Comparison of phasing methods for sulfur-SAD using in-house chromium radiation: case studies for standard proteins and 69-kDa protein

Nobuhisa Watanabe,[a*] Yu Kitago[a] and Isao Tanaka,[a]

and

Jia-wei Wang,[b] Yuan-xin Gu,[b] Chao-de Zheng[b] and Hai-fu Fan[b]

[a]Graduate School of Science, Hokkaido University, Sapporo 0600810, Japan, and [b]Institute of Physics, Chinese Academy of Sciences, Beijing 100080, People's Republic of China. E-mail: nobuhisa@sci.hokudai.ac.jp

Synopsis Three phasing methods were compared for data sets collected with a Cr $K\alpha$ X-ray source using loop-less data collection. Phasing with *OASIS-2004* gave the best results for four standard proteins and the 69-kDa protein, TT0570.

Abstract Phasing of the crystal structures of four standard proteins (lysozyme, trypsin, glucose isomerase, and thaumatin) and a novel 69-kDa protein from *Thermus thermophilus*, TT0570, was performed using the single-wavelength anomalous diffraction of sulfur atoms intrinsically present within the native protein molecules. To utilize the sulfur anomalous diffraction, the data sets were collected using the loop-less data collection method with chromium $K\alpha$ X-rays of 2.29Å. Three phasing methods, *MLPHARE*, *SHARP*, and *OASIS-2004*, were tested in combination with the *DM* or *SOLOMON* density modification method. The results showed that the solvent contents are still an important factor for phasing with the S-SAD method, even when longer wavelength Cr $K\alpha$ radiation is used. Among the three procedures, the improved direct phasing of *OASIS-2004* with its implemented fragment feedback to the direct-method probability calculation gave the best results in determining the initial phases. For all five proteins, almost the entire models could be built automatically.

Keywords: single-wavelength anomalous diffraction; sulfur SAD; chromium X-ray radiation

1. Introduction

In the last decade, the multi-wavelength anomalous diffraction (MAD) method (Hendrickson, 1985) in conjunction with selenomethionine derivatization has become a powerful and commonly used tool to solve novel protein structures. In addition to selenium, many other anomalous scatterers have also been incorporated into protein molecules for phasing (Dauter *et al.*, 2000; Nagem *et al.*, 2001; Evans & Bricogne, 2002; Usón *et al.*, 2003). In most cases, MAD phases provide electron density maps of excellent quality immediately after data collection. However, the MAD method requires synchrotron X-rays to measure the intensities at different wavelengths to utilize the

anomalous differences of heavy atoms, such as Se, Pt, Au, Hg, *etc.* On the other hand, single-wavelength anomalous diffraction (SAD) phasing at peak wavelength of such heavy atoms (Rice *et al.*, 2000) is also becoming increasingly common. The SAD method makes more efficient use of beam time and reduces the extent of radiation damage. A new technique that utilizes longer wavelength X-rays for phasing (Stuhrmann *et al.*, 1995; Chayen *et al.*, 2000; Liu *et al.*, 2000; Cianci *et al.*, 2001; Weiss *et al.*, 2001; Micossi *et al.*, 2002; Ramagopal *et al.*, 2003) provided the possibility of using in-house chromium *K*α radiation (2.29 Å), which was first reported by Blow (1958). The choice of X-ray wavelength is one of the major decisions required for SAD data collection. The decision is made following evaluation of several wavelength-dependent factors and data processing methods (Teplyakov *et al.*, 1998; Mueller-Dieckmann, *et al.*, 2004). Use of a wavelength of 2.1 Å was suggested as the best choice to obtain the highest anomalous signal-to-noise ratio using standard means of data collection and processing (Mueller-Dieckmann, *et al.*, 2005). However, changing the wavelength is not possible for in-house X-ray sources. We may only choose from Cu *K*α (1.54 Å), Co *K*α (1.79 Å) and Cr *K*α (2.29 Å) as an anticathode target for longer wavelengths in the laboratory.

Phasing using the anomalous signal of sulfur alone was first reported in structure analysis of crambin (Hendrickson & Teeter, 1981) using Cu *K*α X-rays. Despite this accomplishment, there has been a revival in interest in the use of sulfur anomalous signals only in the past few years. To utilize an anomalous signal from light atoms, such as sulfur, it is advantageous to perform data collection using longer wavelengths of Cr *K*α radiation where the anomalous intensity differences or the Bijvoet ratio becomes 1–2% of total reflection intensity, because the $\Delta f''$ value of sulfur becomes 1.14 e^- as compared to 0.56 e^- at 1.54 Å, as suggested by Wang (1985). Successful applications of the sulfur SAD (S-SAD) technique were reported using a longer wavelength from a chromium target (Chen *et al.*, 2004; Phillips *et al.*, 2004; Rose *et al.*, 2004; Madauss *et al.*, 2004) with an X-ray apparatus optimized for protein crystallography (Yang *et al.*, 2003).

Recently, the use of selenomethionine-derivatized protein in combination with Cr *K*α radiation has been proposed as an efficient routine pathway for in-house high-throughput protein crystallography (Xu *et al.*, 2005). However, using sulfur atoms as anomalous scatterers has the advantage that the sulfur atoms are naturally present in proteins as methionine or cysteine residues, and thus neither modification nor heavy atom soaking is necessary for structure analysis. The average frequency of sulfur atoms in 24 bacterial proteomes is 3.3 sulfurs per 100 amino acids (Jáuregui *et al.*, 2000; Micossi *et al.*, 2002), and genomic-scale analysis shows that many gene products of bacterial genomes will have a Bijvoet difference higher than 1.5% with Cr *K*α radiation (Nagem *et al.*, 2005). Higher sulfur contents are expected in eukaryotic genomes (Micossi *et al.*, 2002).

In the case of light atoms, such as sulfur, the Bijvoet difference is still very small even when a longer wavelength is used. Therefore, highly accurate data collection and powerful methods for determining initial phases from the diffraction data are essential. One of the experimental difficulties in using longer wavelengths is the increased absorption. Especially, in standard protein crystallography where the crystal is mounted in a cryoloop with cryo-buffer, X-ray absorption by these materials sometimes prevents the detection of tiny anomalous signals. We have developed a novel technique for mounting a protein crystal to eliminate absorption by the cryo-buffer and cryoloop,

and have demonstrated its superiority to the standard cryoloop method (Kitago *et al.*, 2005). This technique increases the precision of the anomalous differences between the Bijvoet pair, and makes the in-house S-SAD method with a Cr $K\alpha$ X-ray source a very useful tool for high-throughput structure determination.

However, the solvent fraction is an important factor in S-SAD phasing, as mentioned by Ramagopal *et al.*, (2003). In this study, several proteins with different Bijvoet differences and solvent contents were chosen as test proteins, and three phasing methods were compared—*OASIS-2004* (Wang *et al.*, 2004a,b), *SHARP* (La Fortelle & Bricogne, 1997), and *MLPHARE* (Otwinowski, 1991)—in combination with density improvement by *DM* (Cowtan, 1994) or *SOLOMON* (Abrahams & Leslie, 1996) and auto-model building by *RESOLVE* (BUILD only) (Terwilliger, 2003a,b) and/or *ARP/wARP* (Perrakis *et al.*, 1999).

2. Methods

2.1. Sample preparation, diffraction data collection, and processing

Hen egg-white lysozyme (Seikagaku Kogyo, 100940), bovine pancreas trypsin (Sigma, T7309), *Thaumatinococcus daniellii* thaumation (Sigma, T7638), and *Streptomyces rubiginosus* glucose isomerase (Hampton Research, HR7-100) were purchased and used without further purification. Assuming Cr $K\alpha$ X-ray wavelength, the overall Bijvoet difference or $\langle|\Delta F|\rangle/\langle F\rangle$ values calculated from the amino acid sequence and all other possible anomalous scatterers in the crystals were 5.1%, 4.2%, 2.5%, and 1.2%, respectively, according to Hendrickson & Teeter (1981). A novel protein from *Thermus thermophilus*, TT0570, was also used in this experiment (manuscript in preparation). TT0570 consists of 603 amino acid residues with a molecular weight of 69 kDa, and the estimated overall $\langle|\Delta F|\rangle/\langle F\rangle$ of TT0570 was 1.1% at Cr $K\alpha$ wavelength. Crystallization experiments were all performed at 293 K using the hanging-drop vapor diffusion method. The crystallization conditions for these proteins and the data concerning the unit cell contents are summarized in Table 1.

The crystals were flash-cooled under a stream of nitrogen gas at 93 K after they had been transferred into a cryoprotectant solution as described in Table 1, with the exception of trypsin that can be frozen directly. In the flash-cooling process, the crystal was mounted using the novel crystal-mounting device for eliminating X-ray absorption of buffers and cryoloops as described in Kitago *et al.* (2005). All diffraction data sets were collected using an in-house chromium X-ray source. X-rays were generated with a chromium target on a Rigaku FR-E SuperBright, which has a Cu/Cr dual target (40 kV, 40 mA for Cr), and focused through an Osmic Confocal MaxFlux optics optimized for chromium (Cr CMF). Single-pass ϕ-axis scan oscillation images were recorded on a Rigaku R-AXIS VII imaging plate detector modified for using longer wavelengths (Yang *et al.*, 2003; Kitago *et al.*, 2005). A 0.5-mm collimator was used to keep the whole crystal always bathed in the X-ray beam during exposure. To improve the statistics of the data sets, a total of 720 images of 1.0° oscillation were collected for lysozyme and glucose isomerase, 1440 images of 0.5° for thaumatin and TT0570. In the case of trypsin, for which a relatively high Bijvoet ratio is expected by Ca^{2+} and Zn^{2+} as shown in Table 1, only 360 images of 1.0° were collected. The exposure time of each image was 1.0 min

and the crystal-to-detector distance was 80 mm. At this distance, the resolution limit at the edge of the rectangular imaging plate detector was 2.17 Å. All data were indexed, integrated, and scaled with *HKL2000* (Otwinowski & Minor, 1997). Statistics for the diffraction data are summarized in Table 2.

2.2. Phasing and model building

For each test data set, three phasing methods, *MLPHARE* (Otwinowski, 1991), *OASIS-2004* (Wang *et al.*, 2004a), and *SHARP* (de La Fortelle & Bricogne, 1997), in combination with density modification were tested. For all five proteins examined, the anomalous scattering substructures were solved by *SHELXD* (Sheldrick, *et al.*, 2001) and refined by *SOLVE* (Terwilliger, 1999), and the refined positions were input into *MLPHARE*, *OASIS-2004*, and *SHARP*. Only major anomalous scatterer sites from *SHELXD* solution were used without confirming the sites with known structures. For *MLPHARE* and *OASIS-2004*, density modification was performed by *DM* (Cowtan, 1994), and *SOLOMON* (Abrahams & Leslie, 1996) was used in the case of *SHARP*. Auto-model building and refinement by *ARP/wARP* (Perrakis *et al.*, 1999) and *REFMAC* (Murshudov *et al.*, 1997) were used to compare the resulting electron density maps by the three phasing methods with the default mode of CCP4 (Collaborative Computational Project, Number 4, 1994) or *SHARP*, respectively. In the case of *OASIS-2004, RESOLVE* (Terwilliger, 2003a,b) was also used with the BUILD_ONLY option to find structure fragments during the early stages of iteration, because *OASIS-2004* accepts feedback information of structure fragments found from model building and the calculation is implemented in iterative mode (Wang *et al.*, 2004b). *OASIS-2004* is an extensive revised version of the CCP4 supported program *OASIS* (Hao *et al.*, 2000). The latter is dedicated for *ab initio* phasing of SAD/SIR data and is based on the principle proposed by Fan & Gu (1985) and the implementation of Fan *et al.* (1990). In the case of TT0570, combination of *SHARP*, *DM*, and *ARP/wARP* was also tested with NCS found by *SOLVE*, because *SOLOMON* failed to find NCS for TT0570.

3. Results and discussion

3.1. Diffraction data and anomalous signals

For the five test proteins, the experimental $<|\Delta F|>/<F>$ is plotted as a function of resolution in Fig. 1. In the case of lysozyme and trypsin, the experimental values were lower than expected as shown in Table 1, and gave anomalous signals at the same level as thaumatin. This may have been because the expected values listed in Table 1 were calculated using all anomalous scatterers in the crystal with full occupancy. The values for lysozyme and trypsin calculated with only protein sulfur atoms are 2.4% and 2.2%, respectively.

As a typical behavior mentioned by Dauter *et al.* (2002), the experimental estimate of the anomalous signal became significantly higher at high resolution range, because of the reduced accuracy of estimation of the reflection intensities. In the case of TT0570, which has relatively low $<I>/<\sigma I>$ at higher resolution (Table 2), the $<|\Delta F|>/<F>$

plot also indicates that the quality of the dataset was poorer than the others, because the experimental $<|\Delta F|>/<F>$ values of TT0570 increased earlier and became higher with resolution as compared with the other four data sets.

3.2. SAD Phasing and automated model building

The numbers of anomalous scatterer sites found by *SHELXD*, refined by *SOLVE*, and used for phasing are also shown in Table 1. With use of Cr $K\alpha$ radiation and our crystal mounting method, there was no difficulty in solving anomalous scatterer substructures with *SHELXD* for all five proteins. The automatic model-building by *ARP/wARP* using the *MLPHARE* phases after density improvement by *DM* gave 0 of 129 residues for lysozyme, 36 of 223 for trypsin, 203 of 207 for thaumatin, 346 of 388 for glucose isomerase, and 87 of 1206 for TT0570. As mentioned by Ramagopal *et al.* (2003), the solvent fraction is an important factor in S-SAD phasing. The relationships among solvent content, phase error after *DM*, and model building results using *ARP/wARP* were compiled in Table 3. With the exception of TT0570, the density-modified SAD phases were improved for the crystals with high solvent fraction, and *ARP/wARP* could build almost all residues with a side chain. In the case of thaumatin and glucose isomerase, almost all residues were built with side chains: 203 of 207 residues and 312 of 388 residues, respectively. Conversely, in the case of lysozyme and trypsin, which have low solvent contents, model building using *ARP/wARP* was unsuccessful, even though the experimental anomalous signal of $<|\Delta F|>/<F>$ seemed sufficient as shown in Fig. 1.

Detailed comparisons of the results of three phasing methods for each test protein are shown in Tables 4 to 8. In all cases, *SHARP-SOLOMON* and *OASIS-2004-DM* gave better results than *MLPHARE-DM*. A clear difference between *SHARP* and *OASIS-2004* was seen in the two difficult cases of lysozyme and TT0570. In both cases, *OASIS-2004* gave better results than *SHARP*, even if the NCS information was used for density improvement with *DM* (Table 8). As shown in Tables 4 to 8, in the case of proteins with sufficient sulfur contents, almost all of the structure could be constructed in a straightforward manner by *OASIS-2004*: 126 of 129 residues (97.7%) for lysozyme, 220 of 223 (98.7%) for trypsin, 203 of 207 (98.1%) for thaumatin, 381 of 388 (98.2%) for glucose isomerase, and 1173 of 1206 (97.3%) for TT0570. The superiority of *OASIS-2004* arises from its implemented fragment feedback in the direct-method probability calculation (Wang *et al.*, 2004b). No manual intervention was needed to tune controlling parameters during either the phase derivation or automatic model building. As an example, ribbon models built automatically at different cycles of iteration for lysozyme are shown in Fig. 2. The numbers of residues found in each cycle of iteration of *OASIS-2004* are shown in Table 9.

4. Conclusions

The results of the present study show that the solvent contents are still an important factor for phasing with the S-SAD method, even when longer wavelength Cr $K\alpha$ radiation is used. Among the three phasing procedures, the

improved direct phasing by *OASIS-2004* with its implemented fragment feedback in the direct-method probability calculation gave the best results in determining the initial phases.

In the present study, we applied this Cr S-SAD method to a larger target, TT0570, which has a molecular weight of 69 kDa and is crystallized in the space group of orthorhombic $P2_12_12$. The crystal contains two independent molecules in an asymmetric unit, which corresponds to 1206 residues per asymmetric unit. Furthermore, it has a relatively low sulfur content, with 9 Met and 2 Cys residues per molecule, and a calculated Bijvoet ratio corresponding to 1.1% even with use of Cr $K\alpha$ radiation (or 0.57% for Cu $K\alpha$ radiation). Reflecting these characteristics, the data statistics of TT0570 were much poorer than those for the other standard proteins examined here (Table 2). The overall signal to noise ratio of the anomalous differences $<|\Delta F|>/<\sigma(\Delta F)>$ for TT0570 was only 0.80. Nonetheless, the structure could be solved by the S-SAD method using *OASIS-2004*. Although TT0570 has a solvent content of 48% and the data had poor statistics as compared with the other test proteins, the electron density map was sufficiently clear to allow protein models to be built. As shown in Table 8, 1173 of 1206 residues (97.3%) were constructed in a straightforward manner.

High-throughput crystallography in the post-genomic era requires a method by which the native protein structure can be solved both quickly and easily. Several phasing techniques exploiting anomalous signals in the data collected with in-house generators have emerged in recent years. The MAD method using Se-Met substitution is no longer the only solution. The present results showed that the combination of our loop-less data collection method using Cr Kα radiation and the *OASIS-2004* phasing method with density modification is a powerful and routine tool to solve novel structures with the S-SAD method.

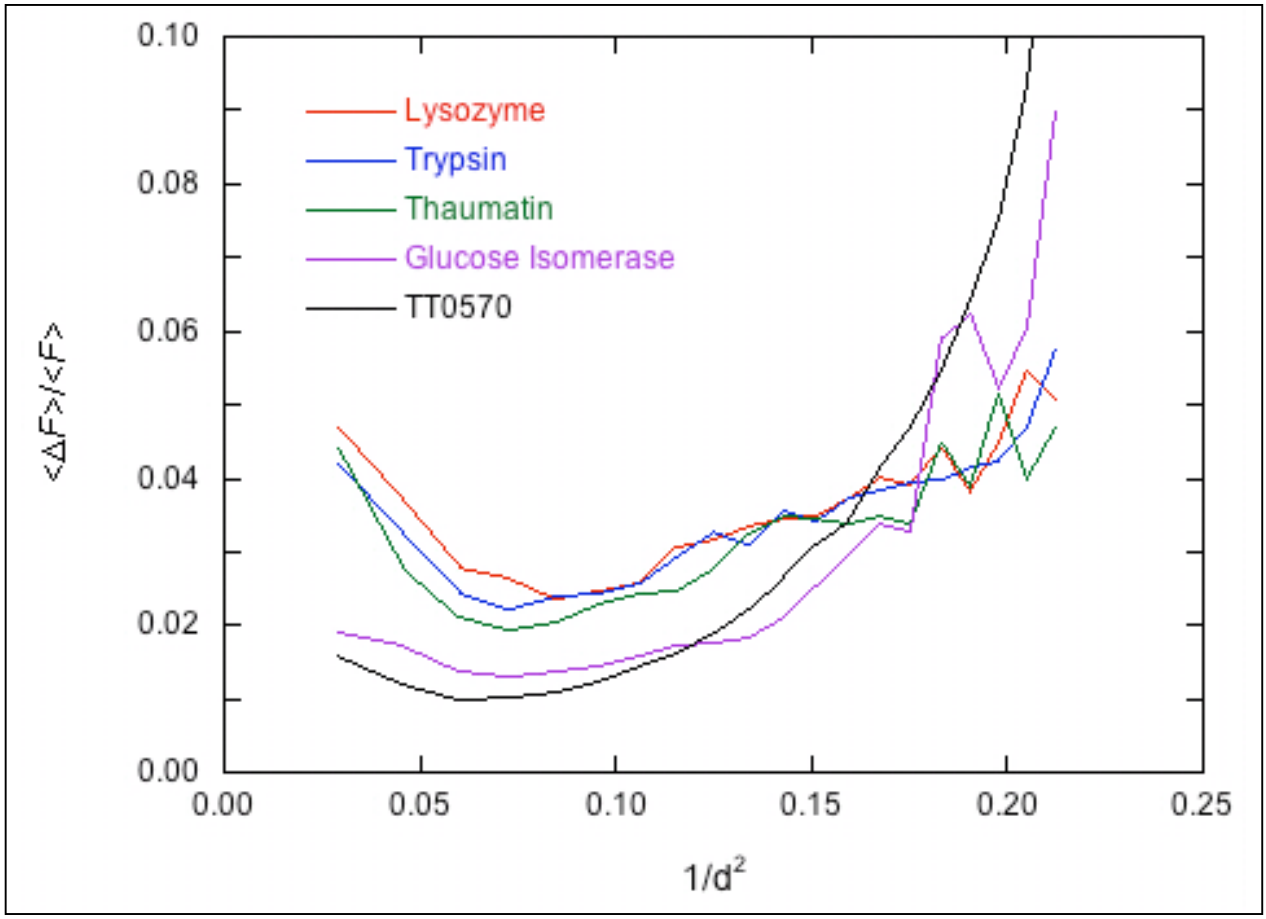

Figure 1 Amounts of anomalous signal in the diffraction data. The Bijvoet ratio of $<|\Delta F|>/<F>$ is plotted as a function of resolution for the five test proteins.

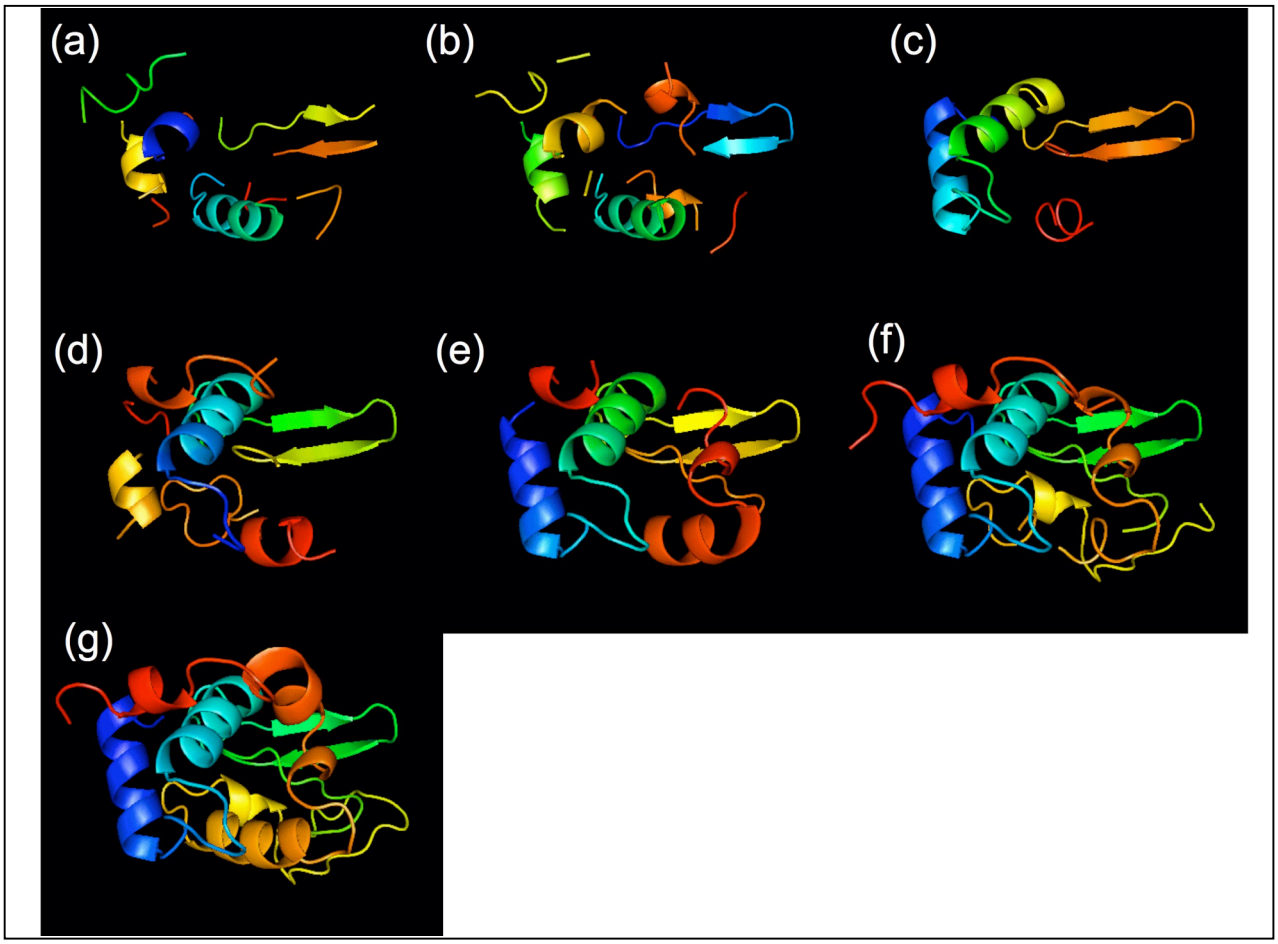

Figure 2 Lysozyme ribbon models built automatically from different cycles of iteration. (a) 0-th cycle by *RESOLVE BUILD*; (b) 1-st cycle by *RESOLVE BUILD*; (c) 2-nd cycle by *ARP/wARP*; (d) 3-rd cycle by *ARP/wARP*; (e) 4-th cycle by *ARP/wARP*; (f) 5-th cycle by *ARP/wARP*; (g) 6-th cycle by *ARP/wARP*. All ribbon models were plotted using *PyMOL* (DeLano, 2002).

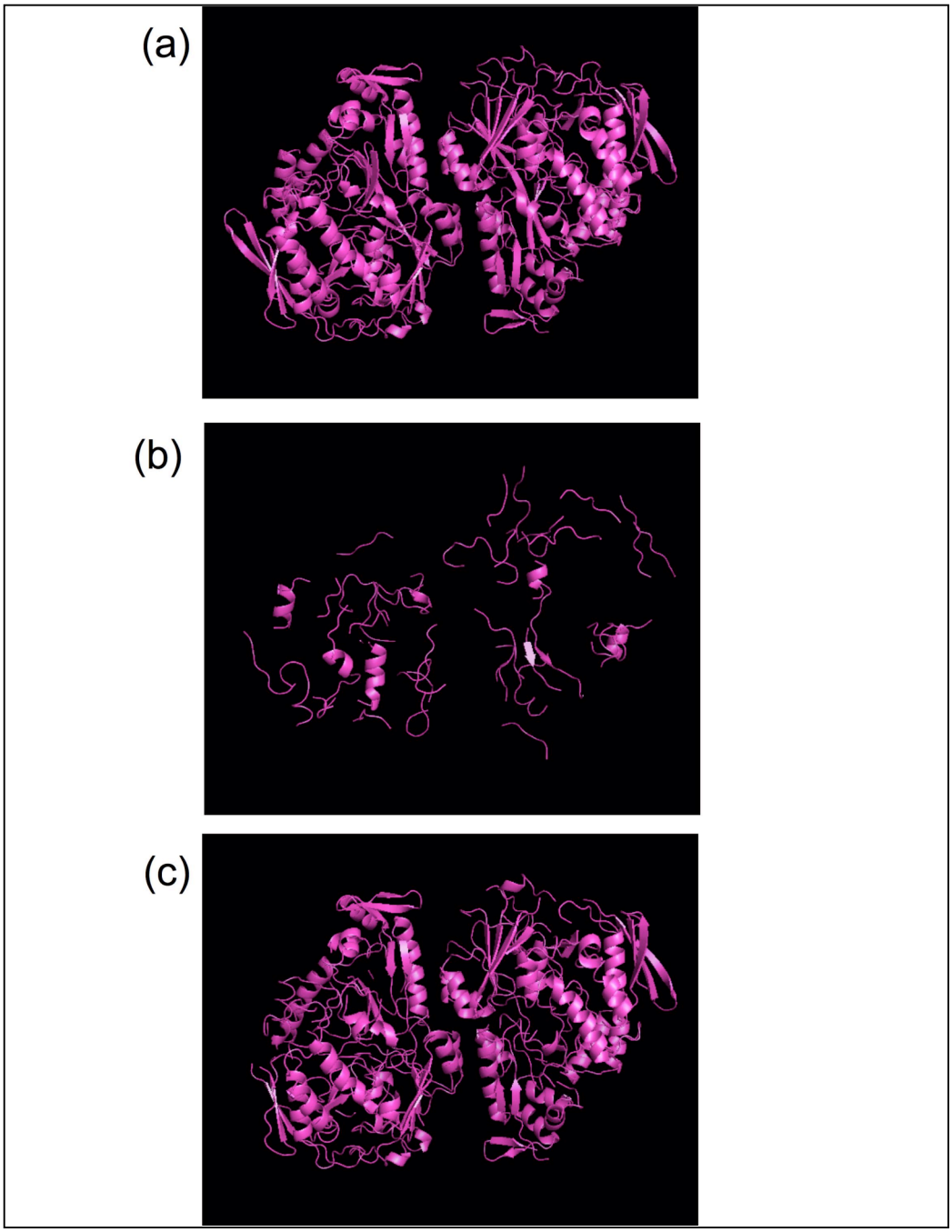

Figure 3 TT0570 ribbon models built automatically using different phasing methods. (a) *OASIS-2004* with *DM*; (b) *SHARP* with *SOLOMON*; (c) *SHARP* with *DM* using NCS information. All ribbon models were plotted using *PyMOL* (DeLano, 2002).

Table 1 Crystallization conditions and contents of the unit cell for the sample proteins.

	Crystallization condition (Cryoprotectant)	Residues per asymmetric unit	Anomalous scatterers per molecule	Estimated $\langle\lvert\Delta F\rvert\rangle/\langle F\rangle$ (%)	Solvent content (%)	Anomalous scatterer sites used for phasing
Lysozyme	60 mg/ml protein 0.1 M sodium acetate pH 4.2 2.0 M NaCl (30% glycerol for cryo)	129	10 S (4 SS*) 8 Cl^-	5.1	38.2	16
Trypsin	60 mg/ml protein 0.1 M Tris-HCl pH 7.8 0.2 M $LiSO_4$ 25% PEG4000 16% Glycerol	223	14 S (6 SS) 1 Ca^{2+} 1 Zn^{2+}	4.2	37.9	15
Thaumatin	30 mg/ml protein 0.1 M Na_2ADA / NaADA pH 6.5 0.8 M K,Na tartrate (30% glycerol for cryo)	207	17 S (8 SS)	2.5	55.7	16
Glucose isomerase	15 mg/ml protein 0.05 M Tris-HCl pH 7.0 11% MPD 0.1 M $MgCl_2$ (30% MPD for cryo)	388	9 S	1.2	53.7	8
TT0570	10 mg/ml protein 0.2 M $Mg(NO_3)_2$ 20% PEG3350 (15% PEG400 for cryo)	1206 (603 × 2)	11 S	1.1	48.0	20

* SS: Super sulfur or disulfide.

Table 2 Diffraction data statistics for the sample data.

	Lysozyme	Trypsin	Thaumatin	Glucose isomerase	TT0570
Unit-cell parameters (Å)	a=78.4, c=37.0	a=54.6, c=107.3	a=57.9, c=150.3	a=92.8, b=98.1, c=102.6	a=100.3, b=109.0, c=114.6
Space group	$P4_32_12$	$P3_121$	$P4_12_12$	$I222$	$P2_12_12$
Resolution limit	50.0 - 2.17 (2.25 - 2.17)				
Reflections measured	301531	196655	704259	675807	1732599
Oscillation angle (°)	1.0	1.0	0.5	1.0	0.5
Total rotation range (°)	720	360	720	720	720
Unique reflections	6383	10388	14202	47046	66273
Completeness (%)	98.3 (85.8)	100.0 (99.9)	99.4 (95.2)	98.4 (93.0)	98.6 (87.7)
Mosaicity (°)	0.88	0.55	0.44	0.46	0.75
R_{sym} † (%)	4.0 (12.3)	5.6 (12.6)	6.4 (13.7)	5.4 (16.8)	8.0 (29.2)
$R_{p.i.m.}$ ‡ (%)	0.6 (2.1)	1.3 (3.4)	0.9 (2.3)	1.1 (3.6)	1.5 (7.8)
$\langle\Delta F\rangle/\langle F\rangle$ (%)	3.2 (5.3)	3.0 (5.2)	2.8 (4.3)	2.5 (7.5)	1.9 (10.7)
Multiplicity §	47.2 (36.0)	18.9 (14.1)	49.6 (38.8)	27.4 (21.5)	26.1 (17.3)
$\langle I\rangle/\langle\sigma(I)\rangle$	104.3 (35.0)	60.3 (26.0)	107.4 (59.0)	36.0 (9.2)	44.4 (7.3)

† $R_{sym} = \sum|I-\langle I\rangle|/\sum I$. ‡ $R_{p.i.m.}$ is the precision-indicating merging R factor (Weiss & Hilgenfeld, 1997). § Counts the Friedel mates as the same reflection.

Table 3 Solvent contents and *MLPHARE-DM* results.

	Lysozyme	Trypsin	Thaumatin	Glucose isomerase	TT0570
Solvent content (%)	38.2	37.9	55.7	53.7	48.0
Phase error (°)	56.5	54.5	50.1	50.0	69.7
Residues found by *ARP/wARP* (%)	0	16	98	89	7
Estimated $\langle\Delta F\rangle/\langle F\rangle$ (%)	5.1	4.2	2.5	1.2	1.1

Table 4 Comparison of three phasing methods for lysozyme.

		OASIS-2004 + DM + ARP/wARP with fragment feedback (default mode)	SHARP + SOLOMON + ARP/wARP (default mode)	MLPHARE + DM + ARP/wARP
Phase error (for top 6,000 reflections)		42.5°	54.3°	56.5°
Map correlation coefficient	main chain	0.76	0.60	0.63
	side chain	0.65	0.52	0.49
Residues found in automatic model building	total number	126	14	0
	number with side chains	126	14	0
Number of residues in ASU		129		

Table 5 Comparison of three phasing methods for Trypsin.

		OASIS-2004 + DM + ARP/wARP with fragment feedback (default mode)	SHARP + SOLOMON + ARP/wARP (default mode)	MLPHARE + DM + ARP/wARP
Phase error (for top 10,000 reflections)		37.9°	49.0°	54.5°
Map correlation coefficient	main chain	0.78	0.65	0.61
	side chain	0.72	0.62	0.56
Residues found in automatic model building	total number	220	221	36
	number with side chains	220	221	7
Number of residues in ASU		223		

Table 6 Comparison of three phasing methods for Thaumatin.

		OASIS-2004 + DM + ARP/wARP with fragment feedback (default mode)	SHARP + SOLOMON + ARP/wARP (default mode)	MLPHARE + DM + ARP/wARP
Phase error (for top 14,000 reflections)		30.3°	28.4°	50.1°
Map correlation coefficient	main chain	0.77	0.77	0.63
	side chain	0.76	0.77	0.61
Residues found in automatic model building	total number	203	201	203
	number with side chains	203	201	203
Number of residues in ASU		207		

Table 7 Comparison of three phasing methods for Glucose isomerase.

		OASIS-2004 + DM + ARP/wARP with fragment feedback (default mode)	SHARP + SOLOMON + ARP/wARP (default mode)	MLPHARE + DM + ARP/wARP
Phase error (for top 24,000 reflections)		27.9°	29.4°	50.0°
Map correlation coefficient	main chain	0.83	0.83	0.69
	side chain	0.76	0.77	0.60
Residues found in automatic model building	total number	381	383	346
	number with side chains	381	383	312
Number of residues in ASU		388		

Table 8 Comparison of three phasing methods for TT0570.

		OASIS-2004 + DM + ARP/wARP with fragment feedback (default mode)	SHARP + DM + ARP/wARP with NCS information (manual intervention)	SHARP + SOLOMON + ARP/wARP (default mode)	MLPHARE + DM + ARP/wARP
Phase error (for top 66,000 reflections)		49.7°	56.7°	60.7°	69.7°
Map correlation coefficient	main chain	0.70	0.64	0.58	0.48
	side chain	0.61	0.55	0.47	0.38
Residues found in automatic model building	total number	1173	1117	340	87
	number with side chains	1167	964	17	0
Number of residues in ASU		1206			

Table 9 Number of residues found automatically in each cycle of *OASIS-2004* iteration.

Sample		Lysozyme	Trypsin	Thaumatin	Glucose Isomerase	TT0570
Residues in the ASU		129	223	207	388	1206
Residues found in each cycle	Cycle 0	67 (R)	113(R)	57 (R)	281 (R)	838 (R)
	Cycle 1	78 (R)	109(R)	99 (A)	381 (A)	881 (R)
	Cycle 2	65 (A)	188(A)	191 (A)		1167 (A)
	Cycle 3	79 (A)	213(A)	203 (A)		
	Cycle 4	80 (A)	220(A)			
	Cycle 5	117 (A)				
	Cycle 6	126 (A)				

(A)– Model built by *ARP*/*wARP* with CCP4 default mode. (R)– Model built by *RESOLVE* with BUILD_ONLY mode.

Acknowledgements This study was supported in part by a research grant from the National Project on Protein Structural and Functional Analysis from the Ministry of Education, Culture, Sports, Science, and Technology of Japan and by the *Innovation Project* of the Chinese Academy of Sciences and the *973 Project* (Grant No. 2002CB713801) and the *863 Project* (Grant No. 2002BA711A12) of the Ministry of Science and Technology of China.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst*. **D52**, 30-42.

Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. Academic Press Inc. London. P. 177.

Blow, D. M. (1958). *Proc. R. Soc. London Ser. A*, **247**, 302-336.

Chayen, N. E., Cianci, M., Olczak, A., Raftery, J., Rizkallah, P. J., Zagalsky, P. F. & Helliwell, J. R. (2000). *Acta Cryst*. D**56**, 1064-1066.

Chen, L., Chen, Li., Zhou, X. E., Wang, Y., Kahsai, M. A., Clark, A. T., Edmondson, S. P., Liu, Z., Rose, J. P., Wang, B.-C., Meehan, E. J. & Shriver, J.W. (2004). *J.Mol. Biol.* **341,** 73-91.

Cianci, M., Rizkallah, P. J., Olczak, A., Raftery, J., Chayen, N. E., Zagalsky, P. F. & Helliwell, J. R. (2001). Acta Cryst. D57, 1219-1229.

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

Cowtan, K. (1994). *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **31**, 34-38.

Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289,** 83-92.

Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst*. D56, 232–237.

Dauter, Z., Dauter, M. & Dodson, E. J. (2002). *Acta Cryst*. D58, 494–506.

DeLano, W.L. The PyMOL Molecular Graphics System (2002) DeLano Scientific, San Carlos, CA, USA.

Evans, G. & Bricogne, G. (2002). *Acta Cryst*. D**58**, 976-991.

Fan, H. F. & Gu, Y. X. (1985). *Acta Cryst*. A**41**, 280-284.

Fan, H. F., Hao, Q., Gu, Y. X., Qian, J. Z., Zheng, C. D. & Ke, H. (1990). *Acta Cryst*. A**46,** 935–939.

La Fortelle E de & Bricogne, G. (1997) *Methods Enzymol*. **276**, 472-494.

Liu, Z. J., Vysotski, E. S., Chen, C. J., Rose, J. P., Lee, J. & Wang, B. C. (2000). Protein Sci. 9, 2085-2093.

Hao, Q., Gu, Y.X., Zheng, C.D. & Fan, H.F. (2000). *J. Appl. Cryst.* **33**, 980-981.

Hendrickson, W. A. (1985). *Trans. Am. Crystallogr. Assoc*. **21**, 11-21.

Hendrickson, W.A. & Teeter, M. M. (1981). *Nature (London),* **290,** 107-113.

Jáuregui, R., Bolivar, F. & Merino, E. (2000). *Microb. Comput. Genomics*, **5**, 7-15.

Kitago, Y., Watanabe, N. & Tanaka, I. (2005). *Acta Cryst.* D**61**, 1013-1021.

Madauss, K., Juzumiene, D., Waitt, G., Williams, J. & Williams, S. (2004). Endocrine Res. 30, 775-785.

Micossi, E., Hunter, W. N. & Leonard, G. A. (2002). *Acta Cryst.* D**58**, 21-28.

Mueller-Dieckmann, C., Polentarutti, M., Carugo, K. D., Panjikar, S., Tucker, P. A. & Weiss, M. (2004). *Acta Cryst.* D**60**, 28-38.

Mueller-Dieckmann, C., Panjikar, S., Tucker, P. A. & Weiss, M. (2005). *Acta Cryst.* D**61**, 1263-1272.

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

Nagem, R. A., Dauter, Z. & Polikarpov, I. (2001). Acta Cryst. D57, 996-1002.

Nagem, R. A., Ambrosio, A. L., Rojas, A. L., Navarro, M. V., Golubev, A. M., Garratt, R. C. & Polikarpov, I. (2005). *Acta Cryst*. D**61**, 1022-1030.

Otwinowski, Z. (1991) in Proceedings of the CCP4 Study Weekend, W. Wolf, P. R. Evans and A. G. W. Leslie, Eds. (SERC Daresbury, Daresbury, UK), pp. 80-86.

Otwinowsi, Z. & Minor, W. (1997). *Methods Enzymol*. **276,** 307-326.

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

Phillips, J. D., Whitby, F. G., Warby, C. A., Labbe, P., Yang, C., Pflugrath, J. W., Ferrara, J. D., Robinson, H., Kushner, J. P. & Hill, C. P. (2004). J. Biol. Chem. 279, 38960-38968.

Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst*. D**59**, 1020-1027.

Rice, L. M., Earnest, T. N. & Brunger, A. T. (2000). Acta Cryst. D56, 1413-1420.

Rose, J. P., Liu, Z. J., Tempel, W., Chen, D., Lee, D., Newton, M. G. & Wang, B. -C. (2004). *Rigaku J*. **21**, 1-9.

Sheldrick, G. M., Hauptman, H. A., Weeks, C. M., Miler, M. & Usón, I. (2001). International Tables for Crystallography, Vol. F, edited by E. Arnold & M. G. Rossmann, pp. 333-351. Dordrecht: Kluwer Academic Publishers.

Sim, G. A. (1959). *Acta Cryst*. **12**, 813-815.

Stuhrmann, S., Hütsch, M., Trame, C., Thomas, J. & Stuhrmann, H. B. (1995). *J. Synchrotron Rad*. **2**. 83-86.

Teplyakov, A., Oliva, G. & Polikarpov, I. (1998). *Acta Cryst*. **D54**,610-614.

Terwilliger, T. C. (1999). *Acta Cryst*. D**55**,1863-1871.

Terwilliger, T. C. (2003*a*). *Acta Cryst*. D**59**, 38-44.

Terwilliger, T. C. (2003*b*). *Acta Cryst*. D**59**, 45-49.

Usón, I., Schmidt, B., von Bulow, R., Grimme, S., von Figura, K., Dauter, M., Rajashankar, K. R., Dauter, Z. & Sheldrick, G. M. (2003). *Acta Cryst*. D**59**, 57-66.

Wang, B.-C. (1985). *Methods Enzymol.* **115,** 90-112.

Wang, J. W., Chen, J. R., Gu, Y. X., Zheng, C. D., Jiang, F. & Fan, H. F. (2004a). *Acta Cryst.* D**60**, 1987-1990.

Wang, J. W., Chen, J. R., Gu, Y. X., Zheng, C. D. & Fan, H. F. (2004b). *Acta Cryst.* D**60**, 1991-1996.

Weiss, M. S. & Hilgenfeld, R. (1997). *J. Appl. Cryst*. **30**, 203-205.

Weiss, M. S., Sicker, T., Djinovic-Carugo, K. & Hilgenfeld, R. (2001). *Acta Cryst.* D**57**, 689-695.

Xu, H., Yang, C., Chen, L., Kataeva, I. A., Tempel, W., Lee, D., Habel, J. E., Nguyen, D., Pflugrath, J. W., Ferrara, J. D., Arendall, W. B., 3rd, Richardson, J. S., Richardson, D. C., Liu, Z. J., Newton, M. G., Rose, J. P. & Wang, B. C. (2005). *Acta Cryst.* D**61**, 960-966.

Yang, C., Pflugrath, J. W., Courville, D. A., Stence, C. N. & Ferrara, J. D. (2003). *Acta Cryst.* **D59,** 1943-1957.