A cDNA clone encoding the peroxisomal enzyme glycolate oxidase (EC 1.1.3.15) was identified by probing a cDNA library of spinach with synthetic oligonucleotides based on the partial amino acid sequence of the enzyme. Determination of the DNA sequence of the 1526-nucleotide cDNA indicated a 1107-nucleotide open reading frame which encodes a polypeptide of 40,282 daltons. The polypeptide produced by in vitro transcription and translation of the cDNA insert had the same apparent subunit molecular mass as the enzyme purified from leaves, indicating that the cDNA encodes a full-length polypeptide and that no cleavage of the polypeptide is required for uptake of the polypeptide by peroxisomes. Comparison of the deduced amino acid sequence with those of two other plant peroxisomal proteins revealed a region of homology which may be involved in directing proteins to the peroxisome.

The flavoprotein glycolate oxidase (EC 1.1.3.15) is a peroxisomal enzyme which catalyzes the oxidation of α-hydroxy acids. In vertebrates the enzyme is believed to be involved in the metabolic production of oxalate by the oxidation of glycolate through glyoxylate (1). The enzyme is also present in the leaves of higher plants where it catalyzes the second reaction of the photosynthetic pathway, the oxidation of glycolate to glyoxylate and H₂O₂ (2). Both the mammalian (3) and plant (4–6) enzymes are generally isolated as tetramers or octamers composed of identical subunits of approximately 43 kDa. Glycolate oxidase is the only FMN-dependent oxidase for which the tertiary structure has been determined (5). The subunit structure is an eight-stranded α/β-barrel, similar to that of triose-phosphate isomerase, in which the coenzyme is bound to the barrel. The structure of the flavin binding domain distinguishes glycolate oxidase from the other flavin-containing enzymes for which structures have been obtained (5). A detailed description of the relationship between structure and function of the enzyme has been delayed by the lack of primary sequence information. The availability of high resolution structural information makes glycolate oxidase an attractive subject for mechanistic studies of protein transport into peroxisomes. It is now apparent that peroxisomal enzymes are transported post-translationally into the preformed organelle (7). Unlike other organelles, peroxisomal proteins are not cleaved during transport into the peroxisome, and the structural features of the polypeptides required for peroxisomal localization are not known. The only peroxisomal enzymes from plants for which amino acid sequence information is available are nodulin 35, a subunit of soybean uricase II (8), and a partial deduced amino acid sequence of the glyoxysomal malate synthase from cucumber (9). Thus, the availability of additional sequences of peroxisomal enzymes may be useful in determining the molecular basis for peroxisomal localization.

In this report we describe the cloning of a full-length cDNA for glycolate oxidase from spinach, the determination of the deduced amino acid sequence from the DNA sequence, and the identification of a short region of amino acid homology shared by the three plant peroxisomal proteins for which sequence is available.

EXPERIMENTAL PROCEDURES

Materials—Glycolate oxidase was purified to homogeneity from spinach leaves as described (5) or was a gift from Dr. C. I. Branden (Uppsala). Rabbit antiserum against glycolate oxidase was prepared by standard methods. A λgt11 library derived from mRNA of spinach leaf (Spinacea oleracea L., cv. American Hybrid 424) was generously provided by W. L. Opren (United States Department of Agriculture, Agricultural Research Service, Urbana, IL). Oligonucleotides were synthesized by the phosphoramidite method (10). Plasmid pKSM13* “Bluescript” was purchased from Stratagene (San Diego, CA).

Nucleic Acid Isolation—DNA was extracted from 5-week-old spinach leaves as described (11). Poly(A)⁺ RNA was isolated from leaves by phenol extraction (12) followed by one round of oligo(dT)-cellulose chromatography (13).

Nucleic Acid Hybridization—Nitrocellulose plaque lifts (13) were probed with a mixture of synthetic oligonucleotides (Fig. 1) which were 5'-end-labeled to an average specific activity of 10⁶ dpm µg⁻¹ with [γ-³²P]ATP (3000 Ci mmol⁻¹) and T4 polynucleotide kinase. Filters were prehybridized 3 to 6 h at 52 °C in 6 × SSC, 50 mM sodium phosphate (pH 6.8), 5 × Denhardt’s solution, 100 µg ml⁻¹ of sonicated herring DNA (13). The hybridizations were carried out at 52 °C for 24 to 30 h in the same solutions with the addition of 10% (w/v) dextran sulfate and 0.5 pmol ml⁻¹ of labeled oligonucleotide. The temperature for hybridization was based on an empirical formula (14) by assuming that all ambiguous positions contained A or T bases. The filters were washed twice in 6 × SSC, 0.1% SDS for 15 min at 22 °C, once for 15 min and twice for 30 min at 52 °C in 3 mM tetramethylammonium chloride, 50 mM Tris-HCl (pH 8.0), 2 mM EDTA, 0.1% SDS.

Hybrid-selected Translation—RNA was hybrid-selected as described (16) except that DNA was immobilized on Zeta-Probe membranes (Bio-Rad) by applying 10 µg of linearized plasmid in 50 µl of 0.4 M NaOH to the well of a dot-blot apparatus (Bethesda Research Laboratories) and drawing the liquid through the membrane under suction. The dots were excised and, after removing loosely bound DNA, were prehybridized for 1 h in a siliconized vial containing 100 µl of 65% (v/v) formamide, 0.4 M NaCl, 0.2% SDS, 0.2% bovine serum albumin, 20 mM Pipes (pH 6.4), 100 µg ml⁻¹ yeast tRNA (Sigma). Fifty µg of poly(A)⁺ RNA was added and the hybridization was continued for 6 h at 37 °C. The filters were washed five times with 20 ml of 50% formamide, 20 mM NaCl, 1 mM EDTA, 0.5% SDS, and 20 mM Pipes (pH 6.4) for 15 min at 37 °C. The filters were blotted...
The oligonucleotide mixture used as a probe for the glycolate oxidase gene and the corresponding amino acid sequence. The methionine residue was not observed on the peptide which was obtained from cyanogen bromide cleavage but was inferred from the mode of action of cyanogen bromide.

dry, and the RNA was eluted as described (16), translated in a reticulocyte lysate system (Promega), and resolved on a 12% SDS-polyacrylamide electrophoresis gel which was dried and autoradiographed.

DNA Sequence Analysis—The fragment to be sequenced was digested with Sau3A, TaqI, AluI, HaeIII, or EcoRI and cloned into the BamHI, AccI, Smal, or EcoRI sites, respectively, of M13mp18 or M13mp19. Cloning and preparation of template DNA was performed using standard protocols (17). The inserts were sequenced by the chain-termination method and resolved on buffer gradient gels (18).

In Vitro RNA Transcription—Conditions for in vitro transcription of pMV2 were essentially as described (19) except that T7 RNA polymerase was used and the DNase treatment was omitted.

RESULTS

From a partial amino acid sequence of the spinach glycolate oxidase we designed a mixture of 64 synthetic oligonucleotides (20-mers) which represented all possible DNA sequences which could encode the amino acid sequence (Fig. 1). The utility of this mixture as a gene-specific probe was suggested by the observation that when used to probe a genomic Southern blot of spinach DNA only one major band of homology was observed under stringent washing conditions (results not presented). The temperature for hybridization was chosen, on the basis of empirical rules (14), to be 2 °C above the predicted melting temperature which would result if all ambiguous nucleotides in the probe were A or T. The washing conditions were based on the observation that replacement of sodium chloride with tetramethylammonium chloride permits relatively accurate prediction of melting temperatures solely on the basis of oligonucleotide length (15).

A Agt11 library derived from spinach leaf poly(A) RNA was probed with the end-labeled oligonucleotide mixture using the same hybridization conditions established for the genomic Southern blot. Of 500,000 plaques screened, three gave a reproducible signal and were retained. One of these, designated cGLO-1, had a 1.5-kilobase insert which was subcloned in both orientations into the EcoRI site of pKSM13 to produce plasmids pMV2 and pMV21.

In order to ensure that the cloned fragment encoded glycolate oxidase, the insert from pMV2 was used to select an mRNA from leaf poly(A) RNA. This hybrid-selected RNA was then translated in vitro and the products were electrophoresed in SDS-polyacrylamide gels before and after immunoprecipitation with a rabbit antibody specific for glycolate oxidase (Fig. 2). Even without immunoprecipitation a labeled polypeptide of about 43 kDa was apparent, and after immunoprecipitation only that polypeptide was seen. The apparent molecular weight of these polypeptides corresponded to a 43-kDa protein which was one of two polypeptides immunoprecipitated from the in vitro transition products of leaf poly(A) RNA. Thus, we concluded that the insert in pMV2 encodes glycolate oxidase. The identity of the 39-kDa polypeptide which is also immunoprecipitated from the translation products of poly(A) RNA by the antibody is not known. Since the protein used to immunize the rabbit was highly purified

\(^2\) C. Bränden, personal communication.

**FIG. 1.** The oligonucleotide mixture used as a probe for the glycolate oxidase gene and the corresponding amino acid sequence. The methionine residue was not observed on the peptide which was obtained from cyanogen bromide cleavage but was inferred from the mode of action of cyanogen bromide.

**FIG. 2.** In vitro translation of hybrid-selected glycolate oxidase mRNA. Lane a, immunoprecipitation of translation products of total leaf poly(A) RNA; lanes b and c, translation products of RNA selected from total leaf poly(A) RNA with pMV2 before (c) and after (b) immunoprecipitation; lane d, control, showing translation products of RNA selected with vector pKSM13. The arrow indicates the position of glycolate oxidase. The molecular mass (kDa) of marker proteins is shown at the right.

**FIG. 3.** Restriction map of the cDNA insert in pMV2 and the strategy used to sequence the insert. The protein-coding region is indicated by the open bar. The direction and extent of sequence determinations are indicated by arrows. Short vertical bars on the arrows represent sequencing reactions initiated with the M13 universal primer (17-mer). Open circles on the arrows represent sequencing reactions primed with oligonucleotides synthesized specifically for the purpose. The scale at the bottom is in nucleotide base pairs. Ac, AccI; Ah, AhIII; A, A1uI; B, BamHI; HaeIII; N, NotI; P, PstI; Pv, PvuII; S, Sau3A; St, SstI. (5), we tentatively conclude that the 39-kDa polypeptide shares an epitope with glycolate oxidase.

The complete DNA sequence of the insert in pMV2 was determined on both strands according to the strategy outlined in Fig. 3. The composite nucleotide and deduced amino acid sequences are shown in Fig. 4. The entire insert was 1526 nucleotides and encoded an open reading frame of 1107 nucleotides which gave a deduced amino acid sequence of 369 residues encoding a protein of 40,282 daltons. The deduced amino acid sequence has an excess of positively charged amino acids (+5) which is consistent with the observation that other peroxisomal matrix proteins have relatively high isoelectric points (7). One potential site for asparagine-linked glycosylation is present in the deduced amino acid sequence (nucleotides 583-591).

The first ATG in the sequence of the cDNA occurs at nucleotide 135. Thus, if translation were to begin at another site in a missing region of the RNA further upstream, the polypeptide would have a molecular mass at least 4.5 kDa greater than if it begins at nucleotide 135. In order to test...
FIG. 4. Composite nucleotide sequence of the cDNA for glycolate oxidase and the deduced amino acid sequence. Nucleotides are numbered in the 5' to 3' direction beginning with the first residue of the initiator codon ATG. Nucleotides on the 5' side of residue 1 are designated by negative numbers.

whether the ATG at nucleotide 135 is the normal start site for translation, we determined the apparent molecular mass of the polypeptide produced by in vitro transcription and translation of the cDNA clone. This was facilitated by the fact that plasmid pMV2 was constructed in such a way that the 5' end of the cDNA was proximal to a site on the parent vector pKSM13 + for initiation of transcription by T7 RNA polymerase. The vector was also designed so that the ATG at nucleotide 135 is the first ATG in the transcript originating from the T7 polymerase initiation site. Comparison of the translation products from the in vitro transcript and poly(A +) RNA from leaves indicated that the polypeptide product of the cloned gene had the same apparent molecular mass as that from poly(A +) RNA (Fig. 5). Since a discrepancy of 4.5 kDa would have been readily apparent, we are confident that the open reading frame indicated in Fig. 4 is correct.

In vitro translation of spinach leaf poly(A +) RNA and the RNA produced by in vitro transcription of pMV2 with T7 polymerase resulted in the accumulation of several polypeptides, in addition to the 43-kDa polypeptide, which were immunoprecipitated by the anti-glycolate oxidase antisemur (Fig. 5). A similar observation by others (20) has been attributed to spurious initiation of translation at AUG codons other

**Primary Structure of Glycolate Oxidase**

**FIG. 6. Homology between the deduced amino acid sequences of plant microbody proteins.** MS, cucumber malate synthase; URICASE, soybean nodulin 35 polypeptide; GLO, spinach glycolate oxidase.
than the normally used initiator-AUG codon. The apparent molecular masses of the minor translation products in Fig. 5 (lane b) are 39, 32, 29, 27, 19, and 9 kDa. These values coincide very well with the predicted molecular masses of 38.1, 33.3, 30.8, 29, 19, and 11 kDa which would be obtained if translation of the in vitro synthesized transcript initiated at all AUG codons at a low frequency. The relatively high level of accumulation of the 39-kDa polypeptide in the translation products of poly(A+) RNA (Fig. 5) may reflect the fact that the poly(A+) RNA is capped whereas the transcript from pMV2 is not.

In order to examine the possibility that peroxisomal proteins share a common structural motif, we compared the deduced amino acid sequence of glycolate oxidase to the deduced amino acid sequences of the soybean nodulin 35 polypeptide (8) and cucumber malate synthase (9). In both comparisons, the FASTP algorithm designed by Lipman and Pearson (21) identified the same region of glycolate oxidase as having significant amino acid sequence homology with the other two polypeptides (Fig. 6). In the comparison of glycolate oxidase and malate synthase, a 19-residue region had 42% homology. In the comparison of glycolate oxidase and nodulin 35 there was 28% homology over a 28-residue region.

**DISCUSSION**

Several lines of evidence indicate that pMV2 contains a cDNA sequence encoding the full-length glycolate oxidase. First, the cDNA uniquely selected mRNA which, when translated in vitro, yielded a polypeptide of the correct apparent molecular mass which was immunoprecipitated by antisera against glycolate oxidase. Similarly, in vitro transcription and translation of the cDNA insert in pMV2 yields a polypeptide with an apparent molecular weight of about 43 kDa which comigrated with an in vitro translated glycolate oxidase and was immunoprecipitated by the anti-glycolate oxidase antibody. In addition, the deduced amino acid sequence of the cDNA insert in pMV2 contained the partial amino acid sequence of glycolate oxidase which was used to design the mixed oligonucleotide probe.

It is generally accepted that peroxisomal proteins are translated on free cytoplasmic ribosomes and post-translationally imported without obligatory cleavage of a signal or transit peptide (7). Thus, the information which directs a polypeptide to the peroxisome must be an intrinsic property of the structural protein. Although glycolate oxidase has one potential glycosylation site, the lack of an apparent difference in the molecular mass of the in vitro translated polypeptide and the subunit molecular mass of the enzyme purified from leaves (22) suggests that it is not glycosylated. Thus, an inherent structural feature of the polypeptide is implicated by default. It has recently been suggested (23) that proteins imported into glycosomes of *Trypanosoma brucei* have common elements on their surface which may serve as a topogenic signal for import into the glycosomes. In this respect it is interesting to note the homology found between a specific region of glycolate oxidase, uricase, and malate synthase (Fig. 6). Since these three plant peroxisomal proteins are functionally dissimilar, the existence of structural homology raises the possibility that this region may be involved in import of plant peroxisomal proteins. Alignment of the primary and tertiary structure of glycolate oxidase

**Acknowledgments**—We are indebted to C. I. Bränden and H. Jornvall for providing unpublished amino acid sequence information for spinach glycolate oxidase and for providing pure enzyme. We are also grateful to J. Wernke and W. L. Ogren for the gift of the spinach cDNA library.

**REFERENCES**

1. Schuman, M., and Massey, V. (1971) *Biochim. Biophys. Acta* **227**, 521-537.
2. Tolbert, N. E. (1981) *Annu. Rev. Biochem.* **50**, 133-157.
3. Schwam, H., Michelson, S., Randall, W. C., Sondey, J. M., and Hirschman, R. (1979) *Biochemistry* **18**, 2829-2833.
4. Behrends, W., Rausch, U., Loffler, H. G., and Kindl, H. (1982) *Planta* (Berl) **156**, 566-571.
5. Lindqvist, Y., and Bränden, C. I. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 8656-8659.
6. Nishimura, M., Akhmedov, Y. D., Strzalka, K., and Akazawa, T. (1988) *Arch. Biochem. Biophys.* **222**, 397-402.
7. Fujiki, Y., and Lazarow, P. B. (1985) *Annu. Rev. Cell Biol.* **1**, 489-530.
8. Nguyen, T., Zeliechowska, M., Foster, V., Bergmann, H., and Verma, D. P. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 5040-5044.
9. Smith, S. M., and Leaver, C. J. (1986) *Plant Physiol.* (Bethesda) **81**, 762-767.
10. Beaucage, S. L., and Caruthers, M. H. (1985) *Tetrahedron Lett.* **22**, 1859-1862.
11. Leutwiler, L. S., Hough-Evans, B. R., and Meyerowitz, E. M. (1984) *Mol. Gen. Genet.* **194**, 15-23.
12. Rezaian, M. A., Heaton, L. A., Pederson, K., Milner, J. J., and Jackson, A. O. (1983) *Virology* **131**, 221-229.
13. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
14. Suggs, S. V., Hirose, T., Miyake, T., Kawashima, E. H., Johnson, M. J., Itakura, K., and Wallace, R. B. (1981) in ICN-UCLA *Symposia in Developmental Biology*, (Brown, D. D., and Fox, C. F., eds) Vol. 23, pp. 683-693, Academic Press, New York.
15. Wood, W. I., Gitschier, J., Lasky, L. A., and Lawn, R. M. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 1585-1588.
16. Williams, J. G., and Mason, P. J. (1985) in *Nucleic Acid Hybridization, A Practical Approach* (Hames, B. D., and Higgins, S. J., eds) pp. 113-137, IRL Press, Oxford.
17. Messing, J. (1983) in *Methods Enzymol.* **101**, 20-79.
18. Biggen, M. D., Gibson, T. J., and Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 3963-3965.
19. Green, M. R., Maniatis, T., and Melton, D. A. (1983) *Cell* **32**, 681-684.
20. Small, G. M., and Lazarow, P. B. (1987) *J. Cell Biol.* **105**, 247-250.
21. Lipman, D. J., and Pearson, W. R. (1985) *Science* **227**, 1435-1441.
22. Kindel, H. (1992) *Annu. N. Y. Acad. Sci.* **866**, 314-326.
23. Wtenenga, R. K., Swinkels, B., Michela, P. A. M., Osinga, K., Misset, O., Van Beeumen, J., Gibson, W. C., Postma, J. P. M., Borst, P., Oppderose, F. R., and Hol, W. G. J. (1987) *EMBO J.* **6**, 215-221.
24. Fujiki, Y., and Lazarow, P. B. (1985) *J. Biol. Chem.* **260**, 5603-5609.