Article No. jmbi.1999.2701 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 288, 609–621

JMB

The Crystal Structure of Rubisco from *Alcaligenes eutrophus* Reveals a Novel Central Eight-stranded β-Barrel Formed by β-Strands from Four Subunits

Sissel Hansen[1], Valentina Burkow Vollan[1], Edward Hough[1]* and Kjell Andersen[2]

[1]*Protein Crystallography Group, Department of Chemistry, Faculty of Science University of Tromsø N-9037 Tromsø, Norway*

[2]*Laboratory of Microbial Gene Technology, IBF The Agricultural University of Norway, N-1432 Ås, Norway*

**Corresponding author*

Ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) is involved in photosynthesis where it catalyzes the initial step in the fixation of carbon dioxide. The enzyme also catalyzes a competing oxygenation reaction leading to loss of fixed carbon dioxide, thus reducing the net efficiency of photosynthesis significantly. Rubisco has therefore been studied extensively, and a challenging goal is the engineering of a more photosynthetically efficient enzyme. Hexadecameric rubiscos fall in two distinct groups, ''green-like'' and ''red-like''. The ability to discriminate between CO_2 and O_2 as substrates varies significantly, and some algae have red-like rubisco with even higher specificity for CO_2 than the plant enzyme. The structure of unactivated rubisco from *Alcaligenes eutrophus* has been determined to 2.7 Å resolution by molecular replacement and refined to R and R_{free} values of 26.6 and 32.2 %, respectively. The overall fold of the protein is very similar to the rubisco structures solved previously for green-like hexadecameric enzymes, except for the extended C-terminal domains of the small subunits which together form an eight-stranded β-barrel which sits as a plug in the entrance to the central solvent channel in the molecule. The present structure is the first which has been solved for a red-like rubisco and is likely to represent a fold which is common for this group. The small subunits in general are believed to have a stabilizing effect, and the new quaternary structure in the oligomer of the present structure is likely to contribute even more to this stabilization of the assembled rubisco protein.

Keywords: Rubisco; photosynthesis; β-barrel; crystal structure

Introduction

Ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) is a bifunctional enzyme which is involved both in photosynthesis and photorespiration in plants. In photosynthesis, the enzyme catalyzes the initial step in the fixation of carbon dioxide. In this reaction, which is the initial step in the Calvin cycle, one molecule of CO_2 is added to a five-carbon sugar, ribulose-1,5-biphosphate (RuBP), to give two molecules of 3-phosphoglycerate. During photorespiration rubisco catalyzes a competing oxygenation reaction where the enzyme utilizes O_2 instead of CO_2 as a substrate. This results in the formation of one molecule of 3-phosphoglycerate and one molecule of 2-phosphoglycolate. Part of the carbon skeleton of the latter product is rescued in the subsequent glycolate metabolism and redirected into the Calvin cycle, whereas up to 50 % is lost into the atmosphere as CO_2, thus significantly reducing the net efficiency of photosynthesis.

Due to its important, but inefficient, role in photosynthesis and thus at the primary level of food production, rubisco has been the subject of extensive studies with a major goal of engineering rubisco mutants with an increased carboxylase/oxygenase ratio compared to the wild-type enzyme (Brändén *et al.*, 1991; Wildner *et al.*, 1996; Hartman & Harpel, 1994). Recently, the enzyme's relevance

Abbreviations used: rubisco, ribulose-1,5-biphosphate carboxylase/oxygenase; RuBP, ribulose-1,5-biphosphate; DTT, dithiothreitol.

E-mail address of the corresponding author: Edward.Hough@chem.uit.no

0022-2836/99/190609–13 $30.00/0

to environmental issues has become more obvious since it is the only quantitatively significant enzyme responsible for assimilation of atmospheric CO_2. Rubisco, which probably is the most abundant enzyme in the world, thus plays an important role in the modulation of CO_2 levels in the atmosphere.

In most bacteria and higher plants the enzyme is built up from eight large and eight small subunits (form I) with molecular mass of ~55 kDa and ~15 kDa, respectively, forming an L_8S_8 complex (~550 kDa). In a few other organisms, including the photosynthetic bacterium *Rhodospirillum rubrum*, the enzyme consists of only two large subunits (form II) with less than 30 % sequence identity with form I large subunits. In addition to higher plants and photosynthetic bacteria, rubisco is found in algae and in chemosynthetic (chemoautotrophic) bacteria, such as *Alcaligenes eutrophus* (synonym: *Ralstonia eutropha*), a soil bacterium which can grown in the dark with CO_2 as its carbon source while obtaining its energy from oxidation of H_2.

Phylogenetic analysis of the amino acid sequences for rubisco large subunits from form I enzymes have shown that they fall into two distinct groups, with proteo bacteria, cyanobacteria, green algae and land plants in one group (''green-like''). The other group (''red-like'') consists of eukaryotic non-green algae and some bacteria, including *A. eutrophus* (Delwiche & Palmer, 1996; Watson & Tabita, 1997). Similar analysis of the small subunit gives the same distribution. It is interesting to note that all the sequences for the small subunits belonging to the latter class, i.e. eukaryotic non-green algae and some proteo bacteria, have a longer C-terminal region (up to 38 residues) compared to the sequences of the first class (see Figure 2). Another important difference is that green-like small subunits from eukaryotes have an insertion of variable size (12 amino acid residues for plants) in the central region of the sequence, which is not present in any of the red-like rubiscos sequenced so far. Sequence identity for the large subunit within each class is above about 70 %, while the sequence identity between members of the two classes is only about 53-60 % (Delwiche & Palmer, 1996). The small subunit exhibits the same pattern, but with lower overall sequence identities.

Even though the amino acids in the active sites of rubiscos from different species are conserved, the catalytic properties of the enzymes vary significantly. In particular, substantial differences exist in the substrate specificity factor, a measure of the enzyme's ability to discriminate between CO_2 and O_2. Low substrate specificity factors are characteristic of the form II rubisco, while the form I enzymes show great variation, from low for cyanobacteria to high for terrestrial plants. Some marine algae have red-like rubiscos with even higher specificity factors than the plant enzyme (Jordan & Ogren, 1981; Read & Tabita, 1994). Rubisco from *A. eutrophus* has an intermediate specificity factor (Lee *et al.*, 1991). Previously, three-dimensional structures have been reported for rubisco from spinach (Andersson *et al.*, 1989), tobacco (Chapman *et al.*, 1988), the cyanobacterium *Synechococcus* PCC6301 (Newman & Gutteridge, 1993) and the photosynthetic bacterium *R. rubrum* (Schneider *et al.*, 1990). In addition, preliminary X-ray crystallographic studies have been reported for rubisco from a purple sulfur photosynthetic bacterium, *Chromatium vinosum* (Nakagawa *et al.*, 1986), and the red alga, *Galdieria partita* (Shibata *et al.*, 1996). An effort was also made several years ago to solve the structure of rubisco from *A. eutrophus* (Holzenburg *et al.*, 1987). However, based on subsequent photon correlation spectroscopy studies combined with sedimentation analysis, the same authors later concluded that the resulting model was wrong (Choe *et al.*, 1989). Furthermore, the rubisco preparation used probably consisted of a mixture of two wild-type enzymes, as it was purified from a strain of *A. eutrophus* that has two sets of functional rubisco *rbcL* and *rbcS* genes (Kusian *et al.*, 1995).

Several structures of activated and non-activated rubisco from different sources complexed with the natural substrate, substrate analogues, reaction intermediate analogue and product have led to a detailed understanding of both the activation mechanism, a reversible process which includes the carbamylation of a lysine side-chain in the active site and subsequent binding of a Mg^{2+}, and the carboxylation reaction of the enzyme. The oxygenation reaction is, on the other hand, not understood in detail. The exact role of the small subunits is still much of a mystery. They are not essential for catalytic activity, but their presence increases the activity of the enzyme considerably, probably by stabilizing the hexadecameric form and by influencing loop regions close to the active site. The structure and degree of order of some of the loops, especially loop 6 depend on the state of the enzyme. In the activated, unliganded state (Taylor & Andersson, 1996) loop 6 is disordered. When the substrate, RuBP, is bound to the activated form of the enzyme (Taylor & Andersson, 1997a), loop 6 is partly ordered and resembles that found in non-activated, unliganded rubisco from tobacco (Curmi *et al.*, 1992). As the reaction proceeds, the loop is believed to close over the active site as found for the activated enzyme complexed with the reaction intermediate analogue 2-carboxy arabinitol bisphosphate (2CABP; Newman & Gutteridge, 1993; Schreuder *et al.*, 1993; Andersson, 1996), and is held in place by a hydrogen bond between Lys334 (spinach numbering) and the C2 carboxyl group in 2CABP. When the reaction is completed, the loop is again flexible as observed for activated rubisco complexed with the reaction product 3-phosphoglycerate (3-PGA; Taylor & Andersson, 1997b). In the non-activated form of the enzyme (Taylor & Andersson, 1997a), RuBP acts as an inhibitor which causes the closure of loop 6. Another pro-

tein, rubisco activase, is needed to release rubisco from this dead-end complex (Portis, 1990). In the non-activated form, 2CABP binds with its phosphate groups in the opposite orientation compared to the activated form complexed with 2CABP (Zhang *et al.*, 1994). In this position there is no possibility for hydrogen bonding between 2CABP and loop 6 which remains in the flexible state. Binding of 4CABP and xylulose bisphosphate also causes closure of this loop and, in addition, release of CO_2 and Mg^{2+} from the active site (Taylor *et al.*, 1996; Newman & Gutteridge, 1994).

Mutational analysis of the rubisco active site has been a valuable tool in elucidating structure/function relationships (Brändén *et al.*, Hartman & Harpel, 1994). The prokaryote *A. eutrophus* fixes CO_2 using a red-like rubisco, and represents a very simple system for studies of rubisco *in vivo* efficiency (King & Andersen, 1980). The *A. eutrophus* rubisco genes have been sequenced and characterized (Andersen & Caton, 1987). Rubisco deletion strains have been constructed that can be complemented by cloned rubisco genes, and this represents a very efficient system for screening/selection of mutant rubiscos with altered catalytic properties. This system has been used to isolate and characterize over 30 *A. eutrophus* rubisco mutants after random *in vitro* mutagenesis of the rubisco genes. These mutant rubiscos exhibit a wide range of catalytic properties, including altered specificity factors (K.A., unpublished results). The structure of the *A. eutrophus* wild-type rubisco presented here will serve as a basis for further design and analysis of mutated forms of the enzyme. This is expected to give a more detailed understanding of the structure/function relationship for rubisco, especially as far as catalytic efficiency is concerned. Furthermore, rubisco from *A. eutrophus* represents the first structure determined for the main evolutionary group of red-like rubiscos. Here, we present the three-dimensional structure of the unactivated form of rubisco from *A. eutrophus* at 2.7 Å resolution.

Results and Discussion

The overall structure

The three-dimensional structure of rubisco from *A. eutrophus* has been solved to 2.7 Å resolution by the molecular replacement method using a modified model of the spinach enzyme as a search model. The present structure (Figure 1) shows that, like the other members of the form I group, rubisco from *A. eutrophus* is built up of eight large and eight small subunits. The large subunits form dimers which are arranged around a local non-crystallographic 4-fold axis. Clusters of four small subunits, also related by 4-fold internal symmetry, from the top and bottom of the complete hexadecamer. A solvent channel of varying radius runs through the entire L_8S_8 molecule along the 4-fold axis.

Except for the C-terminal region of the small subunits, which forms a completely new structural motif not observed in any previous rubisco structure, the overall fold of the subunits and their relative location are very similar to what has been found for rubisco from spinach, tobacco and *Synechococcus*. Since the structures of these enzymes, and especially the spinach enzyme, have been described in great detail elsewhere (Knight *et al.*, 1990), only a brief summary of the general features of the structure will be given here, and the main emphasis will be on the differences. The sequence numbering of the *A. eutrophus* enzyme will be used here. The corresponding residue number in the spinach enzyme will be indicated in parenthesis, where appropriate. To facilitate comparison, the numbering of secondary structure elements used for the spinach enzyme (Andersen, 1979) has been retained, even though the number of α-helices and β-strands are slightly different in the two enzymes. The small subunits are denoted I, J, K and L in an anti-clockwise direction looking along the 4-fold axis into the solvent channel of the enzyme and similarly for the large subunits which are denoted A, C, E and G. Figure 1 shows the arrangement of subunits around the 4-fold axis. Figure 2 shows the alignment of the rubisco sequences from *A. eutrophus, Synechococcus* and spinach. The location of secondary structure elements are also indicated. The Figure demonstrates the high degree of conservation of secondary structural elements between *A. eutrophus* and the two green-like rubiscos. Sequence identities (excluding gaps) between *A. eutrophus* and *Synechococcus* or spinach large subunits are 59 % and 57 %, respectively, while they are 37 % and 38 % for the small subunits.

The large subunit

The large subunit of rubisco from *A. eutrophus* consists of 486 amino acid residues which are folded into two separate domains. The N-terminal domain, which consists of about one-third of the amino acids, is folded into a mixed parallel/antiparallel five-stranded β-sheet, whereas the C-terminal domain, consisting of the remaining residues, has an eight-stranded parallel α/β-barrel as the main structural motif. For the core of the large subunits, especially for the helix and strand segments, the electron density maps were well-defined and only minor adjustments were necessary to fit the structure. In contrast, the first 21 and the last 19 residues of the large subunit could not be located in the electron density maps. In addition, some regions close to the active site have large temperature factors indicating considerable flexibility. Compared to the spinach enzyme there are insertions of two and nine amino acid residues, respectively, in the N, and C termini of the large subunit, but these are, of course, not visible in the present structure since both termini are disordered. In addition there is a single insertion (Glu24) and one deletion (Tyr269 in spinach; Figure 2). The electron

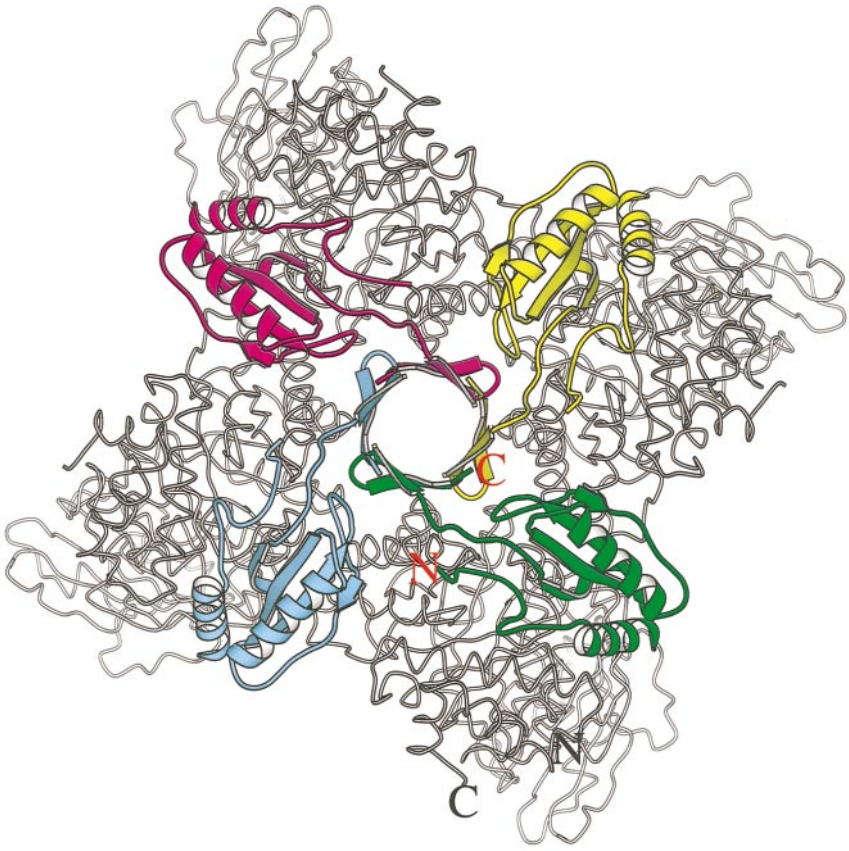

Figure 1. Structure of *A. eutrophus* rubisco. Arrangement of the crystallographic asymmetric unit around the local 4-fold axis is shown. The small subunits (I, J, K, and L) are coloured in magenta, green, yellow and light blue, and the large subunits (A, C, e, and G) in grey. For clarity, only one half (L_4S_4) of the molecule is shown. The C and N termini in a large subunit are marked in black and those in the associated small subunit in red. There are 486 and 139 residues, respectively, in the large and small subunits. The figure was produced with MOLSCRIPT/BOBSCRIPT (Kraulis, 1991).

density of both these regions is well defined, and the structure could easily be rebuilt without any major consequences for the surrounding areas.

As for the amino acids which are different in the *A. eutrophus* and the spinach or *Synechococcus* enzyme, there is no general trend in the distribution. The sequence differences are found both in the buried regions and at the interfaces between the subunits. This is in contrast to the tobacco enzyme where the majority of the mutational differences from the spinach enzyme are on the protein surface (Schruder *et al.*, 1993). It is of course important to note that more than 90% of the sequence of one LS-pair is identical in the two latter enzymes, while the corresponding numbers for the *A. eutrophus* and the spinach enzymes are only about 57 ad 38% (excluding gaps) for the large and the small subunits, respectively.

The active site

Two active sites are located at the interface between the two dimer-related large subunits. Each active site consists partly of amino acid side-chains from the C-terminal barrel domain of one of the monomers and partly of residues from the N-terminal domain of the other dimer-related monomer. The electron density maps show no trace of metal ions in the active site. Neither is there any trace of a carbamoyl moiety at the tip of the essential lysine, Lys204 (201), thus verifying that the enzyme is in its non-activated state. Two phosphate ions are located in the active site region. This is not surprising since the crystals were grown at a relatively high phosphate concentration (up to 1 M). Both bind in what may be called the classical phosphate binding sites for rubisco (Curmi *et al.*, 1992). One of the phosphate ions is within hydrogen-bonding distance of N^{ε} of Arg297 (295) and $N^{\delta 1}$ of His329 (327), whereas the second group fits in nicely within hydrogen-bonding distance of the main-chain nitrogen atoms of Gly405 (403) and Gly406 (404) and the hydroxyl group of Thr68 (65).

All amino acid residues which have been shown to interact with the transition state analogue 2CABP in the spinach enzyme (Andersson, 1996) are conserved in the *A. eutrophus* enzyme (Figure 2). There are, however, several differences in neighbouring residues. Loop 6 is also highly conserved. The ongoing mutagenesis studies are expected to shed more light on which sequence differences determine the differences in catalytic properties between *A. eutrophus* and other rubiscos.

It has been shown previously that in the activated state of the enzyme, the side-chains of the carbamylated Lys204 (201), Asp206 (203) and Glu207 (204) and three water molecules are coordinated to the Mg^{2+} ion (Taylor & Andersson, 1996). In the present structure the side-chain of Lys204 (201) is well defined, whereas the side-chains of Asp206 (203) and Glu207 (204) are completely dis-

ordered. Actually, high temperature factors for the entire Glu207 (204) - Met215 (212) loop in the present structure indicate considerable flexibility (see Figure 3). Most of the regions in the large subunit with high temperature factors are loops located close to the active site. This is especially the case

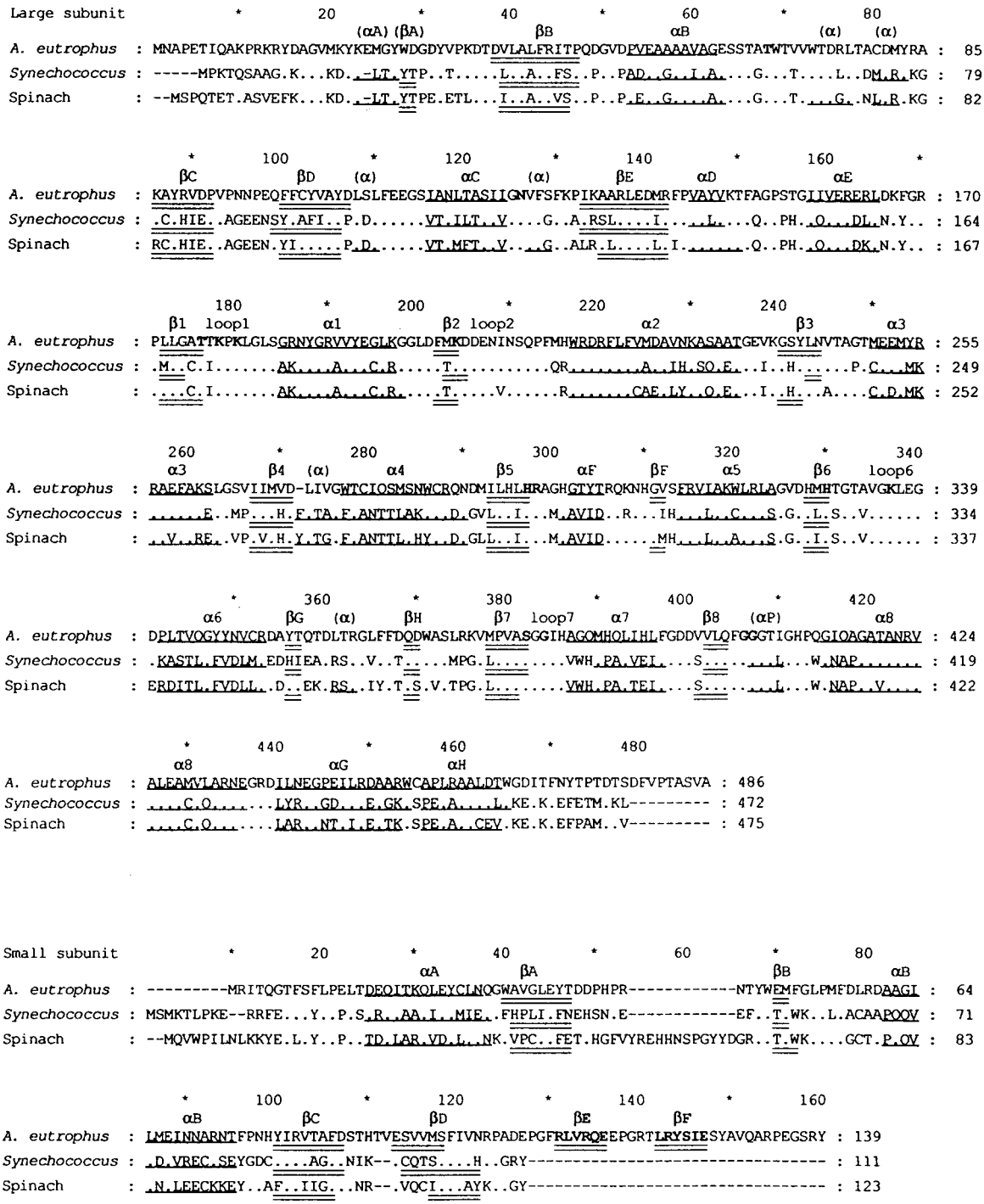

Figure 2. Amino acid sequence and secondary structure comparison of rubisco from *A. eutrophus* to the spinach and the *Synechococcus* enzymes. Residues identical to the *A. eutrophus* sequence are indicated with (.), and gaps with (-). Secondary structure elements are indicated, with single and double underlining denoting α-helical and β-strand regions, respectively. They are named according as described by Andersen (1979), with elements not detected in the *A. eutrophus* protein in parenthesis. Data for the *A. eutrophus* protein is from the present work, with the sequences forming the unique β-barrel structure in the small subunits shown in bold. In the large subunit sequence, residues involved in binding of the transition state analogue 2CAPB in activated spinach rubisco (Andersson, 1996) are shown in bold. Secondary structure for spinach rubisco are from (Knight *et al.*, 1990) and PDB entry 8RUC, and for *Synechococcus* rubisco from (Newman & Gutteridge, 1993) plus PDB entries 1RBL and 1RLC, with the sequence corrections suggested by the authors.

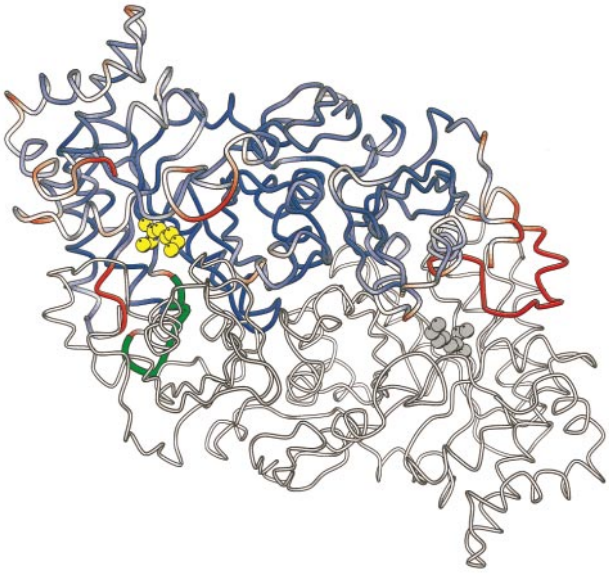

Figure 3. Two dimer-related large subunits viewed down the crystallographic 2-fold axis. The backbone in one subunit is colored according to temperature factors (from dark blue for low, to red for high temperature factors) and shows that the regions around the active sites are the most flexible. The essential lysine residue (Lys204) is shown in yellow. The flexible Glu207 to Met215 loop is emphasised in green. The Figure was produced with MOLSCRIPT/BOBSCRIPT (Kraulis, 1991).

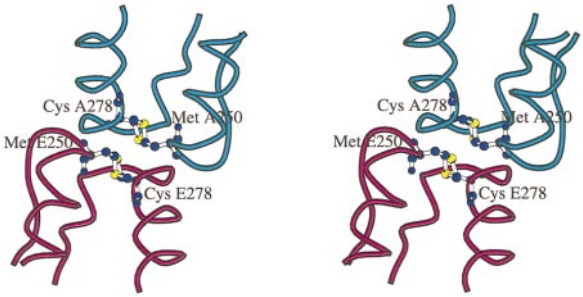

Figure 4. Stereo diagram showing the potential disulfide link connecting two dimer-related large subunits (A in blue and E in green). In the spinach enzyme there are cysteine residues in the 250 positions and alanine residues in the 278 positions. The Figure was produced with MOLSCRIPT/BOBSCRIPT (Kraulis, 1991).

for the loop from 65 to 78 which is highly flexible, thus resembling the situation in non-activated, unliganded rubisco from tobacco where residues 67 to 78 (64 to 68) were completely disordered (Curmi *et al.*, 1992). Also the helix (helix B) preceding this loop has substantially higher temperature factors than the other helices in the large subunit of the *A. eutrophus* enzyme. This could partly be explained by the high content of alanine and other short-chained amino acids, a feature which is conserved in several rubisco sequences. Other flexible regions close to the active site include the loop connecting strand 1 and helix 1 and, not unexpectedly, loop 6. The general impression is that the active site region in this non-activated, unliganded form of the enzyme is very flexible and accessible for the substrates.

Disulfide links

There are fewer cysteine residues in the *A. eutrophus* rubisco than in the spinach, tobacco or *Synechococcus* enzymes. For *A. eutrophus*, the numbers are six and one for the large and small subunits, respectively, *versus* nine and four for the spinach enzyme. In both the tobacco, spinach and the *Synechococcus* enzymes, a disulfide bridge also links two dimer-related cysteine residues (Cys247). This residue is present in most of the green-like rubisco sequences and seems to be conserved in higher plants. The corresponding residue in the *A. eutrophus* enzyme, as well as in most other red-like enzymes, is methionine. In the *A. eutrophus* enzyme, however, the alanine in position 276 in spinach, tobacco, *Synechococcus* and most other L_8S_8 sequences, is replaced by a cysteine residue (Cys278) which could make a disulfide link to the corresponding cysteine 278 in the dimer-related large subunit (see Figure 4). In the present structure the distance between the sulfur atoms is 2.8 Å, a distance which is longer than a normal S-S bond. There is electron density for the side-chain in none of the subunits, but the temperature factors are quite high. For the dimer-related subunit the side-chain of Cys278 is disordered, indicating that at least in the crystal structure there is no disulfide bridge. A possible explanation could be that the enzyme was kept under reducing conditions during purification. An interesting point is that in the spinach enzyme the cysteine residue is located in the N-terminal part of helix 3 in the α/β-barrel, whereas in the *A. eutrophus* enzyme the cysteine is located near the N-terminal end of the neighboring helix 4. During evolution there has then been a double mutation which may conserve the structure-function relationship.

The small subunit

The small subunit of rubisco from *A. eutrophus* consists of 139 amino acid residues. Compared to the spinach enzyme, it has a deletion of seven amino acid in the N terminus, and a deletion of 12 amino acid residues about halfway in the sequence (residues 52 to 63 in the spinach enzyme). It also has an extension of 33-35 amino acid residues at the C terminus (the size depends on the alignment used). The present structure shows that the central part is folded into a four-stranded antiparallel β-sheet and two α-helices, and is very similar to what has been observed for the L_8S_8 rubisco structures solved previously. The region between β-strands A and B is perhaps the most highly conserved part of the red-like small subunit sequences

determined so far (Figure 5). This is in contrast to green-like sequences which exhibit considerable variation in this region (Figure 2). The electron density in the region of the internal deletion (when compared to the spinach enzyme) is well defined, and the new link consisting of residues Arg44 (spinach Val51) and Asn45 (Gly64) makes, together with His42 (Gly49) and Pro43 (Phe50), a classical type I reverse turn (see Figure 6). In the spinach enzyme residues 52 to 63 form a hairpin loop which protrudes into the solvent channel and makes an extensive network of interactions with

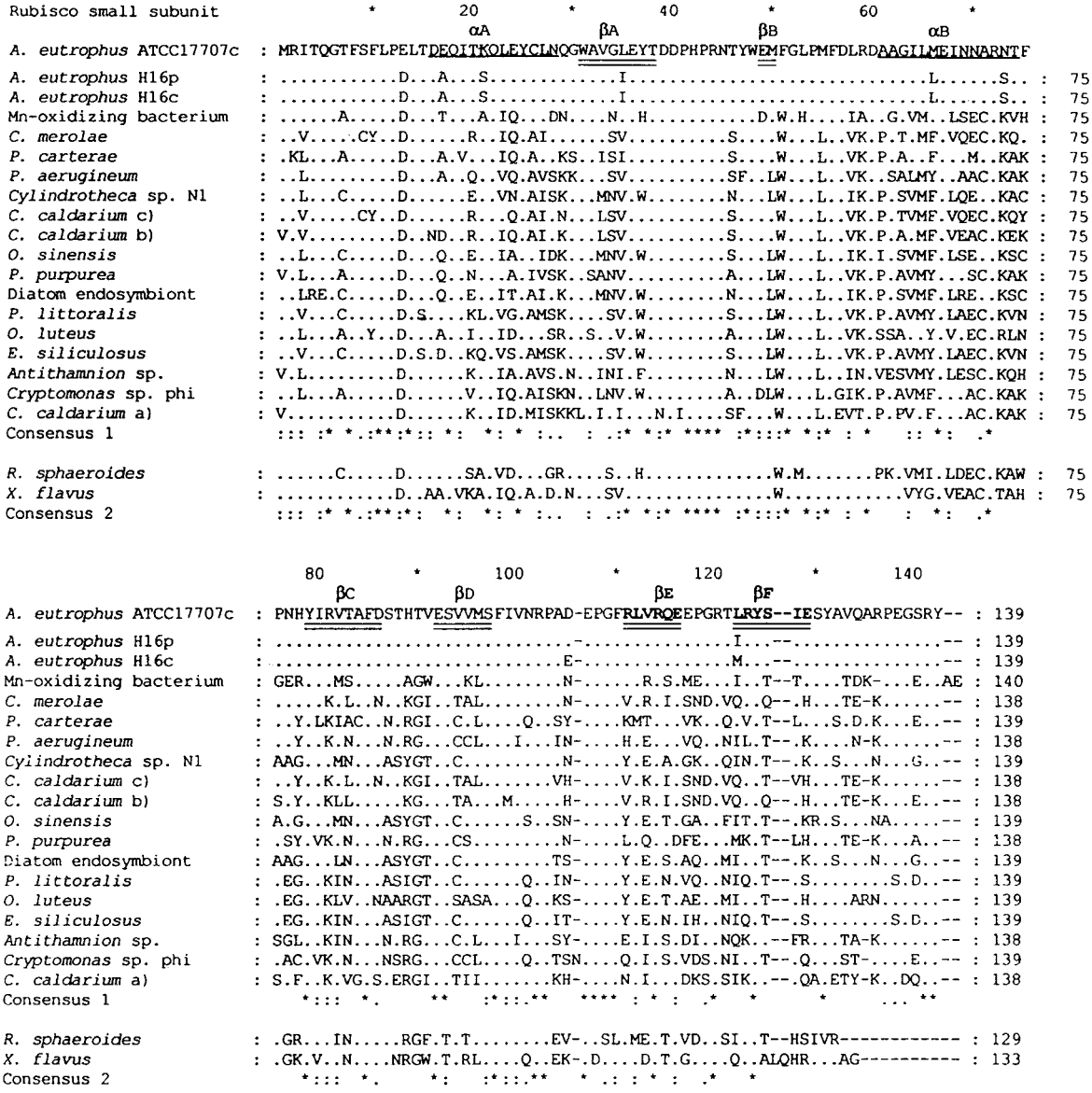

Figure 5. Sequence conservation among small subunits from red-like rubisco proteins. Alignment position numbers are indicated above, and actual sequence numbers at the end of each line. Residues identical to the *A. autrophus* ATCC17707 sequence are indicated with (.), and gaps with (-). Consensus line 1 (produced with CLUSTAL W 1.7; Thompson *et al.*, 1994) is for all except the *R. sphaeroides* and *X. flavus* sequences, and consensus line 2 for all sequences. Symbols in the consensus lines are: (-) gap; (*) identical; (:) highly similar; and (.) weakly similar residues. Secondary structure assignment for the *A. eutrophus* protein (from Figure 1) is indicated, with single and double underlining indicating α-helix and β-strand regions, respectively. The sequences are from the following organisms (with GeneBank sequence GI numbers in parenthesis): Chemoautotrophic bacteria: *A. autrophus* ATCC 17707 chromosomally encoded (gi132131); *A. eutrophus* H16 chromosomally encoded (gi2120966); *A. eutrophus* H16 plasmid encoded (gi2120967); manganese-oxidizing bacterium S185-9A1 (gi567190); and *Xanthobacter flavus* (gi132162). Photosynthetic bacterium: *Rohdobacter sphaeroides* (gi132084). The rest of the sequences are chloroplast encoded from red and other eukaryotic algae: *Anthhamnion* sp. (gi1334342); *Cryptomonas* sp. phi (gi132137); *Cyanidioschyzon merolae* (gi2541884); *Cyanidium caldarium* a) (gi1334348); *C. caldarium* b) (gi1334350); *C. caldarium* c) (gi2541894); *Cylindrotheca* sp. N1 (gi132141); diatom (*Peridinium*) endosymbiont (gi1675202); *Ectocarpus siliculosus* (gi132142). *Odontella sinensis* (gi1352822); *Olisthodiscus luteus* (gi132149); *Pilayella littoralis* (gi132152); *Pleurochrysis carterae* (gi730478); *Porphyra purpurea* (gi2148038); and *Porphyridium aerugineum* (gi730479).

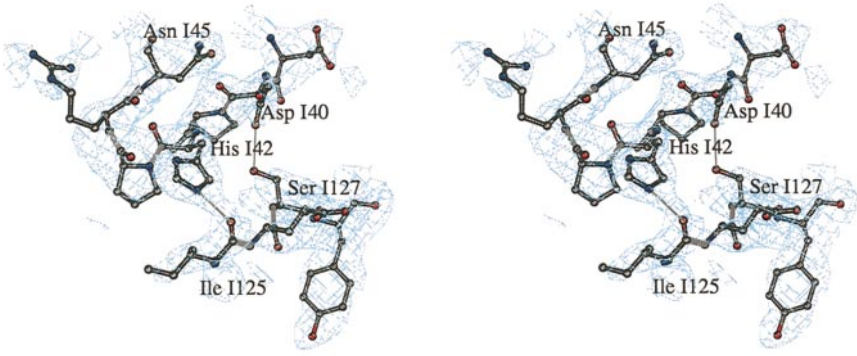

Figure 6. Stereo diagram showing $2F_o - F_o$ electron density for the hairpin loop between His I-42 and Asn I-45 in the small subunit. In rubisco from spinach there is an insertion of 12 amino acid between Arg I-44(Val51) and Asn I-45(Gly64) (see Figure 2). The Figure also shows hydrogen bonds within the small subunit linking previously determined structure (I125-I127) and the ''new'' structure (I40-I45). The Figure was produced with MOLSCRIPT/BOBSCRIPT (Kraulis, 1991).

the C-terminal barrel of the large subunit. In the present structure the shortened loop is mainly located at the surface of the enzyme and makes only one major interaction with the large subunit. Both N^{ε} and NH2 of Arg44 (Val51) (subunit 1) are within hydrogen-bonding distance of the carbonyl oxygen of Glu163 (Asp160) in large subunit A.

The most interesting feature of the small subunit of rubisco from *A. eutrophus* is the extended C-terminal region. Residues Ala104 to Glu106 (125-127 in the alignment in Figure 2), which correspond to the C terminus of the spinach enzyme, are located at the surface of the *A. eutrophus* enzyme, close to the entrance to the solvent channel. The electron density maps showed unassigned, continuous electron density inside the channel which was of such a quality that new structure could be assigned with the correct sequence for residues 104 to 129. From Phe109, the chain extends down into the solvent channel. It then turns and runs back out again forming a two-stranded anti-parallel β-sheet (see Figure 7). The two strands (denoted E and F) of the small subunit I, each consisting of six amino acid residues, are linked by a loop consisting of five residues (denoted the EF loop hereinafter). Strands E and F of the small subunit I run alongside helix 2 of the large subunit A, whereas the EF loop is located close to two of the loops in the C-terminal domain of the large subunit G. The last three residues (127 to 129) are located closer to the surface of the enzyme and are less well ordered. For the last ten residues (130 to 139), there is no trace of electron density at all, suggesting that these residues are highly mobile. We note, however, that the sequence of this region has also been highly conserved during evolution (Figure 5).

Together, the C-terminal regions of four small subunits from an elegant eight-stranded antiparallel β-barrel (see Figure 8) which sits as a plug in the solvent channel, decreasing its radius considerably. The minimum radius between C^{α} atoms is 13 Å. Including the side-chains, the shortest radius is 5.5 Å. In the spinach enzyme, the radius is about 17 Å. The length of the barrel, which has the classical right-handed twist, is approximately 20 Å. The only main-chain hydrogen bond linking strands from neighbouring subunits is between N of Glu115 and O of Tyr123. Inside the barrel, the side-chains of Arg110 and Glu126 are involved in a network of hydrogen bonds which link all eight strands together (see Figure 9). One of the main interactions with the large subunit is between Arg122 and Ser262. Arg122 is actually located on the inside of the barrel, but the side-chain protrudes from the bottom of the barrel and makes a hydrogen bond to the carbonyl oxygen atom of Ser262. Several of the residues which are located on the outside of the barrel are involved in both intra and inter-subunit interactions. For these resi-

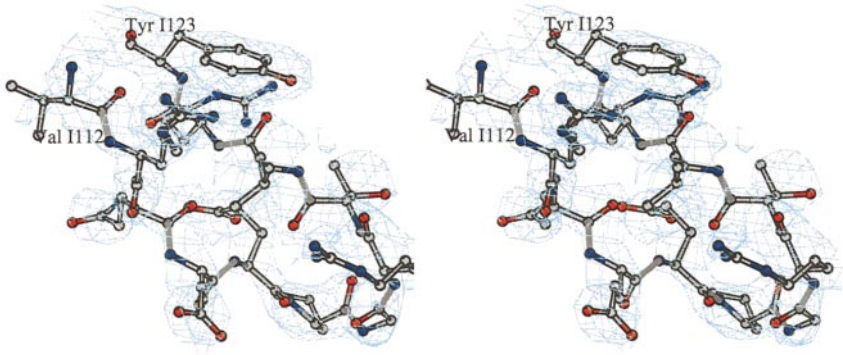

Figure 7. Stereo diagram showing $2F_o - F_o$ electron density for part of the new two-stranded antiparallel β-sheet in the small subunit.

Figure 8. The new eight-stranded antiparallel β-barrel formed by the C-terminal regions of four small subunits. The small subunits are colored as described in the legend to Figure 1. If Figure 1 is defined as a top view, then this Figure shows a side view of a S_4 unit.

dues, the main interaction with the large subunit is a hydrogen bond between the hydroxyl group of Tyr123 and $O^{\delta 1}$ of Asp226 in helix 2. Within the small subunit, Ile125 and Ser127 are hydrogen bonded to with His42 and Asp40, respectively (Figure 6). The tip of the EF loop consists of Pro117, Gly118 and Arg119. The interacts with the large subunit *via* a hydrogen bond between the main-chain nitrogen atom of Gly118 (subunit I) and the main-chain oxygen atom of Lys261 (subunit G). The electron density for the side-chain of Arg119 is less well defined than for the rest of the residues in the β-sheet region, giving a higher degree of flexibility, which could partly be explained by the fact that the δ-guanido group has several possibilities for forming hydrogen bonds with large subunit G, depending on its conformation.

Comparison of the small subunits of the present structure with the spinach structure, showed that the EF loop of the former is located in the same region as the tip of the hairpin loop of the latter. Or more precisely, the EF loop of subunit I occupies the same region as the hairpin loop in the neighbouring subunit J of the spinach enzyme (see Figure 10). It is interesting to note that the *Synechococcus* enzyme, which has a short C-terminal region similar to the spinach enzyme, also lacks the hairpin loop thus resulting in a solvent channel with even larger radius.

The present structure is the first which has been solved for any red-like rubisco. Sequence identities between small subunits from *A. eutrophus* and eukaryotic algae in the red-like group (Figure 5) are 54 to 65 % (excluding gaps) For the large subunit, the corresponding sequence identities are 71 to 76 %. The small subunit C-terminal sequence is also highly conserved, except for the two bacteria *R. sphaeroides* and *X. flavus*. The structure described here is, therefore, likely to represent a fold which is common for other red-like rubiscos.

Accessible and buried surfaces in *A. eutrophus* rubisco calculated with and without the new C-terminal extension are given in Table 1, from which it is apparent that the new structure results in a 4 % reduction of the total accessible surface of the complete L_8S_8 molecule. This is mainly due to increased interactions between the large and small subunits. It is also clear that the new eight-stranded β-barrel results in a major increase in the buried surface between the small subunits, which may well have a stabilizing effect on the assembled L_8S_8 oligomer. It is also possible that long-range interactions between the new structural elements and the active site may modulate substrate specificity.

Materials and Methods

Purification

Chromosomally encoded rubisco from *A. eutrophus* strain ATCC17707 was purified by a modification of a procedure used previously (Andersen, 1979), and will be described in detail elsewhere. Briefly, for crystallization purposes, two extra purification steps were included where the enzyme was first rechromatographed on an anion exchange column at pH 6.8 using a Pharmacia FPLC system, concentrated by $(NH_4)_2SO_4$ precipitation (50 % saturation), and dissolved in a buffer containing 50 mM Tris-HCl (pH 7.8), 0.1 mM EDTA and 1 mM Dithiothreitol (DTT). The enzyme was finally applied to a Pharmacia Superdex 200 (16/60) gel filtration column and then concentrated by $(NH_4)_2SO_4$ precipitation. The

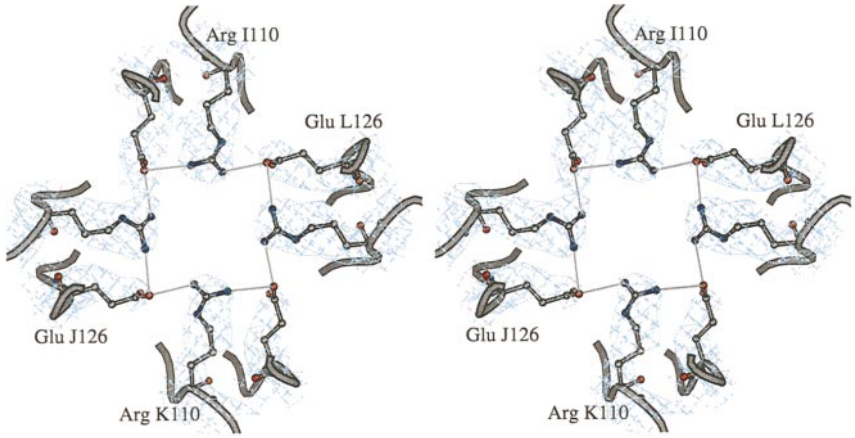

Figure 9. Stereo view into the solvent channel showing the network of hydrogen bonds which link all eight strands (two from each small subunit) of the β-barrel together.

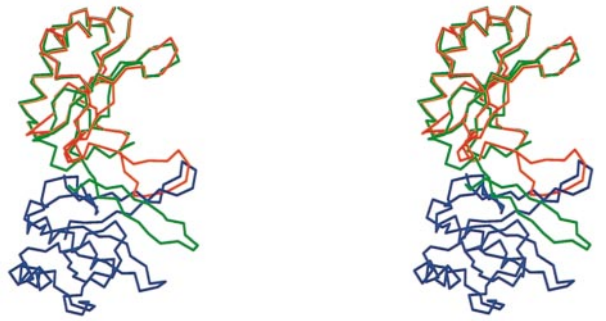

Figure 10. Superimposition (in stereo) of small subunit I of the spinach enzyme (red) on the *A. eutrophus* enzyme (green). The neighboring subunit J of the *A. eutrophus* enzyme is shown in blue. The Figure was produced with O (Jones *et al.*, 1991).

purity of the resulting rubisco preparation was >98-99 %, as judged by SDS-PAGE.

Crystallization and data collection

The very first crystallization experiments yielded well-defined crystals, unfortunately too small for diffraction experiments. It took a further two-and-a-half years before crystals large enough for data collection using synchrotron radiation were obtained (Hansen *et al.*, 1998). The crystals of rubisco from *A. eutrophus* which were used for structure determination, were grown in sitting drops at 4 °C. The drops were prepared by mixing equal volumes of protein solution containing 40 mg/ml rubisco in 20 mM Hepes buffer (pH 7.5) with reservoir solution consisting of 0.7-0.8 M K_2HPO_4 (pH 9.2). Crystals grew within two weeks. The crystals belong to space group $P4_32_12$ with cell parameters $a = 112.0$ Å and $c = 402.7$ Å. Prior to data collection, the crystals were transferred to a cryo protection solution containing 1 M K_2HPO_4 and 25 % (w/v) ethylene glycol. After a few seconds, the crystals were removed from the solution by fibre loops and frozen immediately in liquid nitrogen. The size of the crystal used for data collection was 0.15 mm × 0.15 mm × 0.05 mm. Intensity data were collected at the Swiss-Norwegian Beamline (BM01, I = 0.9117 Å), ESRF, Grenoble, at 100 K using a MAR image plate detector at a distance of 420 mm and an oscillation angle of 0.25 deg. The data were processed using DENZO (Otwinowski, 1993), whereas SCALA and AGROVATA from the CCP4 program suite (CCP4, 1994) was used for scaling and merging. The data set, which consisted of 65,785 unique reflections, was 87.5 % complete in the resolution range 25 to 2.7 Å with an R_{merge} of 12.5 % (average multiplicity 4.4). From the statistics given in Table 2, it is clear that the data are increasingly poor beyond ca 3.1 Å. Nevertheless, since a significant fraction of the experimental data lies between 3.1 and 2.7 Å, we elected to use the whole data set in the structure determination and refinement. It should, of course, be pointed out that the multiplicity over the whole resolution range is greater than four.

Table 1. Effect of the C-terminal extension on accessible areas and surface areas in *A. eutrophus* rubisco

	Accessible surface area (Å^2)	
	Without C-terminal extension	With C-terminal extension
A. *Accessible surface areas in* A. eutrophus *rubisco*		
S	6682	8244
L	19,000	18,548
LS	23,334	23,777
S_4	26,766	29,202
L_2	30,025	29,291
L_4S_4	83,769	80,114
L_8S_8	128,822	123,741
B. *Buried surface areas in* A. eutrophus *rubisco*		
	Buried surface area (Å^2)	
	Without C-terminal extension	With C-terminal extension
L-S		
	1177	1528
	1171	1487
L-L		
	4026	3841
S-S-S-S		
	10	986

Structure determination and refinement

The crystal structure of rubisco from *A. eutrophus* was determined using a combination of the molecular replacement method (program AMoRe; Navaza, 1994), density modification (program DM; Cowtan, 1994) and automatic procedures for tracing continuous electron density (program BONS; Jones *et al.*, 1991). A modified model of rubisco from spinach (Brookhaven pdb code 1aus) where the spinach sequence had been replaced by the *A. eutrophus* sequence, was used as a search model in the molecular replacement calculations. The entire asymmetric unit consisting of four large and four small subunits related by 4-fold non-crystallographic symmetry, was used as a search object. Since the initial difference electron density map was only partly interpretable, 30 cycles of density modification (solvent-flattening, histogram-matching and 4-fold molecular averaging) was performed using the program DM. The mask used during the calculations was generated by NCSMASK and SIGMAA was used for estimating the reliability of the calculated phases (both programs from the CCP4 suite). The resulting electron density map was an incredible improvement compared to the initial map and made it possible to fit most of the model to electron density using the program O (Jones *et al.*, 1991). The crystallographic *R*-factor after manual rebuilding was 48 %, but after 200 cycles of positional refinement using the program X-PLOR, the *R*-factor dropped to 40 % (Brünger *et al.*, 1987; Brünger & Krukowski, 1990). Cycles of rebuilding with subsequent refinement (simulated annealing, positional, overall and grouped *B*-factor) using non-crystallographic symmetry restraints reduced the *R*-factor to 32 % (R_{free} 38 %; Brünger, 1992). The lengthened C terminus of the small subunit was then located using the program BONES. Aided by the Slider and Lego commands in O, the direction of the chain was readily apparent and the correct amino acids were fitted to the electron density map. Subsequent refinement resulted in a final *R*-factor of 26.6 % (R_{free} 32.2 %) using data in the resolution rang e15 to 2.7 Å. The R_{cryst}/R_{free} ratio falls well within range observed for published structures of this size and resolution (Tickle *et al.*, 1988). No water molecules were added because of the limited resolution of the data. Refinement statistics are summarized in Table 3. Refinement of the structure using

Table 2. Merging and intensity statistics for *A. eutrophus* rubisco

N_{unique}	Resolution (Å)	$I/\sigma I$	Completeness (%)	Multiplicity	R_{merge} (%)
2325	7.99	7.7	79.8	4.4	5.3
3821	5.80	8.0	83.3	4.5	6.0
4923	4.78	8.0	85.7	4.4	6.4
5784	4.16	8.6	86.4	4.4	6.7
6606	3.73	6.4	87.8	4.4	9.2
7319	3.41	5.6	88.6	4.4	13.4
7939	3.16	3.7	88.8	4.3	21.7
8512	2.96	2.5	88.9	4.3	32.5
9010	2.79	1.6	88.7	4.3	50.0
9546	2.65	1.1	89.1	4.3	69.5
Total 65,785[a]			87.5	4.4	12.5

The total observations are 286,973 ($I/\sigma(I) > 0$) in the resolution range 25.2.7 Å.
[a] This is the number of unique reflections.

Table 3. Refinement statistics for *A. eutrophus* rubisco

A. *Refinement statistics*	
R-factor[a] (15-2.7 Å resol.)	26.6
R_{free}[b]	32.2
Number of protein atoms	18,112
rms deviation from ideal values:	
Bond lengths (Å)	0.010
Bond angles (°)	1.2
Dihedral angles	22.76
Improper angles	1.11
B. *Ramachandran plot*[c]	
Residues in most favored regions (%)	84.3
Residues in additional allowed regions (%)	12.6

[a] *R*-value $= \Sigma \mid |F_{calc}| - |F_{obs}| \mid / \Sigma |F_{obs}| \times 100$, where F_{calc} and F_{obs} are the calculated and observed structure factors, respectively.
[b] *R*-free: *R*-factor over a set of 10% randomly selected reflections, always omitted from refinement.
[c] Data from the program PROCHECK (Laskowski *et al.*, 1993).

REFMAC (CCP4, 1994) did not yield any new information about the disordered regions.

The quality of the model was monitored at several stages of the rebuilding and refinement process using the programs OOPS (Kleywegt & Jones, 1996) and PROCHECK (Laskowski *et al.*, 1993). The final model had rms deviations for bonds and angles of 0.010 Å and 1.2°, respectively, as calculated by using X-PLOR. When analyzed by PROCHECK, 83.8% of the residues of the large subunit fell within the most favored regions of the Ramachandran plot, 14.1% were in the additional allowed regions and 2.1% in the generously allowed regions. The corresponding values for the small subunit were 72.3, 25.0 and 2.7%.

Protein Data Bank accession number

Coordinates have been deposited in The Brookhaven Protein Data Bank with accession number 1bxn.

Acknowledgments

We thank the Nordic Council/Nordic Energy Research Program and the Norwegian Research Council for financial support.

References

Andersen, K. (1979). Mutations altering the catalytic properties of a plant-type ribulose bisphosphate carboxylase/oxygenase in *Alcaligenes eutrophus*. *Biochim. Biophys. Acta*, **585**, 1-11.

Andersen, K. & Caton, J. (1987). Sequence analysis of the *Alcaligenes eutrophus* chromosomally encoded ribulose bisphosphate carboxylase large and small subunit genes and their products. *J. Bacteriol.* **169**, 4547-4558.

Andersson, I. (1996). Large structures at high resolution: the 1.6 Å crystal structure of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase complexed with 2-carboxy-arabinitol bisphosphate. *J. Mol. Biol.* **259**, 160-174.

Andersson, I., Knight, S., Schneider, G., Lindqvist, Y., Lundqvist, T., Brändén, C.-I. & Lorimer, G. H. (1989). Crystal structure of the active site of ribulose-bisphosphate carboxylase. *Nature,* **337**, 229-234.

Brändén, c.-I. , Lindqvist, Y. & Schneider, G. (1991). Protein engineering of rubisco. *Acta Crystallog. sect. B,* **47**, 824-835.

Brünger, A. T. (1992). Free *R*-value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature,* **355**, 472-474.

Brünger, A. T. & Krukowski, A. (1990). Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallog. sect. A,* **46**, 585-593.

Brünger, A. T., Kurian, J. & Karplus, M. (1987). Crystallographic *R* factor refinement by molecular dynamics. *Science,* **235**, 458-460.

Chapman, M. S., Suh, S. W., Curmi, P. M., Cascio, D., Smith, W. W. & Eisenberg, D. S. (1988). Tertiary structure of plant RuBisCO: domains and their contacts. *Science,* **241**, 71-74.

Choe, H. W., Georgalis, Y. & Saenger, W. (1989). Comparative studies of ribulose-1,5-biphosphate carboxylase/oxygenase from *Alcaligenes eutrophus* H16 cells, in the active and CABP-inhibited forms. *J. Mol. Biol.* **207**, 621-623.

Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D,* **50**, 760-763.

Cowtan, K. (1994). An automated procedure for phase improvement by density modification. In *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, vol. 31, pp. 34-38, , Warrington, SERC Daresbury Laboratory.

Curmi, P. M. G., Cascio, D., Sweet, R. M., Eisenberg, D. & Schreuder, H. A. (1992). Crystal structure of the unactivated form of ribulose-1,5-bisphosphate carboxylase/oxygenase from tobacco refined at 2.0 Å resolution. *J. Biol. Chem.* **267**, 16980-16989.

Delwiche, C. F. & Palmer, J. D. (1996). Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. *Mol. Biol. Evol.* **13**, 873-882.

Hansen, S., Hough, E. & Andersen, K. (1998). Purification, crystallization and preliminary X-ray studies of two isoforms of Rubisco from *Alcaligenes autrophus*. *Acta Crystallog. sect. D,* **55**, 310-313.

Hartman, F. C. & Harpel, M. R. (1994). Structure, function, regulation and assembly of D-ribulose-1,5-bisphosphate carboxylase/oxygenase. *Annu. Rev. Biochem.* **63**, 197-234.

Holzenburg, A., Mayer, F., Harauz, G., van Heel, M., Tokuoka, R., Ishida, T., Harata, K., Pal, G. P. & Saenger, W. (1987). Structure of D-ribulose-1,5-bisphosphate carboxylase/oxygenase from *Alcaligenes eutrophus*. H16. *Nature,* **325**, 730-732.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A,* **47**, 110-119.

Jordan, D. B. & Ogren, W. L. (1981). Species variation in the specificity of ribulose bisphosphate carboxylase/oxygenase. *Nature,* **291**, 513-515.

King, W. R. & Andersen, K. (1980). Efficiency of CO_2 fixation in a glycollate oxidoreductase mutant of *Alcaligenes eutrophus* which exports fixed carbon as glycollate. *Arch. Microbiol.* **128**, 84-90.

Kleywegt, G. J. & Jones, T. A. (1996). Efficient rebuilding of protein structures. *Acta Crystallog. sect. D,* **52**, 829-832.

Knight, S., Andersson, I. & Brändén, C.-I. (1990). Crystallographic analysis of ribulose-1,5-bisphosphate carboxylase/oxygenase from spinach at 2.4 Å resolution. Subunit interactions and the active site. *J. Mol. Biol.* **215**, 113-160.

Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Kusian, B., Bednarski, R., Husemann, M. & Bowien, B. (1995). Characterization of the duplicate ribulose-1,5-bisphosphate carboxylase genes and cbb promoters of *Alcaligenes eutrophus*. *J. Bacteriol.* **177**, 4442-4450.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.

Lee, B. G., Read, B. A. & Tabita, F. R. (1991). Catalytic properties of recombinant octameric, hexadecameric, and heterologous cyanobacterial/bacterial ribulose-1,5-bisphosphate carboxylase/oxygenase. *Arch. Biochem. Biophys.* **291**, 263-269.

Nakagawa, H., Sugimoto, M., Kai, Y., Harada, S., Miki, K. & Kasai, N. (1986). Preliminary crystallographic study of a ribulose-1,5-bisphosphate carboxylase-oxygenase from *Chromatium vinosum*. *J. Mol. Biol.* **191**, 577-578.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A,* **50**, 157-163.

Newman, J. & Gutteridge, S. (1993). The X-ray structure of *Synechococcus* ribulose-bisphosphate carboxylase/oxygenase-activated quaternary complex at 2.2 Å resolution. *J. Biol. Chem.* **268**, 25876-25886.

Newman, J. & Gutteridge, S. (1994). Structure of an effector-induced inactivated state of ribulose-1,5-bisphosphate carboxylase/oxygenase: the binary complex between enzyme and xylulose 1,5-bisphosphate. *Structure,* **2**, 495-502.

Otwinowski, Z. (1993). Oscillation data reduction program. In *Proceedings of the CCP4 Study Weekend. Data Collection and processing* (Sawyer, L., Isaacs, N. & Bailey, S., eds), pp. 56-62, , Warrington, SERC Daresbury Laboratory.

Portis, A. R. (1990). Rubisco activase. *Biochim. Biophys. Acta,* **1015**, 15-28.

Read, B. A. & Tabita, F. R. (1994). High substrate specificity factor ribulose bisphosphate carboxylase/oxygenase from eukaryotic marine algae and properties of recombinant cyanobacterial RubiSCO containing ''algal'' residue modifications. *Arch. Biochem. Biophys.* **312**, 210-218.

Schneider, G., Lindqvist, Y. & Lundqvist, T. (1990). Crystallographic refinement and structure of ribulose-1,5-bisphosphate carboxylase from *Rhodospirillum rubrum* at 1.7 Å resolution. *J. Mol. Biol.* **211**, 989-1008.

Schreuder, H. A., Knight, S., Curmi, P. M. G., Andersson, I., Cascio, D., Sweet, R. M., Brändén, C.-I. & Eisenberg, D. (1993). Crystal structure of activated tobacco rubisco complexed with the reaction-intermediate analogue 2-carboxy-arabinitol 1,5-bisphosphate. *Protein Sci.* **2**, 1136-1146.

Shibata, N., Yamamoto, H., Inoue, T., Uemura, K., Yokota, A. & Kai, Y. (1996). Crystallization and preliminary crystallographic studies of ribulose-1,5-bisphosphate carboxylase/oxygenase from a red alga, *Galdieria partita*, with a high specificity factor. *J. Biochem.* **120**, 1064-1066.

Taylor, T. C. & Andersson, I. (1996). Structural transitions during activation and ligand binding in hexadecameric rubisco inferred from the crystal structure of the activated unliganded spinach enzyme. *Nature Struct. Biol.* **3**, 96-101.

Taylor, T. C. & Andersson, I. (1977a). The structure of the complex between rubisco and its natural substrate ribulose-1,5-bisphosphate. *J. Mol. Biol.* **265**, 432-444.

Taylor, T. C. & Andersson, I. (1997b). Structure of a product complex of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase. *Biochemistry,* **36**, 4041-4046.

Taylor, T. C., Fothergill, M. D. & Andersson, I. (1996). A common structural basis for the inhibition of ribulose-1,5-bisphosphate carboxylase by 4-carboxy-arabinitol 1,5-bisphosphate and xylulose 1,5-bisphosphate. *J. Biol. Chem.* **271**, 32894-32899.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673-4680.

Tickle, I. J., Laskowski, R. A. & Moss, D. S. (1988). R_{free} and the R_{free} ratio of expected values of cross-validation residuals used in macromolecular least-squares refinement. *Acta Crystallog. sect. D,* **54**, 547-557.

Watson, G. M. F. & Tabita, F. R. (1997). Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: a

molecule for phylogenetic and enzymological investigation. *FEMS Microbiol. Letters,* **146**, 13-22.
Wildner, G. F., Schlitter, J. & Muller, M. (1996). Rubisco, an old challenge with new perspectives. *Z. Naturforsch.* **C51**, 273-286.
Zhang, K. Y. J., Cascio, D. & Eisenber, D. (1994). Crystal structure of unactivated ribulose-1,5-bisphosphate carboxylase/oxygenase complexed with a transition state analog, 2-carboxy-arabinitol 1,5-bisphosphate. *Protein Sci.* **3**, 64-69.

Edited by R. Huber

(Received 8 September 1998; received in revised form 7 March 1999; accepted 10 March 1999)