters, 1974; Taylor & Kennard, 1982). Applying this 'cone correction' leads to a distribution indicative of linearity of the CA—H$\cdots$O angle as would be expected for a hydrogen bond, although this is far from convincing (Fig. 8b). In an analysis of the β-sheet of 11 unique protein structures with a resolution better than 1.3 Å, Fabiola *et al.* (1997) found very similar results for the hydrogen-bond lengths and angles: $d_H = 2.38$ (13) Å and 143 (6)°, respectively, showing that these are general properties in protein structure. For strong hydrogen bonds, *e.g.* O—H$\cdots$O in different ice polymorphs, d_H is much shorter, ranging from 1.73 to 1.96 Å, and the bond is linear, with variations up to 15° (Jeffrey & Sänger, 1991).

Two other important angular parameters are the H$\cdots$O—C angle ξ and the 'elevation angle' θ, which describe the approach of the hydrogen donor with respect to the lone pairs of the O atom. In principle, their values should be around 120 and 0°, respectively, for ideal sp^2 hybridization of the O atom. In DhlA, the average ξ value for all 141 C—H$\cdots$O interactions is 137 (15)° (Fig. 8c). The specific β-sheet interactions show a more narrow distribution around the same average, 137 (10)°. The θ values show a very broad distribution around the averages of 1 (26) and 6 (18)° for all interactions and β-sheet interactions, respectively (Fig. 8d). The average values for ξ and θ are in agreement with the study of Derewenda *et al.* (1995), who discriminated between parallel and antiparallel strands and found values of 135 (8) and 140 (9)°, respectively, for ξ and of 5 and −9°, respectively, for θ.

The analysis of C—H$\cdots$O distances in DhlA clearly supports the conclusion that C—H$\cdots$O hydrogen bonds exist in proteins, just as they have been demonstrated to exist in small organic molecules (Desiraju, 1996; Steiner, 1997) and nucleic acids (Leonard *et al.*, 1995). At the same time, it should be noted that most of the angular parameters are rather different from their ideal values. An explanation for this might be that this kind of attractive interaction is weaker than a 'classical' N—H$\cdots$O or O—H$\cdots$O hydrogen bond; Desiraju (1991) reports values ranging from −1.3 to −21 kJ mol^{-1}, although the latter value is an extreme case, irrelevant for protein structures. Unfortunately, no direct quantum-chemical data on the hydrogen-bonding energy of a CA—H$\cdots$O=C interaction are available. However, values of −2.2 and −2.6 kJ mol^{-1}, respectively, have been reported for CH$_4\cdots$OH$_2$ and NH$_2$—CH$_3\cdots$OH$_2$ systems *in vacuo* (Van Mourik & Van Duijneveldt, 1995). This might lead to a situation in a protein structure where it is energetically more favourable to have the strong electrostatic interactions in a near-ideal geometry and the weaker interactions, such as C—H$\cdots$O, in a less ideal configuration. For small compounds, it is observed that these deviations from ideal geometry are more pronounced when the contacts are intramolecular (Taylor & Kennard, 1982). Although we tried to minimize this effect by excluding from the analysis any contacts between

Figure 7
Interatomic distance distribution for C$\cdots$C (dashed line), C$\cdots$O (thin line) and N$\cdots$O (dotted line) distances in haloalkane dehalogenase. The difference curve of C$\cdots$C minus C$\cdots$O distances is represented by a thick line. The number of contacts in a shell is given as the percentage of the total number of that type of contact.

residues which are close in sequence, it is obvious that the geometry of a C—H···O hydrogen bond in, for example, a β-sheet is influenced by neighbouring hydrogen bonds.

One of the most prominent contributions of the attractive C—H···O interactions to a protein structure is found in β-sheets. In a study on hydrogen bonding, McDonald & Thornton (1994) observed that four out of five buried main-chain carbonyl O atoms fail to form two hydrogen bonds and that these atoms were mainly found in β-sheets. If their classical definition of a hydrogen bond is applied in an analysis of the hydrogen-bonding pattern in DhlA, we find that the carbonyl O atoms in the β-sheets accept on average 1.2 hydrogen bonds (Fig. 9). However, if the C—H···O interactions are also taken into account, the saturation level increases and on average 1.7 hydrogen bonds are observed. Inspection of the 18 electron pairs which are not fully satisfied demonstrates that some of these form a classical hydrogen bond to a partially occupied atom or a residue which is too close in the sequence. In nine cases, a C—H group is found at a distance $d_H > 2.7$ Å or with an elevation angle $\theta > 45°$. It should be noted that the $d_H < 2.7$ Å criterion is disputable, as weak hydrogen bonds are long-range interactions which fall off with r^{-1} and might still contribute significantly as an attractive force beyond the sum of the van der Waals radii (Jeffrey & Sänger, 1991).

6. Conclusions

This study describes the refinement at atomic resolution of haloalkane dehalogenase, one of the largest structures for which AR data are currently available, as well as the analysis of selected stereochemical properties. It demonstrates that cryogenic data collection with high-intensity synchrotron radiation in combination with newly developed refinement strategies can be applied successfully to yield a reliable AR protein structure. The structure shows typical features of AR refinement, such as anisotropy of individual atoms, double and triple conformations for an appreciable number of amino acids and error assessment of all individual atomic parameters.

It is remarkable that the geometry of the final DhlA model shows some significant differences from what has been generally accepted as 'ideal geometry', even though these 'ideal' values were applied as restraints in the refinement of the structure. Although the number of observations in one structure is too small to draw firm conclusions about the validity of small-molecule values for protein structures or to propose better values for restraints in macromolecular refinement, the observed deviations in DhlA demonstrate the necessity of a reconsideration of the currently used refinement target libraries.

The distribution of main-chain and side-chain dihedral angles supports the conclusion by Wilson and co-workers
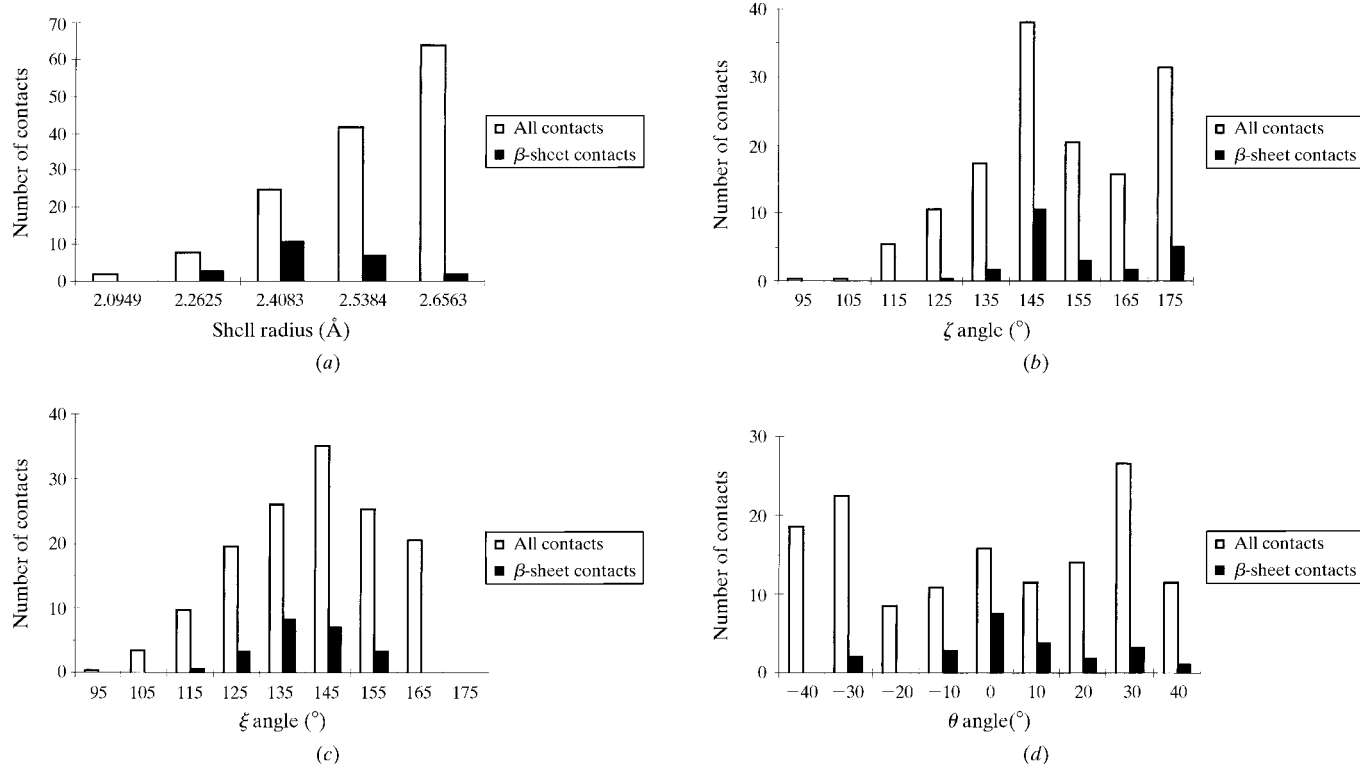
Figure 8
Stereochemical analysis of C—H···O hydrogen-bonding interactions. (a) hydrogen–oxygen distance d_H; (b) C—H···O angle ζ, applying the 'cone correction' (Kroon & Kanters, 1974), i.e. the histogram was corrected by multiplying each bar by $N/\sin\langle\varphi\rangle$, where $\langle\varphi\rangle$ is the average of the upper and lower limits of the bar and N is a normalization constant chosen such that the area under the corrected histogram is equal to that under the uncorrected histogram; (c) H···O—C angle ξ, applying the same 'cone correction'; (d) elevation angle θ; the number of possible C—H···O configurations with elevation angles between $\theta - d\theta$ and $\theta + d\theta$ is proportional to $\cos\theta$ (Taylor & Kennard, 1982).
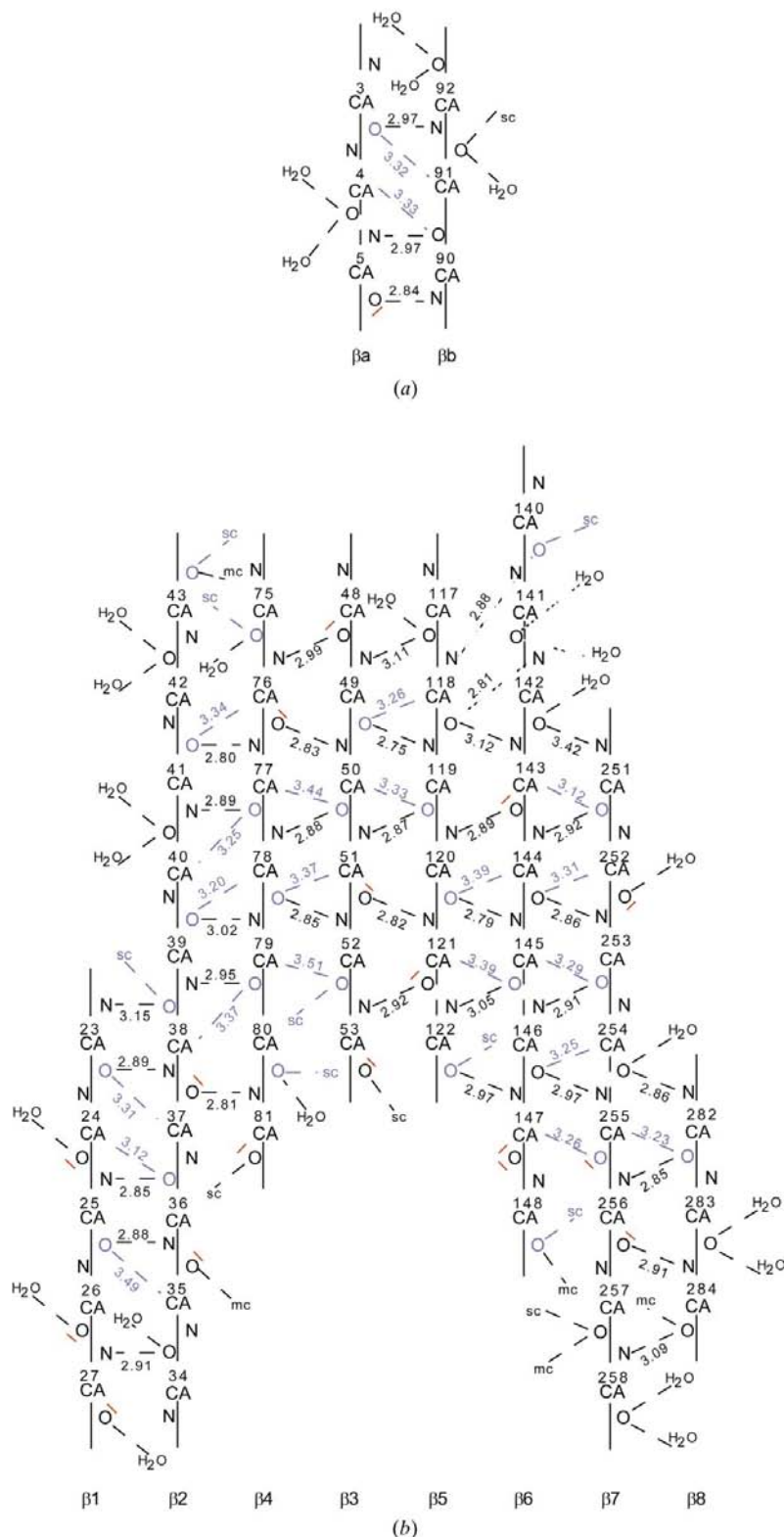
Figure 9
Schematic view of the hydrogen-bonding pattern in the β-sheets of haloalkane dehalogenase; classical N—H···O and O—H···O hydrogen bonds in black, C—H···O hydrogen bonds in blue, unsatisfied electron pairs in red; sc side-chain, mc main-chain hydrogen-bond donor; donor–acceptor distances are given in Å; (a) small two-stranded antiparallel β-sheet; (b) the main eight-stranded β-sheet. Secondary structure as identified by Kabsch & Sander (1983).

(Sevcik *et al.*, 1996; Walsh *et al.*, 1998) that the side-chain χ_1 torsion angles are more tightly clustered around their mean values than was previously observed. Furthermore, the main-chain ω torsion angle might display large deviations from its ideal value of 180°. Strong restraints on the planarity of the peptide bond are applied by default in several refinement packages. If there is any conformational strain in the protein, *e.g.* owing to the functionality of a residue, enforcing planarity induces compensation of this strain in the φ and ψ torsion angles. Asn148 in the haloalkane dehalogenase provides a nice example, as in previous structures the surrounding ω angles were flat, but the residue was outside the favoured regions in the Ramachandran plot. In the AR structure, where the planarity restraint is loosened, ω deviates from 180° and the φ and ψ torsion angles adopt more reasonable values.

A significant number of non-bonded C···O contacts shorter than the sum of the van der Waals radii of the two atoms have been observed in DhlA. This might seem remarkable, since an 'anti-bumping' restraint was used to avoid too short interatomic distances. Although the restraint was relaxed in the final refinement cycles, it might have introduced a bias in the analysis of the C···O contacts. However, in *SHELXL*, the distance at which the restraint is activated is not set at the sum of the van der Waals radii of the atoms, but the hydrogen-bonding capacity has been taken into account. The target minimum used for C···O contacts is 2.8 Å. Only distances shorter than this minimum are affected by the restraint; in total less than 0.5% of the contacts in the analysis. The high resolution of the model enabled the reliable inclusion of H atoms at calculated positions. A careful analysis of the short C···O contacts revealed the presence of C—H···O interactions with all the geometrical characteristics of a hydrogen bond. This concept might be unfamiliar among structural biologists, but it is in accordance with a more general study by Derewenda *et al.* (1995). C—H···O hydrogen bonds are of special interest in β-sheet structures, as they account for 25% of the possible hydrogen bonds to the carbonyl O atoms of residues in the β-sheet of DhlA. These data indicate that C···O contacts should not be treated unthoughtfully as repulsive during crystallographic refinement. Moreover, they constitute a basis for a more complete description of the attractive forces which hold together the β-sheet secondary-structure element.

The rapidly increasing number of AR structures will facilitate a more accurate parametrization of geometrical and stereochemical libraries, from which medium- and low-resolution structure

determinations will undoubtedly benefit most. At the same time, these AR structures also demonstrate the necessity for reconsideration of those libraries. Until this has been implemented, restraints on dihedral angles and van der Waals contacts have to be applied with special prudence.

The atomic resolution has enabled the remodelling of one loop and both terminal residues which were previously ill-defined in haloalkane dehalogenase. The position of two chloride ions in the DhlA structure is determined with high accuracy; one is found in the active site equidistant from Trp125 and Trp175 NE1 atoms. However, from the distance to the calculated H atoms it can be inferred that the ion binds more strongly to Trp175. A second ion is observed with partial occupancy at the surface of the DhlA molecule, indicating the location of the kinetically detected secondary halide-binding site (Schanstra & Janssen, 1996). Clear anisotropy in the electron density allows two possible orientations of Asp124, indicating that at pH 5 the neighbouring His289 side chain is only partially protonated at the NE2 atom. If crystal structures are determined at a resolution where anisotropy and alternative conformations cannot be modelled, any conclusion from them about the preferred state of an equilibrium has to be drawn with great care.

References

Abola, E. E., Sussman, J. L., Prilusky, J. & Manning, N. O. (1997). *Methods Enzymol.* **277**, 556–571.

Allen, F. H. S., Bellard, S., Brice, M. D., Cartwright, B. A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B. G., Kennard, O., Motherwell, W. D. S., Rodgers, J. R. & Watson, D. G. (1979). *Acta Cryst.* B**35**, 2331–2339.

Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*, edited by B. Horecker, N. O. Kaplan, J. Marmur & H. A. Scheraga. New York: Academic Press.

Brünger, A. T. (1992*a*). *Nature (London)*, **355**, 472–475.

Brünger, A. T. (1992*b*). *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR*. New Haven, Connecticut: Yale University Press.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1997). *Curr. Opin. Struct. Biol.* **7**, 681–688.

Derewenda, Z. S., Lee, L. & Derewenda, U. (1995). *J. Mol. Biol.* **252**, 248–262.

Desiraju, G. R. (1991). *Acc. Chem. Res.* **24**, 290–296.

Desiraju, G. R. (1996). *Acc. Chem. Res.* **29**, 441–449.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 133–138.

Fabiola, G. F., Krishnaswamy, S., Nagarajan, V. & Pattabhi, V. (1997). *Acta Cryst.* D**53**, 316–320.

Franken, S. M., Rozeboom, H. J., Kalk, K. H. & Dijkstra, B. W. (1991). *EMBO J.* **10**, 1297–1302.

Gliubich, F., Gazerro, M., Zanotti, G., Delbono, S., Bombieri, G. & Berni, R. (1996). *J. Biol. Chem.* **271**, 21054–21061.

Hirshfeld, F. L. & Rabinovich, D. (1973). *Acta Cryst.* A**29**, 510–513.

Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. (1996). *Nature (London)*, **381**, 272.

Janssen, D. B., Scheper, A., Dijkhuizen, L. & Witholt, B. (1985). *Appl. Environm. Microbiol.* **49**, 673–677.

Janssen, D. B., Van den Wijngaard, A. J., Van der Waarde, J. J. & Oldenhuis, R. (1991). *On-site Bioreclamation*, pp. 92–122. Boston: Butterworth–Heinemann.

Jeffrey, G. A. & Sänger, W. (1991). *Hydrogen Bonding in Biological Structures.* Berlin: Heidelberg; New York: Springer–Verlag.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Keuning, S., Janssen, D. B. & Witholt, B. (1985). *J. Bacteriol.* **163**, 635–639.

Kleywegt, G. J. & Jones, T. A. (1994). *Acta Cryst.* D**50**, 178–185.

Kroon, J. & Kanters, J. A. (1974). *Nature (London)*, **248**, 667–669.

Krooshof, G. H., Ridder, I. S., Tepper, A. W. J. W., Vos, G. J., Rozeboom, H. J., Kalk, K. H., Dijkstra, B. W. & Janssen, D. B. (1998). *Biochemistry*, **37**, 15013–15023.

Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.

Lamzin, V. S. & Wilson, K. S. (1997). *Methods Enzymol.* **277**, 269–305.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Leonard, G. A., McAuley-Hecht, K., Brown, T. & Hunter, W. N. (1995). *Acta Cryst.* D**51**, 136–139.

Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.

MacArthur, M. W. & Thornton, J. M. (1996). *J. Mol. Biol.* **264**, 1180–1195.

McDonald, I. K. & Thornton, J. M. (1994). *J. Mol. Biol.* **238**, 777–793.

Moews, P. C. & Kretsinger, R. H. (1975). *J. Mol. Biol.* **91**, 201–228.

Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). *Proteins Struct. Funct. Genet.* **12**, 345–364.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Nardini, M. (1998). Personal communication.

Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., Sussman, J. L., Verschueren, K. H. G. & Goldman, A. (1992). *Protein Eng.* **5**, 197–211.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pikkemaat, M. G., Ridder, I. S., Rozeboom, H. J., Kalk, K. H., Dijkstra, B. W. & Janssen, D. B. (1999). In preparation.

Pries, F., Kingma, J., Pentenga, M., Van Pouderoyen, G., Jeronimus Stratingh, C. M., Bruins, A. P. & Janssen, D. B. (1994). *Biochemistry*, **33**, 1242–1247.

Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Prot. Chem.* **23**, 283–437.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Rozeboom, H. J., Kingma, J., Janssen, D. B. & Dijkstra, B. W. (1988). *J. Mol. Biol.* **200**, 611–612.

Schanstra, J. P. & Janssen, D. B. (1996). *Biochemistry*, **35**, 5624–5632.

Schanstra, J. P., Kingma, J. & Janssen, D. B. (1996). *J. Biol. Chem.* **271**, 14747–14753.

Schanstra, J. P., Ridder, I. S., Heimeriks, G. J., Rink, R., Poelarends, G. J., Kalk, K. H., Dijkstra, B. W. & Janssen, D. B. (1996). *Biochemistry*, **35**, 13186–13195.

Sevcik, J., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1996). *Acta Cryst.* D**52**, 327–344.

Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.

Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.

Steiner, T. (1997). *J. Chem. Soc. Chem. Commun.*, pp. 727–734.

Steiner, T. (1998). *Acta Cryst.* B**54**, 456–463.

Stucki, G. & Thüer, M. (1994). *Appl. Microbiol. Biotechnol.* **42**, 167–172.

Stucki, G. & Thüer, M. (1995). *Environ. Sci. Technol.* **29**, 2339–2345.

Taylor, R. & Kennard, O. (1982). *J. Am. Chem. Soc.* **104**, 5063–5070.

Tronrud, D. E. (1997). *Methods Enzymol.* **277**, 306–319.

Van Mourik, T. & Van Duijneveldt, F. B. (1995). *J. Mol. Struct. (THEOCHEM)*, **341**, 63–73.

Verschueren, K. H. G., Franken, S. M., Rozeboom, H. J., Kalk, K. H. & Dijkstra, B. W. (1993). *J. Mol. Biol.* **232**, 856–872.

Verschueren, K. H. G., Kingma, J., Rozeboom, H. J., Kalk, K. H., Janssen, D. B. & Dijkstra, B. W. (1993). *Biochemistry*, **32**, 9031–9037.

Verschueren, K. H. G., Seljée, F., Rozeboom, H. J., Kalk, K. H. & Dijkstra, B. W. (1993). *Nature (London)*, **363**, 693–698.

Walsh, M. A., Schneider, T. R., Sieker, L. C., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1998). *Acta Cryst.* D**54**, 522–546.

Watkin, A. (1994). *Acta Cryst.* A**50**, 411–437.

Wilson, A. J. C. (1942). *Nature (London)*, **150**, 152.

Wilson, A. J. C. (1973). *Acta Cryst.* B**29**, 1488–1490.

Wilson, A. J. C. (1976). *Acta Cryst.* A**32**, 781–783.

Wilson, K. S., Dauter, Z., Lamzin, V. S., Walsh, M., Wodak, S., Richelle, J., Pontius, J., Vaguine, A., Hooft, R. W. W., Sander, C., Vriend, G., Thornton, J. M., Laskowski, R. A., MacArthur, M. W., Dodson, E. J., Murshudov, G., Oldfield, T. J., Kaptein, R. & Rullmann, J. A. C. (1998). *J. Mol. Biol.* **276**, 417–436.