Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Derivative manipulation in the structure solution of the integral membrane LH2 complex

S. M. Prince,[a] G. McDermott,[a]† A. A. Freer,[a] M. Z. Papiz,[c] A. M. Lawless,[c] R. J. Cogdell[b] and N. W. Isaacs[a]*

[a]Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, Scotland, [b]Division of Biochemistry and Molecular Biology, University of Glasgow, Glasgow G12 8QQ, Scotland, and [c]CCLRC Daresbury Laboratory, Daresbury, Warrington WA4 4AD, England

† Present address: Macromolecular Crystallography Facility, Lawrence Berkeley National Laboratory, Mailstop 6-2100, 1 Cyclotron Road, Berkeley, CA 94720, USA.

Correspondence e-mail: neil@chem.gla.ac.uk

Received 19 February 1999
Accepted 7 May 1999

The structure of the peripheral light-harvesting complex from *Rhodopseudomonas acidophila* strain 10050 was determined by multiple isomorphous replacement methods. The derivatization of the crystals was augmented by the addition of a backsoaking stage. The soak/backsoaked data comparison had greater isomorphism and showed simpler Patterson maps than the standard native/soak comparison. Amplitudes from the derivatized then backsoaked crystals and from the derivatized crystals were compared in order to extract a subset of heavy-atom sites. Using this information, the full array of sites were found from a derivative/native comparison, eventually leading to excellent electron-density maps.

1. Introduction

The structure determination of integral membrane proteins (IMPs) by multiple isomorphous replacement and anomalous scattering (MIRAS) techniques is generally complicated by detergent-related factors. Partial encapsulation of the protein by a detergent micelle (Garavito *et al.*, 1986) reduces the effective protein surface area for heavy-atom binding, and heavy-atom salts may interact with detergent causing disruption to the crystalline lattice. The surface available for binding a derivative in an isomorphous manner is further reduced as the hydrophilic extra-membraneous domains of IMPs are often involved in crystal contacts.

The light-harvesting complex from *Rhodopseudomonas acidophila* strain 10050 (LH2) (McDermott *et al.*, 1995) has small extra-membraneous domains (Hawthornthwaite & Cogdell, 1991), and a large proportion of the surface of the protein is thought to be encapsulated by a detergent micelle. The crystals are type II IMP crystals (Michel, 1983), in which all crystal contacts are made between hydrophilic regions.

Prior to the structure determination of the LH2 complex, the structures of only two other types of integral transmembrane proteins had been determined to high resolution by MIR methods: the bacterial reaction centre (Diesenhofer *et al.*, 1985) and porin (Weiss *et al.*, 1991), along with the structure of the membrane-associated protein prostaglandin H2 synthase (Picot *et al.*, 1994). In each of these cases, the derivatization problems were alleviated by the presence of extensive hydrophilic domains. In the case of the bacterial reaction centre from *Rps. viridis*, large extramembranous domains are formed on either side of the membrane by a subunit of the reaction-centre complex and a tightly bound cytochrome. The porin trimer has extensive extracellular loop structures and the interior of each monomer is hydrophilic. Prostaglandin H2 synthase has extensive extra-membranous domains.

The structure determination of LH2 was achieved using standard MIRAS techniques, with an additional backsoaking stage. Backsoaking is normally used to deplete a soaked crystal of excess unbound or weakly bound heavy atoms. In this case, a backsoaked/native Patterson map should have less background and show vectors between heavy atoms in the major binding sites. Backsoaking can also be used to selectively remove a subset of sites by making use of differing chemical properties in a matrix of reagents or different binding-site dynamics in the crystal. In this case, when the soaked and backsoaked data sets are subtracted, the partially removed subset of dynamic sites are those present in the difference Patterson map. This presents a simpler problem for difference Patterson solution, as the number of sites can only be less than the complete array. For partially removed sites, a gain in isomorphism relative to the native/soak comparison can also be expected.

Table 1
Data sets and phasing statistics.

Each data set was collected from a single crystal to a resolution of 3.0 Å.

Data set	Completeness (%)	$R_{\rm merge}$† (%)	$R_{\rm iso}$‡ (%)	Number of sites	Phasing power§¶
Native	98.0	3.0	—	—	—
$KHgI_4$††	93.4	3.6	11.7	6	0.9 (0.8)
Combined Pt	90.6	3.7	12.8	6	1.4 (1.0)
K_2HgI_4 (BS‡‡)	90.5	3.6	3.8§§	−1§§	(0.6)§§
Combined Pt (BS‡‡)	86.3	4.0	8.6§§	−3§§	(1.1)§§
$K_2Pt(CN)_4$††	96.8	3.3	13.0	3	1.0 (1.0)
K_2PtCl_4††	95.8	3.8	11.1	3	1.5 (1.2)

† $R_{\rm merge} = \sum_h \sum_j |I(h) - I(h)_j| / \sum_h \sum_j I(h)_j$, where $I(h)$ is the mean intensity for reflection h after rejection of outliers. ‡ $R_{\rm iso} = \sum_h ||F_{\rm PH}| - |F_{\rm P}|| / \sum_h |F_{\rm P}|$, where $|F_{\rm PH}|$ and $|F_{\rm P}|$ are the derivative and native structure-factor amplitudes, respectively. § Phasing power = $\sum_h |F_{\rm Hcalc}| / \sum_h \varepsilon$, where $|F_{\rm Hcalc}|$ is the calculated structure-factor amplitude for the heavy atom and ε is the residual lack of closure (of the vector triangle $|F_{\rm P}|$, $|F_{\rm PH}|$ and $|F_{\rm Hcalc}|$) for reflection h (calculated over the isomorphous resolution interval 16–4.0 Å). ¶ Acentric (centric). †† Derivatives used in the final phase calculation. ‡‡ BS denotes backsoaked. §§ Expressed with respect to soaked data set.

2. Methods

An artificial mother liquor was prepared for the LH2 crystals by modifying the crystallization drop conditions [1 *M* K_2HPO_4, 350 m*M* NaCl, 2.5%(*w*/*v*) benzamidine hydrochloride, 0.75%(*w*/*v*) *β*-octyl glucoside, LH2 protein at 5 mg ml^{-1}]. The artificial mother liquor consisted of 1.5 *M* K_2HPO_4, 350 m*M* NaCl with 0.5%(*w*/*v*) *β*-octyl glucoside. 1 ml of this solution was equilibrated by vapour diffusion against 8 ml 2.3 *M* $(NH_4)_2SO_4$ pH 9.3 for 36–48 h. The pH of the crystallization drop for LH2 is high (11–12) and the equilibration step employed when making artificial mother liquor is primarily a pH adjustment, as ammonia evolved from the well solution dissolves in the drop. The high pH of the LH2 mother liquor (10.6) limited the solubility of many heavy-atom compounds. Heavy-atom salts were dissolved in the equilibrated drop solution and native crystals were added immediately. As the artificial mother liquor contains phosphate, which is known to react with heavy-atom salts, in particular those of platinum (Blundell & Johnson, 1976), it was critical to add the heavy-atom reagent after the equilibration stage.

Native crystals were soaked in mother liquor for 8–12 h prior to data collection. Crystals were derivatized in the standard soaking manner using K_2HgI_4 with KI and a combination of K_2PtCl_4, $K_2Pt(CN)_4$ and $K_2Pt(NO_3)_4$. The derivatization conditions were found from an analysis of data amplitudes collected to a resolution of 4.5 Å [Siemens Multiwire detector with Cu *Kα* radiation from a rotating-anode source; the data were processed with *XDS* (Kabsch, 1988*a*,*b*)]. As difference Patterson maps were difficult to interpret, the degree of non-isomorphism was determined by the variation with resolution of the mean weighted isomorphous fractional residual ($wR_{\rm iso}$)[1] calculated with the *CCP*4 program *SCALEIT* (Collaborative Computational Project, Number 4, 1994). The point at which the gradient of the $wR_{\rm iso}$ against resolution plot became positive was judged to be the isomorphous limit, as the signal arising from an isomorphous derivative should not become more significant with resolution.

The combined Pt salts were screened in terms of soak time and concentration, with all three reagents being held at the same molarity. The K_2HgI_4 derivative was studied in terms of soak time, concentration and the concentration of additional KI required. The best soak conditions were found to be 3 m*M* of each Pt salt, or 1 m*M* K_2HgI_4 with 10 m*M* KI, with a soak time of 12–16 h in both cases. Data were collected at the Daresbury synchrotron at station 9.6 on a 30 cm MAR scanner to a resolution of 3.0 Å. The illumination wavelength was 0.87 Å, giving relatively strong anomalous scattering for both Hg and Pt. As well as native and derivative data sets, data were collected from equivalently derivatized crystals which were then backsoaked in artificial mother liquor (4–6 h). This resulted in crystallographic data sets for native, soaked and backsoaked crystals. Data processing was carried out using the program *MOSFLM* (Leslie, 1992) and all successive computations used the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994).

The dynamic sites observed in the soak/backsoak comparisons are not completely removed. Difference Patterson maps calculated with native and backsoaked amplitudes were very similar to those calculated with native and soaked amplitudes. This indicates that the dynamic sites are only partially removed by the further soak. The comparison of soaked and backsoaked data sets gives difference amplitudes for sites depleted from the derivatized complex by the backsoak.

Solving the simpler soaked/backsoaked comparison difference Pattersons was trivial in both cases (solutions were obtained by the program *SHELX* in Patterson search mode; Sheldrick, 1990). Successive cross-phasing gave the solution for all sites of the derivative soaks and eventually the phase set leading to the determination of the structure of LH2.

The reagents contributing to the combined Pt derivative were later split into their constituents and it was found that the

[1] $wR_{\rm iso} = [\sum_h |(F_P^2 - F_M^2)/\sigma^2(F_P^2)| + |(F_{\rm PH}^2 - F_M^2)/\sigma^2(F_{\rm PH}^2)|] / [\sum_h F_P^2/\sigma^2(F_P^2) + F_{PH}^2/\sigma^2(F_{\rm PH}^2)]$, where $F_M^2 = [F_{\rm P}^2/\sigma^2(F_{\rm P}^2) + F_{\rm PH}^2/\sigma^2(F_{\rm PH}^2)] / [1/\sigma^2(F_{\rm P}^2) + 1/\sigma^2(F_{\rm PH}^2)]$, $\sigma^2(F_{\rm P}^2) = \sigma^2(F_{\rm P}) \times 4F_{\rm P}^2$.

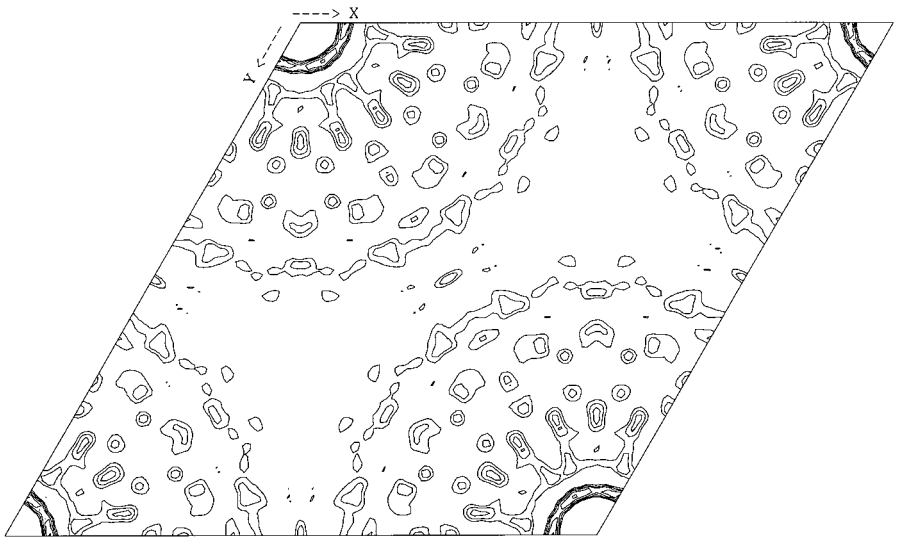
cyano compound provided the sites which were depleted. The remaining sites were provided by the chloride compound. Table 1 summarizes the data sets with their phasing statistics.

3. Results and discussion

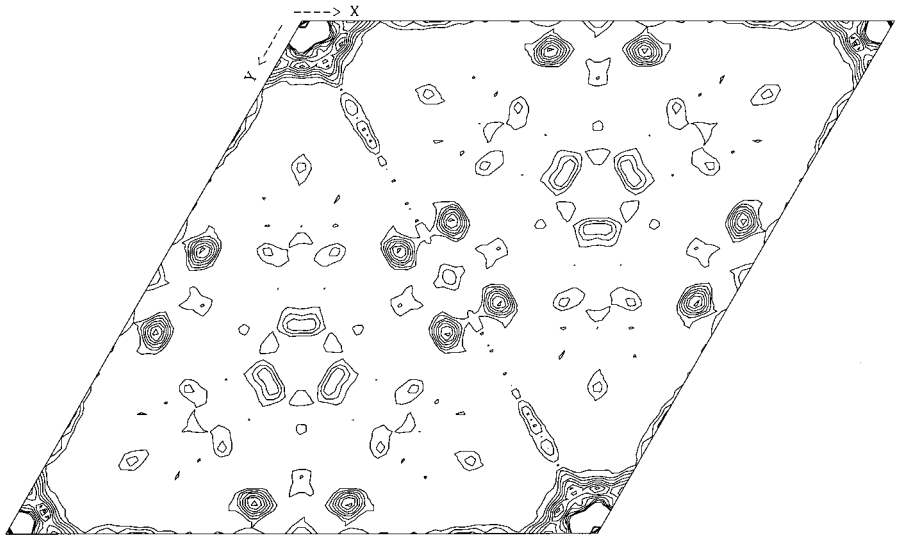
The LH2 complex crystallizes in the rhombohedral space group $R32$. Harker vectors fall at $w = 0$ and $w = 2z$ in the Patterson map. The LH2 complex was found to possess C_9 molecular symmetry (McDermott *et al.*, 1995), the ninefold rotation axis being coincident with the crystallographic threefold axis. This led to an asymmetric unit comprising one-third of the LH2 complex, with each derivative site having two non-crystallographic symmetry (NCS) equivalents in the asymmetric unit. Additionally, the NCS-related derivative sites all had the same z coordinate, leading to cross vectors and Harker vectors falling on the $w = 0$ plane in the Patterson map. The clarity of the difference Pattersons encountered in the soak/backsoak comparisons, with regard to native/derivative comparisons (see Fig. 1), is not entirely a consequence of a gain in isomorphism. The Patterson function is an auto-correlation function; an auto-correlation map resulting from N distinct points will give $N(N-1)$ cross-vectors. The number of cross-vectors will dramatically affect the root-mean-square density of the Patterson map. Therefore, a map contoured at 'σ' will appear to have a greater signal-to-noise ratio for a single-site derivative than a six-site derivative.

In the case of the Pt combined derivative, the chemical difference between the association of the cyano and chloro compounds with LH2 led to the former subset being more dynamic. The Hg dynamic site is found close to a crystal contact (see Fig. 2), causing the binding at this site to be different from the NCS equivalents.

The backsoaked derivative subsets could be used for both self- and cross-phasing. Cross-phasing the Pt soak/backsoak comparison with the single Hg site from the Hg soak/backsoak comparison revealed Pt subsites consistent with an independent difference Patterson solution of the Pt soak/backsoak comparison. Extending the cross-phasing to native/soak difference amplitudes led to the complete derivative site array. Additionally, a single isomorphous replacement phase set calculated using the Hg site from the soak/backsoak comparison with solvent flattening produced a poor map but showed the annular structure of the LH2 molecule.



(*a*)



(*b*)

Figure 1
(*a*) The $w = 0$ (Harker) section of the K_2HgI_4 derivative difference Patterson (origin removed) and (*b*) the equivalent soak/backsoak section (contours at 1σ and integer multiples thereof).

4. Conclusions

In the case of LH2 crystals, the backsoaking protocols outlined here were integral to the solution of the structure. A considerable gain in isomorphism was observed for the soak/backsoak comparison and this is illustrated in the phasing powers of the derivatives. In the soak/backsoak comparisons, the effective number of sites is lower, whilst the phasing power of the derivative is respectable when compared with that from the full array of sites in the native/soak comparison. For the K_2HgI_4 derivative backsoak and soak, an increase in the number of effective sites from one to six ought to give a corresponding $(6)^{1/2}$ increase in phasing

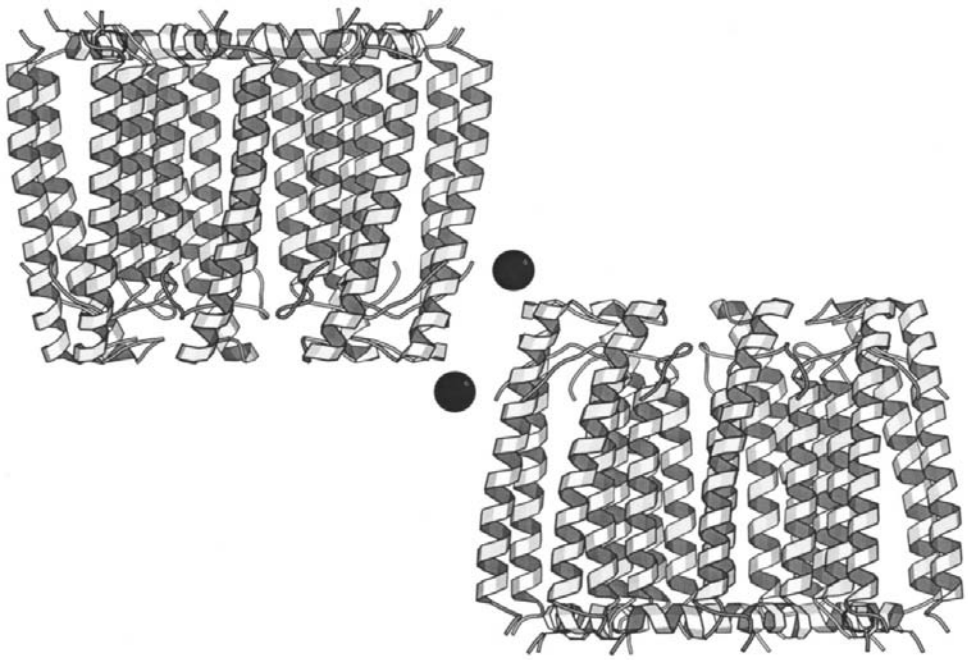


Figure 2
The local environment of the backsoak depleted K_2HgI_4 site, at the N-terminus of the protein (figure produced using *MOLSCRIPT*; Kraulis, 1991).

power (neglecting occupancies); however, only a slight increase is encountered. The phases gained from this technique were improved by the inclusion of the anomalous scattering contribution for the full derivative array. This phase set was improved using solvent flattening (Wang, 1987), which produced a map of sufficient quality for model building. Subsequently, NCS averaging was employed for phase extension to a resolution of 2.5 Å, which gave an electron-density map of exceptional quality.

Many proteins are constructed from identical subunits, and this may be detected at an early stage through self-rotation function analysis. The presence of NCS is often used to improve phase estimates. This study shows that NCS may be useful in the derivatization process. With luck, a site may be located which, whilst being NCS related to others, has different dynamics. As the LH2 structure solution shows, this difference can be used to the crystallographer's advantage.

This work was supported by the (SERC) BBSRC membrane protein initiative. We thank Paul Emsley for help with general computing. We are grateful for the provision of beam time at the Daresbury Synchrotron over the duration of this project.

References

Blundell, T. & Johnson, L. (1976). *Protein Crystallography*, pp. 196–198. London: Academic Press.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Diesenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. (1985). *Nature (London)*, **318**, 618–624.
Garavito, R. M., Markovic-Housley, Z. & Jenkins, J. A. (1986). *J. Cryst. Growth*, **76**, 701–709.
Hawthornthwaite, A. M. & Cogdell, R. J. (1991). *The Chlorophylls*, edited by H. Scheer, pp. 493–528. Boca Raton: CRC.
Kabsch, W. (1988*a*). *J. Appl. Cryst.* **21**, 67–71.
Kabsch, W. (1988*b*). *J. Appl. Cryst.* **21**, 916–924.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACMB Newslett. Protein Crystallogr.* **26**.
McDermott, G., Prince, S. M., Freer, A. A., Hawthornthwaite-Lawless, A. M., Papiz, M. Z., Cogdell, R. J. & Isaacs, N. W. (1995). *Nature (London)*, **374**, 517–521.
Michel, H. (1983). *Trends Biochem. Sci.* **8**, 56–59.
Picot, D., Loll, P. J. & Garavito, R. M. (1994). *Nature (London)*, **357**, 243–249.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Wang, B.-C. (1987). *Methods Enzymol.* **115**, 90–112.
Weiss, M. S., Kreusch, A., Nestel, U., Welte, W., Weckesser, I. & Shultz, G. E. (1991). *FEBS Lett.* **280**, 379–382.