Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Use of an ion-binding site to bypass the 1000-atom limit to structure determination by direct methods

Blaine H. M. Mooers and Brian W. Matthews*

Institute of Molecular Biology, Howard Hughes Medical Institute and Department of Physics, 1229 University of Oregon, Eugene, OR 97403-1229, USA

Correspondence e-mail: brian@uoxray.uoregon.edu

Received 2 April 2004
Accepted 12 July 2004

PDB References: T4 lysozyme, 1sx7, r1sx7sf; 1swz, r1swzsf; 1sx2, r1sx2sf; 1swy, r1swysf.

Proteins with more than 1000 non-H atoms and without heavy-atom prosthetic groups are very difficult to solve by *ab initio* direct methods. T4 lysozyme is being used to explore these limits. The protein has 1309 non-H atoms, seven S atoms, no disulfide bonds and no heavy-atom prosthetic group. It is recalcitrant to structure determination by direct methods using X-ray diffraction data to 0.97 Å. It is shown here that it is possible to obtain a truly *ab initio* structure determination of a variant of the protein that has an Rb^+ ($Z = 37$) binding site. Using diffraction data to 1.06 Å resolution, the direct-methods programs *SIR*2002 and *ACORN* independently solved the structure in about 20 h. The bound Rb^+, which contributes about 1.7% of the total scattering, does not appear to distort the structure or to inhibit refinement (R factor 12.1%). The phases obtained *via* *SIR*2002 or *ACORN* are in good agreement with those from a reference structure obtained from conventional molecular-substitution and refinement procedures (average error in the figure-of-merit-weighted phases of less than 25°). Thus, proteins with more than 1000 atoms that include halide-binding or other such sites may be amenable to structure determination by *ab initio* direct methods. The direct-methods approaches are also compared with structure determination *via* use of the anomalous scattering of the Rb^+ ion. As shown by examples, high-resolution structures determined by direct methods can be useful in highlighting regions of strain in the protein, including short hydrogen bonds and non-planar peptide groups.

1. Introduction

The use of direct methods to determine crystal structures with more than 200 non-H atoms was very difficult prior to the development of dual-space recycling methods (Miller *et al.*, 1993; Weeks *et al.*, 1993). Subsequently, several small protein crystal structures have been solved (Weeks *et al.*, 1995; Deacon *et al.*, 1998). Proteins with more than 1000 atoms remain problematic even when atomic resolution data are available (Uson & Sheldrick, 1999). However, when sufficiently electron-dense heavy atoms are present, *e.g.* within a prosthetic group, proteins of several thousands of atoms have been determined (Jia-xing *et al.*, 2002). By *ab initio* direct methods we mean structure determination directly from the native structure factors without use of isomorphous replacement, anomalous dispersion or molecular replacement and with no prior information on any atomic position. The only allowed prior information is the general chemical features of proteins and the elemental composition of the unit cell.

In principle, it may be possible to introduce electron-dense atoms to extend the application of *ab initio* direct methods to

proteins with more than 1000 atoms and to proteins with less than atomic resolution data. Ideally, these atoms should contribute at least 1–2% of the total scattering power in order to ensure that their positions can be found by direct methods. On the other hand, the contribution to the overall scattering should probably be less than 5% in order to avoid having phase sets that overemphasize the heavy-atom contribution and possibly lead to loss of enantiomorphic information (Sheldrick *et al.*, 1993). Based on these considerations, a protein of 10 000 atoms might be solved by *ab initio* direct methods with three to seven Cs^+ ($Z = 55$) binding sites and with atomic resolution data and in principle there may be no limit to the size of proteins that can be determined in the presence of an appropriate amount of heavy-atom scattering. In addition, the presence of electron-dense atoms can relax the resolution required for an *ab initio* direct-methods determination to as low as 1.55 Å in select cases (Mukherjee *et al.*, 2000, 2001; Burla, Carrozzini, Caliandro *et al.*, 2003; Burla, Carrozzini, Cascarano *et al.*, 2003).

It has been shown that halide-binding sites can sometimes be used to facilitate structure determination by isomorphous-replacement and/or anomalous-scattering methods (Dauter & Dauter, 1999; Dauter *et al.*, 2000). Methods for introducing such sites have also been developed (see, for example, Boggon & Shapiro, 2000; Sun *et al.*, 2002). Here, we use T4 lysozyme to show how the presence of such a heavy-atom binding site can also enhance the power of direct methods.

Wild-type T4 lysozyme (WT) has 1309 non-H atoms and its structure is well known (Weaver & Matthews, 1987). It has seven S atoms (none in disulfide bridges) and no prosthetic group. Each sulfur contributes less than 0.4% of the total scattering, which makes these sites very difficult to locate. Using data to 0.97 Å resolution, we have not been able to obtain a successful *ab initio* structure determination. The following mutant of T4 lysozyme makes it possible to explore the effectiveness of a somewhat more electron-dense site.

Wild-type T4 lysozyme can bind Rb^+ ($Z = 37$) near Glu11 in the active-site cleft with an occupancy of about 0.4, which gives a scattering power similar to that of a single S atom. By changing Arg96 to glutamic acid and Asp72 to alanine, a second Rb^+-binding site is created on the surface of the protein (B. H. M. Mooers & B. W. Matthews, unpublished results). This rubidium ion appears to bind with full occupancy and does not lead to distortion of the protein structure. It is estimated to contribute about 1.7% of the total scattering. As shown below, the presence of this ion, together with the use of data to 1.06 Å resolution, makes it possible to obtain an *ab initio* structure determination. The quality of different *ab initio* structure determinations are evaluated by comparison with an independently refined model of the protein.

The technique of using electron-dense atoms (added by co-crystallization, soaks or selenomethionine incorporation) in direct methods is called augmented direct methods (ADM). ADM may find wider application in systems where there is a need to obtain electron-density maps at atomic resolution that are determined *ab initio* or are free of possible bias from a molecular-replacement model.

Table 1
X-ray data-collection statistics.

The values for the highest resolution shell are given in parentheses.

Crystal	WT	D72A/R96E with RbCl
X-ray source	SSRL BL 7-1	SSRL BL 9-1
Unit-cell parameters		
$a = b$ (Å)	60.04	60.20
c (Å)	95.94	95.24
Wavelength (Å)	1.08	0.773
Resolution range (Å)	20.1–1.06 (1.12–1.06)	23–1.06 (1.12–1.06)
Crystal-to-detector distances (mm)	95, 124, 224	160, 148, 235, 265
No. reflections	839470 (73162)	982963 (38027)
No. unique reflections	73162 (12966)	87349 (11950)
Multiplicity	9.2 (5.6)	11.1 (3.2)
Completeness (%)	99.8 (99.8)	96.3 (91.4)
$R_{\rm merge}$ (%)	0.044 (0.211)	0.052 (0.304)
$R_{\rm anom}$ (%)	N/A	0.015 (0.167)
$\langle I/\sigma(I)\rangle$	7.9 (3.7)	7.4 (4.0)
$B_{\rm Wilson}$ (Å^2)	9.4	8.7

2. Materials and methods

2.1. Mutagenesis, protein expression and purification

As discussed in §3, the double mutant D72A/R96E in T4 lysozyme creates a halide-binding site. As such, it is a convenient subject for the studies described here. The mutant D72A/R96E was made by an iterative two-stage polymerase chain reaction (PCR; Landt *et al.*, 1990) based on the gene for wild-type T4 lysozyme (WT). The *Bam*HI/*Hin*dIII-digested PCR product was ligated into the vector PH1403 and the DNA sequence was confirmed by automated methods incorporating PCR (Perkin–Elmer ABI PRISM 377 DNA sequencer). The vector was transformed into *Escherichia coli* RR1 cells and the mutant protein was overexpressed and purified by standard methods (Alber & Matthews, 1987; Muchmore *et al.*, 1989; Poteete *et al.*, 1991). The buffer used for protein storage was 0.1 *M* sodium phosphate pH 6.5, 0.55 *M* NaCl and 0.02% NaN_3. Since the double mutant caused cell lysis and behaved similarly during purification, we assume that it has activity similar to that of WT.

2.2. Crystallization

The mutant protein D72A/R96E in 0.55 *M* NaCl was crystallized using 2 *M* potassium/sodium phosphate at about pH 7.0 as previously described (Eriksson *et al.*, 1993). The space group is $P3_221$ and the crystals were isomorphous with WT. Crystallization experiments were repeated with solutions containing 0.55 *M* RbCl in place of NaCl. The resulting crystals were also isomorphous with the wild type.

2.3. X-ray data collection

An Rb^+-containing crystal of D72A/R96E measuring $0.3 \times 0.4 \times 0.6$ mm was mounted with Paratone-N in a rayon cryo-loop and flash-cooled to 100 K. X-ray data as described in Table 1 were collected with the c axis aligned within 5° of the spindle axis in order to collect anomalous pairs on the same image. To compensate for saturation of the detector by

Table 2
Refinement statistics for different models of D72A/R96E crystallized in the presence of rubidium.

Source of model	Molecular substitution (reference structure)	*SIR*2002	*ACORN*	SAD
Resolution range (Å)	22–1.06	22–1.06	22–1.06	22–1.06
R factor [$F > 4\sigma(F)$/all data]	0.121/0.130	0.119/0.128	0.116/0.125	0.114/0.123
$R_{\rm free}$ [$F > 4\sigma(F)$/all data]	0.143/0.152	0.143/0.152	0.140/0.149	0.139/0.148
No. non-H atoms	1309	1309	1309	1309
No. solvent atoms	317	319	319	319
No. observations	83026	83026	83026	83026
No. restraints/parameters	18668/14929	18653/15310	18453/15260	18944/15396
R.m.s.d. from ideal geometry				
Bond lengths (Å)	0.015	0.015	0.015	0.015
Angle distances (Å)	0.029	0.030	0.031	0.030
PDB code	1sx7	1swz	1sx2	1swy

intense low-resolution reflections, sets of images were collected at four increasing distances from the crystal with progressively shorter exposure times (Table 1). To improve the multiplicity of the diffraction data at high resolution, 181° of images were collected in the second sweep, which had a resolution limit of 1.15 Å. The data were integrated with *MOSFLM* and scaled by *SCALA* (Leslie, 1992; Evans, 1994) with the anomalous pairs separated.

Similar procedures were used to collect data for WT except that the crystal was not aligned to optimize collection of anomalous pairs and such pairs were merged during the scaling of the intensities.

2.4. Structure determination

As discussed in §3, the structure of D72A/R96E was solved by *ab initio* direct methods using *ACORN* (Foadi *et al.*, 2000; McAuley *et al.*, 2001) and *SIR*2002 (Burla *et al.*, 2002). The phases from *ACORN* and *SIR*2002 were read into *ARP*/*wARP* (Perrakis *et al.*, 1999) and models were built automatically. These models were edited using the molecular-graphics program *XFIT* from the *XtalView* package (McRee, 1992). They were then used to start refinement with *SHELXL* (Sheldrick & Schneider, 1997).

3. Results and discussion

3.1. Engineering of the rubidium-binding site

It was found that the double mutant D72A/R96E in T4 lysozyme creates a halide-binding site. The mutant is soluble and active, although less stable than the wild-type protein. It crystallized isomorphously with the wild-type protein using 2 *M* phosphate solutions containing 0.55 *M* RbCl in place of the standard 0.55 *M* NaCl. [Lower concentrations of Rb^+ were not tested, although it may be noted that Korolev *et al.* (2001) found that 200 m*M* was sufficient to obtain binding in crystals of other proteins.] Large (0.7 × 0.5 × 0.3 mm) crystals grew in a month at 277 K. Difference maps (not shown) confirmed the presence of the bound Rb^+ ion. Subsequent refinement (Table 2) indicated that the occupancy of the ion is 100% and its B factor is 6.6 $Å^2$.

3.2. Determination of a reference structure

As a control and to provide an independent measure of the mean phase error, the structure of D72A/R96E was determined by molecular substitution using the high-resolution refined coordinates of the wild-type protein (B. H. M. Mooers & B. W. Matthews, unpublished results) as the starting model. The model was refined to 1.2 Å resolution using *TNT* (Tronrud *et al.*, 1987; Tronrud, 1997) following the procedures described previously (Eriksson *et al.*, 1993). The structure was refined further to 1.06 Å resolution using *SHELXL* (Sheldrick & Schneider, 1997). The R factor was 12.1% for data with $F > 4\ \sigma(F)$ and 13.0% for all data (Table 2). This is referred to as the reference structure and the phases and coordinates were used as a basis to evaluate the *ab initio* models.

3.3. Phasing by *ab initio* direct methods

3.3.1. Phasing with *SIR*2002. *SIR*2002 (Burla, Carrozzini, Caliandro *et al.*, 2003; Burla, Carrozzini, Cascarano *et al.*, 2003) is a descendent of the traditional direct-methods program *SIR* (semi-invariants representations) that uses the theory of representations to efficiently estimate structure invariants and semi-invariants (Giacovazzo, 1977). *SIR*2002 has extended capabilities including real-space methods that allow the determination of protein structures with minimal user intervention (Burla *et al.*, 2002; Burla, Camalli *et al.*, 2003). *SIR*2002 starts a trial with a random set of phases, which are refined with the triple tangent formula using the reflections with the 4000 largest E values ($E > 1.81$). At the start of the next trial, a new set of random phases is selected. These phases are extended and refined using real-space techniques. *SIR*2002 differs from the classical dual-space recycling methods (*Shake-and-Bake* and *Half-Baked*) by not returning to reciprocal space after starting density modification. A discriminating figure of merit, called RAT (Burla *et al.*, 2002), is used to select promising phase sets. If a set of phases has a low RAT value, the current trial is terminated and a new trial is started with a new set of phases. RAT is computed using the observed and calculated $|E|$ values with the current calculated phases. The 'true' phases from the reference structure were not used in the computation of RAT.

In our case, *SIR*2002 was tested using rubidium-containing T4 lysozyme with data to 1.06 Å resolution (Table 1). The fifth phase set had a RAT value of 3.2, whereas the first four phase sets had values of about 1.0. Next, the atoms (with assignment of element types based on the electron density) were refined for four cycles of least-squares refinement with isotropic B factors and without stereochemical restraints (because the atoms had not been assigned to residue type). The final R factor was 23.9%, suggesting that this was the correct solution. In contrast, incorrect solutions had final R factors of about

33% or higher. A Compaq Alpha Workstation took 22 h to reach the end of the fifth trial. In later trials, structure determination was attempted using the *RELAX* directive, which expanded the data to *P*1 symmetry for promising trials and then returned the structure to $P3_221$ symmetry once the structure has been determined. A correct solution was found on the fifth trial and the process took approximately the same amount of time (18.6 h).

*SIR*2002 wrote out the phases and the atomic coordinates and atom type without assignment of residue type. A map made with coefficients mF_o revealed that the alternative origin (0.0, 0.0, 0.5) had been used (Giacovazzo, 1998). The coordinates were therefore translated to facilitate comparison with the reference structure. The coordinates were also scattered across several asymmetric units and had to be gathered into a connected model. After shifting the phases for the change in origin, they were compared with those from the reference structure (Figs. 1*a* and 1*b*). The figure-of-merit (fom) weighted mean phase error (wMPE) has a minimum of about 15° at medium resolution and increases steadily to about 30° at the high-resolution limit (Fig. 1*b*). The final phase set from a seeded solution that was developed from initial phases derived from the refined Rb^+ position had a wMPE that was essentially identical to that of the *ab initio* solution (Fig. 1*b*).

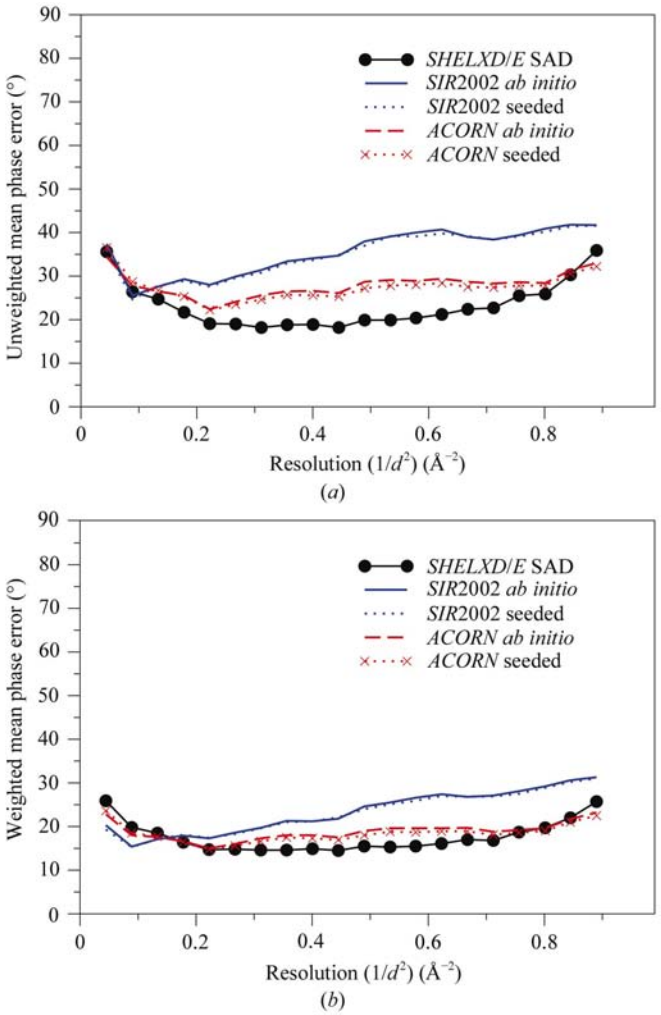

Figure 1
(*a*) Phase error as a function of resolution for *ab initio* and seeded direct-methods solutions obtained with *ACORN* and *SIR*2002. A comparison with the reference structure is also included for phases obtained by SAD using *SHELXD*/*E*. (*b*) Comparison of phase error, as in (*a*), but weighted by the figure of merit of the phase determination. Computations were made using *SHELXPRO* v.2000-1.

Direct-methods structure determinations are not expected to succeed with 'small-molecule' structures if fewer than half of the reflections in the resolution range 1.1–1.2 Å are observed with $|F| > 4\sigma(|F|)$ (Sheldrick, 1990). This rule allows a relaxation of the requirement for 1.2 Å data when heavy atoms are present. This empirical rule has applied to proteins fairly well until recently. An advanced version of *SIR*2002 (*SIR*2003-N) was able to determine the structure of an all-light-atom protein that has 795 non-H atoms and 1.45 Å data (Burla, Carrozzini, Caliandro *et al.*, 2003). *SIR*2003-N introduces the use of molecular-envelope calculations along with other enhancements. Future improvements in direct-methods programs may allow the structure determination of small all-light-atom proteins with data having resolution limits below 1.2 Å, while larger proteins may require progressively higher resolution data. This may be especially true of methods that use tangent-formula refinement, because these methods are very sensitive to the distribution of the *E* values.

The diffraction data for D72A/R96E have 67.4% of the theoretical number of unique reflections in the resolution range 1.1–1.2 Å with $|F| > 4\sigma(|F|)$. Hence, these data pass Sheldrick's empirical rule.

We tested the influence of resolution on the ability to obtain a solution with *SIR*2002 by truncating the data to 1.07 Å. The entire process was repeated in *SIR*2002 starting with Wilson scaling and computation of the normalized structure-factor magnitudes. This trial ran for 16 d over 84 cycles without obtaining a solution. This suggests that a very small increase in resolution, here from 1.07 to 1.06 Å, can be the difference between failure and success. This is not surprising since tangent-formula-based methods succeed when a sufficient number of reliable triplet invariants are found. The number of reliable triplet invariants increases with higher resolution X-ray data.

During the *ab initio* trial, *SIR*2002 was run with the following unit-cell contents: C4878/H10632/N1410/O1422/S42/Rb6. To test for the influence of the heavy atoms in the atom list, *SIR*2002 was rerun with the following unit-cell contents: C7758/H10632. This all-equal-atom trial ran for 200 cycles without reaching a solution. Apparently, the prior knowledge that atoms heavier than carbon are present dramatically reduces the number of trials required to reach a solution.

While the high-resolution limit has received a lot of attention in direct-methods work owing to the need for the data to contain sufficient information about the atomicity of the structure, little attention has been given to the low-resolution limit in structure determinations of proteins with atomic resolution X-ray data. One exception is the inclusion of molecular-envelope calculations in *SIR*2003-N, which is thought to enable this program to determine protein structures by *ab initio* direct methods with X-ray data that have

resolution limits below 1.2 Å (Burla, Carrozzini, Caliandro *et al.*, 2003). Low-resolution data are expected to be important in methods that employ density modification because their absence can lead to Fourier ripples in the electron-density map. *Ab initio* direct-methods trials were attempted with the low-resolution limit set to 5, 8, 12, 13 and 15 Å. In no case was a successful solution reached in 20 or more trials. These results show that the loss of low-resolution reflections can inhibit the success of *ab initio* direct methods.

Although, as noted above, *SIR*2002 typically starts with random phases, it is also possible to use phases generated in other ways. We therefore tested the effectiveness of starting phases based on a single atom located at or close to the Rb^+ site (this was also for comparison with *ACORN*; see below). The phases were generated from a test atom placed first at the refined Rb^+ site and then displaced in increments of 0.04 Å along Cartesian *x*, *y* and *z* axes. Phases that led to successful structure determinations had RAT values of at least 1.5 and *R* factors of less than 25.0% (Figs. 2*a* and 2*b*). Displacements up to about 0.4 Å along the *x* and *y* axes led to solutions. On the other hand, displacements greater than 0.17 Å along the *z* axis prevented structure determination. The anisotropy in the sensitivity of *SIR*2002 to error in the Rb^+-ion position may be related to a cusp of missing data about the *l* axis which arose from the rotation of the crystal around the *c* axis during data collection (see §2). In principle, these results suggest that the systematic sampling of a grid of dimensions $0.8 \times 0.8 \times 0.3$ Å with Rb^+ as a starting site would lead to a solution. In practice, since the anisotropy could not be predicted accurately in advance, a search on a finer grid might be required.

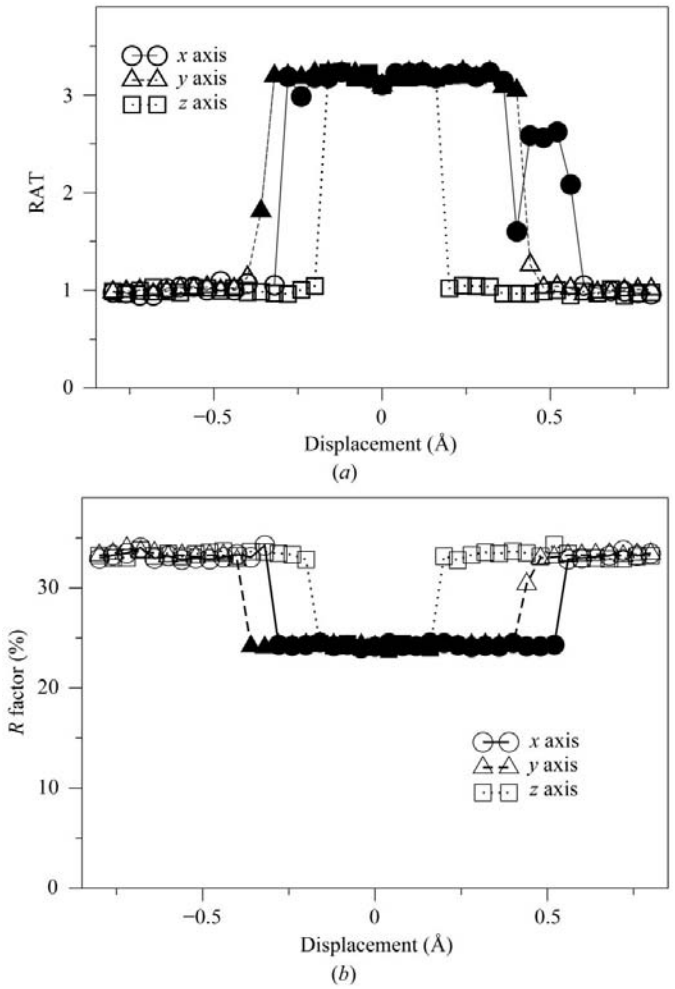

Figure 2
(*a*) The *SIR*2002 figure of merit 'RAT' (Burla *et al.*, 2002) as a function of displacement of the Rb^+ ion from its position in the refined structure along the three Cartesian axes. The Rb^+ ion was used as a test site to initiate direct-methods solutions. Successful structure determinations are indicated by solid symbols; open symbols indicate unsuccessful attempts. (*b*) The crystallographic *R* factor at the end of the *SIR*2002 structure determination as a function of Rb^+ displacement. Solid symbols indicate successful structure determinations.

3.3.2. Phasing with *ACORN*. Using the same 1.06 Å data set, an *ab initio* direct-methods solution of the Rb^+-containing lysozyme mutant was also obtained by *ACORN* (Jia-xing *et al.*, 2002). This program normally follows the strategy of finding a molecular fragment to produce non-random starting phases. In the present case, starting phases were generated from a single atom placed at random at many different sites. The random atom was treated as if it were making the contribution of sulfur. While it is possible to specify other atom types in seeded trials, this was not an option for *ab initio* trials with *ACORN*. Unlike *SIR*2002, *ACORN* ran without any prior information about the unit-cell contents. Trials with 100 000 and 250 000 random positions failed, but trials with 500 000 positions were successful (see below). The correlation coefficient (CC) between the resulting calculated and observed *E* values was computed for each position using the strongest 20% of the data. The randomly generated starting positions were sorted by CC and the CC was computed again for the top 1000 positions but this time using all of the data. On the basis of these CCs, the top 200 positions were then used to provide initial phase sets for dynamic density modification and, if necessary, Sayre equation refinement. Of these 200 positions (*i.e.* 200 phase sets), that ranked 54th led to a successful structure determination after 27 cycles of density modification. Unlike *SIR*2002, *ACORN* does not assign atoms to electron density. The calculated structure factors are instead derived directly from the modified electron density rather than from atomic coordinates. A Compaq Alpha Workstation took 22.4 h to reach and finish trial 54, similar to the time taken by *SIR*2002 to test five trials. The successful solution had an *R* factor of 30.6%, a correlation coefficient of 0.787 and had the same origin as the reference structure. The weighted mean phase error with the reference model was 19.0° and was roughly constant across all resolution shells (Fig. 1*a*).

Some of the important parameters in an *ab initio ACORN* trial include the number of random starting sites, the resolution limits and the grid size used in the FFT calculations. The distribution of the correlation coefficients for 200 top-ranked phase sets is shown in Fig. 3. Site 54, which led to the correct structure, was only 0.038 Å away from the 'true' Rb^+ position, whereas none of the other 200 candidates was within 3 Å. Analysis of the top 53 sites showed that all of these were close to the crystallographic twofold axis which lies along the line $x = y$, $z = 0$ in the $P3_221$ space group. Clearly, putative starting

sites that are on or close to special positions should be regarded with skepticism. The fact that the successful site was so close to the rubidium and no other high-correlation site was close prompted us to investigate more carefully the requirements for a successful solution.

As would be expected, larger numbers of random starting sites improve the chances of obtaining a solution when using a Rb^+ ion to generate starting phases. We obtained successful solutions for 14 out of 15 trials when using 500 000 starting sites, five out of ten trials when using 250 000 starting sites and one out of seven trials when using 100 000 starting sites. This suggests that one needs to test about 500 000 sites to ensure a high probability of having at least one close enough to the Rb^+ position to lead to a successful structure determination.

To explore the sensitivity of *ACORN* to errors in the assumed Rb^+ position, the site was translated from its position in the reference structure in increments of 0.04 Å parallel to the Cartesian axes in the positive and negative directions. For each displaced site, a starting phase set was calculated and a structure determination attempted using the procedure described above. The number of cycles to a solution increased gradually with increasing displacement from the reference position and then increased dramatically with displacements greater than 0.3 Å (Fig. 4*a*). As long as the assumed position of the starting site was within about 0.35 Å of its correct value, a successful structure determination could be achieved. This suggests that starting phases calculated from test sites placed on a grid of spacing 0.7 Å (*i.e.* 2 × 0.35 Å) should include one phase set that would lead to the correct structure. Such a grid for one asymmetric unit of T4 lysozyme would include about 167 000 sites. A more conservative approach (requiring more computer time) would be to use test sites placed on a grid of spacing about 0.4 Å. This would require testing about 900 000 starting phase sets.

Fig. 4(*b*) shows that the correlation coefficient for the starting phase set has its maximum value when the test atom is at its correct position, even though this atom constitutes only 1.7% of the scattering power of the asymmetric unit. This is true whether the CC is based on the strong $|E|$ values or the medium and weak $|E|$ values. The initial CC for the strong $|E|$ values could be as low as 0.008 and still lead to a solution.

The inclusion of high-resolution data is expected to facilitate the correct identification of the starting site for phasing. To test this, *ab initio* trials with *ACORN* were started with 1 000 000 random starting sites and with data that were gradually truncated, starting at 1.06 Å resolution and eventually stopping at 1.50 Å resolution. The entire process was repeated for each truncated data set, starting with the calculation of the normalized structure-factor magnitudes. Trials with resolution limits of 1.14 Å or lower failed because the resolution was insufficient for direct methods even though the correct site was found using data to 1.2 Å resolution. It is noteworthy that *ACORN*, which is not based on the tangent formula, was less sensitive than *SIR*2002 to the truncation of the high-resolution limit.

It is not unusual for the low-resolution data in any given X-ray data set to be incomplete for a variety of possible

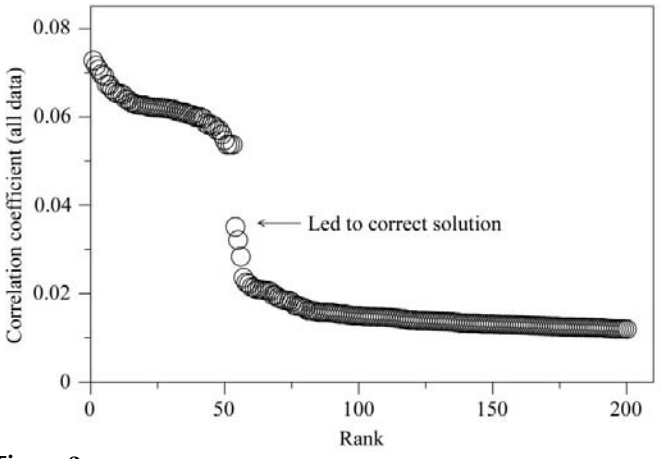

Figure 3
Ranking of the correlation coefficients for the top 200 sites generated by *ACORN* for the Rb^+-containing mutant lysozyme D72A/R96E/WT.

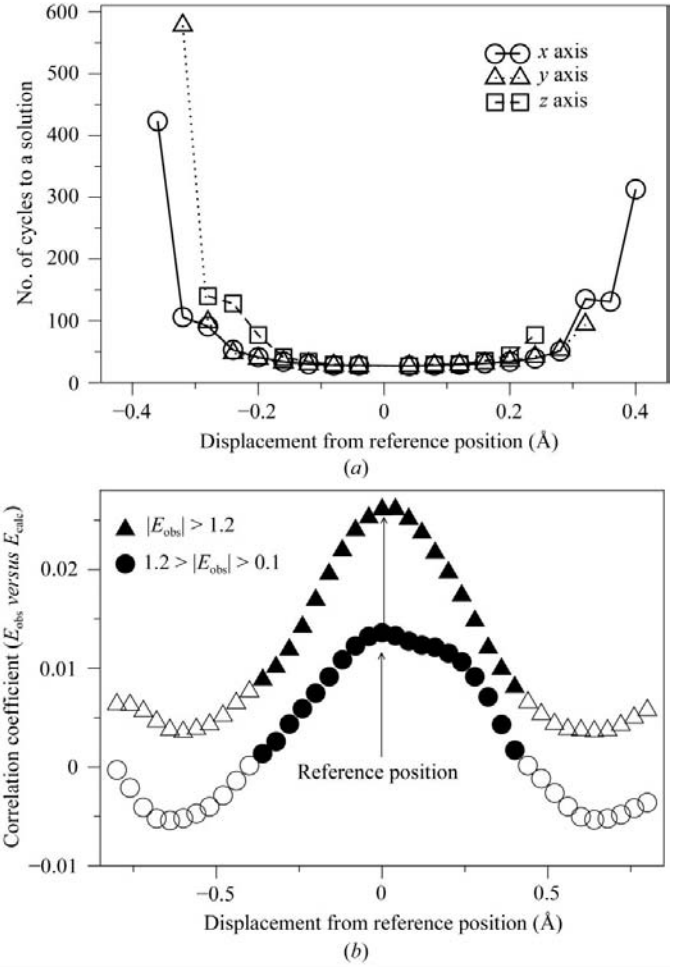

Figure 4
Trials with *ACORN* using 1.06 Å data and starting with Rb^+ displaced from its refined position. (*a*) Number of cycles of dynamic density modification needed to reach a solution as a function of displacement along the Cartesian axes. (*b*) The correlation coefficient between the calculated and observed E values as a function of displacement of the Rb^+ ion from its 'true' position. The solid symbols correspond to trials that were successful, while the open symbols correspond to failed attempts. Only the values for displacements along the x axis are shown.

reasons. To explore the effect of the low-resolution limit on the ability to obtain a successful solution, we varied the low-resolution cutoff from 5 to 15 Å in increments of 1 Å and carried out trials starting with 500 000 random starting sites. *ACORN* failed to obtain a solution when the cutoff was 5 and 6 Å and succeeded when the cutoffs were 7–15 Å. Hence, *ACORN* was less sensitive than *SIR*2002 to the loss of low-resolution data.

Fourier ripples around heavy atoms (in this case sulfur and Rb^+) can be avoided by using a sufficiently small grid size (*e.g.* 0.25 Å). This problem is also avoided in *ACORN* by the use of an upper electron-density cutoff of 5σ during dynamic density modification. The default grid size in *ACORN* is 0.53 Å (*i.e.* half the high-resolution limit). We tried a grid of 0.25 Å, in which case *ACORN* took eight times as long to reach a solution. This is as expected since the grid volume is eight times smaller.

3.4. Quality of the direct-methods maps

The quality of the direct-methods solutions was assessed by calculating the correlation between the direct-methods *E*-map and the *E*-map calculated from the reference structure. The correlation coefficients were determined section by section using the program *OVERLAPMAP* from the Collaborative Computational Project, Number 4 (1994) and had an average value of 0.895. The overall *E*-map correlation coefficient is related to the mean phase error (Lunin & Woolfson, 1993). The correlation coefficients between the *F*-maps were quite similar (average value 0.878). Hence, subsequent comparisons were made using *F*-maps.

F-map correlation coefficients were also computed on a residue-by-residue basis using *OVERLAPMAP* (Figs. 5*a* and 5*b*). The correlation coefficients for the main-chain atoms were higher than those for the side-chain atoms in both the *ACORN* and *SIR*2002 solutions. Both solutions had approximately linear relationships between the average map corre-

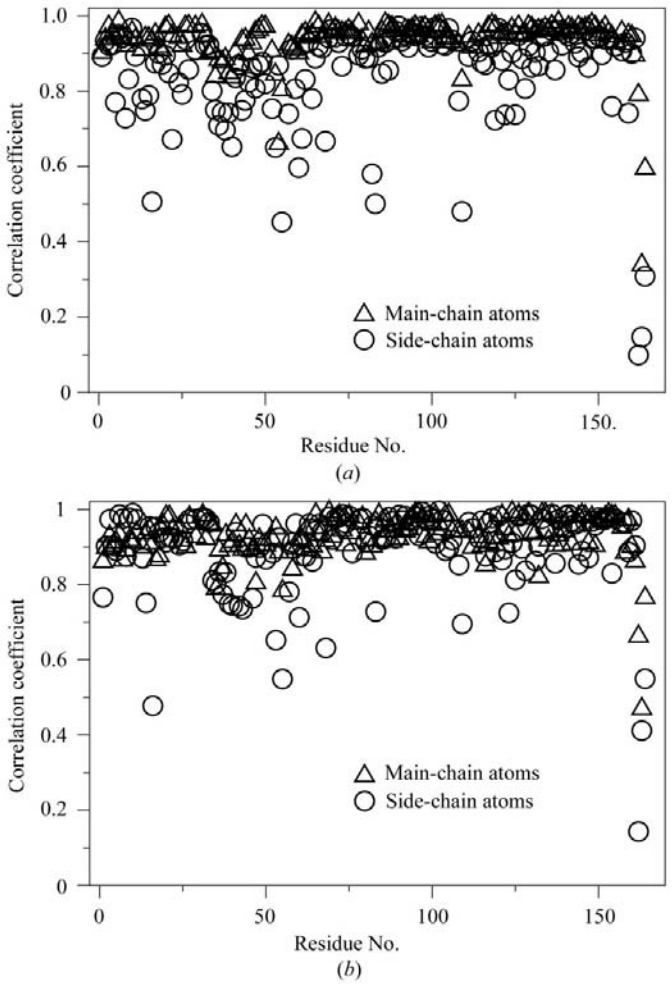

Figure 5
Map correlation plots. (*a*) Correlation coefficients between the *ACORN* map and the reference structure map for each residue in the structure. (*b*) Same as (*a*) but with the *SIR*2002 map.

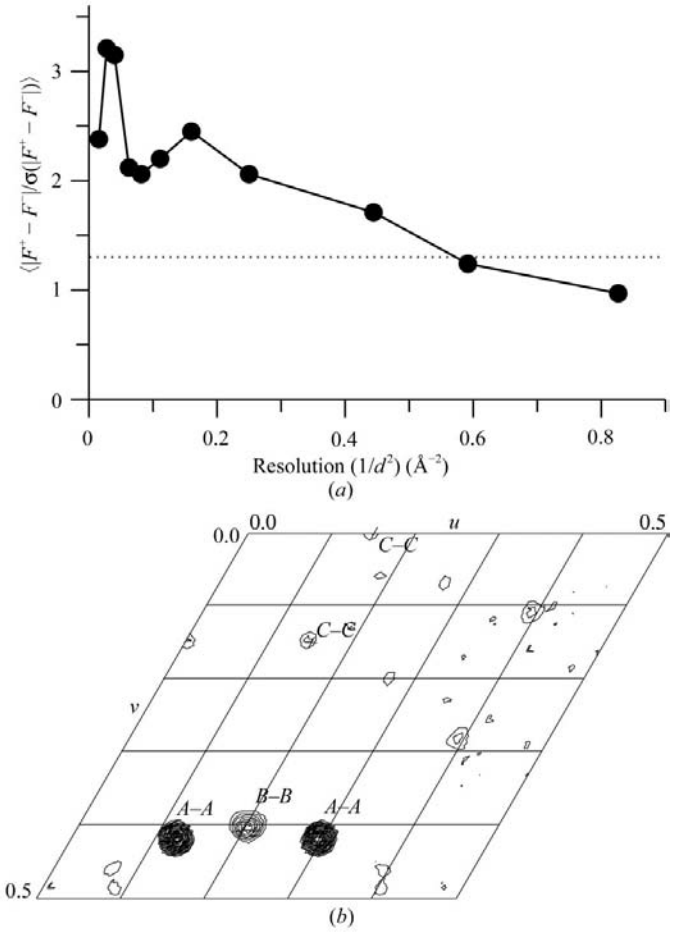

Figure 6
(*a*) Signal to noise for the anomalous pairs for the rubidium-containing lysozyme crystal. The dashed line is at 1.3, which corresponds to the suggested limit for SAD phasing. (*b*) Bijvoet Patterson map generated in *XtalView* using data from 22 to 1.4 Å after excluding Bijvoet pairs with differences greater than 30% of the average of *F*. Shown is part of the Harker section $w = 1/3$. The map was made with the origin peak removed. The contours start at 2σ and increase in intervals of 1σ. The peaks labeled *A–A* arise from the fully occupied site and that labeled *B–B* arises from the partially occupied site (see text). The *B–B* peak is the result of two overlapping self-vectors. The peak labeled *C–C* corresponds to an additional Rb^+ site that is partly occupied, is disordered and occupies two alternative positions.

lation coefficient of the side-chain atoms and the average B value of the side-chain atoms. The *SIR*2002 *ab initio* solution had higher correlation coefficients for the solvent atoms than the *ACORN* solution.

3.5. SAD phasing at atomic resolution

The Bijvoet differences were above the noise level to a resolution of about 1.4 Å (Fig. 6*a*). A Patterson map made with these differences contained one self-vector with a peak height of 24σ and a second self-vector peak of height 12σ (Fig. 6*b*). *SHELXD* (Schneider & Sheldrick, 2002) was used with the Bijvoet differences to solve the heavy-atom substructure. 52 of 100 trials in *SHELXD* found the correct solution. The best solution had a correlation coefficient of 0.392 with all of the data (0.241 for the weak data not used in the substructure solution). The best solution had four sites. The first two of these corresponded to the major peaks in the Bijvoet Patterson map (*A*–*A* and *B*–*B* in Fig. 6*b*). The third and fourth sites were weaker and in retrospect arose from two alternative sites of a disordered Rb^+ ion. These sites correspond to the Patterson peaks labeled *C*–*C* in Fig. 6(*b*). (A Bijvoet Fourier later suggested two additional minor sites with occupancies of about 0.1, but these were not included in any phase calculations.) Using the SAD phases to a resolution of 1.4 Å resolution derived from the Rb^+ sites found by *SHELXD*, *SHELXE* (Sheldrick, 2002) was used with the native data to 1.06 Å to (i) apply the low-density elimination method to resolve the twofold phase ambiguity, (ii) include the contribution of the heavy atoms to the protein phases and (iii) perform 100 cycles of density modification using the sphere-of-influence method.

While SAD and MAD phasing have been performed to atomic resolution (Brodersen *et al.*, 2000; Schmidt *et al.*, 2002), the number of examples is still very limited. This is the first example of SAD phasing to atomic resolution using rubidium.

3.6. Comparison with phases determined from anomalous dispersion

The phases from *ACORN*, *SIR*2002 and *SHELXD*/*E*-SAD are compared with those from the reference structure in Fig. 1(*a*) and in Table 3. The SAD solution agrees best with the reference phases, the *ACORN* solutions are intermediate and the agreement for the *SIR*2002 solutions is poorest. When the phase differences are weighted by the figure of merit of the experimental phases (Fig. 1*b*), the SAD and *ACORN* phases appear to be about the same in terms of their agreement with the reference phases and the *SIR*2002 phases still show the largest disagreement.

The seeded and *ab initio* phase sets from both *SIR*2002 and *ACORN* have very similar mean phase errors (Figs. 1*a* and 1*b*; Table 3). This suggests that the density-modification steps in the later part of structure determinations may have removed to a large degree any differences in the starting phase sets. These two comparisons imply that seeded direct-method solutions may be as accurate as *ab initio* solutions when many cycles of density modification are included.

Table 3
Comparison of phase sets determined by direct methods and by SAD with the reference structure.

Each comparison was made with 83 251 unique reflections in the resolution range 22–1.06 Å. The map correlation coefficient was computed using equation (12) of Lunin & Woolfson (1993). Computations were performed with *SHELXPRO* v.2000-1. MPE is the mean phase error, FOM is the figure of merit, wMPE is the FOM-weighted MPE and FwMPE is the mean phase error weighted by both the figure of merit and the structure amplitude.

Phase set	⟨Unweighted MPE⟩ (°)	⟨FOM⟩	⟨wMPE⟩ (°)	⟨FwMPE⟩ (°)	⟨Map CC⟩
SHELXD/*E*-SAD	21.6	0.811	16.4	13.2	0.940
ACORN, *ab initio*	28.2	0.348	19.0	15.1	0.851
ACORN, seeded	27.4	0.349	18.4	14.9	0.851
*SIR*2002, *ab initio*	36.7	0.374	24.4	16.7	0.900
*SIR*2002, seeded	36.3	0.377	24.2	16.6	0.901

3.7. Automated model building

The experimental electron-density maps from the two direct-methods solutions and from the *SHELXD*/*E*-SAD solution were suitable for automated model building by *ARP*/*wARP* (Perrakis *et al.*, 1999) and *RESOLVE* (Terwilliger, 2000). *RESOLVE* has been optimized for building into maps of about 1.8 Å resolution and worked best when the data were truncated to this resolution. Under these conditions *RESOLVE* built 140 out of 164 residues. These residues had side chains that were largely correctly placed.

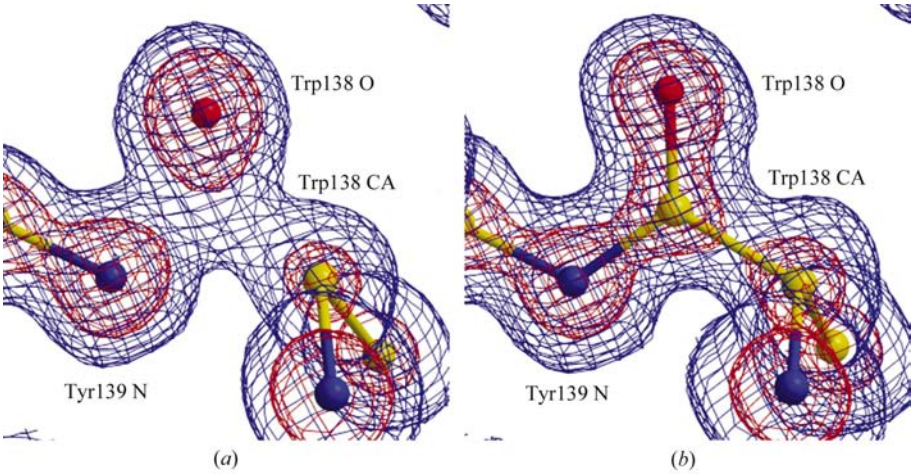

Figure 7
(*a*) The backbone carbonyl C atom of Trp138 was missing from *SIR*2002 *ab initio* direct-methods structure determination of D72A/R96E/Rb^+. The *SIR*2002 F_o map shows low density at the expected position for this atom. (*b*) After refinement in *SHELXL*, the phases improved enough to reveal much stronger density in the $3F_o - 2F_c$ map at the position of Trp138 C. In both maps, the blue density is at the 1σ level and the red density is at the 4σ level.

In the case of *ARP*/*wARP*, all of the data could be used. In the SAD map, *ARP*/*wARP* built residues 1–161 with three small gaps that included residues 53–55, 69–70 and 142–143. The last three residues 162–164 are highly disordered and difficult to model. *ARP*/*wARP* did make several mistakes in the placement of side chains. Trp138 was pointed in the wrong direction and clearly out of density. This was quite surprising since this residue has amongst the lowest *B* values in the structure. The mislocation of Trp138 was correlated with misplacement of Gln105, which was clearly out of density and also in the wrong rotameric state. Likewise, the buried side chains of Trp126 and Arg95 were out of density in a similar correlated fashion. The side chain of Leu7 was in the wrong rotameric state, with the CD1 atom rotated out of density. Thus, there was about a 4% error rate in the placement of side chains. The gaps in the model had continuous main-chain density although the side chains lacked some density. These gaps were easy to complete by manual model building in the SAD map.

ARP/*wARP* also was very successful with main-chain tracing of the *ACORN* and *SIR*2002 maps. However, the same mistakes in side-chain placement were made.

3.8. Reassembly of direct-methods structures

Our goal was to assemble starting models from the *ACORN* and *SIR*2002 determinations for refinement that were as accurate and complete as possible and were unbiased by stereochemical restraints. As well as giving phases, *SIR*2002 gives the coordinates of the atoms after four cycles of unrestrained least-squares refinement. These coordinates were scattered across several asymmetric units and were gathered together with a program that we wrote for this purpose. (The *REGROUP* command in the *SIR*2002 graphical user interface is designed to perform this task, but it gathered only about half of the atoms in this case. An enormous amount of manual intervention would have been required to complete the model.) When the coordinates resulting from the *SIR*2002 solution were gathered into one molecular entity they were found to be remarkably complete for residues 1–161. The one exception was the absence of a number of backbone carbonyl C atoms (Fig. 7), which may have been a consequence of an inappropriate threshold for peak picking during the *SIR*2002 structure determination. After placing these atoms by hand, the main-chain atoms had an r.m.s.d. of 0.2 Å with the refined reference structure.

We also obtained a very complete model for the *ACORN* structure determination. In this case the peak-picking threshold was set low enough to include the backbone carbonyl C atoms. Both solutions had remarkably few misplaced protein atoms and fairly complete solvent structures.

When the *ARP*/*wARP* and peak-picked models were compared with the refined molecular-substitution model (Table 4), the *ARP*/*wARP* models were found to agree much better than the peak-picked models, although the latter were quite accurate. Consequently, we decided to use the *ARP*/*wARP* models to begin refinement at the resolution limit.

Table 4
Comparison to the reference structure of models built into *ACORN ab initio*, *SIR*2002 *ab initio* and *SHELXD*/*E*-SAD maps.

The values quoted are the root-mean-square deviation between main-chain atoms (N, CA, C, O).

	Deviation, residues 1–161 (Å)			Deviation, residues 81–161 (Å)		
	ACORN	*SIR*2002	SAD	*ACORN*	*SIR*2002	SAD
Before refinement						
Peak picked	0.192	0.186	0.179	0.173	0.148	0.013
Automatically built	0.097	0.091	0.051	0.040	0.040	0.040
After refinement						
Refined in *SHELXL*	0.085	0.057	0.072	0.038	0.016	0.025

3.9. Refinement of the *ab initio* determined structures

The recommended procedure for refining models derived from MIR, MAD and molecular substitution is to apply stereochemical restraints, to gradually increase the resolution and to conservatively add parameters related to anisotropic displacement parameters, model disorder, solvent location and so on. The rationale is that the starting model may be inaccurate and that new parameters must be introduced gradually in order to avoid introducing bias into the model.

In contrast, a successful *ab initio* structure determination will automatically provide a quite accurate set of atomic coordinates. The experimental maps showed side-chain disorder and this disorder was modeled manually by building into the experimental maps before beginning refinement with *SHELXL*.

The *SIR*2002, *ACORN* and SAD structures were refined independently using the protocol for atomic resolution refinement with *SHELXL* described by Sheldrick & Schneider (1997). In principle, *SHELXL* allows the use of separated Bijvoet pairs during refinement, but this was not performed in order to keep all comparisons of phase sets and structures on the same basis. Anisotropic atomic displacement parameters were used with all of the atoms. The occupancies of the disordered side chains and ions were refined. Riding H atoms were added to all atoms except hydroxyl and carboxyl O atoms, imidazole N atoms and disordered side chains. Near the end of the refinement, the algorithm was changed from conjugate-gradient least squares to block-matrix least squares. The normal matrix was inverted at the end of refinement in order to obtain accurate estimates of the standard uncertainties of the model's parameters. For the majority of the structure the coordinate uncertainty is about 0.02 Å.

3.10. Biological insights

These precise structure determinations at atomic resolution provide some new insights into the structure of T4 lysozyme.

3.10.1. An unusually short hydrogen bond. The crystal structure of the wild-type protein has an unusually short hydrogen bond (2.48 ± 0.05 Å) between of Ser117 OG and of

Table 5
Geometry of Ser117 OG–Asn132 OD1 hydrogen bond in the structure of mutant D72A/R96E after peak-picking of *E*-maps, after automated model building with *ARP/wARP* and after restrained least-squares refinement in *SHELXL*.

The OG—OD1 distance and the CB—OG—OD1 angle are given.

	ACORN		*SIR*2002		SAD		Reference structure	
	Length (Å)	Angle (°)	Length (Å)	Angle (°)	Length (Å)	Angle (°)	Length (Å)	Angle (°)
Before refinement								
Peak-picked *E*-map	2.47	131.7	2.46	130.6	2.35	125.7	—	—
Automatically built	2.46	127.0	2.46	128.4	2.44	128.8	—	—
After refinement	2.49	129.4	2.47	127.3	2.50	127.9	2.51	129.8

Table 6
Non-planarity of the peptide bond between Arg125 and Trp126.

	ω (°)	Non-planarity (°)
WT at 1.7 Å (Bell *et al.*, 1991)	−173.1	6.9
WT at 1.45 Å (Gassner *et al.*, 2003)	−167.9	12.1
D72A/R96E at 1.06 Å		
Model from peak-picked *E*-map		
ACORN ab initio	−157.3	22.7
*SIR*2002 *ab initio*	−161.5	18.5
SHELXD/*E*-SAD	−166.6	13.4
Model from *ARP*/*wARP*		
ACORN ab initio	−161.7	18.3
*SIR*2002 *ab initio*	−161.5	18.5
SHELXD/*E*-SAD	−161.7	18.3
Model from restrained *SHELXL* refinement		
ACORN ab initio	−162.0	18.0
*SIR*2002 *ab initio*	−163.4	16.6
SHELXD/*E*-SAD	−162.7	17.3
Molecular substitution	−162.5	17.5
Model from *SHELXL* refinement without the peptide-planarity restraint		
ACORN ab initio	−161.7	18.3
*SIR*2002 *ab initio*	−162.9	17.1
SHELXD/*E*-SAD	−162.2	17.8
Molecular substitution	−162.3	17.7
Model from *SHELXL* refinement without the peptide-planarity and bond-angle restraints		
ACORN ab initio	−161.7	18.3
*SIR*2002 *ab initio*	−162.9	17.1
SHELXD/*E*-SAD	−162.2	17.8
Molecular substitution	−162.1	17.9
Model from *SHELXL* refinement without the peptide-planarity, bond-angle and bond-length restraints		
ACORN ab initio	−161.6	18.4
*SIR*2002 *ab initio*	−162.2	17.8
SHELXD/*E*-SAD	−162.2	17.8
Molecular substitution	−162.1	17.9

Asn132 OD. This hydrogen bond is of biological interest because it is in the region of the protein that interacts with the *E. coli* peptidoglycan cell-wall substrate (Kuroki *et al.*, 1993). Three mutations that remove this hydrogen bond, Asn132→Ala (Zhang *et al.*, 1992), Ser117→Ala (Blaber *et al.*, 1995) and Ser117→Phe (Anderson *et al.*, 1993), all increase the stability of the protein. This is consistent with the observed short bond length and suggests that the bond is under strain. The geometry of this hydrogen bond in the different structures of D72A/R96E before and after refinement is shown in Table 5. (During the refinements with riding H atoms in *SHELXL*, hydrogen was not added to the γ-hydroxyl of Ser117.) In all ten structures listed in Table 5, the hydrogen bond is significantly shorter than the normal length of 2.7–2.8 Å. This supports the short bond length seen in WT and suggests that it also occurs in the D72A/R96E mutant. The three models built by peak picking in *E*-maps are free of bias from stereochemical restraints, further suggesting that this short hydrogen bond is a feature of the crystal structure and not an artifact.

3.10.2. Possible role of Na^+ in catalysis. The discovery of Rb^+ bound to the carboxylate of Glu11 may be of relevance to the reaction mechanism. Glu11 is thought to serve as a general acid in catalysis (Kuroki *et al.*, 1993). After donating its proton, the negatively charged Glu11 may be stabilized by the bound cation. The binding of a chemically similar sodium ion at this site would be consistent with the observed reaction-rate enhancement with increasing concentrations of NaCl up to 100 m*M* (Becktel & Baase, 1985). Because of their similar electron density, Na^+ ions are difficult to distinguish from water, especially at lower resolution. This would explain the prior difficulty in identifying such a cation-binding site.

The Rb^+ ion binds electrostatically to the backbone carbonyl of Glu11 and to the hydroxyl of Tyr18, as well as to Glu11 OE2. In addition to neutralizing the charge of Glu11, the Rb^+ may help orient the Glu11 side chain to facilitate its interaction with the guanidinium group of Arg145.

3.10.3. Internal solvent-binding sites. T4 lysozyme has four buried waters in three cavities (Weaver & Matthews, 1987; Xu *et al.*, 2001). Three of these waters each form at least three hydrogen bonds with protein ligands and/or another water molecule. The fourth water, in a cavity near site 149, forms only two hydrogen bonds with protein ligands, which raises some doubt about its identity. In the structures of D72A/R96E it is modeled as an O atom. Its isotropic *B* value is 9.8 $Å^2$, while the corresponding values of its hydrogen-bonding partners Asn101 OD1 and Gln105 OE1 are 8.6 and 9.1 $Å^2$, respectively. The similarity in *B* values supports the hypothesis that this solvent molecule is indeed a water molecule.

3.10.4. Distortion of a peptide plane. The experimental structures built by peak picking in the *E*-maps revealed an ∼18° deviation from planarity in the peptide between Arg125 and Trp126 (Table 6), which is in a tight turn between two helices in the C-terminal domain. The atoms involved in this peptide bond have average *B* values of about 10 $Å^2$. The deviation from planarity is expected to introduce a peptide-bond rotation energy penalty of about 14.6 kJ mol^{-1} by Maxwell–Boltzmann relations for peptides and proteins (MacArthur & Thornton, 1996; Edison, 2001) and about

12.1 kJ mol^{-1} by the relation $A\sin^2(\delta)$, where A is about 125 kJ mol^{-1} and δ is the deviation from planarity in degrees (Corey & Pauling, 1953).

Similar deviations from planarity have been found on occasion in the atomic resolution structures of other proteins (*e.g.* König *et al.*, 2003). However, this is one of the few instances where such distortion has been validated by several *de novo* structure determinations.

The deviation from planarity was not apparent in the structure of WT at 1.7 Å resolution (Bell *et al.*, 1991) and only somewhat apparent in the structure of WT at 1.45 Å resolution (Gassner *et al.*, 2003) (Table 6). At these resolutions, the planarity restraint limited the distortion of the peptide plane.

The ω angles from the automatically built *ARP*/*wARP* models agree quite well with those from the peak-picked structures. Even though these automatically built structures experienced many cycles of refinement with *REFMAC* in which the peptide bonds were restrained, the unusual ω angle between site 125 and 126 persisted. Likewise, the structures refined with stereochemical restraints in *SHELXL* had very similar ω angles to those of the peak-picked structures.

The influence of stereochemical restraints in *SHELXL* on the peptide plane was studied by removing the planarity restraint, the planarity and angle restraints and then the planarity, angle and bond restraints on all of the peptide bonds. In each case, ten cycles of conjugate-gradient least-squares refinement followed the removal of the restraints (Table 6). None of these restraints appear to have a large influence on the ω angle. A comparison of the ω-angle distributions before and after the removal of all of the main-chain restraints showed that the distributions were very similar and suggested that the main-chain restraints did not have a very important influence on most of the structure (not shown). However, it may still be advisable to retain main-chain restraints in atomic resolution refinement in order to improve the model in poorly defined sections of the backbone.

3.11. Quick soaks with halides and ADM

Halides and other ligands can be introduced into crystals by short soaks (Dauter *et al.*, 2000; Sun *et al.*, 2002) and this may be a method of preparing crystals for ADM experiments. This approach has two limitations. Firstly, not all crystals still diffract to atomic resolution after short soaks with halides and some are in fact destroyed by such soaks. This is the case for T4 lysozyme crystals grown under the standard high-phosphate conditions. Secondly, the halide-binding sites must be sufficiently occupied to allow their location by direct methods. After quick soaking, a halide-binding site often has an occupancy of 0.2–0.4 (Dauter *et al.*, 2000). In the case of such a partially occupied Rb^+ or Br^-, the contribution to the total scattering is comparable to that of a protein S atom. This may be insufficient to be useful with ADM for proteins larger than 1000–1200 atoms.

4. Conclusions

Using data to a resolution of 0.97 Å, it was not possible to solve the 1300-atom structure of T4 lysozyme using direct methods. If, however, the crystal included a single bound Rb^+ ion, an *ab initio* structure determination was possible using either *ACORN* or *SIR*2002. Both programs were essentially equivalent in time required. Tests suggest that trial starting sites located on a grid of spacing about 0.7 Å should be sufficient for *ACORN* to find a site close enough to the true Rb^+ position to lead to an *ab initio* solution. A search on a grid for successful 'seeding sites' can also be used for *SIR*2002, but requires a finer grid and is not recommended. Our experience with T4 lysozyme suggested that random starting phases from use of the tangent formula gave phases that were as effective as those based on the known Rb^+ site. In the present case, the Rb atom contributed 1.7% of the total protein scattering power. The equivalence of the two types of solutions is important because it suggests that essentially no bias from the starting Rb^+ phases remains. Such 'seeded' direct-methods solutions require much less computer time and the required resolution of the diffraction is often less. Phases obtained *via* direct methods appeared in this example to be not quite as good as those from SAD. The models that were obtained from *ab initio* solutions and then refined are very similar to those obtained by conventional structure determination. The limitations of the ADM method are that it requires atomic resolution data and the presence of one or more electron dense atoms (*e.g.* Se, Br, I, Rb, Cs, Hg *etc.*).

We thank Hong Xiao for overexpressing and purifying the mutant D72A/R96E, the user-support staff at SSRL for their assistance during our visit to BL 9-1, and Mike Quillin for allowing us to modify his program *RESOLVE* to regroup and relabel protein and solvent atoms.

References

Alber, T. & Matthews, B. W. (1987). *Biochemistry*, **26**, 3754–3758.
Anderson, D. E., Hurley, J. H., Nicholson, H., Baase, W. A. & Matthews, B. W. (1993). *Protein Sci.* **2**, 1285–1290.
Becktel, W. J. & Baase, W. A. (1985). *Anal. Biochem.* **150**, 258–263.
Bell, J. A., Wilson, K. P., Zhang, X.-J., Faber, H. R., Nicholson, H. & Matthews, B. W. (1991). *Proteins*, **10**, 10–21.
Blaber, M., Baase, W. A., Gassner, N. & Matthews, B. W. (1995). *J. Mol. Biol.* **246**, 317–330.
Boggon, T. J. & Shapiro, L. (2000). *Structure*, **8**, R143–R149.
Brodersen, D. E., de La Fortelle, E., Vonrhein, C., Bricogne, G. & Kjeldgaard, M. (2000). *Acta Cryst.* D**56**, 431–441.
Burla, M. C., Camalli, M., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Polidori, G. & Spagna, R. (2003). *J. Appl. Cryst.* **36**, 1103.
Burla, M. C., Carrozzini, B., Caliandro, R., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Polidori, G. (2003). *Acta Cryst.* A**59**, 560–568.
Burla, M. C., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Polidori, G. (2003). *Acta Cryst.* A**59**, 245–249.
Burla, M. C., Carrozzini, G. L., Cascarano, G. L., Giacovazzo, C. & Polidori, G. (2002). *Z. Kristallogr.* **217**, 629–635.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Corey, R. B. & Pauling, L. (1953). *Proc. R. Soc. London Ser. B*, **141**, 10–20.
Dauter, Z. & Dauter, M. (1999). *J. Mol. Biol.* **289**, 93–101.
Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* D**56**, 232–237.
Deacon, A. M., Weeks, C. M., Miller, R. & Ealick, S. E. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 9284–9289.
Edison, A. S. (2001). *Nature Struct. Biol.* **8**, 201–202.
Eriksson, A. E., Baase, W. A. & Matthews, B. W. (1993). *J. Mol. Biol.* **229**, 747–769.
Evans, P. R. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **33**, 22–24.
Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Jia-xing, Y. & Chao-de, Z. (2000). *Acta Cryst.* D**56**, 1137–1147.
Gassner, N. C., Baase, W. A., Mooers, B. H., Busam, R. D., Weaver, L. H., Lindstrom, J. D., Quillin, M. L. & Matthews, B. W. (2003). *Biophys. Chem.* **100**, 325–340.
Giacovazzo, C. (1977). *Acta Cryst.* A**33**, 933–944.
Giacovazzo, C. (1998). *IUCr Monographs on Crystallography*, No. 8. Oxford University Press.
Jia-xing, Y., Woolfson, M. M., Wilson, K. S. & Dodson, E. J. (2002). *Z. Kristallogr.* **217**, 636–643.
König, V., Vértesy, L. & Schneider, T. R. (2003). *Acta Cryst.* D**59**, 1737–1743.
Korolev, S., Dementiera, I., Sanishvili, R., Minor, W., Otwinowski, Z. & Joachimiak, A. (2001). *Acta Cryst.* D**57**, 1008–1012.
Kuroki, R., Weaver, L. H. & Matthews, B. W. (1993). *Science*, **262**, 2030–2033.
Landt, O., Grunert, H. & Hahn, U. (1990). *Gene*, **96**, 125–128.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EAMCB Newsl. Protein Crystallogr.* **26**.
Lunin, V. Y. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 530–533.
MacArthur, M. W. & Thornton, J. M. (1996). *J. Mol. Biol.* **264**, 1180–1195.
McAuley, K. E., Jia-xing, Y., Dodson, E. J., Lehmbeck, J., Ostergaard, P. R. & Wilson, K. S. (2001). *Acta Cryst.* D**57**, 1571–1578.
McRee, D. E. (1992). *J. Mol. Graph.* **10**, 44–46.
Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A. (1993). *Science*, **259**, 1430–1433.
Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E. & Dahlquist, F. W. (1989). *Methods Enzymol.* **177**, 44–73.
Mukherjee, M., Maiti, S., Ghosh, S. & Woolfson, M. M. (2001). *Acta Cryst.* D**57**, 1276–1280.
Mukherjee, M., Maiti, S. & Woolfson, M. M. (2000). *Acta Cryst.* D**56**, 1132–1136.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Poteete, A. R., Dao-pin, S., Nicholson, H. & Matthews, B. W. (1991). *Biochemistry*, **30**, 1425–1432.
Schmidt, A., Gonzalez, A., Morris, R. J., Costabel, M., Alzari, P. M. & Lamzin, V. S. (2002). *Acta Cryst.* D**58**, 1433–1441.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.
Sheldrick, G. M., Dauter, Z., Wilson, K. S., Hopre, H. & Sieker, L. C. (1993). *Acta Cryst.* D**49**, 18–23.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Sun, P. D., Radaev, S. & Kattah, M. (2002). *Acta Cryst.* D**58**, 1092–1098.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Tronrud, D. E. (1997). *Methods Enzymol.* **277**, 306–319.
Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst.* A**43**, 489–501.
Uson, I. & Sheldrick, G. M. (1999). *Curr. Opin. Struct. Biol.* **9**, 643–648.
Weaver, L. H. & Matthews, B. W. (1987). *J. Mol. Biol.* **193**, 189–199.
Weeks, C. M., DeTitta, G. T., Miller, R. & Hauptman, H. A. (1993). *Acta Cryst.* D**49**, 179–181.
Weeks, C. M., Hauptman, H. A., Smith, G. D., Blessing, R. H., Teeter, M. M. & Miller, R. (1995). *Acta Cryst.* D**51**, 33–38.
Xu, J., Baase, W. A., Quillin, M. L., Baldwin, E. P. & Matthews, B. W. (2001). *Protein Sci.* **10**, 1067–1078.
Zhang, X.-J., Baase, W. A. & Matthews, B. W. (1992). *Protein Sci.* **1**, 761–776.