Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The C_1 subunit of α-crustacyanin: the *de novo* phasing of the crystal structure of a 40 kDa homodimeric protein using the anomalous scattering from S atoms combined with direct methods

Elspeth J. Gordon,[a]* Gordon A. Leonard,[a] Seán McSweeney[a] and Peter F. Zagalsky[b]

[a]Macromolecular Crystallography, European Synchrotron Radiation Facility, BP 220, F-38043 Grenoble CEDEX, France, and [b]Department of Biochemistry, Royal Holloway College, University of London, Egham, Surrey TW20 0EX, England

Correspondence e-mail: gordon@esrf.fr

Received 1 March 2001
Accepted 6 June 2001

PDB Reference: C_1 subunit of α-crustacyanin, 1i4u.

The previously unknown crystal structure of the C_1 subunit of the carotenoid-binding protein α-crustacyanin has been determined using the anomalous scattering available at 1.77 Å wavelength to determine the partial structure of the S atoms intrinsic to the native protein. The resulting 'heavy-atom' phases, in conjunction with near-atomic resolution ($d_{\min}$ = 1.15 Å) data, were then used to initiate successful structure determination using a direct-methods approach. This is, to the authors' knowledge, the first time such a small anomalous signal (∼1%) has been used to aid the determination of a macromolecular structure. As well as the structure itself, the methods used during data collection and those used in the elucidation of the sulfur 'heavy-atom' partial structure are described here. As predicted, the C_1 subunit adopts a tertiary structure typical of the lipocalin superfamily: an eight-stranded antiparallel β-barrel with a repeated +1 topology. The β-barrel has a calyx shape with the two molecules in the asymmetric unit interacting in such a way that the open ends of each calyx face each other, although they do not form a single elongated pocket. A comparison of this structure with those of other members of the lipocalin superfamily has allowed speculation as to the nature of carotenoid binding by the protein.

1. Introduction

Except for the special case of metalloenzymes, the most common methods for the *de novo* solution of the phase problem in macromolecular crystallography (isomorphous replacement, MAD) rely on either the preparation of heavy-atom derivatives or the incorporation into native sequences of modified amino acids or nucleotides. When neither of these options is available, macromolecular structure determination becomes notoriously difficult, as the atomic resolution data currently required for *ab initio* direct-methods solution of the phase problem are rarely measurable for macromolecular crystals. Even when such data are available, the special nature of macromolecular crystals means that the success of direct methods is not always guaranteed.

However, the vast majority of proteins and all nucleic acids contain atoms (sulfur and phosphorus, respectively) which have considerable anomalous scattering properties. Could the presence of these properties be used to derive phases without the need for incorporation of heavy atoms or the measurement of atomic resolution data? This question is not new and it has been shown that it is possible to measure data from which phases might be determined from crystals of macro-

molecules at wavelengths close to the sulfur *K* absorption edge (Lehmann *et al.*, 1993; Stuhrmann *et al.*, 1997). However, this is unlikely to become commonplace: the severe absorption and radiation-damage problems encountered at such long wavelengths (~5 Å) are in most instances likely to prove insurmountable. A much more appealing and possibly general method would be the exploitation of the anomalous scattering properties of sulfur and phosphorus at wavelengths more commonly used for macromolecular crystallography data collection.

This idea has also been addressed and it has been shown that one can solve *ab initio* the crystal structures of proteins such as crambin (Hendrickson & Teeter, 1981) or lysozyme (Dauter *et al.*, 1999) by taking advantage of the anomalous scattering of sulfur at 1.54 Å wavelength. Given that f'' for sulfur at this wavelength is only 0.55 e^-, the anomalous signal available is rather small and makes for difficulties in (i) measuring anomalous differences accurately enough to determine the S-atom substructure and (ii) using the substructure as a springboard to full structure solution. However, inspired by these successes, we have solved the previously unknown crystal structure of the C_1 subunit of α-crustacyanin by elucidating the positions of the S atoms on the unit cell using anomalous differences measured at 1.77 Å wavelength and then using these to initiate the successful direct-methods phasing of the entire crystal structure. We report here our results and in particular our experimental method.

The binding of astaxanthin (3,3′-dihydroxy-β,β-carotene-4,4′-dione) by the protein α-crustacyanin is responsible for the blue colouration of the carapace of the lobster *Homarus gammarus*. α-Crustacyanin itself has a molecular weight in the region of 320 kDa, is hexadecameric and consists of eight copies of so-called β-crustacyanin dimers, the formation of which is astaxanthin-dependent (Cheesman *et al.*, 1966; Kuhn & Kuhn, 1967; Buchwald & Jencks, 1968; Quarmby *et al.*, 1977) and which are made up of combinations of type I and II apocrustacyanin monomers. The type I apocrustacyanins are the C_1, C_2 and A_1 proteins, while type II comprises the subunits A_2 and A_3. The sequences of the major member of the two families (C_1, A_2) are known (Keen *et al.*, 1991*a*,*b*) with alignment (Thompson *et al.*, 1994) of the two sequences showing them to be highly (~56%) homologous, although the actual sequence identity is only 35%. It is therefore highly likely that the type I and type II monomers adopt the same three-dimensional structure. Crystals of α-crustacyanin were first grown around 20 years ago, but the size of the problem was at the time deemed unfeasible. Therefore, ten years ago the emphasis was placed on targeting the individual subunits (Wright *et al.*, 1992), crystals of which it was hoped would be more suitable for a crystallographic structure determination. Until now, this has not proved to be the case.

The C_1 subunit consists of 181 amino-acid residues of which six are cysteine and none are methionine. It has been classified as being a member of the lipocalin superfamily of proteins that are responsible for binding small hydrophobic ligands (North, 1989). The protein crystallizes readily with a dimer in the asymmetric unit and crystals can diffract to very high resolution (Gordon & McSweeney, unpublished data). Unfortunately, despite much effort and excellent crystal quality, the structure solution of the C_1 subunit (and the homologous A_1 protein, the structure of which is reported by Cianci *et al.*, 2001) has proved recalcitrant. Although the Protein Data Bank (Berman *et al.*, 2000) contains many crystal structures with the lipocalin fold, exhaustive molecular-replacement searches were unsuccessful. Isomorphous replacement experiments also failed and were probably hindered by the use of high concentrations of ammonium sulfate and a high pH in the crystallization medium. Thus, despite much effort suitable heavy-atom derivatives for the C_1 crystals have never been found. Coupled with these difficulties, the protein is purified from lobster shells and the possibility of engineering a selenomethionine residue into the sequence did not exist. It thus seemed that structure solution using conventional techniques posed an insurmountable problem, hence our attempt to solve the structure using the anomalous signal from the S atoms.

2. Data collection

Orthorhombic crystals (space group $P2_12_12_1$) of the C_1 subunit containing a dimer (362 residues) in the asymmetric unit were grown as previously described (Wright *et al.*, 1992). A crystal was flash-frozen in a stream of gaseous nitrogen at 110 K using 30%(*v*/*v*) 2-methyl-2,4-pentanediol (MPD) as a cryoprotectant and mounted on beamline BM14 of the ESRF. The expected Bijvoet ratio resulting from the presence of sulfurs for this system at 1.54 Å wavelength ($f'' \simeq 0.56$ e^-) is extremely small at 0.8%. Thus, to increase our chances of success the wavelength was set to 1.77 Å, where an f'' of ~0.72 e^- for sulfur would yield an increased Bijvoet ratio of 1.0%.

Two data sets were collected from the same crystal. The first of these was collected to 2.5 Å resolution at $\lambda = 1.77$ Å; 720° of data were collected in the hope that a highly redundant data set would result in extremely accurate anomalous differences being measured. No attempt was made to align the crystal and measure Bijvoet-related reflections on the same or even on neighbouring images. These data were integrated with *DENZO* and scaled with *SCALEPACK* (Otwinowski & Minor, 1997) with, at the latter stage, the *hkl* and $\overline{hkl}$ reflections in an anomalous pair being treated as separate reflections during both scaling and merging. Although the resulting data were of good quality, for reasons that are unclear a large number of data oscillation images around $\varphi = 0$, 360 and 720° had individual merging *R* factors close to double the overall value. These images were excluded from the final data set which consisted of 560° of data. Details of this are given in Table 1.

As it was already known that crystals of the C_1 subunit diffract to very high resolution (see above), the wavelength of BM14 was then set to 0.72 Å to allow a data set with $d_{\min} = 1.1$ Å to be collected. The rationale behind this data-collection strategy was to minimize any error in non-isomorphism between data sets in case any phases derived

Table 1
Data-collection statistics.

Values in parentheses refer to the highest resolution shell for the data set in question (2.54–2.50 and 1.17–1.15 Å, respectively).

Wavelength (Å)	1.77	0.72
Unit-cell parameters (Å)	a = 41.22, b = 79.73, c = 109.59	a = 41.16, b = 79.62, c = 109.45
Space group	$P2_12_12_1$	$P2_12_12_1$
Resolution limits (Å)	40.0–2.5	40.0–1.15
Unique reflections	24155†	127510
Reflections measured	266685	464503
Completeness (%)	99.9 (98.9)	99.6 (99.9)
$\langle I\rangle/\langle\sigma(I)\rangle$	46.7 (14.5)	25.8 (2.7)
$R_{\rm sym}$†	4.4 (8.1)†	5.0 (27.8)

† I_{hkl} and $I_{\overline{hkl}}$ scaled and merged separately.

from the 2.5 Å data set needed to be 'transferred' to the high-resolution data to allow complete solution of the structure. These high-resolution data were collected in two passes to ensure proper data completeness and were integrated with *DENZO* and scaled with *SCALEPACK* (see Table 1 for full details). For both data sets, mean structure-factor amplitudes and anomalous differences were derived using the program *TRUNCATE* (Collaborative Computational Project, 1994; French & Wilson, 1978).

3. Determination of S-atom partial structure

To find the S-atom partial structure, the data collected at λ = 1.77 Å with I_{hkl} and $I_{\overline{hkl}}$ scaled and merged separately were used as input into the direct-methods program *SnB* (Howell *et al.*, 2000). The difference E values (diff$E$) (to a maximum resolution of 2.6 Å) required for this were derived based on the observed anomalous differences using the program *DREAR* (Blessing, 1987). During this procedure, a search was carried out for a total of 12 sites (*i.e.* all S atoms in the protein dimer). 400 reflections (phases) were used to generate 6000 triplet-phase invariants with the maximum diff$E$ set to be 5.0 and the minimum being 0.75. In order to generate a sufficient number of reflections in the starting set, the minimum diff$E$/σ(diff$E$) acceptable was reduced to 2.0 from the *SnB* default value of 3.0. 1000 trials were carried out with starting phases being derived from random atoms. 'Twice baking' was not carried out and all other parameters and choices were left as the defaults provided by the program.

At the end of the procedure, the distribution of the minimum function $R_{\rm min}$ was examined for the bimodal nature that is characteristic of successful partial structure determination (Miller *et al.*, 1994; Howell *et al.*, 2000). As can be seen from Fig. 1, this is clearly in evidence and the top 12 sites were thus used as input for refinement and phase-probability distribution calculation using the anomalous differences in the 1.77 Å wavelength data with the program *SHARP* (de La Fortelle & Bricogne, 1997). Of the 12 sites input to the initial round of refinement, the top six refined to yield occupancies greater than 1 (in the range 1.10–1.40). Inspection of the log-likelihood-based residual maps from *SHARP* revealed that each of these top six sites had an adjacent peak of positive (>4σ) density. The distance between each of the top six sites and its neighbouring residual map density was on average close to the 1.8 Å S—S distance in a disulfide bond. Hence, it appears that using data with $d_{\rm min}$ = 2.6 Å, the *SnB* search had shown up as its top six sites peaks corresponding to a single member of each of six disulfide moieties present in the asymmetric unit and that the second sulfur of each of the disulfides could then be found from the *SHARP* residual maps.

The second set of six positions from the original 12 all refined in *SHARP* to have occupancies around 0.5 and did not have adjacent density on the residual maps. It was thought that these probably corresponded to the positions of bound sulfate ions. A second round of refinement and phase-probability distribution calculations in *SHARP* was then carried out with a model for the S-atom partial structure consisting of 12 'disulfide' sulfurs and six 'free' S atoms. This resulted in clean residual maps indicating a correct S-atom partial structure. The overall FOM for acentric reflections was 0.365 for all data in the resolution range 40.0–2.5 Å.

4. Phasing, model building and refinement

The electron density ($F_{\rm best}$, $\alpha_{\rm best}$, FOM) map calculated by *SHARP* was not interpretable and was clearly very biased towards the partial structure model. Moreover, use of the powerful solvent-flattening programs *DM* (Cowtan, 1994) and *SOLOMON* (Abrahams & Leslie, 1996) did not significantly improve this. Attempts to use different approaches to partial structure refinement and phase-probability distribution calculation, such as those available in the programs *MLPHARE* (Otwinowski, 1991) and *OASIS* (Hao *et al.*, 2000), were no more successful. It therefore seemed that with the programs currently available, the combination of a very low anomalous signal (1%) and relatively low resolution data ($d_{\rm min}$ = 2.5 Å) the determination of the sulfur substructure would not allow full phasing of a crystal structure.

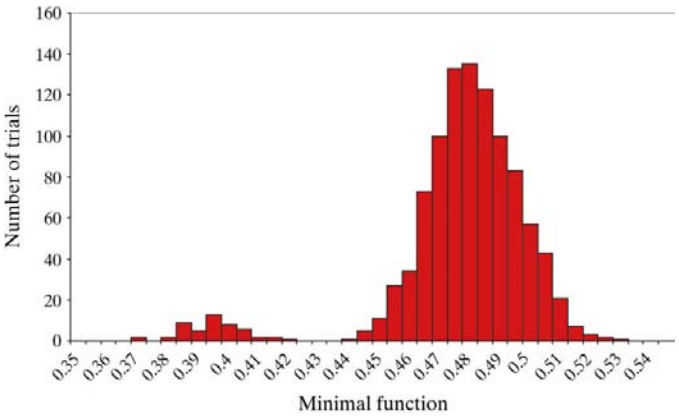

Figure 1
The histogram of the minimal function $R_{\rm min}$ resulting from the *SnB* search for the S-atom partial structure. The bimodal distribution is characteristic of a successful result to the search.

Table 2
Quality of present model.

Values in parentheses are those for data where $F > 4\sigma(F)$.

Crystallographic R value (%)	14.96 (13.90)
No. of reflections	121772 (91740)
Free R value (%)	18.77 (17.54)
No. of reflections	6420 (4854)
Resolution range (Å)	40.0–1.15
No. non-H atoms	
All atoms	3382
Protein	2911
Water	429
MPD	32
SO_4^{2-}	10
Average isotropic B values (Å^2)	
All atoms	19.4
Protein	17.9
Water	28.6
MPD	23.9
SO_4^{2-}	55.6
R.m.s. deviation from standard geometry	
Bond lengths (Å)	0.013
Bond angles (°)	2.4
R.m.s. deviations in B values (Å^2)	
Protein main-chain atoms	1.3
Protein side-chain atoms	3.6

However, the availability of high-resolution data collected from the same crystal as the 1.77 Å wavelength data did allow us to determine the crystal structure. E values for the high-resolution data were calculated for all data to a resolution limit of 1.15 Å [the point where $\langle I/\sigma(I)\rangle$ for the resolution shell fell below 2.5] using the program *DREAR* (Blessing, 1987). Starting phases calculated from the positions of the 12 S atoms in the six disulfide bridges only were then used as a basis for phase refinement using the program *ACORN* (Foadi *et al.*, 2000). Provided that the resolution of the data was not truncated in any way, this procedure worked immediately providing electron-density maps of excellent quality (Fig. 2). The fact that even with some prior phase information direct-methods procedures still required data to a resolution limit of 1.15 Å is significant, suggesting that the 1.2 Å resolution limit often quoted for successful direct-methods structure determination (see, for example, Dauter *et al.*, 1999) is a 'hard' limit that may be difficult to overcome.

The major part of the atomic model for the protein dimer was built automatically using the program *wARP* (Perrakis *et al.*, 1999). Further refinement ($d_{\min}$ = 1.15 Å) and manual rebuilding of the model were carried out with the *SHELX* suite of programs (Sheldrick, 1998) and *QUANTA* (Molecular Simulations Inc., San Diego, USA), respectively. The final

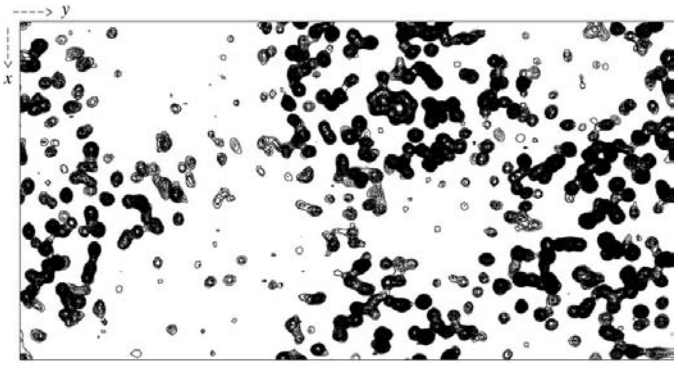

Figure 2
A section of the 1.15 Å experimental electron-density map (F_{obs}, α_{acorn}, FOM_{acorn}) resulting from an *ACORN* procedure using initial phases calculated from the positions of the 12 S atoms involved in the six S—S bridges in the asymmetric unit. Note that the density for Trp35 is clearly visible.

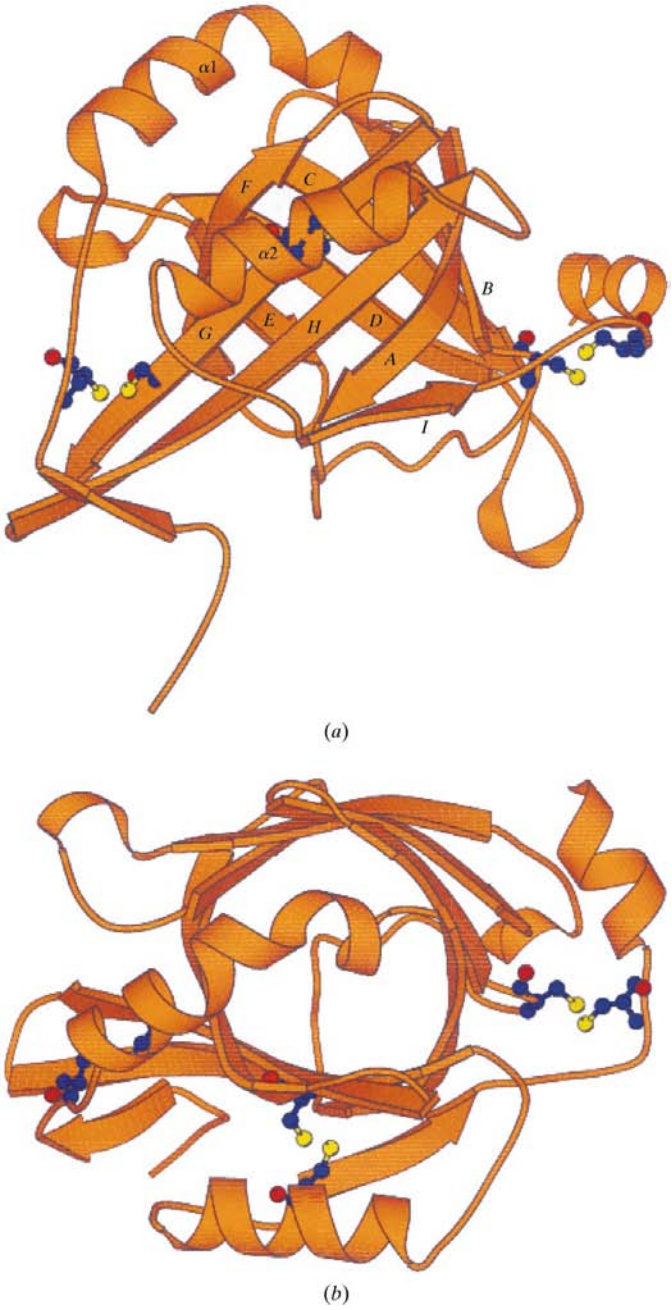

Figure 3
Ribbon diagrams of the three-dimensional structure of the C_1 subunit of α-crustacyanin. (*a*) A 'classical' view of the lipocalin fold shows the orthogonally arranged β-sheets of the C_1 subunit. The cysteine residues involved in disulfide bridges are shown in ball-and-stick representation. (*b*) A view perpendicular to (*a*) giving a clearer view of disulfide positions. This figure and Figs. 4 and 6 were produced using *BOBSCRIPT*/*MOLSCRIPT* (Kraulis, 1991; Esnouf, 1999).

model contains 361 amino-acid residues (residues 1–181 for molecule *A*, 2–181 for molecule *B*), four bound MPD moieties, two sulfate ions and 429 water molecules. Despite the high resolution of the data available, poor electron density in certain regions of the map meant that residue Asp1 in molecule *B* was not included in the model and that Lys61 in the same molecule was truncated to alanine. All C, N, O and S atoms in the model were refined with anisotropic temperature factors. H atoms were included in the refinement procedure by generating their positions based on geometrical criteria and then allowing them to 'ride' on the heavy atoms to which they are attached. The current crystallographic *R* factor is 15.0% for 121 722 reflections with $F > 0$ in the resolution range 40.0–1.15 Å, while the $R_{\rm free}$ (Brünger, 1992) is 18.8% for 6420 reflections in the same resolution range. Full details of the results of the refinement procedure can be found in Table 2. The quality of the final model (see Table 2) was assessed using *PROCHECK* (Laskowski *et al.*, 1993) and *WHATIF* (Vriend & Sander, 1993). The conformations of all the amino acids lie in the allowed region of the Ramachandran plot (Ramakrishnan & Ramachandran, 1965), except for Tyr112 in both monomers. This residue is found in the Thr-Asp-Tyr γ-turn motif and its non-standard conformation is a feature of many lipocalin structures (Huber *et al.*, 1987; Cowan *et al.*, 1990; North, 1989).

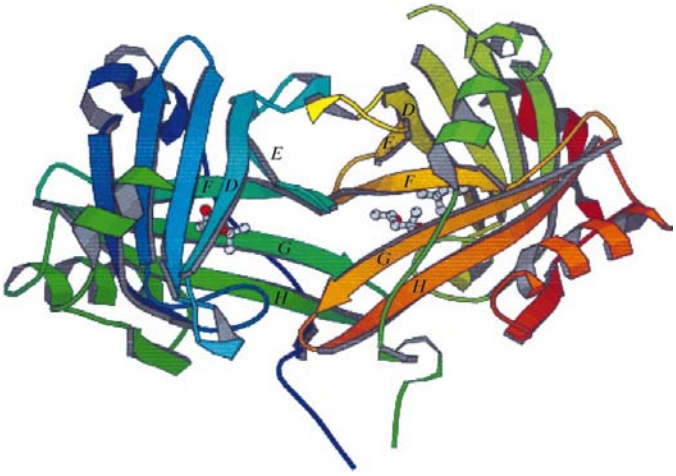

Figure 4
A ribbon diagram of the C_1/C_1 homodimer found in the asymmetric unit. Note that the dimerization is of a 'face-to-face' nature and that monomer–monomer interactions occur *via* the N-terminal and loops *GH*, *EF* and *DE*. Two MPD molecules shown in ball-and-stick representation (carbon, grey; oxygen, red) can be seen in the two binding pockets, which are almost facing one another.

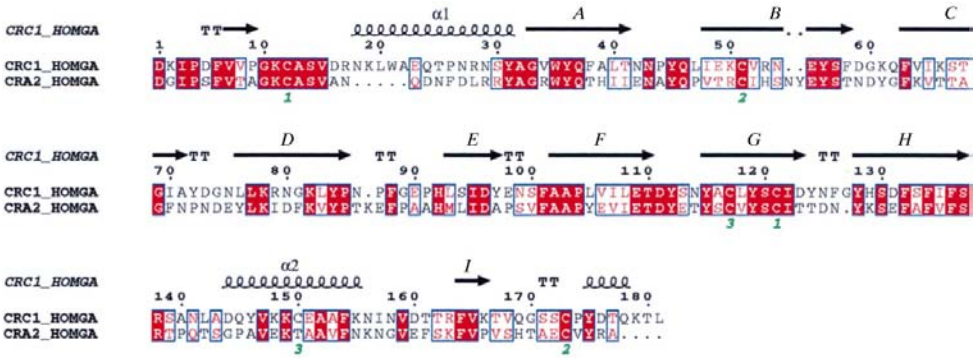

Figure 5
The amino-acid sequence and secondary-structure assignment of the C_1 crustacyanin subunit aligned with the A_2 subunit sequence using *ESPrit* (Gouet *et al.*, 1999). Areas of sequence similarity are boxed with white background and areas of sequence identity boxed with red background. As can be seen, the two sequences are highly homologous, suggesting they have the same overall structure, hence our hypothesis (see main text) that the C_1/C_1 homodimer described here is a good model for the C_1/A_2 heterodimer found in α-crustacyanin.

5. Structure description

As expected, the structure of the C_1 subunit (Fig. 3) has a characteristic lipocalin fold: an eight-stranded antiparallel β-barrel with a repeated +1 topology (North, 1991). The β-barrel, formed by two orthogonally arranged β-sheets, forms a calyx or cup shape. The strands of the barrel are linked by a +1 connection giving a closed β-sheet. There is a conserved 3_{10} helix at the N-terminal end and a longer C-terminal helix which folds back along the barrel and packs against strands *G* and *H*. In the C_1 structure, there are also extra secondary-structural elements that are not characteristic or common to all members of the lipocalin superfamily. At the N-terminus – before the 3_{10} helix – there is a short β-strand (residues 7–10), an extended loop and then an α-helical extension to the 3_{10} helix. At the C-terminus, strand *I* is followed by an additional two-turn helix near the very end of the polypeptide chain (residues 145–156). There are three disulfide bonds per monomer: Cys12–Cys121, Cys51–Cys173 and Cys117–Cys150, which link the N-terminus to strand *G*, strand *B* to the C-terminus and strand *G* to the C-terminal end of helix α2, respectively.

By analogy with the crystal structures of other members of the lipocalin family, the binding cavity of the C_1 subunit is composed predominantly of aromatic and polar residues: nine aromatic residues (Trp35, Tyr45, Tyr56, Tyr97, Phe101, Phe132, Phe134, Phe136, Phe1101), nine aliphatic residues (Leu47, Ile48, Val52, Leu83, Ile95, Ala102, Ala103, Leu105, Leu118), three hydroxyl groups (Ser67, Ser100, Ser120), four amides (Gln7, Asn3, Gln46, Asn54) and two charged residues (Arg79 and Asp96). The cavities in the crystallographic dimer are closed to the solvent and have a volume of 148 Å^3 as calculated using *VOIDOO* (Kleywegt & Jones, 1994) using a 1.4 Å probe. There are two MPD molecules bound per pocket, with the interactions of these moieties and the protein being mediated *via* water molecules. The bottom of the binding

pocket is characterized by the presence of Trp35, a residue that is strictly conserved in all members of the lipocalin superfamily.

The C_1 crystals contain a homodimer (Fig. 4) in the asymmetric unit. The two molecules lie so that the top of the cavities are packed against one another, each facing in the opposite direction. Interactions are through the N-termini and loops *GH*, *EF* and *DE*.

A comparison of the C_1 structure with a database of known protein structures carried out using the DALI server (Holm & Sander, 1995) showed that despite sequence identities of only 23 and 18%, respectively, the structures with the greatest similarity in tertiary structure were bilin binding protein (BBP) (Huber *et al.*, 1987; r.m.s. deviation of 2.6 Å for 158 superimposed C^{α} positions) and pig plasma retinal binding protein (RBP) (Zanotti *et al.*, 1993; r.m.s. deviation of 2.5 Å for 156 C^{α} positions). The magnitudes of these r.m.s. deviations help to explain the failure of molecular-replacement techniques in solving the crystal structure described here. The β-barrel core of the protein appears to be the most strongly conserved feature, with the base of the binding pockets showing a remarkable similarity in side chains and rotamer positions despite differences in the ligand specificities.

Assuming that the structure of the C_1 subunit is very similar to that of A_2 (see Fig. 5 for sequence similarities), it may be that the dimer we see in the crystal structure is a good model for the structure of the heterodimeric β-crustacyanin. In common with the model of the C_1/A_2 heterodimer (Keen *et al.*, 1991*a*,*b*), the dimerization in the crystal is of a near face-to-face nature. However, the residues (N-termini and loops *GH*, *EF* and *DE* of each monomer) and secondary-structural elements found at the dimer interface are markedly different from those predicted in the model dimer, which was based on the dimer found in the crystal structure of BBP (Huber *et al.*, 1987; Keen *et al.*, 1991*b*).

Examination of a van der Waals surface of the C_1 dimer gives little clue as to where the two astaxanthin molecules bound to β-crustacyanin might bind. However, superposition of a C_1 monomer with the structures of BBP and RBP gives two possibilities as to how this might be achieved. The first of these is directly analogous to the binding of retinol in the RBP structure: in each monomer one end of the astaxanthin would bind to the protein with the other end protruding away from the subunit (Fig. 6*a*). This is similar to the binding mode proposed in the C_1/A_2 model. However, differences in dimer formation between the C_1 dimer crystal structure and the previously proposed model mean that in our crystal one end of each astaxanthin would be exposed to solvent, an unlikely possibility given the hydrophobic nature of astaxanthin.

A second possible mode of astaxanthin binding is shown in Fig. 6(*b*) and is analogous to bilin binding in BBP: this results in the carotenoid spanning both molecules in the dimer, one end being buried in the pocket amd the other on the surface of the neighbouring molecule. Here, the residues involved in the binding of *both* ends of the astaxanthin are overwhelmingly highly conserved between the C_1 and A_2 subunit sequences (data not shown), which would allow for both $C_1 \rightarrow A_2$ and $A_2 \rightarrow C_1$ fixing of the substrate. This 'dimer-spanning' binding mode sits well with observations of Keen and colleagues (Keen *et al.*, 1991*b*) that astaxanthin is 'much more firmly fixed' in the β-crustacyanin heterodimer than in either the C_1 or A_2 subunits alone. However, in this mode of binding, the long axes of the two astaxanthin moieties are parallel, contrary to evidence cited in Keen *et al.* (1991*b*) that they should be almost orthogonal to each other.

It should be emphasized that the above speculation as to the nature of astaxanthin binding is precisely that and we await with interest the structure determination of astaxanthin-bound β-crustacyanin to discover exactly how the chromophore is bound.

6. Discussion

We have reported here that a very small 'anomalous signal' of around 1% has been measured sufficiently accurately at 2.6 Å resolution to allow the elucidation of a heavy-atom partial structure. This was in turn used to kick-start the solution of the crystal structure of a 40 kDa homodimeric protein at near-atomic resolution. SAD phase-probability distributions derived from the partial structure and the measured anomalous differences were not of sufficient quality to allow solution of the crystal structure even in conjunction with powerful density-modification procedures. The anomalous differences, however, are much smaller and available to much lower resolution than for the cases of psoriasin (2.0 Å; Brodersen *et al.*, 2000), lysozyme (1.5 Å; Dauter *et al.*, 1999) and crambin (1.5 Å; Hendrickson & Teeter, 1981) the structures of which have been solved directly from observed anomalous differences. Tests of the limit of SAD methods carried out with error-free data indicate that it should be possible to solve a structure using data to 3 Å resolution with an anomalous signal (Bijvoet ratio) in the region of 0.5% (Wang, 1985). However, it may be that we have inadvertently found the current practical limit for the 'general method' of solution of crystal structures using anomalous differences resulting from the presence of atoms that are intrinsic to native macromolecules. Our experiments also show that the requirements for successful determination of the S-atom partial structure need not be as stringent as those used by Dauter *et al.* (1999), who suggest that data to at least 2.2 Å resolution are required for this. We have used data to 2.6 Å resolution and using a combination of direct methods and heavy-atom refinement in *SHARP* have shown that one can 'resolve' both the S atoms in a disulfide bridge. Other tests we have carried out (Micossi & Leonard, unpublished work) have shown that 2.5 Å data are also sufficient to find the positions of individual cysteine or methionine S atoms using this technique.

We have also investigated the effect of multiplicity on the success of partial structure determination using *SnB* by truncating the number of images included in the data set used for partial structure determination. In all cases, using the approach outlined above, an average multiplicity of greater than 11.0 produced the correct solution. This fact, combined

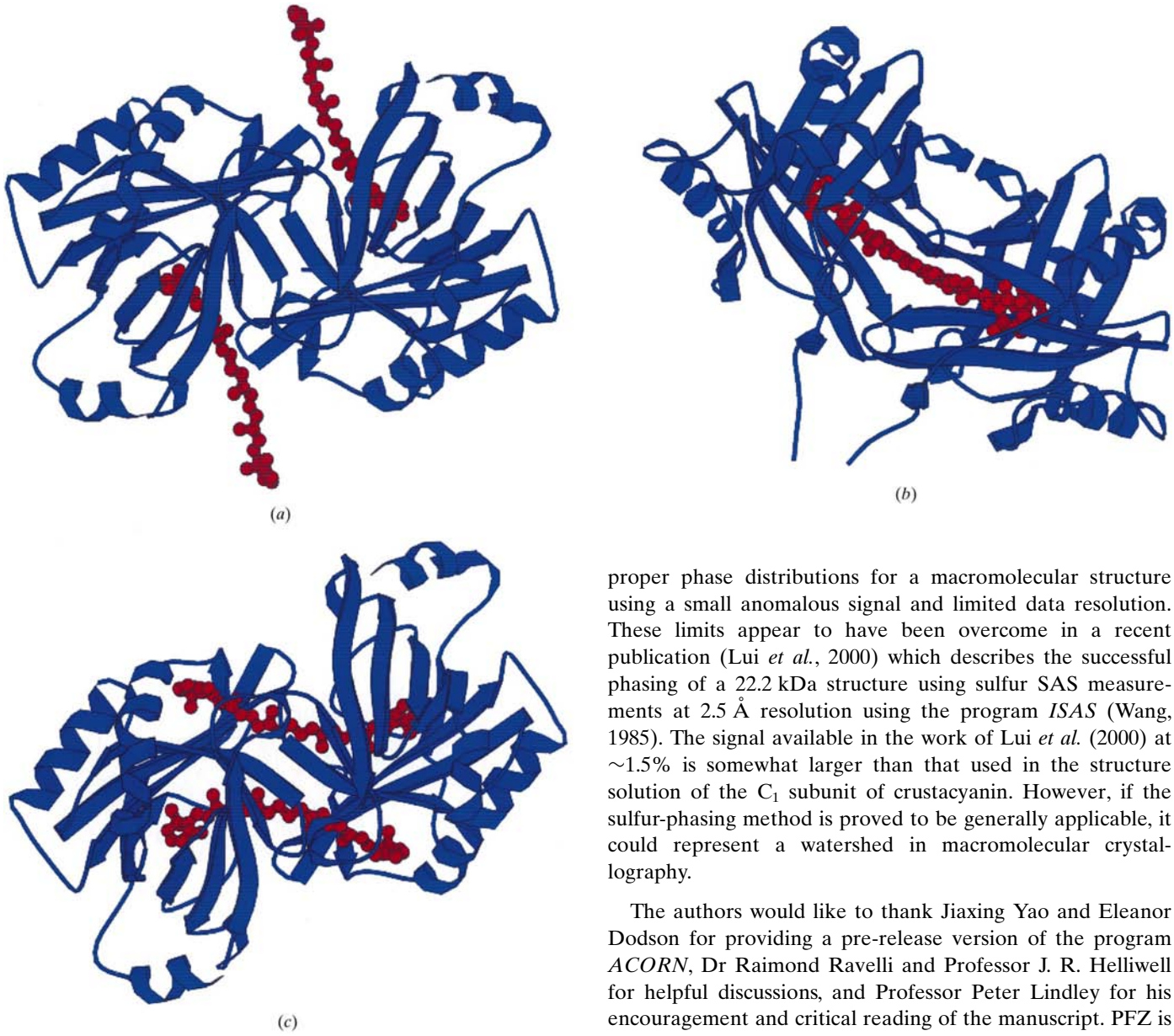

Figure 6
Possible binding modes for the astaxanthin (red ball-and-stick representation) to α-crustacyanin. (*a*) Binding analogous to that found for retinol in the structure of RBP (Zanotti *et al.*, 1993). Here, the astaxanthin protrudes from the binding cavity into the solvent. (*b*) Two perpendicular views of a dimer-spanning binding mode based on an analysis of bilin binding in the structure of BBP (Huber *et al.*, 1987): one end of the astaxanthin is buried deep in the ligand-binding pocket and the other end lies in a surface cleft of the other molecule in the dimer.

with the moderate resolution at which we were successful in resolving the heavy-atom partial structure, suggests that for moderately diffracting crystals one may be able to routinely solve S-atom partial structures without the need for an excessively long data-collection time. In the future, this may thus allow the routine use of S atoms as markers to facilitate protein structure building, if not for use in the full phasing of macromolecular crystal structures.

We have suggested above that the limits of this technique appear to be using this substructure information to calculate proper phase distributions for a macromolecular structure using a small anomalous signal and limited data resolution. These limits appear to have been overcome in a recent publication (Lui *et al.*, 2000) which describes the successful phasing of a 22.2 kDa structure using sulfur SAS measurements at 2.5 Å resolution using the program *ISAS* (Wang, 1985). The signal available in the work of Lui *et al.* (2000) at ∼1.5% is somewhat larger than that used in the structure solution of the C_1 subunit of crustacyanin. However, if the sulfur-phasing method is proved to be generally applicable, it could represent a watershed in macromolecular crystallography.

The authors would like to thank Jiaxing Yao and Eleanor Dodson for providing a pre-release version of the program *ACORN*, Dr Raimond Ravelli and Professor J. R. Helliwell for helpful discussions, and Professor Peter Lindley for his encouragement and critical reading of the manuscript. PFZ is grateful for support from The Leverhulme Trust.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I. & Bourne, P. (2000). *Nucleic Acids Res.* **28**, 235–242.
Blessing, R. H. (1987). *Crystallogr. Rev.* **1**, 3–58.
Brodersen, D., de La Fortelle, E., Vonrhein, C., Bricogne, G., Nyborg, J. & Kjeldgaard, M. (2000). *Acta Cryst.* D**56**, 431–441.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Buchwald, M. & Jencks, W. P. (1968). *Biochemistry*, **7**, 844–859.
Cheesman, D. F., Zagalsky, P. F. & Ceccaldi, H. J. (1966). *Proc. R. Soc. Lond. B*, **164**, 130–151.
Cianci, M., Rizkallah, P. J., Olczak, A., Raftery, J., Chayen, N. E., Zagalsky, P. F & Helliwell, J. R. (2001). *Acta Cryst.* D**57**, 1219–1229.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowan, S. W., Newcomer, M. E. & Jones, T. A. (1990). *Proteins*, **8**, 44–61.

Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
Esnouf, R. M. (1999). *Acta Cryst.* D**55**, 938–940.
Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Jia-xing, Y. & Chao-de, Z. (2000). *Acta Cryst.* D**56**, 1137–1147.
French, S. & Wilson, K. (1978). *Acta Cryst.* A**34**, 517–525.
Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Hao, Q., Gu, Y. X., Zheng, C. D. & Fan, H. F. (2000). *J. Appl. Cryst.* **33**, 980–981.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 478–480.
Howell, P. L., Blessing, R. H., Smith, G. D. & Weeks, C. M. (2000). *Acta Cryst.* D**56**, 604–617.
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.
Keen, J., Caceres, I., Eliopoulos, E. E., Zagalsky, P. F. & Findlay, J. B. (1991*a*). *Eur. J. Biochem.* **202**, 31–40.
Keen, J., Caceres, I., Eliopoulos, E. E., Zagalsky, P. F. & Findlay, J. B. (1991*b*). *Eur. J. Biochem.* **197**, 407–417.
Kleywegt, G. J. & Jones, T. A. (1994). *Acta Cryst.* D**50**, 178–185.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Kuhn, R. & Kuhn, H. (1967). *Eur. J. Biochem.* **2**, 349–360.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lehmann, M. S., Muller, H.-H. & Stuhrmann, H. B. (1993). *Acta Cryst.* D**49**, 308–310.
Lui, Z.-J., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J. & Wang, B. C. (2000). *Protein Sci.* **9**, 2085–2093.
Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
North, A. C. T. (1989). *Int. J. Biol. Macromol.* **11**, 56–58.
North, A. C. T. (1991). *Biochem. Soc. Symp.* **57**, 35–48.
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.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Quarmby, R., Norden, D. A., Zagalsky, P. F., Ceccaldi, H. J. & Daumas, R. (1977). *Comput. Biochem. Physiol. B*, **56**, 55–61.
Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.
Sheldrick, G. M. (1998). *SHELX: Applications to Macromolecules.* Dordrecht: Kluwer Academic Publishers.
Stuhrmann, S., Bartels, K. S., Braunwarth, W., Doose, R., Dauvergne, F., Gabriel, A., Knochel, A., Marmotti, M., Stuhrmann, H. B., Trame, C. & Lehmann, M. S. (1997). *J. Synchrotron Rad.* **4**, 298–310.
Thompson, J., Higgins, D. & Gibson, T. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Vriend, G. & Sander, C. (1993). *J. Appl. Cryst.* **26**, 47–60.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Wright, C. E., Rafferty, J. B., Flower, D. R., Groom, C., Findlay, J. B. C., North, A. C. T., Phillips, S. E. V. & Zagalsky, P. F. (1992). *J. Mol. Biol.* **224**, 283–284.
Zanotti, G., Berni, R. & Monaco, H. L. (1993). *J. Biol. Chem.* **268**, 10728–10738.