Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Extension to 2268 atoms of direct methods in the *ab initio* determination of the unknown structure of bacteriophage P22 lysozyme

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: blaine@uoxray.uoregon.edu

Received 17 August 2005
Accepted 11 November 2005

PDB References: bacteriophage P22 lysozyme L86M mutant, 2anv, r2anvsf; L87M mutant, 2anx, r2anxsf.

The X-ray crystal structure of the previously unknown bacteriophage P22 lysozyme, the product of gene *19*, has been determined *ab initio* by direct methods using the program *SIR*2002. The presence of several partially occupied iodine anions and samarium cations augmented the ability of direct methods to locate all 2268 non-H protein atoms in the asymmetric unit, making this one of the largest structures to date to be determined *ab initio*. The iodides were introduced from a quick soak, which the crystal survived sufficiently well to diffract to 1.04 Å resolution. The complete heavy-atom substructure contributed 6.6% of the total scattering power. The initial determination of the structure assumed that there were two iodide ions in the asymmetric unit, although it was later determined that these sites correspond to partially occupied samarium ions. Tests suggested that it is better to overestimate rather than underestimate the heavy-atom content. While experimental phases from all of the successful tests were of high quality, the best results came from a SAD experiment using the programs *SHELXD* and *SHELXE*. Nonetheless, *ab initio* structure determination by direct methods was found to be a viable alternative to traditional protein crystallographic methods provided that the X-ray data extend to atomic resolution and heavy atoms with sufficient scattering power are present in the crystal.

1. Introduction

Ab initio structure determination is taken to mean the direct determination of a structure from the native structure factors without the use of isomorphous replacement, anomalous dispersion or molecular replacement and with no prior information on any atomic position.

Traditional direct methods rely on reciprocal-space techniques. They are for the most part limited to structures with less than 200 non-H atoms because the probabilistic relationships between the phases become progressively weaker as the number of atoms in the unit cell grows larger (Cochran, 1955). In addition, these methods generally require atomic resolution data, which has also limited their application to proteins.

The prospects for applying direct methods to proteins improved dramatically in the early 1990s with the development of dual-space methods, which cycle between reciprocal space and real space (Miller *et al.*, 1993). Dual-space methods are included in the popular programs *SnB* (Miller *et al.*, 1994; Weeks & Miller, 1999) and *SHELXD* (Sheldrick & Gould, 1995). These programs have been used to determine the structures of a number of proteins (Usón & Sheldrick, 1999; Sheldrick *et al.*, 2001) with up to 1010 non-H protein atoms in

Table 1
X-ray data-collection statistics for the L87M mutant of P22 lysozyme.

Values for the highest resolution shell (1.08–1.04 Å) are given in parentheses.

X-ray source	ALS beamline 8.2.2
Space group	*C*2
Unit-cell parameters	
a (Å)	134.0
b (Å)	50.4
c (Å)	46.6
β (°)	103.8
Wavelength (Å)	0.9184
Resolution range (Å)	42.1–1.04
Distance (mm)/exposure (s)/oscillation (°)	120/8/0.75, 290/1/1.5
No. of reflections	144532 (14367)
Multiplicity	4.4 (3.1)
Completeness (%)	99.6 (99.4)
$R_{\rm merge}$ (%)	5.3 (22.2)
$\langle I\rangle/\sigma(I)$	42.3 (3.9)
$B_{\rm Wilson}$ (Å^2)	8.3

the absence of atoms heavier than sulfur (Liu *et al.*, 2003) and to slightly over 2000 atoms in the case of metalloproteins (Frãzao *et al.*, 1999).

There are more than 250 proteins in the PDB with resolutions beyond 1.2 Å (Dauter, 2003). This is generally considered to be the cutoff for atomic resolution data. About one-sixth of these proteins have been determined by direct methods, largely as part of tests of new versions of direct-methods programs. Only a handful of these determinations are of previously unknown structures. In addition, only five are of light-atom proteins with more than 1000 atoms. These five proteins all have at least two disulfide bonds. The close proximity of the sulfurs in disulfides causes them to contribute more to the total scattering than two isolated sulfurs. As a result, it is much easier to use direct methods to locate a pair of S atoms in a disulfide bond than to locate an isolated S atom. Also, all five structures had resolution limits that extended beyond 1.0 Å. Higher resolution improves the strength of the probabilistic phase relationships and the likelihood of a successful structure determination.

The larger metalloproteins determined by direct methods contain transition metals that have smaller atomic numbers than the heavy atoms such as mercury and platinum traditionally used in MIR. The largest is cytochrome c_3 from the sulfate-reducing bacterium *Desulfovibrio gigas* (Frãzao *et al.*, 1999). This protein forms a cross-linked dimer with 2024 atoms including eight heme irons in the asymmetric unit. It was solved with *SHELXD* using 1.2 Å data. Another interesting example is pseudoazurin, which contains a copper cation and was solved with data to only 1.55 Å resolution using *SIR*2003-*N* (Burla, Carrozzini, Caliandro *et al.*, 2003; Burla, Carrozzini, Cascarano *et al.*, 2003).

These results suggest that the introduction into a light-atom protein of an atom somewhat more electron-dense than sulfur can enhance the probability of obtaining an *ab initio* structure determination. This heavy atom could be introduced by selenomethionine incorporation, chemical modification, cocrystallization or soaking of crystals. We called this approach augmented direct methods (ADM) and showed its successful application to a mutant of T4 lysozyme which had 1308 non-H atoms, no heavy-atom prosthetic groups and no disulfide bonds (Mooers & Matthews, 2004). This mutant bound several rubidium cations with various occupancies.

The opportunity to test ADM on an unknown protein arose when we discovered crystallization conditions for the L87M mutant of bacteriophage P22 lysozyme that allowed the collection of atomic resolution data. Crystals of the wild-type form of this protein had resisted structure determination for 20 y owing to their epitaxic growth habit, generally poor diffraction quality, inadequate cryoconditions, variation in crystal quality between protein preparations and very poor quality of crystals of the selenomethionine-containing protein. The crystal of the L87M mutant used here had two molecules in the asymmetric unit, giving 2268 non-H protein atoms. ADM gave an *ab initio* determination using the program *SIR*2002 (Burla, Carrozzini, Caliandro *et al.*, 2003; Burla, Carrozzini, Cascarano *et al.*, 2003), making this one of the largest structures determined *ab initio*.

To help extend the application of direct methods to larger proteins, we tested several aspects of the application of ADM. Firstly, we describe the successful extension of ADM to 2268 non-H protein atoms. Secondly, we explore the effect of underestimating or overestimating the number of heavy atoms included in the unit cell at the start of a direct-methods trial. Thirdly, we test the effectiveness of treating the Bijvoet pairs separately in the structure-determination trial. We also compare the effectiveness of direct-methods structure determination relative to the more conventional use of anomalous scattering (SAD). The details of the structure of P22 lysozyme will be presented elsewhere.

2. Materials and methods

2.1. Mutagenesis, protein expression and purification

The L87M mutant was made using the QuickChange mutagenesis kit (Stratagene). The template was gene *19* of bacteriophage P22, which had been moved from plasmid pDR118 (Rennell & Poteete, 1985) into the *Nde*I/*Xho*I restriction sites of pET22b. The start site of translation was Met1 rather than Met2 as suggested by amino-acid sequencing analysis (Rennell & Poteete, 1985). The protein was over-expressed in Stratagene BL21-CodonPlus RR1 cells and purified using the protocol for bacteriophage T4 lysozyme (Eriksson *et al.*, 1993), except that the protein was stored in 20 m*M* Tris–HCl pH 7.5. MALDI–TOF analysis of the purified protein showed that the first methionine was still present.

2.2. Crystallization

Large plate-like crystals were grown in 3–10 d by vapor diffusion at 277 K using the hanging-drop method. The protein solution was 20 m*M* Tris pH 7.5 at a concentration of 27 mg ml^{-1}. The reservoir solution had a volume of 1 ml and contained 20 m*M* Tris pH 7.5, 10 m*M* $ZnSO_4$, 125 m*M* Li_2SO_4 and 23% PEG 3400. 5 µl protein solution, 4 µl reservoir solution and 1 µl of 100 m*M* $SmCl_3$ were mixed together to form the crystallization drop.

2.3. X-ray data collection

A crystal with dimensions of 0.5 × 0.4 × 0.05 mm was mounted in a rayon loop and flash-cooled in a nitrogen stream after passage through a drop containing 19%(*v*/*v*) glycerol, 18%(*v*/*v*) polypropylene glycol P400, 65%(*v*/*v*) 2x synthetic mother liquor containing 0.8 *M* NaI. The crystal had the long *a* edge (Table 1) aligned close to the spindle axis in order to minimize the number of overlapped reflections at higher resolution. X-ray diffraction data were collected at ALS beamline 8.2.2 using radiation with energy 13.5 keV and an ADSC Quantum 315 CCD detector. This wavelength was selected as a compromise between maximization of the photon flux and minimization of radiation damage from longer wavelengths. To compensate for saturation of the detector by intense low-resolution reflections, data were collected at two distances. At the longer distance, 180° of images were collected in the inverse-beam mode in wedges of two images. This ensured that the Bijvoet pairs were measured accurately. At the shorter distance, 180° of data were collected. The high-resolution data were collected as quickly as possible to avoid a loss of resolution arising from radiation damage.

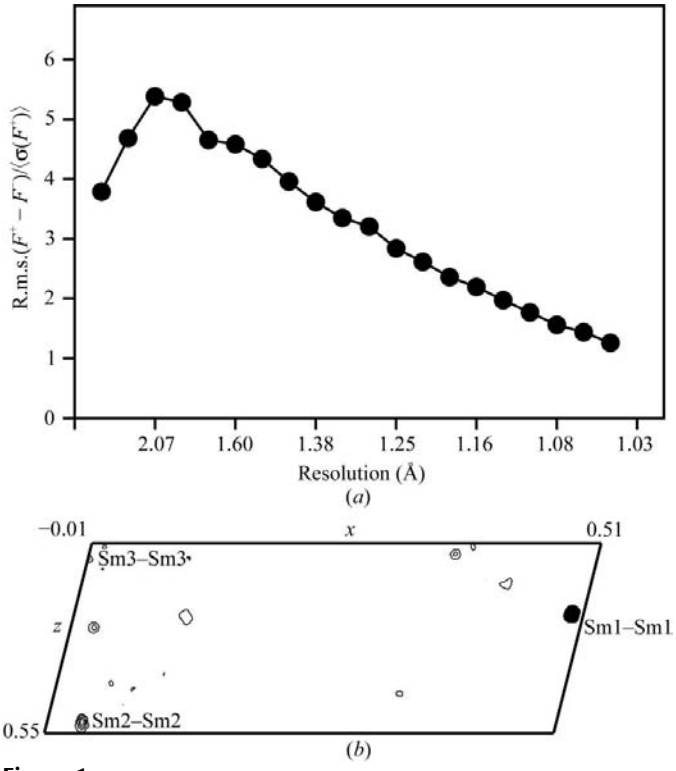

Figure 1
Detection of anomalous signal for the mutant L87M of P22 lysozyme. (*a*) Plot of r.m.s.$(F^+ - F^-)/\langle\sigma(F^+)\rangle$ *versus* resolution, where r.m.s.$(F^+ - F^-)$ is the r.m.s. difference between the Bijvoet-related reflections and $\langle\sigma(F^+)\rangle$ is the average uncertainty of F^+ in the same resolution shell. (*b*) Harker section $y = 0$ from the Patterson map made with the anomalous differences using all of the data. The origin peak was removed. The contours start at 4σ and continue in 2σ increments. The self-vectors associated with the three partially occupied samariums in the asymmetric unit have labels placed to their right. The unlabeled peaks arise from Sm–I cross-vectors. The iodines were of low occupancy and did not generate obvious peaks in the Harker section.

The X-ray data were integrated and scaled with *HKL*2000 (Otwinowski & Minor, 1997). The intensities were put on an approximately absolute scale using the *CCP*4 program program *TRUNCATE* and were then used in structure determination and refinements.

2.4. Structure determination

Using merged data, the structure was determined *ab initio* using the direct-methods program *SIR*2002 (Burla *et al.*, 2002).

The structure was also determined *de novo* by SAD. In this case the program *SHELXD* (Schneider & Sheldrick, 2002) was used to determine the heavy-atom substructure using the anomalous differences to 1.2 Å. The program *SHELXE* (Sheldrick, 2002) was used to improve the phases by density modification and to extend the phases to the resolution limit of the native data.

The experimental maps from *SIR*2002 and *SHELXD/E* were used for automated model building with *ARP*/*wARP* (Perrakis *et al.*, 1999). The resulting models were corrected and completed manually. Refinement of the models began at the resolution limit using the program *SHELXL* (Sheldrick & Schneider, 1997).

3. Results and discussion

3.1. X-ray data

The single mutant L87M of P22 lysozyme was produced to augment the sulfur content of the P22 lysozyme, which is relatively low, consisting of three methionines and one cysteine. It has previously been suggested that leucine is the optimum residue for substitution with methionine (Lipscomb *et al.*, 1998; Gassner & Matthews, 1999). Our original goal was to make selenomethionine-containing protein and to perform either a MAD experiment with synchrotron radiation or an MIR experiment with in-house copper radiation using other selenomethionine-containing mutants. Unfortunately, crystals of the selenomethionine protein did not diffract, so we took the alternative approach of trying to crystallize the methionine-containing L87M mutant in the presence of a wide range of transition metals. Crystals grew as large monoclinic plates in the presence of zinc and yttrium as well as in the presence of samarium. Successful cryocooling was difficult, but about one out of ten crystals froze well and diffracted X-rays to atomic resolution.

Individual diffraction spots were seen to 0.95 Å, but the scaled and merged data had a resolution limit of 1.04 Å. The Wilson *B* value of only 8.2 Å^2 and mosaic spread of 0.2° suggested that this crystal may have yielded higher resolution data with longer exposure times.

The data are of high quality and completeness (Table 1). A plot comparing the observed Bijvoet differences with the estimated uncertainty in the measurements (Fig. 1*a*) suggested that a significant anomalous signal was present to high resolution.

Table 2
Effect of the assumed number of heavy atoms on the number of trials to a successful solution.

RAT is the figure of merit defined by Burla *et al.* (2002). wMPE is the weighted mean phase error relative to the reference structure (see text). The light-atom composition of the asymmetric unit is $C_{1418}H_{2320}N_{412}O_{430}S_8$.

No. of iodines assumed to be present in the asymmetric unit	Trials	*R* factor (%)	RAT	⟨wMPE⟩ (°)
0	727	31.0	1.97	58.5
2	12	23.9	2.74	24.9
3	66	23.8	3.02	24.6
4	90	23.7	2.73	24.5
5	28	23.8	2.90	24.6
6	25	23.7	2.94	24.3
10	7	24.8	1.97	27.8
15	128	24.4	2.24	26.9
20	22	24.6	2.30	26.7
30	137	24.9	2.05	27.3
50	13	23.3	2.81	23.8
100	54	23.8	2.52	24.5
200	8	24.2	2.50	23.3

Table 3
Refinement statistics for P22 lysozyme (mutant L87M) determined by *ab initio* direct methods (*SIR*2002) and by SAD (*SHELXD/E*).

Source of model	*SIR*2002	*SHELXD/E*
Resolution range (Å)	1.04–42	1.04–42
R factor (%)	11.4	11.4
R_{free} (%)	14.6	14.5
Non-H protein atoms (summed by occupancy)	2265.7	2262
Solvent atoms	696	651
Observations/restraints	141033/32273	141033/33096
Parameters	26563	27162
R.m.s.d. from ideal geometry		
Bond lengths (Å)	0.014	0.015
Angle distances (Å)	0.030	0.031
PDB code	2anv	2anx

3.2. *Ab initio* phasing by augmented direct methods

3.2.1. Phasing with *SIR*2002. *SIR*2002 (Burla, Carrozzini, Caliandro *et al.*, 2003; Burla, Carrozzini, Cascarano *et al.*, 2003) is based on the successful direct-methods program *SIR* for small-molecule structure determination. *SIR* implements the theory of semi-invariant representations, developed by Giacovazzo (1977), which makes full use of space-group symmetry in estimating structure invariants (origin-independent) and structure semi-invariants (origin-dependent). Density modification and other phase-refinement techniques have been incorporated to allow the extension of *SIR*2002 to proteins. *SIR*2002 starts with a set of random phases for the 4000 strongest *E* values. These phases are modified in reciprocal space using tangent-formula-based techniques. The refined phase set is then used to produce an electron-density map for further development in real space. This development includes multiple cycles of density modification, the assignment of atomic species to peaks in the electron-density map based on peak height, four cycles of refinement of the thermal factors of the heaviest atomic species and six cycles of isotropic displacement parameter refinement for all atoms. At this point, the figure-of-merit RAT described below is computed. If it is greater than the RAT value for previous trials, the trial is advanced to the next step, which includes four cycles of least-squares refinement alternated with $2F_o - F_c$ map calculations to refine and complete the structure. Eventually, peaks in the electron-density map are picked and assigned atom types based on peak height. The assigned positions are refined without any stereochemical restraints.

The crystallographic *R* factor is used along with the figure of merit RAT (Burla *et al.*, 2002) to identify the correct solution. RAT is defined by the ratio

$$\mathrm{RAT} = \mathrm{CC}/\langle E_{\mathrm{calc}}^2 \rangle.$$

The numerator (CC) is the correlation coefficient between the observed and calculated *E* values (including Sim-type weighting) and is calculated using $|E_{obs}|$ ranging in value from 0.3 to 1.2. These E_{obs} are too weak for use in the tangent-formula refinement, but they are included in the subsequent real-space techniques. $|E_{obs}|$ in this size range make up most of the *E* values in the top 70 percentile. The denominator includes the remaining 30% of the reflections. These are weak and not used in phasing. RAT should be numerically larger for more promising phase sets, but its value can change from structure to structure. In the present case RAT was generally close to 1.0 for incorrect solutions and 1.5–3.0 for correct ones. The identification of promising trials during phase extension in direct methods is often very difficult when a multi-solution algorithm is applied to macromolecular data (Gilmore, 1998), but RAT has proven to be a strong discriminator of promising phase sets (Burla *et al.*, 2002). Phase sets with high RAT values often have low crystallographic *R* factors. If the current trial structure has a crystallographic *R* factor greater than a preset threshold (25% in our case), *SIR*2002 generates a new set of random phases and repeats the procedure just outlined. The program continues to cycle through these

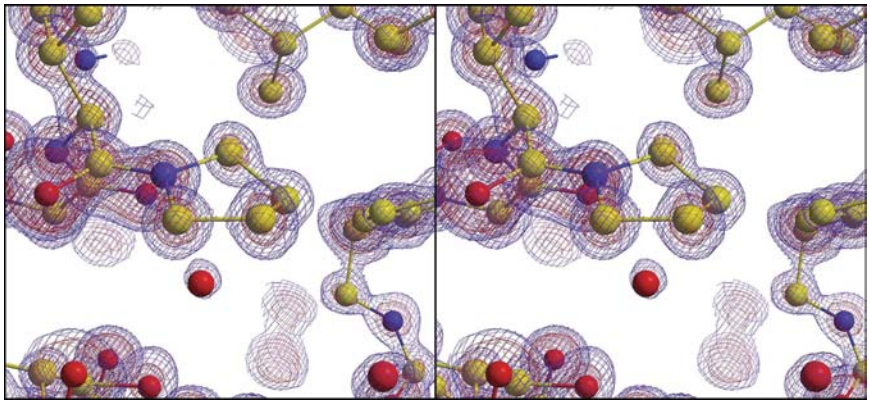

Figure 2
Direct-methods electron-density map at 1.04 Å resolution from the *ab initio* structure determination using *SIR*2002. The blue contours are at 2σ and the red contours are at 4σ. The figure shows Pro30 from molecule *A*, with the pucker of the pyrrolidine ring clearly visible.

steps until a solution is found or until the maxium number of trials is reached. *SIR*2002 differs from *SnB* and *SHELXD* in that it does not cycle between direct and reciprocal space while developing a set of phases.

3.2.2. Initial structure determination. A successful phase set was found on the 12th trial, assuming two iodide ions in the asymmetric unit (Table 2). The crystallographic *R* factor was 23.9% for data with $F > 4\sigma(F)$ and the figure of merit RAT had a value of 2.74. The next most favorable trial had a RAT value of 1.65 and the least favorable trial had a RAT value of 1.01. The successful trial was reached in 14.5 h with an ADM Athlon 2400+ processor and 2 Gb of random-access memory. This phase set had a weighted mean phase error of 24.5° compared with the refined reference structure determined independently by SAD at a later time (Tables 3 and 4) (see §3.3).

Part of the experimental map made with phases from the *SIR*2002 solution is shown in Fig. 2. The map was quite complete and was sufficiently accurate to reveal details such as the pucker in the pyrrolidine ring of Pro30. This map was used with *ARP*/*wARP* to automatically build a model that included 226 of the possible 292 residues.

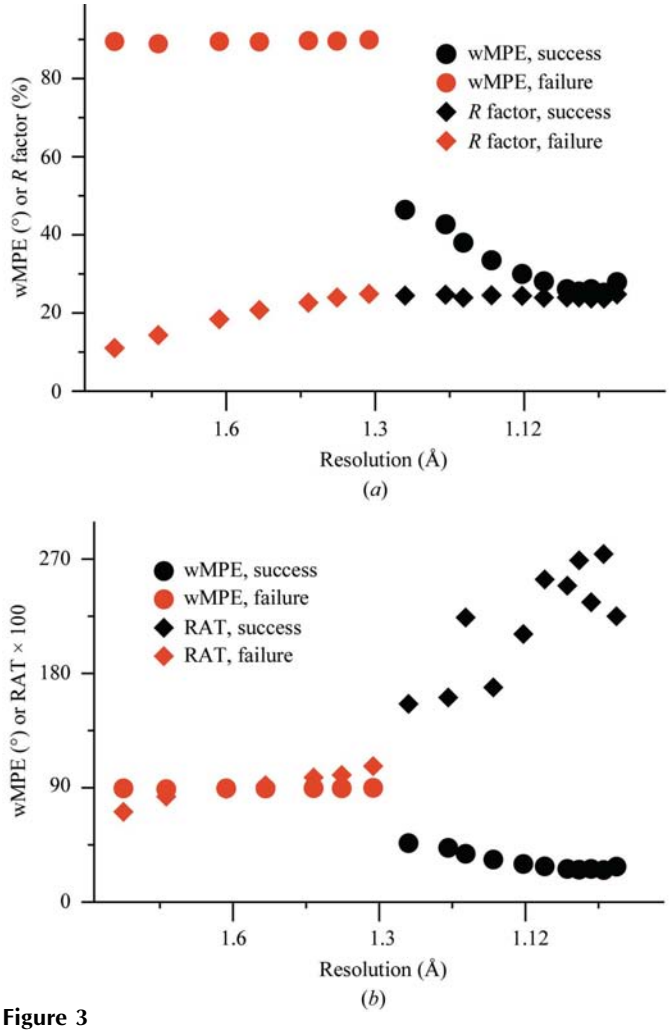

Figure 3
Quality of phase sets from direct-methods experiments with data truncated to various high-resolution limits. (*a*) Comparison of figure-of-merit weighted mean phase error (wMPE) with the *R* factor. (*b*) Comparison of wMPE with the *SIR*2002 figure-of-merit RAT. These figures show that the RAT value is much more effective than the *R* factor in discriminating between failed and successful phase determinations.

Table 4
Comparison with the reference structure of phase sets determined by direct methods and by SAD.

The map correlation coefficient was computed using equation (12) of Lunin & Woolfson (1993). Computations were performed with *SHELXPRO* v.2000-1. Cos(err) is the cosine of the difference in phase angle, FOM is the figure of merit, wMPE is the FOM-weighted mean phase error and FwMPE is the mean phase error weighted by both the figure of merit and the structure amplitude.

Solution	*Ab initio*	SAD
Resolution (Å)	42.1–1.04	42.1–1.04
Reflections	141033	141033
⟨Map CC⟩	0.894	0.903
⟨Cos(err)⟩	0.724	0.852
⟨FOM⟩	0.525	0.827
⟨wMPE⟩ (°)	24.5	18.1
⟨FwMPE⟩ (°)	19.3	16.6

3.2.3. Trials with *RELAX*. The *RELAX* procedure (Burla *et al.*, 2002) was tested to determine whether fewer trials could be used to reach a solution by reconsidering trial structures with molecular fragments that were correctly oriented but incorrectly located. Such trial structures originate from a systematic error in phase assignment during an early stage of the phasing process. This error is extended into other phases. Such trials are identified by their intermediate RAT values. The reflections are expanded to *P*1 to relax the symmetry constraints. A search is then made for the correct origin translation. The correct origin shift is applied to the *P*1 phases and the modified phases are applied to the unique reflections in the original *C*2 space group.

A correct solution was found on the 13th trial in 14 h 23 min with an *R* factor of 24.5% and a RAT value of 1.75. Thus, there was no real advantage in using the *RELAX* procedure with these data.

3.2.4. Sensitivity to the assumed content of the unit cell. In the first step of refinement of the phases in real space, *SIR*2002 uses the user-supplied number and type of atoms expected in the unit cell in the assignment of atoms to peaks in the electron-density map. The importance of the completeness and accuracy of this list has not been explored extensively. Even with the aid of anomalous scattering or isomorphous replacement data, it can be difficult to assess the number of heavy atoms in a protein crystal. In addition, a purely *ab initio* solution by ADM precludes the use of such data.

In the present case, if no heavy atom was included a successful solution was not obtained until trial 727 (967 h; Table 2). Even after this number of cycles, the *R* factor was high (31.3%) and the RAT value was marginal (1.97). The phase set had a wMPE of 58.5°. This illustrates the importance of including at least some heavy atoms in the unit-cell contents.

The assumption of two iodide ions made in the first trial with *SIR*2002 lead to a successful solution, as mentioned above (Table 2). The assumption of a single I^- in the asymmetric unit led to a solution after 578 trials. Guesses closer to

Table 5
The effect of using different resolution ranges with the program *SIR*2002 on the number of cycles to a successful solution.

RAT is defined in the text. The value in parentheses is the highest value obtained for a wrong solution. The trials were carried out with the Bijvoet pairs merged and kept separate. At the time of these tests, the heavy-atom substructure was assumed to be three Y^{3+} and five I^- per asymmetric unit. This was used for all trials except for the full data set (1.04–42.1 Å), in which the heavy-atom substructure consisted of two iodines per asymmetric unit.

	Bijvoet pairs merged				Bijvoet reflections treated separately			
Resolution range (Å)	Trials	RAT	*R* factor (%)	⟨wMPE⟩ (°)	Trials	RAT	*R* factor (%)	⟨wMPE⟩ (°)
High-resolution truncations								
1.04–42.1	12	2.74 (1.66)	23.9	27.9	4	2.44 (1.25)	24.3	29.6
1.12–42.1	141	2.11 (1.50)	24.4	30.0	216	2.10 (1.33)	24.7	34.2
1.18–42.1	36	1.69 (1.24)	24.5	33.5	110	1.92 (1.16)	24.9	40.4
1.25–42.1	55	1.56 (1.10)	24.4	46.4	39	1.65 (1.07)	24.7	48.6
1.35–42.1	8	1.00 (N/A)	24.0	89.6	7	1.26 (0.98)	23.7	60.5
Low-resolution truncations								
1.04–1.8	37	2.08 (1.65)	19.1	18.3				
1.04–2.0	157	2.39 (1.90)	18.9	18.0				
1.04–3.0	47	2.31 (1.72)	21.7	23.4				
1.04–6.0	193	2.86 (2.20)	19.0	22.8				
1.04–10.0	11	2.65 (1.76)	22.0	23.6				
High- and low-resolution truncations								
1.09–10.0	70	2.48 (1.62)	22.2	26.5				
1.20–10.0	67	1.85 (1.39)	23.9	37.5				
1.25–10.0	85	1.51 (1.25)	24.9	43.4				
1.30–10.0	5	1.13 (1.07)	24.9	85.8				

the actual heavy-atom content (eight iodide and three samarium ions) tended to reduce the number of trials required to reach a solution. On the other hand, gross overestimation of the number of heavy atoms also led to solutions in a reasonable number of trials. In spite of a large number of heavy atoms in the atom list, only a handful of the highest peaks were assigned as heavy atoms in the final coordinates written out by *SIR*2002. As a result, the phase sets were of similar high quality as reflected in their high RAT values and their low *R* factors and mean phase errors. The insensitivity of the final phases to the overestimation of the heavy-atom content suggests that it may be better to overestimate rather than underestimate the number of heavy atoms in the unit cell.

3.2.5. Sensitivity to the high-resolution limit. With small-molecule structures, direct methods are not expected to succeed 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 empirical rule appears to apply to proteins, although the presence of heavy atoms may allow it to be relaxed. The diffraction data for L87M are almost 100% complete (Table 1), with 86.7% in the range 1.1–1.2 Å having $|F| > 4\sigma(|F|)$, which easily satisfies Sheldrick's rule.

The ability of *SIR*2002 to obtain a solution was expected to be sensitive to the resolution limit, since tangent-formula-based methods only succeed when a sufficient number of reliable triplet invariants have been found. The reliability of the phase-triplet invariants is known to decline with reduction of the high-resolution limit (Cochran, 1955). The high-resolution limit was successively reduced from 1.04 to 2.0 Å (Fig. 3) and the structure-determination trials were repeated starting with Wilson scaling and computation of the normalized structure factors. The phase error (wMPE) between the *SIR*2002 phases and the phases calculated from the final refined structure increases as the high-resolution limit is truncated to a resolution of 1.25 Å (Fig. 3*a*). At lower resolutions, the phase error jumps to 90°, indicating random phases and hence a false solution. The results of these trials demonstrate that high-resolution data are critical to a successful solution and that the quality of the solution improves as the resolution limit increases.

We also compared the ability of the *R* factor and RAT to discriminate between true and false solutions as the high-resolution limit was decreased (Figs. 3*a* and 3*b*). The *R* factor remained constant until false solutions were selected and then dropped steadily, suggesting that a correct solution has been found which was not the case. Clearly, the *R* factor is unreliable as a figure of merit at less than atomic resolution. On the other hand, RAT (Fig. 3*b*) dropped to a value of about 1.0 with the first false solution, thereby demonstrating its reliability in detecting correct solutions with atomic resolution data (1.2 Å or less). However, it should be noted that the maximum RAT declined with decreasing resolution, as did the gap between the maximum RAT and the highest RAT value for a false solution. In other words, the discriminating power of RAT declined as the high-resolution limit decreased. It should also be noted that the magnitude of the RAT is meaningless without knowledge of the RAT values of previous trials. For example, the RAT value for the correct solution with the data truncated to 1.25 Å is 1.56, which is similar to the highest RAT value (1.50) for a false solution with the data truncated to 1.12 Å (Table 5).

Since a new set of random phases is chosen at the start of each trial, the number of trials required to reach a solution is expected to be stochastic. We found that the number of cycles varied substantially, with the 1.12 Å truncated data requiring 141 trials to reach a solution and the 1.18 Å truncated data requiring 36 trials. These results suggest that a strategy for obtaining a solution in difficult cases may be to launch multiple experiments in parallel with different high-resolution limit cutoffs.

3.2.6. Sensitivity to the low-resolution limit. Low-resolution reflections are thought to play an important role in structure determination because their absence can lead to Fourier ripples in the electron-density map. Such reflections are insensitive to fine structural detail, but they do contain information about the solvent and the location and shape of the molecular envelope. Before the start of the phasing process, *SIR*2002 uses the solvent model arising from Babinet's principle (Tronrud, 1997) to subtract the solvent contribution. For some reflections the original F_{obs} are too small to be included in the phasing process, but the modified F_{obs} are large enough to be included. The presence or absence of such reflections can alter the direction of the phasing process.

Ab initio direct-methods trials were attempted with the low-resolution limit set to a series of values between 1.8 and 30 Å (Table 5). The highest low-resolution limit that led to a solution in a reasonable number of trials was 1.8 Å. The resulting electron-density map was quite noisy owing to Fourier ripples from the absence of low-angle terms, but the peaks at the atomic sites were clear and the electron density along the bonds was weak. The map resembled an E_{obs} map more than an F_{obs} map as would be expected owing to the lack of medium-resolution data. Nonetheless, *ARP*/*wARP* was able to automatically locate the atomic positions and trace the chain.

Minima in the *R* factor occurred with cutoffs of 6.0 Å (19.0%) and 2.0 Å (18.9%). The 6.0 Å cutoff has the highest RAT value. The exclusion of the lower resolution data also led to a smaller wMPE.

Taken together, these results suggest that *SIR*2002 does not require low- to medium-resolution data to be effective. The medium-resolution data contain information about 1,3-distances (the distance between atoms linked *via* an intervening atom; 2.2–2.5 Å). Reflections from Miller planes with interplanar distances of less than 1.8 Å contain information about bond lengths and individual atoms. This information is sufficient for *SIR*2002 to succeed.

The ability of *SIR*2002 to obtain solutions with high-resolution truncations in the presence of a low-resolution cutoff of 10 Å was tested in a series of experiments (Table 5). It was still possible to obtain a solution with the high-resolution data truncated to 1.25 Å, although the phases were of low quality. In general, the RAT values for the successful solutions declined as the high-resolution data were truncated, while the phase error increased. These same trends were seen when low-resolution data were included to 42.1 Å. Interestingly, the test with a low-resolution cutoff of 10 Å led to a solution on the 11th trial with an *R* factor of 21.97% and an RAT of 2.65. This is the most promising figure of merit of all of the trials explored here. It suggests that in difficult cases phasing experiments with low-resolution truncations may lead to easier identification of a correct phase set and that further improvement of the modeling of the bulk solvent in *SIR*2002 may improve the chances of a successful structure determination. In summary, with a 10 Å inner resolution limit solutions were still possible to a resolution of 1.25 Å and the map accuracy depends strongly on the outer resolution limit.

Further enhancements of *SIR*2002 or similar programs may be achieved by improved solvent models and by the incorporation of *a priori* chemical information about 1,3-distances. The incorporation of such information may help extend direct methods to lower resolution data and to larger structures without heavy atoms present. In addition, ADM may work using *SIR*2002 in cases where the low-resolution reflections are not available because of detector saturation, presence of ice rings or other reasons.

3.2.7. Influence of the heavy atoms. How do heavy atoms in a protein crystal influence the probability of success with *SIR*2002? The presence of electron-dense ions such as Sm^{3+} or I^- violates the equal-atom assumption that underlies direct methods and at first glance would be expected to hinder progress. The presence of heavy atoms amongst a modest number of light atoms in a centrosymmetric crystal has been shown to skew the distribution such that more *E* values have large magnitudes (Shmueli, 1982). This skewing can be dramatic enough to cause such a centrosymmetric crystal to give an acentric distribution of *E* values to that expected from a non-centrosymmetric crystal. The presence of the few heavy atoms amongst over 2000 light atoms in the structure reported here did not appear to skew the probability density function (pdf; not shown). The same was true for the pdf of *E* values calculated from the crystal structure with the heavy atoms removed. The main difference between the two pdfs was the presence of more *E* values in the histogram bins near the peak of the distribution. These *E* values are well below the threshold for the top 4000 *E* values used in calculating the triplets.

The number of large *E* values decreases with high-resolution limit truncations. For example, the threshold for the top 4000 largest *E* values was 1.927, 1.745 and 1.388 at resolution limits of 1.04, 1.25 and 1.8 Å, respectively. This clearly demonstrates the importance of atomic resolution data. In this

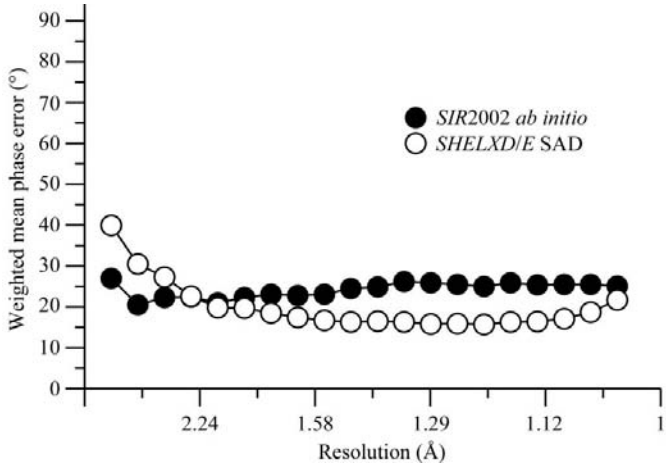

Figure 4
Comparison of figure-of-merit-weighted phases with the refined reference structure (see text). Solid circles, phases obtained by direct methods. Open circles, phases obtained by SAD. In addition, the experimental phase sets are cross-validated by comparison with the calculated phases angles from the refined structure determined by the alternative method.

context, the number of very large E values is expected to increase with increasing resolution in the presence of heavy atoms because the scattering factor of electron-dense atoms decreases more slowly than those of the lighter atoms. For example, the unitary scattering factor of the lanthanides at atomic resolution is twice that of carbon. Since the effective contribution to the overall scattering is proportional to the number of electrons squared, this effect is magnified. For example, at high resolution an Sm^{3+} cation makes, in effect, a 400-fold larger contribution than a C atom.

3.2.8. Trials using Bijvoet pairs. *SIR*2002 has the ability to use separated Bijvoet pairs in *ab initio* direct methods in a way that is very similar to how it uses merged (*i.e.* Bijvoet-averaged) data. The probability relationships in the presence of anomalous scatterers were developed over two decades ago (Hauptman, 1982; Giacovazzo, 1983). Briefly, in the case of merged data, the most common phase relationship is the well known triple-phase relation ($\varphi_{\mathbf{h}} = \varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}$; Cochran, 1955), where **h**, **k** and (**h** − **k**) are three related reflections. In the case of separated Bijvoet pairs, the basic phase relationship depends on six reflections, *i.e.* **h**, **k**, (**h**−**k**) and their Friedel mates. The user must supply the real and imaginary parts of the complex scattering factor of the anomalous scatterers. Because the use of separated Bijvoet pairs almost doubles the number of intensities used by *SIR*2002, some of the computational procedures take longer. However, it is hoped that the use of the separated Bijvoet pairs makes fuller use of the information content of the X-ray data and thus increases the chance of obtaining a successful solution. Since the Bijvoet

Table 6
Triplet statistics for successful and unsuccessful trials.

Resolution range (Å)	1.04–42	1.25–42	1.35–42	1.8–42
E_{min} for top 4000 E values	1.93	1.74	1.68	1.39
Initial strong triplets	169400	257053	300000	300000
Strong triplets after P_{10} formula	83595	119871	149752	44230
Trial successful?	Yes	Yes	No	No

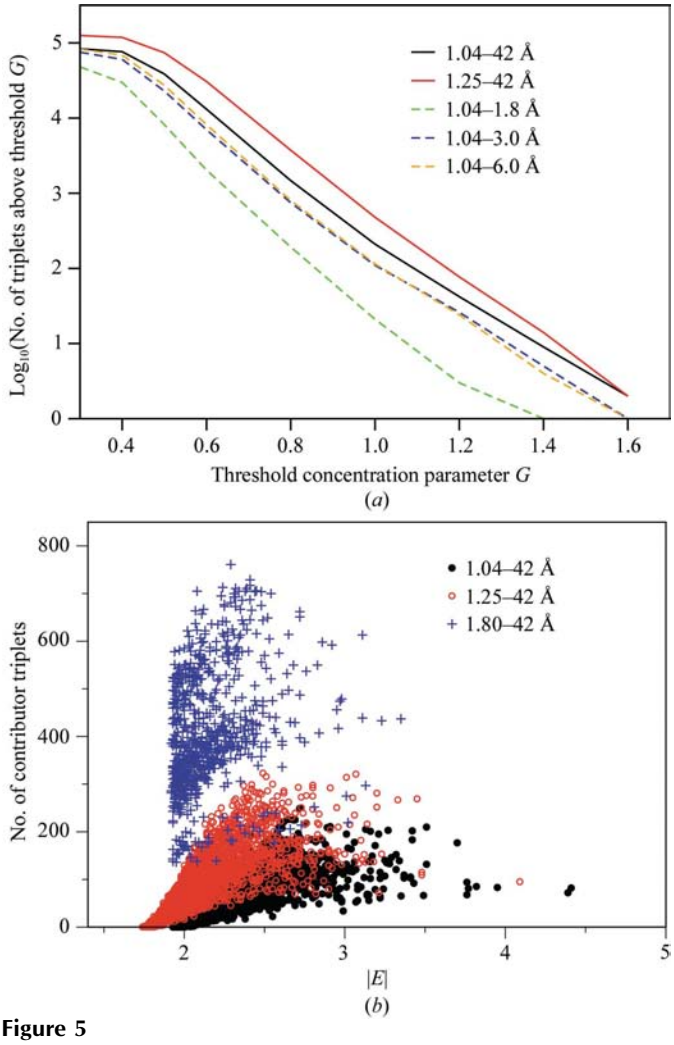

Figure 5
Effect of resolution cutoff on the triplets used in tangent-formula refinement of the starting phases. (*a*) The log of the number of triplets *versus* the concentration parameter threshold (see text). The reliability of the triplet increases with larger concentration parameters. (*b*) The scatter of the number of triplets to which each E belongs *versus* the magnitude of the E. The triplets from the data truncated to 1.8 Å failed to lead to a solution.

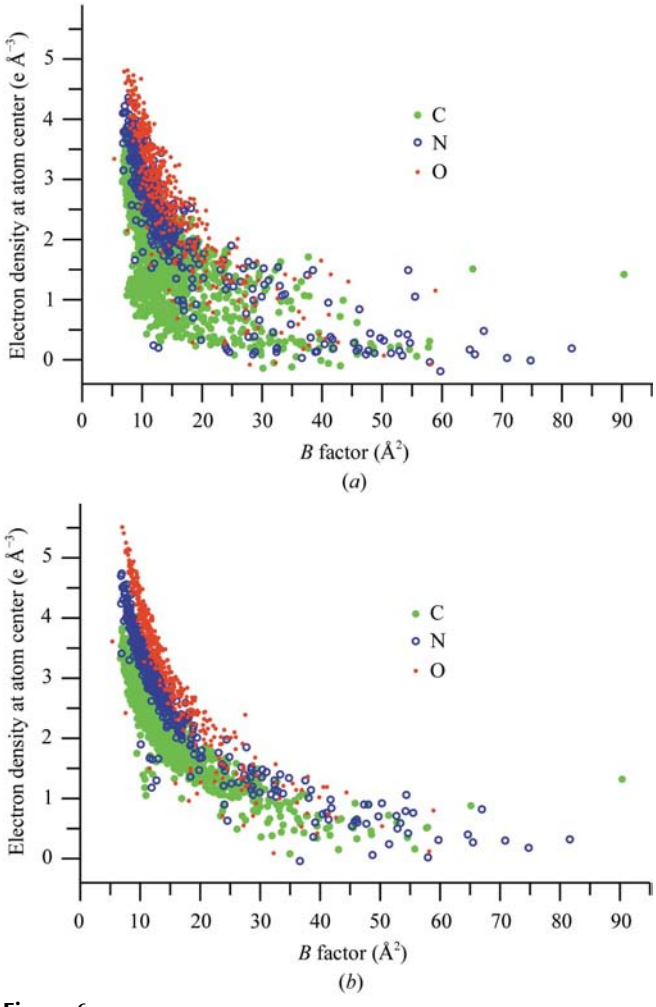

Figure 6
(*a*) Plot of peak density in the *ab initio* electron-density map *versus* the B factor in the fully refined model of P22 lysozyme. C atoms green, N blue and O red. (*b*) Plot similar to (*a*) but for the electron density in a σ_A-weighted $2F_o - F_c$ map based on the final refined P22 lysozyme model.

differences are not used directly to locate the anomalous scatterers (*e.g. via* Patterson methods), this approach still qualifies as a pure *ab initio* direct-methods technique.

When this approach was used with the initial atom list including two iodines, a solution was reached in four rather than 12 trials (Table 5). However, when a complete atom list was used there was little if any reduction in the number of trials required to reach a solution and there was no apparent improvement in the figures of merit for the final phase sets. (It was possible to obtain a solution with data truncated to 1.35 Å, 0.1 Å less than for the merged data.) Thus, the separation of the Bijvoet pairs may make the difference between success and failure in marginal cases, but the results suggest that this should only be tried as a last resort.

3.2.9. Resolution limits and the quality of phase triplets. The key to a successful solution with *SIR*2002 is the use of a favorable set of triplets in the tangent-formula refinement of the starting phases. A triplet can be estimated with the Cochran formula (Cochran, 1955), which depends on three moduli: $|E_{\mathbf{h}}|$, $|E_{\mathbf{k}}|$ and $|E_{\mathbf{h}-\mathbf{k}}|$. This formula often overestimates the reliability of the triplet. Improved identification of good and bad triplets is accomplished in *SIR*2002 by exploiting all the information in reciprocal space through the P_{10} formula (Cascarano *et al.*, 1984). The subscript refers to the consideration of ten moduli at a time in the triclinic case. When the space group has symmetry higher than triclinic, a greater number of moduli are considered. The P_{10} formula gives a new von Mises distribution which is related to the well known Cochran distribution (Cochran, 1955). The concentration parameter of the new distribution is $G = C\,(1 + Q)$, where Q is a function of all of the magnitudes of the E values in the second representation of the triplet (Cascarano *et al.*, 1984) and C is the concentration parameter of the Cochran distribution. The latter formula tends to overestimate the probably of the invariants, while the P_{10} formula is superior in identifying useful sets of invariants. The triplets are sorted by their G values and those with G values greater than a threshold value (usually 0.3) are used in tangent-formula refinement of the random phases. The quality of a set of triplets is expected to depend on the reliability of the individual triplets and on

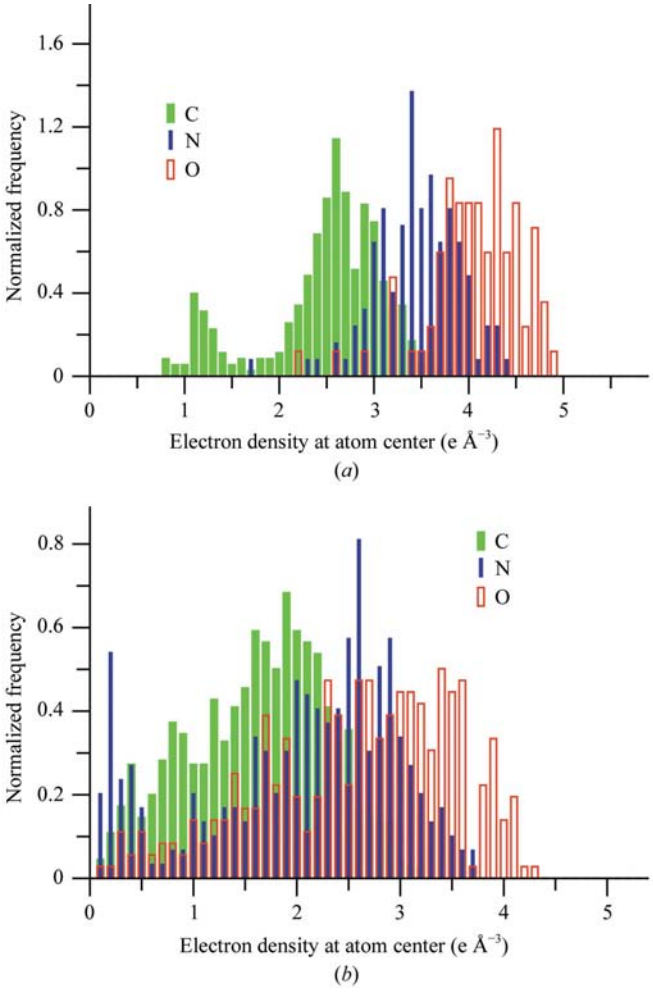

Figure 7
Normalized frequency of electron density at atomic centers in the *SIR*2002 *ab initio* map. The final refined model was the source of the atomic positions. The atoms were divided into (*a*) those with $B < 10$ Å^2 and (*b*) those with $B > 10$ Å^2.

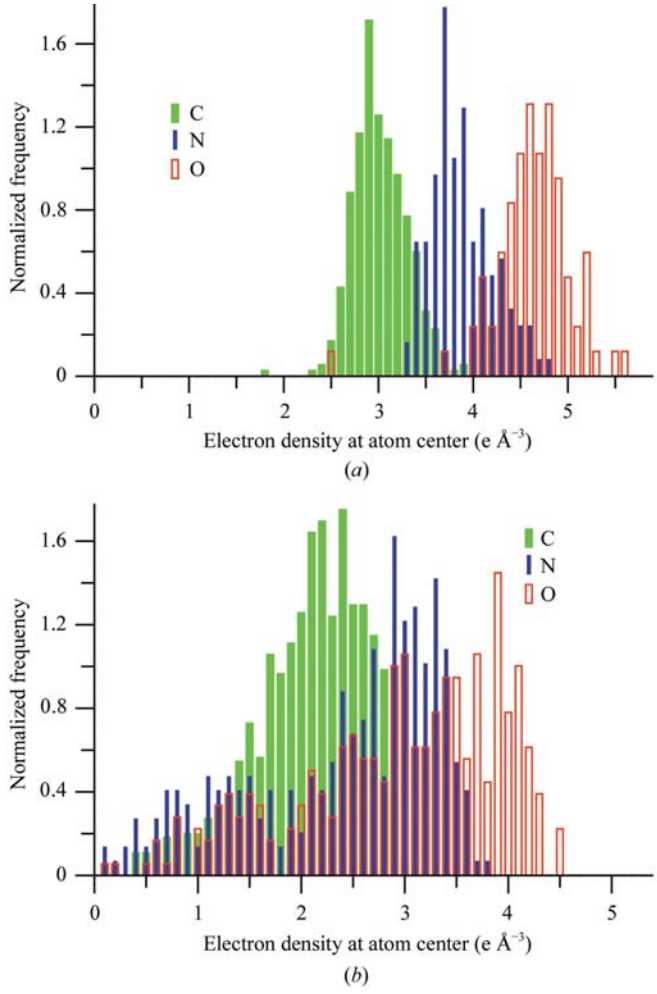

Figure 8
Normalized frequency of electron density at atomic centers in in the σ_A-weighted $2F_o - F_c$ map. The final model was the source of the atomic positions. The atoms were divided into (*a*) those with $B < 10$ Å^2 and (*b*) those with $B > 10$ Å^2.

the number of triplets to which each E belongs because this reflects the degree to which the phase angle of the E is over-determined.

There is an inverse relation between the number of triplets and the threshold concentration parameter G (Fig. 5*a*). The number of triplets is larger when the high-resolution limit is truncated to 1.25 Å (Table 6). In other words, the inclusion of data beyond 1.25 Å led to fewer triplets. On the other hand, low-resolution truncations reduced the number of triplets. The deleted low-resolution E values contributed to a large number of triplets. A severe truncation of the data to 1.8 Å led to a greater loss of the more reliable triplets.

The inclusion of higher resolution data is associated with higher cutoffs for the minimum E value in the top 4000 E values used to generate the initial triplets (Table 6). After the application of the P_{10} formula, there were 3400–3600 strong E values contributing to one or more triplets in the data sets that led to successful structure determination. On the other hand, there were only 2774 and 740 strong E values in the data sets truncated to 1.35 and 1.8 Å, respectively. Both of these latter cases failed to give the correct structure. The E values in the higher resolution triplet sets tended to belong to fewer triplets (Fig. 5*b*) and yet gave more accurate final phases. The largest E values did not necessarily contribute to the largest number of triplets. On the other hand, each of the E values in the lower resolution sets often contributed to a larger number of triplets (Fig. 5*b*) than when all of the data were used. The population of larger E values associated with higher resolution data led to more useful sets of triplets as reflected in the lower mean phase error but not reflected in the number of triplets above the threshold G value nor the number of contributing triplets per large E.

3.3. *De novo* determination of reference structure by SAD

After the determination of the structure by ADM as described below, a SAD experiment was performed to obtain a reference structure for comparison purposes. A Patterson map made with the Bijvoet differences revealed several significant peaks (Fig. 1*b*). (The f'' values for Sm^{3+} and I^- are 5.3 and 2.8 e, respectively.) Using *SHELXD* with data to a resolution of 1.2 Å, 11 'heavy-atom' sites were found. After heavy-atom refinement, density modification and phase extension in *SHELXE*, the phase set with the correct hand had an average map correlation coefficient (Lunin & Woolfson, 1993) of 0.94, a map contrast of 0.703, a map connectivity of 0.859 and a pseudo-free correlation coefficient of 85.75%. This phase set was used with *ARP*/*wARP* (Perrakis *et al.*, 1999) to automatically build 226 of 292 residues in 50 cycles. There were four gaps, the longest being in the N-terminal domain, where parts of the main-chain lacked continuous density. The model was completed manually and refined against the merged data with *SHELXL* following the standard protocol (Sheldrick & Schneider, 1997). The final refined model had an R factor of 12.3% and acceptable geometry (Table 2). The SAD phase set had an average figure-of-merit weighted phase error of 18° compared with the final refined model (Fig. 4).

3.4. Effectiveness of quick soaks with halides

The quick-soak procedure was successful in introducing eight iodide ions into the asymmetric unit. At the same time, the fact that Sm^{3+} cations were already present made it difficult to assess the effectiveness of the iodide ions. However, the fact that the crystals still diffracted to atomic resolution following a quick soak in a high concentration of NaI does suggest that this can be a useful way to improve the effectiveness of *ab initio* phase determination.

3.5. Metals as crystallization additives

Some transition metals are already included in commercial additive screens (*e.g.* Hampton Research Additive Screens). Our success with samarium suggests that a broad range of transition metals and lanthanides should be explored as crystallization additives. Metal–protein cocrystals have the advantage of having electron-dense atoms ready for phasing by SAD, MAD or ADM.

A similar proposal was made by Qiu & Janson (2004). They proposed engineering surface Asp and Glu mutants and crystallizing them in the presence of transtion metals such as zinc from the fourth row of the periodic table. Our results suggest that lanthanides should also be considered.

3.6. Model validation at atomic resolution: MAD, SAD or ADM?

Structures refined with atomic resolution data are in principle defined by the X-ray data alone. In practice, however, stereochemical restraints are often retained for those parts of the structure that are not well ordered. In addition, the poorly ordered parts of the structure are more susceptible to errors in interpretation of the electron density. Consequently, these parts can suffer from model bias. The model can also acquire errors from the application of incorrect restraints during model refinement (Podjarny *et al.*, 2003)

There are now several examples of experimental phases determined to atomic resolution from anomalous scattering (MAD or SAD; Brodersen *et al.*, 2000; Schmidt *et al.*, 2002; Podjarny *et al.*, 2003). The experimentally phased maps are of quality comparable to the final maps from refined models. These maps have been used to validate features of the solvent and poorly ordered main-chain and side-chain atoms (Podjarny *et al.*, 2003).

To assess the potential value of the *ab initio* electron-density map for model validation, the agreement with the final model was evaluated by checking the electron density at atomic sites in the model (Fig. 6*a*). The *ab initio* peak densities for N and O atoms decrease exponentially with the atomic temperature factor in the final model. The distribution of peak densities of C atoms deviate somewhat from exponential decay because the *ab initio* map had either low electron-density values or larger positional error for some C atoms. The deviation of the data points for carbon is striking when compared with the

same correlation between peak density values at atomic positions in an electron-density map made with the final model (Fig. 6*b*). C, N and O atoms are easier to distinguish if more accurate phases, lower *B* factors and higher resolution data are all available. As shown in Fig. 7(*a*), oxygen and carbon sites can be readily distinguished when their *B* factors are less than 10 Å^2. However, nitrogen sites could not be readily distinguished from carbon and oxygen sites (Fig. 7*a*). In contrast, all three atomic species were reasonably distinguishable in the well ordered parts of the σ_A map (Fig. 8*a*). In contrast, little if any differentiation between the three atom types was possible in the less-ordered parts of either map (Figs. 7*b* and 8*b*).

The *ab initio* map contained several errors. The distribution of electron density for the well ordered C atoms had an unexpected secondary peak that was absent in the refined map (Figs. 7*a* and 8*a*). In addition, the less-ordered N atoms had a similar secondary peak (Figs. 7*b* and 8*b*). These secondary peaks are a consequence of errors in atomic position and misassignment of the atom type during the real-space phasing in *SIR*2002. Notwithstanding such errors, the *ab initio* map is independent of stereochemical information and restraints and so has value for model validation.

In contrast, a traditional experimentally phased electron-density map can have errors arising from low heavy-atom occupancy, weak anomalous scattering and heavy-atom pseudosymmetry among many other factors. Because the *ab initio* and experimental maps have different sources of error, they may complement each other in model validation. Also, ADM does not require carefully measured Bijvoet pairs at precisely defined wavelengths so the data-collection requirements may not be as stringent as those for MAD or SAD. Thus, ADM may provide a useful source of phases where the structure is already known but a set of unbiased phases is desired.

3.7. Comparison with other large structures determined by direct methods

Following a strict definition in which the phasing procedure starts from random phases, the largest structure determined *ab initio* by direct methods is bovine pancreatic ribonuclease (PDB code 1dy5), which has 1910 non-H protein atoms, 25 sulfurs (16 are in eight disulfide bonds) and 0.87 Å data (Burla, Carrozzini, Caliandro *et al.*, 2003; Burla, Carrozzini, Cascarano *et al.*, 2003). The largest light-atom protein without disulfides is cutinase with 1441 non-H protein atoms, five sulfurs and 1.0 Å data (Foadi *et al.*, 2000; Burla, Carrozzini, Caliandro *et al.*, 2003; Burla, Carrozzini, Cascarano *et al.*, 2003).

4. Summary

The presence of heavy atoms has allowed the extension of direct methods to determine the structure of a 2268 atom non-metalloprotein using data to 1.04 Å resolution.

Where the number of heavy atoms is unknown, as is typically the case, relatively inaccurate estimates can still lead to successful solutions. Overestimating the number of heavy atoms seems preferable to underestimation.

Trials with truncated data sets showed no clear relation between the number of trials and resolution limit, but did suggest that in difficult cases a range of resolution limits should be tested.

Keeping the Bijvoet pairs separate in direct-methods trials in general did not have advantages over using the merged data.

The use of Bijvoet pairs in a SAD structure determination by *SHELXD/E* was faster and gave somewhat higher quality phases than ADM. Nonetheless, *ab initio* structure determination by direct methods is a viable alternative to traditional protein crystallography methods provided that the data extend to atomic resolution and that there are heavy atoms of sufficient scattering power present in the crystal.

We thank Leslie Gay for overexpressing and purifying the mutant L87M and the user-support staff at ALS, especially Dr Cory Ralston, for their assistance with our visit to beamline 8.2.2. Supported in part by NIH grant GM21967 to BWM.

References

Brodersen, D. E., de La Fortelle, E., Vonrhein, C., Bricogne, G. & Kjeldgaard, M. (2000). *Acta Cryst.* D**56**, 431–441.
Burla, M. C., Carrozzini, B., Caliandro, R., Cacarano, 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, B., Cascarano, G. L. & Polidori, G. (2002). *Z. Kristallogr.* **217**, 629–635.
Cascarano, G., Giocovazzo, C., Camalli, M., Spagna, R., Burla, M. C., Nunzi, A. & Polidori, G. (1984). *Acta Cryst.* A**40**, 278–283.
Cochran, W. (1955). *Acta Cryst.* **8**, 473–478.
Dauter, Z. (2003). *Methods Enzymol.* **368**, 288–337.
Eriksson, A. E., Baase, W. A. & Matthews, B. W. (1993). *J. Mol. Biol.* **229**, 747–769.
Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Yao, J.-X. & Chao-de, Z. (2000). *Acta Cryst.* D**56**, 1137–1147.
Frãzao, C., Sieker, L., Sheldrick, G., Lamzin, V., LeGall, J. & Carrondo, M. A. (1999). *J. Biol. Inorg. Chem.* **4**, 162–165.
Gassner, N. C. & Matthews, B. W. (1999). *Acta Cryst.* D**55**, 1967–1970.
Giacovazzo, C. (1977). *Acta Cryst.* A**33**, 933–944.
Giacovazzo, C. (1983). *Acta Cryst.* A**39**, 585–592.
Gilmore, C. J. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 159–167. Dordrecht: Kluwer Academic Publishers.
Hauptman, H. A. (1982). *Acta Cryst.* A**38**, 632–641.
Lipscomb, L. A., Gassner, N. C., Snow, S. D., Eldridge, A. M., Baase, W. A., Drew, D. L. & Matthews, B. W. (1998). *Protein Sci.* **7**, 765–773.
Liu, Q., Huang, Q., Teng, M., Weeks, C. M., Jelsch, C., Zhang, R. & Niu, L. (2003). *J. Biol. Chem.* **278**, 41400–41408.
Lunin, V. Y. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 530–533.
Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A. (1993). *Science*, **259**, 1430–1433.
Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
Mooers, B. H. M. & Matthews, B. W. (2004). *Acta Cryst.* D**60**, 1726–1737.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Podjarny, A., Schneider, T. R., Cachau, R. E. & Joachimiak, A. (2003). *Methods Enzymol.* **374**, 321–341.
Qiu, X. & Janson, C. A. (2004). *Acta Cryst.* D**60**, 1545–1554.
Rennell, D. & Poteete, A. R. (1985). *Virology*, **143**, 280–289.
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. & Gould, R. O. (1995). *Acta Cryst.* B**51**, 423–431.
Sheldrick, G. M., Hauptman, H. A., Weeks, C. M., Miller, R. & Usón, I. (2001). *International Tables for Crystallography, Vol. F*, edited by E. Arnold & M. G. Rossmann, pp. 333–351. Dordrecht: Kluwer Academic Publishers.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Shmueli, U. (1982). *Acta Cryst.* A**38**, 362–371.
Tronrud, D. E. (1997). *Methods Enzymol.* **277**, 306–319.
Usón, I. & Sheldrick, G. M. (1999). *Curr. Opin. Struct. Biol.* **9**, 643–648.
Weeks, C. M. & Miller, R. (1999). *J. Appl. Cryst.* **32**, 120–124.