

High-resolution experimental phases for tryptophanyl-tRNA synthetase (TrpRS) complexed with tryptophanyl-5'AMP

Pascal Retailleau,^a Yuhui Yin,^a Mei Hu,^a Jeffrey Roach,^a Gérard Bricogne,^{b,c} Clemens Vornrhein,^b Pietro Roversi,^b Eric Blanc,^b Robert M. Sweet^d and Charles W. Carter Jr^{a*}

^aDepartment of Biochemistry and Biophysics, CB# 7260, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, ^bMRC-LMB, Hills Road, Cambridge CB2 2QH, England, ^cLURE, Bâtiment 209D, 91405 Orsay CEDEX, France, and ^dDepartment of Biology, Brookhaven National Laboratory, Upton, NY 11937-500, USA

Correspondence e-mail: carter@med.unc.edu

Native data, anomalous data at three wavelengths and an independent peak-wavelength data set for SeMet-substituted protein have been collected from cryoprotected crystals of the TrpRS–adenylate product (TAM) complex to a resolution limit of 1.7 Å. Independent phase sets were developed using *SHARP* and improved by solvent flipping with *SOLOMON* using molecular envelopes derived from experimental densities for, respectively, peak-wavelength SAD data from four different crystals, MAD data and their M(S)IRAS combinations with native data. Hendrickson–Lattman phase-probability coefficients from each phase set were used in *BUSTER* to drive maximum-likelihood refinements of well defined parts of the previously refined room-temperature 2.9 Å structure. Maximum-entropy completion followed by manual rebuilding was then used to generate a model for the missing segments, bound ligand and solvent molecules. Surprisingly, peak-wavelength SAD experiments produced the smallest phase errors relative to the refined structures. Selenomethionylated models deviate from one another by 0.25 Å and from the native model by 0.38 Å, but all have r.m.s. deviations of ~1.0 Å from the 2.9 Å model. Difference Fourier calculations between amplitudes from the 300 K experiment and the new amplitudes at 100 K using 1.7 Å model phases show no significant structural changes arising from temperature variation or addition of cryoprotectant. The main differences between low- and high-resolution structures arise from correcting side-chain rotamers in the core of the protein as well as on the surface. These changes improve various structure-validation criteria.

Received 26 March 2001

Accepted 17 July 2001

PDB References: TrpRS–TAM, native, MIRAS phases, 1i6m; TrpRS–TAM, seleno derivative, SAD phases, 1i6l; TrpRS–TAM, seleno derivative, MAD phases, 1i6k.

1. Introduction

Faithful translation of genetic information requires accurate aminoacylation of tRNAs by their cognate aminoacyl-tRNA synthetases (aaRSs). Although the tRNAs have similar structures, the 20 aaRSs show considerable diversity in molecular size, amino-acid sequence, quaternary structure and modes of action. Particularly striking is the partitioning of enzymes for the 20 amino acids into two classes on the basis of comparative sequence (Eriani *et al.*, 1990) and structural (Cusack *et al.*, 1991; Delarue & Moras, 1992; Ruff *et al.*, 1991) analyses. These two classes were subsequently subdivided (Arnez & Moras, 1997).

Because they utilize ATP for carboxyl-group activation prior to its transfer to the 3'-terminal ribose of the tRNA, these enzymes are suspected of making use of conformational coupling to ensure specificity (Carter, 1993). The outlines of such coupling have begun to be revealed in the structures of different *Bacillus stearothermophilus* TrpRS conformers

(Doubl   *et al.*, 1995; Ilyin *et al.*, 2000), which cocrystallize with many different ligands in crystal forms that depend on the ligands present during crystal growth. Crystal characterization (Carter *et al.*, 1994) and previous structure solutions, together with small-angle X-ray scattering (Ilyin *et al.*, 2000), provide evidence for major ligand-dependent domain movements in solution as well as in crystals. TrpRS is the smallest aaRS monomer (328 residues) and belongs with TyrRS in subclass Ic. Both are functionally homodimers and the TrpRS conformational cycle might therefore be relevant only for this subclass. However, similar domain movements have been observed for LeuRS (Cusack *et al.*, 2000), confirming their importance in other monomeric class I aaRSs.

Strong circumstantial links between conformational changes and catalysis of amino-acid activation are provided by comparing the TrpRS structures with the elegant mutagenic and pre-steady-state kinetic analysis of TyrRS (Fersht *et al.*, 1988). These links in turn emphasize the importance of analysing the structural changes in considerable detail in order to understand the mechanistic basis of translation and hence justify renewed interest in higher resolution structural data for all of the TrpRS conformations.

Conformational flexibility may present an entropic barrier to crystallization and especially to the growth of well ordered crystals. Thus, although structures are now available for nearly all of the 20 aaRSs, only the 1.55   structure of truncated class II ThrRS (Sankaranarayanan *et al.*, 2000) has been solved at high resolution (<1.8  ). Moreover, with the exception of the recent report of different LeuRS conformations (Cusack *et al.*, 2000), crystallization strategies using appropriate ligands to select different stable conformers (Carter *et al.*, 1994) have found limited application. Crystals obtained for other aaRSs may represent only a limited subset of their possible biochemical states and may therefore be insufficient for fully describing conformational changes that occur during catalysis. An important example is the case of the class Ic TyrRS, for which only a single conformational state is known, but for which there is conclusive evidence for large-scale conformational changes during amino-acid activation (Fersht *et al.*, 1988).

Tetragonal crystals of the tryptophanyl-5'AMP (TAM) complex of TrpRS (Doubl   *et al.*, 1995) diffract much better than other TrpRS crystals. We have exploited this high-resolution diffraction pattern to provide more accurate information on the TrpRS structure. Substitution of sulfur by selenium for the ten methionines (~3% of the 328 residues) and the existence of only one monomer per asymmetric unit in this crystal form made it perfectly suited to anomalous dispersion phasing. Accordingly, a native data set and a three-wavelength MAD data set were recorded, followed by a second run with additional crystals dedicated to careful measurement of the data for the peak absorption wavelength. All crystals diffracted to a resolution better than 1.7   with mosaicities of 0.2–0.7 . Data collection was limited in practice to ~1.7   both by beam intensity and by the inappropriate match between the asymmetric unit cell and the solid angle coverage of the detectors.

In this report, we evaluate the three different sources of experimental phases and compare the three resulting structures. Although multiwavelength anomalous dispersion (MAD) has become the standard experimental source of phase information (Hendrickson, 1999), we find that SAD consistently produced comparable electron-density maps from substantially fewer experimental measurements, consistent with a similar observation of Rice *et al.* (2000). More importantly, all experimental phasing methods obviate the need to rely on model-based phases to image higher resolution diffraction patterns and hence provide essentially unbiased electron-density maps (Burling *et al.*, 1996).

The three independent but rather similar solutions of the TrpRS–TAM complex comprise all of the non-H atoms for 326 of 328 residues and provide unbiased evidence for the solvation structure. They provide independent estimates of coordinate errors for comparison with those estimated, for example, by the method of Luzzati (1952). These r.m.s. coordinate differences (~0.25–0.38  ) are substantially smaller than the r.m.s. deviations of the three new structures from the previous 2.9   model. Nevertheless, the three different ($|F_{\text{obs, 300 K}}| - |F_{\text{obs, 100 K}}|$) difference Fourier maps indicate that despite differences between the models themselves, there are few if any significant differences between the structures represented by the lower resolution room-temperature and higher resolution cryogenic amplitudes. The new structures allow us to specify more confidently side-chain rotamers in the non-polar packing arrangements affected by the TrpRS rigid-body domain motion. Additionally, we have identified six bound cryoprotectant glycerol molecules and seven sulfate ions. Coordinates and structure factors for all three structures (native protein, MIRAS phases; seleno derivative, SAD phases and MAD phases) have been deposited in the PDB with the codes 1i6m, 1i6l and 1i6k, respectively. Functional implications of the high-resolution structures will be discussed in a subsequent publication.

2. Experimental

2.1. Crystallization

TrpRS from *B. stearothermophilus* was crystallized by microdialysis in the presence of tryptophan (2 mM), ATP (10 mM) and magnesium (10 mM) as previously described (Carter & Carter, 1979). Systematic response-surface analysis (Carter & Yin, 1994) revealed a crystal-growth optimum in ~2.0 M K₂HPO₄ at pH 7.6 and at 310 K. The four-sided bipyramidal crystals have edge lengths of ~0.25 mm and belong to the space group *P*4₃2₁2. Unit-cell parameters for a cryoprotected crystal are $a = b \simeq 63$, $c \simeq 233$   and the resolution limit, even using synchrotron sources, is around 3.0  . Soaking such crystals in a solution of 3.5 M ammonium sulfate (AS) prior to cryoprotection led to a rapid phase transition to a slightly smaller unit cell with unit-cell parameters $a = b \simeq 60$, $c \simeq 232$   and improved the diffraction resolution from 3.0   to better than 1.7   (Carter *et al.*, 1994). The decreased volume of the unit cell must be associated with

Table 1
Data-collection statistics.

Values in parentheses represent the outermost resolution shell.

Crystal type	Native	Se derivative	Se derivative	Se derivative	Se derivative	
	$\lambda_0 = 1.0 \text{ \AA}$ (bn1)	$\lambda_3 = 0.985 \text{ \AA}$ (bn1)	$\lambda_1 = 0.978 \text{ \AA}$ (bn1)	$\lambda_2 = 0.979 \text{ \AA}$ (bn1)	4 crystals (bn2)	1 crystal (bn2)
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$
Unit-cell parameters (\AA)	$a = b = 59.60,$ $c = 231.27$	$a = b = 60.12,$ $c = 232.66$			$a = b = 59.84,$ $c = 232.62$	$a = b = 59.77,$ $c = 232.65$
Observations	332053	546675	518574	502151	1045314	384870
Unique reflections	45713	46732	46498	46545	51522	44261
Resolution limits (\AA)	25–1.72	25–1.72	25–1.72	25–1.72	50–1.72	25–1.72
Completeness (%)	83.4 (75.7)	94.5 (80.9)	92.8 (60.6)	91.5 (64.4)	95.7 (68.8)	87.4 (45.6)
$\langle I/\sigma(I) \rangle$	15.9	6.1	6.8	7.0	8.1	12.6
Reflections with $I > 3\sigma(I)$ (%)	82.9	65.0	55.2	61.4	71.4	72.7
Overall R_{sym}^\dagger	0.042 (0.315)	0.083 (0.354)	0.089 (0.437)	0.080 (0.322)	0.078 (0.571)	0.066 (0.388)
(Redundancy)	4.9	2.43‡	2.45‡	2.48‡	5.15‡	3.15‡

$^\dagger R_{\text{sym}}$ is defined as $\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the i th observation of reflection hkl and $\langle I(hkl) \rangle$ is the weighted mean of all observations (after rejection of outliers). ‡ Friedel mates separate.

some modification of active-site ligands evidenced elsewhere (Carter *et al.*, 1994; Coleman & Carter, 1984) and with crystal packing rearrangement, as crystals sometimes crack. The protein conformation differs substantially from that observed in crystals isomorphous to the original tetragonal crystals grown in the presence either of high ATP concentrations (Carter *et al.*, 1994) or ATP together with non-reactive tryptophan analogs (Yin *et al.*, 2001) and which belong to a third conformational state (Carter *et al.*, 1999).

Although the solubility of the TrpRS complex in AS remains compatible with crystallization, crystallizing directly from this salt was not straightforward. Moreover, cryoprotection of the transformed crystals was problematic, as the high glycerol concentrations necessary for cryoprotection simultaneously increased protein solubility while reducing ammonium sulfate solubility. Transfer from the original phosphate mother liquor first to AS and then to a 15% (v/v) glycerol, 2.5 M AS mixture nevertheless produced cryoprotection with the desired crystal phase transition. Gradual change of glycerol concentrations further decreased ice formation and the post-refined mosaicity (0.2 *versus* 0.7°). Flash-cooling also could be applied successfully to selenomethionyl-TrpRS, despite the previous observation that selenomethionyl-TrpRS crystals were temperature sensitive.

2.2. Data collection

All native and SeMet data sets were collected on the synchrotron beamline X12C at NSLS, Brookhaven. Data for the peak-wavelength SAD experiment were collected using the single-module CCD detector provided by W. Phillips and M. Stanton of the Rosenstiel Center. Native and MAD data sets were collected on the four-module successor to the single-module CCD detector (Phillips *et al.*, 2000). In both cases, the detector was offset along 2θ to measure high-resolution data and the κ -axis goniometer (Enraf–Nonius, The Netherlands)

was crucial for efficient recording of reciprocal space to the limiting resolution of 1.7 \AA from the very asymmetric unit cell of TrpRS crystals. Control over experimental setup, data gathering and reduction at X12C is provided *via* a graphical user interface (Sweet & Skinner, 1997).

A three-wavelength experiment (bn1) was performed at the absorption-edge peak, the inflection point and a remote higher energy wavelength, as determined by XANES excitation measurements. A native data set (bn1) was also collected at $\lambda = 1.0 \text{ \AA}$. All bn1 data sets were collected in 2–3 h using a single $2 \times 70^\circ$ pass

suggested by the user interface to give optimal coverage of reciprocal space. An oscillation range of 0.75° was recorded in 2 min exposures, with Friedel-flip geometry to improve the quality of the Bijvoet differences. In retrospect, a more complicated strategy segmenting the 70° range into smaller sweeps with alternating wavelengths and beam geometry might have minimized the impact of radiation damage on the MAD phasing calculations.

For the SAD (bn2) experiment we aligned the longest c axis along the spindle axis so as to record Bijvoet pairs related by a central mirror plane on the same frame. A high-resolution sweep was collected followed by a lower 3 \AA resolution sweep. Three other crystals were used to improve reciprocal-space coverage and redundancy in the bn2 experiment (see Table 1).

2.3. Data reduction and scaling

Data were reduced and integrated by *DENZO* and scaled independently using *SCALEPACK* (Otwinoski & Minor, 1997; Otwinowski, 1993), as reported in Table 1. Native and derivative crystals are very nearly isomorphous, although native unit-cell parameters are slightly shorter. Structure-factor amplitudes were calculated using *TRUNCATE* (French & Wilson, 1978) and then scaled to a reference data set using *SCALEIT* (Howell & Smith, 1992; Collaborative Computational Project 4, Number 4, 1994).

The sources of phase information are the anomalous and dispersive differences, which are weak for selenium and are more sensitive to data quality and outlier removal than are the amplitudes themselves. Resolution-dependent properties of the amplitudes and anomalous differences at the peak-absorption wavelength for both runs are illustrated in Fig. 1 (adapted from Brodersen *et al.*, 2000). Crystals used for bn2 were evidently superior to those used for bn1, which may have resulted from the gradual introduction of glycerol for cryoprotection: mosaicity was reduced (0.2 *versus* 0.7°), $\langle I/\sigma(I) \rangle$

Table 2
Scattering factors and phasing statistics.

(a) Refined scattering factors by *SHARP*.

	SAD								
	Ref†	<i>CHOOCH</i> ‡		bnl2, 1 crystal				MAD bnl1	MIRAS bnl1/bnl2
		bnl2	bnl2, 4 crystals	High-resolution sweep	Low-resolution sweep	bnl1			
<i>f'</i> λ ₂	-7.35	-7.7	-8.61	-6.24	-5.09	-7.80	-6.29	-6.30	
<i>f''</i> λ ₂	5.92	5.38	5.30	5.15	5.87	5.62	5.67	5.63	
<i>f'</i> λ ₁	-9.52	-9.61					-8.80	-8.73	
<i>f''</i> λ ₁	3.15	3.1					4.31	4.23	
<i>f'</i> λ ₃	-2.19	—					-2.19§	-2.19§	
<i>f''</i> λ ₃	3.46	—					3.76	3.71	

(b) Phasing statistics by *SHARP*. Upper resolution limit is 1.72 Å, except for values in italics, where it is 1.66 Å.

Phasing Data	SAD (bnl2)/(bnl1)				MAD (bnl1)			
	Peak λ ₂		Peak λ ₂		Edge λ ₁		Remote λ ₃	
	Isom.	Anom.	Isom.	Anom.	Isom.	Anom.	Isom.	Anom.
<i>R</i> _{Cullis} ¶								
Centric	—	—	0.389	—	0.306	—	—	—
Acentric	—	0.553 (0.670)	0.504	0.642	0.467	0.704	—	0.749
<i>R</i> _{kraut} ††								
Centric	—	—	0.197	0.381	0.225	0.375	0.239	0.391
Acentric	0.005 (0.012)	0.049 (0.066)	0.034	0.065	0.037	0.062	0.040	0.065
Phasing power‡‡								
Centric	—	—	4.38	—	4.44	—	—	—
Acentric	—	3.34 (2.40)	4.44	2.84	4.33	2.38	—	2.18
Figure of merit§§								
Centric		—			0.483			0.458
Acentric		0.531 (0.461)			0.489			0.524
After <i>SOLOMON</i>		0.920 (0.877)			0.919			0.865

† From an SeMet crystal on beamline X12C (Ramakrishnan & Biou, 1997). ‡ Estimated from the program *CHOOCH* by Evans & Pettifer (2001). § Kept constant during the refinement as reference. ¶ $R_{\text{Cullis}} = \sum_{hkl} (|F_{PH}| \pm |F_P|) - |F_H(\text{calc})| / \sum_{\text{cent}} (|F_{PH}| \pm |F_P|)$. †† $R_{\text{kraut}} = \sum_{hkl} (|F_{PH}| - |F_P + F_H(\text{calc})|) / \sum_{hkl} |F_{PH}|$. ‡‡ Phasing power = $(|F_H(\text{calc})|) / \text{phase-integrated lack of closure}$, where $F_H(\text{calc})$ is the calculated anomalous difference and E is the phase-integrated lack of closure $|F_{PH}(\text{obs}) - |F_{PH}(\text{calc})|$. §§ The figure of merit for a given reflection hkl is calculated as $\text{FOM} = |F(hkl)_{\text{best}}| / |F(hkl)|$, where $F(hkl)_{\text{best}} = \sum_{\alpha} P(\alpha) F_{hkl}(\alpha) / \sum_{\alpha} P(\alpha)$ and $P(\alpha)$ is the phase-angle probability distribution.

was higher and measurement noise indicated by R_{sym} was significantly reduced over the entire resolution range. Further, Friedel pairs were recorded on the same image for bnl2. As a result, $\Delta F/\sigma\Delta F$ was 1.78 and the signal-to-noise remained higher than 1.0 out to 2.0 Å for the bnl2 SAD data, compared with only 1.3 and 2.4 Å for the corresponding peak-wavelength bnl1 set. These advantages of the bnl2 set were reduced somewhat by the fact that the monochromator was offset from the absorption edge, as indicated by the higher r.m.s. (ΔF)/r.m.s. (F) for bnl1. At 1.7 Å, the signal represents almost 25% of the intensities for MAD (bnl1) against only 15% for SAD (bnl2).

2.4. Phasing

Positions of the Se atoms taken from the previous coordinates refined by Doublé *et al.* (1995) were input to *SHARP* (de La Fortelle & Bricogne, 1997) for heavy-atom parameter refinement. Since f'' is dependent on the environment of the heavy atom in the structure and on the X-ray energy (and

therefore on the characteristics of the beamline), it is worthwhile to start from reasonable values for anomalous scattering factors that will be further refined by *SHARP*. For our crystals, values of f' and f'' at the absorption edge and the inflection point were estimated using the program *CHOOCH* (Evans & Pettifer, 2001) from the X-ray fluorescence absorption spectrum obtained for bnl2. Values for the remote wavelength were taken from Sasaki (1989). Table 2(a) reports the values refined by *SHARP* for different SeMet crystals. Higher values of f'' (around 6 e⁻) are usually reported, suggesting that Se atoms have been inhomogeneously (or partially) oxidized.

Phase information was developed independently by *SHARP* for the MAD experiment and its SAD (bnl1) subset, two SAD (bnl2) experiments using either merged complete data from four different crystals or just high and low-resolution sweeps from a single crystal, a SIRAS experiment with the high-resolution sweep data set as the reference set together with the native data set and, finally, for a MIRAS experiment combining all bnl1 and bnl2

data. Phasing statistics are given in Table 2(b) for the MAD, SAD and MIRAS refinements, on which we will focus.

For the MAD experiment, the data set corresponding to the remote wavelength was taken as the reference data set for estimating scale and lack-of-isomorphism parameters, because its scattering-factor values are rather well estimated and stable and also because of its higher completeness. The common attributes (site positions, occupancies and temperature factors) are refined appropriately by *SHARP*. In a SAD refinement, the occupancies and the atomic scattering factors of the Se atoms are strongly correlated, so occupancies were assumed equal to 1 and were not refined during the first round, allowing the refinement of both f' and f'' . These values were then kept fixed while occupancies and anisotropic B factors were refined.

Refinement of the lack-of-isomorphism (LOI) parameters in *SHARP* was not entirely straightforward. As the low-resolution sweep was carried out much later than the high-resolution data collection on the single crystal, these two sets were safely treated as coming from distinct crystals. SIRAS

phases were also calculated, taking care that the lack of isomorphism between derivative and native data refined by *SHARP* remained within the acceptable range (<3.0). The risk of lack of isomorphism was even greater for the MIRAS experiment, which was performed by combining the MAD protocol with the native data set. The peak-wavelength data sets from bnl1 and bnl2 (single crystal) were combined for MIRAS phasing, as the relevant LOI parameters for each crystal data set refined against the reference remote wavelength did not reveal strong non-isomorphism. Merging and scaling intensities from the various crystals for the SAD (bnl2) experiment in *SCALEPACK* were also more appropriate than refining their LOI separately in *SHARP*.

Solvent flipping by *SOLOMON* (Abrahams & Leslie, 1996; Collaborative Computational Project, Number 4, 1994) provides dramatic phase improvement, especially when large numbers of reflections have bimodal experimental phase probabilities. We used 130 cycles, during which the radius of the solvent sphere was reduced from 2.4 Å to the maximum resolution of the data. The flipping factor was -1 and the optimal disordered solvent content was estimated at 46.9% by maximizing the correlation coefficient on E^2 from *SIGMAA* (Read, 1986). This value is lower than the 55% estimated for maximum-entropy solvent flattening at 2.9 Å resolution (Doublé *et al.*, 1995) and presumably reflects ordering of the first solvation sphere in cryoprotected crystals. Electron-density maps were calculated with *FFT* and *MAPMASK* (Collaborative Computational Project, Number 4, 1994) and displayed in the program *O* (Kleywegt & Jones, 1996) *via* the interface with *SHARP*.

2.5. Refinement

The previously solved 2.9 Å structure was taken as the starting model for these and other refinements used below in compiling statistics using the program *BUSTER* developed by GlobalPhasing Ltd. The core of this program implements statistical treatment of both experimental and model structure factors according to the maximum-likelihood and maximum-entropy methods (Bricogne, 1997). Its purpose is to generate and exploit quantitative descriptions of two main sources of random noise: errors in the current partial model and uncertainty arising from atoms missing from the model. Properly representing these errors overcomes many of the major shortcomings encountered by least-squares and difference Fourier methods when dealing with refinement and completion of partial structures (Bricogne & Irwin, 1996). Difference maps are then filtered according to non-uniform prior knowledge of where missing atoms are likely to be (Roversi *et al.*, 2000), by enforcement of a maximum-entropy condition which increases the signal-to-noise ratio of the final reconstruction of density for missing atoms.

The first part of the process is driven by refinement of the atomic parameters for the known fragment. External phase information output by *SHARP* as the Hendrickson–Lattman coefficients (Hendrickson & Lattman, 1970) are input to and constrain the refinement. At each cycle, approximations to an

elliptic Rice maximum-likelihood function and of its gradient and Hessian are passed on to *TNT* (Tronrud, 1997) to generate Agarwal–Lifchitz-type parameter gradients and curvatures and to carry out positional refinement with respect to stereochemical restraints. The fragment corresponds to those parts of the previous model in best agreement with the experimentally phased map output by *SHARP*.

By keeping an appropriate distance from the data, maximum-likelihood methods limit overfitting the observed amplitudes at phases too close to those of the initial fragment, which leads to the model bias to which least-squares is vulnerable. The detailed differences discussed below between the high-resolution structures, the previous 2.9 Å model and the improved definition of solvent atoms (Roach *et al.*, 2001) suggest that subtle model biases were avoided by using *BUSTER/TNT* and high-resolution experimental phase information.

2.6. Delaunay tessellation

Delaunay tetrahedra are space-filling simplices that minimize tetrahedron edge lengths between points in three dimensions, which in our case are given by residue centers of mass (Singh *et al.*, 1996; Tropsha *et al.*, 1996). Analysis of packing volumes associated with methionine S/Se atoms was performed using the Protein Core Alignment Map (ProCam; <http://mmlsun2.pha.unc.edu/~procam.html>), which performs Delaunay tessellation and likelihood weighting of the resulting tetrahedral simplices and provides volumes for each simplex and estimates of their Z scores. Volumes for tetrahedra involving methionine residues were identified from the output list and analyzed as discussed in §3.3.

2.7. Validation criteria

The computer programs *PROCHECK* (Laskowski *et al.*, 1993) and *SFCHECK* (Vaguine *et al.*, 1999) were used to evaluate the quality of all resulting models. A useful complementary structure-validation tool was provided by running the programs *REDUCE* and *PROBE* (Word, Lovell, LaBean *et al.*, 1999; Word, Lovell, Richardson *et al.*, 1999) to generate placement of all H atoms including their van der Waals interactions in order to highlight stereochemical clashes. For purposes of coordinate comparisons, all the models were superimposed using 42 C^α atoms taken from the first half of the Rossmann fold domain, which has been used as a reference orientation for previous comparisons among class I aaRSs (Doublé *et al.*, 1995; Ilyin *et al.*, 2000).

3. Results and discussion

The MAD and SAD intensity data recorded for the TrpRS–TAM complex represent the highest resolution to which experimental phasing has yet been achieved for any full-length aaRS. Since we have increased the number of amplitudes for this structure roughly fourfold, we wished to solve the high-resolution cryogenic TrpRS complex structure free of bias from the initial model, solved at room temperature and in the

absence of glycerol. This study is distinguished by the availability of three distinct phasing paths and by the consistent use of maximum-likelihood procedures implemented in *SHARP* and the *BUSTER/TNT* refinement combination. We discuss each of these briefly in §§3.1 and 3.2, followed by comparisons of native and SeMet TrpRS in §3.3 and of the low-resolution, high-temperature and high-resolution, low-temperature models in §3.4 and conclude with discussions of stereochemical validation (§3.5) and solvent structure (§3.6).

3.1. Comparisons between SAD, MAD and MIRAS phasing

Envelopes used for solvent flipping were derived from the respective experimental electron densities by the *SOLOMON* script from *SHARP* and were therefore unbiased toward any model and do not compromise the independence of distinct phase sets. Because map quality is influenced by data completeness and other factors besides phase error, the mean differences between phases calculated by each method and

those from the relevant refined models are probably the most meaningful error estimates for comparing phase sets (Fig. 2).

Not surprisingly, the MAD experimental phases are the most reliable in the absence of solvent flipping, especially at low resolution, while the MIRAS method eventually produced the largest phase-error estimates. However, phase improvement by solvent flipping with both SAD experiments actually produced the lowest phase errors, especially at high resolution. The evident superiority of the bnl2 SAD phase set relies in part on inherent crystal diffraction properties (Fig. 1 and §2.3), but in fact the bnl1 SAD phase set is the next best phase set ($\sim 25.0^\circ$ overall phase error with respect to the corresponding *SOLOMON* map, compared with 22.3° for bnl2, 26.7° for MAD and 29.6° for MIRAS). Thus, it is comparable to and perhaps better than the MAD phase set from the same crystal.

Correlation coefficients calculated by *rs_fit* in *O* between model coordinates and the respective maps confirm that solvent flipping improves the SAD phases dramatically to a map quality comparable with those obtained by solvent flipping of maps from the MAD and MIRAS phases (Fig. 3, Table 3). Smaller phase errors for SAD phasing are also reflected in superior refinement statistics and reduced coordinate errors (see §3.2).

Maps of the least reliable loop, residues 112–119 (Figs. 4a and 4b), underscore the value of having performed multiple phase determinations which provide reliable low-resolution envelopes and reduced false interpretations during model rebuilding. Despite the statistical superiority of the SAD (bnl2) phases, both MAD and MIRAS phases show more continuous density in this region and therefore probably benefit from higher completeness as well as the additional anomalous, dispersive and isomorphous differences at lower resolution: acentric reflections to 3 \AA have figures of merit (FOM) > 0.8 . Continuous density in experimental maps,

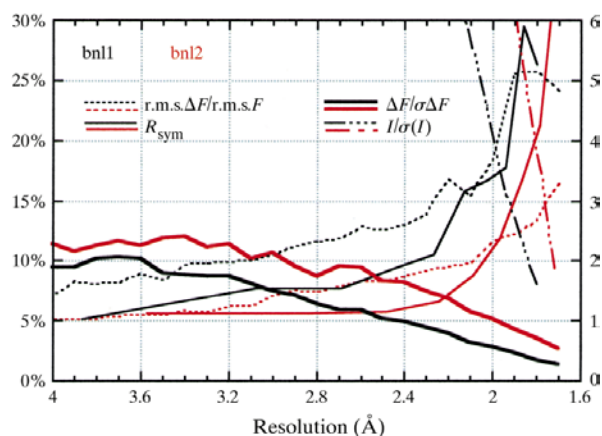


Figure 1 Resolution-dependent behaviour of quality indicators for two different data sets. R_{sym} and $r.m.s.(\Delta F)/r.m.s.(F)$ measured in percent on the left y axis, whereas $I/\sigma(I)$ and $|\Delta F|/\sigma(\Delta F)$ on the right are absolute numbers. The $r.m.s.(\Delta F)/r.m.s.(F)$ shows the average relative contribution of the anomalous signal in resolution bins and $|\Delta F|/\sigma(\Delta F)$ the signal-to-noise level of the anomalous differences.

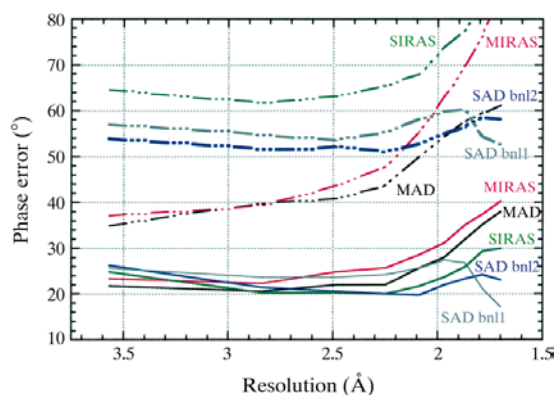


Figure 2 Average non-weighted difference between the phases calculated from the final refined model and the relevant phases output first by *SHARP* (dotted) then by *SOLOMON* (solid) versus resolution (\AA) for each of the different phasing methods.

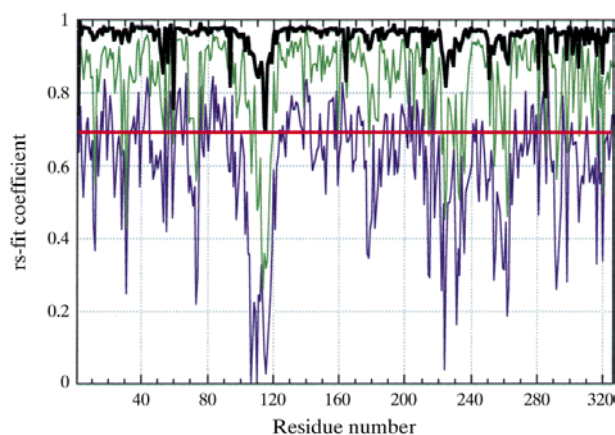


Figure 3 Real-space correlation coefficients, calculated versus residue number by *rs_fit* in the program *O*, between TAM-TrpRS structure coordinates and successive electron-density maps calculated using experimental phases output by *SHARP* (purple), then after solvent modification with *SOLOMON* (green) and using refined centroid phases for the SAD (bnl2) experiment (black). The red line indicates the cutoff used for pruning the initial 2.9 \AA model.

subsequently improved by solvent flipping, is accompanied by higher real-space correlation with the model compared with the SAD maps (see Table 3).

In summary, the most accurate TrpRS–TAM complex model produced by this study came from a SAD (bnl2) experiment, which also benefitted from crystals with the finest

diffraction quality of any used in this study. SAD phasing from the peak-wavelength subset gave nearly identical results to the intact MAD data set and may even have outperformed it, suggesting that radiation damage became important during the multiwavelength experiment. Additional phase errors may also be introduced by using reflections from three different crystals (one native and two derivatives), despite the fact that global non-isomorphism and model imperfection parameters, first estimated and then refined by *SHARP*, remain acceptable for the MIRAS parameter refinement.

Anomalous scattering has become the foremost source of experimental phase information and the MAD experiment remains the conventional path to a satisfactory least-squares determination of the phase angle from this weak signal. However, statistical phasing and density-modification methods have enhanced the quality of phasing that is routinely possible from weaker phasing signals (Abrahams & Leslie, 1996; Doublé *et al.*, 1994). The higher signal-to-error ratio of Bijvoet differences with respect to dispersive differences (see §2.3) has been exploited quite effectively by others from relatively light atoms at a single wavelength far from the scattering edge (Dauter & Dauter, 1999; Dauter *et al.*, 1999) and even at 2 Å resolution (Brodersen *et al.*, 2000).

A delicate balance must be achieved in order to realize the intrinsic advantages of the MAD algorithm in the context of the increased error propagation in the system of non-linear phasing equations with multiple wavelengths and non-isomorphism arising from radiation damage (Hendrickson & Teeter, 1981; Minor *et al.*, 2000). Comparison of our phase errors, map correlations and refinement statistics substantiates the somewhat counter-intuitive claim (Rice *et al.*, 2000) that optimizing the anomalous difference signal near the absorption edge for SAD phasing (Harvey *et al.*, 1998) should be the

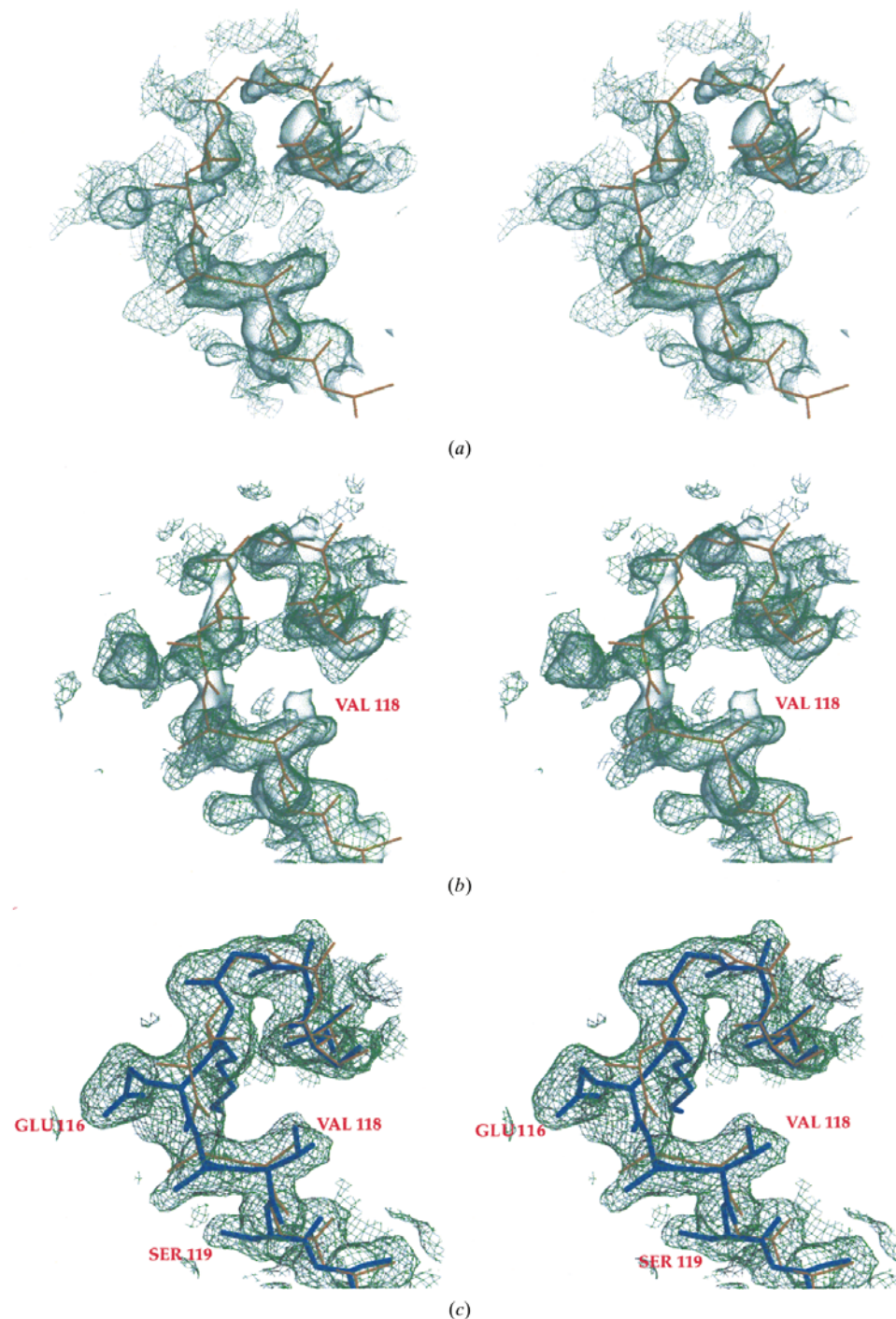


Figure 4

Electron-density maps for the loop, residues 111–120 around the former room-temperature 2.9 Å model (orange) or the SAD refined model (blue). (a) and (b) are, respectively, the (bnl2) SAD and (bnl1) MAD maps contoured at 1σ . The mesh corresponds to the experimental (*SHARP*) maps. The transparent maps are after solvent flipping. Discontinuities remain for SAD, while Val118 appears in MAD. (c) is the centroid final SAD (bnl2) *BUSTER/TNT* map contoured at 0.5σ .

Table 3

Overall real-space correlation coefficients between maps and corresponding refined models at 1.7 Å (at 2.9 Å room temperature).

MAPS	MAD (bnl1)	SAD (bnl1)	MIRAS (bnl1–2)	SAD (bnl2, 4 crystals)	SAD (bnl2, 1 crystal)	SIRAS (bnl2)
<i>SHARP</i>	0.772 (0.689)	0.538 (0.442)	0.792 (0.747)	0.644 (0.540)	0.488 (0.413)	0.670 (0.605)
<i>SOLOMON</i>	0.907 (0.810)	0.837 (0.693)	0.895 (0.833)	0.878 (0.752)	0.904 (0.779)	0.909 (0.830)
<i>SIGMAA</i>	0.901 (0.770)	0.880 (0.776)	0.837 (0.702)	0.894 (0.785)	0.896 (0.801)	0.935 (0.804)
<i>BUSTER</i>	0.960 (0.811)	0.926 (0.753)	0.903 (0.743)	0.950 (0.821)	0.949 (0.823)	0.972 (0.820)

method of choice in the context of limited availability of synchrotron time and become a preferred alternative for *de novo* and high-throughput structural genomic structure solution.

3.2. The quality of higher resolution phasing and the refined models

Table 4 lists refinement statistics for three refined TAM complex structures (native and two SeMet structures) and for the original room-temperature 2.9 Å structure. Fig. 5 illustrates the quality of the final 1.7 Å ($2|F_{\text{obs}}| - |F_{\text{calc}}|, \varphi_{\text{calc}}$) SAD centroid map, which displays features expected from a map at this resolution. Density is provided for all non-H protein atoms of residues 1–326 in each of the three final maps at this resolution when contoured at 1σ .

The overall real-space correlation coefficients with the respective final ($2|F_{\text{obs}}| - |F_{\text{calc}}|, \varphi_{\text{calc}}$) electron-density maps are superior to 0.90 for all three structures (Table 3). Surprisingly, the best agreement between model and map has been obtained for the SIRAS centroid map (0.97), in spite of the relatively less accurate experimental SIRAS phases (Fig. 2). For comparison, correlation coefficients for the previous room-temperature 2.9 Å model are substantially lower, ranging from 0.74 for the MIRAS final centroid map to 0.82 for the final SAD (bnl2) centroid map.

Overall *B*-factor values are 25.8 Å² for the SAD structure, 28.7 Å² for the MAD structure and 32.8 Å² for the native MIRAS structure, suggesting a mean *B* value of 29.5 ± 3.1 Å² for the 1.7 Å structures. The *B*-factor values correspond to the mean values of the pair of *B* values evaluated by *SFCHECK* from the respective Wilson plots (Wilson, 1942) and by Patterson origin peak (Rogers, 1965) (28.4/24.2 Å² for all SAD data, 32.0/26.2 Å² for the MAD and 34.3/27.4 Å² for MIRAS). The Patterson origin peak method is more suitable to limited resolution data, providing a *B* factor of 41.9 Å² for the room-temperature 2.9 Å structure.

Side-chain rotamers are mostly identical in the three structures, except for Arg59, Arg60, Arg164 and Arg258 and also a handful of lysine side chains. Note that among the latter Lys111, Lys191 and Lys195 are located within the active site and their positional variation between native and derivative models may reflect chemical changes associated with the phase transition on transfer to AS. The last important difference between structures concerns the conserved Glu116 within the

most delicate region to refine, loop 112–119 (Fig. 4), which adopts different positions in all three structures.

After refinement, the three high-resolution models are in substantial agreement with one another, providing an independent basis for estimating overall coordinate errors. For comparison, the coordinate

error estimated by the method of Luzzati output by *SFCHECK* (Fig. 6) for the 2.9 Å structure was 0.52 Å, while for the 1.7 Å data it was 0.21 ± 0.005 Å for three structures refined to bnl1 data and 0.19 ± 0.007 Å for three structures refined to bnl2 data. Increasing the resolution of the data therefore reduced overall coordinate uncertainty to ~40% of its previous value. Moreover, the difference between values for bnl1 and bnl2 structures is three times the standard deviation of individual determinations, highlighting the superior amplitude data provided by the bnl2 crystals.

3.3. Differences between native and selenomethionylated TrpRS

Flash-cooling apparently makes SeMet-TrpRS crystals diffract essentially as well as native crystals, whereas cooling SeMet crystals below 285 K abolishes all diffraction. An explanation for the loss of diffraction previously observed at temperatures below 285 K was destabilization of hydrophobic interactions made by a ‘methionine zipper’ involving Met314, Met318 and Met322 from the C-terminal long helix, and Met92 that helps to hold the long helix to the three α -helices at the bottom of the tryptophan-binding site (Doublé *et al.*, 1995). The long helical motif contributes to the high value of coordinate differences between the room-temperature and the new 100 K solved structures (r.m.s.d. 1.38–1.47 Å *versus* ~0.8 Å for the Rossmann fold domain and ~1 Å for the helical bundle; Fig. 7). However, side-chain rearrangement and especially new ionic interactions account for most of this higher value. Curiously, native and selenomethionylated protein also appear to differ in this region.

The r.m.s. coordinate difference is ~0.25 Å between the two SeMet structures, consistent with Luzzati estimates, but is ~0.36 Å between the MAD and MIRAS, and ~0.40 Å between the SAD and MIRAS structures. The difference between these two groups of values is roughly 20 times the standard deviation of the coordinate error estimates. It is therefore statistically significant and is likely to reflect subtle differences induced by the presence of selenium rather than sulfur in the ten methionine side chains. Evidence for this conclusion was derived from the packing volumes associated with methionine side chains. Delaunay tessellation (Singh *et al.*, 1996; Tropsha *et al.*, 1996) (§2.6) provided a decomposition of residue-based tertiary interactions into tetrahedra of nearest neighbors for the three structures, 27 of which involve methionine. Three of these formed different neighbors in the

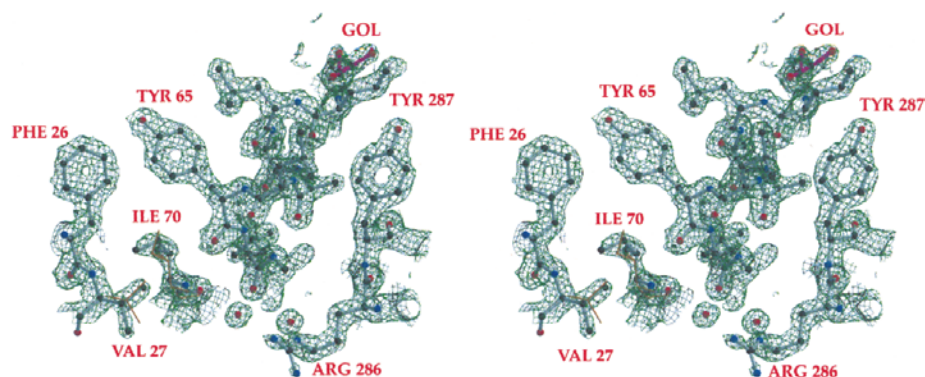


Figure 5
Final centroid *BUSTER/TNT* electron-density map from the SAD (bnl2) data-set contoured at 2σ . Val27 and Ile70 of the former model (orange) have been reoriented in the 1.7 Å resolution map. A glycerol molecule (GOL) has been identified in this region of the map.

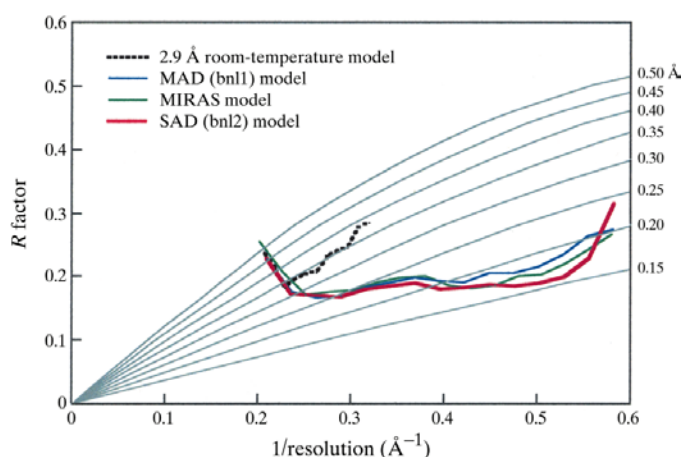


Figure 6
Luzzati plots for the three deposited structures. *R*-factor values were output by *SFCHECK* and truncated below 5 Å. *R* factors for the former 2.9 Å coordinates lie along the curves of estimated errors 0.35–0.40 Å, while those for the new structures are between 0.15–0.20 Å. Estimated coordinate errors reported in §3.3 are higher because *SFCHECK* uses the entire range of resolution, despite the application of a ‘soft’ low-resolution cutoff to structure factors in order to reduce the disordered solvent contribution.

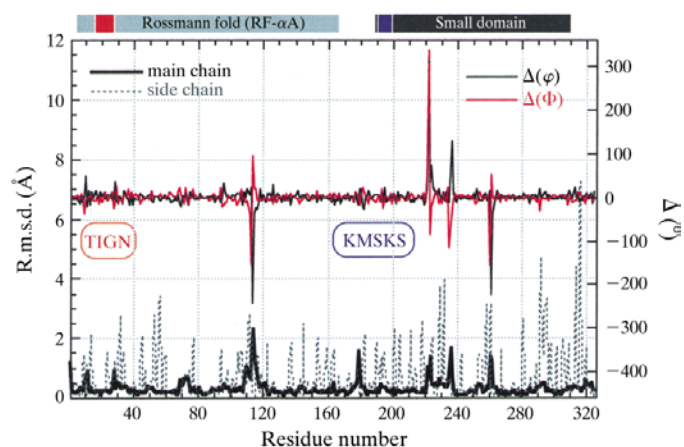


Figure 7
Conformation changes between the former room-temperature 2.9 Å model and the SAD 1.7 Å model *versus* residue number. Upper part: absolute differences between corresponding Ramachandran (φ , ψ) angles. Lower part: r.m.s. coordinate differences for main-chain (solid dark line) and side-chain (dashed grey line) residues.

native and SeMet structures. Of the 21 remaining simplices, 19 had identical vertices in all three structures, permitting evaluation of standard deviations for volumes in the two SeMet structures, and these were examined for packing-volume differences.

Comparison of all 19 tetrahedra failed to reveal significant changes between native and SeMet TrpRS. However, volumes associated with methionines in the C-terminal ‘methionine zipper’ are larger by an average of 0.32 \AA^3 in SeMet TrpRS, compared with a corresponding mean change of -0.002 \AA^3 for methionine-

containing tetrahedra elsewhere in the molecule and an average uncertainty of 0.25 \AA^3 in the volumes of equivalent tetrahedra in the two SeMet structures. They are consistent with the observation (not shown) that residue-by-residue structural deviations between native and SeMet structures are conspicuously higher near the C terminus. Statistical significance was also assessed by multivariate regression of simplex volumes, in which the 19 tetrahedra were divided into two groups according to the presence of one or more of the four ‘zipper’ methionines. The Student’s *t*-test probability for the grouping was 0.025, supporting the inference that packing volumes of methionine-containing tetrahedra from the C-terminal methionine zipper are drawn from different populations in native and SeMet TrpRS.

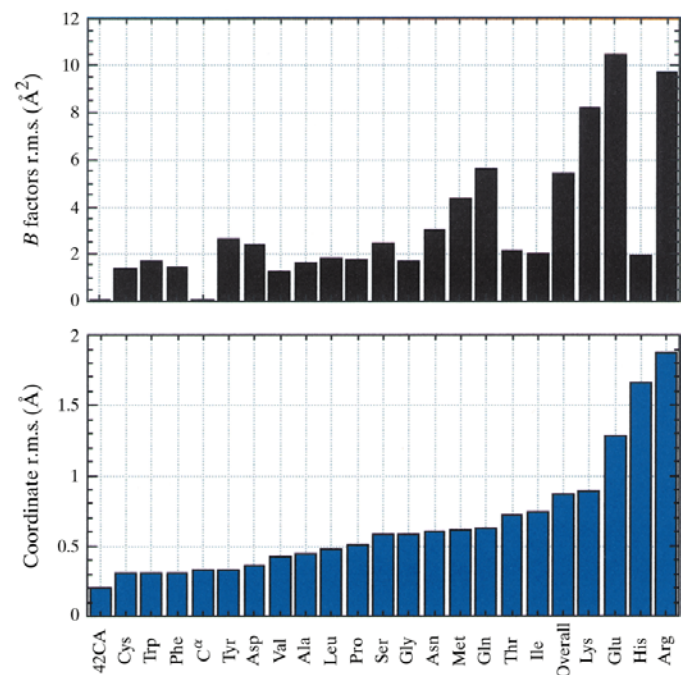


Figure 8
Variance characteristics of the refined TrpRS models. (a) R.m.s. *B*-factor values from the SAD (bnl2) 1.7 Å refined model. (b) R.m.s. coordinate difference after superposition of the new SAD model onto the 2.9 Å model using 42 C^α atoms from the RF.

3.4. Differences between room-temperature and low-temperature structures

Coordinate differences between the 2.9 Å room-temperature structure and the higher resolution structures are ~1 Å or nearly five times the estimated error for the two SeMet structures and roughly twice the coordinate error estimated for the 2.9 Å room-temperature model. Using the variance of the three high-resolution structures to estimate their mean coordinate errors suggests that the discrepancies between the low- and high-resolution structures are 3–4 times larger than the r.m.s. error of the new models. Thus, the high-resolution structures are significantly different from the structure solved previously at 2.9 Å. It is, therefore, of some importance to identify differences between the two sets of structures and attribute them to the appropriate sources where possible.

The TrpRS monomer consists of two main domains, a canonical dinucleotide-binding fold which includes the dimer interface and a helical domain, which were described earlier by Doublie *et al.* (1995) and whose relative position rearrangements were detailed in subsequent crystal structures (Ilyin *et al.*, 2000; Yin *et al.*, 2001). Despite their overall 1 Å r.m.s. coordinate differences with the previous room-temperature 2.9 Å model, the cryoprotected TrpRS structures complexed with the adenylate product differ only by 0.4 Å r.m.s. for the backbone atoms and hence have essentially the same overall conformation.

Three loops can be identified whose Ramachandran angles φ and ψ deviate significantly from the room-temperature structures (Fig. 7): loop 112–119, located at the entrance of the active site in the Rossmann fold domain, and two loops (loops 220–235 with a Pro236 peptide flip and 259–266) that link successive α -helices in the small domain. Each loop is likely to be crucial for cognate tRNA^{Trp} interaction (Doublie *et al.*, 1995), the former to anchor the acceptor stem of the tRNA and the others for anticodon recognition. In the absence of tRNA, they all present a high solvent accessibility and are likely to fluctuate, rendering positional refinement difficult. Finally the short belt Pro10–Gly12 (Ilyin *et al.*, 2000) that allows the motion of the α A-helix relative to the Rossmann fold is apparently also rearranged. The large variance in positions of these loops seems to be because of the limiting resolution of the 2.9 Å data.

The main differences between low- and high-resolution structures, however, are in the side-chain rotamers, which differ enough that the r.m.s. coordinate difference for side-chain coordinates is 1.2 Å. Fig. 8 displays the ascending distribution of r.m.s. coordinate differences between high- and low-temperature structures for each residue type. Aromatic residues, in particular, have moved little from their positions in the original model. Long charged side chains, mostly on the surface, have the largest r.m.s. values, as glutamate and arginine side chains exposed new rotamers to the solvent. As a consequence, several ionic hydrogen bonding interactions in the original model had to be revised: those between Lys201/Glu295–302, Arg160/Glu101 and Arg182/Glu145 no longer

Table 4
Refinement statistics.

	SeMet SAD λ_2 (bnl2)	SeMet MAD λ_3 (bnl1)	Native MIRAS	Native 2.9 Å (300 K)†
Final model parameters				
Resolution range (Å)	20.0–1.72	20.0–1.72	13.1–1.72	7–2.86
No. of reflections ($I > 3\sigma$)	44725	45149	38006	8859
No. of atoms	31909	32467	29968	($I > 2\sigma$)
No. of protein atoms	3065	3066	3025	2617
No. of water/NH ₄ ⁺	2599	2599	2599	2580
No. of sulfate/glycerol	351/1	368/1	316/1	0/0
Solvent content (%)	7/6	7/5	7/4	0/0
R factor‡/all ref§ (%)	48.8	48.8	49.5	57.9
$R_{\text{free}}\ddagger$ (10% of ref; %)	20.9/20.0	23.3/21.1	22.5/20.9	19.7
B factors (Å ²)	24.1	26.8	25.7	27.6
All atoms	25.8	28.7	32.8	17.1
Average main-chain B factor	24.0	27.4	30.7	114.9
Average side-chain B factor	28.0	30.3	35.2	19.1
Average and solvent B factor	40.6	42.7	46.8	—
Quality indicators				
DPI§ (Å)	0.107	0.115	0.130	—
Bond lengths‡ (Å)	0.019	0.021	0.020	016
Bond angles‡ (°)	1.409	1.418	1.462	2.85
Torsion angles‡ (°)	15.54	15.85	15.54	22.80
Trigonal planes‡ (Å)	0.010	0.011	0.012	—
General planes‡ (Å)	0.016	0.016	0.018	—
Ramachandran¶ (%)	95.5	95.5	94.8	92.4
Overall G factor¶	0.14	0.10	0.11	–0.08
No. of bad contacts‡¶	0/66	0/76	0/46	0/—
Hydrogen-bond energy std deviation (kJ M^{-1})	2.4	2.8	2.8	2.8
Ω -angle deviation¶ (°)	3.1	4.4	3.2	4.4
ξ -angle deviation¶ (°)	2.0	2.8	4.3	0.6

† Doublie *et al.* (1995). ‡ Output by TNT (Tronrud, 1997). § Output by SFCHECK (Vaguine *et al.*, 1999). ¶ Output by PROCHECK (Laskowski *et al.*, 1993).

exist in the newly refined structures, while new ones between Arg175/Glu28 and Arg316/Glu313 have been identified. The salt bridge between Arg182 and Glu145 was previously proposed to be an electrostatic switch stabilizing ground-state conformations during the amino-acid activation step (Ilyin *et al.*, 2000), an interpretation that must be modified in light of the apparent loss of this interaction at higher resolution.

Differences occur throughout the TrpRS non-polar core. Histidines are surprisingly divergent from the 2.9 Å model, leading to the second largest r.m.s. differences (after arginine). Further, many of the hydrophobic residues in the core: Ile14, Ile39, Val40, Ile45, Val47, Ile70, Ile77, Ile140, Val143, Ile151, Ile183, Leu186, Ile204 and Ile228 had to be remodeled by choosing new rotamers to reposition C^γ and/or C^{δ1} to fit the high-resolution experimental electron density. In the same way, the high-resolution electron-density map also suggested different rotamers for the hydroxyl groups of threonine residues Thr106, Thr138, Thr190 and Thr205. Thus, the β -branched side chains isoleucine, valine and threonine all contribute disproportionately large coordinate errors (Fig. 8).

Coordinate differences between low- and high-temperature structures could reflect either subtle but real differences associated with cryoprotection and 4.7% shrinking of the unit-cell volume or errors in the low-resolution modeling and refinement. To distinguish between these two possibilities, we

calculated a 3.1 \AA ($|F_{\text{obs, cryo}}| - |F_{\text{obs, room temperature}}|$, φ_{final}) difference Fourier map, a section of which is shown in Fig. 9. This map is somewhat noisy, as a result of the evident lack of isomorphism. Fig. 9 shows, however, that lack of isomorphism does not prevent imaging of (selenium – sulfur) peaks in the methionine residues, which are the highest in the map. Elsewhere, the difference map is essentially featureless, suggesting few if any differences between these two types of structures. Fig. 10 displays an example of peaks in the difference Fourier map which are associated with minor modifications (a slight twist of Asp223 and Ile238 side chains). In contrast, major structural changes in the newly refined models, such as the peptide flip at Pro236 and reorientation of Thr222, which remove an unfavourable interaction, are not indicated in this 220–236 loop region of the map. The relative flatness of the difference Fourier map throughout the structure therefore suggests that model differences are actually errors in the low-resolution model and not consequences of flash-cooling.

3.5. Stereochemical validation

Consistent with the notion that the newly refined structures correct errors in the previous model, the number of severe clashes between pairs of distinct residues identified using *PROBE* decreased from 104 for the 2.9 \AA model to 30 for the 1.7 \AA SAD structure. The histogram (Fig. 11) reports the occurrences of residue types involved in the clash lists for both former and newly refined models. The most serious clash in the 2.9 \AA model concerned the ϵ -methyl group of Met193, which collided severely with the C^β of Thr15. In the high-resolution map, the extremity of Met193 is clearly seen to be inverted, alleviating this clash (Fig. 9). Side chains that were misplaced in the new high-resolution electron-density map were actually making the worst contacts in the initial model. Proline is the residue implicated in the greatest number of bad contacts, owing to difficulty in assigning the ring pucker (Word, Lovell, LaBean *et al.*, 1999). Finally, note that refinement did not relieve clashes between water molecules and solvent-exposed residues (ten bad contacts implicate Arg and Lys) and that some of these clashes are intra-residue, particularly in Arg (total of 12) and Ile (total of five), which suggests that there may be problems with existing side-chain rotamer libraries.

Several residues appear to adopt alternate conformations, even at 1.7 \AA resolution. Two conformations have been identified for His31, whose second ring position interacts with Asp32, pulling that side chain to a new orientation with respect to the room-temperature structure. Electron density for two possible rotamer positions of Lys195 within the active site suggests a dynamic situation in which Lys195 sweeps an area that is potentially occupied by the leaving pyrophosphate (PP_i) group of ATP. This dynamic situation is likely to be associated with the density in both $2F_o - F_c$ and $F_o - F_c$ electron-density maps located on the position occupied by PP_i in other crystal structures and which could be occupied by glycerol or sulfate in the new structures. Alternate orientations for Ser312 O^γ provide a new hydrogen bond with

Arg309 CO in addition to the pre-existing bond with Asn308 CO. Additionally, the native MIRAS electron-density map suggests that Asp313, Ile16 $C^{\gamma 1}$ and Ile20 $C^{\delta 2}$ may adopt two rotamer conformations.

Finally, Gln94, which is located on the α C-helix in the dimer interface, was refined with alternate conformations for the amide group to optimize interaction with the amide group of its twofold-related counterpart. Indeed, the two amide groups face each other on the same plane (with a shortest distance of 2.7 \AA between atoms of the same type), while they make orthogonal ($>3.6 \text{ \AA}$) interactions in the ligand-free structure (Ilyin *et al.*, 2000). The pair of alternately flipped conformations manages both to comply with the strict crystallographic symmetry of the dimer (one monomer per asymmetric unit in $P4_32_12$) and to suggest, by promoting a single CO–NH interaction for any one dimer, the asymmetric functionality of dimeric TrpRS shown by catalysis experiments (Merle *et al.*, 1986). This residue, which is always either Asn or Gln in TrpRS sequences, may therefore play a key role in intramolecular communication through the dimer interface.

3.6. Solvent structure

When experimental phases are determined accurately to high resolution, the solvent structure near protein molecules in crystals can be located without the bias associated with model refinement (Burling *et al.*, 1996). The rule-of-thumb ‘one water molecule per protein residue’ for $\sim 2 \text{ \AA}$ resolution structures is confirmed for the SAD and MAD structures, which have more than 350 bound water molecules. Note that the native structure binds ~ 50 fewer water molecules. As a matter of fact, refinement against the less complete set of observed structure factors, constrained with MIRAS phases, eventually leads to a much noisier final refined map with regard to organized solvent. In general, water and other bound solute molecules superimpose between the three different maps but some differences occur related either to the variation of side-chain positions or map clarity. The average water molecule B factors for the different models lie between 40.6 and 46.8 \AA^2 and none of the individual values exceeds 85 \AA^2 .

Among the solvent molecules, four to six features in the asymmetric unit of density were assigned to be glycerol molecules rather than three non-aligned water molecules. An assumption was made that a glycerol molecule could be introduced into the active site, filling the place left by the pyrophosphate group. Seven sulfate ions were identified by the size and tetrahedral shape of the density and a suitable environment, such as the proximity of positively charged arginine or lysine side chains and/or the alignment of dipolar moments at the amino termini of α -helices, generally near a glycine residue (Chakrabarti, 1993). High B -factor values prompted us to lower their respective occupancies by half. Partial substitution of sulfate ions by selenate anions in the soaking solution in a later experiment did not allow visualization in either isomorphous or anomalous difference Fourier maps, suggesting that the sulfates are not tightly bound.

One bound molecule was identified as an ammonium ion on the basis of its distorted octahedral environment with three protein carbonyl O atoms and three bound solvent molecules and this interpretation was confirmed by peaks in both isomorphous and anomalous difference Fourier maps after substitution with caesium ion. This cationic site appears to be ubiquitous in TrpRS crystal structures and could therefore be exploited for model-bias-free solution by detecting even weak anomalous signal from such a cationic scatterer (Brodersen *et al.*, 2000; Dauter & Dauter, 1999; Dauter *et al.*, 1999).

Neither temperature variation nor the presence of cryoprotectant affect the phase transition observed when the composition of the mother liquor is switched from potassium phosphate to ammonium sulfate. Similarly, the TrpRS

conformation itself does not seem to be modified by variations of these parameters, even though glycerol and sulfate interact directly with the protein. Details of bound solvent and its correlation with the catalytic function will be analyzed in a subsequent paper.

4. Conclusions

Several reasons motivated us to record new data sufficient for high-resolution experimental phasing from a crystal whose structure had already been solved (Doublé *et al.*, 1995). Firstly, many conclusions suggested at lower resolution raised important questions involving accurate placement of amino-acid side chains in the active site and elsewhere in the hydrophobic core, as well as the nature of interactions with bound ligands and solvent molecules. Secondly, the previous structure was determined using considerably fewer data than we knew to be available from the high-quality diffraction diagram evident from previous data-collection efforts. Thirdly, experimental phasing was accessible directly from the availability of selenomethionylated TrpRS. In the course of data collection, it became apparent that comparing the various phasing strategies in a meaningful way might be useful in a wider context. Finally, we were anxious to investigate the consequences arising from the model bias inherent in using models refined at low resolution for the purpose of phasing higher resolution structures.

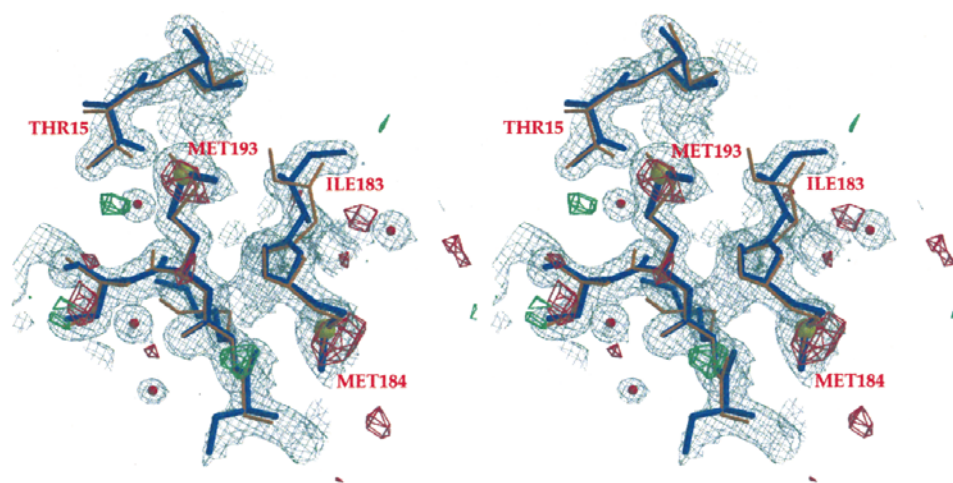


Figure 9

A 3.1 Å resolution $F_{\text{obs, derivative, 100 K}} - F_{\text{obs, native, 300 K}}$ synthesis contoured at 3.3σ shows positive peaks (red) around Se positions (yellow) for Mse184 and 193 in the 1.7 Å model (blue). The program *PROBE* pointed out the need for reorienting the ϵ -methyl group of Met193, which made a severe clash with Thr15 in the former model (orange). The Ile183 side chain has been reoriented to relieve the clash. The absence of difference Fourier peaks suggests that the original model was incorrect. Water molecules are displayed as red balls.

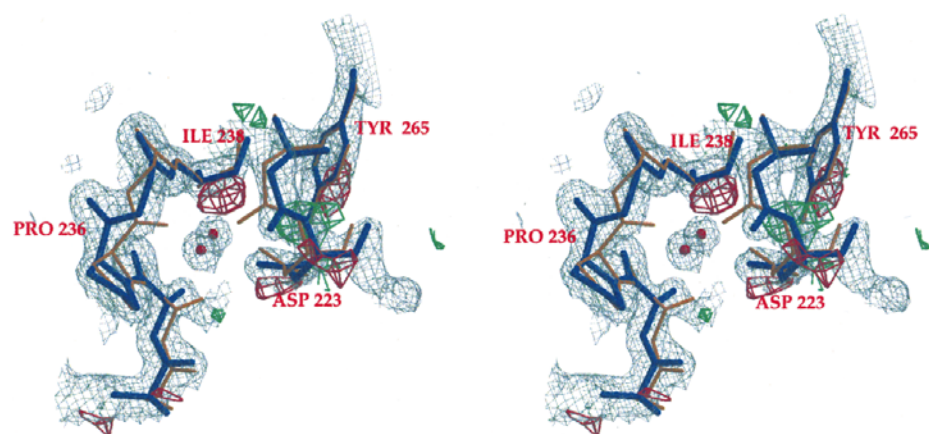


Figure 10

Positive peaks (red) and negative (green) of a 3.1 Å resolution $F_{\text{obs, derivative, 100 K}} - F_{\text{obs, native, 300 K}}$ are associated with minor rearrangements of Asp223 and Ile238. In contrast, the Pro236 peptide-bond flip avoids a 3 Å carbonyl-carbonyl interaction with Thr222 in the former model (orange). This flip does not generate peaks in the difference Fourier map. Reinterpretation of the map in this region places two water molecules (red) bridging the two sides of the loop in the new 1.7 Å model (blue).

Higher resolution data can probably be recorded from these crystals at a more intense beam-line, as diffraction spots to 1.6 Å were recorded in the limited exposures employed in this study. However, the asymmetric TrpRS unit cell will pose increasingly difficult problems with current detectors as the radius of the data increases. Moreover, by testing the oxidation effect on an SeMet crystal, we might also exploit the 40% enhancement of the anomalous signal observed in a recent study (Sharff *et al.*, 2000) and secure even higher resolution experimental phases. Development of Sayre-type equations (Roach *et al.*, 2001) could limit the

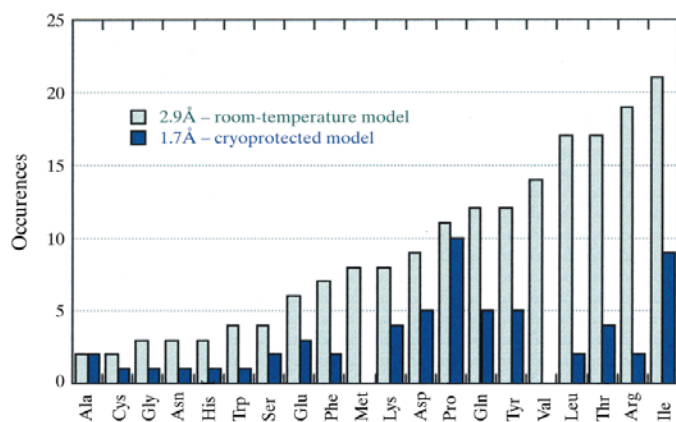


Figure 11

Stereochemical clashes between pairs of H atoms involving distinct residues output by *REDUCE* and *PROBE*. The light grey histogram represents the 2.9 Å room-temperature model. The dark blue histogram represents the 1.7 Å SAD (bnl2) model.

need to a single higher resolution data set by extrapolating the current experimental phases to higher resolution.

High-resolution data provide essential new information about the (TrpRS–TAM) complex itself. We were able to correct and refine the positions of many side chains, in particular those which compose the protein core. This model will be used to solve isomorphous crystal structures of several TrpRS mutants, Ile16Leu and Ile16Val, complexed with the adenylate (Longo *et al.*, 2001), whose resolution limits approach 2 Å, in order to study hydrophobic core interactions. The more accurate structural characterization of the KMSKS loop and evidence for the presence of solvent compounds near the α phosphate of the adenylate are expected to clarify the chemistry involved in adenylation and in the unusual crystal phase transition. These questions will be pursued in a subsequent publication.

This work was supported at UNC by NIH GM48519. We thank John Sondek and Brenda Temple (UNC) and Gwyndaf Evans (Global Phasing Ltd) for advice and assistance. Diffraction data for this study were collected at Brookhaven National Laboratory in the Biology Department single-crystal diffraction facility at beamline X12-C in the National Synchrotron Light Source. This facility is supported by the United States Department of Energy Offices of Health and Environmental Research and of Basic Energy Sciences under prime contract DE-AC02-98CH10886, by the National Science Foundation and by National Institutes of Health Grant 1P41 RR12408-01A1. All molecular graphics were created using *BOBSCRIPT* (Kraulis, 1991; Esnouf, 1997) and rendered using *RASTER3D* (Merritt & Murphy, 1994).

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* **D52**, 30–42.
Arnez, J. G. & Moras, D. (1997). *Trends Biochem. Sci.* **22**, 211–216.

- Bricogne, G. (1997). *Methods Enzymol.* **276**, 361–423.
Bricogne, G. & Irwin, J. J. (1996). In *Proceedings of the CCP4 Study Weekend. Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey. Warrington: Daresbury Laboratory.
Brodersen, D. E., de La Fortelle, E., Vornrhein, C., Bricogne, G., Nyborg, J. & Kjeldgaard, M. (2000). *Acta Cryst.* **D56**, 431–441.
Burling, F. T., Weis, W. I., Flaherty, K. M. & Brünger, A. T. (1996). *Science*, **271**, 72–77.
Carter, C. W. Jr (1993). *Annu. Rev. Biochem.* **62**, 715–748.
Carter, C. W. Jr & Carter, C. W. (1979). *J. Biol. Chem.* **254**, 12219–12223.
Carter, C. W. Jr, Doublé, S. & Coleman, D. E. (1994). *J. Mol. Biol.* **238**, 346–365.
Carter, C. W. Jr, Ilyin, V., Huang, X., Yin, Y., Retailleau, P., Vachette, P., Longo, A. & Hu, M. (1999). *Acta Cryst.* **A55**, Supplement, Abstract P05.04.009.
Carter, C. W. Jr & Yin, Y. (1994). *Acta Cryst.* **D50**, 572–590.
Chakrabarti, P. (1993). *J. Mol. Biol.* **234**, 463–482.
Coleman, D. E. & Carter, C. W. Jr (1984). *Biochemistry* **23**, 381–385. Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
Cusack, S., Hartlein, M. & Leberman, R. (1991). *Nucleic Acids Res.* **19**, 3489–3498.
Cusack, S., Yaremchuk, A. & Tukalo, M. (2000). *EMBO J.* **19**, 2351–2361.
Dauter, Z. & Dauter, M. (1999). *J. Mol. Biol.* **289**, 93–101.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
Delarue, M. & Moras, D. (1992). *Nucleic Acids and Molecular Biology*, Vol. 6, edited by F. Eckstein & D. M. J. Lilley, pp. 203–224. Berlin/Heidelberg: Springer-Verlag.
Doublé, S., Gilmore, C. J., Bricogne, G. & Carter, C. W. Jr (1995). *Structure*, **3**, 17–31.
Doublé, S., Xiang, S., Gilmore, C. J., Bricogne, G. & Carter, C. W. Jr (1994). *Acta Cryst.* **A50**, 164–182.
Eriani, G. G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990). *Nature (London)*, **347**, 203–206.
Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.
Evans, G. & Pettifer, R. F. (2001). *J. Appl. Cryst.* **34**, 82–86.
Fersht, A. R., Knill Jones, J. W., Bedouelle, H. & Winter, G. (1988). *Biochemistry*, **27**, 1581–1587.
French, G. S. & Wilson, K. S. (1978). *Acta Cryst.* **A34**, 517–525.
Harvey, I., Hao, Q., Duke, E. M., Ingledew, W. J. & Hasnain, S. S. (1998). *Acta Cryst.* **D54**, 629–635.
Hendrickson, W. A. (1999). *J. Synchrotron. Rad.* **6**, 845–851.
Hendrickson, W. A. & Lattman, E. E. (1970). *Acta Cryst.* **B26**, 136–143.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Howell, L. & Smith, D. (1992). *J. Appl. Cryst.* **25**, 81–86.
Ilyin, V. A., Temple, B., Hu, M., Li, G., Yin, Y., Vachette, P. & Carter, C. W. Jr (2000). *Protein Sci.* **9**, 218–231.
Kleywegt, G. A. & Jones, T. A. (1996). *Acta Cryst.* **D52**, 826–828.
Kraulis, P. (1991). *J. Appl. Cryst.* **24**, 946–950.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). *J. Mol. Biol.* **231**, 1049–1067.
Longo, A., Retailleau, P. & Carter, C. W. Jr (2001). In preparation.
Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.
Merle, M., Trezeguet, V., Graves, P. V., Andrews, D., Muench, K. H. & Labouesse, B. (1986). *Biochemistry*, **25**, 1115–1123.
Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* **D50**, 869–873.
Minor, W., Tomchick, D. & Otwinowski, Z. (2000). *Structure Fold. Des.* **8**, R105–R110.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.

- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Phillips, W. C., Stanton, M., Stewart, A., Qian, H., Ingersoll, C. & Sweet, R. M. (2000). *J. Appl. Cryst.* **33**, 243–251.
- Ramakrishnan, V. & Biou, V. (1997). *Methods Enzymol.* **276**, 538–557.
- Read, R. J. (1986). *Acta Cryst.* **A42**, 140–149.
- Rice, L. M., Earnest, T. N. & Brünger, A. T. (2000). *Acta Cryst.* **D56**, 1413–1420.
- Roach, J., Retailleau, P. & Carter, C. (2001). *Acta Cryst.* **A57**, 341–350.
- Rogers, D. (1965). *Computing Methods in Crystallography*, edited by J. S. Rollett, pp. 133–148. Oxford: Pergamon Press.
- Roversi, P., Blanc, E., Vornrhein, C., Evans, G. & Bricogne, G. (2000). *Acta Cryst.* **D56**, 1316–1323.
- Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C. & Moras, D. (1991). *Science*, **252**, 1682–1689.
- Sankaranarayanan, R., Dock-Bregeon, A. C., Rees, B., Bovee, M., Caillet, J., Romby, P., Francklyn, C. S. & Moras, D. (2000). *Nature Struct. Biol.* **7**, 461–465.
- Sasaki, S. (1989). *KEK Report*, pp. 88–14. National Laboratory for High Energy Physics, Japan.
- Sharff, A. J., Koronakis, E., Luisi, B. & Koronakis, V. (2000). *Acta Cryst.* **D56**, 785–788.
- Singh, R. K., Tropsha, A. & Vaisman, I. I. (1996). *J. Comput. Biol.* **2**, 213–221.
- Sweet, R. M. & Skinner, J. M. (1997). *Crystallographic Computing 7. Proceedings from the Macromolecular Crystallography Computing School*, edited by P. E. Bourne & K. D. Watenpugh. <http://www.sdsc.edu/Xtal/IUCr/CC/School96>.
- Tronrud, D. E. (1997). *Methods Enzymol.* **277**, 306–319.
- Tropsha, A., Singh, R. K., Vaisman, I. I. & Xheng, W. (1996). In *Pacific Symposium on Biocomputing 1996, Hawaii, USA, 3–6 January, 1996*, edited by L. Hunter & T. E. Klein. River Edge, NJ, USA: World Scientific Publishing Co.
- Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* **D55**, 191–205.
- Wilson, A. J. C. (1942). *Nature (London)*, **150**, 151.
- Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S. & Richardson, D. C. (1999). *J. Mol. Biol.* **285**, 1711–1733.
- Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. (1999). *J. Mol. Biol.* **285**, 1735–47.
- Yin, Y., Hu, M., de La Fortelle, E., Ilyin, V. & Carter, C. W. Jr (2001). In preparation.