worth mentioning is that, in the absence of osmotic stress, *R. meliloti* but not *E. coli* actively catabolizes glycine betaine, salvaging carbon and nitrogen. Interestingly, this catabolic degradation system is "repressed" by increasing the osmotic strength of the medium, preventing a futile cycle of uptake and degradation and ensuring that glycine betaine is preserved to function as an osmoprotectant.

The bacterial experiment served as a bridge to the whole-plant study in which glycine betaine as well as proline betaine supplied to the roots of nodulated alfalfa seedlings enhanced symbiotic nitrogen fixation. For example, symbiotic N_2 fixation by control plants exposed to $0.2M$ NaCl was almost completely blocked for N_2 fixation activity. On the other hand, stressed plants treated with 10 mM betaines recovered up to 30 percent of maximum levels observed with the untreated controls. It is interesting to speculate that betaines are behaving as osmoprotectants at the whole-plant level, perhaps protecting the N_2-fixing apparatus in the nodule tissue against stress.

Conclusion

Cellular adaptation to osmotic stress is a fertile area for basic research, with possible future applications in agriculture, medicine, and industry (*33*, *34*). The concept of osmosensory proteins linking environmental changes in osmotic strength to dynamic changes in membrane, biochemical, and genetic activities of the cell should generate further investigations by molecular biologists and biochemists. The discovery of osmoprotective molecules and new classes of *osm* genes may lead to genetic enhancement of drought and salinity tolerance in crop plants. To achieve this challenging goal the emerging technology must be closely integrated with established procedures of plant improvement.

References and Notes

1. J. S. Boyer, *Science* **218**, 443 (1982).
2. E. Epstein *et al.*, *ibid.* **210**, 399 (1980).
3. P. H. Yancey, M. E. Clark, S. C. Hand, R. D. Bowlus, G. N. Somero, *ibid.* **217**, 1214 (1982).
4. D. Aspinall and L. G. Paleg, in *Physiology and Biochemistry of Drought Resistance in Plants*, L. G. Paleg and D. Aspinall, Eds. (Academic Press, New York, 1981), pp. 205–241.
5. T. J. Flowers, B. F. Troke, A. R. Yeo, *Annu. Rev. Plant Physiol.* **28**, 89 (1977).
6. J. A. Hellebust, *ibid.* **27**, 485 (1976).
7. R. L. Jefferies, in *Genetic Engineering of Osmoregulation*, D. W. Rains *et al.*, Eds. (Plenum, New York, 1980), pp. 135–154.
8. A. R. Stewart and J. A. Lee, *Planta* **120**, 279 (1974).
9. R. G. Wyn Jones and R. Storey, in *Physiology and Biochemistry of Drought Resistance in Plants*, L. G. Paleg and D. Aspinall, Eds. (Academic Press, New York, 1981), pp. 171–204.
10. A. D. Hanson and D. Rhodes, *Plant Physiol.* **71**, 692 (1983).
11. W. D. Hitz, J. A. R. Ladyman, A. D. Hanson, *Crop Sci.* **22**, 47 (1982).
12. R. Storey and R. G. Wyn Jones, *Plant Sci. Lett.* **4**, 161 (1975).
13. T. Mizuno, M.-Y. Chon, M. Inouye, *J. Biol. Chem.* **258**, 6932 (1983).
14. M. N. Hall and T. J. Silhavy, *Annu. Rev. Genet.* **15**, 91 (1981).
15. L. A. Laimins, D. B. Rhoads, W. Epstein, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 464 (1981).
16. E. P. Kennedy, *ibid.* **79**, 1092 (1982).
17. D. Le Rudulier and R. C. Valentine, *Trends Biochem. Sci.* **7**, 431 (1982).
18. A. R. Strom, D. Le Rudulier, M. W. Jakowec, R. C. Bunnell, R. C. Valentine, in *Genetic Engineering of Plants*, T. Kosuge *et al.*, Eds. (Plenum, New York, 1983), pp. 39–59.
19. D. Le Rudulier, S. S. Yang, L. N. Csonka, *Biochim. Biophys. Acta* **719**, 273 (1982).
20. L. Bouillard and D. Le Rudulier, *Physiol. Veg.* **21**, 447 (1983).
21. D. Le Rudulier, T. Bernard, G. Goas, J. Hamelin, *Can. J. Microbiol.*, in press.
22. B. Schobert and H. Tschesche, *Biochim. Biophys. Acta* **541**, 270 (1978).
23. R. T. Vinopal *et al.*, in *Genetic Engineering of Osmoregulation*, D. W. Rains *et al.*, Eds. (Plenum, New York, 1980), pp. 59–72.
24. T. Kohno and J. Roth, *Biochemistry* **18**, 1386 (1979).
25. L. N. Csonka, *Mol. Gen. Genet.* **182**, 82 (1981).
26. B. Perroud, thesis, University of Nantes, Nantes, France (1983).
27. R. J. Britten and F. T. McClure, *Bacteriol. Rev.* **26**, 292 (1962).
28. L. N. Csonka, *J. Bacteriol.* **151**, 1433 (1982).
29. B. Landfald and A. R. Strom, in preparation.
30. A. R. Strom, P. Falkenberg, T. Bjørnsen, O. B. Styrvold, B. Landfald, in preparation.
31. D. Le Rudulier, T. Bernard, J.-A. Pocard, G. Goas, *C. R. Acad. Sci.* **297**, 155 (1983).
32. J.-A. Pocard, thesis, University of Rennes, Rennes, France (1983).
33. A. Hollaender, Ed., *The Biosaline Concept: An Approach to the Utilization of Underexploited Resources* (Plenum, New York, 1979).
34. D. W. Rains *et al.*, Eds., *Genetic Engineering of Osmoregulation* (Plenum, New York, 1980).
35. This article was prepared with the support of NSF grant PFR 77-07301. D.L.R. was supported by a grant from Centre National de la Recherche Scientifique and A.S. by a grant from the Norwegian Research Council of Fisheries. We thank B. Perroud, T. Bernard, J.-A. Pocard, L. Bouillard, G. Goas, B. Landfald, P. Falkenberg, T. Bjørnsen, O. B. Styrvold, M. Jakowec, and M. Eshoo for theses and preprints.

RESEARCH ARTICLE

Structure and Expression of a Complementary DNA for the Nuclear Coded Precursor of Human Mitochondrial Ornithine Transcarbamylase

Arthur L. Horwich, Wayne A. Fenton, Kenneth R. Williams
Frantisek Kalousek, Jan P. Kraus, Russell F. Doolittle
William Konigsberg, Leon E. Rosenberg

Systems of compartmentation in eukaryotic cells direct specific polypeptides synthesized in the cytoplasm to destinations that include the extracellular space, cellular membranes, lysosomes, and the mitochondria. The first three of these systems share a common feature; polypeptides are synthesized on membrane-bound polyribosomes and are cotranslationally inserted into the cisternae of the endoplasmic reticulum (*1*). The system responsible for the compartmentation of nuclear-coded mitochondrial proteins appears to be fundamentally different; these polypeptides are synthesized on free polyribosomes and are subsequently released into the cytoplasm for posttranslational import by the mitochondria (*2*).

This import process is a major determinant of mitochondrial biogenesis. Fewer than 20 mitochondrial polypeptides are encoded by mitochondrial DNA and synthesized on mitochondrial ribosomes (*3*). The remainder (more than 200) are encoded in the nucleus and synthesized in the cytoplasm. Most nuclear-coded polypeptides destined for mitochondria are synthesized in the cytoplasm as larger precursors, containing amino-terminal leader sequences not present in their mature mitochondrial counterparts. Import of these precursors involves specific binding by receptor molecules present in the outer mitochondrial membrane, translocation through one or both mitochondrial membranes

---

A. L. Horwich, W. A. Fenton, F. Kalousek, J. P. Kraus, and L. E. Rosenberg are in the Department of Human Genetics, Yale University School of Medicine, New Haven, Connecticut 06510. K. R. Williams and W. Konigsberg are in the Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine. R. F. Doolittle is in the Department of Chemistry, University of California, San Diego, La Jolla, California 92093.

by a process requiring energy at the step of inner membrane passage, and proteolytic cleavage of the NH_2-terminal leader from the mature portion (*2*).

The mammalian hepatic enzyme ornithine transcarbamylase (OTC) is a useful model for study of the biogenesis of nuclear-coded mitochondrial proteins. This enzyme, a trimer of identical subunits located in the mitochondrial matrix, catalyzes the second step of the urea cycle, the condensation of carbamyl phosphate with ornithine to form citrulline. The structural gene for OTC is encoded on the X chromosome in both man and mouse (*4*). Studies of OTC biogenesis in rat and mouse liver indicate that its subunit is synthesized in the cytoplasm as a precursor (40 kilodaltons), which is processed posttranslationally to its mature form (36 kD) during mitochondrial import (*5*). The import of OTC is energy-dependent (*5*), and cleavage of the leader sequence is catalyzed by a Zn^{2+}-dependent matrix protease (*5*).

As a further step in defining more precisely the events controlling the biogenesis of OTC, we have deduced from cloned complementary DNA (cDNA) sequences the complete primary structure of the human OTC precursor. Further, we have used the cloned cDNA to demonstrate expression of the OTC precursor and its mature mitochondrial counterpart in cultured human cells that do not ordinarily express this enzymatic function.

Abstract. *Most mitochondrial proteins are encoded in the nucleus and are translated on free cytoplasmic ribosomes as larger precursors containing amino-terminal "leader" sequences, which are removed after the precursors are taken up by mitochondria. We have deduced the complete primary structure of the precursor of a human mitochondrial matrix enzyme, ornithine transcarbamylase (OTC), from the nucleotide sequence of cloned complementary DNA. The amino-terminal leader peptide of OTC is 32 amino acids in length and contains four arginines but no acidic residues. Cleavage of the leader peptide from the "mature" protein occurs between glutamine and asparagine residues. The sequence of mature human OTC resembles that of the subunits of both OTC and aspartate transcarbamylase from* Escherichia coli. *The biological activity of the cloned OTC complementary DNA was tested by joining it with SV40 (an animal virus) regulatory elements and transfecting cultured HeLa cells, which do not normally express OTC. Both the precursor and mature forms of the OTC subunit were identified; in stable transformants, enzymatic activity was also detected.*

Synthesis of Human OTC Precursor

The events responsible for recognition, translocation, and cleavage of a nuclear-coded mitochondrial precursor can be reconstituted in vitro by programming its synthesis in a cell-free translation system, and then adding isolated intact mitochondria (*6*). The data in Fig. 1 show such a reconstitution experiment for the subunit of human OTC. When total messenger RNA (mRNA) isolated from human liver is translated in a cell-free system and the products are precipitated with antiserum to OTC, a 40-kD OTC precursor (pOTC) is observed (lane 1). After the addition of fresh intact rat liver mitochondria, most of the pOTC was converted to a 36-kD polypeptide (lane 2), corresponding to the mature form of OTC in the active mitochondrial trimer.

We previously reported the isolation of a plasmid containing cDNA sequences encoding the carboxyl-terminal portion of rat OTC (*7*). A 389-base-pair (bp) Hind III restriction fragment was derived from the plasmid, and it contained a segment of 221 bp of COOH-terminal coding sequence. This fragment was nick-translated and used to screen a cDNA library prepared from mRNA of adult human liver (*8*). Because one of the three tryptic peptides cleaved from the COOH-terminal portion of human mature OTC subunits migrated differently from that of the rat subunit (*9*), we used reduced stringency of washing after the hybrid-

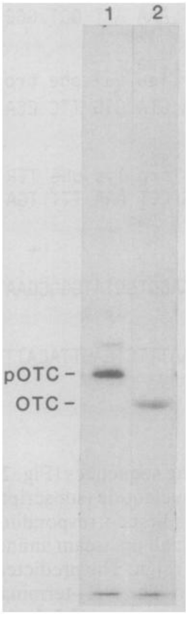

Fig. 1. Cell-free synthesis of human OTC precursor and its processing by intact rat liver mitochondria. Total cellular RNA was prepared (*33*) from human liver from autopsy within 3 hours of death and the poly(A) (polyadenylate)–containing fraction was selected by oligo(dT)-cellulose chromatography (dT, deoxythymidylate). Conditions for cell-free protein synthesis, posttranslational incubation with mitochondria, immunoprecipitation with rabbit antiserum to rat OTC, SDS - polyacrylamide gel electrophoresis, and fluorography have been described (*5*). (Lane 1) Immunoprecipitation of in vitro synthesized $[^{35}S]$methionine-labeled pOTC; (lane 2) immunoprecipitation of OTC from a mitochondrial pellet after incubation of translation mixture with intact mitochondria.

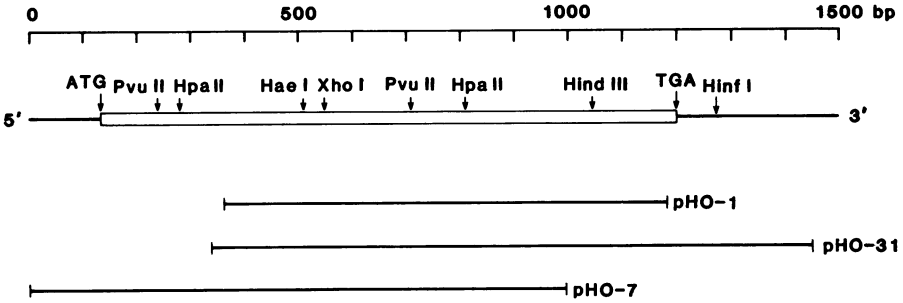

Fig. 2. Restriction endonuclease cleavage map of plasmids containing human OTC cDNA sequences. The upper line and boxed region represent the entire region of human OTC cDNA contained in the plasmids pHO-1, pHO-31, and pHO-7. The base pair scale begins arbitrarily at zero, which designates the beginning of the nucleotide sequence. The boxed region encloses the open reading frame. The probable site of translation initiation (*ATG*) is indicated, as is the termination codon (*TGA*). Those restriction sites used to generate fragments for DNA sequence analysis are shown. Plasmid pHO-1 was isolated from 40,000 cDNA clones synthesized from human hepatic mRNA, which were hybridized on nitrocellulose filters with a 389-bp nick-translated rat cDNA segment in a solution containing 5× Denhardt's, 5× SSC (standard saline citrate), 0.005*M* TES, *p*H 6.8, 0.5 percent SDS, and denatured salmon sperm DNA (10 μg/ml), and washed in 6× SSC at 68°C. The plasmid giving a positive hybridization signal was digested with Pst I or with Hpa II, and the products were separated on a 1 percent agarose gel, followed by transfer to a nitrocellulose filter. Hybridization with the 389-bp labeled rat OTC segment was performed exactly as in colony screening. The plasmid insert released by Pst I and a 490-bp fragment cleaved by Hpa II gave positive signals. The Hpa II fragment was completely sequenced by the Maxam-Gilbert method (*34*). Plasmids pHO-31 and pHO-7 were isolated from 40,000 additional clones from the human hepatic cDNA library, screened with the use of the nick-translated insert of pHO-1 and stringent washing conditions.

ization reaction. Screening of 40,000 colonies yielded two clones whose plasmid inserts appeared identical on restriction analysis. These plasmids were digested and subjected to Southern blot analysis (*10*) with the same rat OTC probe used for colony screening. A portion of the plasmid insert hybridized with the probe; this portion was isolated and subjected to DNA sequence analysis. Its nucleotide sequence was nearly identical to the rat OTC coding sequence present in the probe. The plasmid was designated pHO-1; its insert (Fig. 2), purified after digestion with Pst I, was used to screen additional colonies under stringent conditions of washing. From 40,000 colonies screened, eight additional positive clones were identified. After restriction analysis of small amounts of their plasmid DNA's, two plasmids, pHO-7 and pHO-31, were selected for further analysis. The three inserts, whose restriction maps are shown in Fig. 2, were predicted

```
5' AAGCTGAAGGGTGATATTACCTTTGCTCCCTCACTGCAACTGAACACATTTCTTAGTTTTTAGGTGGCCCCCGCTGGCTAACTTGCTGTGGAGTTTTCAAGGGCATAGAATCGTCC
             -220                -200                -180                -160                -140                -120

                            -30                                     -20                                     -10
                    met leu phe asn leu arg ile leu leu asn asn ala ala phe arg asn gly his asn phe met val arg asn phe
TTTACACAATTAAAAGAAG ATG CTG TTT AAT CTG AGG ATC CTG TTA AAC AAT GCA GCT TTT AGA AAT GGT CAC AAC TTC ATG GTT CGA AAT TTT
              -100                      -80                        -60                       -40

                         -1  1                                   10                                      20
 arg cys gly gln pro leu gln asn lys val gln leu lys gly arg asp leu leu thr leu lys asn phe thr gly glu glu ile lys tyr
 CGG TGT GGA CAA CCA CTA CAA AAT AAA GTG CAG CTG AAG GGC CGT GAC CTT CTC ACT CTA AAA AAC TTT ACC GGA GAA GAA ATT AAA TAT
-20                       -1 1                       20                         40                        60

                         30                                      40                                      50
 met leu trp leu ser ala asp leu lys phe arg ile lys gln lys gly glu tyr leu pro leu leu gln gly lys ser leu gly met ile
 ATG CTA TGG CTA TCA GCA GAT CTG AAA TTT AGG ATA AAA CAG AAA GGA GAG TAT TTG CCT TTA TTG CAG GGG AAG TCC TTA GGC ATG ATT
             80                         100                       120                        140

                         60                                      70                                      80
 phe glu lys arg ser thr arg thr arg leu ser thr glu thr gly phe ala leu leu gly gly his pro cys phe pro thr thr gln asp
 TTT GAG AAA AGA AGT ACT CGA ACA AGA TTG TCT ACA GAA ACA GGC TTT GCA CTT CTG GGA GGA CAT CCT TGT TTT CCT ACC ACA CAA GAT
160                       180                        200                        220                       240

                         90                                      100                                     110
 ile his leu gly val asn glu ser leu thr asp thr ala arg val leu ser ser met ala asp ala val leu ala arg val tyr lys gln
 ATT CAT TTG GGT GTG AAT GAA AGT CTC ACG GAC ACG GCC CGT GTA TTG TCT AGC ATG GCA GAT GCA GTA TTG GCT CGA GTG TAT AAA CAA
             260                        280                       300                        320

                         120                                     130                                     140
 ser asp leu asp thr leu ala lys glu ala ser ile pro ile ile asn gly leu ser asp leu tyr his pro ile gln ile leu ala asp
 TCA GAT TTG GAC ACC CTT GCT AAA GAA GCA TCC ATC CCA ATT ATC AAT GGG CTG TCA GAT TTG TAC CAT CCT ATC CAG ATC CTG GCT GAT
340                       360                        380                        400                       420

                         150                                     160                                     170
 tyr leu thr leu gln glu his tyr ser ser leu lys gly leu thr leu ser cys phe gly asp gly asn asn ile leu his ser ile met
 TAC CTC ACG CTC CAG GAA CAC TAT AGC TCT CTG AAA GGT CTT ACC CTC AGC TGT TTC GGG GAT GGG AAC AAT ATC CTG CAC TCC ATC ATG
             440                        460                       480                        500

                         180                                     190                                     200
 met ser ala ala lys phe gly met his leu gln ala ala thr pro lys gly tyr glu pro asp ala ser val thr lys leu ala glu gln
 ATG AGC GCA GCG AAA TTC GGA ATG CAC CTT CAG GCA GCT ACT CCA AAG GGT TAT GAG CCG GAT GCT AGT GTA ACC AAG TTG GCA GAG CAG
520                       540                        560                        580                       600

                         210                                     220                                     230
 tyr ala lys glu asn gly thr lys leu leu leu thr asn asp pro leu glu ala ala his gly gly asn val leu ile thr asp thr trp
 TAT GCC AAA GAG AAT GGT ACC AAG CTG TTG CTG ACA AAT GAT CCA TTG GAA GCA GCG CAT GGA GGC AAT GTA TTA ATT ACA GAC ACT TGG
             620                        640                       660                        680

                         240                                     250                                     260
 ile ser met gly arg glu glu glu lys lys lys arg leu gln ala phe gln gly tyr gln val thr met lys thr ala lys val ala ala
 ATA AGC ATG GGA CGA GAA GAG GAG AAG AAA AAG CGG CTC CAA GCT TTC CAA GGT TAC CAA GTT ACA ATG AAG ACT GCT AAA GTT GCT GCC
700                       720                        740                        760                       780

                         270                                     280                                     290
 ser asp trp thr phe leu his cys leu pro arg lys pro glu glu val asp asp glu val phe tyr ser pro arg ser leu val phe pro
 TCT GAC TGG ACA TTT TTA CAC TGC TTG CCC AGA AAG CCA GAA GAA GTG GAT GAT GAA GTC TTT TAT TCT CCT CGA TCA CTA GTG TTC CCA
             800                        820                       840                        860

                         300                                     310                                     320
 glu ala glu asn arg lys trp thr ile met ala val met val ser leu leu thr asp tyr ser pro gln leu gln lys pro lys phe TER
 GAG GCA GAA AAC AGA AAG TGG ACA ATC ATG GCT GTC ATG GTG TCC CTG CTG ACA GAT TAC TCA CCT CAG CTC CAG AAG CCT AAA TTT TGA
880                       900                        920                        940                       960


 TGTTGTGTTACTTGTCAAGAAAGAAGCAATGTTGGTCAGTAACAGAATGAGTTGGTTTATGGGGAAAAGAGAAGAGAATCTAAAAAATAAACCAATCCCTAACACGTGGTATGGGCGAAT
          980                 1000                1020                1040                1060                1080

 CGTACGATATGCTTTGCCATTGTGAAACTTTCCTTAAGCCTTCAATTTAAGTGCTGATGCACTGTAATACGTGCTTAACTTTGCTTAAACTCTCTAATTCCCAATTTCTGAGTTACATTT
          1100                1120                1140                1160                1180                1200

 AGATATCATATTAACTATCATATA     3'
          1220
```

Fig. 3. Nucleotide and amino acid sequences of the human OTC precursor. The cDNA sequence was derived from overlapping sequences (Fig. 2) by the procedure of Maxam and Gilbert (*34*); the following convention was used to number the amino acid (superscript) and nucleotide (subscript) residues in their respective sequences. The asparagine residue found at the NH_2-terminus of the mature OTC subunit, and its first corresponding nucleotide base (adenine) were denoted +1; all downstream amino acid residues (322) and nucleotides (1233) carry (+) signs; all upstream amino acid residues (the 32 from predicted residues of the leader sequence) and nucleotide sequences (231 bp) are denoted by the (−) sign. The predicted amino acid sequence is supported by direct amino acid sequences of mature OTC subunits purified from human and rat liver. The NH_2-terminal sequence of the human subunit was obtained (see text). Reversed-phase HPLC [a Water's C-18 μ-Bondapak column equilibrated at *p*H 2.5 and eluted with increasing concentrations of acetonitrile (*35*)], was used to isolate several tryptic peptides from human and rat OTC. Three of these human tryptic peptides were subjected to solid phase sequencing (*36*), which directly confirmed the predicted amino acid sequence of residues 190 to 206 and 246 to 257. In addition, several tryptic peptides from rat OTC were similarly analyzed, and their sequences exactly matched those predicted in the human subunit at residues 63 to 76, 80 to 97, 190 to 194, 199 to 206, and 275 to 299. Carboxypeptidase A digestion confirmed that phenylalanine is at the COOH-terminus of human OTC.

to share portions of their sequences. This was confirmed by DNA sequence analysis. From the three sequences, a single continuous sequence was derived. Because the coding sequence of the COOH-terminal portion of the human subunit, contained in pHO-1, had been determined, it was possible to identify the translational reading frame that encoded the additional portions of the subunit. The DNA sequences and predicted amino acid sequence are shown together in Fig. 3.

Amino Acid Sequence Analysis of Mature OTC Subunit

The amino acid sequence of the OTC subunit predicted from the cloned cDNA was consistent with several, independently derived, direct amino acid sequence determinations. The sequence of the NH_2-terminal 56 residues of the mature subunit (+1 to +56 as shown in Fig. 3) was determined by sequential Edman degradation of the purified mature human subunit and high-performance liquid chromatography (HPLC) of the phenylthiohydantoin derivatives. This sequence agreed precisely with that predicted by the DNA sequence except for lysine-14, which was identified as arginine. This discrepancy may represent an artifact of the molecular cloning procedure. Alternatively, it may identify a polymorphism at this position.

Solid phase sequencing of three tryptic peptides from human OTC directly confirmed the predicted sequence spanning residues 190 to 206 and 246 to 257. The amino acid sequences of six tryptic peptides derived from the mature rat subunit were similarly determined; each was nearly identical to that of corresponding tryptic peptides deduced from the predicted human amino acid sequence (data not shown). Only seven amino acid differences out of 85 residues analyzed were identified. All occurred within the internal domain of the peptides; with a single exception, all could be accounted for by a single nucleotide change. The amino acid sequence of the COOH-terminal portion of rat OTC, predicted from the sequence of the rat cDNA, matched the corresponding predicted human sequence at 102 out of 108 positions. Five of these six differences are the result of single base changes. Overall, when coding and amino acid sequences of the human mature subunit were compared with the available rat sequences, there was 90 percent identity of amino acid sequences and 95 percent homology of the DNA sequences.

Predicted Amino Acid Sequence of the Leader

Immediately upstream from the AAT (A, adenine; T, thymine) codon for the asparagine residue, which demarcates the NH_2-terminus of mature OTC, a sequence of 31 in-frame codons was preceded by an ATG (Met) (G, guanine) codon, at base −96. No additional ATG triplets were detected in 151 bp of additional insert sequence further upstream (Fig. 3). The latter upstream sequence also contained multiple translational termination codons in each of the three potential translational reading frames. A second ATG codon was identified at base −36, positioned in the same translational reading frame as the mature subunit. It was excluded as the translational initiation codon because it would predict a leader of only 1.3 kD, far less than the 4 kD predicted from comparison of the OTC precursor and mature subunits in Fig. 1. In contrast, the combined molecular mass of the 32 amino acid residues predicted by initiation with the ATG codon at base −96 totals 3.8 kD.

Examination of the 32 residues comprising the leader revealed that (i) the leader contains a uniform distribution of charged residues. Specifically, it does not contain a long, "membrane spanning" hydrophobic region; therefore, it bears no obvious resemblance to the "signal" sequences of secreted proteins (*11*). Second, the leader contains no acidic residues. This contrasts sharply with the composition of the mature subunit, which contains 13 percent acidic residues, and with an average for eukaryotic proteins, which contain 20 percent acidic residues (*12*). Third, in contrast with the absence of acidic residues, the leader contains 12 percent basic residues (4 of 32), compared with 14 percent in the mature subunit, and approximately 12 percent as an average for eukaryotic proteins (*12*).

We next compared the composition and sequence of the OTC leader with three previously reported leader sequences from chloroplast and mitochondrial precursors—one each from pea, *Neurospora*, and yeast. The proteins involved and their origin, cellular localization, and leader sequences are shown in Table 1. Of these three proteins, the subunit of ribulose-1,5-bisphosphate carboxylase (RBPCase) from pea chloroplast resembles OTC most closely, because its destination, the thylakoid space, is the homolog in plant chloroplasts of the mitochrondrial matrix. The leader of this protein contains 57 residues (*13*), nearly twice as many as that predicted for the OTC leader. Its composition is striking—seven basic residues and only a single acidic residue. The same is true for the leader sequences of the proteolipid subunit of adenosinetriphosphatase from *Neurospora* (*14*), an inner membrane protein, and of cytochrome c peroxidase of yeast (*15*), an intermembrane space component. These leader sequences are composed of 19 and 10 percent basic residues, respectively, while neither contains a single acidic residue. A tally of the compositions of the four leader sequences (Table 1) indi-

Table 1. Leader sequences of mitochondrial and chloroplast precursors. Precursors whose leader sequences have been analyzed are indicated with their species of origin, cellular localization, and NH_2-terminal amino acid sequence. Sites of cleavage between leader sequences and mature subunits are indicated by a vertical arrow. Basic residues are underlined; acidic residues are overlined. A hydrophobic domain identified in the cytochrome c peroxidase leader is continuously underlined. The leader of the small subunit of ribulose-1,5-bisphosphate carboxylase (*13*) (RBPCase) contains a single acidic residue while those of the proteolipid subunit of adenosine triphosphate (ATP) synthase (*14*), cytochrome c peroxidase (*15*), and OTC are devoid of acidic residues. See text for discussion. Single letter amino acid designations were used: 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; Y, Tyr.

Polypeptide	Organism	Location	Amino acid sequence
Small subunit of RBPCase	Pea	Thylakoid space	MASMISSSAVTTVSRASRGQSAAVAPFGGLKSMTGFPVKKVNTDIT SITSNGGRVKC↓MQVWPPI...
Proteolipid subunit of ATP synthase	Neurospora crassa	Inner membrane	MASTRVLASRLASQMAASAKVARPAVRVAQVSKRTIQTGSPLQTLK RTQMTSIVNATTRQAFQKRA↓YSSEIAQAM...
Cytochrome c peroxidase	Yeast	Intermembrane space	MTTAVRLLPSLGRTAHKRSLYLFSAAAAAAAAAATFAYSQSHKRSS SSPGGGSNHGWNNWGKAAALAS↓TTP...
Ornithine transcarbamylase	Human	Matrix space	MLFNLRILLNNAAFRNGHNFMVRNFRCGQPLQ↓NKVQLKGRD...

cates that there is only a single acidic residue out of 223, less than 0.5 percent, compared with 30 basic residues, or 13.5 percent.

Three general possibilities for the role of the leader bear consideration. First, its sequence could function as a biological "zip code" or "address," directing delivery of the precursor to mitochondria via binding of the leader to outer membrane receptors. If this general model pertains, other leader sequences might be expected to resemble that of OTC. The sequences analyzed up to now do not bear such a resemblance, but they are derived from different species, where quite conceivably, the address used may be entirely different. Evidence for such an address system has been provided by the recent analysis of a nuclear-coded elongation factor, designated Ef TU, which localizes to the matrix of yeast mitochondria (*16*). The amino acid sequence of the NH_2-terminal portion of Ef TU is nearly identical to that of the NH_2-terminal portion of the leader of yeast cytochrome c peroxidase (see Table 1).

A second general possibility is that the charge of the leader plays a role in the interaction of the precursor with the outer membrane receptor. The striking lack of acidic residues in the leader suggests that absence of negative charge is important. Equally important, however, may be the relative excess of basic residues, which contribute an overall positive charge. The significance of this imbalance of charged residues is supported by the comparison of leader sequences of the small subunit of RBPCase derived from soybean (*17*), pea (*13*), and wheat (*18*). Four basic residues appear at nearly identical positions in these sequences, and only a single acidic residue appears in two of the sequences. Other residues, with the interesting exceptions of proline, and the sequence at −4 to −8 residues from the site of leader cleavage, are not highly conserved among these sequences (*18*).

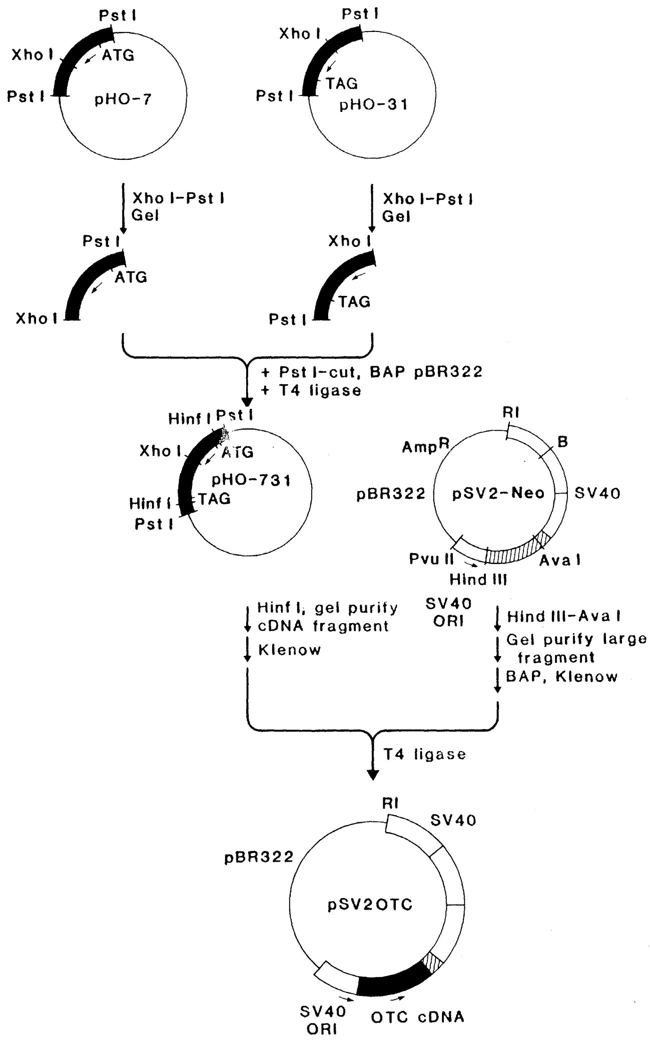

Fig. 4. Construction of plasmid joining human OTC sequence with SV40 regulatory elements. Plasmids pHO-7 and pHO-31 (Fig. 2), containing 5′ and 3′ OTC cDNA sequences, respectively, were digested and the DNA fragments were joined to create a plasmid with the entire OTC coding sequence, pHO-731. A Hinf fragment of pHO-731 was joined with a portion of the plasmid pSV2Neo as shown. In the derived plasmid, pSV2OTC, OTC sequences are joined at their 5′ terminus with SV40 sequences containing the early promoter and enhancer; and at the 3′ terminus with Tn 5 sequences (see text) and SV40 sequences containing the small-t intron and polyadenylation signal. Restriction endonuclease fragments were fractionated in 5 percent acrylamide gels, electroeluted, extracted with phenol, and precipitated with ethanol; creation of blunt-ended termini with the use of the Klenow fragment of polymerase I, and joining reactions using T4 DNA ligase, were performed according to the commercial supplier of the enzymes (New England Biolabs); bacterial transformations were performed according to Wensink *et al.* (*37*), with HB101 as the recipient; desired recombinants were identified by colony filter hybridization with nick-translated DNA fragments.

The third general possibility is that the basic composition of the leader alters the conformation of the mature portion of the precursor in such a way that recognition elements in the mature portion become "exposed." This formulation would predict that the leader sequence alone does not supply sufficient information for import of an adjoined polypeptide by the mitochondria, a hypothesis now testable by ligating nucleotide sequences for mitochondrial leader peptides to coding sequences for nonmitochondrial proteins.

The COOH-terminal amino acid residue of the leader is glutamine, and the NH_2-terminal residue of the mature subunit is asparagine. This sequence, acted on by the mitochondrial protease responsible for posttranslational cleavage, differs from those at the sites of proteolysis of the three other precursors shown in Table 1. Their cleavage sites resemble those of signal peptides, in which a residue with a short side chain, such as alanine, glycine, serine, or cysteine, is situated on the NH_2-terminal side (*11*). The cleavage site for OTC also differs from prohormone cleavage sequences, in which one or sometimes two basic residues are situated on the NH_2-terminal side (*11*). In fact, we have been unable to identify any previously reported cleavage site that contains glutamine on the NH_2-terminal side.

To test the biological activity of the cloned OTC sequences, a cDNA segment encoding the entire precursor was joined in a plasmid with SV40 regulatory elements, and introduced into HeLa cells. The steps performed to construct the plasmid are shown in Fig. 4. Because the OTC coding sequence was not present in a single cloned segment, it was necessary to join 5′ sequences contained in plasmid pHO-7 with 3′ sequences contained in pHO-31 (Fig. 2) through the single Xho I restriction site identified in the sequence shared by the two plasmids (Fig. 2). In this manner, a recombinant designated pHO-731, containing the entire OTC coding sequence was prepared (Fig. 4). To isolate a segment containing the complete OTC coding sequence and a minimum amount of flanking sequence, we digested plasmid pHO-731 with Hinf I, which cleaves at a site 16 bp upstream from the translational initiation codon (base −124, Fig. 3) and at a site 98 bp downstream from the translational terminator (base 1045, Fig. 3). This segment was then inserted into the plasmid pSV2Neo (*19*) which had been digested with Hind III and Ava I to remove all but 167 bp of the neomycin resistance (Tn 5) segment, a sequence containing an unassigned open reading frame, with no potential polyadenylation signal (*20*). The construct pSV2OTC joins the SV40 origin, including the enhancer and early promoter sequences, with the 5′ end of the OTC coding sequence; at the 3′ end, Tn 5 sequences are joined with SV40 sequences encoding the small-t intron and a polyadenylation signal (Fig. 4).

The plasmid pSV2OTC was introduced into HeLa cells by calcium phosphate precipitation (*21*). Forty-eight hours later the cells were incubated with labeled methionine for 1 hour, then harvested by lysis. The lysates were subjected to immune precipitation with antiserum to OTC. The OTC-specific products (Fig. 5, lanes 2 and 3), which were not present when pBR322 was used (lane 1), were clearly visible. When rhodamine 6G, an inhibitor of oxidative phosphorylation and mitochondrial processing, was added to pSV2OTC-transfected cells during the period of labeling, the immunoprecipitable OTC product (lane 2) comigrated with the human OTC precursor. In the absence of rhodamine 6G, the product (lane 3) comigrated precisely with the mature form of OTC.

Cotransfection of HeLa cells with pSV2OTC and the plasmid pSV2Neo, and subsequent selection by means of the drug G418 (*19*), permitted the establishment of a number of stable cell lines from which the mature form of OTC was precipitable with specific antiserum (not shown). Assay of extracts of these cell lines for OTC enzymatic activity revealed a range of specific activities from 0.49 to 1.5 μmole/hour per milligram of cell protein. This corresponds to 3 to 10 percent of OTC activity in human liver, and is particularly noteworthy because nontransfected HeLa cells contain no measurable OTC activity.

These expression studies establish beyond doubt both the identity and integrity of the cloned cDNA sequences. They show not only an OTC precursor conforming to its natural counterpart, but also its correctly localized and active mature subunit. Parenthetically, these results support previous findings both from our laboratory and from that of Mori *et al.* (*5*) indicating that liver-specific expression of OTC reflects transcriptional control, not mitochondrial diversity. Indeed, it seems likely that the mitochondrial portion of the pathway of biogenesis is ubiquitous in mammalian tissues. With our expression system, it should be possible to test various molecular constructs directed to provide understanding of the role of leader sequences and other, still obscure facets of the import process.

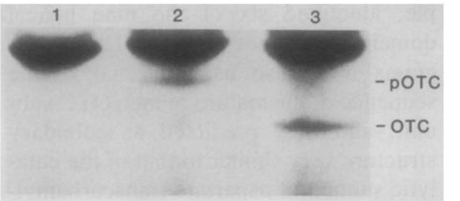

Fig. 5. Expression of OTC in HeLa cells transfected with pSV2OTC. HeLa cells were transfected with plasmid DNA, 1 μg per 10-cm dish, by the calcium phosphate coprecipitation method (*21*). Forty hours following transfection, cells were incubated with [^{35}S]-methionine (25 μCi/ml) in 4 ml of Puck's saline F, plus 10 percent dialyzed fetal calf serum and 50 m*M* glucose, for 1 hour, harvested by lysis with 0.25 percent SDS, and incubated with antiserum overnight. Immunocomplexes were recovered with *Staphylococcus aureus* cells, washed, and boiled in sample buffer; the solubilized labeled products were subjected to SDS-polyacrylamide gel electrophoresis, and the gels were subjected to fluorography. (Lane 1) Immunoprecipitate of cells transfected with pBR322; (lane 2) immunoprecipitate of cells transfected with pSV2OTC and labeled in the presence of rhodamine 6G (*38*) (1 μg/ml); (lane 3) immunoprecipitate of cells transfected with pSV2OTC. The heavy band present in all lanes corresponds in mobility with the 42-kD subunit of actin.

Because arginine is the penultimate product of the urea cycle, transcarbamylation of ornithine in mammals is a step in the pathway of arginine biosynthesis. This function appears to be highly conserved in evolution because *Escherichia coli* and yeast have analogous enzymatic activity, and become auxotrophic for arginine as a consequence of mutation at the OTC locus (*22*). Interestingly, the enzyme isolated from these organisms, like that of mammals, is composed of three identical subunits of approximately 36 kD (*23*). Therefore, we compared the predicted structure of the mature human OTC subunit with that of the two nearly identical and interchangeable OTC subunits, Arg F and Arg I (*23*), from *E. coli*. When Arg I (*24*) was aligned with the human sequence as shown in the top two lines of Fig. 6, 25 percent identity of the amino acid sequences was observed. A region of near identity (11 of 14 consecutive residues) was detected, correspond-

```
                                                    40
            1                  20                   |                     60                   80
            |                  |                   KKSG                   |                    |
                                                     V
E. coli OTC         SGFYHKHFLKLLDFTPAELNSL LQLAAKLKADKEQAKLTGKNIALIFEKDST RTRCSFEVAAYDOG
                         | |          |    |    | |       | ||    |||| || |||  | |      |
Human OTC   NKVQLKGRDLLTLKNFTGEEIKYMLWLSAD LKFRIKQKGEYLP LLQGKSLGMIFEKRST RTRLSTETGFALLG GHPCFPTTQDIHLG
                      |  |            |    |    | |    | ||  |     |   || |||||  |    ||     |       ||
E. coli ATC        ANPLYQKHIISINDLSRDDLNLV LATAAKLKANPQPELLKHKVIASCFFEAST RTRLSFQTSMHRLGASVVGFSDSANTSLGK
                                                                       |-------
                                                                          H2

                       100                  120                    140                  160
                        |                    |                      |                    |
Human OTC   VNESLTDTARVLSSMADAVLARVYKQSDLD TLAKEAS   IPIINGLSDLYHPIQILADY LTLQEHYSSLKGLTLSCFGDGNNILHSIMM
              | | ||  | |   ||   |   |      || | |        |     || | | |   | ||    |              ||  |
E. coli ATC KGETLADTISVISTYVDAIVMR HPQEGAA RLATEFSGNVPVLNAGDGSNQHPTQTLLDL FTIQETQGRL            DNLHVAMV
              |----------                                        |-------------
                  H3                                                   H5
                180                 200                  220                     240
                 |                   |                    |                       |
Human OTC   SAAKFGMHLQAATPKGYEPDASVTKLAEQY AKENGTKLLLTNDPLEAA   HGGNVLITD TWISMGREEEKKKRLQAFQGYQV      T
               | |      |      |           |       |   |    |   |                    | ||       |
E. coli ATC GDLKTGRTVHSLTQALAKFDGNRFYFIAPD ALAMPQYILDMLDEKGIAWSLHSSIEEVMV EVDILYMTRVQKERLDPSEYANVKAQFVLR


               260                 280                  300                  320
                |                   |                    |                    |
Human OTC   MKTAKVAASDWTFLHCLPRKPEEVDDEVFY SPRSLVFPEAENRKWTIMAVMVSLLTDYSP QLQKPKF
                  |      || |||  |   | |    |    |  | |      |     |
E. coli ATC ASDLHNAKANMKVLHPLPRVDEIATD VDK TPHAWYFQQAGNGIFARQALLALVLNRDLV L
                         --|
                         S10
```

Fig. 6. Comparison of amino acid sequence of human mature OTC subunit with the sequence of *E. coli* OTC subunit Arg I, and with *E. coli* aspartate transcarbamylase (*ATC*) catalytic subunit. The sequence of the mature subunit of human OTC is designated human OTC; that of the NH$_2$-terminal portion of the Arg I subunit (*24*) by *E. coli* OTC; and that of the catalytic subunit of ATC (*29*) by *E. coli* ATC. Sequences (single-letter abbreviations) are displayed in horizontal blocks. Each block contains three sets of 30 consecutive positions of comparison, and each set is separated from its neighbor (or neighbors) by a single space. The numbers shown above the blocks of compared sequences designate positions of amino acids in the human mature OTC sequence, numbered beginning with +1 at the NH$_2$-terminal asparagine residue. The sequence of the Arg I subunit was aligned with the human OTC sequence by visual inspection, with placement in the human sequence of a single gap of four residues. This gap is denoted in the Arg I sequence by placement of the corresponding residues, represented as KKSG, above the line of the Arg I sequence. The sequence of the catalytic subunit of *E. coli* ATC was aligned with the human sequence with the use of the computer program of Doolittle (*28*). Gaps introduced in this procedure of alignment are represented as such at positions of comparison. Vertical lines extending between the compared sequences indicate positions of amino acid identity. Where primary structural similarity involves a defined secondary structural domain in the ATC catalytic subunit (*31*), the region of similarity is underlined, and the domain is designated. Where such a region of similarity includes the border of a domain, the border is indicated by a perpendicular vertical mark.

ing in the human subunit to residues 50 to 63. Not only are NH_2-terminal sequences shared but additional regions of near identity, extending the length of the polypeptides, have been found using the DNA sequence of the Arg I subunit (*25*). These structural similarities could be the result of either convergent or, more likely, divergent evolution. Interestingly, OTC is cytoplasmic in *Saccharomyces cerevisiae* (*26*) and mitochondrial in *Neurospora crassa* (*27*). It should be revealing to determine if OTC from these lower eukaryotes contains a leader sequence, and to compare the mature subunit from them with those from *E. coli* and man.

A comparison of the human OTC sequence with other amino acid sequences by computer analysis (*28*) revealed an additional surprising similarity; a 25 percent match was detected when the sequence of the mature human OTC subunit was compared with that of the catalytic subunit of aspartate transcarbamylase of *E. coli* (*29*). A number of sequence "gaps" were introduced in this analysis (*28*), as shown in Fig. 6. Once again, both transcarbamylases have been noted to be composed of trimers of catalytic subunits of 36 kD; OTC contains one set of catalytic trimers, aspartate transcarbamylase has two sets (*30*), as well as three dimers of regulatory subunits. Both enzymes catalyze condensations that use carbamyl phosphate as substrate. Comparison of the primary amino acid sequences of the two proteins revealed regions containing nearly identical amino acid sequences (Fig. 6). As in the comparison with the Arg I subunit, a nearly identical sequence was identified in the region corresponding to residues 50 to 65 (Fig. 6) in the human protein. Other regions containing extensive "matching" included residues 96 to 103, 141 to 154, and 283 to 297 (Fig. 6).

Many of the secondary and tertiary structural features of aspartate transcarbamylase from *E. coli* have been defined (*31*), and many of the secondary structural features revealed by x-ray diffraction studies were predicted from the primary amino acid sequence by the computer programs of Chou and Fasman and of Garnier (*32*). The programs, for example, identified six of the nine helical domains (data not shown). These programs were also used to analyze the sequence of the mature human OTC subunit, and they predicted a secondary structure very similar to that of the catalytic subunit of aspartate transcarbamylase, including the same number of turns and similar regions of α-helix (data not shown).

Because of the possibility that the secondary structures of the two enzymes may be similar, it seems conceivable that functional elements of the tertiary structures may also be similar. The highly conserved region corresponding to residues 50 to 65 in the OTC subunit has been shown, in aspartate transcarbamylase, to lie in a helical domain that both comprises part of the catalytic site and provides a point of contact with a neighboring catalytic subunit (Fig. 6) (*31*). The conserved region corresponding to residues 141 to 154 of the human subunit appears in aspartate transcarbamylase to be involved in contact between catalytic subunits (Fig. 6) (*31*). These similarities leave little doubt that the two polypeptides are evolutionarily linked, and bring forth questions such as whether they evolved from a primal bifunctional ancestor or one from another, and how and why OTC acquired a leader sequence and mitochondrial localization.

References and Notes

1. G. Blobel and B. Dobberstein, *J. Cell Biol.* **67**, 852 (1975).
2. G. Schatz and R. A. Butow, *Cell* **32**, 316 (1983); M. Teintze and W. Neupert, in *Cell Membranes: Methods and Reviews*, E. Elson, W. A. Frazier, L. Glaser, Eds. (Plenum, New York, 1984), vol. 1, pp. 89–114.
3. S. Anderson *et al.*, *Nature (London)* **290**, 457 (1981).
4. F. C. Ricciuti, T. D. Gelehrter, L. E. Rosenberg, *Am. J. Hum. Genet.* **28**, 332 (1976); R. DeMars, S. L. LeVan, B. L. Trend, L. B. Russell, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 1693 (1976).
5. J. G. Conboy and L. E. Rosenberg, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3073 (1981); M. Mori, S. Miura, T. Marita, M. Takiguchi, M. Tatibana, *Mol. Cell. Biochem.* **49**, 97 (1982); L. E. Rosenberg, F. Kalousek, M. D. Orsulak, *Science* **222**, 426 (1983); D. M. Kolansky, J. G. Conboy, W. A. Fenton, L. E. Rosenberg, *J. Biol. Chem.* **257**, 8467 (1982); J. G. Conboy *et al.*, *Biochem. Biophys. Res. Commun.* **105**, 1 (1982).
6. M. L. Maccecchini, Y. Rudin, G. Blobel, G. Schatz, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 343 (1979).
7. A. L. Horwich, J. P. Kraus, K. Williams, F. Kalousek, W. Konigsberg, L. E. Rosenberg, *ibid.* **80**, 4258 (1983).
8. S. Orkin kindly supplied a cDNA library derived from adult human liver.
9. F. Kalousek and K. Williams, unpublished observations.
10. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975).
11. D. F. Steiner, P. S. Quinn, S. J. Chan, J. Marsh, H. S. Tager, *Ann. N.Y. Acad. Sci.* **343** 1 (1980).
12. G. D. Fasman, Ed., *Handbook of Biochemistry and Molecular Biology* (CRC Press, Cleveland, ed. 3, 1976).
13. A. R. Cashmore, in *Genetic Engineering of Plants*, T. Kossugge, C. P. Meredith, A. Hollaender, Eds. (Plenum, New York, 1983), pp. 29–38.
14. A. Viebrock, A. Perz, W. Sebald, *EMBO J.* **1**, 565 (1982).
15. J. Kaput, S. Goltz, G. Blobel, *J. Biol. Chem.* **257**, 15054 (1982).
16. S. Nagata, Y. Tsunetsugu-Yokota, A. Naito, Y. Kaziro, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6192 (1983).
17. S. L. Berry-Lowe, T. D. McKnight, D. M. Shah, R. B. Meagher, *J. Mol. Appl. Genet.* **1**, 483 (1982).
18. R. Broglie, G. Coruzzi, G. Lamppa, B. Keith, N.-H. Chua, *Biotechnology* **1**, 55 (1983).
19. P. J. Southern and P. Berg, *J. Mol. Appl. Genet.* **1**, 327 (1982).
20. E. Beck, G. Ludwig, E. A. Auerswald, B. Reiss, H. Schaller, *Gene* **19**, 327 (1982).
21. F. L. Graham and A. J. van der Eb, *Virology* **52**, 456 (1973); M. Wigler, A. Pellicer, S. Silverstein, R. Axel, G. Urlaub, L. Chasin, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1373 (1979).
22. A. Kikuchi, D. Elseviers, L. Garni, *J. Bacteriol.* **122**, 727 (1975); M. Crabeel, F. Messenguy, F. LaCroute, N. Glansdorff, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5026 (1981).
23. C. LeGrain, P. Halleux, V. Stalon, N. Glansdorff, *Eur. J. Biochem.* **27**, 93 (1972); M. Penninckx, J. P. Simon, J.-M. Wiame, *ibid.* **49**, 429 (1974).
24. J. Piette, D. Gigot, M. Lauwereys, R. Cunin, N. Glansdorff, A. Pierard, *Arch. Int. Physiol. Biochim.* **90**, B211 (1982); S. K. Moore, thesis, University of South Carolina (1981).
25. D. A. Bencini, J. E. Houghton, T. A. Hoover, K. F. Foltermann, J. R. Wild, G. A. O'Donovan, *Nucleic Acids Res.* **11**, 8509 (1983).
26. Urrestarazu, S. Vissers, J.-M. Wiame, *Eur. J. Biochem.* **79**, 473 (1977).
27. R. L. Weiss and R. H. Davis, *J. Biol. Chem.* **248**, 5403 (1973).
28. R. F. Doolittle, *Science* **214**, 149 (1981).
29. T. A. Hoover, W. D. Roof, W. D. Foltermann, G. A. O'Donovan, D. A. Bencini, J. R. Wild, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2462 (1983); W. H. Konigsberg and L. Henderson, *ibid.*, p. 2467.
30. J. A. Cohlberg, V. P. Pigiet, Jr., H. K. Schachman, *Biochemistry* **11**, 3396 (1972).
31. H. L. Monaco, J. L. Crawford, W. N. Lipscomb, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5276 (1978); R. B. Honzatka *et al.*, *J. Mol. Biol.* **160**, 219 (1982); R. B. Honzatka and W. N. Lipscomb, *ibid.*, p. 265.
32. P. Y. Chou and G. D. Fasman, *Annu. Rev. Biochem.* **47**, 51 (1978); J. Garnier, D. J. Osguthorpe, B. Robson, *J. Mol. Biol.* **120**, 97 (1978).
33. J. M. Chirgwin, A. E. Przybyla, R. J. MacDonald, W. J. Rutter, *Biochemistry* **18**, 5294 (1979).
34. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
35. K. R. Williams *et al.*, in *Methods in Protein Sequence Analysis*, M. Elzinga, Ed. (Humana Press, Clifton, N.J., 1982), pp. 499–507.
36. J. J. L'Italien and J. E. Strickler, *Anal. Biochem.* **127**, 198 (1982).
37. P. C. Wensink, D. J. Finnegan, J. E. Donelson, D. S. Hogness, *Cell* **3**, 315 (1974).
38. A. R. L. Gear, *J. Biol. Chem.* **249**, 3628 (1974).
39. Francis Collins, Ron Evans, Mike Liskay, Jack Rose, and Peter Southern for helpful suggestions; Adelle Hack, Michelle Orsulak, and Cynthia Williamson for technical assistance. Supported by Public Health Service grants GM 31539, AM 09527, AM 12579, and GM 32156.

24 January 1984; accepted 22 March 1984