

Tertiary Structure of Plant RuBisCO: Domains and Their Contacts

MICHAEL S. CHAPMAN,* SE WON SUH,† PAUL M. G. CURMI,
DUILIO CASCIO, WARD W. SMITH,‡ DAVID S. EISENBERG§

The three-dimensional structure of ribulose-1,5-biphosphate carboxylase-oxygenase (RuBisCO), has been determined at 2.6 Å resolution. This enzyme initiates photosynthesis by combining carbon dioxide with ribulose biphosphate to form two molecules of 3-phosphoglycerate. In plants, RuBisCO is built from eight large (L) and eight small (S) polypeptide chains, or subunits. Both S chains and the NH₂-terminal domain (N) of L are antiparallel β, “open-face-sandwich” domains with four-stranded β sheets and flanking α helices. The main domain (B) of L is an α/β barrel containing most of the catalytic residues. The active site is in a pocket at the opening of the barrel that is partly covered by the N domain of a neighboring L chain. The domain contacts of the molecule and its conserved residues are discussed in terms of this structure.

STRUCTURAL STUDIES OF RUBISCO have been motivated by its importance in photosynthesis and its massive terrestrial biosynthesis, estimated at 4×10^{13} g year⁻¹ (1), or more than a million grams every second. Work on tobacco RuBisCO started with the definition of the molecular symmetry and L₈S₈ stoichiometry (2), continued with studies of the molecular shape (3), and more recently reached near atomic resolution that permitted conclusions on the quaternary structure (4). These conclusions are that S subunits cluster as two tetramers, one near the top and the other near the bottom of the molecule, and that L subunits are elongated and bridge between the two clusters of S subunits. Refinement of the structure has now progressed to permit the first analysis of the tertiary structure and detailed domain contacts in plant RuBisCO. The present structure is of “unactivated” RuBisCO, in the absence of the CO₂ activator. Earlier, the backbone structure of the unactivated bacterial *Rhodospirillum rubrum* RuBisCO, which has only two L subunits and no S chains, was reported (5). Other x-ray studies of RuBisCO molecules under way include activated tobacco RuBisCO in the presence of a transition-state analog (6), spinach RuBisCO, both complexed with a transition-state analog (7) and without (8), and bacterial RuBisCOs from *Alcaligenes eutrophus* (9) and *Chromatium* (10).

Much attention has also been focused on the stability and assembly of plant RuBis-

COs (11, 12). Our structure offers a preliminary opportunity for a detailed examination of the organization of the RuBisCO molecule. To interpret the involved pattern of interdomain contacts, we have computed the surface area of each domain that is shielded from the solvent by other domains (13, 14). These areas effectively define the residues in neighboring domains that are near each amino acid side chain of the protomer. This information can be used to interpret RuBisCO stability and is also helpful in understanding patterns of conserved amino acid residues.

The structure was determined by x-ray crystallographic methods, including isomorphous replacement, solvent-flattening, atomic model-building, and refinement

(15). Published amino acid sequences for the S (16) and L (17) chains were fit into OMIT (18) electron density maps with the program FRODO (19). In OMIT maps, the electron density of each region is computed in the usual way, except that the atoms of that region are omitted from the phasing model. Thus a match of the atomic model to the electron density is a stringent test of the adequacy of the model (Fig. 1).

Our current Model 7 includes all 4724 non-hydrogen atoms of the 596 amino acid side chains, and for 2σ data, has an R-factor of 28% at 2.6 Å resolution (refined coordinates to be submitted to the Brookhaven Data Bank). The fit is generally good; however, a few segments run through poor OMIT electron density. These include in S subunits, residues 1 to 6, 38 to 42, and 114 to 123 at the COOH-terminus; and in L subunits, residues 1 to 4 that probably have been posttranslationally removed (17), residues 5 to 19, residues 146 to 150 in the connector between the N and B domains, residues 302 to 309 in the loop between β strand 5 and helix 5 of the barrel, and residues 402 to 405 in the loop between β strand 8 and helix 8 of the barrel. Some minor changes in the model, particularly in these segments, can be expected as refinement proceeds, but the overall model is likely to remain unchanged. These poor segments (other than residues 1 to 4 of L that are probably posttranslationally cleaved) are retained in the model to restrain it during refinement. Omitting the worst segments (residues 146 to 150 in L and 1 to

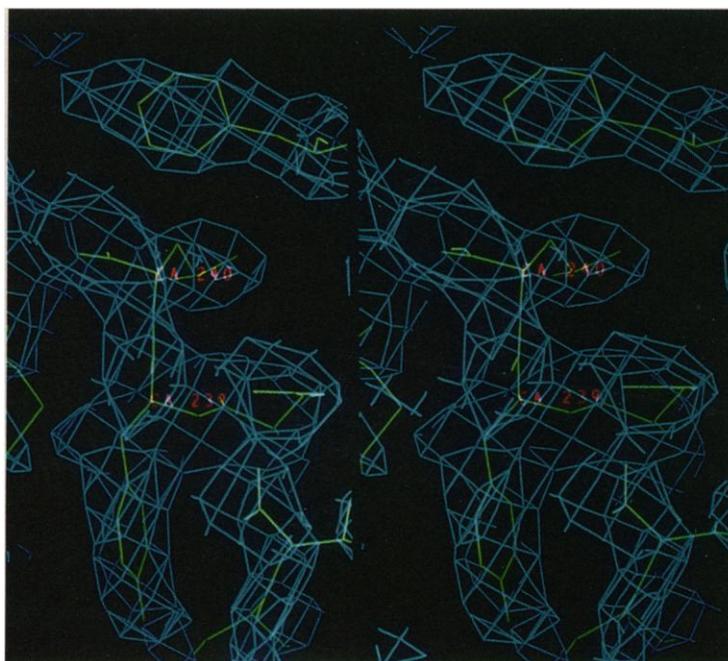


Fig. 1. An example of the fit of the RuBisCO atomic model to an OMIT electron density map, shown in stereo. The protein model is in green and the OMIT density is in blue, with two α-carbon atoms labeled in red. At the top, the side chain of Phe²¹⁸ is visible, and below it are Leu²⁴⁰ and Tyr²³⁹.

Molecular Biology Institute and Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90024.

*Present address: Lilly Hall of Biological Sciences, Purdue University, West Lafayette, IN 47907.

†Present address: Department of Chemistry, College of Natural Sciences, Seoul National University, Seoul, South Korea.

‡Present address: Agouron Pharmaceuticals, P.O. Box 12209, La Jolla, CA 92037.

§To whom correspondence should be sent.

6 and 38 to 42 in S) raises the *R*-factor by 1.3%.

The 123 residues (16) of the S chain fold into a brain-shaped domain with a COOH-terminal tail (Fig. 2A), which extends parallel and near to the fourfold axis, about which four S chains cluster. A four-turn helix at the NH₂-terminus (α S1 in Fig. 3A) precedes a four-stranded β sheet of topology (+1, +2 \times , -1) (20). Between strands 2 and 3 there is an extended cross-connection including two short α helices (α S2 and α S3), much of which packs against the neighboring barrel domain of L. [This is shown in Fig. 3A: the 1's in row B indicate that these S residues contact the B domain of protomer 1 (Fig. 3C).]

The 477 residues of the L chain (17) can be considered as forming two domains, each having, like the S subunit, long COOH-terminal extensions (Fig. 2B). The first 168 residues are the N domain. Residues 5 to 134 form a four-stranded, antiparallel β sheet of the "open-face sandwich" type (20) of topology (-2 \times , +1, +2 \times), with one surface exposed to solvent. The extended loop between strands 1 and 2 contains a two-turn α helix (α N1 in Fig. 3B), as does the loop between strands 3 and 4 (α N2). Helix α N2 packs against the barrel domain of the twofold related L subunit (as shown in Fig. 3B by the 5's in the B row under α N2), where it seems to form part of the active site (see below). The COOH-terminal

connector (residues 135 to 168) leads from inside the sheet of the N domain, passing completely underneath the barrel, where it forms a short helix (α N3) before linking to the barrel domain.

Residues 169 to 477 fold into the α/β barrel, or B domain, of the type found in more than a dozen enzymes to date (21). The barrel is formed from eight hydrogen-bonded, twisted, but roughly parallel β strands, which are surrounded by eight helices (Fig. 2B). The mouth of the barrel at the COOH-termini of the β strands open onto a pocket that is covered partially by the N domain of a twofold-related L subunit (N5 in Fig. 3C). This pocket at the mouth of the barrel is the active site. Three residues implicated in catalysis by chemical studies [Lys¹⁷⁵, Lys²⁰¹, and Glu⁶⁰ (12, 22)] are within 10 Å of the mouth. Glu⁶⁰ is from the N domain of the neighboring L subunit. In all, there are three segments of chain from this N domain that pass within 13 Å of the Ne atom of Lys²⁰¹, including residues 20 to 22, 57 to 65, and 113 to 121. Thus in RuBisCO, as in at least four other enzymes (23), the active site is formed at the interface of two subunits. Essentially this conclusion had been reached earlier on the basis of site-directed mutagenesis (24).

The B domain of RuBisCO is similar to other α/β barrels, but there are variations. After leaving the connector from the N domain, the polypeptide chain forms the first β strand of the barrel, followed by the first α helical connector to the next strand, as is usual. However, after forming six of these β - α units as it coils into the barrel, the chain makes an excursion from the barrel (at the lower right of the red B domain in Fig. 2B) to form a two-stranded β ribbon (β B6a and β B6b). This ribbon lies near to the β N4 strand of the β sheet of the N domain, but is just displaced from the position to extend the sheet. The chain then returns to complete the barrel with the seventh and eighth β - α units. The overall conformations of both the N and B domains are similar to these domains in the subunit of *R. rubrum* RuBisCO (5).

Study of Fig. 3B reveals a wide variation in the helices and intervening loops of the eight units. For example, the fourth helix is only 5 residues long, whereas the second has 17 residues, and the seventh is actually a 3₁₀ helix. These variations and chain excursions in the barrel are probably what make the recognition of barrels purely from their amino acid sequences such a difficult task (25).

The final 45 residues of the B domain form its COOH-terminal extension, which sits on the side of the barrel opposite to the N domain (Fig. 2B). At the start of this extension is a short α helix (α C1 in Fig. 3B)

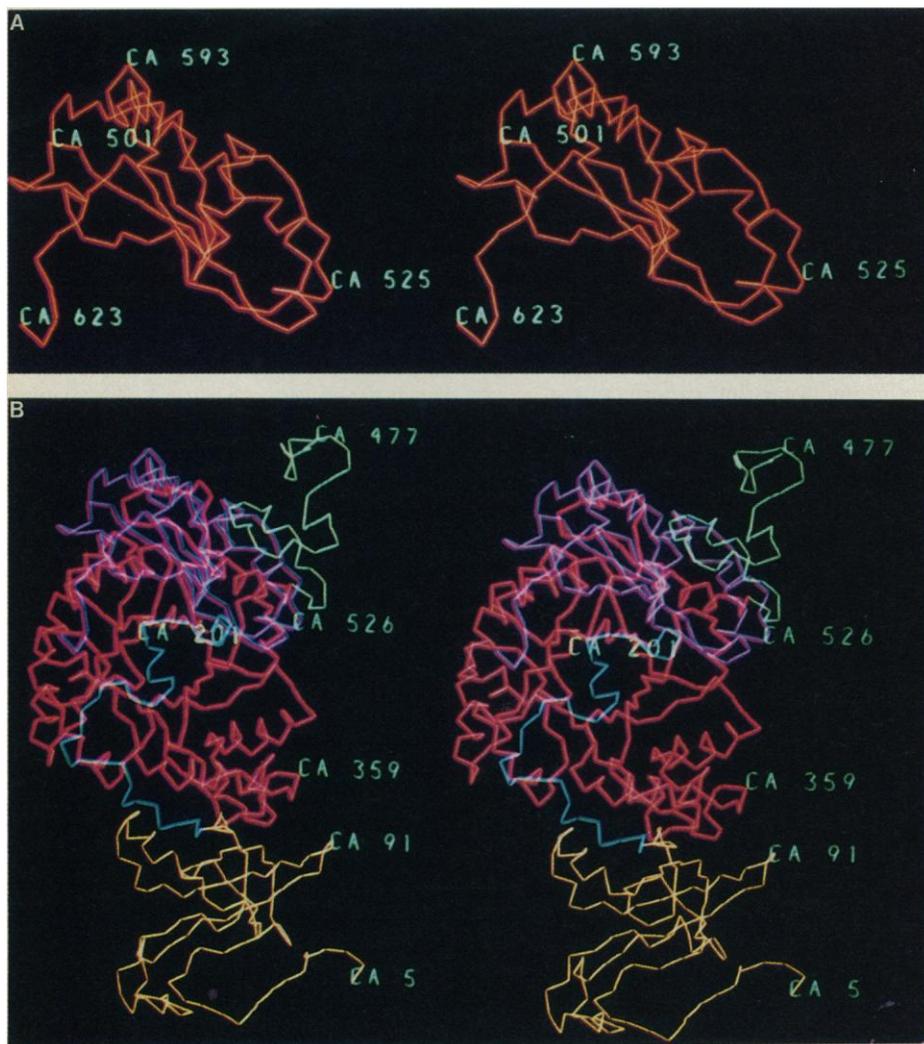


Fig. 2. The α -carbon backbone of RuBisCO. (A) Stereopair photograph of the S subunit, viewed roughly down the *x*-axis (see Fig. 3C), which happens to be approximately parallel to the four-stranded β sheet. The fourfold (*z*) axis is nearly vertical, with the COOH-terminal tail pointing downward, roughly along negative *z*. Four α -carbon atoms are labeled with the residue number +500 (to distinguish them from L subunit numbers). (B) Stereopair photograph of the entire LS protomer, viewed down the axis of the barrel of the B domain (residues 169 to 432 in red). The COOH-terminal extension of the B domain (residues 433 to 477) is in green; the N domain (residues 5 to 134) is in yellow, and its connection to B (residues 135 to 168) is in blue; S is in purple, behind the barrel. This protomer can be oriented in the L₈S₈ molecule by noting that the barrel axis is almost the same as the *x*-axis in Fig. 3C. Notice that the orientation of S is the same as in (A). Labels for several α -carbon atoms, including the active site Lys²⁰¹, are labeled just to the right of the atomic position.

that runs antiparallel to the final helix of the barrel. Four COOH-terminal extensions run between and above S subunits to form the extreme top of the RuBisCO molecule, and four form the bottom.

The functions of residues conserved among species of RuBisCOs (Fig. 3, A and B) are illuminated to some extent by the structure. In the L subunit, the most conserved segments are the loops at the

COOH-terminal sides of the β strands of the barrel. These segments surround the barrel opening, which contains the active site. They include: residues 174 to 180, 198 to 204 (which include β B2), 209 to 211,

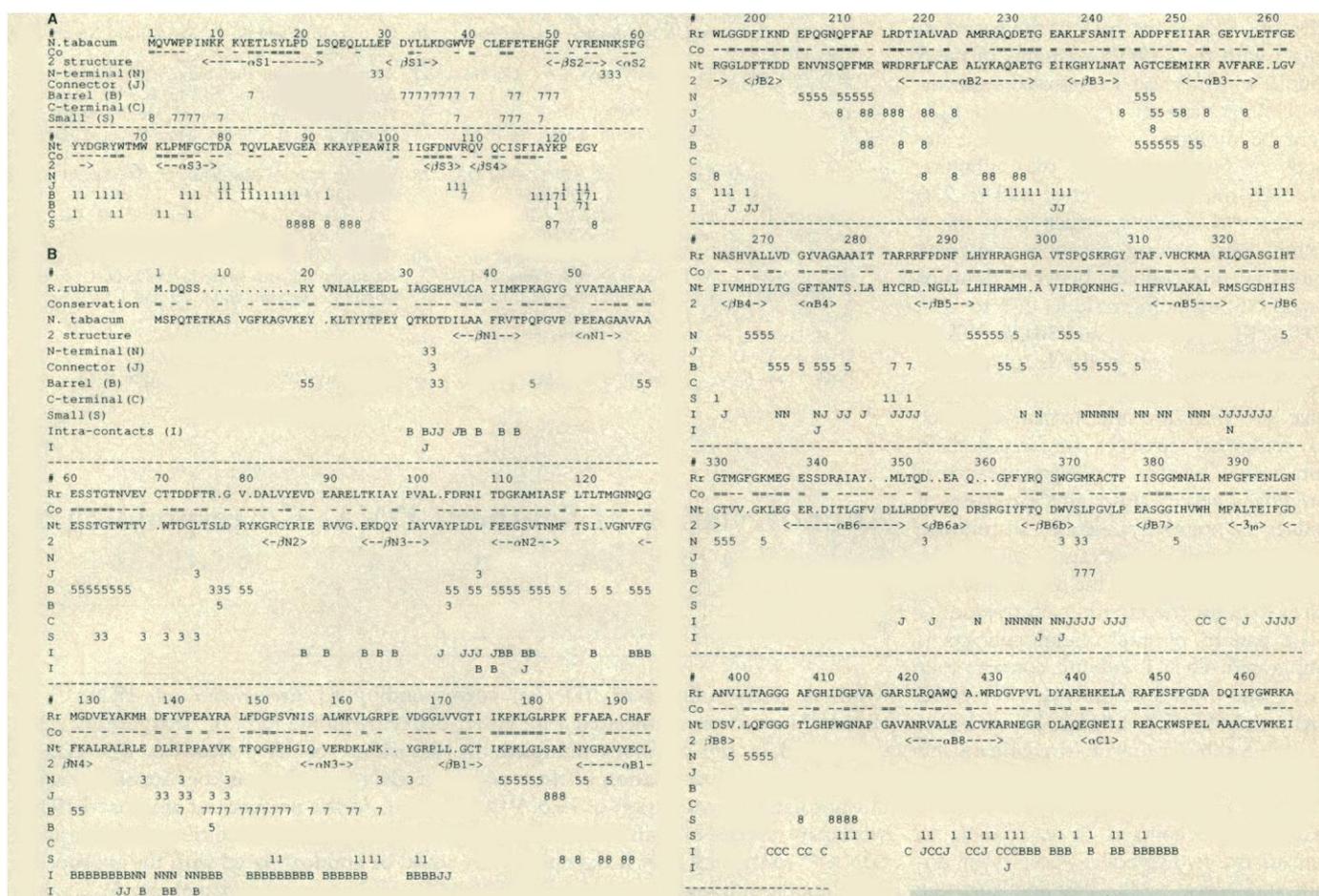
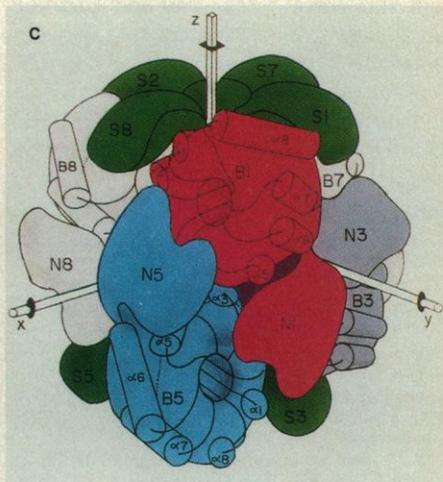


Fig. 3. Secondary structure, sequence conservation, and interdomain contacts of RuBisCO. **(A)** S subunit. The top line of both blocks contains sequence numbers of tobacco RuBisCO; the next line gives the sequence (16, 27); the next gives conserved residues (Co): the symbol “-” denotes residues identical in higher plant RuBisCOs, and the symbol “=” denotes residues identical in higher plant RuBisCOs plus *Anabaena* and *Anacystis* according to (28). The next line (2 structure) gives the type of secondary structure, the domain (S), and the number of the secondary structure element within the domain; rows N, J, B, C, and S give the interdomain contacts. Numbers denote the symmetrically equivalent domain or domains contacted by the protomer 1 at that residue. Protomer numbers are defined in (C). The symbols for the lines are: N for N domain (residues 5 to 134), J for COOH-terminal extension of N (residues 135 to 168), B for B domain (residues 169 to 432), and C for COOH-terminal extension of B (residues 433 to 477). In the lower block, where three residues of S each contact two barrel domains, two B lines are required. **(B)** L subunit. Each block has about 12 lines: (i) sequence numbers of tobacco RuBisCO; (ii) the sequence of *R. rubrum* (29), labeled Rr; (iii) conserved residues, labeled Co; the symbol “-” denotes identical residues in *L8S8* RuBisCOs according to (10) and “=” denotes identical residues in *R. rubrum* and *L8S8* RuBisCOs; (iv) the sequence of *Nicotiana tabacum* (17), labeled Nt; and (v) secondary structure designation (2 structure). The label gives the type of secondary structure, its domain location, and its number within the domain; (rows N to C) the interdomain contacts, as in (A). The final two lines, labeled I, give intrasubunit domain contacts. Each letter denotes the domain of the same subunit contacted by that residue. Where one residue contacts more than one neighboring domain, the number of lines increases. **(C)** A schematic drawing of the RuBisCO molecule with all visible domains labeled, which defines the protomer numbers used in (A) and (B). Protomers 1, 2, 7, and 8 have their S subunits and B domains at positive z; those of protomers 3, 4, 5, and 6 are at negative z. The α helices visible in the B domains of protomers 1 and 5 are labeled. For clarity, the drawing omits the COOH-terminal extensions of the B domain and the full extent of the overlap of the N domain over the barrel. S subunits are green. Subunit L1 (red) is related to subunit L5 (blue) by a twofold axis that bisects the x- and y-axes. Amino acid residues that contact neighboring protein domains were detected as follows: for each domain, the solvent-accessible surface area (13) of each non-hydrogen atom was computed (14) both in the presence and absence of each neighboring domain. Any residue that shows a difference in accessible surface area, on addition of the neighboring domain, is defined as being in contact with the domain added. In computing the surface areas, we used a probe sphere of 1.4 Å radius. This means that our operational definition of “contact” is generous, including any pairs of residues having nuclei within 2.8 Å plus the sum of their van der Waals radii (roughly 5.6 Å).



291 to 300, 329 to 335, and 403 to 405, as well as residues 379 to 381, which actually form a β strand. Another conserved segment is 58 to 65, which is one of the segments of the N domain that is part of the active site. However, another of the conserved segments, residues 228 to 236, is not near the active site. It is at the back of the barrel, in contact with the COOH-terminus of the S subunit, which is also a conserved segment. The conservation of this region in L₈S₈ might be thought to reflect stringent demands of bonding of the two subunits. However, the conservation of 228 to 236 extends to *R. rubrum*, which lacks the S subunit. Also, among the more conserved regions are the contacts between domains B1 and N5, including residues 175 to 180, 209 to 213, 291 to 297, and 301 to 303. In general, conservation in the L subunit appears stronger for loops that surround the active site than for interdomain contacts, and somewhat stronger for interdomain contacts than for structural elements such as β strands or helices, although these classes of residues are not always distinct. In contrast, in the S subunit, conservation tends to be somewhat stronger in the β strands and helices than for the interdomain residues.

The pattern of intersubunit contacts is complicated. Each L subunit contacts four other L and three S subunits (L1 in Fig. 2C touches L3, L5, L7, L8, S1, S3, and S8); and each S subunit touches two other S and three L subunits (S1 contacts S7, S8, L1, L3, and L7). From the accessible atomic areas of these contacts, it is possible to estimate the hydrophobic energies of domain interaction (26). We find that the strongest hydrophobic interaction between domains is between the barrel of one subunit and the N domain that forms part of the same active site. An example is the B1-N5 interaction in Fig. 3C. These interaction energies and other aspects of RuBisCO structure will be discussed elsewhere.

REFERENCES AND NOTES

1. R. J. Ellis, *Trends Biochem. Sci.* **4**, 241 (1979).
2. T. S. Baker, D. Eisenberg, F. A. Eislerling, L. J. Weissman, *J. Mol. Biol.* **91**, 391 (1975).
3. T. S. Baker *et al.*, *Science* **196**, 293 (1977).
4. M. Chapman *et al.*, *Nature* **329**, 354 (1987).
5. G. Schneider, U. Lindqvist, C.-I. Branden, G. Lorimer, *EMBO J.* **5**, 3409 (1986).
6. S. W. Suh, D. Cascio, M. S. Chapman, D. Eisenberg, *J. Mol. Biol.* **197**, 363 (1987).
7. I. Andersson, A.-C. Tjader, E. Cedergren-Zeppezauer, C.-I. Branden, *J. Biol. Chem.* **258**, 14088 (1983).
8. J. A. Barcena, R. W. Pickersgill, M. J. Adams, D. C. Phillips, F. R. Whatley, *EMBO J.* **2**, 2363 (1983).
9. A. Holzenberg *et al.*, *Nature* **325**, 730 (1987).
10. H. Nakagawa *et al.*, *J. Mol. Biol.* **191**, 577 (1986).
11. J. T. Andrews, *Methods Enzymol.* **118**, 410 (1986); B. A. McFadden, J. Torres-Ruiz, H. Daniell, G. Sarojini, *Philos. Trans. R. Soc. London Ser. B* **313**, 347 (1986); T. Takabe, A. K. Rai, T. Akazawa, *Arch. Biochem. Biophys.* **229**, 202 (1984).

12. H. M. Miziorko and G. H. Lorimer, *Annu. Rev. Biochem.* **52**, 507 (1983).
13. B. Lee and F. M. Richards, *J. Mol. Biol.* **55**, 379 (1971).
14. T. J. Richmond, *ibid.* **178**, 63 (1984).
15. M. S. Chapman *et al.*, *Philos. Trans. R. Soc. London Ser. B* **313**, 367 (1986).
16. K.-D. Muller, J. Salnikow, J. Vater, *Biophys. Biochem. Acta* **742**, 78 (1983).
17. K. Shinozaki and M. Sugiura, *Gene* **20**, 91 (1982); I. Amiri, J. Salnikow, J. Vater, *Biophys. Biochem. Acta* **784**, 116 (1984).
18. T. N. Bhat and G. H. Cohen, *J. Appl. Cryst.* **17**, 244 (1984).
19. T. A. Jones, *ibid.* **11**, 268 (1978).
20. J. S. Richardson, *Adv. Protein Chem.* **34**, 167 (1981).
21. C.-I. Branden, *Q. Rev. Biophys.* **13**, 317 (1980); D. W. Banner *et al.*, *Nature* **255**, 609 (1975); A. Goldman, D. L. Ollis, T. A. Steitz, *J. Mol. Biol.* **194**, 143 (1987); J. P. Priestle *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 5690 (1987); Xia Zong-Xiang *et al.*, *ibid.*, p. 2629; L. W. Lim *et al.*, *J. Biol. Chem.* **261**, 15140 (1987); J. Sygusch, D. Beaudry, M. Allaine, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7846 (1987).
22. G. H. Lorimer, *Biochemistry* **20**, 1236 (1981); F. C. Hartman, C. D. Stringer, S. Milanez, E. H. Lee, *Philos. Trans. R. Soc. London Ser. B* **313**, 447 (1986); F. C. Hartman *et al.*, *J. Biol. Chem.* **262**, 3496 (1987); F. C. Hartman, F. W. Larimer, R. J. Mural, R. Machanof, T. S. Soper, *Biophys. Biochem. Res. Commun.* **145**, 1158 (1987).
23. R. J. Almassy, C. A. Janson, R. Hamlin, N.-H. Xuong, D. Eisenberg, *Nature* **323**, 304 (1986).
24. F. W. Larimer, E. H. Lee, R. J. Mural, T. S. Soper, F. C. Hartman, *J. Biol. Chem.* **262**, 5327 (1987).
25. M. R. Hurle *et al.*, *Proteins Struct. Funct. Genet.* **2**, 210 (1987).
26. D. Eisenberg and A. D. McLachlan, *Nature* **319**, 199 (1986).
27. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
28. F. Nagy *et al.*, *Philos. Trans. R. Soc. London Ser. B* **313**, 409 (1986).
29. F. C. Hartman, C. D. Stringer, E. H. Lee, *Arch. Biochem. Biophys.* **232**, 280 (1984).
30. We thank S. Wildman, the discoverer of RuBisCO, for suggesting this study in 1970, D. C. Rees for helpful discussions, and NIH and USDA for support.

22 January 1988; accepted 20 April 1988

Aminoacylation of Synthetic DNAs Corresponding to *Escherichia coli* Phenylalanine and Lysine tRNAs

AKBAR S. KHAN AND BRUCE A. ROE

Synthetic DNA oligomers (tDNAs) corresponding to *Escherichia coli* tRNA^{Phe} or tRNA^{Lys} have been synthesized with either deoxythymidine (dT) or deoxyuridine (dU) substituted in the positions occupied by ribouridine or its derivatives. The tDNAs inhibited the aminoacylation of their respective tRNAs with their cognate amino acids, but not the aminoacylation of tRNA^{Leu} with Leu. In the presence of aminoacyl-tRNA synthetase, species of both a tDNA^{Phe} synthesized with a 3' terminal riboadenosine and a tDNA^{Lys} containing only deoxynucleotides could be aminoacylated with the appropriate amino acids, although the Michaelis constant K_m and observed maximal rate V_{max} values for aminoacylation were increased by three- to fourfold and decreased by two- to threefold, respectively. The aminoacylation of synthetic tDNAs demonstrates that the ribose backbone of a tRNA is not absolutely required for tRNA aminoacylation.

TRANSFER RNAs PLAY A PIVOTAL role in protein synthesis through their adaptor function and also serve as modulators of various cellular biosynthetic pathways (1). Because tRNAs are the most highly posttranscriptionally modified nucleic acid polymers (2), our laboratory has attempted to gain insight into the structural features important for tRNA function by determining how various modified nucleotides affect tRNA function (3). Recent advances in automated DNA synthesis (4) have increased our understanding of both the physical structure and the biological function of nucleic acids (5) as well as the mechanism of their interactions with themselves, with other nucleic acids, and with various cellular proteins (6). In an attempt to determine if the ribose backbone of tRNA is necessary for its biological functions, as well as to generate substrates that

might be active with tRNA posttranscriptional modifying enzymes and to gain insight into the secondary and tertiary structure of tRNA genes, we synthesized DNA polymers that corresponded to the sense strand of the *Escherichia coli* tRNA^{Phe} gene (7, 8) and the tRNA^{Lys} gene (9, 10) to produce the corresponding full-length tDNA^{Phe} and tDNA^{Lys}.

Initially we synthesized two 76-nucleotide DNA oligomers, corresponding to the full-length tDNA^{Phe} with either dT or dU in the positions occupied by ribouridine or its derivatives in *E. coli* tRNA^{Phe} (7). Since earlier studies demonstrated that the 3' terminus of *E. coli* tRNA^{Phe} requires the 2' hydroxyl for aminoacylation (11, 12), it was not surprising that both of these tDNAs

Department of Chemistry, University of Oklahoma, Norman, OK 73019.