Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Ping-pong cross-validation in real space: a method for increasing the phasing power of a partial model without risk of model bias

John F. Hunt[a,b]*† and Johann Deisenhofer[a]

[a]Howard Hughes Medical Institute and Department of Biochemistry, The University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-9050, USA, and [b]Department of Biological Sciences, 702A Fairchild Center, MC2434, Columbia University, New York, NY 10027, USA

† Present address: Department of Biological Sciences, 702A Fairchild Center, MC2434, Columbia University, New York, NY 10027, USA.

Correspondence e-mail: hunt@sid.bio.columbia.edu

Received 11 April 2002
Accepted 15 October 2002

PDB References: nucleotide-free SecA, 1m6n, r1m6nsf; Mg-ADP-bound SecA, 1m74, r1m74sf.

Experimental phases could only be obtained to 4.4 Å resolution for crystals of the SecA translocation ATPase. Density modification of these phases exploiting the 65% solvent content of the crystal produced a map from which an approximate backbone model could be built for 80% of the structure. Combining the phases inferred from this partial model with the MIR phases and repeating the density modification produced an improved map from which a more complete backbone model could be built. However, this procedure converged before yielding a map, that allowed unambiguous sequence assignment for the majority of the protein molecule. In order to avoid the likely model bias associated with a speculative attempt at sequence assignment, a real-space cross-validation procedure was employed to facilitate completion of the crystal structure based on partial model phasing. The protein was partitioned into two disjoint sets of residues. Models in which the side chains were built for residues in one of the two sets were used for phase combination and density modification in order to produce improved electron density for interpretation of residues in the other set that had not been included in the model. Residues in the two sets were therefore omitted from the model in alternation except at sites where the side chain could be identified definitively based on phasing with the other set. This ping-pong cross-validation procedure allowed partial model phasing to be used to complete the crystal structure of SecA without being impeded by model bias. These results show that the structure of a large protein molecule can be solved with exclusively low-resolution experimental phase information based on intensive use of partial model phasing and density modification. Real-space cross-validation can be applied to reduce the risk of model bias associated with partial model phasing, streamlining this approach and expanding its range of applicability.

1. Introduction

The phase problem in crystallography derives from the fact that both the amplitude and relative phase offset of the diffracted radiation waves need to be known for Fourier synthesis of an electron-density map, but only the amplitudes can feasibly be measured experimentally (Bricogne, 1984; Hauptman, 1986). While powerful techniques have been developed for *ab initio* phasing of crystal structures when very high resolution data are available (Hauptman, 1986), these techniques are rarely applicable to the solution of macromolecular crystal structures. Various techniques have therefore been developed for 'experimental phasing' of macromolecular crystal structures which involve perturbation of the native diffraction amplitudes either by derivatizing the crystal with a small number of strongly diffracting atoms (Crick & Magdoff, 1956; Blow & Rossmann, 1961) or by exploiting the X-ray resonant properties of a small subset of atoms in the lattice (Hendrickson, 1991). The partial structure of the atoms giving rise to the perturbation is determined using *ab initio* techniques (either Patterson analysis or direct methods) and the relative phase offset of all of the diffracted waves is then estimated based on the magnitude of the perturbation of their amplitudes by the partial structure. Because the experimental phases determined in this way are typically of limited accuracy, the final stage of the phase-determination process generally makes use of iterative electron-density modification procedures (Wang, 1985; Cowtan & Main, 1996, 1998; Abrahams & Leslie, 1996) which involve computational manipulation of the electron-density map in real space in order to satisfy empirical constraints and then inverse Fourier transformation for refinement of the estimates of the relative phase offsets.

In the final stages of crystallographic refinement, accurate phases are inferred by inverse Fourier transformation of the refined atomic model itself. With a reasonably good atomic model, the phases derived from the model are generally so accurate that they can reveal errors in the model when combined with the appropriately weighted experimentally measured amplitudes and used for Fourier synthesis of a new electron-density map (Read, 1986). Incorrect and incomplete models can also provide substantial phase information. However, the utility of this information can be compromised by a phenomenon called 'model bias', which results in the phases inferred from such a model being distorted in order to force the map to match the model even in regions where it is incorrect; the magnitude of this problem is exacerbated considerably as the accuracy of the atomic model decreases (Read, 1986). Therefore, model bias limits the ability to use very incomplete partial models to bootstrap phase improvement when the available phase information is insufficient for definitive atomic interpretation.

The molecular-replacement technique exploits the ability to derive phase information from an imperfect atomic model by positioning a 'search model' comprising a closely related molecular structure in the unit cell of the crystal to be phased (Rossmann, 1990). This technique has been widely applied to the solution of macromolecular crystal structures and in many cases is immediately successful in generating phases of sufficient accuracy for the electron-density map calculated from the molecular-replacement phases to faithfully represent the new structure even in regions where it differs from the atomic model used to generate the phases. However, in cases where the search model is remotely related to the new structure, the molecular-replacement phases are substantially less accurate even when the model is correctly positioned in the unit cell (Read, 1990; Adams *et al.*, 1999). In this situation, the model bias can be so severe that it is impossible to detect the differences in the new structure because the electron-density map calculated from the molecular-replacement phases slavishly reproduces the features of the search model even in regions where the new structure is different. Therefore, model bias can also be a severe impediment to solving a crystal structure using the molecular-replacement technique when the search model is either substantially incomplete or inaccurate.

Fourier cross-validation in reciprocal space has been used successfully as a means of preventing over-refinement of crystal structures when the parameters of the atomic model are underdetermined by the diffraction data, a situation commonly encountered in macromolecular crystallography (Brünger, 1992*a*, 1993). This technique exploits the fact that the electron-density value at any single point in the real-space map is given by Fourier summation of the amplitudes and phases of all of the diffraction data in reciprocal space. In this context, a reasonably accurate electron-density map can still be synthesized even if a set of reflections comprising a small fraction of the reflection data is omitted from the Fourier synthesis. Given the fact that an atomic model for a crystal structure is essentially a representation of the real-space electron-density map, computational refinement of the atomic model is not significantly impeded by consistently omitting the same 5–10% of the diffraction data when mathematically optimizing the parameters of the model to fit the data. However, assuming that the refinement is improving the quality of the model rather than over-fitting the data, inverse Fourier transformation of the model should give improved agreement with all of the diffraction data – *i.e.* even with the reflections in the 'free' set that was omitted from the mathematical optimization process (Brünger, 1992*a*, 1993). The use of such a free set of reflections in reciprocal space to monitor the convergence of the refinement of a model in real space represents Fourier cross-validation in reciprocal space.

Because the Fourier transformation relating the reciprocal-space diffraction data to the real-space electron-density map is bilateral, Fourier cross-validation techniques can also be applied in real space by omitting regions of the map or model prior to inverse Fourier transformation in order to update the phase estimate. This technique has been widely applied in calculating 'omit maps' (Bhat & Blow, 1982; Bhat & Cohen, 1984) in which a local segment of the model is removed prior to calculation of a new phase set which is used to synthesize a map that is de-biased in the omitted region. The atoms that are not omitted are ideally subjected to simulated-annealing prior to phase calculation in order to reduce the bias potentially encoded in their atomic coordinates (Brünger, 1988; Rice & Brünger, 1994; Adams *et al.*, 1997; Brünger *et al.*, 1997). In 'composite omit maps' (Bhat, 1988; Brünger *et al.*, 1997, 1998), this technique is systematically applied to an entire model by generating omit maps for small segments which are then combined to yield a globally de-biased map. Molecular-replacement projects often employ a Fourier cross-validation approach by default in that side-chain atoms are generally removed from the search model when attempting to solve the structure of a homologous but non-identical protein, and a correct solution with strong phasing power can be validated when interpretable electron density appears for the omitted side chains (Adams *et al.*, 1999). However, while these examples show the power of real-space Fourier cross-validation, this technique has not been systematically applied in the early stages of a crystal structure determination as a means of overcoming the model bias encountered when using an inaccurate or incomplete atomic model for phase determination or phase improvement.

In the course of our efforts to solve the crystal structure of the SecA translocation ATPase from *Bacillus subtilis* (Hunt *et al.*, 2002), we were only able to obtain experimental phases to a limiting resolution of 4.4 Å, permitting construction of at most an approximate backbone model for the α-helices in the 96 kDa protein molecule. Therefore, we attempted to use iterative partial model building and density modification to extend the phases to the 2.7 Å resolution limit of the diffraction data and enable construction of a complete molecular model for SecA. The use of a simple real-space Fourier cross-validation technique allowed this process to proceed efficiently to conclusion without being impeded by model bias.

Table 1
Diffraction data statistics for MIR phasing of SecA.

Space group $P3_112$. Unit-cell parameters ~131 × 131 × 151 Å, 90, 90, 120° at 100 K. Data were integrated and scaled with the programs *DENZO* and *SCALEPACK*, respectively (Otwinowski & Minor, 1997). All individual observations with $I \geq -3\sigma(I)$ were merged and included in the calculation of $R_{\rm sym}$. Bijvoet pairs were merged separately for the TAMM, $KAuCl_4$ and SeMet derivates, which were collected at the corresponding anomalous absorption edges.

Protein	Metal†	Beamline	Resolution‡ (Å)	$R_{\rm sym}$ (%)	ESF§	EE§	No. of sites	$\langle B\rangle$¶
Wild type		NSLS-X12B	2.9	9	1.82	0.070		
Wild type		ESRF-BL4	2.7	7.3	1.17	0.060		
Wild type	UO_2	Rotating anode	4.6	21	1.55	0.030	3	562
Wild type	$KAuCl_4$	NSLS-X12B	4.0	15	1.72	0.030	3	146
Wild type	Me_3Pb	Rotating anode	3.7	13	1.55	0.040	3	144
N96C	TAMM	NSLS-X12B	4.0	15	1.68	0.040	2	249
N96C	Thmsl	Rotating anode	4.0	19	1.55	0.050	2	118
N96C	CH_3HgCl	Rotating anode	4.0	13	1.67	0.020	1	103
SeMet††		NSLS-X4A	3.9	8.2	1.41	0.045	33/35	

† UO_2 represents uranyl acetate; Me_3Pb represents trimethyl lead acetate; TAMM represents *tetrakis*-acetoxymercurimethane; Thmsl represents thimerosal (ethylmercurithiosalicylate). ‡ The limiting resolution was defined as the last shell which was at least 1/3 complete for reflections with $I \geq 2\sigma(I)$ after merging. § ESF and EE represent the Error Scale-Factor and Estimated Error parameters from *SCALEPACK*, which were empirically varied to yield χ^2 values of approximately 1 upon merging symmetry-related reflections. The ESF is a linear scale factor for the experimentally estimated $\sigma(I)$, while the EE is a systematic error parameter expressed as a fraction of the measured intensity that is added to the product of ESF*$\sigma(I)$ to give a corrected estimate of the standard deviations of the intensities. ¶ The parameter $\langle B\rangle$ represents the average of the equivalent isotropic thermal B factors from all of the heavy-atom sites as refined using the program *MLPHARE* (Otwinowski, 1991) from *CCP*4 (Collaborative Computational Project, Number 4, 1994) (see §2). †† The SeMet crystals had methionine biosynthetically substituted with selenomethionine (Hendrickson *et al.*, 1990); this derivative was not used for MIR phasing.

2. Experimental

2.1. Heavy-atom derivatization

Wild-type protein and the engineered N96C mutant were purified and crystallized as previously described (Weinkauf *et al.*, 2001). Crystals between two weeks and seven months in age were dispersed in a stabilizing solution comprising 54% ammonium sulfate, 30% glycerol, 20 m*M* dithiothreitol, 100 m*M* *N*,*N*-bis(2-hydroxyethyl)-2-aminoethanesulfonic acid pH 6.9. Following incubation in this reducing environment overnight, crystals were transferred to the same solution without DTT for 2 h before initiating derivatization in the same DTT-free solution plus the heavy-atom reagent. Crystals were removed from the derivatization solution using a microloop and frozen in liquid propane for low-temperature data collection after incubating with the indicated concentrations of the reagents for the following periods of time: 0.05% saturated uranyl acetate for 6 h, 2 m*M* $KAuCl_2$ for 6 h, 5 m*M* trimethyl lead acetate for one week, 0.01% saturated *tetrakis*-acetoxymercurimethane (TAMM) for 28 h, 1 m*M* thimerosal for 24 h or 0.5 m*M* methyl mercury chloride for 24 h. Heavy-atom reagents from commercial sources were used without further purification.

2.2. Labeling with selenomethionine

Protein was biosynthetically labeled with selenomethionine as previously described (Hendrickson *et al.*, 1990) with the addition of the Kao and Michaluk vitamin mixture. However, attempts to overproduce *B. subtilis* SecA under T7 promoter control in the standard DL41 methionine auxotroph led to a low yield of heavily proteolyzed protein. Transduction of this strain to carry a *ClpX* mutation (by D. B. Oliver of Wesleyan University) allowed isolation of full-length SecA labeled with selenomethionine, although at lower levels than expected based on experience of producing other well expressed proteins under T7 promoter control in the same strain background and growth medium.

2.3. Crystallographic data collection

Diffraction measurements were made on frozen crystals at 100–130 K using either a rotating-anode radiation source equipped with focusing mirrors (MSC, Woodlands, TX, USA), beamlines X4A and X12B at the National Synchrotron Light Source at Brookhaven National Laboratory or beamline BL4 at the European Synchrotron Radiation Facility (Table 1). Diffraction intensities were measured using an R-AXIS II area detector on the rotating anode and on beamline X12B, a manually loaded Fuji imaging-plate reader on beamline X4A and a MAR Research imaging-plate area detector on beamline BL4. Data were integrated and scaled using the programs *DENZO* and *SCALEPACK*, respectively (Otwinowski & Minor, 1997).

2.4. Heavy-atom derivative screening

Derivatives were initially screened based on having significant anomalous differences when measured at the synchrotron with the X-ray energy tuned to slightly above the expected anomalous absorption edge. Given the proximity of the target energies at the Pt (11 579 eV), Au (11 934 eV) and Hg (12 299 eV) absorption edges, candidate derivatives containing any one of these metals could be screened without major realignment of beamline optics. Following calibration of the *SCALEPACK* (Otwinowski & Minor, 1997) error-scaling parameters for the crystal/beamline/detector system based on a native data set, each potential derivative was mounted and 20° of data were collected in inverse-beam geometry (*i.e.* 10° at each of two spindle settings 180° apart). Following integration and scaling treating Bijvoet pairs independently, the anomalous χ^2 for this first data sector was determined by merging the reduced Bijvoet pairs together in *SCALEPACK*. Data collection was only continued if the anomalous χ^2 was substantially greater than 1 and the $R_{\rm merge}$ against the native data set was at least 9% at low resolution for this first data sector. In this manner, the likelihood of productive derivatization of the crystal could be determined in substantially less than 1 h. The two crystals fulfilling the target criteria both turned out to be heavy-atom derivatives with useful phasing properties (Tables 1 and 2).

2.5. Location of heavy-atom sites

An anomalous difference Patterson was calculated for the TAMM derivative using data from 25.0 to 7.5 Å resolution in *CCP*4 (Collaborative Computational Project, Number 4, 1994). Although this data set had only 90% anomalous completeness, the analysis showed one correct and unambiguous heavy-atom site in both the vector-verification map and the Harker sections. The heavy-atom sites in the other five derivatives were located in cross difference Fourier maps synthesized at 5 Å using the single-isomorphous replacement/anomalous scattering (SIRAS) phases from the TAMM derivative.

2.6. MIR phasing

Phasing calculations were performed using the program *MLPHARE* (Otwinowski, 1991) from *CCP*4 (Collaborative Computational Project, Number 4, 1994), with simultaneous refinement of position, occupancy, anomalous occupancy and anisotropic *B* factors (based on isomorphous differences) for all sites in all derivatives. The average isotropic thermal *B* factor ⟨*B*⟩ given in Table 1 represents the mean value of the equivalent isotropic thermal *B* factors for all of the heavy-atom sites in the derivative; this parameter is calculated within *MLPHARE* to give a thermal sphere whose volume is equivalent to that of the thermal ellipsoid corresponding to the refined anisotropic thermal *B* factors. A low-resolution cutoff of 25 Å was used for the phasing calculation, which employed the 2.9 Å native data set from NSLS beamline X12B (Table 1). Isomorphous differences were included to 5.5 Å for the uranyl acetate derivative, to 5.2 Å for the $KAuCl_4$ derivative and to 4.4 Å for the other derivatives. Anomalous differences were included to 10 Å for the $KAuCl_4$ derivative, to 9 Å for the trimethyl lead acetate derivative, to 7 Å for the TAMM derivative and to 8 Å for the other derivatives. As the partial model phasing of the structure progressed, heavy-atom phases were routinely redetermined following heavy-atom parameter refinement against external phases calculated from latest refined protein model from *X-PLOR* 3.851 (Brünger, 1992*b*; Engh & Huber, 1991); however, this procedure produced only a minor improvement in map quality either before or after phase combination and density modification.

Table 2
Phasing-power statistics for the SecA heavy-atom derivatives.

The abbreviations for the heavy atoms are defined in Table 1. The final row in the table gives the mean figure-of-merit (⟨FOM⟩) for simultaneous phasing with all derivatives. Standard definitions were used for the parameters (Drenth, 1994). Phasing calculations employed the 2.9 Å native data set from NSLS beamline X12B (Table 1) and were performed using the program *MLPHARE* (Otwinowski, 1991) from *CCP*4 (Collaborative Computational Project, Number 4, 1994) with a low-resolution cutoff of 25 Å as described in §2. The SeMet derivative was not used for MIR phasing.

	14.1 Å	9.8 Å	7.5 Å	6.1 Å	5.2 Å	4.4 Å	Overall
UO_2	2.2	2.0	1.4	0.9	0.5		1.2
$KAuCl_4$	0.9	0.7	0.6	0.6	0.4		0.6
Me_3Pb	1.2	0.8	0.7	0.7	0.5	0.3	0.5
TAMM	4.8	2.4	2.2	1.6	0.8	0.4	1.3
Thmsl	2.3	1.3	1.3	1.3	1.0	0.7	1.1
CH_3HgCl	2.4	1.5	1.3	1.3	1.0	0.6	1.0
⟨FOM⟩	0.92	0.81	0.78	0.68	0.50	0.33	0.53

2.7. Density modification

A shell script *AUTO_DM_KSH* was written in order to drive sequential density-modification runs using the program *DM* (versions 1.6, 1.8 and 2.0; Cowtan, 1994; Cowtan & Main, 1996, 1998). MIR phases, either alone or combined with the latest partial model phases using *SIGMAA* (Read, 1986; Collaborative Computational Project, Number 4, 1994), were used as the starting point for 20–50 successive runs of *DM* employing progressively higher resolution data and assuming progressively higher solvent content. Because test runs with a fixed envelope calculated from complete coordinates yielded inferior results, the envelope used for density modification was calculated automatically within *DM* and therefore did not include any direct contribution from the side-chain atoms omitted during the ping-pong procedure. Solvent flipping was used intermittently during the density-modification sequence (Abrahams & Leslie, 1996). Cross-validated phase combination was performed within *DM* using the 'Combine Omit' mode, and the number of steps in each individual run of *DM* was determined automatically based on monitoring the real-space free *R* factors (*i.e.* by using 'NCYCLE AUTO'). The output map from each run of *DM* was used as the starting point for the subsequent run but not for phase combination, which was performed instead initially with the combined MIR/partial model phases and subsequently with the MIR phases alone. The density-modification sequence concluded with at least one run of *DM* without any phase combination. Maps from each step of the sequence were examined at several diagnostic sites in the structure, and two to four maps were used to guide subsequent model building. The maps with the lowest combined real-space free *R* factor generally had the best overall electron density. However, the maps from runs performed without any phase combination were often useful for interpreting the structure in some regions even though they showed degradation of the map quality in other regions.

Sequential runs performed at constant resolution and solvent content often initially produced an increase in real-space free *R* factor but later produced a decrease accompanied by an improvement in map quality. For density modification of the MIR phases (Fig. 1*b*), 27 sequential *DM* runs were performed with the first assuming a solvent content of 40% and using data only to 5 Å but later runs eventually assuming a solvent content of 65% and using data to 3.5 Å. For density modification of model 7 (Table 3), 25 sequential *DM* runs were performed with the first assuming a solvent content of 54% and using data only to 5 Å but the eighth assuming a solvent content of 64% and using data to 2.6 Å. In this case, phase combination was performed initially with combined MIR/partial model phases but with MIR phases alone after run 10 and was omitted entirely after run 15. The maps from runs 15 and 16 were used for model building. Because the density-modified maps produced by *DM* were of sufficient quality to allow the ping-pong procedure to move forward, no effort was made to compare the relative efficiencies of different density-modification procedures.

2.8. Crystallographic refinement

The computationally refined model was never used for rebuilding in order to avoid the propagation of model bias during the ping-pong cross-validation procedure and also to

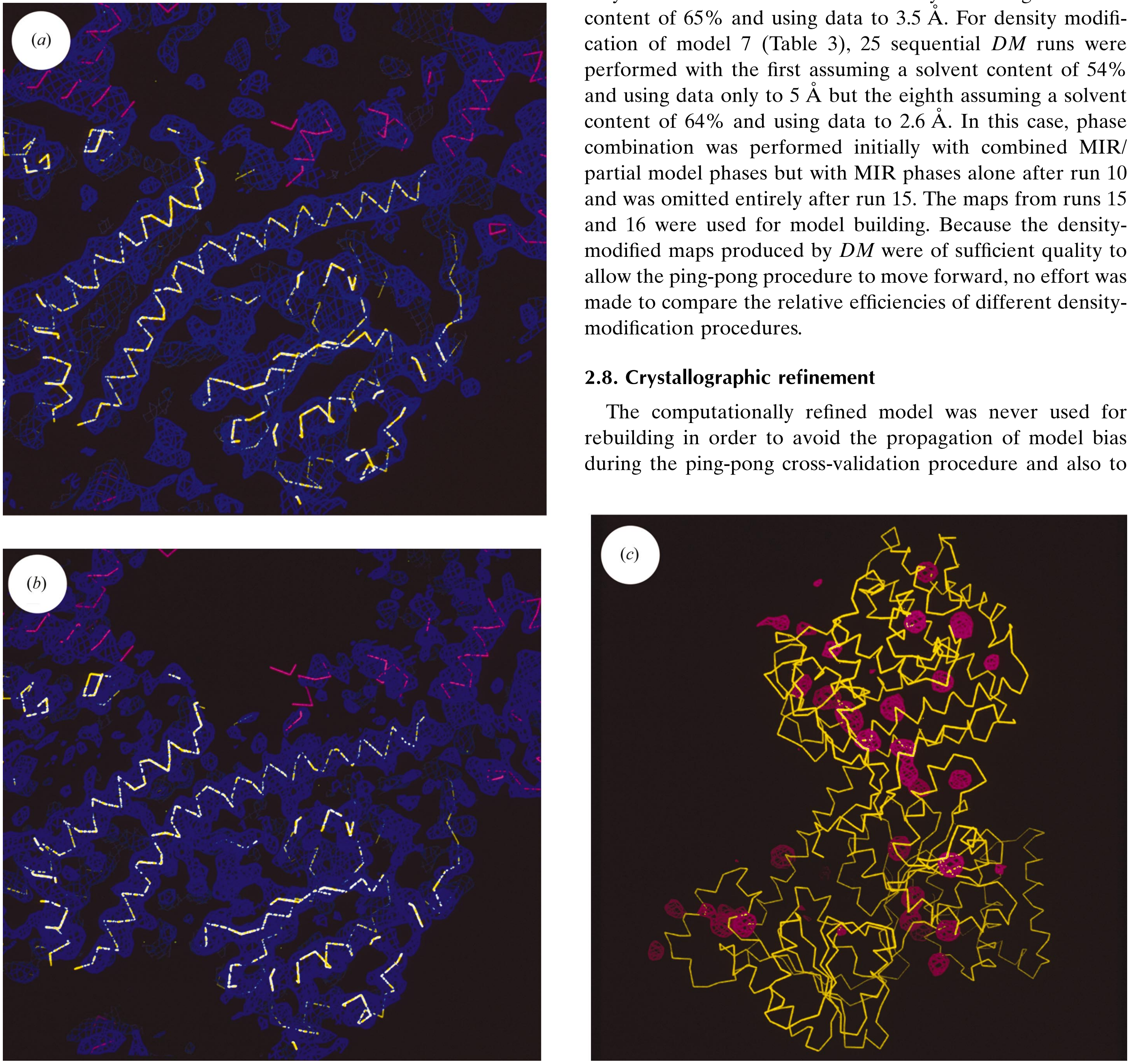

Figure 1
Experimental electron-density maps. (*a*) The final MIR map at 4.4 Å is shown in blue. The map was synthesized using phases from the calculation reported in Table 2. A section through the first nucleotide-binding fold and the α-helical wing domain of SecA is shown (Hunt *et al.*, 2002). (*b*) The MIR map after density modification is shown in blue. An identical view is shown as in (*a*). The MIR map from (*a*) was used as the starting point for an extended density-modification sequence using the program *DM* (Cowtan, 1994) (see §2 for details). The nominal resolution of this map is 3.5 Å, but its effective resolution is obviously worse based on the failure to observe significant side-chain electron density. Polyalanine model 1 (Table 3) was built from this map and C^{α} traces of this model are shown here and in (*a*). The yellow trace shows one SecA protomer (*i.e.* the contents of one asymmetric unit in the lattice), while the magenta traces show protein molecules related by crystallographic symmetry. (*c*) The anomalous difference Fourier map for the SeMet derivative of SecA is shown in magenta. Density-modified phases from polyalanine model 2 (Table 3) were employed to synthesize this map using data from 12.0 to 4.5 Å. A C^{α} trace of the complete SecA molecule is shown in yellow.

Table 3
SecA partial model phasing.

The values in the '% of total' column represent the number of protein residues in the model either with or without side chains divided by the 802 residues in the final model (*i.e.* not including the ordered solvent molecules or the 39 residues that are disordered in the crystal lattice). For the refinement, mean figure-of-merit (⟨FOM⟩) and *R*-factor calculations, atoms beyond the C^{β} position were omitted for the side chains being cross-validated in each step. Standard definitions were used for all parameters. The ⟨FOM⟩ data were calculated using *SIGMAA* (Read, 1986) from *CCP*4 (Collaborative Computational Project, Number 4, 1994). The *R* factors were calculated using *X-PLOR* 3.851 (Brunger, 1992*b*) after using the same program to refine the partial model at the indicated limiting resolution. Models 1–4 were refined against the 2.9 Å native data set from beamline X12B that was also used for phasing, while models 5–10 were refined against a higher resolution native data set from BL4 at the ESRF (Table 1). This latter data set is only complete to 2.7 Å, although refinement was conducting using all data to 2.6 Å with $F \geq 2\sigma(F)$. Different refinement protocols were used for the different models as outlined in §2. Models 8–10 each involved multiple refinement/rebuilding cycles but without adding residues or changing their register. The FOMs were calculated from the refined models using the higher resolution data set with a low-resolution cutoff of 15 Å.

Model	No. of fragments	No. in polyalanine	No. with side chains	% of total	⟨FOM⟩, 4.0 Å	⟨FOM⟩, 3.0 Å	⟨FOM⟩, 2.7 Å	⟨FOM⟩, overall	$R_{\rm free}$ (%) (resolution)
1	44	652	0	81	0.28	0.08	0.04	0.20	51.0 (4.0 Å)
2	37	707	0	88	0.45	0.14	0.06	0.29	49.1 (3.5 Å)
3	6	372	412 (ping)	98	0.55	0.31	0.16	0.41	43.8 (2.9 Å)
4	3	419	383 (pong)	100	0.67	0.39	0.18	0.49	42.2 (2.9 Å)
5	3	392	407 (ping)	100	0.71	0.40	0.25	0.54	38.5 (2.6 Å)
6	1	335	464 (pong+)	100	0.74	0.44	0.28	0.56	36.1 (2.6 Å)
7	1	56	743 (ping+)	100	0.75	0.47	0.31	0.58	34.4 (2.6 Å)
8	1	0	799	100	0.76	0.50	0.35	0.60	32.7 (2.6 Å)
9	1	0	802	100	0.79	0.55	0.38	0.63	32.1 (2.6 Å)
10	1	0	802	100	0.78	0.56	0.38	0.63	30.5 (2.6 Å)

minimize the problems associated with over-refinement of invariant protein segments. Instead, rebuilding was always performed on the hand-built model which was propagated into successive steps without computational refinement. The updated hand-built model (with the side-chain atoms beyond C^{β} omitted for the residues to be cross-validated) was subjected to an automated sequence of refinement steps using *X-PLOR* 3.851 driven by the shell script *AUTO_XPL* (Hunt *et al.*, 1998), yielding the results summarized in Table 3. The refinement of model 1 comprised one step of overall rigid-body refinement, one step of grouped rigid-body refinement of the individual secondary-structural elements and one step of positional refinement using data from 6.5 to 4.0 Å resolution. The intervening models were refined using progressively lengthier and more comprehensive protocols at progressively higher resolution (Table 3). Overall isotropic *B*-factor refinement was included starting with model 2, grouped isotropic atomic *B*-factor refinement was included starting with model 3, a bulk-solvent correction was included starting with model 4, individual isotropic atomic *B*-factor refinement was included starting with model 8 and overall anisotropic *B*-factor refinement was included starting with model 9. The refinement of model 10 comprised 20 sequential steps of refinement, starting with data from 6.5 to 3.2 Å resolution and ultimately including data from 50.0 to 2.6 Å resolution. The isotropic atomic *B* factors were all reset to 45.0 Å^2 after 14 steps in this sequence and then refined again to convergence in alternation with further positional refinement. During this phase of the procedure, grouped *B*-factor refinement effected a 5.6% drop in the free *R* factor, while subsequent individual *B*-factor refinement effected a 0.9% drop in the free *R* factor. Extensive sets of 1400 K simulated-annealing omit maps were generated and used to guide rebuilding starting with model 9.

2.9. Molecular-graphics figures

The images are screen captures from the program *O* (Jones *et al.*, 1991).

3. Results and discussion

3.1. Diffraction data characteristics

The soluble form of the SecA translocation ATPase from *B. subtilis* was crystallized in space group $P3_112$ (Weinkauf *et al.*, 2001), yielding crystals that diffracted to approximately 3.7 Å using a rotating-anode X-ray source or a maximum of 2.7 Å using the strongest synchrotron-radiation sources (Table 1). The diffraction data collected using a rotating-anode X-ray source were quite weak, as were the synchrotron data collected from the heavy-atom derivatives. In order to maximize the quality and information content of the merged diffraction data, highly redundant data sets were collected from the weakly diffracting crystals and all individual observations with $I \geq -3\sigma(I)$ were included in merging as recommended by Otwinowski & Minor (1997). This data-processing strategy improves the signal-to-noise ratio of the merged data set [*i.e.* $\langle I/\sigma(I)\rangle$] based on the averaging of multiple observations of weak reflections, but results in the inclusion of a large number of weak observations in the calculation of $R_{\rm sym}$. These weak observations with low values of $[I/\sigma(I)]$ are expected to have large fractional variations in intensity compared with their symmetry mates and therefore tend to inflate the $R_{\rm sym}$ value of data sets processed in this manner (Table 1).

Nonetheless, several factors indicate that the quality of the resulting data sets is high in spite of the elevated $R_{\rm sym}$ values. Firstly, the merging process yields χ^2 values near 1.0 in all resolution shells using error-scaling parameters (Otwinowski & Minor, 1997) equivalent to those used for stronger data sets with lower $R_{\rm sym}$ values collected on the same detector (Table 1 and additional data not shown). Secondly, the $R_{\rm sym}$ values in the low-resolution shells of these data sets are consistent with the mean intensity in each shell [*i.e.* $R_{\rm sym} \simeq \langle\sigma(I)\rangle/\langle I\rangle$] (data not shown). Thirdly, the number of rejected reflections was generally similar to that expected owing to statistical fluctuations and was always substantially lower than 1% (data not shown). Finally, these data sets yielded very good phases when used for multiple isomorphous replacement (MIR) calculations (see below).

3.2. MIR phasing

Given the large size of the protein molecule and the weak diffraction properties of the crystal, a simple heavy-atom derivative with strong phasing power was sought as a starting point for MIR phasing. Preliminary soaking experiments showed that the diffraction properties of the SecA crystals were not significantly perturbed by overnight incubation in the presence of mercury reagents, suggesting that site-directed mutagenesis to introduce a surface-exposed cysteine residue could be used to obtain a single-site mercury derivative. Therefore, residues 90–220 of SecA were aligned with two proteins of known structure that were predicted to have a similar F1-like ATP-binding fold (RecA and nitrogenase) and sites were identified that were both surface-exposed in these molecules and substituted by cysteine in at least one SecA homologue. This exercise led to the construction of an N96C mutant of *B. subtilis* SecA, which crystallized isomorphously with the native protein and immediately yielded a high-quality mercury derivative.

Rapid screening of potential heavy-atom derivatives was conducted on a synchrotron beamline tuned so that the X-ray energy was slightly above the predicted absorption edge of the candidate atom. After calibrating the error-scaling parameters of the crystal/detector system, a pair of 10° data wedges were collected in inverse-beam geometry (*i.e.* at 0 and 180° settings of the spindle) from each potential derivative and the data was reduced in *SCALEPACK* (Otwinowski & Minor, 1997) keeping anomalous pairs separated during merging. A high anomalous χ^2 at low and intermediate resolution was taken to be indicative of likely derivatization of the protein, in which case a complete data set was taken (see §2 for details.) This protocol led to the identification of two heavy-atom deriva-

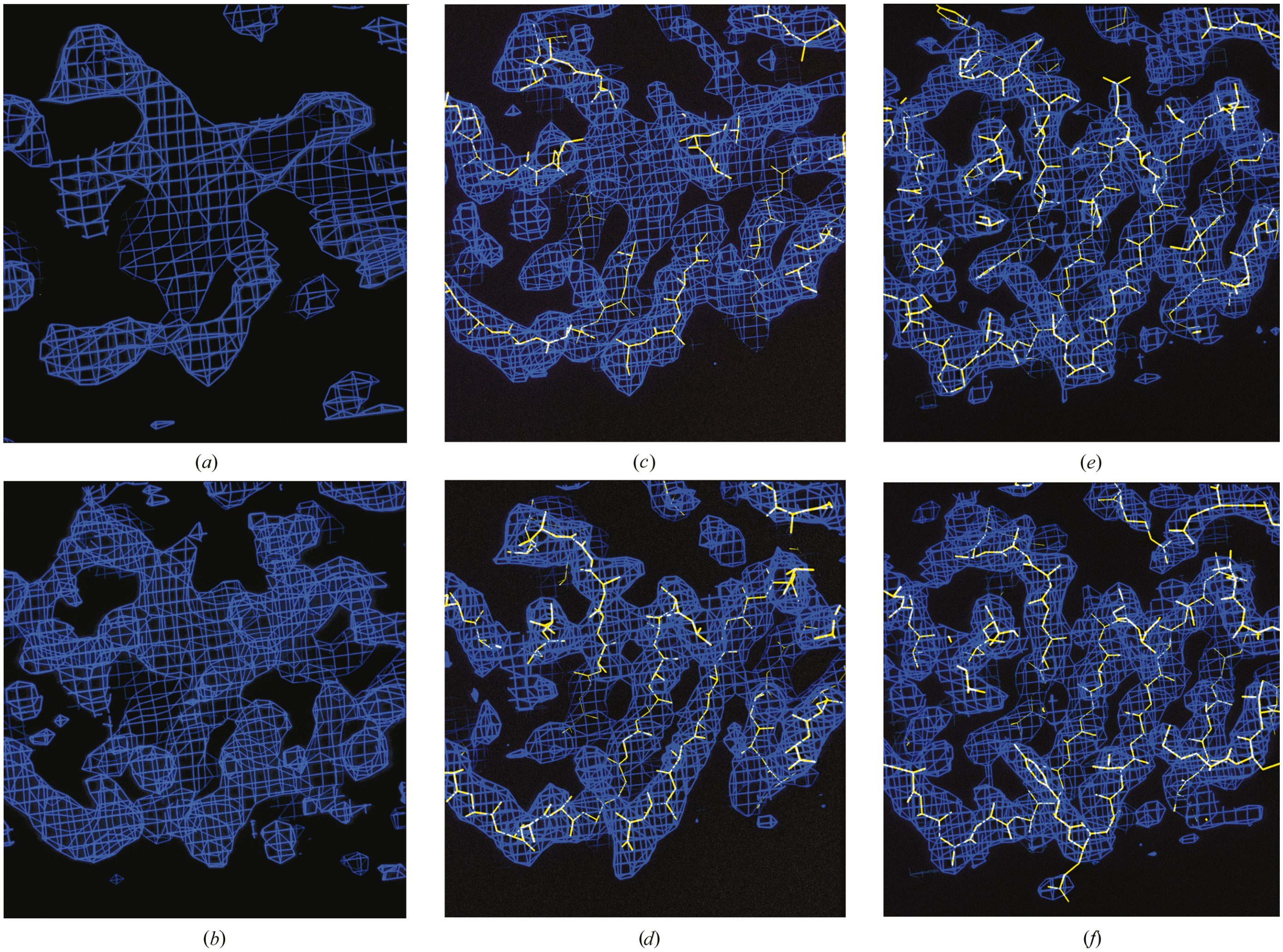

Figure 2
Evolution of the electron density during MIR and partial model phasing. (*a*) MIR, (*b*) *DM*, (*c*) polyalanine 1, (*d*) polyalanine 2, (*e*) ping, (*f*) pong. A section of the map containing the central β-sheet in the first nucleotide-binding fold is shown. Except for the MIR panel (*a*), the best map produced by an extended density-modification sequence is shown (see §2 for details). The yellow traces show the models used to produce each map, *i.e.* the model whose phases were combined with the MIR phases as the starting point for the density-modification sequence that produced the map. The MIR (*a*) and *DM* (*b*) panels show the same maps as Fig. 1. The polyalanine 1 (*c*), polyalanine 2 (*d*), ping (*e*) and pong (*f*) panels show the maps obtained by density modification of the combined phases from models 1, 2, 3 and 4 in Table 3, respectively. The nominal resolution of the maps is as follows: 4.4 Å for MIR, 3.5 Å for *DM*, 3.0 Å for polyalanine 1 and polyalanine 2, and 2.9 Å for ping and pong.

tives (Table 1), a *tetrakis*-acetoxymercurimethane (TAMM) derivative of the N96C protein and a $KAuCl_4$ derivative of the native protein. TAMM contains four covalently linked Hg atoms, and the corresponding derivative was dominated by a single diffuse heavy-atom site at the N96C position which produced an anomalous difference Patterson map that was immediately interpretable.

The SIR phases from the TAMM derivative were used to locate the heavy-atom sites in other potential derivatives based on isomorphous and anomalous difference Fourier analyses. In addition to the $KAuCl_4$ derivative from the synchrotron, four derivatives were identified using rotating-anode data sets (Table 1). Two of these were non-mercury derivatives of the native protein, while two others were additional mercury derivatives of the N96C protein. These latter two derivatives both contained an Hg atom bound to the N96C site, but differed considerably in their occupancy characteristics from the TAMM derivative because the reagent molecules contained only a single Hg atom; the derivative made using the reagent thimerosal also showed a second site in the unit cell with significant mercury occupancy (Table 1).

This set of six heavy-atom derivatives produced very high quality MIR phases (Otwinowski, 1991; Collaborative Computational Project, Number 4, 1994), but extending only to a resolution of 5 Å before losing meaningful phasing power (Table 2). The corresponding electron-density map showed a very clear molecular boundary and smooth cylinders at the positions of most of the α-helices in the structure (Fig. 1*a*). However, the electron density for the β-sheets in the protein was not interpretable in the MIR map (see Fig. 2*a*).

The limiting resolution of the derivative data sets ranged from 3.7 to 4.6 Å and the low resolution of these data sets could potentially have contributed to the loss of phasing power beyond 5 Å. However, additional analyses suggested that this problem was instead a consequence of the inherent disorder of the bound metal atoms in the derivatives. Firstly, collection of higher resolution synchrotron data sets from the mono-mercury derivatives yielded no improvement in their phasing power or its fall-off with resolution (data not shown). Furthermore, using phases from the fully refined model to calculate isomorphous difference Fourier maps from these heavy-atom derivatives shows peaks approximately 5 Å in diameter, consistent with the *B* factors for these sites being greater than 100 Å^2 as estimated during heavy-atom parameter refinement (Table 1). Even the Hg atoms covalently bound to the engineered cysteine in these derivatives appear to have substantial mobility, potentially arising from χ_2 dihedral rotation. The dominant heavy-atom site in the TAMM derivative gives an asymmetrical peak approximately 8–11 Å in diameter in an isomorphous difference Fourier map synthesized with phases from the refined model (data not shown), consistent with the *B* factors for this site being greater than 200 Å^2 as estimated during heavy-atom parameter refinement (Table 1) and suggesting that the four individual Hg atoms in the compound are rotationally disordered. Finally, the dominant heavy-atom site in the uranium derivative appears as a cylindrical peak approximately 15 Å long and 6 Å in diameter in an isomorphous difference Fourier map synthesized with phases from the refined model (data not shown). This interaction site is located in a weakly acidic groove on the surface of the protein in a region with backbone thermal *B* factors of the order of 100 Å^2 in the refined structure. While the extraordinarily high *B* factors estimated for this site during heavy-atom parameter refinement (Table 1) could be biased by the low resolution limit of the diffraction data from the derivative, the difference Fourier analysis shows that the uranyl ion(s) has a very diffuse interaction site that will prevent it from contributing to the diffraction pattern at high resolution owing to the steep fall-off in the Fourier transform of its spatial distribution. Similarly, all of the well occupied heavy-atom sites in the derivatives listed in Table 1 appeared to be substantially disordered, making it very unlikely that they could be used to obtain MIR phases at higher resolution.

A selenomethionyl derivative (Hendrickson *et al.*, 1990) of SecA was prepared in the hope that the experimental phases could be extended to higher resolution using multiple-wavelength anomalous diffraction (MAD) measurements (Hendrickson, 1991). Although this protein crystallized isomorphously with the native, the crystals grew to a substantially smaller limiting size and diffracted to only 3.9 Å on a MAD beamline (Table 1). In this context, it was not possible to use MAD to obtain experimental phases to higher resolution than the MIR phases. However, an anomalous difference Fourier map from this derivative synthesized with the MIR phases showed the locations of most of the methionine residues in the structure (Fig. 1*c*), and this information provided useful constraints in assigning the protein sequence (see below).

Given the observed disorder in all of the heavy-atom sites in the first six derivatives and the inability to obtain a selenomethionyl derivative diffracting to high resolution, we concluded that it would be very difficult to obtain experimental phases to higher resolution and decided to attempt to solve the structure using the existing low-resolution MIR phases.

3.3. Density modification

The MIR phases were improved using an extended density-modification sequence involving iterative runs of the program *DM* (Cowtan & Main, 1996, 1998; Cowtan, 1994). Solvent-flattening and histogram-matching calculations were performed with a gradual increase in the assumed solvent content of the crystal conducted in parallel with a gradual expansion in the resolution of the data included in the calculation (see §2 for more details). Solvent-flipping was used during some of the density-modification steps in the sequence (Abrahams & Leslie, 1996). The high quality of the low-resolution MIR map (Fig. 1*a*) combined with the 65% solvent content of the crystal allowed robust determination of the protein envelope and substantial improvement in the electron-density map during the density-modification procedure (Fig. 1*b*). While the well ordered α-helices in the protein appeared as smooth cylinders in the MIR map (Fig. 1*a*), the helical contour of the polypeptide backbone in these regions was detectable in the density-modified map (Fig. 1*b*). However, the β-strands in the protein were not readily interpretable nor were there interpretable side-chain electron-density features anywhere in the map either before or after density modification of the MIR phases (Figs. 2*a* and 2*b*).

3.4. Partial model phasing from backbone models

Nonetheless, the map produced by intensive density modification of the MIR phases (Figs. 1*b* and 2*b*) allowed an approximate backbone model to be constructed for 81% of the well ordered protein residues in the SecA crystal, comprising most of the α-helices in the molecule plus a limited number of β-strands (Table 3). Refinement of this model followed by σ_A-weighted combination of the partial model phases and the MIR phases (Read, 1986) and then another round of intensive density modification produced a new electron-density map which permitted construction of an improved backbone model containing an additional 7% of the protein molecule (Fig. 2*c* and Table 3). Another round of model refinement, phase combination and density modification produced a map in which electron density was visible for many side chains but which was not of sufficient quality to allow confident sequence assignment in most regions of the protein molecule (Fig. 2*d*).

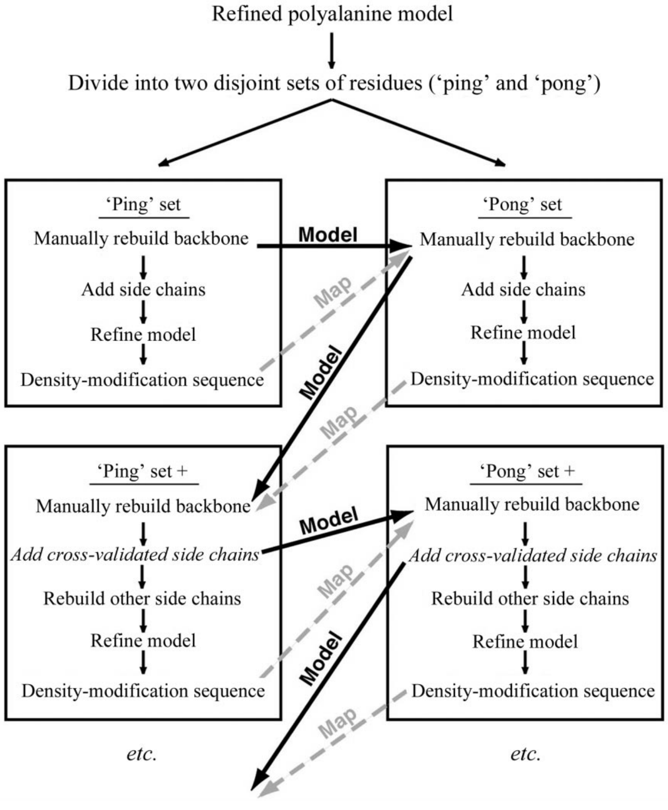
Refined polyalanine model
Divide into two disjoint sets of residues ('ping' and 'pong')
'Ping' set
Manually rebuild backbone
Add side chains
Refine model
Density-modification sequence
'Pong' set
Manually rebuild backbone
Add side chains
Refine model
Density-modification sequence
Model
Map
Model
Map
'Ping' set +
Manually rebuild backbone
Add cross-validated side chains
Rebuild other side chains
Refine model
Density-modification sequence
'Pong' set +
Manually rebuild backbone
Add cross-validated side chains
Rebuild other side chains
Refine model
Density-modification sequence
Model
Map
Model
Map
etc.
etc.

Figure 3
Flowchart describing the ping-ping cross-validation procedure. See text for explanation.

3.5. Sequence assignment and ping-pong cross-validation

Given the fact that the backbone model was nearly complete but unambiguous sequence assignment was still not possible, a procedure that we call 'ping-pong cross-validation' (Fig. 3) was developed in order to produce further improvements in the phasing power of the partial model while avoiding the risk of acute model bias that would be associated with a speculative attempt at sequence assignment. The residues in the protein sequence were partitioned into two disjoint sets that we called 'ping' and 'pong' (Table 3), and the partial model was expanded to include side-chain atoms only for the residues in the ping set. The partitioning was performed at random, although it was biased toward the inclusion of residues with more readily interpretable electron density in the ping set. Refinement of the new partial model containing the ping set was followed by phase combination and density modification in order to produce an improved map (Fig. 2*e*) which was used to attempt to assign the sequence for the pong set. At this point, the side chains from the ping set were removed from the partial model, and side-chain atoms were added for the residues in the pong set. This real-space cross-validation procedure was repeated iteratively until convincingly interpretable side-chain density was obtained for each individual residue in a map derived from a partial model that did not include any side-chain atoms from that residue beyond the C^{β} position (Fig. 3). Only after this criterion had been met were the side-chain atoms for a specific residue retained in the evolving partial model on a permanent basis.

The phasing power of the partial model improved dramatically upon addition of the side-chain sets (Table 3). In practice, the ping-pong cross-validation procedure converged quite rapidly after first two cycles, and only five cycles were required to complete the atomic model (*i.e.* models 3–7 in Table 3). Standard protein refinement techniques were used starting with model 8 (see below).

In order to prevent model bias from being introduced indirectly into the electron-density maps by the computational refinement of the backbone in models containing unvalidated side chains, the computationally refined model was never propagated into subsequent steps during the ping-pong cross-validation procedure (Table 3). Instead, a manually built reference model was maintained that had never been subjected to computational refinement, and side chains were added to this unrefined backbone model. Following the completion of each new model, an extended series of automated computational refinement steps was performed using *X-PLOR* 3.851 (Brünger, 1992*b*) driven by the shell-script *AUTO-XPL*, and the refined model was used to derive phases for combination with the MIR phases prior to input into density-modification calculations (Fig. 2). However, the computationally refined structure was not used to construct the next model, which was instead constructed from the reference manually built model. Even after the convergence of the ping-pong cross-validation procedure, rebuilding was still performed on the hand-built model during the terminal stages of the refinement process because this strategy reduced problems associated with over-refinement of the correctly modeled parts of the structure during standard iterative rebuilding cycles.

The sequence-assignment process was facilitated by the use of an anomalous difference Fourier map derived from the SeMet derivative of SecA which showed the location of 33 of the 36 methionine residues in the protein (Fig. 1*c*). This map played a minor role in assigning the sequence in the more easily interpretable regions of the structure but a more prominent role in the regions of the structure with more ambiguous electron density, which all turned out to be located in weakly ordered protein domains with very high backbone atomic *B* factors in the fully refined structure (Hunt *et al.*, 2002). However, we believe that the ping-pong cross-validation procedure would have converged only slightly more slowly if the SeMet anomalous difference Fourier map had not been available to guide the sequence-assignment process.

3.6. Completion of the refinement

After convergence of the ping-pong cross-validation procedure (*i.e.* starting with model 8 in Table 3), the refinement proceeded along traditional lines involving iterative rebuilding cycles guided by $F_o - F_c$ maps from the refined structure, simulated-annealing omit maps and geometric criteria (see §2.8). Only minor changes in polypeptide-chain registration occurred during this phase of the refinement, indicating that the real-space cross-validation procedure had allowed correct protein-sequence assignment throughout most of the protein structure. The final SecA model was refined to a free *R* factor of 30.5% using data with $F \geq 2\sigma(F)$ to 2.6 Å resolution (Table 3; although the 2.6–2.7 Å resolution shell is less than 1/3 complete for reflections satisfying the cutoff criterion). The geometric details of this refinement are reported elsewhere along with the crystal structure itself (Hunt *et al.*, 2002).

Upon completion of the refinement, the details of the SecA model remained ambiguous in certain regions of the protein structure characterized by unusually high backbone thermal *B* factors (Hunt *et al.*, 2002). The value of the mean backbone atomic *B* factor varies tremendously in the six different domains/subdomains included in the structure, ranging from a minimum of 50 Å^2 to a maximum of 129 Å^2. These variations reflect the fact that the different domains of SecA have different degrees of static or dynamic disorder in the crystal lattice (Ringe & Petsko, 1986). Ultimately, 148 residues of the 802 in the final protein model were designated as 'unreliable' because the atomic coordinates in these regions were not considered to be definitive, and even local errors in polypeptide chain registration seemed possible. These unreliable residues are located exclusively in the areas with the highest backbone atomic *B* factors (including essentially the entire subdomain with the 129 Å^2 mean backbone *B* factor). While the ping-pong cross-validation procedure did not allow construction of a definitive atomic model in these regions, it seems unlikely that a definitive interpretation would have been possible even with very high quality experimental phases given the extraordinarily high *B* factors in these regions.

3.7. Conclusions

While powerful techniques exist for experimental phasing of macromolecular crystal structures, the electron-density maps produced using these techniques are often imperfect owing to various experimental limitations. In this context, phase-improvement and phase-extension techniques play an important role in most macromolecular crystal structure-determination projects (Wang, 1985; Cowtan & Main, 1996, 1998; Abrahams & Leslie, 1996). In attempting to solve the structure of the 96 kDa SecA translocation ATPase, we could only obtain experimental phases to a limiting resolution of 4.4 Å. Nonetheless, using partial model phasing combined with intensive density modification, we were able to complete the determination of this protein structure containing 6402 non-H atoms, even with experimental phases with a mean figure of merit of 0.53 (Table 2) available for only 8300 reflections. A real-space Fourier cross-validation technique called ping-pong cross-validation was used to mitigate the problems of model bias often associated with attempts to use incomplete and/or inaccurate models for phase improvement. This technique involved partitioning the side chains in the protein into two disjoint sets (called 'ping' and 'pong') and then using the residues in one set to cross-phase the residues in the other. Each individual side chain was only incorporated in the final protein model when unambiguously interpretable electron density for it was observed in a map produced without assuming its identity. The expansion of the partial model to include a large number of side chains produced very substantial improvement in the phasing power of the partial model, while the ping-pong cross-validation technique allowed efficient completion of the crystal structure of SecA without being impeded by model bias.

The ping-pong cross-validation procedure could readily be computationally automated. Furthermore, extension of the method to include real-space cross-validation of backbone segments would be straightforward. Although real-space cross-validation procedures have routinely been used in the terminal phases of crystal structure refinement (Brünger, 1988; Rice & Brünger, 1994; Adams *et al.*, 1997; Brünger *et al.*, 1997), they have not been employed systematically at the early stages of the model-building process. The work presented in this paper shows that such techniques can also facilitate the use of partial model phasing to bootstrap phase extension and improvement in the early phases of a crystal structure-determination project. Systematic application of real-space cross-validation techniques could therefore be helpful in overcoming model bias when using a remotely related model for structure determination by molecular replacement and could even play a role in attempting to use partial model phasing as an adjunct to *ab initio* phasing of macromolecular crystal structures (Bricogne, 1984, 1988).

The authors would like to thank M. Machius, L. Esser, E. Lee, Z. Otwinowski, D. Turk and P. Smith for helpful discussions, and acknowledge L. Henry and R. Krishnaraj for help with synchrotron data collection. D. B. Oliver provided the all of the protein-expression strains employed in this project.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Adams, P. D., Pannu, N. S., Read, R. J. & Brünger, A. T. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.
Adams, P. D., Pannu, N. S., Read, R. J. & Brunger, A. T. (1999). *Acta Cryst.* D**55**, 181–190.
Bhat, T. N. (1988). *J. Appl. Cryst.* **21**, 279–281.
Bhat, T. N. & Blow, D. M. (1982). *Acta Cryst.* A**38**, 21–29.
Bhat, T. N. & Cohen, G. H. (1984). *J. Appl. Cryst.* **17**, 244–248.
Blow, D. M. & Rossmann, M. G. (1961). *Acta Cryst.* **14**, 1195–1202.
Bricogne, G. (1984). *Acta Cryst.* A**40**, 410–445.
Bricogne, G. (1988). *Acta Cryst.* A**44**, 517–545.
Brünger, A. T. (1988). *J. Mol. Biol.* **203**, 803–816.
Brünger, A. T. (1992*a*). *Nature (London)*, **355**, 472–475.
Brünger, A. T. (1992*b*). *X-PLOR Version 3.1, A System for Crystallography and NMR.* New Haven, CT, USA: Yale University Press.
Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Brünger, A. T., Adams, P. D. & Rice, L. M. (1997). *Structure*, **5**, 325–336.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Cowtan, K. & Main, P. (1998). *Acta Cryst.* D**54**, 487–493.
Cowtan, K. D. & Main, P. (1996). *Acta Cryst.* D**52**, 43–48.
Crick, F. H. C. & Magdoff, B. S. (1956). *Acta Cryst.* **9**, 901–908.
Drenth, J. (1994). *Principles of Protein X-Ray Crystallography*. New York: Springer-Verlag.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Hauptman, H. (1986). *Science*, **233**, 178–183.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Hunt, J. F., Mixon, M. B. & Berghuis, A. M. (1998). *J. Appl. Cryst.* **31**, 491–495.
Hunt, J. F., Weinkauf, S., Henry, L., Fak, J. J., McNicholas, P., Oliver, D. B. & Deisenhofer, J. (2002). *Science*, **297**, 2018–2026.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Read, R. J. (1990). *Acta Cryst.* A**46**, 900–912.
Rice, L. M. & Brünger, A. T. (1994). *Proteins*, **19**, 277–290.
Ringe, D. & Petsko, G. A. (1986). *Methods Enzymol.* **131**, 389–433.
Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Weinkauf, S., Hunt, J. F., Scheuring, J., Henry, L., Fak, J., Oliver, D. B. & Deisenhofer, J. (2001). *Acta Cryst.* D**57**, 559–565.