Mammalian Carbamyl Phosphate Synthetase (CPS)

cDNA SEQUENCE AND EVOLUTION OF THE CPS DOMAIN OF THE SYRIAN HAMSTER MULTIFUNCTIONAL PROTEIN CAD*

(Received for publication, January 31, 1990)

James P. Simmer, Ruth E. Kelly*, Austin G. Rinker, Jr., Joshua L. Scully, and David R. Evans

From the Department of Biochemistry, Wayne State University School of Medicine, Detroit, Michigan 48201

Glutamine-dependent carbamoyl-phosphate synthetase (EC 6.3.5.5) catalyzes the first step in de novo pyrimidine biosynthesis. The mammalian enzyme is part of a 240-kDa multifunctional protein which also has the second (aspartate carbamoyltransferase, EC 2.1.3.2), and third (dihydroorotase, EC 3.5.2.3) activities of the pathway. Shigesada et al. (Shigesada, K., Stark, G. R., Maley, J. A., and Davidson, J. N. (1985) Mol. Cell Biol. 175, 1-7) produced a truncated cDNA clone from a Syrian hamster cell line that contained most of the coding region for this protein. We have completed sequencing this clone, known as pCAD142. The cDNA insert contained all of the coding region for the glutaminase (GLN) and carbamyl phosphate synthetase (CPS) domains but lacked a short amino-terminal segment. By comparing the primary structure of the mammalian chimeric to multifunctional proteins we have identified the borders of the functional domains. The GLN domain is 21 kDa, close to the size of the functionally similar polypeptide products of the Escherichia coli pabA and hisH genes. The domain has the three regions of homology common to trpG-type glutamine amidotransferases, as well as a fourth region specific to the carbamyl phosphate synthetase (CPS) domain. The CPSase domain is similar to other reported CPSases in size (120 kDa), primary structure (37-67% amino acid identity), and homology between its amino and carboxyl halves. Analysis of the nucleotide and amino acid sequence identities among the various carbamyl phosphate synthetases suggests that the gene fusion which joined the GLN and CPS domains was an early event in the evolution of eukaryotic organisms and that the Saccharomyces cerevisiae enzyme consisting of separate subunits arose by defusion from an ancestral multifunctional protein.

Glutamine-dependent carbamoyl phosphate synthetase

* This work was supported by National Cancer Institute Grants CA27674 and T32 CA09531 (to J. P. S.) and an Office of Naval Research Contract N00014-87-K-0081. The costs of publication of this article were defrayed in part by the payment of page charges. The sequencing studies of Lusty and co-workers (33-36) and others (37) have shown that the carbamyl phosphate synthetase domain is typically 120 kDa in size and comprised of highly homologous amino and carboxyl halves, the result

1 The abbreviations used are: GlnCPSase, glutamine-dependent carbamoyl phosphate synthetase; ATCase, aspartate transcarbamylase; CACCcase, chicken acetyl-CoA carboxylase; CAD, polypeptide having CPSase, ATCase, and DHOase activities; CPS, carbamyl phosphate synthetase domain; CPSase, carbamyl phosphate synthetase activity; CPSI, mitochondrial ammonia-dependent carbamyl phosphate synthetase; DHOase, dihydroorotase; GLN, glutaminase domain, GLNase, glutaminase, GinCPS, polypeptide including the glutaminase and carbamyl phosphate synthetase domains; YPCase, yeast pyruvate carboxylase.

10395
of an ancestral gene duplication. Studies of the *E. coli* enzyme (38, 39), and CPSI (40, 41), demonstrated the existence of two distinct ATP binding sites. This symmetry of structure and function suggested (33, 42) that the two homologous halves act separately to catalyze the partial reactions, the phosphorylation of bicarbonate to carboxy phosphate (43, 44) and the phosphorylation of carbamate to carbamyl phosphate, formulated by Jones (45).

**EXPERIMENTAL PROCEDURES**

The plasmid pCAD142 (9), a pHBluescript construct containing a 6.5-kilobase abbreviated Syrian hamster CAD cDNA insert, was kindly provided by George Stark (Imperial Cancer Research Fund Laboratories, London) and Jeffrey Davidson (University of Kentucky). The insert carries 6215 base pairs of coding sequence, including the stop codon located about 300 base pairs from the 3' end, but lacks an estimated 600 base pairs from the 5' end of the message.

The plasmid was isolated from the *E. coli* host (UC6C4517) using the method of Davis et al. (46). Supercoiled DNA was purified by

**FIG. 1. Nucleotide sequencing strategy.** The arrangement of the coding regions on pCAD142 is shown above a partial restriction map of the region sequenced. The map units represent the distance in kilobases from the start of the cDNA insert. Arrows represent regions read from different subclones or using different oligonucleotide primers.
cesium chloride centrifugation (47). The purified plasmid was restricted with various restriction endonucleases and selected fragments purified by electrophoresis in 0.7% low melting point agarose (48) and ligated into the polylinker region of pBS+ (Stratagene). Using the hamster unique primer these subclones produced sequence in the rightward direction starting at the restriction sites: Acel (885), KpnI (2783), PstI (-223, 2230, 3138), PvuII (303, 1165, 2668), SalI (3742), (3162), Smal (630), and SalI (1686, 2019). A nested set of nine subclones were made starting with KpnI (2782) inserted in pKS+(Stratagene). The recombinant was digested with SstI, HindIII, (3162), SmaI (630), and SstI (1686, 2019). A nested set of nine subclones were made starting with KpnI (2782) inserted in pKS+(Stratagene). The recombinant was digested with SstI, HindIII, SmaI, and SstI (1686, 2019). The DNA sequencing reactions were analyzed by electrophoresis in both 6 and 8% polyacrylamide gels (33 3441. Oligonucleotides priming in the leftward direction started at nucleotide positions 118, 416, 717, 1007, 1333, 1566, 1847, 2384, and 2668. Oligonucleotides priming in the rightward direction began at the restriction sites: AccI (885), KpnI (2782), PstI (2230, 3138), PvuII (3162), SmaI (630), and SalI (3742). These oligonucleotides began at nucleotide positions 206, 416, 775, 2658, 2936, 3382, 3573, 3899, and 4036.

The DNA sequencing reactions were determined by the Sanger method (49) using 5'-O-[32P]thiotriphosphate (Du Pont-New England Nuclear; 600 Ci/mmol) following the Bethesda Research Laboratories (BRL) protocol (50). Oligonucleotides, 20 bases in length, were synthesized for priming in regions distant from convenient restriction sites. In the rightward direction, these oligonucleotides began at nucleotide positions 118, 416, 717, 1007, 1333, 1566, 1847, 2384, and 3441. Oligonucleotides priming in the leftward direction started at positions 206, 416, 775, 2658, 2936, 3382, 3573, 3899, and 4036.

The sequence of the GLNase and CPSase domains of CAD was determined by sequencing both strands of the pCAD142 insert following the strategy outlined in Fig. 1. The nucleotide sequence (Fig. 2) begins at the end of the poly(G) segment created during cloning (53) and extends for over 3900 nucleotides to the end of the CPS coding region and confirms approximately 800 nucleotides reported at the junction of the glutaminase and synthetase domains (54). This complete coding sequence found on pCAD142 but leaves undetermined the short 5' region of the mRNA coding for the amino terminus. This latter region of the protein has no known catalytic activity but is necessary for the association of the E. coli GLNase and CPSase subunits (55).

The deduced amino acid sequence of the glutaminase (Fig. 3) and carbamyl phosphate synthetase (Fig. 4) was aligned to those of other CPSases. Based upon the size of the E. coli pabA gene product the GLNase domain of CAD starts at glycine 19 and ends at threonine 201. The abrupt appearance of strong homology at lysine 240 marks the amino terminus of the CUff. Alignment of the amino half of the CAD CPSase (Lys-240 to Pro-778) with the carboxyl half (His-779 to Cys-1300) gave a 26% sequence identity and identified histidine 779 as the start of the carboxyl half. The CPSase extends to cysteine 1300 where the dihydroorotase domain begins (12). The synthetase, as defined, is 8 residues shorter on the amino terminus and 6 residues shorter on the carboxyl terminus than the CAD product (Ecoli). The con-}

To avoid bias toward a particular sequence, the homology plots were calculated by summing the number of identities for each residue in the aligned sequences using each of the proteins, in turn, as the reference sequence. The total number of identities for all of the comparisons was then summed and the values smoothed with a window size of five. The resulting homology index was plotted against the residue number deleting any residue not found in the hamster sequence so as to maintain a consistent numbering system. The amidotransferase homology plot compared hamster CAD, ura2, pab1, and E. coli carA gene products. The CPS plot compared hamster CAD, rat CPSI, and the pab2, and E. coli carB gene products.

## RESULTS

The sequence of the GLNase and CPSase domains of CAD was determined by sequencing both strands of the pCAD142 insert following the strategy outlined in Fig. 1. The nucleotide sequence (Fig. 2) begins at the end of the poly(G) segment created during cloning (53) and extends for over 3900 nucleotides to the end of the CPS coding region and confirms approximately 800 nucleotides reported at the junction of the glutaminase and synthetase domains (54). This complete coding sequence found on pCAD142 but leaves undetermined the short 5' region of the mRNA coding for the amino terminus. This latter region of the protein has no known catalytic activity but is necessary for the association of the E. coli GLNase and CPSase subunits (55).

The deduced amino acid sequence of the glutaminase (Fig. 3) and carbamyl phosphate synthetase (Fig. 4) was aligned to those of other CPSases. Based upon the size of the E. coli pabA gene product the GLNase domain of CAD starts at glycine 19 and ends at threonine 201. The abrupt appearance of strong homology at lysine 240 marks the amino terminus of the CPSase. Alignment of the amino half of the CAD CPSase (Lys-240 to Pro-778) with the carboxyl half (His-779 to Cys-1300) gave a 26% sequence identity and identified histidine 779 as the start of the carboxyl half. The CPSase extends to cysteine 1300 where the dihydroorotase domain begins (12). The synthetase, as defined, is 8 residues shorter on the amino terminus and 6 residues shorter on the carboxyl terminus than the carB gene product from E. coli. The connecting region between the GLNase and CPSase domains is 39 amino acids long. There are remnants of sequence identity in this linker (Fig. 5) among the hamster CAD, Drosophila CAD, yeast ura2 protein products and, to a lesser extent, the Dictyostelium pyr1-3 gene product.

The domain structure of the CAD CPSase was probed by controlled proteolysis. A 60-kDa trypsin fragment was isolated and its amino terminus sequenced by Edman degradation. Based upon the size of the polypeptide it would extend from isoleucine 600 to lysine 1140 and include the junction...
**Fig. 4.** Amino acid alignment of selected CPSases. Identical residues in the hamster CAD (HAM), rat mitochondrial (CPS1) (86), D. melanogaster CAD (DROS) (37), S. cerevisiae pyrimidine-specific (URA2) (66), and S. cerevisiae arginine-specific (CPA2) (35), and E. coli (CARB) (33), CPSases are boxed. The shaded areas highlight the peaks from the homology plot (Fig. 9). The numbers identify the residue at the beginning of each row. The serine phosphorylated in CAD (622) is marked with an asterisk.

Labeled brackets mark the start of the carboxyl half of the CPSase, the extent of the ATP binding domains based upon homology to the YPCase and CACCase, and the borders of the 60-kDa proteolytic fragment. Parentheses mark positions of possible experimental artifacts that caused the enclosed conserved sequences to be published in alternate reading frames.
between the two homologous halves of the CPSase. The resistance of the 60-kDa fragment to further proteolysis suggests that no appreciable exposed linker connects the amino and carboxyl halves of the CPSase.

Nucleotide sequence alignments were directly derived from the aligned amino acid sequence and both were used to calculate the identities matrices and dendrograms showing the phylogenetic relationship of the proteins (Fig. 6). The CPSases were compared to homologous regions of yeast pyruvate carboxylase (56) and chicken acetyl-CoA carboxylase (57), two enzymes which also phosphorylate bicarbonate to produce carboxy phosphate and can synthesize ATP from carbamyl phosphate and ADP (Fig. 7). These two proteins are more closely related to each other than to the CPSases (36% versus 18% amino acid sequence identity). A dendrogram based upon amino acid identity (not shown) indicates that the ancestral gene common to the yeast pyruvate carboxylase (YPCase) and chicken acetyl-CoA carboxylase (CACCase) diverged before the duplication that produced the homologous amino and carboxyl halves of the CPSases. A homology plot of the GLNase domain, computed as described above, revealed four distinct peaks. (Fig. 7). In addition to the three homologous regions characteristic of trpG-like amidotransferases, there is a fourth cluster of conserved residues (threonine 136 to aspartate 145) which appears only in the carbamyl phosphate synthetases. The 10 regions of strict conservation that map within the synthetase domain (Fig. 8) highlight areas most likely to contain residues important for structure and function.

**DISCUSSION**

Mammalian carbamyl phosphate synthetase is part of a chimeric protein formed by the fusion of genes encoding...
pleated sheet (60). However, these structural components do not require a highly specific primary structure for function. By comparing the primary structure of CAD to its well-characterized homologues, we are able to assign likely functional units. These domains have act as modulating factors to many regions of the protein. The functions of these three homologous segments (Fig. 3) common to the trpG amidotransferases are assigned by analogy to other members of this group. Region I probably serves a structural role (23) while regions II and III form the glutaminase active site. Cysteine 97 in Region II forms a thioester intermediate in the hydrolysis of the 5-glutamyl thioester intermediate (59). ATP binding sites typically contain three segments: an a-helix, and a hydrophobic parallel p-sheet and glycine-rich loop, an a-helix, and a hydrophobic parallel p-sheet. The glutaminase was compared from glycine 19 to valine 202 and the synthetase from lysine 240 to cysteine 1300. The calibration at the left of the dendrograms marks percent differences. The numbers in the dendrograms represent the average number of changes required to make a sequence on one branch identical to a sequence on another. The top row provides the number of identities (above right) over the length of the region compared (counting dashes). These identities are given in the legends to Figs. 3 and 4.

The likely ATP binding sites of the CPSases have been proposed based upon comparisons to other well-characterized ATP binding proteins (36, 41). These sites fall within the region homologous to YPCase (56) and CACCase (57) as well as the domains photolabeled by 8-azido-ATP (41). These sites fall within the region homologous to YPCase (56) and CACCase (57) as well as the domains photolabeled by 8-azido-ATP (41). These sites fall within the region homologous to YPCase (56) and CACCase (57) as well as the domains photolabeled by 8-azido-ATP (41).

Because the region between the CPSase and ATCase domains, which includes the DHO domain in CAD, is homologous in hamster, Drosophila, Dictyostelium, and Saccharomyces (12, 65, 66), the 240-kDa multifunctional protein initiating pyrimidine biosynthesis was probably formed prior to the divergence of these organisms. The dendrograms based upon the amino acid comparisons show the CPSI and cpa2 gene products diverging before the formation of the chimeric proteins. The dendrogram derived from the nucleotide alignment is quite different. It shows that upo2 is significantly closer to cpa2 than to the other chimeric proteins. Similarly the nucleotide sequence of CPSI is more closely related to hamster CAD than to cpa2. Part of the reason for the differences is that codon preferences may tend to cluster nucleotide sequences from closely related organisms, while a change in ligand specificity, or other function, can increase the rate of ...
Amino acid alignment of the CPSases with pyruvate carboxylase from yeast and acetyl-CoA carboxylase from chicken. Residues identical with either of the carboxylases are boxed. The segments from the amino half of the CPSases are above while those from the carboxyl half are below the carboxylases (shaded). The solid bar marks the region shown (36) to be homologous to porcine adenylate kinase (72) and the β-subunit of F1-ATPase (73). The dashed bar marks the region homologous to equine phosphoglycerate kinase (74) and bovine glutamate dehydrogenase (75). The first line of asterisks marks a proposed glycine-rich loop and the second a β-pleated sheet based upon a consensus of ATP binding sites (41). The abbreviations are: hamster CPS (CAD), rat mitochondrial (CPA), S. cerevisiae pyrimidine-specific CPS (URAP), and arginine-specific CPS (CPAZ), E. coli CPS (CARB), yeast pyruvate carboxylase (YPC), and chicken acetyl-CoA carboxylase (CACC).

FIG. 8. Homology plot of the glutaminase and synthetase regions of the CPSases. The peaks (I, II, and III) correspond to the three classical regions found in type-I amidotransferases. The peak marked IV is unique to the carbamyl phosphate synthetases. The 10 peaks in the synthetase domain where at least five consecutive amino acids are conserved are numbered starting at the amino end. The domain structure of the glutaminase (GLN) and the two homologous halves of the synthetase are shown in the bar above the graph. The numbers on the abscissa correspond to the hamster CAD residue number while the ordinate is number of identities.

divergence of proteins. The loss of glutaminase activity and the acquisition of N-acetylglutamate regulation in CPS1, and the lack of allosteric regulation of cpaz should have allowed mutations that would make these proteins appear more distant in the amino acid comparison. Although it has been suggested that CPS1 was formed by the fusion of the glutaminase with the synthetase (35, 67), our analysis raises the possibility that CPS1, cpaz and cpaz arose by defusion from the pyrimidine chimera. It is noteworthy that the carboxyl terminus of the cpaz and amino terminus of the cpaz gene products are longer than expected based upon the length of the E. coli carA and carB gene products. If the homology between these extensions is the result of a common origin they too may have split from a fused protein (Fig. 5). These comparisons therefore suggest a complex evolutionary history for this family of proteins involving multiple gene duplications, fusions, defusions, and extinctions.

Acknowledgments—We thank Drs. George Stark (Imperial Cancer Research Fund, London) for the cell line 165-23 and the plasmid pADJ42 and Grace Hannon, Hong Wan, and Kathleen Benoit for expert assistance.

REFERENCES
1. Jones, M. E. (1980) Annu. Rev. Biochem. 49, 253–279
2. Shoaf, W. T., and Jones, M. E. (1973) Biochemistry 12, 4039–4051
3. Mori, M., Ishida, H., and Tatsibana, M. (1975) Biochemistry 14, 2622–2630
4. Coleman, P. F., Suttle, D. P., and Stark, G. R. (1977) J. Biol. Chem. 252, 6379–6385
5. Mally, M. I., Grayson, D. R., and Evans, D. R. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 6647–6651
Mammalian carbamyl phosphate synthetase (CPS). DNA sequence and evolution of the CPS domain of the Syrian hamster multifunctional protein CAD.

J P Simmer, R E Kelly, A G Rinker, Jr, J L Scully and D R Evans

*J. Biol. Chem.* 1990, 265:10395-10402.

Access the most updated version of this article at [http://www.jbc.org/content/265/18/10395](http://www.jbc.org/content/265/18/10395)

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 0 references, 0 of which can be accessed free at [http://www.jbc.org/content/265/18/10395.full.html#ref-list-1](http://www.jbc.org/content/265/18/10395.full.html#ref-list-1)