Acta Cryst. (1993). D**49**, 548-560

Structure Determination and Refinement of Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase from *Synechococcus* PCC6301*

BY JANET NEWMAN† AND CARL-IVAR BRÄNDÉN

Department of Molecular Biology, Swedish University of Agricultural Sciences, Box 590, Biomedical Centre, 751 24 Uppsala, Sweden

AND T. ALWYN JONES

Department of Molecular Biology, University of Uppsala, Box 590, Biomedical Centre, 751 24 Uppsala, Sweden

(*Received* 25 *January* 1993; *accepted* 25 *May* 1993)

Abstract

The structure of an activated quaternary complex of ribulose 1,5-bisphosphate carboxylase/oxygenase (rubisco) from *Synechococcus* PCC6301 has been solved by molecular replacement. The protein crystallizes in an orthorhombic $P2_12_12_1$ unit cell with a complete L_8S_8 complex consisting of 4608 residues (37 680 non-hydrogen atoms) in the asymmetric unit. Data were collected both on film and image plate using synchrotron radiation; there were 218 276 unique reflections in the final 2.2 Å data set. The eightfold non-crystallographic symmetry could be used both to improve map quality and to reduce the computing requirements of refinement. The coordinates were refined using strict non-crystallographic symmetry constraints. The stereochemistry of the final model is good, and the model has an R value of 20.0% for the reflections between 7 and 2.2 Å.

Introduction

Ribulose 1,5-bisphosphate carboxylase/oxygenase (E.C. 4.1.1.39) is perhaps the world's most abundant protein, and certainly the most abundant enzyme. The enzyme catalyses the addition of carbon dioxide to ribulose 1,5-bisphosphate in the first step of the photosynthetic Calvin cycle. This is the only known way that large quantities of carbon can be incorporated into the biosphere (Andrews & Lorimer, 1987). However, the addition of molecular oxygen to the same substrate is also catalyzed by the enzyme, and this then leads to a net loss of carbon from the system and a concomitant use of energy and reducing power. In order to be competent for either reaction, the enzyme must undergo activation. This involves the condensation of a molecule of CO_2 with an active-site lysine. The labile adduct is then stabilized by the binding of Mg^{2+}. There are two well characterized forms of rubisco: form II is a simple homodimer (L_2) which is found in anaerobic bacteria, and form I is a hexadecamer of two types of subunit (L_8S_8) which is found in cyanobacteria, algae and plants. An L_2 dimer is the smallest catalytically competent entity (Larimer, Lee, Mural, Soper & Hartman, 1987). The L_2 dimer can be regarded as a building block of the more complicated enzyme, which consists of a core of four L_2 dimers capped at each end by a tetramer of S subunits and can

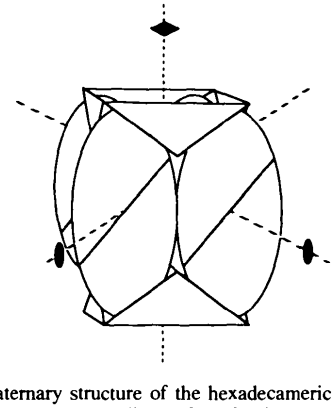


Fig. 1. The quaternary structure of the hexadecameric form I rubiscos. The egg shapes represent a dimer of L subunits, the triangles represent the small subunits. The symmetry associated with this arrangement of subunits is shown by the dotted lines.

* Abbreviations used in this paper: rubisco, ribulose 1,5-bisphosphate carboxylase/oxygenase; 2CABP, 2-carboxyarabinitol bisphosphate - a tight-binding transition-state analogue; L, large subunit of rubisco; S, small subunit of rubisco; DESY, Deutsches Elektronen Synchrotron, Hamburg, Germany; I, intensity $\alpha|F_o|^2$; R_{merge}, $\Sigma(\Sigma|I_{i,h} - \langle I_h\rangle|/\langle I_h\rangle)$, $I_{i,h}$ = individual measurement of reflection h, $\langle I_h\rangle$ = mean intensity of that reflection summed over all measured symmetry-related reflections; R value, $\Sigma||F_o| - |F_c||/|F_o|$, $|F_o|$ = observed structure-factor amplitude, $|F_c|$ = calculated structure-factor amplitude; r.m.s.d., root-mean-square deviation $[\Sigma(\langle x\rangle - x_i)^2/n]^{1/2}$, $\langle x\rangle$ = arithmetic mean of sample, x_i = sample point; n.c.s., non-crystallographic symmetry; SGI, Silicon Graphics Industries 310 VGX graphics workstation; ESV, Evans and Sutherland ESV 10 graphics workstation; w_A, the weighting factor used during refinement by the program *X-PLOR* (this number is given by the check procedure in *X-PLOR*); rsfit, real-space residue fit of the model coordinates to the electron density, as described in Jones, Zou, Cowan & Kjeldgaard (1991); $\langle B_L\rangle$, average value of the temperature factor (Å^2) for the main-chain atoms of the large subunit; $\langle B_S\rangle$, average value of the temperature factor (Å^2) for the main-chain atoms of the small subunit; CPU, central processor unit.

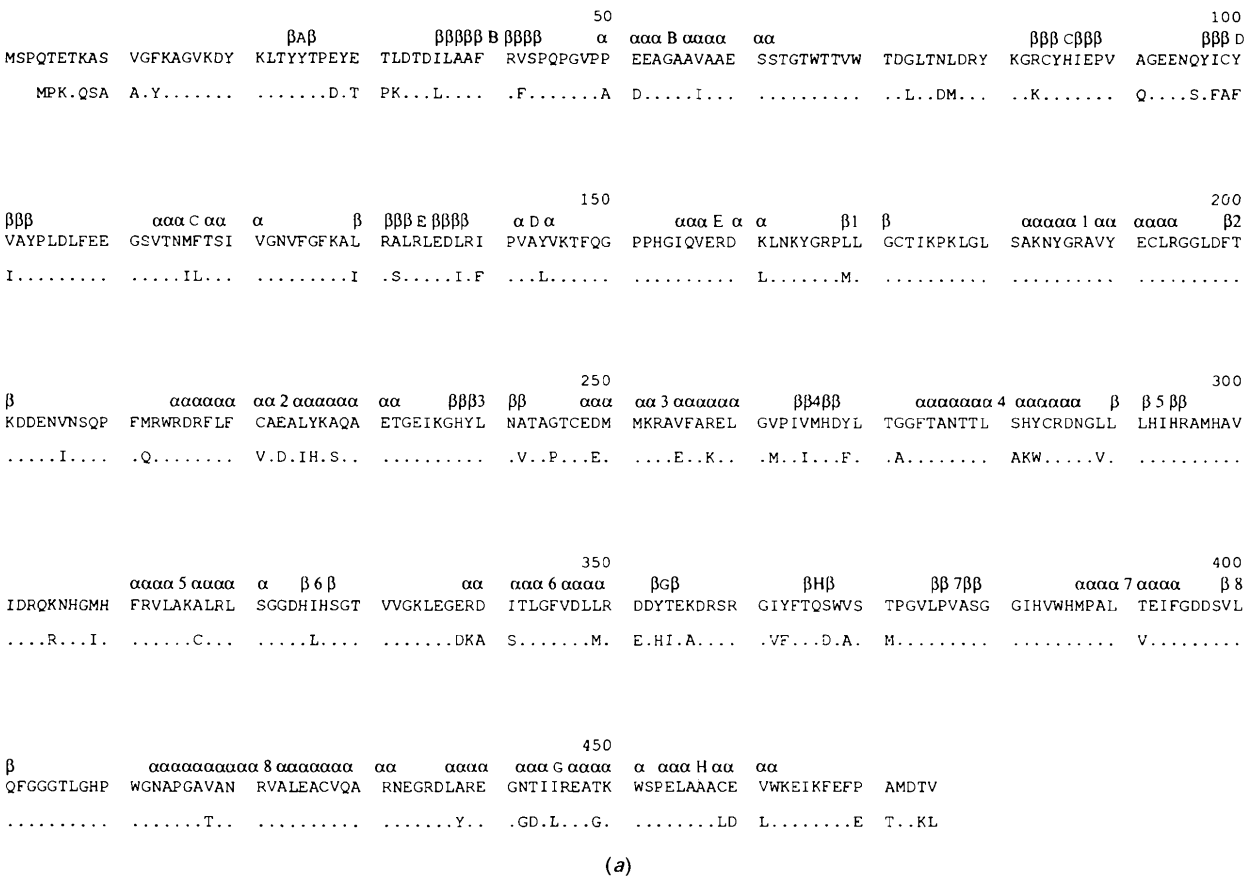
† Author to whom correspondence should be addressed.

ISSN 0907-4449

thus be written more descriptively $(L_2)_4(S_4)_2$. The L_8S_8 complex has local 422 symmetry associated with it, as a result of this subunit arrangement (Fig. 1).

The sequence homology between the *L* proteins of the form I rubiscos is very high, but there is less conservation of sequence in the *S* subunits. The *L* subunits of rubisco from *Synechococcus* and spinach share 80% identity, compared to the 40% identity in the *S* subunits. Fig. 2 shows a sequence alignment of the *L* and *S* chains from spinach and *Synechococcus*. Despite the similarities in sequence, there is significant variation in the kinetic and specificity parameters amongst the form I rubiscos (Andrews & Lorimer, 1987).

This enzyme is to be the target for protein engineering, with the ultimate aim of reducing or excluding the oxygenase activity, and thus it is crucial that there



(*a*)

(*b*)

Fig. 2. Alignment of spinach (upper row) and *Synechococcus* (lower row) sequences of the (*a*) *L* and (*b*) *S* subunits. The secondary structure assignment is from Knight *et al.* (1990).

exists a system for producing modified enzyme. Ideally, an expression system for producing and modifying rubiscos from plants should be used; unfortunately, a number of factors - including a strict requirement for a chaperonin to fold and assemble the plant enzyme - forbid this. Although there is sequence similarity between the chaperonin used to aid assembly of rubisco in higher plants and *E. coli* chaperonin 60, correctly folded and active higher plant rubisco is not produced in *E. coli* (van der Vies & Ellis, 1991). However, there are efficient *E. coli* expression systems for a higher plant-like hexadecameric enzyme from cyanobacteria. The *E. coli* chaperonin 60 protein is apparently similar enough to the cyanobacterial chaperonin that it can be used in the folding and assembly of cyanobacterial rubisco. Thus the structure of rubisco from a cyanobacterium is a necessity for the intelligent planning and interpretation of mutation studies of the hexadecameric enzyme.

Experimental

Purification and crystallization of rubisco from *Synechococcus* PCC6301 has been described elsewhere (Newman & Gutteridge, 1990). The crystals are of the activated enzyme, with the tight-binding transition-state analogue, 2CABP. The space group, although orthorhombic, was not known exactly as crystal morphology made it difficult to collect data around the $0k0$ axis - the ambiguity was limited, however, to $P2_122$ or $P2_12_12_1$. Not until the translation functions were solved could the space group be unequivocally assigned to $P2_12_12_1$.

Data collection

Data were collected on film at station 9.6 of SRS using a wavelength of 0.89 Å. Oscillation steps of 0.5° were chosen to minimize overlaps given the unit cell of 223.9 × 111.9 × 199.7 Å and the requirement to collect data to a nominal resolution of 2.2 Å. One crystal was rotated through 45° giving data which were processed using an Uppsala modification of the program *OSCGEN* (Rossmann, 1979). Each digitized film was inspected on a RAMTEK display to ensure that the predicted pattern matched that observed. Approximately 70% of a 3 Å data set was collected on the MAR image plate on beamline X31 at the EMBL outstation, DESY, using four crystals. The wavelength used for that data collection was 1.48 Å. These data were processed using the in-house modification of the *MOSFLM* film processing software (Leslie, Brick & Wonacott, 1986). Higher resolution data were collected using three more crystals at DESY, this time with a shorter wavelength of 1.01 Å. The three data sets were scaled and merged with the programs *ROTAVATA* and *AGROVATA* from the *CCP*4 program package (SERC Daresbury Laboratory, 1979). A total of 852 051 measurements from eight crystals were used to produce an 86.2% complete data set with an R_{merge} of 10.5% for the 218 276 unique reflections between 100 and 2.2 Å. Of these, 76.4% have $I > 3\sigma(I)$ (51.7% in the resolution shell 2.26-2.20 Å) (Fig. 3).

Structure determination

The model. Coordinates of spinach rubisco with an R value of 24% to 2.4 Å resolution (Brookhaven PDB code 8RUB) were kindly provided by Drs S. Knight and I. Andersson, and were used as the search model in the molecular-replacement procedure. These coordinates were of activated enzyme, with bound 2CABP. In order to facilitate comparison the numbering used for the spinach model was retained. There is an insertion of 12 residues in the spinach small subunit at residue 52; thus the *Synechococcus* numbering of the small subunit extends from 2 to 122, with a gap between residues 51 and 64, see Fig. 2.

The search model was created by first translating the L_4S_4 complex which constitutes the asymmetric unit in the unit cell of the spinach crystal form such that the centre of gravity of the x and y coordinates were zero, applying a small rotation matrix (to align the fourfold exactly with the z axis) then applying a twofold rotation around x to generate the other half of an L_8S_8 complex. The side chains which differ from the *Synechococcus* sequence were trimmed to alanine (or glycine, if either sequence has a glycine).

Rotation searches. Self-rotation functions (Rossmann & Blow, 1962) were run using *POLARRFN* in the *CCP*4 package (see Newman & Gutteridge, 1990). Using data between 10 and 6 Å, and a radius of 35 Å, a $\kappa = 90°$ search gave one clear peak at $\varphi = 175$, $\psi = 180°$. This suggested that the fourfold axis of the L_8S_8 molecule is almost parallel to the z direction. The peak, which is decidedly non-circular, also shows that this fourfold axis is not perfectly aligned. Whether or not the tilt could be

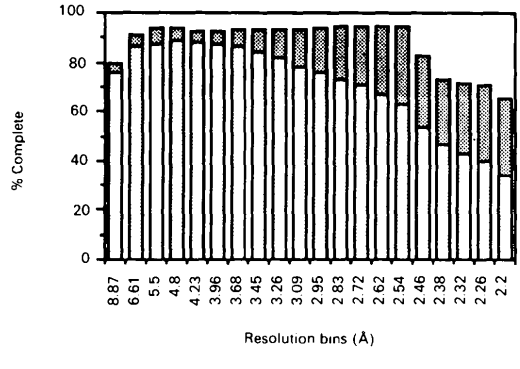


Fig. 3. Completeness of the native *Synechococcus* data as a function of resolution. The completeness is the ratio of the number of measured reflections to the number of possible unique reflections. Overall the data is 86.2% complete to 2.2 Å. The lighter shaded region shows the data in each resolution bin with $I > 3\sigma(I)$.

seen in the rotation function was very dependent on the radius/resolution range used in the search - a search run using a radius of 15 Å showed no signs of any tilting at all (Fig. 4). A κ = 180° search using the resolution limits 10-6 Å as above gave three peaks - those expected from the space group alone.

As the fourfold axes could be superimposed, only a one-dimensional cross-rotation function was needed to completely define the rotation of the *Synechococcus* enzyme with respect to the spinach enzyme. Such a search was run using the same program, and gave two very clear peaks, at κ = 45 and 135°. Given the local fourfold symmetry of the molecule these two peaks are equivalent. In an attempt to pinpoint more exactly the tilt in the *Synechococcus* enzyme, further rotation functions were run using Fitzgerald's molecular-replacement program suite *MERLOT* (Fitzgerald, 1988). Initially, the implementation of the fast-rotation function (Crowther, 1972) was used. The highest peak from a κ = 90° self-rotation function was a 2σ peak at φ = 5, ψ = 175° suggesting that the fourfold axis is tilted in two planes. Cross-rotation functions were run first in *CROSUM*, then in *LATSUM* (Lattman & Love, 1972), eventually yielding the rotations (α = -40.7, β = 173.9 and γ = 84.7°) required to map the spinach coordinates onto the *Synechococcus* enzyme.

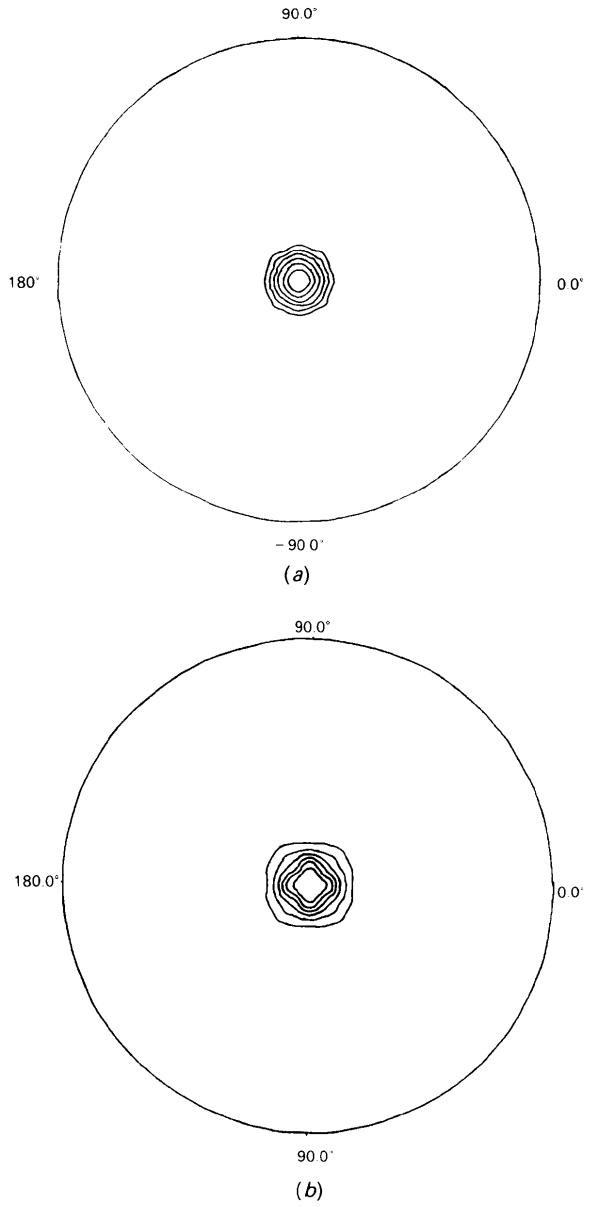


Fig. 4. Self-rotation functions (κ = 90°) with *Synechococcus* data. The sensitivity of the function to the displacement of the local fourfold axis is dependent on the radius/resolution range chosen. The peak is quite smooth in (*a*), where the radius = 15 Å (8-3 Å data) and is clearly non-uniform in (*b*), where the radius = 35 Å (8-6 Å data).

Translation searches. Possible packing arrangements of the hexadecamer within the *Synechococcus* unit cell were examined using *PACMAN* (Bergdoll, unpublished results). When the space group $P2_122_1$ was used, it could be seen that there was only one degree of freedom in a translation search; the x and z translations were fixed at $\frac{1}{4}$ by packing constraints. Similarly, in $P2_12_12_1$, the z coordinate constrained to 0, requiring a two-dimensional search only. The program package *X-PLOR* (Brünger, 1992) was used to calculate one- and two-dimensional translation searches in $P2_122_1$ and $P2_12_12_1$, respectively. As translation searches are notorious for being resolution sensitive, each translation function was repeated at various permutations of resolutions between 8 and 3 Å. Each of the trials in $P2_12_12_1$ yielded a high ($\geq 10\sigma$) peak at the same position (0.375, 0.19, 0.0), whereas the highest peaks in the $P2_122_1$ trials were only of the order of 2σ, and were not always found at the same position. Thus these searches suggested that the molecule should be positioned at 0.375, 0.19, 0.0 in a $P2_12_12_1$ cell. This was checked by running a three-dimensional translation search in *X-PLOR* in this space group (resolution range 8-5 Å). The translation search was run over the limits 0-$\frac{1}{2}$, 0-$\frac{1}{2}$, 0-1. The highest peak from this three-dimensional search had a correlation of 0.42, and confirmed these values of the translation. Indeed, the top ten peaks of the translation function were clustered around the peak at 0.375, 0.190, 0.000.

Refinement

Initial refinement. A 5 Å $2|F_o| - |F_c|$ map calculated using the spinach model coordinates in the *Synechococcus* orientation was disappointing. Although secondary structural elements could be seen, they did not match the coordinates. *CORELS* (Sussman, Holbrook, Church & Kim, 1977) was used to refine the protein as a single rigid entity using reflections in the resolution range 10-8 Å. Five cycles of refinement brought about a reduction in the R value from 46 to 37%, with a

corresponding increase in the correlation coefficient from 55 to 68%. No further improvement in the R value was obtained either by increasing the resolution of the refinement or by increasing the number of rigid groups in the refinement. A 5 Å $2|F_o| - |F_c|$ map produced from the coordinates after rigid-body refinement in *CORELS* was of vastly improved quality. This truncated spinach model was minimized in *X-PLOR*, giving an R value of 24% for the reflections between 8 and 3 Å (model 0).

In the L_2 rubisco from *Rhodospirillum rubrum*, there are significant differences between the two L subunits (Schneider, Linqvist & Lundqvist, 1990). Thus $\frac{1}{4}$ of model 0 (*i.e.* a dimer of large subunits and two associated small subunits) was initially rebuilt on the assumption that similar differences would be found in the four dimers that make up the *Synechococcus* enzyme. This rebuilding and all subsequent interactive manipulations of the coordinates were performed using the program *O* (Jones & Kjeldgaard, 1991). There was density for most of the pruned side chains in a 3.0 Å $2|F_o| - |F_c|$ map generated in *X-PLOR* from model 0 (Fig. 5), as well as density for the phosphate groups of the bound transition-state analogue, 2CABP. The complete L_8S_8 was generated by rotating the rebuilt quarter of model 0 onto the position of the other three quarters. The transformations used were generated in *O* from the initial model 0. A standard *X-PLOR* slow-cool procedure from 3273 K using reflections between 8 and 2.8 Å was run using strict fourfold n.c.s. constraints and a quarter of the L_8S_8 complex. The resulting coordinates had an R value of 23.2% (model 1). This slow-cool procedure required 17 d of CPU time (35 d real time) using one processor of an ALLIANT FX40. A $2|F_o| - |F_c|$ map to 3.0 Å was generated using the *CCP*4 package, and showed continuous density for the 2CABP, which was built in by analogy to the 2CABP in the original spinach coordinates. The projected CPU requirements of continued refinement to higher resolution in *X-PLOR* were awe inspiring, so refinement was continued in *PROLSQ* (Hendrickson & Konnert, 1981). One cycle of the fast-Fourier transform version of *PROLSQ* took 40.5 CPU h on a VAX 8800. Eventually, the 12 cycles of *PROLSQ* were run on a Cray Y-MP where the cycle time was 2.5 CPU h. However, incorrect weighting factors were used which gave a distorted model with a higher R value of 33% to 2.5 Å (model 2). To achieve more acceptable computing times, the refinement was continued (in *X-PLOR*) with strict n.c.s. constraints applied to a smaller non-crystallographic unit of one large and one small subunit.

Averaging. Averaging was performed initially to improve map quality by exploiting the eightfold non-crystallographic symmetry to carry out phase refinement (Bricogne, 1976). The real-space averaging package *A* (Jones, 1992) was used throughout the averaging process. A mask of an *LS* monomer was created from the coordinates of model 2. This mask was manually edited in *O* to produce a smooth envelope with no aberrant points inside the mask. The transformations which map the *LS* monomer onto the other seven n.c.s.-related monomers were generated in *O*. The r.m.s.d. of the overlap of the C_α's from the *LS* monomer with the C_α's from each of the other seven was approximately 0.5 Å in all cases. This suggests that the differences seen in the large subunits that make up the *R. rubrum* dimer are not as apparent in the catalytic dimers of the *Synechococcus* complex. Three cycles of phase refinement by averaging to a resolution of 2.2 Å were enough to achieve convergence. Each cycle of averaging consisted of a structure-factor calculation from a previous Fourier-inverted averaged map (program *SFALL* - CPU time on an SGI $\simeq$ 15 min); scaling of the calculated and observed structure factors (program *RSTATS* - CPU time on an SGI $\simeq$ 2 min); a Fourier transformation using $2|F_o| - |F_c|$ amplitudes and calculated phases (program *FFT* - CPU time on an SGI $\simeq$ 12 min); and averaging to produce a new map (program *A* - CPU time on an ESV $\simeq$ 40 min). Both mask and maps were sampled at 0.7 Å which corresponds to a third of the highest resolution data. The three cycles of averaging reduced the R value of the

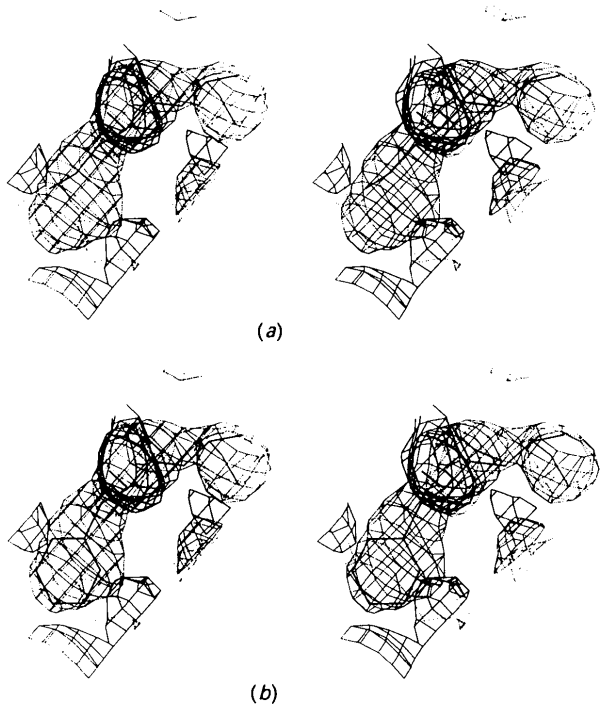


Fig. 5. Density appeared for many of the truncated side chains in a 3.0 Å $2|F_o| - |F_c|$ map calculated after rigid-body refinement in *CORELS* and energy minimization in *X-PLOR*. (*a*) The isoleucine residue found in the spinach structure at position 140 of the *L* subunit, (*b*) the phenylalanine found in *Synechococcus* rubisco. The density in both (*a*) and (*b*) is that of the cyanobacterial enzyme calculated with an alanine in that position.

inverted averaged map from 27.1 to 21.1% (Fig. 6). This high-quality averaged map was used to guide a rebuild of the model, and occasionally, missing side chains could be built into density. After rebuilding, the value of the model was 33%. Water molecules were added to the *LS* monomer at this stage where there was density and a suitable hydrogen-bonding environment. This rebuilt model (including 110 waters) was minimized using strict n.c.s. constraints and the reflections between 5 and 2.2 Å to give model 3.

Refinement of the model with strict n.c.s. constraints. According to a recent report (Weis, Brünger, Skehel & Wiley, 1990) the weight given by the check procedure in *X-PLOR* tends to overweight the contribution of the X-ray terms at the expense of stereochemistry. Nine sets of 25 cycles of minimization were run on the rebuilt model 2 where the weighting factor was reduced in steps of 10% from the 'ideal' weight given by the check procedure. After each run the final coordinates were compared to the starting coordinates and were also analysed for poor stereochemistry. The results are summarized in Table 1. A weight of 50% of the 'ideal' gave a model with more tightly restrained stereochemistry. Unless otherwise stated, the weight used in all refinement was 50% of the 'ideal' weight from the *X-PLOR* check procedure. The parameter files used in the refinement were parhcsdx.pro and tophcsdx.pro, in which the force constants have been calculated from the variation in bond lengths found in molecules deposited in the Cambridge Structural Database (Engh & Huber, 1991).

Rebuilt model 2 was refined using strict n.c.s. constraints for 100 cycles of Powell minimization, with a drop in *R* value from 33.2 to 27.2%. Refinement of the overall and subsequently individual temperature factors also using n.c.s. constraints resulted in a further decrease of the *R* factor to 22.6%. However, some waters from the monomer clashed with symmetry-related waters. Therefore, new waters were carefully checked for symmetry clashes. This eventually yielded 112 self-consistent 'monomeric' water molecules. A complete cycle of reciprocal-space refinement with this protocol required 80 CPU h on the SGI.

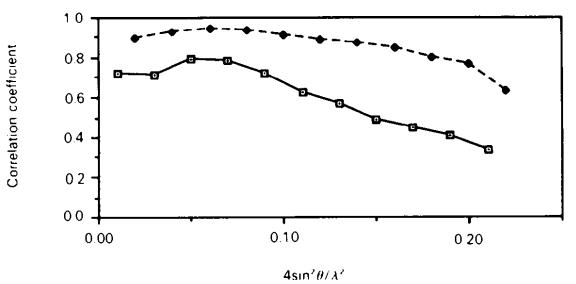


Fig. 6. Correlation coefficient *versus* $4\sin^2\theta/\lambda^2$ of the initial *Synechococcus* coordinates (solid line) and a $2|F_o| - |F_c|$ map after three cycles of real-space averaging (broken line).

Table 1. *Effect of w_A (the weight given by the check option of X-PLOR) on the stereochemistry of the LS monomer after rebuilding model 2*

The *R* value was essentially constant (32.8%) for all of the trials. The column labelled coords gives the r.m.s.d. of the C_α positions before and after minimization; dist is the r.m.s.d. from ideal bond lengths (Å); angles is the r.m.s.d. from ideal bond angles (°); dihedrals is the r.m.s.d. from ideal dihedral angles (°); and impr is the r.m.s.d. from ideal improper dihedral angles (°).

			R.m.s.d.				
Trial	w_A	%	Coords	Dist	Angles	Dihedrals	Impr
1	465980	100	0.0266	0.138	6.244	35.079	5.169
2	419382	90	0.0265	0.138	6.242	35.080	5.170
3	372784	80	0.0265	0.138	6.242	35.080	5.170
4	326186	70	0.0265	0.138	6.242	35.080	5.170
5	279588	60	0.0265	0.138	6.244	35.081	5.171
6	232990	50	0.0424	0.078	5.710	35.037	4.615
7	186392	40	0.0424	0.078	5.712	35.037	4.617
8	139794	30	0.0422	0.078	5.712	35.039	4.619
9	93196	20	0.0421	0.078	5.714	35.041	4.620

One of the weak points in the refinement was the assumption that the transformation matrices used in averaging and in refinement were correct. These transformations were refined by performing 100 cycles of minimization on the complete, unconstrained L_8S_8 molecule using the reflections between 5 and 2.2 Å. The transformation operators which mapped the C_α positions of the *LS* monomer onto the other seven n.c.s.-related monomers after this refinement were used in the subsequent constrained refinements. The model after the refinement of the n.c.s. operators, model 5, had an *R* value of 21.4%. A comparison of the C_α positions of model 3 and model 5 showed that the errors in the original n.c.s. operators were small - the r.m.s.d.'s of the *L* and *S* subunits of these two models were 0.079 and 0.13 Å, respectively. Further refinement included modification of the averaging mask, minor rebuilding, the location of further waters and more minimization in *X-PLOR*. Model 6 has an *R* factor of 20% for all reflections between 7.0 and 2.2 Å, and contains 229 waters per n.c.s. asymmetric unit.

Discussion

Comparative packing

Rubisco from *Synechococcus* crystallizes in a $P2_12_12_1$ cell with dimensions 223.9 × 111.9 × 199.7 Å. The spinach model crystallizes in $C222_1$ with dimensions 157.2 × 157.2 × 201.3 Å. The two cells have approximately the same dimension in the *z* direction, and the dimension of the $P2_12_12_1$ cell is the diagonal of the $C222_1$ cell; thus the cells have the same volume (Fig. 7). In both cases there are four complete molecules in the unit cell, so clearly the density of the crystals must be the same: V_m = 2.32 Å Da^{-3} (Matthews, 1968). The spinach L_8S_8 sits at 0, $\frac{1}{4}$, $\frac{1}{4}$, with the non-crystallographic fourfold axis almost parallel to the *z* direction. The fourfold axis is inclined 1.8° in the *xz* plane. The *Synechococcus* rubisco hexadecamer sits at 0.38, 0.19, 0 again with the local

fourfold axis approximately parallel to z. Here the tilting in the fourfold axis is greater (6.3°) and is perpendicular to the tilt seen in the spinach model. The spinach fourfold is close enough to the crystallographic twofold along z that the low-resolution reflections display pseudo tetragonality (Knight, Andersson & Brändén, 1990). No indication of any tetragonality is seen at any resolution in the *Synechococcus* data, doubtless the size of the tilt in the local fourfold precludes this. The axis is only very slightly (0.3°) inclined in the xz plane. Also, the local twofold axes show a 49.5° deviation from the crystallographic screw axes.

The question arises of why the cyanobacterial enzyme crystallizes in the lower symmetry space group. In spinach, the major contacts between the molecules in different layers are from residues 24-28 in the *S* subunit, which see the same residues in a crystallographically related L_8S_8 molecule. There is an unusual contact between symmetry-related glutamic acid residues at position 24. The *Synechococcus* sequence shows little conservation of the region 24-28 in the *S* subunit (Fig. 2) and in particular, Glu24 has been replaced with an arginine. Thus it appears that the tilting seen in the packing of the cyanobacterial enzyme is the result of minimizing the contacts between two symmetry-related arginines. The contacts between symmetry-related L_8S_8 molecules are primarily mediated through this arginine, which makes extensive contacts to residues 21-24 of another small subunit. To accommodate these interactions, the molecule has shifted from the singular position it occupies in the spinach unit cell to the more general position seen in the *Synechococcus* unit cell.

Structure

The overall fold of the cyanobacterial enzyme is identical to that of the spinach enzyme, indeed the r.m.s.d. of the C_α positions is 0.33 Å for the 467 residues of the large subunits. The small subunits show greater deviation in the position of the main-chain atoms (r.m.s.d. C_α = 2 Å for 110 residues), but this is predominately caused by movement of the chain around the 12 amino-acid deletion. In brief, the *L* subunit of rubisco is a two-domain protein: the N-terminal domain is a 150-residue mixed α/β-domain and the remaining 300 or so residues make up a C-terminal β/α-barrel. The *S* subunit has a single domain consisting of a central four-stranded β-sheet covered on one side by two helices (Fig. 8). The extra 12 amino acids found in the spinach small

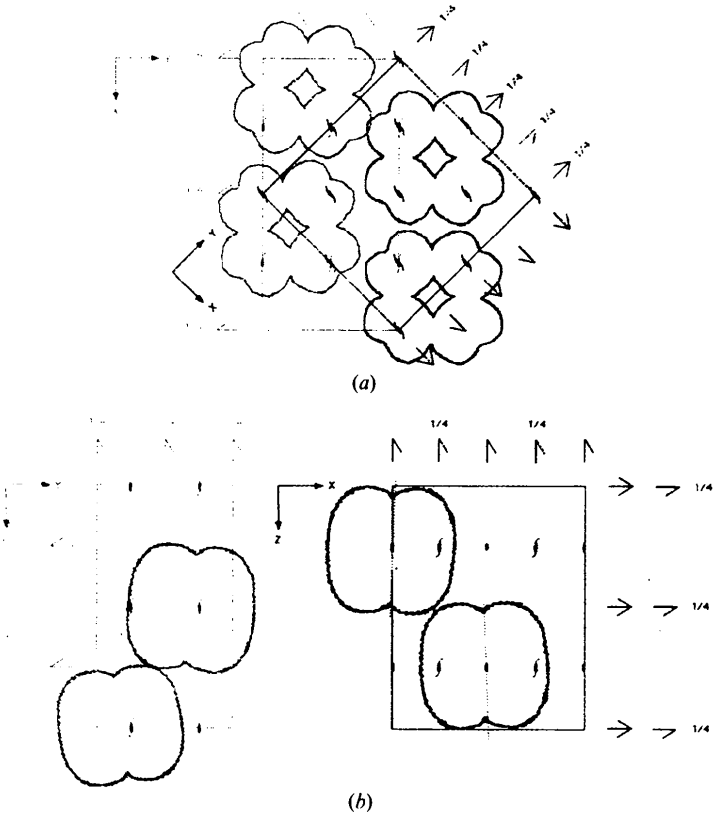
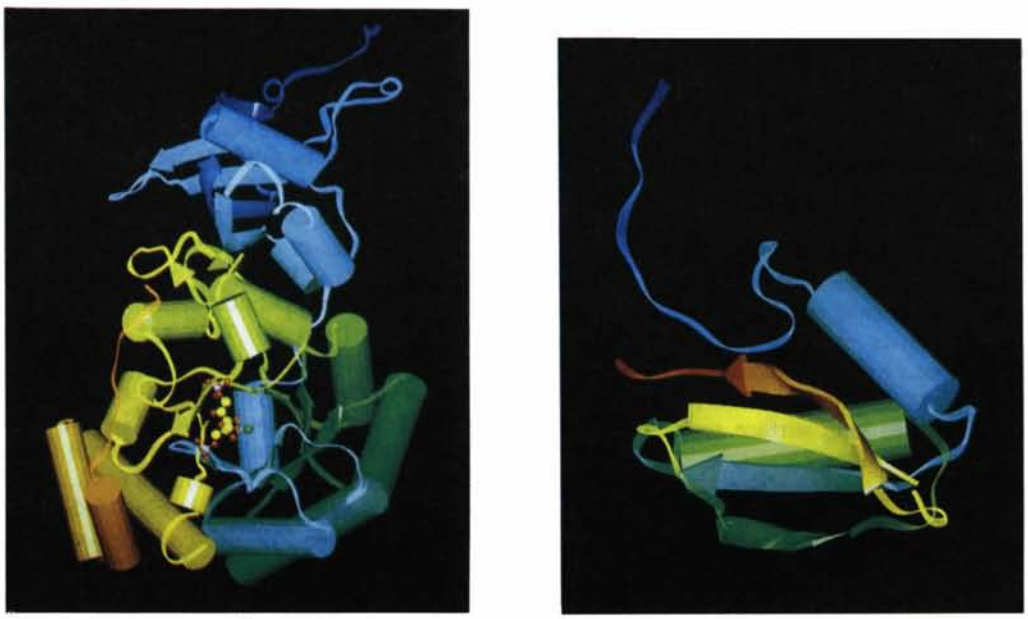


Fig. 7. (*a*) The view down the z direction of the *Synechococcus* (red) and spinach (black) unit cells. The *Synechococcus* L_8S_8 is centred at 0.38, 0.19, 0 in a cell with dimensions a = 224, b = 112 and c = 200 Å. The spinach L_8S_8 is centred at 0, 0.25, 0.25 in a $C222_1$ cell with dimensions a = b = 158 and c = 200 Å. (*b*) View perpendicular to (*a*). The dotted lines show the local symmetry axes.

subunit extend the loop between strands *A* and *B*. This loop protrudes into the central cavity of the *L* core, and reduces the diameter of the channel from 19 to 17 Å. The active site is made up of residues of the large subunit only, the small subunits, situated at the ends of the L_8 core, are too distant to contribute. Residues of both domains are found in the active site. The barrel of one *L* subunit and the N-terminal domain of the other *L* subunit



(*a*) (*b*)

Fig. 8. (*a*) The large subunit of *Synechococcus* rubisco. This subunit has two domains; the N-terminal domain consists of a five-stranded mixed β-sheet with two helices covering one side of the sheet. A third small helix is makes up part of the connection between the domains The C-terminal domain is an eight-stranded β/α-barrel with an extension containing two short β-strands between $\alpha 6$ and $\beta 7$. There are a further two helices at the C terminus. The molecule is coloured so that the N terminus is blue and the C terminus red. The position of the active site is shown by the sugar molecule. (*b*) The secondary structure of the *S* subunit. This is a four-stranded antiparallel β-sheet covered on one side by two helices. This molecule is also coloured blue to red from the N terminus to the C terminus.

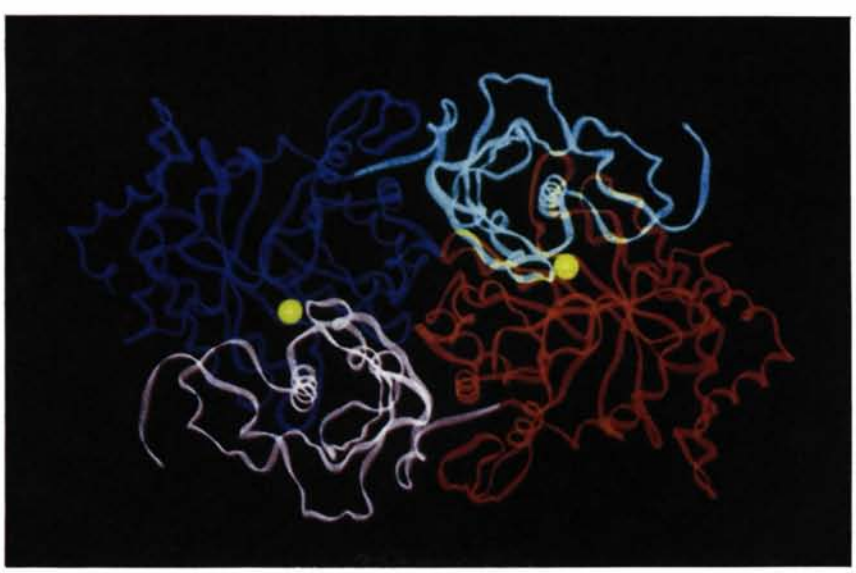


Fig. 9. A ribbon diagram of an L_2 dimer. The barrel of one *L* subunit is coloured blue, and the N-terminal domain of the same subunit, cyan. The barrel domain of the second *L* subunit is red, and its N-terminal domain is magenta. The yellow spheres are Mg^{2+} ions, and show the positions of the active sites. Both this figure and the previous figure were generated in *O*.

in the dimer make up one active site, and the barrel of the second *L* subunit combines with the N-terminal region of the first to create the other active site (Fig. 9).

Quality of the model

An upper limit for the estimate of the error in the coordinates can be determined by plotting the *R* value from the coordinates *versus* the inverse of the Bragg spacing (Luzzati, 1952). A Luzzati plot of the coordinates is shown in Fig. 10; the error in the coordinates can be estimated from this plot to be close to 0.20 Å.

The quality of the electron-density maps produced from later models of the coordinates was very high. A typical aromatic ring is shown in Fig. 11(*a*). A $2|F_o| - |F_c|$ map calculated using reflections between 10 and 2.2 Å showed continuous density for all main-chain atoms, excepting the first three residues of the large subunit, the last two residues of the small subunit and the region around residue 75 of the small subunit (Fig. 11*b*). With the exception of some long surface residues, the side chains are also well defined.

The real-space fit of the coordinates to the averaged density was used to gauge the quality of the model at each step of the refinement (Jones, Zou *et al.*, 1991). A plot of the fit of each residue to the density, Fig. 12(*a*), indicated the regions of the model which needed attention. Modification of the mask used in the averaging procedure coupled with minor rebuilding greatly improved the rsfit of the coordinates to the density (Fig. 12*b*). Some of the troughs in the rsfit plot were the result of having the wrong sequence at that point. Four such errors were discovered in this way: three were introduced while building the initial *Synechococcus* model and one is a putative sequence error. This involved residue 415 of the *L* subunit, which the primary sequence showed as a leucine, but the density suggests is a proline. The spinach sequence has a proline in this position, as do most of the form I enzymes.

Plots of $\langle B_L \rangle$ and $\langle B_S \rangle$ *versus* residue number were generated using a local set of data-visualization programs (Kleywegt, unpublished results) and compared to secondary structure (Fig. 13). The average temperature factor for the main-chain of the *S* subunit was higher than that of the *L* subunit; this agrees with the quaternary structure, which shows that the small subunits are situated on the extremities of the *L* core. The peaks in the plots of $\langle B_L \rangle$ and $\langle B_S \rangle$ occur in loop regions, with the exception of valine 113 of the small subunit. This residue is situated in the middle of the last β-strand in the small subunit. Inspection of the density suggested that the residue could have been mis-sequenced since rotomer 2 of serine fits beautifully into the density (Fig. 14). There is a precedent for this substitution, the rubiscos from two algae have serine in position 113 of the small subunit (Knight *et al.*, 1990).

The pepflip option in *O* was used to calculate the r.m.s. fit of each peptide carbonyl. In this procedure, the residues from $x - 2$ to $x + 2$ for each residue x are compared to a cluster of pentapeptides of similar structure from a database of well refined structures. The agreement between the carbonyl O atom of residue x and the equivalent O atom of the cluster is then the pepflip value for that residue (Jones, Zou *et al.*, 1991). A plot of the pepflip values is shown in Fig. 15. Each residue with a pepflip value of ≥ 2.5 was examined more closely - the 15 residues with a pepflip value of ≥ 2.5 are summarized in Table 2. In more detail, the proline at position *L*46 is flanked on one side by a glycine (position $x + 1$) and the subset of pentapeptides not containing a glycine in position $x + 1$ have the carbonyl O atom of the x position pointing in the opposite direction, and this gives the high pepflip value seen for proline *L*46. The same reasoning

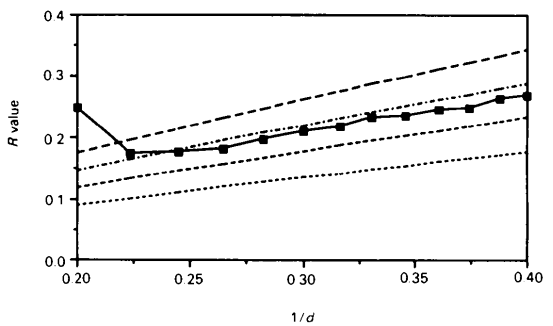


Fig. 10. A Luzzati plot giving an estimation of the error in the coordinates. This plot suggests an error of between 0.2 and 0.25 Å in the coordinates.

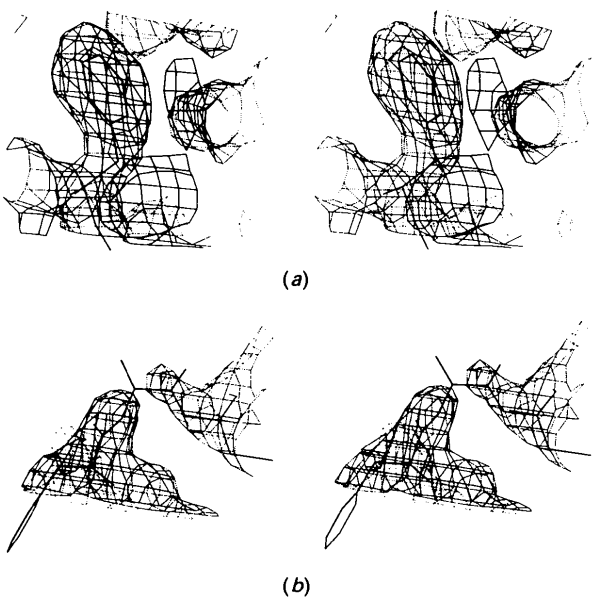


Fig. 11. (*a*) Typical density for an aromatic side chain (Phe135 of the *L* subunit). The region of weakest density (*b*) is the loop of the small subunit (residues 73–75).

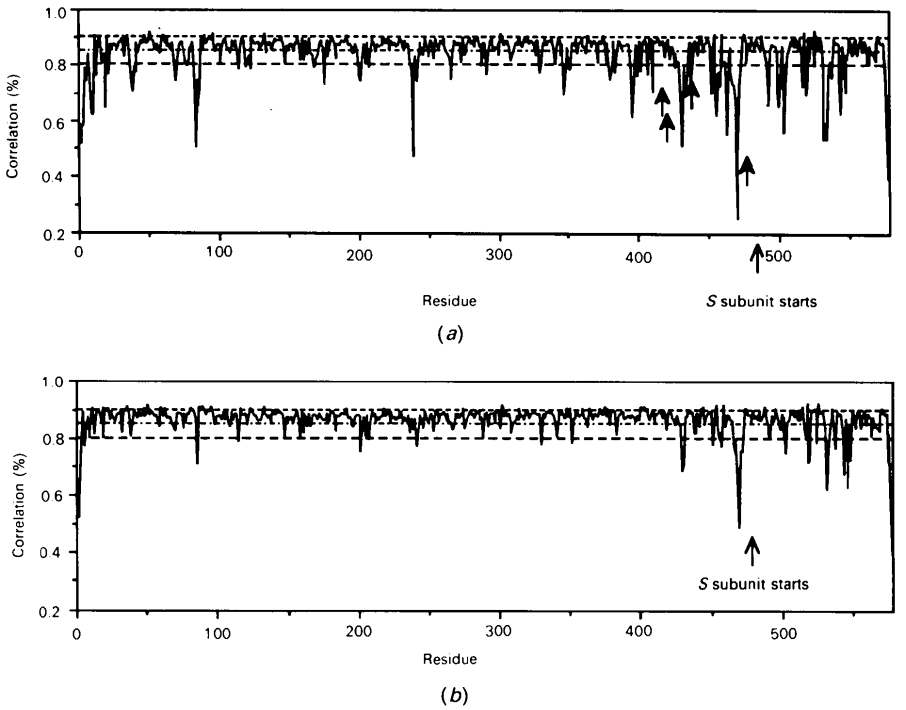


Fig. 12. Real-space fit of the *LS* monomer to averaged density before (*a*) and after (*b*) mask modification. Open-headed arrows show the positions of incorrect side chains in the sequence. The dotted lines are at 80, 85 and 90% levels. The average correlation is 84% for (*a*) and 86% for (*b*).

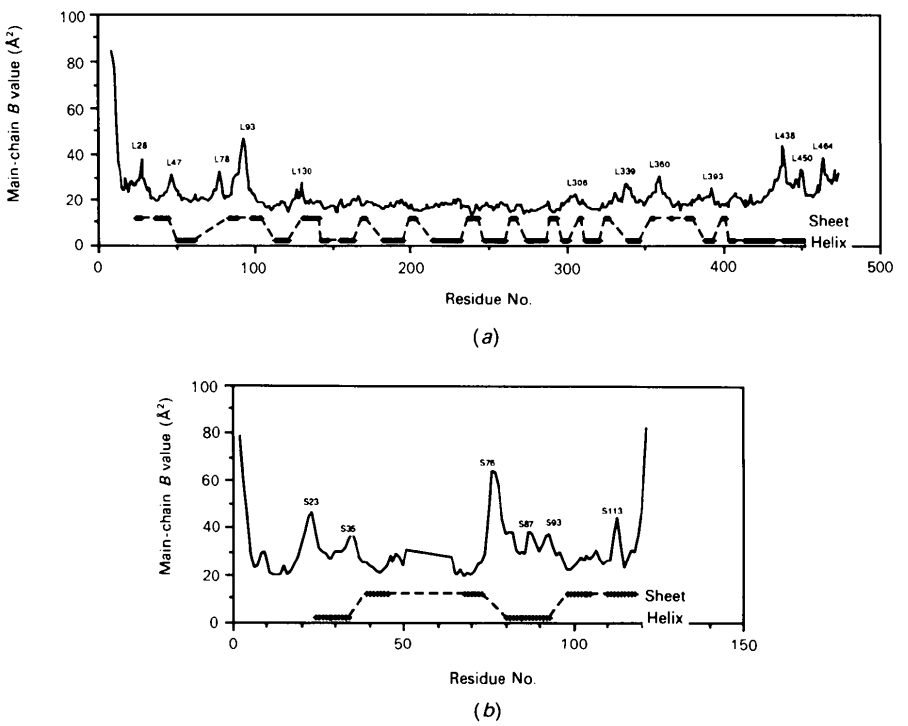


Fig. 13. Averaged main-chain temperature factors *versus* residue number. The large subunit *L* is shown in (*a*) and the small subunit *S* in (*b*). The lower curve indicates secondary structure.

can be applied to the high pepflip value for leucine *L*178. Only four of the 20 pentapeptides extracted from the database had a glycine in position $x + 1$ and the only two peptides which had a carbonyl pointing in the same direction as *L*178 both had a glycine in the critical $x + 1$ position. The carbonyl O atom of glycine *L*122 adopts a non-standard conformation so that it can hydrogen bond to the main-chain N atom of residue 297 of another large subunit. Also, the carbonyl O atom of *L*296 embraces an unusual conformation to allow the main-chain N atom of *L*297 to make contact with the main-chain O atom of residue 122 in this second large subunit. The carbonyl O atom of *L*149 is involved in a solvent-mediated hydrogen bond to the side chain of arginine *L*285. Similarly, the O atom of *L*207 is also involved in the stabilization of the buried arginine *L*217. The carbonyl O atom of histidine *L*153 points away from the average position so that the two flanking main-chain N atoms can stabilize the buried side chain of aspartic acid 324. The carbonyl O atom of tyrosine *L*368 is involved in two hydrogen bonds to the main-chain N atoms of residues *L*370 and *L*371. The main-chain N atom of residue *L*328 is part of the cluster of charged atoms which make up the binding site for the second phosphate group of 2CABP. The carbonyl O atom of this residue assumes a non-standard orientation so that the N atom can interact, through a water, with one of the phosphate O atoms. A comparable interaction is seen between the main-chain N atom of *L*380 and an O atom of the first phosphate group of 2CABP. Here also, the carbonyl O atom is atypical so as to allow this interaction. The carbonyl O atom of the interesting threonine residue *L*330 is the first residue of loop 6 of the α/β-barrel. It is believed that this loop is only locked into position after the sugar substrate has bound; in the structure of rubisco from *R. rubrum*, which has no bound inhibitor, this loop is disordered. Thus this residue is part of a hinge region, and the unusual conformation of the carbonyl O atom may reflect this.

Table 2. *Summary of the residues with a pepflip value of greater than* 2.5

Residue No.	Pepflip value	Description
*L*11	3.045	N-terminus – no density at all
*L*46	2.710	Proline in a tight turn
*L*122	3.190	Glycine at the end of helix *C*
*L*149	3.319	Glutamine in a turn
*L*153	3.027	Histidine at the start of the helix *E*
*L*178	3.275	Leucine at the dimer interface
*L*207	2.890	Aspartate at the dimer interface
*L*296	2.565	Alanine in a loop at the dimer interface
*L*328	2.941	Part of the P2 of CABP binding site
*L*330	3.279	Part of the P2 of CABP binding site
*L*368	2.699	Tyrosine which stabilizes buried Arg L319?
*L*380	2.635	Both flanking residues bind CABP
*S*12	2.677	—
*S*13	2.616	This loop is part of the small/large interface
*S*14	2.966	—

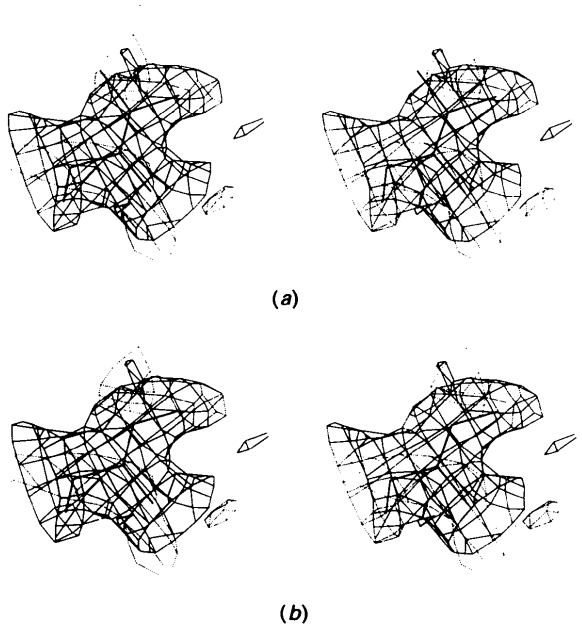


Fig. 14. Density of *S*113 showing (*a*) the original valine and (*b*) the suggested serine. This putative sequencing error was discovered by examination of the plots shown in Fig. 13.

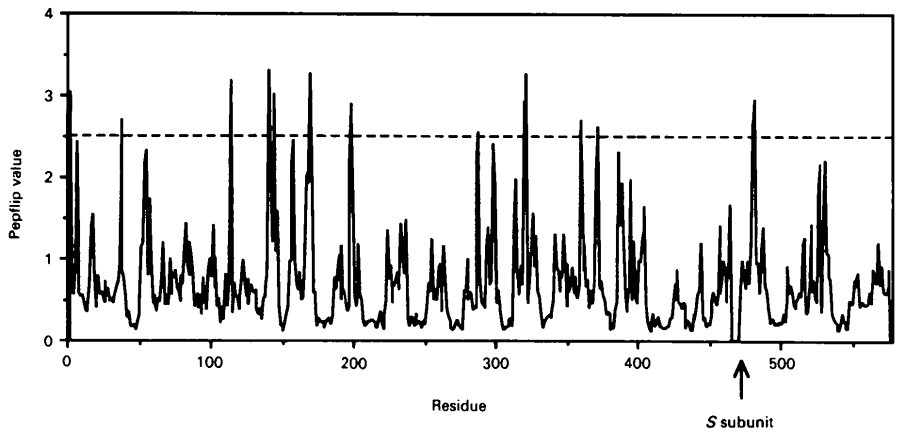


Fig. 15. Pepflip values *versus* residue number. The dashed line shows the 2.5 threshold suggested by Jones *et al.* Table 3 summarizes the residues which have a pepflip ≥ 2.5.

A scatter plot of the φ and ψ angles of every residue (Ramachandran & Sasisekharan, 1968) showed nine non-glycine residues which lie outside the allowed regions (Fig. 16). Of these, eight are also found as outliers in a Ramachandran plot of the spinach enzyme. Of these eight, seven are involved in subunit-subunit interactions, and only one is found on a loop near the active site (*cf.* Herzberg & Moult, 1991). Table 3 summarizes these contacts. The outlier which is not conserved is valine *L*331, which is found at the beginning of the flexible loop 6. Studies by Chen & Spreitzer (1989) show this residue to be important for the specificity of the enzyme, and in particular the contacts that this residue makes with aspartic acid *L*338, which is on the other end of this loop, are interesting. Also noteworthy is that proline *L*46 is found as an outlier on the Ramachandran plot of the spinach enzyme; this perhaps reflects the high pepflip value seen for the corresponding residue in the *Synechococcus* enzyme.

Table 3. *Summary of the residues that lie outside the allowed regions of Ramachandran plot*

	φ	ψ	Description
*S*3	60	140	Small/small subunit contacts
*S*13	60	−145	—
*S*15	75	−1	All on small/large subunit interfaces
*S*71	50	−125	
*L*62	−135	−77	On a loop near the active site
*L*207	123	−86	Forms part of the dimer interface
*L*297	94	−9	Forms part of the dimer interface
*L*331	71	−97	Starts loop 6 – important for specificity
*L*370	71	18	Makes a contact to a neighbouring dimer

Table 4. *Deviations from ideal stereochemistry of the LS monomer*

R.m.s. bond lengths (Å)	0.008	No. of deviations > 0.06 Å	6
R.m.s. bond angles (°)	1.259	No. of deviations > 10°	5
R.m.s. dihedral angles (°)	26.121	No. of deviations > 90°	5
R.m.s. improper dihedral angles (°)	2.446	No. of deviations > 20°	2*

* Both residues which show > 20° deviation in the improper dihedral angles are arginine residues (*L*215, *L*285).

Concluding remarks

The ability of rubisco to consume ribulose bisphosphate in two competing reactions, one desirable and the other wasteful has been the stimulus for many biochemical and structural studies of the enzyme. The partitioning of the substrate between the two reactions is dependent on the source of the protein: bacterial rubisco is faster, but less choosy than its higher plant relatives. The structures of rubisco available have not been ideal for detailed comparison. In the structure of the form II rubisco from *R. rubrum* the critical loop 6 is held open by crystallographic contacts, so that although the enzyme from this source shows large differences in kinetics and partitioning from the form II rubiscos, it was not possible to make a direct comparison with the structures of the quaternary complexes available from higher plants. Determining the structure of the quaternary complex of a cyanobacterial enzyme will allow the direct comparison of a less effective enzyme to those of higher plants.

Synechococcus PCC6301 has been refined to an *R* value of 20% with very tight stereochemistry (Table 4). The refinement proceeded quickly, despite the poor initial model. The strict non-crystallographic symmetry was essential, in that it reduced the CPU requirement for refinement from the nearly impossible to the merely excessive. The agreement between the overlapped C_α positions of the monomer *LS* and all other non-crystallographically related chains after unconstrained refinement was 0.06 Å, suggesting that the enzyme adheres quite well to the non-crystallographic symmetry; this is also seen in the very reasonable value for the final model.

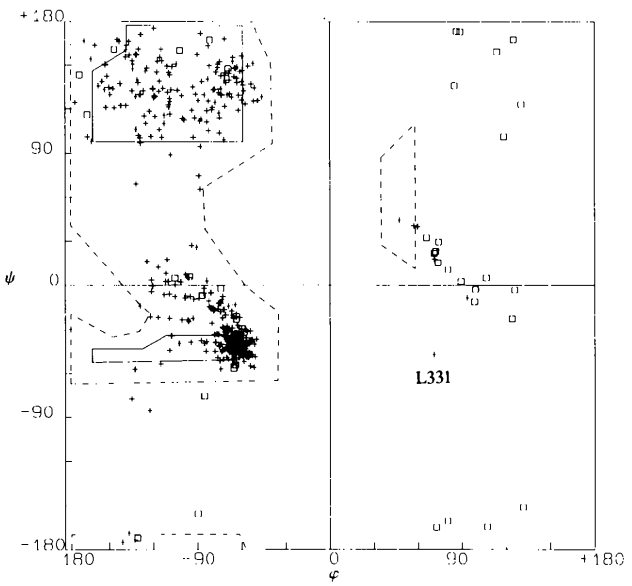


Fig. 16. Ramachandran plot of the *LS* monomer. The outliers are conserved in the spinach structure (data not shown), except for Val *L*331 (marked).

We would like to thank Dr Steve Gutteridge for providing both the enzyme and endless enthusiasm. Thanks also to the staff at the EMBL outstation, DESY, the Uppsala rubsico group and the Dupont experimental station computing service for access and help with the Cray Y-MP. This project was supported in part by Dupont and the Swedish Natural Science Research Council. The coordinates have been deposited with the Protein Data Bank.*

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory. Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Supplementary Publication No. SUP 37089). A list of deposited data is given at the end of this issue.

References

ANDREWS, J. & LORIMER, G. (1987). *The Biochemistry of Plants*. Vol. 10, edited by M. D. HATCH, pp. 131-218. Orlando: Academic Press.

BRICOGNE, G. (1976). *Acta Cryst.* **A32**, 832-847.

BRÜNGER, T. A. (1992). *X-PLOR Manual*. Version 3.0. Yale Univ., New Haven, USA.

CHEN, Z. & SPREITZER, R. J. (1989). *J. Biol. Chem.* **193**, 839-844.

CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. A. ROSSMANN, pp. 173-178. New York: Gordon & Breach.

ENGH, R. A. & HUBER, R. (1991). *Acta Cryst.* **A47**, 392-400.

FITZGERALD, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273-278.

HENDRICKSON, W. A. & KONNERT, J. H. (1981). *Biomolecular Structure: Conformation, Function and Evolution*, Vol. 1, edited by R. SRINIVASAN, pp. 43-57. Oxford: Pergamon Press.

HERZBERG, O. & MOULT, J. (1991). *Proteins Struct. Funct. Genet.* **11**, 223-229.

JONES, T. A. (1992). In *Proceedings of the CCP4 Study Weekend on Molecular Replacement*, edited by E. J. DODSON. SERC Daresbury Laboratory, Warrington, England.

JONES, T. A & KJELDGAARD, M. (1991). *O Manual*. Univ. of Uppsala, Sweden.

JONES, T. A., ZOU, J. Y., COWAN, S. W. & KJELDGAARD, M. (1991). *Acta Cryst.* **A47**, 110-119.

KNIGHT, S., ANDERSSON, I. & BRÄNDÉN, C. I. (1990). *J. Mol Biol.* **215**, 113-160.

LARIMER, F. W., LEE, E. H., MURAL, R. J., SOPER, T. S. & HARTMAN, F. C. (1987). *Biol Chem.* **262**, 15327-15329.

LATTMAN, E. E. & LOVE, W. E. (1972). *Acta Cryst.* B**26**, 1065-1068.

LESLIE, A. G. W., BRICK, P. & WONACOTT, A. J. (1986). *CCP4 Newsl.* **18**, 33-39.

LUZZATI, V. (1952). *Acta Cryst.* **5**, 802-810.

MATTHEWS, B. W. (1968). *J. Biol Chem.* **33**, 491-497.

NEWMAN, J. & GUTTERIDGE, S. (1990). *J. Biol Chem.* **265**, 15154-15159.

RAMACHANDRAN, G. N. & SASISEKHARAN, V. (1968). *Adv. Protein Chem.* **28**, 283-437.

ROSSMANN, M. G. (1979). *J. Appl. Cryst.* **12**, 225-238.

ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24-31.

SCHNEIDER, G., LINDQVIST, Y. & LUNDQVIST, T. (1990). *J. Mol. Biol.* **211**, 989-1008.

SERC Daresbury Laboratory (1979). *The SERC Collaborative Computing Project in Crystallography*. SERC Daresbury Laboratory, Warrington, England.

SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S. H. (1977). *Acta Cryst.* **A33**, 800-804.

VIES, S. M. VAN DER & ELLIS, R. J. (1991). *Annu. Rev. Biochem.* **60**, 321-347.

WEIS, W. I., BRÜNGER, T. A., SKEHEL, J. J. & WILEY, D. C. (1990). *J. Mol. Biol.* **212**, 737-761.