Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 2244–2246, April 1985
Biochemistry

The gene coding for carbamoyl-phosphate synthetase I was formed by fusion of an ancestral glutaminase gene and a synthetase gene

(gene fusion/enzyme evolution/cDNA clones/urea cycle enzymes)

HIROSHI NYUNOYA, KAREN E. BROGLIE, AND C. J. LUSTY

Molecular Genetics Laboratory, The Public Health Research Institute of The City of New York, Inc., New York, NY 10016

Communicated by Sarah Ratner, November 28, 1984

ABSTRACT **A near full-length cDNA copy of rat carbamoyl-phosphate synthetase I (EC 6.3.4.16) mRNA has been cloned. The cDNA insert in the recombinant plasmid pHN234 is 5.3 kilobases long. Analysis of the sequence coding for carbamoyl-phosphate synthetase I indicates that the gene has arisen from a fusion of two ancestral genes: one homologous to *Escherichia coli carA*, coding for a glutaminase subunit, and the second homologous to the *carB* gene that codes for the synthetase subunit. A short amino acid sequence previously proposed to be part of the active site involved in glutamine amide nitrogen transfer in the *E. coli* and yeast carbamoyl-phosphate synthetases (EC 6.3.5.5) is also present in the rat enzyme. In the mammalian enzyme, however, the glutaminase domain lacks a cysteine residue previously shown to interact with glutamine. The cysteine is replaced by a serine residue. This substitution could, in part, account for the inability of mammalian carbamoyl-phosphate synthetase I to catalyze the hydrolysis of glutamine to glutamic acid and ammonia.**

Carbamoyl-phosphate synthetase I [carbon-dioxide:ammonia ligase (ADP-forming, carbamate-phosphorylating); EC 6.3.4.16] provides carbamoyl phosphate for arginine and urea biosynthesis. In ureotelic animals, the enzyme occurs only in the liver and small intestine (ref. 1 and J. Ryall, M. Nguyen, M. Bendayan, and G. C. Shore, personal communication) where it is localized in the matrix compartment of the mitochondria. Mitochondrial carbamoyl-phosphate synthetase consists of a single subunit of M_r 160,000 (2, 3). The protein is encoded in the nuclear genome and is synthesized on free polysomes as a precursor with $M_r \approx 165{,}000$ (4). The precursor is incorporated post-translationally by mitochondria, where the mature enzyme constitutes 20–25% of the matrix protein (2, 3, 5). Levels of carbamoyl-phosphate synthetase I mRNA have been shown to be regulated developmentally, by hormones, and by diet (6).

Carbamoyl-phosphate synthetase I catalyzes the synthesis of carbamoyl phosphate from HCO_3^-, 2 ATPs, and NH_3 in a reaction that requires the allosteric activator *N*-acetyl-L-glutamate (7–9). This enzyme differs in its allosteric regulation, substrate specificity, and subunit structure from carbamoyl-phosphate synthetases [carbon-dioxide:L-glutamine amidoligase (ADP-forming, carbamate-phosphorylating); EC 6.3.5.5] of most bacteria and lower eukaryotes (yeast and *Neurospora crassa*). These latter enzymes utilize glutamine or NH_4^+ as nitrogen donor for carbamoyl phosphate synthesis. They are composed of two distinct subunits (10–13). A smaller subunit (M_r 42,000) transfers glutamine amide nitrogen to a larger subunit (M_r 118,000) that utilizes enzyme-bound NH_3 for carbamoyl phosphate synthesis from HCO_3^- and ATP (10, 12, 13). NH_4^+ has been shown to be a substrate for carbamoyl phosphate synthesis by the large subunit (10, 12, 13).

Sequence analyses of the genes of the prokaryotic and yeast arginine-specific carbamoyl-phosphate synthetases have shown these enzymes to be evolutionarily related (14–17). The present studies were undertaken with the general aim of establishing the relation of the mammalian ammonia-dependent carbamoyl-phosphate synthetase to the glutamine-dependent enzymes. We report the successful cloning of a cDNA complementary to rat liver carbamoyl-phosphate synthetase I mRNA. Nucleotide sequence analysis of the rat cDNA indicates that the mammalian gene is related to both the prokaryotic and fungal genes for the glutaminase and synthetase subunits of carbamoyl-phosphate synthetase.

MATERIALS AND METHODS

DNA and Enzymes. Enzymes were purchased from New England Biolabs, P-L Biochemicals, and Amersham. Reverse transcriptase was a gift from J. Beard (Life Sciences, St. Petersburg, FL). Plasmid DNA from *Escherichia coli* RR1 was prepared according to the method of Birnboim and Doly (18) and purified by chromatography on Sepharose 6B.

Rat Liver mRNA. Rat liver polysomes were isolated from livers of Wistar strain rats that had been maintained for 7 days on a 60% protein diet. The polyribosomal RNA was extracted with guanidinium thiocyanate by the procedure of Chirgwin *et al.* (19) and enriched for poly(A)$^+$ RNA by passage of the total polysomal RNA through oligo(dT)-cellulose. The isolated mRNA preparation was confirmed to contain functional, full-length transcripts by *in vitro* translation and immunoprecipitation of the translation products with antibodies specific for rat liver carbamoyl-phosphate synthetase I. Autoradiographs showed a single radioactive band corresponding in size to the precursor form of the rat enzyme (unpublished data).

Construction of a cDNA Library. A cDNA library of the purified rat liver mRNA was made by the method of Okayama and Berg (20). The cDNA recombinant library in *E. coli* RR1 consisted of $\approx 1.5 \times 10^6$ independent clones.

Selection of Carbamoyl-Phosphate Synthetase I cDNA Clones. The cDNA library was screened by colony hybridization according to the method of Hanahan and Meselson (21). The hybridization probe used in the screen was a nick-translated *Xho* I/*Eco*RI fragment encompassing the sequence coding for the 345 carboxyl-terminal residues of carbamoyl-phosphate synthetase I (unpublished data). This clone, pKB21, was obtained from a conventional pBR322 cDNA library that had been screened with a 770-base-pair (bp) cDNA insert from a clone, pCPSr1, kindly provided by W. E. O'Brien (Baylor College of Medicine). An initial screening of the Okayama–Berg cDNA library (120,000 transformants) yielded 34 confirmed positive clones. The longest clone, pHN107, was ascertained to have a cDNA insert of 4.4 kilobase pairs (kbp). A restriction fragment (*Xho* I/*Bst*EII) from the 5′-proximal end of pHN107 was used as a

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: kbp, kilobase pair(s); kb, kilobase(s).

probe for a larger screening involving 1.8×10^6 clones. This screening produced another 135 clones, of which the longest (pHN234) had an insert of 5.3 kbp. Based on the estimated size of the mRNA, 5.7 ± 0.3 kilobases (kb) (unpublished data), pHN234 should encode almost the entire sequence of the mRNA. The sequence data reported in this paper were obtained from pHN234.

DNA Sequence Analysis. The method of Maxam and Gilbert (22) was used for sequencing selected regions of pHN234. All of the sequence data were obtained from 5'-end-labeled fragments after separation of the single strands on polyacrylamide gels (22).

RESULTS AND DISCUSSION

Restriction Analysis of pHN234. pHN234 was established to have a cDNA insert of 5.3 kbp. A restriction map of the cloned cDNA is presented in Fig. 1. The restriction map agreed with that of the shorter clones pHN107 (4.4 kbp) and pKB21 (1.5 kbp), both of which were used for the selection of pHN234.

The primary objective of this study was to determine whether mammalian carbamoyl-phosphate synthetase I is a gene-fusion product consisting of a synthetase subunit homologous to the large synthetase subunit of the prokaryotic carbamoyl-phosphate synthetase and to a smaller subunit, which in *E. coli* catalyzes the hydrolysis of glutamine and transfer of ammonia.

Sequence analysis of pKB21 indicated that it codes for the 381 carboxyl-terminal amino acids of the rat liver carbamoyl-phosphate synthetase protein (unpublished data). Furthermore, the nucleotide sequence of the cDNA clone pKB4 showed that the coding frame terminates 905 nucleotides from the poly(A) tail (6). Since the amino acid sequence encoded in the two clones was found to be homologous to the synthetase subunits of both yeast and *E. coli*, it was possible to estimate that the site of fusion with the glutaminase gene would be approximately 4.2 kb upstream of the poly(A) tail.

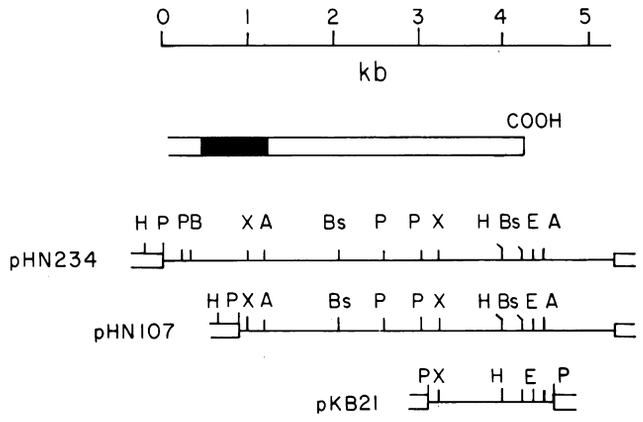

FIG. 1. Restriction maps of cDNA clones. The smallest clone, pKB21, was used to obtain the *Xho* I/*Eco*RI probe for the selection of pHN107. The longest clone, pHN234, was isolated by using the 5' proximal *Xho* I/*Bst*EII fragment of pHN107 as a probe. The cDNA insert in pHN234 contains almost the entire coding sequence plus ≈1 kb of 3' flanking sequence of the mRNA. The location of the coding region of the gene is indicated by the bar above the restriction maps of the three clones; COOH denotes the carboxyl-terminal end of the encoded polypeptide. The cDNA region sequenced (indicated by the shaded region of the bar) includes 800 nucleotides spanning the most 5' *Xho* I site. The restriction sites for *Hin*dIII (H), *Pst* I (P), *Bam*HI (B), *Xho* I (X), *Acc* I (A), *Bst*EII (Bs), and *Eco*RI (E) are indicated above each restriction map. pHN234 and pHN107 were isolated from the cDNA library constructed by the method of Okayama and Berg (20). pKB21 was obtained from a library constructed by inserting double-stranded cDNA into the *Pst* I site of pBR322 (23).

This region of pHN234 was isolated on three separate restriction fragments and the entire sequence of both strands was determined [*Hin*dIII/*Xho* I, *Xho* I/*Bst*EII, and *Bam*HI/*Acc* I (Fig. 1)].

Evidence for a Fusion of Ancestral Glutaminase and Synthetase Subunits in the Rat Carbamoyl-Phosphate Synthetase I Gene. The region of the rat liver cDNA sequenced is shown in Fig. 1. This region of approximately 800 nucleotides starts 400 nucleotides from the *Pst* I site of the vector in pHN234. Based on the length of the synthetase subunits of *E. coli* and yeast, the predicted site of fusion should be close to the upstream *Xho* I site, 960 nucleotides from the start of the insert. The amino acid sequence deduced from the nucleotide sequence of this region is presented in Fig. 2. Also shown in Fig. 2 are the sequences of the carboxyl-terminal ends of the yeast and *E. coli* glutaminase subunits and of the amino-terminal regions of the respective synthetase subunits. The region of nonhomology at the junction of the small and large subunits in the rat sequence comprises only 15 amino acid residues, suggesting that the gene for carbamoyl-phosphate synthetase I was formed by a simple gene fusion event.

Evolutionary Changes in the Glutamine Catalytic Domain. Our studies of the small subunits of *E. coli* and yeast carbamoyl-phosphate synthetase, together with previously reported sequences of other amidotransferases, indicated that these enzymes have a common domain consisting of 13 amino acids (17). One of the salient features of this domain, proposed to be part of the glutamine catalytic site, is the presence of a cysteine residue, which has been shown in the cases of anthranilate synthase (EC 4.1.3.27) (24, 25) and phosphoribosylformylglycinamidine synthetase (EC 6.3.5.3) (26, 27) to be directly involved in the binding of the substrate glutamine. Furthermore, Buchanan (28) and Nagano *et al.* (29) have proposed that this cysteine participates in the formation of a γ-glutamyl thioester as part of the catalytic mechanism. Since mammalian carbamoyl-phosphate synthetase I utilizes free NH_3 but not glutamine as the source of ammonia nitrogen, we were especially interested in any changes that might have occurred in the glutamine catalytic domain of the enzyme. This domain is still recognizable in the rat carbamoyl-phosphate synthetase I sequence (underlined in Fig. 2). Out of 13 amino acid residues, there are five identities and two conservative substitutions among the three enzymes. The most significant difference noted between the yeast, *E. coli*, and rat sequences is that the rat sequence contains serine in place of the reactive cysteine residue.

The conservation of sequence homology flanking the serine in carbamoyl-phosphate synthetase I suggests that this domain may have retained a function in either acetylglutamate or NH_3 binding. Even though the large subunits of both *E. coli* and fungal carbamoyl-phosphate synthetases have been shown to utilize NH_4^+ for carbamoyl phosphate synthesis, the affinity for NH_4^+ is low. The K_m for free NH_4^+ has been estimated to be 90 mM for the *E. coli* enzyme (30) and 16.6 mM for the *N. crassa* large subunit (13). These values are 17- to 90-fold higher than that of the mammalian enzyme. If the proposed glutamine catalytic domain in the small subunit is involved not only in hydrolysis of glutamine but also in transfer of the liberated ammonia to the synthetase site, the absence of the cysteine residue in the mammalian enzyme would be expected to result in a loss of hydrolytic activity but not necessarily of NH_3 binding or of its transfer to an acceptor site in carbamoyl-phosphate synthetase I. The modified glutamine domain may also have acquired a new function as an allosteric site for *N*-acetylglutamate, a known effector of ammonia-dependent carbamoyl-phosphate synthetases. These questions can be clarified from studies of carbamoyl-phosphate synthetase III of invertebrates (31) and elasmobranch and teleost fishes (32–34). This enzyme is

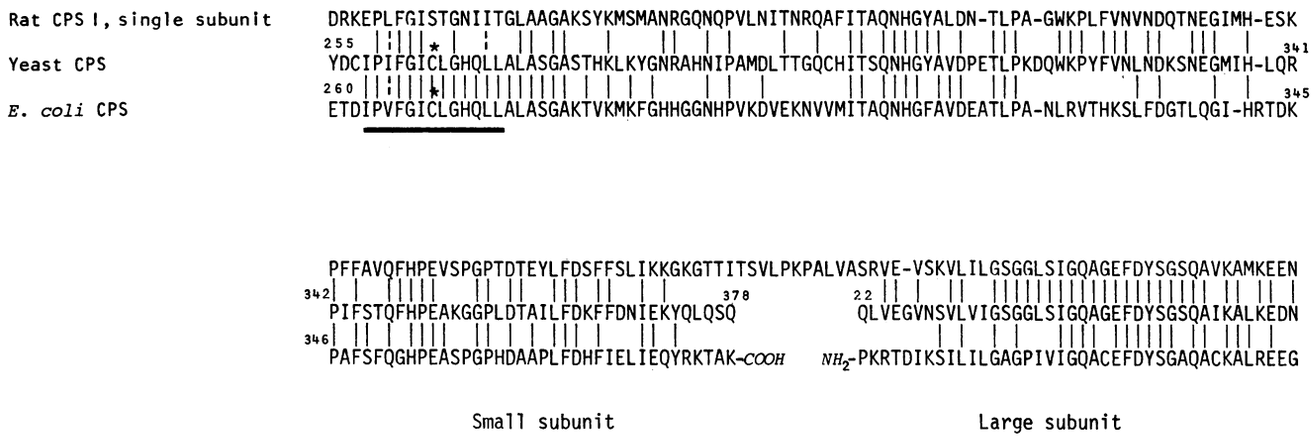

FIG. 2. Amino acid sequence homology between the yeast and *E. coli* small and large subunits and the single subunit of rat carbamoyl-phosphate synthetase I (CPS I). The alignments of the yeast and *E. coli* subunits are those proposed previously (16, 17). The amino acid sequences of rat, yeast, and *E. coli* carbamoyl-phosphate synthetases (CPS) are each continuous between the upper and lower parts of the figure. The numbers above the yeast and *E. coli* sequences indicate the position of the amino acid residues in the previously published sequences (14–17). The yeast small subunit has an additional 33 amino acids beyond Gln-378. The NH_2 terminus of the large subunit of yeast carbamoyl-phosphate synthetase is 21 amino acids before the sequence shown. The COOH terminus (Lys) of the *E. coli* small subunit and the NH_2 terminus (Pro) of the large subunit are so indicated. The glutamine active site of the yeast and *E. coli* small subunits proposed previously (17) is underlined. The reactive cysteines, Cys-264 and Cys-269, of the yeast and *E. coli* proteins, respectively, are indicated by asterisks. The vertical lines indicate identical amino acids; the broken lines denote functionally conserved hydrophobic residues in the glutamine domain. Dashes indicate spaces inserted to maximize homology. The single-letter amino acid abbreviations 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.

known to utilize glutamine as the nitrogen donor for carbamoyl phosphate synthesis, but, like mammalian carbamoyl-phosphate synthetase, it is a single polypeptide ($M_r \approx$ 160,000) requiring *N*-acetylglutamate for enzymatic activity.

We thank Roy Smith for the computer analyses of the sequences. These studies were supported by Grant GM25846 from the National Institutes of Health.

1. Jones, M. E., Anderson, A. D., Anderson, C. & Hodes, S. (1961) *Arch. Biochem. Biophys.* **95,** 499–507.
2. Lusty, C. J. (1978) *Eur. J. Biochem.* **85,** 373–383.
3. Clarke, S. (1976) *J. Biol. Chem.* **251,** 950–961.
4. Shore, G. C., Carignan, P. & Raymond, Y. (1979) *J. Biol. Chem.* **254,** 3141–3144.
5. Raijman, L. & Jones, M. E. (1976) *Arch. Biochem. Biophys.* **175,** 270–278.
6. Ryall, J., Rachubinski, R. A., Nguyen, M., Rozen, R., Broglie, K. E. & Shore, G. C. (1984) *J. Biol. Chem.* **259,** 9172–9176.
7. Allen, C. M., Jr., & Jones, M. E. (1966) *Arch. Biochem. Biophys.* **114,** 115–122.
8. Guthöhrlein, G. & Knappe, J. (1968) *Eur. J. Biochem.* **7,** 119–127.
9. Alonso, E. & Rubio, V. (1983) *Eur. J. Biochem.* **135,** 331–337.
10. Trotta, P. P., Burt, M. E., Haschemeyer, R. H. & Meister, A. (1971) *Proc. Natl. Acad. Sci. USA* **68,** 2599–2603.
11. Matthews, S. L. & Anderson, P. M. (1972) *Biochemistry* **11,** 1176–1183.
12. Piérard, A. & Schröter, B. (1978) *J. Bacteriol.* **134,** 167–176.
13. Davis, R. H., Ristow, J. L. & Hanson, B. A. (1980) *J. Bacteriol.* **141,** 144–155.
14. Nyunoya, H. & Lusty, C. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 4629–4633.
15. Piette, J., Nyunoya, H., Lusty, C. J., Cunin, R., Weyens, G., Crabeel, M., Charlier, D., Glansdorff, N. & Piérard, A. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 4134–4138.
16. Lusty, C. J., Widgren, E. E., Broglie, K. E. & Nyunoya, H. (1983) *J. Biol. Chem.* **258,** 14466–14472.
17. Nyunoya, H. & Lusty, C. J. (1984) *J. Biol. Chem.* **259,** 9790–9798.
18. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7,** 1513–1523.
19. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
20. Okayama, H. & Berg, P. (1982) *Mol. Cell. Biol.* **2,** 161–170.
21. Hanahan, D. & Meselson, M. (1983) *Methods Enzymol.* **100,** 333–342.
22. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
23. Land, H., Grez, M., Hauser, H., Lindenmaier, W. & Schütz, G. (1981) *Nucleic Acids Res.* **9,** 2251–2266.
24. Tso, J. Y., Hermodson, M. A. & Zalkin, H. (1980) *J. Biol. Chem.* **255,** 1451–1457.
25. Kawamura, M., Keim, P. S., Goto, Y., Zalkin, H. & Heinrikson, R. L. (1978) *J. Biol. Chem.* **253,** 4659–4668.
26. Dawid, I. B., French, T. C. & Buchanan, J. M. (1963) *J. Biol. Chem.* **238,** 2178–2185.
27. Ohnoki, S., Hong, B.-S. & Buchanan, J. M. (1977) *Biochemistry* **16,** 1070–1076.
28. Buchanan, J. M. (1973) in *The Enzymes of Glutamine Metabolism*, eds. Prusiner, S. & Stadtman, E. R. (Academic, New York), pp. 387–408.
29. Nagano, H., Zalkin, H. & Henderson, E. J. (1970) *J. Biol. Chem.* **245,** 3810–3820.
30. Anderson, P. M., Wellner, V. P., Rosenthal, G. A. & Meister, A. (1970) *Methods Enzymol.* **17A,** 235–243.
31. Tramell, P. R. & Campbell, J. W. (1971) *Comp. Biochem. Physiol. B* **40,** 395–406.
32. Anderson, P. M. (1980) *Science* **208,** 291–293.
33. Anderson, P. M. (1981) *J. Biol. Chem.* **256,** 12228–12238.
34. Casey, C. A. & Anderson, P. M. (1983) *J. Biol. Chem.* **258,** 8723–8732.