Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Practical experience with the use of halides for phasing macromolecular structures: a powerful tool for structural genomics

Zbigniew Dauter,[a]* Mi Li[b,c] and Alexander Wlodawer[c]

[a]Synchrotron Radiation Research Section, Macromolecular Crystallography Laboratory, Program in Structural Biology, National Cancer Institute and NSLS, Brookhaven National Laboratory, Building 725A-X9, Upton, NY 11973, USA, [b]Intramural Research Support Program, SAIC Frederick, National Cancer Institute–FCRDC, Frederick, MD 21702, USA, and [c]Protein Structure Section, Macromolecular Crystallography Laboratory, Program in Structural Biology, National Cancer Institute–FCRDC, Frederick, MD 21702, USA

Correspondence e-mail: dauter@bnl.gov

Received 21 August 2000
Accepted 25 October 2000

PDB Reference: pepstatin-insensitive carboxyl proteinase, 1ga1.

The crystal structure of pepstatin-insensitive carboxyl proteinase (PCP) from *Pseudomonas* sp. 101, an enzyme with no overall sequence similarity to any other proteinases of known structure, was solved using crystals soaked in sodium bromide solution and then cryocooled. A data set collected at the bromine peak absorption wavelength was sufficient for calculation of an excellent map and the entire process of phasing and tracing the maps required almost no direct human intervention. The process of structure solution using single-wavelength data was compared with three-wavelength multi-wavelength anomalous diffraction (MAD); although the latter resulted in slightly better maps, the use of this much more labor-intensive approach did not significantly improve the ability to solve the structure. The successful phasing approaches are compared with several less successful attempts utilizing other crystal forms of the enzyme and the practical aspects of the use of bromine as a heavy-atom derivative are discussed. In conclusion, the use of halides with single-wavelength diffraction data fulfills the requirements of being a first-choice method of high-throughput structure solution for the emerging field of structural genomics.

1. Introduction

Structural genomics, an approach to solving macromolecular structures on a massive scale in order to categorize all protein folds and to derive functional clues from structural data, requires the ability to determine novel structures at extremely rapid rates. A number of such efforts have already been reported or are under way (see, for example, Volz, 1999; Burley *et al.*, 1999; Kim, 2000). With many sequences of complete genomes now becoming available, the ability to solve crystal structures in a very fast predictable manner is more important than ever. A number of bottlenecks to structure solutions can be identified, including protein cloning, expression, purification and crystallization, as well as data collection and phasing. Although it is generally accepted that crystallization is still the least predictable stage of the process of structure solution, phasing of diffraction data often presents another difficult step.

The traditional method of solving the phase problem for proteins with unknown fold has been multiple isomorphous replacement (MIR) and its several variants (Blundell & Johnson, 1976; Boggon & Shapiro, 2000). In this approach, protein crystals are derivatized by soaking in solutions of heavy atoms, usually metal salts or organometallic compounds. The choice of proper derivatives is heuristic and very dependent on the nature of the protein under study and thus not

Table 1
Various crystal forms of PCP.

Crystal form	I	I	I	II	III	III	III
Derivative	Native	KI	Hg	NaBr	Native	Native long λ	NaBr
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P6_122$	$P6_2$	$P6_2$	$P6_2$
Application	MR + refinement	Not used	Tests only	MR + refinement	*DM* + refinement	Locating Ca + I	*SHELXD* + *SHARP*

easy to generalize. Non-isomorphism between the derivative and native crystals also often creates problems. In the last 10 y, multiwavelength anomalous diffraction (MAD), most often utilizing selenomethionine (Se-Met) modifications of the protein of interest (Se-MAD), has been the preferred method of structure solution (Hendrickson *et al.*, 1990; Yang *et al.*, 1990; Hendrickson, 1999). In principle, only a single crystal is required and thus no errors caused by non-isomorphism between crystals of the native protein and the derivative are introduced. This method, however, shifts the burden from finding and solving derivatives to making modified proteins by recombinant means and is not always easily applicable. Typical difficulties in performing Se-MAD experiments are encountered when systems other than bacteria are used for protein expression or when the protein under study contains no (or very few) methionine residues. Thus, modifications of the MIR and/or MAD techniques that would work with atoms either already present in all proteins or easily attached to them appear to be most universal and therefore applicable to large-scale structure-determination efforts.

Anomalous scattering of S atoms, present in almost all known proteins, is one such technique that is being developed currently. Although the structure of crambin was solved by this route 20 y ago (Hendrickson & Teeter, 1981), only a few structures have been successfully solved by this approach since then (Dauter *et al.*, 1999; Liu *et al.*, 2000; Newton *et al.*, 2000; Wang *et al.*, 2000; Chen *et al.*, 2000). No other atoms with significant anomalous scattering are universally present in unmodified proteins, so that either their incorporation during protein synthesis or subsequent chemical modification of proteins or crystals is still necessary. However, recent studies of derivatization of proteins by short cryosoaking with halides have shown that various proteins could be modified in a minute or less by soaking in a solution containing bromide or iodide ions (Dauter & Dauter, 1999; Dauter *et al.*, 2000). The use of bromine for the purpose of phasing is particularly important, since its absorption edge is at the wavelength of 0.92 Å, easily accessible on most tunable synchrotron beamlines. A number of known structures were re-solved in order to test this method (Dauter *et al.*, 2000) and a few new protein structures were also determined, including yeast hypothetical protein (Ho *et al.*, 2000) and human β defensin II (Hoover *et al.*, 2000). The largest protein structure solved so far with this approach was that of human acyl-protein thioesterase (Devedjiev *et al.*, 2000), with a molecular mass of 56 kDa for a dimer in the asymmetric unit. However, no systematic studies that would evaluate the limits of this technique and provide practical guidance for its future use have been reported.

We have now solved the structure of an enzyme known as pepstatin-insensitive carboxyl proteinase (PCP) from *Pseudomonas* sp. 101 (Ito *et al.*, 1999; Oyama *et al.*, 1999). We consider this protein to be a good choice for showcasing the use of bromine as a universal heavy atom. Although this enzyme has been studied for more than 10 y (Oda *et al.*, 1987) and a number of aspartic and glutamic acid residues were implicated in its catalytic activity (Oyama *et al.*, 1999), its lack of sequence similarity to known aspartic proteases (Davies, 1990) clearly showed that it was not a typical member of this family (Oyama *et al.*, 1999). However, besides a short signature sequence characteristic of serine proteases (Kwon *et al.*, 1999; Rawlings & Barrett, 1999), no other unambiguous sequence similarly between PCP and other proteolytic enzymes has been detected (Oda *et al.*, 1994) and thus its fold has not been predicted. Although crystals of PCP have been available for a number of years, all previous attempts to make useful derivatives using heavy metals have failed (K. Oda, personal communication). While PCP was successfully expressed in *Escherichia coli*, preparation of Se-Met derivatives would require further mutagenesis, as only one methionine is present in the native sequence consisting of 372 amino acids for the mature PCP (Oda *et al.*, 1994) and a single Se atom would be unlikely to phase the structure of a 40 kDa protein (Hendrickson & Ogata, 1997).

In this paper, we discuss the practical aspects of solving the crystal structure of PCP and analyze both the successful experiments and the failed ones that utilized different crystal forms (Table 1). We also compare the process of structure solution using only single-wavelength data (the approach first used for PCP) with the full three-wavelength MAD that was performed subsequently.

2. Materials and methods

2.1. Crystallization and derivatization of PCP

PCP used in this study was provided by Dr Kohei Oda (Kyoto Institute of Technology, Kyoto, Japan) in the form of complexes with either a specific inhibitor tyrostatin (Oda *et al.*, 1992) or its iodinated derivative in which the tyrosine residue was replaced with iodophenylalanine. Tyrostatin binds to PCP with K_i = 2.6 n*M* (Oda *et al.*, 1992) and iodotyrostatin binds with K_i = 15 n*M* (K. Oda, personal communication). The original purpose of this substitution was to insert an intrinsic heavy atom into a compound known to bind to this enzyme.

We have obtained three crystal forms of PCP under different conditions: two by direct crystallization and one by soaking of the pre-grown crystals. Crystal form I was obtained by hanging-drop vapor diffusion using a stock solution of a 1:1 complex of PCP with tyrostatin. We used a protein sample with a concentration of 6 mg ml^{-1} in 0.1 *M* sodium acetate buffer pH 4.8. This solution was mixed in a 1:1 ratio with well solution consisting of 1 *M* ammonium sulfate, 0.005 *M* guani-

Table 2
Details of X-ray data collection.

Crystal form	Derivative	Space group	Unit-cell parameters (Å)			Wavelength (Å)	Resolution (Å)	Measured reflections	Unique reflections	R_{merge} (%)	R_{anom} (%)	$I/\sigma(I)$	Completeness (%)
			a	*b*	*c*								
I	Native	$P2_12_12_1$	128.3	219.5	86.0	0.98	2.3	463766	106230	5.4 (23.1)	—	22.0 (5.4)	98.0 (99.9)
I	KI	$P2_12_12_1$	127.9	219.5	85.7	1.54	2.3	418192	106274	7.5 (28.6)	4.9 (17.8)	16.4 (4.0)	99.0 (93.0)
I	Hg peak	$P2_12_12_1$	128.5	220.3	85.7	1.0071	2.5	321333	82285	7.9 (28.6)	6.0 (19.9)	15.4 (3.7)	97.2 (89.3)
I	Hg inflection	$P2_12_12_1$	128.5	220.3	85.7	1.0086	2.5	326219	82595	6.8 (30.9)	4.7 (19.9)	16.9 (3.6)	97.5 (89.6)
I	Hg remote	$P2_12_12_1$	128.5	220.3	85.7	0.9686	2.5	330884	83044	8.6 (43.8)	6.1 (27.2)	14.7 (2.8)	98.0 (95.5)
II	NaBr peak	$P6_122$	127.4	127.4	86.9	0.9198	2.2	142186	21560	8.9 (38.5)	5.4 (19.1)	14.0 (3.4)	99.7 (98.9)
II	NaBr inflection	$P6_122$	127.4	127.4	86.9	0.9202	2.2	106102	21324	9.3 (52.4)	5.9 (25.7)	11.4 (2.5)	97.6 (97.2)
II	NaBr remote	$P6_122$	127.4	127.4	86.9	0.9164	2.2	106318	21336	9.7 (52.0)	6.3 (24.3)	10.3 (2.5)	97.6 (97.8)
III	Native	$P6_2$	97.2	97.2	83.4	0.92	1.4	677827	87947	3.6 (21.6)	—	50.2 (7.0)	99.9 (99.5)
III	Native long	$P6_2$	97.7	97.7	82.9	1.54	1.7	1046501	49410	6.1 (25.1)	3.4 (19.3)	37.6 (7.0)	99.9 (99.3)
III	NaBr peak	$P6_2$	97.3	97.3	83.0	0.9191	1.8	228827	41335	4.1 (32.7)	4.1 (19.3)	23.4 (3.2)	98.9 (96.7)
III	NaBr inflection	$P6_2$	97.3	97.3	83.0	0.9196	1.8	229034	41369	4.1 (33.5)	3.6 (20.7)	23.8 (3.0)	98.9 (96.9)
III	NaBr remote	$P6_2$	97.3	97.3	83.0	0.9164	1.8	228898	41370	4.4 (37.1)	3.9 (22.4)	22.4 (2.8)	98.9 (96.8)

dine and 10% glycerol in 0.1 *M* sodium citrate buffer pH 3.32. The resulting 4 µl drops were suspended over 0.7 ml of well solution. Crystals appeared after 2 d and grew to maximum dimensions of 0.15 × 0.15 × 0.6 mm in 5 d.

Crystal form I was found to be orthorhombic, space group $P2_12_12_1$, unit-cell parameters a = 128.3, b = 219.5, c = 86.0 Å, with six molecules in the asymmetric unit. When soaked for 20 s in a solution consisting of 1 *M* ammonium sulfate, 0.005 *M* guanidine, 18% glycerol, 1 *M* sodium bromide in 0.1 *M* sodium citrate buffer pH 3.32, these crystals reproducibly changed their symmetry to hexagonal, space group $P6_122$, unit-cell parameters $a = b$ = 127.4, c = 86.9 Å, with one molecule in the asymmetric unit (crystal form II). The relationship between crystal forms I and II is discussed in detail below.

A very different crystal form (form III) was obtained from a sample consisting of a 1:1 complex of PCP with modified tyrostatin in which the putative P3 tyrosine was replaced with *p*-iodophenylalanine. We used protein at a concentration of 10 mg ml^{-1}, 0.1 *M* sodium chloride, 5 m*M* calcium chloride in 50 m*M* sodium acetate buffer pH 4.8. This stock solution was mixed in a 1:1 ratio with well solution consisting of 1.4 *M* lithium sulfate in 0.1 *M* Tris buffer pH 7.5, producing drops with a final volume of 6 µl. The process of crystal growth was not sensitive to pH, with crystals found at pH as low as 5.0. The crystals usually appeared after 5 d and grew to final dimensions of 0.5 × 0.5 × 0.75 mm within 10 d. These crystals were found to be hexagonal, space group $P6_2$, unit-cell parameters $a = b$ = 97.3, c = 83.0 Å, with one molecule in the asymmetric unit. They were very well ordered and diffracted to beyond 1.4 Å resolution using a synchrotron source. The Br derivative was prepared by soaking the crystals for 30 s in a solution containing 1 *M* sodium bromide, 1.2 *M* lithium sulfate, 14% glycerol in Tris buffer pH 7.5.

A putative heavy-atom derivative of crystal form I was prepared by soaking crystals for two weeks in a solution of saturated mercury ethyl phosphate also containing 0.8 *M* ammonium sulfate, 0.005 *M* guanidine, 10% glycerol in 0.1 *M* sodium citrate buffer pH 3.32. Crystals subjected to this treatment were stable and diffracted well.

2.2. X-ray data collection

X-ray diffraction data were collected on beamline X9B, NSLS, Brookhaven National Laboratory using the ADSC Quantum4 CCD detector; reflections were integrated and merged using the *HKL*2000 suite (Otwinowski & Minor, 1997), with the results summarized in Table 2. The initially obtained crystal form I diffracted to 2.3 Å and merged in an orthorhombic system with systematic absences indicating the space group $P2_12_12_1$. These crystals resisted most heavy-atom derivatization trials and were eventually subjected to derivatization by soaking briefly in a cryosolution containing halides, NaBr or KI (Dauter *et al.*, 2000). The soaking in 1 *M* KI did not produce a useful derivative with any significant anomalous scattering signal. Soaking in 1 *M* NaBr, however, caused the crystals to change their symmetry from orthorhombic to hexagonal form II in the $P6_122/P6_522$ space group (later proved conclusively to be $P6_122$). This phenomenon has been confirmed by collecting diffraction images from a freshly frozen crystal displaying the orthorhombic symmetry, removing it from the goniostat, soaking in NaBr for about 1 min, freezing it again and collecting more diffraction images, which this time clearly showed the phase change from orthorhombic to hexagonal. The relationship between the two cells is discussed below. Diffraction data collected from NaBr-soaked hexagonal form II crystals to a resolution of 2.2 Å showed a significant amount of anomalous signal, but all attempts to identify the bromide sites by direct methods or Patterson search failed.

A three-wavelength MAD data set for a Hg derivative of the form I crystals was collected to a resolution of 2.5 Å (Table 2). Since the positions of Hg atoms could not be identified directly, these data were not used for structure solution, but the presence of heavy atoms was subsequently confirmed (see below).

Diffraction data from a native crystal in the hexagonal form III were collected to a resolution of 1.4 Å. Three-wavelength MAD data were collected on a crystal cryosoaked in the presence of 1 *M* NaBr to a resolution of 1.8 Å (Table 2) and

showed a significant anomalous signal. In addition, data were collected on another native crystal by using a longer wavelength of 1.54 Å, aiming at the identification of the I atom within the inhibitor from its anomalous dispersion signal.

3. Results

3.1. Structure solution using form III crystals

The diffraction data from the hexagonal form III (space group $P6_2$ or enantiomorph) were of high quality for both the native and brominated crystals (Table 2). The single data set from the NaBr-soaked crystal, collected at the peak wavelength, was presumed to have the highest amount of anomalous signal and served as input data to the program *SHELXD* (Sheldrick, 1998) after extracting Bijvoet differences by *XPREP* (Bruker Analytical X-ray Systems). *SHELXD* was run with 974 of the highest *E*s looking for six anomalous sites; 24% of all phase sets gave the correct solution and the nine highest peaks were accepted for further phasing by *SHARP* (de La Fortelle & Bricogne, 1997). The refinement of bromide coordinates and *B* factors and the protein phase estimation in *SHARP* were also performed against the same single data set collected at the peak wavelength to a resolution of 1.8 Å. The final overall figure-of-merit (FOM) value was 0.213 and the maps were partially interpretable, as shown in Fig. 1(*a*).

The phases from *SHARP* were input to density modification by solvent flattening and histogram matching and extended to the full resolution of the native data (1.4 Å) using the program *DM* in the *COMBINE PERTURB* mode (Cowtan & Main, 1996). The resulting overall FOM was 0.739. The electron-density map calculated with *DM* phases and FOM as weights (Fig. 1*b*) was interpretable in large parts, as judged by visual inspection and the overall correlation coefficient (CC)

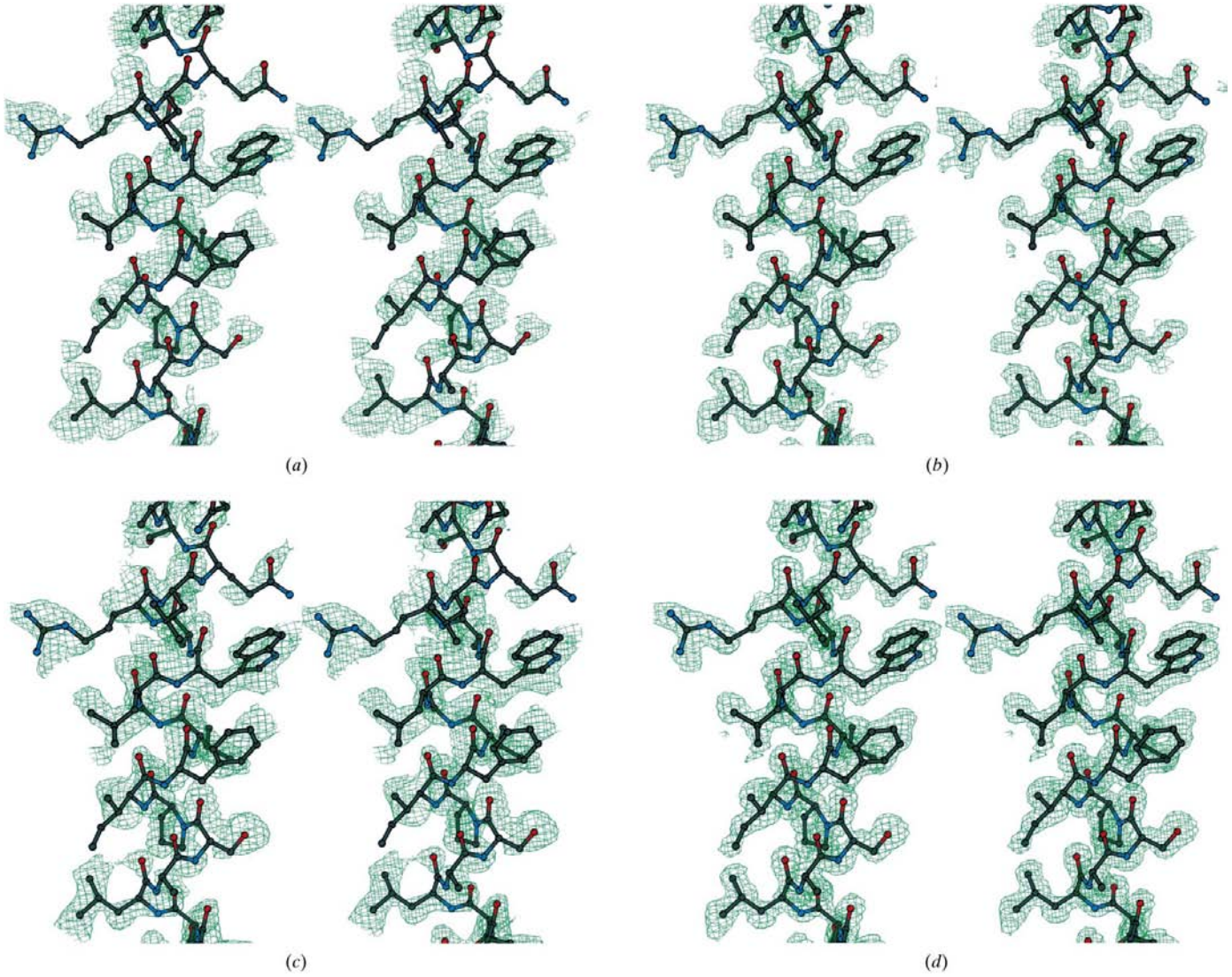

Figure 1
Electron-density maps for a fragment of PCP solved in crystal forms III (*a*–*f*) and II (*g*–*h*). The coordinates are those of a largely buried helix, residues 285–304, and are taken from the final refined model. All maps are contoured at the 1σ level. (*a*) Map at 1.8 Å resolution based on phases resulting from the single-wavelength Br experiment after refinement with *SHARP*. (*b*) Map after *SHARP* and *DM* at 1.4 Å resolution. (*c*) Map based on three-wavelength Br MAD experiment after refinement with *SHARP* at 1.8 Å resolution. (*d*) Map after *SHARP* and *DM* at 1.4 Å resolution.

between this map and the F_{obs} map obtained later after the refinement, which was 0.653 (Table 3). The same procedure involving *SHARP* and *DM* was repeated in the alternative enantiomeric space group $P6_4$ (after inverting all coordinates), but the resulting map was not interpretable.

The map obtained by *DM* was submitted to the automatic *wARP* procedure (Perrakis *et al.*, 1999), based on filling the electron-density map by water O atoms, refining them without restraints and iteratively discarding and acquiring new ones on the basis of electron density and geometrical criteria. The quality of this map was excellent (Fig. 1*e*). In the *warpNtrace* mode, the program automatically builds pieces of the peptide main chain and fits amino-acid side chains according to the protein sequence into the interpretable portions of the map. As a result of this procedure, the model consisting of 363 of 372 amino acids of the protein (residues 4–176 and 181–370) was built automatically with most of the correct side chains, with an R factor of 19.9% and an R_{free} of 22.6%. This model resulted from single-wavelength data after automatic computations without any human interpretation at the graphics display. The remaining side chains and a few missing residues not interpreted by *warpNtrace* were introduced manually using the graphics display program *O* (Jones & Kieldgaard, 1997). The electron density calculated after *warpNtrace* was clearly interpretable, even in places that the automatic procedure had failed to represent correctly and that had been left filled by water atoms. The correlation with the final F_{obs} map was 0.882.

The model was subsequently refined by *CNS* (Brunger *et al.*, 1998) to an R factor of 18.97% and an R_{free} of 20.31%, the latter based on 10% of randomly selected reflections. Because the native data were of high quality and the resolution was 1.4 Å, the refinement was continued with *SHELXL* (Sheldrick & Schneider, 1997) using the same group of reflections for calculating R_{free} and with *SHELXL* default restraints applied to geometrical and atomic displacement parameters. Each

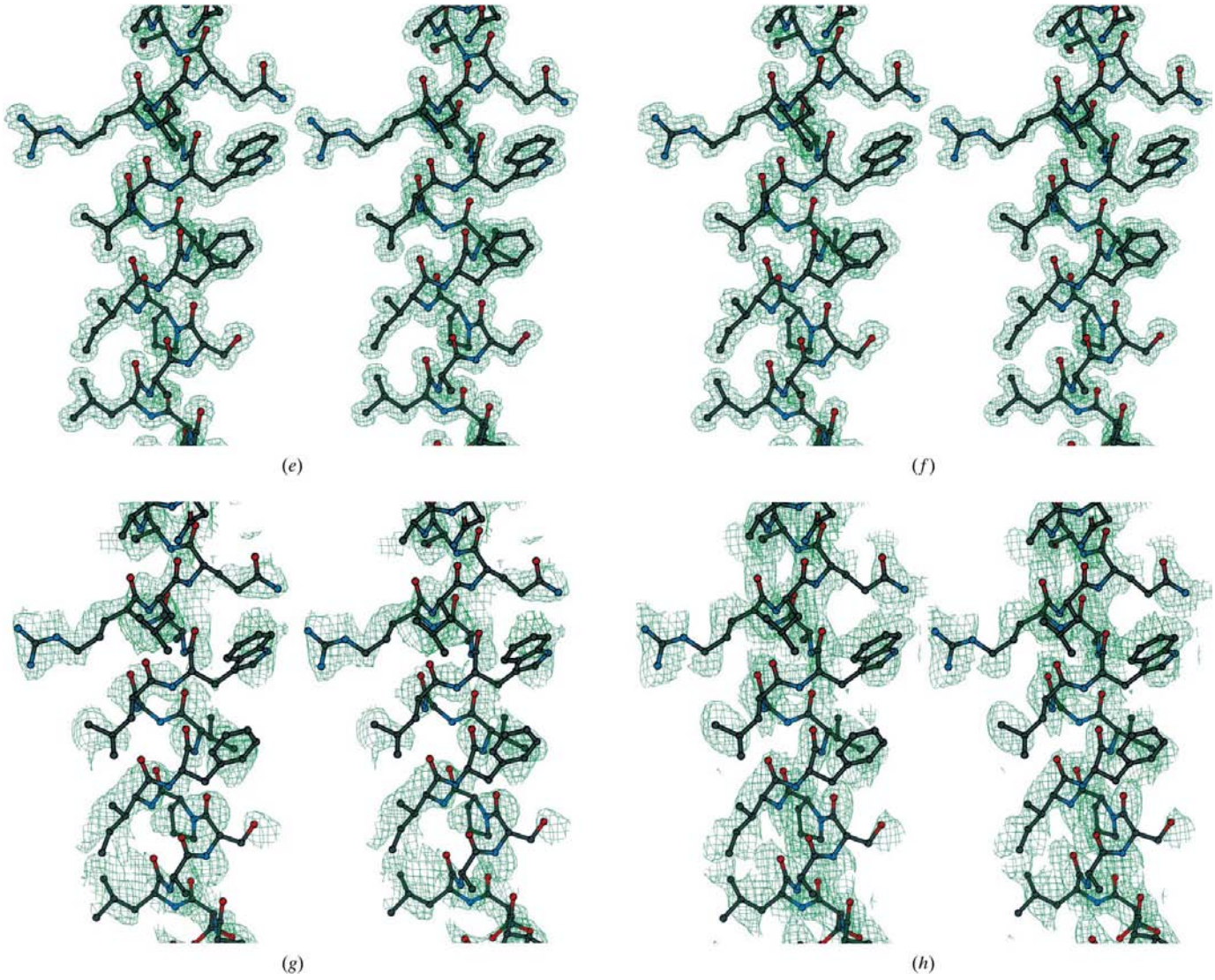

Figure 1 (continued)
(*e*) The same region at 1.4 Å resolution after refinement with *wARP*. (*f*) The final $3F_{obs} - 2F_{calc}$ refined 1.4 Å map. (*g*) Map at 2.2 Å resolution based on phases resulting from a MAD experiment after refinement with *SHARP*. (*h*) Map after *SHARP* and *DM* at 2.2 Å resolution.

Table 3
Correlation coefficients between the final refined F_{obs}, α_{calc} map and electron-density maps at various stages of structure solution.

The total value was calculated for all grid points within the asymmetric unit; values for main and side chains were based on grid points around main- or side-chain atoms of individual residues.

	Map CC		
	Total	Main chain	Side chains
Crystal form III, $P6_2$			
Final $3F_{obs} - 2F_{calc}$	0.971	0.988	0.975
ARP	0.882	0.950	0.898
DM, single wavelength	0.653	0.759	0.638
SHARP, single wavelength	0.282	0.449	0.337
DM, MAD	0.768	0.866	0.762
SHARP, MAD	0.374	0.564	0.430
Crystal form II, $P6_122$			
SHARP, MAD	0.341	0.457	0.367
DM, MAD	0.477	0.556	0.457

round of *SHELXL* refinement consisted of ten cycles of conjugate-gradient minimization followed by the inspection of the model and electron-density maps with the program *QUANTA* (Molecular Simulations Inc., San Diego).

The introduction of the anisotropic atomic displacement parameters proved to be valid, as corroborated by the drop in the *R* factor from 16.26 to 12.45% and in R_{free} from 17.96 to 15.60%. The introduction of H atoms as 'riding' on their parent atoms lowered the *R* factor to 11.13% and R_{free} to 14.20%. Only a small number of regions had uninterpretable electron density and were assigned zero occupancy. Subsequent rebuilding sessions led to the introduction of 19 side chains in double conformations and their relative occupancies were refined. About 430 water molecules were introduced to the model and were refined with unit occupancies. Some of them in fact must have lower occupancy, but at 1.4 Å resolution their occupancies and displacement parameters are too correlated to warrant independent refinement. The presence and location of the solvent were judged by the appearance of the electron density and the participation of putative water molecules in the hydrogen-bonding networks.

Two solvent water molecules were refined to a low displacement-parameter value. The inspection of the first site revealed its almost perfectly octahedral coordination by O atoms of two aspartic acid carboxyl groups, three peptide carbonyl O atoms and one water, with distances of 2.3–2.4 Å characteristic for the calcium ion, which was subsequently introduced to the model. The second site was located at the protein surface at a distance of 3.1 Å from a guanidinium group and was classified as a Cl^- anion. This classification was confirmed unequivocally by the anomalous difference Fourier map computed with Bijvoet differences and phases from the data collected at the long (1.54 Å) wavelength, where both calcium and chlorine have a significant anomalous signal (see below). Near the end of refinement, when the *R* factor dropped to below 12%, some continuous electron density appeared near the protein surface and it was possible to interpret part of it as an aromatic amino acid accompanied on both sides by the main chain of two residues. One of the anomalous difference map peaks was positioned ideally at the *para* position of the aromatic ring. This feature was therefore modeled as a part of the iodotyrostatin, an inhibitor present in the crystallization medium. The occupancy of this fragment was refined to about 0.3. It was impossible to interpret the remaining parts of the inhibitor, which were apparently highly disordered. The electron-density features corresponding to the inhibitor coincide with solvent waters, also at partial occupancies, making the interpretation difficult.

At the end of refinement, when the process reached convergence, a round of blocked full-matrix least-squares refinement was performed using all reflections, including those previously used for R_{free}, to obtain as accurate a model as possible. The final *R* factor was 10.63%. This mode of refinement also gave a proper estimation of the standard uncertainties of all individually refined parameters from the inversion of the least-squares matrix. The accuracy of the positional parameters of the atoms in well defined regions of the protein is about 0.02–0.03 Å, with considerably higher uncertainties for atoms in flexible regions and with high *B* factors.

The structure of PCP comprises a single compact domain with a diameter of about 55 Å, consisting of the seven-stranded parallel β-sheet flanked on both sides by a number of helices (Fig. 2). The depression on one side, where the inhibitor iodotyrostatin was identified, contains the active site. The detailed description of the structure and the discussion of its biological context will be published elsewhere (Wlodawer *et al.*, manuscript in preparation).

3.2. Three-wavelength Br MAD on form III crystals

The structure of this form was solved using only single-wavelength data collected on the Br-derivatized crystal, as described above. However, the full three-wavelength diffraction data were collected at the peak, at the inflection point and at the high-energy near-remote point of the fluorescence spectrum. The region around the absorption edge of bromine lies at the end of the accessible range of the monochromator at the X9B beamline; therefore, the high-energy remote wavelength was set to only 50 eV beyond the bromine edge.

In parallel with phasing by *SHARP* based on the single (peak) wavelength, an analogous run was performed using all three data sets, with the same initial conditions, *i.e.* the same number of anomalous scatterers (nine) and their parameters. The resulting protein phase set was clearly superior, as judged from the overall FOM of 0.454 *versus* the previous 0.213. This phase set was then used as input to the program *DM* with the same protocol as before, *i.e.* with option *COMBINE PERTURB*, 60% solvent content (as estimated from the Matthews coefficient of 2.85 $Å^3\ Da^{-1}$; Matthews, 1968) and automatically selected number of cycles. After four cycles, the procedure converged and the resulting FOM was 0.626, a value lower than that obtained from the single-wavelength data. However, the resulting electron-density map was superior, with a higher degree of continuity and more easily interpretable features (Fig. 1*c*). This result is also reflected by

the higher CC with the final F_{obs} map (Table 3). In fact, the single-wavelength *DM* performed seven cycles before convergence; when *DM* was forced to run for seven cycles as with the single-wavelength data, the resulting FOM rose to 0.789 with the resulting map almost identical to the previous four-cycle run.

The above results show first, as expected, that phasing based on full three-wavelength MAD data is superior over the single-wavelength procedure. However, the three-wavelength *SHARP* phase refinement took almost five times longer than the single-wavelength one (76 h *versus* 15 h, respectively, on an SGI Octane computer under the IRIX 6.4 operating system). Secondly, the density modification to some extent leveled out the quality of the resulting map. In both cases the maps were interpretable and the single-wavelength map, albeit slightly inferior, was good enough to be successfully used for the automatic *wARP* procedure. Although it was not performed, it can be safely assumed that *wARP* performed on the three-wavelength MAD map would provide a result comparable to that obtained with the single-wavelength data. Taking into account the time involved in collecting two additional data sets and phasing against three times as many reflections, we conclude that the use of single-wavelength data was significantly more economical in this case.

3.3. Longer wavelength data on form III crystals

In addition to the MAD experiment on the Br-soaked crystal of the $P6_2$ form and the high-resolution (1.4 Å) data collected on an untreated crystal of the same form using short-wavelength radiation (0.92 Å), we collected diffraction data to 1.7 Å resolution on an untreated crystal using a longer wavelength (1.54 Å) corresponding to the Cu $K\alpha$ characteristic peak. These crystals were expected to indicate the position of the inhibitor iodotyrostatin, initially assumed to be bound to PCP. After solving the PCP structure, we refined its model in ten restrained cycles against these long-wavelength data and used the phases obtained to compute the anomalous difference Fourier synthesis. As stated above, six significant peaks were present in the map (Fig. 2*c*), corresponding to the calcium ion (32.7σ), partially occupied I atom of the inhibitor (22.4σ), chloride anion at the protein surface (12.3σ) and three S atoms of the protein (12.9, 11.0 and 8.9σ). The remaining peaks, below the 8σ level, constituted noise.

Because the quality of the diffraction data was high, we attempted to find the positions of the anomalous scatterers independently by direct methods. The *SHELXD* run at 2.5 Å, with 610 highest *E*s, gave a clear solution, identifying the two highest sites (Ca and I) and two more peaks, one corresponding to methionine sulfur and another encompassing two S atoms of the disulfide bridge. The two most prominent sites were submitted to phasing by *SHARP* against the corresponding single-wavelength data. However, the signal appeared to be too weak and this phasing attempt did not lead to an interpretable map.

3.4. Molecular-replacement (MR) solution of crystal forms I and II

After the PCP model was constructed and refined isotropically in the $P6_2$ crystal form III, it was used to solve the structures of the other crystal forms by MR with the program *AMoRe* (Navaza, 1994). For crystal form II, the data collected at the remote wavelength were used for solution and refinement of the structure. The solution was obtained easily, with a very clear contrast between the correct and false solutions in both the rotation and translation function (but only in the $P6_122$ space group, not in $P6_522$). The resulting CC was 0.713 and the *R* factor was 35.5% at 3.0 Å resolution after rigid-body fitting. This model was subjected to restrained refinement by *REFMAC* (Murshudov *et al.*, 1997) combined with automatic selection of solvent waters by *ARP* (Lamzin & Wilson, 1997). The final *R* factor was 27.0% and R_{free} was 36.7%. These values are somewhat high, presumably related to the crystal phase change that was forced on the orthorhombic crystals by soaking them in the cryosolution containing NaBr. Such a rapid rearrangement of individual molecules with subsequent freezing could have led to imperfect ordering within the whole crystal, although it satisfactorily displayed the hexagonal symmetry of the crystal lattice and of the merged intensities.

In spite of the presence of six PCP molecules in the asymmetric unit of the orthorhombic crystal (form I), the solution of the structure by *AMoRe* proceeded smoothly owing to the high quality of the search model. Three orientations with a CC of 0.23–0.25 appeared in the rotation function on a background of orientations with a CC below 0.1. There was also a high contrast between the correct one-molecule solutions in the translation function with a CC above 0.18 *versus* false solution values below 0.12. After positioning of all six molecules in the cell and rigid-body fitting at 3 Å, the CC was 0.757 and the *R* factor was 32.9%. There were only three independent peaks in the rotation function because the constellation of six molecules possessed an approximate non-crystallographic twofold screw axis (in fact, a 6_1 axis) parallel to the unit-cell *c* direction; hence, two molecules were represented by the same peak in the rotation function. This model was refined with the same protocol as the previous structure in $P6_122$, except for the use of non-crystallographic restraints applied to the six independent molecules in the asymmetric unit. The final *R* factor was 19.4% and the R_{free} was 24.7%.

3.5. Relationship between crystal forms I and II

From the comparison of their unit-cell parameters and from the fact that the crystal form II (space group $P6_122$) was obtained by a short soaking of the crystal form I (space group $P2_12_12_1$) in the solution containing NaBr, it was evident that the packing of PCP molecules in both cells must be closely related. The crystal phase transition caused by soaking was manifested by the disappearance of half of the reflections present in the diffraction pattern of the orthorhombic cell (those with $h + k = 2n + 1$), in effect leaving the *C*-centered orthorhombic cell with the ratio of cell parameters $b/a = 219.5/$

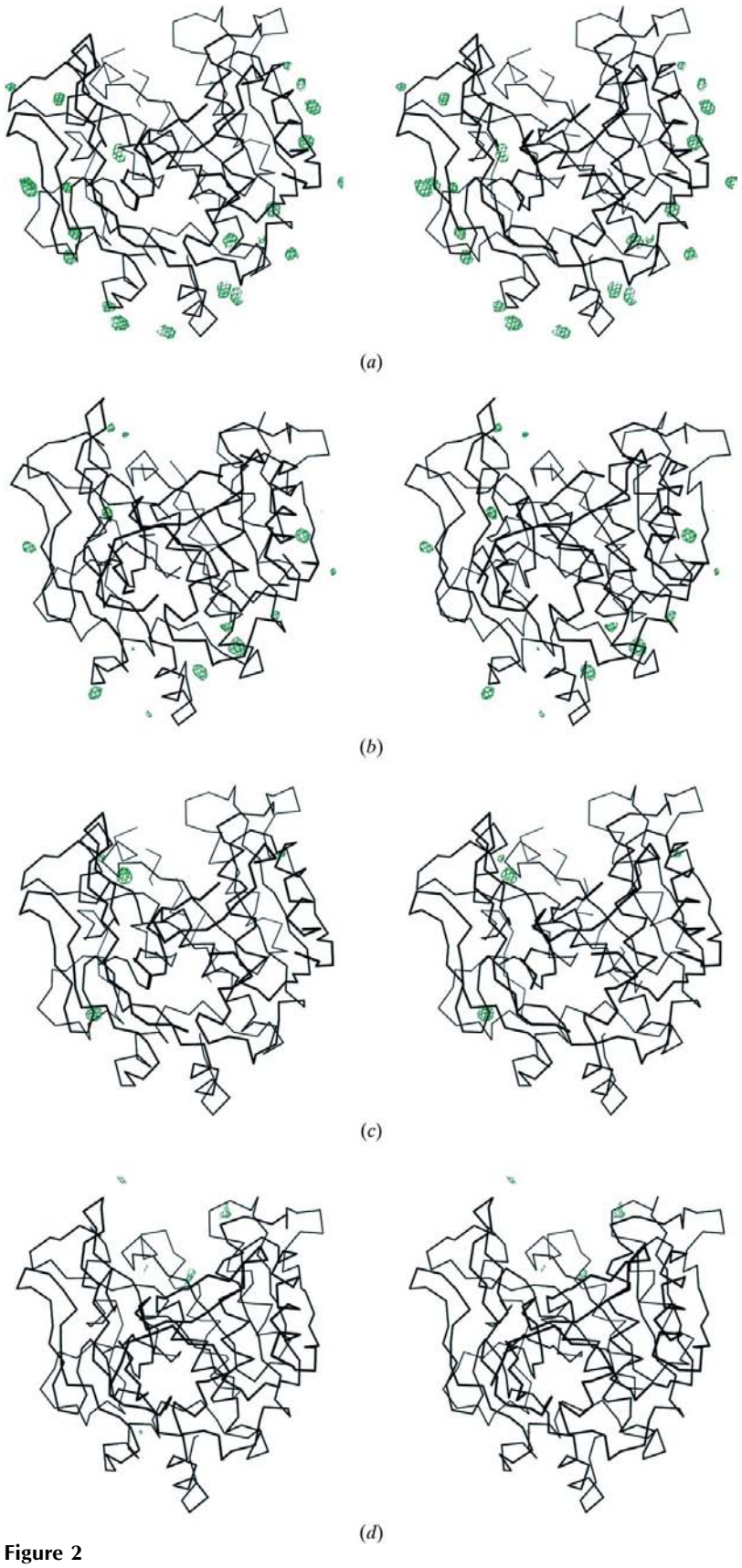

Figure 2
Anomalous difference Fourier maps for the derivatives of PCP, contoured at the 5σ level (*a*, *b* and *d*) or the 7σ level (*c*). (*a*) Crystal form III, Br-peak wavelength. (*b*) Crystal form II, Br-peak wavelength. (*c*) Crystal form III, wavelength 1.54 Å. (*d*) Crystal form II, Hg near-remote wavelength.

128.3 $\simeq 3^{1/2}$. The unit-cell *c* parameters of both cells were almost identical. Such a cell can be indexed as hexagonal and, indeed, the distribution of intensities conformed to the crystal class 622. The systematic absences along the 00*l* line clearly showed the presence of only every sixth reflection, corresponding to the space group $P6_122$ or $P6_522$; the former was subsequently confirmed by MR solution (see above).

The presence of the pseudohexagonal packing of molecules in the orthorhombic form was confirmed by computing the self-rotation function. It had very clear peaks corresponding to the sixfold axis parallel to *c* and six twofold axes in the *ab* plane, 30° apart, almost reaching the height of the origin peaks. However, the average intensity of reflections in the two parity groups, $h + k = 2n$ and $h + k = 2n + 1$, was identical.

The orthorhombic space group $P2_12_12_1$ with such a ratio of *a*:*b* is a subgroup of the hexagonal space group $P6_122$, where the orthorhombic cell axes *a* and *b* are parallel to both diagonals of the hexagonal cell and the origin is shifted appropriately. The relationship between both cells is illustrated in Fig. 3. The packing of molecules in both structures agrees with this relationship. The orthorhombic unit cell is only twice as large as the hexagonal, but the former has six molecules in the asymmetric unit whereas the latter has only one.

The degree of rearrangement necessary for the transition from the less symmetric orthorhombic cell to the more symmetric hexagonal cell was checked by overlapping pairs of corresponding molecules in the two structures. Each of the six individual molecules in the orthorhombic cell required a rotation in the range 1–3° and translation in the range 1–3 Å to adopt a perfect hexagonal constellation, as in the $P6_122$ cell. It can be noted that with the average radius of this rather globular molecule of about 20 Å, a 2° rotation corresponds to a distance of about 0.8 Å at the surface. Therefore, most of the atoms within the orthorhombic cell lie not farther than about 2 Å from the positions corresponding to the more symmetric hexagonal symmetry. It is thus somewhat surprising that this phenomenon was not reflected in the statistics of the intensities of even and odd $h + k$ reflections, even at low resolution.

3.6. Attempts to phase crystal form II using Br data

After solving the structure of PCP in form II (space group $P6_122$) by MR and refining it, we were able to use the calculated phases to

compute the anomalous-difference synthesis based on measured Bijvoet differences in the peak-wavelength data from the NaBr-soaked crystal (Fig. 2*b*). Judging from the statistics of data processing and the appearance of the map, these data contained a certain amount of anomalous signal. There were 19 sites in the map with peak height higher than 5σ, but none above the 10σ level. Those sites correspond to the bromide sites identified around the PCP molecule in the $P6_2$ form, but their occupancies are lower.

Direct-methods attempts to find bromide sites from Bijvoet differences within these data were unsuccessful. We were interested in finding out, however, whether the anomalous signal was sufficient for solving the PCP structure in this crystal form, even though this was clearly not a practical approach to this problem. The ten highest sites identified from the anomalous-difference synthesis were input to *SHARP* and the resulting phase set (FOM 0.372) was submitted to density modification by *DM*, yielding an FOM of 0.767. The resulting map was for the most part interpretable. The corresponding CC values for the final map of this crystal form are given in Table 3 and a portion of the Fourier maps is illustrated in Figs. 1(*g*) and 1(*h*).

The quality of data collected for this crystal form was poorer than for other forms (Table 2) and the refinement of the structure was not satisfactory. As stated above, it was probably caused by the imperfect rapid crystal phase transition. In addition, the occupancies of the bromide sites were lower than in the crystal of form III. The combined effect of imperfect data and the low amount of anomalous signal explains why the attempts to directly identify anomalous scatterers in this form were unsuccessful.

3.7. Analysis of a mercury derivative of crystal form I

The diffraction data collected on the orthorhombic crystal (space group $P2_12_12_1$) soaked in the solution containing mercury also contained a detectable anomalous signal, but attempts to find the anomalous scatterer sites by direct methods failed. The phases obtained from model refinement were used to calculate the anomalous difference Fourier map, which contained a number of significant peaks (Fig. 2*d*). Inspection of the anomalous difference map revealed a high degree of regularity in the mercury sites around the six molecules present in this crystal form. The first six highest peaks within the level 18–15σ correspond to the same site in the six molecules, about 3 Å away from the carboxylate group of a glutamic acid; the next five peaks and peak 18 (13.5–10σ) correspond to the common site near the carboxylate of an aspartic acid; peaks 13–17 and 19 (10.5–10σ) are located again near the same glutamic acid on the other side of the carboxylate group. Among the next group of peaks are five sites near the carboxylate of an aspartic acid. All of these sites are located within the concave cleft containing the putative active site of PCP. There are no free cysteine sulfhydryls in PCP; therefore, none of the usual sites for derivatization with mercury compounds are present. The phasing attempt that was based on the strongest of these sites failed to produce meaningful phases and an interpretable electron-density map; the occupancies of the mercury sites were evidently too low for a solution of this 6 × 40 kDa structure.

4. Conclusions

For more than 5 y, the crystal structure of PCP resisted concerted attempts at solution by molecular replacement or by heavy-atom derivatization. Eventually, the structure was solved by using the anomalous signal of bromide ions soaked into the crystal of one of the existing crystal forms of PCP. It was not possible to use molecular replacement for structure solution, as the amino-acid sequence of PCP did not have a high enough overall similarity to any other proteinases and thus no suitable starting model could be constructed.

PCP resisted derivatization by several different heavy-metal compounds. This was to some extent exacerbated by the large size of the asymmetric unit of the original orthorhombic crystals, which contained six molecules of PCP and comprised more than 200 kDa of protein in total. As shown by the *a posteriori* analysis, when soaked in mercury ethyl phosphate this crystal form bound Hg atoms in several specific sites, but with low occupancy, not sufficient to provide enough phasing power to solve such a large crystal structure.

In this situation, the novel approach of introducing anomalously scattering atoms was used, based on soaking protein crystals in the cryosolution containing bromide ions. Bromide ions diffuse quickly into solvent channels within protein crystals, where they share sites with the solvent waters with partial occupancies. They do not require any chemical reaction for binding and do not require specific coordination around their sites. Their contact with the protein is either through the hydrophobic van der Waals interactions or through hydrogen bonds, mainly involving amide NH proton donors or positively charged arginine or lysine side chains, as well as solvent waters.

The first attempt of bromide cryosoaking with the original crystals of form I gave the unexpected result of a change from orthorhombic $P2_12_12_1$ to the related hexagonal $P6_122$ crystal form II, coupled with the rearrangement of PCP molecules, which adopted more symmetric packing. The new form contained only one molecule of PCP in the asymmetric unit. The three-wavelength MAD data collected from this crystal showed a certain amount of anomalous signal that was not significant enough to lead to structure solution. Direct methods and Patterson searches failed to find the sites of anomalous scatterers. This structure, as well as the structure of form I, was subsequently solved by molecular replacement and it was then possible to inspect the anomalous difference Fourier synthesis. Several bromide sites were identified around the protein surface, but with relatively low occupancies, as judged by the height of corresponding peaks, which did not exceed the 10σ map level. The bromide sites with occupancy approaching unity usually appear at the level 30–40σ (Dauter *et al.*, 2000). The difficulty in phasing this crystal form may be related to the fact that the refinement of its eventually solved structure converged with an *R* factor of

27%, a rather high value for data collected at the synchrotron and with a CCD detector. These effects are probably caused by imperfect ordering of the molecules within the crystal after forced rearrangement of the molecules during a short soak and rapid crystal freezing. The possibility of crystal twinning, which could have a similar detrimental effect on the phasing and refinement of the structure, can be excluded in this case because the original orthorhombic arrangement of molecules does not support the presence of the polar or threefold axes which would be required if the crystal class was lower than 622. Moreover, the intensity statistics do not show significant deviations from the theoretical distribution.

The bromide cryosoaking led to very efficient solution of the structure of crystal form III. Despite the fact that three-wavelength MAD data were collected, only a single-wavelength data set was used for phasing and the atomic model of the virtually complete PCP molecule was obtained in an almost automatic manner by the sequential use of the programs *SHELXD*, *SHARP*, *DM* and *ARP* (with *REFMAC*), all of them run with default parameters. The electron-density map obtained after *ARP* was excellent, with its CC to the final map after careful anisotropic refinement exceeding 0.9.

The phasing based on full three-wavelength MAD data gave intermediate maps and statistical FOM of higher quality than the previous procedure, but required a significantly longer time both for data collection and for phasing. The PCP crystals of form III diffracted to high resolution and did not deteriorate significantly on X-ray irradiation, but the use of a single data set collected at only one wavelength may be beneficial for crystals that show a tendency to suffer from radiation damage during data collection. Moreover, it may be envisaged that collecting single-wavelength data with higher multiplicity of measurements will lead to higher accuracy of estimated intensities, providing a significantly more efficient phasing. The very small anomalous signal of sulfur, of the order of 1.5% of the protein scattering (Hendrickson & Teeter, 1981; Dauter *et al.*, 1999), can be measured accurately and used successfully for phasing, and such an accuracy can be achieved by averaging multiple measurements of intensities of equivalent reflections, thereby minimizing their uncertainties.

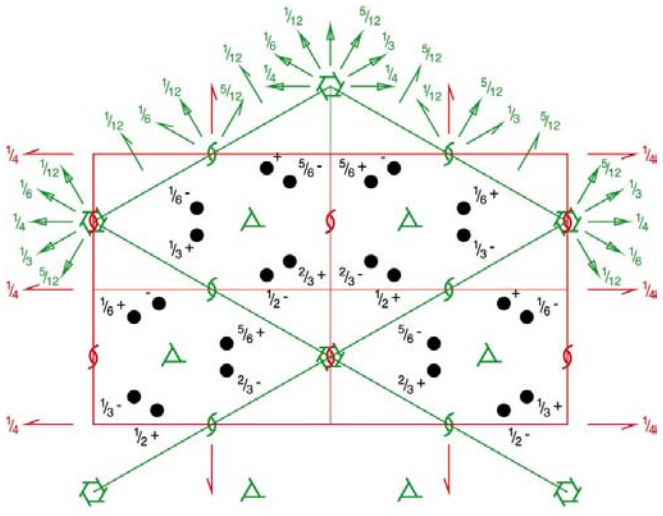

Figure 3
Relationship between crystal forms I (red, space group $P2_12_12_1$) and II (green, space group $P6_122$). A comparison with International Tables Vol. A shows that the $P2_12_12_1$ cell with $b/a = 3^{1/2}$ is a subgroup of the hexagonal $P6_122$ cell. The arrangement of PCP molecules in both cells is analogous and is marked by black dots.

The rapid and efficient phasing of PCP against single-wavelength data is to a large extent the effect of using a number of powerful programs which have been introduced to routine use by the crystallographic community in recent years. In spite of the very sophisticated algorithms involved, contemporary programs such as *HKL*2000, *SHELXD*, *SnB* (Weeks & Miller, 1999), *SHARP*, *SOLVE* (Terwilliger & Berendzen, 1999), *DM* or *ARP* are very user-friendly and easy to run. They require a minimum amount of input and have built-in optimal default parameters, so that their use is almost automatic. However, total reliance on the fully automatic procedures is not recommended, because careful inspection of various stages of structure solution is necessary to avoid mistakes and misinterpretations. Nevertheless, the savings in human time and effort clearly will allow structural biologists to concentrate on other aspects of their work.

The approach involving short cryosoaking in bromide (or iodide) salts and the use of single-wavelength data can be recommended especially for high-throughput projects requiring the rapid solution of many crystal structures. However, it should be stressed that this approach, in common with all methods based on the intrinsically weak anomalous signal, still requires accurately measured intensities. This also applies to the standard MAD method, where the more accurate intensity measurements always lead to more precisely estimated phases and more interpretable experimental maps. Judging from the present and other successful Br-SAD applications (Dauter *et al.*, 2000; Devedjiev *et al.*, 2000), the requirements specified for MAD that $R_{\rm merge}$ should be in the range 3–6% (Hendrickson & Ogata, 1997) also hold for this case.

We wish to thank Dr Kohei Oda for the samples of PCP used in crystallization. We are grateful to Anne Arthur for her editorial comments.

References

Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. New York: Academic Press.
Boggon, T. J. & Shapiro, L. (2000). *Structure*, **8**, R143–R149.
Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Burley, S. K., Almo, S. C., Bonanno, J. B., Capel, M., Chance, M. R., Gaasterland, T., Lin, D., Sali, A., Studier, F. W. & Swaminathan, S. (1999). *Nature Genet.* **23**, 151–157.
Chen, C.-J., Rose, J. P., Rosenbaum, G. & Wang, B. C. (2000). Am. Crystallogr. Assoc. Meet. Abstr., p. 76.
Cowtan, K. D. & Main, P. (1996). *Acta Cryst.* D**52**, 43–48.
Dauter, Z. & Dauter, M. (1999). *J. Mol. Biol.* **289**, 93–101.

Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* D**56**, 232–237.
Davies, D. R. (1990). *Annu. Rev. Biophys. Biophys. Chem.* **19**, 189–215.
Devedjiev, Y., Dauter, Z., Kuznetsov, S. R., Jones, T. L. Z. & Derewenda, Z. S. (2000). *Structure*, **8**, 1137–1146.
Hendrickson, W. A. (1999). *J. Synchrotron Rad.* **6**, 845–851.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Ho, Y.-S. J., Burden, L. M. & Hurley, J. H. (2000). *EMBO J.* **19**, 5288–5299.
Hoover, D., Rajashankar, K. R., Blumenthal, R., Puri, A., Oppenheim, J. J., Chertov, O. & Lubkowski, J. (2000). *J. Biol. Chem.* **275**, 32911–32918.
Ito, M., Narutaki, S., Uchida, K. & Oda, K. (1999). *J. Biochem. (Tokyo)*, **125**, 210–216.
Jones, T. A. & Kieldgaard, M. (1997). *Methods Enzymol.* **277**, 173–208.
Kim, S. H. (2000). *Curr. Opin. Struct. Biol.* **10**, 380–383.
Kwon, H. K., Kim, H. & Ahn, T. I. (1999). *Korean J. Biol. Sci.* **3**, 221–228.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Lamzin, V. S. & Wilson, K. S. (1997). *Methods Enzymol.* **277**, 269–305.
Liu, Z. J., Vysotski, E. S., Rose, J., Rosenbaum, G., Lee, J. & Wang, B. C. (2000). Am. Crystallogr. Assoc. Meet. Abstr., p. 99.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Newton, M. G., Rose, J. P., Liu, Z. J., Foundling, S., Sparks, R. & Wang, B. C. (2000). Am. Crystallogr. Assoc. Meet. Abstr., p. 52.
Oda, K., Nakatani, H. & Dunn, B. M. (1992). *Biochim. Biophys. Acta*, **1120**, 208–214.
Oda, K., Sugitani, M., Fukuhara, K. & Murao, S. (1987). *Biochim. Biophys. Acta*, **923**, 463–469.
Oda, K., Takahashi, T., Tokuda, Y., Shibano, Y. & Takahashi, S. (1994). *J. Biol. Chem.* **269**, 26518–26524.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Oyama, H., Abe, S., Ushiyama, S., Takahashi, S. & Oda, K. (1999). *J Biol. Chem.* **274**, 27815–27822.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Rawlings, N. D. & Barrett, A. J. (1999). *Biochim. Biophys. Acta*, **1429**, 496–500.
Sheldrick, G. M. (1998). *Direct Methods for Solving Macromolecular Structures*, pp. 401–411. Dordrecht: Kluwer Academic Publishers.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Volz, K. (1999). *Protein Sci.* **8**, 2428–2437.
Wang, B. C., Chen, C.-J., Liu, Z. J., Wu, C. K., Schubot, F. D., Rosenbaum, G., Vysotski, E. S., Lee, J., Dailey, H. A., Ferrara, J., Schiffer, M., Pokkulur, P. R., Joachimiak, A., Zhang, R., Howard, A., Chrzas, J., Robbins, A. H. & Rose, J. P. (2000). Am. Crystallogr. Assoc. Meet. Abstr., p. 66.
Weeks, C. M. & Miller, R. (1999). *J. Appl. Cryst.* **32**, 120–124.
Yang, W., Hendrickson, W. A., Crouch, R. J. & Satow, Y. (1990). *Science*, **249**, 1398–1405.