Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Advantages of high-resolution phasing: MAD to atomic resolution

Andrea Schmidt,[a]* Ana Gonzalez,[a]† Richard J. Morris,[a]‡ Marcelo Costabel,[b] Pedro M. Alzari[b] and Victor S. Lamzin[a]

[a]EMBL Hamburg c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany, and [b]Unite de Biochimie Structurale, Institut Pasteur, 25 Rue du Dr Roux, F-75724 Paris, France

† Present address: Stanford Linear Accelerator Laboratory, PO Box 4349 MS 99, Stanford, CA 94309-210, USA.
‡ Present address: Global Phasing Ltd, Sheraton House, Castle Park, Cambridge CB3 0AX, England.

Correspondence e-mail:
andrea@embl-hamburg.de

Received 14 March 2002
Accepted 26 June 2002

PDB Reference: endoglucanase A, 1is9, 1is9sf.

The structure of the endoglucanase A from *Clostridium thermocellum* CelA was re-solved by three-wavelength MAD. Experimental phases were obtained in the resolution range 25–1.0 Å. Various structure-solution approaches were tested in order to quantify the contribution of each wavelength. Two-wavelength MAD phasing was sufficient to obtain excellent experimental phases. SAD at the remote wavelength also resulted in interpretable maps. The three-wavelength MAD electron-density map was of excellent quality: for parts of the structure, atom types and bond types could be easily assigned. Double bonds in peptide links and side chains could be located owing to their increased electron density indicating their π character. Comparison with a previously determined structure of CelA at 1.65 Å showed that, apart from a few additional multiple conformers and differently oriented side chains, major differences occur at the protein–solvent interface. A complete additional solvent shell could be observed and the inner shells have been completed. The high accuracy of the structure allowed unambiguous assignment of the protonation state for the active-site catalytic carboxylates.

1. Introduction

The solution of protein crystal structures with the multi-wavelength anomalous dispersion (MAD) method has become increasingly popular. In many respects, the method is superior to molecular-replacement (MR) or multiple isomorphous replacement (MIR) techniques. MAD avoids the direct introduction of model bias into the structure refinement and problems with non-isomorphism between heavy-atom derivative(s) and native crystals. Data are typically collected from one crystal and this reduces the number of systematic errors in the structure-solution step.

The aim of a protein crystal structure analysis is to obtain information on the protein's properties and function. Interactions between protein and its substrate, cofactor, surrounding solvent or with other proteins can be assigned from the structure and in many cases the mechanism can be deduced. In order to assign interactions unambiguously, high accuracy in the atomic positions is required. An important factor for this accuracy is the resolution of the X-ray data. For mechanistic studies, the crystallographic resolution should therefore be as high as possible.

For the MR or MIR techniques, atomic resolution data (1.2 Å or higher; Dauter *et al.*, 1997) are not needed at the structure-solution stage and can even hinder success, yet they contain most of the information needed for obtaining an accurate final structure. For MR, differences between search model and target structure manifest themselves in the high-resolution data. In the MIR case, non-isomorphism often makes detection of any heavy-atom contribution at high resolution practically impossible.

With MAD, these problems are largely overcome. No external search model is needed and the anomalous signal is present to high resolution (Nagem *et al.*, 2001) as shown in Fig. 1 for the CelA MAD data, provided that the data are accurately measured (Weiss, 2001). The method is applicable to any type of heavy-atom derivative produced by selective labelling or soaking, even if the heavy atoms do not bind to specific sites (Dauter *et al.*, 2000).

High-quality MAD data can be collected routinely on tuneable synchrotron beamlines, which provide high X-ray intensity, a well focused beam and sufficient energy resolution. Advanced cryocrystallography techniques facilitate data collection as secondary radiation damage is largely avoided.

In contrast to other structure-solution approaches, MAD allows inclusion of all data already at the phasing stage. The initial model can already be built into an atomic resolution electron-density map phased from the anomalous signal only. The introduction of errors arising from incomplete low-resolution start models and several cycles of phase extension and model rearrangement during refinement can be avoided (Burling *et al.*, 1996).

In this study, we present the structure of endoglucanase A from *Clostridium thermocellum* (CelA; PDB code 1cem; Alzari *et al.*, 1996), re-solved by MAD on a mercury derivative to atomic resolution. This is one of the first protein structures solved by atomic resolution MAD. The results of different structure-solution approaches are presented. The advantages of the method itself, structural details and the differences to the known CelA structure are discussed.

2. Materials and methods

2.1. Preparation of crystals

CelA was crystallized as described elsewhere (Souchon *et al.*, 1996). The mercury derivative was prepared by soaking the crystals in a solution containing 1 m*M* mercury (II) chloride, as outlined by Alzari *et al.* (1996).

2.2. Data collection

MAD X-ray data on the mercury derivative of CelA were collected on the BW7A beamline equipped with a MAR imaging-plate scanner at EMBL Hamburg. Three data sets were acquired at wavelengths of 0.8006 Å (high-energy remote), 1.0063 Å (peak) and 1.0084 Å (inflection point) at a temperature of 100 K. To appropriately record strong (low-resolution) diffraction intensities, data for each wavelength were collected in three passes to resolution cutoffs of 0.95, 2.0 and 3.0 Å, with successively increasing oscillation angles and reduced exposure times.

2.3. Data processing

Data were processed using programs *DENZO* and *SCALEPACK* from the *HKL* suite (Otwinowski & Minor, 1996). Bijvoet pairs of reflections were kept separate during scaling. Intensities were converted to structure-factor amplitudes using the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) program *TRUNCATE* (French & Wilson, 1978). Data sets were subsequently combined and scaled isotropically, using the remote-wavelength data set as the reference. This combined and scaled data set was then used for all subsequent procedures within the common resolution range of 25 to 1.0 Å.

2.4. Heavy-atom search

Anomalous and isomorphous difference Patterson maps were calculated with *FFT* (Ten Eyck, 1973; Read & Schierbeck, 1988) from the *CCP*4 suite using the full resolution range. Reflections with structure-factor amplitudes of less than 1σ were excluded. The set of four strong sites from these maps was refined in *MLPHARE* and further positions were found in difference Fourier maps. The resulting set of six Hg atoms was used as the heavy-atom model for phasing.

2.5. Structure solution – phasing

Several ways of obtaining starting phases were tried in order to quantify the contribution of each wavelength and to test different protocols. *MLPHARE* was used as the phasing program.

2.5.1. Three-wavelength MAD. The six initial mercury positions were assigned an initial *B* factor of 20 Å^2 and occupancies of 0.5 and were first refined against centric data only (ten cycles). Subsequently, all data were included and anomalous occupancies (starting value 0.5) refined (2 × 10 cycles). Afterwards, solvent flattening was applied using *DM* (Cowtan & Main, 1996). The phases obtained from this procedure were used for building the protein structure.

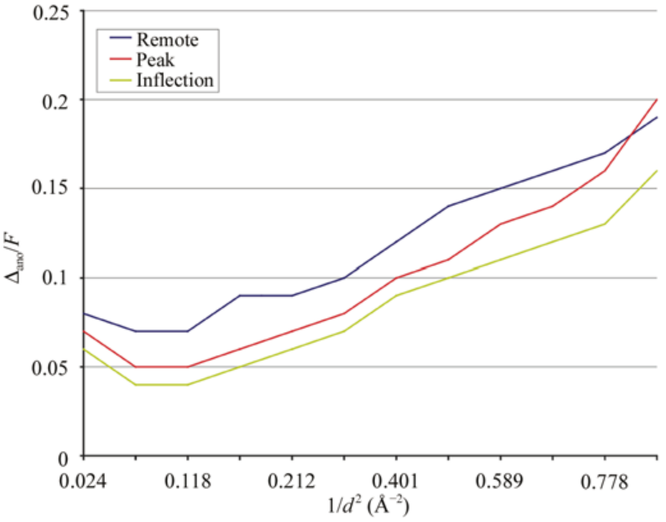


Figure 1
Contribution of the anomalous signal in the CelA MAD data at each wavelength (Δ_{ano}/F), as a function of resolution.

2.5.2. SAD and two-wavelength MAD. SAD on each of the three data sets and two-wavelength MAD with all possible wavelength pairs were carried out using the same protocol as for the three-wavelength case. Again, *DM* was used for subsequent phase improvement.

2.6. Definitions

The phases obtained directly from phasing programs were regarded as 'true experimental phases', since they were obtained from the heavy-atom contributions only with no other models imposed. Differences in phasing were masked out by subsequent solvent flattening. Phases from *DM* are referred to as 'solvent-flattened phases'. Although used for initial model building, these were treated separately in the comparison of structure-solution methods.

2.7. Model building

A free-atom model was built into the electron density calculated from $F_{\rm remote}$ and solvent-flattened phases from the three-wavelength MAD. 5% of the data were set aside for cross-validation using $R_{\rm free}$. To this free-atoms model, the six Hg atoms were added. *ARP*/*wARP*5.1 (Perrakis *et al.*, 1999) was launched in warpNtrace mode and the CelA protein built in five autotracing cycles. This model was essentially complete, except for some residues which were disordered and remained undefined also in the final maps. Side chains were built automatically using side_dock with default settings. The resulting protein molecule was checked visually for additional features in the electron density ($2mF_o - DF_c$ maps) in *O* (Jones *et al.*, 1991). Some manual adjustments of side chains, the insertion of double conformers and Cl atoms were carried out and the structure was then subjected to several cycles of solvent building and refinement in *REFMAC* (Murshudov *et al.*, 1997). The occupancies for the heavy atoms and the double conformers were adjusted between refinement cycles. The criterion for occupancy adjustment was the behaviour of *B* factors for the given atoms. Occupancies were considered correct when atomic displacement parameters were in the range of those for the surrounding groups or atoms. *R*, $R_{\rm free}$ and the number of atoms found/deleted were monitored throughout the refinement. When the *R* factors had converged and the protein model was complete (the number of atoms was constant), anisotropic atomic displacement parameters were refined for the protein molecule and the heavy atoms (Hg, Cl).

2.8. Map-quality assessment

The quality of the maps calculated from experimental or solvent-flattened phases was checked visually for interpretability. The final refined model was used as the reference for calculating map correlations in order to assess the quality of the starting phases from each structure-solution protocol. As a further test, *ARP*/*wARP* was used to generate a free-atoms model from the starting phases and run automatic tracing of the polypeptide chain. The completeness of these models was then compared.

Table 1
Data-processing statistics.

Values in parentheses correspond to the outer shell: 0.96–0.95 Å for the remote-wavelength, 1.04–1.03 Å for the peak-wavelength and 1.01–1.00 Å for the inflection-point wavelength data.

Wavelength (Å)	0.8006 (remote)	1.0063 (peak)	1.0084 (inflection)
Resolution range	30–0.95	30–1.03	30–1.00
No. of observations	1680709	1605139	1579432
No. of unique reflections	201040	158384	171811
No. of reflections used in statistics†	382040	304029	328152
$I/\sigma(I)$	13.3 (3.2)	16.0 (3.9)	13.5 (2.9)
$R_{\rm sym}$	0.055 (0.427)	0.049 (0.355)	0.051 (0.262)
$R_{\rm p.i.m.}$	0.045 (0.331)	0.031 (0.252)	0.025 (0.291)
$R_{\rm r.i.m.}$	0.104 (0.594)	0.080 (0.482)	0.060 (0.414)
Completeness	97.3 (97.6)	98.2 (90.9)	96.9 (66.3)
Anomalous completeness	95.7	98.7	97.5
Anomalous redundancy†	2	2	2
Wilson *B* factor	4.4	5.5	5.6

† 'Scale anomalous' option in *SCALEPACK*, Bijvoet pairs kept separate.

2.9. Comparison to a previously determined CelA structure

Structural comparison focused mainly on the solvent regions, where most of the differences occurred. Solvent molecules were sorted into three shells. The coordinates were compared and their r.m.s.d.s to the previously determined CelA structure were calculated. Shells were assigned as follows: shell1 contains water molecules directly attached to the protein, shell2 contains those water molecules attached directly to the solvent from shell1 and shell3 consists of all solvent further outside.

2.10. Phase-probability distribution functions, averaged peptide maps, atom-type assignment

Phase-probability distribution functions were calculated from Hendrickson–Lattman coefficients (Hendrickson & Lattman, 1970). Calculation of average peptide maps was performed as described in Lamzin *et al.* (1999). For the analysis of the electron density *versus B* factor (Sevcik *et al.*, 1996) and atom-type assignment, the atoms were sorted on their *B* factor and then divided into *B*-factor bins consisting of 350 atoms each.

3. Results and discussion

3.1. Data processing

A summary of statistics is given in Table 1. Despite the fact that these data were collected over the standard asymmetric unit of reciprocal space for space group $P2_12_12_1$ (90°), the quality and completeness of the anomalous signal were satisfactory. Unit-cell parameters were determined to be a = 49.44 (3), b = 62.83 (2), c = 103.89 (3) Å.

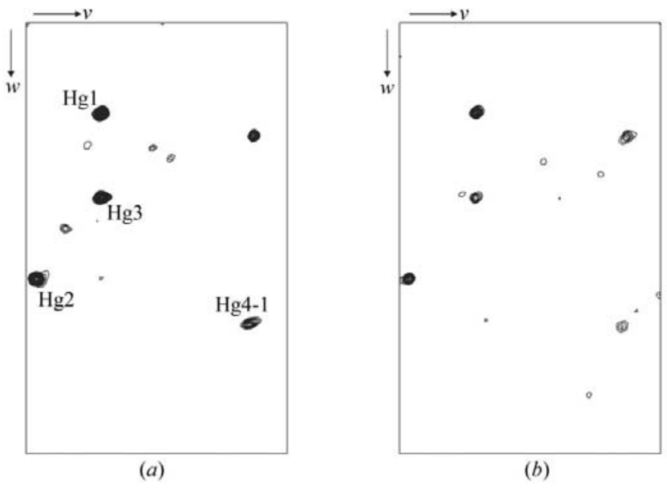
3.2. Heavy-atom search and refinement

Four mercury positions were found directly from either isomorphous or anomalous Patterson maps with peak heights between 15σ and 30σ (Fig. 2). The two remaining sites could be detected in double-difference Fourier maps ($F_{\rm peak} - F_{\rm remote} - F_{\rm heavy}$) after the first four had been refined in *MLPHARE*. All the heavy atoms are located in hydrophobic pockets buried inside the protein and bind to the five cysteine residues in the CelA structure. One of the binding sites was split, with the Hg atoms having about the same occupancies and a distance between the two heavy-atom positions of 2.3 Å. The corresponding cysteine SH group also shows two conformations. Chloride anions were found attached to all Hg atoms.

3.3. Phasing

A summary of the structure-solution results and an assessment of map quality is given in Table 2. All phasing approaches except SAD at the inflection point and the two-wavelength MAD with peak and inflection data sets gave interpretable maps that allowed straightforward model building. Electron density obtained from three-wavelength MAD or the other two two-wavelength cases were of exceptional quality even before solvent flattening (Fig. 3). For most starting phase sets the autotracing routine in *ARP/wARP* could straightforwardly build most of the protein model, thus indicating the high quality of the electron density more objectively than a visual test. At least 70% of the main chain was immediately traced; solvent flattening and five cycles of unrestrained refinement with *REFMAC* (Murshudov *et al.*, 1997) of a free-atoms model increased this percentage to 85% of the total of 363 residues (see Table 2). It was possible to build a largely complete model at a very early stage, without the need for phase extension or repeated rearrangements. Therefore, the resulting protein structure can essentially be considered to be unbiased.

Collecting high-resolution MAD data is time-consuming, but this experiment shows that at least one wavelength (not remote) can be omitted for the benefit of extending the X-ray data to the highest possible resolution. Excellent starting phases could be obtained from two-wavelength MAD combining either peak and remote or inflection-point and remote wavelengths (see Table 2). This has also been shown previously by González *et al.* (1999) for various test cases at medium resolution. It seems that in the case of CelA the peak and inflection-point data sets are equally important in the phasing procedure when combined with the remote wavelength. The combination of peak- and inflection-point wavelength data gave relatively weak experimental phases. In this case, no useful phases could be obtained for a resolution range higher than 2 Å, which leaves the high-resolution data with undetermined phases. The reason for this could be the small dispersive differences between these two data sets. Also, the difference in the strength of the anomalous signal is small (see Fig. 1) in this case.



(*a*) (*b*)

Figure 2
(*a*) Isomorphous difference (peak − remote) and (*b*) anomalous difference Patterson maps (peak) in the resolution range 25.0–1.0 Å, section u = 0.5. Maps are contoured at 2σ, 4σ, 6σ *etc.*

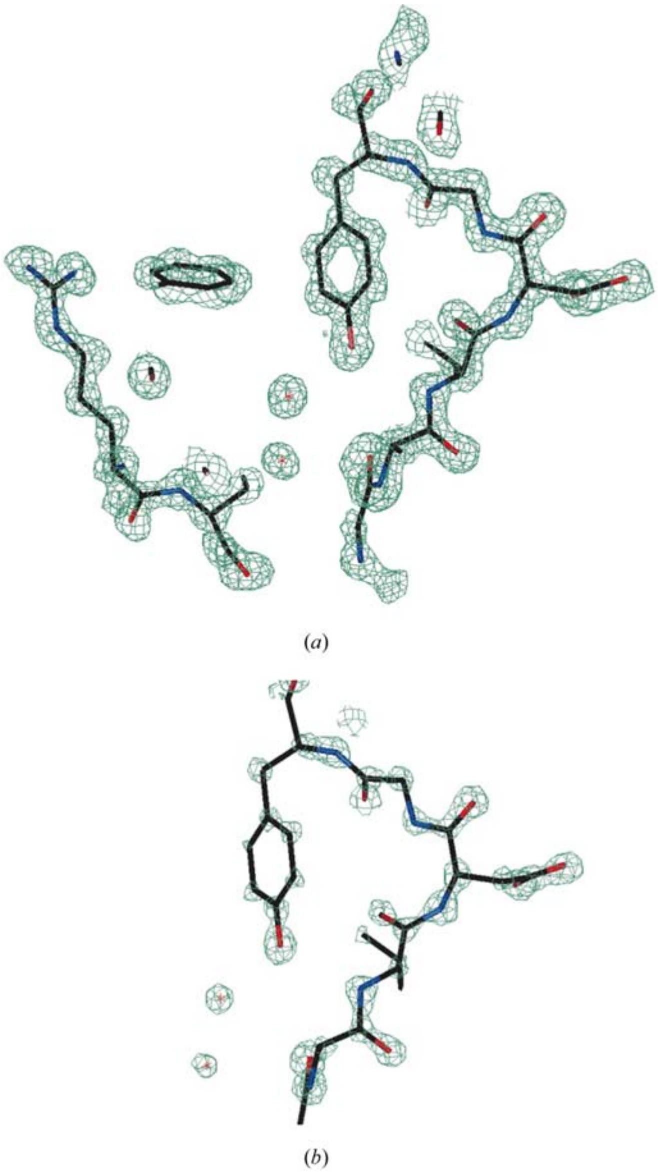


Figure 3
(*a*) Electron density calculated from three-wavelength MAD experimental phases contoured at the 2σ level, superimposed onto the final model. (*b*) The same map, but contoured at 3.8σ. The stronger electron density indicates N and O atoms as well as π-electrons.

Good results could be obtained from SAD phasing, with the remote wavelength giving the best experimental phases (Table 2). The autotracing test in the experimental and solvent-flattened map gave 1/3 and 2/3 of the model, respectively, thus indicating good connectivity in the density (Table 2). Phasing with only the inflection-point wavelength data, where the anomalous signal is lower, failed.

MAD and SAD experiments resulted in a different distribution of phase errors over the resolution range. In the SAD cases, the low-resolution phases were better defined, while at high resolution the difference to the final model phases increased rapidly (see Fig. 4). The lower accuracy of the anomalous signal in the high-resolution shells did not allow reliable phases to be obtained from a SAD experiment alone. For MAD, the experimental phases were of very good quality throughout the whole resolution range. The map correlation decreases slightly with resolution, but does not show the rapid fall-off which is seen for the SAD cases. However, in a region from 1.4 to 1.2 Å ($1/d^2$ = 0.51–0.69 Å^2) the map correlation for the remote-wavelength SAD is better than for the multiple-wavelength cases (Fig. 4). This, in combination with the differences at low resolution, explains the different appearance of the maps in visual inspection in spite of the fact that they have similar overall map correlation to the final map (see Table 2).

In order to obtain an improved signal-to-noise ratio for the anomalous signal, averaging of peak- and remote-wavelength data was tried to simulate increased redundancy. The resulting data set was used as a SAD case, but gave only marginally better phasing statistics than for peak-wavelength SAD (a map correlation to the final map of 0.58 before and 0.65 after solvent flattening) and poorer electron-density maps (visual test) than remote- and peak-wavelength SAD.

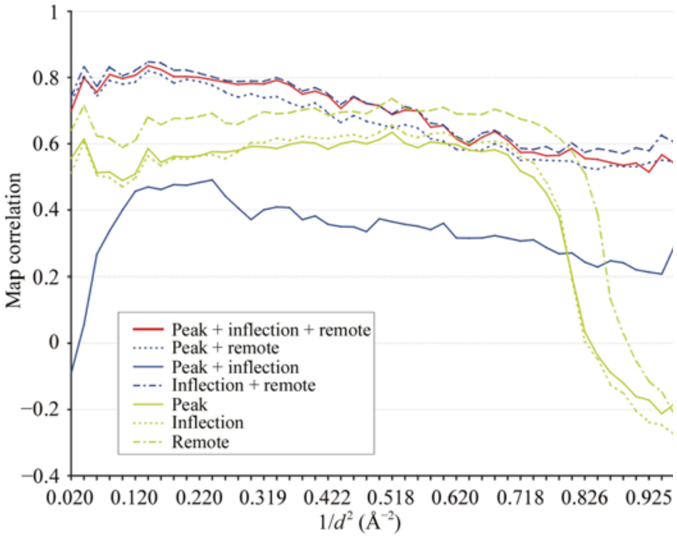


Figure 4
Correlation coefficients of the map calculated from experimental phases to the final density map as a function of resolution for different structure-solution approaches.

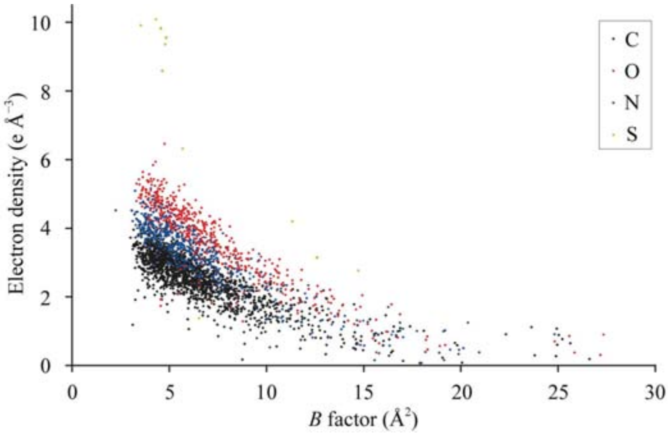


Figure 5
Observed electron density in the experimental map at the atomic centre as a function of the atomic displacement parameters in the final structure.

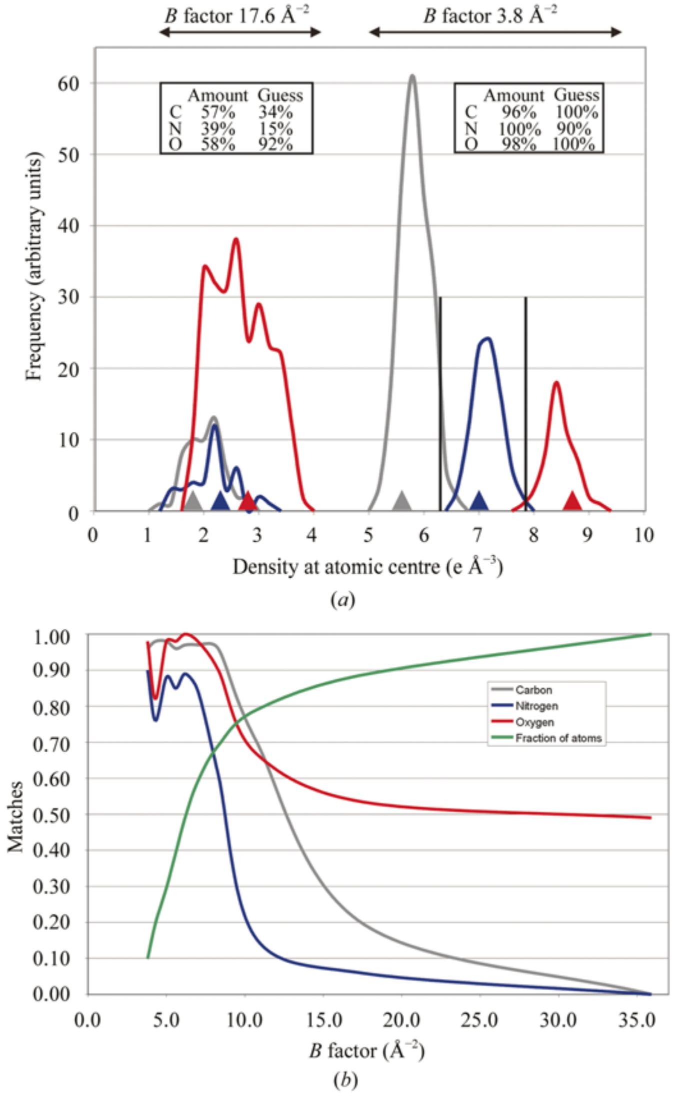


Figure 6
Atom-type assignment. (*a*) Electron-density distribution for atom types with a given *B* factor. C, grey; N, blue; O, red. The triangles show the theoretical values. The vertical black lines show the C/N and N/O boundaries. Peaks are well separated and allow reliable identification of atom types at low temperature factors. (*b*) Fraction of identified atoms *versus* *B* factor. The limit for correct assignment is a *B* factor of 10 Å^2.

Application of a σ-cutoff of 4 to the data for the map calculation did not enhance the signal for H atoms or other weak features. The map calculated with all data was always better. We suppose that a high σ-cutoff reduces the effective resolution of the data and the resulting map accumulates additional series-termination errors.

For most phasing methods, the majority of the reflections have unimodal probability distributions for their experimental phases (see Table 2). For the three best cases, three-wavelength MAD, two-wavelength MAD peak/remote and peak/inflection, more than 90% of the reflections have a unimodal phase-probability distribution.

For the single-wavelength cases, about 1/3 of the reflections are unambiguously phased experimentally. Only few of the bimodal distributions are symmetrical. Solvent flattening had a large effect.

3.4. Model building and refinement

Automatic building and refinement of the protein structure was straightforward. Close inspection of the protein model showed that histidine, asparagine or glutamine side chains had been modelled in their correct orientations, according to peak heights in the electron density, hydrogen-bonding patterns and *B* factors of the involved atoms.

Table 2
Results from different structure-solution approaches.

Map quality was ranked +++, excellent; ++, good; +, interpretable; +−, hardly interpretable. 1, remote; 2, peak; 3, inflection.

Method	SAD 1	SAD 2	SAD 3	MAD 12	MAD 13	MAD 23	MAD123
Phasing power, centric	1.27	1.24	1.20	1.36	1.43	0.29	1.42
Phasing power, acentric	2.21	2.33	2.25	1.50	1.53	0.24	1.51
FOM, centric	0.31	0.28	0.23	0.48	0.46	0.12	0.46
FOM, acentric	0.28	0.23	0.17	0.49	0.46	0.15	0.44
Unimodal reflections (%)	38	33	26	92	94	87	94
Correlation to final map	0.64	0.53	0.53	0.74	0.77	0.36	0.77
Correlation (+ *DM*) to final map	0.72	0.58	0.63	0.80	0.81	0.65	0.81
Map quality	+	+	+−	+++	+++	+−	+++
Map quality (+ *DM*)	++	++	+	+++	+++	+−	+++
Residues traced in map	111	22	20	277	261	18	257
Residues traced in solvent-flattened map	252	146	75	318	309	130	300

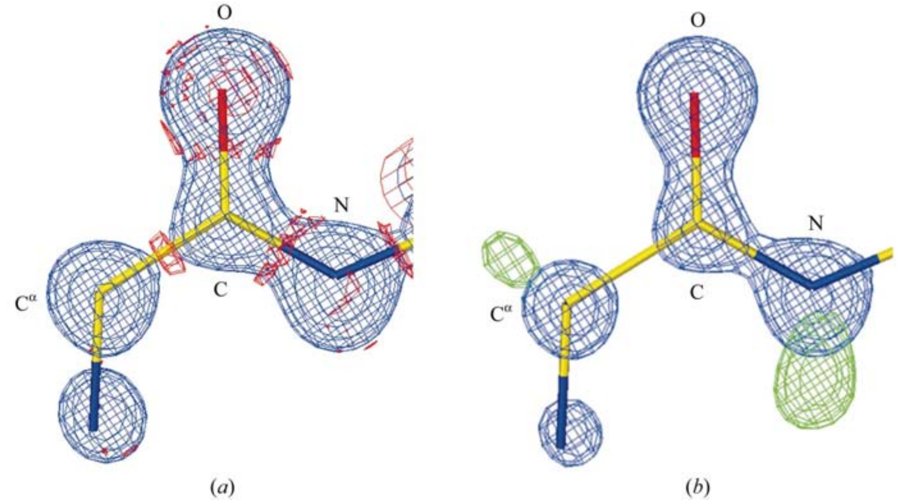


Figure 7
Averaged density over peptides. (*a*) Blue: average density. Red: uncertainty in the average density which can be attributed to anisotropic motion and insufficient modelling of bonding electrons. (*b*) Blue: average density. The stronger electron density indicates the π-character of the peptide bond is shown. Green: average difference map, indicating the H atoms.

The final *R* factor after anisotropic refinement was 0.132 (free *R* factor = 0.149) for 2818 protein atoms, 760 waters, six Hg atoms and eight Cl atoms.

The assignment of atom types in side chains and determination of side-chain orientation is normally performed by investigation of the hydrogen-bonding patterns around these residues. Inclusion of the high-resolution data in phasing made unambiguous assignment of most side chains possible. In the map calculated from the three-wavelength MAD experimental phases (Fig. 3), atom types could be distinguished in the protein as well as in the solvent regions (see Fig. 5), even if the difference in the number of electrons was small, *e.g.* for oxygen, nitrogen and carbon. The accuracy of the assignment depends mainly on the temperature factors of the atoms involved, but can be considered to be very high for atoms with *B* factors lower than 10 Å^2 (see Fig. 6). The electron-density distributions for a given atom type with a certain *B* factor are narrow and well separated (Fig. 6*a*) at low temperature factors. The slight shift to density higher than the theoretical value for C and N atoms may be explained by the fact that these atoms mostly belong to the main chain of the protein and therefore have stronger electron density. More than 90% of C, N and O atoms can be identified correctly up to a *B* factor of 5 Å^2; 70% are still assigned their correct type at a *B* factor of 10 Å^2 (Fig. 6*b*). Altogether, this accounts for about 75% of the total number of atoms in the structure that can be unambiguously assigned, including the first solvent shell. This implies that fine electronic detail and small differences in the electron density can be distinguished in essentially the whole structure.

H atoms, however, could not be assigned reliably. This might be a consequence of the fact that data was only available to 1.0 Å, with the data in the highest resolution range already showing some noise. Some peptide H atoms did appear in the final difference maps after refinement.

Peptide bonds show increased electron density covering the amide groups (Fig. 3), which is further pronounced in averaged peptide maps (see Fig. 7). Continuous density covering the peptide bond can be observed, which proves that different bond types can be distinguished throughout the whole main chain of the protein. This information is particularly valuable for assessment of the electronic state.

Peptide H atoms could be located in the averaged difference electron-density maps (see Fig. 7). The peak volume for H atoms in the average peptide density calculated from the final map corresponds to 0.4 e and is at a distance of 0.8 Å from the peptide nitrogen. The presence of these peaks in the difference map averaged over 340 peptides indicates that the signal from the H atoms is present throughout the polypeptide chain. Discrepancies in the averaged peptide maps can be attributed to anisotropic motion of the main-chain O atoms (see Fig. 7) and the insufficient modelling of bonding electrons.

In other cases, the protonation state can be deduced. Aspartic acid side chains show differences in the electron-density distributions over their carboxylic groups (Fig. 8). Some appear symmetric, indicating the deprotonated state (aspartate). The deprotonated carboxylate is a resonance hybrid, which makes the two O atoms indistinguishable and delocalizes the charge. Other aspartic acid side chains show clear asymmetry in the electron density of their side-chain functional groups, thus indicating the presence of a double bond between the C and one of the O atoms. The proton would then be attached to the other O atom and be involved in hydrogen-bonding contacts, therefore making the acidic group asymmetric. Inspection of the immediate surroundings of these residues confirm these observations. Some of the aspartates (or glutamates) are involved in salt bridges or are solvent-exposed, while the aspartic acid residues are located mostly in a neutral environment and are engaged in a very intricate hydogen-bonding pattern.

π-Electrons could be seen in aromatic side chains (Fig. 8). In tryptophan residues, the aromatic character of the side chain can be seen from the increased electron density in the rings (Fig. 8), indicating the presence of delocalized π-electrons. Another example is the imidazole side chain of histidine where, upon protonation, the change in the distribution of electrons in the ring should be detectable.

Multiple conformers in the protein could be observed in the experimental map and were inserted into the protein model at the side-chain-building stage. This strategy has been employed in the solution of the structure of rat

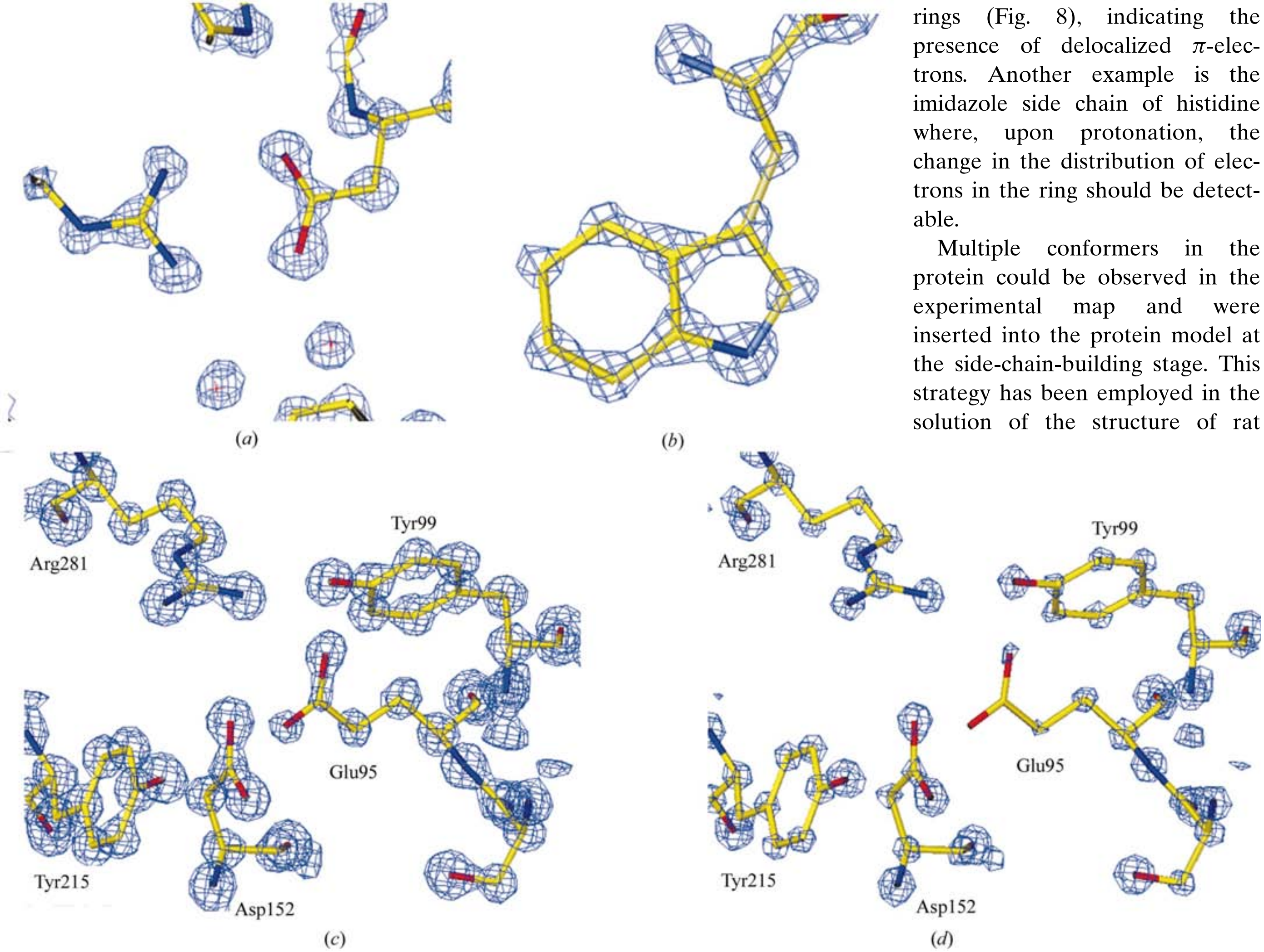


Figure 8
Some features of the electron-density maps contoured at 3σ calculated with experimental phases from three-wavelength MAD. (*a*) An aspartate involved in a salt bridge shows symmetric electron density for its carboxyl group, as expected for the deprotonated state. (*b*) A tryptophan side chain with increased electron density in its aromatic rings caused by the delocalized π-electron system. (*c*) and (*d*) View of the active site of CelA, map contoured at 2σ (*c*) and 3σ (*d*). The residues involved in catalysis are Glu95 and Asp152. Both show asymmetric electron density in their side-chain carboxylic groups.

mannose-binding protein A (Burling *et al.*, 1996). Typically, side chains are fitted in one conformation first, then refined and the new structure and new electron-density map checked for additional features. This can introduce bias in the refinement, as the already present conformer will always be well defined in the maps, while its counterpart might only appear weakly. By building both conformers into the density at a very early stage of the refinement, it is possible to account properly for the relative occupancies of both conformers, thus enhancing the accuracy of the model.

In the present CelA structure, the positions of the active-site residues, Glu95, Asp152 and Asp278 (Alzari *et al.*, 1996), are very accurately defined. Glu95 shows asymmetric electron density in its side-chain carboxyl group, which indicates the protonated state of this residue (Fig. 8). Glu95 was proposed to be the proton donor during catalysis (Alzari *et al.*, 1996). For Asp152 and Asp278, the assignment of the protonation state becomes even more important. Both residues were proposed to play the role of the nucleophile in glycolysis (Alzari *et al.*, 1996). For Asp152, asymmetry in the carboxyl group can be observed (Fig. 8), which indicates protonation and makes it less likely that this residue has the function of the nucleophile in the enzymatic reaction. Asp278 appears symmetric and is therefore the better candidate for the proposed role in catalysis. Asp278 shows two equal conformations of its side chain which could be observed even in the map from experimental phases. This disorder is not observed in the 1.65 Å CelA structure and seems to be lost completely upon contact with the substrate (Alzari *et al.*, 1996), which also indicates this residue's involvement in substrate cleavage. The substrate-binding channel is filled with solvent water molecules, some of which are involved in an intricate hydrogen-bonding network with the proposed catalytic residues.

In a recently determined structure of a complex of a CelA E95Q mutant with cellopentaose the roles of Glu95 and Asp278 have been confirmed. In this structure, Gln95 forms a strong hydrogen bond to the glycosidic O atom in the substrate, while Asp278 was found to be the general base activating the nucleophilic water (Guerin, 2002).

3.5. Comparison with the previously determined CelA structure

The main differences between the previously determined 1.65 Å CelA structure (Alzari *et al.*, 1996) and the structure obtained from atomic resolution MAD are observed in the solvent regions and weakly defined residues, *e.g.* multiple conformers. In the protein itself there are hardly any changes. The two structures show an overall average r.m.s.d. of 0.26 Å for the main-chain atoms and 0.46 Å for the side-chain atoms. These rather large deviations are attributed to the local changes around the heavy-atom binding sites, which cause shifts in the surrounding loops.

The positions of the Hg atoms are shown in Fig. 9. All Hg atoms show different coordination in terms of the number of ligands as well as the geometry. Two of them (Hg1 and Hg3) show square-planar coordination with the Hg atom slightly out of plane, while two others (Hg2 and Hg5) show pyramidal geometry. In one of the pyramids (around Hg2) the ligand at one corner is substituted by a hydrophobic residue, Ile211. Independent of the overall geometry of the coordination, the chloride ion is always located opposite to the cysteine S^{γ}, hence forming a straight line with the central ion. The remaining contacts are mostly made by main-chain O atoms. The geometry around Hg3 is especially interesting, as one of the four ligands is the aromatic ring of Phe184 which seems to function as an η_6 ligand. The mercury ion sits almost perfectly centrally above the ring at a distance from the plane of ~3.5 Å and therefore interacts with the π-electron system of the phenyl ring. This is unusual, as mercury does not normally have π-acceptor properties. Hg4-1 and Hg4-2 which form the split site both show linear coordination by two ligands only. The distance between the two partially occupied mercury sites is 2.25 Å. Each of these mercury ions has one chloride ion as a ligand opposite to the cysteine SH group in a very hydrophobic environment.

The most obvious difference in the solvent region is the number of observed water molecules. In the atomic resolution structure, 760 solvent molecules were included in the model, while the comparison structure contained only 266. The positions of mercury and chloride in the MAD structure are

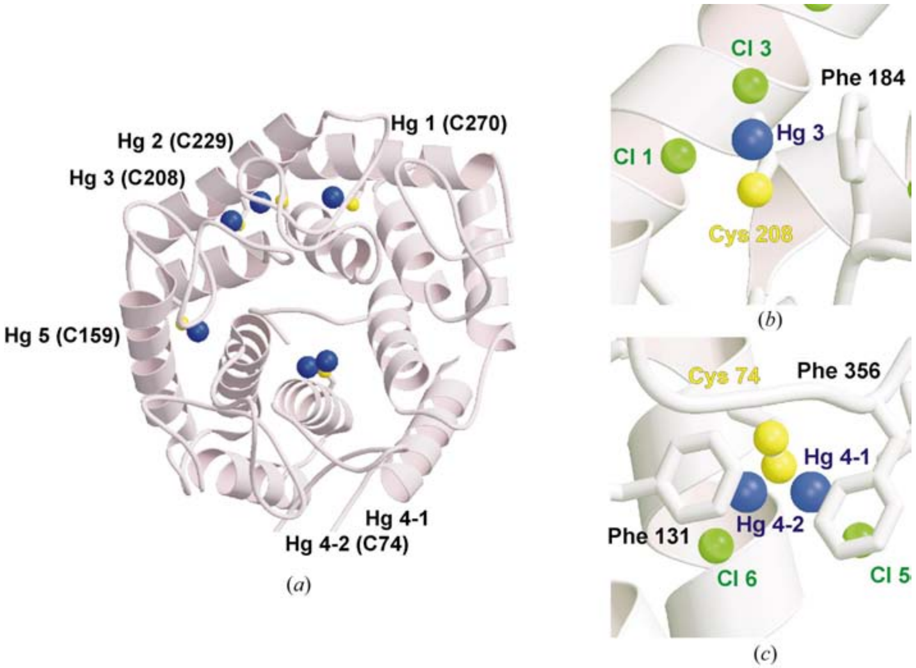


Figure 9
Mercury sites in CelA. (*a*) Overview. (*b*) Close-up of site Hg3 coordinated by a phenyl ring. (*c*) Close-up of site Hg4-1/4-2 in a double conformation.

not occupied by water molecules in 1cem; these sites are empty. Two shells of solvent water molecules can be found in both crystal structures. Shell1 is defined as water molecules directly attached to the protein molecule, while shell2 contains those water molecules attached to the water molecules of shell1. In 1cem, most of the water molecules, 234 out of 266 (88%), belong to shell1, while in the MAD-structure these numbers change to 434 out of 760 (57%). A third solvent shell (shell3), which contains all the solvent outside shell2, can only be assigned in the atomic resolution structure; it is missing completely in 1cem.

4. Conclusions

Atomic resolution MAD structure solution and refinement resulted in a highly accurate protein structure. A two-wavelength experiment is easily sufficient to obtain excellent experimental phases, thus allowing the data-collection time to be shortened. Fast and straightforward automated model building and refinement accomplished the structure determination. Differences in atom types and bond types were detected and accounted for early on, during the phasing step. Valuable information on the protonation state and the electronic state of residues of interest could be extracted even from the experimental maps. If applied to protein–substrate complexes, this approach could lead to a very accurate picture of the protein–substrate and protein–solvent interactions.

References

Alzari, P. M., Souchon, H. & Dominguez, R. (1996). *Structure*, **4**, 265–275.

Burling, F. T., Weis, W. I., Flaherty, K. M. & Brunger, A. T. (1996). *Science*, **271**, 72–77.

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

Cowtan, K. D. & Main, P. (1996). *Acta Cryst.* D**52,** 43–48.

Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* D**56**, 232–237.

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

French, G. S. & Wilson, K. (1978). *Acta Cryst.* A**34**, 517–525.

González, A., Pedelacq, J.-D., Sola, M., Gomis-Ruth, F. X., Coll, M., Samama, J. P. & Benini, S. (1999). *Acta Cryst.* D**55**, 1449–1458.

Guerin, D. M., Lascombe, M.-B., Costabel, M., Souchon, H., Lamzin, V. S., Beguin, P. & Alzari, P. M. (2002). *J. Mol. Biol.* **316**, 1061–1069.

Hendrickson, W. A. & Lattman, E. E. (1970). *Acta Cryst.* B**26**, 136–143.

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

Lamzin, V. S., Morris, R. J., Dauter, Z., Wilson, K. S. & Teeter, M. M. (1999). *J. Biol. Chem.* **274**, 20753–20755.

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

Nagem, R. A. P., Dauter, Z. & Polikarpov, I. (2001). *Acta Cryst.* D**57**, 996–1002.

Otwinowski, Z. & Minor, W. (1996). *Methods Enzymol.* **276**, 307–326.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Read, R. J. & Schierbeck, A. J. (1988). *J. Appl. Cryst.* **21**, 490–495.

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

Souchon, H., Beguin, P. & Alzari, P. M. (1996). *Proteins*, **25**, 134–136.

Ten Eyck, L. (1973). *Acta Cryst.* A**29**, 183–191.

Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.