Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Triiodide derivatization and combinatorial counter-ion replacement: two methods for enhancing phasing signal using laboratory Cu *Kα* X-ray equipment

Gwyndaf Evans[a,b] and Gérard Bricogne[a,b]*

[a]Global Phasing Limited, Cambridge, England, and [b]MRC Laboratory of Molecular Biology, Hills Road, Cambridge, England

Correspondence e-mail: gb10@globalphasing.com

Received 19 December 2001
Accepted 25 March 2002

PDB References: xylose isomerase (KI/I_2), 1gw9; elastase (KI/I_2), 1gwa; hen egg-white lysozyme (KI/I_2), 1gwd; apoferritin (KI/I_2), 1gwg.

A series of experiments performed at Cu *Kα* wavelength on in-house X-ray equipment are presented which investigate two possibilities for enhancing the experimental phasing signal by means of (i) triiodide/iodide soaks using KI/I_2 and (ii) combinations of counter-ions introduced using the short cryosoak method. Triiodide-derivative crystal structures for five test proteins have been refined and reveal that iodine can bind as polyiodide and single iodide ions through hydrophobic and hydrogen-bonding interactions both at the molecular surface and in intramolecular and intermolecular cavities. In three cases, the structures could be automatically determined with *autoSHARP* using in-house SAD and SIRAS data. The investigation of combinatorial counter-ion replacement using multiple salts with Na^+ and Cs^+ as cations and I^- and Cl^- as anions reveals that, for the case of hen egg-white lysozyme, significant improvement in phasing signal is obtained by the combined use of salts compared with SIRAS methods using native and single short-soak derivative data sets.

1. Introduction

The isomorphous replacement (IR [1]) method (Green *et al.*, 1954; Crick & Magdoff, 1956), which played a key role in the birth of protein crystallography, has seemed over the last decade to inexorably lose ground to the MAD method.

There is, however, evidence that the very same advances which made the implementation of the MAD method possible could now support the re-emergence of IR by allowing the exploitation of versatile chemical interactions which were previously unsuitable for derivatization.

The work of Dauter & Dauter (1999), Dauter *et al.* (2000) and Nagem *et al.* (2001) has demonstrated that short soaks in solutions containing halide ions (Br^-, I^-) and certain cations (Cs^+, Gd^{3+}) can produce sufficient derivatization to support *de novo* structure determination by exploiting IR and/or anomalous scattering (AS) effects on a synchrotron source tuned to the Cu *Kα* wavelength.

In this paper, we present an investigation of the potential usefulness of these methods using conventional laboratory equipment (Cu rotating-anode generators and image-plate or CCD detectors) and extend them in two main directions: (i) towards the use of combinatorial counter-ion replacement, which mimics the MAD method by chemical means, and (ii)

[1] Abbreviations used throughout: IR, isomorphous replacement; AS, anomalous scattering; SAD, single-wavelength anomalous diffraction; S/MIRAS, single/multiple isomorphous replacement with anomalous scattering; MAD, multiple-wavelength anomalous diffraction; ADF, anomalous difference Fourier; HEWL, hen egg-white lysozyme; XI, xylose isomerase; APO, horse spleen apoferritin; ABF, antibody fragment; PPE, porcine pancreatic elastase.

the systematic investigation of the triiodide ion, a versatile derivatization agent which can display a spectrum of behaviours ranging from ionic to hydrophobic, mimicking Xe gas in the latter.

It is concluded that a renaissance of IR can be expected in the near future, which ought to yield high-quality phase information at low cost on in-house equipment and may create new demands for medium-energy synchrotron beamlines.

2. Logistics of heavy-atom phasing: key factors and trends

The classical image of heavy-atom derivatization of protein crystals in the isomorphous replacement and related methods has been superseded in the last few years. Traditionally, the preparation of isomorphous derivatives would involve long soaks in typically sub-millimolar solutions of heavy-atom compounds. The eligible compounds needed to be non-aggressive over long time periods, as samples would be used later for X-ray data collection at room temperature (at least before the introduction of crystal freezing). In addition, the heavy-atom substitution had to be sparse and clean because early heavy-atom detection methods could not routinely determine the positions of large numbers of sites. To ensure a high signal-to-noise ratio in isomorphous and anomalous differences, the substitution also had to be strong and the sites were required to be well ordered. These stringent requirements were difficult to meet and were a significant motivation for the development of alternative methods of introducing heavy atoms into protein crystals, such as the incorporation of Se atoms into proteins by expressing them in selenomethionine-containing media (Hendrickson *et al.*, 1990; Doublié, 1997).

Over the last decade, however, advances in the field of macromolecular crystallography have allowed the problem of producing heavy-atom derivatives to be viewed from a fresh perspective. These advances have been (i) the fast freezing of crystals to cryogenic temperatures, (ii) the improvements in detector technology and X-ray sources, (iii) the dramatic improvements in crystallographic software for the determination of heavy-atom substructures with large numbers of atoms and (iv) the appearance of powerful maximum-likelihood heavy-atom refinement and phasing programs and of automated density-modification procedures.

2.1. Crystal freezing and detectors

The ability to routinely freeze crystals (Hope, 1988) has had a profound effect on the scope of soaking methods for preparing heavy-atom derivatives. The macromolecular crystals need no longer be stable in heavy-atom solutions over long time periods. Crystals may therefore be soaked in heavy-atom compounds and flash-frozen before the crystal integrity is challenged. Crystal freezing essentially halts the sometimes vigorous effects that heavy-atom compounds may have on protein crystals, so that higher concentrations of heavy-atom compounds, as well as potentially more aggressive compounds, become eligible.

Furthermore, the freezing of crystals makes it possible to collect highly redundant diffraction data, which leads to an increased isomorphous and/or anomalous signal-to-noise ratio. Similarly, the improvements in detector technology, through the use of CCDs and image plates, have made the measurements of diffraction intensities more accurate, so that smaller intensity differences have become potential sources of phasing signals.

2.2. Heavy-atom detection methods

Software for the detection of heavy atoms using isomorphous and/or anomalous signals in diffraction data has recently improved to the point where heavy-atom substructures of over 100 atoms may be solved. A decade ago, derivatives yielding numerous weak binding sites would have been deemed to be 'non-specific' and of little use in structure determination. Nowadays, such heavy-atom substructures are solvable by dual-space direct-methods approaches, as implemented in *Shake-and-Bake* (Miller *et al.*, 1994; Smith *et al.*, 1998) and *SHELXD* (Usón & Sheldrick, 1999).

2.3. Heavy-atom refinement and phasing methods

Finally, developments in heavy-atom refinement and phasing software have allowed weak signals to be utilized through the use of maximum-likelihood (ML) methods and statistical descriptions of lack of isomorphism and heavy-atom model imperfection (de La Fortelle & Bricogne, 1997). Log-likelihood gradient maps (Bricogne, 1993) resulting from ML refinement can also identify additional heavy-atom sites, starting from very few initially detected sites. The inclusion of such information over several iterations can considerably improve the quality of experimental phases, which in turn enhance the effectiveness of density-modification programs such as *DM* and *SOLOMON*.

In recent years, this methodology has been shown to be effective in the extreme case of sulfur single-wavelength anomalous diffraction (S-SAD), in which the sole AS signal from S atoms present in cysteine and methionine residues acts as a source of phasing information. This was illustrated by S-SAD results on three test cases: thaumatin (de La Fortelle, Graves & Bricogne, unpublished results), hen egg-white lysozyme (HEWL; Dauter *et al.*, 1999) and β-trypsin (Yang & Pflugrath, 2001).

2.4. Short-soak experiments

A key observation made by Dauter *et al.* (1999) was that a significant additional source of AS in HEWL, besides the S atoms, arose from bound Cl^- ions at the surface of the protein. Indeed, experiments showed that the anomalous signal from HEWL could even be enhanced by the introduction of Br^- ions (Dauter & Dauter, 1999). The accessibility of the Br *K* absorption edge to synchrotron radiation meant that a MAD experiment could be performed, although a single wavelength was actually sufficient to produce electron-density maps of a

very high quality. Further studies ensued and led to the development of the short-cryosoak method of halide incorporation into protein crystals (Dauter *et al.*, 2000). The number of bound halide ions was typically large and the sites had varying degrees of occupation.

The short-cryosoak method of heavy-atom derivative preparation takes full advantage of the technical advances discussed above. The method involves soaking the crystal for 10–60 s in a solution containing both cryoprotectant and a high concentration of halide salts. The crystals are then frozen and are ready for immediate data collection or storage.

3. New strategies for in-house heavy-atom phasing

The SAD and SIRAS methods are well suited to synchrotron sources, where the wavelength may be chosen to optimize AS contributions from the substructure. In fact, the use of SAD phasing with synchrotrons is well described in the literature (Rice *et al.*, 2000; Brodersen *et al.*, 2000; Dauter & Dauter, 1999; Yu-dong *et al.*, 1999; Harvey *et al.*, 1998). There are, on the other hand only, a few examples in the literature of successful SAD experiments performed in-house with Cu $K\alpha$ radiation (Chen *et al.*, 1991; Yang & Pflugrath, 2001). The lack of tunability of Cu-anode laboratory sources is somewhat compensated for by the magnitudes of AS factors at 1.5418 Å for many elements. All elements with atomic number > 46 have $f'' > 3.7$ e. In addition, Cr, Mn, Fe and Co have $f'' > 2.4$ e. Thus, a large number of candidate heavy atoms, such as some first transition-series metals, Cd, I, Xe, Cs, the lanthanides, the third transition-series metals, Tl and Pb, may act as significant sources of anomalous scattering using Cu $K\alpha$ radiation. However, the inability to optimize AS signals and the poorer average signal-to-noise ratios of in-house data impose limitations on the in-house SAD method. This can be overcome to some extent by including native crystal data (SIRAS) in the phase determination. As long as the non-isomorphism between native and derivative samples is small, it can be statistically modelled by phasing programs such as *SHARP*.

In this paper, two possible strategies for enhancing phasing signals using in-house Cu $K\alpha$ radiation are investigated. Firstly, the modification of the AS signal in a predictable way by the replacement of counter-ions is examined and the results of phasing using multiple counter-ion combinations are compared with those obtained by conventional SIRAS methods. Secondly, we extend previous studies using triiodide (I_3^- introduced into the sample as KI/I_2) to protein derivatization for the specific purpose of phasing in-house with Cu $K\alpha$ radiation, where iodine provides significant AS power. These strategies take advantage of the methodology described in the previous section in order to enhance the phasing signal in SAD, SIRAS and MIRAS phasing.

3.1. Combinatorial counter-ion replacement

The binding of cations and/or anions to the surface of proteins has been observed and studied by a number of authors (Badger *et al.*, 1994; Gursky *et al.*, 1992). Fourme *et al.* (1995) observed ordered binding when using high concentrations of heavy-atom salts to induce changes in the scattering properties of the solvent region in MASC (multiple-wavelength anomalous solvent contrast) experiments. In-house low-resolution contrast-variation studies (Evans *et al.*, 2000) performed on HEWL using high concentrations of NaCl, NaI, CsCl and CsI (unpublished results) have shown that some surface-binding sites are shared among different counter-ions of like charge. Indeed, recent work by Nagem *et al.* (2001) has demonstrated that NaI, CsCl and $GdCl_3$ salts at high concentrations can be used to produce phase information using the exchange of one type of counter-ion for another. This only produces relatively weak amplitude differences, but these authors were able to overcome this limitation by collecting highly redundant data sets on synchrotron sources tuned to the Cu $K\alpha$ wavelength.

We decided to investigate an alternative possibility of enhancing the phasing power of such weak differences which would be applicable to data collected on laboratory X-ray sources. We chose to exploit the similarities of chemical behaviour between different alkali-metal cations on the one hand and different halide anions on the other, which result in their binding at essentially the same sites on the protein surface. In this way, even when using fixed-wavelength radiation, site-specific changes in heavy-atom scattering power can be induced by purely chemical means, which mimics the physically induced changes of f' and f'' on which the MAD method is based.

The studies were performed on crystals of HEWL and the results reported in §5.1 show that phase information of very high quality can indeed be obtained in this way, with great economy of means.

3.2. Use of triiodide as a broad-spectrum derivatization agent

In studies described by Dauter & Dauter (2001), the iodide ion – the most promising halide ion for work at the Cu $K\alpha$ wavelength – was unfortunately found to be the most deleterious to protein crystals. The work reported here shows that this difficulty may be overcome by exploiting a specific feature of the chemistry of iodine: the existence of the I_3^- ion and higher polyiodide ions I_n^-.

One of the earliest uses of iodine in labelling protein crystals was reported by Bluhm *et al.* (1958). They described the effect of I_3^- ions on myoglobin. Their aim was in fact to iodinate the tyrosine residues, but instead they observed binding of I_3^- in a hydrophobic cavity bounded on one side by the haem group. This site coincided with the observed position of a bound HgI_3^- group in myoglobin co-crystallized with K_2HgI_4 as reported by Bluhm *et al.* (1958) and Kretsinger *et al.* (1968). The latter authors were able to determine that HgI_3^- binds to metmyoglobin in two sites of differing character: an internal site, being the hydrophobic cavity near the haem group, and an external site, which was interacting through electrostatic forces. The observation of Schoenborn *et al.* (1965) that xenon was binding in the same position in myoglobin as one of the I atoms of the HgI_3^- molecule led them to suggest that the

binding of both Xe and I atoms in the internal site is mediated by dispersion forces, which are a consequence of the large polarizabilities of these atoms.

Iodine has been used in the derivatization of protein crystals in numerous forms. Among the most common forms of iodine compounds used are KI (Li *et al.*, 1997), KI/I_2 (Sigler, 1970), HgI_3^- (Kretsinger, 1968; Kretsinger *et al.*, 1968) and the replacement of thymine by 5-iododeoxyuracil (Nolte *et al.*, 1998). Iodine has also been used in crystallographic studies of peroxidases (Fukuyama *et al.*, 1995). The iodination of tyrosine residues in a crystal structure has been successfully attempted by Sigler (1970) and Ghosh *et al.* (1999). In the latter case, the presence of three almost fully occupied I atoms per 207 amino acids allowed the structure of acetylxylan esterase to be determined by SIRAS to 1.8 Å. An example of the potential of iodine as a significant source of AS on conventional Cu $K\alpha$ sources was provided by Chen *et al.* (1991), who used an iodinated dipeptide substrate to solve the structure of bovine neurophysin II with single-wavelength anomalous data. These various authors were only able to identify two cases in which iodine was bound as triiodide ions (I_3^-), namely in the crystal structures of *Arthromyces ramosus* peroxidase (Fukuyama *et al.*, 1995) and sperm-whale metmyoglobin (Kretsinger, 1968).

The following sections are devoted to the presentation of a series of experiments which assess the potential benefits of combinatorial counter-ion replacement and triiodide derivatization as phasing methods using conventional X-ray equipment.

4. Materials and methods

The work was carried out using five test proteins: hen egg-white lysozyme (HEWL), apoferritin from horse spleen (APO), xylose isomerase from *Streptomyces rubiginosus* (XI), elastase from porcine pancreas (PPE) and an antibody fragment (ABF) kindly supplied by Leo James of the MRC Centre for Protein Engineering. Hanging-drop vapour diffusion (McPherson, 1982) was used to crystallize all the proteins, apart from PPE which was crystallized using sitting drops.

4.1. Crystallization

HEWL was crystallized according to protocols developed by N. Duke of the Structural Biology Center, Argonne National Laboratory (personal communication). The drop consisted of 7.5 µl reservoir solution (2 *M* NaCl, 0.025 *M* CH_3COONa pH 4.7, 25% ethylene glycol) and 7.5 µl protein solution (100 mg ml^{-1} HEWL, 0.1 *M* CH_3COONa, 0.1% NaN_3). The drop was usually seeded with a dilute solution of fragments of a crystal grown in similar conditions. A crystal was broken up into seeds by vortexing it in 0.5 ml stabilizing solution (4 *M* NaCl, 0.1 *M* CH_3COONa) for 3 min in a 1.5 ml Eppendorf tube along with a 3 mm PTFE ball (Luft & DeTitta, 1999). Crystals of space group $P4_32_12$ (unit-cell parameters $a = b = 78.84$, $c = 36.81$ Å), approximately 0.2–0.3 mm in size, appeared within 24 h.

APO was purchased from Sigma Chemicals (A3641; Lot No. 88H70152) and crystallized using $CdSO_4$ as a crystallizing agent. The conditions described by Granier *et al.* (1997) were used as guidelines for the preparation. Drops were set up as a matrix of varying protein concentration (10–40 mg ml^{-1} protein in 0.1 *M* NaCl) and varying Cd^{2+} concentration [7–12 m*M* $CdSO_4$, 1 *M* $(NH_4)_2SO_4$, 3 m*M* NaN_3]. 10 µl drops consisting of equal volumes of protein solution and reservoir solution were suspended above 1 ml reservoirs. Small crystals belonging top space group *F*432 (unit-cell parameters $a = b = c = 181.74$ Å) and size <50 µm appeared within a few days but grew to 200 µm or larger over a period of weeks. The crystals were cryoprotected prior to mounting in the liquid-nitrogen gas stream by soaking in a solution of 1.25 *M* $(NH_4)_2SO_4$, 10 m*M* $CdSO_4$, 25% ethylene glycol.

XI was purchased from Hampton Research as a 33 mg ml^{-1} protein solution. This was further diluted to produce drops of 25 mg ml^{-1} protein, 0.9 *M* $(NH_4)_2SO_4$, 10 m*M* Tris–HCl pH 7.5. 10 µl drops were suspended above 1 ml of reservoir solution [1.8 *M* $(NH_4)_2SO_4$, 10 m*M* Tris–HCl pH 7.5]. Crystals of space group *I*222 (unit-cell parameters $a = 92.27$, $b = 98.46$, $c = 102.38$ Å) and size 0.15–0.25 mm appeared within a few days and were cryoprotected by soaking in a solution of 1.8 *M* $(NH_4)_2SO_4$, 10 m*M* Tris–HCl pH 7.5, 2 m*M* $MgCl_2$ and 1.86 *M* L-xylose. In this case the L-xylose acted as an effective cryoprotectant, while $MgCl_2$ was necessary in order to maintain the stability and integrity of the crystals (Carrell *et al.*, 1989).

PPE was purchased from SERVA GmbH, Heidelberg and crystallized by sitting-drop vapour diffusion using 5% PPE dissolved in 0.1 *M* CH_3COONa pH 5.0 and 0.02% NaN_3, with 10 m*M* CH_3COONa pH 5.0 and 10 m*M* Na_2SO_4 as precipitant. Crystals of space group $P2_12_12_1$ (unit-cell parameters $a = 50.15$, $b = 57.90$, $c = 74.56$ Å), having approximate dimensions $0.3 \times 0.1 \times 0.1$ mm, appeared within 48 h. The cryosolution contained 1 m*M* Na_2SO_4, 10 m*M* CH_3COONa pH 5.0 and 28%(*v*/*v*) glycerol.

ABF was crystallized in space group $P2_12_12_1$ (unit-cell parameters $a = 36.54$, $b = 70.74$, $c = 82.12$ Å) in 1.0 *M* $LiSO_4$, 0.5 *M* $(NH_4)_2SO_4$ and 0.1 *M* sodium citrate buffer pH 7.5. The crystals used were ~0.2 mm in each dimension.

4.2. HEWL counter-ion derivative preparation

Three HEWL derivatives were prepared using the short-soak method. In each case, the cryosoaking solution contained 2.5 *M* NaCl, 50 m*M* CH_3COONa and 25% ethylene glycol plus 0.5 *M* NaI, CsCl or CsI. Each crystal was soaked for 30 s in its respective cryosolution. The soak time was determined to be the upper limit after which time the integrity of the crystals, as judged by their appearance, was challenged by the heavy-atom salts. The presence of 2.5 *M* NaCl along with the other salts was found to be essential for maintaining good diffraction quality. This was also observed by Dauter & Dauter (2001) when using KI to soak HEWL and xylanase. In the case of HEWL, this is explained by the known involvement of Na^+ and Cl^- ions in maintaining the structural stability of the molecule and crystal lattice (Sauter *et al.*, 2001; Dauter *et al.*, 1999).

Table 1
Short-soak conditions for all the crystals used in the triiodide study.

Conc. KI/I_2 is expressed as a (*v*/*v*) percentage of the stock solution.

Data set	Solution	Soak time (min)
XI (KI/I_2)	1.8 *M* $(NH_4)_2SO_4$, 1.86 *M* L-xylose, 1% KI/I_2	8
ABF (KI/I_2)	$(NH_4)_2SO_4$, $LiSO_4$, 10% glycerol, 2% KI/I_2	5
HEWL (KI/I_2)	2.5 *M* NaCl, 50 m*M* CH_3COONa, 25% ethylene glycol, 8% KI/I_2	10
PPE (KI/I_2)	1 m*M* Na_2SO_4, 10 m*M* CH_3COONa, 30% glycerol, 10% KI/I_2	5
APO (KI/I_2)	1.25 M $(NH_4)_2SO_4$, 10 m*M* $CdSO_4$, 25% ethylene glycol, 2% KI/I_2	5

4.3. Triiodide-derivative preparation

For HEWL, APO, XI and PPE, the triiodide derivatives were prepared by cryosoaking the crystals in cryosolutions containing a small concentration of KI/I_2. A stock solution of KI/I_2 was prepared by dissolving 1 g KI in 4 ml of water and then adding 0.54 g I_2. This gave approximate final concentrations of 0.47 *M* I_2 and 0.67 *M* KI. Typically, this solution was then added to protein mother liquor or cryosoaking solution diluted 50–100 times such that iodine was present in ~5–10 m*M* quantities. The cryosoak drop size was typically 15–20 µl. The cryosoak derivatization conditions are given in Table 1.

For ABF, the KI/I_2 derivative was prepared by preparing a 10%(*v*/*v*) solution of KI/I_2 in glycerol and then adding 0.8 µl of this directly to a 4 µl drop containing crystals. The drop was left until a significant change from colourless crystal to dark yellow/brown had been observed; the crystal was then mounted directly in the Cryostream.

In all cases, the uptake of iodine from the drop into the crystal was clearly visible as a darkening of the crystal from clear to pale yellow through to brown, while the drop generally lightened in colour. This was a strong indicator of binding of iodine, in whatever form, to the macromolecules in the crystal as opposed to mere penetration by the iodine into the crystal's solvent channels. The crystal was generally left soaking until the colour change in the crystal was seen to stabilize or as long as the crystal could tolerate the solution without visibly starting to degrade. No back-soaking of the crystals to an iodine-free solution was attempted because no significant increase in background scatter was observed in the derivative crystal diffraction compared with that of the native. All samples were mounted directly from the cryoderivatization solution into the nitrogen-gas stream in preparation for data collection.

4.4. Data collection and analysis

All the experiments were performed on rotating Cu anode X-ray sources. Three experimental setups were used: (i) an Elliot GX-13 with Supper mirrors (Charles Supper Company, Natick, MA, USA) and MAR 30 cm IP, (ii) a Rigaku RU-200 fitted with Osmic multilayer optics and a MAR345 IP and (iii) a RigakuMSC RU-300/Osmic multilayer/AFC9 goniometer and Jupiter CCD detector. These three setups are respectively referred to as GXIP, RIP and RCC, respectively, throughout this paper.

Table 2
Computer programs used for data analysis and presentation.

Program name	Reference
ARP/wARP	Perrakis *et al.* (1999)
BUSTER	Bricogne (1997)
SHARP	de La Fortelle & Bricogne (1997)
autoSHARP	Vonrhein & Bricogne (2001)
SHELXD	Usón & Sheldrick (1999)
SOLOMON	Leslie (1987), Abrahams & Leslie (1996)
DM	Cowtan (1994)
TNT	Tronrud *et al.* (1987)
*d*TREK*	Pflugrath (1999)
MOSFLM	Leslie (1992)
*CCP*4 suite	Collaborative Computational Project, Number 4 (1994)
RANTAN	Yao (1981)
SCALA	Evans (1997)
XtalView	McRee (1999)
PROCHECK	Laskowski *et al.* (1993)
PyMOL	DeLano (2001)

Diffraction images were integrated using either *MOSFLM* or *d*TREK* and scaling was performed with *SCALA* from the *CCP*4 suite. The conversion from *d*TREK* integrated and profile-fitted reflection lists to a multi-record MTZ format file suitable for input into *SCALA* was performed with an in-house modified version of the *CCP*4 program *ABSURD*. The modifications allow all of the goniometric information to be carried through into the *SCALA* program and therefore permit full use of the spherical harmonic correction feature within *SCALA* to deal with systematic errors arising from sample absorption. Heavy-atom detection, heavy-atom refinement, phase calculation and solvent flattening were performed using the *autoSHARP* procedure, which makes use of *RANTAN*, *SHARP*, *SOLOMON* and the *CCP*4 suite. Literature references for the computer programs used are given in Table 2.

For the combinatorial counter-ion experiments on HEWL, highly complete and redundant diffraction data were measured from each sample between ~20 and 1.9 Å resolution. For the native crystal, a complete set of diffracted intensities from 55.8 to 1.9 Å was measured. The results of data analysis with *MOSFLM* and *SCALA* are summarized in Table 3.

For the triiodide derivatives, diffraction data measured from native and KI/I_2-soaked derivatives of XI, ABF, HEWL, PPE and APO were measured using conventional X-ray apparatus. The data were analysed as described in §4.4 and the results are given in Table 4.

5. Phase-determination results

5.1. Phasing by combinatorial counter-ion replacement

In order to verify initially that the data were of sufficient quality to allow *de novo* structure determination, the native and three derivative sets were input into *autoSHARP*. The

procedure determined a total of 28 anomalous scattering sites: 12 for NaI, 18 for CsCl and seven for CsI. A number of sites were common to the three derivatives. After determination of the correct hand and solvent flattening, *ARP*/*wARP* was able to build and dock into the sequence 123 residues out of 129 in three peptide chains. The r.m.s. deviation in atomic coordinates between the *ARP*/*wARP*-built structure and the refined native structure [see HEWL (NaCl) in Table 5 for the results of this refinement] was 0.95 Å, illustrating that the model was essentially complete and correctly interpreted.

To further simplify tests of the phasing strategy and allow better conditioning of the *SHARP* input, the heavy-atom positions for the phasing tests were taken from ADF maps calculated using calculated phases from a model refined against the native data set. It was felt that the use of sites from ADF maps was justified in this case to circumvent inessential complications such as differing origin choices in the various *autoSHARP* runs. This also enabled us to more easily identify the species of individual anomalous scatterers. From the ADF maps a total of 36 sites were identified: ten S and 17 I sites for NaI, ten S, nine Cl and six Cs sites for CsCl and ten S, 17 I and four Cs sites for CsI.

The electron-density map calculated from phases determined by *SHARP* (before solvent flattening) with all four data sets included was of a very high quality, having a phased correlation coefficient (PCC) of 0.796 (mean phase difference, $\langle\Delta\Phi\rangle = 26.0^\circ$) calculated against phases from the refined native structure. An essentially complete model (125/129 residues in two chains) was built into this experimental *SHARP* map and docked into the known sequence using the *warpNtrace* mode of *ARP*/*wARP*. The PCC after solvent flattening with *SOLOMON* was 0.891 ($\langle\Delta\Phi\rangle = 17.8^\circ$). Fig. 1 shows a region of the HEWL refined model along with the electron-density map calculated from *SHARP* experimental phases and from phases after solvent flattening.

Table 3
Data integration and scaling statistics for HEWL combinatorial phasing experiment.

Values in parentheses refer to the highest resolution shell. All data sets were measured using the GXIP setup.

Data set	Resolution (Å)	Multiplicity	Completeness (%)	$R_{\rm merge}$	$I/\sigma(I)$
HEWL (NaCl)	55.78–1.90 (2.00–1.90)	13.2 (11.6)	100.0 (100.0)	3.2 (6.9)	16.2 (9.8)
HEWL (NaI)	19.15–1.89 (2.00–1.89)	8.0 (7.1)	98.3 (94.6)	3.9 (8.4)	14.3 (8.1)
HEWL (CsCl)	19.74–1.90 (2.00–1.90)	14.6 (14.8)	99.9 (100.0)	3.2 (5.6)	14.8 (11.9)
HEWL (CsI)	22.22–1.90 (2.00–1.90)	25.7 (23.6)	99.9 (100.0)	5.6 (11.3)	9.8 (5.9)

Table 4
Data integration and scaling statistics for XI, ABF, APO and HEWL derivative data sets and PPE native and derivative sets.

Also given are mean anomalous difference ratios ($\langle|\Delta F|/F\rangle$) for the KI/$I_2$-soaked data sets. Values in parentheses refer to the highest resolution shell.

Data set	Apparatus	Resolution (Å)	Multiplicity	Completeness (%)	$R_{\rm merge}$	$I/\sigma(I)$	$\langle|\Delta F|/F\rangle$
XI (KI/I_2)	RIP	32.11–1.55 (1.63–1.55)	18.8 (13.0)	98.6 (96.3)	6.5 (26.4)	8.4 (2.7)	0.074 (0.126)
ABF (KI/I_2)	RIP	27.32–1.86 (1.96–1.86)	16.9 (12.3)	99.7 (98.3)	4.2 (9.7)	12.8 (6.7)	0.041 (0.070)
PPE (native)	RCC	29.88–1.63 (1.74–1.63)	3.0 (1.2)	77.7 (17.7)	5.7 (20.9)	10.3 (1.9)	n/a
PPE (KI/I_2)	RCC	26.68–1.85 (1.95–1.85)	12.5 (7.4)	99.2 (95.0)	10.5 (34.4)	6.6 (2.1)	0.106 (0.181)
APO (KI/I_2)	GXIP	22.14–2.01 (2.12–2.01)	9.1 (6.2)	98.9 (93.0)	7.3 (32.1)	10.1 (2.4)	0.093 (0.181)
HEWL (KI/I_2)	RIP	15.09–1.77 (1.87–1.77)	12.1 (11.7)	99.8 (100.0)	2.1 (4.6)	22.7 (15.5)	0.033 (0.047)

Table 5
Refinement statistics from *BUSTER*/*TNT* for the KI/I_2-soaked structures and the native HEWL structure.

Ramachandran plot statistics from *PROCHECK*.

	XI	ABF	PPE	APO	HEWL	HEWL (NaCl)
Protein atoms	3050	1743	1836	1398	1001	1001
I atoms	29	13	23	9	5	0
Solvent waters	347	139	180	145	99	152
Resolution range (Å)	32.05–1.55	27.34–1.86	26.68–1.85	22.14–2.01	15.09–1.77	55.78–1.9
No. reflections	66821	18574	19184	17272	11832	9595
Overall completeness (%)	98.6	99.7	99.2	98.9	99.9	99.8
$R_{\rm work}$	16.4	18.5	16.9	18.0	17.2	16.1
$R_{\rm free}$ (5% data omitted)	18.8	22.8	20.5	20.5	20.8	21.2
Geometry: r.m.s. deviations in						
Bond lengths (Å)	0.018	0.018	0.019	0.019	0.017	0.019
Bond angles (°)	1.619	1.386	1.663	1.283	1.353	1.350
Ramachandran plot: % residues in						
Most favoured regions	92.1	91.1	85.9	94.7	86.7	89.4
Additional allowed regions	7.3	8.4	13.6	5.3	13.3	10.6
Generously allowed regions	0.6	0.0	0.5	0.0	0.0	0.0
Disallowed regions	0.0	0.5	0.0	0.0	0.0	0.0

5.2. Triiodide as a heavy-atom derivative

For each of the five test samples, *de novo* phase determination was attempted by SAD and SIRAS methods using the *autoSHARP* procedure to first detect the positions of bound I atoms and subsequently determine phases. Where SAD was not successful, SIRAS was used. It should be emphasized that no supplementary phase information was used at any stage.

Phase determination for KI/I_2-soaked data sets was performed using the *autoSHARP* protocol whereby the integrated and scaled diffraction amplitudes were input as either a SAD or SIRAS data set. The protocol automatically prepared the data files, performed outlier rejection, calculated Patterson maps, determined heavy-atom sites using *RANTAN* and launched *SHARP* jobs. Based on *SHARP* residual map features, heavy-atom sites were either rejected or added to the model and further jobs were launched. After solvent flattening using *SOLOMON* and/or *DM* the final map was subjected to analysis by *ARP*/*wARP* for automatic chain tracing and model

Table 6
Isomorphous and anomalous phasing power (PP) from *SHARP* and phased map correlation coefficients (PCC) calculated with triiodide-derived phases directly from *SHARP* and after applying *SOLOMON*.

The mean absolute phase difference ($\Delta\varphi$) is also given in each case. The *ARP*/*wARP* results are given as $n/m(c)$: the number of correctly built residues, n out of a total of m, that *ARP*/*wARP* was able to auto-build and dock into the amino-acid sequence using c separate chains if breaks occurred. Also given are the r.m.s. deviations between coordinates built and refined by *ARP*/*wARP* and those refined with *BUSTER*/*TNT*.

		PP-centric	PP-acentric		*SHARP*		*SOLOMON*		*ARP*/*wARP*	
Crystal	Resolution (Å)	Iso	Iso	Ano	PCC	$\Delta\varphi$ (°)	PCC	$\Delta\varphi$ (°)	No. residues	R.m.s. difference (Å)
XI	1.55	—	—	1.69	0.56	42.6	0.91	15.5	376/388 (3)	0.76
ABF	1.86	—	—	2.10	0.51	45.8	0.76	29.5	210/236 (5)	0.66
PPE	1.71	0.99	1.29	1.23	0.50	45.3	0.62	41.9	219/240 (6)	0.87

building. The results of these steps are summarized in Table 6 and the steps are described in more detail below.

In each case, the success of the *autoSHARP* procedure was based on its ability to build correct portions of the model into the final solvent-flattened map. The correctness of heavy-atom site determination by *RANTAN* was judged on the ability of *SHARP* to refine the heavy-atom site parameters to sensible and physical values, *i.e.* *B* factors not significantly greater that the average protein temperature factor and substantial fractional occupancies.

5.2.1. XI. The scaled structure factors from the KI/I_2-soaked XI crystal were fed into *autoSHARP* as a SAD data set. A correct heavy-atom structure solution was found only when the data was truncated from 1.55 to 2.0 Å. In this way, ten heavy-atom sites were found and used as input to *SHARP*. Automatic interpretation of *SHARP* residual maps rejected two sites and identified 17 new sites. *SHARP* jobs using 25 sites were then run using both possible hands. Solvent flattening using *SOLOMON* as implemented in *SHARP* was run on experimental electron-density maps from each hand. The correlation coefficients on E^2 clearly indicated the correct solution and the solvent-flattened map was finally used with *warpNtrace* (*ARP*/*wARP*) to automatically build the structure to 2 Å using the known amino-acid sequence.

Using the 25 sites found by *autoSHARP*, *SHARP* was run with all data between 32 and 1.55 Å. The *SHARP* residual maps identified 11 additional sites. One of the new sites was positioned directly between two existing sites at about 2 Å from each. This cluster of three was interpreted as a triiodide molecule. Solvent flattening of the resulting experimental phases resulted in a map into which *ARP*/*wARP* traced 376 out of 388 residues in three sections running from 4–186, 189–243 and 247–386. The Ramachandran plot showed only five non-glycine residues lying outside the allowed regions.

5.2.2. ABF. All data between 26.82 and 1.86 Å resolution were used as input to *autoSHARP*. Initially, *RANTAN* found 12 heavy-atom sites which were used in a first round of heavy-atom refinement in *SHARP*. Automatic inspection of the residual maps from *SHARP* resulted in the rejection of five sites and the inclusion of four new sites. A second *SHARP* refinement using these 11 sites was followed by solvent-flattening optimization, which determined the correct hand of the heavy-atom solution. The solvent-flattened electron-density map allowed *ARP*/*wARP* to build 142 residues out of a total of 236 in 16 separate chains.

At this stage, the *SHARP* residual maps were inspected by hand. After adding new sites and rejecting sites with very low occupancy or very high *B* factors (>100 $Å^2$), the total number of sites was brought to 20. Although no new significant features appeared in the residual maps, it was possible to identify two of the existing anomalous scattering sites as S atoms involved in a disulfide bridge. The atoms were 2.01 Å apart. Their refined occupancies (*B* factors) of 0.065 (9.8 $Å^2$) and 0.098 (22.7 $Å^2$) supported this conclusion. The sites were therefore included as S atoms and further heavy-atom refinement was performed.

A final experimental map produced using a total of 20 heavy-atom sites in *SHARP* was input to *SOLOMON* and then *ARP*/*wARP*. A model consisting of 210 residues in five separate chains was automatically built and docked into the protein sequence. The final *ARP*/*wARP* conventional and free *R* factors obtained with *REFMAC* were 20.1 and 24.6%, respectively.

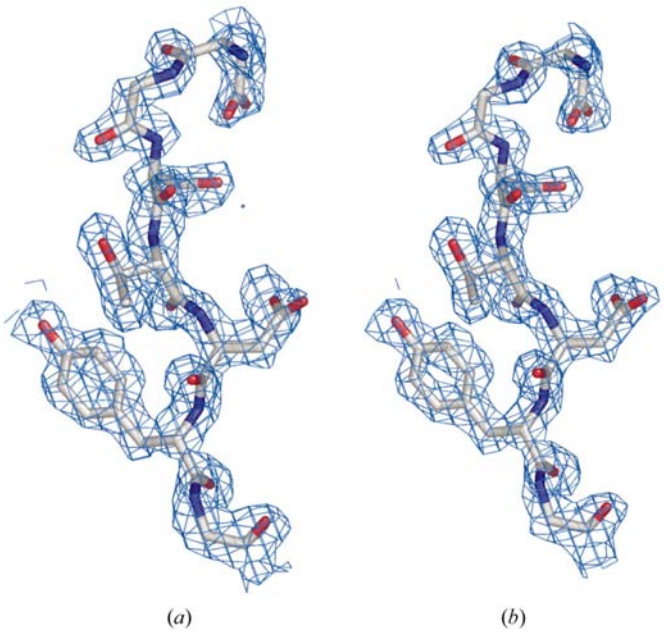

Figure 1
HEWL electron-density map contoured at the 1σ level calculated using phases (*a*) from *SHARP* and (*b*) after solvent flattening with *SOLOMON*. The region from Asp48 to Gly54 is shown. (All molecular figures were generated using *PyMOL*.)

5.2.3. PPE. It was necessary in this case to use a native data set of PPE along with data from the KI/I_2-soaked crystal to perform SIRAS phase determination. Again, the *autoSHARP* protocol was used with native data extending to 1.63 Å and derivative data to 1.85 Å. *RANTAN* initially found two heavy-atom sites using the anomalous differences for the derivative data and, after an initial *SHARP* run, three additional sites were added to the heavy-atom model based on residual map peaks. The hand of the heavy-atom model was correctly selected on the basis of E^2 correlation coefficients, but the resulting map was not of sufficient quality to allow automatic building with *ARP*/*wARP*. Manual inspection of the residual maps showed evidence of anisotropy in the three strongest iodine sites. Thus, an additional round of heavy-atom refinement using anisotropic B values was performed, after which two new sites were identified in the isomorphous residual map. Solvent flattening with these phases produced an electron-density map into which *ARP*/*wARP* was able to build 219/240 residues as six polypeptide chains.

6. Crystallographic model refinement

In order to determine accurate positions and occupancies of bound I atoms, the structures of the various KI/I_2 derivative crystals were refined using the *BUSTER*/*TNT* program. The following general protocol was used in each case. After initial rounds of rigid-body refinement to position the molecule correctly, successive rounds of positional and B-factor refinement were performed. In general, the occupancies of all atoms except iodine were fixed to unity and all B factors were refined. The iodides and polyiodides were refined as independent atoms without the use of I—I bond-distance restraints. After each run, water molecules were added to the model in $F_o - F_c$ electron-density peaks greater than 4σ until all clearly identifiable and chemically sensible water molecules had been placed.

At each stage of refinement, the current phase set was used to calculate anomalous difference Fourier (ADF) maps. Peaks in the ADF maps greater than 4σ were classified as I atoms and included in the model. The results of all five refinements are summarized in Table 5.

6.1. XI KI/I_2

The protein atom coordinates of the 1.6 Å refined structure of Carrell *et al.* (1989) retrieved from the PDB (code 1xib) were used to provide an initial model. A total of 347 water molecules were added to this model and two L-xylose molecules were also identified in the electron density. The first of these was bound in the active site in the position where D-xylose is thought to bind (Whitaker *et al.*, 1995). The second xylose molecule was found bound to the surface of the protein, making hydrogen-bond interactions from O3 to Asp295 OD2 and from O1 to Ala332 O. The O2 atom is hydrogen bonded to Wat126 and Wat802. The final model had an $R_{\rm free}$ of 18.8% and an $R_{\rm work}$ of 16.4%.

Given the crystallization conditions, the only significant anomalous scattering signal was expected to come from either S atoms [f''(S) = 0.558 e] in Cys and Met residues and perhaps from solvent sulfate ions, from potassium cations [f''(K) = 1.07 e] or from I atoms [f''(I) = 6.843 e]. It was therefore relatively straightforward to identify I atoms owing to their large f'' value. In total, 29 I atoms were found as two clusters of five atoms each, one four-atom cluster, one three-atom cluster, two pairs of atoms (probably mutually exclusive) and eight single sites. The occupancies of the iodine sites ranged from 0.61 to 0.21. The average B factor for the 29 I atoms was 29 Å^2.

Analysis of the *autoSHARP* heavy-atom positions compared with the refined positions of anomalous scattering centres showed that 31 sites agreed with an r.m.s. deviation of 0.37 Å. The positions corresponded to 24 I atoms, five S atoms in Met and Cys residues and two active-site Ca^{2+} ions.

Table 7
Bond distances and angles for all putative polyiodide groups.

Distances (Å)		Angles (°)	
XI (KI/I_2)			
11—12	2.27	11—12—13	163.3
12—13	2.67	12—13—14	144.6
13—14	2.36		
21—22	2.76	21—22—23	165.7
22—23	3.26	22—23—24	120.2
23—24	2.56	23—24—25	127.8
24—25	3.06		
31—32	2.49	31—32—33	169.2
32—33	3.66	32—33—34	88.2
33—34	3.54		
32—35	2.92	31—32—35	82.4
41—42	2.90		
91—92	3.09	91—92—93	175.7
92—93	2.62		
APO (I_2)			
8—9	3.06	8—9—10	130.2
9—10	2.94		
9—11		8—9—11	73.5
ABF (KI/I_2)			
11—12	2.71	11—12—13	166.0
12—13	2.85		
21—22	2.84	21—22—23	158.8
22—23	3.48		
HEWL (KI/I_2)			
11—12	2.73	12—11—12†	176.3
12†—11	2.73		
PPE (KI/I_2)			
11—12	2.55	11—12—13	143.3
12—13	3.20		
12—14	3.11	11—12—14	153.6
21—22	2.72	21—22—23	174.4
22—23	3.21		
31—32	2.91	31—32—33	172.2
32—33	3.06		
41—42	2.92	41—42—43	174.4
42—43	2.87		
51—52	2.68	51—52—53	170.8
52—53	2.80		
61—62	2.73		
71—72	3.33		

† Symmetry related atoms.

6.2. ABF KI/I_2

The starting model for the analysis was the refined native structure of the antibody fragment (Leo James, personal

communication). A total of 13 iodine sites were identified using ADF maps calculated with model phases. The final model had an R_{free} of 22.8% and an R_{work} of 18.5%. The Ramachandran plot showed one residue, ThrL53, in the disallowed region. However on inspection, the residue was modelled in very clear electron density but was positioned at the end of a tight turn in the main chain.

Iodine was bound as two triiodide ions and seven single iodide ions. The triiodides lie very close to one another and are bound through hydrophobic interactions with GluL40, LysL41, HisL44, ProL42, GluL85, AlaL86, IleL87 and LysL105 from the light chain and ProH75 from the heavy chain. Both triiodides are unsymmetrical and deviate from a linear arrangement (see Table 7). The iodine occupancies range from 0.23 to 0.41 and the average B factor is 32.2 Å^2.

Analysis of the *autoSHARP* heavy-atom positions showed that all of the 13 refined iodine sites had been found as well as all seven S atoms in Met and Cys residues. The r.m.s. deviation between the *autoSHARP* and *BUSTER*/*TNT* refined positions was 0.26 Å.

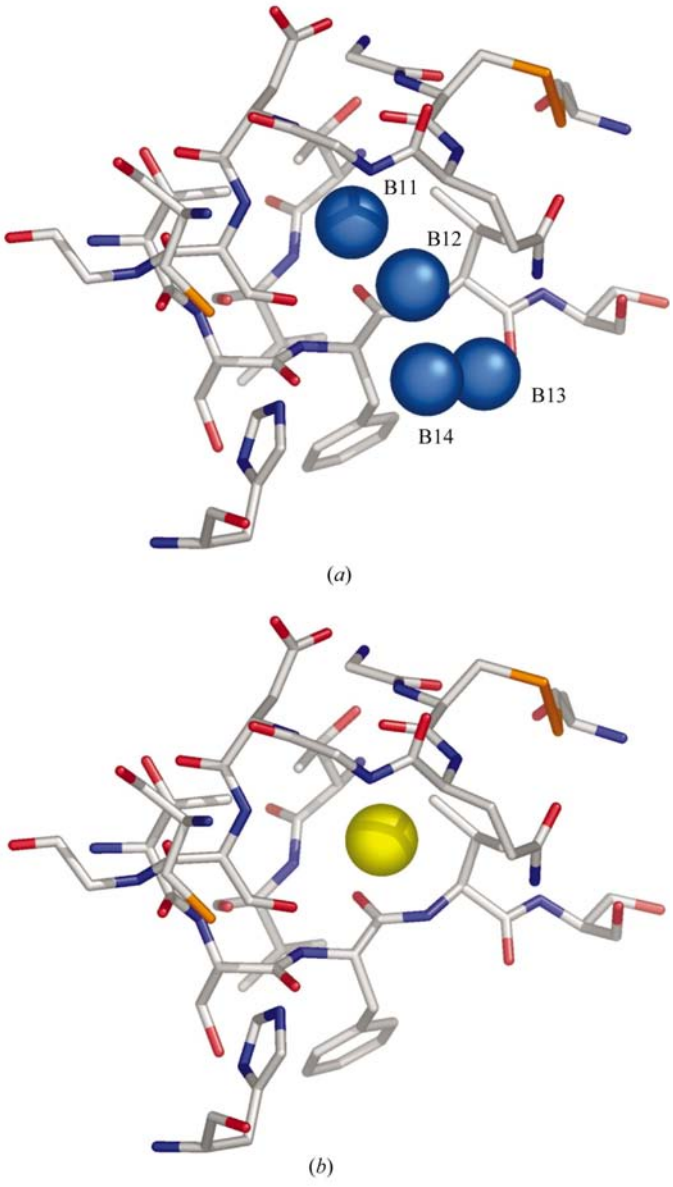

Figure 2
Region of the elastase molecule showing the common binding site for (*a*) iodine and (*b*) xenon.

6.3. HEWL KI/I_2

After the addition of 99 waters to the starting model (Evans *et al.*, 2002), five iodine sites were identified. Sites B11 and B12 are the central and outer I atoms of an I_3^- ion which sits across a special position such that the outer atoms of the triiodide are related by symmetry. The central atom, B11, has fractional coordinates −0.201, 0.201, 0.250 and has a refined occupancy of 0.23 which, because of its twofold nature, corresponds to a physical occupancy of 0.46 shared between two HEWL molecules. The triiodide sits between two molecules, making contacts to the hydrophobic parts of Lys13 and Arg14 and to the backbone carboxyl group of Ala10 and Leu129. The third iodine site, B21, is identical to a known chlorine-binding site in HEWL. It was identified as iodine on the basis of its refined occupancy and of the height of its ADF peak relative to other S and Cl atoms. Sites B31 and B41 are also single iodide sites but have low occupancies of ∼0.20. The final model incorporating three ethylene glycol molecules, seven Cl atoms and three Na atoms had a R_{free} of 20.8% and an R_{work} of 17.2%.

6.4. PPE KI/I_2

A refined model of PPE retrieved from the PDB (code 1c1m; Prangé *et al.*, 1998) was used as a starting model for refinement against the KI/I_2-soaked PPE data. The final model had an R_{free} of 20.5% and an R_{work} of 16.9%. A total of four triiodide molecules were identified in the structure, along with three more sites where pairs of I atoms were found together as indicated by a very weak signal in the anomalous difference Fourier map. One pair of iodine positions was found close to and roughly collinear with the position of a bound sulfate ion from the native structure. The ADF peak height at the sulfate position was roughly twice that of other cysteinyl S atoms in the structure and its refined B factor was approximately double. It was likely therefore that an I atom occupied the position of the 'native' sulfate ion.

The triiodide molecule with the highest refined occupancy was in the catalytic site of the proteinase. In fact, four distinct peaks were observed (B11–B14) with refined occupancies of 0.87, 0.84, 0.50 and 0.60. Site B11 was only 0.83 Å away from the Xe site in PPE identified by Schiltz *et al.* (1995). Fig. 2 shows the PPE catalytic site with Xe and triiodide bound. The sum of the occupancies of B13 and B14 is higher than the individual occupancies of B11 and B22, which suggests that these two atoms are both present together in the structure; thus, the four-atom cluster does not represent two conformations of a triiodide molecule.

Retrospective analysis of the *autoSHARP* heavy-atom positions for PPE showed that only seven of the 23 final

refined iodine sites had been used in the phase determination. Their positions agreed with an r.m.s. deviation of 0.56 Å. The relatively low number of heavy-atom positions found by *autoSHARP* reflects the on average lower quality of the PPE native and derivative data sets, as indicated by their merging statistics in Table 4. This may also explain the dependence of the phase determination on the isomorphous signal derived from inclusion of the native data.

6.5. APO KI/I_2

Model refinement against the APO-KI/I_2 data led to the identification of nine bound I atoms and produced a final R_{free} of 20.5% and R_{work} of 18.0%. Iodine site B11 is bound with an occupancy factor of 0.64 to the first Cd^{2+} ion (B101), which is responsible for stabilizing the intermolecular crystal contacts through Asp80. In fact, two symmetry-related iodines bind a single Cd^{2+} ion forming coordination bonds. The bond length of 2.79 Å is in excellent agreement to within ±0.005 Å with the average Cd—I distance calculated from a variety of structures containing Cd—I bonds (Svensson *et al.*, 1998), whereas the bond angle of 95.05° is at the lower end of the range of observed I—Cd—I bond angles (95.44–121.82°). The coordination of the Cd atom by two Asp80 side chains and two B11 I atoms is approximately tetrahedral. The iodine binding did not hinder the ability of Cd atom to stabilize the crystal contacts and in fact appeared to have a positive effect on the crystal stability as witnessed by an increase in the maximum resolution obtained with KI/I_2-soaked crystals (2.01 Å) compared with that for the native crystals (2.6 Å). A reduction in the unit-cell parameters of 0.56 Å (0.31%) on addition of KI/I_2 suggests that the molecules pack together more tightly and this could be responsible for the higher resolution diffraction.

Site B21 is hydrogen bonded to Leu93 N in a small pocket on the outer surface of the spherical shell of the apoferritin molecule. B31 also sits on a twofold axis between two molecular subunits and interacts with Leu81 and two symmetry-related I atoms B32 which are 3.07 Å away. The B32–B31–B32 triplet represent a triiodide molecule. B41 is 8.25 Å away from the second Cd^{2+} site (B102) near the outer entrance to the narrow threefold channel. The site is 4.88 Å from the threefold axis. Finally, a group of four weak iodine sites B51–B54 were observed in a hydrophobic pocket on the inner surface of the apoferritin spherical shell.

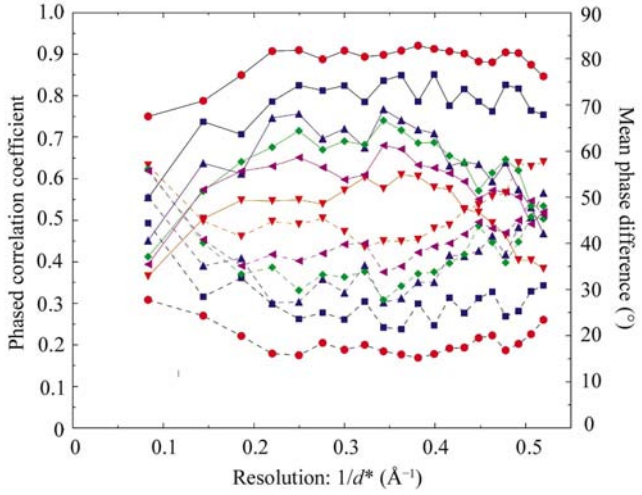

Figure 3
Plot of phased correlation coefficient (solid) and mean phase error (dashed) against resolution for HEWL combinatorial counter-ion replacement (CCR) phases calculated directly from *SHARP* (blue squares) and after solvent flattening with *SOLOMON* (red circles). Also shown are results for *SHARP* experimental phases calculated for SIRAS cases using NaI (purple triangles), CsCl (orange inverted triangles) and CsI (green diamonds). Results from SIRAS phasing using CsCl and native and CsI as derivative are shown as magenta left-pointing triangles.

7. Discussion

7.1. Combinatorial counter-ion replacement

The benefits of using combinations of counter-ions in phasing HEWL as opposed to using conventional SIRAS were investigated by examining quantitative measures of phasing power which are presented in Fig. 3.

Heavy-atom refinement and phase calculation were performed for the conventional SIRAS approach using four pairs of HEWL data sets: NaCl–NaI, NaCl–CsCl, NaCl–CsI and CsCl–CsI. These SIRAS results were compared with those obtained using all four data sets simultaneously in a combinatorial phasing experiment. The best SIRAS phases are produced by NaCl–NaI (PCC = 0.676) where only iodine provides isomorphous signal. The results from the NaCl–CsI pair (PCC = 0.657) are similar and suggest that the Cs atoms contribute little to the phasing. Indeed, the NaCl–CsCl pair, where isomorphous differences arise from Cs alone, provides the least phase information (PCC = 0.536). Phasing from CsCl–CsI, where again iodine is expected to contribute most to isomorphous differences, gives a PCC of 0.605.

Although the iodine is clearly the main source of phase information, a significant gain in the quality of the phases is observed when all four data sets are combined for heavy-atom refinement and phasing. The average PCC for experimental phases increases from 0.676 (the best SIRAS result) to 0.796 when the full combinatorial design is used. The improvement is more marked as a function of resolution (see Fig. 3). Beyond 2.5 Å (0.4 $Å^{-1}$), the increase in PCC for the combinatorial case over individual SIRAS results is between 0.1 and 0.3.

The combined use of a few counter-ions species compared with a single species offers the benefits of improved phasing power and potentially better experimental phases to pass on to density-modification schemes.

7.2. Structure and binding of polyiodide ions to protein molecules

It has long been thought that solutions of molecular iodine dissolved in KI contain polyiodide species such as I_3^-, I_5^- *etc*. However recent evidence suggests that I^-, I_3^-, I_2 and possibly other protonated polyiodide complexes such as I_2H^+ are the dominant species (Calabrese & Khan, 2000). In solution,

triiodide ions are, on average, symmetric and linear. However, in crystalline environments I_3^- may be unsymmetrical and may deviate slightly from being linear. In fact the two I—I bond distances may vary by as much as 0.3 Å. Iodine's van der Waals radius is 2.15 Å, implying that I—I distances of as much as 3.5 Å may still represent real bonds (Cotton & Wilkinson, 1988). Furthermore, in crystalline environments iodine may form longer chain ions such as I_5^-, I_8^{2-} and I_{16}^{4-} (see Greenwood & Earnshaw, 1997 for a description of these structures). Typically, these polyiodide species have the structures I—I···I···I—I···I··· *etc*., where the solid lines represent shorter bond lengths of <3.0 Å on average and ··· represents longer bond lengths of up to 3.5 Å. The upper limit on bond length is not well defined, however. Bond-angle values congregate around 180 ± 15° for I—I···I, but may deviate significantly from 180° for I···I···I, often approaching right angles.

The KI/I_2-soaked protein structures presented here show evidence for the formation of polyiodide chains which bind both on the surface of protein molecules and in intermolecular cavities with triiodide being the predominant species. Table 7 lists I—I bond distances and I—I—I bond angles for all putative polyiodide species found in the five crystal structures. The longest polyiodide observed here was five atoms long and bound to XI. The large variation of bond length and the alternating nature of shorter and longer bonds observed in smaller molecule structures of polyiodides is in general also observed here. The shortest bond length observed is 2.27 Å and the longest is 3.66 Å. The former bond length, however, is much shorter than the smallest bond length of 2.67 Å observed for molecular iodine, suggesting that in this particular case two alternate positions of an I atom are observed. Indeed, there are six cases where the refined I—I bond length is <2.67 Å. These small discrepancies are most likely a consequence of the absence of restraints on I—I bond distances, which result in larger positional errors for I atoms.

Table 8
Classification of the bound I atoms in KI/I_2-soaked structures based on definitions and terminology used by Prangé *et al.* (1998) in their analysis of Xe- and Kr-binding sites.

Where an atom of a given residue is specifically named, a hydrogen-bonding interaction is being described. († indicates residues from a symmetry-equivalent molecule.)

Iodine site(s)	Interaction	Residues involved	Classification
XI			
B11–B14	Hydrophobic	Arg10, Thr12, His49, Lys85, Pro87, Pro283, Ser281, Gly282	Surface pocket
B21–B25	Hydrophobic and H-bond	Gln234 NE2, Leu200, Trp237, Ala238, Ala201†, Phe202†, Arg205†, Glu204, Ile203, Lys240 NZ	Deep intermolecular pocket
B31–B35	Hydrophobic	B33–Trp137 N, B35–Phe104 N	Intramolecular cavity
B41–B42	Hydrophobic	Asp24, Pro25, Tyr254, Pro25, Glu186	Intermolecular cavity
B61	H-bond	Leu241 N	Surface pocket
B71/B72	Hydrophobic	π interaction to His382	Surface pocket
B81	Hydrophobic	Trp16, His54, Phe94, Wat167, Wat198	Active-site channel
B91–B93	Hydrophobic	Gln41, Arg42, Glu45, Glu38	Surface pocket
B101	H-bond	Leu193 N	Intermolecular cavity
B111	H-bond	Wat68	Intramolecular cavity
B121	H-bond	Gly263 N, Arg266 NH2, Asn250 ND2	Intermolecular cavity
B131	Hydrophobic	π interaction to His71	Surface pocket
B141	H-bond	Asn227 ND2	Intermolecular cavity
B151	H-bond	Wat233	
PPE			
B11–B14	Hydrophobic and H-bond	Ser195 OG, Gln192 NE2	Catalytic site (see Fig. 2)
B21–B23	Hydrophobic and H-bond	Asn115, Ser116, Gln119, Wat153	Surface pocket
B31–B33	Hydrophobic	Ser217, Arg217A, Leu218, Asp60†, Arg61†, Glu62†, Val59†, Val90†	Intermolecular cavity
B41–B43	H-bond	Ser244 OG, Asn245 ND2	Intermolecular pocket
B51–B53	Hydrophobic	Trp38, Thr64, Phe65, Gly84, Asn95†, Asp97†, Asp98†, Arg65A, Tyr82	Intermolecular pocket
B61–B62	Hydrophobic	Leu143, Leu151, His40, Thr41, Gly193	Surface pocket
B71–B72	H-bond	Wat117, Asn178 ND2	Surface pocket
B81	H-bond	Arg230 NH2	Surface pocket (SO_4^{2-} site)
B91	H-bond	Asn74 ND2	Surface pocket
B101	H-bond	Ser29 OG	Surface pocket
APO			
B11	Bound to Cd^{2+}		Intermolecular cavity
B21	H-bond	Leu93 N, Thr91, Thr92, Ile151, Glu163	Surface pocket
B31–B32	Hydrophobic and H-bond	Leu81, Ser27, Tyr28, Leu31	Intermolecular cavity
B41	Hydrophobic	Gly117, Ser118, Ala121, Asp122, Pro123, Cys126, His114†	Surface pocket
B51–B54	Hydrophobic	B8–Cys48 SG	Surface pocket
ABF			
B11–B13	Hydrophobic	GluL40, LysL41, HisL44, ProH75†, ProL42, IleL87, GluL85, AlaL86, LysL105	Intermolecular cavity
B21–B23	Hydrophobic	HisL44, GlnH39, ProH41, ProH75†, HisL44, GlnH39, ValH93, AlaH92, AlaL28†	Intermolecular cavity
B31	H-bond	TrpH47 N	Surface pocket
B41	H-bond	SerL25 OG	Intermolecular surface pocket
B51	H-bond	TyrH60 N	Surface pocket
B61	H-bond	SerH17 OG	Surface pocket
B71	H-bond	TrpH100 NE1	Surface pocket
B81	H-bond	PheH107 N	Intramolecular cavity
B91	H-bond	PheH105 N	Surface pocket
HEWL			
B11–B12	Hydrophobic	Lys13, Leu129, Ala10, Arg14	Intermolecular cavity
B21	H-bond	Asn65 ND2	Intermolecular pocket
B31	H-bond	Leu25 N	Surface
B41	H-bond	Asn74 ND2	Surface

The distribution of observed bond lengths peaks between 2.8 and 2.9 Å and the average bond length is 2.89 Å, which is in good agreement with average I—I bond lengths observed in other small triiodide complexes formed with the cations $AsPh_4^+$, $(PhCONH_2)_2H^+$, NEt_4^+, Cs^+ and NH_4^+ and with recent XAFS results on the triiodide ion in ten different solvents,

where an average bond length of 2.915 Å was observed (Sakane *et al.*, 2001).

Our results show that I atoms bind predominantly to hydrophobic pockets as triiodide molecules or by hydrogen-bonding interactions with main-chain and side-chain atoms. Table 8 lists the polyiodide- and iodide-binding sites for the five structures and classifies their interaction with the macromolecule as hydrophobic and/or hydrogen bonding at surface pockets (intermolecular pockets if formed by one or more molecules) and in inter- or intramolecular cavities.

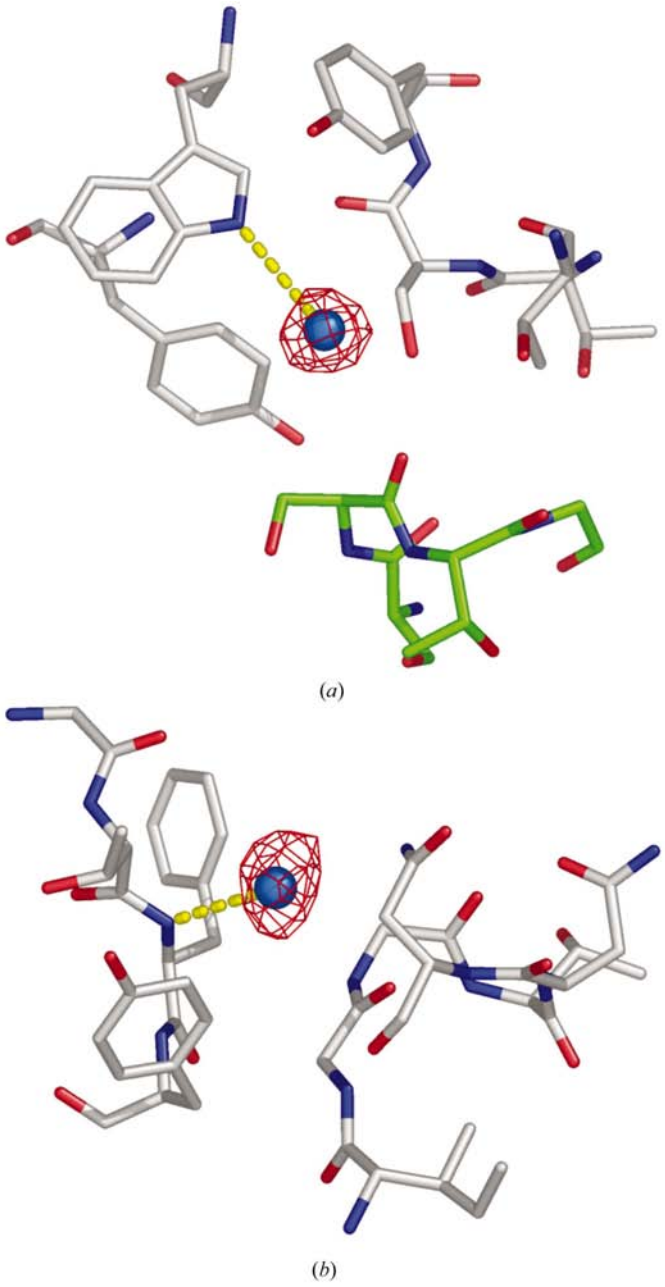

Figure 4
Hydrogen bond (yellow dashes) in ABF between (*a*) iodide B71 and TrpH100 NE1 and (*b*) iodide B91 and the main-chain N atom of PheH105. The anomalous difference Fourier map contoured at 4σ is shown as red mesh.

Hydrophobic binding is generally observed for polyiodide molecules and is characterized by van der Waals interactions at distances greater than 3.8 Å on average with mostly C^{α} atoms of the main chain, side-chain C atoms and less often with main-chain carbonyl groups.

In some cases, apparent water–iodine interactions are visible but difficult to interpret because the iodine sites are only partially occupied. Thus, water molecules that, by virtue of their proximity, could form electrostatic interactions with iodine may themselves exclude the presence of iodine. Without higher resolution data allowing a better determination of the occupancies of water molecules and better interpretation of the partially occupied iodine/water sites, it is difficult to identify conclusively the extent of water–iodine interactions except for a few cases where the local order of the molecules allows a clear interpretation.

Iodide ions interact predominantly through hydrogen bonds with the main-chain amide groups of Trp, Tyr, Phe, Leu, Gly, Arg and Asn residues. Hydrogen-bond interactions are also formed with side-chain N atoms of Gln, Lys, Arg and Asn, the hydroxyl groups of Ser and in a single case (ABF) through the N atom of the indole ring of Trp. Fig. 4 shows two examples of hydrogen bonding on the surface of ABF from iodine B71 to TrpH100 NE1 and from B91 to PheH105 N. These sites were both one quarter occupied. Hydrogen-bond formation between amide groups and halide ions is commonly observed and exploited in supramolecular chemistry (Schneider & Yatsimirsky, 2000). Iodide ions also bind through hydrophobic interactions as observed by Dauter *et al.* (2000), but this occurs much less frequently in the five structures discussed here.

The binding properties of polyiodides are sometimes a mixture of van der Waals and hydrogen bonding. This can be seen from Fig. 5, which shows a triiodide binding at the surface of PPE through a combination of van der Waals interactions with protein atoms at one end and electrostatic interactions with an ordered water molecule at the other. A further example is shown in Fig. 6, where a pentaiodide molecule is bound between two symmetry-related molecules in XI. The pentaiodide interacts through van der Waals forces along its length, but is anchored by hydrogen bonds from each of its end atoms to amide groups of Gln234 and Lys240. This would suggest that the majority of the charge on the pentaiodide molecule resides at its ends. *Ab initio* calculations on the triiodide ion indicate a charge distribution in solution of just less than −0.5 e on each end atom (Lynden-Bell *et al.*, 1998). Hydrogen-bond formation by the end I atoms would indicate that in our case the charge is qualitatively distributed in a similar manner along the pentaiodide molecule.

The PPE structure illustrates triiodide's ability to bind at hydrophobic sites. In this case, it binds in the same position that both Xe and Kr occupied in previous work on PPE (Schiltz *et al.*, 1995, 1997). Xe bound to PPE, obtained using a pressure of 1.5 MPa, had a refined occupancy of 0.78 and a *B* factor of 15.5 $Å^2$. Using a pressure of 5.6 MPa, Kr was found to bind with occupancy 0.5 and *B* factor 18.9 $Å^2$. The main iodine-binding site in PPE determined here has a final refined occupancy of 0.87 and a *B* factor of 16.9 $Å^2$. The observation

that I and Xe may bind at the same position in PPE agrees with that of Schoenborn *et al.* (1965) for myoglobin.

In work performed by Prangé *et al.* (1998) on the binding of Xe to HEWL (PDB code 1c10), three Xe-binding sites were observed, none of which is in common with iodide- or triiodide-binding sites in HEWL-KI/I_2. At Xe site 1 in 1c10, a Wat60 is seen in HEWL-KI/I_2. At site 2, a small intermolecular cavity, no atom is observed in HEWL-KI/I_2 and at site 3 an ethylene glycol molecule Etg131 is seen. The absence of any triiodide binding in HEWL-KI/I_2 at site 2 is a consquence of the small size of the intermolecular cavity. I^- would be small enough to gain access to this site, but this is energetically unfavourable.

The HEWL Xe work was carried out at room temperature and therefore in the absence of cryoprotectant. It is interesting to note that Xe and ethylene glycol share a binding site in HEWL. This site is spacious enough to host a triiodide molecule; however, in HEWL-KI/I_2 binding of ethylene glycol is favoured. This may suggest some competition between ethylene glycol and triiodide for hydrophobic binding sites and might explain the low number of triiodide-binding sites in the HEWL and APO structures, since these were the only two structures where ethylene glycol was used as a cryoprotectant. The other cryoprotectants used, L-xylose and glycerol, would tend to interact through hydrogen bonding and would therefore be less likely to inhibit the binding of triiodide.

In retrospect, the most striking result of the study is the apparent promiscuity of triiodide and its ability to form larger polyiodide chains on encountering a protein molecule. This ability to bind so readily to proteins is a consequence of its physical characteristics. Firstly, the flexibility of polyiodides both in terms of variation in bond length and bond angle enable them to yield to the numerous potential binding motifs at the surface of proteins. Secondly, iodine's high polarizability enables it to form stable bonds involving electrostatic interactions *via* dipole and charge-induced dipole moments and through London interactions (Schoenborn *et al.*, 1965). Finally, the low average charge per atom in polyiodide chains gives the molecule access to hydrophobic regions of protein molecules. An added bonus comes from the fact that where triiodide binding is not favoured either because of its size or its preference for oily surfaces, the iodide ions alone may be able to form ordered sites through electrostatic interactions.

7.3. Scope and limitations of triiodide phasing

The HEWL and APO cases were less promising that ABF, XI and PPE in terms of phasing using the triiodide derivatives.

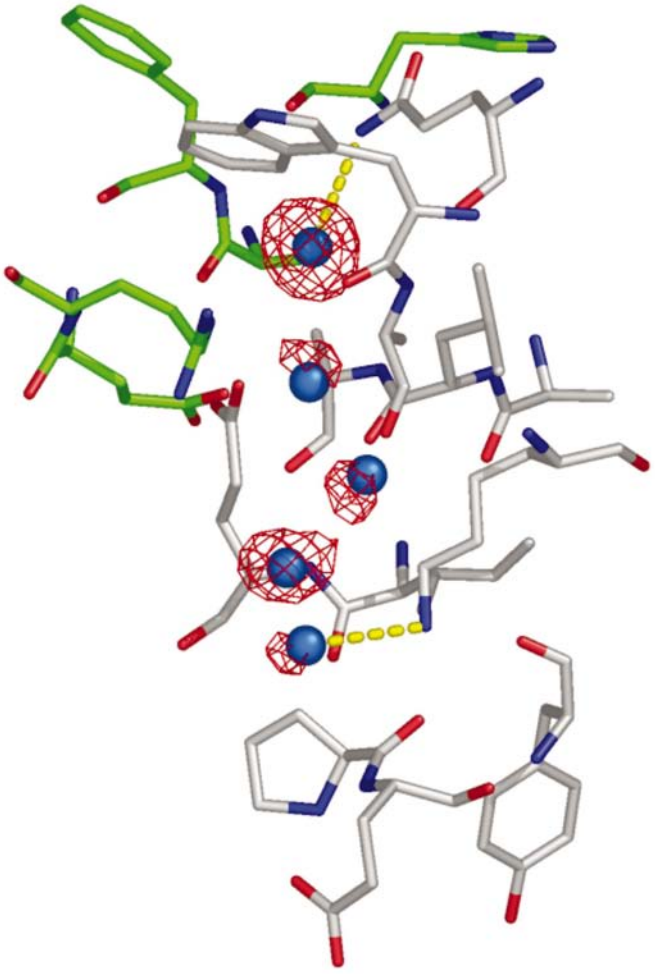

Figure 5
Triiodide B21–B23 (blue spheres) in a surface pocket of PPE. The triiodide make van der Waals contacts with backbone atoms of Asn115 and Ser116 on one side and the side chain of Gln119 on the other. The close proximity of Wat153 (red sphere) which is 3.06 Å from B21 and 3.34 Å from B22 suggests that it binds electrostatically to the triiodide. The anomalous difference Fourier map contoured at 4σ is shown in red mesh.

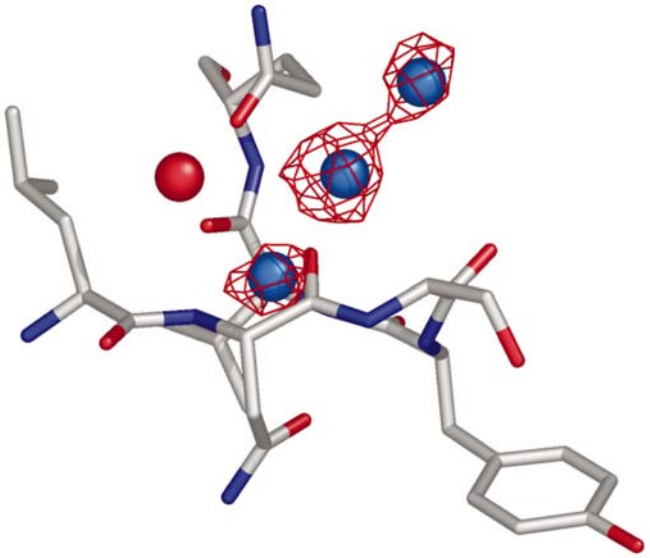

Figure 6
Example of a pentaiodide molecule (sites B21–B25) binding in a partially solvent-exposed pocket formed between two symmetry-related molecules of XI. The pentaiodide is tethered at each end by hydrogen bonds (yellow dashes) formed with Gln234 NE2 and Lys240 NZ. The central three atoms contact through van der Waals interactions. The symmetry-related protein molecule is shown in green. The anomalous difference Fourier map contoured at 4σ is shown in red mesh.

Table 9
Unit-cell axes lengths (Å) of the five KI/I_2-soaked protein crystals together with their respective native crystal cell dimensions recorded from samples crystallized from identical conditions and frozen using identical cryosolutions.

The last columns contain the relative differences in Å between native and derivative cell dimensions.

	Native			KI/I_2-soaked			Difference (Å)		
Protein	*a*	*b*	*c*	*a*	*b*	*c*	Δa	Δb	Δc
XI	92.27	98.46	102.38	92.35	98.54	102.51	+0.08	+0.08	+0.13
ABF	36.54	70.74	82.12	36.63	70.85	82.19	+0.09	+0.11	+0.07
HEWL	78.84	78.84	36.81	79.05	79.05	36.82	+0.21	+0.21	+0.01
PPE	50.15	57.90	74.56	50.57	57.90	74.76	+0.42	+0.00	+0.20
APO	181.74	181.74	181.74	181.19	181.19	181.19	−0.55	−0.55	−0.55

The anomalous signal present in the data suggested significant iodine binding and the structure refinement confirmed this by identifying triiodide/iodide sites. The failure of the phasing in both cases arose from the unfortunate positioning of major iodine sites (and in the APO case the Cd^{2+} ion) on special positions in the unit cell. This complicated both the heavy-atom determination and the phasing steps. The APO case was extreme in that all of the stronger heavy-atom sites lay on or very near special positions, giving rise to systematic absences of phase information in reciprocal space and hence aliasing of electron-density maps. The success and quality of the phase determination for ABF, XI and PPE shows that the concentration of iodine binding achieved with cryosoaks of a few minutes is certainly sufficient for the structure solution of macromolecular crystal structures with up to 40 kDa per asymmetric unit.

Table 9 shows the unit-cell parameters of native and KI/I_2-soaked crystals for all five cases and the differences arising from soaking. On average, the change in unit-cell length was 0.165 Å, with a maximum change of 0.42 Å occurring along the *a* axis of PPE. On a residue-by-residue basis in the refined structures, local r.m.s. deviations from the native structures of up to a few ångströms were observed predominantly in the vicinity of iodine binding. The largest changes occurred in the vicinity of triiodide binding. The usefulness of native plus derivative data for phasing should therefore not be greatly restricted by non-isomorphism effects. Indeed, for PPE, where the largest cell-edge changes were observed, the use of native diffraction data to provide isomorphous signal was essential for the determination of sufficiently good phases to allow automatic model building.

The use of triiodide as a 'heavy-molecule' derivative provides the added advantage that the results of heavy-atom structure determination can, in the first place, be assessed on the basis of whether some sites lie within expected I—I distances of each other and, in the future, be performed using the triiodide molecule as a search model. During these studies, the identification of groups of three atoms approximately 2.7 Å apart played a significant role in increasing our level of confidence in the the heavy-atom solutions generated by *RANTAN* and in the positions of extra sites determined with the help of *SHAR*P residual maps.

It may well be that in some cases the binding of counter-ions and/or triiodide ions will be too weak to support *ab initio* heavy-atom detection and *de novo* phasing. If an initial model for the protein is available, however, it may allow detection to take place and the phase information thus generated may be of value in helping overcome model bias from the initial model during the refinement of the structure. This could be especially useful in ligand-binding studies.

7.4. A tool for general use?

The studies performed here demonstrate triiodide's ability to bind at pHs ranging from 4.7 to 7.5 and, together with the triiodide-binding studies of Fukuyama *et al.* (1995) on peroxidase performed at pH 6.5 and 5.0, suggest that the triiodide ion is stable over this range and able to bind to macromolecules. Ammonium sulfate and sodium chloride are two of the most commonly used crystallization reagents for macromolecules (Gilliland *et al.*, 1994). It is therefore our belief that triiodide derivatization should be applicable in a large number of cases. The relationship between the number of bound iodine sites and the molecular weight of the five test proteins was well described by a straight line, suggesting that the larger the protein, the greater the level of binding one could expect from triiodide and iodide. Similar observations have been made in short halide-soak investigations (Dauter *et al.*, 2000). This bodes well for the determination of much larger structures using KI/I_2 as a derivatizing agent.

8. Conclusions

We have shown that KI/I_2 solutions can be successfully used with short-soak methods to generate derivatives with heavy-atom sites of varying structural character which allow phase determination by the SAD and SIRAS methods using conventional Cu $K\alpha$ radiation. The nature of the bound polyiodide molecules makes them easily identifiable at the heavy-atom detection stage. The strong colouration seen in KI/I_2 derivative crystals also provides valuable assurance that binding has occurred.

We have also demonstrated that improvements in the quality of experimental phases from in-house X-ray apparatus may be achieved by systematic variation of the heavy-atom substructure using combinatorial counter-ion replacement. Thus, at fixed wavelengths one can produce systematic variations in the scattering power generated at particular binding sites through chemical means.

Both these techniques should be useful sources of inexpensive phase information, either for *de novo* structure determination or for bias removal in structures from homologous starting models.

The triiodide work could be extended through the use of synchrotron radiation, which would allow selection of longer wavelengths to further enhance the anomalous signal of iodine. Recent studies by Weiss *et al.* (2001) have shown that routine diffraction measurement at wavelengths of 2.6 Å are indeed a reality on suitably designed and equipped beamlines.

Similarly, combinatorial counter-ion replacement can take advantage of other counter-ion species such as Rb^+ (Korolev *et al.*, 2001) and Br^- and could therefore be used to enhance the quality of phasing from fixed-wavelength synchrotron beamline side-stations tuned to shorter wavelengths.

It still remains to be seen whether triiodide is of use with other crystallization reagents. Recent observations suggest that its use in combination with PEG may be hindered by PEG's affinity for binding iodine. However, we believe that the large flexibility in iodine's chemistry may provide numerous ways of mediating its incorporation into PEG-grown protein crystals. This will be the subject of further investigation.

We would like to thank Leo James for providing us with crystals of his antibody fragment, Clemens Vonrhein for help with *SHARP* and *autoSHARP*, Pietro Roversi and Eric Blanc for help with *BUSTER*, Marc Schiltz and John Berriman for helpful discussions, Professor David M. Blow for helpful comments on the manuscript and Astex Technologies for access to their data-collection facilities. This work was partly supported by a Sponsored Research Agreement from Pfizer Central Research to GB and by European Commission grant No. HPRI-CT-1999-50015 within the EXMAD project.We would like to dedicate this article to the memory of Max Perutz, who created protein crystallography thanks to the use of isomorphous replacement.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Badger, J., Li, Y. & Caspar, D. L. D. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 1224–1228.
Bluhm, M. M., Bodo, G., Dintzis, H. & Kendrew, J. C. (1958). *Proc. R. Soc. London A*, **246**, 369–389.
Bricogne, G. (1993). *Acta Cryst.* D**49**, 37–60.
Bricogne, G. (1997). *Methods Enzymol.* **276**, 361–423.
Brodersen, D. E., de La Fortelle, E., Vonrhein, C., Bricogne, G., Nyborg, J. & Kjeldgaard, M. (2000). *Acta Cryst.* D**56**, 431–441.
Calabrese, V. T. & Khan, A. (2000). *J. Phys. Chem. A*, **104**, 1287–1292.
Carrell, H. L., Glusker, J. P., Burger, V., Manfre, F., Tritsch, D. & Biellmann, J.-F. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 4440–4444.
Chen, L., Rose, J. P., Breslow, E., Yang, D., Chang, W.-R., Furey, W. F. Jr, Sax, M. & Wang, B.-C. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 4240–4244.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cotton, F. A. & Wilkinson, G. (1988). *Advanced Inorganic Chemistry*, 5th ed. New York: John Wiley & Sons.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Crick, F. H. C. & Magdoff, B. S. (1956). *Acta Cryst*, **9**, 901–908.
Dauter, Z. & Dauter, M. (1999). *J. Mol. Biol.* **289**, 93–101.
Dauter, Z. & Dauter, M. (2001). *Structure*, **9**, 21–26.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. (1999). *J. Mol. Biol* **289**, 83–92.
Dauter, Z., Dauter, M. & Rajashankar, K. (2000). *Acta Cryst.* D**56**, 232–237.
DeLano, W. L. (2001). *The PyMOL User's Manual.* DeLano Scientific, San Carlos, CA, USA.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.
Evans, G., Roversi, P. & Bricogne, G. (2000). *Acta Cryst.* D**56**, 1304–1311.
Evans, G., Westbrook, E., Joachimiak, A. & Walsh, M. A. (2002). In the press.
Evans, P. R. (1997). In *Proceedings of CCP4 Study Weekend. Recent Advances in Phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey. Warrington: Daresbury Laboratory.
Fourme, R., Shepard, W., Kahn, R., l'Hermite, G. & Li de La Sierra, I. (1995). *J. Synchrotron Rad.* **2**, 36–48.
Fukuyama, K., Kunishima, N., Amada, F., Kubota, T. & Matsubara, H. (1995). *J. Biol. Chem.* **270**, 21884–21892.
Ghosh, D., Erman, M., Sawicki, M., Lala, P., Weeks, D. R., Li, N., Pangborn, W., Thiel, D. J., Jörnvall, H., Gutierrez, R. & Eyzaguirre, J. (1999). *Acta Cryst.* D**55**, 779–784.
Gilliland, G. L., Tung, M., Blakeslee, D. M. & Ladner, J. (1994). *Acta Cryst.* D**50**, 408–413.
Granier, T., Gallois, B., Dautant, A., d'Estaintot, B. L. & Précigoux, G. (1997). *Acta Cryst.* D**53**, 580–587.
Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *Proc. R. Soc. London A*, **225**, 287–307.
Greenwood, N. N. & Earnshaw, A. (1997). *Chemistry of the Elements*, 2nd ed. Oxford: Butterworth-Heinemann.
Gursky, O., Li, Y., Badger, J. & Casper, D. L. D. (1992). *Biophys. J.* **61**, 604–611.
Harvey, I., Hao, Q., Duke, E. M. H., Ingledew, W. J. & Hasnain, S. S. (1998). *Acta Cryst.* D**54**, 629–635.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Hope, H. (1988). *Acta Cryst.* B**44**, 22–26.
Korolev, S., Dementieva, I., Sanishvili, R., Minor, W., Otwinowski, Z. & Joachimiak, A. (2001). *Acta Cryst.* D**57**, 1008–1012.
Kretsinger, R. H. (1968). *J. Mol. Biol.* **31**, 315–318.
Kretsinger, R. H., Watson, H. C. & Kendrew, J. C. (1968). *J. Mol. Biol.* **31**, 305–314.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leslie, A. G. W. (1987). *Acta Cryst.* A**43**, 134–136.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EAMCB Newsl. Protein Crystallogr.* **26**.
Li, H., Dunn, J. J., Luft, B. J. & Lawson, C. L. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 3584–3589.
Luft, J. R. & DeTitta, G. T. (1999). *Acta Cryst.* D**55**, 988–993.
Lynden-Bell, R. M., Kosloff, R., Ruhman, S., Danovich, D. & Vala, J. (1998). *J. Chem. Phys.* **109**, 9928–9937.
McPherson, A. (1982). *Preparation and Analysis of Protein Crystals.* New York: John Wiley & Sons.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
Nagem, R. A. P., Dauter, Z. & Polikarpov, I. (2001). *Acta Cryst.* D**57**, 996–1002.
Nolte, R. T., Conlin, R. M., Harrison, S. C. & Brown, R. S. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 2938–2943.
Perrakis, A., Morris, R. J. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.
Prangé, T., Schiltz, M., Pernot, L., Colloc'h, N. & Longhi, S. (1998). *Proteins Struct. Funct. Genet.* **30**, 61–73.
Rice, L. M., Earnest, T. N. & Brunger, A. T. (2000). *Acta Cryst.* D**56**, 1413–1420.
Sakane, H., Mitsui, T., Tanida, H. & Watanabe, I. (2001). *J. Synchrotron Rad.* **8**, 674–676.
Sauter, C., Otálora, F., Gavira, J.-A., Vidal, O., Giegé, R. & García-Ruiz, J. M. (2001). *Acta Cryst.* D**57**, 1119–1126.
Schiltz, M., Fourme, R., Broutin, I. & Prangé, T. (1995). *Structure*, **3**, 309–316.

Schiltz, M., Shepard, W., Fourme, R., Prangé, T., de La Fortelle, E. & Bricogne, G. (1997). *Acta Cryst.* D**53**, 78–92.
Schneider, H.-J. & Yatsimirsky, A. (2000). *Principles and Methods in Supramolecular Chemistry*. New York: John Wiley & Sons.
Schoenborn, B. P., Watson, H. C. & Kendrew, J. C. (1965). *Nature (London)*, **207**, 28–30.
Sigler, P. B. (1970). *Biochemistry*, **9**, 3609–3617.
Smith, G. D., Nagar, B., Rini, J. M., Hauptman, H. A. & Blessing, R. H. (1998). *Acta Cryst.* D**54**, 799–804.
Svensson, P. H., Bengtsson-Kloo, L. & Persson, P. (1998). *J. Chem. Soc. Dalton Trans.*, pp. 1425–1429.
Tronrud, D. E., Eyck, L. F. T. & Matthews, B. W. (1987). *Acta Cryst.* A**43**, 489–501.
Usón, I. & Sheldrick, G. (1999). *Curr. Opin. Struct. Biol.* **9**, 643–648.
Vonrhein, C. & Bricogne, G. (2002). In the press.
Weiss, M. S., Sicker, T., Djinovic Carugo, K. & Hilgenfeld, R. (2001). *Acta Cryst.* D**57**, 689–695.
Whitaker, R. D., Cho, Y., Cha, J., Carrell, H. L., Glusker, J. P., Karplus, P. A. & Batt, C. A. (1995). *J. Biol. Chem.* **270**, 22895–22906.
Yang, C. & Pflugrath, J. W. (2001). *Acta Cryst.* D**57**, 1480–1490.
Yao, J.-X. (1981). *Acta Cryst.* A**37**, 642–644.
Yu-dong, L., Harvey, I., Yuan-xin, G., Chao-de, Z., Yi-zong, H., Hai-fu, F., Hasnain, S. S. & Hao, Q. (1999). *Acta Cryst.* D**55**, 1620–1622.