Application of maximum-entropy maps in the accurate refinement of a putative acylphosphatase using 1.3 Å X-ray diffraction data

Eiji Nishibori,[a]* **Takahiro Nakamura,**[a] **Masanori Arimoto,**[a] **Shinobu Aoyagi,**[a] **Hideo Ago,**[b] **Masashi Miyano,**[b] **Toshikazu Ebisuzaki**[c] **and Makoto Sakata**[a]

[a]Deparment of Applied Physics, Nagoya University, Nagoya 464-8603, Japan, [b]Structural Biophysics Laboratory, RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan, and [c]Computational Astrophysics Laboratory, RIKEN Discovery Research Institute, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

Correspondence e-mail:
eiji@mcr.nuap.nagoya-u.ac.jp

Accurate structural refinement of a putative acylphosphatase using 1.3 Å X-ray diffraction data was carried out using charge densities determined by the maximum-entropy method (MEM). The MEM charge density clearly revealed detailed features of the solvent region of the putative acylphosphatase crystalline structure, some of which had never been observed in conventional Fourier maps. The structural model in the solvent region was constructed as distributions of anisotropic water atoms. The omit-difference MEM maps and the difference MEM maps were effective in revealing details of the protein structure, such as multiple conformations of the side chains of amino-acid residues, anisotropy of atoms and H atoms. By model building using the MEM charge densities, the reliability factors R_1 and R_{free} in the SHELX refinement were dramatically improved from 17.9% and 18.3% to 9.6% and 10.0%, respectively. The present results prove the usefulness of MEM in improving the accuracy of refinement of protein crystal structures.

1. Introduction

Atomic resolution and ultrahigh-resolution protein structures obtained using X-ray crystallography have become a powerful tool in the investigation of the relationships between the structure and function of biological macromolecules. Atomic resolution and ultrahigh-resolution structures allow more accurate views of protein structure such as the structure of hydrated water molecules and ligand molecules, the statistical or thermal disorder of amino acids and even the existence of hydrogen bonds. Charge-density maps from atomic resolution and ultrahigh-resolution data should make it possible both to make model alternative conformations and to obtain precise directions of the side chain for amino-acid and domain movements.

Diffraction data extending to a resolution of higher than 1.2 Å are considered atomic resolution (Sheldrick, 1990) and data to 0.8 Å are considered as ultrahigh (subatomic) resolution. Recent progress in synchrotron X-ray diffraction techniques, including third-generation synchrotron X-ray sources, advanced detector systems and cryocrystallography, have provided us with the ability to measure atomic resolution and ultrahigh-resolution data. More than 600 protein structures at atomic and ultrahigh resolution have been deposited in the Protein Data Bank (PDB; Berman et al., 2000) to date. Among these are approximately 50 high-resolution structures with crystallographic R values below 10%. These structures

doi:10.1107/S0907444907065663 **237**

have been determined by precise refinement including aniso-tropic displacement parameters and H atoms. Some of them were analyzed using more advanced models such as the dummy-bond electron approach (DBE; Afonine *et al.*, 2004) and the multipolar model (Jelsch *et al.*, 2000). These structures have provided crucial information in understanding the function of biological macromolecules (Howard *et al.*, 2004; Liu *et al.*, 2002; Schmidt *et al.*, 2003).

The number of near-atomic resolution data sets, which are defined as those with resolutions between 1.2 and 1.4 Å, has increased even more dramatically. The number of deposited structures in this resolution range is more than 1000, which is two times more than the total number of atomic resolution and ultrahigh-resolution data sets. In contrast to the cases of atomic resolution and ultrahigh-resolution data, there are almost no very accurate refinements with R values less than 10% in this resolution range. Only two such data sets (PDB codes 1hq2 and 1q0n; Blaszczyk *et al.*, 2000, 2003) with an R value of less than 10% have been deposited in the PDB. Their structures are very similar. For near-atomic resolution data, structural refinement using anisotropic displacement para-meters and H atoms is extremely difficult. Charge-density maps obtained using the Fourier method are unavoidably distorted in a sense owing to the truncation effect, which prevents accurate structural information including anisotropic thermal vibrations and the positions of H atoms being obtained. In principle, refinement using anisotropic displace-ment parameters and H atoms can be performed using the *SHELXL* program (Sheldrick, 2008). However, this does not mean that one can reach a physically meaningful solution by introducing anisotropic parameters because the *SHELXL* program cannot tell what kind of anisotropy is needed for a certain atom. The refinements often fall into local minima such as physically meaningless negative anisotropy of atoms. In addition, X-ray data represent a time- and space-averaged electron density. The lower resolution data are often asso-ciated with disordered structures. It is very reasonable to consider that near-atomic resolution data do actually include detailed structural information such as atom anisotropy and H atoms. In order to extract such detailed information from the near-atomic resolution data, analytical methods require much better experimental charge-density distribution maps without truncation effects.

The maximum-entropy method (MEM) for the crystallo-graphic Fourier inversion problem is a powerful tool for charge-density determination from a limited number of structure factors. High-resolution charge density without a truncation effect can be obtained by the MEM. The MEM has been successfully applied to a wide variety of materials such as inorganic materials (Nishibori *et al.*, 2007), fullerene-related compounds (Nishibori *et al.*, 2006) and metal–organic complexes (Kitaura *et al.*, 2002; Kubota *et al.*, 2005) and even from a limited number of powder diffraction data owing to peak overlaps. The MEM charge densities are able to reveal detailed structural features such as bond character, disordered atoms and H atoms, especially in lower density regions. For example, the MEM charge densities of silicon and diamond

obtained from SPring-8 powder data revealed the covalent bonding electrons very clearly. The deviations of charge densities at the bond mid-point from several theoretical calculations were within 0.05 e Å^{-3} (Nishibori *et al.*, 2007). In the case of high-performance thermoelectrics, the disorder of the Zn atoms, which have less than 10% occupancy, in Zn_4Sb_3 was determined from the MEM charge density (Snyder *et al.*, 2004). In addition, the charge densities of H atoms were clearly revealed in MgH_2 (Noritake *et al.*, 2002). The MEM analysis of protein crystals requires huge computational resources. In 2002, Tanaka and coworkers developed the MEM program package system *ENIGMA*, which is applicable to huge systems such as proteins (Tanaka *et al.*, 2002). The MEM analysis for limited resolution protein data has the potential to determine high-resolution structures even from near-atomic resolution data of protein crystals.

In this study, we carried out the high-resolution structural refinement of a protein molecule based on the MEM charge density. The 1.3 Å near-atomic resolution X-ray diffraction data of a putative acylphosphatase was used for the evaluation and development of the method.

2. Experimental and analytical methods and procedures

2.1. Protein expression, purification, crystallization and synchrotron-radiation X-ray experiments

Escherichia coli BL21(DE3) was transformed by the plasmid pET11a (Novagen) carrying the cDNA for TT0497 (UniProtKB/TrEMBL ID Q5SKS6_THET8; Boeckmann *et al.*, 2003) from *Thermus thermophilus* HB8 (Masui *et al.*, unpublished results). For the overexpression of TT0497, the transformed *E. coli* was cultured at 310 K for 20 h in LB medium and harvested. The harvested *E. coli* was disrupted by sonication in lysis buffer (20 m*M* Tris–HCl pH 8.0, 0.5 *M* NaCl, 5 m*M* 2-mercaptoethanol). The supernatant of the cell lysate was heated for 11.5 min in a water bath maintained at 343 K and then centrifuged at 277 K to remove insoluble constituents. The supernatant was applied onto four columns in series: Super Q Toyopearl 650M (Tosoh) equilibrated with 20 m*M* Tris–HCl pH 8.0, Resource S (GE Healthcare) equi-librated with 20 m*M* MES–NaOH pH 6.0, CHT2-I (Bio-Rad) equilibrated with 10 m*M* sodium phosphate buffer pH 7.0 and HiLoad 16/60 Superdex 75 (GE Healthcare) equilibrated with 20 m*M* Tris–HCl pH 8.0, 0.2 *M* NaCl. The TT0497 bound to the ion-exchange columns and to the CHT2-I column was eluted with a sodium chloride gradient to 0.3 *M* and a sodium phosphate gradient from 0.01 to 0.15 *M*, respectively. The purity of the final sample was confirmed by the presence of a single band on SDS–PAGE stained with Coomassie Brilliant Blue. An N-terminal sequence analysis of the purified TT0497 showed that the first methionine in the amino-acid sequence coded by the cDNA was removed in the purified protein.

The TERA crystallization robot, which performs crystal-lization setup and observation of crystal growth automatically (Sugahara & Miyano, 2002), was used for the initial screening and subsequent refinement of crystallization conditions for

Table 1
Statistics of diffraction data and conventional refinement.

(a) Statistics of diffraction data. Values in parentheses are for the outer shell.

Beamline	BL45XU, SPring-8
Space group	$P2_12_12_1$
Unit-cell parameters	$a = 29.9, b = 45.7, c = 50.0$
Resolution (Å)	33.7–1.30
Wavelength (Å)	1.0000
No. of measured/unique reflections	57078/15934
Completeness (%)	90.5 (89.4)
$I/\sigma(I)$, outer shell	2.7
R_{merge} (%)	11.3 (21.4)

(b) Statistics of the conventional structural refinement and the MEM-based refinement. Values in parentheses are for the outer shell.

	Conventional	MEM-based
Resolution (Å)	33.7–1.30	
R	0.190 (0.241)	0.0956
R_{free}	0.216 (0.231)	0.0999
Protein atoms	677	677
No. of waters	151	203
Chloride ions	—	1
H atoms	—	743
R.m.s.d. in distances (Å)	0.027	0.018
R.m.s.d. in angles (°)	1.858	0.036

TT0497. Purified TT0497 was concentrated to 2.9 mg ml^{-1} for crystallization. A crystal suitable for X-ray crystallographic analysis was grown at 291 K during concentration of the mother liquor: 0.5 µl TT0497 solution (2.9 mg ml^{-1} in 20 mM Tris–HCl pH 8.0, 0.2 M NaCl) and 0.5 µl solution containing 27.5%(w/v) polyethylene glycol 4000 and 0.1 M MES pH 6.9 were mixed under a sealing oil composed of silicone and paraffin oils. The crystal was treated with Paratone-N (Hampton Research) in order to avoid ice formation at cryogenic temperatures during the X-ray diffraction experiment.

Diffraction-intensity data were measured to 1.30 Å at 100 K using an R-AXIS V (Rigaku) detector on BL45XU at SPring-8 (Yamamoto et al., 1998). The diffraction images collected were processed using CrystalClear (Rigaku). The R_{merge} values for the overall data (33.7–1.30 Å) and the data in the highest resolution shell (1.38–1.30 Å) were 0.113 and 0.214, respectively (Table 1).

2.2. Structure determination and conventional structural refinement

The first structure analysis of TT0497 was performed by the molecular-replacement method with AMoRe (Navaza, 1994) using the coordinates of acylphosphatase from bovine testis (PDB code 2acy; Thunnissen et al., 1997) as the search model. The amino acids of the resultant model from the molecular-replacement method were replaced with those of TT0497 to produce the initial model for subsequent conventional structural refinement using O (Jones et al., 1991), CNS (Brünger et al., 1998) and REFMAC5 (Winn et al., 2001; Collaborative Computational Project, Number 4, 1994). The structural refinement was performed using diffraction-intensity data from 33.7 to 1.3 Å resolution. The refined model contains 677

protein atoms and 151 water molecules and its R and R_{free} values were 0.19 and 0.24, respectively (Table 1). Anisotropic displacement parameters and H atoms were not introduced in the refinement. There were ten zero-occupancy sites for the residues Glu70, Glu73, Glu78 and Lys83. The disorder of residues, including side chains, was not considered during this process. The geometrical quality of the refined model was confirmed by Ramachandran analysis (Laskowski et al., 1993), which showed that all amino-acid residues were in favoured regions of the plot. This structure was deposited in the PDB (PDB code 1ulr).

2.3. The MEM-assisted structural refinement

The structure of acylphosphatase deposited in the PDB described above was used as a starting model for MEM-based refinement. In this procedure, the structure was refined using the SHELXL program. Initially, bulk-solvent correction using the SWAT switch of SHELXL was applied to the refinement. The R_1 and R_{free} values after scale-factor refinement were 0.1785 and 0.1833 for the starting model, respectively. The phases of the structure factors were calculated from the model mentioned above and observed structure factors with these phases were used in the MEM analysis. We used all of the structure factors in the MEM analysis because the absence of diffraction data sometimes affects the MEM charge density (Takata & Sakata, 1996). The MEM analysis was carried out using the computer program ENIGMA (Tanaka et al., 2002). The unit cell was divided into $128 \times 256 \times 256$ pixels. The program Coot was used to inspect the charge densities
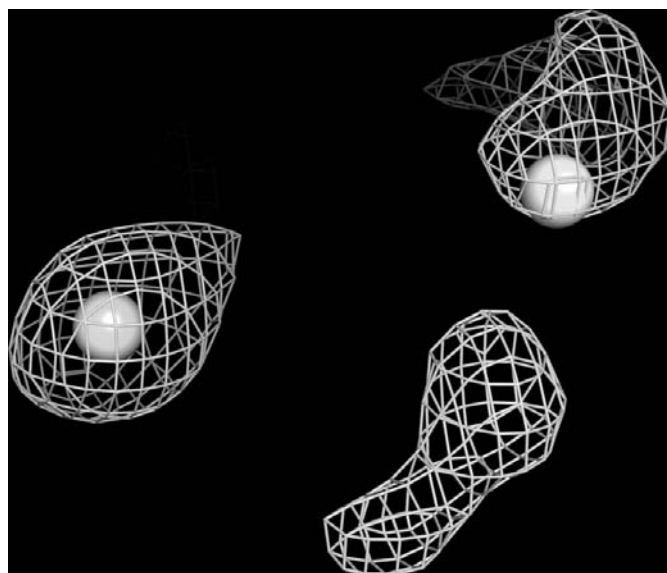
Figure 1
The MEM charge densities of the solvent region for the putative acylphosphatase based on the structure model without water molecules. The MEM map is shown as a grey mesh. The level of charge density is 0.75 e Å^{-3}. The water molecule assumed to be present in the PDB-deposited structure is shown as a white sphere. The charge-density peaks which are not assumed to be water molecules in the PDB structure are clearly recognized in the MEM charge density.
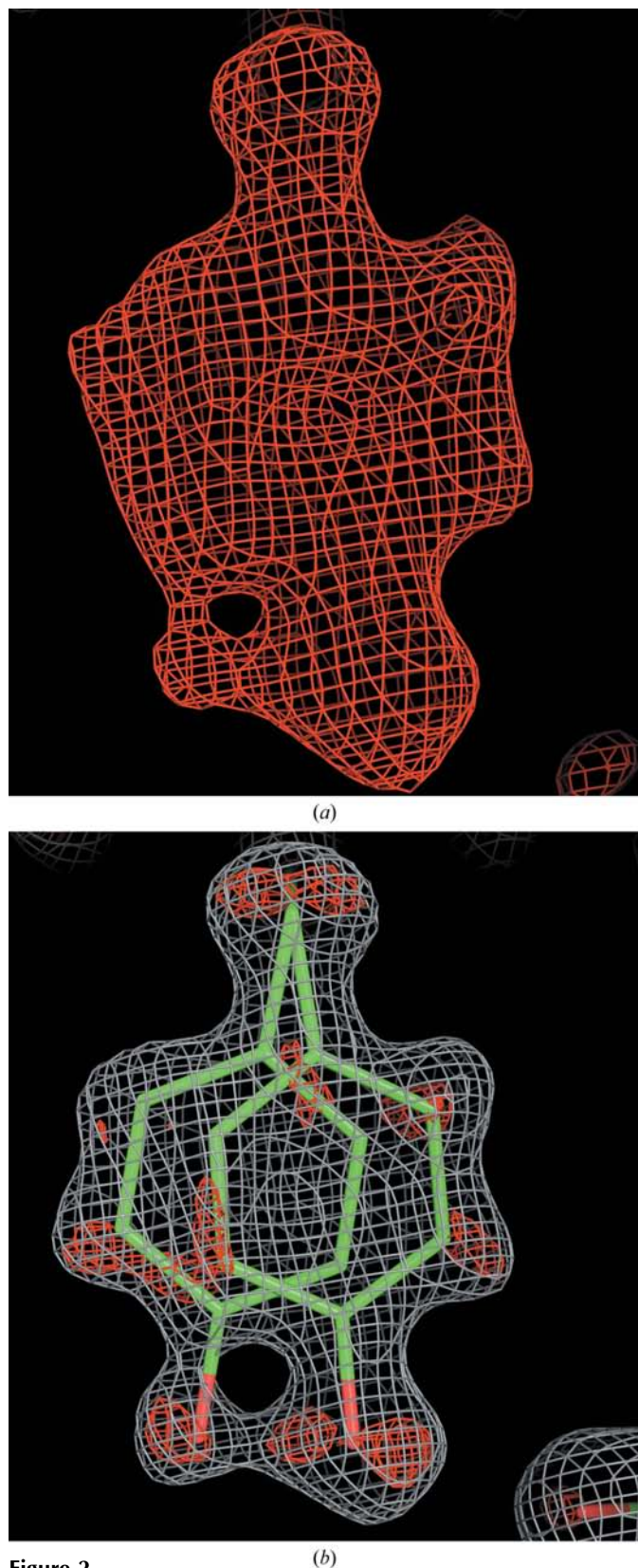
Figure 2
(a) The omit-difference MEM charge density of Tyr33. The charge-density level is 0.44 e Å^{-3}. (b) The multiple-conformer model of Tyr33 with the MEM charge density. The MEM charge density is shown as a grey mesh. The difference MEM charge density is shown as a red mesh. The charge-density levels are 0.70 e Å^{-3} for the MEM map and 0.40 e Å^{-3} for the difference map.

(Emsley & Cowtan, 2004). Figures were prepared using *PyMOL* (DeLano Scientific LLC).

The main differences between the MEM charge density and the starting model at this stage were found in the solvent regions. It was found that there were many undefined water peaks in the MEM charge density. We reconstructed the structure model for the solvent region based on the MEM charge density. We removed all the water molecules from the PDB-deposited structure. The MEM charge density was calculated based on this model. The MEM charge density in the solvent regions in this model is shown in Fig. 1. The level of the isocharge-density surface was 0.75 e Å^{-3}, which corresponds to 1.2σ for the MEM map. The figure shows that unidentified peaks were clearly recognized in the MEM map. Next, we reconstructed the structure model for solvent water molecules based on the MEM charge density. By using the revised model for solvent water molecules, we refined the structure model again. The parameters for bulk-solvent correction were decreased to approximately zero in the *SHELXL* refinement during this process. We did not use bulk-solvent correction after this stage. In the final stage of this process, we assumed the presence of partially occupied water molecules at the weak charge-density peaks. In such cases, only the positional and thermal parameters of the partially occupied waters were refined in the *SHELX* refinement. The R_1 and R_{free} values were improved to 0.149 and 0.166, respectively. The total number of water molecules in the asymmetric unit was increased to 204, including six disordered molecules, compared with 151 in the deposited model. The occupancies of the water molecules were 1.0, except for six water molecules that had occupancies of less than 1.0.

Omit-difference MEM maps (Nishibori *et al.*, 2006) were used in the model rebuilding of the protein molecule in order to reduce model-bias effects. To perform the omit-difference MEM map analysis, a series of phases were calculated from the structure model by removing three successive residues at a time. Three residues correspond to 3.45% of the atoms of the protein molecule on average. The corresponding F_c maps were also calculated by the MEM. A total number of 29 omit-difference maps were produced in this process. Although they were omitted in the calculation of structure factors, ten atoms that were undetected in the deposited model were found in these 29 omit-difference MEM maps. We then added these ten atoms to the structure models. In addition, disorder of residues was recognized in the charge density. Such processes were repeated until no new atoms were found in the 29 omit-difference maps. Examples of omit-difference MEM maps are shown in Fig. 2. The model-building process of multiple conformers of the side chain for Tyr33 can be understood from Fig. 2. Fig. 2(a) is an omit-difference MEM map which clearly shows the disorder of the phenyl ring. There are no difficulties in making a disordered model as shown in Fig. 2(b).

After completing the above process, approximately 77% of the charge-density peaks for H atoms were recognized in the difference MEM charge densities. The HFIX command in the *SHELXL* program was used to add H atoms to the structure model. The H atoms were refined as riding on connected

atoms. The thermal displacement parameters of H atoms were fixed to 1.5 times that of the connected atom for methyl and hydroxy groups and 1.2 times that of the connected atom for other H atoms. The positional and thermal parameters of the protein molecule were refined. The R_1 and R_{free} values in the refinement changed from 0.149 and 0.166 to 0.142 and 0.168, respectively, by the above model building.

There were both positive and negative differences around the solvent water molecules in the difference MEM charge density at this stage. In order to reduce the differences, we used anisotropic displacement parameters for water molecules in the refinement. The default value of the ISOR restraint in the *SHELXL* program was used in the refinement. It was impossible to determine reasonable anisotropic thermal parameters in the positional and thermal parameter refinement of whole water molecules. The determined parameters were not consistent with the difference MEM charge density. We therefore estimated initial anisotropic thermal parameters from the difference MEM charge density. These parameters were used in the refinement. We gradually introduced the anisotropy of water molecules such that the number of anisotropically modelled molecules at a particular time was typically ten. We repeated the parameter estimation from the difference MEM charge density and refinement in order to minimize the difference MEM charge density. The R_1 and R_{free} values decreased from 0.142 and 0.168 to 0.116 and 0.127, respectively, during the refinement and the charge-density differences were decreased in the difference MEM charge density. Similar charge-density differences were also found at atoms in the protein molecule such as the peptide carbonyl O atom, as shown in Fig. 3. The anisotropic displacement parameters were also used for atoms in the protein molecule. The value of the ISOR restraint for the atoms in protein was set to
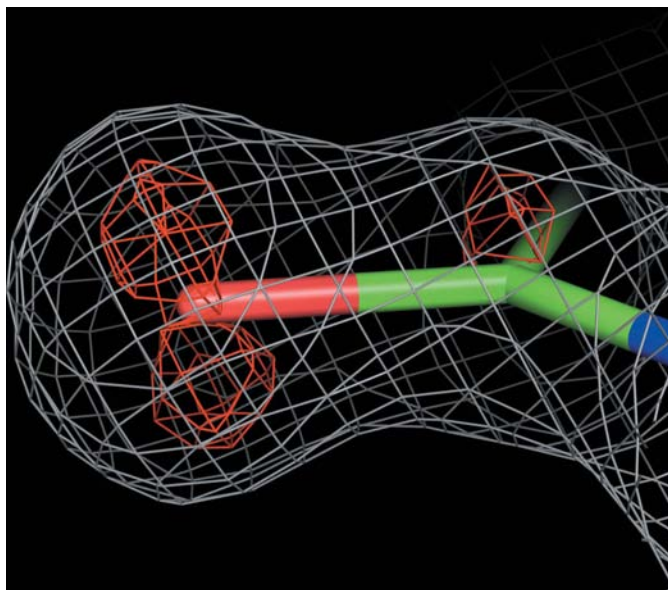
0.02 and the SIMU restraint was also used in the refinement. The σ values in the bonding distance and angles restraints were set to 1.3 times larger than the default values. There were many inconsistencies between the refined structure and the difference MEM charge density in the case of full parameter refinement for atoms in the protein. Thus, a similar procedure to the anisotropic refinement of water molecules, using initial parameter estimation and gradual multi-step refinement, was carried out. The R_1 and R_{free} values decreased from 0.1160 and 0.127 to 0.0979 and 0.1047, respectively, and the charge-density differences also decreased in the protein molecules.

We carefully investigated the whole of the MEM and difference MEM charge density. Almost all the values of the difference charge densities were smaller than $1.0 \, e \, Å^{-3}$, corresponding to 1.3σ in the MEM charge density. An exception was found at one position, which showed a large difference charge density of $+3.10 \, e \, Å^{-3}$ between the water molecules. This indicated that this peak must be arise from another ion than oxygen. HCl and NaCl were used in the crystallization buffer, suggesting that the peak may be interpreted as either an Na^+ or Cl^- ion. The peak was surrounded by five protein N atoms, which would favour a negative ion. Considering that there was a disordered water molecule with 40% occupancy located 1.99 Å from the peak position, it was assumed that a Cl^- ion with 60% occupancy was the origin of

Figure 3
The MEM and difference MEM charge densities of the carbonyl O atom of Lys24. The MEM and difference-MEM charge densities are shown as a grey and a red mesh, respectively. The charge-density levels are $1.00 \, e \, Å^{-3}$ for the MEM map and $0.40 \, e \, Å^{-3}$ for the difference map.
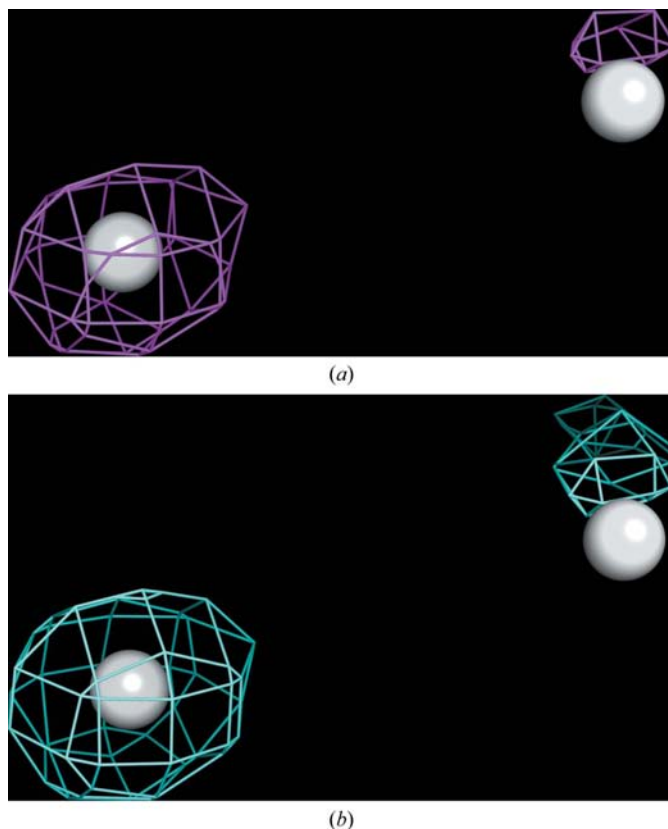

Figure 4
(a) The $F_o - F_c$ and (b) the maximum-likelihood weighted difference maps of the solvent region for the putative acylphosphatase based on the structure model without water molecules. The charge-density level is $0.75 \, e \, Å^{-3}$, the same as in Fig. 1.

the major peak. A fully occupied Cl^- ion located at the peak position would be too close to the disordered water molecule. In this model, the nearest water molecule and Cl^- ion cannot coexist in the same unit cell. If there is no coexistence of the water molecule and Cl^- ion in the same unit cell, the nearest interatomic distance between the Cl^- ion and the water molecule becomes 3.05 Å, which is consistent with other atomic resolution structures such as phosphoserine amino-transferase (PDB code 1w23; Dubnovitsky *et al.*, 2005) and C-α-formylglycine-generating enzyme (PDB code 1z70; Roeser *et al.*, 2005). The value of the difference charge density decreased to -1.54 e Å^{-3} on the addition of a Cl^- ion with 60% occupancy. In the final refinement, we refined the parameters of the Cl^- ion and the disordered water molecule. The R_1 and R_{free} values in the refinement became 0.0956 and 0.0999, respectively.

2.4. Structural refinement based on the Fourier method

In order to confirm the usefulness of the above MEM-assisted refinement, we carried out a parallel refinement using conventional Fourier maps. The computer program *FFT* from the *CCP*4 suite was used to calculate difference Fourier ($F_o - F_c$) and $2F_o - F_c$ maps. The *SHELX* program was used in these procedures. The acylphosphatase structure deposited in the PDB was used as a starting model. We removed all the water molecules from the PDB-deposited structure and calculated $F_o - F_c$ and $2F_o - F_c$ maps using this model. The $F_o - F_c$ map based on the starting model is shown in Fig. 4(*a*) and can be compared with the MEM charge density based on the same model as shown in Fig. 1. The maximum-likelihood weighted difference map calculated using *REFMAC* is also shown in Fig. 4(*b*). These figures show that one of three charge-density peaks in the MEM charge density cannot be recognized in either the $F_o - F_c$ map or the maximum-likelihood weighted difference map. It is also seen that the charge-density distribution of the $F_o - F_c$ map is similar to that of the maximum-likelihood weighted difference map in this case. The features of the $2F_o - F_c$ map are also similar to those of the $F_o - F_c$ map. We performed model reconstruction of the water structure in the solvent region based on the $F_o - F_c$ and $2F_o - F_c$ maps. The R_1 and R_{free} values were improved to 0.158 and 0.163, respectively. The total number of water molecules in the asymmetric unit increased to 178 from 151 in the PDB-deposited model. Approximately 30 water molecules found in the MEM-assisted refinement were not found in this process.

Omit $F_o - F_c$ maps were used for model rebuilding of the protein molecule. A series of phases were calculated in an identical way as for the MEM-assisted refinement. A total number of 29 omit-difference Fourier maps were produced in this process. Five disordered residues were recognized in the omit $F_o - F_c$ map. The other four disordered residues found in the MEM-assisted refinement were not found in this process. An example of the differences between the omit $F_o - F_c$ and omit-difference MEM maps is shown in Fig. 5. The level of the isocharge-density surface in these figures was 0.75 e Å^{-3},

which corresponds to 1.2σ for the MEM map and to 2.0σ for the Fourier map. The figure clearly shows that undetected peaks in the Fourier map were clearly recognized in the MEM map. Multiple conformers of the side chain of Glu51 can be clearly recognized from the omit-difference MEM map
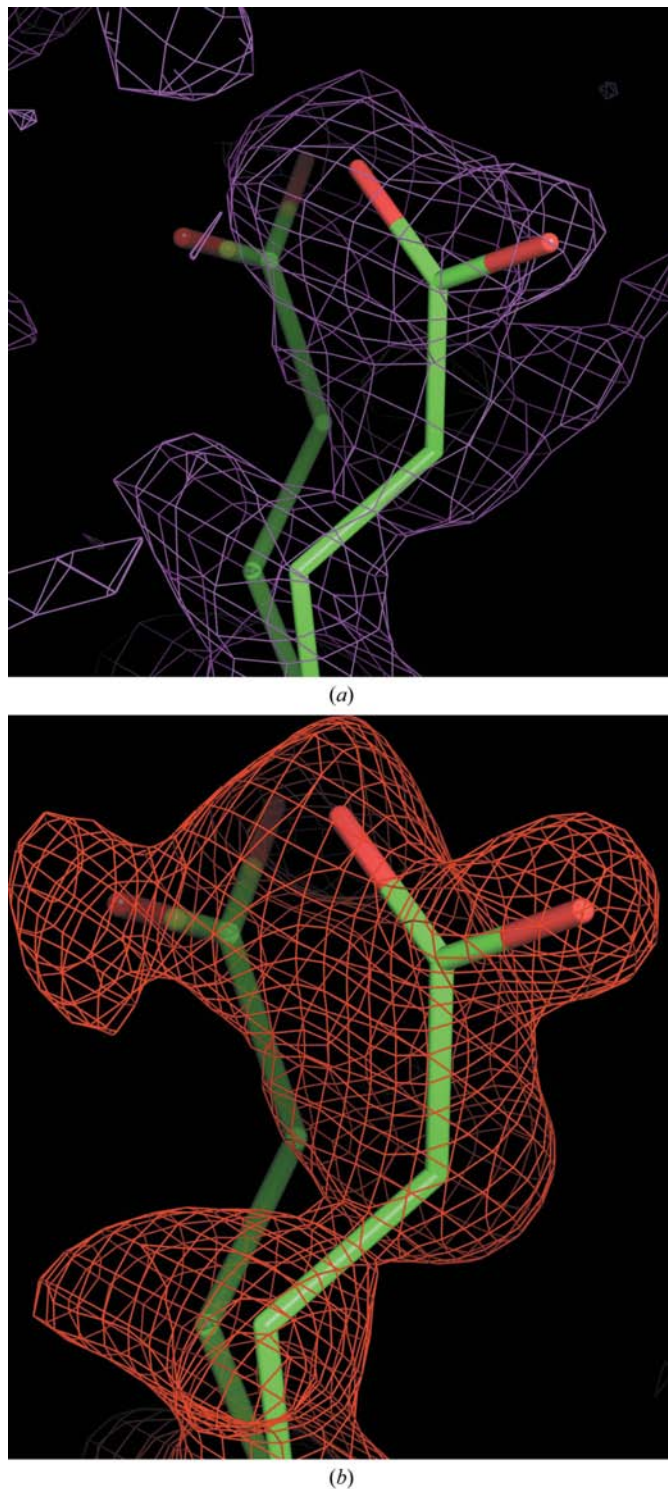

(*a*)

(*b*)

Figure 5
(*a*) The omit $F_o - F_c$ map for Glu51. The level of charge density is 0.28 e Å^{-3}. (*b*) The omit-difference MEM map for Glu51. The charge-density level is 0.35 e Å^{-3}. The multiple-conformer model for Glu51 with the MEM charge density.

(Fig 5b). On the other hand, it is not possible to find charge-density peaks that indicate multiple conformers in the omit $F_o - F_c$ map shown in Fig. 5(a). We performed model rebuilding for the five disordered residues found by the omit-difference Fourier map.

The HFIX command in the *SHELXL* program was used to add H atoms to the structure model, although only 20.9% of the charge-density peaks for H atoms were recognized in the $F_o - F_c$ and $2F_o - F_c$ maps. The H atoms were refined in an identical manner as for the MEM-assisted refinement. The R_1 and R_{free} values in the refinement were 0.146 and 0.173 after protein model building and the addition of H atoms.

It was very difficult to detect anisotropy of atoms in water molecules in the $F_o - F_c$ map at this stage. Nonetheless, we introduced anisotropic displacement parameters for water molecules in the refinement to make it possible to compare the results based on the conventional Fourier method with those of the MEM-assisted refinement. The default value of the ISOR restraint in the *SHELXL* program was used in the refinement. There was no indication of how to estimate the initial anisotropic parameters from the $F_o - F_c$ map. Therefore, the anisotropies were gradually introduced in order to maintain consistency with the MEM-assisted refinement. The R_1 and R_{free} values decreased from 0.146 and 0.173 to 0.130 and 0.147, respectively.

It was also difficult to recognize anisotropy of atoms for protein molecules in the $F_o - F_c$ and $2F_o - F_c$ maps, as shown in Fig. 6. The $F_o - F_c$ map containing the carbonyl O atom of Lys24, the same residue as in Fig. 3, is shown. Although there
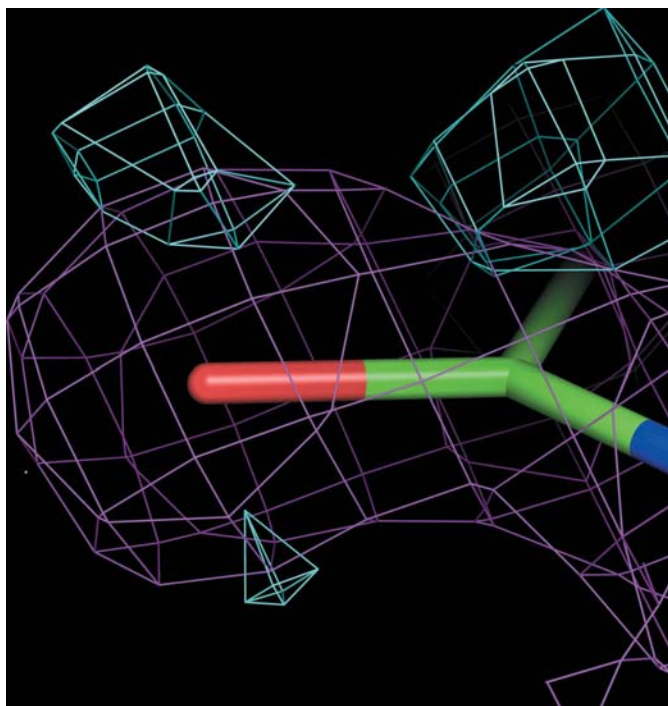

Figure 6
The F_o and $F_o - F_c$ maps for the carbonyl O atom of Lys24. The F_o and $F_o - F_c$ maps are shown as a purple and a sky-blue mesh, respectively. The charge-density levels are 1.00 e Å^{-3} for the F_o map and 0.10 e Å^{-3} for the $F_o - F_c$ map.

were no positive or negative differences in charge densities indicating anisotropy of the O atom, anisotropic displacement parameters were introduced for the atoms in the protein molecule in order to compare the results based on the conventional Fourier method with those of the MEM-assisted refinement. The same values of the ISOR, SIMU and σ parameters were used in the MEM-assisted refinement. It was impossible to determine reasonable parameters when the positional and anisotropic thermal parameters were refined for whole atoms in the protein simultaneously. Therefore, anisotropies of atoms were gradually introduced during the refinement. The R_1 and R_{free} values decreased from 0.130 and 0.147 to 0.120 and 0.131, respectively. From an $F_o - F_c$ map, we found large difference charge-density peaks indicating the Cl$^-$ ion. The Cl$^-$ ion and disordered water molecule were added to the structure model. The R_1 and R_{free} values of the final refinement based on the Fourier map were 0.1146 and 0.1206, respectively.

3. Results

3.1. Structure and charge density

The overall structure was almost identical to the structure deposited in the PDB (PDB code 1ulr). The structure of the putative acylphosphatase molecule consisted of five anti-parallel β-strands and two α-helices. There is one protein molecule in the asymmetric unit. The asymmetric unit includes 677 non-H atoms, one partially occupied Cl$^-$ ion and 203 water molecules.

The protein molecule was mainly ordered, with the exception of the C- and N-termini and several side chains. The charge densities of several atoms in the side chains of residues Glu73, Glu78 and Lys83 were relatively weak compared with those of the other well ordered side chains. These residues were located at the protein–protein contact interfaces in the crystal packing. These residues included zero-occupancy atoms in the deposited structure model. We were able to assign all the atoms from the MEM charge density.

There were nine multiple-conformer residues: Pro2, Lys23, Tyr33, Lys50, Glu51, His58, Gln62, Glu70 and Tyr88. The ratio of the multiple-conformer residues to the total of 87 residues in the protein molecule was 10.3%. Both the main chain and the side chain of Pro2 at the N-terminus were disordered. The C$^\alpha$ atom of Glu70 was disordered along with the side chain. Only the atoms in the side chains were disordered in the other seven residues. The nine disordered residues are all located on the protein surface with packing contacts.

At the chain termini the N-terminal residue Pro2 was slightly disordered, but the residues after Arg3 were not disordered. The side chain of the C-terminal residue Tyr88 showed two conformers with two directions in the χ_1 angle, although the main chain of the residue was not disordered.

3.2. H atoms

A total of 698 charge-density peaks were observed at the positions of H atoms in the protein molecule in the difference

MEM charge density, as shown in Fig. 7. The 698 atoms correspond to 93.9% of the total of 743 H atoms. A total of 599 ordered H atoms, corresponding to 96.3% of the total 622 atoms, were observed together with 99 disordered H atoms, corresponding to 81.8% of the total 121 atoms. A histogram of the percentage of observed H atoms in the difference MEM charge density as a function of B factor is shown in Fig. 8. The average B factor for the observed H atoms in the structure was 7.92 $Å^2$, while the average for non-observed H atoms was 12.76 $Å^2$.

3.3. Intramolecular hydrogen bonds in the protein

A total of 71 hydrogen bonds were detected in the one molecule using the program *XtalView* (McRee, 1999), including 15 hydrogen bonds that were formed between main-chain atoms in the α-helix. A further 25 hydrogen bonds are formed between main-chain atoms in the β-strands. Another nine hydrogen bonds are formed between main-chain atoms not related to the α-helix and the β-strands, 14 hydrogen bonds are formed between main-chain atoms and side-chain atoms and eight hydrogen bonds are formed between side-chain atoms. The hydrogen bonds are also clearly revealed in the MEM charge density, as shown in Fig. 9.

The distribution of bonding distances for hydrogen bonds is shown in Fig. 10 as a function of distance. The distance of the shortest hydrogen bonds was 2.50 Å between heavy atoms. This was observed between the amino group of Lys50 conformer A (Lys50A) and the carboxyl of Glu54 and between the ring imino of Pro2A and the carboxyl O atom of Glu79. These represent ionic bonding and the donors in these bonds come from one conformer of multiple-conformer residues. There are nine hydrogen bonds between multiple-conformer residues in the protein. The shortest hydrogen-bond distance excluding multiple-conformer atoms was 2.68 Å between the hydroxyl of Tyr17 and the backbone carbonyl O atom of Leu60.
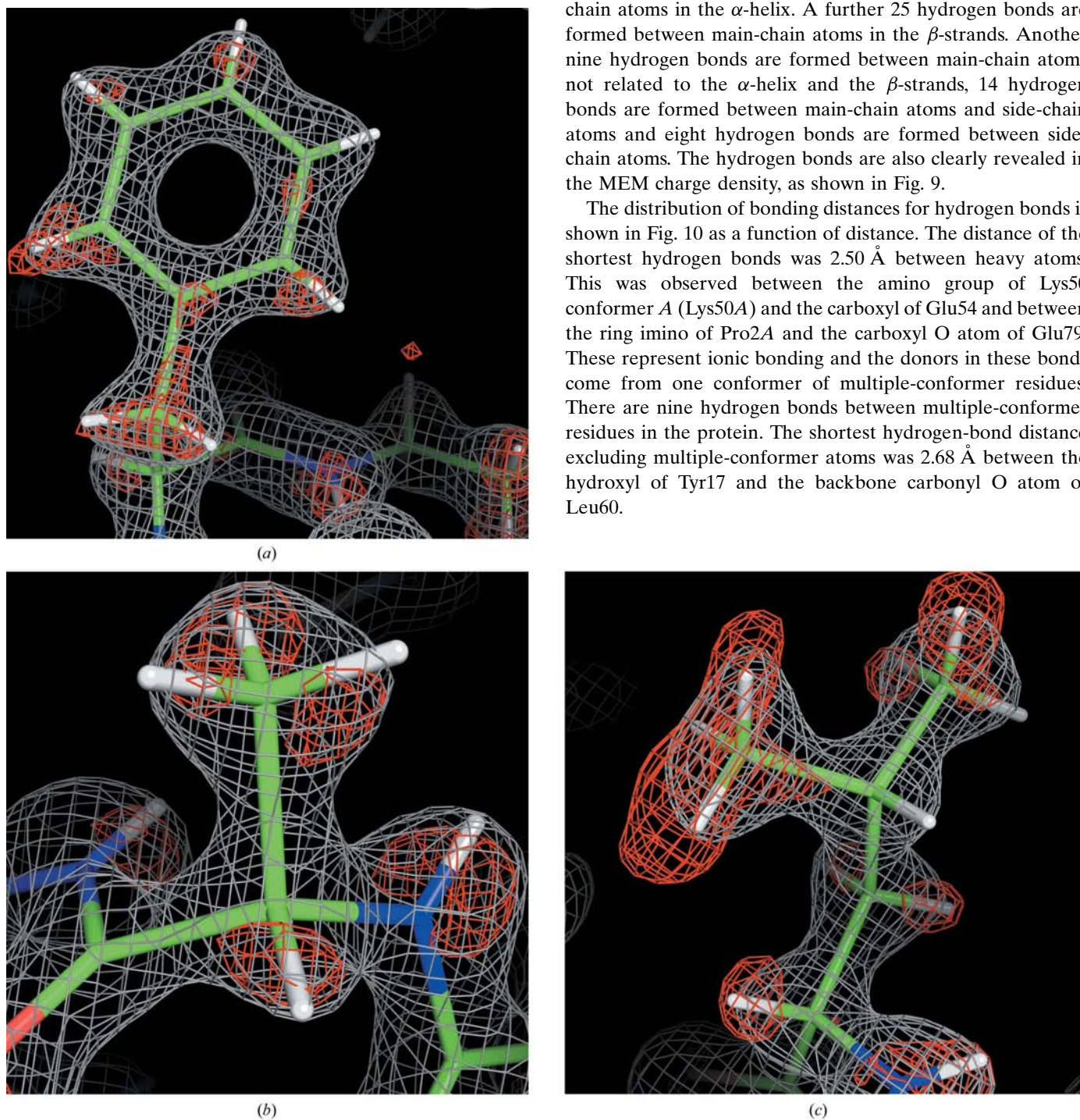
Figure 7
The MEM and difference MEM charge densities of H atoms. (*a*) Phe20, (*b*) Ala25 and (*c*) Leu26. The MEM maps are shown as a grey mesh. The difference MEM maps are shown as red lines. The MEM charge-density levels are 1.00 e $Å^{-3}$ for (*a*), (*b*) and (*c*). The difference MEM charge-density levels are 0.30 e $Å^{-3}$ in (*a*), 0.32 e $Å^{-3}$ in (*b*) and 0.25 e $Å^{-3}$ in (*c*).

3.4. Solvent

A total of 203 water molecules were located in the asymmetric unit, of which nine were disordered. Four water molecules had occupancies of less than 1.0. No H atoms for waters were added to the model. The Matthews coefficient (Matthews, 1968) was estimated as 1.54 $\text{Å}^3 \text{Da}^{-1}$ and corresponds to a solvent content of 20.1%. The volume of a water molecule was calculated as 25.83 Å^3 and is consistent with reported values (Hubbard & Argos, 1994). A total of 128 water molecules were in the first hydration shell and 71 were in the second hydration shell. Four were in the third hydration shell. The specific gravity of the water region was 1.161 based on the solvent content, when the specific gravity of the protein region was taken to be the constant value of 1.351 for protein structures.

3.5. Bond lengths and angles inside the protein molecule

The structure was validated by a Ramachandran plot calculated using the program PROCHECK (Laskowski et al., 1993), which showed 93% of residues to be in the most favoured regions and the remaining 7% to be in the additional allowed regions. The result is identical to that for the deposited structure. The residues Arg11, Val12, Val15, Glu62 and Glu68 were in the additional allowed regions. Arg11 and Glu68 were located at the edge of β-strands, Glu62 was located at the edge of the α-helix and Val12 and Val15 were located in the turn loop. The charge densities of the residues, including the multiple-conformer side chain of Glu68, were clearly observed in the MEM charge density. There was no 'distorted geometry' indicated by bonds differing by >0.05 Å and bond angles differing by >10.0° from the ideal small-molecule values (Engh & Huber, 1991).

The bond distances in the main chain were investigated using PROCHECK. There were ten bond distances that deviated from the values for small molecules (Engh & Huber, 1991) by more than 0.028 Å. The largest deviation in the distances is for the C—N bond in Ala80. The deviation in the

distance is 0.042 Å. Ala80 is located on the turn loop and on the surface of the protein. The main-chain bond angles were also investigated using PROCHECK, which showed that 49 bond angles deviate from the values for small molecules (Engh & Huber, 1991) by more than 3.80°. The largest deviation of
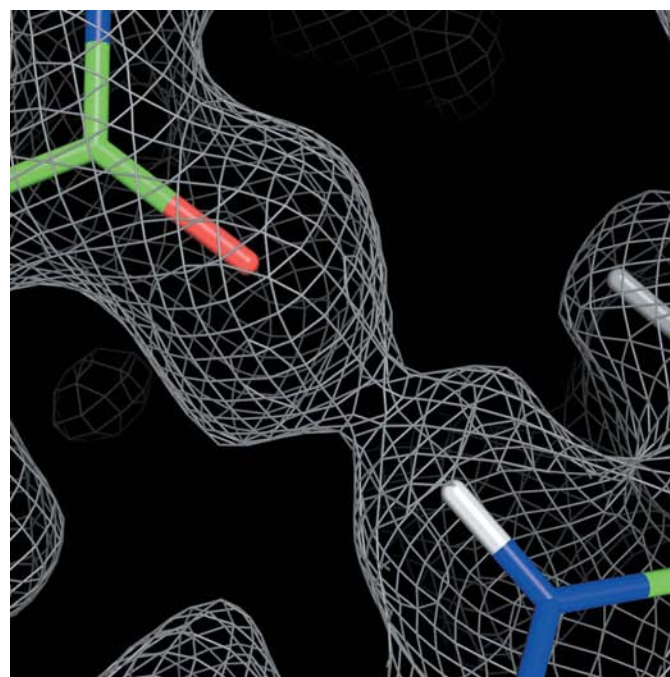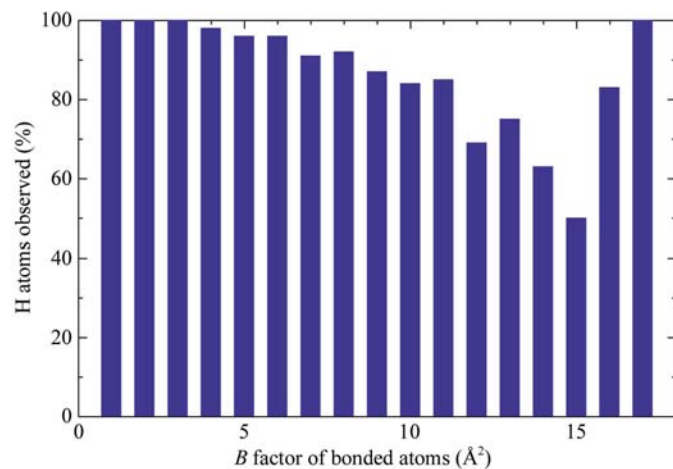

(a)
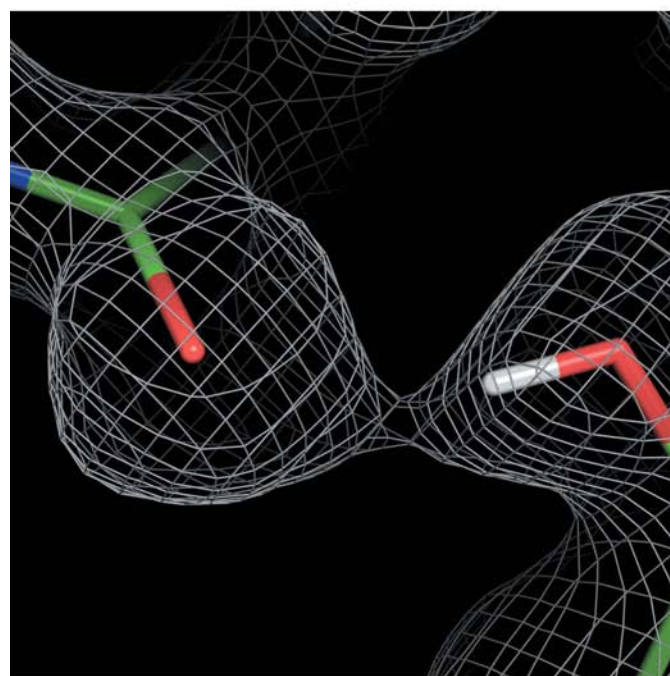
(b)

Figure 9
The MEM charge densities of hydrogen bonds. (a) The MEM charge density of hydrogen bonds in β-strands. The MEM map with 0.25 e Å^{-3} between Leu4 N and Ala46 O is shown. (b) The MEM map (0.32 e Å^{-3}) between the hydroxyl of Tyr17 and the backbone carbonyl O atom of Leu60 is shown.

Figure 8
Histogram of the fractions of H atoms observed as a function of the B factor of the hydrogen-bonded heavy atom.

Nishibori et al. · Refinement using the maximum-entropy method   **245**

the angles is in $C^\beta—C^\alpha—C$ for His86, which is located on the β-strand and close to the C-terminus.

The average ω value is 179.1°, with a standard deviation σ_ω of 5.8°. There are six regions in which the deviation from the ideal value of 180° is two times larger than the standard deviation. These are located on the turn loop. The most deviating ω angle is between residues Pro64 and Arg65, where the ω angle is 196.7°. We investigated the ω values of 11 ultrahigh-resolution structures, crambin (PDB code 1ejg; Jelsch *et al.*, 2000), subtilisin (1gci; Kuhn *et al.*, 1998), syntenin (1r6j; Kang *et al.*, 2004), HiPIP (1iua; Liu *et al.*, 2002), trypsin (1pq7; Schmidt *et al.*, 2003), aldose reductase (1us0; Howard *et al.*, 2004), antifreeze protein (1ucs; Ko *et al.*, 2003), CBM36 (1w0n; Jamal-Talabani *et al.*, 2004), PAK pilin (1x6z; Dunlop *et al.*, 2005), rubredoxin (1yk4; Bönisch *et al.*, 2005) and HFBII (2b97; Hakanpää *et al.*, 2006), deposited in the PDB. The minimum value of ω, 153.7°, was found in aldose reductase and the maximum value of ω, 204.5°, was found in subtilisin. There are several non-ideal ω values which deviate more than $2\sigma_\omega$ from the ideal value. The distribution of the residues with non-ideal ω values in secondary structures is 62.8% in loops, 30.6% in β-strands and 6.6% in α-helices. These facts indicate that the ω values of the present study are comparable to those from ultrahigh-resolution data.

3.6. Anisotropy of atoms

The anisotropy of atoms was validated using the program *PARVATI* (Merritt, 1999). There is no perfectly isotropic atom in the protein molecule. The mean anisotropy and its σ for the protein molecule were 0.411 and 0.199, which are close to the average values, 0.466 and 0.143, of the atomic resolution and ultrahigh-resolution structures reported by Merritt (1999). The mean anisotropy of the O atoms in the protein molecule is slightly larger than that of other atoms. This reflects the larger anisotropy of the carbonyl O atom in the main chain, as shown in Fig. 3. The largest anisotropy was found for the N atom of Gly32. The mean anisotropy and its σ of the solvent water molecules were 0.344 and 0.179, respec-
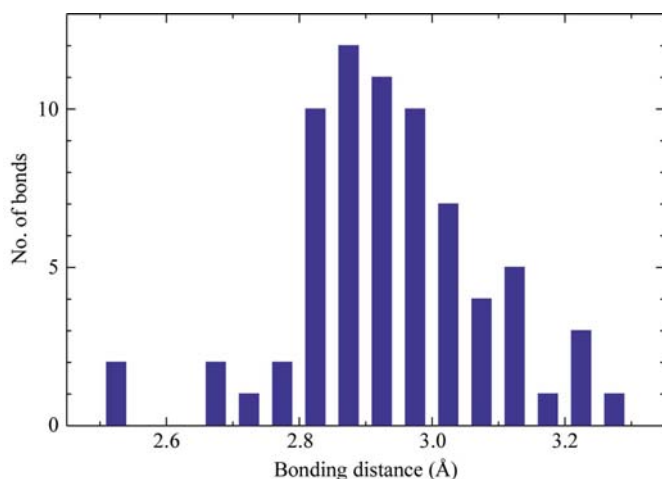

Figure 10
Distance distribution of the hydrogen bonds.

tively. These values were also close to the average values of the atomic resolution and ultrahigh-resolution structures reported by Merritt (1999). These facts confirm that the MEM charge-density-based refinement, assisted by the SIMU and ISOR restraints in *SHELX*, provide us with a reasonable solution for the anisotropy of atoms even from near-atomic resolution data.

4. Conclusion

In this study, we carried out high-resolution refinement that gave values of 0.096 and 0.100 for R_1 and R_{free}, respectively, for 1.3 Å data from a putative acylphosphatase when assisted by the use of MEM charge density. The high resolution and the least-biased MEM charge densities play a crucial role in the detection of water molecules and multiple conformers of residues. Approximately 30 water molecules and four multiple-conformer residues in the MEM-based refinement were not found by conventional Fourier-based refinement. In addition, the anisotropy of atoms and the H atoms were clearly revealed by the MEM charge density. These were important clues for high-resolution refinement using aniso-tropic displacement parameters and H atoms. These abilities of the MEM charge density are very useful in macromolecular crystallography.

The omit-difference MEM map procedure was necessary to reveal detailed structural features such as multiple confor-mations in the protein. The phases of structure factors for the MEM calculation have to be calculated from a particular structural model. Model bias of the phases cannot be completely avoided in the MEM charge-density method, so conventional protein refinement should be performed exhaustively prior to the application of MEM-based refine-ment using at least a near-atomic resolution data set. Four of the nine residues modelled as multiple conformers in the MEM-based structure were not detected without the omit-difference MEM charge density. The proportion that should be omitted from the whole protein molecule is an important issue in revealing detailed disordered structure. In the present study, we determined the ratio of the omitted part to the protein molecule through many trial calculations. If we remove ten residues, which correspond to 10% of the protein molecule, three of nine disordered parts could not be detected. It should be mentioned that it was not possible to construct the present model structure for solvent waters unless all the waters in the PDB-deposited model were completely removed at the initial stage of the refinement. In fact, 30 solvent water molecules were only detected after the removal of all the water molecules.

As for computational times, MEM calculations need approximately 3000 times the CPU time of a conventional Fourier calculation. A total of 1000 MEM calculations were needed to reach the final solution. The total CPU time at SX-6i at Nagoya University and SX-7 at the RIKEN vector supercomputer exceeded 650 h in the present case. The typical CPU time for one MEM calculation was 40 min with 1500 cycle iterations. To obtain the difference MEM map we need

twice as many MEM calculations, since F_o and F_c maps must be individually calculated each time by MEM. The difference $F_o - F_c$ map is obtained by subtracting the F_c map from the F_o map. The calculation costs of the MEM depend linearly on the size of the unit cell. For medium-sized proteins deposited in the PDB, approximately ten times greater computational resources are needed compared with the present calculation. The continual improvement of the calculation speed of computers will probably lead to routine MEM calculations for medium-sized proteins in the near future. We believe that a recent high-end PC with a quad-core and dual CPU with more than 8 Gb memory may perform the MEM calculation for the present protein with a realistic computational time.

In the present study, we have shown the validity and capability of the MEM charge-density approach for macromolecular crystallography. The MEM-based refinement for protein crystallography is competent in determining an atomic resolution structure and in overcoming the termination effects of the Fourier method, even for near-atomic resolution data.

References

Afonine, P. V., Lunin, V. Y., Muzet, N. & Urzhumtsev, A. (2004). *Acta Cryst.* D**60**, 260–274.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Blaszczyk, J., Li, Y., Shi, G., Yan, H. & Ji, X. (2003). *Biochemistry*, **42**, 1573–1580.

Blaszczyk, J., Shi, G., Yan, H. & Ji, X. (2000). *Structure*, **8**, 1049–1058.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. & Schneider, M. (2003). *Nucleic Acids Res.* **31**, 365–370.

Bönisch, H., Schmidt, C. L., Bianco, P. & Ladenstein, R. (2005). *Acta Cryst.* D**61**, 990–1004.

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.

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

Dubnovitsky, A. P., Kapetaniou, E. G. & Papageorgiou, A. C. (2005). *Protein Sci.* **14**, 97–110.

Dunlop, K. V., Irvin, R. T. & Hazes, B. (2005). *Acta Cryst.* D**61**, 80–87.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

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

Hakanpää, J., Linder, M., Popov, A., Schmidt, A. & Rouvinen, J. (2006). *Acta Cryst.* D**62**, 356–367.

Howard, E. I., Sanishvili, R., Cachau, R. E., Mitschler, A., Chevrier, B., Barth, P., Lamour, V., Van Zandt, M., Sibley, E., Bon, C., Moras, D., Schneider, T. R., Joachimiak, A. & Podjarny, A. (2004). *Proteins*, **55**, 792–804.

Hubbard, S. J. & Argos, P. (1994). *Protein Sci.* **3**, 2194–2206.

Jamal-Talabani, S., Boraston, A. B., Turkenburg, J. P., Tarbouriech, N., Ducros, V. M.-A. & Davies, G. J. (2004). *Structure*, **12**, 1177–1187.

Jelsch, C., Teeter, M. M., Lamzin, V., Pichon-Pesme, V., Blessing, R. H. & Lecomte, C. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 3171–3176.

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

Kang, B. S., Devedjiev, Y., Derewenda, U. & Derewenda, Z. S. (2004). *J. Mol. Biol.* **338**, 483–493.

Kitaura, R., Kitagawa, S., Kubota, Y., Kobayashi, T. C., Kindo, K., Mita, Y., Matsuo, A., Kobayashi, M., Chang, H. C., Ozawa, T. C., Suzuki, M., Sakata, M. & Takata, M. (2002). *Science*, **298**, 2358–2361.

Ko, T.-P., Robinson, H., Gao, Y.-G., Cheng, C.-H. C., DeVries, A. L. & Wang, A. H.-J. (2003). *Biophys. J.* **84**, 1228–1237.

Kubota, Y., Takata, M., Matsuda, R., Kitaura, R., Kitagawa, S., Kato, K., Sakata, M. & Kobayashi, T. C. (2005). *Angew. Chem. Int. Ed.* **44**, 920–923.

Kuhn, P., Knapp, M., Soltis, M., Ganshaw, G., Thoene, M. & Bott, R. (1998). *Biochemistry*, **37**, 13446–13452.

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

Liu, L., Nogi, T., Kobayashi, M., Nozawa, T. & Miki, K. (2002). *Acta Cryst.* D**58**, 1085–1091.

McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Merritt, E. A. (1999). *Acta Cryst.* D**55**, 1109–1117.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Nishibori, E., Sunaoshi, E., Yoshida, A., Aoyagi, S., Kato, K., Takata, M. & Sakata, M. (2007). *Acta Cryst.* A**63**, 43–52.

Nishibori, E., Terauchi, I., Sakata, M., Takata, M., Ito, Y., Sugai, T. & Shinohara, H. (2006). *J. Phys. Chem. B*, **110**, 19215–19219.

Noritake, T., Aoki, M., Towata, S., Seno, Y., Hirose, Y., Nishibori, E., Takata, M. & Sakata, M. (2002). *Appl. Phys. Lett.* **81**, 2008–2010.

Roeser, D., Dickmanns, A., Gasow, K. & Rudolph, M. G. (2005). *Acta Cryst.* D**61**, 1057–1066.

Schmidt, A., Jelsch, C., Ostergaard, P., Rypniewski, W. & Lamzin, V. S. (2003). *J. Biol. Chem.* **278**, 43357–43362.

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

Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Snyder, G. J., Christensen, M., Nishibori, E., Caillat, T. & Iversen, B. B. (2004). *Nature Mater.* **3**, 458–463.

Sugahara, M. & Miyano, M. (2002). *Tanpaku Kakusan Koso*, **47**, 1026–1032.

Takata, M. & Sakata, M. (1996). *Acta Cryst.* A**52**, 287–290.

Tanaka, H., Takata, M., Nishibori, E., Kato, K., Iishi, T. & Sakata, M. (2002). *J. Appl. Cryst.* **35**, 282–286.

Thunnissen, M. M., Taddei, N., Liguri, G., Ramponi, G. & Nordlund, P. (1997). *Structure*, **5**, 69–79.

Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.

Yamamoto, M., Kumasaka, T., Fujisawa, T. & Ueki, T. (1998). *J. Synchrotron Rad.* **5**, 222–225.