Protein Science (1993), *2*, 1136–1146. Cambridge University Press. Printed in the USA.

Crystal structure of activated tobacco rubisco complexed with the reaction-intermediate analogue 2-carboxy-arabinitol 1,5-bisphosphate

HERMAN A. SCHREUDER,[1,4] STEFAN KNIGHT,[2] PAUL M.G. CURMI,[1,5] INGER ANDERSSON,[2] DUILIO CASCIO,[1] ROBERT M. SWEET,[3] CARL-IVAR BRÄNDÉN,[2] AND DAVID EISENBERG[1]

[1] Molecular Biology Institute and Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90024
[2] Swedish University of Agricultural Sciences, Uppsala Biomedical Center, Department of Molecular Biology, S-751 24 Uppsala, Sweden
[3] Department of Biology, Brookhaven National Laboratory, Upton, New York 11973

(RECEIVED June 9, 1992; REVISED MANUSCRIPT RECEIVED April 8, 1993)

Abstract

The crystal structure of activated tobacco rubisco, complexed with the reaction-intermediate analogue 2-carboxy-arabinitol 1,5-bisphosphate (CABP) has been determined by molecular replacement, using the structure of activated spinach rubisco (Knight, S., Andersson, I., & Brändén, C.-I., 1990, *J. Mol. Biol. 215*, 113–160) as a model. The *R*-factor after refinement is 21.0% for 57,855 reflections between 9.0 and 2.7 Å resolution.

The local fourfold axis of the rubisco hexadecamer coincides with a crystallographic twofold axis. The result is that the asymmetric unit of the crystals contains half of the L_8S_8 complex (molecular mass 280 kDa in the asymmetric unit). The activated form of tobacco rubisco is very similar to the activated form of spinach rubisco. The root mean square difference is 0.4 Å for 587 equivalent C^{α} atoms. Analysis of mutations between tobacco and spinach rubisco revealed that the vast majority of mutations concerned exposed residues. Only 7 buried residues were found to be mutated versus 54 residues at or near the surface of the protein.

The crystal structure suggests that the Cys 247–Cys 247 and Cys 449–Cys 459 pairs are linked via disulfide bridges. This pattern of disulfide links differs from the pattern of disulfide links observed in crystals of unactivated tobacco rubisco (Curmi, P.M.G., et al., 1992, *J. Biol. Chem. 267*, 16980–16989) and is similar to the pattern observed for activated spinach tobacco.

Keywords: CABP; disulfide links; photosynthesis; protein crystallography; rubisco

Photosynthetic carbon dioxide fixation in plants and other organisms starts with the condensation to one molecule of CO_2 to a five-carbon sugar, ribulose-1,5-bisphosphate, to yield two molecules of 3-phosphoglycerate (see Fig. 1). This carboxylation reaction is catalyzed by the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) (see Andrews & Lorimer [1987], Gutteridge [1990], and Brändén et al. [1991] for reviews).

In a side reaction, the oxygenation reaction, the enzyme can also utilize O_2 instead of CO_2 as a substrate, resulting in the formation of one molecule of 3-phosphoglycerate and one molecule of 2-phosphoglycolate. About three-quarters of the carbon present in the glycolate molecules can be salvaged at the expense of energy and reducing equivalents via the glycolate pathway, the remaining carbon being lost as CO_2. This oxygenation reaction is a major factor limiting crop yield and improving the carboxylase/oxygenase ratio is therefore of substantial interest.

Three-dimensional structures are essential to fully understand the mechanism of the carboxylase and oxygenase reaction and to allow rational changes to be made by site-directed mutagenesis. The first crystal structure of ru-

Reprint requests to: David Eisenberg, Molecular Biology Institute and Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90024.
[4] Present address: Marion Merrell Dow Research Institute, 16 Rue d'Ankara, 67009 Strasbourg Cedex, France.
[5] Present address: School of Physics, University of New South Wales, PO Box 1, Kensington, NSW 2033, Australia.

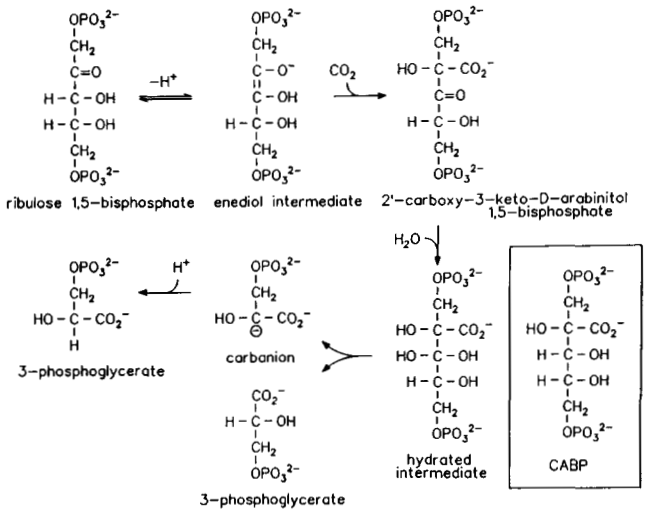

Fig. 1. The carbon-fixation reaction catalyzed by rubisco. The reaction-intermediate analogue 2-carboxy-arabinitol 1,5-bisphosphate (CABP) is shown in the box.

bisco to be reported was the structure of rubisco from the photosynthetic bacterium *Rhodospirillum rubrum*, consisting of two identical subunits (Schneider et al., 1986). However, rubisco from plants and most other photosynthetic organisms forms a complex of eight large (L, 55,000 M_r) and eight small (S, 15,000 M_r) subunits with a total relative molecular mass around 550,000. Crystal structures have been reported for the L_8S_8 tobacco rubisco (Chapman et al., 1988) and spinach rubisco (Andersson et al., 1989; Knight et al., 1989, 1990). Preliminary studies have further been reported on L_8S_8 rubiscos from the hydrogen bacterium *Alcaligenes eutrophus* (Pal et al., 1985), the purple sulfur photosynthetic bacterium *Chromatium vinosum* (Nakagawa et al., 1986), and the cyanobacterium *Synechococcus* PCC6301 (Newman & Gutteridge, 1990).

Comparison of the dimeric *R. rubrum* rubisco and the L_8S_8 spinach rubisco revealed a very similar active site, despite only 28% sequence identity between the spinach large subunits and *R. rubrum* rubisco and the absence of small subunits in the latter enzyme (Schneider et al., 1990). Rubisco needs to be activated before it becomes catalytically active (Lorimer et al., 1976; Lorimer, 1981). In this process, an active-site lysine (Lys 201 in tobacco rubisco) reacts with CO_2 to form a carbamate. This carbamate, together with active-site aspartate and glutamate residues, binds an Mg^{2+} ion to form a complex capable of performing the carboxylation and oxygenation reaction.

In this paper, we describe the crystal structure of activated tobacco rubisco, cocrystallized with the reaction-intermediate analogue 2-carboxy-arabinitol 1,5-bisphosphate (CABP) (see Fig. 1 and Kinemage 1). CABP is an extremely tight-binding inhibitor, with a binding constant $\leq 10^{-11}$ M (Pierce et al., 1980).

Results

The first electron density map

The correctness of the molecular replacement solution, and the activation state of rubisco in the crystals, was checked by calculating a difference map based on an oriented, truncated model of spinach rubisco. In this truncated model, all residues that are different between spinach and tobacco rubisco had been changed to alanine. Also the carbamylated side chain of Lys 201, the Mg^{2+} ion, and the inhibitor CABP, also present in the spinach model, had been removed. This difference map was improved by averaging over the fourfold noncrystallographic symmetry (Bricogne, 1976).

The averaged $F_o - F_c$ map (Fig. 2) showed, in the absence of any model information about the parts that had been removed, clear density for the carbamylated lysine, the CABP molecule, and most of the truncated side chains. These features and the low initial R-factor of 32% confirm that the crystals contain activated rubisco, complexed with CABP, and that the molecular replacement solution is correct.

Crystal packing

Figure 3 shows how the keg-shaped rubisco hexadecamers are packed next to each other around the 3_1 axis. The L_8S_8 units are packed as in space group $P3_1$, which becomes $P3_121$ because the local fourfold axis of the hexadecamer coincides with a crystallographic twofold axis. The result is that the asymmetric unit contains half of the hexadecamer, the other half being related by crystallographic symmetry (Kinemage 1).

Crystal contacts

Two types of crystal contacts exist: (1) contacts within the hexadecamer, across the crystallographic twofold axis, and (2) "true" crystallographic contacts between different hexadecamers. The L_4S_4 unit present in the asymmetric unit has 968 contacts (atoms within 4.0 Å) within the hexadecamer and only 98 crystallographic contacts.

Only two distinct crystallographic contact regions exist; these are listed in Table 1 and highlighted in Kinemage 1. Region I links the hexadecamers around the 3_1 axis and involves residues 30, 33, 86, and 89 from the N-terminal domain; residues 439, 442, 446, 450, 451, and 452 from the barrel domain; and residues 43 and 85 from the small subunit. Region II links the hexadecamers as they are stacked on top of each other along the c-axis. Involved in these contacts are the N-terminal residues 25, 27, 30, 31, and 89. Gln 30 plays a crucial role, because it is involved in 6 out of the 13 contact residues listed in Table 1.

Although only 31 out of 472 large subunit residues are different between spinach and tobacco rubisco, as many

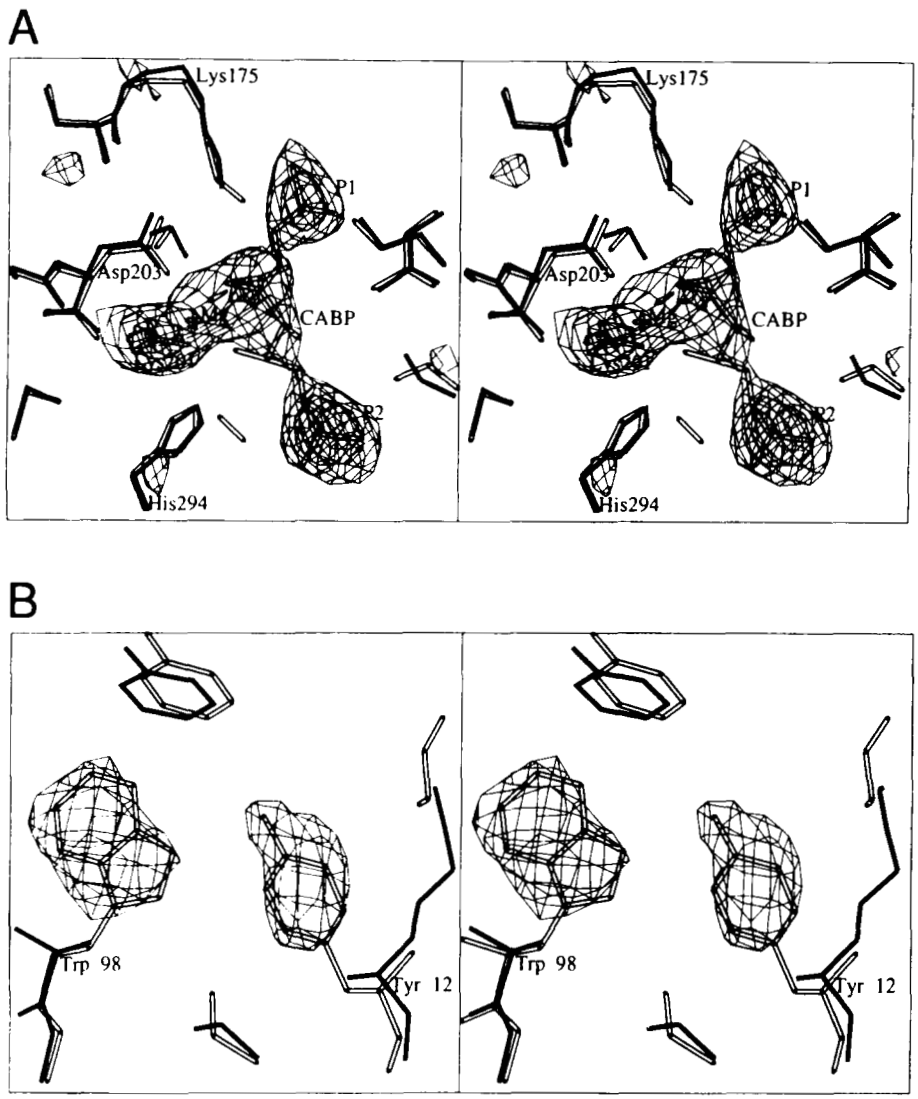

Fig. 2. Fourfold averaged $F_o - F_c$ map, based on the truncated model of spinach rubisco (see text). The truncated spinach model, used for the calculation of F_c and phases, is indicated with solid bonds. The final model of tobacco rubisco is indicated with open bonds. The map is contoured at 3σ and 6σ. **A:** Active-site region, showing the difference density of the CABP molecule. The difference density is 7.4σ near the P1 phosphate, 8.9σ near the P2 phosphate, and 6.5σ near the carbamyl carbon attached to Lys 201. **B:** Difference density of Trp 98 and Tyr 12 of the small subunit. These two side chains have been truncated because they are both phenylalanines in spinach rubisco. The peak height is 4.4σ for Tyr 12 and 4.7σ for Trp 98. The OH of Tyr 12 in the model is close enough to the NE1 of Trp 98 to be able to form a hydrogen bond.

Table 1. *Unique contacts ($d < 4.0$ Å) between L_8S_8 complexes*[a]

Contact region I		Contact region II	
Between subunits A and B		Between subunits D and D	
Gln 30	Lys 450	Tyr 25	Pro 27
	Trp 451	Pro 27	Pro 27
	Ser 452	Gln 30	Arg 89
	Pro 43 (S subunit)	Thr 31	Arg 89
Asp 33	Arg 446		
	Lys 450		
Between subunits A and A			
Arg 89	Ala 85 (small subunit)		
Between subunits C and B			
Gln 30	Asn 442		
Arg 86	Gln 439		

[a] The independent subunits are numbered A–D. They compare as follows to the numbers given by Chapman et al. (1988): A = 7, B = 1, C = 5, and D = 3.

as 4 out of the 13 contact residues mentioned above are mutated. These residues are (spinach residues within parentheses): Gln 30 (Glu), Arg 86 (His), Arg 89 (Pro), and Gln 439 (Arg). It is therefore no surprise that tobacco and spinach rubisco crystallize in different space groups, despite their great sequence similarity.

Temperature factors

The temperature factor distribution is shown in Figure 4. Except for residues 468–472 of the large subunit and the last three residues of the small subunit, no regions are present that have extremely high temperature factors like the N-terminal domain in the unactivated form of the protein (Curmi et al., 1992). Residues 1–12 of the large subunit have not been built because no clear density is present to fit them. These residues are most likely either disordered or not present due to proteolysis. In the 2.0 Å struc-

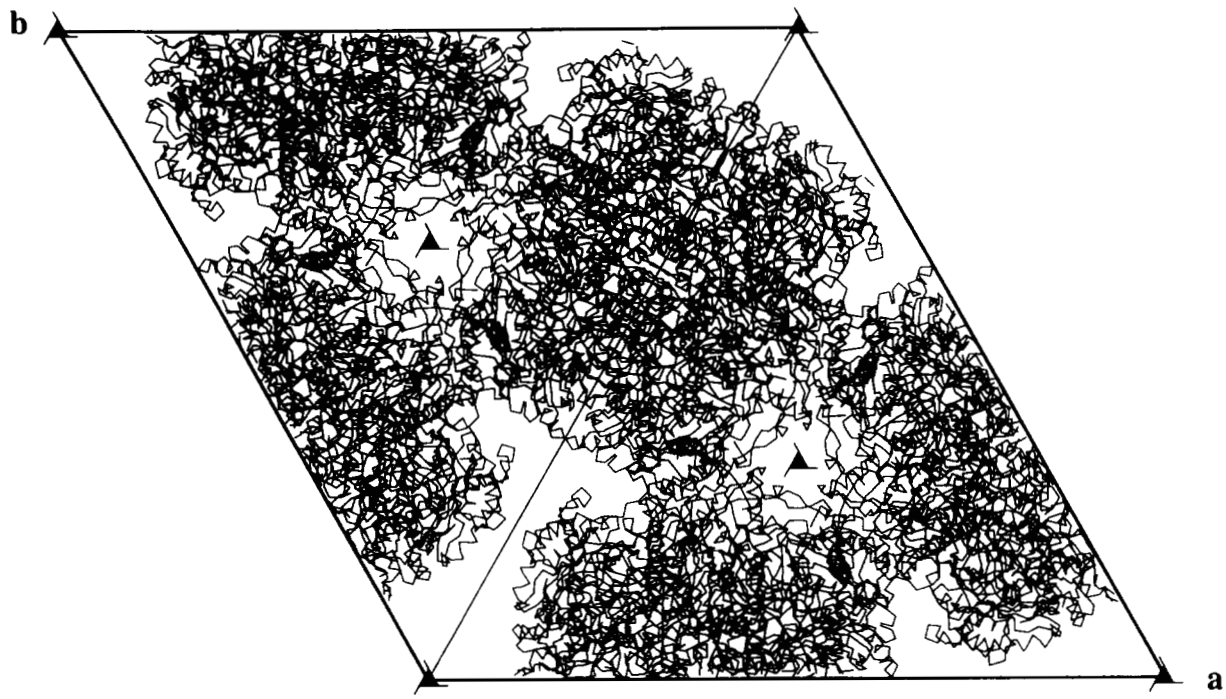

Fig. 3. Packing of crystals of activated tobacco rubisco. The view is down the *c*-axis. All molecules present in the unit cell are shown as C^{α} tracings. The next layer of molecules falls exactly on top of the layer shown. Two regions of intermolecular contacts (see text) exist: contacts around the 3_1 axis (region I) and contacts between molecules as they are stacked along the *c*-axis (region II). The triangular symbols indicate 3_1 axes.

ture of unactivated tobacco rubisco, clear density for the large subunit starts only at residue 22. Most other residues, especially in low temperature regions such as the barrel domain, are well defined.

The structure

The structure of activated tobacco rubisco is very similar to the structure of activated spinach rubisco. The root mean square (RMS) difference is 0.4 Å for 587 equivalent C^{α} atoms (Kinemage 2). Although it cannot be excluded that this similarity is in part due to the molecular replacement procedure, it would be very surprising if the proteins were different, because they share more than 90% sequence identity and have identical functions. The structure of activated spinach rubisco has been described in very great detail by Knight et al. (1990). We will not repeat this whole description here, but will instead concentrate on the effects of mutations between the tobacco and spinach enzyme.

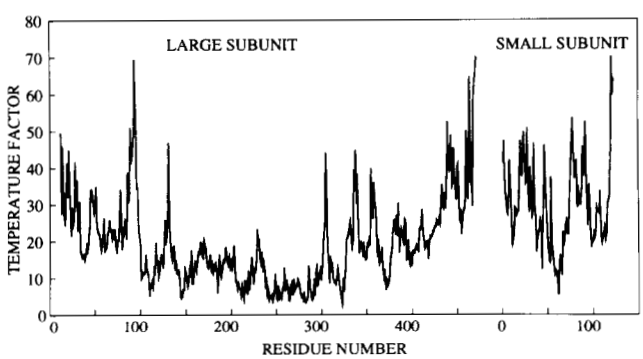

Fig. 4. Isotropic temperature factor distribution of main chain and side chain atoms, averaged per residue, in units of $Å^2$. The distributions of all four independent large and small subunits are plotted, but only a single line is visible because the distributions are so similar that they completely overlap.

The models of activated tobacco and spinach rubisco contain 61 different amino acids per LS pair. These are listed in Table 2 and shown in Figure 5. Only seven buried amino acids are mutated. The other 54 mutated amino acids are at or near the surface of the protein. Striking examples of specific mutations of surface residues are: (1) the β-peptide 466–472, where all even-numbered, solvent exposed residues are mutated, while the odd-numbered, buried residues are conserved; and (2) the helices αA (residues 23–35) and αB (residues 80–93) of the small subunit, where, again, the solvent exposed sides are mutated, while the buried sides are, with one exception, conserved (Kinemage 2).

The four buried mutations in the large subunit are not compensated by other mutations. The Ala 99 Cys and Ala 281 Ser mutations cause the difference of one SH or

Table 2. *Residues that are different between tobacco and spinach rubisco*

Sequence number	Tobacco	Spinach	Acc. area[a]	Comments
				A: Large subunit
19	Glu	Asp	99	Surface
30	Gln	Glu	160	Surface
32	Lys	Leu	83	Surface; Lys 32 interacts with the side chains of Asp 35 and Tyr 29
43	Thr	Ser	13	Surface
76	Ser	Asn	21	Surface
86	Arg	His	95	Surface
89	Arg	Pro	189	Surface
91	Val	Ala	143	Surface
94	Lys	Glu	165	Surface; Lys 94 interacts with side chain of Glu 93; the Glu 94 side chain in spinach rubisco points in the opposite direction
95	Asp	Asn	73	Surface
99	Ala	Cys	1	BURIED; hardly any changes!
142	Pro	Val	27	Surface
249	Glu	Asp	46	Surface
251	Ile	Met	37	Surface
279	Ser	Thr	1	Near surface
281	Ala	Ser	0	BURIED; Pro 151 moves 0.4 Å closer to Ala 281
309	Ile	Met	0	BURIED
320	Met	Leu	2	BURIED; no changes
353	Phe	Tyr	100	Surface
354	Val	Thr	6	Near surface
356	Gln	Lys	105	Surface
367	Asp	Ser	50	Surface
371	Leu	Thr	17	Hardly any changes
429	Lys	Gln	10	Near surface; Lys 429 interacts with Glu 433 and small subunit Glu 29
439	Gln	Arg	168	Surface; Arg 439 interacts with Glu 440 in spinach rubisco
443	Glu	Thr	108	Surface
449	Cys	Thr	2	Near surface; Cys 449 allows disulfide bond with Cys 459
466	Val	Lys	102	Surface; solvent-exposed side of β-peptide[b]
468	Asn	Glu	139	Surface; solvent-exposed side of β-peptide
470	Ala	Pro	87	Surface; solvent-exposed side of β-peptide
472	Val	Met	92	Surface
				B: Small subunit
6	Pro	Ile	16	Surface
7	Ile	Leu	42	Surface
8	Asn	Gly	146	Surface
9	Lys	Leu	85	Surface
12	Tyr	Phe	30	BURIED; Tyr OH interacts with NE1 of Trp 98, which is Phe 98 in spinach rubisco
20	Asp	Pro	128	Surface; Asp 20 interacts with Lys 119
22	Ser	Thr	49	Surface
23	Gln	Thr	132	Surface
28	Ser	Ala	47	Surface; solvent-exposed side of α-helix
31	Glu	Asn	80	Surface; solvent-exposed side of α-helix
35	Lys	Val	145	Surface; solvent-exposed side of α-helix
36	Asn	Lys	73	Surface; Lys 35/Asn 36 pair replaced by the Val 35/Lys 36 pair
39	Val	Ile	15	Surface
46	Thr	Val	27	Surface; OG of Thr 46 points toward the solvent
47	Glu	Lys	125	Surface
48	His	Asp	69	Surface; Glu 47/His 48 pair replaced by the Lys 47/Asp 48 pair
55	Asn	His	25	Surface
56	Asn	Asp	23	Surface; Asn 55/Asn 56 pair replaced by the His 55/Asp 56 pair
80	Ala	Pro	8	Near surface; solvent-exposed side of α-helix
81	Thr	Ala	77	Surface; solvent-exposed side of α-helix
84	Leu	Val	16	Near surface; solvent-exposed side of α-helix
85	Ala	Asn	60	Surface; solvent-exposed side of α-helix
88	Gly	Glu	33	Surface; solvent-exposed side of α-helix
90	Ala	Val	0	BURIED; compensated by Ile 99 Val mutation in spinach rubisco
96	Gln	Asp	81	Surface
98	Trp	Phe	15	Near surface; Trp 98–Tyr 12 interaction becomes Phe 98–Phe 12 interaction in spinach rubisco
99	Ile	Val	0	BURIED; compensates Ala 90 Val mutation
107	Val	Lys	87	Surface
109	Gln	Glu	14	Surface
121	Glu	Ala	189	Surface

[a] Solvent-accessible surface area (in Å^2) as calculated with the program DSSP (Kabsch & Sander, 1983). The values listed are calculated on the basis of the complete L_8S_8 complex and are the average of the four independent subunits.

[b] Extended piece of polypeptide chain with ϕ,ψ angles in the β-region ($-161° \leq \phi \leq -57°$, $59° \leq \psi \leq 165°$).

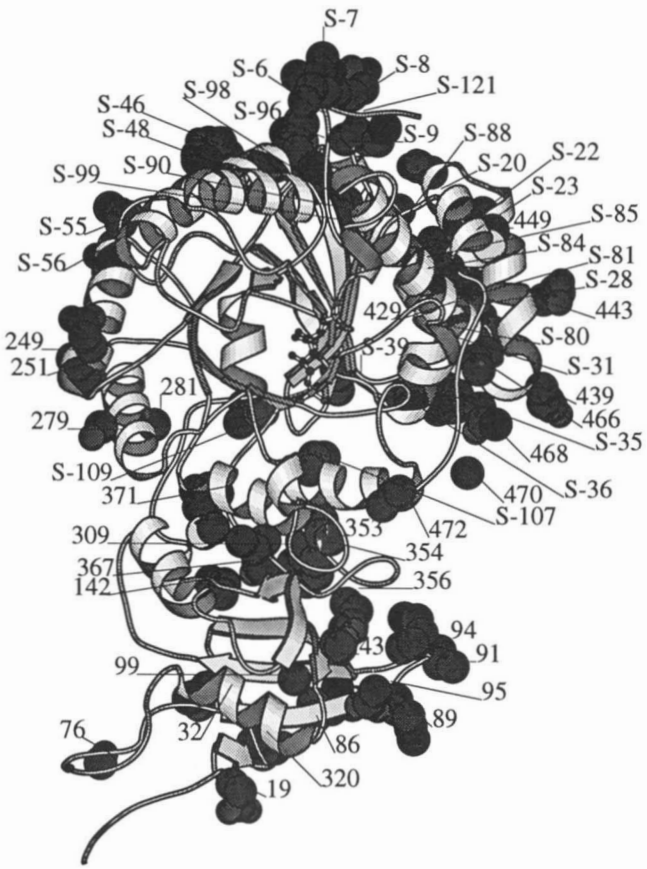

Fig. 5. Ribbon diagram of one large and one small subunit of activated tobacco rubisco. Below in the figure is the N-terminal domain of the large subunit, in the center is the barrel domain, and in the upper right part is the small subunit. Residues that are different between spinach and tobacco rubisco are drawn as space-filling models. The figure shows that the mutations are concentrated at the surface of the molecule and at the solvent-exposed interface between the N-terminal domain and the barrel domain. The figure was produced using the program MOLSCRIPT (Kraulis, 1991).

OH group, respectively. These kinds of differences can apparently be accommodated by the protein without significant changes. The Ile 309 Met and Met 320 Leu mutations do not change the number of side chain atoms, but the presence of the sulfur leads to a somewhat larger volume for the methionine side chain. The differences in shape and small differences in volume do not seem to disturb the folding of the protein.

Two of these buried substitutions have been identified with the C4 to C3 transition in higher plants. In a study of closely related C3 and C4 species from three genera (Hudson et al., 1990), the only common substitution was Met 309 Ile (C3 to C4). A second substitution, Ala 281 Ser (C3 to C4) was found in two of the three genera.

Residue 309 is well situated to affect the kinetic parameters of rubisco. It is located near the interface between the two large subunits that form a pair of active sites (Kinemage 3). Its side chain is in a hydrophobic pocket that is largely made up of conserved residues (Val 300, Met #297 [residue numbers prefixed with a # refer to the dimer-related subunit], Phe 125, Val 121, Leu 135, Phe 117, Ile 301, Val 313, and Leu 314). In particular, it makes van der Waals contact with Val 121 and contacts Phe 125 via Val #300. These residues close a loop that holds Asn 123 in the symmetry-related active site. Asn 123 makes contact with the metal in the activated state (Lundqvist & Schneider, 1991) and with the transition state analogue, CABP, in the quaternary complex (Knight et al., 1990). Mutations of Asn 123 in *R. rubrum* (Asn 111) affect both the kinetic parameters and the CO_2/O_2 specificity (Chène et al., 1992). Residue 309 is also in van der Waals contact with Met #297 and Ile 301. These residues are proximal to His 298 and Arg 295, which form the P2 binding site (note: residue 309 is thus able to affect both P2 sites in the dimer). Therefore, residue 309 appears to be in a prime position to alter the enzyme characteristics.

A superposition of the spinach and tobacco coordinates using backbone plus CB atoms of residue 309 and its neighbors (Phe 117, Val 121, Gly 122, Phe 125, Leu 135, Ile 301, Val 313, and Leu 314) was used to examine the structural effects of this substitution. A comparison of the two structures shows very little change (RMS deviation of 0.24 Å for the 44 atoms used in the superposition). The loop containing Asn 123 is essentially unaltered. The methionine and isoleucine residues at position 309 closely follow each other (Fig. 6). Met CG lies close to Ile CG1 (0.6 Å), and Met CE is near Ile CD (0.7 Å). The Met S lies between the two Ile CG atoms (1.9 Å and 2.2 Å from the CGs, respectively). The arrangement of these alternative side chains preserves the sides of the pocket that interact with the active site residues. The position of the Met S is accommodated by a side chain rotation of Phe 117, which in turn is accommodated by rotations of Phe 274 and Tyr 269. In the presence of Ile 309, the terminal CE of Met #297 is moved away from Ile 309 by a rotation about the methionine CG–SD bond. Thus, the structural impact of the Met 309 Ile substitution is either absorbed in local adjustments or directed away from the proximal active sites.

An examination of the Ala 281 Ser site shows even fewer alterations. In spinach rubisco, the Ser 281 side chain makes hydrogen bonds with the carbonyl of Ser 321 and the side chain of Asn 149. There are no apparent changes in the tobacco structure to compensate for the loss of the OH moiety. It is possible that it is replaced by a water molecule, as residue 281 is near the surface of the internal cavity in the L_8S_8 hexadecamer (no explicit water molecules have been fit to the tobacco map due to its limited resolution).

In summary, the comparison of the spinach and tobacco structures fails to explain why Met 309 Ile and Ala 281 Ser appear to be linked to the kinetic changes observed in related C3 and C4 species. In fact, the protein appears to accommodate these substitutions in a way that minimizes their effect on the neighboring active sites.

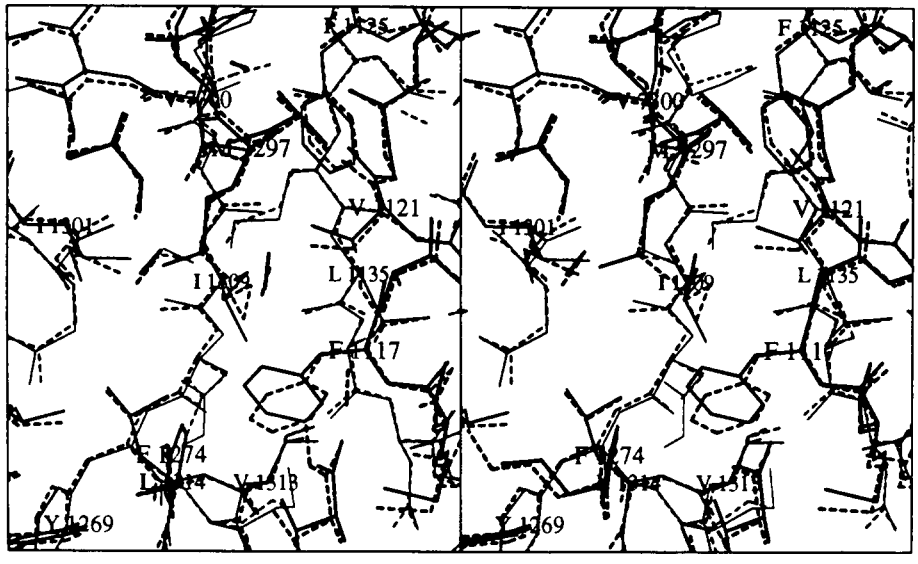

Fig. 6. Stereo diagram of the hydrophobic pocket occupied by residue 309. The continuous lines represent tobacco coordinates, and the broken lines represent spinach. The two residues at position 309 (Met and Ile) are configured so as to occupy essentially the same space. The largest changes observed are the rotation of the terminal side chain bond in Met #297 (residue numbers prefixed with a # refer to the dimer-related subunit) and the rotation of Phe 117. The latter cascades into rotations of Phe 274 and Tyr 269 as seen in the lower left foreground. The labels refer to tobacco coordinates with the first digit (1 or 7) as a chain identifier, chain 7 being dimer-related to chain 1. The figure was produced using the program MOLSCRIPT (Kraulis, 1991).

In the small subunit, the buried mutations are compensated. The Tyr 12 Phe mutation is compensated by the Trp 98 Phe mutation, replacing a polar Tyr–Trp interaction with a hydrophobic Phe–Phe interaction. The Ala 90 Val mutation is compensated by the Ile 99 Val mutation, because the total volume of the 90/99 pair stays approximately constant.

None of the 123 residues within 15 Å of the Mg^{2+} ion (the active site region) are different. The RMS difference after superposition of these residues is 0.4 Å for all 946 atoms and 0.2 Å for the 123 C^{α} atoms. These differences are of the same magnitude as the expected mean coordinate error of 0.35 Å and show that the active sites of tobacco and spinach rubisco are identical.

Disulfide links

Tobacco rubisco contains three unique pairs of cysteines close enough to potentially form disulfide links. A minimum contact distance of 2.0 Å between sulfur atoms was used during refinement. In this way the refinement programs were free to position the sulfur atoms, where dictated by the data, without enforcing either covalent links or van der Waals distances. Table 3 shows that, within experimental error, the SG atoms of the dimer-related Cys 247 residues and of the Cys 449–Cys 459 pair are at covalent distance, and the SG atoms of the Cys 172–Cys 192 pair are at van der Waals distance (Kinemage 1). In spinach rubisco, the same disulfide pattern has been observed for the first two cysteine pairs: Cys 247 is linked to the dimer-related Cys 247, while the Cys 172–Cys 192 pair does not form a disulfide link. A third disulfide link is not possible because in spinach rubisco residue 449 is a threonine and not a cysteine.

Table 3. *Cysteine SG atoms closer than 5.0 Å*

Residue numbers		SG–SG distance (Å)			
		Subunit A	Subunit B	Subunit C	Subunit D
172	192	3.6	3.5	3.5	3.3
247	247	2.0	1.7	1.7	2.0
449	459	2.3	2.1	2.1	2.1

Discussion

The results described in this paper clearly establish that the activated forms of tobacco rubisco and spinach rubisco are very similar indeed, and that the active sites are virtually identical. It is therefore very likely that these models will also be relevant to many other higher plant rubiscos.

The mutations between tobacco and spinach rubisco are either conservative or at the surface of the protein. Only 7 out of 61 mutations involve completely buried residues. This observation is in line with results of Dao-Pin et al. (1991), who showed that even replacing charged residues at the surface of T4 lysozyme hardly affected the stability and activity of this enzyme.

Several chloroplast enzymes, such as malate dehydrogenase and glucose-6-phosphate dehydrogenase, are regulated by light via oxidation/reduction of SH groups (Scheibe & Anderson, 1981). The SH groups are reduced under light conditions and become oxidized in the dark. Although rubisco is not regulated this way (Campbell & Ogren, 1990), it is conceivable that the same regulatory mechanism causes disulfide links to be formed in the dark that are of no consequence for the catalytic activity. Ranty et al. (1991) showed that the disulfide link present between dimer-related Cys 247 residues does not influence

the catalytic activity. This disulfide link is present in the crystal structure of spinach rubisco (Knight et al., 1990), *Synechococcus* rubisco (Newman & Gutteridge, 1990), and tobacco rubisco.

The pattern of disulfide bonds found in activated tobacco rubisco (Cys 449–Cys 459 linked, and the SGs of Cys 172 and Cys 192 at van der Waals distance) is quite interesting, because in unactivated rubisco (Curmi et al., 1992), Cys 172–Cys 192 were found to be linked, whereas Cys 449–Cys 459 were found to be at van der Waals distance. The Cys 247–Cys 247 link is present in both crystal structures. Different disulfide links are present in the two crystal forms. These differences can therefore not be attributed to general differences in crystallization conditions, like the redox potential. It is not yet clear whether these observed differences are an artifact of the preparation of enzyme or crystals, or whether they are linked to the activation state of the protein. The cysteines mentioned are not conserved in all rubisco sequences, so they are not essential for the activity of rubisco.

Materials and methods

Purification and crystallization

Rubisco from *Nicotiana tabacum* var. Turkish samsun was purified and crystallized by the hanging drop as described by Se Won Suh et al. (1987). Difficulties in reproducing the crystals forced us to use slight variations on the published methods. The crystals eventually used for data collection at the synchrotron were obtained under the following conditions: 0.692 mL of a solution of 28 mg/mL protein with an E_{280}/E_{260} ratio of 1.9 in activation buffer, containing 50 mM tris-HCl, pH 8.38, 50 mM sodium chloride, 1 mM sodium azide, 10 mM magnesium chloride, and 20 mM sodium bicarbonate, was heated for 30–60 min to 37 °C in order to activate the protein. Then 28 μL 0.1 M CABP solution was added, followed by the addition of 80 μL 10% (w/v) β-octyl glucoside in water and 480 μL 13% (w/v) PEG 8000. Drops of 15 μL of this solution were placed on siliconized coverslips. The wells contained 1 mL of 16–17% (w/v) PEG 8000 in activation buffer. Because tobacco rubisco is known to deactivate at lower temperature (Kawashima et al., 1971) and also because better crystals were obtained at elevated temperatures, the crystallizations were performed at 30–35 °C in incubators. It was nevertheless extremely difficult to reproduce the crystals (the best results were obtained after mild earthquakes in the Los Angeles area). From 120 trays with 24 drops each, only 2 trays contained a few wells with crystals of sufficient quality for data collection. These crystals grew in 2–3 weeks and were hexagonal rods 1.5–2.0 mm long and 0.3–0.5 mm thick. The space group was $P3_121$ with $a = b = 204.6$ Å and $c = 117.4$ Å. The asymmetric unit contains four large and four small subunits, with total molecular mass of 280 kDa.

Data collection

Two data sets were used in the structure determination. The first data set was collected on a crystal grown under conditions as published by Suh et al. (1987) using a Xuong-Hamlin multiwire area detector (Hamlin, 1985) and Cu K-α radiation from a Rigaku rotating anode generator. Data collection and processing software was written by Larry Weissmann (unpubl.). The final data set was 61.8% complete to 5.0 Å with an R_{merge} of 10.0% on intensities.

The second data set was collected on Kodak DEF film at the NSLS in Brookhaven, using 1.1-Å radiation. The crystals used were grown as described in this paper. Two crystals were used for data collection. The long dimension (1.5–2 mm) of the rod-shaped crystals coincides with the crystallographic *c*-axis. After mounting, this axis is parallel to the capillary axis and therefore parallel to the spindle axis. The length of the crystal allowed 10–15 translations of 0.1–0.2 mm, parallel to the spindle axis. In total, 147 oscillation photographs of 0.7° were taken.

The films were processed with the MOSFLM package (Nyborg & Wonacott, 1977). Each film was individually indexed using Kabsch's (1988) autoindexing program. However, the large cell dimensions and sometimes weak films caused the program to misindex quite frequently. H.A.S. wrote a small program to check whether the orientation matrix of a given film closely matched the orientation of the previous film from a consecutive series of oscillation pictures. When necessary, the program applied 60° rotations until a match with the previous orientation matrix was obtained. This procedure ensured the same choice of axes for all films. As a further check on the indexing, the fulls of the raw data from each film were scaled against the 5.0-Å detector data set. Films that turned out to be misindexed were indexed again using a different threshold for the selection of spots. Finally, 13 films had to be discarded because of indexing problems or too-high *R*-sym values.

The orientation matrices were further refined using the Winkler et al. (1979) method of postrefinement. Scaling and merging of the data were done according to Fox and Holmes (1966). Reflections with a partiality greater than 0.6 were scaled to full intensity. Only observations greater than 1σ were used. The resolution per film varied from 2.6 Å for fresh crystals to 3.2 Å at the end of a run. The final data set contains 59,608 reflections, based on 206,055 observations, and is 76.7% complete to 2.7 Å. The *R*-sym on intensities is 10.6%.

Molecular replacement

The structure of rubisco form IV crystals was solved by molecular replacement. The process was started using the 5.0-Å detector data and the model of form III tobacco rubisco, as reported by Chapman et al. (1988).

Rotation functions

The orientation of the search model in the form IV crystals was found using the local implementation of Crowther's (1972) fast rotation function, by searching for the 422 symmetry (Baker et al., 1977) of the L_8S_8 tobacco rubisco molecule. Self-rotation functions were calculated using different resolution ranges and different integration ranges. The best results were obtained using 5–12-Å data and an integration range of 30 Å. Self-rotation functions were plotted in polar coordinates and showed strong peaks in the $\kappa = 90°$ section parallel to the a-axis and b-axis, repeated every 60°. This indicates that the local fourfold axis runs parallel to the a- and b-axes. The $\kappa = 180°$ section showed two sets of peaks, repeated every 45°, representing two possible solutions. One set consisted of peaks at ψ equal to 0°, 45°, 90°, etc.; the other set consisted of peaks at ψ equal to 19°, 63°, −27°, and −71°.

The cross-rotation function, using an L_8S_8 search model based on the structure of nonactivated tobacco rubisco, revealed a strong peak in the $\beta = 90°$ section at $\alpha = 60°$ and $\gamma = 17°$ plus all peaks related by the crystallographic symmetry and by the 422 symmetry of the search model, supporting the second set of solutions of the self-rotation function. The orientation of the model was optimized by comparing the self-rotation function of the oriented model with the self-rotation function of the observed data and by adjusting the orientation of the model to get an optimal match.

Positioning the search model in the unit cell

Suh et al. (1987) concluded on the basis of packing considerations that the asymmetric unit of form IV crystals contained half of the L_8S_8 complex. This means that an internal symmetry axis of the L_8S_8 molecule coincides with a crystallographic twofold axis, and that the center of the L_8S_8 rubisco molecule lies on this crystallographic twofold axis. A packing program using spheres with radii of 40 Å and 50 Å revealed only one possible solution with the center of the L_8S_8 molecule on a crystallographic twofold axis, at $\frac{1}{3}$ or $\frac{2}{3}$ from the origin, depending of the choice of axes.

However, the classical translation function of Crowther and Blow (1967) failed to produce a clear solution. One-dimensional translation searches were done by translating the full L_8S_8 model along the x-axis in the lower symmetry space groups $P3_1$ and $P3_2$. The best solutions were further refined by a rotational R-factor search at the best position from the translation search. The lowest R-factor of 47.2% was obtained in space group $P3_1$. The best solution in $P3_2$ had an R-factor of 49.0%, which is only slightly worse than the $P3_1$ solution.

The searches were repeated with the program BRUTE (Fujinaga & Read, 1987), because the correlation coefficients used in this program give a much improved signal-to-noise ratio. The large number of atoms (>18,000) forced us to add FFT routines to the program in order to be able to do the calculation in a reasonable time. The first translation search was done using 1.0-Å steps. The orientation and position were subsequently refined by applying 0.1° rotation steps and 0.1-Å translation steps.

A clear maximum correlation of 0.24, 6σ above the next highest peak, was obtained with BRUTE at the same position as previously found in the R-factor search. However, the R-factor was still 48% and the 5-Å electron density maps were not readily interpretable. Rigid body refinements did not give a significant improvement. Higher resolution synchrotron data were therefore collected. The 2.7-Å synchrotron data gave an improved correlation coefficient of 0.32, but the R-factor was close to 50% and the electron density maps were difficult to interpret.

The reason for these problems became immediately clear when a paper by Knight et al. (1989) appeared, indicating that the model we were using contained a serious error in the tracing of the small subunit. The structure of activated spinach rubisco as a search model gave the same solution as before, but the maximum correlation was 0.79 and the R-factor was 32%, and the electron density map was easily interpretable.

Refinement

The starting model was rebuilt on an Evans and Sutherland PS390 system, running FRODO software (Jones, 1985), using fourfold averaged $2F_o - F_c$ and $F_o - F_c$ maps. The program used to do the averaging was written by H.A.S. and takes the complete map into (virtual) memory, thereby eliminating the tedious double sorting techniques of Bricogne's (1976) original program, which had been written for computers with very limited memories. The maps, based on a truncated model of spinach rubisco, showed clear density for most of the left-out side chains. It is unlikely that any major differences between tobacco and spinach rubisco would not have shown up in this map.

Subsequently, the model was refined with the program XPLOR (Brünger et al., 1987), imposing strict noncrystallographic symmetry. This refinement consisted of one cycle of positional refinement, reducing the R-factor from 32.5% to 27.4%, followed by molecular dynamics refinement, first heating the molecule to 1,400 K and slowly cooling it back to 300 K. The R-factor was at this stage reduced to 26.4%. Positional refinement reduced the R-factor to 26.0%.

The model was inspected and rebuilt, and 15 cycles of PROFFT refinement (Hendrickson, 1985; Finzel, 1987; Sheriff, 1987) interrupted with rebuilding were done with four independent large and small subunits. The noncrystallographic symmetry was tightly restrained, both for positions and for temperature factors. We did not find any indication that the noncrystallographic symmetry matrices, as obtained by rotation/translation searches with steps of 0.1° and 0.1 Å, respectively, needed to be up-

Table 4. *Refinement statistics of activated tobacco rubisco (form IV)*[a]

Statistic	Value	Target σ
R[b]	21.0	
Reflections used	57,855	
Distance		
Bond distance	0.021	0.020
Angle distance	0.050	0.030
Planar 1–4 distance	0.054	0.040
Planes	0.016	0.020
Chiral volumes	0.177	0.150
Nonbonded contacts		
Single torsion contacts	0.221	0.500
Multiple torsion contacts	0.265	0.500
Possible hydrogen bonds	0.230	0.500
Torsion angles		
Planar	3.1	3.0
Staggered	23.7	15.0
Orthonormal	30.3	20.0
Noncrystallographic symmetry[c]		
RMS delta L_A	0.119	0.050
RMS delta L_B	0.118	0.050
RMS delta L_C	0.120	0.050
RMS delta L_D	0.131	0.050
RMS delta S_A	0.110	0.050
RMS delta S_B	0.097	0.050
RMS delta S_C	0.119	0.050
RMS delta S_D	0.112	0.050
Thermal parameter correlations		
Main chain bond	0.870	1.000
Main chain angle	1.546	1.500
Side chain bond	2.313	2.000
Side chain angle	3.835	3.000

[a] Values are root mean square differences from ideal values (in Å for distances, in Å^2 for thermal parameter correlations, and in Å^3 for chiral volumes).

[b] $$R = \frac{\Sigma \big| |F_{obs}| - |F_{calc}| \big|}{\Sigma |F_{obs}|} \times 100\%.$$

[c] Root mean square deviations from the average structure. L_A, first large subunit; S_A, first small subunit; etc.

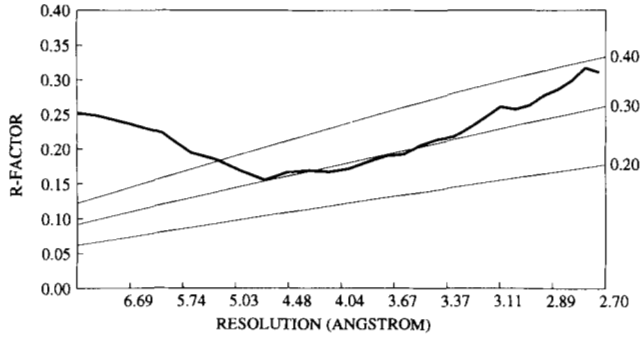

Fig. 7. Luzzati plot indicating a mean coordinate error of about 0.35 Å. The thick line shows the R-factor as a function of resolution, and the thin lines show theoretical curves corresponding to mean coordinate errors of 0.2, 0.3, and 0.4 Å (Luzzati, 1952). The R-factor increases at low resolution because no correction has been made for the scattering of the bulk solvent.

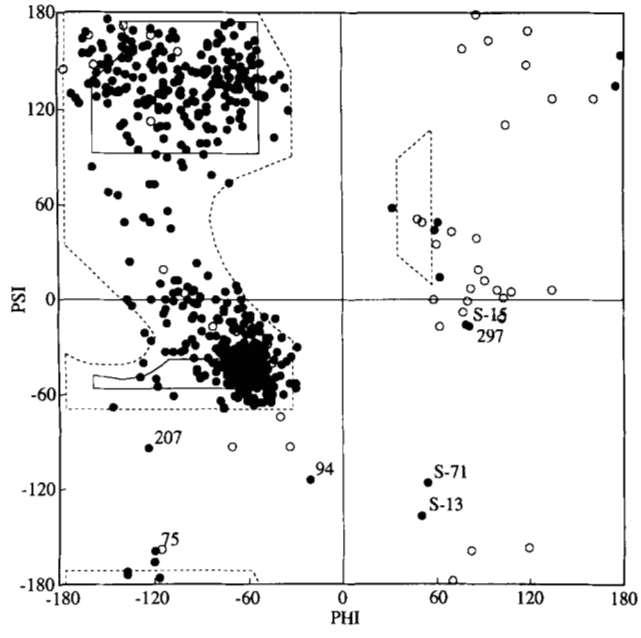

Fig. 8. Plot of the main chain torsion angles ϕ, ψ in the model of activated tobacco rubisco complexed with CABP. Allowed regions according to Ramachandran and Sasisekharan (1968) are indicated by broken lines. Open circles indicate glycines; filled circles indicate non-glycine residues. Non-glycine residues more than 10° outside allowed regions are labeled. These residues also have deviating ϕ, ψ angles in the 2.0 Å structure of unactivated tobacco rubisco (Curmi et al., 1992).

dated, and the original matrices were used throughout the refinement. No waters were added because of the limited resolution of the data. The final R-factor was 21.0%. The final refinement statistics are listed in Table 4. The expected mean coordinate error is 0.35 Å, as deduced from the Luzzati plot (Luzzati, 1952) shown in Figure 7. The Ramachandran plot (Fig. 8; Ramachandran & Sasisekharan, 1968) shows that seven residues per LS pair have ϕ, ψ angle combinations more than 10° outside allowed regions. These residues also have deviating ϕ, ψ angles in the 2.0-Å structure of unactivated tobacco rubisco (Curmi et al., 1992).

Acknowledgments

We thank Dr. Kam Zhang for help in the preparation of this manuscript. We thank NIH for support and the San Diego Supercomputer Center for computer time.

The coordinates have been deposited with the Protein Data Bank in Brookhaven.

References

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

Andrews, T.J. & Lorimer, G.H. (1987). Rubisco: Structure mechanisms, and prospects for improvement. In *The Biochemistry of Plants*, Vol. 10 (Hatch, M.D. & Boardman, N.K., Eds.), pp. 131–218. Academic Press, Orlando, Florida.

Baker, T.S., Eisenberg, D., & Eiserling, F. (1977). Ribulose bisphosphate carboxylase: A two layered, square-shaped molecule of symmetry 422. *Science 196*, 293–295.

Brändén, C.-I., Lindqvist, Y., & Schneider, G. (1991). Protein engineering of rubisco. *Acta Crystallogr. B 47*, 824–835.

Bricogne, G. (1976). Methods and programs for direct-space exploitation of geometric redundancies. *Acta Crystallogr. A 32*, 832–847.

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

Campbell, W.J. & Ogren, W.L. (1990). A novel role for light in the activation of ribulose bisphosphate carboxylase/oxygenase. *Plant Physiol. 92*, 110–115.

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

Chène, P., Day, A.G., & Fersht, A.R. (1992). Mutation of Asn111 of rubisco from *R. rubrum* alters the carboxylase/oxygenase specificity. *J. Mol. Biol. 225*, 891–896.

Crowther, R.A. (1972). The fast rotation function. In *The Molecular Replacement Method* (Rossmann, M.G., Ed.), pp. 173–178. Gordon & Breach, New York.

Crowther, R.A. & Blow, D.M. (1967). A method of positioning a known molecule in an unknown crystal structure. *Acta Crystallogr. 23*, 544–548.

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.

Dao-Pin, S., Söderlin, E., Baase, W.A., Wozniak, J.A., Suaer, U., & Matthews, B.W. (1991). Cumulative site-directed charge-charge replacements in bacteriophage T4 lysozyme suggest that long range electrostatic interactions contribute little to protein stability. *J. Mol. Biol. 221*, 873–887.

Finzel, B. (1987). Incorporation of fast Fourier transforms to speed restrained least-squares refinement of protein structures. *J. Appl. Crystallogr. 20*, 53–55.

Fox, G.C. & Holmes, K.C. (1966). An alternative method of solving the layer scaling equations of Hamilton, Rollett and Sparks. *Acta Crystallogr. A 20*, 886–891.

Fujinaga, M. & Read, R. (1987). Experiences with a new translation-function program. *J. Appl. Crystallogr. 20*, 517–521.

Gutteridge, S. (1990). Limitations of the primary events of CO_2 fixation in photosynthetic organisms: The structure and mechanism of rubisco. *Biochim. Biophys. Acta 1015*, 1–14.

Hamlin, R. (1985). Multiwire area X-ray diffractometers. *Methods Enzymol. 114*, 416–452.

Hendrickson, W.A. (1985). Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol. 115*, 252–270.

Hudson, G.S., Mahon, J.D., Anderson, P.A., Gibbs, M.J., Badger, M.R., Andrews, T.J., & Whitfeld, P.R. (1990). Comparisons of rbcl genes for the large subunit of ribulose-bisphosphate carboxylase from closely related C3 and C4 plant species. *J. Biol. Chem. 256*, 808–814.

Jones, T.A. (1985). Interactive computer graphics: FRODO. *Methods Enzymol. 115*, 157–171.

Kabsch, W. (1988). Automatic indexing of rotation diffraction patterns. *J. Appl. Crystallogr. 21*, 67–71.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 2577–2637.

Kawashima, N., Singh, S., & Wildman, S.G. (1971). Reversible cold inactivation and heat reactivation of RuDP carboxylase activity of crystallized tobacco fraction I protein. *Biochem. Biophys. Res. Commun. 42*, 664–668.

Knight, S., Andersson, I., & Brändén, C.-I. (1989). Reexamination of the three-dimensional structure of the small subunit of rubisco from higher plants. *Science 244*, 702–705.

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

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

Lorimer, G.H. (1981). Ribulose bisphosphate carboxylase: Amino acid sequence of a peptide bearing the activator carbon dioxide. *Biochemistry 20*, 1236–1240.

Lorimer, G.H., Badger, M.R., & Andrews, T.J. (1976). The activation of ribulose-1,5-bisphosphate carboxylase by carbon dioxide and magnesium ions. Equilibria, kinetics, a suggested mechanism, and physiological implications. *Biochemistry 15*, 529–536.

Lundqvist, T. & Schneider, G. (1991). Crystal structure of the ternary complex of ribulose-1,5-bisphosphate carboxylase, Mg(II) and activator CO_2 at 2.3 Å resolution. *Biochemistry 30*, 904–908.

Luzzati, V. (1952). Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallogr. 5*, 802–810.

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.

Newman, J. & Gutteridge, S. (1990). The purification and preliminary X-ray diffraction studies of recombinant *Synechococcus* ribulose-1,5-bisphosphate carboxylase/oxygenase from *Escherichia coli*. *J. Biol. Chem. 265*, 15154–15159.

Nyborg, J. & Wonacott, A.J. (1977). Computer programs. In *The Rotation Method in Crystallography* (Arndt, U.W. & Wonacott, A.J., Eds.), pp. 139–151. North-Holland Publishing Company, Amsterdam.

Pal, G.P., Jocab, R., Hahn, U., Bowien, B., & Saenger, W. (1985). Single and twinned crystals of ribulose-1,5-bisphosphate carboxylase-oxygenase from *Alcaligenes eutrophus*. *J. Biol. Chem. 260*, 10768–10770.

Pierce, J., Tolbert, N.E., & Barker, R. (1980). Interaction of ribulose bisphosphate carboxylase/oxygenase with transition-state analogues. *Biochemistry 19*, 934–942.

Ramachandran, G.N. & Sasisekharan, V. (1968). Allowed conformations of polypeptide chains. *Adv. Protein Chem. 23*, 325–342.

Ranty, B., Lorimer, G., & Gutteridge, S. (1991). An intra-dimeric cross-link of large subunits of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase is formed by oxidation of cysteine 247. *Eur. J. Biochem. 200*, 353–358.

Scheibe, R. & Anderson, L.E. (1981). Dark modulation of NADP-dependent malate dehydrogenase and glucose-6-phosphate dehydrogenase in the chloroplast. *Biochem. Biophys. Acta 636*, 58–64.

Schneider, G., Knight, S., Andersson, I., Brändén, C.-I., Lindqvist, I., & Lundqvist, Y. (1990). Comparison of the crystal structures of L_2 and L_8S_8 rubisco suggests a functional role for the small subunit. *EMBO J. 9*, 2045–2050.

Schneider, G., Lindqvist, Y., Brändén, C.-I., & Lorimer, G. (1986). Three-dimensional structure of ribulose-1,5-bisphosphate carboxylase/oxygenase from *Rhodospirillum rubrum* at 2.9 Å resolution. *EMBO J. 5*, 3409–3415.

Sheriff, S. (1987). Addition of symmetry-related contact restraints to PROTEIN and PROLSQ. *J. Appl. Crystallogr. 20*, 55–57.

Suh, S.W., Cascio, D., Chapman, M.S., & Eisenberg, D. (1987). A crystal form of ribulose-1,5-bisphosphate carboxylase/oxygenase from *Nicotiana tabacum* in the activated state. *J. Mol. Biol. 197*, 363–365.

Winkler, F.K., Schutt, C.E., & Harrison, S.C. (1979). The oscillation method for crystals with very large unit cells. *Acta Crystallogr. A 35*, 901–911.