Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of lobster apocrustacyanin A_1 using softer X-rays

M. Cianci,[a] P. J. Rizkallah,[b] A. Olczak,[a] J. Raftery,[a] N. E. Chayen,[c] P. F. Zagalsky[d] and J. R. Helliwell[a]*

[a]Department of Chemistry, University of Manchester, Manchester M13 9PL, England, [b]CLRC, Daresbury Laboratory, Daresbury, Warrington, Cheshire WA4 4AD, England, [c]Biological Structure and Function Section, Division of Biomedical Sciences, Sir Alexander Fleming Building, Imperial College School of Medicine, London SW7 2AZ, England, and [d]Department of Molecular Biology and Biochemistry, Royal Holloway College, University of London, Egham, Surrey TW20 0EX, England

Correspondence e-mail: john.helliwell@man.ac.uk

Received 16 February 2001
Accepted 6 June 2001

PDB Reference: lobster apocrustacyanin A_1, 1h91.

The molecular basis of the camouflage colouration of marine crustacea is often provided by carotenoproteins. The blue colour of the lobster carapace, for example, is intricately associated with a multimacromolecular 16-mer complex of protein subunits each with a bound astaxanthin molecule. The protein subunits of crustacyanin fall into two distinct subfamilies, CRTC and CRTA. Here, the crystal structure solution of the A_1 protein of the CRTC subfamily is reported. The problematic nature of the structure solution of the CRTC proteins (both C_1 and A_1) warranted consideration and the development of new approaches. Three putative disulfides per protein subunit were likely to exist based on molecular-homology modelling against known lipocalin protein structures. With two such subunits per crystallographic asymmetric unit, this direct approach was still difficult as it involved detecting a weak signal from these sulfurs and suggested the use of softer X-rays, combined with high data multiplicity, as reported previously [Chayen *et al.* (2000), *Acta Cryst.* D**56**, 1064–1066]. This paper now describes the structure solution of CRTC in the form of the A_1 dimer based on use of softer X-rays (2 Å wavelength). The structure solution involved a xenon derivative with an optimized xenon $L_{\rm I}$ edge f'' signal and a native data set. The hand of the xenon SIROAS phases was determined by using the sulfur anomalous signal from a high-multiplicity native data set also recorded at 2 Å wavelength. For refinement, a high-resolution data set was measured at short wavelength. All four data sets were collected at 100 K. The refined structure to 1.4 Å resolution based on 60 276 reflections has an R factor of 17.7% and an $R_{\rm free}$ of 22.9% (3137 reflections). The structure is that of a typical lipocalin, being closely related to insecticyanin, to bilin-binding protein and to retinol-binding protein. This A_1 monomer or dimer can now be used as a search motif in the structural studies of the oligomeric forms α- and β-crustacyanins, which contain bound astaxanthin molecules.

1. Introduction

The lipocalins are a superfamily of proteins binding small lipophilic ligands (North, 1989; Akerstrom *et al.*, 2000). The crystal structures of an increasing number of lipocalins are known (reviewed by Flower, 2000) and include human plasma retinol-binding protein (RBP; Newcomer *et al.*, 1984; Cowan *et al.*, 1990; Zanotti *et al.*, 1993; Newcomer & Ong, 2000), β-lactoglobulin (BLG; Papiz *et al.*, 1986; Monaco *et al.*, 1987; Brownlow *et al.*, 1997; Sawyer & Kontopidis, 2000), odorant-binding protein (Monaco & Zanotti, 1992; Blanchet *et al.*, 1996; Tegoni *et al.*, 1996, 2000), rat epididymal retinoic acid binding protein (RABP; Newcomer, 1993; Ong *et al.*, 2000),

bilin-binding protein (BBP; Huber, Schneider, Epp *et al.*, 1987; Huber, Schneider, Mayr *et al.*, 1987), insecticyanin (ICYB; Holden *et al.*, 1987), rat urinary α-2-globulin and mouse major urinary protein (Bocskei *et al.*, 1992; Cavaggioni & Mucignat-Carretta, 2000). They possess a characteristic β-barrel structure consisting of eight antiparallel β-strands of $+1+1+1$ repeated topology arranged in two orthogonal β-sheets with a calyx (cavity) between them. The likely location for ligand binding within the calyx of the lipocalins has been confirmed in the crystal structure of several of the proteins. This site, however, may not be exclusive, since surface ligand-binding pockets have also been identified for oligomeric forms of β-lactoglobulin (Monaco *et al.*, 1987; Sawyer & Kontopidis, 2000) and odorant-binding protein (Monaco & Zanotti, 1992; Tegoni *et al.*, 1996)

The blue colouration of the lobster *Homarus gammarus* is provided by the astaxanthin (3,3′-dihydroxy-β,β-carotene-4,4′-dione) binding protein α-crustacyanin. The absorption spectrum of the carotenoid in the carotenoprotein is red shifted by some 150 nm (Wald *et al.*, 1948). α-Crustacyanin is an octamer of dimeric units. The dimers, termed β-crustacyanins, are combinations of two types of protomers [apoproteins type I (CRTC) and type II (CRTA)] of about 20 kDa. The nomenclature of the apoproteins is based on their electrophoretic mobilities (Quarmby *et al.*, 1977): CRTC comprises apocrustacyanins C_1, C_2 and A_1, and CRTA comprises apocrustacyanins A_2 and A_3.

The amino-acid sequences of the two major subunits A_2 and C_1 have been determined (Keen *et al.*, 1991*a*,*b*) and shown to possess the two consensus sequences characteristic of the lipocalin superfamily (Pevsner *et al.*, 1988) as well as a third consensus sequence containing a conserved Arg (Flower *et al.*, 1993, 2000). On this basis, the tertiary structure of A_2 has been modelled on the known coordinates of RBP and that of C_1 modelled on the resultant A_2, with the disulfide-bond arrangement suggested corresponding to that found in BBP and insecticyanin (Keen *et al.*, 1991*a*,*b*). The N-terminal amino-acid sequences of A_2 and A_3 are identical for at least 47 residues; likewise, at least 30 residues are identical at the N-terminal sequences of A_1, C_1 and C_2 (J. B. C. Findlay, personal communication). Significantly, only a single gene has so far been isolated for each of CRTA and CRTC (A. Cox and J. B. C. Findlay, personal communication). The differences between apo proteins within the subsets CRTA and CRTC are therefore likely to be the result of post-translational modification or alteration during isolation, for example by amide hydrolysis.

We recently reported (Chayen *et al.*, 2000) the space group of crystals of apocrustacyanin A_1 and suggested a novel structure-determination approach (see §5 of that paper). A new approach was needed for apocrustacyanin because the different avenues of 'standard' protein structure determination had not been successful, namely molecular replacement involving a lipocalin protein search model and a heavy-atom derivative search. Moreover, the protein contains no methionine for selenomethionine protein production for use with the MAD technique. Chayen *et al.* (2000) reported encouraging preliminary measurements utilizing softer X-rays to enhance the anomalous scattering of the expected disulfide S atoms so as to locate their positions as a substructure of the protein. If successful, this model could then have been combined with a 0.95 Å data set measured on the ESRF by our collaborators (Gordon, McSweeney & Leonard, personal communication). At that stage, our measured sulfur anomalous signal-to-noise ratio was not high enough to yield clear positions for these S atoms; the need for increased data multiplicity was identified in Chayen *et al.* (2000).

Table 1
Diffraction data statistics for (*a*) the native data set to 2.3 Å resolution, (*b*) the xenon data set to 2.3 Å resolution, (*c*) the high-multiplicity native sulfur data set to 2.7 Å resolution, (*d*) the high-resolution native data set to 1.35 Å resolution.

Values in parentheses are for the appropriate highest resolution shell.

Data set	Native (*a*)	Xe derivative (*b*)	Sulfur (high multiplicity) (*c*)	High-resolution native (*d*)
Station at SRS	7.2	7.2	9.5	9.6
Wavelength (Å)	2.045	2.045	2.0	0.87
Detector	MAR345	MAR345	MAR CCD	ADSC CCD
Distance (mm)	109.4	109.4	83.3	118.6
Oscillation range per image (°)	1.0	1.0	1.0	1.0
No. of images	187	200	411	100 × 2
Resolution (Å)	64.5–2.3	64.5–2.3	64.5–2.7	64.5–1.35
Total No. of reflections	87610	118157	153210	330300
No. of unique reflections	16278	16703	10392	72445
Data multiplicity	5.4	7.1	14.7	4.6
Highest resolution shell (Å)	2.38–2.30	2.38–2.30	2.80–2.70	1.38–1.35
Overall completeness (%)	98.0 (99.0)	99.7 (100.0)	99.1 (99.7)	87.1 (79.4)
$R_{\rm merge}$ (%)	7.6 (15.0)	7.3 (14.5)	5.8 (14.4)	4.7 (13.1)
Overall $I/\sigma(I)$	19.7 (13.4)	23.4 (12.3)	48.5 (22.3)	31.1 (7.5)
Reflections with $I/\sigma(I) > 3$ in the highest resolution shell (%)	84.1	80.4	86.5	76.6
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)				
a	41.06	41.11	41.14	41.65
b	79.46	79.81	79.75	80.68
c	109.66	109.86	109.73	110.84

2. Methods

2.1. Crystallization

Crystals of the A_1 subunit of lobster crustacyanin were grown in hanging drops at 291 K over reservoirs containing 0.1 *M* Tris–HCl pH 9.0, 5% MPD (2-methyl-2,4-pentanediol), 1 m*M* EDTA and 1.9 *M* ammonium sulfate (Wright *et al.*, 1992; Chayen *et al.*, 2000). The final volume of the drops was 2 µl, made up by mixing equal volumes of the protein solution (containing 20 mg ml^{-1} A_1 subunit in 0.1 *M* Tris–HC_1 pH 7.0 and 1 m*M* EDTA) and of the reservoir solution.

It is interesting to note that in different drops under identical conditions (described above) three different forms of crystals were obtained, yet X-ray analysis showed that all the forms belong to the same space group with essentially identical unit-cell parameters. The various forms may have

occurred owing to a different area of contact between the crystal drops and their supporting cover slips, thereby generating different nucleation rates as they grew.

2.2. Data collection and analysis

2.2.1. Strategy. A total of four data collections (Table 1) were carried out at the SRS in Daresbury (England). The first was a native data set, the second was a Xe derivative, the third was a native with a high data multiplicity so as to obtain significant sulfur anomalous differences (henceforth called the sulfur data set) and the fourth was a native data set to high resolution. One phasing option included the SIROAS approach involving xenon and native with the wavelength chosen to optimize the xenon $L_{\rm I}$ edge f'' (wavelength 2.27 Å). Alternatively, the high-multiplicity sulfur data set, recorded at a wavelength around 2 Å, could yield the expected disulfide positions and partial phasing information, which then could be combined with atomic resolution data for a combined direct-methods approach. A wavelength of 2 Å was the longest practicable to allow reasonable diffraction resolution for the detector apertures available. We have termed 2 Å wavelength 'softer' X-rays, since the tradition for diffraction involves Cu $K\alpha$ radiation (1.54 Å) or harder X-rays delivered by synchrotron sources; soft is usually taken to mean 5 Å or longer wavelength. The high-resolution data set used a short wavelength to access the highest resolution possible with these crystals, to be used for the detailed protein model refinement.

2.2.2. Experimental details.

(i) *Native*. The crystal was mounted in a nylon loop and flash-cooled with a nitrogen-gas stream at approximately 100 K. No cryoprotectant was used. X-ray data collection was undertaken at station 7.2 (Helliwell *et al.*, 1982), where the Ge(111) single-crystal monochromator was set to reflect X-rays at λ = 2.045 Å so as to match the wavelength used in the xenon data set. A MAR 345 image-plate detector system was used and 187 images of 1.0° each were collected with exposure times of ∼7 min per frame in dose mode.

(ii) *Xe derivative*. The crystal was pressurized in 1419 MPa xenon gas using an OxfordXcell for 1 h and then flash-cooled. No cryoprotectant was used. The same data-collection protocol as the native data set was followed, also on station 7.2, at 2.045 Å wavelength. A total of 200 × 1° images were collected with exposure times of ∼7 min per frame in dose mode.

(iii) *High-multiplicity sulfur data set*. The crystal was mounted in a nylon loop and flash-cooled. No cryoprotectant was used. It was exposed to X-rays at station 9.5, where the double-crystal Si(III) monochromator was set to reflect X-rays at λ = 2.000 Å. To remove the 0.67 Å harmonic, a post-mirror was used to reflect only the 2 Å wavelength beam. A MAR CCD detector system was used to collect 411 high-quality images of 1.0° each with exposure times of 50 s.

(iv) *High-resolution data*. The crystal was mounted in a nylon loop and flash-cooled. No cryoprotectant was used. It was exposed to X-rays at station 9.6 with λ = 0.87 Å. An ADSC CCD detector system was used to collect 100 images of 1.0° each with exposure times of 90 s and the same 100 images with 10 s exposures (quick pass).

2.2.3. Data processing. The *HKL* package (Otwinowski & Minor, 1997) was used for the integrating, scaling and merging of each of the four data sets. The unit-cell parameters (Table 1)

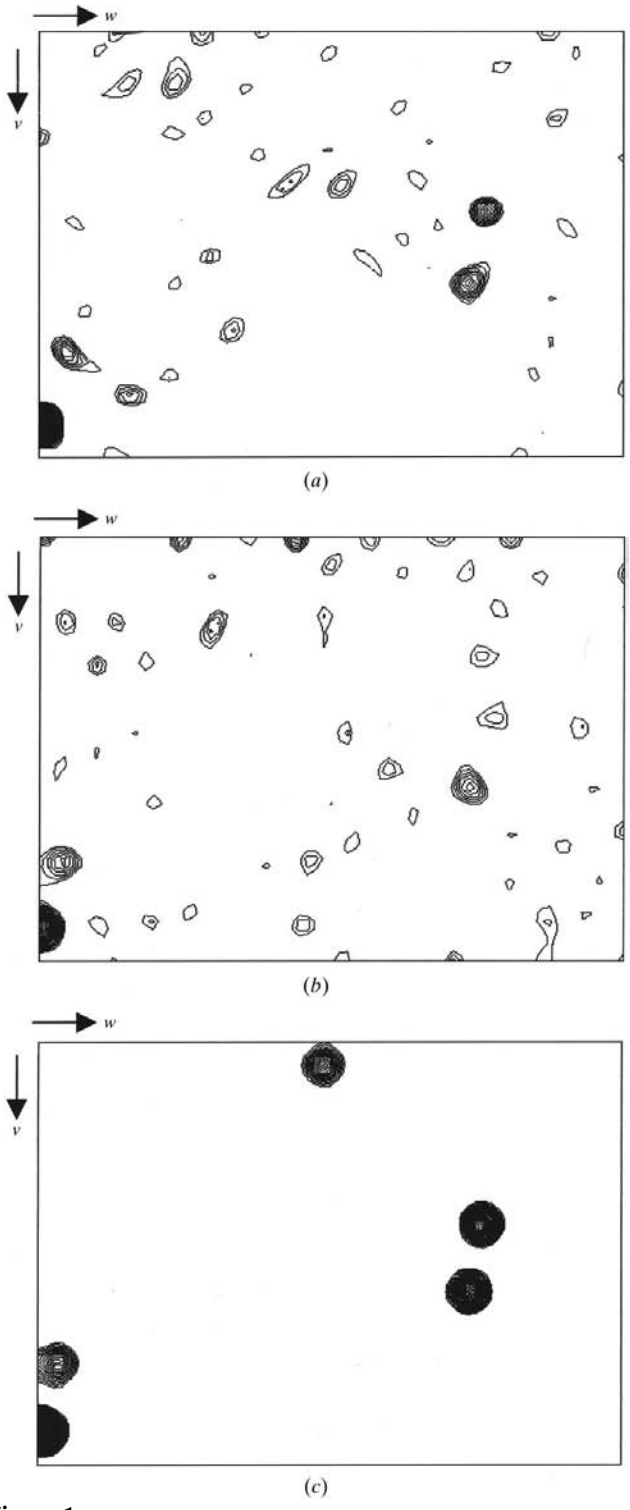

Figure 1
Xenon difference Patterson Harker sections $u = \frac{1}{2}$: (*a*) anomalous, (*b*) isomorphous, (*c*) calculated.

were similar to each other and to those of apocrustacyanin A_1 at room temperature (Chayen *et al.*, 2000). The isomorphism of the native and the Xe derivative proved to be successful for SIROAS phasing (see below). Table 1 summarizes the data-collection and processing statistics.

2.3. Locating the xenon sites, phasing using the xenon SIROAS data, determining the hand and locating the disulfides

The anomalous and isomorphous difference Patterson maps Harker sections clearly indicated the presence of Xe atoms with good agreement between the two. There were strong peaks on $u = 0$ and $u = \frac{1}{2}$ which suggested one of the two space groups $P2_12_12_1$ or $P22_12_1$. For space group $P2_12_12_1$, using the xenon anomalous differences as input to *Shake-and-Bake* (Weeks & Miller, 1999), the program produced the coordinates of three sites with a very clear bimodal distribution having in excess of 250 correct solutions for 5000 trials. There was no solution for the other space group. The calculated xenon Patterson was in accord with the observed difference Harker sections (Fig. 1).

Phases were calculated with *MLPHARE* (Otwinowski, 1991; Collaborative Computational Project, Number 4, 1994) to 2.3 Å resolution using the *Shake-and-Bake* solution xenon sites. A repeat calculation with the xenon coordinates on the inverted hand was also made. Each of these two phase sets was then used with the high-multiplicity sulfur data set anomalous differences. The disulfides were clearly found (Figs. 2*a*, 2*c* and 2*d* *versus* 2*b*) in the opposite hand to that selected by *Shake-and-Bake*. Thus, the hand of the phases was established. On the correct hand, a xenon isomorphous difference Fourier map revealed a fourth site of xenon substitution. This fourth Xe

Table 2
Xenon-derivative and sulfur-site parameters.

S atoms and S—S bridges were refined against anomalous data only, so the occupancy factor (OCC) for them is zero. AOCC is the anomalous occupancy. Temperature factors, *B*, were constrained to 10.0 for the S atoms and 40.0 for the unresolved S—S bridges. In the apo A_1 crystal structure, the two Xe atoms are 6.65 Å apart and the closest residues are Arg28 and Tyr31 (α-helix), Phe63 (β-strand *C*), Leu83 and Pro85 (β-strand *D*), Leu93 (β-strand *E*) and Ile107 (β-strand *F*2).

	OCC	AOCC	*B*	*x*, *y*, *z*		
Xe1	0.318	0.455	13.7	−0.62360	−0.52335	−0.50331
Xe2	0.419	0.590	21.6	−0.87037	−0.55929	−0.99017
Xe3	0.268	0.337	18.2	0.65256	−0.76443	−0.87725
Xe4	0.080	0.122	12.4	0.47078	0.20858	0.15206
S1	−0.022	1.294	10.0	0.47920	0.38996	0.33749
S2	0.012	1.663	10.0	0.48712	0.36670	0.34990
S3	−0.059	1.611	10.0	0.21479	0.32502	0.49103
S4	0.005	1.444	10.0	0.24172	0.33534	0.50628
S5	−0.074	1.334	10.0	0.26966	0.12002	0.89390
S6	−0.043	1.239	10.0	0.26590	0.11325	0.91316
S7	−0.067	1.218	10.0	0.10455	0.41033	0.63125
S8	−0.001	1.173	10.0	0.11466	0.42150	0.64724
SS1	0.012	0.910	40.0	0.36220	0.08760	0.61820
SS2	0.007	0.910	40.0	0.24651	0.25846	0.78495

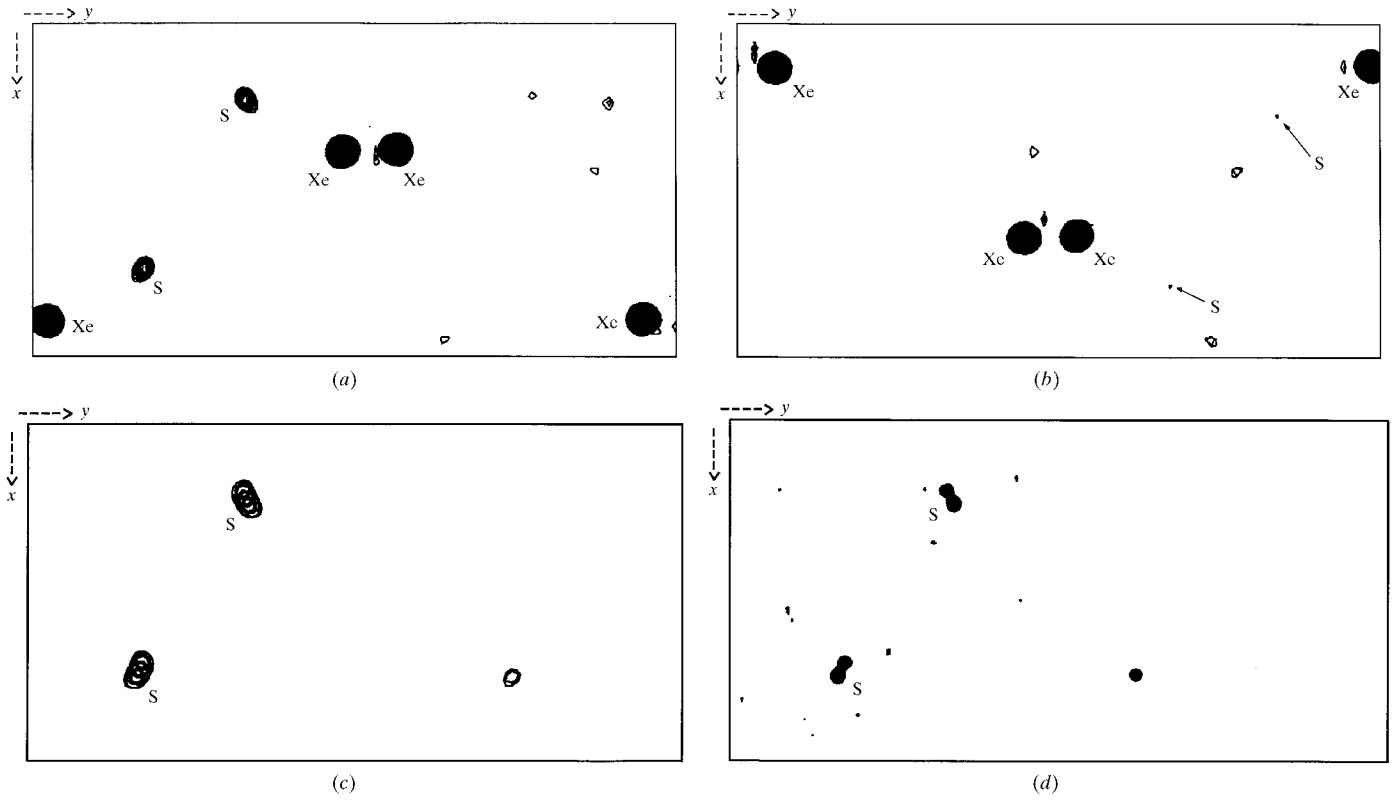

Figure 2
Fourier maps calculated for the same slabs of the unit cell. (*a*) Using the xenon data set anomalous differences (which includes the disulfide anomalous signal) at 3 Å resolution. (*b*) Using the xenon data set anomalous differences with phases on the opposite (*i.e.* incorrect) hand at 3 Å resolution. (*c*) Using the native 'high-multiplicity' sulfur anomalous data set at 2.7 Å resolution. (*d*) Calculated based on the final refined S-atom positions at 2 Å resolution.

Table 3
Statistics for side-chain docking and building with *ARP/wARP*.

Based on the sequence of A_1 monomer	Apocrustacyanin C_1		Apocrustacyanin A_2	
	A	*B*	*A*	*B*
Average confidence level (%)	88	91	22	24
Side chains built completely	88	91	22	35
Percentage of chains built completely	54.7	57.2	13.7	22.0

site was added into the phase calculation using *MLPHARE* (Table 2). The figure of merit for the phasing was 0.46 to 2.3 Å resolution. The Xe and the disulfide-atom sites (Table 2) were used to generate the full contents of a unit cell, which were then examined on the graphics. The calculated non-crystallographic twofold axis for the six disulfides agreed with the observed one (Chayen *et al.*, 2000). The orientation and position of the molecular twofold non-crystallographic axis was clearly established.

2.4. Phase improvement

A final set of phases was calculated with *MLPHARE* using four Xe atoms and the disulfide atoms (Table 2). Phase improvement was undertaken using *DM* (Collaborative Computational Project, Number 4, 1994) all at 2.3 Å. The SIROAS + *DM* mean figure of merit was 0.76. At this resolution, some attempts were made to automatically trace the main chain using the *ARP/wARP* suite (Perrakis *et al.*, 1999) but were not successful, presumably owing to the observations-to-parameters ratio being only around 1.5. The situation was also complicated, however, by uncertainty over the amino-acid sequence (see below). As soon as the data set at high resolution was available, the phases were extended to 1.35 Å using *DM*; the following run of *ARP/wARP*, with a starting ratio of observations-to-parameters now close to 8.0, provided the main-chain trace.

2.5. Model building and refinement

The output model from *ARP/wARP* for the expected dimer consisted of nine chains and 320 residues (out of 360) with an

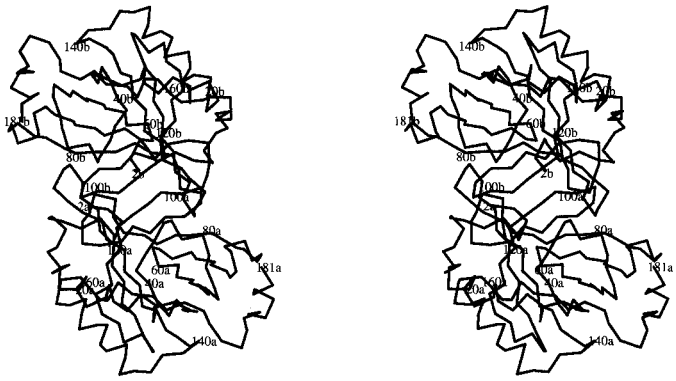

Figure 3
Apocrustacyanin A_1 dimer (Cα trace in stereo) viewed down the twofold non-crystallographic axis.

Table 4
Summary of the structure-fitting and refinement cycles.

Program	Job	No. of waters	*R* factor (R_{free}) (%)
DM	Phase extension to 1.35 (Å)	—	—
ARP/wARP	Automatic main-chain tracing	—	—
ARP/wARP	Automatic side-chain tracing	—	—
LSQKAB	Manual rebuilding of the dimer	—	—
REFMAC	Refinement	—	37.6 (39.2)
O/*OOPS*2	Adjustment main/side chains	—	—
REFMAC	Refinement	—	33.4 (35.1)
O/*OOPS*2	Minor adjustment of side chains	—	33.1 (34.8)
ARP/wARP	Solvent addition	676	26.5 (30.3)
O/*OOPS*2	Adjustment of main/side chains, removal of waters with $B > 24$	—	—
REFMAC	Refinement	291	28.5 (30.7)
ARP/wARP	Solvent addition	616	27.1 (29.8)
O/*OOPS*2	Adjustment of main/side chains, removal of waters with $B > 26$	—	—
REFMAC	Refinement	426	26.2 (28.3)
ARP/wARP	Solvent addition	619	26.7 (28.2)
O/*OOPS*2	Adjustment of main/side chains, removal of waters with $B > 28$	—	—
CNS	Refinement	427	26.1 (29.0)
O/*OOPS*2	Adjustment of main/side chains	—	—
CNS (resolution 1.51 Å)	Refinement	427	25.8 (28.0)
O/*OOPS*2/ omit maps	Adjustment of main/side chains	—	—
CNS	Refinement	420	25.3 (27.7)
O/*OOPS*2	Adjustment of main/side chains	—	—
CNS (ice-ring reflections removed)	Refinement	420	23.5 (27.3)
O/*OOPS*2	Adjustment of main/side chains, removal of waters	—	—
ARP/wARP	Solvent addition	490	21.9 (26.7)
*SHELX*97 (iso, resolution 1.51 Å)	Refinement	490	19.8 (24.2)
O/*OOPS*2	Adjustment of main/side chains	—	—
*SHELX*97 (aniso, resolution 1.4 Å)	Refinement, aniso refinement on main-chain atoms only	479	18.1 (23.6)
O/*OOPS*2	Multiple conformations, adding residues 2 (*A*/*B*)	—	—
*SHELX*97	4 MPD molecules added	442	17.7 (22.9)

R factor of 27.3%. The side chains were automatically built using the *ARP/wARP* scripts for sequence docking and side-chain building. The sequence of apocrustacyanin A_1, which is unknown, was postulated to be close to C_1 (Keen *et al.*, 1991*b*), since both proteins are of the type I of this family according to amino-acid composition, size estimation and peptide mapping (Zagalsky, 1983). Therefore, two tests were run using as templates firstly the deposited apocrustacyanin C_1 sequence (Keen *et al.*, 1991*b*) and secondly the deposited apo A_2 sequence (Keen *et al.*, 1991*a*). *ARP/wARP* provided a clear solution consistent with the C_1 sequence (Table 3).

The reconstruction of the two monomers continued manually using *O* (Jones *et al.*, 1991) and *OOPS*2 (Kleywegt & Jones, 1996). After using the refinement program *REFMAC* (Collaborative Computational Project, Number 4, 1994) and *ARP/wARP* for solvent building, the *R* factor and R_{free} were 26.7 and 28.2%, respectively, to 1.35 Å. To speed up the calculations, the resolution was then limited to 1.51 Å and refinement was continued using the *CNS* package (Brunger *et al.*, 1998) to take advantage of the simulated-annealing

approach, always alternating sessions on the graphics with *O*/*OOPS*2.

At this point, several problems with the refinement procedure were addressed. Firstly, there were ice-ring reflections in the resolution ranges 1.93–1.97 and 2.25–2.32 Å, which had very poor *R* factors and $R_{\rm free}$ values. Therefore, 600 reflections in these rings which had $|F_o - F_c| > 5\sigma$ were eliminated.

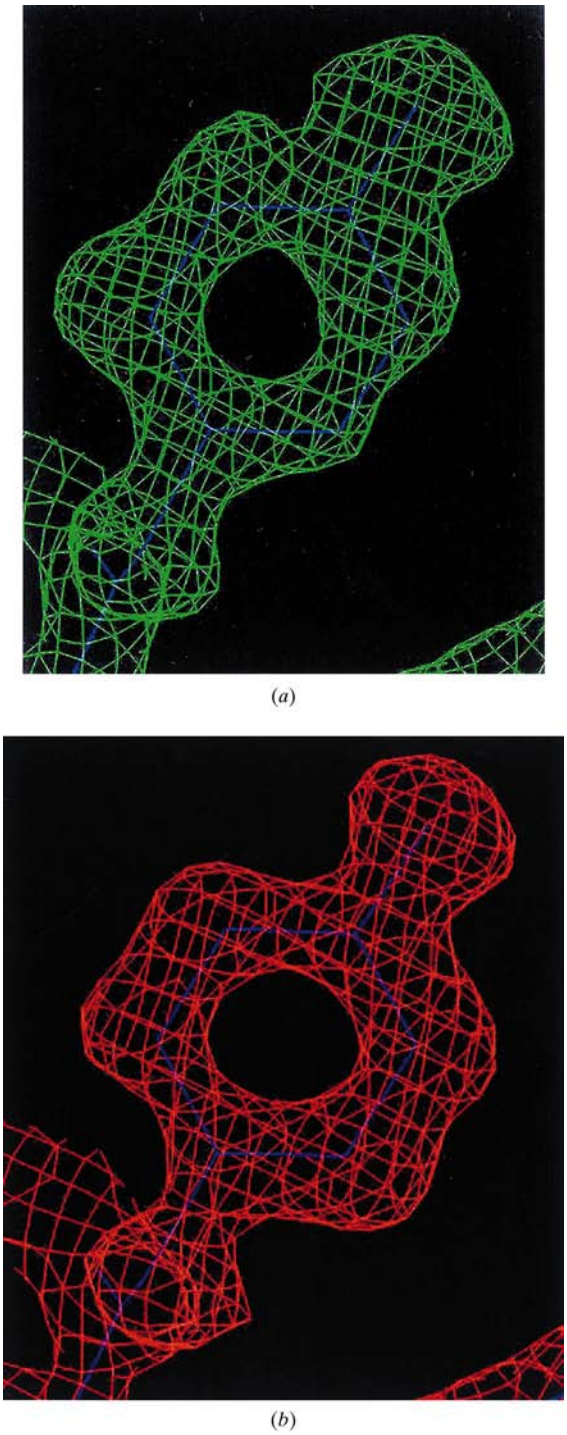

Figure 4
The electron-density maps for Tyr112, contoured at 2 r.m.s. (*a*) SIROAS + *DM* phases to 2.3 Å and phase extension by *DM* to 1.35 Å resolution. (*b*) $2F_o - F_c$ after refinement at 1.4 Å resolution.

Table 5
Final refinement statistics.

Values in parentheses are for the highest resolution shell.

No. of atoms	
Protein	3030
Solvent	442
Non-solvent	32
Resolution range (Å)	1.40–36.3 (1.40–1.46)
R factor for all data (%)	17.7 (18.4)
Free *R* factor for all data (%)	22.9
Goodness of fit (*SHELX*)	2.815
Restrained goodness of fit	2.313
R.m.s. deviations	
Bond length (Å)	0.011
Bond angles (Å)	0.028
Isotropic $B_{\rm eq}$ values† (Å^2)	
Main chain	1.51
Side chain	3.90
Average *B* values† (Å^2)	
Main chain	13.7
Side chain	17.9
NCS r.m.s.‡ (Å)	0.53
Double conformations modelled	31

† R.m.s. and average *B* values calculated with *BAVERAGE* (Collaborative Computational Project, Number 4, 1994). ‡ NCS r.m.s.d. calculated with *LSQKAB* (Collaborative Computational Project, Number 4, 1994).

Secondly, the problem stretches of polypeptide found at the loops at residues 14–16 and 59–61 in both chains were examined in detail *via* omit maps. Thirdly, a significant number of further bound waters were expected for a protein dimer of this molecular weight (Carugo & Bordo, 1999). More bound waters were added with *ARP*/*wARP*. As a result of these three amendments to the refinement, the *R* factor and $R_{\rm free}$ fell to 21.9 and 26.7%, respectively. At this point, the data were extended to 1.4 Å resolution and the refinement continued

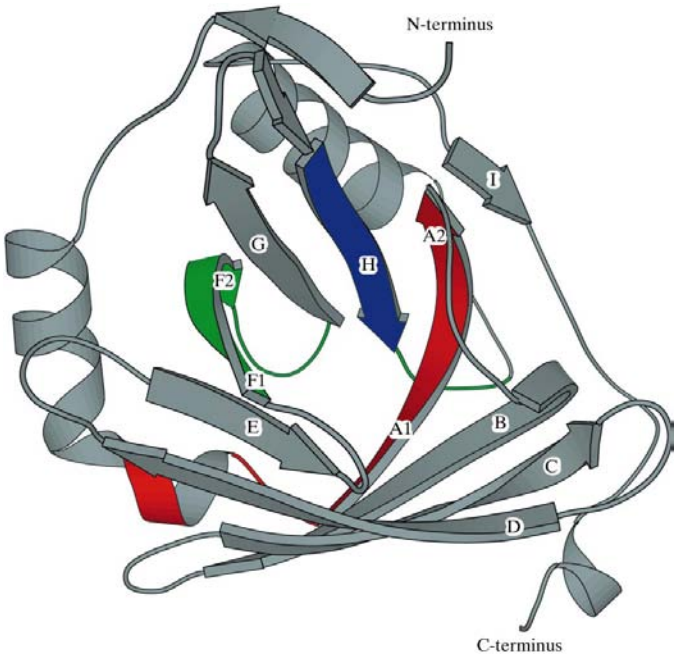

Figure 5
Ribbon diagram of a monomer with the consensus sequences highlighted: red for consensus sequence 1, green for stretch 2, blue for stretch 3.

using *SHELX*97 (Sheldrick, 1997, 1998). Finally, the N-terminus and the C-terminus regions were rebuilt adding residues Lys2 and Val181 in each monomer and multiple conformations for 31 residues were added.

A summary of the refinement is given in Table 4. The final R factor and R_{free} following anisotropic temperature-factor refinement (main-chain atoms only) using data to 1.4 Å were 17.7 and 22.9%, respectively, calculated on all reflections. The final model calculated phases were compared with the SIROAS + *DM* phases to 2.3 Å and the mean phase error was 38.4° (pre-*DM* it was 53.5°). Table 5 lists the final refinement statistics.

3. Description of the molecule

3.1. Amino-acid sequence

The model of apocrustacyanin A_1 in the asymmetric unit consists of a homodimer with a monomer of approximately 40 Å in diameter related by a twofold non-crystallographic axis (Fig. 3). The complete A_1 amino-acid sequence was determined from the high-resolution model and omit electron-density maps. To simplify the cross-referencing between the structure of apo A_1 presented in this paper and the structure of apo C_1 presented in the companion paper (Gordon *et al.*, 2001), the numbering of the sequence has been kept the same as that of apo C_1. Hence, in the A_1 electron-density map, the N-terminus starts at residue 2. Residue 1 is not seen in either subunit of the dimer. There are two further apparent changes to the A_1 sequence determined from the electron-density map compared with that published for the C_1 protein (Keen *et al.*, 1991*b*). The first discrepancy from the apo C_1 sequence is an apparent mutation of residue 5 from Asp to Asn. The $2F_o - F_c$ map shows a difference in the electron density between O and N suggesting this change. Similarly, another change appears to have been made of Ile181 to Val181 (the C-terminus).

There are two loop regions, residues 14–16 and 59–61, both of which were better ordered in the '*B*' chain. Hence, on the basis of this one monomer only, the amino-acid assignment was confirmed as that of C_1 for these two loops. Overall, the amino-acid sequence is in good agreement with that deposited for apocrustacyanin C_1 (Keen *et al.*, 1991*b*) apart from the three amino acids detailed above.

3.2. Structure

3.2.1. The monomer. The final model for the dimer consists of 360 amino-acid residues, 442 ordered waters and four MPD (2-methyl-2,4-pentanediol) molecules. The deviation from ideality of the protein is 0.011 Å on bond lengths, 0.028 Å on distances derived from angles and 0.0292 Å on distances from restrained planes. The average r.m.s. deviation between all Cα atoms of the two monomers is 0.53 Å. The Ramachandran (Ramachandran *et al.*, 1963) plot of apo A_1, calculated with *PROCHECK* (Laskowski *et al.*, 1993), has 293 non-glycine residues in the most favoured regions (91.6%), 25 residues in the additional allowed regions (7.8%), no residues in generously allowed regions (0.0%) and two residues in the disallowed regions (0.6%). The outlier residues are Tyr112 visible on both chains *A* and *B*. It is a feature of other members of the lipocalin family to have torsion angles of this residue in a disallowed region of the Ramachandran plot (Cowan *et al.*, 1990; Zanotti *et al.*, 1993). The high quality of the electron density around Tyr112 is shown in Fig. 4.

Table 6
Secondary-structure pattern of apocrustacyanin A_1.

Residues	Type of secondary structure	Consensus region	Label	Belonging to β-sheet
7–10	β-Strand			
11–16	Hairpin			
17–31	α-Helix	First		
33–41	β-Strand	First	*A*1 (33–36)	I
			*A*2 (36–41)	II
42–47	Hairpin		*L*1	
48–58	β-Strand		*B*	I
59–62	Hairpin		*L*2	
63–71	β-Strand		*C*	I
72–76	Hairpin		*L*3	
77–85	β-Strand		*D*	I
86–92	Hairpin		*L*4	
93–97	β-Strand		*E*	I
98–101	Hairpin		L5	
102–110	β-Strand	Second	*F*1 (102–106)	I
			*F*2 (107–110)	II
111–114	Hairpin	Second	*L*6	
115–123	β-Strand		*G*	II
124–127	Hairpin		*L*7	
128–137	β-Strand	Third	*H*	II
138–143	Hairpin	Third		
144–156	α-Helix			
164–166	β-Strand		*I*	
167–174	Hairpin			
175–178	α-Helix			

The A_1 monomer is one single domain made up of an N-terminal β-strand, one α-helix, eight antiparallel β-strands (*A*–*H*), one α-helix, one β-strand (I) and at the C-terminus

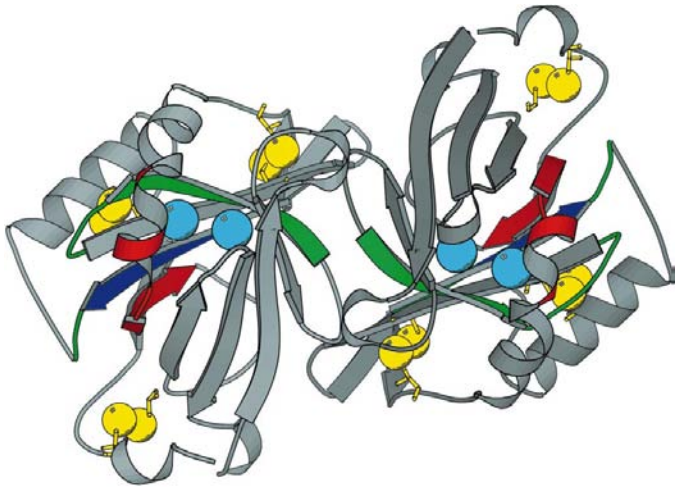

Figure 6
Apocrustacyanin A_1 dimer showing the positions of the bound Xe atoms (in cyan) and the disulfide bonds (in yellow) viewed down the molecular twofold-symmetry axis. The consensus regions use the same colour code as in Fig. 5.

Table 7
Agreement between related lipocalins.

Apo A_1 crustacyanin *versus*	R.m.s. Å (No. of Cα atoms compared)
Apo C_1 predicted model	2.5 (157)
Bilin-binding protein (BBP)	2.5 (158)
Insecticyanin (ICYP)	2.4 (163)
Retinol-binding protein (RBP)	2.3 (154)
β-Lactoglobulin (BLG)	2.8 (133)

ending with one turn of α-helix (Table 6). The common topology for lipocalins of a calyx-shaped antiparallel β-barrel is formed by β-sheets I and II, which are almost orthogonal to each other (Flower *et al.*, 1993). In Fig. 5, β-sheet I is seen to consist of β-strands *A*1, *B*, *C*, *D*, *E* and *F*1, while β-sheet II is made up of strands *A*2, *F2*, *G* and *H*. The sharing of strand *A* between β-sheet I and II is realised at residues Trp35, Tyr36 and Gln37, with a smooth curvature that preserves the hydrogen-bonding pattern. These residues are part of the 'first structurally conserved region' (SCR1) of the lipocalin-family members (Flower *et al.*, 1993). The 'second structurally conserved region' (SCR2) starts with the end of strand *F*1 and continues through *F*2 and the following hairpin. The curvature in the *F* strand defines *F*1 and is caused by Pro104, Leu105 and Val106. The 'third structurally conserved region' (SCR3) coincides with strand *H* and part of the following hairpin; it is this strand which forms the pattern of hydrogen bonds with strand *A*2. The three structurally conserved regions characteristic of lipocalins are highlighted in Fig. 5.

The N-terminus wraps around β-sheet II, where it is firmly bound to strand *G* by the disulfide bridge between Cys12 and Cys121. The C-terminus wraps around β-sheet I and is fixed to strand *B* with the disulfide bridge between Cys51 and Cys173. The third disulfide bridge between Cys117 and Cys150 connects strand *G* of β-sheet II to the α-helix (residues 144–156) present in the C-terminus. The α-helix (residues 17–31) present in the N-terminus lies beside β-strand *F* close to one side and part of the bottom of the pocket created by β-sheets I and II.

3.2.2. The dimer. The dimer (Fig. 6) is formed by facing the β-strand *F*1 of both of the monomers *A* and *B* to form a weakly interacting continuum of antiparallel β-sheet with hydrogen bonds involving only the carbonyl O atom and amide H atom of Ala102 in both chains. Additionally, the interaction is reinforced by hydrogen bonds created with the side chains of Asp90 in one chain and Gln99 in the other. Moreover, a network of shared waters between the monomers extends the hydrogen-bonding pattern. It is also worth noting that a large number of interface hydrophobic residues (N-terminus, Phe6 and Val7; hairpin *L*1, Tyr45; β-strand *D*, Tyr84; hairpin *L*4, Phe88; hairpin *L*5, Phe101; β-strand *G*, Ile122; hairpin *L*7, Tyr124 and Phe126; β-strand *H*, Phe132) provide additional interaction between monomers.

The crystal packing of the dimer of apo A_1 might provide clues as to the possible kind of oligomerization that is expected to form the α-crustacyanin, but in fact has not. The unit cell of apo A_1 is made up of four dimers whose basic interactions are generated by the N-terminus (residues 9, 10), α-helix 22–26, hairpin *L*2 (59,60), *L*3 (73), *L*4 (91), *L*6 (113), *L*7 (127), α-helix 139–141 and the C-terminus (residues 175–178).

4. Discussion

4.1. Techniques

The combination of softer X-rays and Xe derivatization *via* the SIROAS approach yielded high-quality anomalous differences from the xenon-binding sites, of which three were well substituted. The quality of the xenon anomalous difference Patterson was high and somewhat superior to the isomorphous difference Patterson map. The determination of the hand of the xenon SIROAS + *DM* phases was clearly established in the anomalous difference Fourier maps based on the high-multiplicity sulfur anomalous difference data set, which even showed partially resolved S atoms.

The ease of working at the softer X-ray wavelength of 2 Å was notable, as also recently noted by Weiss *et al.* (2001). Possible systematic errors of absorption could have been occurring at several stages. Firstly, at room temperature a xenon atmosphere of 1419 MPa could have been an obstacle, but by the use of cryocooling, the path of the X-ray beam was simply that of the crystal. The size and shape of the crystal in the loop was thus not a limitation for the xenon-bound data set. Secondly, the path to the detector and the transmission through the detector-entrance window was not a problem either for the MAR 345 on station 7.2 (as was reported in Chayen *et al.*, 2000) nor on the MAR CCD on station 9.5. (The precise details of the transmission of 2 Å X-rays through the detector entrance windows are not known but possibly could be optimized somewhat, *i.e.* the windows made thinner.)

The separate step of automatic tracing of the electron-density map required higher resolution data and the phase-extension capability of *DM*. The available detector apertures, in a flat normal-incidence geometry, for 2 Å wavelength, restricted the data resolution to 2.3 Å. In this study, we opted to measure a new data set on a new crystal at short wavelength to obtain high-resolution data to 1.35 Å. The availability of the high-resolution data was also very important in discerning the small post-translational sequence changes that could not be conclusively resolved with biochemical methods.

4.2. Structural and sequence comparisons

4.2.1. Overall. The r.m.s.d. Cα coordinate difference values (calculated using the DALI server; Holm & Sander, 1995) between apocrustacyanin A_1, the apo C_1 predicted model and the other lipocalin proteins whose structures have been solved are given in Table 7. The apo C_1 model offered a good prediction for β-strands *A*1 and *A*2, *B*, *D*, *F*1 and *F*2, *G* and *H* (Fig. 7). The X-ray structure of retinol-binding protein (RBP) appears to be the most similar to A_1, followed by insecticyanin. Other similar candidate protein structure folds are bilin-binding protein (BBP) and β-lactoglobulin (BLG). Fig. 8

shows an overlay of the Cα-atom traces of the A_1 and the insecticyanin monomers.

The amino-acid sequence comparison (Fig. 9) among members of the lipocalin family shows three consensus regions, despite a very low (around 20%) overall sequence homology. Both Figs. 5 and 6 show the consensus regions colour coded; Fig. 5 (and Table 6) use the standard lipocalin secondary-structure labels. The general liopcalin structural motif (double β-sheets formed by eight β-strands) is seen again here with apocrustacyanin A_1.

4.2.2. Disulfide bridges. During our preliminary studies in the search for the likely disulfide bridges based on the anomalous signal of sulfur, it was important to know how many disulfide bridges might be present. Therefore, since all the three-dimensional structures of the solved members of the lipocalin family contain disulfide bridges, all the structures were investigated to check the possibility of disulfide bridges in apo C_1 and apo A_1.

BLG has five cysteine residues (66, 106, 119, 121, 160) in its amino-acid sequence and there are S—S bridges between residues 106 and 119 and between residues 66 and 160, with no evidence of the Cys106–Cys121 bond previously suggested (McKenzie *et al.*, 1968). Therefore, Cys121 in BLG remains with its side chain free (Monaco *et al.*, 1987). RBP has six cysteine residues (4, 70, 120, 129, 160, 174); the bridge 4–160 connects the N-terminus region to an α-helix, the bridge 70–174 connects a β-strand to the C-terminus region and the bridge 120–129 connects two β-strands (Cowan *et al.*, 1990; Zanotti *et al.*, 1993). BBP has four cysteine residues (9, 43, 119, 175) and forms two S—S bridges; bridge 9–119 connects the N-terminus region to a turn of a β-strand and the S—S bridge 43–175 connects a β-strand to the C-terminus region (Huber, Schneider, Epp *et al.*, 1987; Huber, Schneider, Mayr *et al.*, 1987). Insecticyanin (ICYB) is similar to BBP in having S—S bonds between Cys9 and Cys119 and between Cys43 and Cys175 (Holden *et al.*, 1987).

Table 8
Position of the third putative S—S disulfide bridge of lobster apo A_1 in other lipocalins, with respective Cys Cα separation distances.

Protein	Apo A_1	RBP	ICYB	BBP	BLG
Residue	Cys117	Val116	Ile115	Ile115	Leu104
Residue	Cys150	Gln149	Val149	Vla149	Phe136
Cα—Cα distances (Å)	5.62	10.91	8.20	7.92	6.77

All these similarities in the lipocalin disulfide-bond patterns suggested that apo C_1 and also apo A_1 may have disulfide bridges between Cys12 and Cys121, relating the N-terminus region to a β-strand, and Cys173 and Cys51 between the C-terminus region and the beginning of a β-strand (Keen *et al.*, 1991*b*; Zagalsky, 1994). However, there were no indications about what the two remaining Cys residues, *i.e.* Cys117 and Cys150, might do. To check the possibility of an S—S bridge between Cys117 and Cys150, an overall alignment (Fig. 9) was created between the amino-acid sequences of apocrustacyanin A_1, β-lactoglobulin (PDB code 1bsq; Quin *et al.*, 1999), apolipoprotein D (SWISS-PROT code P51910; Seguin *et al.*, 1995), insecticyanin (SWISS-PROT code Q00630; Li & Riddiford, 1992), bilin-binding protein (PDB code 1bbp; Huber, Schneider, Epp *et al.*, 1987; Huber, Schneider, Mayr *et al.*, 1987) and retinol-binding protein (PDB code 1brq; Zanotti *et al.*, 1993) using the *BLAST* 2.0 server (Altschul *et al.*, 1997) and following previous work (Cowan *et al.*, 1990; Huber *et al.*, 1987; Flower *et al.*, 1993; Zagalsky, 1994).

The results (Table 8) showed that in all the proteins checked, the residues that match in the sequence with the Cys117 and Cys150 residues of apo C_1 are on the same side of the protein and without any interfering polypeptide chain coming between them. However, each protein required a small adjustment for the Cα—Cα distance to be optimal for disulfide-bond formation, *i.e.* for BLG, BBP and insecticyanin the distances between the Cα atoms are respectively 6.77, 7.92 and 8.20 Å. These should be compared with an average distance between the Cα atoms of cysteine residues involved in a disulfide bridge of about 5.5 Å. In RBP the

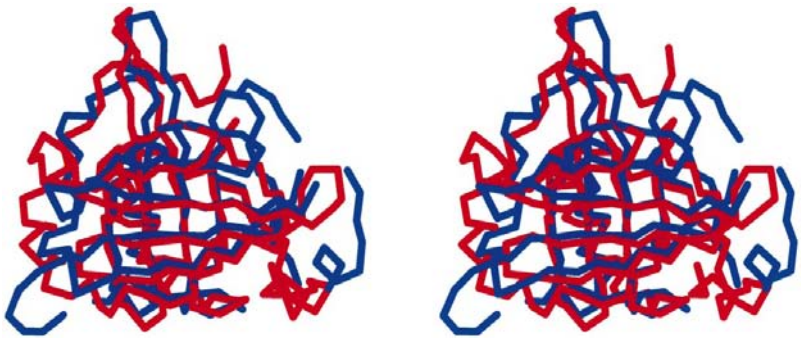

Figure 7
Overlay of the apocrustacyanin A_1 monomer structure (red) and the predicted model (blue) of Keen *et al.* (1991*b*).

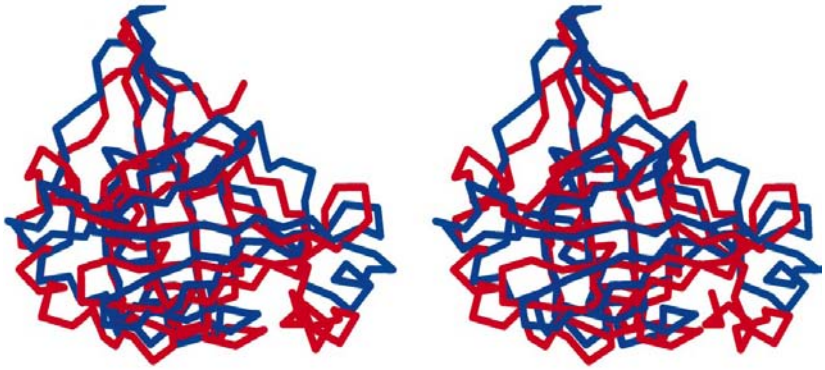

Figure 8
Overlay of the apocrustacyanin A_1 monomer structure (red) with that of an insecticyanin (blue) monomer (Holden *et al.*, 1987).

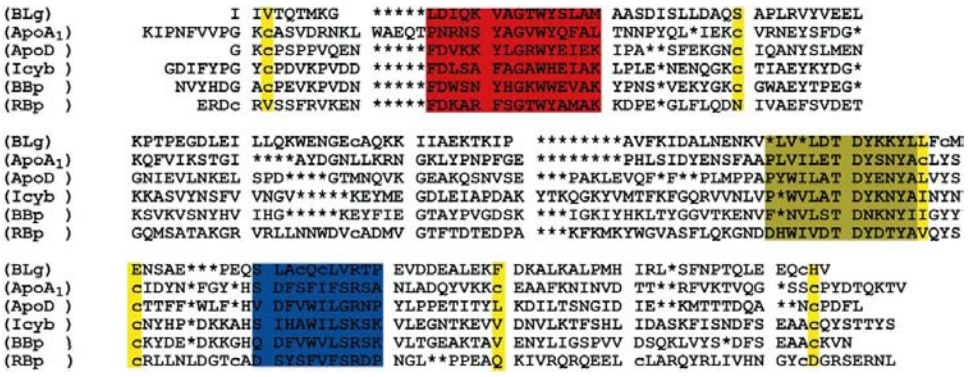

Figure 9
Amino-acid sequence alignment for A_1 and other lipocalins. The consensus regions are highlighted in red for consensus sequence 1, in green for stretch 2 and in blue for stretch 3 (as in Figs. 5 and 6). The cysteines are shown in yellow.

residues considered are 10.91 Å apart, so the required adjustment is much more difficult. We concluded that the existence of the third disulfide bridge had the same probability as that of the other two very likely S—S bridges. The presence of a third disulfide in apo A_1 was also supported chemically by testing apo C_1 for a free SH group using the Ellman assay with DTNB [5,5′-dithio-bis(2-nitrobenzoic acid)] and none was found. Eventually, this evidence was confirmed by the X-ray analysis reported here, with a Cα—Cα distance of 5.62 Å.

4.2.3. Ligand binding. Some of the cavity aromatic residues binding retinol in RBP are conserved or conservatively substituted in the A_1 structure (and in ICYB), *i.e.* Tyr133 (RBP) to His139 (ICYB) to Phe132 (A_1) and Phe45 (RBP) to Tyr55 (ICYB) to Tyr56 (A_1). However, it is important to bear in mind the warning of Newcomer (1993) that structurally conserved residues in homologous proteins do not necessarily have homologous functions. Their observation was based on the finding that the five binding-site residues in RABP that are conserved in RBP are not part of the ligand-binding site in RABP. The crystal structure of a ligand-bound form of A_1 or C_1 or in the crustacyanin β-complex still needs to be determined.

4.2.4. Oligomeric form in nature. The fact that A_1 crystallizes in its dimer form is especially interesting because it may resemble the heterodimer types that form the β-crustacyanins. It is these various β-proteins which assemble to make up the oligomeric α-crustacyanin. Knowledge of this dimer basic building unit might then provide the vital starting point for studying the macroassembly.

5. Conclusions

An innovative structure-solution approach involving softer X-rays, cryotrapped xenon and enhanced sulfur anomalous differences has been successful. Additionally, the phase extension (from 2.3 Å) to 1.35 Å using *DM* yielded a high-quality electron-density map which allowed the automatic chain tracing using *ARP*/*wARP* and fitting of the amino-acid sequence of the highly homologous C_1 protein. A high-multiplicity native data set was also measured at 2 Å wavelength to yield significant sulfur anomalous differences. This very clearly gave the hand of the xenon single isomorphous replacement with optimized anomalous scattering (SIROAS) phases *via* the sulfur peaks.

The lobster apocrustacyanin A_1 structure verifies that it belongs, as proposed for C_1 (Keen *et al.*, 1991*b*) to which it is highly homologous, to the lipocalin family of protein structures. Within the lipocalin family members, the A_1 structure is closest to the retinol-binding protein and insecticyanin structures. The biological function of the protein complexes in the lobster carapace (A_1) and in insects (insecticyanin) is camouflage. Nature has achieved this *via* similar protein three-dimensional structures. The differences in loop structure, cavity geometry and residues, which are all likely to determine ligand-binding specificity, will be reported when the structure of β-crustacyanin is solved.

We thank SRS Daresbury Laboratory and the Joint Biology Programme of the UK Research Councils for synchrotron-radiation beamtime. We are grateful for the award of a research grant from The Leverhulme Trust. We thank the EPSRC and the University of Manchester for a PhD studentship for M. Cianci. NC and JRH thank the EU (Biocrystallogenesis award BI04-CT98-0086) for funds that supported AO. MC thanks the Structural Chemistry Section of the University of Padua for the encouragement to pursue studies in protein X-ray crystallography. The Wellcome Trust and the BBSRC funded the Manchester Structural Chemistry Laboratory computing and graphics suite. We are very grateful to Hazel Holden for supplying the coordinates of the insecticyanin structure and to Elias Eliopoulos, Jeff Keen and John Findlay for supplying the coordinates of the Keen *et al.* (1991) homology model.

References

Akerstrom, B., Flower, D. R. & Salier, J. P. (2000). *Biochim. Biophys. Acta*, **1482**, 1–8.
Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Blanchet, M. A., Bains, G. & Pelosi, P. (1996). *Nature Struct. Biol.* **3**, 934–939.
Bocskei, Z., Groom, C. R., Flower, D. R., Wright, C. E., Phillips, S. E. V., Cavaggioni, A., Findlay, J. B. C. & North, A. C. T. (1992). *Nature (London)*, **360**, 186–188.
Brownlow, S., Cabral, J. H. M., Cooper, R., Flower, D. R., Yewdall, S. J., Polikarpov, I., North, A. T. C. & Sawyer, L. (1997). *Structure*, **4**, 481–495.
Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, K. 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.

Carugo, O. & Bordo, D. (1999). *Acta Cryst.* D**55**, 479–483.
Cavaggioni, A. & Mucignat-Carretta, C. (2000). *Biochim. Biophys. Acta*, **1482**, 218–228.
Chayen, N. E., Cianci, M., Olczak, A., Raftery, J., Rizkallah, P. J. & Helliwell, J. R. (2000). *Acta Cryst.* D**56**, 1064–1066.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowan, S. W., Newcomer, M. E. & Jones, T. A. (1990). *Proteins Struct. Funct. Genet.* **8**, 44–61.
Flower, D. R. (2000). *Biochim. Biophys. Acta*, **1482**, 46–56.
Flower, D. R., North, A. C. T. & Attwood, T. K. (1993). *Protein Sci.* **2**, 753–761.
Flower, D. R., North, A. C. T. & Sansom, C. E. (2000). *Biochim. Biophys. Acta*, **1482**, 9–24.
Gordon, E. J., Leonard, G. A., McSweeney, S. & Zagalsky, P. F. (2001). *Acta Cryst.* D**57**, 1230–1237.
Helliwell, J. R., Greenhough, T. J., Carr, P. D., Rule, S. A., Moore, P. R., Thompson, A. W. & Worgan, J. S. (1982). *J. Phys. E*, **15**, 1363–1372.
Holden, H. M., Rypniewski, W. R., Law, J. H. & Rayment, I. (1987). *EMBO J.* **6**, 1565–1570.
Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 478–480.
Huber, R., Schneider, M., Epp, O., Mayr, I., Messerschmidt, A., Pflugrath, J. & Kayser, H. (1987). *J. Mol. Biol.* **195,** 423–434.
Huber, R., Schneider, M., Mayr, I., Muller, R., Deutzmann, R., Suter, F., Zuber, H., Falk, H. & Kayser, H. (1987). *J. Mol. Biol.* **198**, 499–513.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Keen, J. N., Caceres, I., Eliopoulos, E. E., Zagalsky, P. F. & Findlay, J. B. C. (1991*a*). *Eur. J. Biochem.* **197**, 407–417.
Keen, J. N., Caceres, I., Eliopoulos, E. E., Zagalsky, P. F. & Findlay, J. B. C. (1991*b*). *Eur. J. Biochem.* **202**, 31–40.
Kleywegt, G. J. & Jones, T. A. (1996). *Acta Cryst.* D**52**, 829–832.
Laskowski, R. A., McArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Li, W. & Riddiford, L. M. (1992). *Eur. J. Biochem.* **205**, 491–499.
McKenzie, H. A., Ralston, G. B. & Shaw, D. C. (1968). *Biochemistry*, **11**, 4539–4547.
Monaco, H. L. & Zanotti, G. (1992). *Biopolymers*, **32**, 457–465.
Monaco, H. L., Zanotti, G., Spadon, P., Bolognesi, M., Sawyer, L. & Eliopoulos, E. E. (1987). *J. Mol. Biol.* **197**, 695–706.
Newcomer, M. E. (1993). *Structure*, **1**, 7–18.
Newcomer, M. E., Jones, T. A., Åqvist, J., Sundelin, J., Eriksson, U., Rask, L. & Peterson, P. A. (1984). *EMBO J.* **3**, 1451–1454.
Newcomer, M. E. & Ong, D. E. (2000). *Biochim. Biophys. Acta*, **1482**, 57–64.
North, A. C. T. (1989). *Int. J. Biol. Macromol.* **11**, 56–58.
Ong, D. E., Newcomer, M. E., Lareyre, J. J. & Orgebin-Crist, M. C. (2000). *Biochim. Biophys. Acta*, **1482**, 209–217.
Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Scattering and Anomalous Replacement*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Papiz, M. Z., Sawyer, L., Eliopoulos, E. E., North, A. C. T., Findlay, J. B. C., Sivaprasadarao, R., Jones, T. A., Newcomer, M. E. & Kraulis, P. J. (1986). *Nature (London)*, **324**, 383–385.
Perrakis, A., Morris, R. J. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Pevsner, J., Reed, R. R., Feinstein, P. G. & Snyder, S. H. (1988). *Science*, **241**, 336.
Quarmby, R., Norden, D. A., Zagalsky, P. F., Ceccaldi, H. J. & Dumas, R. (1977). *Comput. Biochem. Physiol. B*, **56**, 55–61.
Quin, B. Y., Bewley, M. C., Creamer, L. K., Baker, E. N. & Jameson, G. B. (1999). *Protein Sci.* **8**, 75.
Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.
Sawyer, L. & Kontopidis, G. (2000). *Biochim. Biophys. Acta*, **1482**, 136–148.
Seguin, D., Desforges, M. & Rossart, E. (1995). *Mol. Brain Res.* **30**, 242–250.
Sheldrick, G. M. (1997). In *Proceedings of the CCP4 Study Weekend. Recent Advances in Phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey. Warrington: Daresbury Laboratory.
Sheldrick, G. M. (1998). *SHELX: Applications to Macromolecules. Direct Methods for Solving Macromolecular Structures.* Dordrecht: Kluwer Academic Publishers.
Tegoni, M., Pelosi, P., Vincent, F., Spinelli, S., Campanacci, V., Grolli, S., Ramoni, R. & Cambillau, C. (2000). *Biochim. Biophys. Acta*, **1482**, 229–240.
Tegoni, M., Ramoni, R., Bignetti, E., Spinelli, S. & Cambillau, C. (1996). *Nature Struct. Biol.* **10**, 863–867.
Wald, G., Nathanson, N., Jenksand, W. P. & Tarr, E. (1948). *Biol. Bull. Woods Hole*, **95**, 249–250.
Weeks, C. M. & Miller, R. (1999). *J. Appl. Cryst.* **32**, 120–124.
Weiss, M. S., Sicker, T., Djinovic Carugo, K. & Hilgenfeld, R. (2001). *Acta Cryst.* D**57**, 689–695.
Wright, C. E., Rafferty, J. B., Flower, D. R., Groom, C. R., Findlay, J. B. C., North, A. C. T., Phillips, S. E. V. & Zagalsky, P. F. (1992). *J. Mol. Biol.* **224**, 283–284.
Zagalsky, P. F. (1983). *Comput. Biochem. Physiol. B*, **73**, 997–1000.
Zagalsky, P. F. (1994). *Pure Appl. Chem.* **66**, 973–980.
Zanotti, G., Ottonello, S., Berni, R. & Monaco, H. L. (1993). *J. Mol. Biol.* **230**, 613–624.