Phil. Trans. R. Soc. Lond. B **313**, 367–378 (1986)
Printed in Great Britain

Structural studies of Rubisco from tobacco

By M. S. Chapman[1], W. W. Smith[1]†, S. W. Suh[1], D. Cascio[1], A. Howard[1]‡, R. Hamlin[2], N.-h. Xuong[2] and D. Eisenberg[1]

[1] *Molecular Biology Institute and Department of Chemistry and Biochemistry, University of California, Los Angeles, California* 90024, *U.S.A.*

[2] *Departments of Biology, Chemistry, and Physics, University of California, San Diego, California* 92093, *U.S.A.*

[Plates 1 and 2]

An electron density map of ribulose 1,5-bisphosphate carboxylase–oxygenase (Rubisco) from tobacco (*Nicotiana tabacum*) has been obtained by X-ray crystallography at a nominal resolution of 0.34 nm. Phases were determined by multiple isomorphous replacement with three heavy atom derivatives and then refined by solvent flattening.

Rubisco is barrel-shaped, and has (422) symmetry. The fourfold axis runs down an open central channel, concentric with the barrel. The molecule measures 10.5 nm along the fourfold axis, and has a diameter of 13 nm perpendicular to the fourfold axis at the widest point. The diameter of the central channel is 2.8 nm at the centre of the molecule, and 0.6 nm at its narrowest constriction.

Portions of the polypeptide backbone of the protomer have been traced and some 127 residues have been assigned to 14 alpha-helices.

The amino acid sequences of Rubisco from *Rhodospirillum rubrum* and from the large subunit of tobacco are sufficiently similar to suggest that the two chains are folded in the same general way.

Abbreviations

EDTA	ethylenediamine tetraacetate
ESP	electronic stationary picture method
MAD	multiwire area detector
MIR	multiple isomorphous replacement
Rubisco	Ribulose 1,5-bisphosphate carboxylase–oxygenase
Tris	Tris(hydroxymethyl)aminomethane

Introduction

Ribulose 1,5-bisphosphate carboxylase–oxygenase (Rubisco) (EC 4.1.1.39), catalyses the initial steps in both the pentose phosphate cycle of photosynthesis and its competing pathway, photorespiration (Miziorko & Lorimer 1983). This dual function presents a tantalizing challenge for enzyme modification. Partly to provide a structural basis for such work, and also to illuminate general features of Rubisco function, regulation and assembly, we have undertaken a structural study of plant Rubisco. We report progress, and some initial results.

† Present address: Monsanto Corporation, 800 N. Lindbergh Blvd., St Louis, Missouri 63166, U.S.A.
‡ Present address: Genex Corporation, 16020 Industrial Drive, Gaithersburg, Maryland 20877, U.S.A.

The crystals used in this work were grown from tobacco (*Nicotiana tabacum*), in the third of three crystal forms (Baker *et al.* 1977; Johal *et al.* 1980). Crystals suitable for X-ray analysis have also been prepared from potato (Johal *et al.* 1980), *Alcaligenes eutrophus* (Bowien *et al.* 1980), spinach (Andersson *et al.* 1983; Barcena *et al.* 1983), and *Rhodospirillum rubrum* (Schloss *et al.* 1979; Schneider *et al.* 1984; Janson *et al.* 1984).

Tobacco Rubisco, typical of higher plants, consists of eight protomers arranged in (422) symmetry (Baker *et al.* 1977). Each protomer is comprised of one small and one large subunit in tobacco of molecular mass 52.935 kDa and 14.877 kDa respectively. The amino acid sequences of both are known (Amiri *et al.* 1984; Shinozaki *et al.* 1982; Muller *et al.* 1983).

Materials

Rubisco was isolated as described earlier (Johal *et al.* 1980; Baker *et al.* 1977). About 150 g of fresh leaves of tobacco were harvested when 2–3 months old. After washing, the leaves were blended in a Waring blender for 30 s in 150 ml of 0.1 M Tris–HCl (pH 8.5) 0.2 M NaCl, to which had been added 0.2 ml 2-mercaptoethanol, 0.04 g phenylmethylsulphonyl fluoride, and 40 g Dowex-1 anion exchange resin. Cell debris was eliminated by filtering through cheesecloth, and by centrifugation at 30000 *g* for 1 h. The crude protein was concentrated threefold by ultrafiltration through an Amicon PM30 membrane at a pressure of 30 lbf in^{-2}†. The following purification by precipitation was repeated three times. The protein was precipitated by dialysis for 18 h against 4 l of buffer C, (25 mM Tris–HCl (pH 7.8) 0.5 mM EDTA, 0.1 mM azide, and 2 mM 2-mercaptoethanol). The precipitate was suspended in buffer C and centrifuged for 10 min at 480 *g*. The pellet was dissolved in about 6 ml buffer C, with a minimal addition of 2 M $(NH_4)_2SO_4$, and redialysed. After centrifugation on the third cycle, the pellet was dissolved in 50 mM phosphate (pH 7.2) 0.5 mM EDTA and 0.1 mM azide with the addition of up to 200 µl $(NH_4)_2SO_4$. The solution was diluted in this buffer to yield about 400 mg of Rubisco at 20 mg ml^{-1}.

Crystals of up to 1 mm^3 in size were obtained from the protein solution by dialysing against 0.2 M phosphate, 0.3 M $(NH_4)_2SO_4$ and 1 mM azide at pH 5.2 for 1 month under nitrogen in pairs of 0.8 ml Zeppezauer tubes (Zeppezauer *et al.* 1968). Rubisco crystallized in space group *I*422 with cell dimensions 14.85 × 14.85 × 13.75 nm.

Three heavy atom derivatives were prepared as follows:

(i) 4 mg of potassium platinum (II) cyanide, $K_2Pt(CN)_4$ (purchased from ICN) was added to the diffusate outside the Zeppezauer tube. Crystals were soaked for three days.

(ii) 0.18 mg of thiomersal, $C_2H_5HgSC_6H_6COONa$, (purchased from Sigma) was added to the diffusate and the crystals were soaked for 24 h.

(iii) Dimethyl mercury, $(CH_3)_2Hg$, (Ventron Corporation, P.O. Box 299, 152 Andover Street, Danvers, Massachusetts 01923) was added with a syringe about 12 h before data collection after mounting the crystal. One end of the capillary contained a drop of mother liquor, and was sealed with wax. Into the open end was inserted a small piece of filter paper. This was carefully wetted with a small drop of dimethyl mercury and the capillary was sealed with wax before the volatile liquid could evaporate.

† 1 lbf in^{-2} (= 'p.s.i.') ≈ 6895 Pa.

Data collection

All data were collected on multiwire area detectors (Hamlin *et al.* 1981; Hamlin 1982), both at UCLA and at UCSD. Two software packages were used to collect data, the electronic stationary picture method (ESP) (Xuong *et al.* 1978), and the multiwire area detector (MAD) package (Weissman 1986*a*). The algorithms differ, but both packages integrate the data to produce Lorentz-corrected intensities.

Details of data collection are given in table 1. Standard methods were used to collect medium-resolution data. Frame exposures of about one minute enabled measurement of symmetry-related reflections up to six times before excessive crystal decay. This facilitates correction for systematic errors, such as absorption, during scaling (Weissman *et al.* 1986). Unlike data for thiomersal and dimethyl mercury, some platinum and native data were observed to a resolution of 0.245 nm. The accuracy of these weak reflections was limited by counting statistics. A less orthodox strategy was required to maximize the number of counts per observation. Starting with the fourfold axis along the X-ray beam, and the *a* or *b* axis 22.5° from the rotation axis, one asymmetric unit could be collected with a rotation about phi of 90°. Anomalous data were not always collected, as the signal was weak. High X-ray flux was obtained with a high current (120 mA). A full data set could be collected from one crystal, with frame exposures of 2.5 min, in 40 h.

Data reduction

Examination of the intensity distributions as a function of exposure time and resolution showed that there was a differential decay of high-resolution reflections with time. To avoid the compromise of well-determined reflections by averaging with poorly determined symmetry equivalents, stringent selection criteria were applied before scaling. Thus data were rejected according to signal: noise, resolution and X-ray exposure. As a minimum, all data for which $I/\sigma(I)$ was less than 2 were discarded, where I is the intensity of an observation, and $\sigma(I)$ its standard deviation. Additionally, up to half of the observations were removed because they were collected in frames that scaled poorly to others, presumably because of progressive radiation damage and slight crystal slippage. The tracking procedure of MAD (Weissman 1986*a*), showed that crystals sometimes slipped by several tenths of a degree, even with capillary cooling (Hamlin 1982). This results in the clipping of reflections. Removal of poor frames and weak observations had a striking effect. For example, discarding DMM1 observations with signal:noise less than 2 reduced the weighted internal R-factor on I from 9 to 6.4%. The apparent del F of DMM5 relative to native was reduced from 19.5 to 14.5% after similar treatment of both native and derivative data sets.

Data were scaled by Fourier scaling, an extension of the Fourier–Bessel method (Weissman *et al.* 1986), suitable for application to rectangular (x,y) detectors. Data were subdivided into frames of 2–5 degrees of rotation during collection. Determination of scale constants for individual frames was alternated with determination of a two-dimensional cosine scaling surface to minimise R_{int} (table 1). The surface is given by:

$$A(x,y)_{hi} = \sum_{m=0}^{M} \sum_{n=0}^{N} a_{mn} \cos(mx_{hi}) \sin(ny_{hi}),$$

in which n and m are integers, h is the reflection index, and i the frame number. Scaling converged to give R-factors of 4.5–8.1 %, (table 1), within three iterations, requiring at most second-order terms in m and n. Because of a lack of redundancy of reflections, scaling was more successful when other data sets were normalized by native set N03, the scaling of which had been boot-strapped from another data set.

Both native data sets were incomplete after scaling. N04 was better at low resolution. At resolutions lower than 0.37 nm, missing reflections were filled from N03. At higher resolutions, N03 was superior, so reflections missing from N03 were filled with those from N04. A small number of reflections, especially at low resolution, were added from other data sets. All other data sets had effectively been scaled to N03. Additionally, before merging, other native data sets were scaled to N03 according to the method of Weissman *et al.* (1986). All derivative data were similarly scaled to the merged native data.

The search for heavy atoms

Rubisco is a large protein that reacts with many heavy metal reagents. It was found that most sulphydryl reagents destroy crystal order, but that many other reagents were unreactive. This led us to follow the suggestions of Sigler & Blow (1965), and Petsko *et al.* (1978) for diminishing competitive binding from solvent. Ammonium ions were removed by dialysis before addition of precious metal complexes to the protein, and phosphate was removed before the addition of actinides. However, many tetra- and hexa-coordinate precious metal complexes, (e.g. K_2PtCl_6), organomercurials, as well as lead, thallium, mercury, thorium and uranyl salts, bind and destroy the crystal order, even at substoichiometric concentrations.

This problem forced us to use the original ammonium sulphate and phosphate buffer for all heavy atoms, some of which were reported earlier (Eisenberg *et al.* 1978). The following derivatives were prepared in the original buffer. Dimethyl mercury was less destructive than many other organomercurials, perhaps because of its hydrophobicity. The cyanides, unlike other platinum and gold complexes, do not react with proteins by SN2 substitution, but rather by electrostatic attraction, and binding is likely to be moderated by phosphate (Blundell & Johnson 1976). $K_2Pt(CN)_4$ was found to have identical binding sites to K_3IrCl_6. $KAu(CN)_2$ had the same sites as thiomersal. $K_2Pt(CN)_4$ and thiomersal were selected because of superior isomorphism with the native enzyme. $K_2Pt(CN)_4$ had a single binding site, and the three-dimensional difference Patterson could be solved by inspection. Multiple-site substitution of the others, coupled with the high crystallographic symmetry, precluded their solution by Patterson methods. Sites were found by difference Fourier syntheses, by using platinum phases. Solutions were consistent with the difference Pattersons, and produced mutually consistent cross-Fouriers (Dickerson *et al.* 1967).

Heavy atom refinement

Heavy atom occupancies were low, because high substitution produced disorder or lack of isomorphism with the native enzyme. Data in table 1 suggest that a significant proportion of the small del Fs are the result of random error in native and derivative data. Heavy atom parameter refinement was attempted in several ways: phased refinement (Weissman 1986*b*), manual flattening of residual maps, and origin-removed Patterson refinement (Terwilliger &

TABLE 1. STATISTICS FOR DATA COLLECTION AND REDUCTION

	native		thiomersal		dimethyl mercury		$K_2Pt(CN)_4$	
data set	N03	N04	HG1	HG3	DMM5	DMM1	PT4	PT9
detector	UCLA	UCLA	UCSD	UCSD	UCSD	UCLA	UCLA	UCLA
X-ray source[1]/kW	4.8	2	1.1	1.2	1.2	2.4	1.6–4.8[7]	2.2
crystal size/mm^2	0.9	0.27	0.67	—	—	0.65	0.52	0.83
collimator/mm^2	0.8	0.8	1.0	0.5	0.5	1.0	1.0	1.0
crystal-detector/mm	682	645	635	660	660	650	635	640
detector half-angle/deg[3]	22	15.8	10	18.2	18.2	13.3	4.2–19.0	16
rotation step[2]/deg	0.1	(0.1)	0.09	0.09	0.09	(0.1)	0.09	(0.1)
exposure/frames s^{-1}	150	59	45	40	40	45	23–150	57
count rate/kHz	5.5	18	25	14	10	18	7–19	22
resolution/mm	0.8–0.25	1.5–0.275	3.5–0.4	1.12–0.295	1.16–0.295	8.0–0.33	8.0–0.25	2.4–0.29
number of reflections:								
collected	50154	59987	34994	68221	67326	42375	55868	33099
input to scaling	29808	34885	29099	31065	28472	29563	39032	23292
scaled and unique	12306	9397	5616	6952	8332	7665	6992	10831
phased to 0.34nm	—	—	5575	2294	7013	1409	5908	4335
phased (total)[4]	—	—	5575	2294	7997	1417	6576	4378
R_{int}[5]/%	8.1	4.5	5.8	7.7	6.3	6.4	6.5	5.4
del F[6]/%	—	—	9.3	10.6	14.5	12.0	14.0	10.0

[1] Cu Kα 40kV graphite monochromatized rotating anode X-rays were used, except for HG1, for which Cu Kα Ni-filtered X-rays from a Picker sealed tube were used. X-ray attenuation was reduced by a helium chamber between the crystal and the detector.

[2] Microstepping was used with MAD.

[3] The Bragg angle to the centre of the detector.

[4] If a reflection was present in one data set, it was excluded from the other data set of the same derivative. 9763 of a possible 15188 native reflections had $I/\sigma(I) > 2$ to 0.34 nm as did 15304 reflections at all resolutions.

[5]
$$R = \sum_{h}^{\text{sym}} \sum_{i=1} w_i \frac{(A_{hi} I_{hi} - \langle I_h \rangle)^2}{\langle I_h \rangle^2}$$

where I_i is the intensity of the ith reflection, of weight w; $\langle I_h \rangle$ is the mean intensity of set h of symmetry-related reflections, and A_{hi} is the local value of the scaling surface.

[6]
$$\text{del}\, F = \frac{\sum_i |F_{\text{nat},i} - F_{\text{der},i}|}{\frac{1}{2}\sum_i |F_{\text{nat},i} + F_{\text{der},i}|}$$

where F_{nat}, F_{der} are the native and derivative structure factor magnitudes respectively.

[7] Power cut during data collection; low-resolution data were collected as the power was brought back up.

Eisenberg 1983). All three methods tended to fit the calculated heavy atom scattering to the observed del *F*s, but overestimated the occupancies. Patterson refinement was least susceptible to this problem, but still suffered from both unacceptably high and correlated occupancies and temperature factors.

Our method of refinement was to adopt Patterson refinement, using data only in the resolution range 0.95–0.4 nm, where heavy atom scattering approximated a gaussian dependence upon resolution. Initial 'average' temperature factors for each derivative were estimated from Wilson plots (Wilson 1949). Individual site temperature factors were refined while manually restraining their 'average'. The parameters obtained were quite stable when occupancies and temperature factors were refined in alternate cycles without constraint. Subsequent phased refinement according to the method of Dickerson *et al.* (1968) was unsuccessful. Phasing from just two derivatives was too poor to refine a third, and inclusion of the refining derivative in the phasing led to exaggeration of its occupancy, as discussed by Dickerson *et al.* (1967).

TABLE 2. HEAVY ATOM PARAMETERS

	overall B[a]		fractional coordinates			B
data set	Å²[b]	occupancy[c]	*X*	*Y*	*Z*	Å²
HG1	0.5760	0.0039	0.1211	0.1761	0.2209	10.6562
		0.0044	0.2758	0.4578	0.2077	10.6416
		0.0012	0.2255	0.4620	0.1960	6.2349
		0.0014	0.2918	0.2852	0.1939	30.8594
HG3	0.5520	0.0027	0.1213	0.1751	0.2194	13.4821
		0.0038	0.2737	0.4580	0.2067	17.9865
		0.0011	0.2173	0.4651	0.1837	6.2300
		0.0021	0.2953	0.2847	0.1970	25.0154
DMM5	−0.3340	0.0052	0.2263	0.1443	0.1486	4.8384
		0.0052	0.1117	0.2249	0.0019	5.3963
		0.0039	0.1603	0.1781	0.0000	11.5711
DMM1	−0.0200	0.0037	0.2256	0.1459	0.1504	4.5324
		0.0042	0.1109	0.2244	0.0008	10.6779
		0.0028	0.1561	0.1774	0.0020	29.0644
PT4	−1.8530	0.0069	0.0750	0.2180	0.2280	18.5160
PT9	−0.4790	0.0052	0.0747	0.2117	0.2306	18.5000

[a] *B* is the isotropic temperature factor of the derivative data relative to the native. Scaling constants ranged from 0.982 to 0.993.
[b] 1 Å = 10^{-10} m = 10^{-1} nm.
[c] Occupancy is on an arbitrary scale.

Final heavy atom parameters are given in table 2, and were used to phase reflections at all resolutions. The phasing statistics are shown in table 3 and figure 1, and reveal a low signal:noise. HG1, DMM5 and PT4 were the best derivative data sets, but were incomplete. Missing reflections were filled from HG3, DMM1, and PT9, but were retained as separate data sets with their own parameters for phasing. Thus derivative reflections were not duplicated. Details are shown in table 1. Despite the low signal:noise, the 0.35 nm multiple isomorphous replacement (MIR) map, calculated from best phases (Blow & Crick 1959), showed the molecular envelope, although it resembled a lower resolution map. This encouraged us to proceed to phase refinement by the method of density modification.

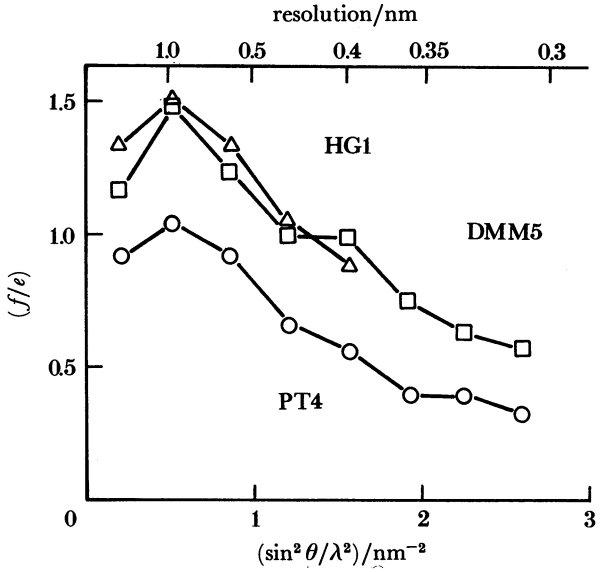

FIGURE 1. Phasing power of the three principal derivative data sets. f is the calculated heavy atom scattering, and e the lack of closure error (Blow & Crick 1959).

TABLE 3. PHASING STATISTICS

resolution/nm	1.092	0.685	0.534	0.453	0.399	0.361	0.333	0.310	all
figure of merit[a]	0.58	0.58	0.53	0.50	0.39	0.26	0.19	0.09	0.37
fraction of native with $I/\sigma(I) > 2$	68	78	71	75	58	53	50	30	57

[a] All native reflections are included regardless of the presence of derivatives.

DENSITY MODIFICATION

Phases were refined by the application of solvent flattening to the MIR map, by the reciprocal-space method devised by Cascio and Williams, extending the real-space procedure of Wang (1986). A description of the method, and of its application to Rubisco will be given elsewhere. The molecular envelope of Rubisco was determined from a blurred electron density map, according to the volume fraction of protein in the crystal. In the usual real-space method, the blurred map is calculated by computing a weighted average of each point with its neighbours. The reciprocal-space method replaces this step with a Butterworth lowpass filter (Gonzales & Wintz 1977), applied to the structure factors. Also, the electron density of the protein–solvent boundary was determined, not from a sorted list of electron density, but from a density histogram.

A summary of the application and results of solvent flattening is shown in table 4. The experimentally determined volume fraction of solvent in the crystal is 56% (Baker *et al.* 1977). In the initial cycles, lower estimates were used. Earlier attempts had shown that the automatic boundary determination was slow to flatten large noise peaks in the solvent region. Therefore, the first envelope or mask was drawn manually. Heavy atom sites were also flattened. This procedure of phase refinement converged within three cycles of automatic envelope determination. There were still some noise peaks in the solvent, but the initial manual mask could not be reapplied, because an alpha-helix had appeared during automatic refinement in

TABLE 4. PHASE REFINEMENT

cycle[a]...	1	2	3	4	5	6a	6b	7a	7b
envelope	man	auto	auto	auto	man	auto	man	auto	man
solvent fraction (percent)	45	47	47	47	35	47	35	56	35
⟨phase change⟩/deg	25.9	19.9	9.2	6.0	21.3	11.6	15.0	13.6	15.1
⟨cumulative phase change⟩/deg	25.9	31.7	34.8	36.6	36.4	37.7	37.5	38.3	38.7
⟨figure of merit⟩	0.62	0.68	0.71	0.73	0.72	0.73	0.74	0.72	0.75
R/percent	46	37	33	30	32	30	29	32	27
$r(F_o,F_c)$[b] (~correlation coefficient)	0.85	0.91	0.93	0.94	0.93	0.94	0.94	0.93	0.95

[a] On cycles 6 and 7, both automatically determined and manually determined masks were applied, and combined with previous phases in the same cyle.

[b] $r = \langle F_o \times F_c \rangle / \sqrt{(\langle F_o \times F_c \rangle \times \langle F_c \times F_c \rangle)}$, where F_o, F_c are observed and calculated structure factors, respectively.

a region that had been flattened by the first manual mask. A more conservative manual mask was then made (35 % solvent) so that it could be applied repeatedly without risk of flattening protein peaks. Phase recombination of previous phases with the results of both manual and automatic masking on the same cycle proved to be very effective. Although phasing statistics could be improved, after two cycles there was no further improvement of the map with additional cycles.

RESULTS

The general shape of the Rubisco molecule is shown by the wooden model in figure 2. This model represents the molecular envelope of the 16-subunit oligomer (8 large and 8 small subunits) at low resolution. The molecule is shaped like a barrel, with the fourfold molecular axis running down a central aqueous channel concentric with the barrel. Perpendicular to the barrel axis at the molecular centre are four twofold axes, which relate the 8 protomers (each containing one large and one small subunit) in (422) symmetry. The placement of each subunit is not yet known.

The height of the Rubisco molecule along the barrel (fourfold) axis is 10.5 nm, and its diameter at the widest point is about 13.2 nm. This point is 1.8 nm above (and also below) the centre of the barrel. The molecular diameter is slightly smaller at the centre. The diameter of the aqueous channel varies from 2.8 nm at the centre of the molecule to 0.6 nm at its narrowest constriction, about 3.3 nm above and below the molecular centre. Near the top and bottom of the molecule, 4.9 nm above and below the centre, the diameter of the channel is 1.4 nm. Thus the channel flares out from the constriction, as can be seen in figure 2*b*. Near the top and bottom of the barrel, the molecule is about 4.9 nm in diameter.

A section of the electron density of Rubisco at the nominal resolution of 0.34 nm is shown in figure 3*a*. The section is 0.3 nm thick, perpendicular to the fourfold (barrel) axis, and 0.7 nm above the centre of the molecule. Surrounding the central solvent channel are the four protomers. The plane parallel to this one, but 0.7 nm below, contains the four twofold axes: the horizontal and vertical twofold axes run through gaps that clearly separate the four protomers and twofold axes also run diagonally at an angle of 45°. Two alpha-helices are visible in this section of electron density; one is shown enlarged in figure 3b. About three turns of the helix are visible, and from a fit of a polyalanine backbone into the helix (figure 3*c*), it appears that the direction of the helix is toward the central channel. Some 13 other alpha-helices have

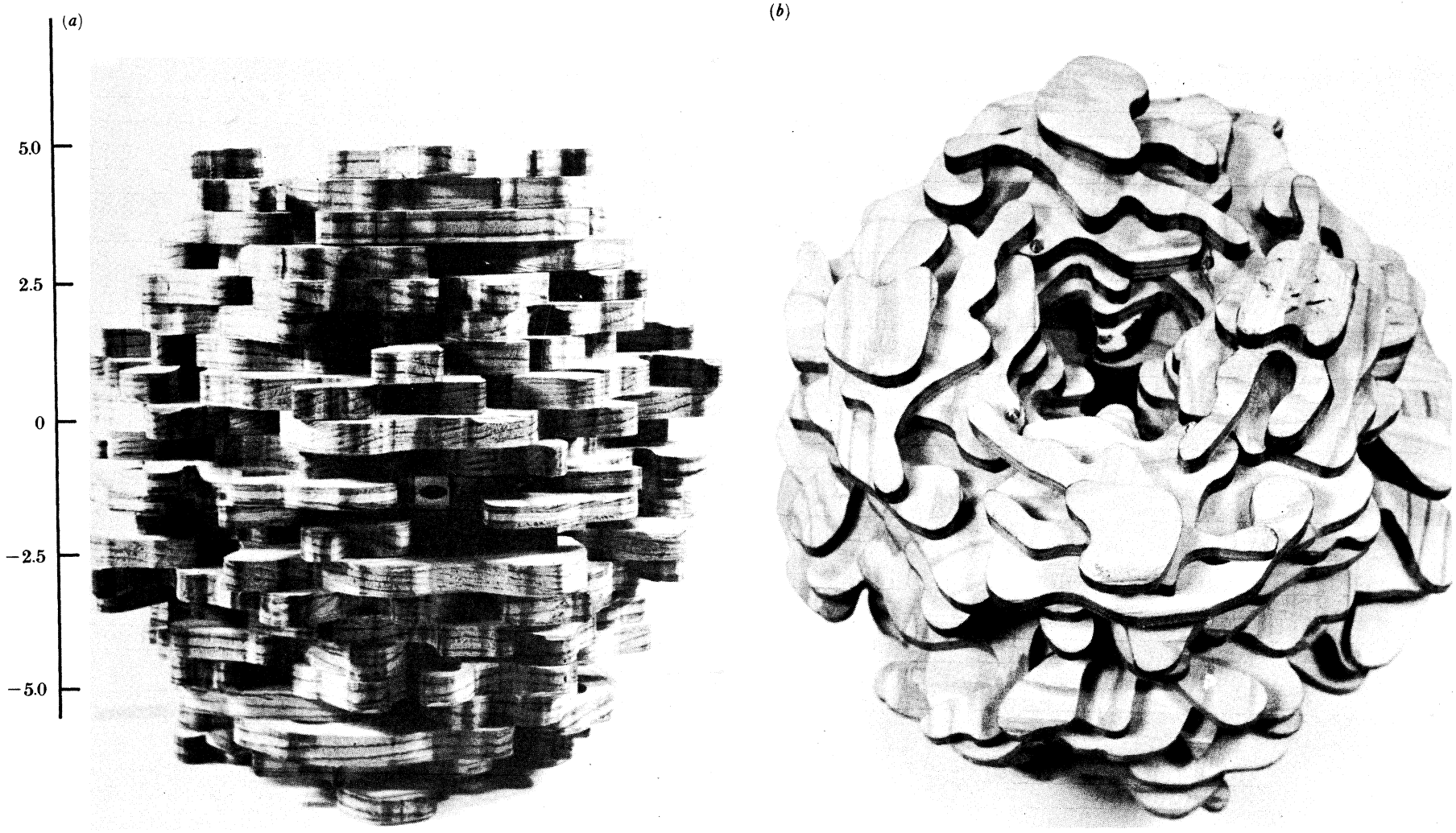

FIGURE 2. A wooden model of Rubisco representing the envelope of the electron density of 0.6 nm resolution. (*a*) View along a twofold axis, showing the barrel shape, about 13.2 nm in diameter at its widest point, and 10.5 nm high. The fourfold axis is vertical. (*b*) View down the fourfold axis showing the central channel, with its constriction not far below the top of the barrel.

(*Facing p. 374*)

FIGURE 3. The solvent-flattened electron density map at a nominal resolution of 0.34 nm. (*a*) A slice of electron density, 0.3 nm thick, 0.7 nm above the centre of the molecule. The dashed grid lines are 2.96 nm apart. The view is down the fourfold axis, showing the nearly empty channel centred among four protomers. One alpha-helix is marked by an arrow. (*b*) A more detailed view of the same slice, showing portions of several alpha-helices. (*c*) The alpha-helix marked in (*a*) and (*b*), displayed with FRODO (Jones 1978). Some of the side chains are apparent, as is the upward direction of the helix. An ideal polyalanine helix is superimposed.

been tentatively identified in each of the eight protomers, three of them lying near to this one with similar orientations. Above these helices sits a small three-stranded beta-sheet. Still further above and near to the central channel, other alpha-helices are oriented more nearly parallel to the fourfold axis.

Although we have not yet been able to trace the full polypeptide backbone, some 200 residues of the total 603 of Rubisco have been found in segments of regular secondary structure.

Discussion

(*a*) *Phase determination*

Our major difficulty in determining the structure of Rubisco from tobacco has been the sensitivity of the crystals to virtually every bound heavy atom that we have introduced for phase analysis. Over the past eight years we have tested about 120 heavy-atom-containing compounds, under a variety of solvent conditions, with mainly negative results. The three most useful compounds are those reported in table 2, and even these are poor by the usual standards of protein crystallography.

Despite the poor quality of our heavy-atom derivatives, we have been able to determine an electron density map that is at least partly interpretable in terms of a polypeptide chain. Among the factors that led to better phases were the following:

(i) Area-sensitive X-ray detectors were used for all data collection.

(ii) During X-ray data collection, frame exposure times were increased until the accuracy of weak reflections was not limited by counting statistics.

(iii) Raw data were screened thoroughly before and during scaling. Reflections collected early during the lifetime of each crystal were not compromised by averaging with decaying symmetry equivalents collected later.

(iv) Care was taken to prevent overestimation of heavy-atom occupancies, which is a possible consequence of feedback in phased refinement (Dickerson *et al.* 1967). Thus, while the phasing statistics we report here are poor, the calculated phase probability distributions are probably not oversharpened. In trials with phased refinement, in which heavy-atom occupancies were permitted to refine upwards, the phasing statistics were improved, but the phases seemed to be degraded.

(v) Heavy-atom parameters were refined only at medium resolution, where heavy-atom scattering approximated gaussian dependence on resolution. Refinement at medium resolution avoided attenuation of the high-resolution signal.

(vi) Solvent flattening was crucial for phase improvement. As a prelude to automatic refinement, a manual mask was used to reduce large noise peaks that were clearly in the solvent. Then combined application of a manual mask with a less conservative automatic mask proved effective; whereas in the multiple isomorphous replacement map one alpha-helix was apparent and two others discernible, 14 helices were apparent in the solvent-flattened map.

Further phase refinement (Wang 1986) is being attempted, based on the partial structure of the current map.

(*b*) *Molecular shape*

The molecular shape of plant Rubisco is similar to that depicted earlier in models built from 16 spheres (Baker *et al.* 1977; Eisenberg *et al.* 1978), but smoother in overall outline and with no distinctive subunit boundaries. The overall shape also is consistent with the small-angle neutron scattering studies of Donnelly *et al.* (1984).

(*c*) *Relationship of amino acid sequence to structure*

Earlier we suggested (Janson *et al.* 1984) that the overall folding of the large subunit of plant Rubisco may be similar to that of Rubisco from *R. rubrum*. The basis of this speculation is the quantitative similarity (Sweet & Eisenberg 1983) of the amino acid sequences of the tobacco large subunit (Shinozaki & Sugiura 1982) and of *R. rubrum* Rubisco (Nargang *et al.* 1984). From the overall shape of the Rubisco molecule reported here, it appears that the large subunit must have an elongated, prolate shape. This is the shape of the *R. rubrum* subunit reported by Donnelly *et al.* (1984), from small-angle neutron scattering.

One feature of the Rubisco structure evident from our electron density map is an abundance of alpha-helices. As noted above, about 14 helices are present. Is the amino acid sequence consistent with a large number of alpha-helices? Recently we have found that the algorithm of Lim (1974) gave a reaonably accurate prediction for the number and placement of the alpha-helices of the enzyme glutamine synthetase (Janson *et al.* 1986). Applying the Lim algorithm to the known sequence of tobacco Rubisco, we find that 13 alpha-helices are predicted.

A final sequence–structure relationship may be worthy of notice. Wierenga *et al.* (1985) have proposed a sequence 'fingerprint' for dinucleotide binding sites in proteins. The chief characteristic of the fingerprint is three invariant glycyl residues within six amino acid residues on the polypeptide backbone (in relative positions, 1, 3 and 6). This characteristic of the fingerprint is present in the sequences of *N. tabacum* large subunit (starting at residue 403) and of *R. rubrum* (starting at residue 393), as well as in Rubisco from five other species. Although other features of the fingerprint are not perfectly obeyed, it is conceivable that this portion of the sequence forms a dinucleotide binding site.

We acknowledge with thanks discussions with Dr L. Weissman and support from NIH (GM 31299) and USDA (83–CRCR–1–1324).

References

Amiri, I., Salnikow, J. & Vater, J. 1984 Amino-acid sequence of the large subunit of D-ribulosebisphosphate carboxylase/oxygenase from *Nicotiana tabacum*. *Biochim. biophys. Acta* **784**, 116–123.

Andersson, I., Tjader, A.-C., Cedergren-Zeppezauer, E. & Brändén, C.-I. 1983 Crystallization and preliminary X-ray studies of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase complexed with activator and a transition state analogue. *J. biol. Chem.* **258**, 14088–14090.

Baker, T. S., Suh, S. W. & Eisenberg, D. 1977 Structure of ribulose-1,5-bisphosphate carboxylase-oxygenase: Form III crystals. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1037–1041.

Barcena, J. A., Pickersgill, R. W., Adams, M. J., Phillips, D. C. & Whatley, F. R. 1983 Crystallization and preliminary X-ray data of ribulose-1,5-bisphosphate carboxylase from spinach. *EMBO J.* **2**, 2363–2367.

Blow, D. M. & Crick, F. H. C. 1959 The treatment of errors in isomorphous phase analysis. *Acta Cryst.* **12**, 794–802.

Blundell, T. L. & Johnson, L. N. 1976 *Protein crystallography*, p. 230. London: Academic Press.

Bowien, B., Mayer, I., Spiess, E., Pahler, A. Englisch, U. & Saenger, W. 1980 On the structure of crystalline ribulosebisphosphate carboxylase from *Alcaligenes eutrophus*. *Eur. J. Biochem.* **106**, 405–410.

Dickerson, R. E., Weinzierl, J. E. & Palmer, R. A. 1968 A least squares refinement method for isomorphous replacement. *Acta Cryst.* B **24**, 997–1003.

Dickerson, R. E., Kopka, M. L., Varnum, J. C. & Weinzierl, J. E. 1967 Bias, feedback and reliability in isomorphous phase analysis. *Acta Cryst.* A **23**, 511–522.

Donnelly, M. I., Hartman, F. C. & Ramakrishnan, V. 1984 The shape of ribulose bisphosphate carboxylase/oxygenase in solution as inferred from small angle neutron scattering. *J. biol. Chem.* **259**, 406–411.

Eisenberg, D., Baker, T. S., Suh, S. W. & Smith, W. W. 1978 In *Photosynthetic carbon assimilation* (ed. H. W. Siegelman & G. Hing), pp. 271–281. New York: Plenum Press.

Gonzales, R. C. & Wintz, P. 1977 *Digital image processing*, p. 145. Reading, Massachusetts: Addison–Wesley Publishing Co.

Hamlin, R. 1982 Multiwire proportional counters for use in area X-ray diffractometers. *Transactions of the American Crystallographic Association* **18**, 95–123.

Hamlin, R., Cork, C., Howard, A., Neilson, C., Vernon, W., Matthews, D. & Xuong, N.-h. 1981 Characteristics of a flat multiwire area detector for protein crystallography. *J. appl. Cryst.* **14**, 85–93.

Janson, C. A., Smith, W. W. & Eisenberg, D. 1984 Preliminary structural studies of ribulose 1,5-bisphosphate carboxylase/oxygenase from *Rhodospirillum rubrum*. *J. biol. Chem.* **259**, 11594–11596.

Janson, C. A., Kayne, P. S., Almassy, R. J., Grunstein, M. & Eisenberg, D. 1986 DNA-encoded sequence of glutamine synthetase from *Salmonella typhimurium*: Implications for the protein structure. *Gene* (Submitted.)

Johal, S., Bourque, P. B., Smith, W. W., Suh, S. W. & Eisenberg, D. 1980 Crystallization and characterization of ribulose-1,5-bisphosphate carboxylase/oxygenase from eight plant species. *J. biol. Chem.* **255**, 8873–8880.

Jones, T. A. 1978 A graphics model building and refinement system for macromolecules. *J. appl. Cryst.* **11**, 268–272.

Lim, V. I. 1974 Algorithms for prediction of alpha-helical and beta-structural regions in globular proteins. *J. molec. Biol.* **88**, 873–894.

Miziorko, H. M. & Lorimer, G. H. 1983 Ribulose-1,5-bisphosphate carboxylase/oxygenase. *A. Rev. Biochem.* **52**, 507–535.

Muller, K.-D., Salnikow, J. & Vater, J. 1983 Amino acid sequence of the small subunit of D-ribulosebisphosphate carboxylase/oxygenase from *Nicotiana tabacum*. *Biochim. biophys. Acta* **742**, 78–83.

Nargang, F., McIntosh, L. & Somerville, C. 1984 Nucleotide sequence of the ribulosebisphosphate carboxylase gene from *Rhodospirillum rubrum*. *Molec. gen. Genet.* **193**, 220–224.

Petsko, G. A., Phillips, D. C., Williams, R. J. P. & Wilson, I. A. 1978 On the problem of crystal chemistry of chloroplatinite ions: General principles of interactions and interactions with triose phosphate isomerase. *J. molec. Biol.* **120**, 345–359.

Schloss, J. V., Phares, E. F., Long, M. V., Norton, I. L., Stringer, C. D. & Hartman, F. C. 1979 Isolation, characterization, and crystallization of ribulosebisphosphate carboxylase from autotrophically grown *Rhodospirillum rubrum*. *J. Bact.* **137**, 490–510.

Schneider, G., Brändén, C.-I. & Lorimer, G. 1984 Preliminary X-ray diffraction study of ribulose-1,5-bisphosphate carboxylase from *Rhodospirillum rubrum*. *J. molec. Biol.* **175**, 99–102.

Shinozaki, K. & Sugiura, M. 1982 The nucleotide sequence of the tobacco chloroplast gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Gene* **20**, 91–102.

Sigler, P. B. & Blow, D. M. 1965 A means of promoting heavy atom binding in protein crystals. *J. molec. Biol.* **14**, 640–644.

Sweet, R. M. & Eisenberg, D. 1983 Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. *J. molec. Biol.* **171**, 479–488.

Terwilliger, T. C. & Eisenberg, D. 1983 Unbiased three-dimensional refinement of heavy atom parameters by correlation of origin-removed Patterson functions. *Acta Cryst.* A **39**, 813–817.

Wang, B. C. 1986 Resolution of the phase ambiguity in macromolecular crystallography. In *Diffraction methods for biological macromolecules*. Methods in Enzymology (ed. H. W. Wycoff, C. H. W. Hirs & S. N. Timashiff) New York: Academic Press.

Weissman, L. W. 1986*a* (In preparation.)

Weissman, L. W. 1986*b* (In preparation.)

Weissman, L., Stauffacher, C. & Eisenberg, D. 1986 Errors in film data collection: Fourier-Bessel scaling. (In preparation.)

Wierenga, R. K., De Maeyer, M. C. H. & Hol, W. G. J. 1985 Interaction of pyrophosphate moieties with alpha-helices in dinucleotide binding proteins. *Biochemistry, Wash.* **24**, 1346–1357.

Wilson, A. J. C. 1949 The probability distribution of X-ray intensities. *Acta Cryst.* **2**, 318–322.

Xuong, N.-h., Freer, S., Hamlin, R., Nielsen, C. & Vernon, W. 1978 The electronic stationary picture method for high speed measurement of reflection intensities from crystals with large unit cells. *Acta Cryst.* A **34**, 289–296.

Zeppezauer, M., Eklund, H. & Zeppezauer, E. S. 1968 Micro-diffusion cells for the growth of single protein crystals by means of equilibrium dialysis. *Archs Biochem. Biophys.* **126**, 564–573.

Discussion

R. W. PICKERSGILL (*Laboratory of Molecular Biophysics, Rex Richards Building, Oxford, OX1 3QU*).
The *Nicotiana tabacum* form III crystals are grown at a pH (about 5.5) that would tend to wash off the small subunits in solution. The fourfold crystallographic symmetry of form III crystals may result from the absence of small subunits. Berlin, Oxford and Uppsala spinach enzymes

are crystallized at a pH greater than 7.0 and the crystals do not have a crystallographic fourfold symmetry. This difference could be due to the presence of heterogeneous small subunits, which form different contacts in the crystal.

D. EISENBERG. Form III crystals have been washed, redissolved and subjected to SDS–polyacrylamide gel electrophoresis. The gels display bands corresponding to both large and small subunits. We conclude that small subunits remain after crystallization.

In any case, it is not clear how heterogeneous small subunits might lower the symmetry of the crystals. To do so, one subtype would have to occupy the same sites in all Rubisco molecules. It seems more likely that Rubisco molecular assembly places each subtype of the small subunits in random positions. Then crystallization might be expected to be accompanied by a small disorder, perhaps giving rise to a blurring of the electron density at the positions of the heterogeneous side chains.