Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Use of complementary cation and anion heavy-atom salt derivatives to solve the structure of cytochrome P450 46A1

Mark Andrew White,[a,b]* Natalia Mast,[c] Ingemar Bjorkhem,[d] Eric F. Johnson,[e] C. David Stout[f] and Irina A. Pikuleva[c]

[a]Sealy Center for Structural and Molecular Biophysics, UTMB Galveston, TX 77555, USA, [b]Department of Biochemistry and Molecular Biology, UTMB Galveston, TX 77555, USA, [c]Department of Pharmacology and Toxicology, UTMB Galveston, TX 77555, USA, [d]Department of Clinical Chemistry, Karolinska Institute, S-141 88 Huddinge, Sweden, [e]Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA, and [f]Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

Correspondence e-mail: white@xray.utmb.edu

Received 29 November 2007
Accepted 7 February 2008

Human cytochrome P450 46A1 (CYP46A1) is one of the key enzymes in cholesterol homeostasis in the brain. The crystallization and heavy-atom structure solution of an active truncated CYP46A1 in complex with the high-affinity substrate analogue cholesterol-3-sulfate (CH-3S) is reported. The 2.6 Å structure of CYP46A1–CH-3S was solved using both anion and cation heavy-atom salts. In addition to the native anomalous signal from the haem iron, an NaI anion halide salt derivative and a complementary CsCl alkali-metal cation salt derivative were used. The general implications of the use of halide and alkali-metal quick soaks are discussed. The importance of using isoionic strength buffers, the titration of heavy-atom salts into different ionic species and the role of concentration are considered. It was observed that cation/anion-binding sites will occasionally overlap, which could negatively impact upon mixed RbBr soaks used for multiple anomalous scatterer MAD (MMAD). The use of complementary cation and anion heavy-atom salt derivatives is a convenient and powerful tool for MIR(AS) structure solution.

1. Introduction

Cytochrome P450s (P450 or CYP) constitute a huge superfamily of more than 6500 enzymes that are found in almost all organisms (http://drnelson/utmem.edu/CytochromeP450.html). These proteins play key roles in the biotransformation of various xenobiotics and endobiotics and catalyze several types of reactions, with the most common being monooxygenation (Guengerich, 2001). There are 57 P450 genes in humans, which belong to 18 families and 43 subfamilies[1]. 15 human P450s primarily catalyze the transformation of various xenobiotics, such as drugs, carcinogens and environmental pollutants. 14 are involved in cholesterol, bile acid or steroid synthesis, four are primarily involved in the metabolism of vitamins A and D and nine are involved in the metabolism of eicosanoids and fatty acids. The roles of the remaining 14 have not yet been established (Guengerich, 2005). Currently, the structures of 11 different mammalian P450s have been deposited in the PDB. Of these, one belongs to family 1 (1A2), eight to family 2 (2A6, 2A13, 2B4, 2C5, 2C8, 2C9, 2D6 and 2R1), one to family 3 (3A4) and one to family 8 (8A1). The structurally characterized P450s have a similar overall fold, but have varying topologies of the active site in order to accommodate different

[1] Designation into families and subfamilies is based on amino-acid sequence identity (Nelson *et al.*, 1996). Enzymes that share >40% identity are assigned to the same family, designated by an Arabic numeral followed by a letter indicating the subfamily to which the P450 belongs. Enzymes within subfamilies are also numbered, usually in the order in which they were discovered. Sequence identity within subfamilies is >55%.

substrates (Li & Poulos, 2004; Johnson & Stout, 2005; Graham & Peterson, 1999). Even with the same P450, the shape of the active site can vary significantly with different ligands, the quintessential example being the several structures of CYP2B4 (Zhao *et al.*, 2006, 2007; Scott *et al.*, 2003, 2004). The CYP46 gene family has never been crystallized before and is only distantly related to other structurally characterized P450s. The crystal structure of CYP46A1 is the first of a P450 for which the primary role is cholesterol hydroxylation.

1.1. CYP46A1

CYP46A1 catalyzes the conversion of cholesterol to 24(*S*)-hydroxycholesterol (24OH-CH), the first step in the major pathway for cholesterol elimination from the brain (Lund *et al.*, 1999; Lutjohann *et al.*, 1996). CYP46A1 was also found to be expressed in the retina (Bretillon *et al.*, 2007). However, the physiological role of CYP46A1 in the retina still remains to be established. Evidence has recently been obtained in mice indicating that side-chain oxidized oxysterols, including 24OH-CH, may be important *in vivo* ligands for the liver X receptors, which function as transcription factors for a number of genes that are involved in lipid metabolism (Chen *et al.*, 2007; Chawla *et al.*, 2001). Individuals with a complete lack of cholesterol 24-hydroxylation have not been identified. The phenotype of CYP46A1 gene-knockout mice provides some insight. These animals show severe deficiencies in spatial, associative and motor learning (Kotti *et al.*, 2006), suggesting that CYP46A1 may also be involved in cognitive function in humans. Polymorphisms in the CYP46A1 gene are frequent and some could be associated with Alzheimer's disease. 13 published papers have suggested that CYP46A1 is a risk factor for Alzheimer's disease, while eight papers do not support this association (reviewed in Pikuleva, 2006; Ma *et al.*, 2006; Helisalmi *et al.*, 2006; Li *et al.*, 2006; Tedde *et al.*, 2006; Wang & Jia, 2007). In healthy people, CYP46A1 immunoreactivity is predominantly confined to neurons, whereas in patients with Alzheimer's disease its expression is also detected in astrocytes (Brown *et al.*, 2004; Bogdanovic *et al.*, 2001). The mechanism and significance of this shift in the expression pattern are currently unclear. Cholesterol-3-sulfate (CH-3S), which we used for cocrystallization with CYP46A1, is a naturally occurring steroid that is present in many tissues, including the brain (Liu *et al.*, 2003; Iwamori *et al.*, 1976; Strott & Higashi, 2003). This steroid sulfate is suggested to have many cellular functions, including regulation of lipid metabolism (Strott & Higashi, 2003). Until this study, it was not known that CYP46A1 could bind and metabolize CH-3S *in vitro*. Further research is needed to evaluate whether this *in vitro* activity is of physiological relevance.

1.2. Halide and alkali-metal salt derivatives

The use of halide salts as anionic heavy-atom derivatives has been well documented by Dauter and coworkers (Dauter *et al.*, 2000; Nagem *et al.*, 2003; Dauter & Dauter, 2006). However, the use of halide-salt quick soaks does not seem to have become as widespread as one might expect given the simplicity of the method. The Rb salt used by Korolev and coworkers in their MAD solution of the hsp60 apical domain (Korolev *et al.*, 2001) indicated that a heavy-atom cation alkali-metal salt could be used as a general and nonspecific heavy-atom derivative in the same way that halide salts have been used. The efficacy of cation salts for quick-soak derivatives has previously been explored using lysozyme test crystals (Nagem *et al.*, 2001). The use of exceptionally well diffracting test crystals such as lysozyme in the published test cases for halide and cation quick-soak derivatization and the use of high concentrations (1 *M*) of heavy-atom salts still leaves an important question unanswered: is this a general technique which can be used with samples of varying diffraction quality? Here, we explore a practical example of the use of complementary cation and anion heavy-atom quick soaks of moderate ionic strength (300 m*M*) on a protein crystal of average diffraction quality, with the MIRAS structure solution of CYP46A1 at 2.6 Å resolution.

2. Experimental procedures

2.1. Cloning and expression

The Δ(2–50)CYP46A1H4 truncated variant was generated from the Δ(2–23)CYP46A1H4 expression construct (Mast *et al.*, 2004) using mutagenic primers and the QuikChange site-directed mutagenesis kit. The forward 5′-primer contained the initiator Met codon, Ala as the second codon and a sequence complementary to codons 51–57 (KKDEVGG) of wild-type CYP46A1. The reverse 3′-primer was complementary to the forward 5′-primer. The C-terminal 4×His tag was incorporated unmodified from the template sequence. The construct was transformed into GC5 cells (PGC Scientifics Corp.) in the presence of the chaperones GroEL and GroES (Sansen *et al.*, 2007) and expressed for 48 h at 299 K with shaking at 200 rev min^{-1}. The expression of P450 and chaperones was induced with 1 m*M* isopropyl β-D-thiogalactopyranoside and 0.18 g ml^{-1} L-arabinose, respectively, 5 h after the overnight culture had been diluted 100-fold with Terrific Broth medium. Simultaneously, 0.5 m*M* α-aminolevulinic acid was added to maximize P450 expression.

2.2. Purification of Δ(2–50)CYP46A1H4

CYP46A1 was purified as described previously (Mast *et al.*, 2004) with several modifications. Briefly, the harvested *Escherichia coli* cells were suspended in 100 m*M* potassium phosphate buffer (KP_i) pH 7.2 containing 0.2 *M* KCl, 10 m*M* β-mercaptoethanol (BME), 10 µ*M* CH-3S and 20% glycerol. The cell suspension was sonicated on ice using six 20 s pulses at 1 min intervals and subjected to centrifugation at 106 000*g* for 1 h. The supernatant was diluted 1/3-fold with 10 m*M* KP_i and 20% glycerol and applied onto a DEAE-cellulose column equilibrated with 75 m*M* KP_i pH 7.2 containing 135 m*M* KCl, 10 m*M* BME, 10 µ*M* CH-3S and 20% glycerol. The flow-through fraction was collected and applied onto Ni-agarose resin equilibrated with 155 m*M* KP_i pH 7.2 containing 325 m*M* KCl, 0.5% CHAPS, 10 m*M* BME, 10 µ*M* CH-3S, 5 m*M*

Table 1
Crystallographic data and MIRAS statistics.

Values in parentheses are for the highest resolution shell.

	Native (Fe)	NaI	CsCl
Detector	DIP2030	SMART 2K	DIP2030
No. of frames	194 [×2]	182	101
Frame width (°)	0.5	0.25	0.5
Exposure (min)	35	10	35
Net rotation (°)	97	45.5	50.5
Space group	$I4_122$		
Unit-cell parameters			
a (Å)	121.73	121.81	121.85
c (Å)	144.21	143.56	143.94
Resolution	30–2.62	70–2.80	50–2.62
No. of observations	225798	44106	67008
Unique observations	16668	13419	16908
Redundancy	13.4 (6.8)	3.3 (2.1)	4.0 (3.6)
Completeness	99.7 (98)	98.3 (88)	99.4 (92)
$R_{\rm merge}$	13.2 (55)	5.6 (21)	8.7 (49)
$I/\sigma(I)$	9.5 (3.6)	27.7 (4.2)	10 (2.8)
No. of sites	1	14	5
Phasing power	1.0	0.93	0.77
FOM	0.38 (0.10)		
FOM after DM	0.67 (0.30)		

imidazole and 20% glycerol. The resin was washed with five volumes of equilibration buffer and five volumes of 50 m*M* KP_i pH 7.2 containing 0.5% CHAPS, 10 m*M* BME, 10 µ*M* CH-3S, 5 m*M* imidazole and 20% glycerol. The P450 was eluted with the latter buffer containing 300 m*M* imidazole and then diluted tenfold with 50 m*M* KP_i pH 7.2 containing 0.5% CHAPS, 0.2 m*M* DTT, 1 m*M* EDTA, 10 µ*M* CH-3S and 20% glycerol (the dilution buffer) and applied onto a hydroxyapatite column equilibrated with the same buffer. The column was washed with ten column volumes of the dilution buffer and the P450 was eluted with the same buffer containing 300 m*M* KP_i. The red-colored fractions were pooled, diluted tenfold with the dilution buffer and applied onto CM-cellulose equilibrated with the same buffer. The resin was washed with five volumes of the dilution buffer followed by 50 resin volumes of the same buffer without detergent. The P450 was eluted with 50 m*M* KP_i pH 7.2 containing 400 m*M* NaCl, 0.2 m*M* DTT, 1 m*M* EDTA, 50 µ*M* CH-3S and 20% glycerol. The purified protein was then diluted with 50 m*M* KP_i pH 7.2 containing 0.2 m*M* DTT, 1 m*M* EDTA, 50 µ*M* CH-3S and 20% glycerol to decrease the NaCl concentration to 25 m*M* and concentrated using an Amicon Ultra centrifugal filter.

2.3. Crystallization

The complex of Δ(2–50)CYP46A1H4 with CH-3S was crystallized in Emerald Biosystems clover-leaf sitting-drop trays. The crystals were grown at 291 K by the sitting-drop vapour-diffusion method in which 1 µl of an ∼40 mg ml^{-1} protein solution (as assessed by the reduced CO-difference spectrum; Omura & Sato, 1964) in 50 m*M* KP_i pH 7.2, 20% glycerol, 25 m*M* NaCl, 1 m*M* EDTA, 0.2 m*M* DTT and 50 µ*M* CH-3S was mixed with an equal volume of 8% PEG 8000, 50 m*M* KP_i pH 4.7 and 20% glycerol. Microseeding with a cat whisker was then employed to improve the morphology of the initial crystals. Before flash-cooling in liquid nitrogen, all crystals were soaked briefly in a cryo-buffer consisting of well solution with 30% glycerol. Heavy-atom cation and anion quick-soak salt derivatives were prepared by the addition of 300 m*M* CsCl or NaI salt to the cryo-buffer (Nagem *et al.*, 2001).

2.4. Data collection and processing

All data were collected at 105 K using a 100 µm source MacScience M06X SRA Cu rotating-anode X-ray generator (Table 1). The 2.6 Å native data, merged from two crystals for higher redundancy and greater completeness, and those from the CsCl derivative were collected on a MacScience DIP2030 image-plate detector equipped with MSC Blue confocal optics. The exposure time roughly matches the image-decay time in the storage phosphors ($\tau \simeq 30$ min; White *et al.*, 1999) and the oscillation range of 0.5° was selected to maximize collection efficiency while still providing good signal to noise without fine-slicing (Pflugrath, 1999). The NaI-derivative data set was collected on a Bruker-AXS SMART-2K CCD detector equipped with XENOCS Fox-2D elliptical optics. The exposure time of 10 min per 0.25° frame is limited by the build-up of dark current and zingers in the CCD. The Bruker CCD was used to rapidly screen all of the heavy-atom derivatives and various cryo-buffers and then to collect a preliminary native data set to 2.8 Å. The image-plate data were indexed, integrated and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). All data collected on the CCD were processed using the *PROTEUM* suite (Bruker-AXS, Madison, Wisconsin, USA). Potential heavy-atom derivatives were evaluated *via* cross-χ^2 and cross-$R_{\rm merge}$ values (Gewirth, 2003) with *HKL*-2000.

2.5. Structure solution

Phaser (Collaborative Computational Project, Number 4, 1994; McCoy *et al.*, 2005) was used to search for possible molecular-replacement (MR) solutions. Homology models for molecular-replacement searches were created using *Swiss-Model* (Schwede *et al.*, 2003) based on CYP3A4 (PDB code 1tqn; Yano *et al.*, 2004) and CYP2C8 (PDB code 1pq2; Schoch *et al.*, 2004). In addition to the homology models, the original PDB entries were also used in MR searches.

Heavy-atom structure solution, refinement and density modification were carried out using the *SOLVE*/*RESOLVE* program package (Terwilliger & Berendzen, 1999). Initial verification of the *SOLVE* heavy-atom sites was performed using isomorphous and anomalous difference cross-Fourier maps created in *XTALVIEW* (McRee, 1999*b*) with the experimental phases from *RESOLVE*. *TEXTAL* (Gopal *et al.*, 2007) was utilized for automated model building into the *RESOLVE* density-modified maps, while manual model building was performed using *Coot* (Emsley & Cowtan, 2004) and *XFIT* (McRee, 1999*b*).

2.6. Post-structure solution analysis

Final isomorphous difference cross-Fourier maps and figures were generated using phases calculated from the

refined 2.6 Å native model. These maps, and σ_A-weighted $2F_o - F_c$ and $F_o - F_c$ maps from *PMB*, were viewed in *XFIT*. After solving the 2.6 Å resolution structure, the model was re-refined and a few missing residues were added using a subsequently obtained 1.9 Å resolution data set collected at SSRL BL11-1. Detailed analysis of this structure (PDB code 2q9g) will be reported elsewhere, but it is used here in the analysis of the heavy-atom structures instead of the less accurate 2.6 Å native structure.

The heavy-atom-containing models were refined against the derivative data using default *PMB* techniques (Singh *et al.*, 2006; Scott *et al.*, 2004), such as variable-σ *B*-factor restraints (Tickle *et al.*, 1998) and stereochemical bond-length variation targets. *PMB*, a suite of utilities and *CNS* patches for semi-automated structure refinement using *CNS*, is available from the author MAW *via* the WWW (http://xray.utmb.edu/PMB). No water molecules were modelled owing to the limited resolution. The individual occupancies of the heavy-atom sites were refined after resetting their *B* factors to the average protein value (I, 40 Å^2; Cs, 45 Å^2). However, at this resolution the occupancy and *B* factors are highly correlated.

3. Results and discussion

3.1. Crystallization and crystal morphology

Many proteins, including most P450s, are stabilized by the binding of a substrate or inhibitor molecule. Cocrystallization with a ligand will frequently aid in the crystallization of a protein. Therefore, CYP46A1 was purified and crystallized in the presence of the tight-binding substrate analogue CH-3S. Cocrystals of CYP46A1 and CH-3S grew within two weeks from microseeded sitting drops. The crystals appeared as clusters of rectangular rods of approximately 40 µm in width and 300 µm in length (Fig. 1). The CYP46A1–CH-3S crystals diffracted to ∼2.6 Å on a Cu *Kα* home source and were found to belong to the tetragonal space group $I4_122$, with unit-cell parameters $a = b = 121.7$, $c = 144.2$ Å. The Matthews coefficient (Matthews, 1962) of 2.3 Å^3 Da^{-1} indicated that the crystal had one molecule per ASU and a solvent content of 47%.

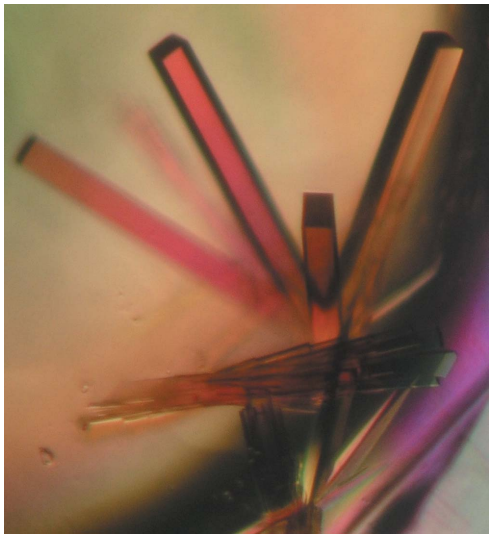

Figure 1
Typical tetragonal crystals of CYP46A1. The long tetragonal rods are approximately 40 µm in width and 300 µm in length. The red color is typical of the cytochrome P450s and is the consequence of optical absorbance by the haem prosthetic group. The cytochrome P450 46A1 packs in space group $I4_122$ with the *c* axis parallel to the long axis of the rods.

3.2. Molecular replacement and derivative screening

Molecular replacement was attempted with several models, using two closely related PDB structures and homology models based on them. None of these MR searches produced a solution. The conformational flexibility of the P450s can make structure solution by MR problematic. Therefore, at this point we decided to proceed with a strategy that had proven successful in our previous CYP2B4 structure solution (Scott *et al.*, 2004): the use of NaI anionic halide salt derivatization. In addition to the NaI derivative, a cation heavy-atom salt derivative, CsCl, was also screened. This choice was based on our experience using various monovalent cation and anion solutions for phasing. Other commonly used heavy-atom compounds of Pt and Hg were screened at the same time and none were found to produce a useable heavy-atom derivative. We did not attempt to use the Fe position, determined using an anomalous difference Patterson map, to boot-strap an MR solution. We felt that the time and effort spent in obtaining derivative data and an unbiased experimentally phased electron-density map *via* the quick-soak method is significantly less than that required to rebuild a poor MR solution. In a post-structure-solution analysis, the best MR search model, a *Swiss-Model* homology model based on CYP3A4, had a 4.7 Å C$^\alpha$ r.m.s.d. to the final CYP46A1 structure, although the C$^\alpha$ r.m.s.d. to the palm domain (residues 290+) was only 2 Å.

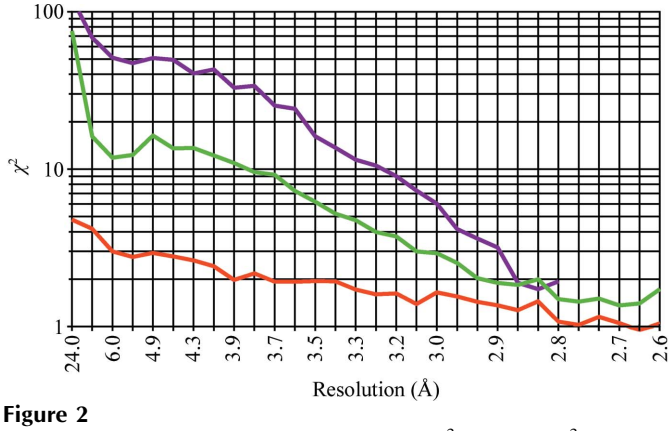

Figure 2
The isomorphous and anomalous differences $\{\chi^2 \simeq [\Delta I/\sigma(I)]^2\}$ from the three derivatives. The NaI (purple) and CsCl (green) derivatives both have a large solvent contribution at very low resolution. At the Cu *Kα* wavelength the anomalous Fe signal (red) is smaller than either of the isomorphous differences. The NaI data were limited to 2.8 Å resolution by the detector geometry. The χ^2 values were calculated in *SCALEPACK* using 30 bins from 30 to 2.6 Å.

3.3. MIRAS structure solution

Data were collected to 2.6 Å resolution from two native crystals and a CsCl quick-soak derivative crystal (Table 1). The NaI quick-soak derivative crystal data were limited to 2.8 Å resolution by the detector geometry. The structure was then solved using the MIRAS technique in *SOLVE*. The weak anomalous signal from the haem iron was combined with the isomorphous differences from the two heavy-atom salt soaks (NaI and CsCl). Although the anomalous signal from the iron ($f' \simeq -1.1$ and $f'' \simeq 3.2$ at the Cu $K\alpha$ wavelength) was weaker than the isomorphous differences for either the NaI or CsCl soaks ($\Delta f \simeq 36$ for I and $\Delta f \simeq 40$ for Cs) based on the χ^2 statistic (Fig. 2), the anomalous Patterson map showed clear and strong peaks (Figs. 3*a* and 3*b*). The isomorphous difference Patterson maps for the two salt derivatives also displayed strong peaks (Figs. 3*c* and 3*d*). Even though the χ^2 values for the caesium and iodide derivatives are significantly greater than that of the iron anomalous data, their isomorphous difference Patterson maps are noisier with many weak peaks, requiring automated peak-search algorithms, while the Fe anomalous difference Patterson could be solved by inspection.

The (300 m*M*) NaI and CsCl derivatives were nearly isomorphous to the low ionic strength native crystal (25 m*M* NaCl) without the addition of NaCl to its cryo-buffer to create an isoionic strength buffered native crystal. The NaI derivative displayed a strong anomalous and isomorphous derivative signal as measured by the χ^2 statistic in *SCALEPACK*, with an overall isomorphous difference χ^2 of 28. The CsCl derivative also had a moderately good signal, with an overall isomorphous difference χ^2 of 8.4. The anomalous (Bijvoet) differences in the native data showed a small but significant signal at low resolution, with an overall Bijvoet pair χ^2 of only 1.9 (Fig. 2). However, the phasing power of the iron was significantly higher than that of either isomorphous derivative (Table 1) and the phasing power of the iodide derivative was significantly higher than that of the caesium derivative.

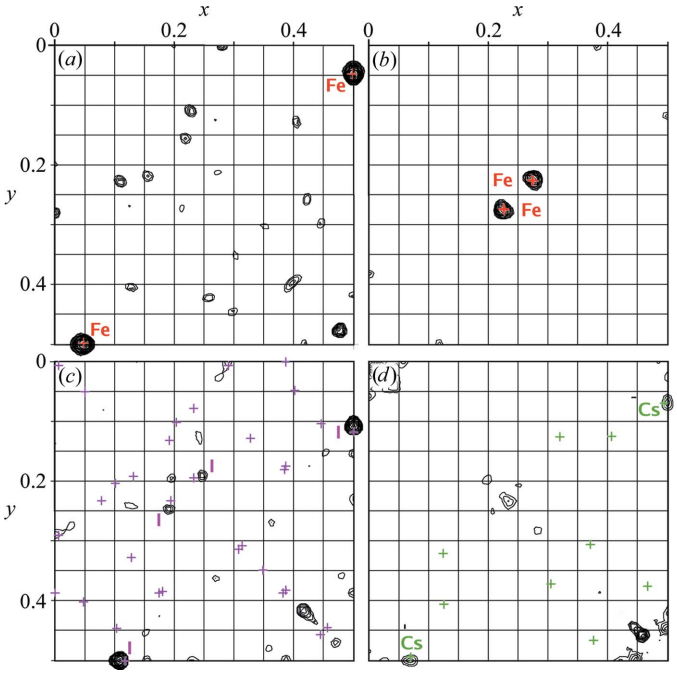

Figure 3
Patterson maps. These are *xy* Harker sections showing the asymmetric unit. Bijvoet difference Patterson maps at (*a*) $z = 0.5$ and (*b*) $z = 0.25$ are shown for the native data. Isomorphous difference Patterson maps at $z = 0.5$ are shown for (*c*) the NaI and (*d*) the CsCl quick-soak derivatives, respectively. Maps are contoured starting at 3σ in 0.5σ steps. Predicted Patterson peaks are labelled with a cross; observed peaks are labelled with the atom name. The iron anomalous and caesium isomorphous difference maps are at 2.6 Å resolution, while the iodide isomorphous derivative map is limited to 2.8 Å resolution. The origin peak at (1/2, 1/2, 1/2) has been removed.

The heavy-atom structure solution in *SOLVE* found the single iron site, 14 iodide sites and five caesium sites. The iron site was treated as a separate derivative using the native data and was refined by the anomalous signal alone, while the Cs and I sites were refined using only the isomorphous differences, because the anomalous signals of all the data have a significant contribution from the haem iron. However, the anomalous signal from each of the isomorphous derivatives was used in the phase calculations. The anomalous and isomorphous difference signals of the data all become insignificant beyond about 3 Å resolution (Fig. 2) owing to the lower signal to noise of the high-resolution data. However, all data collected to about 2.6 Å resolution are useful in the structure-solution process when solvent flattening is used to improve and to extend the phases to the limit of the data.

The observed decrease in the isomorphous difference signals with resolution (Fig. 2) arising from the resolution-dependence of the atomic scattering factors and the measurement accuracy is consistent with expectations. The small but slightly elevated χ^2 values in the highest resolution bins is an indication that the derivatives are not completely isomorphous to the native crystal. Non-isomorphism, the effect of which increases with resolution (Garman & Murray, 2003), will result in significant intensity differences in all resolution bins. This measurement of the isomorphism of the derivatives is a more reliable initial indicator than the cell-dimension changes alone, as the accuracy of the cell dimensions determined by typical protein crystallography programs is limited and may not be indicative of changes in protein conformation or orientation within the cell. The resolution-dependence of ΔF is a better measure of isomorphism and should be examined once a full data set has been collected (McRee, 1999*a*). Note that at very low resolution both the cation and anion derivatives have a large erroneous signal from differences in their solvent scattering, a problem inherent to the quick-soak procedure. This effect is not observed in traditional heavy-atom derivatives that are back-soaked (Garman & Murray, 2003) and which use significantly lower heavy-atom concentrations than either anion or cation quick soaks.

3.4. Analysis of the heavy-atom sites

In contrast to traditional single-compound heavy-atom derivatives, both of the derivatives and the native crystal each contained the anomalously scattering haem iron. The anom-

Table 2
Crystallographic refinement of heavy-atom sites.

	Native (Fe)		NaI		CsCl	
	Occupancy	B (Å^2)	Occupancy†	B (Å^2)	Occupancy†	B (Å^2)
Protein	1.00	48	1.00	40	1.00	46
Fe	1.00	26	1.00	21	1.00	28
Site 1			0.97	48	0.34	52
Site 2			0.71	50	0.30	52
Site 3			0.68	53	0.26	63
Site 4			0.65	53	0.21	52
Site 5			0.62	49	0.20	60
Site 6			0.52	53		
Site 7			0.52	49		
Site 8			0.49	51		
Site 9			0.47	57		
Site 10			0.47	53		
Site 11			0.46	58		
Site 12			0.42	60		
Site 13			0.37	57		
Site 14			0.22	58		

† Occupancy and B factors are correlated (see §2.6).

alous signal from the iron ($f'' \simeq 3.4$ e at Cu $K\alpha$) was exploited as a single anomalous-only derivative using the native data. The Fe atom is tightly bound to the protein and is present as a single copy per ASU or eight Fe atoms per unit cell. Refinement of the B factor, but not the occupancy (which is assumed to be one), in *CNS* showed that the tightly bound iron had a low B factor, much lower than the average protein B factor (Table 2). The high occupancy and low B factor means that even the relatively low anomalous signal available from a single Fe atom per 500 residues still makes a useful contribution to phasing (Fig. 2). In contrast, the heavy atoms in the salt-derivative soaks are not tightly bound to the protein; they are dynamic and are loosely bound on the protein surface (Fig. 4). Their occupancies vary in a rough continuum from a maximum of 1.0 to zero occupancy (Table 2). The iodide ions can find many surface-binding sites of significant affinity (Fig. 4). In the case of CYP46A1, *SOLVE* found 14 iodide sites with occupancies ranging from about one to a lower limit of ∼0.2. The dynamic nature of surface-binding residues is also mirrored in the B-factors of the halide ions, which are higher than that of the average protein atom. The caesium derivative had a slightly higher overall B factor than the iodide derivative. However, it is the lower caesium occupancies, in concert with the lower number of sites (five), that combined to reduce the phasing power of the caesium derivative relative to those of the iron and iodide derivatives.

An examination of the dual nature (Evans & Bricogne, 2002) of the iodide-binding sites is informative as to why there are more sites for the iodide derivative than there are for the caesium derivative. Iodide can bind to either solvent-exposed basic residues or in hydrophobic pockets (Figs. 5*a* and 5*b*) and in addition to having extensive hydrophobic interactions frequently forms at least one hydrogen bond. In the CYP46A1 iodide derivative, four of the 14 iodides were observed without any bonding partners (Fig. 5*b*). Other derivative salts seem to require more than one bonding partner, such as a carbonyl or carboxylate group in the case of caesium ions. The caesium-ion sites in CYP46A1 typically required three ordered oxygen-binding partners (Fig. 5*a*). The coordination of the caesium seems to be less than ideal, as in most cases the occupancy is much lower than that observed for the iodide sites. Another factor favouring the NaI derivative is the protein pI. CYP46A1 is a basic protein (pI 9.1) with many more basic sites suitable for anion binding than acidic sites available for cations. It is interesting to note that the initial halide-ion test crystals were also basic proteins (Dauter *et al.*, 2000; Nagem *et al.*, 2003). The role of protein pI in halide binding has not been extensively explored. However, our experience with Langat virus E protein domain III (pI 7.0) has shown that halide ions are still a useful derivative at lower protein pIs (White *et al.*, 2003). Based on solution chemical properties, Nagem and coworkers suggested that the pH of the buffer solution may also be important. The binding of cation salts may be favoured at high buffer pH, while the anion salts were observed to be favoured at lower pH values.

3.5. Implication of results for the phasing of other systems

When evaluating a potential derivative, there are several critical items to consider. What makes a good derivative and why do the apparently large signals not correspond to phasing power? Heavy atoms that are bound tightly to the protein, such as a Hg bound to a cysteine, the Fe of a haem or a selenium incorporated into a methionine, will produce strong sharp Patterson peaks. The loosely bound solvent atoms of a cation or anion quick soak, with their moderate occupancy and higher than average B factors, will have smaller and much more diffuse Patterson peaks (Buerger, 1959; Dauter & Dauter, 2001). The difference in the solvent scattering between the native and the quick-soak derivative is also a

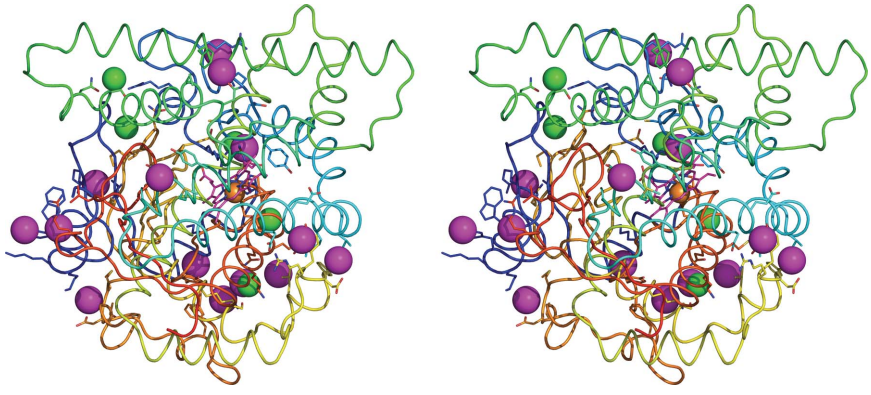

Figure 4
Complementary cation and anion heavy-atom salt-derivative sites. Cross-eyed stereoview of the incomplete N-to-C-termini rainbow-colored CYP46A1 protein backbone and heavy-atom sites (colored spheres: I, purple; Cs, green) showing the nearby interacting residues as sticks. This figure was created using *PyMOL* (DeLano, 2003).

major source of non-isomorphism or noise (Fig. 2), even though the protein and crystal lattice are less likely to be affected by the soaking (assuming an isoionic strength buffer). Solvent contributions can be significant at 3–4 Å resolution (Afonine *et al.*, 2005) and will affect both anomalous and isomorphous differences. In addition, the SIR/MIR(AS) techniques suffer from inter-crystal differences, such as absorption, differential solvent contrast and classical crystal non-isomorphism (lattice changes and protein rearrangement), which are not present in the single-crystal MAD and SAD methods. This was amply demonstrated in our case, where an apparently weak (SAD) anomalous signal from a well ordered Fe provided superior phasing power to the demonstrably stronger (MIRAS) isomorphous differences from the iodide and caesium quick soaks. However, the protein structure could not be solved with any one derivative alone.

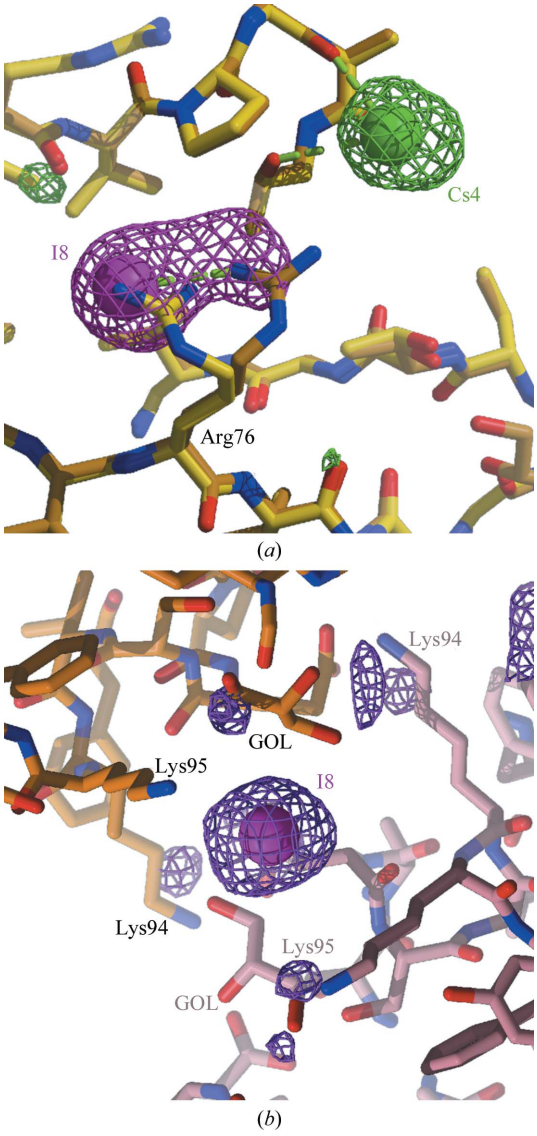

Figure 5
(*a*) Superposition of the 2.6 and 2.8 Å isomorphous difference cross-Fourier maps contoured at 3σ showing the interaction of caesium (green) or iodide (purple) ions with the corresponding CYP46A1 model (yellow or brown, respectively). Interactions are displayed as green dashes. Close caesium and iodide sites are observed near Arg76, which has moved in the NaI model (brown) and makes a contribution to the isomorphous difference signal (magenta mesh). The third interaction of Cs4, with the carbonyl of Asn388, is obscured in this view. (*b*) View of a twofold crystallographic symmetry-generated protein interface. The iodide (I8) sits on the twofold axis, flanked by four symmetry-related lysine side chains and two symmetry-related glycerol (GOL) molecules; the symmetry-generated molecules are colored lavender. This is one of four iodide sites that lacks hydrogen bonds; the lysine N^{ζ}–I distance is 6 Å. This figure was drawn using *XFIT*/*RASTER3D* (Merritt & Bacon, 1997).

Although this structure was solved on a fixed-wavelength home source, the general technique is applicable to synchrotron radiation using a modified MAD, SAD or enhanced MIRAS technique. Either lighter heavy-atom salts that have accessible absorption edges can be used, such as NaBr and RbCl, or in the enhanced MIRAS and SAD methods the X-ray wavelength can be chosen to provide increased anomalous signal from the CsCl and NaI salts.

The simultaneous use of anion–cation heavy-atom salts (RbBr) has been considered as a possible tool for MAD structure solution (MAW, personal communication). This multiple-wavelength multiple anomalous scatterer diffraction (MMAD) technique offers the possibility of creating two perfectly isomorphous anomalous derivatives, thereby increasing the number of sites and also the anomalous signal for phasing. The CYP46A1 structure solution provides an example of an overlapping cation and anion site near Asn444 and there is another location near Arg76 where the iodide and caesium sites are in close proximity (Fig. 5). This would have a negative impact on the expected anomalous signal when using the MMAD technique. In a mixed heavy-atom soak these sites could either have low occupancy of both anomalous species or favour a single species. In either case the anomalous signal for phasing would be reduced in comparison with separate derivatives. In general, for these reasons mixed heavy-atom soaks such as CsI or RbBr may be less advantageous than separate derivatives.

Which technique is better when using anion and cation quick soaks, Cs/I MIRAS or Rb/Br MAD? Given their large anomalous and isomorphous ($f''_{\rm Cu} \simeq 7$, $\Delta f \simeq 40$) difference signals, higher atomic number salts, such as Cs and in particular I with its dual nature, may provide superior phasing power when using either MIRAS or possibly even SAD than the strong anomalous signal from the lighter heavy-atom salts Rb and Br, which have accessible absorption edges suitable for MAD or SAD (Br, $f' = -8.3$, $f'' = 3.8$ at 13 474 eV; Rb, $f' = -10.4$, $f'' = 3.8$ at 15 200 eV). The ability to solve structures quickly using only a readily available home source is an advantage of the Cs/I salts with their large Δf values. However, when crystal-to-crystal non-isomorphism is a problem then it may be easier to solve the structure with data from a single crystal using either the MAD or SAD technique.

3.6. General application of complementary salt derivatives

When using the MIR(AS) technique, cell non-isomorphism can be a major impediment to structure solution. This problem

is exacerbated in the quick-soak method by the high ionic strength of the heavy-atom salts. Osmotic stress-induced crystal damage can be reduced by lowering the salt concentration, but at the expense of phasing power. A possible solution to this problem is to use isoionic strength crystal buffers by the addition of similar salt concentrations to each condition: cation heavy atom (CsCl), anion heavy atom (NaI) and native (NaCl). This isoionic strength buffer technique was applied in a recent MIRAS structure solution of the closed form of CYP2B4 (Scott *et al.*, 2004). In this example, the cryo-buffer for the native 2B4 crystal contained 300 m*M* NaCl and a single NaI derivative was prepared by replacing the NaCl with the isoionic strength heavy-atom quick-soak cryo-buffer containing 300 m*M* NaI. The importance of using isoionic strength buffers was demonstrated recently in the 1.4 Å SIRAS structure solution of *Schizosaccahromyces pombe* Pop2p deadenylase subunit (Jonstrup *et al.*, 2007). When the native proved to be non-isomorphous with either derivative, the CsCl quick-soak derivative was used as an isomorphous 'native' to find the iodide sites in the NaI derivative. It is not known whether isoosmotic strength or isoionic strength solutions are better at reducing crystal non-isomorphism. In the case of monovalent salts this distinction is irrelevant as both are proportional to the salt concentration.

In other cases, we have titrated halide salts up to the maximum concentration permitted by the crystal before non-isomorphism or mosaicity (crystal damage) became problematic. With crystals of Langat virus E protein domain III crystallized in 1.8 *M* ammonium sulfate (White *et al.*, 2003), we were able to titrate up to 250 m*M* KI, simultaneously reducing the ammonium sulfate to 1.6 *M*, while still maintaining the crystal's suitability as an isomorphous derivative. This was a difficult problem because the sulfate ions formed bridging crystal contacts that were disrupted at higher halide salt concentrations. This case represents the lowest concentration of halide salt we have used to obtain an isomorphous derivative, but does not necessarily represent the lowest concentration that will produce a useful derivative: the successful SAD structure solution of SOD was accomplished with only a 200 m*M* NH_4I crystallization solution (Yogavel *et al.*, 2007).

We have demonstrated the basic application of both Cs cation and I anion heavy-atom salts as traditional MIR(AS) derivatives. The structure of CYP46A1 with CH-3S bound in the active site has been determined to 2.6 Å by the MIRAS method using complementary cation and anion salt quick soaks. This demonstrates the applicability of the quick-soak technique in the practical case of a moderately well diffracting crystal. Given the simplicity and ease of applying this method, it is recommended as a routine approach to be tried whenever attempting a MIR(AS) structure solution. Heavy-atom salt quick soaks can also be used to rescue a poor MAD or SAD solution (Evans & Bricogne, 2002; Nagem *et al.*, 2003).

3.7. Final CYP46A1 structure refinement

New CYP46A1–CH-3S data were collected to 1.9 Å resolution at SSRL BL11-1 and refinement of the preliminary 2.6 Å model against these high-resolution data has been performed. The details of the refinement and analysis of the 1.9 Å structure of CYP46A1–CH-3S (PDB code 2q9f) will be reported elsewhere.

We wish to acknowledge the support of the Sealy and Smith Foundation *via* a grant to the Sealy Center for Structural and Molecular Biology. This work was supported by grants from NIH, GM 62882, AG 024336 and EY018383, Merck Research Laboratories and the Sealy and Smith Foundation (to IAP) and from NIH GM 031001 (to EFJ). We are grateful to David A. Konkel for carefully reviewing the manuscript and providing helpful comments.

References

Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *Acta Cryst.* D**61**, 850–855.
Bogdanovic, N., Bretillon, L., Lund, E. G., Diczfalusy, U., Lannfelt, L., Winblad, B., Russell, D. W. & Bjorkhem, I. (2001). *Neurosci. Lett.* **314**, 45–48.
Bretillon, L., Diczfalusy, U., Bjorkhem, I., Maire, M. A., Martine, L., Joffre, C., Acar, N., Bron, A. & Creuzot-Garcher, C. (2007). *Curr. Eye Res.* **32**, 361–366.
Brown, J. III, Theisler, C., Silberman, S., Magnuson, D., Gottardi-Littell, N., Lee, J. M., Yager, D., Crowley, J., Sambamurti, K., Rahman, M. M., Reiss, A. B., Eckman, C. B. & Wolozin, B. (2004). *J. Biol. Chem.* **279**, 34674–34681.
Buerger, M. J. (1959). *Vector Space and its Application in Crystal Structure Investigation.* New York: Wiley.
Chawla, A., Repa, J. J., Evans, R. M. & Mangelsdorf, D. J. (2001). *Science*, **294**, 1866–1870.
Chen, W., Chen, G., Head, D. L., Mangelsdorf, D. J. & Russell, D. W. (2007). *Cell Metab.* **5**, 73–79.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dauter, M. & Dauter, Z. (2006). *Methods Mol. Biol.* **364**, 149–158.
Dauter, Z. & Dauter, M. (2001). *Structure*, **9**, R21–R26.
Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* D**56**, 232–237.
DeLano, W. L. (2003). *The PyMOL Molecular Graphics System.* http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, G. & Bricogne, G. (2002). *Acta Cryst.* D**58**, 976–991.
Garman, E. & Murray, J. W. (2003). *Acta Cryst.* D**59**, 1903–1913.
Gewirth, D. (2003). *The HKL Manual*, v.1.0. Charlottesville, USA: HKL Research.
Gopal, K., McKee, E., Romo, T., Pai, R., Smith, J., Sacchettini, J. & Ioerger, T. (2007). *Bioinformatics*, **23**, 375–377.
Graham, S. E. & Peterson, J. A. (1999). *Arch. Biochem. Biophys.* **369**, 24–29.
Guengerich, F. P. (2001). *Chem. Res. Toxicol.* **14**, 611–650.
Guengerich, F. P. (2005). *Cytochrome P450*, edited by P. R. Ortiz de Montellano, pp. 377–530. Dordrecht: Kluwer Academic Publishers.
Helisalmi, S., Vepsalainen, S., Koivisto, A. M., Mannermaa, A., Iivonen, S., Hiltunen, M., Kiviniemi, V. & Soininen, H. (2006). *J. Neurol. Neurosurg. Psychiatr.* **77**, 421–422.
Iwamori, M., Moser, H. W. & Kishimoto, Y. (1976). *Biochim. Biophys. Acta*, **441**, 268–279.
Johnson, E. F. & Stout, C. D. (2005). *Biochem. Biophys. Res. Commun.* **338**, 331–336.
Jonstrup, A. T., Andersen, K. R., Van, L. B. & Brodersen, D. E. (2007). *Nucleic Acids Res.* **35**, 3153–3164.
Korolev, S., Dementieva, I., Sanishvili, R., Minor, W., Otwinowski, Z. & Joachimiak, A. (2001). *Acta Cryst.* D**57**, 1008–1012.

Kotti, T. J., Ramirez, D. M., Pfeiffer, B. E., Huber, K. M. & Russell, D. W. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 3869–3874.
Li, H. & Poulos, T. L. (2004). *Curr. Top. Med. Chem.* **4**, 1789–1802.
Li, Y., Chu, L. W., Chen, Y. Q., Cheung, B. M., Leung, R. Y., Yik, P. Y., Ng, K. M., Mak, W., Jin, D. Y., St George-Hyslop, P. & Song, Y. Q. (2006). *Dement. Geriatr. Cogn. Disord.* **22**, 399–404.
Liu, S., Sjovall, J. & Griffiths, W. J. (2003). *Anal. Chem.* **75**, 5835–5846.
Lund, E. G., Guileyardo, J. M. & Russell, D. W. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 7238–7243.
Lutjohann, D., Breuer, O., Ahlborg, G., Nennesmo, I., Siden, A., Diczfalusy, U. & Bjorkhem, I. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 9799–9804.
Ma, S. L., Tang, N. L., Lam, L. C. & Chiu, H. F. (2006). *Int. Psychogeriatr.* **18**, 37–45.
McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.
McRee, D. E. (1999*a*). *Practical Protein Crystallography*, 2nd ed. San Diego: Academic Press.
McRee, D. E. (1999*b*). *J. Struct. Biol.* **125**, 156–165.
Mast, N., Andersson, U., Nakayama, K., Bjorkhem, I. & Pikuleva, I. A. (2004). *Arch. Biochem. Biophys.* **428**, 99–108.
Matthews, B. W. (1962). *J. Mol. Biol.* **33**, 491–497.
Merritt, E. & Bacon, D. (1997). *Methods Enzymol.* **277**, 505–524.
Nagem, R. A. P., Dauter, Z. & Polikarpov, I. (2001). *Acta Cryst.* D**57**, 996–1002.
Nagem, R. A., Polikarpov, I. & Dauter, Z. (2003). *Methods Enzymol.* **374**, 120–137.
Nelson, D. R., Koymans, L., Kamataki, T., Stegeman, J. J., Feyereisen, R., Waxman, D. J., Waterman, M. R., Gotoh, O., Coon, M. J., Estabrook, R. W., Gunsalus, I. C. & Nebert, D. W. (1996). *Pharmacogenetics*, **6**, 1–42.
Omura, T. & Sato, R. (1964). *J. Biol. Chem.* **239**, 2370–2378.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.
Pikuleva, I. A. (2006). *Pharmacol. Ther.* **112**, 761–773.
Sansen, S., Yano, J. K., Reynald, R. L., Schoch, G. A., Griffin, K. J., Stout, C. D. & Johnson, E. F. (2007). *J. Biol. Chem.* **282**, 14348–14355.
Schoch, G. A., Yano, J. K., Wester, M. R., Griffin, K. J., Stout, C. D. & Johnson, E. F. (2004). *J. Biol. Chem.* **279**, 9497–9503.
Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. (2003). *Nucleic Acids Res.* **31**, 3381–3385.
Scott, E. E., He, Y. A., Wester, M. R., White, M. A., Chin, C. C., Halpert, J. R., Johnson, E. F. & Stout, C. D. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 13196–13201.
Scott, E. E., White, M. A., He, Y. A., Johnson, E. F., Stout, C. D. & Halpert, J. R. (2004). *J. Biol. Chem.* **279**, 27294–27301.
Singh, R., White, M. A., Ramana, K. V., Petrash, J. M., Watowich, S. J., Bhatnagar, A. & Srivastava, S. K. (2006). *Proteins*, **64**, 101–110.
Strott, C. A. & Higashi, Y. (2003). *J. Lipid Res.* **44**, 1268–1278.
Tedde, A., Rotondi, M., Cellini, E., Bagnoli, S., Muratore, L., Nacmias, B. & Sorbi, S. (2006). *Neurobiol. Aging*, **27**, 773.e1–773.e3.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Tickle, I. J., Laskowski, R. A. & Moss, D. S. (1998). *Acta Cryst.* D**54**, 243–252.
Wang, F. & Jia, J. (2007). *Brain Res.* **1147**, 34–38.
White, M. A., Liu, D., Holbrook, M. R., Shope, R. E., Barrett, A. D. T. & Fox, R. O. (2003). *Acta Cryst.* D**59**, 1049–1051.
White, M. A., Watowich, S. J. & Fox, R. O. (1999). *J. Appl. Cryst.* **32**, 65–70.
Yano, J. K., Wester, M. R., Schoch, G. A., Griffin, K. J., Stout, C. D. & Johnson, E. F. (2004). *J. Biol. Chem.* **279**, 38091–38094.
Yogavel, M., Gill, J., Mishra, P. C. & Sharma, A. (2007). *Acta Cryst.* D**63**, 931–934.
Zhao, Y., Sun, L., Muralidhara, B. K., Kumar, S., White, M. A., Stout, C. D. & Halpert, J. R. (2007). *Biochemistry*, **46**, 11559–11567.
Zhao, Y., White, M. A., Muralidhara, B. K., Sun, L., Halpert, J. R. & Stout, C. D. (2006). *J. Biol. Chem.* **281**, 5973–5981.